
BIOCATALYST
DESIGN
FOR
STABILITY AND
SPECIFICITY
Edited by
Michael E. Himmel and George Georgiou
ACS Symposium Series 568

Biocatalyst Design for Stability and Specificity

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

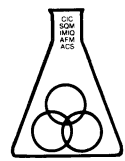
ACS SYMPOSIUM SERIES 516

Biocatalyst Design for Stability and Specificity

Michael E. Himmel, EDITOR
National Renewable Energy Laboratory

George Georgiou, EDITOR
The University of Texas

Developed from a symposium sponsored
by the Division of Biochemical Technology
of the American Chemical Society
at the Fourth Chemical Congress of North America
(202nd National Meeting of the American Chemical Society),
New York, New York
August 25–30, 1991



American Chemical Society, Washington, DC 1993

Chemical Congress of North America 1991 : Biocatalyst design for stability and specificity

Library of Congress Cataloging-in-Publication Data

Chemical Congress of North America (4th: 1991: New York, N.Y.)
Biocatalyst design for stability and specificity / [edited by] Michael E. Himmel, George Georgiou.

p. cm.—(ACS Symposium Series, 0097–6156; 516).

"Developed from a symposium sponsored by the Division of Biochemical Technology of the American Chemical Society at the Fourth Chemical Congress of North America (202nd National Meeting of the American Chemical Society) New York, New York, August 25–30, 1991."

Includes bibliographical references and index.

ISBN 0–8412–2518–4

1. Enzymes—Biotechnology—Congresses. 2. Protein engineering—Congresses. 3. Protein folding—Congresses. I. Himmel, Michael E., 1952– . II. Georgiou, George, 1959– . III. American Chemical Society. Division of Biochemical Technology. IV. American Chemical Society. Meeting (202nd: 1991: New York, N. Y.). V. Title VI. Series.

TP248.65.E59C5 1993
660'.634—dc20 92–38473
CIP

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48–1984. ♾

1993 Advisory Board

ACS Symposium Series

M. Joan Comstock, *Series Editor*

Foreword

THE ACS SYMPOSIUM SERIES was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of this series is to publish comprehensive books developed from symposia, which are usually "snapshots in time" of the current research being done on a topic, plus some review material on the topic. For this reason, it is necessary that the papers be published as quickly as possible.

Before a symposium-based book is put under contract, the proposed table of contents is reviewed for appropriateness to the topic and for comprehensiveness of the collection. Some papers are excluded at this point, and others are added to round out the scope of the volume. In addition, a draft of each paper is peer-reviewed prior to final acceptance or rejection. This anonymous review process is supervised by the organizer(s) of the symposium, who become the editor(s) of the book. The authors then revise their papers according to the recommendations of both the reviewers and the editors, prepare camera-ready copy, and submit the final papers to the editors, who check that all necessary revisions have been made.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previously published papers are not accepted.

M. Joan Comstock
Series Editor

Preface

ENZYME-BASED PROCESSES ARE CURRENTLY RECOGNIZED as crucial to many industrial interests, including pharmaceuticals, specialty chemicals, and the production of commodity chemicals and fuels. Enzymes are often the catalysts of choice when issues of specificity, mild operating conditions, and adaptability are considered. However, they are less robust under many industrial process conditions than are their chemical equivalents, especially where heat cycling, pumping shear, protease exposure, and general aging conspire to denature proteins. Also, enzymes in nature often have limited activity as a result of adaptation pressure imposed by requirements for survival. The successful use of enzymes as tools, therefore, may require an amalgam of properties not important in nature, such as the combination of catalytic and polymer binding domains to suit individual requirements.

Biocatalyst Design for Stability and Specificity features 25 chapters written by many of the leading international authorities in the fields of protein biochemistry and molecular biology. These chapters not only highlight the diversity of approaches used in understanding the function of proteins; they illustrate how this understanding is applied to improving protein design for biotechnology applications. We believe the authors collectively bring a unique perspective to enzyme biochemistry and engineering in their presentations on studies of protein structure, folding (refolding), and stability, and in their examinations of the methods (both chemical and recombinant) currently used to improve the usefulness of these enzymes.

The book is grouped into 5 major areas of interest and concern. The first section, "Protein Stability", describes ongoing research into the interactions that underly protein stability and discusses the effects of mutations on the thermodynamic properties of proteins, the relative contributions of the hydrophobic effect and hydrogen bonding on protein stability, the assessment of protein structure by circular dichroism, studies of *Trichoderma reesei* cellobiohydrolase I stability, the stabilization of proteins via metal binding, and the factors that affect the function of proteins in organic solvents. The effect of protein structure on the stability of proteins from thermophilic and psychrophilic microorganisms is discussed in Chapters 4 and 5.

"Protein Folding" emphasizes the renaturation of functional proteins from the aggregated state. Topics covered include: mutational analysis of protein aggregation in vivo, characterization of inclusion bodies, function

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

of the GroE chaperonins, and finally the effects of detergents and cosolvents on the efficiency of protein folding.

An area of considerable interest in enzyme engineering is the relationship between the structure and function of multifunctional proteins and the design of artificial multifunctional enzymes by genetic engineering. "Multifunctional Proteins" is comprised of 4 excellent chapters on the function of natural multifunctional proteins, the genetic engineering of bifunctional enzymes, and the use of fusion to a cellulose binding domain for protein purification and immobilization.

"Design of Cellulases by Recombinant Methods" contains 4 chapters on recent advances in improving the design of cellulases through recombinant technology. Chapter 18 describes important properties of *T. reesei* cellobiohydrolase II, following carefully designed site directed mutagenesis studies, and Chapter 19 reports a successful recombinant project with another *T. reesei* enzyme, β-D-glucosidase. Chapter 20 reviews progress in the field of *Thermomonospora fusca* and bacterial cellulase cloning, and Chapter 21 describes a new model for the mechanism of cellulose hydrolysis exhibited by the *Clostridium thermocellum* cellulosome.

The last section, "Improving Natural Enzymes by Chemical Cross-Linking" is dedicated to the use of chemical modification to improve enzymes. These 4 chapters include an overview of the subject, a review of glutaraldehyde modification chemistry, and examples of successful applications of the technology.

Acknowledgments

We are grateful to the Division of Biochemical Technology of the ACS for supplying the forum for this work and for their support, along with that of Genencor International, Inc., and Merck and Company in partially funding the symposium or its functions. The program coordinators for the Division of Biochemical Technology to the ACS meeting, Jonathan Woodward and Douglas C. Cameron, deserve special commendation. We also wish to thank the organizers and moderators of the four sessions around which the symposium and book were constructed: Anne Lee, Session I; John O. Baker, Session II; Michael E. Himmel, Session III; and Sharon P. Shoemaker, Session IV. Their contribution was essential to the success of the symposium and the book. The editors wish to bring special attention to participation in the symposium by Rainier Jaenicke and Minishwar Gupta; their expense and sacrifice incurred in order to participate was greater than our own, and we respect this dedication to the field. Our Acquisitions Editor at ACS Books was Barbara C. Tansill. Her efficient processing of the manuscripts and her guidance were key to the timely publication of the book.

Last, and probably most important, we appreciate the timely and thoughtful reports of research progress or reviews of research contributed by the authors. We wish them continued good fortune in their research and in their lives.

MICHAEL E. HIMMEL
Applied Biological Sciences Branch
National Renewable Energy Laboratory
Golden, CO 80401

GEORGE GEORGIOU
Department of Chemical Engineering
The University of Texas
Austin, TX 78712

August 20, 1992

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

Chapter 1

Effects of Mutations on Thermodynamic Properties of Proteins

Julian M. Sturtevant

Departments of Chemistry, Molecular Biophysics, and Biochemistry, Yale University, New Haven, CT 06511

Differential scanning (DSC) and isothermal titration (ITC) microcalorimetry are employed in studying the effects of amino acid replacements on the thermodynamic properties of proteins. The proteins studied by DSC include the lysozyme of phage T4 (17 mutations), staphylococcal nuclease (9 mutations) and the lambda repressor of *E. coli* (15 mutations). ITC is used in the study of the effects of replacements in S-peptide on the binding of S-peptide to S-protein to form ribonuclease-S. Although partial explanations, based on crystallographic structures, of some of the results are presented, it seems evident that we are very far from being able to rationalize the results in any really satisfactory manner. One outstanding feature of the results is the lack of correlation between the free energy changes and the enthalpy changes produced by the replacements. The enthalpy changes are generally much larger than the free energy changes, and frequently of opposite sign.

A central problem in biophysics is to achieve a quantitative understanding of the intra- and intermolecular forces which cause biopolymers to fold into their specific native structures, and to retain these structures through various more or less severe perturbations. It seems evident that a useful way of attacking at least the second part of this problem with respect to proteins would be to make small known changes, such as single amino acid replacements, in a protein and to measure the energetic consequences thereof (*1*). Our present state of ignorance concerning the quantitative aspects of intermolecular forces does not permit fully satisfactory interpretations of the thermodynamic results of such experiments (*2*). Even a small protein of only 100 amino acid residues will probably have hundreds of energetically significant intramolecular interactions. Nevertheless, I am convinced that it is worthwhile to obtain such experimental results, if for no other

0097–6156/93/0516–0002$06.00/0

reason than to supply data with which theoretical calculations can be compared.

One of the most interesting features of many proteins is that they have perfectly cooperative structures. As nearly as we can tell, when such a structure is thermally unfolded, only two states, the initial folded state and the final unfolded state, are significantly populated. This is illustrated in Figure 1 where differential scanning calorimetric (DSC) curves of the apparent excess specific heat as a function of temperature are shown for wild type (WT) staphylococcal nuclease (SNase) at two different values of the pH. In each case the experimental data (solid curves) are very well reproduced by data (dashed curves) calculated for the transition on the basis that it obeys the van't Hoff equation

$$\frac{d \ln K}{dT} = \frac{\Delta H_{vH}}{RT^2} \quad (1)$$

where K is the equilibrium constant for the reaction folded ⇌ unfolded, T is the absolute temperature and ΔH_{vH} is the van't Hoff enthalpy, which in the case of a truly two state process will equal the calorimetric, or true, enthalpy, ΔH_{cal}. For SNase, the all-or-none model was extended to include dimerization of the unfolded protein. It is evident that at each pH there is a substantial permanent change in specific heat, which means that ΔH_{cal} is a strong function of the temperature, and it is assumed that ΔH_{vH} shows the same variation with temperature. Of course, this variation must be included in the curve fitting calculation. In the present case, at pH 3.84, ΔH_{cal} is near zero at 20°C and equal to 60 kcal mol^{-1} at 47°C.

Two important methods for obtaining biothermodynamic data for which excellent instruments are commercially available involve differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC). We employ for DSC studies the MC-2 microcalorimeter (Microcal, Inc., Northampon, MA) and the DASM-4 microcalorimeter (Biopribor, Puschchino, Moscow Region, Russia), and for ITC studies the Omega microcalorimeter (also Microcal, Inc.).

Differential Scanning Calorimetry

I will first review some of our recent DSC methods and results, some published and some unpublished. We generally evaluate the thermodynamic parameters for the unfolding transition by the curve fitting procedure indicated in Figure 1, generalized to permit ΔH_{vH} to be different from ΔH_{cal}, and to accommodate transitions made up of several two-state steps (*3*).

A convenient summary of the evaluated thermodynamic parameters for a mutant protein is in terms of $\Delta\Delta G^{o}_{d}$, $\Delta\Delta S^{o}_{d}$ and $\Delta\Delta H_{d}$ where ΔG^{o}_{d} is a standard free energy of unfolding, ΔS^{o}_{d} is the standard unfolding entropy, and ΔH_{d} is the enthalpy of unfolding, and the significance of $\Delta\Delta$ is as shown in equation 2,

$$\Delta\Delta J_{d} = \Delta J_{d}(\text{mutant}) - \Delta J(\text{WT}) \quad (2)$$

with the thermodynamic parameters evaluated at the temperature of

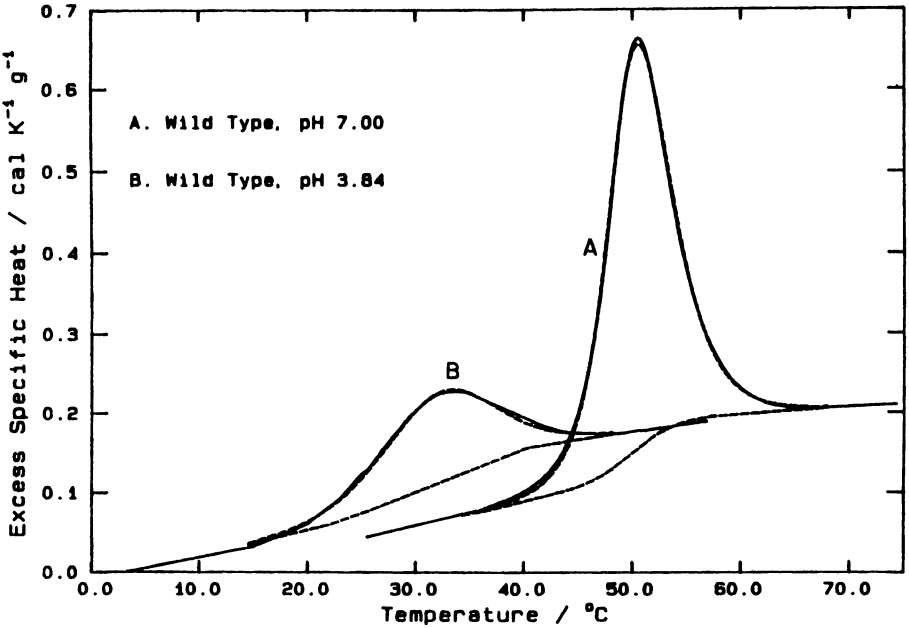


Figure 1. Differential scanning calorimetric scans showing the variation with temperature of the apparent excess specific heat of staphylococcal nuclease at pH 7.00 (curve A) and pH 3.84 (curve B). In each case the solid curve represents the observed data after subtraction of the instrumental baseline, and the dashed curves show the calculated chemical baseline and data calculated for an all-or-none model modified to include dimerization of the unfolded protein.

half-denaturation, $t_{1/2}$, of the WT protein by means of the equations

$$\Delta H_d(\text{at } T_1) = \Delta H_d(\text{at } T_2) - \Delta Cp(T_2 - T_1) \quad (3)$$

$$\Delta G^o_d(\text{at } T_1) = \Delta H_d(\text{at } T_2)\frac{T_2 - T_1}{T_2} - \Delta Cp\left[T_2 - T_1 + T_1 \ln\frac{T_1}{T_2}\right] \quad (4)$$

$$\Delta S^o_d = (\Delta H_d - \Delta G^o{}_d)/T \quad (5)$$

Here T_1 and T_2 are equal to $t_{1/2}$ + 273.15 for the WT and mutant proteins respectively and ΔCp is the change in molar heat capacity accompanying the transition of the mutant protein. As mentioned above, since ΔCp is frequently as large as 2 to 3 kcal $K^{-1}mol^{-1}$, it is important that it be included in calculating $\Delta\Delta J_d$. According to

the definition in equation 2, a negative value for $\Delta\Delta G^{o}_{d}$ indicates apparent destabilization (destabliization of the folded form or stabilization of the unfolded form or both) by the mutation. (To conform with the current usage in the literature, the definition of $\Delta\Delta J_{d}$ given in equation (2) is reversed from the form used in papers from this laboratory published prior to March, 1992, so that care must be used in reading the earlier papers to avoid confusion.)

DSC of T4 Lysozyme. As mentioned earlier, we have studied a total of 17 mutant forms of the lysozyme of phage T4, kindly supplied to us by Brian Matthews and Joan Wozniak of the University of Oregon. The results for the WT protein and 8 of the mutants have been published (*4,5,6*). In all cases, the thermal unfolding, which is reversible, occurs at a temperature which is strongly dependent on pH, and with a large change in heat capacity. These features are qualitatively similar to those illustrated in Figure 1 for SNase. For each form of T4 lysozyme the values for $t_{1/2}$ can be expressed as

$$t_{1/2} = A + B\ pH \tag{6}$$

and those for ΔH_{cal} as

$$\Delta H_{cal} = \Delta H_{o} + \Delta Cp\ t \tag{7}$$

the four constants being evaluated by linear least squaring. These constants are employed in the calculation of $\Delta\Delta J_{d}$ by equations 3, 4 and 5.

The results calculated for WT and the mutant T4 lysozymes at pH 2.5, where $t_{1/2}$ for the WT protein is 46.2°C, are summarized in Table I. Similar results were calculated at pH 2.0 and 3.0. The largest change in stability is an apparent destabilization of 4 kcal mol^{-1} caused by the R96H mutation, which also causes a substantial change in the entropy of denaturation. An important feature of the data in Table I is the lack of correlation with respect to either sign or size between the free energy changes and the changes in the other thermodynamic quantities. This is illustrated in Figure 2 for the seven replacements of Thr 157 which we have studied. The largest compensating entropy changes, those for I3F and I3E, amount to 49 cal $K^{-1}mol^{-1}$. It would seem that any fully satisfactory theoretical interpretation of the effects of these mutations should include an elucidation of the origins of such significant entropy and enthalpy effects.

In a recent paper Tidor and Karplus (*7*) undertook a simulation analysis of the effects of the R96H replacement. The complexity of such an analysis is illustrated by the fact that 44 contributions to the free energy change of converting Arg to His in the folded protein and 30 to the conversion in the unfolded form were included, giving a result in fair agreement with our experimental value. No indication of the enthalpy and entropy make-up of these free energy values was given.

Table I. Changes in Thermodynamic Parameters at pH 2.5 Produced by Various Mutations of T4 Lysozyme

($t_{1/2}$ = 46.2°C, ΔH_{cal} = 108 kcal mol^{-1}, for WT Protein at pH 2.5)

Protein	$\Delta\Delta G_d^o$	$\Delta\Delta H_d$	$\Delta\Delta S_d^o$	$\Delta\Delta Cp$
A82P	0.5	5	15	-110
A93P	0.5	6	18	-530
R96H	-4.0	6	32	90
G113A	0.5	5	15	310
I3F	-1.5	14	49	-470
I3E	-1.8	13	49	-60
I3L	0.85	9	25	-290
I3P	-3.0	3	19	390
I3T	-2.5	11	42	220
T157A	-0.9	3	13	170
T157E	-1.3	-12	-33	-300
T157I	-1.9	1	9	-150
T157L	-1.7	-9	-23	-260
T157N	-1.1	6	21	280
T157R	-0.6	2	7	-230
T157V	-1.6	-1	2	60
C54T:C79A	-0.8	-2	-4	-630

$\Delta\Delta G_d^o$, $\Delta\Delta H_d$, kcal mol^{-1}; $\Delta\Delta S_d^o$, $\Delta\Delta Cp$, cal $K^{-1}mol^{-1}$.
Estimated uncertainties: $\Delta\Delta G_d^o$, ± 0.4 kcal mol^{-1}; $\Delta\Delta H_d$, ± 4 kcal mol^{-1} (average value); $\Delta\Delta S_d^o$, ± 10 cal $K^{-1}mol^{-1}$; $\Delta\Delta Cp$, ± 200 cal $K^{-1}mol^{-1}$.

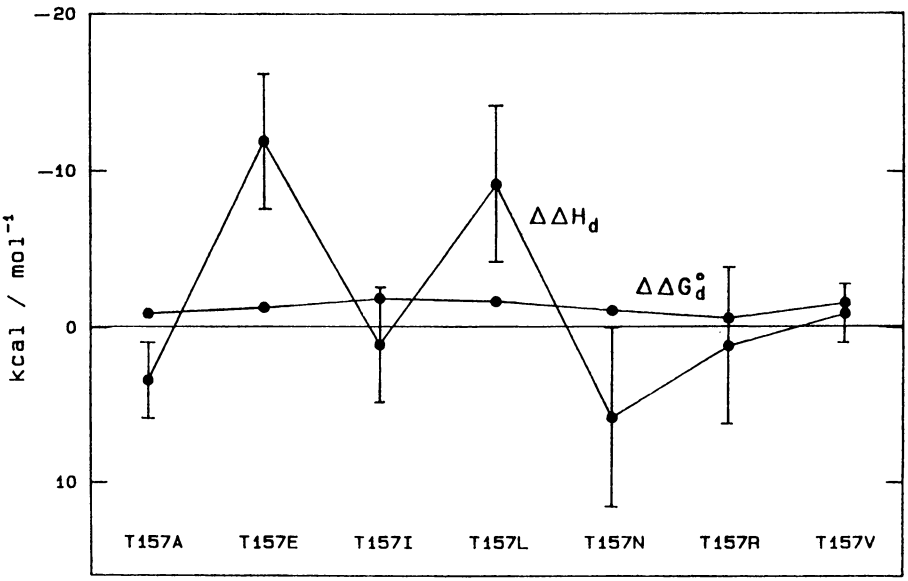


Figure 2. Values of $\Delta\Delta G_d^o$ and $\Delta\Delta H_d$ as determined by DSC at pH 2.5 for seven different replacements of T157 in T4 lysozyme, illustrating the lack of correlation frequently observed between changes in the standard free energy and enthalpy of unfolding resulting from amino acid replacements. (Adapted from ref. 5. Copyright 1991 American Chemical Society.)

Weaver and Matthews (*8*) determined by x-ray crystallography the structure of R96H to a resolution of 1.7Å. The authors concluded on the basis of comparisons with the structure of the WT protein, that the two most important sources of the apparent destabilization of the mutant were the loss of a helix-dipole interaction involving the C-terminus of helix 82-90, and significant strain caused by the introduction of the imidazole ring of histidine. Tidor and Karplus, on the other hand, state that the helix-dipole model is inappropriate for this case.

In the Thr 157 series, it is interesting that T157I and T157L have nearly equal values of $\Delta\Delta G_d^o$ but values of $\Delta\Delta H_d$ and $\Delta\Delta S_d^o$ differing by 10 kcal mol^{-1} and 32 cal $K^{-1}mol^{-1}$ respectively. The x-ray data of Alber et al. (*9*) show that the introduction of Ile at position 157 forces the side chain of Asp 159 to move 1.1Å from its position in the WT protein, a much larger motion than caused by any of the other T157 substitutions studied. It is accordingly difficult to understand why the substitution of Leu at position 157 causes a much larger change in enthalpy than the introduction of either Ile or Val.

Because of the large values of $\Delta\Delta Cp$ in some cases, the thermodynamic parameters will vary significantly with temperature and also with pH. In the case of T157R, for example, $\Delta\Delta G_d^o$ changes from an apparent destabilization at pH 2.0 and 38.8 °C ($\Delta\Delta G_d^o$ = -1.1 kcal mol^{-1}) to a small apparent stabilization at pH 3.0 and 53.6°C ($\Delta\Delta G_d^o$ = +0.3 kcal mol^{-1}). Alber et al. (*9*) observed on the basis of CD melting curves that the apparent destabilization of the T157R mutant was less at pH 6.5 than at pH 2.0, probably because of the ionization of Asp 159 and the resulting formation of the Arg-Asp hydrogen-bonded ion pair indicated by the crystallographic data. The DSC result suggests that significant ionization of Asp 159 has already taken place by pH 3.0.

Dang et al. (*10*) have applied molecular dynamics/free energy perturbation calculations to the T157V mutation to obtain the value $\Delta\Delta G_d^o$ = -1.9±1.1 kcal mol^{-1}, in reasonable agreement with the DSC result, $\Delta\Delta G_d^o$ = -1.6±0.4 at pH 2.0. It happens that in this case the enthalpy and free energy changes are of similar magnitude.

DSC of Staphylococcal Nuclease. We have studied by means of DSC (Engelman et al., manuscript in preparation) several mutants of SNase, namely L25A, V66L, G79S, G88V, A90S and H124L, the multiple mutants V66L:G88V and V66L:G79S:G88V (*11*) and an unusual mutant, ΔNase, (*12*) formed by removing a six-residue segment from the Ω-loop of the WT protein. It was found with all these proteins that $t_{1/2}$ decreases significantly as the concentration is increased, indicating that the unfolded form of the protein is more aggregated than the folded form. In fact, good curve fits were obtained in most cases on the assumption that the folded form is monomeric while the unfolded form is nearly completely dimerized. We have arbitrarily selected 500 μM as a standard concentration at which quantities such as $\Delta\Delta G_d^o$ are calculated and compared, interpolation being based on least squared van't Hoff plots of ln concentration *vs* $(t_{1/2} + 273.15)^{-1}$ for $t_{1/2}$ and on least squared linear plots of ΔH_{cal} *vs* $t_{1/2}$ for ΔH_{cal}.

The changes in thermodynamic quantities resulting from the various mutations are listed in Table II. The first point which may be made here is that the data for the double and triple mutations indicate approximate additivity of $\Delta\Delta G^o_d$ and $\Delta\Delta H_d$. For the double mutant the calculated and observed values for $\Delta\Delta G^o_d$ are respectively +1.3 and +1.0 for the $\Delta\Delta H_d$ -31 and -24 (all in kcal mol^{-1}); similarly for the triple mutant the figures are 0 and +0.4 and -45 and -52. Approximate additivity in the effects of mutations has been observed in other cases as well, as will be discussed below.

Table II. Changes in Thermodynamic Parameters at pH 5.0 and 7.0 Produced by Various Mutations of Staphylococcal Nuclease ($t_{1/2}$ = 47.0°C, ΔH_{cal} = 64 kcal mol^{-1}, at pH 5.0, and 51.4°C, 73 kcal mol^{-1} at pH 7.0 for the WT Protein)

	pH 7.0			pH 5.0			
Protein	$\Delta\Delta G^o_d$	$\Delta\Delta H_d$	$\Delta\Delta S^o_d$	$\Delta\Delta G^o_d$	$\Delta\Delta H_d$	$\Delta\Delta S^o_d$	$\Delta\Delta Cp$
L25A	-2.6	11	-40	-2.8	10	40	300
V66L	0.8	-14	-45	0.6	-13	-45	-130
G79S	-1.3	-14	-40	-1.2	-13	-40	-250
G88V	0.5	-17	-55	0.6	-18	-60	270
A90S	-2.5	13	50	-2.6	12	45	220
H124L	1.2	2	3	1.9	3	4	-210
V66L:G88V	1.0	-24	75	1.4	-28	-90	(-1110)*
V66L:G79S:G88V	0.4	-52	-160	0.7	-41	-130	(10)*
ΔNase	1.8	7	16	2.1	4	6	-280

$\Delta\Delta G^o_d$, $\Delta\Delta H_d$, kcal mol^{-1}; $\Delta\Delta S^o_d$, $\Delta\Delta Cp$, cal $K^{-1}mol^{-1}$. Estimated uncertainties: $\Delta\Delta G^o_d$, ±0.4 kcal mol^{-1}; $\Delta\Delta H_d$, ±4 kcal mol^{-1} (average value); $\Delta\Delta S^o_d$ ± 10 cal $K^{-1}mol^{-1}$; $\Delta\Delta Cp$, ± 200 cal $K^{-1}mol^{-1}$.
*Based on the values observed in individual experiments at pH 5 and pH 7.

As is the case with the mutants of T4 lysozyme, there is no discernible correlation between the values for $\Delta\Delta G^o_d$ and those for $\Delta\Delta H_d$. An extreme example of this is shown by the triple mutant, for which the entropy of unfolding at pH 7.0 is reduced from 228 to 68 cal $K^{-1}mol^{-1}$. What might appear to be a very minor disturbance of the energetics of the molecule, as judged by the value of $\Delta\Delta G^o_d$ = 0.4 kcal mol^{-1}, obviously involves profound changes in enthalpy and entropy which are probably distributed throughout the molecule. In fact, it seems safe to assume that this huge entropy effect is mainly an increase in the entropy of the folded state due to a severe loosening of its three dimensional structure, certainly an effect which could not be predicted to result from the replacement of three amino acid residues in a protein which contains 149 residues. These considerations lead us again to emphasize the importance of including such significant thermodynamic changes in any theoretical analysis of the effects of mutations on protein energetics.

David Shortle (*11*), who originally prepared most of the mutants used in our study and supplied us with plasmids, investigated the thermal stability of some of these mutants and certain other ones by means of fluorescence melting curves. There are significant differences between the results obtained by Shortle et al. and our results, due at least in part to the facts that the fluorescence method can determine only van't Hoff enthalpies, and that DSC experiments require much higher protein concentrations.

Shortle et al. argue, largely on the basis of the apparent enthalpy-entropy compensation exhibited by these mutants, that the mutations cause alterations in the hydration of the denatured state, and that these alterations in hydration are the primary cause of the mutational change in stability. If this argument is valid, then it would appear to follow that for other proteins the primary effects of mutations are on the unfolded rather than the folded state, since it appears to be a general rule that largely compensating enthalpy and entropy effects are characteristic of protein thermal unfolding. The apparently random distribution of $\Delta\Delta H_d$ between positive and negative values makes this seem quite unlikely, as do also the many instances of significant mutational changes observed by x-ray crystallography in the structure of the folded form. It seems evident, as discussed above, that the massive enthalpy and entropy effects shown by V66L:G79S:G88V must be almost entirely due to changes in the folded protein.

DSC of λ Repressor. In collaboration with Robert Sauer and his colleagues at M.I.T. we have looked at 14 single replacements and 1 double replacement in the N-terminal domain of the λ repressor of *E. coli* (*13*, *14*, *15*). The results observed with 6 of these mutants are shown in Figure 3 and the complete results are summarized in Table III. It is evident in the figure that the unfolded N-terminal domain has no significant effect on the unfolding of the C-terminal domain, as would be expected on the basis of Brandts' model for domain interactions (*16*). The one apparent exception, A66T, was found after the DSC result became available to have a previously undetected mutation, Y21OH, in the C-terminal domain. The mutations causing the largest stabilizations are at buried sites. All the N-terminal peaks, except that for A66T, and all the C-terminal peaks are hypersharp and unsymmetrical for unknown reasons; only the A66T peak satisfies the two-state criterion of $\Delta H_{vH} = \Delta H_{cal}$.

Table III lists the data for 8 mutants causing apparent stabilization and 7 destabilizing mutants. Two of the latter cause unusually large drops in $t_{1/2}$. It is interesting that replacing His by Phe in the second of these causes a weak stabilization. Although any correlation between $\Delta t_{1/2}$ and ΔH_{cal} is very weak, it may be noted that all the mutants which lower $t_{1/2}$ also lower ΔH_{cal} by as much as 37 kcal mol^{-1}. This suggests that part of the mutational effect on ΔH_{cal} may in fact be a ΔCp effect. It is not possible to estimate values of ΔCp for individual mutants from complex traces such as those in Figure 3. Linear least squaring of the values of ΔH_{cal} *vs* $t_{1/2}$ gives $\Delta Cp = 1.13$ kcal $K^{-1}mol^{-1}$, with a standard deviation in ΔH_{cal} of ±7.2 kcal mol^{-1} and a coefficient of determination of only 0.7. Values of $\Delta\Delta G^o_d$, $\Delta\Delta H_d$ and $\Delta\Delta S^o_d$ are given

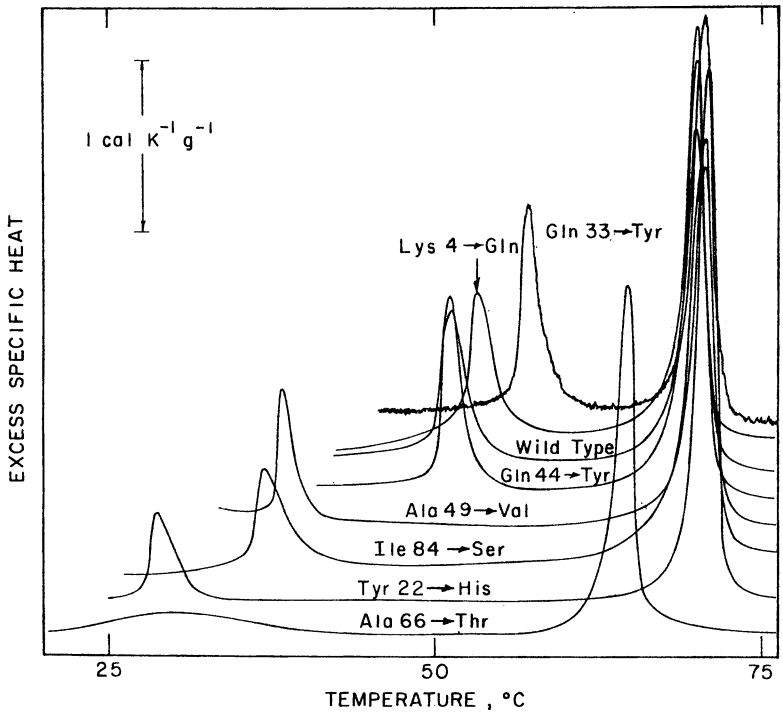


Figure 3. The DSC traces for wild type and seven mutant forms of the λ repressor of *E. coli*. This protein at pH 8.0 shows two clearly separated peaks, the low temperature one resulting from the unfolding of the N-terminal domain and the high temperature one from the unfolding of the C-terminal domain. All the amino acid replacements were located in the N-terminal domain of the protein except for an accidental one, Y210H, in the C-terminal domain. It is shown in Table III that the replacements caused a wide range of changes in the denaturational thermodynamics. (Reproduced from ref. 13. Copyright 1984 J. M. Sturtevant.)

Table III. Changes in Thermodynamic Parameters at pH 8.0 Produced by Various Mutations in the N-terminal Domain of λ Repressor ($t_{1/2}$ = 53.4°C, ΔH_{cal} = 66 kcal mol^{-1} for WT Protein at pH 8.0)

Protein	$\Delta t_{1/2}$, °C	ΔH_{cal}, kcal mol^{-1}	$\Delta\Delta G_d^o$ kcal mol^{-1} A	$\Delta\Delta G_d^o$ kcal mol^{-1} B	$\Delta\Delta H_d$, kcal mol^{-1} A	$\Delta\Delta H_d$, kcal mol^{-1} B	$\Delta\Delta S_d^o$ cal $K^{-1}mol^{-1}$ A	$\Delta\Delta S_d^o$ cal $K^{-1}mol^{-1}$ B
Stabilizing Mutations								
Y22F	1.9	56	0.3	0.3	-10	-12	-32	-38
K4Q	2.0	70	0.4	0.4	4	2	11	4
G46A	3.1	70	0.7	0.6	4	1	10	0
G48S	4.0	55	0.7	0.6	-11	-16	-36	-50
G48N	4.1	62	0.8	0.7	-4	-9	-15	-29
G48A	4.7	62	0.9	0.8	-4	-9	-15	-31
Q33Y	5.9	74	1.3	1.3	8	1	20	0
G46A:G48A	6.2	59	1.1	1.0	-7	-14	-25	-46
Destabilizing Mutations								
Q44Y	-0.1	60	-0	-0	-6	-6	-18	-18
E34K	-1.9	51	-0.3	-0.3	-15	-13	-45	-38
E83K	-3.6	59	-0.7	-0.7	-7	-3	-19	-7
A49V	-13.0	29	-1.2	-1.5	-37	-22	-110	-64
I84S	-14.3	49	-2.2	-2.6	-17	-1	-43	6
A66T	-22.5	40	-2.8	-3.9	-26	-1	-71	10
Y22H	-22.7	30	-2.2	-3.2	-36	-10	-103	-22

A: Calculated assuming $\Delta Cp = 0$; B: calculated assuming ΔCp = 1.13 k cal $K^{-1}mol^{-1}$.
Estimated uncertainties: $\Delta\Delta G_d^o$, ± 0.4 kcal mol^{-1}, $\Delta\Delta H_d$, $\Delta\Delta S_d^o$, see discussion in text.

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

in Table III based on $\Delta Cp = 0$ and $\Delta Cp = 1.13$ kcal $K^{-1}mol^{-1}$ as indicated. Introducing ΔCp has a negligible effect on $\Delta\Delta G_d^o$ except in the last two mutants in the table, but rather large effects on $\Delta\Delta H_d$ and $\Delta\Delta S_d^o$. In view of the large uncertainty in ΔCp, the values obtained taking $\Delta Cp = 0$ are probably to be preferred.

The largest apparent stabilization observed in this series of experiments is that of Q33Y. It was suggested to us by Gregory Petsko that this stabilization is due at least in part to the fact that the aromatic ring of Tyr 33 is oriented perpendicularly to that of Tyr 22, in the minimal energy orientation seen in crystalline benzene.

Replacement of Gly 46 by Ala in the third α-helix in the N-terminal domain led to an apparent stabilization of 0.7 kcal mol^{-1}, and of Gly 48 by Ala, Asn or Ser to stabilization of 0.9, 0.8 or 0.7 kcal mol^{-1}, respectively (*15*). The double mutant, with both of these Gly's replaced by Ala, showed stabilization of 1.1 kcal mol^{-1}, with rough additivity. These results are consistent with the usual view (*17*) that the Gly residue is a poor helix former while the Ala residue is one of the best, and Asn and Ser residues are intermediate.

Prediction of Mutational Effects; Protein Engineering.

In the above discussion we have emphasized the difficulty in accounting for the thermodynamic effects of mutations in terms of molecular energetics. However, there are many known cases where, at least in qualitative terms, changes in the thermal stability of a protein can be understood and where predicted changes in stability have been realized (*18*). Several examples of apparently successful explanations or predictions will be given here.

Stearman et al. (*19*), working with the separated dimeric N-terminal domain of λ-repressor, introduced the G46A and G48A mutations mentioned earlier, which in the holoprotein produced a total increase of 6.2°C in $t_{1/2}$ (*15*), and in addition a Y88C mutation which permitted formation of an intersubunit disulfide bond. It had previously been shown (*20*) that this mutation caused an increase in $t_{1/2}$ of 8°C. The protein containing all three mutations showed $\Delta t_{1/2} = 16$°C, a very substantial apparent stabilization, probably including a contribution from decreased entropy of the unfolded state. A similar engineered disulfide bond introduced into T4 lysozyme (*21*) was found to enhance the stability of the protein.

Presumably there would be pharmacological or other importance in engineering an increase in thermostability for certain proteins. In this connection it should not be overlooked that nature has already designed some very stable proteins. An outstanding example of this is given by the tryptophan aporepressor of *E. coli*, hardly a thermophilic bacterium (*22*). This protein is a dimer of molecular weight 24.7 kDa, containing no metal or other non-amino acid component, and of ordinary appearing amino acid sequence. Yet when it is heated at pH 7.5 in 10 mM phosphate buffer containing 100 mM KCl, it undergoes unfolding and dissociation into monomeric units in a transition centered at 93°C. If the heating is interrupted at 100°C and the solution cooled to 50°C during 30 min, the unfolding-

dissociation transition is found to be quantitatively repeated on reheating.

A striking case of additivity of mutational effects on stability as measured by thermal changes in the degree of helicity was reported by Merutka and Stellwagen (*23*) for the model peptide $CH_3COAEAAAKEAAAKEAAAKCONH_2$. Replacing the central Ala in an Ala_3 group by Ser lowered $t_{1/2}$, the temperature at which the content of α-helix was half the maximal value for the original peptide, by 11°C regardless of which Ala_3 group was changed. Making simultaneous replacements in two groups lowered $t_{1/2}$ by 22°C regardless of which two groups were affected, and changing all three groups caused a 31°C lowering. Replacements of the central Ala by Met were also additive, with effects approximately two-thirds as large as those produced by Ser.

Andersen et al. (*24*) have presented a convincing picture of an important stabilizing salt bridge in T4 lysozyme (see also *25*). The x-ray structure for the C54T:C97A mutant, designated WT*, shows that His 31 and Asp 70 are well placed to form a salt bridge. Andersen et al. determined by NMR that the pK_a of His 31 is 9.05 in the folded form and 6.8 in the unfolded form, and that of Asp 70 is 0.5 in the folded and 3.5 - 4.0 in the unfolded protein. The pK_a of His 31 is 6.9 in the mutant D70N, showing that the change in this pK_a in WT* is due to the presence of Asp 70. When WT* unfolds the change in the free energy of ionization of His 31 is approximately 3 kcal mol^{-1} and that of Asp 70 is approximately 3.9 kcal mol^{-1}. Thus the salt bridge contributes at least 4-5 kcal mol^{-1} to the stabilization free energy of the protein. Correspondingly, $t_{1/2}$ for D70N is about 12°C lower than that of WT*.

Isothermal Titration Calorimetry

ITC is a very useful method for the determination of the thermodynamics of reactions, having as its main advantage over other isothermal methods that enthalpies are directly determined. Titration calorimeters have been developed to the point that a total enthalpy change of as little as a hundred microcalories is adequate for obtaining 20 points on a titration curve in a time duration of 60-90 min, provided the reaction being studied has a half time of less than a minute, and the equilibrium constant for the reaction lies in the range $10^3 M^{-1}$ to $10^8 M^{-1}$. Reaction half times greater than a minute can be handled with longer periods between injections of titrant, but a value greater than 10 min would probably lead to a serious loss of accuracy.

A curve fitting analysis of the titration data leads to a value for the equilibrium constant, K, and therefore the standard free energy change

$$\Delta G^o = -RT \ln K \qquad (8)$$

the enthalpy change, ΔH_{cal}, and therefore also the standard entropy change, ΔS^o, and a value for the number of interaction sites, N, per molecule of substrate. Titrations over a temperature range permit evaluation of the heat capacity change in the reaction, ΔCp, and the van't Hoff enthalpy, ΔH_{vH}.

In 1959, Richards and Vithayathil (*26*) observed that mild proteolysis of ribonuclease A (RNaseA) severed one peptide bond between residues 20 and 21 near the N-terminus. The product, RNase S, had full enzymic activity and unchanged x-ray structure. At low pH the peptide and protein can be separated, and at neutral pH they recombine to form RNaseS which is indistinguishable from RNase S. We have used the Omega microcalorimeter (*27*) to study in detail the binding of modified S-peptides to S-protein to form modified RNaseS. In our work (*28, 29*) we have used abbreviated peptides lacking the residues 16 to 20, which do not appear to have much effect on the binding, with Met 13 replaced by various other residues.

It may be noted here that the effects of mutations on this type of reaction might well be more susceptible to quantitative analysis than has been possible for protein thermal denaturations. Here one is starting with the same protein in each case and is adding 15-residue peptides none of which has significant specific structure in solution. Thus all variations in thermodynamic properties resulting from changes at position 13 can with some confidence be attributed to differences in the products formed. The crystal structures of these products are currently being determined to high resolution.

The results of our study are summarized in Table IV. The large decreases in entropy accompanying these reactions can be

Table IV. Changes in the Thermodynamic Parameters for the Modified S-peptide S-protein Interaction at pH 6.0, 5° - 25°C, Produced by Replacing M13 in the Peptide by the Indicated Residues

(WT at pH 6.0, 25°C: ΔG^o = -9.4 kcal mol^{-1}; ΔH = 41.3 kcal mol^{-1}; ΔS^o = 108.1 cal K^{-1}mol^{-1}; ΔCp = 910 cal K^{-1}mol^{-1}

Peptide	$\Delta\Delta G^o$(25°C) kcal mol^{-1}	$\Delta\Delta H$(25°C) kcal mol^{-1}	$\Delta\Delta S^o$(25°C) cal K^{-1}mol^{-1}	$\Delta\Delta Cp$ cal K^{-1}mol^{-1}
M13A	4.0	5.0	3	-50
M13ANB[a]	1.3	9.5	28	200
M13V	-0.4	4.8	17	270
M13I	-0.3	5.8	21	250
M13L	0.1	5.8	19	100
M13F	2.7	4.4	6	20

[a] α-Amino-n-butyrate residue.

largely attributed to a general tightening of the structure of the protein on addition of the peptide, as is also indicated by the observation of Rosa and Richards (*30*) that the rate of proton exchange in S-protein is 1000 times that in RNase S.

It has been argued that the heat capacity increment on protein unfolding is associated primarily with the ordering of water molecules around newly exposed non-polar groups (*31, 32, 33*). However, the differences in heat capacity shown in Table IV cannot be accounted for on that basis. With the exception of M13A and M13F, the observed values for $\Delta\Delta Cp$ are at least an order of

magnitude larger than those calculated on the basis of the treatment proposed by Spolar et al.(*32*), assuming that the only differences in the non-polar surface areas that become buried are those between Met13 and the groups which replace it. It seems likely that the different peptides cause differing degrees of tightening of the S-protein, which could also be detected by measurements of proton exchange rates. This sort of effect may not be detectable by x-ray crystallography. It should be added that an additional significant contribution to ΔCp, which can become very large at temperatures above 25°C, arises from the difference between the thermal unfolding behavior of S-protein and that of RNase S′ as demonstrated by DSC (*29,34*). This contribution would also probably vary between the various RNase S′ products.

It is evident that, as seen for the thermal denaturation of mutant proteins, there is no correlation between the values for $\Delta\Delta G^{o}$ and $\Delta\Delta H$. Actually in this system the values for $\Delta\Delta G^{o}$ cover a much wider fractional range, -0.3 kcal mol^{-1} for M13I to 4.0 kcal mol^{-1} for M13A, than do the values for $\Delta\Delta H$. It will probably be very difficult to account quantitatively for a 4.3 kcal mol^{-1} difference between the binding free energies for these two peptides. It seems likely that the changes, and the differences in the changes, in the S-protein caused by the binding of the various S-peptides are distributed widely in the protein molecule.

Acknowledgements

The individuals from whom we have received generous supplies of proteins have been mentioned above. The author is greatly indebted to Drs. John Brandts (University of Massachusetts), Brian Matthews (University of Oregon), and Frederic Richards (Yale University) for very useful discussions and other aids. Most of the actual experimentation was carried out by my colleagues S.-J. Bae, P. Connolly, L. Ghosaini, C.-Q. Hu, N. Kishore, S. Kitamura, J. Ladbury, A. Tanaka, J. Thomson and R. Varadarajan.

Literature Cited

1. Matthews, B. W. *Biochemistry* **1987**, *26*, 6885-6888.
2. Avbelj, F.; Moult, J.; Kitson, D.H.; James, M.N.G.; Hagler, A.T. *Biochemistry 1990, 29*, 8658-8676.
3. Sturtevant, J.M. *Ann. Rev. Phys. Chem.* **1987**, *38*, 463-488.
4. Kitamura, S.; Sturtevant, J.M. *Biochemistry* **1989**, *28*, 3788-3792.
5. Connelly, P. R.; Ghosaini, L.; Hu, C.-Q.; Kitamura, S.; Tanaka, A.; Sturtevant, J.M. *Biochemistry* **1991**, *30*, 1887-1892.
6. Hu, C.-Q.; Kitamura, S.; Tanaka, A.; Sturtevant, J.M. *Biochemistry* **1992** *31*, 1643-1647.
7. Tidor, B.; Karplus, M. *Biochemistry* **1991**, *30*, 3217-3228.
8. Weaver, L.H.; Matthews, B.W. *J. Mol. Biol.* **1977**, *193*, 189- 199.
9. Alber, T.; Dao-pin, S.; Wilson, K.; Wozniak, J.A.; Cook, S.P.; Matthews, B.W. *Nature* **1987**, *330*, 41-46.
10. Dang, L.X.; Merz, K.M.; Kollman, P.A. *J. Am. Chem. Soc.* **1989**, *111*, 8505-8508.
11. Shortle, D.; Meeker, A.K.; Freire, E. *Biochemistry* **1988**, *27*, 4761-4768.

12. Poole, L.B.; Loveys, D.A.; Hale, S.P.; Gerlt, J.A.; Stanczyk, S.M.; Bolton, P.H. *Biochemistry* **1991**, *30*, 3621-3627.
13. Hecht, M.H.; Sturtevant, J.M.; Sauer, R.T. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 5685-5689.
14. Hecht, M.; Hehir, K.; Nelson, H.; Sturtevant, J.; Sauer, R. *J. Cellular Biochem.* **1985**, *29*, 217-224.
15. Hecht, M.H.; Sturtevant, J.M.; Sauer, R.T. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 43-46.
16. Brandts, J.F.; Hu, C.-Q.; Lin, L.N. *Biochemistry* **1989** *28*, 8588-8596.
17. Chou, P.Y.; Fasman, G.D. *Advanced Enzymol.* **1978**, *47*, 45-148.
18. Matthews, B.W.; Nicholson, H.; Becktel, W.J. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6663-6667.
19. Stearman, R.S.; Frankel, A.D.; Freire, E.; Lui, B.; Pabo, C.O. *Biochemistry* **1988**, *27*, 7571-7574.
20. Sauer, R.T.; Hehir, K.; Stearman, R.S.; Weiss, M.A.; Jeitler-Nikson, A.; Suchanek, E.G.; Pabo, C.O. *Biochemistry* **1986** *25*, 5992-5998.
21. Matsumara, M.; Matthews, B.W. *Science* **1989** *243*, 792-794.
22. Bae, S.-J.; Chou, W.-Y.; Matthews, K.S.; Sturtevant, J.M. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6731-6732.
23. Merutka, G.; Stellwagen, E. *Biochemistry* **1990**, *29*, 894-898.
24. Andersen, D.E.; Becktel, W.J.; Dahlquist, F.W. *Biochemistry* **1990**, *29*, 2403-2408.
25. Dao-pin, S.; Sauer, U.; Nicholson, H.; Matthews, B.W. *Biochemistry* **1991**, *30*, 7142-7153.
26. Richards, F.M.; Vithayathil, P.J. *J. Biol. Chem.* **1959** *234*, 1459-1465.
27. Wiseman, T.; Williston, S.; Brandts, J.; Lin, L. *Anal. Biochem.* **1989**, *179*, 131-137.
28. Connelly, P.R.; Varadarajan, R.; Sturtevant, J.M.; Richards, F.M. *Biochemistry* **1990**, *29*, 6108-6114.
29. Varadarajan, R.; Connelly, P.R.; Sturtevant, J.M.; Richards, F.M. *Biochemistry*, in press.
30. Rosa, J.J.; Richards, F.M. *J. Mol. Biol.* **1981**, *145*, 834-851.
31. Baldwin, R.L. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8069-8072.
32. Spolar, R.S.; Ha, J.; Record, T.M. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8382-8385.
33. Murphy, K.P.; Privalov, P.L.; Gill, S.J. *Science* **1990**, *247*, 559-561.
34. Hearn, R.P.; Richards, F.M.; Sturtevant, J.M.; Watt, G.D. Biochemistry **1971**, *10*, 806-817.

RECEIVED March 31, 1992

Chapter 2

Contribution of Hydrogen Bonding and the Hydrophobic Effect to Conformational Stability of Ribonuclease T1

C. Nick Pace, Ketan Gajiwala, and Bret A. Shirley

Departments of Medical Biochemistry and Genetics, Biochemistry and Biophysics, Center for Macromolecular Design, Texas A&M University, College Station, TX 77843

When RNase T1 folds, 86 intramolecular hydrogen bonds are formed and 82% of the nonpolar side chains are buried. Twelve mutants of RNase T1 (Tyr --> Phe (5), Ser --> Ala (3), and Asn --> Ala (4)) have been prepared that remove 17 of the hydrogen bonds. Based on urea and thermal unfolding studies of these mutants, the average decrease in conformational stability due to hydrogen bonding is 1.3 kcal/mole per hydrogen bond. This estimate is in good agreement with results from several related systems. Thus, we estimate that hydrogen bonding contributes about 112 kcal/mole to the conformational stability of RNase T1, and that hydrophobic interactions make a comparable contribution to the stability. Accepting the idea that intramolecular hydrogen bonds contribute 1.3 ± 0.6 kcal/mole to the stability of systems in an aqueous environment makes it easier to understand the stability of the "molten globule" states of proteins, and the α-helical conformations of small peptides.

Most of the important tasks in living cells are carried out by proteins in which the polypeptide chain is tightly folded into a globular conformation that is essential for their biological activity. Consequently, there is great interest in the forces that stabilize globular proteins. In early discussions of protein structure, hydrogen bonding was thought to be the most important force contributing to the conformational stability. The main proponent of this view was Linus Pauling, who wrote with Mirsky in 1936 (*1*): "The importance of the hydrogen bond in protein structure can hardly be overemphasized". However, by the 1950s the emphasis had shifted and the importance of hydrophobic interactions was stressed first by Kauzmann (*2*) and later by Tanford (*3*), who used model compound data and calculations based on a simple model to conclude: " ··· the stability of the native conformation in water can be

0097–6156/93/0516–0018$06.00/0

explained ··· entirely on the basis of the hydrophobic interactions of the non-polar parts of the molecule". The view that hydrophobic interactions make a more important contribution than hydrogen bonding to globular protein stability is still widely held today. As examples from 1990, Kim and Baldwin state (*4*): "Stripping H_2O from nonpolar side chains to form a hydrophobic core provides the main source of free energy stabilizing a folded protein"; Creighton states (*5*): "Nevertheless, the hydrophobic interaction is probably the major stabilizing factor"; and, in a recent review that gives an excellent overview of the forces contributing to globular protein stability, Dill states (*6*): "More than 30 years after Kauzmann's insightful hypothesis, there is now strong accumulated evidence that hydrophobicity is the dominant force of protein folding, provided that "hydrophobic" is operationally defined in terms of the transfer of nonpolar amino acids from water into a medium that is nonpolar and preferably capable of hydrogen bonding". This is the definition for "hydrophobic" used here.

In this article, we suggest that hydrogen bonding and hydrophobic interactions make comparable contributions to the conformational stability of ribonuclease T1 and other small globular proteins. This is based on studies reported here of mutants of RNase T1 designed to improve our understanding of hydrogen bonding, and on results from a number of other laboratories published in the last five years (*7-14*).

Hydrogen Bonding in RNase T1

RNase T1 contains 293 nitrogen and oxygen atoms capable of donating or accepting hydrogen bonds, 205 are in the backbone and are in the side chains. These atoms can form a total of 503 hydrogen bonds, 312 by the backbone and 191 by the side chains, based on the hydrogen bonding capabilities given by Baker and Hubbard (*15*). Using the geometrical criteria suggested by these authors, folded RNase T1 contains 86 intramolecular hydrogen bonds with an average length of 2.95 Å. (See Table I for references and information on the hydrogen bonding analysis of RNase T1). Fifty two of these hydrogen bonds are between peptide groups in the polypeptide backbone, mostly in the secondary structure, 22 are between peptide groups and side chains, and 12 are between side chains. Another 168 hydrogen bonds to water molecules have been identified.

Our main interest is to estimate how much the intramolecular hydrogen bonds contribute to the conformational stability of RNase T1. As one approach, we have selected 12 side chains that participate in intramolecular hydrogen bonds and used site-directed mutagenesis to substitute an amino acid that removes the hydrogen bonding capability. To disrupt the folded structure as little as possible, the substitutions made were: 5 (Tyr → Phe), 3 (Ser → Ala), and 4 (Asn → Ala) substitutions. These groups are capable of participating in 2, 3, and 4 hydrogen bonds, and can serve as either hydrogen bond donors or acceptors. Table I lists the side chains selected, and relevant information about the hydrogen bonds of each in wild type RNase T1. Note that nine of the groups form one hydrogen bond, one forms two hydrogen bonds, and two form three hydrogen bonds. The average donor-acceptor distance for these 17 hydrogen bonds is 2.93 Å.

Table I. RNase T1 Hydrogen Bonds Removed by Amino Acid Substitutions[a]

Residue	H-Bond & partner		Length(Å)	Angle(°)	%Buried	
Tyr 11	OH ••• OD2	Asp 76	2.69	132	99	99
Tyr 42	OH ••• OD1	Asn 44	2.97	133	98	99
Tyr 56	OH ••• O	Val 52	2.83	157	90	73
Tyr 57	OH ••• OE2	Glu 82	2.56	124	85	70
Tyr 68	OH ••• O	Gly 71	2.68	139	99	98
Ser 12	OH ••• OH	Ser 14	2.71	131	54	88
	OH ••• N	Asp 15	3.24	134		88
Ser 17	OH ••• O	Ser 13	3.19	136	26	24
Ser 64	OH ••• N	Asp 66	3.17	115	53	91
Asn 9	ND2 ••• OD1	Asp 76	2.89	137	66	57
Asn 36	OD1 ••• OH	Ser 35	3.01	115	28	30
Asn 44	ND2 ••• O	Phe 48	2.90	137	81	76
	OD1 ••• N	Phe 48	3.16	102		100
	OD1 ••• OH	Tyr 42	2.97	133		100
Asn 81	OD1 ••• N	Asn 83	2.97	114	100	100
	OD1 ••• N	Gln 85	2.91	143		100
	ND2 ••• OE1	Gln 85	2.94	126		99

[a]The analysis of the hydrogen bonding was done with a program written by Presta and Rose (*62*) and is based on the 1.5 Å crystal structure determined by Martinez-Oyanedel et al. (*51*). Only hydrogen bonds with lengths greater than 3.5 Å and angles greater than 90° were included. Groups that form three center hydrogen bonds, such as some of the amide NH groups in the α-helix, were counted only once.

[b]The accessibility (% buried) of the side chain (e.g., for Ser, -CH_2-OH), and of the individual group that actually forms the hydrogen bond (e.g., for Ser, -OH) are given in the last two columns. The accessibilities were estimated using the Lee and Richards Program (*25*).

Measured Changes in Stability

The unfolding of RNase T1 has been shown to closely approach a two-state folding mechanism (*16, 17*), and the conformational stability has been carefully measured under a variety of conditions (*17, 18*). For our previous studies, wild type RNase T1 had Gln at position 25, Gln 25-RNase T1. This is the RNase T1 first isolated in 1957 by Sato and Egami (*19*) and used for most of the subsequent studies in countries other than Germany. However, all of the crystal structures determined by Saenger's group (*20*) have been determined with a RNase T1 variant with Lys at position 25, Lys 25-RNase T1. For this reason and also because the solubility and conformational stability of Lys 25-RNase T1 are greater than for Gln 25-RNase T1 (*21*), we plan to denote Lys 25-RNase T1 as wild type in the future. Thus, all of the RNase T1 mutants discussed here have Lys at position 25.

The differences in conformational stability between wild type RNase T1 and the hydrogen bonding mutants were determined using both urea and thermal unfolding experiments. As explained elsewhere (*22*), an analysis of urea unfolding experiments yields ΔG as a function of urea concentration, and an analysis of thermal unfolding experiments yields ΔG as a function of temperature. From plots of these data, the midpoints of the transitions (where $\Delta G = 0$), $(urea)_{1/2}$ or T_m, and measures of the steepness of the transitions, m or ΔS_m (the slopes of plots of ΔG versus urea, m, or temperature, ΔS_m), can be determined. These parameters are listed in Table II along with estimates of the difference in stability, $\Delta(\Delta G)$, from both the urea and thermal unfolding experiments. It can be seen that the estimates from the two different experiments are in excellent agreement. This need not have been the case because the results apply to different sets of conditions: 25°C in the presence of 2.9 to 6.3 M urea for the results from urea unfolding, and 42 to 55°C in the absence of urea for the results from thermal unfolding. The close agreement between the two sets of data gives us considerable confidence that the measured differences in stability are quite reliable.

Contribution of Hydrogen Bonding

A difficult problem in interpreting results from stability studies of mutant proteins is how to correct for the contribution of conformational entropy to the observed differences in stability. Conformational entropy can effect the $\Delta(\Delta G)$ values in several ways. First, the amino acid side chain can effect rotation around the N-C_α, ϕ, and C_α-$C_{carbonyl}$, ψ, bonds in the unfolded protein (*23*). This is especially important for substitutions involving Pro residues where rotation around ϕ is largely eliminated, and for substitutions involving Gly residues where rotation around both ϕ and ψ is much less restricted. This effect has been investigated experimentally with mutants of T4 lysozyme (*24*). This effect should make an insignificant contribution for the three types of mutants studied here. Second, conformational entropy can contribute to $\Delta(\Delta G)$ for any amino acid substitution that changes the number of bonds in the side chain since rotation around these bonds will generally be more restricted in the folded

Table II. Parameters Characterizing the Urea and Thermal Unfolding of Wild Type RNase T1 (Lys 25-Rnase T1) and Twelve Hydrogen Bonding Mutants in 30 mM MOPS Buffer, pH 7 [a]

	Urea Unfolding			Thermal Unfolding			
Protein	m	$(urea)_{1/2}$	$\Delta(\Delta G)$	ΔH_m	ΔS_m	T_m	$\Delta(\Delta G)$
	cal/M	M	kcal	kcal	cal/°C	°C	kcal
RNase T1	1210	5.30		110	339	50.9	
Tyr 11 Phe	1270	3.56	-2.11	101	317	44.9	-2.03
Tyr 42 Phe	1170	6.24	1.14	106	325	54.3	1.15
Tyr 56 Phe	1260	4.65	-0.78	99	306	48.8	-0.71
Tyr 57 Phe	1285	4.88	-0.50	107	332	49.6	-0.44
Tyr 68 Phe	1320	4.17	-1.36	89	279	46.9	-1.35
Ser 12 Ala	1275	4.29	-1.23	99	309	47.7	-1.08
Ser 17 Ala	1215	5.85	0.67	109	334	52.6	0.57
Ser 64 Ala	1375	4.11	-1.44	104	324	46.3	-1.56
Asn 9 Ala	1275	4.56	-0.90	101	313	48.8	-0.71
Asn 36 Ala	1310	5.31	0.03	111	344	50.9	0.00
Asn 44 Ala	1300	3.59	-2.08	91	287	45.4	-1.86
Asn 81 Ala	1435	2.92	-2.87	91	287	42.3	-2.91

[a] m and ΔS_m are the slopes of plots of ΔG versus urea and T, respectively; and $(urea)_{1/2}$ and T_m are the midpoints ($\Delta G = 0$) of urea and thermal denaturation curves (22). For the urea data, $\Delta(\Delta G) = \Delta[(urea)_{1/2}] \times m$ (wild type). For the thermal data, $\Delta(\Delta G) = \Delta T_m \times \Delta S_m$ (wild type). Based on many independent experiments, the uncertainties are estimated to be ±50 in m, ±0.05 in $(urea)_{1/2}$, ±15 in ΔS_m, and ±0.5 in T_m. These give rise to an uncertainty of about ±0.25 kcal/mole in the $\Delta(\Delta G)$ values.

protein than in the unfolded protein. Finally, conformational entropy can contribute to $\Delta(\Delta G)$ for any mutation that fills or creates a hole in the folded protein. Little is known about these latter two effects. Richards (*25*) has shown that Tyr and Phe residues occupy almost the same volume in folded proteins, that Ser residues are ≈ 8 Å^3 larger than Ala residues, and that Asn residues are ≈ 44 Å^3 larger than Ala residues. For comparison, adding a -CH_2- or CH_3 group increases the volume of the residue by ≈ 26 Å^3. On reflection, it seems likely that conformational entropy will make a smaller contribution to the measured $\Delta(\Delta G)$ values for Tyr --> Phe and Ser --> Ala mutants than for any other of the mutations that can be used to study hydrogen bonding or hydrophobic interactions, and may well be less than the experimental error for our $\Delta(\Delta G)$ values. For the Asn --> Ala mutants, the contribution of conformational entropy should be larger, but the magnitude is unknown and there is presently no way to make corrections.

For each of the three types of amino acid substitutions reported here, the amino acid side chain in wild type RNase T1 is less hydrophobic than the side chain in the mutant. This effect by itself is expected to increase the stability of the mutants relative to wild type RNase T1, and will surely contribute to the measured $\Delta(\Delta G)$ listed in Table II. We will be able to get a more accurate assessment of the contribution of hydrogen bonding to the differences in stability if we can correct, at least approximately, for this effect. Measurements in several different solvent systems have been used to assess the magnitude of hydrophobic interactions for the amino acid side chains (*26-30*). In Table III (A), we give estimates of the free energy of transfer, ΔG_{tr}, of a -CH_2- group from several different solvent systems to water. The question is which of these solvents is the best model for the interior of a protein? Hexane and cyclohexane are surely too nonpolar, and formamide and methanol are surely too polar. All of the other solvents give comparable ΔG_{tr} values and, consequently, we have chosen to use the complete set of data available for n-octanol (*28*) to make the corrections described in Table IV, and for the calculations presented below in Table V. Note that the octanol data are consistent with the definition suggested by Dill (*6*) (see the Introduction). The octanol phase in the partitioning experiments used to determine the ΔG_{tr} values is capable of hydrogen bonding since it contains 2.5 M H_2O in addition to the octanol. Thus, the ΔG_{tr} values should reflect only changes in hydrophobicity and not hydrogen bonding.

In support of this choice, studies of hydrophobic interactions based on experimental results from mutant proteins generally show reasonable correlations with either the octanol data (*31-33*) or the ethanol data (*34, 35*), although the magnitude of the effect may be larger than predicted (*32, 34, 36, 37*). The best results to consider are for Ile --> Val mutants since they differ by only a -CH_2- group and this will minimize the size of the hole in the mutant, and possible conformational entropy effects. (Leu --> Val mutants are more likely to have unfavorable steric effects). Results for nine Ile --> Val mutants from four different proteins are given in Table III (B). The average $\Delta(\Delta G)$ is 1.1 ± 0.3 kcal/mole. This is 50% higher than the ΔG_{tr} value for a -CH_2- group based on the octanol data. Thus, these results point out the possibility that the ΔG_{tr} values from the octanol data might underestimate the contribution of hydrophobic interactions. Some possible explanations for this have been discussed (*5, 32, 38*).

Table III. ΔG_{tr}(solvent ---> water) Values for a -CH_2- Group Based on Studies with Model Compounds (A), and $\Delta(\Delta G)$ Values for Ile ---> Val Mutants (B) (kcal/mole)

A		B		
Solvent	ΔG_{tr}	Protein	Residue	$\Delta(\Delta G)$
Formamide[26]	0.32	S. Nuclease[36]	15	0.8
N-methylacetamide[a,27]	0.74		18	1.1
Acetone[26]	0.67		72	1.8
Methanol[26]	0.60		92	0.5
Ethanol[26]	0.67		139	1.5
Butanol[26]	0.73	Barnase[32]	88	1.3
Heptanol[26]	0.73		96	1.2
Octanol[a,28]	0.73	Gene 5 protein[37]	47	1.2
Cyclohexane[a,29]	0.88	T4 lysozyme[31]	3	0.5
Hexane[a,30]	1.00	Average = 1.1 ± 0.3 kcal/mole		

[a]The average of the difference between the ΔG_{tr} values for Leu & Val and for Ile is used for these solvents.

Table IV. Estimate of the Contribution of Hydrogen Bonding to the Measured Differences in Stability

Protein	H-bonds[a]	Δ(ΔG)[b]	Δ(ΔG)[c]	Δ(ΔG)/H-bond
	number	kcal/mol	kcal/mol	kcal/mol
Tyr 11 Phe	1	-2.07	-3.2	-3.2
Tyr 42 Phe	1	1.15	0	0
Tyr 56 Phe	1	-0.75	-1.8	-1.8
Tyr 57 Phe	1	-0.47	-1.4	-1.4
Tyr 68 Phe	1	-1.36	-1.5	-2.5
Ser 12 Ala	2	-1.16	-1.4	-0.7
Ser 17 Ala	1	0.62	0	0
Ser 64 Ala	1	-1.50	-1.7	-1.7
Asn 9 Ala	1	-0.81	-1.6	-1.6
Asn 36 Ala	1	0.02	-0.6	-0.6
Asn 44 Ala	3	-1.97	-3.0	-1.0
Asn 81 Ala	3	-2.89	-4.1	-1.4

[a]The number of intramolecular hydrogen bonds in wild type RNase T1 (see Table I).
[b]This is the average of the Δ(ΔG) values in Table III from urea and thermal unfolding experiments.
[c]The Δ(ΔG) values in the preceding column have been corrected for the effect of differences in hydrophobicity between the amino acid in wild type RNase T1 and the mutants by multiplying the following $\Delta(\Delta G_{tr})$ values: -1.1 (Tyr ---> Phe), -0.4 (Ser ---> Ala), and -1.2 (Asn ---> Ala) (in kcal/mole) by the accessibility of the side chain in the wild type protein given in Table I, as described in the text. In addition, for the Ser 17 Ala mutant a -0.5 kcal/mole correction was applied to correct for the differential effect that the two residues have on the stability of the α-helix, as described in the text.

Table V. Contribution of Hydrophobic Interactions to the Conformational Stability of RNase T1

Side Chain	#Present	#Buried[a]	ΔG_{tr}[b]	#Buried x ΔG_{tr}
Trp	1	1.0	3.1	3.1
Phe	4	3.8	2.4	9.2
Ile + Leu	5	4.3	2.4	10.3
Val	8	6.7	1.7	11.4
Tyr	9	7.7	1.3	10.0
Cys	4	3.2	1.3	4.2
Pro	4	3.2	1.0	3.2
Thr	6	3.3	0.4	1.3
Ala	7	4.7	0.4	1.9
His	3	2.2	0.2	0.4
Ser	15	3.9	0.0	0.0
Gln	2	1.3	-0.3	-0.4
Asn	9	5.0	-0.8	-4.0
$-CH_2-$[c]	29	19.4	0.7	13.6
				64.3 kcal/mol

[a]Calculated with the Lee & Richards accessibility program (*25*).
[b]Fauchere and Pliska (*28*) (ΔG_{tr}: n-octanol ---> water).
[c]$-CH_2-$ groups from the Lys, Arg, Glu, and Asp side chains.

The corrections made in Table IV take into account the difference between the ΔG_{tr} values for the pairs: Tyr-Phe (1.1 kcal/mol), Asn-Ala (1.2 kcal/mol), and Ser-Ala (0.4 kcal/mol); and the accessibility of the side chain in the wild type protein estimated using the Lee and Richards procedure (*25*). (The accessibility of each of the side chains in wild type RNase T1 is given in Table I). One further correction was made for the Ser 17 Ala mutant because it occurs at a largely exposed site in the α-helix of RNase T1. Recent studies of peptides have provided quantitative estimates of the helix-forming tendencies of the amino acids (*39-42*). The Ser 17 Ala substitution is expected to stabilize the mutant because Ala has a greater helix-forming tendency than Ser. The predicted stabilization is 0.51 kcal/mole based on the data of Lyu et al. (*40*), 0.46 kcal/mole based on the data of Merutka and Stellwagen (*41*), and 0.42 kcal/mole based on the data of O'Neil and DeGrado (*42*). A value of 0.5 kcal/mole was used for the correction of the Ser 17 Ala data in Table IV.

Before the corrections, three of the Δ(ΔG) values were positive, i.e., the mutants were more stable than wild type RNase T1. After the corrections, two of the Δ(ΔG) values are zero and the rest are negative, ranging as high as -4.1 kcal/mole for the Asn 81 Ala mutant that potentially removes three hydrogen bonds. On a per hydrogen bond basis, the values range from 0 to -3.2 kcal/mole with an average value of -1.3 kcal/mole. Thus, we think the Δ(ΔG) values given in the last two columns in Table IV are due mainly to changes in the hydrogen bonding in the mutants.

The results in Table IV cannot be interpreted with any certainty until three-dimensional structures are available for the mutant proteins. (All of the mutants have enzyme activity suggesting that their conformations do not differ too much from wild type RNase T1). A key question is what happens to the hydrogen bonding partner in the mutants: does it form a new intramolecular hydrogen bond, does it form a hydrogen bond to a water molecule, or does it fail to form any hydrogen bonds? In the first case, the Δ(ΔG) value should be close to zero if the new hydrogen bond is approximately equivalent to the hydrogen bond in wild type RNase T1. This may be the case for the two mutants for which $\Delta(\Delta G) \approx 0$. In the second case, the Δ(ΔG) value would give the estimate that we are most interested in, namely, what the formation of a single intramolecular hydrogen bond contributes to the conformational stability of the protein. We argue below that this is the case for most of the mutants studied here. In the final case, the Δ(ΔG) values should be considerably more negative because now both the loss of the intramolecular hydrogen bond, and the formation of hydrogen bonds to water in the unfolded protein by the unpaired partner will contribute to the decrease in conformational stability. This may be the reason that the Δ(ΔG) value for Tyr 11 Phe is considerably more negative than any of the other values.

For the Tyr 11 hydrogen bond, neither the hydrogen bond donor nor acceptor forms a hydrogen bond with water in wild type RNase T1. Furthermore, the side chains of Tyr 11 and Asp 76 are almost completely buried (99%). Thus, it is certainly possible that the carboxyl group of Asp 76 may not be able to form a hydrogen bond in the Tyr 11 Phe mutant, and that this contributes to the very negative Δ(ΔG) value that is observed. In addition, the hydrogen bond lost is a good hydrogen bond and it is probably to a charged acceptor since Asp 76 is likely to be ionized at pH 7. There is no evidence from the titration curve for RNase T1 (*43*) or from the pH dependence of the stability (*44*) for the presence of a carboxyl group with a pK

of 7 or higher. Note that the hydrogen bonds of Tyr 57 and Asn 9 are also probably to charged acceptors.

For all of the other hydrogen bonds considered here, either the hydrogen bond donor or acceptor forms a hydrogen bond with a water molecule in addition to the intramolecular hydrogen bonds described in Table I. For this reason and because more room is available for hydrogen bonding to water molecules in the mutants, we think that the hydrogen bonding partner will most often form a hydrogen bond to a water molecule in the mutant. Since the same group should also hydrogen bond to water molecules in the unfolded protein, they should no longer contribute significantly to the conformational stability through hydrogen bonding. On this basis, we assume that the eight $\Delta(\Delta G)$ values between -0.6 and -1.8 kcal/mole in the last column of Table IV give the contribution of single hydrogen bonds to the conformational stability of RNase T1. The average of these eight values is -1.3 kcal/mole which does not differ significantly from the average of -1.3 kcal/mole for all twelve of the $\Delta(\Delta G)$ values. We will see below that this estimate is in good agreement with estimates from completely different systems.

Related Studies of Hydrogen Bonding

The longstanding question is whether the formation of intramolecular hydrogen bonds in folded proteins makes a net favorable contribution to protein stability, i.e., is ΔG favorable for the reaction:

$$(>C{=}O\cdot\cdot(H_2O)_2 + H_2O\cdot\cdot H\text{-}N<)_{water} \longleftrightarrow (>C{=}O\cdot\cdot H\text{-}N<)_{protein}$$

Pauling and Corey (*45*) thought so: "With proteins in an aqueous environment the effective energy of hydrogen bonds is not so great, inasmuch as the difference between the energy of the system with the N-H···O hydrogen bonds surrounded by water and a system with the N-H group and the O atom forming hydrogen bonds with water molecules may be no more than around 2 kcal/mole". However, a variety of model compound studies failed to give an unequivocal answer, and some even suggested that the hydrogen bonds with water would be favored (See Dill (*6*) for a review of this literature). These studies strengthened the idea that hydrophobic interactions are the dominant force in protein folding.

With the advent of site-directed mutagenesis, it has been possible to estimate ΔG for the hydrogen bonding reaction shown above more directly. Fersht (*7*) summarized these studies and concluded "···an individual uncharged hydrogen bond contributes some 0.5 to 1.8 kcal/mole to binding energy···." Similar estimates had been obtained much earlier in studies with nucleic acids. In 1964 Crothers and Zimm (*46*) suggested that each hydrogen bond between the bases in double-helical nucleic acids contributes 1 kcal/mole to the stabilizing free energy. This estimate has since been substantiated by extensive studies by Turner's group (*8*). More recent studies of enzyme-substrate (*13*), protein-sulfate (*14*), and protein-base (*9*) hydrogen bonds give comparable estimates. Why the intramolecular hydrogen bonds are more favorable than the hydrogen bonds to water is not clear. It could be due to stronger electrostatic

interactions in the more nonpolar, solid-like environment of the intramolecular hydrogen bonds, or differences in entropy effects (*5, 6*). In any event, the estimates of the contribution of individual hydrogen bonds to the conformational stability of RNase T1 given in Table IV are in good agreement with other studies of hydrogen bonding in an aqueous environment.

The Conformational Stability of RNase T1

The major force favoring the unfolding of proteins is conformational entropy. In the folded protein, rotation around the bonds in the backbone and side chains will generally be more limited than in the unfolded protein. Kauzmann (*47*) made a rough guess that this would favor unfolding by about 1.2 kcal/mole per residue, and Privalov (*48*) has pointed out that this is consistent with one interpretation of experimental results on the thermodynamics of unfolding of small, monomeric globular proteins. (See Dill et al. (*49*) for a method of estimating the conformational entropy based on a different approach that gives a considerably lower estimate). Using this estimate, conformational entropy would favor unfolding by 125 kcal/mole for RNase T1. Experimental studies (*50*) have shown that this would be lowered by about 7 kcal/mole due to the restraints imposed by the two disulfide bonds in RNase T1. Thus, this admittedly crude estimate suggests that conformational entropy favors unfolding by ≈ 118 kcal/mole. How is this overcome?

The approach used to estimate the contribution of hydrophobic interactions to the conformational stability of RNase T1 is given in Table 5. We assume that all of the side chains are fully exposed in the unfolded protein and estimate the number of side chains buried in the folded protein using the Lee and Richards algorithm (*25*) and the 1.5 Å resolution crystal structure of RNase T1 (*51*). The ΔG_{tr} values are based on the octanol data. This analysis suggests that hydrophobic interactions contribute 64 kcal/mole favoring folded RNase T1. Based on the experimental studies of Ile --> Val mutants (Table 3 (B)), these ΔG_{tr} values may be too low. If the ΔG_{tr} values are increased by 50%, the contribution of hydrophobic interactions becomes 96 kcal/mole. Spolar et al. (*52*) have used a related approach based on a different set of model compound data (ΔC_p measurements of a series of hydrocarbons) to conclude that the contribution of hydrophobic interactions can be estimated using: ΔG (hydrophobic interactions) = (80 ± 10) x ΔC_p. Values of ΔC_p = 1220 (*53*), 1240 (*54*), 1590 (*54*), and 1650 (*55*) cal/mole/K have been determined for RNase T1 unfolding using different approaches. These ΔC_p values lead to estimates of the contribution of hydrophobic interactions to the stability of 98 to 132 kcal/mole, which is larger than the estimate based on the octanol data (Table III (A)), but comparable to the estimate based on the larger ΔG_{tr} values (Table III (B)). These should be regarded as upper estimates of the contribution of hydrophobic interactions to the stability because they assume that the nonpolar side chains are completely accessible to solvent in the unfolded protein. There is mounting evidence that RNase T1 (*44, 56*) and other proteins (*57*) do not unfold completely after urea or thermal denaturation.

The 17 intramolecular hydrogen bonds described in Table I are typical of the 86 intramolecular hydrogen bonds observed in folded RNase T1. Based on the results

in Table IV, we suggest that the average hydrogen bond in RNase T1 contributes 1.3 kcal/mole to the conformational stability. This leads to the conclusion that hydrogen bonding contributes 112 kcal/mole to the conformational stability of RNase T1.

So, we suggest that hydrogen bonding and hydrophobic interactions make comparable contributions to the conformational stability of RNase T1. Note that the sum of these stabilizing interactions is considerably greater than our estimate of the destabilizing contribution from conformational entropy. There are, of course, a number of other forces that will contribute to the conformational stability that we have not considered. We will mention two that will favor unfolding.

As noted above, unfilled hydrogen bonds in the folded protein should decrease the conformational stability because hydrogen bonding to water molecules will occur when the protein unfolds. In terms of groups, 41 of the 293 groups capable of hydrogen bonding do not form intramolecular hydrogen bonds and do not appear to form hydrogen bonds to water molecules in the crystal structure. (In terms of hydrogen bonds, 165 of the 503 possible hydrogen bonds are not formed). However, the extent of hydrogen bonding with water molecules in solution is sure to be greater because only those water molecules that are largely immobilized will be observed in the crystal structure. Out of the 41 groups that are not hydrogen bonded, only one (the N-H group of Val 52) is completely buried, and seven are actually hyperexposed compared to the model tripeptide used to estimate the accessibility (*25*). Thus, it seems likely that most of these 41 groups will be at least partially hydrogen bonded to water molecules in the folded protein. Nevertheless, unfilled hydrogen bonds in the folded protein will surely contribute unfavorably to the conformational stability, but it is not possible to estimate the magnitude of this contribution at present.

Another factor that will favor unfolding is the transfer of peptide groups from a more nonpolar environment to water on unfolding. For RNase T1, about 70% of the peptide groups are buried in the folded protein. Based on model compound data, the ΔG_{tr} of a hydrogen bonded peptide group from octanol to water is about -1.1 kcal/mole (*58*). Thus, this is potentially a large contribution favoring unfolding that is generally not considered.

Concluding Remarks

On folding, RNase A forms 129 (*59*) and α-chymotrypsin 217 (*60*) intramolecular hydrogen bonds. Using 1.3 kcal/mole per hydrogen bond as above, suggests that hydrogen bonding contributes 168 and 282 kcal/mole to the stability of these two proteins. Using the approach outlined in Table V, we estimate that hydrophobic interactions will contribute 111 kcal/mole to the conformational stability of RNase A, and 205 kcal/mole to the conformational stability of α-chymotrypsin. Thus, RNase T1 is not unique, and we do not agree that " ··· hydrophobicity is the dominant force of protein folding" (*6*). We suggest instead that hydrogen bonding and hydrophobic interactions make comparable contributions to the conformational stability of small, monomeric globular proteins.

The evidence is strong that most intramolecular hydrogen bonds contribute 1.3 ± 0.6 kcal/mole to the stability of structures such as globular proteins and

double-helical nucleic acids in an aqueous environment. Accepting this makes it easier to understand the stability of "molten globules" (*4, 5, 61*) and of the α-helical conformations observed with some small peptides (*4*).

Acknowledgments

This research was supported by grants from NIH (GM 37039), the Robert A. Welch Foundation (A-1060), the Tom & Jean McMullin Professorship, and the Bill & Wanda Pace Trust. We thank Udo Heinemann & Wolfram Saenger for providing us with the 1.5 Å crystal structure of RNase T1, and Leonard Presta and George Rose for providing us with their program for analyzing hydrogen bonding.

Literature Cited

1. A. E. Mirsky; L. Pauling, *Proc. Nat. Acad. Sci. U.S.A.* **22**, *439* (1936).
2. W. Kauzmann, *Adv. Prot. Chem.* **14**, *1* (1959).
3. C. Tanford, *J. Am. Chem. Soc.* **84**, *4240* (1962).
4. P. S. Kim and R. L. Baldwin, *Annu. Rev. Biochem.* **59**, *631* (1990).
5. T. E. Creighton, *Biochem. J.* **270**, *1* (1990).
6. K. A. Dill, *Biochemistry* **29**, *7133* (1990).
7. A. R. Fersht, *Trends Biochem. Sci.* **12**, *310* (1987).
8. S. M. Freier et al., *Biochemistry* **25**, *3214* (1986).
9. D. R. Lesser, M. R. Kurpiewski, L. Jen-Jacobson, *Science* **250**, *776* (1990).
10. B. L. Bass and T. R. Cech, *Nature* **308**, *820* (1984).
11. N. K. Tanner and T. R. Cech, *Biochemistry* **26**, *3330* (1987).
12. I. P. Street, C. R. Armstrong, S. G. Withers, *Biochemistry* **25**, *6021* (1986).
13. J. Steyaert, C. Opsomer, L. Wyns, P. Stanssens, *Biochemistry* **30**, *494* (1991).
14. J. J. He and F. A. Quiocho, *Science* **251**, *1479* (1991).
15. E. N. Baker and R. E. Hubbard, *Prog. Biophys. Molec. Biol.* **44**, *97* (1984).
16. J. A. Thomson, B. A, Shirley, G. R. Grimsley, C. N. Pace, *J. Biol. Chem.* **264**, *11614* (1989).
17. T. Kiefhaber et al., *Biochemistry* **29**, *8250* (1990).
18. C. N. Pace, *Trends Biochem. Sci.* **15**, *14* (1990).
19. K. Sato and F. Egami, *J. Biochem. (Tokyo)* **44**, *753* (1957).
20. C. N. Pace, U. Heinemann, U. Hahn, W. Saenger, Angew. *Chem. Int. Ed.* **30**, *343* (1991).
21. B. A. Shirley, P. Stanssens, J. Steyaert, C. N. Pace, *J. Biol. Chem.* **264**, *11621* (1989).
22. C. N. Pace, B. A. Shirley, J. A. Thomson, in Protein Structure: a practical approach, T. E. Creighton, Ed. (IRL Press, Oxford, 1989), p.311.
23. G. Nemethy, S. J. Leach, H. A. Scheraga, *J. Phys. Chem.* **70**, *998* (1966).
24. B. W. Matthews, H. Nicholson, W. J. Becktel, *Proc. Nat. Acad. Sci. U.S.A.* **84**, *6663* (1987).

25. F. M. Richards, *Ann. Rev. Biophys. Bioeng.* **6**, *151* (1977).
26. E. J. Cohn and J. T. Edsall, Proteins, Amino Acids, and Peptides (Hafner Publishing, New York, 1965) p. 212.
27. S. Damodaran and K. B. Song, *J. Biol. Chem.* **261**, *7220* (1986).
28. J. -L. Fauchere and V. E. Pliska, *Eur. J. Med. Chem.- Chem. Therm.* **18**, *369* (1983).
29. A. Radzicka and R. Wolfenden, *Biochemistry* **27**, *1644* (1988).
30. J. H. Fendler, F. Nome, J. Nagyvary, *J. Mol. Evol.* **6**, *215* (1975).
31. M. Matsumura, W. J. Becktel, B. W. Matthews, *Nature* **334**, *406* (1988).
32. J. T. Kellis, K. Nyberg, A. R. Fersht, *Biochemistry* **28**, *4914* (1989).
33. A. A. Pakula and R. T. Sauer, *Nature* **344**, *363* (1990).
34. K. Yutani, K. Ogasahara, T. Tsujita, Y. Sugino, *Proc. Nat. Acad. Sci. U.S.A.* **84**, *4441* (1987).
35. M. Matsumura et al., *Eur. J. Biochem.* **171**, *715* (1988).
36. D. Shortle, W. E. Stites, A. K. Meeker, *Biochemistry* **29**, *8033* (1990).
37. W. S. Sandberg and T. C. Terwilliger, *Science* **245**, *54* (1989).
38. J. Bello, *J. Theor. Biol.* **67**, *335* (1977).
39. S. Padmanabhan et al., *Nature* **344**, *268* (1990).
40. P. C. Lyu, M. I. Liff, L. A. Marky, N. R. Kallenbach, *Science* **250**, *669* (1990).
41. G. Merutka and E. Stellwagen, *Biochemistry* **29**, *894* (1990).
42. K. T. O'Neil and W. F. Degrado, *Science* **250**, *646* (1990).
43. S. Iida and T. Ooi, *Biochemistry* **8**, *3897* (1969).
44. C. N. Pace, D. L. Laurents, J. A. Thomson, *Biochemistry* **29**, *2564* (1990).
45. L. Pauling and R. B. Corey, *Proc. Nat. Acad. Sci. U.S.A.* **37**, *729* (1954).
46. D. M. Crothers and B. H. Zimm, *J. Mol. Biol.* **9**, *1* (1964).
47. W. Kauzmann, in The Mechanism of Enzyme Action, W. D. McElroy and B. Glass, Eds., The Johns Hopkins Press: Baltimore, MD, 1954.
48. P. L. Privalov, *Adv. Prot. Chem.* **33**, *167* (1979).
49. K. A. Dill, D. O. V. Alonso, K. Hutchinson, *Biochemistry* **28**, *5439* (1989).
50. C. N. Pace, G. R. Grimsley, J. A. Thomson, B. J. Barnett, *J. Biol. Chem.* **263**, *11820* (1988).
51. J. Martinez-Oyanedel, U. Heinemann, W. Saenger, *J. Mol. Biol.* **222**, *335* (1991).
52. R. S. Spolar, J. -H. Ha, M. T. Record, Jr., *Proc. Nat. Acad. Sci. U.S.A.* **86**, *8382* (1989).
53. T. Kiefhaber et al., *Biochemistry* **29**, *8250* (1990).
54. C. Q. Hu and J. M. Sturtevant, unpublished observations.
55. C. N. Pace and D. V. Laurents, *Biochemistry* **28**, *2520* (1989).
56. C. N. Pace, R. Erickson, and D. V. Laurents, unpublished observations.
57. K. A. Dill and D. Shortle, *Ann. Rev. Biochem.* **60**, *795* (1991).
58. C. N. Pace, unpublished observations.
59. G. W. Harris et al., *Biochim. Biophys. Acta* **912**, *348* (1987).
60. J. J. Birktoft and D. M. Blow, *J. Mol. Biol.* **68**, *187* (1972).
61. K. Kuwajima, *Proteins: Struc. Func. Gen.* **6**, *87* (1989).
62. L. G. Presta and G. D. Rose, *Science* **240**, *1632* (1988).
63. B. A. Shirley and D. V. Laurents, *J. Biochem. Biophys. Methods* **20**, *181* (1990).

RECEIVED May 5, 1992

Chapter 3

Protein Structure and Stability Assessment by Circular Dichroism Spectroscopy

Mark C. Manning

School of Pharmacy, Campus Box 297, University of Colorado, Boulder, CO 80309

With the advent of biotechnology and recombinant DNA methodology, there has been a renewed interest in protein folding and protein stability. One of the most powerful techniques available for the study of changes in protein structure is circular dichroism (CD) spectroscopy. Various applications of CD spectroscopy to the study of proteins are described, including analysis of denaturation curves, calculation of secondary structure composition, observation of changes in both secondary and tertiary structure, characterization of folding intermediates, and evaluation of the effects of site-directed mutagenesis. Finally, a case study on fibrolase, a metalloprotease from snake venom, is presented.

Advances in recombinant DNA technology have lead to an increased interest in the structure, function, and stability of proteins. Coupled with a growing understanding of protein folding processes, there has been a renaissance in the use of spectroscopic methods in the study of protein structure. One of the most powerful techniques for investigating the overall secondary and tertiary structure of proteins is circular dichroism (CD) spectroscopy. While CD spectroscopy has some limitations in that it can be used only for proteins in dilute solution and and cannot provide site-specific information (*1*), it is highly sensitive to changes in both the secondary and tertiary structure of proteins (*2,3*). In addition, semiquantitative estimates of the secondary structure composition of a protein can be obtained as well (*4-6*). While a number of reviews on the CD of proteins and other macromolecules are available (*1-12*), little has been published regarding the use of CD spectroscopy to assess protein stability. This article summarizes some of the more common current applications of CD to the study of protein stability and folding problems. Considering the number of examples reported on the use of CD to assess changes in stability and structure, this article will not seek to be exhaustive. However, the work cited herein should provide an adequate description of the current state of the use of CD spectroscopy in the study of proteins. Finally, the article will conclude with a detailed examination of the stability of fibrolase, a metalloprotease isolated from snake venom, in order to more completely demonstrate the capabilities of CD spectroscopy.

0097–6156/93/0516–0033$06.00/0

Denaturation Curves as an Indicator of Protein Stability

One of the primary mechanisms of protein degradation is the loss of globular structure (*13,14*). This process, termed denaturation, leads to a partially or completely unfolded species which usually lacks any of the biological activity of the native protein. A variety of methods have been employed to monitor the denaturation of proteins, including fluorescence, infrared, nuclear magnetic resonance (NMR), and CD spectroscopy. As CD is very sensitive to changes in both secondary and tertiary structure, its application to the study of protein folding and unfolding has been widespread.

Denaturation of a protein can be induced by a number of conditions, including subjecting the protein to extremes in pH, increasing or decreasing the temperature, or by the addition of salts, organic solvents, chelating agents, or denaturants (such as guanidinium hydrochloride, GnHCl, or urea). One advantage of CD spectroscopy is that these changes in solvent composition are essentially transparent to the CD experiment (provided the concentrations are relatively low), allowing examination of most denaturation processes. The primary exceptions are in the case of urea and GnHCl, where concentrations usually exceed 1 M in order to denature most proteins, limiting CD measurements in the far UV region.

Thermal denaturation of proteins is one of the most common determinations of protein stability. By monitoring the change in CD intensity at a particular wavelength, one can determine the degree of unfolding as a function of temperature. For example, it has been shown that there exists a correlation between α helix content of a protein and its molar ellipticity at 222 nm (*15*). A plot of the CD signal as a function of temperature provides the raw data for construction of a denaturation curve. In the vicinity of the melting temperature, T_m, defined as the temperature where half of the protein is unfolded, there is a rapid, cooperative transition from a relatively rigid, folded structure to a flexible, unfolded conformation. This transition may involve denaturation of the entire protein, unfolding of a single domain, or even denaturation of just a small, independent portion. An increase in the T_m value is then taken to indicate an increase in protein stability (*14,16-26*).

Analysis of a thermal melting curve can provide thermodynamic information regarding the unfolding process (*14,19-26*). From the relative fractions of folded and unfolded protein at a given temperature near T_m, one can determine an equilibrium constant and, thus, a free energy of unfolding (ΔG_u) at that temperature. A plot of ΔG_u versus temperature provides an estimate (from the slope) of the entropy of denaturation. The enthalpy of unfolding can then be calculated from the Gibbs-Helmholz equation.

Lowering the temperature can also cause a protein to unfold in a process termed 'cold' denaturation (*27-30*). This arises primarily from the thermodynamics of hydrophobic interactions which lead to globular protein structures (*14,24*). A number of examples of this phenomenon have been described (*27-29*) and have been detected using CD spectroscopy. Frequently, because many proteins only cold denature below the freezing point of water, nonaqueous solvents and solvent mixtures must be employed to maintain fluidity of the solution (*27,28*). In the case of staphylococcal nuclease (*27*), the unfolding was measured in an aqueous solution containing 2 M urea. Spectra taken at -7° C indicated a greater degree of unfolding than for heat-denatured material at 55° C.

Another common indication of protein stability is the concentration of either urea or GnHCl required to unfold half of the protein available. This concentration, given the symbol $[D]_{1/2}$, is analogous to the T_m value from thermal denaturation curves. Increase or decrease in $[D]_{1/2}$ is presumed to indicate a corresponding increase or decrease in protein stability, respectively. Analysis of these curves can also provide thermodynamic information (*20,21,26,31*). As these experiments can be done at any

temperature, they are more useful in that they can provide information regarding stability at or near room temperature.

Denaturation resulting from either heating the sample or the addition of a denaturant can be followed observing changes in either the near or far UV region of the spectrum. In the near UV, the signals arise from aromatic groups and, to a lesser extent, disulfides. Typically, changes in this region are thought to reflect an alteration of the tertiary structure (*8,9*). Conversely, changes in the far UV represent changes in secondary structure, although contributions from side chains and prosthetic groups may not be negligible at these wavelengths (*1*). Not only do these two regions possess information on different aspects of the globular structure of a protein, they also differ in the intensities of the signals. Typically, the signals in the near UV are one to three orders of magnitude weaker than those in the far UV. As a result, the pathlengths and concentrations used for work in the near UV must be significantly greater. Usually, pathlengths of 5 to 10 mm must be used with protein concentrations in the range of 1 to 5 mg/ml. The sample volumes are between 600 μl and 3000 μl. In order to obtain reasonable far UV CD spectra, the demands are more modest, with pathlengths ranging from 0.1 mm to 1 mm and concentrations typically between 0.05 and 0.50 mg/ml with sample volumes of 100 μl to 1.5 μl. It should be noted that at very low protein concentrations, adsorption to the cell may cause diminished CD intensity (*32*). Still, coupled with its high sensitivity and its nondestructive nature, its minimal sample requirements have been among the primary reasons for CD spectroscopy being the method of choice in the study of protein structure and stability.

Protein Stability as Determined by Changes in Secondary Structure Composition

Often, it is necessary to analyze the entire spectrum rather than just monitor changes at a single wavelength. This is particularly true if changes in secondary structure composition occur. Certainly, CD spectra can be qualitatively assessed for differences upon changes in conditions. However, numerous algorithms exist to deconvolute far UV CD spectra in order to determine quantitatively the relative amounts of α helix, β sheet, β turns, and other structures (*4-6,33*). While early methods only used CD data down to 200 nm, newer approaches (*4,5,33*) have shown that improved accuracy can be obtained if data is available to wavelengths as short as 176 nm, near the limit of commercial instrumentation. While these methods make numerous assumptions (*1*), they have been shown to be reasonably accurate for a wide range of proteins. Therefore, whether qualitative or quantitative analysis of the changes in secondary structure is performed, CD is an extremely sensitive tool for monitoring alterations in globular structure.

Secondary structure has been found to affected not only by changes in solution conditions (i.e, pH, salt concentrations, temperature, etc.), but also by mutagenesis of specific amino acid residues, the concentration of the protein, and by the introduction of nonaqueous solvents, reducing agents, and chelators. Examples of the use of CD to examine each of these effects can be found in the literature.

Interaction with lipids, surfactants, or membranes can induce the formation of secondary structure (*34-38*), particularly in peptides, where they may exist as highly flexible species in solution (*39*). Provided that the lipids form micelles which are relatively small and uniform, CD can be used to measure secondary structure in the presence and absence of lipids. Glucagon is a peptide hormone found in the gastrointestinal tract. In dilute solution, glucagon possesses only a small amount of ordered structure. However, in the presence of phospholipids, there is a significant increase in α helix content from 15% to 30-35% (*40*). Similar increases have been observed for calcitonin gene-related peptide. Both CGRP (*41,42*) and $CGRP^{8-37}$ (*42*, M.C. Manning, unpublished results) display marked increases in α helix content upon addition of anionic detergents such as sodium dodecyl sulfate. Fluorinated alcohols,

such as trifluoroethanol (TFE) and hexafluoroisopropanol (HFIP), also stabilize α helices (*41-48*). Stabilization by these solvents and by lipids and detergents (*34-38,40*) appears to be a general phenomenon for peptides and can be used to assess secondary structure stability (*44,46*), although there may be some difficulties in obtaining reliable estimates of secondary structure composition for small peptides in nonaqueous solvents such as TFE and HFIP (*45*).

Protein Stability as Determined by Changes in Tertiary Structure

It is possible to assess protein stability by observing changes in the near UV CD spectrum as well. As stated above, these signals arise from aromatic residues and disulfides (*8,9*). Decrease in CD intensity in the near UV is usually associated with a loss of tertiary structure or a localized unfolding of the protein in the vicinity of a particular aromatic group. Also, through the use of site-directed mutagenesis, it is possible to systematically replace individual aromatic groups in order to assess their contributions to the near UV CD spectrum. Examples of each of these types of studies are given below.

Acid treatment of ovalbumin does not change the far UV CD, but it does alter the near UV CD dramatically (*49*). This indicates that the mobility of the aromatic groups (and most likely, the mobility of all of the side chains) is increased. At lower pH values, the protein displays an increased rate of unfolding in urea and behaves like a 'molten globule' type of species (*50,51*).

The denaturation of insulin and proinsulin have been examined in detail by Brems et al. (*52*). They have shown that GnHCl-induced denaturation is a reversible, cooperative process for insulin, whether detected by either far or near UV CD spectroscopy. The ΔG_u for the unfolding of insulin was determined to be 4.5 ± 0.5 kcal/mole. The unfolding of proinsulin, on the other hand, appears to be multiphasic as monitored by far UV CD. However, when the contributions of the C-terminal connecting peptide are considered, the denaturation of proinsulin is nearly identical to that of insulin itself. Apparently, the connecting peptide acts as an individual folding unit, complicating the analysis of the denaturation of proinsulin.

Characterization of Folding Intermediates by CD Spectroscopy

A major focus in biophysics and biochemistry is to understand the protein folding process, that is, the mechanism by which a newly synthesized protein adopts its native conformation. The kinetics of protein folding have been measured using a number of techniques, including CD. As proteins are usually completely folded within seconds, this requires either fast detection techniques, such as stopped flow methodology or means to trap folding intermediates. Stopped flow CD has been used although it is limited to single wavelength detection (*53*). One promising new approach is the development of ultrafast CD equipment (*54,55*). With such an apparatus, CD signals can be measured in the microsecond time frame, fast enough to observe all but the fastest steps of protein folding. It should be noted that the major problem with measuring protein folding kinetics is not instrumental, but resides in the difficulty of initiating the folding process. Mixing solutions in order to change solute concentrations, a common method of beginning the folding process, cannot be done faster than the millisecond time scale. Only if protein folding (or unfolding) can be initiated by changing some other condition, such as temperature or pressure, could microsecond CD measurements provide information on folding kinetics.

Refolding of acid- and base-denatured proteins has lead to the observation of compact, non-native structures termed 'molten globules' (*49-51,56,57*). These structures possess a significant amount of secondary structure, but little, if any, tertiary structure. As CD can detect the presence of both types of higher order structure in proteins, it is ideal for characterization of 'molten globule' structures.

Acid-treated β-lactamase is known to lose both secondary and tertiary structure as determined by CD measurements (*56*). However, upon addition of 0.5 M KCl, it refolds into a compact structure (termed A-type) which possesses some secondary structure, but still lacks any near UV CD signals (see Table I). Similar types of structures (B-type) can be obtained following base treatment. Lack of a near UV spectrum along with a native-like far UV CD indicates that the new conformation has some globular structure, but does not have any organized tertiary structure, as in the native state. This represents a utilization of CD spectroscopy to characterize globular, non-native states of proteins. These states are proposed to be similar to early folding intermediates. Similar behavior has been seen for ovalbumin. Lowering the pH produces little change in the far UV CD spectrum, but the near UV displays loss of intensity, suggesting increased flexibility in the side chains (*49*).

Table I. Circular Dichroism of Various Conformations of β-Lactamase (adapted from reference 56)

Conformation	pH	Ionic Strength[a]	FUV CD[b]	NUV CD[c]	α Helix %
Native	7	I = 0.05	+	+	26
	7	0.5 M KCl	+	+	26
Unfolded	2	I = 0.05	-	-	4
A-Form	2	0.5 M KCl	+	-	25
Unfolded	12	I = 0.05	-	-	6
B-Form	12	0.5 M KCl	+	-	20

a- Ionic strength of buffer or of added KCl
b- Observation (+) or lack (-) of significant CD intensity in the far UV
c- Observation (+) or lack (-) of significant CD intensity in the near UV

The kinetics of protein folding have been measured using conventional CD instrumentation. For example, upon dilution of bovine growth hormone (bGH) from 6.0 M GnHCl to 2.2 M, the protein spontaneously refolds (*58*). Following the molar ellipticity at 222 nm (indicating changes in secondary structure) and the absorbance at 290 nm (indicating changes in tertiary structure), it has been shown that the formation of secondary structure is more rapid than formation of tertiary structure, consistent with the framework model of protein folding (*59*). Both recombinant and pituitary-derived bGH folded at the same rate, suggesting that differences in the N-terminal sequence do not affect folding of bGH.

Effects of Amino Acid Replacements on Protein Stability and Structure

Site-directed mutagenesis is a powerful tool for altering the primary structure of a protein in a controlled fashion. The effect of such changes on protein stability has been widely studied (*60,61*). However, it is also of interest to determine the effects of these point mutations on the folding of the protein.

Bovine growth hormone (bGH) is a 191 amino acid residue protein which adopts a four-helix bundle motif. Changes in the sequence within this bundle can greatly affect the conformational stability of the protein, as well as its propensity to aggregate. Replacement of lysine-112 with leucine (Leu) produces a species which is much more prone to aggregate than the wild type (*62*). The aggregate of the mutant displays reduced CD intensity in the far UV.

In our laboratory, we are investigating the effects of replacement of both Tyr and Phe in bovine pancreatic trypsin inhibitor (BPTI) (D.P. Goldenberg, M.C. Manning, unpublished results). Replacement of phenylalanine (Phe) residues, which are not expected to contribute strongly to the near UV CD (relative to tyrosine), can have a dramatic effect. Figure 1 shows the near UV CD spectra of wild type BPTI and the nutant where Phe-22 has been replaced by leucine. The result is a CD signal which has been significantly attenuated. Theoretical studies suggest that this may be a result of modifying the interactions of other aromatic groups (such as tyrosine) in close proximity to that particular group.

Effects of Aggregation on Protein Stability and Structure

While certain proteins are only active after the correct assembly of the various subunits, other proteins exhibit aggregation phenomena which may diminish their biological function. In fact, aggregation may proceed precipitation of the protein from solution (*13, 63*).

Insulin is well known to exist in a number of different aggregation states, ranging from monomeric to hexameric. Recently, a study of the far UV CD of various insulin forms was reported using an instrument which was capable of obtaining CD spectra into the vacuum UV (*64*). Comparison of the CD spectra of zinc free preparations, the 2-Zn form, and two engineered monomeric insulin showed striking differences in the far UV. The structure of the C-terminal portion of the B-chain undergoes conformational change upon association, especially in the formation of hexameric insulin. In the insulin which is engineered to remain monomeric, this section has been modified or deleted. Secondary structure analysis indicated that monomeric insulins possess 10-15 % more β sheet than hexameric insulin.

Interferon-γ (IFN-γ) is a pharmaceutically important protein whose stability is particularly sensitive to pH (*65*). Both rapid denaturation and aggregation have been observed at both acidic and basic pH values. Using light scattering methods, Mulkerrin and Wetzel were able to monitor the rate of aggregation at different temperatures. Onset of aggregation was always delayed by a period of seconds to minutes. Using CD spectroscopy, they were able to demonstrate that the lag time corresponded to an initial unfolding of IFN-γ. This unfolded species then went on the associate and form soluble aggregates.

Unfolded intermediates have been observed for bovine antithrombin (*66*). Biphasic unfolding was observed in the presence of GnHCl, with $D_{1/2}$ values of 0.8 M and 2.8 M determined by monitoring the signal at 220 nm. The second $D_{1/2}$ value is typical for complete unfolding of a protein. Renaturation experiments demonstrated that this process was reversible. However, the first step was not reversible. It appears to arise from exposure of certain hydrophobic groups, as both the near UV and far UV CD spectra are affected. This partial unfolding leads to rapid aggregation which cannot be reversed by removal of the GnHCl.

During unfolding of bGH, a self-associated, partially unfolded intermediate is formed (*67,68*). Characterization of this intermeidate has been accomplished using hydrodymanic measurements, gel permeation chromatography, as well as CD spectroscopy. The intermediate displays increased CD intensity in the near UV relative to the native state. A significant concentration dependence of the near UV CD is observed, consistent with it associating to form soluble aggregates, particularly dimers.

Effects of Metal Binding on Protein Stability and Structure

Interactions with metal ions, whether intrinsic or extrinsic, can alter the stability of a protein. Loss of an intrinsic metal atom often leads to localized unfolding of the protein and significant, if not total, loss of biological activity. Removal of intrinsic

metals can be accomplished by the addition of chelating agents, by increase in the temperature, or by lowering the pH. These effects have been studied in detail for fibrolase (see below), a zinc-containing metalloenzyme (*69,70*).

Calcium binding is an important aspect of the physiological function of many proteins. Loss of calcium may lead to a structure which cannot function properly (*71,72*). For example, apo-actin, which has been depleted of calcium, adopts a structure which is nearly indistinguishable from that of heat-denatured actin as determined by CD (*71*). Upon removal of calcium, the α helix content drops from 33 % to 19 %. Further loss of structure was not observed, even upon heating to 95° C. Similar behavior has been noted for fibrolase (see below).

Conversely, the native structure of a protein can be induced upon addition of calcium. Bayley and co-workers have shown that calmodulin adopts its native conformation only in the presence of calcium (*72*). Lack of calcium produces a species which appears to be partially unfolded as determined by far UV CD spectroscopy.

Replacement of intrinsic metals with other metals which can be used as spectroscopic probes is a technique which has been widely employed in biophysical investigations, especially for fluorescence and NMR studies. Cobalt, because it possesses absorption bands in the visble region of the spectrum and is known to exhibit CD sgnals when placed in chiral environments (*73*), has been used to probe structural changes in insulin (*74,75*). The T -> R transition in zinc-containing hexameric insulin is difficult to detect spectroscopically. However, replacement of zinc with cobalt provides an additional spectroscopic probe. Induced CD from the cobalt occupying a discrete chiral environment produces visible CD bands which correlate to the T and R states of insulin (*74*).

Stability Study of Fibrolase, a Metalloprotease from Snake Venom

Effects of pH. From the discussion above, it is clear that CD can be employed in a variety of ways to assess protein stability and structure. However, it might be instructive to provide a detailed example of the use of CD in a comprehensive stability study. The stability of fibrolase, a small (203 amino acids) proteolytic enzyme isolated from snake venom, was examined. Fibrolase contains one intrinsic zinc atom per molecule of fibrolase (*76*) and is a direct fibrinolytic agent (*77,78*).

The total protein concentration in solution was determined over the pH range of 1 to 10 (Figure 2). Protein concentrations were nearly unchanged in the neutral to alkaline pH range (6.5 to 10.0). However, as the pH became acidic, the total amount of fibrolase in solution decreased with an apparent minimum at pH 5.0. At even lower pH values, the overall concentration increased to substantial levels (30 to 90% of initial protein at pH 1 to 4).

The effect of pH on specific proteolytic activity of fibrolase was evaluated over the pH range 1 to 10 using azocasein and insulin B-chain substrates (Figure 3). At neutral to alkaline pH values (6.5 to 9.0), no significant change in proteolytic activity of fibrolase was observed. At pH 10.0, the data suggested a slight loss of activity. This decrease was greater for the azocasein substrate relative to the insulin B-chain substrate. As the pH was lowered, the activity of fibrolase decreased. At pH 5.0, approximately 85% of the initial activity remained. However, at pH 4.0, only 20% of the initial activity was detected. At pH 1.0 and 2.5, azocaseinolytic activity was less than 10%. Activity versus insulin B-chain was also low at these pH values, but showed high variability in the assay results.

Gel electrophoresis results (data not shown) correlated well with activity and concentration results. SDS-PAGE gels of reduced and non-reduced samples showed no change in the location of the fibrolase band over the entire pH range 1 to 10. Band intensity corresponded qualitatively to protein concentration results, as expected. Native gels of samples at pH 5 to 10 showed no variation in migration of the fibrolase

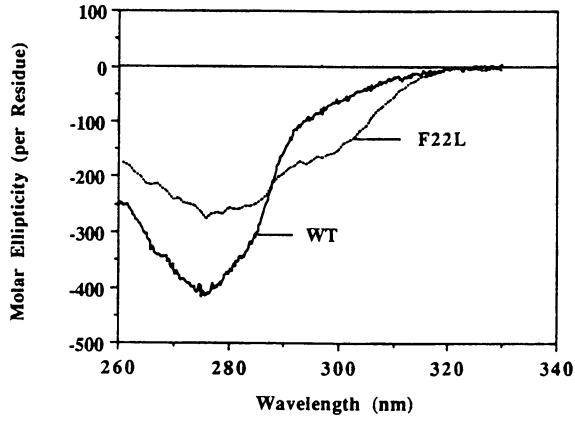


Figure 1. Near UV CD spectrum of wild type bovine pancreatic trypsin inhibitor (WT) and its Phe-22 -> Leu (F22L) mutant.

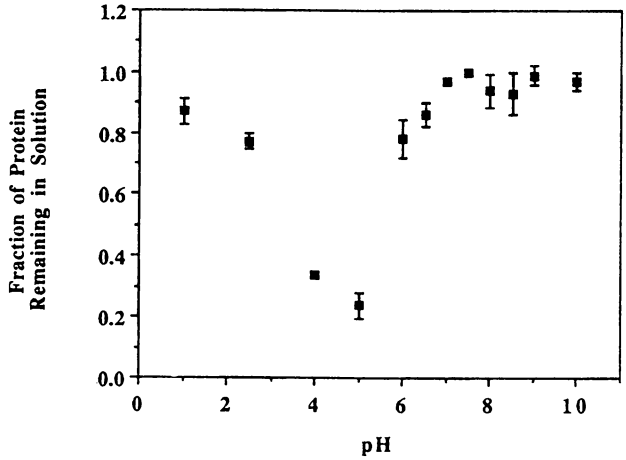


Figure 2. Protein concentration (as percentage of initial concentration) as a function of pH. The error bars represent ± one standard deviation.

band. However, at more acidic pH values, no fibrolase band was seen. Instead, staining was observed as a streak or as multiple bands with increased mobility compared to fibrolase. These results were consistent with the lack of proteolytic activity at acidic pH.

Consideration of specific activity data together with the protein concentration data over the pH range 1 to 10 indicates several interesting effects as a result of changing pH. In the neutral to alkaline pH range, both activity and total protein in solution were not significantly changed. As the pH was lowered, protein concentration decreased to an apparent minimum at pH 5.0, yet the protein remaining in solution had a specific activity of about 0.85. As the pH was lowered even further, most of the proteolytic activity of the protein was lost, yet substantial amounts of protein remained in solution.

The apparent minimum solubility of fibrolase at pH 5.0 does not correspond to the isoelectric point of 6.7 (*69,70*). Minimum protein solubility would be predicted to occur near the isoelectric pH, if the native state of the protein underwent aggregation (*13*). This solubility behavior suggests the formation of an unfolded state which is more prone to aggregate than the native conformation. It should be noted that these studies were not designed as solubility measurements per se (i.e., an excess of protein was not present, thus equilibrium might not have been attained), therefore, the maximum solubility can not be deduced from the data.

Far UV CD spectra of fibrolase displayed essentially no change in band shape or position over the pH range 5 to 9 (Figure 4). However, some variations in intensity were observed. While some differences in intensity were observed in the far UV CD spectra of fibrolase with change in pH (from 5 to 9), the band positions and shapes were invariant. Assuming a 10% error in determination of the protein concentrations, most of the variations in intensity fall within experimental error. Otherwise, these differences could be accounted for by small increases (one or two turns) in the α helix content of the protein. Together, it appears that small variation or error in protein concentration and/or CD measurements is sufficient to account for any intensity variations of the far UV CD spectra within the pH 5 to 9 range. At pH 2 and 3, significant differences were found compared to the pH range 5 to 9 (Figure 5). The spectral changes at pH 2 and 3 were consistent with loss of α-helical structure and were similar to changes observed upon thermal denaturation (see below).

Effects of Temperature. At 37° C, rapid loss of proteolytic activity was observed. Greater than 10% of the activity was lost in less than one day, and essentially no activity against either substrate remained after 10 days. SDS-PAGE gels of non-reduced samples showed loss in intensity of the fibrolase band and lower molecular weight bands. Similar results were observed for reduced samples. In contrast to the activity and gel results, total protein remained unchanged in solution over 16 days.

The observed changes could be due to chemical or physical degradation mechanisms (*13*). Fibrolase contains potential sites for deamidation, oxidation, and hydrolysis, although in order to account for the observed loss, the rates of these reactions would have to be greater than predicted for the experimental conditions (*13*). One other mechanism specific to proteases which could account for the observed instability of fibrolase at 37° C would be autoproteolysis. Activity results were not inconsistent with an autoproteolytic mechanism assuming that degradation products would be inactive. Electrophoresis did not detect smaller protein species, however, autoproteolysis at multiple cleavage sites might produce low molecular weight fragments which would not be visible on the gels.

Physical instability such as adsorption, aggregation, denaturation, etc., could account for the observed loss of activity at 37° C. Significant physical changes commonly result in precipitation of protein, which was not observed. In addition, significant conformational changes at 37° C were not detected by CD, although the CD

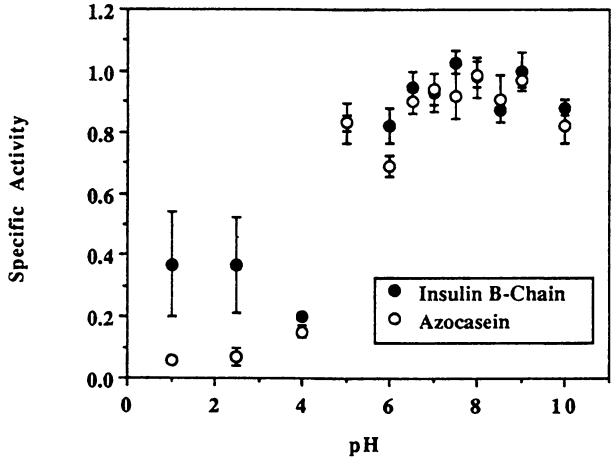


Figure 3. Specific proteolytic activity of fibrolase (expressed as fraction of initial activity) as a function of pH for both azocasein and insulin B-chain as substrates. The error bars represent ± one standard deviation.

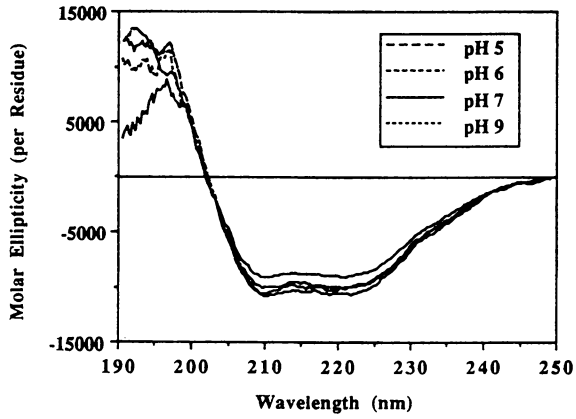


Figure 4. Far UV CD spectra of fibrolase at pH 5, pH 7, pH 8, and pH 9. The solutions contained 10 mM phosphate buffer and fibrolase concentrations ranged from 100 to 400 μg/ml.

samples were maintained at 37° C for a maximum time of only about 10 minutes. Visible absorbance did increase somewhat upon heating to 37° C, indicating the formation of soluble aggregates of fibrolase.

The far UV CD spectrum of fibrolase at pH 8 is typical of a protein possessing a significant amount of α-helical structure (Figure 4). Secondary structure analysis indicates that approximately 25% of the residues in fibrolase are in an α-helical conformation (Figure 6). The globular structure appears to be stable over the temperature range of 15° to 40° C. Above 40° C, some of the secondary structure of fibrolase is rapidly lost in a highly cooperative, irreversible transition with a T_m of ~50° C (Figures 6, 7, and 8). At 70° C, only 8-10 % of the α-helix content remains. Little variation between lots was observed in the T_m (Figure 8). Changes in the CD spectrum were consistent with loss of only α-helical structure, as shown by the loss of intensity at 222 nm and the appearance of a negative band near 200 nm (Figure 7). Presence of an isodichroic point at 207 nm suggested that the thermal denaturation of fibrolase could be well approximated by a two-state model, meaning no evidence for partially unfolded intermediates was found. However, it is important to note that even at 70° C, a significant portion of fibrolase remains folded (as indicated by the residual ellipticity at 222 nm) and fibrolase can only be completely denatured by 6 M GnHCl (results not shown).

Varying pH also affected the thermal stability of fibrolase. At pH 5, the T_m dropped to 43° C, compared to 50° C at pH 8 (Figure 9). In addition, at pH 5, the cooperativity of the transition was less. Some destabilization of fibrolase with decreasing pH is indicated by the lower T_m observed at pH 5 compared to pH 8 (Figure 9). Destabilization resulting in the lower T_m value would not necessarily be apparent in CD spectra taken at room temperature. The destabilization could be a result of unfavorable changes in the electrostatic network of the protein (*65*).

Spectral changes induced by elevated temperatures are similar to those resulting from acid treatment (Figure 10). At 48° C, where the protein is 30-40 % unfolded, the far UV CD spectrum of fibrolase at pH 8 greatly resembles that of fibrolase at pH 2 and 27° C.

Effects of Zinc Removal. The presence of an essential zinc ion (*64*) is a crucial consideration in evaluating the stability of fibrolase. Loss of structure and activity upon exposure to extremes of pH and temperature can be rationalized on the basis of removal of the zinc atom. Denaturation of fibrolase under these conditions leads to a marked tendency to aggregate and precipitate (see below). Various studies were performed in order to examine the strength of zinc binding, the effect of zinc on the structure and activity and secondary structure, and to determine the metal binding site.

The effect of EDTA on the activity and conformation of fibrolase was of interest due to the classification of fibrolase as a zinc metalloprotein (*69,75*). Other investigators had previously observed inhibition of azocaseinolytic activity of fibrolase by EDTA, and, in fact, these studies contributed to the characterization of fibrolase as a metalloprotein (*76,77*). It was of interest to investigate what concentrations of EDTA affected fibrolase activity, and what other changes (e.g., structural) were caused by EDTA.

The effect of EDTA on proteolytic activity of fibrolase was dependent on the EDTA concentration added (see Table II). Partial inhibition was observed for EDTA concentrations as low as 0.1 mM. Complete inhibition occurred at 0.9 mM EDTA for activity versus the insulin B-chain. Further increases in EDTA concentration had no effect. Activity for the insulin B-chain was inhibited to a greater extent than was the azocaseinolytic activity. About half of the azocaseinolytic activity remained in the presence of 1 mM EDTA. A higher EDTA concentration of 5 mM was required for essentially complete (98%) inhibition of azocaseinolytic activity.

The reason for the difference in effective inhibitory concentrations of EDTA might be related to the difference in molecular size of the two substrates. The

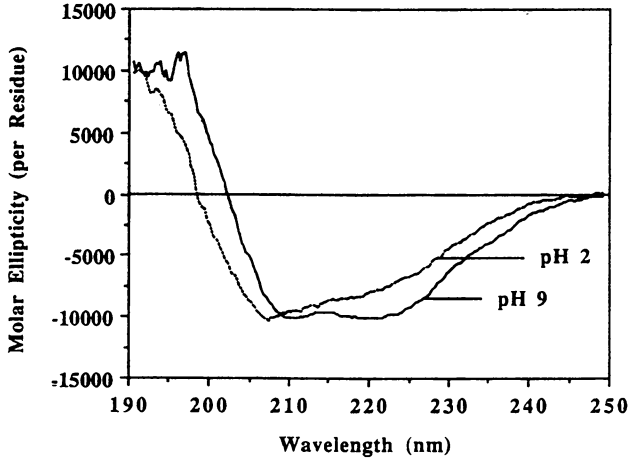


Figure 5. Far UV CD spectra of fibrolase at pH 2 and pH 9.

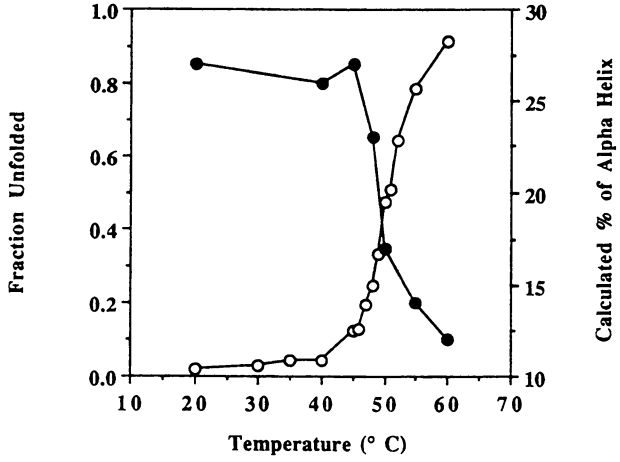


Figure 6. Plot of α helix content (●) and degree of unfolding (○) as a function of temperature.

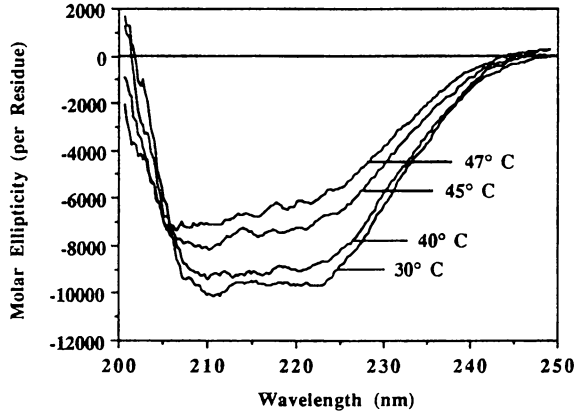


Figure 7. Effects of temperature on the far UV CD spectrum of fibrolase (pH 8). The pathlength was 1.0 mm.

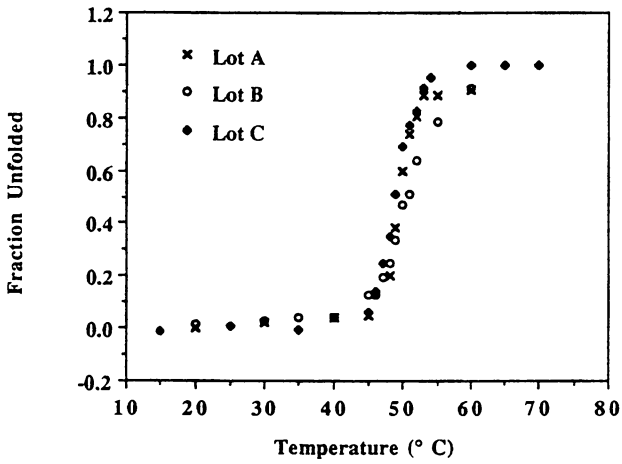


Figure 8. Variation in T_m for three lots of fibrolase as monitored by CD spectroscopy at 222 nm.

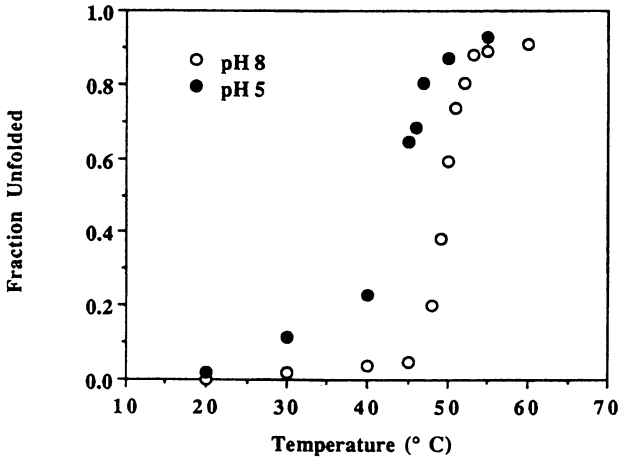


Figure 9. Thermal denaturation curves for fibrolase at pH 5 and pH 8, as determined by molar ellipticity at 222 nm.

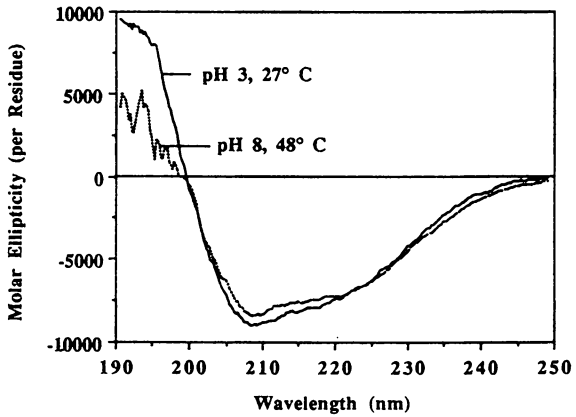


Figure 10. Far UV CD spectra of fibrolase at pH 3 and 27° C and at pH 8 and 48° C.

molecular weight of insulin B-chain is about 3800, while that of azocasein is about 25,000. Because of the size difference, the substrates might be subject to different constraints around the active site of fibrolase. The data indicated that the active site constraints, or importance of metal binding, might be greater for the insulin B-chain, since this activity was inhibited at lower EDTA concentrations. It is interesting to note that these observations are in contrast to activity results for previous studies which suggested that azocaseinolytic activity was lost more readily than was activity for the insulin B-chain at pH extremes and after extended storage at room temperature. These results suggested that different effects might occur in the protein upon EDTA addition. The time course of EDTA inhibition was rapid (see Table III). Inhibition was complete at the first measurable time point for 6 and 20 mM EDTA, and remained complete over one hour. The time course of inhibition was not followed for other EDTA concentrations.

Table II. Proteolytic Activity of Fibrolase in the Presence of EDTA

[EDTA] (mM)	% Activity Remaining Relative to Control[a] vs. Insulin B-Chain	vs. Azocasein
0.1	51.4 (9.3)	55.2 (12.0)
0.2	51.8 (15.8)	50.1 (12.5)
0.5	32.7 (6.9)	49.6 (14.5)
0.9	0	nd
1.0	nd	42.1 (12.5)
2.0	nd	22.6 (3.8)
5.0	nd	2.3 (0.1)
6.0	0	nd
20.0	0	nd

a- No EDTA was added to the control samples. Initial fibrolase concentration was ~100 mg/ml. Numbers reported are the mean of at least three determinations, with standard deviations given in parentheses.
nd- Not determined.

While submillimolar concentrations of EDTA (< 1 mM) resulted in partial loss of activity versus insulin B-chain and azocasein, changes in CD spectra were also noted at these EDTA concentrations (Figure 11). These amounts of EDTA represent only a small molar excess of EDTA compared to fibrolase. At 0.1 mM EDTA, the molar excess of EDTA over fibrolase was only about 25 for studies examining the effect on activity and 10-15 in the case of the CD work. The low concentration and molar excess of EDTA effective in inducing activity and structural changes in fibrolase indicated that the zinc ion might be bound quite loosely to the protein.

CD results (ellipticity at 222 nm) were correlated with the observed activity changes. At submillimolar EDTA concentrations, structural changes were evident by CD as indicated by decreasing elliptical intensity at 222 nm (Figure 11), and partial inhibition of activity was also observed (Table II). At EDTA concentrations greater than or equal to 0.9 mM, structural changes observed by CD appeared to be complete, and correlated well with inhibition of activity for the insulin B-chain. Changes similar to those observed upon addition of EDTA were noted upon addition of DTT (Figure 12). While it is a well known reducing agent, it is also a good chelating agent (79), and it removed the zinc from fibrolase rapidly and easily. The resulting structure is similar to that found upon EDTA addition.

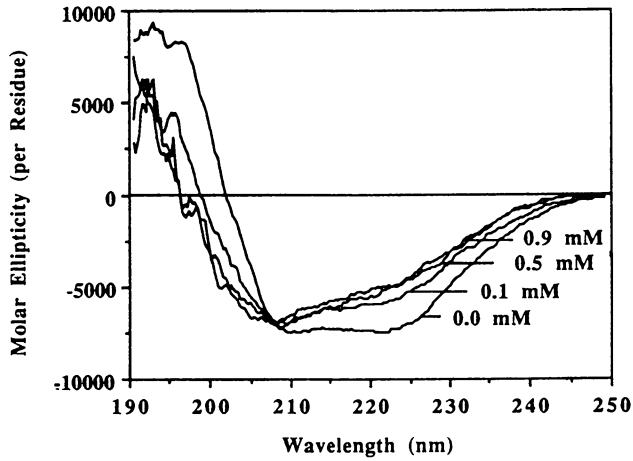


Figure 11. Far UV CD spectra of fibrolase in 10 mM phosphate buffer containing 0.0 mM, 0.1mM, 0.5mM, and 0.9mM.

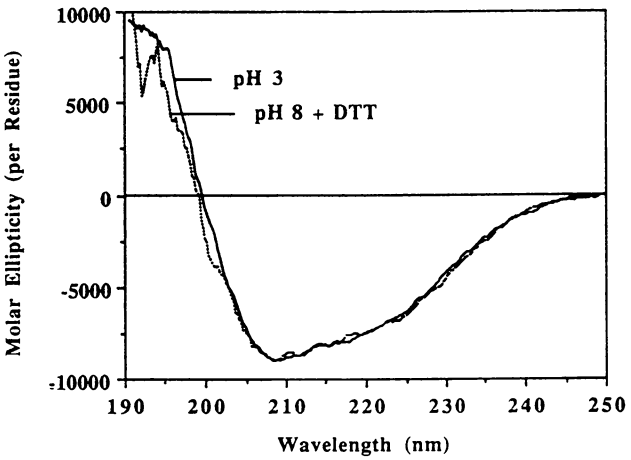


Figure 12. Far UV CD spectrum of fibrolase at pH 3 and of fibrolase at pH 8 with 10 mM DTT added.

Table III. Time Course of EDTA Inhibition of Fibrolase Proteolytic Activity[a]

Time (min) EDTA	PEAK AREA (x 10^{-2})[b] 0.0 mM EDTA	6 mM EDTA	20 mM
0	8.16	8.48	8.69
5	4.78	-----	8.84
10	4.07	-----	8.73
15	-----	8.13	-----
20	2.89	-----	8.52
30	2.48	9.16	8.55
45	2.06	-----	7.82
60	1.49	8.00	8.63

a- Proteolytic activity measured vs. insulin B-chain.
b- Peak area of insulin B-chain substrate. Decreasing peak area indicates proteolysis by fibrolase. Peak areas are reported for one determination. Typical assay variability was ~5%.

The structural changes observed by CD upon EDTA addition to fibrolase were corroborated by native gel electrophoresis. Fibrolase samples containing EDTA (> 1 mM) consistently migrated further on native gels relative to fibrolase without EDTA present. These observed changes in the activity and conformation of fibrolase with EDTA supported involvement of zinc in the active site of the protein. The data showed that not only is the presence of the metal ion critical for proteolytic activity, it is also important to the structure of the protein. The metal ion may bind to several remote sites of the protein, and when the ion is removed by EDTA treatment, these residues might not retain the driving force for association, resulting in extensive structural perturbation.

Overall, EDTA treatment, acidic pH, and elevated temperatures produce species which are similar in secondary structure. In each of these, a significant portion of the secondary structure remains intact, but most of the helical structure is lost. In addition, the loss of secondary structure seems to correlate with the loss of zinc and enzymatic activity. It is believed that each of these conditions (elevated temperature, acidic pH, addition of chelators) leads to loss of zinc with concomittant unfolding of the metal binding site. The remainder of the fibrolase molecule appears to remain folded, as there is still signficant intenisty at 222 nm and the near UV CD spectrum does not change. Therefore, any efforts aimed at ensuring the stability of fibrolase with respect to denaturation should focus on avoiding conditions which would lead to loss of the intrinsic zinc atom.

Conclusions

The ability to accurately evaluate the stability of proteins is an important aspect of protein engineering. Certainly, specific descriptions of protein stability need to be defined, as proteins can decompose via a number of mechanisms. However, if the resistance of a protein to denaturation is taken to indicate stability, then CD spectroscopy should be the primary tool for the evaluation of protein stability. Changes in both overall and local structure can be detected and characterized. In

addition, the effects of atering the sequence and aggregation state of a protein can be investigated with this technique.

Acknowledgements

The support of the University of Kansas General Research Fund and Marion Merrell Dow is gratefully acknowledged. The work on fibrolase was done in collaboration with James Mitchell and Brenda Schiulteis (Marion Merrell Dow Research Institute), David Vander Velde and Chris Smith (University of Kansas) and Denise Pretzer (Merck Sharp and Dohme). Mutants of BPTI were provided by David Goldenberg (University of Utah).

Literature Cited

1. Manning, M.C. *J. Pharm. Biomed. Anal.* **1989**, *7*, 1103-1119.
2. Woody, R.W. *J. Polymer Sci.: Macromol. Rev.* **1977**, *12*, 181-321.
3. Woody, R.W. in *The Peptides: Analysis, Synthesis, Biology*; Hruby, V.J., Ed.; Academic Press: Orlando, Florida, 1985, Volume 7; pp. 16-114.
4. Johnson, W.C., Jr. *Ann. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 145-166.
5. Johnson, W.C. Jr. *Proteins: Structure, Function, Genetics* **1990**, *7*, 205-214.
6. Yang, J.T.; Wu, C.-S. C.; Martinez, H. *Meth. Enzymol.* **1986**, *130*, 208-269.
7. Bayley, P.M. *Prog. Biophys. Mol. Biol.* **1973**, *27*, 3-76.
8. Strickland, E.H. *CRC Crit. Rev. Biochem.* **1974**, *2*, 113-174.
9. Kahn, P.C. *Methods Enzymol.* **1978**, *61*, 339-378.
10. Schellman, J.A. *Chem. Rev.* **1975**, *75*, 323-331.
11. Sears, D.W.; Beychok, S. in *Physical Principles and Techniques of Protein Chemistry, Part C*; Leach, S.J., Ed., 1973, pp. 445-593.
12. Bannister, W.H.; Bannister, J.V. *Int. J. Biochem.* **1974**, *5*, 673-677.
13. Manning, M.C.; Patel, K.; Borchardt, R.T. *Pharmaceutical Research* **1989**, *6*, 903-918.
14. Tanford, C. *Adv. Protein Chem.* **1968**, *23*, 121-282.
15. Chen, Y.-H.; Yang, J.T.; Martinez, H.M. *Biochemistry* **1972**, *11*, 4120-4131.
16. Hecht, M.H.; Sturtevant, J.M.; Sauer, R.T. *Proteins: Structure, Function, Genetics* **1986**, *1*, 43-46.
17. Hecht, M.H.; Hehir, K.M.; Nelson, H.C.M.; Sturtevant, J.M.; Sauer, R.T. *J. Cell. Biochem.* **1985**, *29*, 217-224.
18. Nicholson, H.; Becktel, W.J.; Matthews, B.W. *Nature* **1988**, *336*, 651-656.
19. Schellman, J.A. *Ann. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 115-137.
20. Becktel, W.J.; Schellman, J.A. *Biopolymers* **1987**, *26*, 1859-1877.
21. Goodenough, P.W.; Jenkins, J.A. *Biochem. Soc. Trans.* **1991**, *19*, 655-662.
22. Privalov, P.L. *Ann. Rev. Biophys. Biophys. Chem.* **1978**, *18*, 47-69.
23. Privalov, P.L. *Curr. Trends Life Sci.* **1984**, *12*, 17-21.
24. Privalov, P.L. and Gill, S.J. *Adv. Protein Chem.* **1988**, *39*, 191-238.
25. Tombs, M.P. *J. Appl. Biochem.* **1985**, *7*, 3-24.
26. Shirley, B.A. in *Pharmaceutical Biotechnology*; Volume 2, Ahern, T.J., Manning, M.C., Ed.; Plenum Press, New York, 1992, pp. 000-000.
27. Griko, Y.V.; Privalov, P.L.; Sturtevant, J.M.; Venyaminov, S.Y. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 3343-3347.
28. Fink, A.L.; Anderson, W.D.; Antonio, L. *FEBS Lett.* **1988**, *229*, 123-126.
29. Hatley, R.H.M.; Franks, F. *Cryo-Letters* **1986**, *7*, 226-233.
30. Brandts, J.F. *J. Am. Chem. Soc.* **1964**, *86*, 4291-4301.
31. Pace, C.N. *Methods Enzymol.* **1986**, *131*, 266-280.
32. Wu, C.-S.C.; Chen, G.C. *Anal. Biochem.* **1989**, *177*, 178-182.
33. van Stokkum, I.H.M.; Spoelder, H.J.W.; Bloemendal, M.; van Grondelle, R.; Groen, F.C.A. *Anal. Biochem.* **1990**, *191*, 110-118.

34. Schwyzer, R. *EMBO J.* **1987**, *6*, 2255-2259.
35. Kaiser, E.T.; Kézdy, F.J. *Ann. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 561-581.
36. Wu, C.-S. C.; Yang, J.T. *Mol. Cell. Biochem.* **1981**, *40*, 109-122.
37. Schwyzer, R. *Biochemistry* **1986**, *25*, 6335-6342.
38. Schwyzer, R. *Biopolymers* **1991**, *31*, 785-792.
39. Wright, P.E.; Dyson, H.J.; Lerner, R.A. *Biochemistry* **1988**, *27*, 7167-7175.
40. Pasta, P.; Vecchio, G.; Carrea, G. *Biochim. Biophys. Acta* **1988**, *953*, 314-320.
41. Manning, M.C. *Biochem. Biophys. Res. Commun.* **1989**, *160*, 388-392.
42. Hubbard, J.A.; Martin, S.R.; Chaplin, L.C.; Bose, C.; Kelly, S.M.; Price, N.C. *Biochem. J.* **1991**, *275*, 785-788.
43. Mammi, S.; Mammi, N.J.; Peggion, E. *Biochemistry* **1988**, *27*, 1374-1379.
44. Lehrman, S.R.; Tuls, J.L.; Lund, M. *Biochemistry* **1990**, *29*, 5590-5596.
45. Holladay, L.A. *Biophys. Chem.* **1977**, *7*, 41-49.
46. Nelson, J.W.; Kallenbach, N.R. *Proteins: Structure, Function, Genetics* **1986**, *1*, 211-217.
47. Bayley, P.; Martin, S.; Jones, G. *FEBS Lett.* **1988**, *238*, 61-66.
48. Honda, S.; Ohashi, S.; Morii, H.; Uedira, H. *Biopolymers* **1991**, *31*, 869-876.
49. Koseki, T.; Kitabatake, N.; Doi, E. *J. Biochem.* **1988**, *103*, 425-430.
50. Ohgushi, M.; Waada, A. *FEBS Lett.* **1983**, *164*, 21-24.
51. Ptitsyn, O.B. *J. Protein Chem.* **1987**, *6*, 273-293.
52. Brems, D.N.; Brown, P.L.; Heckenlaible, L.A.; Frank, B.H. *Biochemistry* **1990**, *29*, 9289-9293.
53. Bayley, P.M. *Prog. Biophys. Mol. Biol.* **1981**, *37*, 149-180.
54. Einterz, C.M.; Lewis, J.W.; Milder, S.J.; Kliger, D.S. *J. Phys. Chem.* **1985**, *89*, 3845-3853.
55. Lewis, J.W.; Tilton, R.F.; Einterz, C.M.; Milder, S.J.; Kuntz, I.D.; Kliger, D.S. *J. Phys. Chem.* **1985**, 89, 289-294.
56. Goto, Y.; Fink, A.L. *Biochemistry* **1989**, *28*, 945-952.
57. Brems, D.N.; Havel, H.A. *Proteins: Structure, Function, Genetics* **1989**, *5*, 93-95.
58. Brems, D.N.; Plaisted, S.M.; Dougherty, J.J., Jr.; Holzmann, T.F. *J. Biol. Chem.* **1987**, *262*, 2590-2596.
59. Kim, P.S.; Baldwin, R.L. *Ann. Rev. Biochem.* **1982**, *51*, 459-489.
60. Shortle, D. *J. Biol. Chem.* **1989**, *264*, 5315-5318.
61. Alber, T. *Ann. Rev. Biochem.* **1989**, *58*, 765-798.
62. Brems, D.N.; Plaisted, S.M.; Havel, H.A.; Tomich, C.-S.C. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 3367-3371.
63. Glatz, C.E. in *Pharmaceutical Biotechnology* ; Volume 2, Ahern, T.J., Manning, M.C., Ed.; Plenum Press, New York, 1992, pp. 000-000.
64. Melberg, S.G.; Johnson, W.C., Jr. *Proteins: Structure, Function, Genetics* **1990**, *8*, 280-286.
65. Mulkerrin, M.G.; Wetzel, R. *Biochemistry* **1989**, *28*, 6556-6561.
66. Fish, W.W.; Danielsson, A.; Nordling, K.; Miller, S.H.; Lam, C.F.; Björk, I. *Biochemistry* **1985**, *24*, 1510-1517.
67. Havel, H.A.; Kauffman, E.W.; Plaisted, S.M.; Brems, D.N. *Biochemistry* **1986**, *25*, 6533-6538.
68. Brems, D.N.; Plaisted, S.M.; Kauffman, E.W.; Havel, H.A. *Biochemistry* **1986**, *25*, 6539-6543.
69. Pretzer, D.; Schulteis, B.; Vander Velde, D.G.; Smith, C.D.; Mitchell, J.W.; Manning, M.C. *Pharmaceutical Research* **1991**, *8*, 1103-1112.

70. Pretzer, D.; Schulteis, B.; Vander Velde, D.G.; Smith, C.D.; Mitchell, J.W.; Manning, M.C. in *Pharmaceutical Biotechnology* ; Wang, Y.J.; Pearlman, R., Eds.; Plenum Press, New York, 1992, Volume 5, pp. 000-000.
71. Bertazzon, A.; Tian, G.H.; Lamblin, A.; Tsong, T.Y. *Biochemistry* **1990**, *29*, 291-298.
72. Martin, S.R.; Bayley, P.M. *Biochem. J.* **1986**, *238*, 485-490.
73. DeW. Horrocks, W., Jr.; Ishley, J.N.; Holmquist, B.; Thompson, J.S. *J. Inorg. Biochem.* **1980**, *12*, 131-141.
74. Thomas, B.; Wollmer, A. *Biol. Chem. Hoppe-Seyler* **1989**, *370*, 1235-1244.
75. Mirmira, R.G.; Tager, H.S. *Biochemistry* **1991**, *30*, 8222-8229.
76. Markland, F.S.; Reddy, K.N.N.; Guan, L. in: *Hemostasis and Animal Venoms*, Pirkle, H. and Markland, F.S., Ed., Marcel Dekker, New York, 1988, pp. 173-189.
77. Ahmed, N.K.; Tennant, K.D.; Markland, F.S.; Lacz, J.P. *Haemostasis* **1990**, *20*, 147-154.
78. Bajwa, S.S., Kirakossian, H., Reddy, K.N.N., and Marklan, F.S. *Toxicon* **1982**, *20*, 427-432.
79. Miller, J.; McLachlan, A.D.; Klug, A. *EMBO J.* **1985**, *4*, 1609-1614.

RECEIVED March 31, 1992

Chapter 4

Structure–Function Relationship of Hyperthermophilic Enzymes

Rainer Jaenicke

Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, Universitätstrasse 31, Regensburg D–8400, Federal Republic of Germany

The upper limit of thermal adaptation in the biosphere (≈110°C) coincides with the temperature where hydrophobic hydration vanishes and biomolecules start undergoing hydrothermal decomposition. Regarding T_{max}, hyperthermophilic microorganisms come close to this limit. *Thermotoga maritima* ($T_{opt} \leq 90°C$) has adapted its cellular inventory to $T \geq 100°C$. Enzymes purified to homogeneity show intrinsic stability up to ≈110°C. Their overall properties at physiological temperature resemble those of their mesophilic counterparts: Mutative adaptation tends to maintain "corresponding states" regarding structure, flexibility and ligand binding. Physical, enzymatic and folding properties of glyceraldehyde-3-phosphate dehydrogenase, lactate dehydrogenase and amylase are discussed. Enhanced stability may be ascribed to improved packing and enhanced ligand and/or subunit interactions. Due to the minute adaptive changes in ΔG no general strategy of thermophilism can be given.

There is life on the entire globe. The limits of viability for the biologically relevant variables are: for temperature ≤ 110°C, for hydrostatic pressure ≤ 120 MPa, for water activity ≥ 0.6 (corresponding to salinities up to 6 M), and for pH, values below 2 and beyond 10. Organisms exposed to extreme conditions may respond either by "avoidance", or they have to adapt their cellular components. In the first case, the stress parameter is excluded or compensated, for example, by neutralizing the anomalous external pH, or by levelling a salt gradient through isosmotic concentrations of "compatible" cytoplasmic components; in the second, mutations have resulted in the increased stability requirements, so that all the cell's constituents are intrinsically resistant against the specific variable. Evidently, external structural elements facing the outside world cannot avoid stress parameters. Similarly, because of the isothermal and isobaric conditions in a given habitat, the entire cell inventory of organisms adapted to extremes of

0097–6156/93/0516–0053$06.00/0

temperature and pressure is expected to show enhanced intrinsic stability. However, there are extrinsic effects and effectors which may assist cellular components to overcome stress (*1*).

Organisms which have evolved such that they *require* extreme conditions for their whole life cycle are called "extremophiles", or, more specific for extremes of temperature, pressure, salinity and pH: *thermo-* or *psychrophiles*, *barophiles*, *halophiles* and *acido- or alkalophiles*. In the present paper we are dealing with high temperature, focusing our attention to the upper limit. "Hyperthermophilic" organisms require temperatures close to the boiling point of water for full metabolic activity and reproduction. Correspondingly, their protein inventory is expected to exhibit anomalous thermal stability. In the following, we address the question how this is brought about, and whether general mechanisms of heat stabilization can be deduced from physicochemical studies on the most extreme examples of thermophilic proteins that are presently known. It is obvious that the answer to the question has technological, as well as biological implications: *technological*, because the molecular mechanism of thermal adaptation might be applicable to produce engineered thermostable proteins in connection with bioreactors or biosensors; *biological*, because of the evolutionary aspect of adaptive response reactions to physiological stress.

Organisms and Representative Proteins

Proteins discussed in this article were isolated (i) from *Thermotoga maritima*, an early descendant of the eubacterial evolutionary "tree" which can be grown to high cell mass under strictly anaerobic conditions, and (ii) from *Pyrodictium occultum*, one of the most extreme representatives among the thermophilic archaea known today (*2*).

Classical chromatographic methods, including (metal-)affinity chromatography and immuno-adsorption chromatography, were applied for purification (*3-5*, A. Tomschy, H. Schurig, R. Glockshuber, R. Jaenicke, unpublished results). After establishing (partial) sequences, homology modelling and energy minimization were applied for structure predictions (*6*). In the case of D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from *Thermotoga maritima*, where the full amino acid sequence has been determined both at the protein and the DNA level (*7*; A. Tomschy, unpublished results), cloning and expression of the gene in *Escherichia coli* clearly indicates that the hyperthermophilic enzyme is expressed in active form under non-thermophilic conditions. Obviously, the acquisition of the native structure in the heterologous host does not require the physiological temperature conditions of the guest. This is not trivial since the weak intermolecular interactions stabilizing the native structure of proteins show characteristic temperature dependences which, far below the optimum temperature, may cause cold denaturation (*8,9*; see below). In the present case, the temperature difference amounts to ca. 60°C. Less dramatic examples have been reported in (*1*).

Limits of Growth - Limits of Stability of the Covalent Structure

As indicated, hyperthermophilic GAPDH from *Thermotoga maritima* still forms its native three-dimensional structure when it is expressed about 60 degrees

below its optimum temperature. As will be shown, upon reconstitution at 0°C, or in the presence of low concentrations of chaotropic agents, the enzyme does show a certain degree of low-temperature destabilization, probably due to the involvement of hydrophobic interactions upon forming the native tertiary and quaternary structure (*10-12*).

Apart from temperature-induced perturbations of the native three-dimensional structure, common Arrhenius behavior leads to a more or less complete loss of catalytic function at low temperature. In the case of *Thermotoga maritima* or *Pyrodictium occultum*, significant cell growth requires temperatures beyond 50°C (*2,13*). Considering the temperature effect on the metabolic network, it is obvious that temperature adaptation not only requires directed alterations of the intrinsic stability of cell constituents, but also tuning of the kinetics. Depending on the activation energies of the reactions involved, shifts in the optimum temperature from mesophiles to thermophiles may cause dramatic kinetic dislocations (Table I).

Table I. Alterations of Relative Reaction Rates Normalized to 20°C

T (°C)	Activation Energy (kJ/mol)		
	16	32	64
20	1	1	1
60	3	10	90
100	14	186	30 000

On the high-temperature extreme, one may ask why there seems to be a definite limit of growth and reproduction at temperatures around 120°C. Considering this limit and the reversible deactivation of enzymes at water contents below the normal degree of hydration of proteins (≤ 0.25 g H_2O/g protein), the ultimate requirement for life seems to be the presence of liquid water. In testing this hypothesis, stabilization of the liquid state of water at high hydrostatic pressure proves that the temperature limit of viability (T_{max}) cannot be shifted significantly (*14*). Life under "Black Smoker" conditions (26 MPa and 250°C) must be science fiction: the susceptibility of the covalent structure of the polypeptide chain toward hydrolysis, and the hydrothermal degradation of essential biomolecules (*15,16*) would require compensatory "anaplerotic reactions" which are incompatible with the relatively low metabolic activity of most extremophilic organisms.

It is worth mentioning that the hydrophobic hydration also seems to vanish close to the above temperature limit (*17-19*) so that one may assume that both the stability of the covalent structure of the polypeptide chain and its sidechains and their weak interactions determine the upper limit of protein stability. In this context it is important to note that the building blocks of proteins from extremophiles (including hyperthermophiles) are exclusively the canonical 20 natural amino acids. Thus, molecular adaptation to altered environmental conditions

can merely depend on the mutative change in the local and global distribution of the amino acids along the polypeptide chain[1]. From studies on synthetic polypeptides it became clear that the natural amino acids basically allow to build protein molecules with stabilities beyond those commonly observed for natural proteins (*20*). The conclusion we may draw from this observation is that molecular adaptation obviously results in optimum protein flexibility rather than maximum stability.

Conformational stability of proteins

The central issue in the adaptation of biomolecules to extreme conditions is the conservation of their functional state which is characterized by a well-balanced compromise of stability and flexibility. There exist proteins with exceedingly high intrinsic thermal stability and, as has been mentioned, the given repertoire of natural amino acids would have allowed evolution to develop even more stable proteins if they were advantageous (*11,20*).

Basic mechanisms of molecular adaptation are changes in packing density, charge distribution, hydrophobic surface area, and in the ratio of polar/non-polar or acidic/basic residues. Evidently, dramatic changes in stabilization energy do not occur, even under the most extreme conditions. In general, the overall free energy of stabilization of globular proteins in their natural solvent environment is only marginal. The requirement for flexibility is fulfilled by the subtle compensation of attractive and repulsive interactions. Their individual contributions to the free energy are exceedingly large; however, their superposition yields a small difference between big numbers. Considering the energy contributions of the relevant weak intermolecular forces involved in protein stabilization, on balance, stability is attributable to the equivalent of a few hydrogen bonds, ion pairs or hydrophobic interactions. Taking a known three-dimensional structure such as ribonuclease T1 as an example, its native secondary structure and the packing of its hydrophobic inner core is stabilized by 2 disulfide bridges, 87 intramolecular hydrogen bonds and hydrophobic interactions involving ≈ 85% of its non-polar residues; nevertheless, the free energy of stabilization, ΔG_{stab}, at pH 7 and 25°C is no more than 24 kJ/mol, corresponding to the equivalent of just one ion pair or 2-3 hydrogen bonds. The question whether H-bonds do contribute to protein stability, i.e., whether there is a difference in bond strength between water-water and water-protein hydrogen bonds, has been a long-standing controversy (*11,12,18,21*). Obviously, the answer turns out to be positive: Using again ribonuclease T1 as an example, elimination of single hydrogen bonds by site-directed mutagenesis has shown that the stability increment per H-bond is of the order of 5 kJ/mol, which would clearly indicate that H-bonding in ribonuclease T1 contributes equally, or even to a higher extent, to the conformational stability than hydrophobic interactions (C.N.Pace, personal communication).

The above ΔG_{stab} value observed for ribonuclease T1 can be generalized: the free energies of stabilization for a wide variety of globular proteins with totally

[1] In contrast, membrane lipids of extremophiles have been shown to contain a whole spectrum of unusual components (*2*).

unrelated structures have been shown to cluster in a narrow range between 25 and 65 kJ/mol, independent of both the size of the protein and the mode of denaturation (*1,22*). Determining the stability increment per residue, the resulting figure is one order of magnitude below the thermal energy (kT). This proves clearly, that the overall stability of a polypeptide chain must involve cooperativity because the addition of stability increments per amino-acid residue in the process of structure formation would not allow to overcome the thermal energy. In cases where the molecular mass of a protein is too small to provide the necessary size of a cooperative unit, the structure may be stabilized by covalent cross-linking or ligand binding. Both improve the stability of the entire molecule by decreasing the entropy of unfolding. For a discussion of the stability of local structures in proteins, see (*11,23-26*).

Extremophilic proteins do not exhibit properties qualitatively different from non-extremophilic ones. The essential adaptive alteration tends to shift the normal characteristics to the respective extreme in the sense that under the mutual physiological conditions the molecular properties are comparable; with other words, adaptation to extremes of physical conditions tends to maintain "corresponding states" regarding overall topology, flexibility and solvation. Considering the parabolic profile of the temperature dependence of the free energy of stabilization (*18,19,22*), it is evident that enhanced thermal stability may be accomplished by a variety of possible mechanisms. As depicted schematically in Fig. 1A, the profile of the mesophilic wild-type could either be shifted to higher temperature, or it could be lowered, or flattened; in all three cases, the "melting temperature" of the protein would be shifted to a (hyper-)thermophilic higher value. Experimental evidence seems to indicate that the third alternative holds true (Fig. 1B). Thus, the temperature dependence of ΔG_{stab} of thermophilic proteins is less pronounced than the one of the corresponding wild-type protein. A similar independence of temperature has been found for the free energy of ligand binding to enzymes; in this case, entropy-enthalpy compensation renders thermal adaptation unnecessary (Fig. 1C).

In summarizing the thermodynamics of thermophilic adaptation in quantitative terms, even for *hyper*thermophiles, the changes in free energy, $\Delta\Delta G_{stab}$, generally do not exceed the equivalent of a few additional weak interactions. At present, no comparative X-ray analysis of any protein from a mesophile, thermophile and hyperthermophile has been put forward at sufficiently high resolution to pin down specific residues responsible for the gradual increase in protein stability. The most thorough studies in this context refer to X-ray crystallographic and thermodynamic studies on pointmutants of bacteriophage T4 lysozyme (*1,27*). The results illustrate how subtle the different weak interactions are balanced in native globular proteins, and how much, even marginal local strain in the three-dimensional structure may affect protein stability. From the point of view of protein engineering and the attempt to provide deeper insight into the mechanisms of thermophilic adaptation, the results are a test case to prove that presently no unambiguous predictions or general strategies of protein stabilization can be given. One problem in this context is the fact that high-resolution X-ray structures may not be adequate to detect the real structural changes accompanying "isomorphous single point mutations", because the lattice forces in protein crystals may freeze or shift positions of the polypeptide backbone which are not representative for the relevant *local* solution structure.

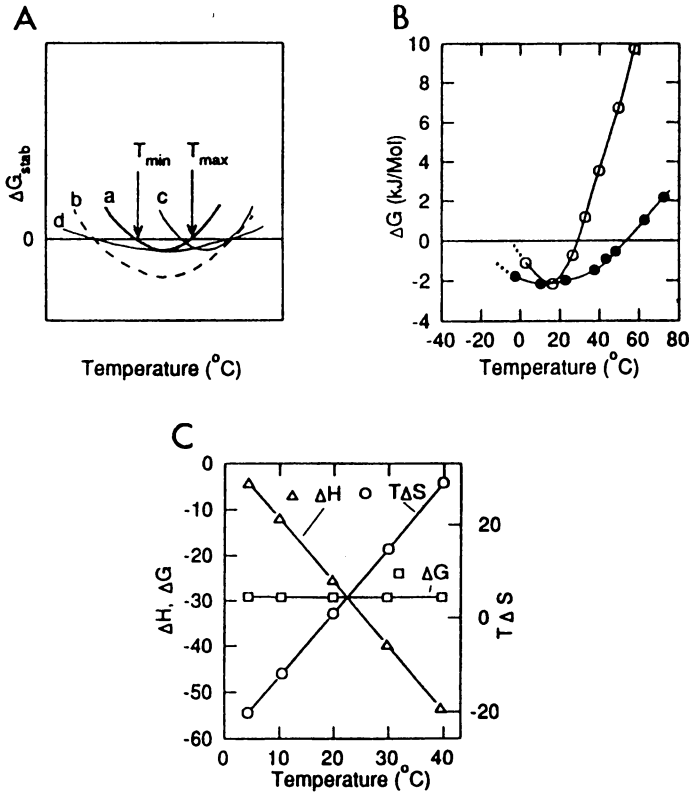


Figure 1. Temperature dependence of the free energy of stabilization and ligand binding. (A) Hypothetical ΔG_{stab} profiles for mesophilic (a) and thermophilic (b-d) proteins. (B) ΔG_{stab} for phosphoglycerate kinase from yeast (o) and *Thermus thermophilus* (●), in the presence of 0.5 M (o) and 2.3 M GdmCl (●). (C) Entropy/enthalpy compensation upon binding of NADH to porcine muscle LDH-M_4.

Selected Examples : Enzymes from *Thermotoga maritima*

Thermophilic adaptation of proteins occurs at the DNA level leading to specific local alterations in structural genes and, subsequently, amino-acid sequences. For almost a decade, such changes have been assumed to be manifested in certain well-defined shifts in the amino-acid composition. The significance of the proposed "rules" for the "gross traffic" in the case of thermal stabilization cannot be considered unambiguous (*1,3,4,28*). The reason is obvious: The previously mentioned low values of the extrafree energy of stabilization ($\Delta\Delta G_{stab}$) required for the increase in stability can be provided by an astronomically large number of subtle changes in local weak interactions. Evidently, summarizing these changes in terms of preferred amino-acid exchanges, minimization of crevices, or enhanced ion-pair formation etc. is an oversimplification. What seems to be well-established are the following observations:
1. Proteins from thermophilic sources commonly exhibit high *intrinsic* stability which becomes marginal at their respective physiological temperatures. At optimum temperature, homologous enzymes show similar conformational flexibility. The increase in rigidity of thermophilic proteins at room-temperature is evident from decreased exchange rates of amide protons, as well as increased resistance against proteolysis and denaturation using mesophilic homologs as references. 2. Available structural data prove that thermophilic proteins are closely similar to their mesophilic counterparts as far as their basic topology, activity and enzyme mechanism are concerned. Correspondingly, virtually all the amino acids constituting the active sites are conserved. 3. The correlation of sequence alterations with changes in thermal stability is expected to be highly complex. It is closely related to the protein folding problem, as increased stability is equivalent to the difference in minimum energy of the thermophilic *vs.* the mesophilic polypeptide chains. So far, no general strategy of thermal stabilization is available.

In the following, specific characteristics regarding compartmentation, enzymatic and solution properties, as well as self-organization of amylases, glyceraldehyde-3-phosphate dehydrogenase and lactate dehydrogenase from *Thermotoga maritima*, will be discussed.

Amylases. *Thermotoga maritima* can be grown on minimal medium with starch as carbon source. The uptake of glucose is not accomplished by secreting amylases into the outside medium: less than 1% of the total amylase activity is detected in the supernatant of the cells. Similarly, there is no release of activity from the periplasm after lysozyme treatment of the cells. Thus, starch as a nutrient must be degraded on the cell surface in order to be able to pass the outer sheath and the periplasm. As shown by Percoll gradient centrifugation, in fact, more than 85% of the total α-, ß- and glucoamylase activity are found to be associated with the "toga". The expression of the three isoenzymes is too low to isolate the pure proteins in mg quantities which would be required for a detailed characterization. Compared with α-amylase from *Bacillus licheniformis* (T_{max} = 75°C), the amylases from *Thermotoga* show high intrinsic stability with upper temperature limits at ≥ 95°C (Table II). Significant turnover of the enzyme occurs only beyond 70°C, i.e., under optimal growth conditions.

Table II. Amylases from *Thermotoga maritima* and *Bacillus licheniformis*

	Thermotoga maritima		*Bacillus licheniformis*
	Fraction I	Fraction II	
Compartmentation	associated with toga		secreted
Number of active bands	2	1	1
Molecular mass (kDa)	≈ 60	≈ 60	60
Specificity	ß	α and gluco	α
Optimum temperature (°C)	95	90	70-75
Activation energy (kJ/mol)	98	60	14
K_M in % starch at (°C)	0.23 (80)	0.22 (80)	0.15 (45)
pH optimum	5.0	6.0	6.8
Cross-reactivity with H-α*	-	-	+

* Cross-reaction with human salivary α-amylase monoclonal antibody

D-Glyceraldehyde-3-phosphate Dehydrogenase. Among the cytosolic enzymes which have been screened, GAPDH shows the highest expression level. Since a series of GAPDH's from various sources has been investigated in great detail, this enzyme was chosen to compare homologs with widely differing thermal stabilities. The enzyme has been purified to homogeneity (*3*); its amino-acid sequence has been determined at the protein (*7*) and DNA level (A. Tomschy, unpublished), and its physicochemical and enzymological properties have been compared with the enzyme from mesophiles and other thermophiles (*3,6*). Considering the amino-acid compositions (Table III), it is obvious that only the exchanges Lys → Arg, Ser → Thr and Val → Ile agree with the results of previous statistical analyses (*28*). The primary structure is highly homologous with the enzyme from other thermophilic bacteria: comparing the sequences of the enzyme from *Thermotoga*, *Bacillus stearothermophilus* and *Thermus aquaticus*, 63 and 59% identity are observed, only 8% of the exchanges are non-conservative; taking all known structures together, about 33% of the residues are identical. Homology modelling yields equal topologies (*6*), in accordance with the observation that the quaternary and secondary structure are closely similar and that the essential residues in the active site are conserved in all species. Marked differences regarding the physical properties of hyperthermophilic GAPDH and its non-thermophilic counterparts are (cf. Table IV and Fig. 2): (a) enhanced intrinsic stability; (b) decreased exchange rates in H-D exchange experiments (*3*); (c) a smaller difference in the change in hydrodynamic volume upon coenzyme binding (Rehaber, V.; Jaenicke, R. *J. Biol. Chem.* 1992, in press); (d) low-temperature intermediate(s) on the pathway of reconstitution with a shift of the tetramer-monomer equilibrium at 0°C (*29*); (e) activation of the enzyme at low denaturant concentrations (Rehaber, V; Jaenicke, R. *J. Biol. Chem.* 1992, in press). The characteristics confirm the tight packing of the *Thermotoga* enzyme and the inhibitory effect of the decreased flexibility on the catalytic function at low temperature; in addition they suggest a significant contribution of hydrophobic interactions to the tertiary and quaternary structure of the enzyme.

Table III. Amino-acid Composition of GAPDH's from *Thermotoga maritima* and Other Sources (given in numbers of residues per polypeptide chain)

Amino acid	*Thermotoga*	Yeast	Thermophiles	Mesophiles
G	24	27	25-26	26-35
A	25	32	38-44	32-34
V	36	37	28-42	33-38
L	26	20	26-31	18-20
I	30	19	19-22	18-21
M	4	6	5-7	6-10
F	8	10	5-6	10-15
W	2	3	2-3	2-3
P	14	12	11-12	11-13
S	15	29	13-18	18-29
T	27	23	18-22	18-23
B	35	36	34-40	29-39
Z	26	19	24-28	18-24
C	3	2	1-3	2-5
Y	9	11	8-10	9-11
K	27	26	23-24	24-28
R	14	11	14-16	9-11
H	7	8	9-10	5-11
Sum	332	331	333-335	332-336

Lactate Dehydrogenase. In contrast to GAPDH, LDH shows an extremely low expression level. Due to its structural similarity and its anomalous sequence homology with GAPDH this causes extreme difficulties in the purification and cloning of the enzyme (*4,30,31*). Therefore, the sequence of the enzyme has yet to be elucidated. Table V summarizes the amino-acid composition. At the present stage, again, it is not possible to detect significant trends.

As in the case of GAPDH, the structural properties of the hyperthermophilic enzyme and its mesophilic homologs are closely similar (Table IV). The enzyme exhibits undiminished activity up to the boiling point of water; in the Arrhenius plot neither cold-denaturation nor any significant curvature at maximum temperature can be detected (*4*) (Fig. 2). NAD^+ has a strong stabilizing effect: deactivation of the *apo*enzyme becomes significant at temperatures below 85°C. Fructose-1.6-bisphosphate (in the absence of divalent ions) also serves as an extrinsic stabilizing factor; both ligands bind with high affinity, their effect is additive (*4,32*). In this connection, two points deserve mentioning: 1. Considering the various ligands involved in catalysis and activation, it is obvious that enzymes under physiological conditions do not necessarily operate with maximum efficiency: K_M for NADH and pyruvate show a dramatic *increase* with increasing temperature (*32*). The reason may be that proteins have multiple

Table IV. Properties of homologous mesophilic and thermophilic GAPDHs and LDHs

Property	GAPDH		LDH	
	Thermotoga	Yeast	*Thermotoga*	Pig (H_4)
Molecular mass (M_4) (kDa)	145	144	144	140
Subunit mass (M_1) (kDa)	37	36	36	35
Change in $s_{20,w}$ (holo minus apo)	3.5%	5.3%	n.d.	7.0%
Thermal denaturation (T_m holo °C)	109	40	91	42
Denaturation in GdmCl ($c_{1/2}$ 20°C)	2.1M	0.5M	1.8M	0.5M
Activation at 0.5 M GdmCl (70°C)	300%	0%	n.d.	n.d.
Half-time of thermal denaturation at 100°C (h)	>2	$<10^{-2}$	≈1	$<10^{-2}$
Rel. H-D exchange rate (β_{rel} 25°C)	64	100	n.d.	n.d.
Specific activity U/mg at (°C)	200 (85)	70 (20)	≈1000 (55)	400 (25)
Michaelis constant K_M (μM)				
μM substrate (°C)	400 (60)	160 (25)	60 (55)	76 (25)
μM NAD^+ /*NADH* (°C)	79 (60)	44 (25)	*27* (55)	*10* (25)
Activation energy of enzyme reaction limiting values (kJ/mol)	33-79	25-80	57-104	38-63
Opt. reactivation at 0°C/≥30°C	0%/85%	35%/85%	n.d.	50%/95%

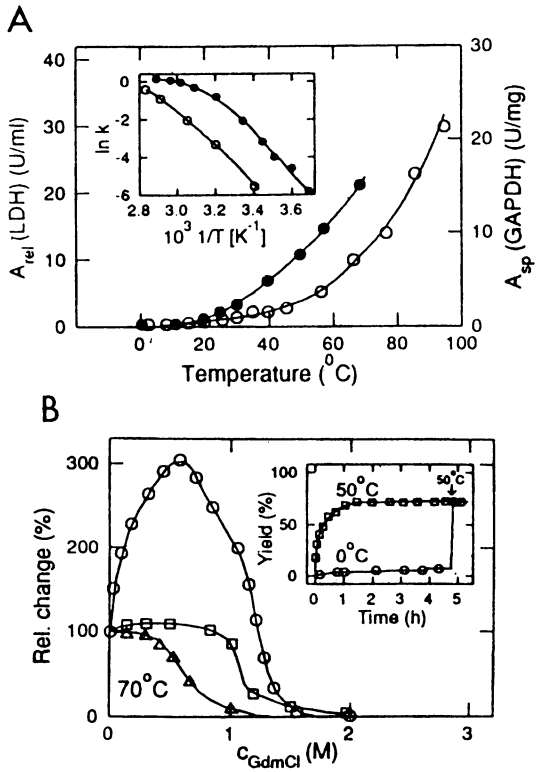


Figure 2. Temperature dependence and denaturation/renaturation of LDH and GAPDH from *Thermotoga maritima*. (A) T-dependence of the enzymatic activity of LDH (o) and GAPDH (●) under standard conditions (*3,4*, Rehaber, V; Jaenicke, R. *J. Biol. Chem.* **1992**, in press). Insert: Arrhenius plot: LDH in the presence of 3 mM F-1.6-BP; in the case of GAPDH, the assay at T>60°C is perturbed by the instability of NADH and the substrates. (B) GdmCl-dependent denaturation of GAPDH monitored by activity (o), fluorescence (□) and circular dichroism (Δ) (Rehaber, V.; Jaenicke, R. *J. Biol. Chem.* **1992**, in press). Insert: Reactivation after denaturation in 6 M GdmCl at 50°C (□) and 0°C (o); after final state was reached, T was shifted to 50°C (*29*).

Table V. Amino-acid Composition of LDHs from *Thermotoga maritima* and Other Sources (given in mol%)

Aminoacid	*Thermotoga maritima*	*Thermus aquaticus*	*Bac. stearo-thermophilus*	*Bac. sacch-arolyticus*	*Pig heart*
G	9.6	10.2	8.8	8.5	7.2
A	8.6	13.9	10.7	8.2	6.0
V	9.5	12.4	9.5	6.3	11.7
L	10.2	11.3	7.3	8.5	10.8
I	7.5	3.8	8.8	9.7	6.9
M	2.0	0.4	1.9	2.2	2.7
F	4.0	3.4	4.4	4.1	1.5
W	n.d.	1.1	1.0	0.9	1.8
P	3.5	3.0	2.8	3.1	3.3
S	4.1	4.2	3.8	4.1	7.8
T	4.9	4.5	3.8	6.0	4.2
B	10.0	5.3	11.7	12.3	10.8
Z	10.7	11.3	8.2	10.1	9.9
C	n.d.	0	0.6	0.9	1.5
Y	2.8	1.9	3.8	3.8	2.1
K	6.5	2.6	4.4	5.7	7.2
R	4.3	8.7	5.7	4.1	2.4
H	2.0	1.9	2.8	1.6	2.1

(Reproduced with permission from ref. 4. Copyright 1990 Springer–Verlag.)

functions which cannot be simultaneously optimized, especially under extreme physical conditions. 2. The fact that *in vitro* studies commonly do not include extrinsic factors may be misleading: Ions, metabolites, coenzymes or specific protectants, such as polyphosphates or chaperons may affect the stability properties of cellular components in a significant way.

Concluding remarks

As has been mentioned, the excess free energy of stabilization for a highly stable protein does not involve more than a few percent of the total number of weak interactions involved in the secondary, tertiary and quaternary contacts within the molecule. Considering the three forms of GAPDH from the thermophilic bacteria *Thermotoga maritima*, *Methanothermus fervidus* and *Bacillus stearothermophilus*, with their respective T_m-values 110, 80 and 68°C, about 100 of the 330 amino acids per subunit differ between the three enzymes; obviously, the answer to the question of which of these is responsible for the increment in thermal stability is not trivial. One way out of this dilemma would be to investigate point mutations yielding proteins of known structures with differing thermal stabilities (cf. *23-27*). For example, single, double and triple mutants of creatinase (a homodimer of 110 kDa molecular mass) exhibit additive increments of stabilization caused by a decrease in hydrophobic surface area and improved packing density; obviously, in this case, subunit interactions do not play a significant role (J. Schumann, unpublished results). In contrast, thermostable point mutations of pyruvate oxidase cause improved coenzyme binding and subunit association, this way promoting the native tetrameric quaternary structure (B. Risse, unpublished results). These investigations try to extend previous studies on the monomeric phage T4 lysozyme (*1,27*) at the next higher level by including ligands and subunit association into the consideration.

Ligands as "extrinsic effectors" have briefly been mentioned. One example is cyclic 2.3 diphosphoglycerate which has been shown to enable non-thermostable proteins to gain thermal stability (*33,34*). Because of their functional analogy to "compatible solutes" and anti-freeze compounds, it should be remembered that similar cellular components allow halotolerant organisms, as well as plants and fish, to survive salt and low-temperature stress without halophilic or psychrophilic adaptation. Further stabilizing factors involving the nascent or folding-competent rather than the native protein, refer to components or parameters "assisting" in structure formation or counteracting misfolding and misassembly (*11,35*). How this "assistance" works, and what is the intermediate state of the nascent or refolding polypeptide chain is not understood. The phenomenon was originally detected as *heat-shock response*, devised to counteract misaggregation as the major consequence of protein heat denaturation. The occurrence of heat-shock proteins in all cells, and their high degree of conservation during evolution led to the view that they must have a fundamental function as "chaperons" for the nascent or infant polypeptide chain (*36,37*). The idea that binding of early folding intermediates to hydrophobic surfaces competes with misfolding, and that additional heat-shock components and/or ATP-hydrolysis lead to conformational changes and subsequent release of the polypeptide chain is a reasonable working hypothesis for future experiments. That such two-step mechanisms may be meaningful has been observed in a number of cases, where reconstitution of

thermophilic proteins was arrested at low temperature; after switching to high temperature folding proceeds to high yields at greatly enhanced rate (*29*). Since aggregation and association are endothermic, low temperature allows the folding polypeptide chain to make the correct tertiary contacts, thus transforming the "molten globule" (*11,38*) into the native three-dimensional structure.

Acknowledgments

I should like to thank the John E. Fogarty International Center for Advanced Study at the NIH, Bethesda, for generous support and hospitality. Fruitful discussions with Drs. G. Böhm, P.L. Privalov, V. Rehaber, B. Risse, R. Rudolph, F.X. Schmid, J. Schumann, H. Schurig, R. Seckler, A. Wrba and P. Závodszky are gratefully acknowledged. Work performed in the author's laboratory was generously supported by grants of the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie and the BAP Program of the European Community.

Literature Cited

1. Jaenicke, R. *Eur.J.Biochem.* **1991**. *202*, 715-728.
2. *Thermophiles*; Brock, T. D., Ed.; Wiley: New York, **1986**.
3. Wrba, A.; Schweiger, A.; Schultes, V.; Jaenicke, R.; Závodszky, P. *Biochemistry* **1990**, *29*, 7585-7592.
4. Wrba, A; Jaenicke, R.; Huber, R.; Stetter, K. O. *Eur. J. Biochem.* **1990**, *188*, 195-201.
5. Schumann, J.; Wrba, A.; Jaenicke, R.; Stetter, K. O. *FEBS Lett.* **1991**, *282*, 122-126.
6. Böhm, G. *Strukturmodellierung extremophiler Proteine*: Dissertation, Universität Regensburg,**1992**.
7. Schultes, V.; Deutzmann, R.; Jaenicke, R. *Eur. J. Biochem.* **1990**, *192*, 25-31.
8. Jaenicke, R. *Phil. Trans. R. Soc.* **1990**, *B326*, 535-553.
9. Privalov, P. L. *CRC Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 281-305.
10. Jaenicke, R. *Progr. Biophys. Mol. Biol.* **1987**, *49*, 117-237.
11. Jaenicke, R. *Biochemistry* **1990**, *30*, 3147-3161.
12. Dill, K. A. *Biochemistry* **1990**, *29*, 7133-7155.
13. Huber, R.; Langworthy, T. A.; König, H.; Thomm, M.; Woese, C. R.; Sleytr, U. B.; Stetter, K. O. *Arch. Microbiol.* **1986**, *144*, 324-333.
14. Bernhardt, G.; Jaenicke, R.; Lüdemann, H.-D.; König, H.; Stetter, K. O. *Appl. Environ. Microbiol.* **1988**, *54*, 2375-2380.
15. Bernhardt, G.; Lüdemann, H.-D.; Jaenicke, R.; König, H.; Stetter, K. O. *Naturwissenschaften* **1984**, *71*, 583-586.
16. White, R. H. *Nature (London)* **1984**, *310*, 430-432.
17. Baldwin, R. L. *Proc. Natl Acad. Sci. U.S.A.* **1986**, *83*, 8069-8072.
18. Privalov, P. L.; Gill, S. J. *Adv.Prot.Chem.* **1988**, *39*, 193-231.
19. Privalov, P. L.; Gill, S. J. *Pure Appl.Chem.* **1989**, *61*, 1197-1104.
20. De Grado, W. F. *Adv. Prot. Chem.* **1988**, *39*, 51-124.
21. Tanford, C. *Adv. Prot. Chem.* **1968**, *23*, 122-282; **1970**, *24*, 1-95.
22. Privalov, P. L. *Adv. Prot. Chem.* **1979**, *33*, 167-241.

23. Pace, C. N. *Trends Biotechnol.* **1990**, *8*, 93-98.
24. Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631-660.
25. Dobson, C. M. *Curr. Opin. Struct. Biol.* **1991**, *1*, 22-27.
26. Matthews, C. R. *Curr. Opin. Struct. Biol.* **1991**, *1*, 28-35.
27. Matthews, B. W. *Curr. Opin. Struct. Biol.* **1991**, *1*, 17-21.
28. Argos. P.; Rossmann, M. G.; Grau, U. M.; Zuber, H.; Frank, G.; Tratschin, J. D. *Biochemistry* **1979**, *18*, 5698-5703.
29. Schultes, V.; Jaenicke, R. *FEBS Lett.* **1991**, *290*, 235-238.
30. Wrba, A. *Enzyme aus dem extrem thermophilen Eubakterium Thermotoga maritima: Reinigung und Charakterisierung der LDH und GAPDH*: Dissertation, Universität Regensburg, **1989**.
31. Tomschy, A. *Versuch zur Klonierung extrem thermophiler Enzyme aus Thermotoga maritima*: Diplomarbeit, Universität Regensburg, **1990**.
32. Hecht, K.; Wrba, A.; Jaenicke, R. *Eur. J. Biochem.* **1989**, *183*, 69-74.
33. Hensel, R.; König, H. *FEMS Microbiol. Lett.* **1988**, *49*, 75-79.
34. Huber, R.; Kurr, M.; Jannasch, H. W.; Stetter, K. O. *Nature (London)* **1989**, *342*, 833-834.
35. Fischer, G.; Schmid, F. X. *Biochemistry* **1990**, *29*, 2205-2212.
36. Ellis, R. J. *Semin. Cell Biol.* **1990**, *1*, 1-9.
37. Ellis, R. J.; van der Vies, S. *Annu. Rev. Biochem.* **1991**, *60*, 321-347.
38. Christensen, H.; Pain, R. *Eur. Biophys. J.* **1991**, *19*, 221-229.

RECEIVED March 31, 1992

Chapter 5

Psychrophilic Proteinases from Atlantic Cod

J. B. Bjarnason[1], B. Asgeirsson[1], and J. W. Fox[2]

[1]Science Institute, University of Iceland, 107 Reykjavik, Iceland
[2]Department of Microbiology, University of Virginia Medical School, Charlottesville, VA 22908

One mechanism of cold adaptation in ectothermic organisms involves alterations of the efficiency of enzyme catalyzed reactions at lower temperatures. Adaptation of ectothermic organisms is required so that the normal physiological/metabolic pathways operate at a flux level necessary to sustain life at the ambient temperature of the organism's environment. To meet these physiological requirements organisms may simply increase the levels of the enzymes. A more long-term adaptation is the structural evolution of the enzymes such that their catalytic properties are optimized for function at low ambient temperatures. The Atlantic cod, *Gadus morhua*, is a cold-water poikilothermic fish living in waters with temperatures ranging from 0 to 8°C. We have isolated and characterized the trypsins, chymotrypsins and elastases from the digestive tract of this fish and we have demonstrated that these enzymes have altered biophysical and functional properties compared with analogous enzymes from endotherms. Essentially, the cod enzymes have an increased catalytic efficiency accompanied by a decrease in thermal stability. This observation leads one to the hypothesis that the enhanced catalytic characteristics of the fish enzymes is structurally interfaced with the decreased thermal stability. Both of these characteristics, increased catalytic efficiency and decreased thermal stability have some practical value in industrial applications and the structural principles underlying these properties may in the future be exploited by protein engineering of enzyme systems other than the proteases.

Biochemical attributes of organisms that have adapted to low temperatures must include the preservation of structural integrity of the macromolecules as well as the macromolecular assemblies for function in a particular environment (*1*). In the case of the adaptation of ectothermic organisms to low temperature conditions, the evolutionary pattern is such that there is often observed a compensation of catalytic rates either by varying enzyme

0097–6156/93/0516–0068$06.00/0

concentrations or increasing the catalytic capacity of enzymes (*1*). Changes in enzyme concentrations are often seen in response to seasonal variations in temperature (*1,2*), however, alteration in catalytic ability is typically more common as an evolutionary mechanism of enzymatic adaptation to low temperature. This adaptation can take the form of changes in the interaction of the enzyme with its substrate as measured by K_m values but in the case of intracellular enzymes the K_m's tend to be conserved in relation to the physiological levels of the substrates.

Alteration of the catalytic efficiency of enzymes of ectothermic, cold adapted organisms is generally the result of enzyme's ability to lower the free energy of activation for the particular reaction. Low and Somero (*3*) have proposed that this can be achieved structurally by an increase in the weak interactions between a substrate and the enzyme which results in further stabilization of the transition state of the reaction. Complimentary to this would be a reduction of weak intramolecular interactions which normally serve to stabilize the protein structure with the result of a generally less-stable but more flexible structure (*1*). On an evolutionary time scale these changes must occur at the level of the primary structure of the enzymes.

We have chosen to study the digestive enzymes of the Atlantic cod, *Gadus morhua*, a poikilothermic teleost fish which inhabits waters with temperatures ranging from 0 to 8°C. These serine proteinases well serve our requirements in trying to unravel the structural basis of the cold-activity of these enzymes since there is abundant literature on the kinetic and biophysical properties of serine proteinases from warm-blooded animals. This allows for ready comparison with the analogous enzymes from low temperature adapted species. It is our recent work on these fish enzymes which we will review in this chapter.

Cod Trypsins

Isolation and biochemical characterization. Three isozymes of trypsin were isolated from the pyloric caeca of the Atlantic cod using a combination of affinity and gel filtration chromatography and chromatofocusing (*4*). All three trypsins had a similar molecular mass of 24.2 kDa. The three isozymes had isoelectric points of 6.6, 6.2 and 5.5. The 6.6 pI species was the most abundant of the three species. The amino acid composition of the primary trypsin is shown in Table I. The composition is similar to the other trypsin compositions shown. The average residue hydrophobicity of the cod trypsin is 3.68 kJ compared to 3.60 kJ/residue for the Greenland cod trypsin, both of which are considerably lower than the 4.35 kJ/residue calculated for bovine trypsin (*4*). The N-terminal amino acid sequence of the major cod trypsin is compared with that of several other trypsins in Table II. Over the 37 residues determined the cod trypsin had 30 identical residues with rat and porcine trypsin, 29 with bovine trypsin, 26 with dogfish trypsin and 6 with crayfish trypsin.

Activity characteristics. The esterolytic and amidolytic kinetic properties of the three cod trypsins were determined with the respective substrates TosArgOMe and BzArg-NH-Np[1]. Table III shows the results of these studies. As can be observed, some differences in the catalytic efficiency (k_{cat}/K_m) between the three cod species as well as bovine trypsin are apparent. Two critical points can be taken from these results. Firstly, all of

Table I. The amino acid composition of Atlantic cod trypsin (enzyme 1) compared with trypsins from other species

	Atlantic cod (enzyme I)	Greenland cod	Bovine	Porcine	Human	Shrimp	Sardine	Catfish
Aspartic acid	19	23	22	18	21	30	31	25
Glutamic acid	27	19	14	17	21	24	16	22
Serine	23	32	33	24	24	24	17	23
Glycine	28	28	25	26	20	28	27	24
Histidine	6	7	3	4	3	5	3	7
Arginine	6	5	2	4	6	3	6	5
Threonine	10	10	10	11	10	10	19	9
Alanine	16	16	14	16	13	16	15	13
Proline	6	10	9	10	9	11	10	13
Tyrosine	10	7	10	8	7	10	13	9
Valine	18	16	17	16	16	18	18	18
Methiomine	6	3	2	2	1	2	1	6
Cysteine	12	8	12	12	8	8	8	12
Isoleucine	8	8	15	15	12	14	11	13
Leucine	16	14	14	16	12	10	10	13
Phenylalanine	3	4	3	4	4	6	4	6
Lysine	8	6	14	10	11	5	17	11
Tryptophan	3	2	4	6	3	3	4	7
Total	225	218	223	219	201	237	230	236

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

Table II. The N-terminal amino acid sequence of Atlantic cod trypsin (enzyme I) compared to trypsins from other species

	1				5				10				15				20				25				30				35				40
1)	I	V	G	G	Y	Q	C	E	A	H	S	Q	A	H	Q	V	S	L	N	S	G	Y	H	Y	C	G	G	S	L	I	N	-	-	W	V	V	S	A	A
2)	I	V	G	G	Y	T	C	G	A	N	T	V	P	Y	Q	V	S	L	N	S	G	Y	H	P	C	G	G	S	L	I	N	S	Q	W	V	V	S	A	A
3)	I	V	G	G	Y	T	C	A	A	N	S	V	P	Y	Q	V	S	L	N	S	G	Y	H	F	C	G	G	S	L	I	N	S	Q	W	V	V	S	A	A
4)	I	V	G	G	Y	T	C	P	E	H	S	V	P	Y	Q	V	S	L	N	S	G	Y	H	F	C	G	G	S	L	I	N	D	Q	W	V	V	S	A	A
5)	I	V	G	G	Y	E	C	P	K	H	A	A	P	W	T	V	S	L	N	V	G	Y	H	F	C	G	G	S	L	I	A	P	G	W	V	V	S	A	A
6)	I	V	G	G	T	D	A	V	L	G	E	F	P	Y	Q	L	S	F	Q	E	H	F	L	G	F	S	F	H	F	C	G	A	S	I	Y	N	E	N	Y

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

Table III. Kinetic parameters of trypsins purified from Atlantic cod (enzymes I, II and III) compared to those measured for bovine trypsin. TosArgOMe esterolytic activity and BzArg-NH-Np amidolytic activity were measured at 25°C and pH 8.1.

Substrate	K_m(mM)	$k_{cat}(s^{-1})$	$k_{cat}/K_m(s^{-1}mM^{-1})$
TosArgOMe			
Enzyme I	0.029	136.7	4713
Enzyme II	0.021	57.6	2781
Enzyme III	0.049	62.8	1281
Bovine	0.046	86.1	1870
BzArg-NH-Np			
Enzyme I	0.077	4.0	51.9
Enzyme II	0.094	1.9	20.2
Enzyme III	0.102	0.7	6.8
Bovine	0.650	2.0	3.1

the trypsins examined were significantly better esterases than amidases. Secondly, the cod trypsin, at 25°C, has a 17 fold greater amidase and 2.5 fold greater esterase catalytic efficiency than bovine trypsin (*4*). This is clearly demonstrated in Figure 1, where a comparison is made of the kinetic parameters of bovine trypsin and cod trypsin I at various temperature. The cod enzyme had maximal amidase activity at a pH of approximately 8 which is similar for most serine proteinases. The temperature dependence of the esterolyic activity for the cod trypsin as compared to bovine trypsin is seen in Figure 2. Assay of enzymatic activity of the bovine and cod trypsins at temperatures up to 55°C resulted in an increase in enzymatic activity with increasing temperatures. At temperatures approaching 65°C the bovine trypsin begins to decline in activity. The cod trypsin began to lose activity at a significantly lower temperature (55°C), thus suggesting a greater thermal stability for bovine trypsin as compared to the cod enzyme (Figure 2).

Stability characteristics. The effect of pH on the stability of the cod enzyme was measured by incubating the enzyme for either 30 min or 18 h at 5°C in buffers with increasing pH followed by assay of esterase activity at 25°C, pH 8.1 (*4*). As seen in Figure 3, the cod trypsin was stable at alkaline pH but becomes rather unstable at pH values less than 5. Although this is contrary to similar studies on bovine trypsin which is very stable at low pH values (*5*), the results are in accordance with those of trypsins from other fishes (*6,7*). The effect of temperature on the stability of the cod trypsin was not directly determined, however, certain trends can be inferred from the temperature/activity studies described in the above section. Obviously, the cod enzyme is more sensitive to temperature denaturation than bovine trypsin in that a decline in activity with increasing temperature is observed at a much lower temperature compared to the bovine enzyme (Figure 2) suggesting an enhanced temperature dependent denaturation for the cod trypsin. Our recent studies have indicated that the loss of activity of the cod enzyme with increasing temperature is due to both denaturation and increased autolysis.

Summary. These studies comparing the structural and functional properties of the endothermic bovine trypsin and the ectothermic cod trypsins have brought forth several interesting observations. The cod enzyme is a trypsin analog based upon its substrate specificity. From a structural aspect the cod enzyme is a trypsin homolog, to a degree, as seen from its amino acid composition and N-terminal amino acid sequence, however certain critical structural differences must exist which give rise to the different activity and stability characteristics of the cod trypsin compared to the bovine trypsin. Essentially, the cod trypsin is 2 to 20 fold more active than the bovine enzyme at any temperature in terms of the kinetic parameters. The temperature for the maximal activity for the cod enzyme is approximately 10°C lower than the maximal activity temperature for bovine trypsin.

Some indications for the structural basis of the increase catalytic efficiency of cod trypsin are suggested from the data. Firstly, the amino acid composition of the cod enzyme gives a decreased per residue hydrophobicity compared to the bovine composition which could indicate decreased intramolecular interactions at increasing temperatures. This would translate into a decreased relative stability for the cod molecule. The lower stability is reflected in the temperature/activity study where the cod enzyme is seen to lose activity at a lower temperature compared to the bovine enzyme. To

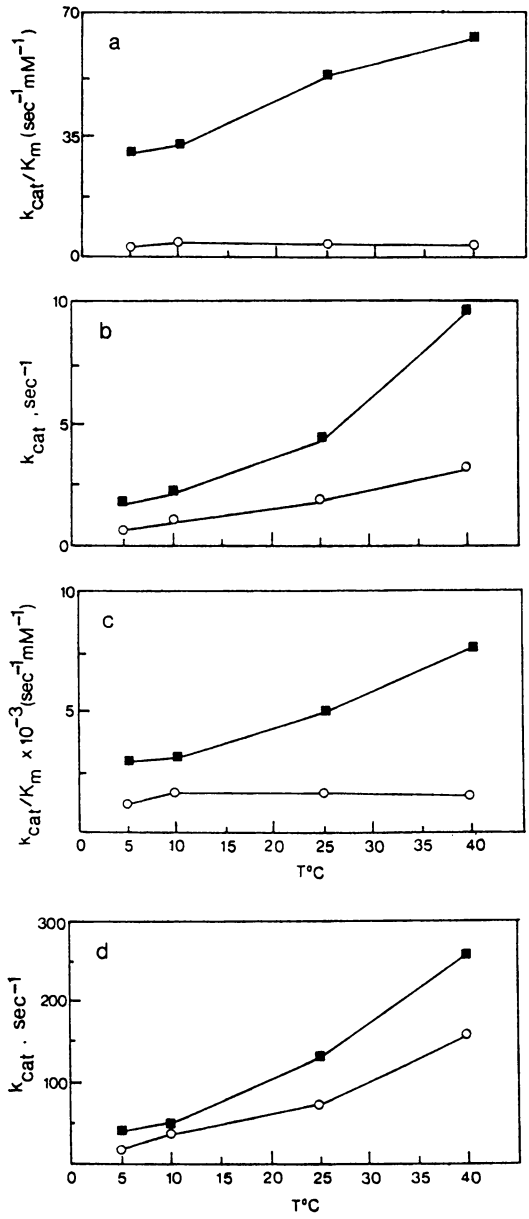


Figure 1. Comparison of catalytic efficiency and turnover number between cod (dark squares) and bovine (circles) trypsins at different temperatures. Amidolysis was determined with the substrate benzoyl-L-arginine p-nitroanilide (a and b) and esterolysis with p-toluenesulfonyl-L-arginine methyl ester (c and d) at pH 8.1.

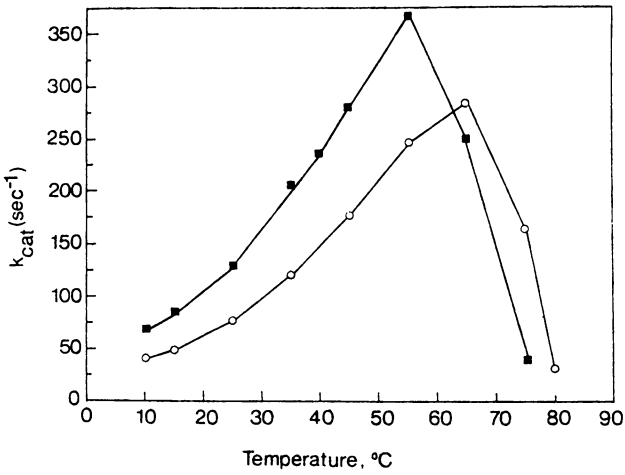


Figure 2. Effect of temperature on the activity of trypsin from Atlantic cod (dark squares) and bovine (circles). The enzymes were added to a preheated thermostatted cuvette, and the average rate of TosArgOMe hydrolysis subsequently determined during a 3 minute period.

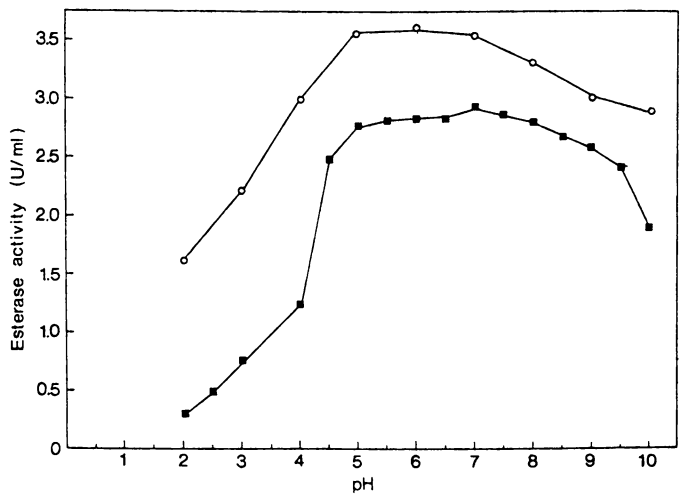


Figure 3. The dependence of cod trypsin stability on pH. Residual activity was determined with TosArgOMe as substrate at pH 8.1 and 25^{0}C after incubation at 5^{0}C for 30 min (circles) or 18 h (dark squares).

relate these observations to an increase in the flexibility of the cod enzyme as compared to the bovine we are currently pursuing studies on the structural dynamics of the two enzymes.

Cod Chymotrypsins.

Isolation and biochemical characterization. Two chymotrypsins were isolated from the pyloric caeca of the Atlantic cod using similar chromatography techniques as exploited in the isolation of the cod trypsins, including affinity, hydrophobic interaction and chromatofocusing chromatography (*8*). These steps yielded two chymotrypsins, A and B, having isoelectric points of 6.2 and 5.8 respectively. The estimated molecular mass of the chymotrypsins was 26 kDa. The N-terminal amino acid sequence of chymotrypsin A is seen in Table IV. The cod sequence is obviously homologous to the bovine chymotrypsin sequence. Furthermore, sequence studies on the non-reduced "native" chymotrypsin B gave an indication that as in the case of bovine chymotrypsin an internal cleavage occurs in the region of amino acid position 15. These data taken in conjunction with SDS-PAGE studies under both reducing and nonreducing conditions suggests that the cod chymotrypsins exist in both a "one chain" and "two chain" form. A similar, although not identical situation is found in the case of bovine chymotrypsin which is often found in either the "two chain" or the ultimate "three-chain" form (*9*).

Activity characteristics. The effect of temperature on the activity of the cod chymotrypsins as compared to bovine chymotrypsin using both ester and amide substrates is shown in Table V. As typically observed for chymotrypsins, the amide substrate gave much lower k_{cat} values, however the K_m values were similar to those found with the ester derivative substrate. It is important to note that the catalytic efficiency (k_{cat}/K_m) of the cod enzymes with the amide derivative substrate at 25°C is approximately two to three fold higher than the bovine enzyme. In fact, all across the temperature range assayed (5-30°C) the cod enzymes were demonstrated to have greater catalytic efficiencies than the bovine enzyme (*8*).

The pH activity profile (not shown) of the cod chymotrypsins shows a typical bell shaped distribution with the optimal pH at 7.8 which is essentially identical to the pH optimum for bovine chymotrypsin (*8,10*).However, the cod chymotrypsins were irreversibly inactivated at pH values less than 5 which is in contrast to the bovine enzyme which is very stable in acidic solutions of pH 3 (*8,10*).

Stability characteristics. As in the case of the cod trypsins, we have investigated the effect of pH and temperature on the stability of cod chymotrypsins as compared to the bovine enzyme. When the cod enzyme was incubated for 60 min at 5°C at various pH values followed by assay for esterase activity at pH 8 the enzyme was seen to be stable over the alkaline pH range but increasingly unstable at pH values less than 5 (Figure 4). When considered together with the pH/activity study mentioned above it appears that the cod enzyme in fact becomes denatured at pH values less than 5 with a concomitant loss of activity, whereas, at pH values greater than 8 there is not an irreversible loss of activity suggesting that the enzyme remains stable at these alkaline pH values.

Table IV. N-terminal sequence of Atlantic cod chymotrypsin compared with bovine α-chymotrypsin where residues 14 and 15 have been removed as the result of autolysis. Underlined residues have not been determined with absolute certainty. The filled circles indicate differences in the two sequences.

	1				5				10				15				20				25				30
Cod	C	G	S	P	A	I	Q	P	V	I	S	G	I	V	N	G	E	E	A	V	P	H	T	W	Y	W	Q	V
Bovine	C	G	V	P	A	I	Q	P	V	L	S	G	L	I	V	N	G	E	E	A	V	P	G	S	W	P	W	Q	V

Table V. Comparison of the effect of temperature on kinetic constants for ester or amide hydrolysis catalysed by chymotrypsin from Atlantic cod or cattle. The BzTyrOEt activity was determined in 10%(v/v) methanol and BzTyr-NH-Np activity in 10%(v/v) Me_2SO

BzTyrOEt	°C	K_m(mM)	k_{cat}(sec^{-1})	k_{cat}/K_m($sec^{-1}mM^{-1}$)
Cod-A	35	0.17	329	1935
Cod-B	35	0.10	276	2760
Bovine	35	0.24	142	591
Cod-A	25	0.14	207	1479
Cod-B	25	0.20	214	1070
Bovine	25	0.27	100	370
Cod-A	10	0.12	117	975
Cod-B	10	0.20	120	600
Bovine	10	0.27	43	159
BzTyr-NH-Np	**°C**	**K_m(mM)**	**k_{cat}(sec^{-1})**	**k_{cat}/K_m($sec^{-1}mM^{-1}$)**
Cod-A	35	0.14	0.82	5.9
Cod-B	35	0.11	0.72	7.2
Bovine	35	0.21	0.41	1.9
Cod-A	25	0.08	0.41	5.1
Cod-B	25	0.08	0.42	5.3
Bovine	25	0.14	0.30	2.2
Cod-A	10	0.09	0.10	1.1
Cod-B	10	0.11	0.10	0.9
Bovine	10	0.11	0.05	0.4

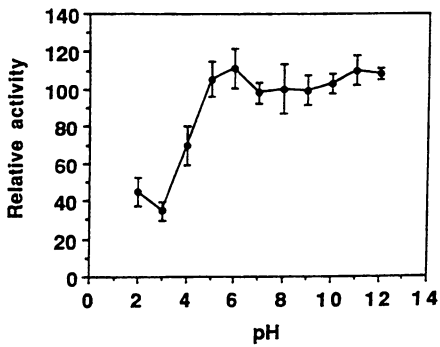


Figure 4. The effect of pH on the stability of Atlantic cod chymotrypsin. The effect of pH on the stability of cod chymotrypsin was estimated by measuring the residual BzTyrOEt activity in 10% (v/v) methanol at 25°C after incubation at 5°C for 60 minutes. The following buffers were used at a final concentration of 0.1M; acetate (pH 2-5); Hepes (pH 6-8) and glycinate (pH 9-12). The calcium concentration was 25mM. Each point is the mean of six determinations from three experiments.

To examine the effect of temperature on the stability of the cod chymotrypsin compared to bovine chymotrypsin the enzymes were incubated for 5 min at increasing temperatures, cooled and then assayed at 25°C for esterase activity. From Figure 5 it can be concluded that the cod enzyme is rather more sensitive to increasing temperature than the bovine enzyme. The loss of half maximal activity occurred at 52°C for the cod enzyme compared to 57°C for the bovine chymotrypsin.

Summary. These studies on the two cod chymotrypsins have corroborated our work with the cod trypsins on certain activity and stability properties of this class of enzymes. The cod chymotrypsins are homologous to the bovine chymotrypsin but not identical. These structural differences are reflected by their differing functional characteristics. The cod chymotrypsins have a similar pH/activity profile as the bovine enzyme but the activity is irreversibly lost at low pH values in the case of the cod enzyme whereas the bovine enzyme remains stable. On amidase substrates the cod enzymes were demonstrated to have a significant increase in catalytic efficiency compared to the bovine enzyme and although maximum rate of catalysis for the cod chymotrypsin is greater than that observed for the bovine enzyme the optimal temperature for catalysis by the cod enzyme is lower. With regard to thermal stability the cod chymotrypsin is less stable than compared to the bovine enzyme

Cod Elastase.

Isolations and biochemical characterization. We have purified an elastase from cod pyloric caeca using phenyl Separose hydrophobic chromatography followed by gel filtration and ion exchange chromatorgraphy (*11*). The enzyme has an apparent molecular mass of 25 kDa based upon SDS-PAGE. The isoelectric point of the protein was estimated to be approximately 9.5. These values are similar, although not identical to another elastase isolated from the Atlantic cod which was characterized to have a molecular weight of 28,000 and an isoelectric point greater than 9.3 (*12*). Whether these two enzymes are actually identical remains to be determined.

Activity characteristics. The effect of pH and temperature on the cod elastase was assayed. In the case of pH and activity the porcine and cod elastases have essentially the same pH optimum, pH 8.2 (results not shown). The comparative effect of temperature on the elastase activity of the two species was quite different both in terms of the position and level of maximal activity. The temperature of maximum activity for the cod elastase was observed at approximately 40°C compared to 50°C as observed for the porcine enzyme (Figure 6).

Stability characteristics. The relative stability of the cod elastase to temperature and pH denaturation has also been examined. In Figure 7 the effect of pH on the stability of the cod elastase at two different incubation at two different incubation temperatures is observed. The results of this experiment were similar to the those of the other cod proteinases. Incubation at either 5°C or 25°C did not significantly affect the pH value below which a loss of activity began (approximately pH 5). The cod elastase, similar to the other cod proteinases, was stable at alkaline pH values.

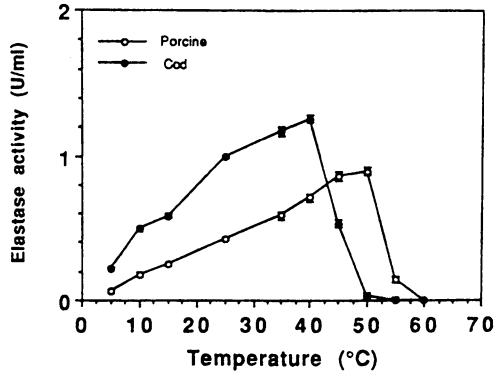


Figure 5. The effect of temperature on the catalytic activity of elastase from Atlantic cod and porcine. The enzyme was in each case added to a preheated medium (0.1 Tris, pH 8.0, containing 1 mM Suc-Ala-Ala-Ala-p-nitroanilide as substrate) and the average enzyme activity determined for the following one minute period.

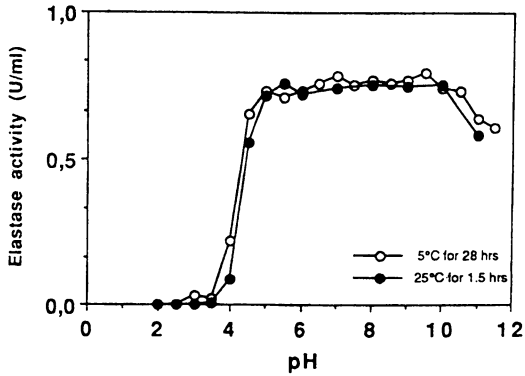


Figure 6. The effect of pH on the stability of elastase from Atlantic cod. The enzyme was incubated in 25 mM sodium phosphate in the pH range of 5.0 to 8.6, and in 25 mM sodium glycine in the pH range of 8.2 to 11.2. After incubation, either at 25^0 C for 1.5 h or at 4^0 C for 28 h, the residual activity was measured at 25^0C with Suc-Ala-Ala-Ala-p-nitroanilide as substrate in 0.1 M Tris at pH 8.0.

In the temperature denaturation experiments the enzymes were incubated for varying lengths of time after which they were cooled and then assayed for elastase activity at 25°C. At 45°C the cod elastase does not maintain full activity longer than 15 seconds whereas the porcine elastase retains full activity for extended periods (data not shown). Comparable to the cod elastase inactivation at 45°C is the inactivation of the porcine enzyme at 55°C; a 10 degree difference.

Summary. As has been demonstrated for the other cod digestive proteinases, the cod elastase also has an overall greater enzymatic activity than the analogous enzyme from a warm-blooded organism however, the temperature of maximal activity is significantly lower than that observed for the porcine elastase. The cod enzyme has a similar stability to pH relationship as that found for the other cod proteinases with these enzymes being relatively stable at alkaline pH values but rapidly and irreversibly losing activity under acidic conditions. Yet with regards to temperature stability, the enzymes from the two species are quite different with the cod elastase being much more sensitive to temperature denaturation.

Biotechnological Applications of Cold Active Proteinases.

As we and others have demonstrated, the cod digestive proteinases all have higher catalytic activities at all temperatures at which they are active, thus permitting the use of less amounts of enzymatic additives necessary for various processes. Associated with the use of enzymes for most industrial processes, particularly in the food industry is the necessity to eliminate the enzymatic activity once the desired results have been attained. Typically this is accomplished by either heat or acid denaturation. For the cod enzymes heat denaturation could be readily performed at lower temperatures than is necessary with many of the enzymes currently in use. This could result in lower energy costs and perhaps higher quality product. Also these cod enzymes appear to be more sensitive to denaturation at acid conditions under which most bovine proteinases remain stable. Mixtures of the digestive enzymes from cod, such as Cryotin contains a different profile of enzymatic activities compared to Pancreatin from bovine pancreas. In the case of Cryotin, the mixture lacks lipase activity which is present in Pancreatin, however, Cryotin has collagenolytic activity, which is not found in Pancreatin. Thus with certain applications the Cryotin may more readily produce the desired result.

Cold-active proteinases, purified or in mixture have many potential uses in industry, medicine and research. The applications in the food processing industrial appear to be especially promising since the ability to inactivate the enzyme adjunct with mild conditions which do not adversely affect the product is critical. The use of these enzymes have proven useful in various fish processing applications such as the skinning of fish, removal of membranes and ripening of herring. These proteinases also have shown potential as digestive aids, both for humans and animals. They are currently undergoing testing as adjuncts in microdiets for fish larvae. In Japan, the pepsin from cod is now in use as an enzyme adjunct in fish feed. Additional food processing applications where use may be found for cold-active enzymes from fish include chill-proofing of beer, biscuit manufacturing, tenderization of meats, hydrolysis of whey, casein and fish proteins,

hydrolysis of gelatins and vegetable protein as well as many other proteins, particularly the collagens.

Future Developments.

Our laboratories are currently near completing the cDNA sequence analysis of the cod proteinases discussed in the proceeding sections. Our next step will be to carefully analyze the amino acid sequence differences between the cod and bovine enzymes with the aim of potentially identifying residues which may play a role in the "destabilization" of the cod enzymes relative to the bovine enzyme giving rise to an overall increase in the catalytic efficiency of the cod enzymes compared to their bovine counterparts. We will use this information in the alteration of the bovine enzyme structure by site specific mutagenesis coupled with appropriate assays to determine whether we have been successful in identifying the important residues in psychrophilic cod enzymes. Ultimately, we hope that these studies will contribute to the general knowledge of protein structure and function and that this information will be of potential use in the engineering of other enzymes which could best be utilized under cold temperature applications.

Acknowledgments

We would like to recognize support for this research from the Nordisk Industrifond (JBB), The Icelandic Science Foundation (JBB) and the National Science Foundation (JWF and JBB).

Literature Cited

1. Hochachka, P.W. ; Somero, G.N. *Biochemical Adaptation*; Princeton University Press: Princeton, MA, 1984, 3-15.
2. Owen, T.G.; Wiggs, A.J. *Comp. Biochem.Physiol.* **1971**, *40B*, 465-473.
3. Low, P.S.; Somero, G.N. *Comp. Biochem. Physiol.* **1974**, *49B*, 307-312.
4. Asgeirsson, B.; Fox, J.W.; Bjarnason, J.B. *Eur. J. Biochem.* **1989**, *180*, 85.
5. Lazdunski, M.; Delaage, M. *Biochim. Biophys. Acta*, **1967**, *140*, 417-434.
6. Hjelmeland, K.; Raa, J. *Comp. Biochem. Physiol.* **1982**, *71B*, 557-562.
7. Simpson, B.K.; Haard, N.F. *Can. J. Biochem. Cell Biol.* **1984**, *62*, 894-900.
8. Asgeirsson, B.; Bjarnason, J.B. *Comp. Biochem. Physiol.* **1991**, *99B*, 327.
9. Bender, M.L.; Killheffer, J.V. *CRC Critical Reviews in Biochemistry*, April 1973.
10. Wilcox, P.E . In *Methods in Enzymology*; Colowick, S.P. and Kaplan, N.O., Eds. 1970, *19*, 64.
11. Bjarnason, J.B.; Mantyla, E.O.; Asgeirsson, B . *Sarsia*, **1992**, in press.
12. Gildberg, A.; Overbo, K. *Comp. Biochem. Physiol.*, **1990**, *97B*, 775.

RECEIVED September 10, 1992

Chapter 6

Thermal and pH Stress in Thermal Denaturation of *Trichoderma reesei* Cellobiohydrolase I

Supporting Evidence for a Two-Transition Model

John O. Baker and Michael E. Himmel

Applied Biological Sciences Branch, Alternative Fuels Division, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO 80401

The structure and thermal denaturation of *Trichoderma reesei* cellobiohydrolase I (CBH I) have been investigated using fluorescence, chemical modification, and differential scanning calorimetry (DSC) techniques. The results of both fluorescence-quenching with cesium ion and chemical modification with N-bromosuccinimide indicate that at least seven, and possibly eight, of the nine tryptophan residues in the CBH I catalytic core region are in "exposed" positions at or near the surface of the native molecule. A biphasic perturbation of the CBH I intrinsic fluorescence reveals that the CBH I core region is capable of binding more than one molecule of cellobiose and suggests that this additional bound molecule may be important for the stabilization of the core region against thermal denaturation. When the temperature of a solution of CBH I (pH 7.5 in 50 mM phosphate) is ramped through its denaturation zone (approximately 28°C - 48°C at this pH), a sharp, sigmoidal change, centered at approximately 36°C, is observed in the polarization of the tryptophan fluorescence of the protein. This polarization change precedes both the endothermic peak maximum (40.15°C) observed in DSC under the same conditions and the second (40.3°C) and larger of two component peaks invoked to explain the asymmetrical shape of the DSC peak. The midpoint of the fluorescence polarization change is much more closely correlated with the first (37.2°C) and smaller of the deconvoluted component peaks. The fluorescence-polarization data thus provide supporting evidence for the component transitions, the existence of which has heretofore rested only on mathematical inference, and thereby for the two-transition model proposed earlier for the denaturation.

0097–6156/93/0516–0083$06.00/0

Cellulose-depolymerizing enzymes have been of considerable industrial interest for years because of their utility in the conversion of cellulosic materials (agricultural residues, paper waste, or crops specifically grown for the purpose) into sugars fermentable into fuel ethanol (*1,2*). In recent years major advances have been made in determining the structures of cellulolytic enzymes and then relating these structures to the observed activities. What might be described as the first benchmark in this process of discovery was drawn from comparisons between the primary sequence of representative enzymes (*3*) and then from small-angle X-ray scattering (SAXS) studies of the overall shape of the enzymes (*4,5*). This benchmark consisted of the realization that cellulolytic enzymes in general, whether exo- or endo-acting, and whether from fungal or bacterial sources, tend to have a bilobed, two-domain structure, with one domain capable of binding tightly to crystalline cellulose, and the other containing the hydrolytic active site (*6-9*).

The second level of understanding followed upon the determination of the three-dimensional structure of a synthetic cellulose-binding domain (CBD) by solution nuclear magnetic resonance (NMR) (*10*) and the crystallization of the proteolytically separated catalytic domain of *T. reesei* cellobiohydrolase II (CBH II) (*11*), which made possible the determination of the tertiary structure of the catalytic domain and the modeling of the interactions between the catalytic domain and substrates and inhibitors (*12*). The most striking revelation provided by these studies of three-dimensional structure is that the actual bond-cleaving site of *T. reesei* CBH II is located in a closed tunnel through the catalytic domain (*12*). This has led to the suggestion that the "exo"-type cellulose depolymerases attack crystalline cellulose by a mechanism in which the cellulose-binding domain of the enzyme molecule pries the end of a cellulose chain away from the crystalline surface, and the chain is then fed across the CBD into the tunnel in the catalytic domain, where successive cellobiosyl fragments are cleaved from the chain (*8*) (and presumably ejected through the far end of the tunnel). In this model for substrate hydrolysis, the enzyme may be considered to hitch its way along the polyglucosyl chain, removing successive cellobiosyl fragments without diffusing away from the chain between hydrolytic events.

Despite the dramatic nature and extremely high value of the three-dimensional protein structures obtained by X-ray crystallography and the molecular-modeling studies they make possible, the existence of such information does not remove the rationale for other approaches, such as circular dichroism (CD), DSC, and fluorescence studies, that are also aimed at revealing the relationship between the structure of enzymes on the one hand and their activities and stabilities on the other. In the most trivially obvious case that could be cited, such alternate methods of structural analysis are indisputably of value for proteins for which detailed three-dimensional structures are not yet available (which happens to be the case, at this writing, for *T. reesei* CBH I, the enzyme with which the present paper is concerned). Even in the case of proteins for which high-resolution X-ray diffraction structures are available, the alternate methods of structural analysis differ from X-ray crystallography in that instead of providing an essentially static picture of the

conformation in which the protein happens to crystallize, CD, DSC, and fluorescence are capable of analyzing other forms that may exist in solution, and of monitoring the transitions between forms; i.e., they are capable of providing a dynamic rather than static picture. The types of information provided by the different methods of structural analysis are therefore complementary, rather than redundant.

In an earlier study using DSC to analyze the thermal denaturation process for *Trichoderma* CBH I, we found indications (based on the asymmetry of the overall endothermic peaks) that the unfolding proceeded through a mechanism involving at least two distinct structural transitions (*13*). We proposed that the two transitions are tightly linked, one occurring closely upon the heels of the other, even when the temperature at which the overall denaturation occurs is varied over the range from approximately 64°C down to approximately 33°C, by varying the pH from 4.8 to 8.3 (Figure 1). The overall denaturation process was strongly affected by the presence of the competitive product inhibitor, cellobiose. Cellobiose at 100 mM, a saturating concentration some 5000 times the binding constant for the inhibitor as measured by inhibition of the enzyme activity, increased the denaturation temperature of CBH I by some 8°C at pH 4.8, at which pH the enzyme is at near-maximal stability in the absence of cellobiose, and by almost 19°C under conditions of pH stress at pH 8.34. It appeared that cellobiose should be a useful probe for the analysis of protein structural changes affecting the integrity of the active site.

In the present report, we describe results of a study in which the polarization of the fluorescence emission of the tryptophan residues of the catalytic core region of CBH I was used to monitor changes in the mobility of these residues, and therefore in the state of folding of the core region. These most recent results are complementary to the previous DSC results and appear to provide support for the two-transition model proposed for the thermal unfolding of CBH I core region on the basis of the DSC results.

Materials and Methods

CBH I Purification. The purification procedure used in this study was that developed earlier (*13*) following the general size exclusion chromatography/anion exchange chromatography protocol described by Shoemaker for the purification of CBH I (*14*).

Differential Scanning Microcalorimetry. Denaturation thermograms were obtained using a Microcal MC-2 Scanning Calorimeter (Microcal, Northampton, MA), interfaced though a DT 2801 A/D converter to an IBM PC-XT microcomputer. Instrument control and data acquisition were by means of the DA-2 software package (Microcal). The sample cell capacity is 1.130 mL, and runs were made with an overpressure of 30 psig (N_2), at scan rates of 0.5 deg/min and 0.21 deg/min. Protein samples for DSC typically contained 1.0 mg/mL protein in 50 mM sodium phosphate buffer, pH 7.5. The protein samples, along with reference

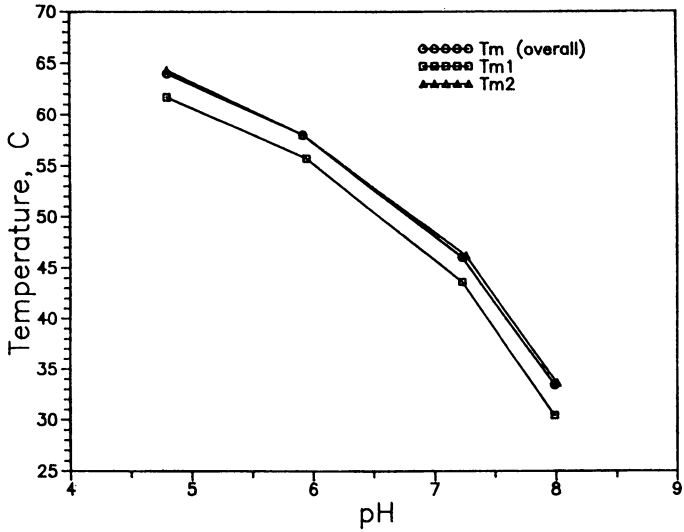


Figure 1. pH-dependence of CBH I structural stability. Ordinate values plotted are from deconvolution analysis of DSC thermograms (ref. *13*). $T_{m(overall)}$ represents the temperature of maximum differential power input for the overall process as observed; T_{m1} and T_{m2} represent the temperatures of the two deconvoluted component transitions invoked to explain the shape of the overall peak. The pH values as plotted have been corrected, at the corresponding temperatures, for the temperature-dependence of the pK_a values of the buffers employed.

samples consisting of the same buffer system without protein, were equilibrated for 1 h at room temperature and atmospheric pressure, with gentle stirring, before loading into the sample and reference cells, respectively. After filling of the calorimeter cells, the protein concentration of the sample as loaded was determined spectrophotometrically, using the remainder of the solution used to load the sample cell. The extinction coefficient used for CBH I was 1.42 $(g/L)^{-1}$ at 280 nm (*15*).

Deconvolution Analysis of DSC. Data analysis was carried out using the DECONV section of the DA-2 software package. This software, which is based on the deconvolution procedure of Biltonen and Freire (*16*), allows deconvolution of differential heat-capacity peaks either as the result of simple addition of multiple independent transitions or as the result of more complex mathematical processes representing the combination of transitions that interact in such a way that an obligatory reaction sequence is imposed (sequential transitions).

Each thermogram was normalized on scan rate, the corresponding (scan-rate-normalized) buffer-buffer baseline was subtracted, and the differential heat capacity values were divided by the number of moles of protein in the sample, to yield ordinate values in terms of calories $mol^{-1}deg^{-1}$. The resulting files were then analyzed using the deconvolution software.

DSC Nomenclature. The term T_m as used here is not a "melting temperature," but is instead standard DSC usage for the temperature at which the maximum differential heat capacity is observed during the denaturation of a protein sample. ("T_{max}" would be a more extended expression of this term.) For a two-state (single-transition) process, T_m approximates, but is not identical to, the value of T_d or $T_{1/2}$, the temperature at which the transition is half-complete. Detailed discussions of these terms can be found in references (*17-21*).

Fluorescence Measurements. Fluorescence measurements were made with a FLUOROLOG 2 spectrofluorometer (Spex Industries, Inc., Edison, NJ) fitted, when appropriate, with Glan-Thompson excitation and emission polarizers (Spex Model 1935B manual L-format polarization kit). Temperature control of the sample was by water-jacketed cuvette holder with an external circulating water bath; sample temperature was monitored by means of a thermocouple immersed in the stirred sample. Data collection and analysis were carried out with a Spex DM1B Spectroscopy Laboratory Coordinator interfaced to the spectrofluorometer.

Enzyme Kinetic Measurements. The glycohydrolytic activity of CBH I was measured using the fluorogenic substrate 4-methylumbelliferyl-β-D-cellobioside (Sigma Chemical Company, St. Louis, MO). Enzyme and substrate were incubated at pH 7.5 and the temperature of interest in a stirred cuvette in the sample compartment of the spectrofluorometer, and the generation of the aglycone product, 4-methylumbelliferone, was continuously monitored by the change in the intensity of the fluorescence of its anionic form. The excitation wavelength used was 380 nm (bandpass 4.5 nm) and the emission was measured at 455 nm (bandpass 4.5 nm). The extent of hydrolysis was quantitated by means of a standard fluorescence

curve for 4-methylumbelliferone, run at the same session, with the same temperature, buffer, and instrumental settings as those used for the enzyme activity measurements. The enzyme concentration was chosen for each set of experiments so that less than 2% (and usually less than 1%) of the substrate was hydrolyzed in a 10-minute assay.

Results and Discussion

T. reesei CBH I contains nine tryptophan residues (*22,23*), all of which are located in the ellipsoidal (*4,5*) catalytic core region of the molecule — none in the cellulose-binding and linker regions. The tryptophan fluorescence of the molecule may therefore be taken as a useful index of changes in the physico-chemical state of the core region alone. (The "tail" of the tadpole-shaped molecule is known from SAXS studies to be in a relatively extended average position in solution (*4,5*), and therefore should not significantly affect the fluorescence of even the surface residues in the core region.)

In the interpretation of changes in the tryptophan fluorescence of the core in terms of changes in the structure of that region, it is advantageous to have as much information as possible about the locations of the tryptophan residues in the native conformation of the protein molecule. To this end, the present study was opened with fluorescence-quenching and chemical-modification experiments designed to estimate the degree of exposure of the tryptophan residues to the external (aqueous) environment at 25°C and pH 4.8, under which conditions the stability of the native form is near-maximal. In a previous study, changes in the wavelength of the maximum emission intensity were used as a measure of changes in the average polarity of the microenvironments of the tryptophan residues, and presumably, therefore, of changes in their exposure to the aqueous external environment during thermal denaturation at pH 4.8 (thermal-scanning fluorescence studies) (*24*). The low-temperature portion of the scans in this previous study revealed a wavelength of maximum emission of approximately 355 nm, indicating a relatively polar (and presumably exposed) average microenvironment for the tryptophan residues in the native molecule (*25*).

Results obtained in the present study bear out this earlier inference, in addition to providing additional information about the distribution of environments among the nine tryptophan residues contributing to the overall fluorescence. Figure 2 illustrates the quenching of the intrinsic fluorescence of CBH I by cesium ion, which, as a charged quencher, is expected to quench exposed residues, but is not expected to penetrate readily into the hydrophobic interior of proteins (*25*). This Stern-Volmer plot (*25*) indicates that a substantial portion (probably well over half) of the tryptophan fluorescence of the core region is quenchable by cesium ion. A comparison of the slopes of the curves in Figure 2 indicates that the tryptophan residues being quenched are less accessible to the quencher than is the small model compound N-acetyl-L-tryptophanamide, as would be expected if the indole side-chains were partly imbedded in the surface of the protein, but by no means

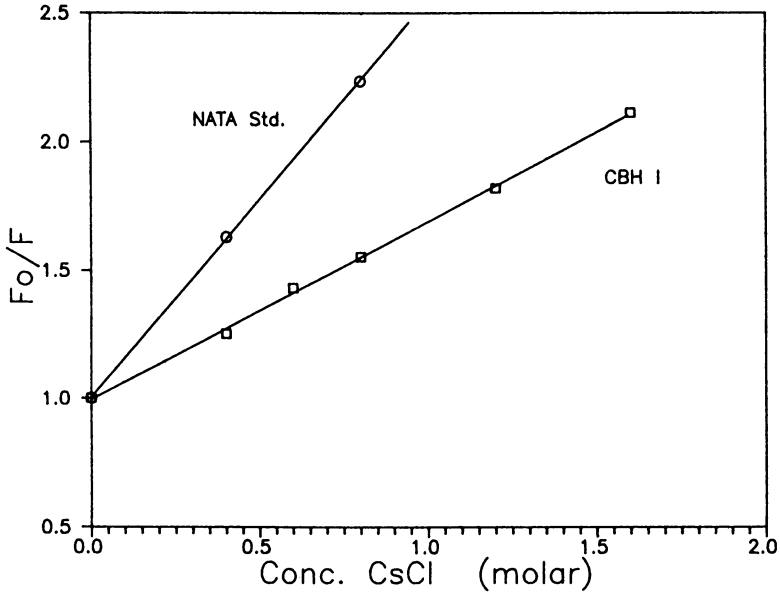


Figure 2. Quenching by cesium ion of the intrinsic fluorescence of native-conformation CBH I at pH 4.8 (50 mM acetate) and 25°C. Excitation at 305 nm (0.9-nm bandpass); emission measured at 353 nm for CBH I and at 367 nm for N-acetyl-L-tryptophanamide (NATA), with an emission bandpass of 9.0 nm in both cases. Quenching of the model compound NATA is presented to allow comparison of slopes.

completely "buried." From the reasonable linearity of the plot for CBH I, we can also infer that the residues being quenched over this range of quencher concentration do not differ drastically from each other in degree of exposure.

The reciprocal of the ordinate intercept in the "modified Stern-Volmer plot" (*25*) of the same data (Figure 3) is an estimator of the fraction of the total protein fluorescence that is susceptible to quenching by cesium ion. The least-squares-fit value for the intercept might seem to indicate that all of the residues are accessible, but in view of the fact that there is some noise in the data, it is more appropriate to conclude from these results only that no more than 10% of the total fluorescence, if that much, is inaccessible to the quencher. Quantitative translation of the fraction of total fluorescence that is accessible to quencher into the fraction of tryptophan residues exposed to solvent is rendered approximate at best by the recognition that different tryptophan residues may have different quantum yields. If we assume, however, that the differences in quantum yield are not large, it seems reasonable to adopt the tentative conclusion that of the nine tryptophan residues, probably no more than one, and certainly no more than two, are "completely buried."

Figures 4 and 5 present a chemical modification approach to the estimation of the exposure of CBH I tryptophan residues to the solution. Both figures describe the titration of CBH I tryptophan side-chains with the reagent NBS, which oxidizes the indole ring to the non-fluorescent oxyindole derivative (*26*). The successive emission spectra recorded during the titration (Figure 4) show that from the beginning, the removal of the contributions from the residues more accessible to NBS results in a blue-shift of the emission spectrum away from the original maximum at 354-355 nm, indicating that the surviving fluorescent residues have an average microenvironment less polar (i.e., less exposed to solvent) than the original average of the microenvironments. The blue-shift is relatively small, however, until more than 60% of the fluorescence has been destroyed. After that the wavelength of maximum fluorescence shifts more rapidly, until it finally slows again near 337 nm, at which point the intensity of the fluorescence (at 337 nm) is approximately 12 % of the intensity originally measured at 354-355 nm.

The peak intensities measured after successive additions of NBS are plotted in Figure 5, as a function of the cumulative molar ratio of added NBS to CBH I. It should be noted that because of competing reactions (notably hydrolysis) that consume reagent, the stoichiometry of NBS added, to tryptophan residues oxidized, is not expected to be exactly 1:1. Under the conditions of our CBH I titration, we have determined that the stoichiometry for the "completely exposed" model compound NATA is approximately 1.2 mol NBS per mol NATA. For somewhat protected (and therefore slower-reacting) tryptophan residues partially buried in proteins, the partitioning of reagent toward solvent rather than tryptophan residues can be expected to occur to greater extents. For this reason, the important quantities to be considered in Figure 7 are the relative values of the slope of the curve at different levels of remaining fluorescence. The shape of the curve can accordingly be rationalized as representing the fairly rapid titration of residues contributing almost 90% of the original intensity, and not differing drastically from each other in accessibility to the reagent, followed by the much slower attack on the residue(s) responsible for the remaining 10%-11% of the intensity. Translation of this

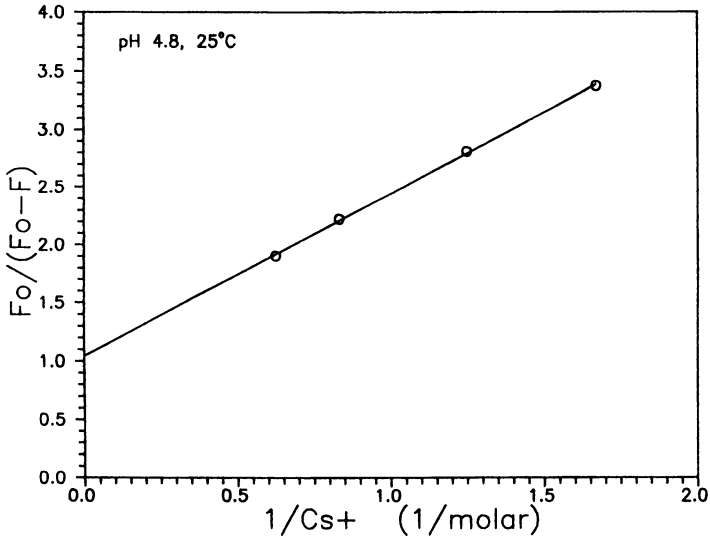


Figure 3. Quenching by cesium ion of the intrinsic fluorescence of native-conformation CBH I. "Modified Stern-Volmer plot" (ref. *25*) of data from Figure 2, in which the ordinate intercept represents the reciprocal of the fraction of total fluorescence accessible to the quencher.

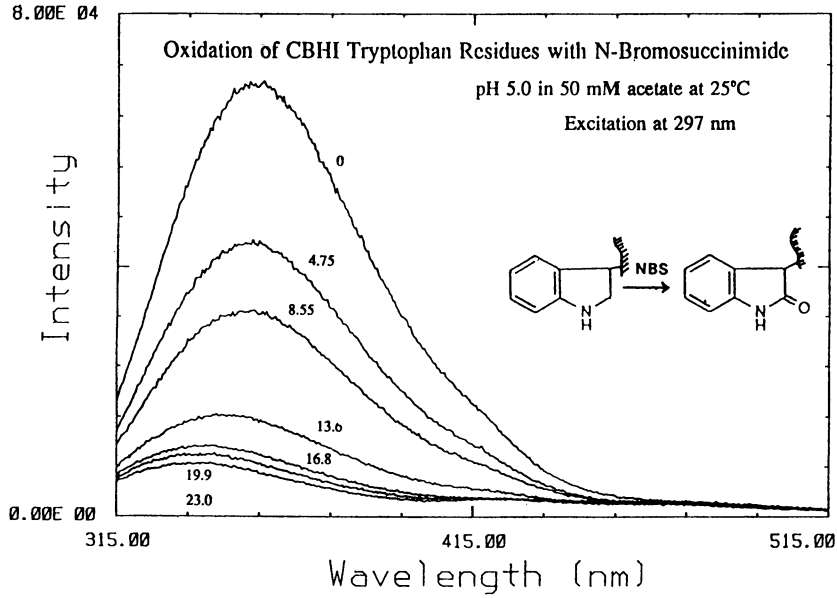


Figure 4. Titration of solvent-accessible residues of CBH I with N-bromosuccinimide. Numbers adjacent to the spectra denote the cumulative molar ratios of NBS to CBH I applied prior to each spectral scan.

description into the average number of residues oxidized per protein molecule at different points of the titration can be only semi-quantitative at best, because (1) for a given excitation intensity, the quantum efficiency of tryptophan residues can vary with the microenvironment (*25*), and (2) the arbitrary selection of the (changing) wavelength of maximum intensity for measurement of the emission will underestimate the contribution of those residues with emission maxima significantly removed from the wavelength of maximum intensity for the overall envelope. These disclaimers notwithstanding, we feel that the data shown in Figures 4 and 5 strongly suggest, at least, that of the nine tryptophan residues of the catalytic core region of CBH I, seven, or perhaps eight, are "surface" residues, relatively exposed to the solution, and one residue is "buried."

Binding of Cellobiose to CBH I Monitored by Perturbation of the Protein Fluorescence. Cellobiose is a potent competitive product inhibitor of the hydrolysis of glycosidic substrates by CBH I, with a K_i value of approximately 20 micromolar at pH 5.0, 25°C (*27*). In the presence of 10 - 100 millimolar cellobiose, CBH I has been shown to be significantly stabilized against thermal denaturation, particularly at pH values at which the enzyme stability, in the absence of cellobiose, is substantially decreased with respect to its maximum stability (*13*). Although the stabilization by cellobiose develops over a range of cellobiose concentrations far higher than the K_i values for inhibition by cellobiose of the hydrolysis of the small substrate 4-methylumbelliferyl-β-D-cellobioside (K_i is 18-22 micromolar at pH 4.8, 25°C, 35-40 micromolar at pH 7.5, 35°C), it is possible to rationalize the stabilization in terms of the micromolar-level interaction that is reflected in the inhibition of the enzyme activity. The stabilization was so rationalized in our previous publication on the subject (*13*).

Recent fluorescence data, however, suggest the possibility that additional interactions may occur between cellobiose and the enzyme at higher cellobiose concentrations, and that these additional interactions may play an important role in the observed stabilization. Figure 6(a-c) illustrates the effect of increasing concentrations of cellobiose on the fluorescence emission spectrum of CBH I at pH 7.5 and 35°C. Increasing the cellobiose concentration from 0 to 40 mM (Figure 6a) brings about a biphasic series of changes in the emission spectrum, the biphasic nature of which may be seen more clearly in Figures 6b and 6c, in which the spectra corresponding to individual cellobiose concentrations have been separated out into low-concentration and high-concentration ranges. In Figure 6b, increasing the concentration of cellobiose from 0 to 5 mM produces primarily an increase in the height of the emission peak, with very little shift in the wavelength of maximum emission as marked by the arrows (although we do begin to see some blue-shifting by the time the cellobiose concentration has increased to 5 mM). In Figure 6c, which covers the range from 5 mM to 40 mM cellobiose, the effect seen is almost entirely a blue-shift, as indicated by the short vertical lines marking the peak positions, with no measurable change in peak height. The short vertical lines represent peak wavelengths for cellobiose concentrations increasing from right to left, with the "zero cellobiose" peak position shown on the right for reference. In Figure 6c the spectral curves are in the same order (concentration increasing from

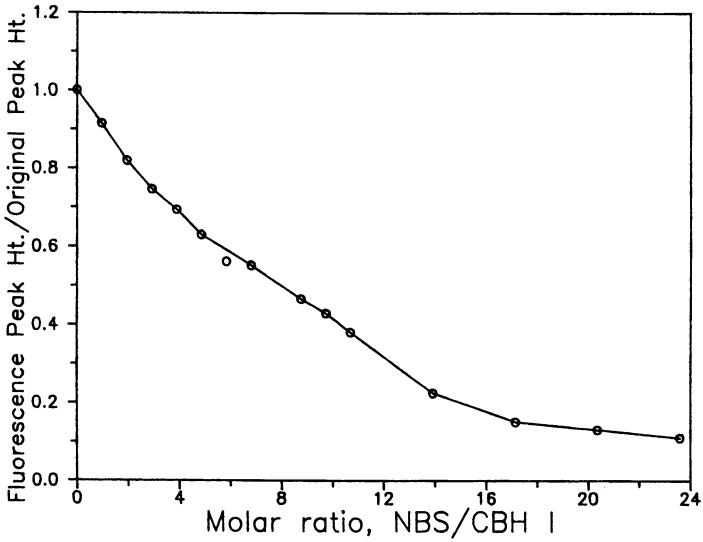


Figure 5. Titration of CBH I tryptophan residues with N-bromosuccinimide. Fluorescence emission peak height as a function of NBS/CBH I ratio for the titration shown in Figure 4. (See text).

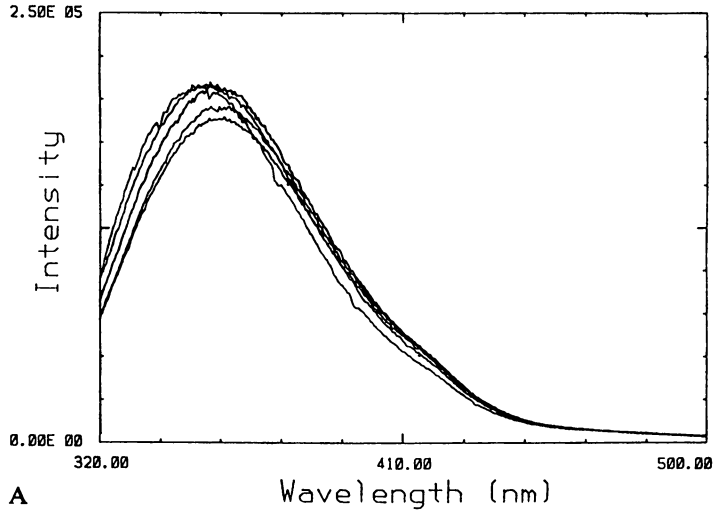


Figure 6. Biphasic effect of cellobiose on the emission spectrum of CBH I at pH 7.5 (50 mM phosphate), 35°C. A: Wide-range composite plot of spectra in the presence of cellobiose concentrations from 0 to 40 mM.

Continued on next page.

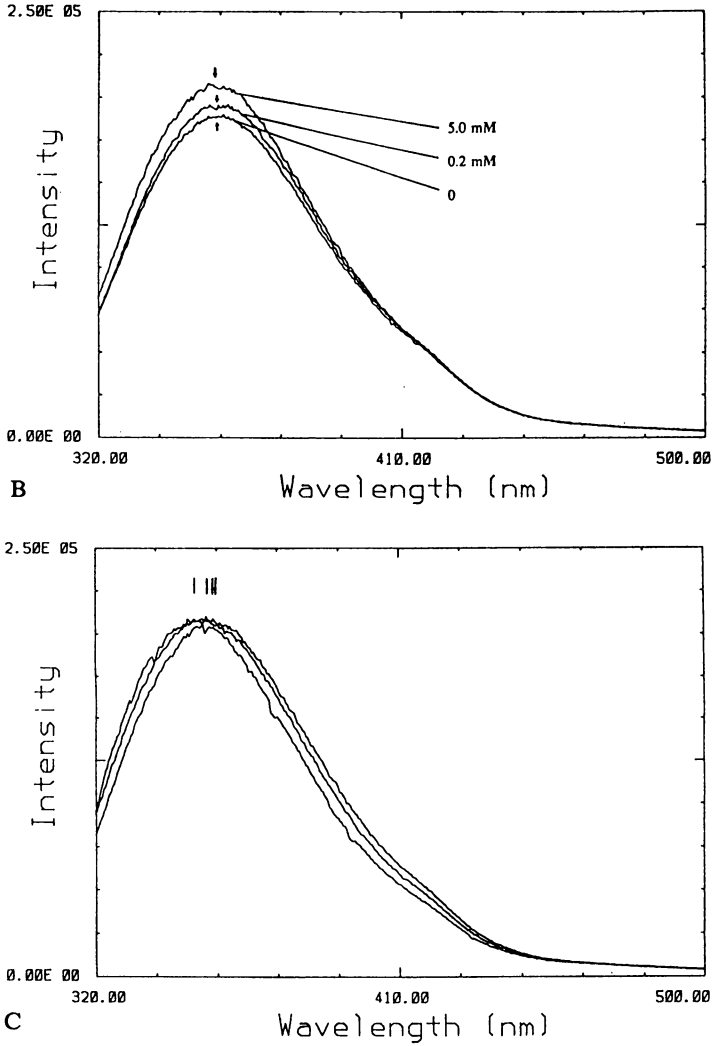


Figure 6. Continued. B: Spectra at a low range of cellobiose concentrations, showing primarily an amplitude effect. (Emission maxima represented by arrows.) C: Spectra at higher cellobiose concentrations, showing a blue-shift of the wavelength of maximum emission with increasing cellobiose concentrations. Peak maxima are indicated by the short vertical line segments above the spectra and represent (from right to left) wavelengths of emission maxima in the presence of 0, 5, 10, and 40 mM celliobose.

right to left) on both sides of the peak maxima (i.e., what is observed here is primarily the result of a simple wavelength shift, not a significant narrowing or broadening of the peak). Similar biphasic fluorescence-perturbation patterns have been obtained at pH 7.5, 25°C and at pH 4.8, 25°C; the data set at pH 7.5, 35°C was chosen for illustration because these conditions fall within the range over which the calorimetrically monitored (DSC) and fluorescence-monitored structural changes to be shown below, and the protection by cellobiose against these changes, are shown to occur. It is significant to point out that under the sets of conditions referred to above, the presence of sucrose in concentrations as high as 40 mM has no detectable effect whatsoever on the emission spectrum of CBH I (data not shown).

The above indications that more than one cellobiose molecule may bind to the CBH I molecule are not particularly surprising in view of the fact that this enzyme binds to and hydrolyzes a homopolymeric substrate and may therefore be expected to have extended binding sites both in the catalytic and the cellulose-binding domains. Rouvinen et al. (*12*), on the basis of X-ray crystallographic and molecular-modeling studies, have identified four clear binding sites for glucosyl residues along the catalytic tunnel through the CBH II core region. If a tunnel of this sort is a general feature of exo-cellulases, as suggested by Rouvinen et al. (*12*), it is not difficult to imagine two cellobiose molecules fitting snugly into a catalytic tunnel in this molecule, occupying the same glucosyl binding sites that would be filled by a cellotetraosyl moiety of a cellulose substrate chain (albeit, perhaps, with some strain between the glucosyl residues in the second and third binding sites). The result may be a substantial bracing of the core region, which would now be effectively a solid molecule.

Actual determination of the specific nature of the different protein/cellobiose interactions indicated here would require data of a more definitive type than that presented here. The conclusions important for the present study are (1) that there are at least two different interactions between cellobiose and CBH I, as indicated by the qualitatively different perturbations of the protein fluorescence in different ranges of cellobiose concentration; and (2) that at least one of these interactions develops over the same range of cellobiose concentrations over which increasing cellobiose concentrations have been shown by DSC (*13*) to stabilize CBH I progressively against major, thermally induced structural changes.

Thermal Denaturation of CBH I as Monitored by the Polarization of Tryptophan Fluorescence Emission. If the locations of most of the tryptophan residues in the core region are as suggested above on the basis of quenching and chemical modification data (i.e., imbedded in the protein surface but with at least one edge of the indole ring effectively exposed to the solution) then one of the most common approaches to using fluorescence to monitor conformational changes, the interpretation of shifts in the wavelength of maximum emission in terms of changing polarity of fluorophore microenvironments upon transfer from hydrophobic protein interior to aqueous solution (*25*), would be expected to be relatively insensitive to the conformational change in this case. The relatively polar environments of the tryptophan side-chains in the native protein simply would not leave much room for large increases in microenvironmental polarity upon unfolding. Indeed, this is what

was found in our earlier study of the thermal unfolding of CBH I at pH 4.8; over the temperature range from 55°C to 70°C, where it had been found by DSC that major calorimetrically detectable structural changes occurred, changes in the wavelength of maximum emission did not rise significantly above the noise level (*24*).

Given these circumstances, it appeared that fluorescence polarization measurements, using the depolarization of the tryptophan fluorescence emission as a measure of the mobility of individual tryptophan residues, might provide a more effective means of monitoring thermally-induced conformational changes in the CBH I core region. The results in Figure 7 show that this proved to be the case. In this wide-range scan, the distinctive feature of the data taken at pH 7.5 is the sigmoid portion of the curve between approximately 25°C and 43°C. The sigmoid portion is overlaid on a broader pattern of general decrease of fluorescence polarization with increasing temperature. Included in Figure 7 for reference purposes is a plot of fluorescence polarization as a function of temperature at pH 5.0, at which pH the enzyme is much more thermally stable than at pH 7.5, DSC studies having shown that it remains essentially native in conformation up to at least 54°-55°C (Figure 1 and reference *13*). The pH 5.0 curve therefore serves, at least up to 55°C, as a measure of that portion of the total polarization decrease (in both curves) that may be ascribed to the net effect of increasingly rapid tumbling of a relatively intact protein molecule with increasing temperature and the temperature-dependence of fluorescence lifetimes for tryptophans in the basically intact protein molecule. The additional (sigmoid) decrease in polarization observed at pH 7.5 over the range 25°-43°C can be ascribed to physical changes associated with unfolding, such as the more rapid rotations of individual tryptophan side-chains once they are freed from the protein structure, effects on lifetimes resulting from exposure of residues to solvent, and possible altered rotational rates for the entire molecule reflecting size changes upon unfolding. (The data point that is off the pH 5.0 curve at 61°C is not the result of experimental noise. By this temperature the unfolding has proceeded to a significant extent, and at this pH at least one unfolded form of the enzyme has a strong tendency to aggregate into larger particles (*13*), resulting in a net decrease in tryptophan mobility relative to the trend shown at lower temperatures. Because of these two changes — unfolding and aggregation — the pH 5.0 curve therefore serves as a reference curve for an intact molecule only up to approximately 55°-56°C.)

In the same type of pattern observed when the unfolding of CBH I is monitored calorimetrically (*13*), increasing concentrations of cellobiose are found to result in progressively greater thermal stabilization of CBH I, shifting the midpoint of the polarization-detected sigmoid transition to higher temperatures, as shown in Figure 8.

Correlations Between DSC-monitored and Fluorescence-polarization-monitored Thermal Unfolding. In Figure 9, the results of a fluorescence-polarization thermal scan of CBH I are co-plotted with a DSC thermogram of the enzyme under exactly the same solution conditions. For this plot, the values for the fluorescence polarization at pH 7.5 have been taken relative to the values interpolated at the same

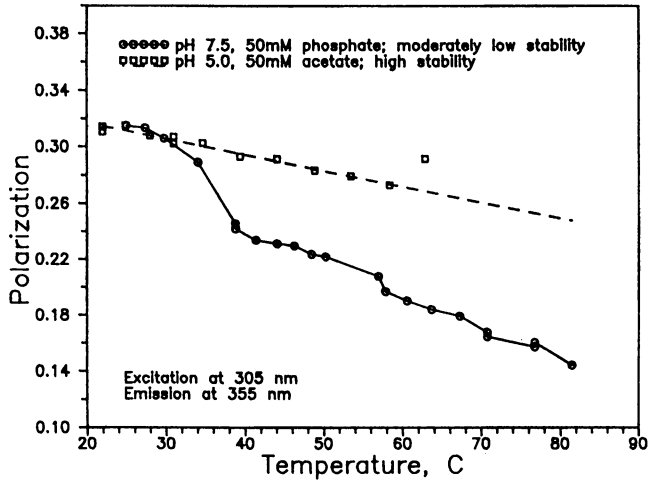


Figure 7. Temperature-dependence of CBH I fluorescence polarization.

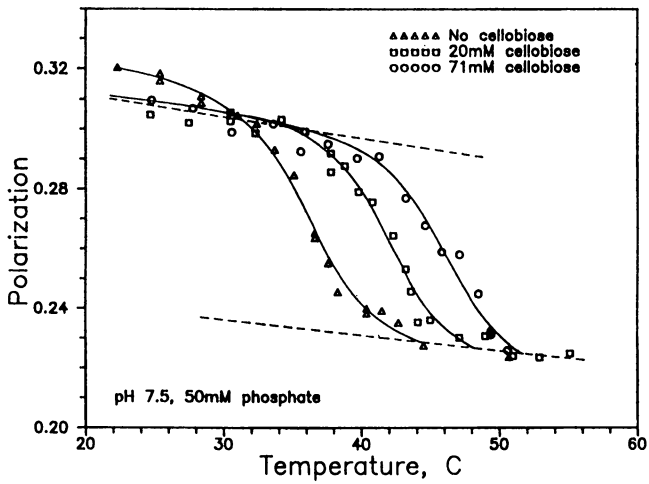


Figure 8. Effect of increasing concentrations of a competitive product inhibitor (cellobiose) on the thermally-induced change in fluorescence polarization.

temperatures from the straight-line approximation to the pH 5.0 data (Figure 7). The temperature-dependent polarization changes shown should therefore be related directly to the unfolding process itself, as explained in the preceding section

The solid line through the data points of the DSC thermogram is the computer-generated two-transition, sequential-model best fit to the data. The peaks shown as dashed lines represent the two mathematically inferred component peaks used to fit the computer model to the observed asymmetrical overall calorimetric envelope. The overall DSC envelope is dominated by the larger of the two deconvoluted component peaks, as shown by the near-coincidence of the maxima for the larger component and the overall envelope (T_{m2} = 40.30°C and $T_{m,overall}$ = 40.15°C). The midpoint of the sigmoid fluorescence-polarization transition at 36.1°C, however, is much more closely correlated with the first (smaller and lower-temperature) of the deconvoluted DSC component transitions (T_{m1} = 37.2°C). An even more telling observation is that by the time the temperature increases to 40°C, the temperature at which both the larger component peak and the overall envelope are nearing their peak differential power inputs (maximum rate of thermally detected structure-breakage), the fluorescence-detected transition is virtually complete.

The temperature-offset between the fluorescence-detected transition and the major calorimetrically detected transition is not the result of a scan-rate effect. The unfolding process is more than 70% reversible under these conditions, and DSC runs at both 0.5°C/min (Figure 8) and at 0.21°C/min (which latter rate is the slowest of which the calorimeter is capable, and is comparable to the "scan rate" used in the fluorescence-polarization studies) showed virtually identical unfolding temperatures.

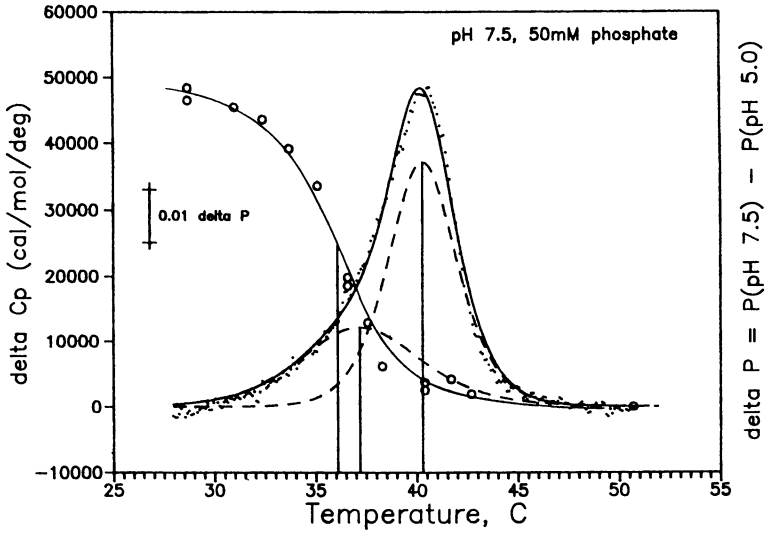


Figure 9. Co-plot of the calorimetrically monitored and fluorescence-polarization-monitored conformational changes as the temperature of a CBH I solution is ramped through the region of denaturation.

Until the present study, the existence of the two component transitions proposed for the thermal denaturation of CBH I, although indicated consistently by deconvolution procedures applied to DSC thermograms obtained under a wide variety of experimental conditions (*13*), has depended solely upon mathematical inference. For the first time, the present results suggest a method of monitoring one transition independently of the other, by dissecting the overall envelope into one transition that can be monitored by fluorescence polarization techniques, and another transition that is silent in terms of these techniques. Given (Figure 10) the known distribution of the nine tryptophan residues of the CBH I core region along the sequence (fairly evenly distributed along the sequence, not bunched at one end) (*22,23*) and the proposed disulfide-bonding pattern (two disulfide-bonded regions, one including four tryptophan residues, the other including the remaining five) (*28*), the construction of a three-dimensional picture to explain the existence of the "silent" transition promises to provide challenging points of application for this and other fluorescence and calorimetric techniques.

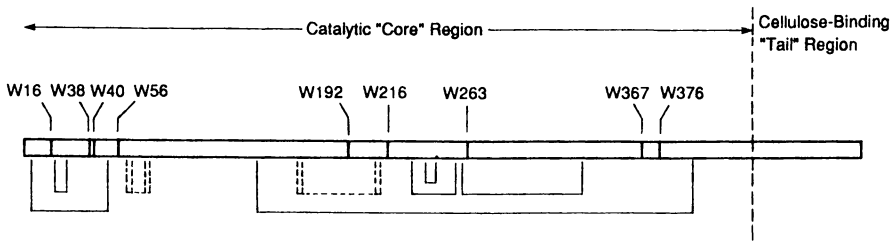


Figure 10. Schematic representation of the distribution of tryptophan residues and disulfide-linked regions along the sequence of the CBH I catalytic core region. The positions of the tryptophan residues (references *22* and *23*) are represented by heavy vertical lines across the horizontal bar representing the peptide chain. The rectangular loops below the peptide chain represent disulfide cross-links as described by Bhikhabhai and Pettersson (ref. *28*), with solid loops representing disulfide bridges determined experimentally, and the dashed loops representing bridges deduced in ref. *28*.

Acknowledgments

This work was funded by the Ethanol from Biomass Program of the Biofuels Systems Division of the U.S. Department of Energy.

Literature Cited

1. Lynd, L.R.; Cushman, J.H.; Nichols, R.J.; Wyman, C.E. *Science* **1991**, *251*, 1318-1323.
2. Grohmann, K.; Wyman, C.E.; Himmel, M.E. In *Emerging Technologies for Materials and Chemicals from Biomass*; Rowell, R.M.; Schultz, T.P.; Narayan, R., Eds.; ACS Symposium Series No. 476; American Chemical Society: Washington, DC, 1992, pp.354-392.
3. Teeri, T.T.; Lehtovaara, P.; Kauppinen, S.; Salovuori, I.; Knowles, J. *Gene* **1987**, *51*, 43-52.
4. Abuja, P.M.; Schmuck, M.; Pilz, I.; Claeyssens, M.; Esterbauer, H. *Eur. Biophys. J.* **1988**, *15*, 339-342.
5. Esterbauer, H.; Hayn, M.; Abuja, P.M.; Claeyssens, M. In *Enzymes in Biomass Conversion*; Leatham, G.F.; Himmel, M.E., Eds., ACS Symposium Series No. 460; American Chemical Society: Washington, DC, 1991, pp. 301-312.
6. Tomme, P.; van Tilbeurgh, H.; Pettersson, G.; van Damme, J.; Vandekerckhove, J.; Knowles, J.; Teeri, T.; Claeyssens, M. *Eur. J. Biochem.* **1988**, *170*, 575-581.
7. Knowles, J.; Lehtovara, P.; Teeri, T. *Trends Biotechnol.* **1987**, *5*, 255-261.
8. Knowles, J.; Teeri, T.T.; Lehtovaara, P.; Penntila, M.; Saloheimo, M. In *Biochemistry and Genetics of Cellulose Degradation*; Aubert, J.-P.; Beguin, P.; Millet, J., Eds.; Academic Press: London, 1988; pp 153-169.
9. Coughlan, M.P. *Animal Feed Science and Technology* **1991**, *32*, 77-100.
10. Kraulis, P.J.; Clore, M.; Nilges, M.; Jones, T.A.; Pettersson, G.; Knowles, J.; Gronenborn, A.M., *Biochem.* **1989**, *28*, 7241-7257.
11. Bergfors, T.; Rouvinen, J.; Lehtovaara, P.; Caldentey, X.; Tomme, P.; Claeyssens, M.; Pettersson, G.; Teeri, T.; Knowles, J.; Jones, T.A. *J. Mol. Biol.* **1989**, *209*, 167-169.
12. Rouvinen, J.; Bergfors, T.; Teeri, T.; Knowles, J.K.C.; Jones, T.A. *Science* **1990**, *249*, 380-386.
13. Baker, J.O.; Himmel, M.E. In *Enzymes in Biomass Conversion*; Leatham, G.F.; Himmel, M.E., Eds., ACS Symposium Series No. 460; American Chemical Society: Washington, DC, 1991, pp. 313-330.
14. Shoemaker, S.; Watt, K.; Tsitovsky, G.; Cox, R. *Bio/Technology* **1983**, *1*, 687-690.
15. Gum, E.K.; Brown, R.D. *Biochim. Biophys. Acta* **1976**, *446*, 371-386.
16. Biltonen, R.L.; Freire, E. *CRC Crit. Rev. Biochem.* **1978**, *5*, 85-124.
17. Privalov, P.L.; Khechinashvili, N.N. *J. Mol. Biol.* **1974**, *86*, 665-684.
18. Pfeil, W.; Privalov, P.L. *Biophys. Chem.* **1976**, *4*, 23-32.
19. Sturtevant, J.M. *Ann. Rev. Phys. Chem.* **1987**, *38*, 463-88.
20. Privalov, P.L.; Potekhin, S.A. *Methods Enzymol.* **1986**, *131*, 4-51.
21. Privalov, P.L.; Gill, S.J. *Advan. Protein Chem.* **1988**, *39*, 191-234.
22. Shoemaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis, M. *Bio/Technology* **1983**, *1*, 691-696.

23. Fagerstam, L.G.; Pettersson, L.G.; Engstrom, J.A. *FEBS Lett.* **1984**, 309-315.
24. Baker, J.O.; Tatsumoto, K.; Grohmann, K.; Woodward, J.; Wichert, J.M.; Shoemaker, S.P.; Himmel, M.E. *Appl. Biochem. Biotechnol.* **1992**, In Press.
25. Lakowicz, J.R. *Principles of Fluorescence Spectroscopy*; Plenum Press: New York, NY; 1983.
26. Cooper, A. *Methods Enzymol.* **1982**, *81*, 285-288.
27. Claeyssens, M.; van Tilbeurgh, H.; Tomme, P.; Wood, T.M.; McCrae, S.I. *Biochem. J.* **1989**, *261*, 819-825.
28. Bhikhabhai, R.; Pettersson, G. *Biochem. J.* **1984**, *222*, 729-736.

RECEIVED April 29, 1992

Chapter 7

Recombinant Protein Stabilization through Engineered Metal-Chelating Sites

Pablo Umaña, James T. Kellis, Jr., and Frances H. Arnold

Division of Chemistry and Chemical Engineering 210–41, California Institute of Technology, Pasadena, CA 91125

Simple metal-chelating sites incorporated into common elements of secondary structure located on a protein surface offer a powerful and general strategy for stabilizing recombinant proteins. By binding with higher affinity to the native state of the protein, a metal ion can shift the folding/unfolding equilibrium toward the native state. To demonstrate this approach, we have engineered metal-chelating sites consisting of pairs of histidine residues into *Saccharomyces cerevisiae* iso-1-cytochrome *c*. 1 mM Cu(II) complexed to iminodiacetate stabilizes the cytochrome c variants by 1-2 kcal/mol, as determined by guanidinium chloride-induced unfolding. The increase in the free energy for unfolding is equal to that calculated from the preferential binding of the metal ion to the native protein. The Cu(II) affinities of di-histidine sites introduced in α-helices of bovine somatotropin indicate a potential for increasing stability by as much as 3.5 kcal/mol with a single di-histidine site. Chelating sites are easily introduced into surface α-helices and ß-sheets while introducing minimal or no disruption of the native structure or biological function.

Improving protein stability is an important goal of protein design and engineering, affecting a wide spectrum of biotechnology operations from large-scale purification to the efficacy of protein therapeutics. More flexible purification schemes could be designed for proteins capable of withstanding room temperatures for long periods of time. Robust proteins that retain their native structures at high temperatures or tolerate organic solvents could be isolated economically by selective heat denaturation and precipitation or extraction of undesirable substances. High stability would also dramatically reduce the costs involved in protein storage and shipping. Effective bioconversion processes can be designed for enzymes that are stable at high temperatures or in organic solvents, with advantages that include higher reaction rates, favorable thermodynamic equilibria for synthetic reactions which require low water concentrations (e.g. peptide synthesis by proteases), new synthetic applications for substrates insoluble in aqueous media, and reduced microbial contamination (*1*).

Stabilizing proteins, however, is a delicate process. A protein's stability, or the free energy by which the folded protein is more stable than its unfolded form, is usually quite small--on the order of 5 to 15 kcal/mol. Furthermore, this relatively

0097–6156/93/0516–0102$06.00/0

small free energy difference is the result of two large, opposing driving forces: attractive intramolecular forces plus the entropy of desolvation, which tend to stabilize the folded state, and the destabilizing decrease in chain entropy upon folding. From all the possible sequences of amino acids, nature has selected those consistent with this delicate balance (*2*). The marginal stabilities of natural proteins fulfill the requirements of the organism; proteins can be turned over and replaced on a regular basis, and this continual turnover gives the organism control over cellular processes. However, marginally-stable proteins are far from optimal for industrial purposes.

One strategy for enhancing protein stability is the substitution of specific amino acid side chains involved in stabilizing the folded state or destabilizing the unfolded state. This is a not a simple task, given our current limited ability to predict the effects on these large and opposing forces of altering the amino acid sequence. Predicting stabilizing mutations in the interior of a protein is particularly difficult, since the interactions arising between groups that become buried during folding are specific to each protein and can involve multiple favorable or unfavorable interactions. We have endeavored to develop general strategies for protein stabilization that are applicable to a wide variety of proteins and rely mainly on the substitution of surface amino acids. Engineering metal-chelating sites is such a strategy (*3*,*4*). By binding with high affinity to sites that optimally satisfy the bonding requirements of the metal ion when the protein is in its folded state, the metal ion shifts the folding/unfolding equilibrium to provide significant stability to the folded state. As illustrated by the thermodynamic cycle of Figure 1, the increase in the free energy of protein unfolding ($\Delta\Delta G_{unf}$) introduced by metal binding is equal to the difference in free energy of binding of the metal to the unfolded and native states ($\Delta\Delta G_{bind}$) (*3*). The preferential binding to the native state relies on structural "scaffolding" provided by elements of secondary structure common to nearly all proteins. This treatment of the effects of ligand binding on protein stability assumes that folding follows a simple two-state transition. If the denaturing conditions lead to formation of an inactive intermediate which retains the ability to bind the ligand (i.e. the intermediate retains the surface element of secondary structure containing the chelating site), then metal ion chelation may not result in any significant stabilization. In this sense, metal ion binding provides a useful probe of the process of protein folding as well as structure. Ghadiri and coworkers (*5*) have demonstrated the utility of metal ion binding in the stabilization of helical peptides. DeGrado and coworkers (*6*) have observed stabilization of de novo synthetic metal-binding proteins in the presence of various metal ions.

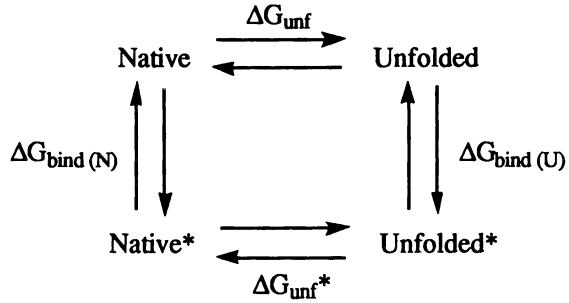


Figure 1. Thermodynamic cycle describing the stabilization of the native state of a protein by preferential binding of a ligand (e.g., a metal ion or metal complex) to the native state. The asterisks denote protein in the ligand-bound state. $\Delta\Delta G_{unf}$ is defined as $\Delta G_{unf}^{*} - \Delta G_{unf}$, and $\Delta\Delta G_{bind}$ is defined as $\Delta G_{bind(U)} - \Delta G_{bind(N)}$.

Basic Principles of Metal Chelation

Metal complexes with vacant coordination sites bind ligating atoms exposed on protein surfaces. Histidine and cysteine are the most important metal-coordinating ligands at neutral pH. Transition metal binding to protein surface residues has been exploited in protein purification by metal-affinity chromatography, extraction, and precipitation (*7*). Proteins that differ in their content of surface histidines, for example, are easily separated by chromatography on supports containing Cu(II)iminodiacetate (Cu(II)IDA) (*8*). We have studied the binding of histidine-containing proteins to Cu(II)IDA-PEG by protein partitioning in a PEG/dextran two-phase system and found an average association constant for the complex formed between a single surface histidine and Cu(II)IDA-PEG in a PEG-rich medium (top phase) to be $2.2 \times 10^3\ M^{-1}$ (*9*).

When two histidines are positioned so that they can form coordinate-covalent bonds simultaneously with a single metal ion, the strength of the association can be dramatically enhanced. A di-histidine metal-binding site engineered into the N-terminal α-helix of yeast cytochrome c binds Cu(II)IDA-PEG with 24 times higher affinity than does a single surface histidine. This metal-binding variant, which has only 3 exposed histidines, partitions in Cu(II)IDA-PEG/dextran as if it had 9 or 10 exposed histidines (*10*). Since few natural proteins have such a large affinity for Cu(II)IDA, these metal-binding sites are extremely useful in purification (*4*).

The fact that there is a large difference between the free energy for the dissociation of a metal complex formed with a polydentate ligand (e.g. the di-histidine binding site) and the free energy of dissociation of a complex formed with two or more monodentate ligands (independent histidines) can be attributed to a greater translational entropy loss for the binding of two monodentate ligands (*11*). Thus, a chelate formed with a polydentate ligand is more stable than the analogous complex with the monodentate ligands because the chelate dissociates into a smaller number of molecules. Metal chelates vary in stability, owing to differences in the size of the chelate rings, the number of rings, and other factors such as resonance and steric effects (*11*). Five-membered rings usually show the greatest chelate effect; increasing the number of atoms in the ring diminishes the chelate stability. This reflects the higher entropic cost of immobilizing a longer, more flexible ligand. A protein with di-histidine chelating site constitutes a bidentate ligand that forms a chelate ring with a large number of atoms. The chelate is stable, however, because the forces that stabilize secondary structure (i.e., hydrogen bonds) provide rigidity to the ring. Therefore, once the protein is folded, bidentate binding occurs at a low entropic cost. The protein's structural rigidity and ability to position the donor atoms to satisfy the stereochemical requirements of the metal constitute the positive contribution of the protein "scaffolding," quantified in the value of $\Delta\Delta G_{bind}$ in Figure 1.

Three variants of bovine somatotropin containing His-X_3-His sites in surface α-helices, $His_{(169),173}$, $His_{15,(19)}$, and $His_{26,30}$, have affinities for Cu(II)IDA that correspond to $\Delta\Delta G_{bind}$ values of 0.9, 2.2, and 3.5 kcal/mol, respectively (*12*). The lower copper affinity of the His-X_3-His site in $His_{(169),173}$ bovine somatotropin may be related to steric hindrance by neighboring arginine-176 and by the C-terminal loop (*12*). The higher Cu(II) affinities of the $His_{15,(19)}$ and $His_{26,30}$ bovine somatotropins, as compared to $His_{4,8}$ cytochrome c, may reflect the rigidity of the helices where these sites are located. Formation of the protein-metal chelate complex is favored by a high, negative enthalpy change and by the lowest possible loss of entropy. If the chelating site is introduced into a very flexible region of the protein, then the entropic cost of forming the complex will be high. However, if the helix backbone or nearby amino acid side chains are distorted upon forming the chelate, the enthalpy of the final complex will be higher. To provide the optimal conditions for metal chelation, the chelating site should be positioned in a region of low flexibility that also minimizes strain.

Engineered Metal-Chelating Proteins

Several di-histidine metal-chelating motifs in common elements of protein secondary structure have been identified, including His-X-His in a β-strand, His-X_2-His in a reverse β-turn, His-X_3-His in an α-helix, and juxtaposed histidines on adjacent strands of a β-sheet (His-X_n-His) (*4*). Here we will describe two particular di-histidine sites engineered into an α-helix and a β-sheet structure. We have replaced lysine-4 and threonine-8 (numbering system based on alignment of tuna and rice cytochromes c) in the N-terminal α-helix of yeast iso-1-cytochrome c with histidines to create the $His_{4,8}$ variant depicted in Figure 2A. The 1.23 Å resolution crystal structure of yeast iso-1-cytochrome c shows a very small antiparallel β-sheet, consisting of residues 37-40 and 57-59 (*13*). Because residue 39 is a histidine in the wild-type protein, we mutated the leucine residue at position 58 on the adjacent strand to histidine to create variant $His_{(39),58}$ with a metal-binding site crosslinking the β-strands (Figure 2B). Both variants also have the wild-type cysteine-102 mutated to serine to eliminate the reactive sulfhydryl group. The incorporation of these surface metal-binding sites has not interfered with the cytochrome c biological activity: both $His_{4,8}$ and $His_{(39),58}$ cytochromes c are functionally expressed in *S. cerevisiae*. Paramagnetic proton NMR spectroscopy of the two variants indicates that Cu(II)IDA binding is significantly enhanced at the putative chelating sites (data not shown).

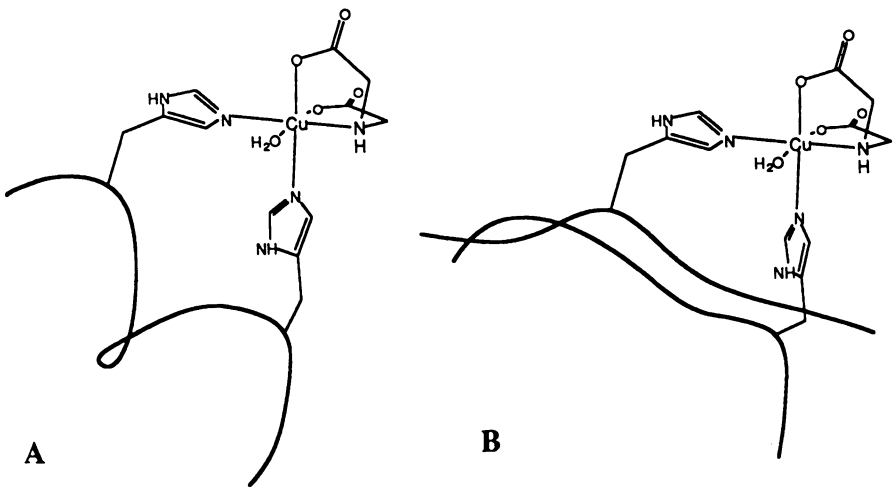


Figure 2. Models of di-histidine metal-binding sites in **A)** $His_{4,8}$ and **B)** $His_{(39),58}$ cytochrome c variants forming complexes with Cu(II)IDA. Thick solid lines depict the protein backbone.

Stabilization Studies

Protein stabilities can be determined through equilibrium studies of the reversible unfolding process. Figure 3 shows the guanidinium chloride (Gdm.Cl)-induced unfolding of the $His_{4,8}$ and $His_{(39),58}$ variants of cytochrome c in the absence and presence of Cu(II)IDA. The two variants (and wild-type enzyme) all exhibit essentially equal stability in the absence of the metal complex. Adding 1 mM Cu(II)IDA increases the amount of Gdm.Cl required to unfold the synthetic metal-binding proteins. This effect can be quantitated in terms of free energy of stabilization provided by the metal complex. The change in free energy for the unfolding process

(ΔG_{unf}) at a given Cu(II)IDA concentration is calculated from the relation $\Delta G_{unf} = - RT \ln K_{unf}$, where K_{unf} = (fraction unfolded / fraction folded). At Gdm.Cl concentrations where the folded and unfolded forms are present at levels that allow reliable calculation of the equilibrium constants in the presence and absence of the metal ion, the stabilization induced by the metal ion ($\Delta\Delta G_{unf}$) can be readily measured: it is 1.0 kcal/mol for $His_{4,8}$ and 1.7 kcal/mol for $His_{(39),58}$ in 1.0 mM Cu(II)IDA. The mutations have a negligible effect on the stability of the protein in the absence of Cu(II)IDA, compared to the protein lacking the engineered histidines.

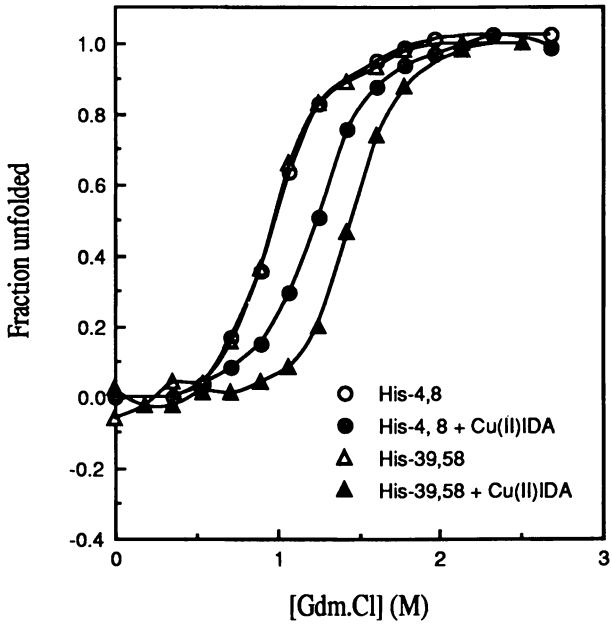


Figure 3. Stabilization of $His_{4,8}$ and $His_{(39),58}$ cytochrome c by 1 mM Cu(II)IDA.

The relation between $\Delta\Delta G_{unf}$ and metal complex concentration can be extrapolated to higher metal concentrations in order to determine the maximal stabilization provided by metal ion chelation. The data for $His_{4,8}$ display saturation behavior, with $\Delta\Delta G_{unf(max)}$ = 1.2 kcal/mol (*3*). The site cross-linking the β-sheet in $His_{(39),58}$ is slightly more effective, with a maximum stabilization of 1.8 kcal/mol.

The value of $\Delta\Delta G_{unf(max)}$ for the His-X_3-His site is equal to that calculated from Cu(II)IDA binding constants, $\Delta\Delta G_{bind}$, as predicted by the thermodynamic cycle in Figure 1. The binding constant K_a for the His-X_3-His chelating site is 5.3 x 10^4 M^{-1} in the folded protein (*8*). The unfolded protein has three histidines available to coordinate Cu(II)IDA, since His-26 becomes exposed upon unfolding. Thus, its binding constant is three times that for a single histidine, or 6.6 x 10^3 M^{-1}. The increase in the change in free energy of binding is calculated using the relation $\Delta\Delta G_{bind} = - RT\ln (K_{a(unf)} / K_{a(nat)})$, which yields 1.2 kcal/mol.

Metal-affinity partitioning studies carried out on synthetic metal-binding bovine somatotropins have shown that the preferential binding to the folded protein can be as high as 3.5 kcal/mol at a surface His-X_3-His site (*12*). The excellent agreement between $\Delta\Delta G_{unf}$ and $\Delta\Delta G_{bind}$ that we have observed for the $His_{4,8}$ cytochrome c

variant indicates that one can achieve up to 3.5 kcal/mol of stabilization using a single di-histidine chelating site and Cu(II)IDA.

Another practically important measure of protein stability is melting temperature, the temperature required to unfold half the protein molecules. Thermal unfolding studies on the metal-chelating cytochrome c variants indicate that 1 mM Cu(II)IDA increases the melting temperature of the $His_{4,8}$ variant by 6°C and the $His_{(39),58}$ variant by a full 10°C in 1.2 M Gdm.Cl, a remarkable degree of thermal stabilization for a single amino acid substitution (data not shown).

Future Perspectives

In order to develop a comprehensive strategy for applying this method of protein stabilization, additional aspects of the approach must be studied. The stabilization provided by other simple metal-chelating motifs, His-X-His in a ß-strand and His-X_2-His in a reverse ß-turn, remain to be studied. In addition, the effects of using different metal ions or metal complexes, general rules for positioning chelating sites in proteins, the stability afforded by metal chelation and sensitivity to environmental conditions have to be determined. One important question is whether stability is further enhanced when multiple chelating sites are introduced: can one engineer two or more chelating sites into a protein and obtain additive contributions to the free energy of stabilization? It is also interesting to investigate whether these metal-chelating proteins are stabilized to denaturation in organic solvents. The association between the metal ion and a ligating histidine should be stronger in organic solvents, since the solvent may not compete for ligation to the metal ion to the same extent as water does. However, this enhanced binding would be present in both the native structure and the unfolded form, and the net effect on binding and stability is difficult to predict.

Metal chelating sites can be applied not only to prevent unfolding, but also perhaps to guide refolding during the recovery of aggregated proteins in industrial operations. Still another application can be found in fundamental studies of protein folding pathways. Substitutions of specific residues have revealed their individual contributions to the folding process (*14*). A metal-chelating site would facilitate the study of the formation of elements of secondary structure through kinetic measurements of folding. The influence of the metal ion on the rate of folding would reveal the point at which the ligands in these elements attain their native structure.

Conclusions

Engineered metal chelation is an excellent tool for protein stabilization; even simple di-histidine chelating sites on a protein surface can provide significant stability enhancements. We have demonstrated that histidine pairs inserted into an α-helix or β-sheet can add 1-2 kcal/mol to the free energy of unfolding in cytochrome c. Previous metal-binding studies on variants of bovine somatotropin indicate that a single α-helical His-X_3-His site should be able to afford as much as 3.5 kcal/mol of stabilization. The method is generally applicable since the structural "scaffolding" needed to coordinate the metal ion at a low entropic cost is provided by common elements of secondary structure. This also allows introduction of a chelating site with minimal or no disruption of native structure or biological function.

Acknowledgments

This work is supported by the National Science Foundation (P.Y.I. award to F.H.A.) and the Office of Naval Research. F.H.A. gratefully acknowledges a fellowship from the David and Lucile Packard Foundation.

Literature Cited

1. Dordick, J. S. *Enz. Microb. Technol.* **1989**, *11*, 194.
2. Jaenicke, R. *Biochemistry* **1991**, *30*, 3147.
3. Kellis, J. T., Jr.; Todd, R. J.; Arnold, F. H. *Bio/Technology* **1991**, *6*, 994.
4. Arnold, F. H.; Haymore, B. L. *Science* **1991**, *252*, 1796.
5. Ghadiri, M. R.; Choi, C. *J. Am. Chem. Soc.* **1990**, *112*, 1630.
6. Handel, T.; DeGrado, W. F. *J. Am. Chem. Soc.* **1990**, *112*, 6710.
7. Arnold, F. H., *Bio/Technology* **1991**, *9*, 151.
8. Hemdan, E. S.; Zhao, Y.; Sulkowski, E.; Porath, J. *Proc. Natl. Acad. Sci., USA* **1989**, *86*, 1811.
9. Suh, S.; Arnold, F. H. *Biotechnol. Bioeng.* **1990**, *35*, 682.
10. Todd, R. J.; Van Dam, M.; Casimiro, D.; Haymore, B. L.; Arnold, F. H. *Proteins: Struct., Funct., Genet.* **1991**, *10*, 156.
11. Martell, A., "The Relationship of Chemical Structure to Metal-Binding Action." In *Metal Binding in Medicine*. Seven, M. J., Ed; J. B. Lippincott Co., Philadelphia, PA, 1959; pp 1-18.
12. Suh, S.-S.; Haymore, B. L.; Arnold, F. H. *Protein Engineering* **1991**, *4*, 301.
13. Louie, G. V.; Brayer, G. D. *J. Mol Biol.* **1990**, *214*, 527.
14. Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Bycroft, M; Fersht, A. R. *Nature* **1990**, *346*, 440.

RECEIVED March 31, 1992

Chapter 8

Engineering Nonaqueous Solvent-Compatible Enzymes

Frances H. Arnold, Keqin Chen, Chara Economou, Wayne Chen, Pascal Martinez, Kyung Pyo Yoon, and Mariana Van Dam

Division of Chemistry and Chemical Engineering 210–41, California Institute of Technology, Pasadena, CA 91125

Given current powerful tools for protein molecular engineering and a continually improving understanding of protein folding and stability, it is possible to construct vastly improved industrial enzyme catalysts. Enzymes are better suited to industrial syntheses if they can be made stable and active in high concentrations of polar organic solvents. Site-directed mutagenesis has been used to test two "design rules" for stabilizing enzymes in organic media. We have shown that surface charge substitution and the introduction of metal-chelating sites provide simple and generally-applicable mechanisms for enzyme stabilization. A variant of α-lytic protease containing two surface charge substitutions is 27 times more stable than wild-type enzyme in 84% DMF. Random mutagenesis and rapid screening techniques have been used to isolate enzyme variants with enhanced catalytic activity in polar organic solvents. Our first experiments resulted in a variant of subtilisin E that is 38 times more active in 85% DMF than the wild-type enzyme. Subsequent rounds of random mutagenesis and screening have yielded a variant 256 times more active in 60% DMF.

Although polar organic solvents provide an excellent medium for many chemical transformations, these solvents often severely compromise enzyme catalysts. Transferring an enzyme from an aqueous environment to a polar organic solvent can drastically reduce both stability and activity (Figure 1). An enzyme in an organic solvent is operating far from the potential it achieved after billions of years of evolutionary optimization. Because the enzyme is "damaged," however, there is considerable room for improvement; the noncovalent forces that contribute to enzyme stability as well as to the stabilization the enzyme imparts to the reaction transition state can be redistributed by making selected substitutions in the amino acid sequence. An evolutionary approach of combining mutations generated by site-directed or random mutagenesis can lead to enzyme variants that exhibit greater stability or activity in the new environment. Because one is starting with a far-from-optimal catalyst, engineering enzymes for use in organic solvents is a particularly promising application of protein engineering (*1,2*).

0097–6156/93/0516–0109$06.00/0

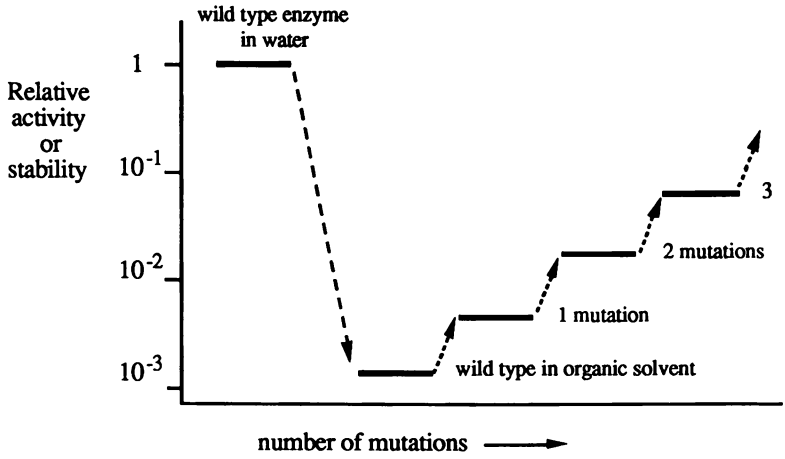


Figure 1. Evolutionary approach to creating enzymes that are optimized for use in nonaqueous solvents. Selected mutations can recover the stability and catalytic activity lost by transfer to the organic solvent.

Our protein engineering studies have been carried out on two proteolytic enzymes, subtilisin E and α-lytic protease. Both are serine proteases that will catalyse specific peptide syntheses and transesterification reactions in organic solvents. α-lytic protease (198 amino acids) is somewhat smaller than subtilisin E (275 amino acids) and contains three disulfide bridges, while subtilisin contains none. Both enzymes are expressed with pre-pro sequences that are eventually cleaved to yield the mature enzymes. Since the pro sequence is required for proper folding, both enzymes unfold irreversibly. As a result, any process that leads to denaturation results in a permanent loss of catalytic activity. If other mechanisms of deactivation (e.g. autolysis, for subtilisin) are excluded by carefully choosing the conditions, the time course of catalytic activity (assayed in any medium) can be interpreted in terms of unfolding events that are related to the protein's thermodynamic stability (*3*).

Enhancing Enzyme Stability in Organic Solvents

Useful approaches to stabilizing enzymes in polar organic solvents should be generally applicable to a wide variety of enzymes. Because the effects of amino acid substitutions on details of interior packing and hydrogen bonding are unique to each enzyme and are difficult to predict, stabilization approaches based on improving these internal interactions are difficult to implement successfully. We have concentrated instead on stabilization mechanisms that depend on the modification of surface amino acids. These mechanisms are: 1) substitution of charged surface residues (*1,3*) and 2) introduction of simple metal-chelating sites such as His-X_3-His in an α-helix (*4*). Because the second mechanism is treated in detail elsewhere in this volume, this paper will discuss only the first.

Desolvation of charged surface residues in organic solvents can contribute to enzyme destabilization. Charged side chains form multiple hydrogen bonds with solvent water molecules; removing water leaves these hydrogen bonding sites unsatisfied. Thus, dehydration could favor conformational changes that allow solvation of charged residues by polar groups elsewhere in the protein. We have proposed that replacing charged side chains renders the surface of the folded protein more compatible with an organic solvent environment and may remove a driving force

for formation of alternate (inactive) structures. The substitution of selected charged residues can effectively stabilize the active enzyme in organic solvents.

This hypothesis was tested by replacing charged surface residues that do not participate in catalysis or favorable noncovalent interactions (salt bridges and hydrogen bonds) in both α-lytic protease and subtilisin E. In α-lytic protease, the targeted residues were substituted with as many of the remaining 19 amino acids as possible (*5*). Seven primarily hydrophobic substitutions at two positions significantly increased enzyme stability in dimethylformamide: replacement of Arg 45 by Glu, Ser, Leu and Ile and Arg 78 by Phe, Leu, and Tyr all stabilize α-lytic protease in 84% DMF. Although Glu is negatively charged, its hydration potential is significantly less than that of Arg (*6*). As shown in Figure 2, single mutations could be combined to yield a double mutant more stable in 84% DMF than either single mutant. The rate at which this double mutant deactivates in 84% DMF is 27 times less than that of wild-type α-lytic protease. These stabilizing effects were only observed in the organic solvent; no substitution significantly improved the stability of α-lytic protease in aqueous buffer.

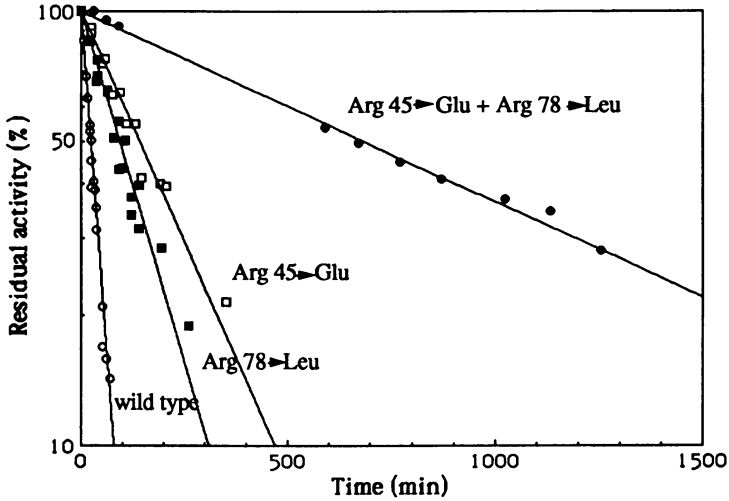


Figure 2. Stabilities of α-lytic protease surface charge variants in 84% (v/v) DMF/16% buffer, pH 8.75, 30 C. o, wild-type; ■, Leu 78; □, Glu 45; ●, Glu 45+Leu 78.

In subtilisin E, Asp 248 was substituted by three amino acids of increasing hydrophobicity, Asn, Ala and Leu, and all three variants were found to be slightly more stable in the presence of high concentrations of the organic solvent (*3*). These studies provide strong evidence that substitution of surface charged residues is a generally useful mechanism for stabilizing enzymes in high concentrations of polar organic solvents. Our current studies are directed towards determining to what extent this mechanism can be applied: how many surface substitutions will be tolerated, and what is the maximum stabilization that can be achieved?

Enhancing Enzyme Activity in Organic Solvents

The loss of catalytic activity that an enzyme experiences upon transfer to an organic solvent often cannot be attributed solely to the solvent's effect on stability and

denaturation--enzymes that are quite stable in organic media also exhibit poor catalytic activity. Because little is known of the exact mechanisms by which polar organic solvents reduce the activity of soluble enzymes, we have relied on a random mutagenesis approach to improve enzyme catalytic performance. Random mutagenesis was carried out on subtilisin E by polymerase chain reaction (PCR) techniques and combined with screening for enhanced activity in the presence of DMF (*7*). Two amino acid substitutions which increase the activity of subtilisin E in the presence of the organic solvent, Q103R and D60N, were identified by screening the randomly mutated bacterial colonies on agar plates containing DMF and casein. In this screening process, a colony secreting a subtilisin variant that is more active than wild-type in the presence of DMF produces a visibly larger halo on the casein plate. The effective substitutions identified in this and subsequent studies are located near the substrate binding pocket or in the active site. The effects of one of the substitutions, D60N, were apparent only in the presence of DMF, a result which highlights the importance of screening in the organic solvent. As with the surface charge substitutions, this mutation improved the enzyme's performance only in organic solvents.

Engineering an enzyme that is truly optimized to function in a polar organic solvent will probably require the accumulation of multiple mutations. If the effects of individual mutations are additive, they can be identified in separate mutagenesis and screening experiments and subsequently combined. A triple mutant was constructed by combining the double random variant D60N+Q103R with N218S, a mutation discovered by random mutagenesis of subtilisin BPN' (*8*). In water, the triple variant is 7 times more efficient towards the hydrolysis of suc-Ala-Ala-Pro-Met-p-nitroanilide than wild-type subtilisin E; it is 38 times more active in 85% DMF (Figure 3). The triple variant also exhibits significantly higher esterase activity than the wild-type in high DMF concentrations (data not shown). Preliminary experiments show that this

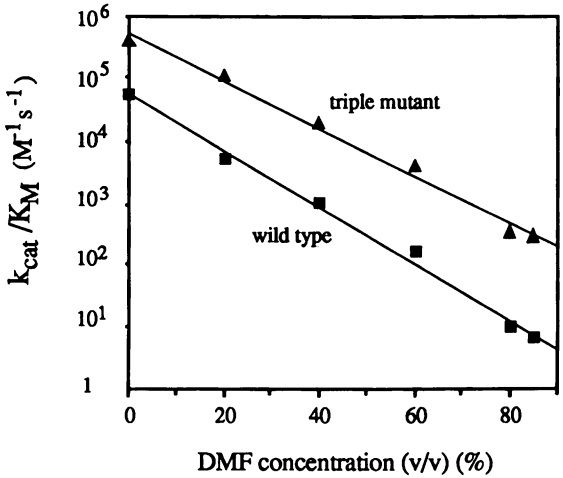


Figure 3. Catalytic efficiencies for hydrolysis of suc-Ala-Ala-Pro-Met-pna by wild-type subtilisin E (■) and triple mutant Q103R+D60N+N218S (▲) in the presence of DMF (*7*). (Reproduced with permission from Chen, K.; Arnold F. H. *Bio/Technology*, **9**, 1073-1077 (1991). Copyright 1991 Nature Publishing Company).

triple mutant is much more effective in kinetically-controlled peptide synthesis than wild-type subtilisin E (Chen, K; Arnold, F. H., unpublished results). Due to the stabilizing effect of the N218S mutation, the triple variant is also slightly more stable than wild-type subtilisin E (*7*).

The combined mutations Q103R+D60N+N218S stabilize the reaction transition state by an incremental free energy that is essentially the sum of all the contributions from each of the single mutants (*7*). These substitutions apparently do not cause more than very localized perturbations in the enzyme's structure. Even if multiple mutations begin to interact in the confined space of the active site, however, effective amino acid substitutions can be accumulated by the "evolutionary" process of sequential mutagenesis in which variant DNA is used as template for generating the next mutation. After three subsequent rounds of mutagenesis and screening in higher organic solvent concentrations, this evolutionary approach has yielded a variant of subtilisin E that is 256 times more active in 60% DMF than the wild-type enzyme. This engineered enzyme represents a dramatic improvement over natural catalysts for applications in peptide and polymer synthesis in organic solvents.

Conclusions

Mechanisms for enhancing both enzyme stability and catalytic activity in organic solvents have been illustrated by site-directed and random mutagenesis techniques. Substitution of selected charged surface residues can stabilize enzymes in high concentrations of organic solvents. Successful implementation of this stabilization strategy requires that the substituted amino acid side chain not be involved in catalysis or important stabilizing interactions. Incorporation of simple metal-chelating sites provides a second general and easy-to-implement strategy for protein stabilization. To improve catalytic activity in organic media, random mutagenesis combined with rapid screening techniques can be very effective. As beneficial mutations accumulate, the screening for enhancements in catalytic activity can be carried out in higher and higher concentrations of the organic solvent. There is clearly great potential for engineering enzymes that are well-suited to synthesis applications in polar organic solvents.

Acknowledgments

This research is supported by the Catalysis and Biocatalysis Program of the Advanced Industrial Concepts Division of the U. S. Department of Energy and by the Office of Naval Research. F.H.A. gratefully acknowledges an NSF PYI award and a fellowship from the David and Lucile Packard Foundation.

Literature Cited

1. Arnold, F. H. *Protein Engineering* **1988**, *2*, 21.
2. Arnold, F. H. *Trends Biotechnol.* **1990**, *8*, 244.
3. Martinez, P.; Van Dam, M. E.; Robinson, A. C.; Chen, K.; Arnold, F. H. *Biotechnol. Bioeng.* **1991**, *39*, 141.
4. Kellis, Jr., J.; Todd, R. J.; Arnold, F. H. *Bio/Technology* **1991**, *9*, 994.
5. Martinez, P.; Arnold, F. H. *J. Am. Chem. Soc.* **1991**, *113*, 6336.
6. Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.
7. Chen, K.; Arnold, F. H. *Bio/Technology* **1991**, *9*, 1073.
8. Bryan, P. N.; Rollence, M. L; Pantoliano, M. W.; Wood, J. F.; Finzel, B. C.; Gilliland, G. L.; Howard, A. J; Poulos, T. L. *Proteins: Struct. Funct. Gen.* **1986**, *1*, 326.

RECEIVED March 31, 1992

Chapter 9

Mutational Effects on Inclusion Body Formation

Ronald Wetzel and Boris A. Chrunyk

Macromolecular Sciences Department, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406

Specific mutations in the cloned cDNAs for two small, human proteins, interferon-γ and interleukin-1β, lead to dramatic alterations in the partitioning of the gene product into inclusion bodies in *E. coli* expression. Examination of the reported three-dimensional structures of these proteins shows that the mutations generating these effects can be found in elements of α-helix, β-sheet, loops, or disordered sequence, in proteins that are either predominantly helical or β-sheet in structure. In both proteins, other mutations are described which lead to loss of stable accumulation of the gene product in any form in the cell. The results are discussed in terms of the possible roles of folding stability and kinetics in IB formation. Practical applications of mutants exhibiting altered inclusion body levels are also discussed.

Bacteria sequester a wide variety of molecules into inclusion bodies (IBs) *(1)*. Inclusion bodies composed of protein were first observed in bacteria grown on amino acid analogues which were misincorporated into protein during ribosomal synthesis *(2)*. These particles can also be formed as a consequence of the synthesis of unnaturally high levels of normal proteins *(3)*. However, IBs became a familiar phenomenon only with the advent of recombinant DNA techniques for the bacterial synthesis of heterologous gene products *(4, 5)*. Despite the observation of IBs in many cases of heterologous gene expression, and despite the practical and fundamental importance of the phenomenon, the process remains little understood *(6-11)*.

An understanding of the mechanisms by which IBs form would be valuable for a number of reasons. First, it should lead to improved control over their formation. In some cases it may be desirable to suppress IB formation, if recovery of active, native protein by refolding *in vitro* proves difficult (see, for example, reference *(12)*). In other cases it may be desirable to encourage IB formation, for instance to stabilize proteins which might otherwise be degraded by cellular proteases *(13)*, or to provide an initial purification step which would remove most *E. coli* proteins *(6)*. Inclusion bodies also are of interest because of their resemblance to a class of human disease called amyloidosis, which is characterized by the formation of insoluble fibrils of selectively aggregated proteins *(14-18)*. The comparative properties of inclusion bodies and amyloid have been discussed *(11)*. Both particles are highly resistant to dissolution in

0097–6156/93/0516–0116$06.00/0

native buffers, normally requiring high levels of denaturing solutes. Both particles are also highly selectively enriched in a particular protein, in spite of the fact that they are formed in milieux which are rich mixtures of proteins. Another aspect of the selectivity with which proteins are aggregated in these processes is the observation of mutational effects on their formation; that is, in both phenomena, the replacement of as little as a single amino acid can dramatically alter the extent to which insoluble aggregate is formed *(11)*. The selectivity of aggregate formation *in vivo* is one of the main reasons for suspecting that these processes are related somehow to protein folding. Although the aggregation of mixtures of proteins is often an amorphous process which, once begun, tends to include many different components of the mixture, aggregation during *in vitro* refolding experiments can be rather selective. This has led to the idea that the aggregation interface in folding-related aggregation resembles one of the protein's normal inter- or intramolecular packing interfaces, giving the aggregation process the same high specificity found in the protein folding process *(2, 19)*. To the extent that IB formation is, in fact, a reflection of inherent folding or stability properties of the protein being expressed, IB formation might be used as a phenotypic marker for folding/stability mutants in a genetic analysis.

The research described here is directed toward understanding the basis for the sequence specificity of the IB formation process. Simple methods for the identification of IB mutations in a collection of mutants have led to the identification of dramatic IB mutations in two simple globular proteins of known three-dimensional structure expressed by recombinant DNA techniques in *E. coli*: human interferon-γ (IFN-γ) and human interleukin-1β (IL-1β). Work is in progress to identify *in vitro* properties which correlate with the ability of various mutants to form IBs. Evidence is accumulating that supports the idea *(20)* that aspects of the folding pathway can influence inclusion body formation.

Solubility, Stability and Folding in Inclusion Body Formation

Figure 1 is a general mechanism for protein folding which can be used to examine possible mechanisms for IB formation and for mutational effects on IB formation. Some abnormal proteins expressed by recombinant DNA techniques might be expected to be poorly soluble in native buffer both *in vivo* and *in vitro*, either because they cannot fold into a native-like structure, or because, once folded, they are still poorly soluble. Possible examples of this type of protein are some integral membrane proteins and some fusion proteins incorporating extensive unnatural sequences. That such proteins might form IBs during expression is not surprising. A second theoretical mechanism for IB formation postulates a higher tendency to aggregate of unfolded states populated at equilibrium due to a diminished thermodynamic stability of the protein under growth conditions of the cell. For example, the inability of a protein to form disulfide bonds in the cytoplasm will correspondingly diminish its ΔG_{stab} and require it to spend most of its time in an unfolded state - potentially of limited solubility. Similarly, mutations which diminish non-covalent interactions required for folding stability will tend to populate such states if the T_m is reduced to or below the growth temperature. In the third theoretical mechanism, transiently populated, kinetic folding intermediates are responsible for IB formation when they become of sufficiently poor solubility or of sufficiently long kinetic lifetimes. It is not known whether all of these possible mechanisms are responsible for IB formation in *E. coli*, nor how widespread any of them might be.

In addition to these mechanisms, it is also possible that mutations may influence IB formation by altering the interaction surface of the protein with a molecular chaperone responsible for facilitating its folding into a soluble, native molecule. Chaperonins have been shown to reduce losses due to aggregation during protein folding *in vivo* and *in vitro* with several proteins *(21)*.

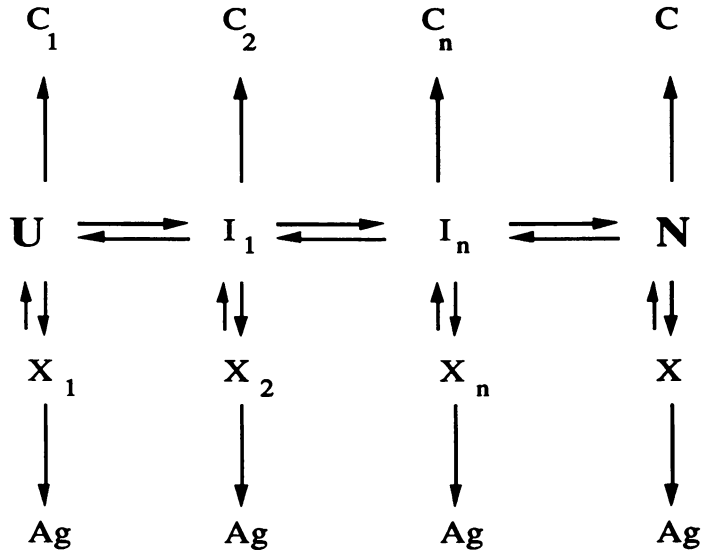


Figure 1. A general mechanism for protein folding accounting for proteolysis and aggregate formation . N = native state, U = unfolded state, I = intermediate, C = covalently modified form, X = associated form, Ag = aggregated form. Reproduced with permission from reference *(11)*. Copyright 1992 Plenum.

Only a few examples of mutational effects on IB formation have been reported. In the work of the King lab, a series of temperature sensitive folding (*tsf*) mutations identified over a decade ago in the tailspike protein of *Salmonella* phage P22 have now been characterized as leading to increased levels of IBs *(20)*. These mutations shift the temperature-dependence of IB formation found for the wild type protein to lower temperature, and also make the temperature-dependence curve more cooperative. The result is that, at the non-permissive temperature of 39° C, where a large percentage of WT tailspike chains fold productively and associate into trimer, *tsf* mutant tailspike chains are exclusively deposited into IBs, and are thus irreversibly lost. Because the *tsf* mutants can be grown at a lower, permissive, temperature to produce folded tailspikes capable of assembling viable phage, these mutations clearly do not simply affect protein solubility. Furthermore, because the T_m values of the folded *tsf* mutants are similar to WT, and much higher than the non-permissive growth temperature, it is clear that they do not influence IB formation *via* large effects on ΔG_{stab}. These mutants are thus presumably defective in folding and/or assembly. Many of the mutations are located in sequences resembling motifs favoring β-turns, consistent with a mechanism involving changes in key folding steps or intermediates related to turn formation.

The characterization of the P22 tailspike *tsf* mutants provided the first example of mutational effects on IB formation *(20)*, and the continuing analysis of these mutants *(22-24)* is establishing a framework of ideas on how to think about mutational effects on IB formation observed in other systems. Several more anecdotal reports of mutational effects have also been reported *(25-28)*. How widespread will the observation of such mutational effects be? The tailspike protein is large and oligomeric (a trimer of 666 amino acid subunits), β-rich, and of unknown three-dimensional structure. Will it be possible to identify and characterize similar mutations in small,

globular proteins of known structure? Are monomeric proteins capable of IB formation and mutational effects on IB formation? Will the locations of IB mutations in the 3D structure, and the *in vitro* measured folding kinetics and stabilities of proteins bearing them, yield clues to the mechanism of formation of IBs and to *in vivo* protein folding in general?

The original observation of the P22 tailspike mutants was possible in spite of their low frequencies of appearance among the classically generated mutagenesis pool because of the availability of a genetic screen for the *tsf* phenotype *(29)*. Modern methods of mutagenesis, built on recombinant DNA techniques, can provide frequencies of one or more mutations per gene segment *(30, 31)*. Such high frequencies of mutation, with correspondingly lower levels of wild type background, allow the use of much less efficient screens for identifying interesting mutants, such as IB mutants, in the pool. This in turn has made possible the introduction of screening methods which do not rely on a functional role of the gene product in the producing organism, thus opening up the possibility of mutational analyses of many heterologous genes. In turn, this makes possible an efficient search for IB mutations in a gene encoding a protein which has been chosen based upon desirable characteristics as a model folding system.

Interferon-γ

Human interferon-γ is a homodimer made up of highly α-helical subunits, each 144 amino acids long when made in *E. coli* (where the initiator Met is not processed off). The X-ray crystal structure shows an extended subunit interface delineated by the intertwining of the subunits *(32)*. This structure does not show the C-terminus beyond residue 120, which is consistent with the disordered structure suggested by the susceptibility of the C-terminus to limited proteolysis both *in vitro* and *in vivo (33)*. The molecule contains no cysteine residues, and exhibits a Tm of 54° C at pH 7 *(34)*. When the wild type sequence is produced in *E. coli* grown at 37° C (17° C lower than its Tm) under control of the *trp* promoter, over 25% of the Coomassie blue staining protein is IFN-γ, and over 95% of this material is found in the insoluble fraction of a native lysate *(35)*. The material is insoluble because it is incorporated into inclusion bodies *(35)*.

In a mutational analysis of the cloned gene directed at elucidating structure/activity relationships *(36, 37)*, mutations were isolated which exhibited significantly higher levels of IFN-γ in soluble cytoplasmic extracts. This phenotype proved to be due, not to an increase in overall expression levels, but rather to an increase in the amount of IFN-γ which partitioned into the soluble fraction *(35)*. Such mutations were conveniently detected by growing and lysing cells in microtiter plate wells, and assaying the soluble fractions by an ELISA using anti-IFN-γ antibodies; very high responses compared to wild type prove to be due to a higher ratio of soluble to insoluble forms of IFN-γ *(35)*. A partial list of sequences which exhibit the phenotype of lower IB formation than WT is listed in Table I. It was of interest to know if mutants similar to the P22 tailspike mutants, yielding higher IB formation than WT, could also be isolated; however, with IB formation already very high in the WT this did not appear feasible. Fortunately, it was possible to overcome this dilemma by adjusting cell culture temperature.

With growth temperature lowered to 30° C, a larger percentage of the wild type protein was found in the soluble fraction *(35)*; this beneficial effect of lower growth temperature on IB formation has been noted with other proteins as well, but is by no means a general rule *(10)*. With a larger amount of wild type protein partitioning into the soluble fraction, it was now possible to screen for mutations, similar to those described for P22 tailspike protein, which directed even more protein into IBs. However, such mutants cannot be unequivocally identified by using an immunoassay

Table I. Levels of Soluble Expression of IFN-γ Variants

Mut. #	Sequence Change from Wild Type	ELISA,μg/ml[a]	Rel.Act.[b]
Cassette 1			
	120 130 140		
WT	EL SPAAKTGKRK RSQMLFRGRR ASQ	2.8	100
71	S	1.6	55
50	R	2.2	174
58	QP	7.5	28
30	S I E G	6.5	6
36	P E W	1.6	124
47	N YC ---- ---	4.8	109
23	LSRR------ ---------- ---	11.0	ND
1	------ ---------- ---	8.0	0.06
25	---------- ---------- ---	3.2	3
88	VDE HPSRS	10.0	204
55	EVRCCVEVDE HASRS	8.6	90
56	Q SE GVRCCLEVDE HPSRS	8.6	0.17
28	L Q SE GVRCC EVDE HPSRS	0.5	8
13	C RLTE SE GVRSCLGADE HPSRLN	5.3	1.1
74	A I LDLNAVVY HS	0.3	3
68	RCCS WTS IPVDLNAVVY HS	23.0	150
7	A SE E GSVS STS IPVDLNAVVY HS	5.3	0.13
87	RL REAK E DAV-	11.8	180
27	AT P Q GVRCC EVDE HPSRS	2.1	95
Cassette 3			
	70 80 90		
WT	-SIQKSV ETIKEDMNVK FFNSNKKKRD DFEKL-	0.9	100
66	A G K	0.4	<<1
7	Q	7	250
5	A	8	200
56	K M	11	<<1
11	S Y	13	50
75	H T	38	70

[a]ELISA on crude lysis supernatants as described *(35)*. [b]Antiviral activity assessed on partially purified material as described *(35)*. Adapted from reference *(35)*.

as described above - which can detect only at the amount of soluble protein in the cell - since an observed reduction in cellular levels might be due to loss of protein due to proteolysis rather than IB formation. To distinguish between these two possible fates of the polypeptide, an assay was devised which is capable of detecting levels of gene product accessible only to denaturant. In this assay, an array of cultures producing different mutant IFNs is lysed in parallel by two methods - one native lysis and one denaturant-based - and the replicate lysates assayed by a variation of the ELISA. This produces values which can be manipulated to give ratios indicating greater-than-wild-type or less-than-wild-type levels of IBs. If no appreciable IFN is made available even from cells lysed by denaturant, then it is clear that no IFN accumulated in the cell, presumably due to loss by proteolysis. SDS PAGE analysis confirmed, in most cases

tested, the characteristics predicted by the differential lysis assay. Table II shows the results of a survey of randomly generated mutations in IFN-γ. Four classes of mutant were thus identified: wild-type-like, no stable accumulation, greater-than-wild-type levels of IBs, and less-than-wild-type levels of IBs.

Table I also shows that there is no correlation between the partitioning of IFN-γ between soluble and insoluble fractions and the antiviral specific activity of the soluble material. Not only are there some highly inactive mutants which are, like wild type, found almost exclusively in IBs, but there are also highly active mutants which are found almost exclusively in the soluble fraction!

One practical implication of the IFN data is that, in an expression system in which a particular wild type gene product is produced in undesirable IBs, it may be possible to identify mutants which retain good biological activity while at the same time gaining improved expression characteristics such as reduced IB levels. It remains to be seen whether mutations giving improved soluble expression will be a general phenomenon. King and co-workers have recently described the characterization of global suppressor mutations of the *tsf* phenotype in the P22 tailspike system which have the same effect of improving soluble levels of product at the expense of IB formation *(24)*.

It is interesting that the wild type C-terminus, although disordered and relatively hydrophilic in character, nonetheless plays a major role in influencing IB formation. One possible explanation, for which there is no other supportive data, is that the C-terminus is important at some point during the folding of the protein, even if its structural role in the native state is not evident.

Another feature IFN-γ has in common with the P22 tailspike is that both are oligomeric proteins. Since many of the proteins known to give severe aggregation on attempts to refold *in vitro* are oligomeric *(7)*, the possibility exists that folding-related IB formation may be a phenomenon limited mostly to expression of oligomeric proteins. For this reason we sought a monomeric protein of known structure exhibiting mutational effects on IB formation during expression in *E. coli*. Human interleukin-1β proved to be such a protein.

TABLE II. Fate of IFN-γ Variants in *Escherichia Coli* by Mutagenesis Region

Sequence Positions:	(1-26)	(26-56)	(51-65)	(65-95)	(95-123)	(119-143)
Phenotype						
No detectable expression, Soluble or insoluble	55	21	18	38	35	0
Detectable expression:						
Very high in inclusion bodies (+ in Gdn-HCl only)	5	64	14	14	37	0
Wild type levels of expression by native lysis	37	14	57	47	28	89
Low in inclusion bodies (high native lysis expression)	3	1	11	1	0	11

Results are listed as % of all mutants tested from each collection, normally a sample of 500-1000 isolates. Reproduced with permission from reference *(35)*. Copyright 1992 Nature Publishing Co.

Interleukin-1β

IL-1β is a monomeric protein of 153 amino acids long containing two unpaired Cys residues and no disulfide bonds. Its three-dimensional structure has been determined both by Xray crystallography *(38, 39)* and by nmr *(40)*. It exists in a β-barrel structure, folding in a motif first observed for soybean trypsin inhibitor. Its Tm at pH 7 was determined by scanning microcalorimetry to be 62° C (C. Brouilette, unpublished). Conditions for its reversible unfolding induced by solute denaturants have been reported *(41)*. Expressed in *E. coli*, the wild type sequence accumulates to about 10% of the Coomassie staining protein, almost all of which is found in the supernatant after centrifugation of crude lysates. The small percentage of IL-1β in the insoluble fraction is presumably localized in inclusion bodies, since characteristic refractile particles are observed in IL-1β producing cells in phase contrast microscopy (unpublished observations).

Using SDS-PAGE, we surveyed a collection of site-directed (ie., not randomly generated) mutants, prepared over the past several years to explore structure - activity relationships *(42)*, for their stability and solubility when expressed in *E. coli*. As was found for IFN-γ mutants, we observed mutants which gave predominantly soluble expression - like the wild type, mutants which did not accumulate stably in any form in the cell, and mutants which, while produced in approximately the same amount as wild type, distributed greater percentages of the accumulated IL-1β into the insoluble fraction. Table III lists representative mutants from the collection of approximately 70 site-directed mutants surveyed *(43)*.

The table shows that residue positions associated with high levels of inclusion bodies are found both in β sheet and in turns. The table also shows that, in some cases, different mutations at the same residue position can give different phenotypes, depending on the nature of the new amino acid. Relatively subtle effects are apparent, as for example the difference between L10S and L10T, where the presence or absence of a single methyl group appears to determine which of the two major off-pathway reactions, proteolysis or inclusion body formation, dominates IL-1β expression. High IB content does not correlate with the degree of exposure of the residue in the native structure: at three of the positions at which mutations can lead to high levels of IBs, position 10 is essentially fully buried, position 9 is partially buried, and position 97 is essentially fully exposed.

The table also shows that several of the mutants have been purified and their stabilities to reversible unfolding determined. Preliminary results suggest a correlation of extent of IB formation in most of the mutants tested with their *in vitro* stabilities to both reversible unfolding and irreversible thermal aggregation *in vitro*. However, there are some problems with taking this trend as being indicative of a mechanism of IB formation dependant on ΔG_{stab}. First, the temperatures at which mutants aggregate *in vitro* are 5-20° C higher than the growth temperature of 42° C at which IL-1β is synthesized *in vivo*. Second, at least one mutant, K97V, is an extreme outlier to the correlated points, in that it is much more prone to IB formation than WT but is at least as stable as WT to Gdn-HCl unfolding (Table III). The K97V mutant on further examination appears to form IBs through a defect in its folding pathway. Recent data shows a temperature dependence to the aggregation of this molecule *in vitro* during its refolding from Gdn-HCl solution (BC and RW, unpublished data). In contrast, the wild type's refolding *in vitro* is relatively insensitive to temperature. These preliminary results suggest that aggregation during refolding *in vitro* may be a valuable model for IB formation, at least for the K97V mutant, allowing further *in vitro* characterization of the aggregation mechanism. In the future we hope to expand the data set now available by conducting a full mutational analysis using screening methods such as those described above for interferon-γ.

TABLE III. Properties of selected IL-1β mutants

Mutation	Total Expression[a]	% IL-1β as IBs	$\Delta\Delta G_{unfold}$[b], kcal/mole	2° Structure
WT	100	8	0	-
T9A	111	8	0	β-sheet
T9Q	66	24	-1.7	β-sheet
T9G	82	18	nd	β-sheet
T9E	33	47	nd	β-sheet
L10T	78	92	nd	β-sheet
L10F	<10	nd	nd	β-sheet
L10S	<10	nd	nd	β-sheet
L80V	<10	nd	nd	β-sheet
K88L	<10	nd	nd	loop
Y90L	<10	nd	nd	loop
Y90S	<10	nd	nd	loop
E96G	49	6	nd	loop
K97R	117	1	-0.1	loop
K97G	96	24	-1.1	loop
K97V	109	68	1.0	loop
IL-1α	116	68	nd	

[a]From densitometry of SDS gels, normalized to a standard *E. coli* protein; WT=100.
[b]From reversible Gdn-HCl unfolding monitored by tryptophan fluorescence, with ΔG values extrapolated to the Cm of the WT for comparison. Adapted from reference *(43)*. nd= not determined.

Conclusions

The combined IFN-γ and IL-1β results suggest that mutations which affect IB formation can be found in all sorts of secondary structures, and the variation in predicted surface exposure in the native structure of the IL-1β IB mutants shows there is no clear correlation with this parameter either. From this small sampling of two globular proteins it is also clear that specific mutational effects on IB formation can be observed both in proteins rich in α helix and proteins rich in β sheet. The IL-1β results show that IB formation, and mutational effects on IB formation, are not restricted exclusively to oligomeric proteins. Although this does not rule out the possibility that assembly intermediates may not sometimes be key players in IB formation, it is clear that simple monomeric proteins are also capable of mutational effects on IB formation.

It is worth noting that in both systems described here, two types of stability mutations are observed: those which lead to loss of material *via* proteolysis, and those which lead to altered levels of inclusion bodies. *In vivo* proteolysis of proteins can be regarded as depending on the impact of a mutation on the interplay between thermodynamic stability *(44)* and the sequence specificities of proteases and/or other factors which mediate degradation pathways *(45-47)*. It will be of great interest to see whether IL-1β stability mutations such as those described here will support these ideas.

The dramatic influences of seemingly subtle mutations on the cellular fate of IL-1β underscores the importance of off-pathway reactions as guiding forces in protein evolution. It is clear that protein sequence changes are tested by natural selection not

only for their contributions to the stability and function of the folded molecule, but also for their influences on the efficiency of the folding process, including the degree to which the folding process can escape off-pathway traps such as proteolysis and aggregation.

That the sequence requirements for a protein's activity can be different from the requirements for stability to proteolysis and inclusion body formation represents an opportunity for biotechnologists, especially in the field of industrial enzymes and other areas where stability and activity are emphasized and immunogenicity is not. One of the principle routes by which proteins become thermally inactivated *in vitro* is *via* aggregation-mediated pathways *(11)*. It is possible that folding intermediates often mediate this aggregation-dependent inactivation. To the extent that such key *in vitro* unfolding intermediates resemble folding intermediates involved in IB formation *in vivo*, mutants which escape IB formation in their cellular production may also prove to have improved *in vitro* stability properties *(9)*. Preliminary data for the IL-1β mutants suggests just such a relationship (BC and RW, unpublished data). Thus, isolation (by serendipity, design, or random mutagenesis and screening) of intragenic mutations associated with altered levels of IB formation may prove to be a useful tool for improving not only the expression characteristics of a desired protein, but also its *in vitro* stability characteristics.

Literature Cited

1. Shively, J.M. *Ann. Rev. Microbiol.* **1974**, 28, 167-187.
2. Prouty, W.F., Karnovsky, M.J. & Goldberg, A.L. *J. Biol. Chem.* **1975**, 250, 1112-1122.
3. Gribskov, M. & Burgess, R.R. *Gene*. **1983**, 26, 109-118.
4. Williams, D.C., Van Frank, R.M., Muth, W.L. & Burnett, J.P. *Science*. **1982**, 215, 687-689.
5. Wetzel, R. & Goeddel, D.V. in *The Peptides: Analysis, Synthesis, Biology;* Meienhofer, J. & Gross, E., Ed.; Academic Press: New York, **1983**; 5; 1-64.
6. Marston, R.A.O. *Biochem. J.* **1986**, 240, 1-12.
7. Mitraki, A. & King, J. *Biotech.* **1989**, 7, 690-697.
8. Schein, C.H. *Biotechnology*. **1989**, 7, 1141-1149.
9. Wetzel, R., Perry, L.J., Mulkerrin, M.G. & Randall, M. in *Protein Design and the Development of New Therapeutics and Vaccines; Proceedings of the Sixth Annual Smith, Kline and French Research Symposium;* Poste, G. & Hook, J.B., Ed.; Plenum: New York, **1990**; 79-115.
10. Wetzel, R. in *Protein Engineering - A Practical Approach;* Rees, A.R., Sternberg, M.J.E. & Wetzel, R., Ed.; IRL Press at Oxford University Press: Oxford, **1992**; In Press.
11. Wetzel, R. in *Stability of Protein Pharmaceuticals: In Vivo Pathways of Degradation and Strategies for Protein Stabilization;* Ahern, T.J. & Manning, M.C., Ed.; Plenum Press: New York, **1992**; In Press.
12. Browner, M.F., Rasor, P., Tugendreich, S. & Fletterick, R.J. *Protein Eng.* **1992**, 4, 351-357.
13. Shortle, D. & Meeker, A.K. *Biochem.* **1989**, 28, 936-944.
14. Glenner, G.G. *New England J. of Med.* **1980**, 302, 1283-1292; 1333-1343.
15. Cohen, A.S. & Connors, L.H. *J. Pathology*. **1987**, 151, 1-10.
16. Benson, M.D. & Wallace, M.R. in *The Metabolic Basis of Inherited Disease;* Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D., Ed.; McGraw-Hill: New York, **1989**; I; 2439-2460.
17. Pepys, M.B. in *Immunological Diseases;* Samter, M.D., Ed.; Little, Brown and Co.: Boston, **1988**; I; 631-674.
18. Stone, M.J. *Blood*. **1990**, 75, 531-45.

19. Buchner, J. & Rudolph, R. in *Current Opinion in Biotechnology;* Freedman, R. & Wetzel, R., Ed.; Current Biology,Ltd.: London, **1991**; 2:4; 532-538.
20. Haase-Pettingell, C.A. & King, J. *J. Biol. Chem.* **1988**, 263, 4977-4983.
21. Gething, M.-J. & Sambrook, J. *Nature*. **1992**, 355, 33-45.
22. King, J., Fane, B., Haase-Pettingell, C., Mitraki, A., Villafane, R. & Yu, M.-H. in *Protein Folding: Deciphering the Second Half of the Genetic Code;* Gierasch, L.M. & King, J., Ed.; American Association for the Advancement of Science: Washington, D.C., **1990**; 225-240.
23. Fane, B., Villafane, R., Mitraki, A. & King, J. *J. Biol. Chem.* **1991**, 261, In press.
24. Mitraki, A., Fane, B., Haase-Pettingell, C., Sturtevant, J. & King, J. *Science*. **1991**, 253, 54-58.
25. Krueger, J.K., Stock, A.M., Schutt, C.E. & Stock, J.B. in *Protein Folding: Deciphering the Second Half of the Genetic Code;* Gierasch, L.M. & King, J., Ed.; American Association for the Advancement of Science: Washington, D.C., **1990**; 136-142.
26. Truong, H.-T.N., Pratt, E.A., Rule, G.S., Hsue, P.Y. & Ho, C. *Biochem.* **1991**, 30, 10722-10729.
27. Fierke, C.A., Calderone, T.L. & Krebs, J.F. *Biochemistry*. **1991**, 30, 11054-11063.
28. Strandberg, L. & Enfors, S.-O. *Appl. Environ. Micro.* **1991**, 57, 1669-1674.
29. Goldenberg, D.P., Smith, D.H. & King, J. *Proc. Natl. Acad. Sci. USA*. **1983**, 80, 7060-7064.
30. Matteucci, M.D. & Heyneker, H.L. *Nucleic Acids Res.* **1983**, 11, 3113-3122.
31. Reidhaar-Olson, J.F. & Sauer, R.T. *Science*. **1988**, 241, 53-57.
32. Ealick, S.E., Cook, W.J., Vijay-Kumar, S., Carson, M., Nagabhushan, T.L., Trotta, P.P. & Bugg, C.E. *Science*. **1991**, 252, 698-702.
33. Rinderknecht, E. & Burton, L.E. in *The Biology of the Interferon System, 1984;* Kirschner, H. & Schellekens, H., Ed.; Elsevier: Amsterdam, **1985**; 397-402.
34. Mulkerrin, M.G. & Wetzel, R. *Biochem.* **1989**, 28, 6556-6561.
35. Wetzel, R., Perry, L.J. & Veilleux, C. *Biotech.* **1991**, 9, 731-737.
36. Wetzel, R., Perry, L.J., Veilleux, C. & Chang, G. *Protein Engineering*. **1990**, 3, 611-623.
37. Wetzel, R., Perry, L.J., Mulkerrin, M.G., Veilleux, C. & Chang, G. in *Protein Engineering '89: The Proceedings of the Second International Conference on Protein Engineering;* Ikehara, M., Oshima, T. & Titani, K., Ed.; Japan Scientific Societies Press: Tokyo, **1990**.
38. Priestle, J.P., Schar, H.-P. & Grutter, M.G. *Proc. Natl. Acad. Sci. USA*. **1989**, 86, 9667-9671.
39. Finzel, B.C., Clancy, L.L., Holland, D.R., Muchmore, S.W., Watenpaugh, K.D. & Einspahr, H.M. *J. Mol. Biol.* **1989**, 209, 779-791.
40. Clore, G.M., Wingfield, P.T. & Gronenborn, A.M. *Biochem.* **1991**, 30, 2315-2323.
41. Craig, S., Schmeissner, U., Wingfield, P. & Pain, R.H. *Biochemistry*. **1987**, 26, 3570-3576.
42. Young, P.R., Lillquist, J.S., Einstein, R., Lee, J., Fenderson, W., Porter, T., Kasyan, K., Green, D., Kumar, V., Sathe, G. & Simon, P.L. *Submitted.*
43. Chrunyk, B.A., Evans, J., Lillquist, J., Young, P. & Wetzel, R. *Ms. in preparation.*
44. Parsell, D.A. & Sauer, R.T. *J. Biol. Chem.* **1989**, 264, 7590-7595.
45. Bachmair, A., Finley, D. & Varshavsky, A. *Science*. **1986**, 234, 179-186.
46. Rogers, S., Wells, R. & Rechsteiner, M. *Science*. **1986**, 234, 364-368.
47. Bowie, J.U. & Sauer, R.T. *J. Biol. Chem.* **1989**, 264, 7596-7602.

RECEIVED May 5, 1992

Chapter 10

Characterization and Refolding of β-Lactamase Inclusion Bodies in *Escherichia coli*

Pascal Valax and George Georgiou

Department of Chemical Engineering, University of Texas, Austin, TX 78712

R_{TEM} β-lactamase was overexpressed in *E. coli* from three different plasmids resulting in the formation of cytoplasmic (plasmid pGB1) and periplasmic (plasmids pKN and pJG108) inclusion bodies. Previous work demonstrated that the inclusion bodies differ in structure and composition according to the cellular compartment in which aggregation occurred (*17*). In this study, we used inclusion bodies purified by sucrose density gradient centrifugation to investigate the effect of the *in vivo* aggregation environment on the properties of the protein within the inclusion bodies. Guanidine hydrochloride and pH solubilization experiments revealed important differences in the interactions involved in the stabilization of the aggregates. In addition, trypsin digestion results suggested a less ordered protein conformation in periplasmic inclusion bodies. The influence of the inclusion body origin on the renaturation of active protein was investigated in detail. The highest recovery was achieved with periplasmic inclusion bodies from RB791(pJG108). The yield of active β-lactamase upon refolding the material obtained from solubilized inclusion bodies was between 20% and 40% of that obtained from the renaturation of the purified protein under identical conditions. Our results suggest that the presence of certain contaminants in the inclusion bodies enhance the reaggregation of the protein during the removal of the denaturant.

High levels of expression of a cloned gene often results in the accumulation of misfolded recombinant protein into large, amorphous aggregates called inclusion bodies. Because of the complexity of intracellular events, very little is known about the mechanism of formation of inclusion bodies, but many growth parameters have been shown to affect *in vivo* aggregation. A high level of expression, and therefore high intracellular protein concentration, enhances aggregation (*1, 2*), whereas low growth temperature favors the production of recombinant proteins in a soluble form (*2-4*). The addition of non-metabolizable sugars, such as sucrose, to the growth medium has been shown to inhibit the formation of β-lactamase periplasmic inclusion bodies in *E. coli* (*1, 17*). Analogous effects of temperature, protein concentration and cosolvents have been

0097–6156/93/0516–0126$06.00/0

observed during the *in vitro* renaturation of purified protein from denaturant solutions (*5-8*). *In vitro* aggregation has been shown to result from the intermolecular interaction of exposed hydrophobic surfaces of a folding intermediate (*9, 10*). Based on analogies with *in vitro* results and recent direct *in vivo* experimental evidence, Mitraki and King suggested that the formation of inclusion bodies follows an similar mechanism (*11, 12*).

Very little information about the conformation of the peptide chains within the aggregates is available. Recent evidence suggests that the protein can be in either a completely misfolded or an active conformation (*13*). In most cases, the recovery of active, correctly folded protein from inclusion bodies involves the complete solubilization of the aggregates under strong denaturing conditions followed by refolding of the protein by dilution or dialysis (*14, 15*). Inhibition of reaggregation during the renaturation step requires careful optimization of the refolding conditions. The properties of the inclusion bodies themselves, such as density, structure and composition, must be taken into consideration for the design of an efficient recovery procedure.

The *E. coli* β-lactamase is a good model protein for the study of both *in vivo* and *in vitro* folding and aggregation (*18*). Overexpression of this protein in *E. coli* from different plasmids has been shown to result in the formation of periplasmic (*1, 16, 17*) or cytoplasmic (*18*) inclusion bodies. The structure and protein composition of β-lactamase inclusion bodies have been shown to depend on the intracellular compartment in which aggregation occurs (*18*). The influence of several growth conditions, such as temperature (*4*), expression level and presence of non-metabolizable sugars in the medium (*1, 17*) on the formation of the periplasmic inclusion bodies have been investigated. Furthermore, the effects of protein concentration, sucrose concentration and temperature on the *in vitro* refolding of pure β-lactamase from denaturant solutions have been shown to be in good agreement with *in vivo* observations (*8*). In this study, experiments were designed to assess differences in aggregated protein conformation among the various types of inclusion bodies. The effect of the origin of the aggregated material on the efficiency of functional protein recovery upon refolding was also investigated.

Materials and Methods.

Materials. Guanidine hydrochloride (GuHCl) was purchased from International Biotechnologies Inc. (New Haven, CT). Dithiothreitol (DTT) and sucrose were purchased from Sigma.

Cell Growth and Inclusion Body Purification. R_{TEM} β-lactamase was produced in *E. coli* RB791 (*22*) harboring the plasmids pGB1(*18*), pJG108 and pKN (*23-24*). pGB1 expresses the protein from a *tac* promoter and contains a deletion of the region encoding the -20 to -1 amino acids of the signal sequence. In pJG108, the native signal sequence of β-lactamase was replaced by the signal sequence of the outer membrane protein A (Omp A). pKN expresses native β-lactamase from a *tac* promoter and contains a kanamycin resistance gene.

Cultures were grown at 37°C in M9 medium supplemented with 0.2% glucose and 0.2% casein. The cultures were induced with isopropylthiolgalactoside (10^{-4} M final concentration) at an optical density ($O.D._{600}$) between 0.35 and 0.4. After overnight growth, the cells were harvested by centrifugation (8,000xg for 10 min). For all experiments, the inclusion bodies were isolated from cell lysates by isopycnic sucrose gradient centrifugation as follows: cells from 50 ml of culture were resuspended in 1 ml of 10 mM Tris-HCl, pH 8.0 containing 0.75 M sucrose and 0.2 mg/ml lysozyme. After 10 min incubation at room temperature, 2 ml of ice cold 3 mM

EDTA solution were added. The cells were then lysed in a French press at 20,000 psia. The lysates were subsequently centrifuged at 12,000xg for 30 min and the pellets, which contained the inclusion bodies, were resuspended in 1.25ml of 10 mM Tris-HCl buffer, pH 8.0, containing 0.25 M sucrose 1 mM EDTA and 0.1% sodium azide. The resuspended pellet material was layered on the top of a sucrose step gradient (40%, 53% and 67% w/w) in 1 mM Tris-HCl buffer, pH 8.0, containing 0.1% sodium azide and 1mM EDTA and centrifuged at 108,000xg for 90 min. The inclusion bodies focussed in a band at the interface between the 53% and 67% sucrose layers which was recovered and resuspended in water. The suspension was subsequently centrifuged at 12,000xg for 30 min. The pelleted material was resuspended in 0.25 M sucrose solution and applied to a second sucrose gradient as above. After recovery from the second gradient, the inclusion bodies were washed with water, resuspended in 50 mM potassium phosphate, pH 7.0 containing 0.1% sodium azide and stored at 4°C.

Solubilization Experiments. The effect of GuHCl concentration on the solubilization properties of the inclusion bodies was determined as follows. A known amount of β-lactamase aggregates was dissolved in 50 mM potassium phosphate buffer, pH 7.0 containing 5 mM DTT and various concentrations of GuHCl. The samples were incubated for three hours at room temperature and centrifuged at 8,000xg for 20 min to precipitate any remaining aggregated material. The protein concentration in the supernatant was determined using the Bio Rad assay.

To determine the effect of buffer pH on solubilization, the inclusion bodies were equilibrated with 0.45 M GuHCl and 5 mM EDTA in the following buffers (*25*): for adjusting the pH between 2.0 and 6.0, buffers were obtained by mixing 0.1 M citric acid solution with a 0.2 M dibasic sodium phosphate solution to specific ratios; buffers with pH values between 6.0 and 7.5 were obtained by mixing a 0.2 M monobasic sodium phosphate solution with a 0.2 M dibasic sodium phosphate solution; 0.2 M Tris buffers were adjusted to pH values between 7.5 and 8.5 and 0.2 M glycine buffers to pH values above 8.5 by addition of concentrated sodium hydroxide. After three hours of incubation at room temperature, the samples were centrifuged at 8,000xg for 20 min and the concentration of solubilized protein in the supernatant was determined as above.

Trypsin Accessibility Experiments. Identical quantities of inclusion bodies were placed in eppendorf centrifuge tubes and dissolved in 50 mM potassium phosphate buffer, pH 7.0. Trypsin was then added to a final concentration of 0.1 mg/ml in all the tubes except the control. The samples were incubated at room temperature for various times. The digestion was then stopped by adding soybean trypsin inhibitor to a final concentration of 0.2 mg/ml. The remaining aggregated material was precipitated by centrifugation at 8,000xg for 20 min. The pellets were resuspended in 50 mM potassium phosphate buffer, pH 7.0, containing 3 M GuHCl and 5 mM DTT and incubated for three hours at room temperature to completely solubilize the aggregated polypeptides (Figure 2). The amount of soluble protein, which corresponds to the remaining undigested material, was measured using the Bio-Rad assay. The tube in which no trypsin was added was used as a reference (100% intact protein).

Refolding of Inclusion Body Protein. β-lactamase inclusion bodies were incubated for three hours at room temperature in 50 mM potassium phosphate buffer, pH 7.0 containing 5 mM DTT and various concentrations of GuHCl. The total protein concentration was either 5 mg/ml or 1 mg/ml as described in the text. The samples were subsequently dialysed against 50 mM potassium phosphate buffer, pH 6.0 for three hours at room temperature in a PIERCE Inc. model 500 microdialyzer apparatus. For all experiments, the final GuHCl concentration was 0.015 M. The remaining aggregated material was precipitated by centrifugation at 8,000xg for 20 min. The total protein concentration and the β-lactamase activity in the supernatant were measured.

Finally, the purified inclusion bodies, the reaggregated protein and the soluble/refolded protein were loaded on a SDS polyacrylamide gel (15% acrylamide).

General Methods. Protein concentrations were measured using the Bio-Rad binding dye assay with bovine serum albumin as the standard. β-lactamase activities were determined spectrophotometrically using 0.5 g/l of penicillin G as the substrate (*19*). The dialysis tubing used in the microdialyzer apparatus was prepared by boiling in 2% sodium bicarbonate, 1 mM EDTA for 10 min, then boiling in 1 mM EDTA for another 10 min (*20*). The membrane was washed with distilled deionized water before and after each boiling. SDS-PAGE (15% acrylamide) was performed according to the method described by Laemmli (*21*). Prior to electrophoresis, the protein was denatured by boiling for 10 min in SDS electrophoresis buffer containing β-mercaptoethanol and bromophenol blue. The gels were stained with Coomassie brilliant blue.

Results and Discussion.

Inclusion Body Purification. Unlike the highly regulated cytoplasm, the periplasmic space of gram-negative bacteria is affected by the composition of the growth medium. Low molecular weight compounds (such as sucrose) can diffuse freely through the outer membrane and affect the folding and aggregation of secreted proteins (*1, 17*). To investigate the influence of the cellular environment on the formation and properties of inclusion bodies, β-lactamase was overexpressed from three different plasmids. In the plasmid pGB1, the signal sequence of β-lactamase has been deleted resulting in the expression of the mature protein preceded by the sequence Met-Arg-Ile. The absence of leader peptide prevents the translocation of the protein across the inner membrane and leads to the formation of cytoplasmic inclusion bodies consisting of the mature β-lactamase. Cells containing the plasmid pKN express β-lactamase with its native signal sequence from the *tac* promoter. Induction of the *tac* promoter results in the aggregation of mature β-lactamase in the periplasmic space. At high expression levels, a fraction of the precursor protein is unable to interact with the secretory apparatus fast enough and remains in the cytoplasm where it forms pre-β-lactamase inclusion bodies. In the plasmid pJG108, the native signal sequence of β-lactamase has been replaced by the leader peptide of outer membrane protein A. The pre-OmpA-β-lactamase gene is transcribed from the inducible lpp-lac promoter. Induction of protein synthesis leads to the formation of periplasmic inclusion bodies consisting of the mature β-lactamase

After overnight growth, the cells were harvested, converted to spheroplasts by treatment with lysozyme and EDTA and then lysed. The insoluble material from the lysates was then precipitated by low speed centrifugation. The pellets contain the protein aggregates along with other contaminants such as membrane debris, nucleic acids, etc. Removal of these contaminants is typically achieved by a series of extraction steps (*26, 18*). DNase I and lysozyme were used to degrade DNA and membrane fragments respectively. Membrane material and proteins adsorbed non-specifically on the surface of the inclusion bodies can be extracted by treatment with detergents such as deoxycholate and Triton X-100. This purification procedure, however, presents some major disadvantages. The detergents can also solubilize part of the inclusion body itself and therefore affect its properties and structure. The use of detergents can be circumvented by using sucrose density gradient centrifugation to separate the inclusion bodies from the cell debris (*18*). This method exploits the density difference between the protein particles and the other cellular components present in the lysates. The protein aggregates and the membrane debris form two distinct visible bands in the gradient. Silver stained polyacrylamide gels of the inclusion bodies did not reveal the presence of any of the major outer membrane proteins in the inclusion body fraction (*18*). Scanning electron micrographs of

cytoplasmic inclusion bodies from RB791(pGB1) purified by both methods are shown in Figure 1. Inclusion bodies purified by sucrose gradient centrifugation were morphologically different from those purified by detergent extraction. The former were more homogeneous and regular in shape. Therefore all inclusion bodies used in this study were purified by sucrose density gradient centrifugation.

Solubilization and Trypsin Accessibility Experiments. The inclusion bodies purified from the three plasmids exhibit different morphologies (*18*). Cytoplasmic inclusion bodies from RB791(pGB1) are highly regular, cylindrical aggregates with a homogeneous surface. Their size can exceed 1.5 μm. Periplasmic inclusion bodies from RB791(pJG108) and RB791(pKN) are small (0.5-1 μm), semi-spherical particles having smooth and rough surfaces. Considering such disparity in structure and surface characteristics, it is tempting to suggest that the conformation of the polypeptide chains within the aggregates and the strength of interchain associations depend on the cellular environment in which the inclusion bodies are formed.

The conformation of proteins in aqueous solvents is dictated by two types of interactions: hydrophobic interactions and electrostatic interactions which include ion pairing, hydrogen bonds and Van der Waals interactions (*27*). The strength of these interactions determines the overall stability of the protein tertiary structure. The intensity of these interactions depends in turn on the environment. The addition of denaturants such as GuHCl or urea to the solvent induces unfolding by weakening the hydrophobic interactions. Changes in pH, on the other hand, affect electrostatic interactions. The extent of inclusion body solubilization in the presence of GuHCl and at different pH values was studied in order to examine the strength of association between the protein molecules within the aggregates. The GuHCl solubilization profiles of the different types of inclusion bodies are shown in Figure 2. Cytoplasmic inclusion bodies from RB791(pGB1) could not be solubilized in the presence of up to 0.75 M GuHCl. At this denaturant concentration, inclusion bodies from RB791(pKN) and RB791(pJG108) experience substantial solubilization (25% and 50% respectively). Complete solubilization occurred at 2.5 M GuHCl, regardless of the origin of the inclusion bodies. These results suggest that the hydrophobic interactions stabilizing the associated protein chains within the cytoplasmic inclusion bodies are stronger than those found in inclusion bodies from RB791(pKN) and RB791(pJG108).

Figure 3 shows the solubilization profiles of the three types of inclusion bodies as a function of the buffer pH. The inclusion bodies were resuspended in buffers of different pH values containing 0.45 M GuHCl and 5 mM DTT. The experiment was conducted in the presence of a moderate concentration of GuHCl to amplify the effect of pH on the stability of the protein aggregates. While the aggregates from RB791(pGB1) show little solubilization over a wide pH range, the inclusion bodies from RB791(pJG108) and to a lesser extent from RB791(pKN) exhibit pH-dependent solubilization. The ionic strength of the buffer also appears to affect the extent of solubilization of the periplasmic inclusion bodies. At pH 7.0, in 50 mM potassium phosphate buffer, the extent of solubilization in 0.45 M GuHCl, 5 mM DTT was 10% for RB791(pKN) inclusion bodies and 30% for RB791(pJG108) aggregates (Figure 2). Under the same conditions but in 0.2 M sodium phosphate buffer, the extent of solubilization was 20% and 50% respectively (Figure 3). The sensitivity of the periplasmic inclusion bodies from RB791(pKN) and especially RB791(pJG108) to pH and ionic strength suggest that electrostatic interactions are involved in stabilizing the aggregate structure. On the other hand, the cytoplasmic inclusion bodies from RB791(pGB1) were not affected by the pH and ionic strength. Therefore it appears that hydrophobic interactions play a dominant role in protein association in the cytoplasm.

The susceptibility of proteins to proteolytic degradation depends on two factors: the presence of the sequences recognized by the protease and the steric accessibility of

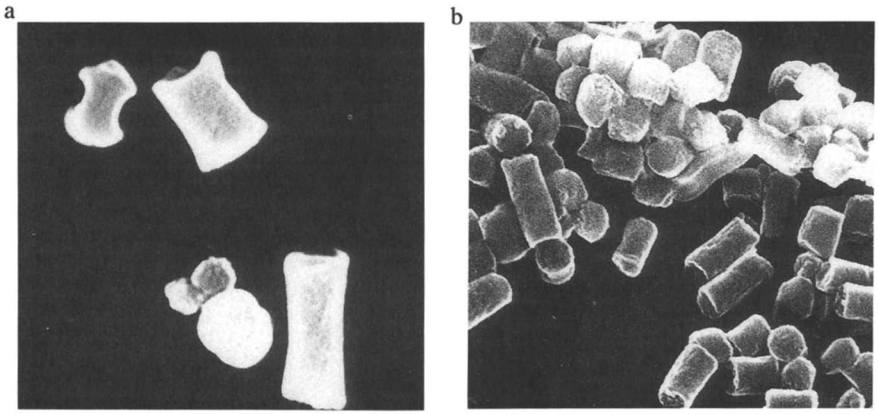


Figure 1. Scanning Electron Microscopy of cytoplasmic inclusion bodies from RB791(pGB1). a. Purified by detergent extraction. b. Purified by sucrose density gradient centrifugation.

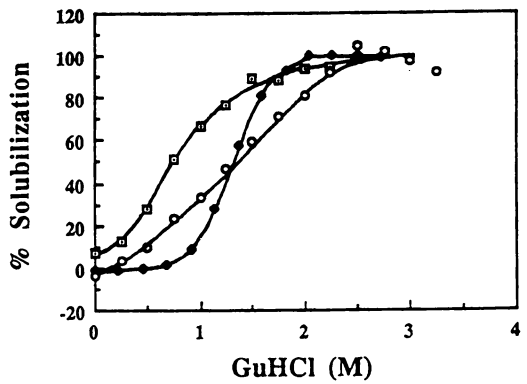


Figure 2. Extent of solubilization of sucrose density gradient centrifugation purified inclusion bodies by incubation with different concentrations of GuHCl. Inclusion bodies from RB791(pGB1) (◆), RB791(pKN) (○), RB791(pJG108) (⊡) were resuspended in 50 mM potassium phosphate, pH 7.0 containing 5 mM DTT and various concentrations of GuHCl. After three hours of incubation at 23°C the remaining aggregates were precipitated by centrifugation and the concentration of soluble protein was determined.

these sequences. Consequently, proteolytic degradation is highly dependent on the protein conformation. The mature β-lactamase has been shown to be extremely resistant to digestion by trypsin whereas the β-lactamase precursor and mutants exhibiting altered stability are readily degraded (*28-30*). Purified inclusion bodies were incubated in a 0.1 mg/ml trypsin solution for various times. The digestion was stopped by adding soybean trypsin inhibitor to a concentration of 0.2 mg/ml. The remaining aggregated material was precipitated by centrifugation, resuspended in potassium phosphate buffer containing 3 M GuHCl and 5 mM DTT and the concentration of solubilized protein was measured by the Bradford assay. Figure 4 shows that inclusion bodies from cells containing the plasmid pGB1 were the most resistant to degradation. The periplasmic inclusion bodies from RB791(pJG108), on the other hand, were digested extremely rapidly. After 20 min of incubation in presence of trypsin, 50% of the initial amount of protein present in the cytoplasmic inclusion bodies (from plasmid pGB1) was digested, compared to more than 70% for RB791(pKN) and about 95% for RB791(pJG108). SDS-PAGE of the aggregated protein remaining after trypsin digestion of the RB791(pKN) inclusion bodies revealed that the precursor was resistant to degradation. On the other hand, the degradation of the mature protein present in periplasmic inclusion bodies was extensive (*18*). These results suggest that the conformation of the polypeptide chains within the periplasmic aggregates is more accessible to trypsin compared to cytoplasmic inclusion bodies.

Raman spectroscopy can provide information on the conformation of insoluble proteins in membranes in the precipitated state (*31*). Preliminary studies showed that the protein in cytoplasmic inclusion bodies from RB791(pGB1) exhibits a high degree of α-helicity similar to the native soluble protein. The periplasmic inclusion bodies, on the other hand, contain a mixture of α-helical, β-sheet and random conformations (T. Przybycien, preliminary results). The higher level of organization of the cytoplasmic inclusion bodies is consistent with their greater resistance to GuHCl solubilization and trypsin digestion.

London et al. (*9*) demonstrated that *in vitro* aggregation of *E. coli* tryptophanase results from the specific intermolecular interactions between exposed hydrophobic surfaces of a folding intermediate. If, as proposed by Mitraki and King (*11, 12*), *in vivo* aggregation proceeds along a similar pathway, then the conformation of the polypeptide chains inside the inclusion bodies should reflect the conformation of the intermediates responsible for aggregation. Our studies show that the conformation of β-lactamase within the inclusion bodies depends on the cellular compartment in which aggregation occurs. This suggest that different association pathways may be responsible for the formation of different inclusion bodies. *In vitro*, β-lactamase aggregation from denaturant solutions has been shown to depend on the redox potential, the pH and the cosolvent composition of the renaturation buffer. Large variations of these three parameters occur across the membranes separating the different compartments of a cell. The cytoplasm is a highly regulated reducing environment of nearly constant pH and composition. On the other hand, the pH and the concentration of low molecular weight solutes in the periplasmic space are very sensitive to external conditions. Small molecular weight compounds can diffuse freely through the outer membrane and directly affect the formation of inclusion bodies (*1*). Differences in the protein composition within cellular compartments might also play an important role. The interaction of folding intermediates with cytoplasmic cellular components such as chaperonins (*32-38*) and other folding catalysts (*10*) could dictate the protein association pathway that leads to inclusion body formation. Other proteins could enhance aggregation by interacting non-specifically with the nascent polypeptide. In addition, cellular components such as nucleic acids and phospholipids could also affect the formation of inclusion bodies.

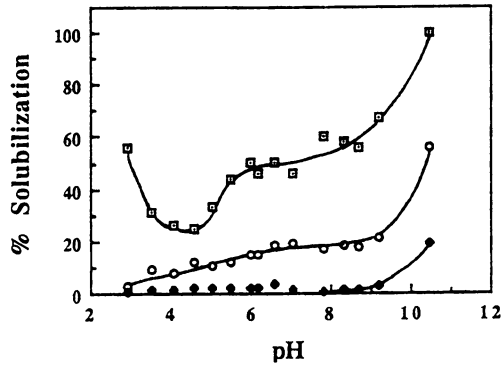


Figure 3. Solubilization of inclusion bodies from RB791(pGB1) (◆), RB791(pKN) (○), RB791(pJG108) (⧈) in buffers of various pH containing 0.45 M GuHCl and 5 mM DTT.

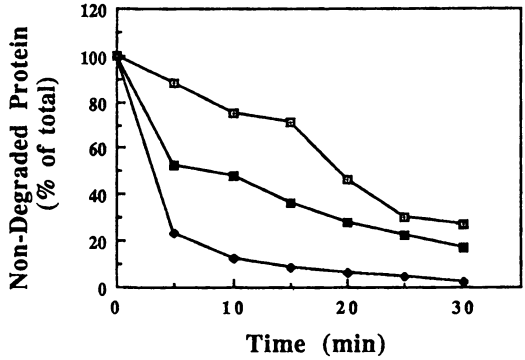


Figure 4. Digestion of inclusion bodies from RB791(pGB1) (⧈), RB791(pKN) (■), RB791(pJG108) (◆) by trypsin. The aggregates were incubated for various times in 0.1 mg/ml trypsin. The degradation was stopped with 0.2 mg/ml of soybean trypsin inhibitor.

Denaturation-Renaturation Experiments. The expression of recombinant protein into inclusion bodies is sometimes considered to be advantageous for several reasons. The aggregates can be easily precipitated from crude cell lysates by centrifugation. They contain the protein of interest at relatively high purity levels (40% or higher). Also aggregated protein is often more resistant to proteases within the cell. The limiting factor in deciding whether a protein should be expressed in an soluble form or as inclusion bodies resides in the difficulty to recover the native functional protein. The efficiency of the refolding process depends on the properties of the inclusion bodies such as density, solubilization characteristics and composition. SDS polyacrylamide gel electrophoresis shows that the protein composition of purified inclusion bodies varies according to their origin (Figure 7). We have showed that the solubilization characteristics of these inclusion bodies is also plasmid dependent. We therefore investigated in some detail the effect of the origin of the inclusion bodies on the efficiency of β-lactamase recovery upon renaturation. In this study, β-lactamase inclusion bodies were incubated for three hours at room temperature in 50 mM potassium phosphate buffer, pH 7.0, containing 3 M GuHCl and 5 mM DTT. The samples were subsequently dialyzed for three hours at room temperature against 50 mM potassium phosphate buffer, pH 6.0. This denaturation-renaturation procedure was completely reversible when purified β-lactamase was used at a concentration lower than 5 mg/ml. The protein recovery profiles for inclusion bodies from all three plasmids are shown in Figures 5 and 6. In all cases, the highest recoveries, both in term of activities and total soluble protein, were obtained with periplasmic inclusion bodies from RB791(pJG108). The activity recovered from RB791(pGB1) inclusion bodies was roughly half those obtained from RB791(pJG108). The lowest activity recoveries were obtained from RB791(pKN). Figures 5a and 6a also reveal that activities recovered at a total protein concentration of 1 mg/ml were about one fifth the activities obtained at 5 mg/ml. Also, the percent recovery of soluble protein obtained after renaturation at a total protein concentration of 1 mg/ml were practically identical to those obtained with 5 mg/ml (between 15% and 45% recovery depending on the plasmid). Our previous work on the refolding of purified β-lactamase clearly showed that the extent of aggregation depends on the protein concentration (*8*). When refolding of denatured β-lactamase was performed at concentrations greater than 5 mg/ml, the activity recovery decreased linearly decreasing as the protein concentration was increased. This type of behavior has been observed with numerous other proteins (*5-7*). The fact that the percent yields obtained from renaturation of inclusion body polypeptide at total protein concentrations of 5 mg/ml and 1 mg/ml were identical therefore seem to indicate that the recoveries achieved at both concentrations are the highest achievable in the conditions of the experiment. The low maximum recovery yield resulting from the solubilization-renaturation of inclusion body protein suggests that some of the contaminants integrated in the inclusion bodies enhance reaggregation of the protein during the refolding step.

To examine the effect of protein contaminants on the renaturation process, the reaggregated and the soluble fractions obtained from the refolding of 5 mg/ml of inclusion body protein from 5 M GuHCl were run on a SDS polyacrylamide gel (Figure 7). The soluble protein fraction obtained by refolding all three types of inclusion bodies was almost exclusively composed of mature β-lactamase. Essentially all the contaminating proteins initially present in the inclusion bodies were sequestered in the aggregates which were formed after the removal of the denaturant. This seems to indicate that the contaminating proteins tend to promote aggregation. Previous work on the refolding of tryptophanase (*9*) clearly showed that the interactions leading to aggregation of this protein are highly specific. Addition of other proteins or even crude cell lysate did not affect the recovery yields. On the other hand, the interactions leading to the aggregation of egg white lysozyme have been shown to be highly non-specific (*10*). The effect of protein contaminants on the refolding yield cannot be assessed at this point. Other types of contaminants such as phospholipids may also influence the

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

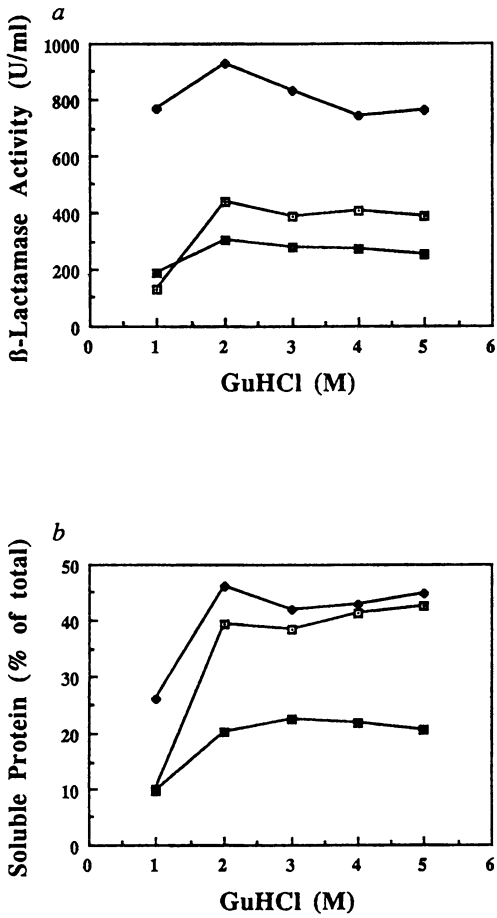


Figure 5. β-lactamase activity (a) and percent of soluble protein (b) obtained upon refolding of inclusion body proteins from RB791(pGB1) (⊡), RB791(pKN) (■) and RB791(pJG108) (◆). The inclusion bodies were first solubilized in 50 mM potassium phosphate buffer, pH 7.0 containing 5 mM DTT and various concentrations of GuHCl. The samples were then dialyzed against 50 mM potassium phosphate buffer, pH 6.0 for three hours at 23°C. The protein concentration was 1 mg/ml. The final GuHCl concentration was 0.015 M.

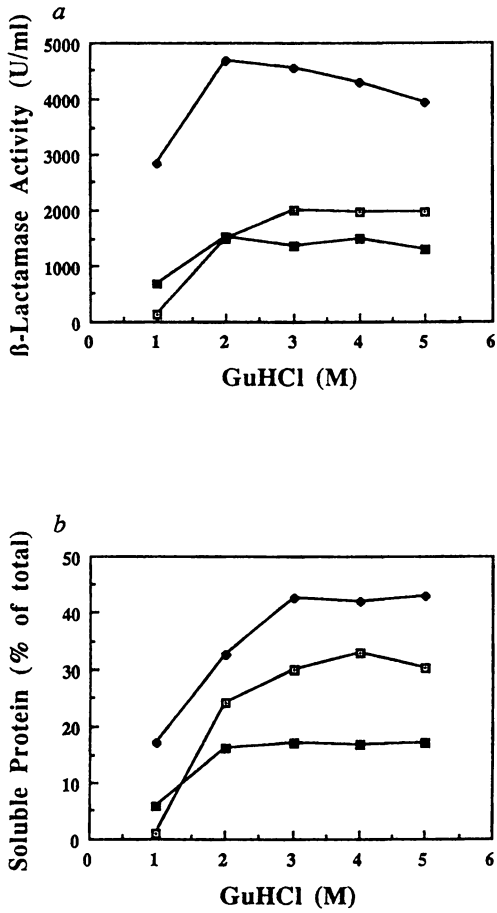


Figure 6. Same as Figure 5 except that the concentration of inclusion bodies was 5 mg/ml.

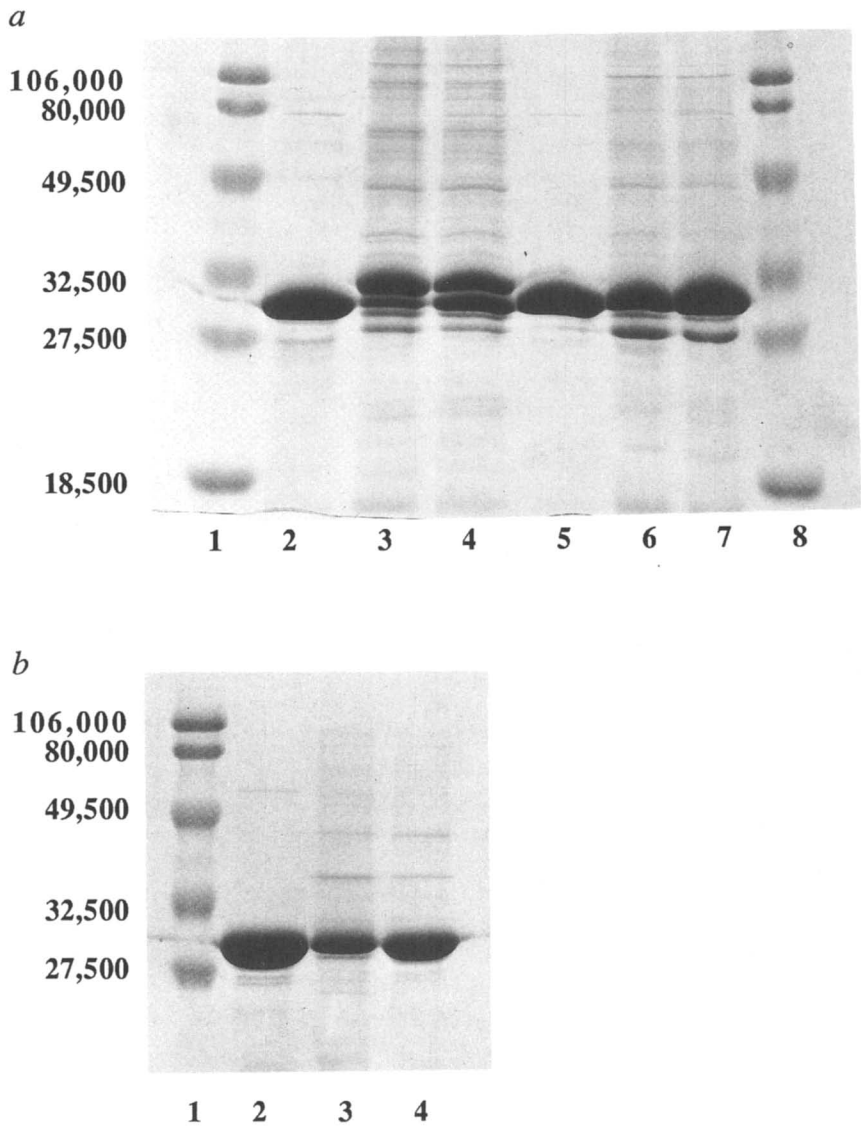


Figure 7. SDS-PAGE analysis of the renaturation of inclusion body protein. The samples correspond to the renaturation of 5 mg/ml of protein from potassium phosphate buffer pH 7.0 containing 5 M GuHCl, 5 mM DTT into the same buffer, pH 6.0. a. Lanes 2-4: RB791(pKN); Lanes 5-7: RB791(pGB1); Lanes 1 and 8 are molecular weight standards. b. Lanes 2-4: RB791(pJG108); Lane 1 is the molecular weight standard. Lanes 2a, 5 and 2b correspond to the soluble protein obtained after renaturation. Lanes 3a, 6 and 3b correspond to the reaggregated protein. Lanes 4a, 7 and 4b show the composition of the purified inclusion bodies.

recovery of active protein. The fact that the β-lactamase precursor of the RB791(pKN) inclusion bodies was found exclusively in the aggregated fraction is not surprising since the leader sequence of β-lactamase is very hydrophobic and has been shown to slow the folding kinetics (*39*), factors which favor aggregation.

In this report, we have shown that the structure of inclusion bodies and the conformation of the polypeptide chains within the inclusion bodies depend on the cellular compartment where aggregation occurs. These results suggest that the association mechanisms leading to the different types of inclusion bodies are different. Furthermore, the renaturation studies indicate that some cellular components present in the inclusion bodies may affect the recovery of active protein upon refolding.

Acknowledgements.

We are grateful to Angel Paredes for his help with Scanning Electron Microscopy. We also thank Gregory Bowden for the useful discussions and Daniel Thomas for his assistance. This work was supported by grant CBT 86-57971.

Literature cited.

1. Bowden, G. A.; Georgiou, G. *Biotechnology progress* **1988**, *4(2)*, 97-101
2. Takagi, H.; Morinaga, Y.; Tsuchiya, M.; Ikemura, H.; Inouye, M. *Bio/technology* **1988**, *6*, 948-950
3. Schein, C. H.; Noteborn, M. H. M. *Bio/Technology* **1988**, *6*, 291-294
4. Chalmers, J. J.; Kim, E.; Telford, J. N.; Wong, E. Y.; Tacon, W. C.; Shuler, M. L.; Wilson B. D. *Applied and Environmental Microbiology* **1990**, *56(1)*, 104-111
5. Mitraki, A.; Betton, J. M.; Desmadril, M.; Yon, J. M. *Eur. J. Biochem.* **1987**, *163*, 29-34
6. Burton, S. J.; Quirk, A. V.; Wood, C. P.; *Eur. J. Biochem.* **1989**, *179*, 379-387
7. Cleland, J. L.; Wang, D. I. C. *Bio/technology* **1990**, *8*, 1274-1278
8. Valax, P.; Georgiou, G. in *Protein Refolding*; Georgiou, G.; De Bernardez-Clark, E., Ed.; ACS Symposium Series 470; American Chemical Society, 1991, 97-109
9. London, J.; Skrzynia, C.; Goldberg, M. E. *Eur. J. Biochem.* **1974**, *47*, 409-415
10. Goldberg, M. E.; Rudolph, R.; Jaenicke, R. *Biochemistry* **1991**, *30*, 2790-2797
11. Mitraki, A.; King, J. *Bio/Technology* **1989**, *7*, 690-697
12. Mitraki, A.; Haase Pettingell, C.; King, J. in *Protein Refolding*; Georgiou, G.; De Bernardez-Clark, E., Ed.; ACS Symposium Series 470; American Chemical Society, 1991, 35-49
13. Tokatlidis, K.; Dhurjati, P.; Millet, J.; Béguin, P.; Aubert, J. P. *FEBS* **1991**, *282(1)*, 205-208
14. Lim, W. K.; Smith Somerville, H. E.; Hardman, J. K. *Applied and Environmental Microbiology* **1989**, *55(5)*, 1106-1111
15. Buchner, J.; Rudolph, R. *Bio/Technology* **1991**, *9*, 157-162
16. Georgiou, G.; Telford, J. N.; Shuler, M. L.; Wilson, D. B. *Applied and environmental microbiology* **1986**, *52(5)*, 1157-1161
17. Bowden, G. A.; Georgiou, G. *J. Biol. Chem.* **1990**, *265(28)*, 16760-16766
18. Bowden, G. A.; Paredes, A. M.; Georgiou, G. *Bio/Technology* **1991**,
19. Sigal, I. S.; DeGrado, W. F.; Thomas, B.J.; Petteway, Jr. S. R. *The Journal of Biological Chemistry* **1984**, *259(8)*, 5327-5332

20. Maniatis, T.; Fritsch, E. F.; Sambrook, J. *Molecular Cloning, A Laboratory Manual*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1982; 456
21. Laemmli, U. K. *Nature* **1970**, *227*, 680-685
22. Brent, R.; Ptashne, M. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4204-4208
23. Ghrayeb, J.; Kimura, H.; Takahara, M.; Hsiung, H.; Masui, Y; Inouye, M. *EMBO J.* **1984**, *3*, 2437-2442
24. Georgiou, G.; Shuler, M. L.; Wilson, D. B. *Biotechnol. Bioeng.* **1988**, *32*, 741-748
25. Gomori, G. In *Methods in Enzymology* ; Academic Press, N.Y., vol. 1, 138-146
26. Marston, F. A. O. In *DNA Cloning: A Practical Approach*; Glover D. M. Ed.; IRL Press, Oxford, 1987; Vol. 3, 59-89
27. Dill, K. A. *Biochemistry* **1990**, *29(31)*, 7133-7155
28. Minsky, A.; Summers, R. G.; knowles, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4180-4184
29. Dalbadie-McFarland, G.; Neitzel, J. J.; Richards, J. H. *Biochemistry* **1986**, *25*, 332-338
30. Georgiou, G.; Baneyx, F. In *Biochemical Engineering IV*; Goldstein, W. E.; DiBiasio, D.; Petersen, H., Eds.; Annals of the New York Academy of Sciences; New York Academy of Science, New York, 1990; Vol. 589,139-147
31. Przybycien, T. M.; Bailey, J. E. *Biochemica et Biophysica acta* **1989**, *995*, 231-245
32. Rothman, J. E. *Cell* **1989**, *59*, 591-601
33. Ellis, R. J.; Hemmingsen, S. M. *TIBS* **1989**, *14(8)*, 339-342
34. Lubben, T. H.; Donaldson, G. K.; Vitanen, P. V.; Gatenby, A. A. *The plant cell* **1989**, *1*, 1223-1230
35. Goulobinoff, P.; Christeller, J. T.; Gatenby, A. A.; Lorimer, G. H. *Nature* **1989**, *342*, 884-889
36. Kusukawa, N.; Yura, T.; Ueguchi, C.; Akiyama, Y.; Ito, K. *The EMBO Journal* **1989**, *8(11)*, 3517-3521
37. Van Dyk, T. K.; Gatenby, A. A.; LaRossa, R. A. *Nature* **1989**, *342*, 451-453
38. Phillips, G. J.; Silhavy, T. J. *Nature* **1990**, *344*, 882-884
39. Laminet, A. A.; Plückthun, A. *The EMBO Journal* **1989**, *8(5)*, 1469-1477

RECEIVED March 31, 1992

Chapter 11

Participation of GroE Heat Shock Proteins in Polypeptide Folding

Anthony A. Gatenby, Gail K. Donaldson, François Baneyx, George H. Lorimer, Paul V. Viitanen, and Saskia M. van der Vies

Central Research and Development, E. I. du Pont de Nemours and Company, Experimental Station, Wilmington, DE 19880–0402

The *in vivo* folding and assembly of numerous proteins, previously considered to be a spontaneous process, appears to be significantly influenced by a class of proteins termed molecular chaperones. The GroE proteins of *Escherichia coli* are the best characterized molecular chaperones, at both the genetical and biochemical level, and have been used extensively to investigate interactions between this class of proteins and a number of target polypeptides. It now appears that molecular chaperones exert their influence by stabilizing protein folding intermediates, thus partitioning them towards a pathway leading to the native state rather than forming inactive aggregated structures. GroE proteins are active during the early steps in folding when aggregation-prone intermediates are abundant.

Chaperonins are a highly conserved family of proteins that are present in most, and probably all, living organisms. They are abundant proteins, and in many microorganisms the levels of synthesis can be dramatically enhanced in response to stress. Thus, they are frequently referred to as heat shock or stress proteins, and one of their cellular functions is probably to provide protection under stress conditions. The chaperonins themselves are members of a larger class of proteins called molecular chaperones that are broadly defined as proteins that influence the folding of other proteins, and yet are not components of the final functional structure (*1*). This expansive definition has resulted in many proteins now being reconsidered as molecular chaperones, for example nucleoplasmin, hsp70, signal recognition particle, SecB and several others (see *2-6* for recent reviews).

Current advances using a number of different experimental systems has led to the realization that protein folding within cells is more complex than was initially thought. It was assumed that since a number of chemically denatured proteins can successfully fold into their functional native structures *in vitro* using information contained in the primary amino acid sequence (reviewed in *7*), then protein folding *in vivo* should also be a spontaneous event. The reality, at least for some proteins in the cell, may be quite different. This is due, in part, to a temporal element, since polypeptide chains emerging from ribosomes in a linear fashion may initiate folding before translation of the mRNA is completed. Therefore, the information in the primary sequence required for successful folding may not be simultaneously available. Similar constraints would also apply to polypeptides during translocation through

0097–6156/93/0516–0140$06.00/0

membranes prior to correct folding. In addition, the high concentrations of proteins in the cell at various stages of folding, and with potentially interactive surfaces, must coexist without mutual impairment of folding pathways (reviewed in *8, 9*). Molecular chaperones appear to have evolved to interact with the non-native states of proteins in cells and to partition polypeptides towards productive folding pathways. In this chapter the role in protein folding of the most structurally complex of the molecular chaperones, the chaperonins, will be described. We shall concentrate on studies that have involved the bacterial GroES and GroEL chaperonins. Several reviews and papers describing chaperonins from other organisms have been published (*1, 2, 4, 10-20*).

Chaperonin Molecules

There are two basic types of chaperonins of quite distinct sizes (Table I). The larger type contains subunits with an approximate molecular mass of 60 kDa that are arranged in a complex structure comprising two rings of seven subunits stacked on each other (*21, 22*). This type of molecule is known as GroEL (in bacteria), hsp60 (in yeast), rubisco subunit binding protein (in plants), P1 protein (in mammalian mitochondria), or collectively as chaperonin 60 (cpn60) on account of its subunit size. The second type of chaperonin (or co-chaperonin) is significantly smaller and contains subunits of about 10 kDa and, at least in bacteria, these are thought to form a single ring of seven subunits (*23*). This smaller heptameric protein isolated from *E. coli* is called GroES and a related protein in mammalian mitochondria is known as cpn10 (*19*), again reflecting its subunit size. GroEL possesses a weak ATPase activity that requires magnesium and potassium ions for maximal activity (*23, 24*). The ATPase is effectively suppressed when GroES binds to GroEL, an event that itself requires the presence of MgATP (*23, 24*). Self-assembly of the oligomeric form of GroEL from monomers also requires MgATP (*25*). Although the quaternary structure of bacterial GroEL oligomers has been observed by electron microscopy (*21, 22*), the tertiary structure is unknown.

The GroES and GroEL chaperonins from *E. coli* were the first to be studied in detail. The GroE proteins were originally identified because mutations in their genes prevented the growth of several bacteriophages (reviewed in *26*). Subsequent studies revealed that the *groES* and *groEL* genes are essential for bacterial survival (*27*), and that they constitute an operon whose expression is enhanced during heat shock. During a stress response the cellular level of GroEL can be increased from about 2% to 10% of cell protein (*28*). The GroES and GroEL proteins functionally interact (*23, 24*) and influence bacteriophage growth at the level of capsid assembly (*26*). Proteins related to either GroES or GroEL have now been identified in numerous prokaryotic organisms, and display a high degree of amino acid sequence homology (*26*).

GroE Chaperonins Influence Protein Folding and Assembly *in vivo*

Early studies with various bacteriophages and *groE* defective strains demonstrated that chaperonins influenced assembly of head or tail structures (depending on the particular phage), and the sites of these interactions were genetically defined (*26*). It was also established that the *E. coli* GroEL protein was related by amino acid sequence homology to the chloroplast cpn60 (*1*), a protein implicated in the assembly of the photosynthetic enzyme ribulose bisphosphate carboxylase/oxygenase (rubisco) (*16-18*). The rubisco holoenzyme has a rather complex structure of eight large (L, 52 kDa) and eight small (S, 12 kDa) subunits. An expression system which directs synthesis and assembly of the L_8S_8 rubisco holoenzyme from the cyanobacterium *Anacystis nidulans* was developed in *E. coli* (*29*). It therefore became possible to test whether folding and assembly of the photosynthetic enzyme in *E. coli* was influenced by the GroE proteins.

Table I. Molecular Properties of *E. coli* GroE Chaperonins

GroEL (cpn60, hsp60)	
	ubiquitous (bacteria, fungi, plants and animals)
	a stress induced protein in some organisms
	essential for bacterial cell growth - no deletion mutants described
	subunit molecular mass - 60 kDa
	native molecular mass - 840 kDa
	tetradecamer - two layers of seven subunits
	weak K^+-dependent ATPase, k_{cat} 0.1 s^{-1} (based on protomer)
	interacts with GroES in the presence of MgATP
	requires GroES to be functional *in vivo*
	stabilizes folding intermediates of several proteins in binary complexes
	groE operon contains *groES* and *groEL* genes
GroES (cpn10)	
	believed to be ubiquitous (bacteria, chloroplasts, mitochondria)
	a stress induced protein in some organisms
	essential for bacterial cell growth - no deletion mutants described
	subunit molecular mass - 10 kDa
	native molecular mass - 70 kDa
	heptamer - single ring of subunits
	inhibits ATPase activity of GroEL
	aids in release of proteins bound to GroEL

To demonstrate that GroE proteins influence the assembly of *A. nidulans* rubisco in *E. coli*, the level of cellular GroE proteins was increased by cloning the *groE* genes on a multicopy plasmid. The *groE* operon was transferred to a chloramphenicol-resistant plasmid (pGroESL) that is compatible with an ampicillin-resistant plasmid (pANK1) that directs the synthesis of rubisco subunits. The presence of the *groE* plasmid increases the levels of GroE proteins in the cell to about 20-30% of total protein (*30*). This high concentration of chaperonins in the cell resulted in a concomitant increase in the level of active L_8S_8 rubisco holoenzyme.

Although chaperonin overexpression resulted in nearly a ten-fold increase in assembled rubisco, there was little change in the amount of rubisco polypeptides. Therefore, the GroE proteins influenced one or more steps in the assembly of the rubisco holoenzyme and not the synthesis or stability of the subunit polypeptides. Additional experiments revealed that overexpression of both *groES* and *groEL* was required for enhanced rubisco assembly, and that *groE* defective strains that prevent bacteriophage morphogenesis also inhibited hexadecameric rubisco assembly (*30*). A structurally simpler dimeric form of rubisco with two large subunits has also been characterized. Assembly of this dimeric form of the rubisco enzyme was similarly influenced by the chaperonin concentration in the cell. Since dimers of L subunits are the basic structural motifs in the more complex L_8S_8 enzyme, assembly of L dimers may be an essential intermediary stage in the formation of the L_8 core, and therefore the influence of GroE proteins on the basic common step of dimer formation can control the subsequent assembly of the more complex form of rubisco. Recent studies have also demonstrated an involvement of GroEL in *nif* gene regulation and nitrogenase assembly (*31*).

GroE Proteins Interact with Many Bacterial Polypeptides

Although a clear involvement of GroE proteins in bacteriophage and rubisco assembly can be demonstrated, these are of course protein targets that are not usually present in *E. coli* cells. In attempts to define the normal role of chaperonins in bacterial cells, the dual techniques of genetic suppression and direct identification of GroEL complexed with target polypeptides have been of value.

Suppression of Heat-Sensitive Mutations. Because successful polypeptide folding can be significantly influenced by temperature, it was suspected that some heat-sensitive mutations in bacteria could be folding mutants. In these types of mutants incubation at non-permissive temperatures might lead to destabilization of folding intermediates, with subsequent aggregation or proteolysis, resulting in the observed growth defects. Increased expression of the GroE chaperonins in a range of heat-sensitive mutants grown at non-permissive temperatures would therefore allow the identification of chaperonin-suppressible mutations.

Initially, heat-sensitive mutations in the *ilv* operon in *Salmonella typhimurium* encoding enzymes of branched chain amino acid biosynthesis were examined. The auxotrophic requirements of one mutation in *ilvGM* and two mutations in *ilvE*, encoding the multimeric enzymes acetolactate synthase II and transaminase B respectively, are suppressed by a multicopy plasmid encoding the *groE* operon from *E. coli* (*32*). These observations were extended by examining the response of numerous mutations to GroES and GroEL overexpression (Table II). Many, but not all, heat-sensitive *hisD*, *hisC* and *hisB* alleles are suppressed by multicopy *groE* plasmids. These genes encode histidinol dehydrogenase, histidinol-phosphate aminotransferase and the bifunctional protein imidazoleglycerol phosphate dehydratase:histidinol phosphate phosphatase, respectively. Suppression of cold-sensitive and temperature-independent alleles was not observed in the *his* system. The presence of the *groE* plasmid also suppresses heat-sensitive growth and secretion defects caused by the *secA51* and *secY24* alleles of *E. coli* (*32*). Thus, mutations causing alterations in enzymes and the protein translocation apparatus are subject to multicopy suppression by *groE*. It should be noted that this suppression requires overexpression of both *groES* and *groEL*. In addition, the levels of GroE proteins required are high, and represent about 20-30% of total cell protein. This is significantly greater than the increased levels of GroE proteins that would be synthesized in cells containing only the chromosomal *groE* operon when plated at the higher non-permissive temperatures.

Table II. Multicopy *groE* Suppression of *Salmonella typhimurium* Heat Sensitive Mutations

Gene	*Enzyme*	*Suppression*[a]
hisB	imidazoleglycerol phosphate dehydratase:histidinol phosphatase	1 of 3
hisC	histidinol-phosphate aminotransferase	2 of 7
hisD	histidinol dehydrogenase	7 of 7
ilvE	transaminase B	2 of 2
ilvGM	acetolactate synthase	1 of 11

[a]Suppression expressed as numbers of strains growing at non-permissive temperatures out of the number of mutations tested. Adapted from (32).

Overexpression of the *groE* operon in *E. coli* is also known to suppress the heat-sensitive phenotype of several *dnaA* alleles (*33, 34*), suggesting an interaction between the *groE* products and the dnaA protein at the origin of replication. These results, taken together with our observations on general *groE* multicopy suppression, suggest a widespread interaction of GroE proteins with other cellular proteins. Although high concentrations of GroE proteins are required to suppress these heat-sensitive mutations, possibly by trapping unstable mutant proteins to allow partitioning towards correct folding, it is possible that normal levels of chaperonins in cells could help correct errors in protein folding and thus alleviate some weak genetic folding defects.

Direct Interaction of GroEL with Many Bacterial Proteins. Evidence to support the genetic indications that many *E. coli* proteins interact with chaperonins has been obtained by forming stable complexes between GroEL and a random assortment of bacterial proteins (*35*). As described later, it is now well established that stable binary complexes can be formed between GroEL and the non-native states of several purified proteins. These complexes are often sufficiently stable to be isolated by gel filtration or non-denaturing gel electrophoresis. Rather than examining the interactions between GroEL and a large range of purified proteins on an individual basis, we chose to directly examine the interactions of proteins from a total *E. coli* cell extract with the purified chaperonin.

E. coli cells were grown for several generations on minimal medium containing [^{35}S]methionine and a soluble protein extract was prepared. When the native mixture of proteins was incubated with GroEL, and then fractionated on a gel filtration column, very little material was associated with the GroEL peak. This indicated that the native states of proteins in a complex mixture were unable to bind to GroEL in a stable fashion. The failure of native states to interact with GroEL has been observed for a number of purified proteins. In contrast, if the total radiolabelled mixture of proteins is denatured with guanidine-HCl followed by dilution of the chaotrope in the presence of GroEL, stable binary complexes are formed with the chaperonin that can be resolved by gel filtration. It appears that about half of the labelled proteins in their unfolded or partially folded states can interact with GroEL, and are efficiently discharged following the addition of MgATP. This may be an underestimate of the number of cellular proteins that can interact with chaperonins, since some proteins may fold so rapidly that they do not have an opportunity to

interact with GroEL, and others may not bind with a sufficiently high affinity to withstand gel filtration. Binary complex formation is reduced when the transient species that interact with GroEL are allowed enough time to progress to more stable states. This suggests that the structural elements that are recognized by GroEL, and that allow complex formation, are only present or accessible in the unfolded or partially folded states of many proteins.

The combined observations of high-affinity binding of many cellular proteins, and the natural abundance of GroEL, taken together with genetic suppression studies, lead us to suspect that the folding of most proteins in *E. coli* does not occur spontaneously while they are in solution, but probably takes place while they are associated with chaperonins. The common interaction between chaperonins and numerous unrelated target proteins has also been observed in eukaryotic organelles when these proteins are either synthesized within chloroplasts (*16-18*) and mitochondria (*36*), or following translocation through the membranes (*13, 20, 37*).

Refolding of Purified Proteins *in vitro* is Influenced by Chaperonins

Protein Folding and Aggregation. In general, the isomerization of proteins from the unfolded (U) state to the native (N) state involves the transient formation of folding intermediates (I). Although initially considered as a highly cooperative single step process, an increasing number of multistep folding pathways have been identified (*38, 39*) that can be populated by distinct partially folded states (*40*). The folding of many proteins analyzed *in vitro* can be considered in its simplest form as a two step process (*39, 41*).

$$U \rightarrow I \rightarrow N$$

The initial fast step involves the conversion of the unfolded (U) polypeptide to the intermediate (I) or molten globule state. The U state will not be highly populated at any one time because of this rapid conversion to the thermodynamically more stable I state. The I or molten globule state is a collapsed yet mobile structure that results from the rapid formation of specific secondary structural elements on a millisecond time-scale. Although the I state is almost as compact as N, it lacks the close packing of the secondary structural elements typical of the N state, and so the amino acid side chains exhibit greater fluctuation and the core residues are accessible to solvent molecules. However, in the second rate-determining step these elements become organized into the specific tertiary structures associated with the N state. In contrast to the rapid and uncooperative transitions between the U and I states, those between the I and N states are both slow and cooperative. The I, or molten globule, state probably has a greater number of hydrophobic residues exposed on its surface than in the N state because the I to N transition is associated with a large change in both enthalpy (ΔH) and in heat capacity (ΔC_p). The I states of proteins are characteristically less soluble than N and have a propensity to aggregate. This predilection for aggregation by the I state most likely results from the interaction of the exposed hydrophobic surfaces with each other. The kinetic partitioning of the I state between the N state or an aggregated state (I_{agg}) is therefore of considerable significance for the efficient folding of a protein under a given set of conditions.

$$\begin{array}{ccccc} U & \rightarrow & I & \rightarrow & N \\ & & \downarrow & & \\ & & I_{agg} & & \end{array}$$

Fortunately, it is possible to control some of the factors that favor aggregation *in vitro* to obtain successful refolding of denatured polypeptides. The simplest ways to suppress aggregation during *in vitro* refolding are to reduce either the concentration

of the protein, the temperature of the refolding reaction, or perhaps both. Although the I to N transition is an isomerization reaction, and is therefore independent of the concentration of I, aggregation is highly dependent on [I] since it is a second or higher order process (*42*). Aggregation can, therefore, occur much faster than first-order folding, and can exceed successful folding as protein concentration increases. Consequently, folding of the maximum number of molecules occurs in very dilute solutions where intermolecular interactions can be reduced (*43*). Interactions between hydrophobic surfaces, such as those found on molten globules, can be minimized by the simple expedient of reducing the temperature. By suppressing these hydrophobic interactions at low temperatures, conditions that would lead to aggregation can be avoided and partitioning to the native state favored (*24*)

In the complex and dynamic cellular milieu it is obviously not feasible to reduce protein concentrations to the levels that would prevent intermolecular aggregation events occurring between unfolded or partially folded molecules. Similarly, a reduction in temperature to reduce hydrophobic interactions is not an approach that has found favor in complex biological systems. Because aggregation reactions are both temperature and concentration dependent, conditions within cells thus appear to be particularly unsuitable for protein folding reactions in the normal physiological range of temperatures and protein concentrations. To resolve this dilemma, cells have apparently evolved molecular chaperones to aid in protein folding reactions *in vivo*.

Interactions Between Chaperonins and Other Proteins. The folding and interactions of several different purified proteins in the presence of GroE chaperonins has now been studied in some detail. Current examples include rubisco (*24, 35, 44, 45*), pre-ß-lactamase (*46*), rhodanese (*47-49*), dihydrofolate reductase (*49, 50*), citrate synthase (*51*), α-glucosidase (*52*), lactate dehydrogenase (*53*) and several thermophilic enzymes (*54*). In addition, the binding and release of proteins to GroEL in cell extracts has been examined (*55, 56*). Some common themes have emerged. The first, and perhaps most striking observation, is that the folding of many different proteins is influenced by chaperonins. These proteins bear little resemblance to each other with regard to size, shape, function or cellular location. The interaction of GroE chaperonins with target proteins during folding is therefore a very general mechanism that enables cells to exert some control over the isomerization of a broad range of molecules.

The Association Step. Clearly, the question of the specificity of the interactions between chaperonins and these target proteins becomes paramount. Evidence obtained to date indicates that once proteins have folded to their N states they do not appear to interact with chaperonins, and by implication the structural motifs that impart recognition are either inaccessible or no longer present. These motifs must consequently be exposed on the non-native I states of proteins. Since exposed hydrophobic residues are a characteristic feature of folding intermediates, perhaps these allow the chaperonin·I state recognition event. Based on the observation that chaperonins can substitute for nondenaturing detergents to obtain successful folding of rhodanese, it was proposed that the interactions of hydrophobic surfaces that lead to aggregation can be prevented by the binding of GroEL to partly folded intermediates (*47*). The presence of hydrophobic surfaces on GroEL itself is suggested by binding of the hydrophobic fluorescent reporter bisANS (*47*). Studies on the chaperonin-dependent folding of the monomeric enzymes dihydrofolate reductase and rhodanese indicate that GroEL stabilizes these proteins in a structure that resembles the molten globule state (*49*). The fluorescence properties of α-glucosidase bound to GroEL also suggests a molten globule state for the target protein (*52*). In contrast, from measurements on the interaction of lactate dehydrogenase with chaperonins it was concluded that GroEL binds to the unfolded and the first transient intermediate in the folding pathway, and not to other later structures such as the molten globule state (*53*).

It has also been suggested that improperly folded proteins are recognized by excessive stretches of solvent-exposed main-chain polar groups rather than binding to hydrophobic patches (*57*). To settle these issues, only the use of high-resolution techniques such as NMR spectroscopy will permit the precise structural analysis of material bound to chaperonins. Experimental evidence obtained thus far with such methods supports a proposal that GroEL interacts with sequences in a non-native polypeptide that have the potential to adopt an amphipathic α-helix, and that the chaperonin binding site promotes formation of a helix (*58*).

The initial partial reaction of chaperonins in the protein folding pathway is the recognition and binding of unstable folding intermediates to form a binary complex. The binding of proteins in non-native states to GroEL does not require the presence of MgATP or the co-chaperonin GroES. These binary complexes are very stable, and in some instances they have been physically isolated by gel filtration (*35, 50*). An apparent exception to the view that only proteins in non-native states will bind to GroEL are the observations that native mouse dihydrofolate reductase and bacterial pre-ß-lactamase are subject to a net unfolding when incubated with GroEL (*46, 50*). It is now known that the native states of these two proteins exist in slow conformational equilibria with species that are recognized by GroEL. As a result, most of the native enzyme can eventually be sequestered on the chaperonin in an inactive form. The stoichiometry of binding to chaperonins appears to be one or two target polypeptides bound to each GroEL tetradecamer, with most studies favoring the lower number. In several documented examples the formation of these binary complexes between GroEL and the non-native states of proteins inhibits the development of aggregates. Under appropriate conditions, rhodanese (*47*), rubisco (*24*), citrate synthase (*51*), and α-glucosidase (*52*) will aggregate following dilution from a solution containing a chaotrope. If GroEL is present during dilution, aggregates are not formed because the partially folded states of these enzymes are trapped by the chaperonin. This stabilizes the I state in a form which not only prevents it from aggregation, but also prevents it from proceeding to the native state.

$$U \rightarrow I + \text{GroEL} \rightarrow \text{GroEL}\cdot I$$

An interesting study on the aggregation of yeast α-glucosidase demonstrated the reconstitution of a heat shock effect *in vitro* (*52*). The yeast enzyme is rapidly inactivated at temperatures above 42°C and will aggregate unless GroEL is present. If the bound enzyme is discharged from GroEL with GroES and MgATP at 42°C it will again aggregate, but if discharge occurs at a lower temperature the released polypeptide will fold into an active structure.

The formation of aggregates during refolding depends on the concentration of the protein solution at a particular temperature. Above a threshhold concentration, defined as the critical aggregation concentration (*45*), aggregation will occur until the concentration is reduced to a value at which the I state can continue to isomerize to the N state. The four enzymes described earlier which aggregate unless GroEL is present, can also refold in the absence of chaperonins at lower temperatures or protein concentrations. However, if GroEL is also present during these spontaneous folding reactions the folding is inhibited (*24, 47, 51, 52*). This inhibition of successful folding is also observed for pre-ß-lactamase (*46*), dihydrofolate reductase (*49, 50*), isocitrate dehydrogenase (*54*) and lactate dehydrogenase (*53, 54*). The chaperonins, therefore, not only suppress aggregation, but also inhibit refolding. The common step in these two apparently distinct mechanisms is the interaction of GroEL with unstable folding intermediates. When unfolded proteins rapidly collapse to the I state, and the critical aggregation concentration is exceeded, they will aggregate. At lower concentrations they may refold. If GroEL is present before the concentration dependent partitioning between the alternative productive or non-productive pathways

is followed, then the I state is physically trapped by the chaperonin and neither pathway can be pursued.

The Dissociation Step. The release, or discharge, of polypeptides bound to GroEL is effected by adenine nucleotides and the co-chaperonin GroES in the presence of potassium ions. Proteins bound to GroEL behave differently in their requirements for nucleotides and GroES in the dissociation reaction. Many proteins, such as dihydrofolate reductase (*49, 50*), pre-ß-lactamase (*46*), citrate synthase (*51*), lactate dehydrogenase (*53*) and α-glucosidase (*52*) can be released and fold to an active form by the addition of MgATP alone. However, it should be appreciated, that in most cases the presence of the co-chaperonin GroES potentiated this ATP-dependent discharge. GroES is therefore not necessarily required for the release process *in vitro*, but instead acts to increase its efficiency, as previously noted (*46*). These GroES-enhanced rates of release, while not essential to obtain the desired product *in vitro*, may be significant to the physiology of cells and account for the simultaneous requirement of both GroEL and GroES for cell viability (*27*). There is also some question on whether ATP hydrolysis is necessary for release. Both lactate dehydrogenase (*53*) and dihydrofolate reductase (*50*) can be dissociated from GroEL in an active form by the addition of the nonhydrolyzable analogue 5'-adenylyl imidodiphosphate (AMP-PNP). It is noteworthy that adenosine 5'-*O*-(3-thiotriphosphate) (ATPγS), which competitively inhibits but cannot support the chaperonin-mediated refolding of rubisco (*44*), is as effective as ATP in releasing bound dihydrofolate reductase from GroEL (*50*). These observations suggest that *in vitro*, discharge of some proteins complexed to the chaperonin is mediated in part through the binding of ATP to GroEL. In the presence of ATP, or a nonhydrolyzable anologue capable of producing a similar conformational change, there is a significant reduction in the affinity between GroEL and some target proteins. This shifts the equilibrium towards free enzyme, and spontaneous folding resumes. The ATPγS and AMP-PNP analogues were also partially effective in the ATP-dependent self assembly of the GroEL tetradecamer itself (*25*). The role of ATP in the discharge process appears to be coupled with a requirement for rapid dissociation of the binary complex. The non-specific affinity of chaperonins for many different target proteins would probably require a dissociation mechanism that leads to a gross conformational rearrangement of GroEL when MgATP is present. Such rearrangements of GroEL have recently been detected by measuring changes in protease sensitivity upon adenine nucleotide addition (*59*).

There are examples in which GroES is essential during the dissociation step for successful recovery of a biologically active proteins. During the chaperonin-dependent refolding of rhodanese (*47-49*) and rubisco (*24, 35, 44, 45*) the complete folding reaction must contain GroEL, GroES, MgATP and potassium ions. It is important to distinguish between a requirement for GroES for efficient release of the polypeptide from GroEL, from one for efficient recovery of the active protein. For example, if a binary complex is prepared between GroEL and radioactive folding intermediates of rubisco, the complex is stable and can be resolved by gel filtration (*35, 59*). The addition of GroES and MgATP results in a substantial dissociation (85-90%) of the complex, and the appearance on the column of two new peaks corresponding to the active rubisco dimer and a small amount of inactive folded monomer. In contrast, when only MgATP is added, a significant proportion (50-75%) of rubisco is discharged from the complex, but it is not resolved on the column and is not catalytically active (*35, 59*). Thus, MgATP alone causes a conformational change in GroEL that weakens its affinity for the bound rubisco, but the species released does not successfully progress to the native state. This indicates that the degree of foldedness of the discharged rubisco differs depending on whether GroES is present or absent. In the presence of GroES, the bound rubisco is able to progress to a state where it is not susceptible to aggregation on release. A similar conclusion was reached for the chaperonin-dependent folding of rhodanese (*49*). In the absence of GroES, the rhodanese that was dissociated from GroEL by MgATP was not active

and formed aggregates. These latter studies on rhodanese suggested that in the absence of GroES there are repeated cycles of release and binding which does not lead to a progression to the N state (*49*). Presumably, for proteins that can fold into active structures following release from GroEL by MgATP alone, there is competition between re-binding to GroEL and progression to the N state, with partitioning to N being favored. This partitioning to the N state would then be further enhanced for most target proteins when GroES is also present to disrupt the energetically wasteful re-binding reaction.

Literature Cited.

1. Hemmingsen, S. M.; Woolford, C.; van der Vies, S. M.; Tilly, K.; Dennis, D.T.; Georgopoulos, C.P.; Hendrix, R. W.; Ellis, R. J. *Nature* **1988**, *333*, 330.
2. Ellis, R. J. *Plant J.* **1991**, *1*, 9.
3. Ang, D.; Liberek, K.; Skowyra, D.; Zylicz, M.; Georgopoulos, C. *J. Biol. Chem.* **1991**, *266*, 24233.
4. Ellis, R. J.; van der Vies, S. M. *Annu. Rev. Biochem.* **1991**, *60*, 321.
5. Gething, M.-J.; Sambrook. J. *Nature* **1992**, *355*: 33.
6. Rothman, J. E. *Cell* **1989**, *59*, 591.
7. Jaenicke, R. *Prog.Biophys. Mol. Biol.* **1987**, *49*, 117.
8. Fischer, G.; Schmid, F. X. *Biochemistry* **1990**, *29*, 2205.
9. Jaenicke, R. *Biochemistry* **1991**, *30*, 3147.
10. McMullin, T. W.; Hallberg, R. L. *Mol. Cell. Biol.* **1988**, *8*, 371.
11. Reading, D. S.; Hallberg, R. L.; Myers, A. M. *Nature* **1989**, *337*, 655.
12. Jindals, S.; Dudani, A. K.; Singh, B.; Harley, C. B.; Gupta, R. S. *Mol. Cell. Biol.* **1989**, *9*, 2279.
13. Cheng, M. Y.; Hartl, F-U.; Martin, J.; Pollock, R. A.; Kalousek, F.; Neupert, W.; Hallberg, E. M.; Hallberg, R. L.; Horwich, A. L. *Nature* **1989**, *337*, 620.
14. Hutchinson, E. G.; Tichelaar, W.; Hofhaus, G.; Weiss, H.; Leonard, K. R. *EMBO J.* **1989**, *8*, 1500.
15. Ostermann, J.; Horwich, A. L.; Neupert, W.; Hartl, F-U. *Nature* **1989**, *341*, 125.
16. Barraclough, R.; Ellis, R. J. *Biochim. Biophys. Acta* **1980**, *608*, 19.
17. Roy, H.; Bloom, M.; Milos, P.; Monroe, M. *J. Cell Biol.* **1982**, *94*, 20.
18. Bloom, M. V.; Milos, P.; Roy, H. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 1013.
19. Lubben, T. H.; Gatenby, A. A.; Donaldson, G. K.; Lorimer, G. H.; Viitanen, P. V. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7683.
20. Lubben, T. H.; Donaldson, G. K.; Viitanen, P. V.; Gatenby, A. A. *Plant Cell* **1989**, *1*, 1223.
21. Hendrix, R. W. *J. Mol. Biol.* **1979**, *129*, 375.
22. Hohn, T.; Hohn, B.; Engel, A.; Wurtz, M.; Smith, P. R. *J. Mol. Biol.* **1979**, *129*, 359.
23. Chandrasekhar, G. N.; Tilley, K.; Woolford, C.; Hendrix, R.; Georgopoulos, C. *J. Biol. Chem.* **1986**, *261*, 12414.
24. Viitanen, P. V.; Lubben, T. H.; Reed, J.; Goloubinoff, P.; O'Keefe, D. P.; G. H. Lorimer. *Biochemistry* **1990**, *29*, 5665.
25. Lissin, N. M.; Venyaminov, S. Yu.; Girshovich, A. S. *Nature* **1990**, *348*, 339.
26. Zeilstra-Ryalls. J.; Fayet. O.; Georgopoulos. C. *Annu. Rev. Microbiol.* **1991**, *45*: 301.
27. Fayet, O.; Ziegelhoffer, T.; Georgopoulos, C. *J. Bacteriol.* **1989**, *171*, 1379.
28. Neidhardt, F. C.; Phillips, T. A.; van Bogelen, R. A.; Smith, M. W.; Georgalis, Y.; Subramanian, A. R. *J. Bacteriol.* **1981**, *145*, 513.
29. Gatenby, A. A.; van der Vies, S. M.; Bradley, D. *Nature* **1985**, *314*, 617.

30. Goloubinoff, P.; Gatenby, A. A.; Lorimer, G. H. *Nature* **1989**, *337*, 44.
31. Govezensky, D.; Greener, T.; Segal, G.; Zamir, A. *J. Bacteriol.* **1991**, *173*, 6339.
32. Van Dyk, T. K.; Gatenby, A. A.; LaRossa, R. A. *Nature* **1989**, *342*, 451.
33. Fayet, O.; Louarn, J.-M.; Georgopoulos, C. *Mol. Gen. Genetic.* **1986**, *202*, 435.
34. Jenkins, A. J.; Marsh, J. B.; Oliver, I. R.; Masters, M. *Mol. Gen. Genetic.* **1986**, *202*, 446.
35. Viitanen, P. V.; Gatenby, A. A.; Lorimer, G. H. *Protein Sci.* **1992**, (in press).
36. Prasad, T. K.; Hack, E.; Hallberg, R. L. *Mol. Cell. Biol.* **1990**, *10*, 3979.
37. Gatenby, A. A.; Lubben, T. H.; Ahlquist, P.; Keegstra, K. *EMBO J.* **1988**, *7*, 1307.
38. Ptitsyn, O. B. *J. Protein Chem.* **1987**, *6*, 273.
39. Kuwajima, K. *Proteins* **1989**, *6*, 87.
40. Smith, C. J.; Clarke, A. R.; Chia, W. N.; Irons, L. I.; Atkinson, T.; Holbrook, J. J. *Biochemistry* **1991**, *30*, 1028.
41. Ptitsyn, O. B.; Pain, R. H.; Semisotnov, G. V.; Zerovnik, E.; Razgulyaev, O. I. *FEBS Lett.* **1990**, *262*, 20.
42. Zettlmeissl, G.; Rudolph, R.; Jaenicke, R. *Biochemistry* **1979**, *18*, 5567.
43. Zettlmeissl, G.; Rudolph, R.; Jaenicke, R. *Biochemistry* **1979**, *18*, 5572.
44. Goloubinoff, P.; Christeller, J. T.; Gatenby, A. A.; Lorimer, G. H. *Nature* **1989**, *342*, 884.
45. van der Vies, S. M.; Viitanen, P. V.; Gatenby, A. A.; Lorimer, G. H., Jaenicke, R. *Biochemistry* **1992** (in press).
46. Laminet, A. A.; Ziegelhoffer, T.; Georgopoulos, C.; Pluckthun, A. *EMBO J.* **1990**, *9*, 2315.
47. Mendoza, J. A.; Rogers, E.; Lorimer, G. H.; Horowitz, P. M. *J. Biol. Chem.* **1991**, *266*, 13044.
48. Mendoza, J. A.; Lorimer, G. H.; Horowitz, P. M. *J. Biol. Chem.* **1991**, *266*, 16973.
49. Martin, J.; Langer, T.; Boteva, R.; Schramel, A.; Horwich, A. L.; Hartl, F.-U. *Nature* **1991**, *352*, 36.
50. Viitanen, P. V.; Donaldson, G. K.; Lorimer, G. H.; Lubben, T. H.; Gatenby, A. A. *Biochemistry* **1991**, *30*, 9716.
51. Buchner, J.; Schmidt, M.; Fuchs, M.; Jaenicke, R.; Rudolph, R.; Schmid, F. X.; Kiefhaber, T. *Biochemistry* **1991**, *30*, 1586.
52. Holl-Neugebauer, B.; Rudolph, R.; Schmidt, M.; Buchner, J. *Biochemistry* **1991**, *30*, 11609.
53. Badcoe, I. G.; Smith, C. J.; Wood, S.; Halsall, D. J.; Holbrook, J. J.; Lund, P.; Clarke, A. R. *Biochemistry* **1991**, *30*, 9195.
54. Taguchi, H.; Konishi, J.; Ishii, N.; Yoshida, M. *J. Biol. Chem.* **1991**, *266*, 22411.
55. Bochkareva, E. S.; Lissin, N. M.; Girshovich, A. S. *Nature* 1988, *336*, 254.
56. Landry, S. J.; Bartlett, S. G. *J. Biol. Chem.* **1989**, *264*, 9090.
57. Hubbard, T. J. P.; Sander, C. *Protein Engineering* **1991**, *4*, 711.
58. Landry, S. J.; Gierasch, L. M. *Biochemistry* **1991**, *30*, 7359.
59. Baneyx, F.; Gatenby, A. A. *J. Biol. Chem.* (in press)

RECEIVED April 9, 1992

Chapter 12

Cosolvent Effects on Refolding and Aggregation

Jeffrey L. Cleland[1] and Daniel I. C. Wang[2]

[1]Pharmaceutical Research and Development, Genentech, Inc., South San Francisco, CA 94080
[2]Biotechnology Process Engineering Center, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

During refolding of many proteins, the off-pathway reaction of aggregation often occurs resulting in a reduced recovery of active protein. To improve the recovery of active protein during refolding, cosolvents have been used to prevent aggregation and enhance refolding. Previous studies on polyethylene glycol (PEG) enhanced refolding of bovine carbonic anhydrase (CAB) have shown that complete recovery of active protein can be achieved under aggregating conditions [Cleland, J. L.; Wang, D. I.C. *Biotechnology* **1990**, *8*, pp. 1274-1278]. The physico-chemical basis for this enhancement in refolding was examined by using cosolvents with different chemical properties. Sugars, polyamino acids, PEG, and PEG derivatives were assessed for their effects on the recovery of active protein during the refolding of CAB at aggregating conditions. The mode of operation of these cosolvents appeared to dictate their ability to improve refolding. Cosolvents, such as sugars, which cause hydration of the protein inhibited aggregation but did not improve the recovery of active CAB. Hydrophobic PEG derivatives which bound tightly to a folding intermediate inhibited the refolding of CAB. On the other hand, both polyamino acids and PEG enhanced the recovery of active CAB and prevented aggregation. These cosolvents probably operated by preventing the aggregation of the molten globule first intermediate in the refolding pathway of CAB. Cosolvents that enhance protein refolding should therefore interact with folding intermediates to prevent association, but they should not inhibit the refolding of these intermediates. The thermodynamics of cosolvent and native protein interactions also provided insight into additional cosolvents which may improve the recovery of active protein during refolding.

The refolding of recombinant proteins has become an essential part of the recovery operations in the biotechnology industry. Many recombinant proteins

0097–6156/93/0516–0151$06.00/0

are produced by *Escherichia coli* in the form of intracellular aggregates or inclusion bodies. These insoluble protein aggregates are usually separated from the host cell and solubilized in a denaturant such as urea or guanidine hydrochloride (GuHCl). During the solubilization process, the protein often becomes completely or partially denatured. The protein is then refolded to its native state by removal of the denaturant. Denaturant removal allows the protein to regain its native state, but it is often accompanied by substantial protein aggregation. The aggregation which occurs during refolding dramatically reduces the recovery of the desired native protein. To increase the yield of active protein from the refolding process, the formation of off-pathway intermediates such as partially folded or aggregated species must be reduced.

Several different approaches have been applied in attempts to avoid the problem of aggregation during refolding (*1-9*). Practical approaches to solving this problem usually involve the addition of a refolding aide or aggregation inhibitor (*1-5*). The addition of sugars has been successfully applied to prevent *in vivo* aggregation and allow recovery of native protein without a refolding step (*3*). Other studies have focussed on enhancing the recovery of native proteins during *in vitro* processing. Several polymeric cosolvents have been used to reduce aggregation during refolding (*1,2,4*). By using a surfactant, lauryl maltoside, in the refolding buffer, the recovery of active rhodanese has been significantly improved (*4*). More recent work has indicated that polyethylene glycol (PEG) significantly enhances the refolding of several proteins (*1,2*). In addition, PEG has been shown to prevent aggregation by binding to the first intermediate molten globule of bovine carbonic anhydrase B (CAB) (*10*). PEG therefore has a very desirable cosolvent characteristic. It can bind to a hydrophobic intermediate, but does not inhibit refolding (*1,10*). Other cosolvents may have similar characteristics. Clearly, the best cosolvent is a molecule which prevents aggregation, but does not alter the rate of refolding. The cosolvent must also be easily recovered and not remain bound to the native protein.

To assess the ability of different cosolvents to enhance refolding, a well-characterized protein, CAB, has been used (*11,12*). Previous research on the refolding and aggregation of CAB has revealed that the molten globule first intermediate aggregates to form dimers and larger aggregates (*13-15*). In addition, this intermediate reversibly associates to form a dimer during refolding at conditions which result in complete recovery of active protein (*15*). To enhance refolding of CAB, a cosolvent must therefore prevent the association of the hydrophobic first intermediate. A variety of different cosolvents were examined in an attempt to determine the physicochemical properties of cosolvents that enhance refolding. The ability of each solvent to enhance refolding was measured by refolding at conditions which previously have resulted in reversible association of the first intermediate or irreversible aggregation with a concomitant reduction in final recovery of native protein (*13,15*).

Experimental Procedures

Materials. Bovine serum albumin (BSA), guanidine hydrochloride (GuHCl), Tris-sulfate, ethylenediaminetetraacetic acid (EDTA), sucrose, glucose, and p-nitrophenol acetate (pNPA) were molecular biology grade and purchased from Sigma Chemical Co. (St.Louis, MO). Polyethylene glycol (PEG) and all polyamino acids were also purchased from Sigma Chemical Co. (St. Louis, MO) and used as supplied. JEFFAMINE polymers were donated by Texaco Chemical Company (Houston, TX). HPLC grade acetonitrile was obtained from J.T. Baker Co. (Phillipsburg, NJ). Decahydranapthalene (Decalin) was a racemic mixture obtained from Aldrich (Milwaukee, WI). The SDS-PAGE materials were obtained from Pharmacia LKB Biotechnology (Uppsala, Sweden). All buffers and samples were prepared with distilled water passed through a MilliQ water purification system (Millipore Corp., Bedford, MA).

Bovine Carbonic Anhydrase B (CAB): Lyophilized bovine carbonic anhydrase B (p*I* = 5.9) was purchased from Sigma Chemical Co. (St. Louis, MO). The purity of each lot was checked by size exclusion high performance liquid chromatography (HPLC) and gel electrophoresis with silver staining. In addition, the native protein in 50 mM Tris-sulfate and 5 mM EDTA at pH 7.5 was measured for activity to ascertain the formation of native protein structure (see Esterase Activity section). For all refolding experiments, the lyophilized protein was solubilized in 5 M GuHCl for at least 6 hours (*11*). Prior to experimental use, the denatured protein was filtered with 0.22 μm syringe filters (Gelman Sciences, Ann Arbor, MI) to remove any remaining insoluble protein.

Protein Concentration. The protein concentration of native CAB in 50 mM Tris-sulfate, 5 mM EDTA, at pH 7.5 was determined by absorbance at 280 ηm using an extinction coefficient of 1.83 $(mg/ml\ protein)^{-1}\ cm^{-1}$ and a molecular weight of 30,000 (*16*). Denatured CAB in 5 M GuHCl was assayed for protein concentration by using a colorimetric dye binding assay, Bio-Rad reagent (Bio-Rad Laboratories, Richmond, CA) or BCA reagent (Pierce Chemical Company, Rockford, IL), with bovine serum albumin (BSA) denatured in 5 M GuHCl as the standard.

Esterase Activity (CAB). The enzymatic activity of CAB was measured by analysis of the esterase reaction as described previously (*17*). The unfolded CAB in 5 M GuHCl was rapidly diluted with buffer containing cosolvent to a desired final protein and GuHCl concentration. An aliquot of the refolded sample was analyzed for its enzymatic activity at various times after dilution as detailed previously (*13*). Recovery of activity was determined using the ester hydrolysis rate constant of the native protein at the same concentration in the dilution buffer (50 mM tris sulfate, 5 mM EDTA, pH 7.5) with cosolvents. The initial rate of refolding, R_{Ref}, was defined as the initial slope of the active protein concentration as function of time as reported previously (*1*).

Quasi-Elastic Light Scattering (QLS) Measurements. Quasi-elastic light

scattering analysis was performed using two different systems. Initial QLS measurements were performed by using Model N4 submicron particle analyzer (Coulter Electronics, Hialeah, FL) instrumented as described previously (*18*). Sample analysis and materials preparation were detailed in previous studies (*13*). Additional QLS measurements were performed using a Brookhaven light scattering system (Brookhaven Instruments, Holtsville, NY). The Brookhaven QLS system consisted of a BI200SM goniometer with a photomultiplier positioned at 90° to the incident laser, 2 W argon ion at 488 ηm (Lexel, Fremont, CA). The goniometer assembly was temperature controlled at 20°C with an external water bath. To reduce flaring from incident laser light, the sample was placed in the goniometer with index matching fluid, *cis,trans* - decahydronapthalene, surrounding the sample. The photon data was collected using a BI2030 autocorrelator with 136 channels and a personal computer collected the data and controlled the system. Prefiltered samples (≥ 1 ml) were placed in disposable glass culture tubes (12 x 75mm, VWR Scientific, San Francisco, CA) which were checked for imperfections and precleaned by using appropriate glassware procedures (*13*).

The particle size distributions from QLS autocorrelation function data were calculated by the method of constrained regularization or CONTIN (*19*). The concentration of each species was then determined by using the previously developed model for CAB aggregation (*13*). Refolding of denatured CAB in 5 M GuHCl was carried out by rapid dilution with buffer (50 mM Tris-sulfate, 5 mM EDTA, pH 7.5) with or without the cosolvent to the desired final protein and GuHCl concentrations. For kinetic studies, the sample was analyzed by QLS immediately after dilution as detailed previously (*13*). Each experiment was repeated several times to provide a significant number of time points for kinetic analysis. The rate of dimer formation, R_D, for each experiment was calculated from the initial slope of the dimer concentration as a function of time as described previously (*13*).

Results

Sugars: Sucrose and Glucose. Significant reductions in *in vivo* aggregation of ß-lactamase have been previously reported using sugars (*3*). In addition, sugars have been used to improve recovery of active ß-lactamase during *in vitro* refolding (*20*). To further assess sugars, refolding of CAB was performed with either sucrose or glucose in the dilution buffer. Previous studies indicated that a sugar concentration of 239 g/l (22% wt/vol) was required to observe significant improvements in the refolding of CAB at aggregating conditions (data not shown). Therefore, denatured CAB in 5 M GuHCl was refolded by rapid dilution to 0.24 mg/ml protein and 0.30 M GuHCl (aggregation conditions, *13*) with 22% wt/vol sugar in the dilution buffer. Under these conditions, formation of large aggregates was not observed by QLS whereas large aggregates had been previously observed during refolding at these conditions without sugars (*1,13*). However, the formation of multimers was detected by QLS for refolding with and without the sugars at these conditions. For each case, the initial rate of dimer formation, R_D, was calculated from the initial slope of the dimer concentration as a function of time (*13*). To measure the competing refolding reaction, the recovery of active protein was measured

as a function of time after dilution to the refolding conditions. The initial rate of refolding, R_{Ref}, was defined as the initial slope of the active protein concentration as a function of time (*1*). For each case, the active protein concentration reached a constant value after twenty minutes and the final active protein concentration was defined as the concentration of active protein after refolding for one hour.

As shown in Table 1, the initial rate of dimer formation, R_D, was reduced in the presence of both sugars. The dimer formation rate was similar for both glucose (MW 180) and sucrose (MW 360). However, the initial rate of refolding, R_{Ref}, with sucrose was two-fold greater than the refolding rate with glucose. Both sugars significantly increased the initial rate of refolding. Unfortunately, the recovery of activity after 1 hour was not substantially improved for refolding in either sugar solution (Table 1). For both sugars, the refolded solution contained approximately 30% monomer as determined by QLS after incubation for 1 hour. Control experiments with native CAB and sugars indicated that sugars did not alter the specific activity. Therefore, some of the monomeric species were not able to assume the active CAB conform-ation since only 20% active protein was observed (Table 1). Additional refolding studies were performed by dilution to aggregating conditions, 1.0 - 1.2 mg/ml CAB and 0.5 M GuHCl, in the presence of the sugars. These studies yielded similar refolding rates and little improvement in the final recovery of activity was observed (data not shown).

The observed increase in the initial refolding rate, R_{Ref}, in the presence of sugars was probably caused by an increase in the hydration of the protein. Previous studies involving sugars and native protein stability have revealed that sugars cause preferential hydration of proteins independent of the chemical properties of their surface (*21*). This preferential hydration of the protein resulted in the formation of compact structural states with hydrophillic surface characteristics analogous to the native protein. The thermodynamic forces which caused the rapid formation of compact native structures also resulted in increased stability of other compact structures such as dimers and trimers. Multimers were the most thermodynamically favorable species since these structures have lower accessible surface area that reduces the overall surface tension of the solution. Therefore, the sugars prevented precipitation of the protein by stabilizing both the compact folded structures and the multimers. The observed reductions in the rate of dimer formation using the two sugars probably resulted from both the reduced rate of large aggregate formation and the increased viscosity. The viscous sugar solutions (22% wt/vol, 2 centipoise) may have also decreased dimerization by introducing diffusional limitations for monomer-monomer collisions.

Other refolding studies with sugars also provided insight into the interaction between sugars and partially folded proteins. For example, sucrose did not alter the refolding of porphyrin c (*22*). On the other hand, the rate of ribonuclease refolding was reduced in the presence of sugars (*23*). Glycerol and glucose have also been observed to decrease the rate of refolding of octopine dehydrogenase, a large multimeric protein (*24*). Other studies have shown that sugars reduced the rate of refolding (*25*). For CAB refolding, only

two thirds of the monomers in solution were active indicating that the sugars interfered with the refolding process. Although these studies contradict the results of ß-lactamase refolding with sugars (*3, 20*), they have revealed that the mechanism of sugar-protein interactions, preferential hydration of the protein, may not be desirable for protein refolding. In general, a cosolvent which only causes preferential hydration of the protein will probably not be effective in increasing the recovery of active protein, but it may prevent the formation of large precipitates.

Polymeric Cosolvents: Polyamino acids. Previous research on interactions between amino acids and native proteins indicate that amino acids may either bind to the protein or cause preferential hydration of the protein (*26*). In addition, protein refolding in the presence of amino acids was previously performed with some success (*27*). Since amino acids could become incorporated into the protein during refolding, polyamino acids were used in these experiments. Several polyamino acids ranging in molecular weight from 2500 to 3500 Daltons were utilized to determine the cosolvent characteristics that enhance refolding of CAB. The refolding of CAB at aggregating conditions (1.0 mg/ml CAB, 0.50 M GuHCl) was previously enhanced by using 3 g/l PEG (3350 MW) in the refolding buffer (*1*). Therefore, the polyamino acid solutions (polyalanine, 2500 MW; polylysine, 3000 MW) were prepared at a concentration of 3 g/l. Polyglycine (3500 MW) was not soluble at 3 g/l and was dissolved to its apparent maximum solubility of 1 g/l. These polyamino acid solutions were then used to dilute CAB in 5 M GuHCl to final concentrations of 1.0 mg/ml protein and 0.50 M GuHCl (aggregation conditions, *13*). The recovery of active protein for each case was then measured as a function of time after dilution as shown in Figure 1. For comparison, refolding was also performed with a dilution buffer containing 3 g/l PEG (3350 MW). The use of polyglycine and polyalanine resulted in a rate of refolding which was similar to the results for 3 g/l PEG. As shown in Figure 1, the initial rate of refolding was much greater when polylysine was used in the refolding buffer. However, refolding with each polymer resulted in complete recovery of active protein after 1 hour. In contrast, when refolding was performed without the cosolvents, only 30% of the active protein was recovered and the protein formed insoluble aggregates within this 1 hour period. Precipitation was not observed for refolding with the polyamino acids or PEG.

For each polyamino acid, the protein rapidly refolded to a high level of activity and continued to recover activity at a slower rate. This phenomenon was previously observed for refolding in the absence of association of the first intermediate (*15*). The first intermediate rapidly folded to form the second intermediate which slowly regained the fully active native conformation (*15*). Therefore, the polyamino acids may have prevented the aggregation of the hydrophobic first intermediate. Previous research indicates that PEG prevented the self association of this intermediate during refolding (*28*). PEG was also observed to bind to this molten globule first intermediate (*10*). In addition to inhibiting aggregation, the more hydrophobic polyamino acids, polyalanine and polylysine, may also have increased the rate of refolding through specific interactions with the hydrophobic first intermediate. These

Table 1. Effect of sugars on aggregation and refolding of CAB. Denatured CAB in 5 M GuHCl was rapidly diluted with buffer to 0.24 mg/ml protein and 0.30 M GuHCl. The solution was then analyzed as described in Experimental Procedures.

Buffer	R_D (μM/min)	R_{Ref} (1/min)	Final % Active Protein
No Sugar	100[a]	0.30	19.0
22% wt/vol Glucose	1.4	1.1	20.4
22% wt/vol Sucrose	1.2	2.0	19.9

[a] Theoretical value calculated from dimer rate equation in Ref. 13.

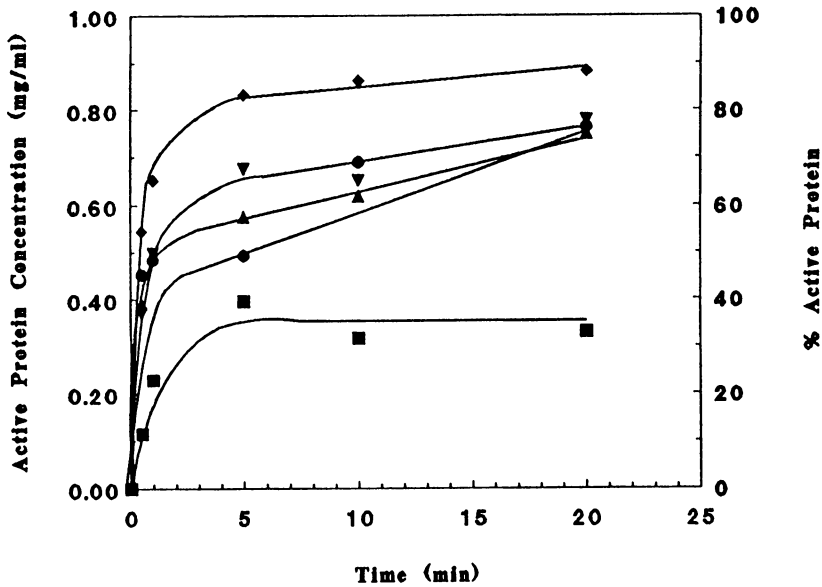


FIGURE 1: Effect of polyamino acids on the CAB aggregation during refolding. The protein in 5 M GuHCl was diluted to 1.0 mg/ml CAB and 0.50 M GuHCl (aggregation conditions) with different polyamino acid solutions. Polyalanine (3000 MW; ▼) and polylysine (2500 MW; ♦) were used at final concentration of 3 g/l. In addition, polyglycine (3500 MW; ●) was used at its solubility limit which was approximately 1 g/l. For comparison, the recovery of activity is displayed for refolding in 3 g/l PEG (3350 MW; ▲) and in the standard dilution buffer (■).

polyamino acids can also form secondary structures which may specifically interact with similar structures on the protein surface. Polylysine will form a beta sheet structure in solution and this structure may interact with hydrophobic beta sheet structures on the surface of an intermediate. In contrast, polyalanine forms a helical structure and polyglycine has a random coil structure. Therefore, the interactions between these amino acid polymers and the folding intermediates may not be structurally related. Further analysis of these cosolvents should be performed to determine if the polyamino acids inhibit aggregation by interacting with specific protein surface structures.

Polyethylene glycol (PEG). As shown in Figure 1, the rate of refolding in the presence of PEG was similar to that obtained with the polyamino acids. In addition, the use of PEG resulted in complete recovery of active protein at these conditions (3 g/l PEG (3350 MW), 1.0 mg/ml CAB, 0.50 M GuHCl). The PEG molecular weight was previously shown to affect the rate of refolding under the same aggregating conditions (*1*). High molecular weight PEG (8000 MW, 3 g/l) resulted in a slightly more rapid rate of refolding than lower molecular weight PEG (1000 or 3350 MW, 3 g/l) as shown in Figure 2 (*1*). When denatured CAB in 5 M GuHCl was rapidly diluted to 1.0 mg/ml protein and 0.50 M GuHCl with buffer containing 3 g/l PEG (8000 MW), complete recovery of active protein was achieved in twenty minutes. If PEG increased the rate of refolding by a catalytic mechanism, higher concentrations of PEG should further increase the rate of refolding. To test this hypothesis, CAB was refolded with both concentrated (30 g/l) and dilute (3 g/l) PEG solutions. CAB in 5 M GuHCl was rapidly diluted to 0.50 mg/ml CAB and 1.0 M GuHCl where transient association of the first intermediate was previously observed (*15*). The rate of refolding with 3 g/l PEG (3350 MW) was slightly greater than 30 g/l PEG (Figure 3). These results suggest that PEG enhanced the rate of refolding in both cases by inhibiting the self association of the first intermediate molten globule (*1,10,28*). Additional research also indicates that PEG did not increase the rate of refolding (*28*). These studies also showed that PEG reversibly bound to the first intermediate during refolding (*10, 28*). In addition, the molecular weight and concentration of PEG have been shown to directly correlate with the extent of recovery (*28*). For each PEG molecular weight, an optimum range of concentrations was observed and, thus, the difference in refolding rates shown in Figure 2 are likely the result of this stoichiometric relationship (*28*). Overall, PEG enhanced the refolding of CAB by preventing the aggregation of the hydrophobic first intermediate (*28*).

The unique physicochemical properties of PEG were probably responsible for the inhibition of aggregation without a reduction in the refolding rate. Unlike the polyamino acids, PEG did not form a rigid structure in solution. Thus, it should be able to interact with the protein by more general chemical mechanisms (ie. hydrophobic interactions). Previous studies of PEG and native proteins have revealed that PEG will cause preferential hydration of the protein when used at high concentrations (>3% wt/vol) and will bind to proteins at dilute concentrations (*29*). Therefore, PEG can interact with proteins by two different mechanisms. PEG has unique physicochemical properties which include both hydrophobic methylene groups and oxygen (ether bonds) which can weakly interact through hydrogen bonding. Since PEG has

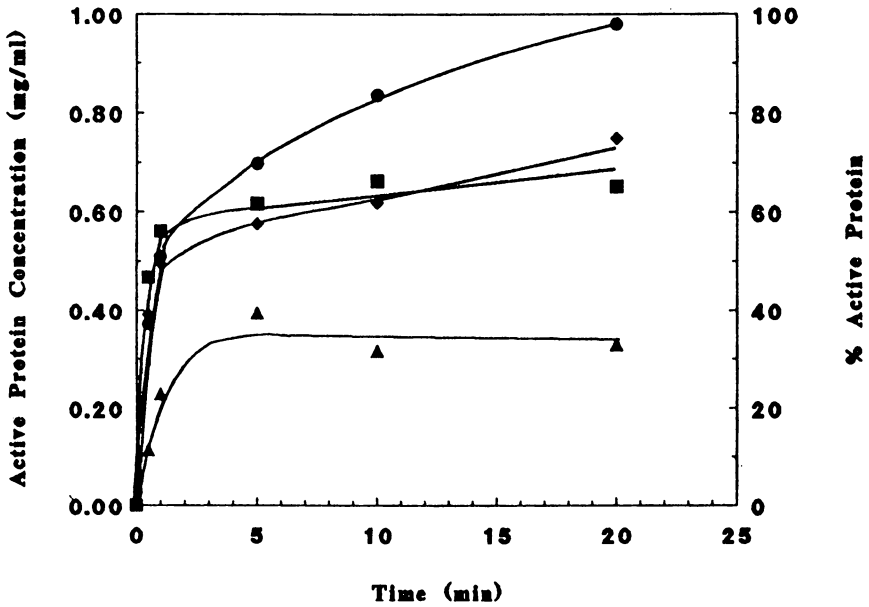


FIGURE 2: CAB aggregation and refolding in the presence of different molecular PEG solutions. CAB in 5 M GuHCl was diluted with 3 g/l PEG to 1.0 mg/ml protein and 0.50 M GuHCl (aggregation conditions). The recovery of activity as a function of time is plotted for the three different molecular weight PEG solutions, 8000 MW (●), 3350 MW (♦), and 1000 MW (■), along with the results for refolding in the absence of PEG in the dilution buffer (▲).

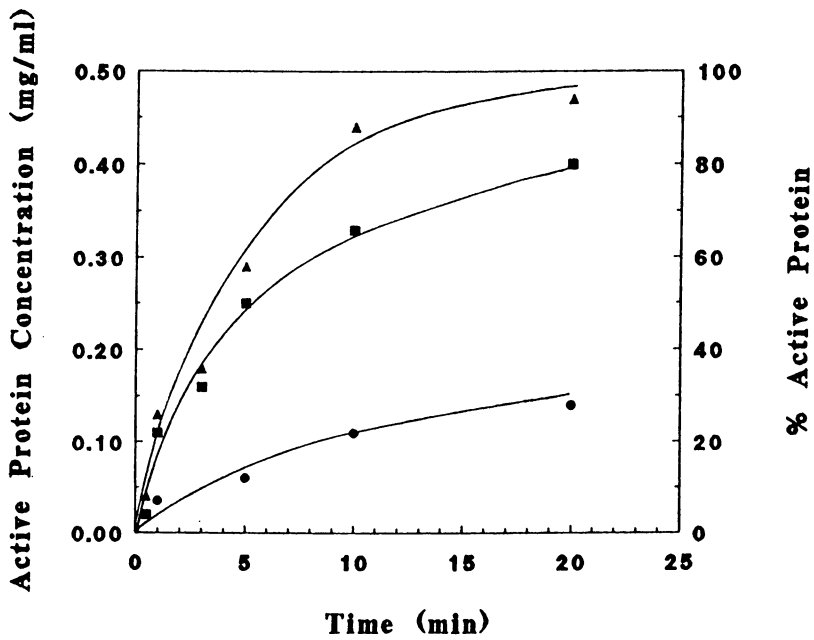


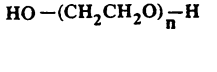
FIGURE 3: Effect of PEG concentration on CAB refolding. Denatured CAB in 5 M GuHCl was diluted to 0.50 mg/ml protein and 1.0 M GuHCl with buffer containing PEG. The recovery of activity is displayed as a function of time for PEG (MW 3350) at 3 (▲) and 30 g/l (■). For comparison, refolding without PEG in the dilution buffer was also measured as a function of time (●).

these unique chemical properties as well as the ability to reversibly bind to a folding intermediate, it should be a useful generic cosolvent for protein refolding. The recovery of active protein for the refolding of two recombinant proteins, deoxyribonuclease and tissue plasminogen activator, was also significantly improved through the use of PEG as a cosolvent (*2*). Further research into the specific interaction between PEG and these proteins should provide insight into the design of cosolvents that inhibit aggregation and enhance refolding of proteins.

Polyethylene Glycol Derivatives. PEG was a successful cosolvent for the inhibition of aggregation during CAB refolding, but it did not catalyze the refolding. Previous research showed that refolding in the presence of PEG proceeds at the same rate as refolding at nonaggregating conditions (*28*). In addition, only a stoichiometric concentration of PEG was required for inhibition of aggregation (*2*, *28*). For CAB refolding, a PEG to protein molar ratio ($[CAB]_t/[PEG]_t$) of 3 to 1 was sufficient to completely inhibit aggregation (1.0 mg/ml CAB, 0.50 M GuHCl, 0.34 g/l PEG (3350 MW); *28*). The stoichio-metric concentrations of PEG required for enhanced refolding (*28*) and the previous PEG binding studies (*10*) indicated that the chemical properties of PEG were responsible for its action. To further study these chemical properties, block copolymers of ethylene oxide and propylene oxide were used in refolding experiments. These copolymers are similar to PEG, which is also referred to as poly(ethylene oxide) as shown in Figure 4. These hydrophobic copolymers possessed primary amines at both termini. However, previous refolding studies with bis(amine)polyethylene glycol yielded the same results as unmodified PEG (data not shown). The M series polymers were random co-polymers with the more hydrophobic propylene group distributed throughout the molecule. In constrast, the ED series polymer (ED-6000, 6000 MW) was a block copolymer with hydrophobic propylene termini and a polyethylene oxide center (Figure 4). Each of these polymers was used at a molar ratio of 3 to 1 ($[Polymer]_t/[CAB]_t$) in the refolding of CAB. Denatured CAB in 5 M GuHCl was rapidly diluted with the polymer solution to 0.50 mg/ml and 1.0 M GuHCl (refolding conditions). These conditions should result in complete refolding of the protein after 1 hour or less as previously demonstrated (Figure 3, *1*, *13*, *15*, *28*). The recovery of active protein was dramatically reduced in the presence of these polymers. Refolding with the least hydrophobic polymer (M-1000) yielded the greatest recovery of active protein (20%) after 1 hour. In the case of the other more hydrophobic polymers, only 10% of active protein was recovered after 1 hour (Figure 4). Therefore, these polymers inhibited the refolding of CAB. Since each of these polymers was more hydrophobic than PEG, it is possible that these molecules are able to bind to the first intermediate molten globule more tightly. The first intermediate could thus be prevented from refolding. Clearly, a more hydrophobic cosolvent than PEG had a detrimental effect on the refolding of CAB. Hydrophobic polymers have been successfully used to refold a membrane protein, rhodanese, but it was suggested that less hydrophobic polymers such as PEG would be more effective (*30*).

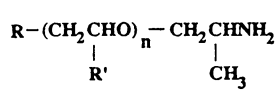
a

Poly(ethylene oxide) (PEO) = Polyethylene glycol (PEG)



MW (Da)	n
1000	22
3350	76
8000	181

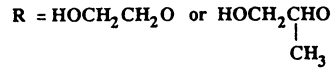
JEFFAMINE - M Series



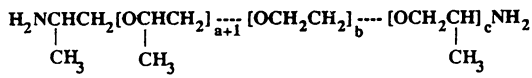
Trade Name	MW (Da)	PO/EO Ratio
M-1000	1000	3/19
M-2070	2000	10/32

R' = H, EO (ethylene oxide)

R' = CH_3, PO (propylene oxide)



JEFFAMINE- ED Series



Trade Name	MW (Da)	b(EO)	a+c (PO)	Overall PO/EO Ratio
ED-6000	6000	132	2.5	3.5/132

b

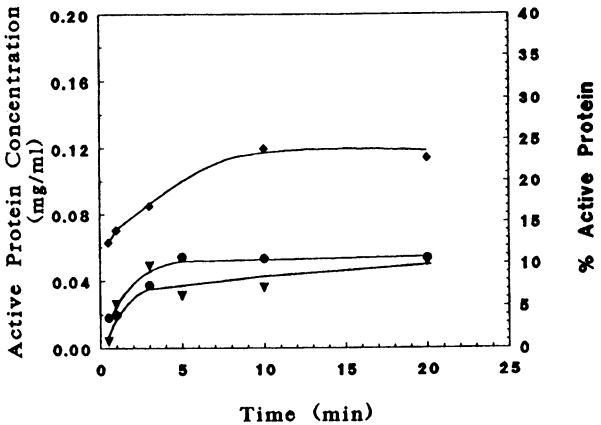


FIGURE 4: (a) Chemical structure of hydrophobic polymers. The hydrophobic block copolymers were provided by Texaco Chemical Company (Houston, TX). These polymers are referred to as the JEFFAMINE polymers by Texaco. The structure and relative ratio of ethylene oxide (EO) and propylene oxide (PO) in each polymer are shown. For comparison, polyethylene glycol (PEG) is also known as polyoxyethylene (PEO). (b) CAB refolding in the presence of hydrophobic polymers. CAB in 5 M GuHCl was refolded by rapid dilution to 0.50 mg/ml (16.7 μm) protein and 1.0 M GuHCl with a final molar ratio of polymer to CAB ($[Polymer]_f/[CAB]_f$) of 3 for three different polymers: M-1000 (♦), M-2000 (●) and ED-6000 (▼).

Discussion

To improve recovery of active protein during refolding, several cosolvents were assessed and preliminary relationships between their effect on refolding and their mode of action were developed. The two primary modes of action of cosolvents are preferential hydration, which is analogous to cosolvent exclusion from the protein surface, and cosolvent binding (*31*). Addition of cosolvents which may act by preferential hydration will stabilize compact protein structures such as the native and multimeric states. This phenomenon can be described by analyzing the thermodynamics of the system. The change in chemical potential of the protein that occurs from cosolvent addition ($\partial\mu_p/\partial m_s$) has been defined as the product of the change in chemical potential of the cosolvent ($\partial\mu_s/\partial m_s$) and the preferential interaction of the cosolvent with the protein ($\partial m_s/\partial m_p$):

$$\left[\frac{\partial \mu_p}{\partial m_s}\right]_{T,P,m_p} = -\left[\frac{\partial m_s}{\partial m_p}\right]_{T,\mu_w,\mu_s}\left[\frac{\partial \mu_s}{\partial m_s}\right]_{T,P,m_p}$$

where μ is the chemical potential, m is the molality, T is the system temperature, P is the system pressure, and the subscripts denote protein (p), cosolvent (s), and water (w). The preferential interaction term ($\partial m_s/\partial m_p$) can then be used to explain the effect of cosolvents on protein refolding.

For sugars such as sucrose and glycerol used in these studies, the preferential interaction term would be positive since sugars cause hydration of the protein (*32*). The sugars did not significantly enhance the recovery of active CAB during refolding at conditions which result in protein aggregation. Both a native-like monomer and multimers were probably stabilized as the result of the preferential hydration. In addition, the sugars could have interfered with the formation of the active native protein. These results indicated that molecules which cause preferential hydration may not be good cosolvents for protein refolding. However, amino acids have also been reported to cause preferential hydration (*26*) and polyamino acids provided complete recovery of active CAB at aggregation conditions (1.0 mg/ml CAB, 0.50 M GuHCl; *13*). Further analysis of the polyamino acids and CAB refolding should be performed to determine the mode of action of these cosolvents. A cosolvent such as PEG which can act through both mechanisms, hydration and binding, was also tested in the refolding of CAB. PEG can bind to the protein at low concentrations (< 10% w/v) and will become excluded from the protein surface at high concentrations (*21,29*). PEG provided complete recovery of active CAB for refolding at aggregation conditions (1.0 mg/ml CAB, 0.50 M GuHCl; *13*). Previous research with PEG and CAB refolding has shown that PEG reversibly bound to the molten globule first intermediate and prevented its aggregation (*10,28*). The refolding of CAB was not inhibited by PEG (*28*). More hydrophobic derivatives of PEG were also tested for their effect on CAB refolding. All of these polymers inhibited the refolding of CAB. These polymers could have been bound to the hydrophobic first intermediate and prevented further refolding. Thus, cosolvents which have

a preferential interaction between hydration and binding ($\partial m_s/\partial m_p \approx 0$) should provide the best results for protein refolding. As shown in Figure 5, many different molecules have been studied for their preferential interaction with native proteins. These molecules should also be assessed for their effect on protein refolding. The results of these studies should provide further insight into the desired cosolvent characteristics for protein refolding and inhibition of aggregation.

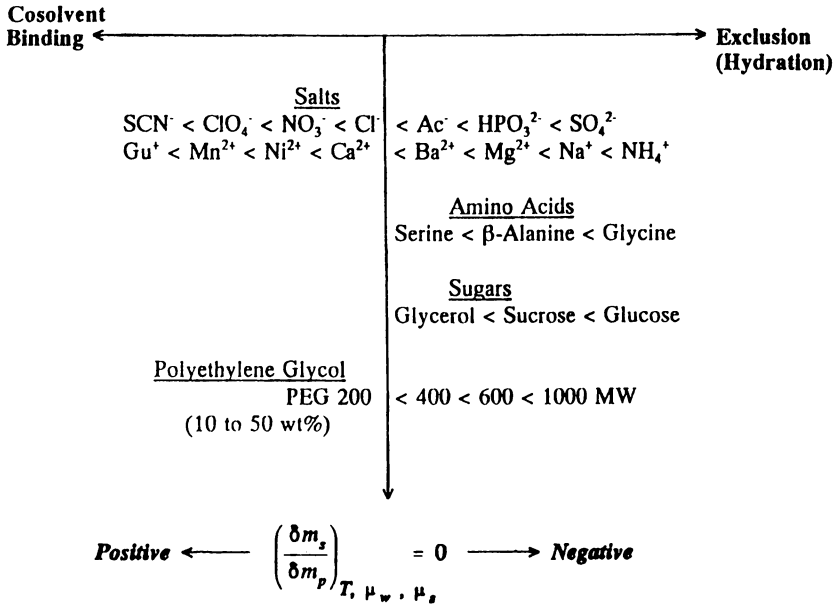


FIGURE 5: Cosolvent interactions with native proteins (*26,29,31-33*). The cosolvents left of the midpoint have been shown to bind nonspecifically to proteins (positive preferential cosolvent interactions). In contrast, the cosolvents to the right of the midpoint have been shown to act by preferential hydration of the protein which involves an increase in the concentration of water at the protein surface (negative preferential cosolvent interaction). The mechanism of interaction for polyethylene glycol (PEG) has been observed to be dependent upon the PEG concentration where high concentrations (10-50 wt%) result in exclusion and low concentrations (< 10 wt%) can result in PEG binding to the protein (*26*).

Acknowledgements

This work was supported by National Science Foundation Grant #CDR-88-03014.

Literature Cited

1. Cleland, J. L.; Wang, D. I. C. *Biotechnology* **1990**, *8*, pp. 1274-1278.

2. Cleland, J. L.; Wang, D. I. C. *Biotechnology* **1992**, *submitted for publication*.

3. Bowden, G. A.; Georgiou, G. *Biotech. Prog.* **1988**, *4*, pp. 97-101.

4. Tandon, S.; Horwitz, P. M. *J. Biol. Chem.* **1987**, *262*, pp. 4486-4491.

5. Brems, D. N.; Stellwagen, E. *J. Biol. Chem.* **1983**, *258*, pp. 3655-3661.

6. Brems, D. N.; Plaisted, S. M.; Havel, H. A.; Tomich, C.-S. C. *P.N.A.S.* **1988**, *85*, pp. 3367-3371.

7. Hagen, A.; Hatton, T. A.; Wang, D. I. C. *Biotech. Bioeng.* **1990**, *35*, pp. 966-975.

8. Creighton, T. E. In *Protein Structure Folding and Design*; Oxender, D. L., Ed.; Alan R. Liss, Inc.: New York, NY, 1985; pp. 249-251.

9. Morris, G. E.; Frost, L. C.; Newport, P. A.; Hudson, N. *Biochem. J.* **1987**, *248*, pp. 53-57.

10. Cleland, J. L.; Randolph, T. W. *J. Biol. Chem.* **1992**, *in press*.

11. Semisotnov, G. V.; Uversky, V. N.; Sokolovsky, I. V.; Gutin, A. M.; Razgulyaev, O. I.; Rodionova, N. A. *J. Mol. Biol.* **1990**, *213*, pp. 561-568.

12. Rodionova, N. A.; Semisotnov, G. V.; Kutyshenko, V. P.; Uverskii, V. N.; Bolotina, I. A.; Bychkova, V. E.; Ptitsyn, O.B. *Mol. Biol.* **1989**, *23*, pp. 683-692.

13. Cleland, J. L.; Wang, D.I.C. *Biochemistry* **1990**, *29*, pp. 11072-11078.

14. Cleland, J. L.; Wang, D. I. C. In *Protein Refolding*; Georgiou, G.; DeBernardez-Clark, E., Eds.; ACS Symposium Series, Vol. 470; American Chemical Society, Washington, D.C., 1991; pp. 169-179.

15. Cleland, J. L.; Wang, D. I. C. *Biotech. Progress* **1992**, *in press*.

16. Wong, K.-P.; Tanford, C. *J. Biol. Chem.* **1973**, *248*, pp. 8518-8523.

17. Pocker, Y.; Stone, J. T. *Biochemistry* **1967**, *6*, pp. 668-678.

18. Yarmush, D. M.; Murphy, R. M.; Colton, C. K.; Fisch, M.; Yarmush, M. L. *Mol. Immun.* **1988**, *25*, pp. 17-32.

19. Provencher, S. W. In *Photon Correlation Techniques*; Schulz-Dubois, E. O., Ed.; Springer-Verlag: New York, NY, 1983; pp. 322-328.

20. Valax, P.; Georgiou, G. In *Protein Refolding*; Georgiou, G.; DeBernardez-Clark, E., Eds.; ACS Symposium Series, Vol. 470; American Chemical Society, Washington, D.C., 1991; pp. 97-109.

21. Arakawa, T.; Bhat, R.; Timasheff, S. N. *Biochemistry* **1990**, *29*, pp. 1924-1930.

22. Brems, D. N.; Lin, Y. C.; Stellwagen, E. *J. Biol. Chem.* **1982**, *257*, pp. 3864-3869.

23. Tsong, T. Y. *Biochemistry* **1982**, *21*, pp. 1493-1497.

24. Teschner, W.; Rudolph, R.; Garel, J.-R. *Biochemistry* **1987**, *26*, pp. 2791-2796.

25. Vaucheret, H.; Signon, L.; Le Bras, G.; Garel, J.-R. *Biochemistry* **1987**, *26*, pp. 2785-2789.

26. Arakawa, T.; Timasheff, S. N. *Biophys. J.* **1985**, *47*, pp. 411-414.

27. Schaffer, S. W.; Ahmed, A.; Wetlaufer, D. B. *J. Biol. Chem.* **1975**, *250*, pp. 8483-8486.

28. Cleland, J. L.; Hedgepeth, C.; Wang, D. I. C. *J. Biol. Chem.* **1992**, *in press*.

29. Arawaka, T.; Timasheff, S. N. *Biochemistry* **1985**, *24*, pp. 6756-6762.

30. Tandon, S.; Horwitz, P. M. *J. Biol. Chemistry* **1986**, *261*, pp. 15615-15618.

31. Timasheff, S. N.; Arakawa, T. *J. Crystal Growth* **1988**, *90*, pp. 39-46.

32. Arakawa, T.; Timasheff, S. N. *Biochemistry* **1982**, *21*, pp. 6536-6544.

33. Arakawa, T.; Timasheff, S. N. *Biochemistry* **1984**, *23*, pp. 5912-5923.

RECEIVED March 31, 1992

Chapter 13

Facilitation of Protein Folding and the Reversibility of Denaturation

P. M. Horowitz

Department of Biochemistry, University of Texas Health Science Center, San Antonio, TX 78284–7760

The enzyme rhodanese that is isolated from tissue sources cannot easily be refolded after denaturation, even though it is monomeric, and contains virtually all of the folding information present in the amino acid sequence at synthesis. Studies indicate that folding difficulties are due to the facts that: 1) rhodanese is very sensitive to oxidation that can occur even in the presence of common reductants; and 2) aggregation kinetically competes with folding to the active enzyme. This aggregation is due to the appearance of exposed, interactive hydrophobic surfaces on folding intermediates. Thus, successful refolding requires appropriate reducing conditions, and procedures that limit protein association. Procedures have been developed that permit high levels of refolding. These approaches include the use of: a) refolding at high dilution; b) enzyme immobilization; c) non-denaturing detergents; or d) heat shock or chaperonin proteins, groES and groEL from *E. coli*. The chaperonin assisted refolding is of particular interest since it is related to the mechanism and control of *in vivo* protein folding and transport, and it makes protein folding dependent on ATP hydrolysis.

Many proteins do not refold efficiently after being unfolded, even though it is accepted that the amino acid sequence contains the information that determines protein structure (*1,2*). This raises the questions: What information is in the code for the sequence and what information is in the mechanism or environment in which the code is expressed? i.e. how does the biological machinery affect folding? Among reasons for the apparent discrepancy between expected and observed protein folding is that competing processes kinetically limit renaturation. Thus, though the native state is most stable, it is not accessible in a biologically reasonable time. Aggregation is one of the major kinetic traps that competes with the folding of many proteins. For example, inappropriate association of partially folded proteins has been suggested as leading to the inclusion bodies that limit the production of correctly folded recombinant proteins in *E. coli* (*3,3a*). My laboratory is interested in the practical control of protein folding and the recovery of active proteins from unfolded, modified or recombinant polypeptides.

0097–6156/93/0516–0167$06.00/0

Every isolated protein has an *in vivo* history. The protein must be synthesized, folded, processed, transported and compartmentalized. The protein rhodanese, whose folding properties will be described, is synthesized in the cytoplasm of eucaryotic cells and it is then transported into the matrix of the mitochondrion. These diverse processes presumably have different conformational requirements, and it is suggested that the protein may have been subjected to different folding pressures at different times of its life. Interactions within the cell can modulate the folding potentials of the sequence. Translation of the genetic information yields a polypeptide chain that sequentially appears from the ribosome. The folding itself is influenced by processing that may include covalent changes such as proteolysis, side chain modification or cofactor interactions. Protein folding may occur in the presence of extensive interactive surfaces, and there may be influences from non-covalent protein association due to the high protein concentration (*4*). Interactions with accessory components such as chaperonins or heat shock proteins may modulate the protein folding and/or assist the transport of the protein to its proper compartment (*4-6*). Further processing and conformational adjustments may then be required to generate the active structure. Finally, every protein is subjected to degradation and turnover which involve further processing steps or conformational changes. Because of the different life histories of isolated proteins, we expect a range of behavior in vitro, from proteins that would readily refold to those that have been extensively processed and will not refold at useful rates under arbitrarily chosen *in vitro* conditions.

The enzyme rhodanese has become an interesting model for studying issues related to protein folding and the recovery of recombinant proteins in their active states. Rhodanese is synthesized in the cytoplasm and then it is transported into the mitochondrial matrix (*7*). Rhodanese is isolated from tissue sources as a soluble, globular protein, and its high resolution X-ray structure is available (*8*).

The protein is folded into two independent, equal-sized domains, and the active site of this enzyme, whose job is to transfer sulfur atoms, is at a cysteine residue that is in the vicinity of the interdomain region. The domains are tightly associated, and the interdomain surfaces are strongly hydrophobic. These surfaces can be made accessible with very little change in the structure of the individual domains.

We expect this protein to refold easily for several reasons. First, there is no significant processing of this single chain protein, and essentially all the amino acids in rhodanese at synthesis are present in the mature protein with the exception of the initiating methionine (*9*). Second, no cofactors, metal ions or side chain modifications are required for the observed structure (*8*). Third, active rhodanese has all four of its sulfhydryl groups reduced, and therefore, there is no requirement for the correct formation of disulfides. However, both intra- and inter-chain disulfide bonds can form under some circumstances, but this requires a conformational change from the X-ray structure (*10,11*). In the mitochondrial matrix, rhodanese can partition into soluble and membrane associated forms, another observation that is not compatible with its observed structure in the crystal (*12*).

Clearly, rhodanese has different folding and conformational requirements at different times of its life, even though all folding information with which the protein was born is present in the mature, isolated enzyme.

Until recently, we were not able to renature unfolded rhodanese. Detailed studies of why refolding failed led to ability to refold this enzyme. To begin, rhodanese is very sensitive to oxidation which can be induced by addition of stoichiometric amounts of hydrogen peroxide (*13*). This oxidation leads to a loss of activity and exposure of hydrophobic surfaces as assessed by the increased fluorescence of the hydrophobic probe, 8-anilino-1-naphthalene sulfonate (ANS) (*10*). There is only a short lag in the appearance of hydrophobic surfaces, and the

hydrophobic exposure continues after complete inactivation. Reducing agents such as dithiothreitol (DTT) can form hydrogen peroxide as they autooxidize (*14*). Therefore, it is the partial reduction of oxygen that gives rise to many of the irreversible effects. Oxidized rhodanese can be reactivated if reductants are added soon after inactivation, but it is necessary to choose the correct reducing agents in the correct amounts (*13*). The oxidation becomes increasingly irreversible on further incubation, so that there is a temporal window during which reduction is effective. These oxidized states have some conformational properties that are similar to those measured for intermediates in rhodanese folding. e.g. new antigenic determinants appear that are not present on the native state and have been mapped to the interdomain region (*15*).

Even without oxidation, structural perturbations expose hydrophobic surfaces and lead to aggregation. For example, guanidinium hydrochloride (GdmHCl) at sub-denaturing concentrations can induce rhodanese aggregation (*16*). The precipitation, which occurs maximally at a sharply defined critical concentration of GdmHCl, removes almost all of the protein from solution. The structural perturbation increases the hydrophobic exposure and increases the oxidative susceptibility. The precipitation produces discrete species that are built from a dimer that is initially formed (*10*). The presence of the non-denaturing detergent, lauryl maltoside completely prevents precipitation under conditions where the detergent has no effect on the activity of the native enzyme.

It has become clear that there are at least three experimental requirements for successful refolding of rhodanese. First, oxidation has to be prevented by including an appropriate reductant at concentrations that avoid effects of reactive oxygen species formed during autooxidation. Second, conditions are needed to reduce aggregation that follows exposure of extensive hydrophobic surfaces during unfolding/refolding. Third, it is important to stabilize the protein after refolding. Rhodanese is particularly sensitive, and efficient renaturation has been hampered in the past, because even under conditions favorable for folding, the protein could be inactivated at specific side chains (especially sulfhydryl groups) whose reactivities are enhanced in the folded structure.

Folding is first order, while precipitation is higher order so that aggregation can be reduced by lowering the protein concentration. Unassisted folding of rhodanese is possible if, in addition to lowering the protein concentration, the substrate, thiosulfate, and the reductant, mercaptoethanol are present (*17*). Activity measurements show that there is an optimum protein concentration that maximizes refolding. In general, lower protein concentrations give more recovery, but when the protein concentration is too low there is inactivation from adventitious solution components and surface adsorption (*18*). At a given protein concentration, there is higher recovery at higher urea concentration as long as the urea concentration is below the unfolding transition. These results are consistent with the hypothesis that urea concentrations producing "weakly native" folding conditions are capable of solubilizing "sticky" intermediates that would rapidly aggregate if the urea concentration were reduced too far or too quickly (*17*). Studies of the denaturation transition at fixed protein concentrations show partial reversibility, and the results are consistent with the hypothesis that during the unfolding/folding transition part of the enzyme reversibly unfolds and part is kinetically trapped. As expected, the fraction of the refolding rhodanese that is kinetically trapped increases with increased protein concentration.

There are alternative ways to achieve the essential conditions of preventing association of protein molecules that possess interactive surfaces on folding intermediates. For example, immobilization of rhodanese on Sepharose permits reversible thermal denaturation, but this procedure tends not to be a facile approach to

general protein folding (*19*). Detergent assisted refolding was developed as an alternative way of preventing aggregation (*20,21*). The same detergents are used to facilitate folding as those mild detergents that would be appropriate to solubilize membrane proteins. This approach produces a clear indication that there are folding intermediates that can be kinetically trapped and studied (*20,22*). After denaturation of rhodanese in 6M GdmHCL, there was no reactivation when denatured enzyme was diluted to concentrations higher than 50 ug/ml in buffer supplemented, singly, with substrate, reductant, or detergent alone. The presence of the pairs of reagents: detergent and thiosulfate; or mercaptoethanol and thiosulfate were not very effective. The combination of detergent and reductant was somewhat effective, and the best results under these conditions led to the recovery of about 30% of the activity when the most effective reductant, thioglycolate, was used. The combined effects of the three components was much greater than the sums of their individual effects, and GdmHCl denatured rhodanese regained more than 90% of its activity when allowed to refold and reactivate in the simultaneous presence of detergent, mercaptoethanol and thiosulfate (*23*).

Refolding can also be made reversible in urea by the combined use of lauryl maltoside, thiosulfate and mercaptoethanol (*22*). The observed reversible unfolding/folding transition is asymmetric and can be described as the sum of two individual two-state transitions which indicates that there are intermediate(s) in the folding process. Significant concentrations of the intermediate(s) could be formed under conditions suitable for study by a number of methods. These species were shown to have properties expected for a so-called "molten globule" state described for folding intermediates (*24*). e.g., they have high secondary structure, are compact, have tryptophan residues with increased accessibility to fluorescence quenchers, and they have increased exposure of hydrophobic surfaces.

Aggregation can be prevented, and folding can be enhanced, by association of the partially folded polypeptide with proteins called chaperonins or heat shock proteins (*4-6*). This association provides a controllable, energy-dependent folding mechanism. Chaperonins have been suggested to play a general role in folding and mitochondrial import. They have been proposed to function by forming complexes with nascent or imported chain thereby maintaining import competent conformation(s). This complexation by chaperonin keeps folding proteins from aggregating, and prevents interactive intermediates from getting kinetically "stuck". Further, by introducing an energy requirement, the biological system gains an element of control. We have been using proteins from *E. coli* which consist of the heat shock proteins, cpn60 and cpn10 (also called groEL and groES) to assist refolding of rhodanese. Cpn60 is related to mammalian hsp60 found in the mitochondrial matrix. This protein is an oligomer of 14 60kD subunits arranged as a double doughnut of two stacked 7-mers. Each doughnut appears to contain a relatively small central hole having restricted access to the solvent. These chaperonins are quite effective in refolding rhodanese (*25,26*). The suggested mode of action of these proteins in folding is that cpn60 sequesters interactive intermediates that are in a state approximating a molten globule. This interaction prevents aggregation. Also, the interaction with cpn60 can maintain a non-native, partially folded conformation in a state that is competent to interact with the other elements of the biological machinery responsible for import or intracellular transport. Subsequent interactions with the second protein, cpn10, and ATP lead to completion of folding and release of the bound protein.

The time course of rhodanese refolding in the presence or absence of chaperonins shows the essential features suggested in the above model (*26*). There is some spontaneous folding (approximately 30%; $t_{1/2}$=4 min), and the rate of this folding is faster than seen with the chaperonins ($t_{1/2}$=10 min). This spontaneous folding is suppressed with cpn60 since it binds and stabilizes an inactive intermediate. Folding can be resumed on addition of the missing components, cpn10 and ATP. Detergent

assisted folding is even slower than the chaperonin assisted case ($t_{1/2}$=35 min) but both reach the same level of refolding which is considerably higher than the unassisted case. This may highlight the fact that the goal of the biological system is not to speed up folding but to control it.

Studies with the fluorescent probe BisANS show that there are a small number of hydrophobic sites on cpn60, and these sites can be modulated by conditions that give the refolding cycle. Conditions that would lead to the release of the bound polypeptide reduce the number of accessible hydrophobic sites on cpn60. This may indicate that hydrophobic interactions between the chaperonins and the partially folded polypeptide chain may be an important component of the mode of folding facilitation.

Low temperature inhibits chaperonin assisted refolding, and this observation can be used to determine the stoichiometry of chaperonin interactions (*27*). Using inhibition of spontaneous folding which can be made efficient at low temperature, it appears that only 1-2 molecules of rhodanese are bound per cpn60 oligomer. This may indicate that a small region of rhodanese may be involved, e.g. its N-terminus which has the characteristics described for the import signal peptides that are at the N-terminus of mitochondrial protein precursors.

A working hypothesis can be formulated for folding of rhodanese which incorporates what is known so far about this system, and it can serve as a basis for further experiments. On dilution from high concentrations of denaturants, unfolded rhodanese rapidly collapses to a relatively compact state, driven to a large extent by hydrophobic interactions with formation of a large amount of secondary structure. This could correspond to nucleation of the individual domains. This state has the characteristics associated with the so-called molten globule state in that it is compact, has increased flexibility compared with the native state, is relatively open to access by solute components, has exposed hydrophobic surfaces, and contains sulfhydryl groups in a state in which they can be easily oxidized. This structure is enzymatically inactive. This state will aggregate or associate with surfaces if it is formed in sufficient concentrations and not stabilized. A second step that is rate limiting would allow the enzyme to regain its final global structure to give a native-like conformer that is still enzymatically inactive. The addition of thiosulfate would then permit final local readjustments at the active site and lead to reactivation. In this way, aggregation is in kinetic competition with the acquisition of the active monomeric state. One way to optimize folding rates to the stable monomer state is by preventing aggregation while still permitting sufficient flexibility for the protein to search accessible conformations i.e. there is a need for controlled flexibility. This can be achieved by using accessory substances such as detergents or chaperonin proteins or by otherwise arranging conditions so that
aggregation is limited.

References

1. Kane,J.F. and Hartly,D.L., *Trends Biotechnol.* **6**, *95* (1988).

2. Marston,F.A.O., *Biochem. J.* **1** *240* (1986).

3. Jaenicke,R., *Prog. Biophys. Mol. Biol.* **49**, *117* (1987).

3a.Mitraki,A. and King,J., *Bio/Technology* **7**, *690* (1989).

4. Rothman,J., *Cell*, **59**, *591* (1989).

5. Randall,L.L., Hardy,S.J.S., and Thom,J.R., *Ann. Rev. Microbiol.* **41**, *507* (1987).

6. Goloubinoff,P., Gatenby,A.A. and Lorimer,G.H., *Nature (London)* **342**, *884* (1989).

7. Horowitz,P.M. and Douglas,M.G., *Fed. Proc. Abstr. 1836, 1859* (1981).
8. Ploegman,J.H., Drent,G., Kalk,K.H., Hol,W.G.J., Heinrikson,R.L., Keim,P., Weng,L., and Russell,J., *Nature* **273**, *124* (1978).
9. Miller,D.M., Delgado,R., Chirgwin,J.M., Hardies,S.J., and Horowitz,P.M., *J. Biol. Chem.* **266**, *4686* (1991).
10. Horowitz,P.M. and Bowman,S., *J. Biol. Chem.* **262**, *8728* (1987).
11. Heinrikson,R.L., Weng,L., and Westley,J., *J. Biol. Chem.* **253**, *8109* (1978).
12. Ogata,K. and Volini,M., *J. Biol. Chem.* **265**, *8087* (1990).
13. Horowitz,P.M. and Bowman,S., *J. Biol. Chem.* **264**, *3311* (1989).
14. Misra,H.P., *J. Biol. Chem.*, **249**, *2151* (1974).
15. Merrill,G.A., Horowitz,P.M., Bowman,S., Bentley,K. and Klebe,R., *J. Biol. Chem.* **263**, *19324* (1988).
16. Horowitz,P.M. and Criscimagna,N. L., *J. Biol. Chem.* **261**, *15652* (1986).
17. Mendoza,J.A., Rogers,E., Lorimer,G.H., and Horowitz,P.M., *J. Biol. Chem.* **266**, *13587* (1991).
18. Aird,B.A. and Horowitz,P.M., *Biochim. Biophys. Acta* **956**, *30* (1988).
19. Horowitz,P. and Bowman,S., *J. Biol. Chem.* **262**, *5587* (1987).
20. Tandon,S. and Horowitz,P.M., *J. Biol. Chem.* **265**, *5967* (1990).
21. Tandon,S. and Horowitz,P.M., *J. Biol. Chem.* **264**, *3311* (1989).
22. Horowitz,P.M. and Criscimagna,N. L., *J. Biol. Chem.* **265**, *2576* (1990).
23. Tandon,S. and Horowitz,P.M., *J. Biol. Chem.* **264**, *9859* (1989)
24. Ptitsyn,O.B.J., *Protein Chem.* **6**, *272* (1987).
25. Martin,J., Langer,T., Boteva,R., Schramel,A., HorwichA.L. and Hartl,F.-U., *Nature* **352**, *36* (1991).
26. Mendoza,J.A., Rogers,E., Lorimer,G.H. and Horowitz,P.M., *J. Biol. Chem.* **266**, *13044* (1991).
27. Mendoza,J.A., Lorimer,G.H. and Horowitz,P.M., *J. Biol. Chem.* **266**, *16973* (1991).

RECEIVED May 5, 1992

Chapter 14

Artificial Bifunctional Enzymes

A Tool To Improve Consecutive Enzyme Reactions and Cell Metabolism

Leif Bülow

Department of Pure and Applied Biochemistry, Chemical Center, P.O. Box 124, Lund S–221 00, Sweden

Two artificial bifunctional enzymes, β-galactosidase/galactokinase and β-galactosidase/galactose dehydrogenase, have been prepared by gene fusion and expressed in *E. coli*. The hybrid proteins are able to catalyze the hydolysis of lactose followed by either phosphorylation or oxidation of the galactose formed. Both hybrid enzymes exhibit favorable proximity effects when the coupled enzyme reactions are analyzed. The effect of the connecting segment on the function and stability of the bifunctional enzyme has been studied *in vivo* and *in vitro*. Artificial bifunctional enzymes are not only used in an isolated form but are also valuable tools in studies of cell metabolism to evaluate the aspects of proximity in metabolic maintenance and cell regulation.

SPATIAL ENZYME ORGANIZATION *IN VIVO*

A living cell is a complex entity containing thousands of enzymes involved in the reactions necessary to maintain the cell status. Most biochemical reactions are organized into multistep pathways in which several enzymes can be involved (*1*). Physical arrangements among enzymes in such pathways are less pronounced in lower organisms but in eucaryotic cells many of the different metabolic pathways are frequently compartmentalized in organelles, for instance, the replication of DNA in the nucleus and the tricarboxylic acid cycle in the mitochondrion. Furthermore, it is believed that the enzymes of a specific reaction pathway within a single compartment are complexed in ordered or at least nonrandom arrangements (*2*). Highly organized enzyme systems in processive pathways such

0097–6156/93/0516–0174$06.00/0

as fatty acid synthesis and oxidation as well as nucleotide metabolism have been demonstrated. The enzyme interactions in these pathways are strong and thus have been easy to demonstrate. Such associations between enzymes are usually confirmed when the enzyme activities copurify and the ratio between the activities remains constant during the purification steps.

Multienzymes capable of catalyzing several separate catalytic reactions have been characterized as either multienzyme complexes or multifunctional enzymes. A multifunctional protein is composed of a polypeptide chain(s) carrying two or more active sites. A multienzyme complex also has several active sites, however each on distinct polypeptide chains. These complexes often have a very low dissociation constant (*3*). Expressions like protein machines, enzyme clusters, supramolecular complexes, aggregates and metabolons are all referring to multienzymes. Recommendations for the nomenclature of these multienzymes have now been established and generally accepted (*4,5*).

Multienzymes are often consecutive enzymes in a reaction path or part of a metabolic pathway, i.e. the product of the first enzyme will serve as a substrate for the second enzyme whose product will serve as the substrate for the subsequent enzyme and so forth:

$$\text{Substrate} \xrightarrow{\text{Enzyme 1}} \text{Product 1} \xrightarrow{\text{Enzyme 2}} \text{Product 2} \xrightarrow{\text{Enzyme 3}} \ldots$$

Most obviously, by aggregation of polypeptide chains into multienzyme complexes like the pyruvate dehydrogenase complex or multifunctional proteins like fatty acid synthase, the enzyme activities are gathered efficiently in a "microcompartment" . There is even evidence for specific interactions between many of the so called "soluble" consecutive enzymes in a pathway. Most attention has been focused on the two major metabolic pathways, the glycolytic pathway occurring in the cytosol and the tricarboxylic acid cycle in the mitochondrial matrix. The complexation between the enzymes in glycolysis tends to be weak and thus difficult to demonstrate; the enzymes are associated but the organization is loose and only transient (*6-8*). For instance, direct transfer of NAD(H) between glycerol-3-phosphate dehydrogenase and lactic dehydrogenase has been shown *in vitro* (*9, 10*) but the interpretation of the data has been debated (*11, 12*). Other enzymes in glycolysis have also been investigated (*13, 14*). It has been claimed that the dynamic assembly and disassembly of the transiently existing complexes provide a regulatory mechanism for catalytic activity of the enzymes involved (*15, 16*). It has further been suggested that the association of glycolytic enzymes to the cytoskeletal proteins and actin containing structural elements of the cell is responsible for some compartmentalization in the cytoplasm (*17, 18*).

The protein concentration in the mitochondrial matrix is over 50% and it has been proposed that the enzymes in the Krebs cycle probably also exist and behave as a multienzyme entity rather than as free enzymes in solution (*19*). Several publications have presented results indicating physical interactions between the enzymes in the cycle and metabolically related enzymes (summarized in (*20*)). It has also been possible to detect the existence of the Krebs cycle enzymes as a sequential complex (*21*). Moreover, organization of a reaction cycle is not restricted to only one compartment. The urea cycle has been shown to be structured in a sequential association that spans two cellular compartments, the cytoplasm and the mitochondria (*22*).

Importance of proximity

The arrangement of sequentially operating enzymes into multienzyme complexes and multifunctional enzymes involves several advantages for the reactions catalyzed. Different functional consequences of the organization of the enzymes into multienzyme systems have been proposed. Particularly features gained in cellular metabolism have been emphasized (*2, 20, 23-26*).

Coordination effects. An entire sequence of enzymes can be coordinately activated or inhibited. An example of this phenomenon has been found for the arom conjugate, a multifunctional protein with five distinct consecutive enzymes residing on a dimer of a single polypeptide chain. Four out of the five enzyme activities are activated by the first substrate. In addition, all five activities appear to be coordinately protected from proteolysis when the first substrate is present (*27*).

Compartmentation. A multienzyme system has the potential of compartmentalizing or containing the substrate of a pathway, which implies that the system prevents interference from an enzyme activity outside the metabolic sequence. This efficient transport of molecules from one enzyme to the next without complete equilibration with the surrounding fluid is also known as substrate channeling. Furthermore, labile intermediates can be protected.

Elimination of lag phases. The intermediate substrate formed does not or only partially diffuses out into the surrounding medium, thereby a high local concentration of substrate for the second enzyme can be obtained. The transient time or the lag phase, defined as the time required to attain a steady-state rate in a series of reactions, is thereby also diminished.

Reduction of diffusion times. The transient time, the time required for the product of one enzymatic reaction to diffuse to the active site of the next enzyme is decreased, particularly if the surrounding medium is viscous (cf. the high protein concentration of the mitochondria).

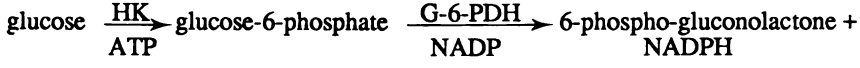
Biosynthesis is more efficient. The most effective coordination of the biosynthesis of enzyme molecules needed for a reaction path can be achieved if the genetic loci are combined into a single unit encoding a single polypeptide chain.

MODEL SYSTEM MIMICING MULTIENZYMES

As previously indicated, the spatial organization of enzymes appears to be the rule rather than the exception in reflecting enzyme sequences *in vivo*. Therefore, it is important to further delineate the potential advantages gained by reconstruction of proximity between enzymes using different model systems. Three main different approaches have been utilized to mimic naturally occuring enzyme systems; co-immobilization of the enzymes to a solid matrix, chemical cross-linking of the enzymes in a random or an oriented fashion, and preparation of artificial multifunctional enzymes by gene fusion.

Co-immobilization

Mosbach and Mattiasson (*27*) co-immobilized a two-enzyme system, hexokinase (HK) and glucose-6-phosphate dehydrogenase (G-6-PDH) which catalyze the sequential reactions:



The initial rate of NADPH formation was considerably enhanced when the two enzymes were co-immobilized, as compared with the situation when they were soluble or immobilized on separate beads (*27*). The two-enzyme system was extended to a three-enzyme system (*28, 29*) and later on the enzymes of the urea cycle were co-immobilized to agarose (*30*). The proximity of the enzymes made the immobilized systems more efficient than the soluble ones since the out-diffusion of the intermediates was hampered due to the unstirred layer surrounding the beads.

Cross-linking of enzymes

Chemical cross-linking between sequentially operating malate dehydrogenase and citrate synthase in the citric acid cycle has been performed to analyze the kinetic behaviour of aggregated enzyme systems (*31, 32*). The kinetic advantages such as increased steady-state rate in the inital phase of the coupled reaction became apparent only when a crowding agent, polyethylene glycol, was included in the assay system. Improvements of the methodology by cross-linking were accomplished when the active sites of two dehydrogenases were spatially arranged face-to-face prior to cross-linking. Diffusion of the product of the first enzyme to the active site of the second enzyme was shown to be facilitated due to the proximity and proper orientation of the active sites (*33*).

Gene fusion

Another efficient approach to obtain proximity between enzymes is to ligate the enzymes on the DNA level. The structural genes of the enzymes of interest are fused in-frame generating an artificial bifunctional enzyme carrying both active sites when the chimeric gene is expressed in a suitable host. Fusions can be made either to the amino- or carboxy-terminal regions of the proteins depending largely on the availability of suitable restriction enzyme sites on the corresponding structural genes. If no such restriction sites are accessible at the 5'- or the 3'- ends of the genes they can be generated by site-directed mutagenesis. By using chemically synthesized DNA fragments in the cloning procedure, special properties in the linker region between the enzymes can be designed by the selection of a certain oligonucleotide sequence. In the construction of artificial fusion enzymes, the three-dimensional structure of the enzymes is most often unknown, but frequently the C- and N-terminus are surface located. A gene fusion therefore normally does not or only to a minor extent interfere with the folding of the protein. If subunit interactions are disturbed or disrupted the fusion can often simply be made at the other end of the gene. The construction prepared is then inserted into a proper expression vector and transformed into a suitable host cell. The effects caused by the bifunctional enzymes on enzyme catalysis can then be analyzed either *in vivo* or *in vitro*.

In many instances, the use of this genetic approach is advantageous over the immobilized and cross-linked enzyme systems. Large amounts of homogeneous bifunctional protein can be produced whereas the degree of crosslinking and homogeneity may vary between different preparations of chemically prepared enzyme conjugates. Additionally, much of the enzyme activity is often lost in the immobilization or cross-linking procedure which is normally not the case for gene fusion. Most often at least 50 % of the wild-type enzyme activity is retained if the entire primary structure of the native enzyme is maintained in the hybrid enzyme

prepared. 2-5 amino acid residues can frequently be removed from the terminus of the enzymes without affecting their activity. However, the ratio between the enzyme activities is more or less fixed in a genetically prepared system while it is easy to change it using the other two methods.

ARTIFICIAL BIFUNCTIONAL ENZYMES

Over the years we and others have prepared a number of different artificial bi- and polyfunctional enzymes (for review see (*34, 35*)). In this paper I will focus on two constructions, β-galactosidase/galactokinase and β-galactosidase/galactose dehydrogenase, which both illustrate several important properties associated with such hybrid enzymes.

β-galactosidase/galactokinase

In the first attempt to construct an artificial bifunctional enzyme *in vitro* a fusion was made between β-galactosidase and galactokinase of *E. coli*, a tetrameric and monomeric enzyme, respectively (*36*). The 5'-end of *galK* was fused to the 3'-end of the structural gene of β-galactosidase, *lacZ*. The bifunctional gene product carried both activities although the β-galactosidase activity was reduced substantially due to 14 missing amino acids in the C-terminus. By adding those residues in a later construction the β-galactosidase activity could be restored (*37*). The bifunctional enzyme β-galactosidase/galactokinase had a tetrameric configuration (Figure 1) and could efficiently catalyze the reaction sequence:

$$\text{lactose} \longrightarrow \underset{+\ \text{glucose}}{\text{galactose}} \xrightarrow{\text{ATP}} \text{galactose-1-phosphate} + \text{ADP}$$

This system was initially used as a model in studies of proximity effects between consecutive enzymes *in vitro*. By utilizing a third enzyme, galactose dehydrogenase, which is competitive to galactokinase, it was demonstrated that the galactose formed is channeled between the enzymes. The vicinity effect became more pronounced relative to a control of native enzymes when polyethylene glycol was used as a crowding agent in the reaction medium.

The linker region

In naturally occuring multifunctional enzymes no homology between the linker regions has been observed (*38*) and the importance of these regions is still unclear. It has been suggested that the correct folding of the protein domains in the polyfunctional *arom* enzyme (*S. cerevisiae*) depends on the linker due to pauses induced by rare combinations of codons near the domain boundaries (*39*). Suitable oligopeptides as candidates for general gene fusion have been presented after the examination of pronounced characteristics of linker peptides joining domains in known tertiary protein structures (*40*). It has been reasoned that the embodied amino acid residues should give the linker some flexibility and allow it to interact with the solvent.

In order to investigate the role of the linker to the function of an artificial bifunctional enzyme, further constructions of β-galactosidase/galactokinase fusion enzymes were made starting from the previous construction (*37*). Oligopeptides of different length and character were introduced between the enzymes. The choice of linker proved to be very important for both the expression of the fusion protein and its stability. Fusion proteins with polyglycine and polyproline linkers encoding polyproline and polyglycine, respectively, were not expressed to the same extent as fusion proteins with amino acids more randomly chosen.

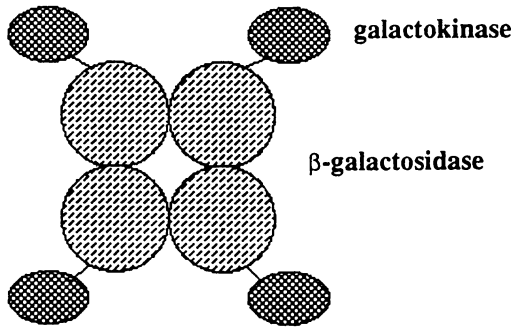


Figure 1. Schematical configuration of the β-galactosidase/galactokinase complex.

Oligonucleotide stretches of the same bases such as oligodC or oligodG were found to be difficult to express in an *E. coli* host. Proteolytic degradation took place both intracellularly as well as extracellularly during purification. This proteolysis occured mainly in the linker region between the catalytic parts. The same phenomenon has been reported for other β-galactosidase fusion proteins as well, e.g. protein A/β-galactosidase (*41*). Linkers with multiple glycine residues in the β-galactosidase/galactokinase hybrids were shown to be the most susceptible ones. Sequences such as gly-gly-X have been shown to be proteolytic processing sites in some biological systems (*42*). Furthermore, glycine is believed to be a flexible residue and thus might destabilize the linker. Several independent strategies to stabilize proteolytically sensitive fusion proteins have been presented (*43*).

Artificial bifunctional enzymes *in vivo*

From all the reports dealing with enzyme interactions *in vitro* a lot of information about cellular organization has been derived. However, as the knowledge of cell metabolism becomes more detailed, the inherent complexity of living systems also becomes more apparent. To understand the importance of a certain enzyme in a metabolic pathway, further studies *in vivo* will no doubt be required. When enzymes are removed from the cells several differences arise that can change the enzyme reactions (*26*):

- There is a potential loss of organization and compartmentation.
- Enzymes are diluted which can affect a number of interactions between macromolecules and small molecules and ions.
- The relative concentration of enzymes and substrates will change. *In vivo* the enzymes are often present in concentration comparable with or greater than their substrates.
- The kinetic parameters, K_m and v_{max}, could be significantly altered *in vivo* (*13*).

Furthermore, the environment of the *in vitro* experiments is quite different from the *in vivo* conditions. For instance, the protein concentration is often much higher. It has been estimated to be 200-400 mg/ml (*44*). This high intracellular protein concentration promotes protein-protein interactions that would not occur in diluted solutions. This volume exclusion effect has been mimicked by the use of polyethylene glycol *in vitro* as previously described.

Therefore, *in vitro* studies have to be extended to include studies on *in vivo* systems. Cell permeabilization is a widely used procedure where it is generally assumed that the protein organization is intact. A simple approach that allows noninvasive studies of cellular metabolism is NMR (*45, 46*). ^{13}C-NMR studies of yeast cells have provided evidence that substrate channeling between the tricarboxylic acid cycle enzymes does occur supporting previous *in vitro* results (*47*).

The different β-galactosidase/galactokinase fusion proteins were used as model systems to evaluate the importance of proximity between two metabolically related enzymes. β-galactosidase and galactokinase catalyze the first steps of lactose metabolism in *E. coli*. To study the function of the different hybrid enzymes *in vivo, E. coli* harbouring various β-galactosidase/galactokinase fusion enzymes were grown on minimal media with lactose as carbon source. Almost no differences in growth rate were visible. However, on introduction of a second plasmid encoding galactose dehydrogenase that scavenges on galactose, pronounced differences became apparent. The product of galactose

dehydrogenase, galactono-lactone, is a metabolic end-product. Therefore, faster growth rates were expected for *E. coli* cells in which the intermediate galactose can be efficiently transferred or channeled to galactokinase. It was demonstrated that *E. coli* cells carrying β-galactosidase/ galactokinase fusion enzymes with short linkers (10 amino acid residues) spacing the two catalytic parts displayed faster growth rates as compared to the cells with bifunctional enzymes with longer linkers (20 residues) or two separate wild-type enzymes. The same behavior due to proximity was also observed by *in vitro* studies on the purified enzymes and crude extracts from cells producing the different β-galactosidase/galactokinase fusion proteins (*48*).

β-galactosidase/galactose dehydrogenase

In order to further characterize artificial bifunctional enzymes a fusion between two oligomeric enzymes was carried out. In this case there is a potential risk of protein polymer formation when the assembly of subunits takes place. However, the fusion of tetrameric β-galactosidase (*E. coli*) to the C-terminus of dimeric galactose dehydrogenase (*Psedomonas fluorescens*) resulted in completely soluble protein complexes (*49*). The oligomeric structure followed that of native β-galactosidase and consisted mainly of tetramers and hexamers. The hinge region, spacing the two enzymes, was only three amino acid residues and it is possible that a longer linker would change the quaternary structure.The galactose dehydrogenase moiety of the complex proved to be more thermostable than its native counterpart while the opposite was true for β-galactosidase.

The potential proximity effects caused by the fusion were investigated by analyzing the kinetics of the coupled enzyme reaction using purified enzymes. The fused enzyme system carried out this reaction more efficiently than a corresponding system composed of native enzymes with the same activities. The transient time of the coupled reaction was hence markedly reduced with the bifunctional enzyme. Furthermore, the overall reaction rate turned out to be higher in the fused system compared with the native system. The reason for the latter effect is not fully understood. A plausible explanation can of course be that substrate channeling is taking place between the active sites. The efficiency of the channeling was found to be dependent on the ratio of the activities between the two enzymes. When the activities of galactose dehydrogenase to β-galactosidase was equal only very small differences in catalytic behavior in comparison with the wild-type enzymes were observed. However, if the ratio was increased (8:1) the difference became pronounced resulting in a two-fold increase in the steady-state rate of the coupled reaction for the fused system compared to the native. This shift in channeling could be achieved simply by changing the pH of the reaction buffer since galactose dehydrogenase has an alkaline pH optimum while β-galactosidase has its optimum at neutral pH. In this context the kinetic parameters K_m and v_{max} must also be considered. A shift in K_m for both lactose and galactose was observed. However, these changes can contribute only partly to the observed differences in kinetic behavior between the fused and native enzyme systems. Similar changes in K_m upon complexation have been observed in natural systems as well. For instance, a decrease in K_m values has been observed when the α-ketoglutarate dehydrogenase complex and succinate thiokinase were mixed and interacted (*50*).

In order to analyze further the subunit interactions of the β-galactosidase/ galactose dehydrogenase fusion protein we decided to express this conjugate together with native galactose dehydrogenase in *E. coli*. When crude extracts were

subjected to gel filtration it was observed that galactose dehydrogenase monomers could interact with the fusion enzyme to form a predominantly tetrameric enzyme complex. This complex cannot be obtained simply by mixing native galactose dehydrogenase and β-galactosidase/galactose dehydrogenase implying that the formation of the new complex is associated with protein translation/folding *in vivo*. This suggests that a relatively well-defined enzyme complex can be formed having one subunit of galactose dehydrogenase attached to each β-galactosidase/galactose dehydrogenase hybrid polypeptide chain. The relative galactose dehydrogenase activity of this complex was increased approximately two-fold compared to the original fusion protein. Moreover, enzyme kinetics experiments revealed that the new complex displayed further reduced transient times for the overall reaction. The thermostability of the β-galactosidase part of the fusion protein was also improved indicating that the embodied proteins behave more native-like.

DISCUSSION

The development of molecular biology has opened new possibilities to structurally link consecutive enzymes by means of in-frame gene fusion, thus creating well-defined artificial bifunctional enzymes. Initially, we fused two sequentially operating enzymes, β-galactosidase and galactokinase, with the aim of studying whether proximity between the enzymes in the system affected the overall reaction kinetics in the conversion of lactose to galactose-1-phospahte. When this fusion enzyme was employed together with a competitive enzyme, galactose dehydrogenase, it was possible to detect a small preference for the pathway utilizing galactokinase. This conclusion was drawn since the added galactose dehydrogenase could not fully compete with the galactokinase of the fusion enzyme for the galactose formed by β-galactosidase. This effect was even more pronounced when polyethylene glycol was added to the solution. Similar conclusions could be made when another bifunctional enzyme, β-galactosidase/galactose dehydrogenase was prepared. This conjugate exhibited a markedly decreased transient time and a higher steady state rate compared to a corresponding reference system of native enzymes.

The use of hybrid bifunctional enzymes is not only useful for basic enzymatic studies but frequently also offers several practical advantages. Their purification is often simplified compared to the isolation of two wild-type proteins since most often the affinity of only one part of the protein is required during a chromatographic procedure. β-galactosidase fusion proteins can easliy be purified using either ion exchange chromatography on DEAE-agarose or affinity chromatography on p-aminobenzyl 1-thio-β-D-galactopyranoside agarose. Such ready-made enzyme conjugates are very attractive in enzymatic analysis where, for example, a dehydrogenase is often coupled to an enzymatic reaction to form an easily monitorable product, NAD(P)H. There are also several biotechnological processes which utilize consecutive enzyme reactions, for instance the enzymatic conversion of starch to high fructose syrup. The use of artificial multienzymes in these cases would be highly advantageous. The multienzyme will speed up the coupled reaction and "guide" the substrates along the enzyme sequence with little loss of intermediates by diffusion. This would be particularly valuable if the enzymes or substrates are expensive or if the intermediates are labile. A fusion protein composed of bovine P450 monooxygenase and yeast reductase was shown to have great potential for hydroxylation of progesteron (*51*).

The introduction of a fused enzyme system in a living cell would also be very valuable to regulate the production of a certain metabolite. By fusing two or more enzymes at a metabolic branch-point the intermediate substrates could be channeled efficiently to the desired pathway. This form of metabolic engineering will no doubt be a valuable tool in the utilization of transgenic microorganisms and plants in the future.

ACKNOWLEDGEMENTS

Helén Carlsson, Christer Lindbladh, Peter Ljungcrantz, Klaus Mosbach and Mats Persson are gratefully acknowledged. This project was supported by grants from the National Science Research Foundation (NFR) and the National Technical Research Foundation (TFR).

Literature Cited

1. Zubay, G.L. *Biochemistry*, MacMillan Publishing Company, New York, **1988**.
2. Srere, P.A. *Ann. Rev. Biochem.* **1988,** *56*, 89-124.
3. Srere, P.A. ; Mathews, C.K. *Methods in Enzym.* **1990,** *182*, 539-551.
4. von Döhren, H. *Trends Biochem. Sci.* **1980,** *5*, VIII.
5. Nomenclature for Multienzymes *Eur. J. Biochem.* **1989,** *185*, 485-486.
6. Keleti, T.; Ovadi, J.; Batke, *J. Prog. Biophys. Mol. Biol.* **1989,** *53*, 105-152.
7. Batke, J. *FEBS Lett.* **1989,** *251*, 13-16.
8. Batke, J. *Trends Biochem. Sci.* **1989,** *14*, 481-482.
9. Srivastava, D.K.; Bernard, S.A. *Science* **1986,** *234*, 1081-1086.
10. Srivastava, D.K.; Smolen, P.; Betts, G.F.; Fukushima, T.; Spivey, H.O.; Bernard, S.A. *Proc. Natl. Acad. Sci. U.S.A.* **1989,** *86*, 6464-6468.
11. Chock, P.B.; Gutfreund, H. *Proc. Natl. Acad. Sci. U.S.A.* **1988,** *85*, 8870-8874.
12. Wu, X.; Gutfreund, H.; Lakatos, S.; Chock, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1991,** *88*, 497-501.
13. Kvassman, J.; Pettersson, G. *Eur. J. Biochem.* **1989,** *186*, 261-264.
14. Kvassman, J.; Pettersson, G. *Eur. J. Biochem.* **1989,** *186*, 265-272.
15. Masters, C.J.; Reid, S.; Don, M. *Molec. Cell. Biochem.* **1987,** *76*, 3-14.
16. Ovadi, J. *Trends Biochem. Sci.* **1988,** *13*, 486-490.
17. Knull, H.R. (1990) in: *Structural and Organizational Aspects of Metabolic Regulation;* Srere, P.A.; Mathews, C.K., Eds.; Liss, New York, 1990, pp. 215-228.
18. Shearwin, K.; Nanhua, C.; Masters, C. *Biochem. Int.* **1989,** *19*, 723-729.
19. Srere, P.A. *Trends Biochem. Sci.* **1980,** *5*, 120-121.
20. Robinson Jr., J.B.; Inman, L.; Sumegi, B.; Srere, P.A. *J. Biol. Chem.* **1987,** *262*, 1786-1790.
21. Watford, M. *Trends Biochem. Sci.* **1989,** *14*, 313-314.
22. Friedrich, P. *Supramolecular Organization,* Pergamon Press, Oxford, U.K., 1984.
23. *Organized Multienzyme Systems;* Welch, G. R., Ed., Academic Press, New York, 1985.
24. Spivey, H.O.; Merz, J.M. *Bioessays* **1989,** *10*, 127-130.
25. Coggins, J.R.; Duncan, K.; Anton, I.A., Boocock, M.R.; Chaudhuri, S.; Lambert, J.M.; Lewendon, A.; Millar, G.; Mousdale, D.M.; Smith, D.D.S. *Biochem. Soc. Trans.* **1987,** *15*, 754-759.
26. *Fundamentals of Enzymology;* Price, N.C.; Stevens, Eds., Oxford University Press, New York, 1989.
27. Mosbach, K.; Mattiasson, B. *Acta Chem. Scand.* **1970,** *24*, 2093-2100.

28. Srere, P.A.; Mattiasson, B.; Mosbach, K. *Proc. Natl. Acad. Sci. U.S.A.* **1973,** *70,* 2534-2538.
29 Mattiasson, B.; Mosbach, K. *Biochim. Biophys. Acta* **1971,** *235,* 253-257.
30. Siegbahn, N.; Mosbach, K. *FEBS Lett.* **1982,** *137,* 6-10.
31, Mattiasson, B.; Johansson, A.-C.; Mosbach, K. *Eur. J. Biochem.* **1974,** *46,* 341-349.
32. Koch-Schmidt, A.-C.; Mattiasson, B.; Mosbach, K. *Eur. J. Biochem.* **1977,** *81,* 71-78.
33. Månsson, M.-O.; Siegbahn, N.; Mosbach, K. *Proc. Natl. Acad. Sci. U.S.A.* **1983,** *80,* 1487-1491.
34. Bülow, L.; Mosbach, K. *TIBTECH* **1991,** *9,* 226-231.
35. Bülow, L. *Biochem. Soc. Symp.* **1991,** *57,* 123-133.
36. Bülow, L.; Ljungcrantz, P.; Mosbach, K. *Bio/Technology* **1985,** *3,* 821-823.
37. Bülow, L. *Eur. J. Biochem.* **1987,** *163,* 443-448.
38. Zalkin, H.; Paluh, J.L.; van Cleemput, M.; Moye, W.; Yanofsky, C. *J. Biol. Chem.* **1984,** *259,* 3985-3992.
39. Purvis, I.J.; Bettany, A.J.E.; Santiago, T.C.; Coggins, J.R.; Duncan, K.; Eason, R.; Brown, A.J.P. *J. Mol. Biol.* **1987,** *193,* 413-417.
40. Argos, P. *J. Mol. Biol.* **1990,** *211,* 943-958.
41. Hellebust, H.; Veide, A.; Enfors, S.-O. *J. Biotechnol.* **1988,** *7,* 185-198.
42. Lopez-Otin, C.; Simon-Mateo, C.; Martinez, L.; Vinuela, E. *J. Biol. Chem.* **1989,** *264,* 9107-9110.
43. Hellebust, H.; Murby, M.; Abrahmsén, L.; Uhlén, M.; Enfors, S.-O. *Bio/Technology* **1989,** *7,* 165-168.
44. Srivastava, D.K.; Bernhard, S.A. *Curr. Top. Cell. Reg.* **1986,** *28,* 1-68.
45. Malloy, C.R. In *Structural and Organizational Aspects of Metabolic Regulation;* Srere, P.A.; Mathews, C.K. Eds.; Liss, New York, 1990, pp. 363-374.
46. Lundberg, P.; Harmsen, E.; Ho, C.; Vogel, H.J. *Anal. Biochem.* **1990,** *191,* 193-222.
47. Sümegi, B.; Sherry, A.D.; Malloy, C.R. *Biochemistry* **1990,** *29,* 9106-9110.
48. Bülow, L. *Biochem. Soc. Symp.* **1991,** *57,* 123-133
49. Ljungcrantz, P.; Carlsson, H.; Månsson, M.-O.; Buckel, P.; Mosbach, K.; Bülow, L. *Biochemistry* **1989,** *28,* 8786-8792.
50. Sümegi, B.; Gyocsi, L.; Alkonyi, I. *Biochim. Biophys. Acta* **1980,** *749,* 172-179.
51. Shibata, M.; Sakaki, T.; Yabasaki, Y.; Muramaki, H.; Ohkawa, H. *DNA* **1990,** *9,* 27-36.

RECEIVED April 9, 1992

Chapter 15

Proteins Designed for Adherence to Cellulose

Edgar Ong, Jeffrey M. Greenwood, Neil R. Gilkes, Robert C. Miller, Jr., R. Anthony J. Warren, and Douglas G. Kilburn

Department of Microbiology, University of British Columbia Vancouver, British Columbia V6T 1Z3, Canada

Molecular genetic techniques have been used to produce fusion proteins containing the cellulose-binding domain (CBD) of the cellulases CenA or Cex from the bacterium *Cellulomonas fimi*. CBD_{CenA} is at the N-terminus of CenA, whereas CBD_{Cex} is at the C-terminus of Cex. Using appropriate cloning vectors, a CBD can be fused either to the N- or the C-terminus of a desired protein, an advantage if fusion at one terminus but not the other inactivates the heterologous protein. Vectors can be modified further to allow construction of fusion proteins containing sites for proteolytic removal of the cellulose-binding domain. The fusion proteins bind tightly to cellulose under normal physiological conditions but can be easily eluted with water or elevated pH, or digested with protease *in situ*, allowing purification almost to homogeneity in a single step. Under appropriate conditions, adsorption of hybrid enzymes to cellulose is effectively irreversible. CBDs thus provide a generic system for enzyme immobilization on an inexpensive, convenient matrix.

Cellulose Binding Domains

Cellulomonas fimi produces an endoglucanase (CenA, 418 amino acids) and an exoglucanase (Cex, 443 amino acids) which bind tightly to cellulose (Figure 1) (*1-3*). Each enzyme comprises a conserved sequence of about 100 amino acids separated from a non-conserved sequence of about 300 amino acids by a linker of 20 proline and threonine residues. The conserved sequence is at the amino terminus of CenA and at the carboxyl terminus of Cex. The conserved sequences are cellulose-binding domains (CBDs) which function independently of the catalytic domains of the enzymes (*4*). Small-angle x-ray scattering analysis shows that CenA is a tadpole-shaped molecule (*5*); Cex has a similar tertiary structure (M. Schmuck and N.R. Gilkes, unpublished results). The CBD and linker form an extended tail region. The CBD forms a hair-pin loop stabilized by a disulfide bond between cysteines near each end.

0097–6156/93/0516–0185$06.00/0

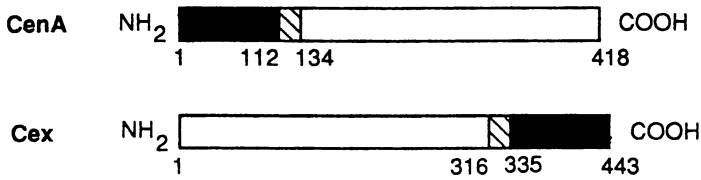


Figure 1. Domain arrangement in CenA and Cex. The catalytic domain (unshaded area) is separated from the cellulose-binding domain (black area) by a Pro-Thr linker (striped area). The numbers refer to the amino acid residues of the mature protein.

Fusion Proteins

The purification of polypeptides produced from cloned genes is a very important aspect of biotechnology. Methods are required which yield highly purified proteins by the simplest, most direct routes. Affinity chromatography takes advantage of the interaction of a protein with a specific ligand coupled to an inert support. The affinity matrices are usually complex and expensive, requiring covalent attachment of the ligand to the support. Enzyme immobilization is another important area of biotechnology. Enzymes can be coupled covalently to inert supports, adsorbed to supports, or trapped within matrices. The preparation of the support or matrix, and the immobilization process itself may require complex and time-consuming steps and may result in a significant loss of catalytic activity. In many instances, the enzyme must be purified before immobilization. The use of CBDs to confer specific adhesive properties to proteins represents a convenient generic solution which allows purification and immobilization in a single step.

The gene fragment encoding the CBD can be fused to the gene for a protein of interest. Using this approach we have generated hybrid proteins which bind tightly to cellulose and exhibit the biological activity of the protein partner. The adsorption of such fusion proteins to cellulose through the CBD provides a simple technique for affinity purification or immobilization. A cleavage site, e.g. for a specific protease such as factor Xa, can be incorporated into the linker region to facilitate recovery of the fusion partner without the CBD. Factor Xa is a member of the blood clotting cascade. It recognizes the tetrapeptide ile, glu, gly, arg (IEGR in the single letter code) and cleaves on the carboxyl side of the arginine residue. The IEGR recognition sequence is found relatively infrequently in proteins. When this sequence is located between the CBD and the NH_2-terminus of the target polypeptide, proteolysis yields the product without extraneous amino acids from the affinity tag.

Protein Purification

A number of constructs have been made to demonstrate these applications. TnPhoA, a transposon derivative containing the alkaline phosphatase gene, was used to generate a series of gene fusions in which portions of CBD_{CenA} gene were linked to the alkaline phosphatase gene. Fusion proteins produced from constructs which contained the entire CBD_{CenA} gene could be readily adsorbed to cellulose and purified by affinity chromatography (*6*). Fusions containing only the N-terminal half of CBD_{CenA} did not bind to cellulose.

The CBD of CenA was fused to the N-terminus of human interleukin 2 (IL2) through a linker incorporating a factor Xa cleavage site. To facilitate cloning, a convenient restriction site within the catalytic domain of CenA was used to generate the CBD fragment. The resulting fusion protein (Figure 2a) contained a short portion of the CenA catalytic domain in addition to the factor Xa site. Cleavage of this protein with factor Xa revealed two spurious cleavage sites within the CenA catalytic domain in addition to the IEGR consensus site These extra cleavage sites were not recognized in a CenA construct containing a factor Xa cleavage site immediately after the Pro-Thr linker (CenA-IEGR). Presumably, unfolding of the CenA catalytic domain sequence as a consequence of its incomplete structure in CenA'-IL2 exposes these sites. A precise fusion of the factor Xa site to the C-terminus of the Pro-Thr linker (Figure 2b) solved this problem, highlighting the need to retain discrete, folded domain structures in hybrid proteins.

Fusion of the CBD_{Cex} to the carboxyl terminus of a ß-glucosidase from *Agrobacterium* (Abg) yielded a hybrid protein, Abg-CBD_{Cex}, with undiminished Abg activity (7). This protein was readily purified from *E. coli* lysates by adsorption

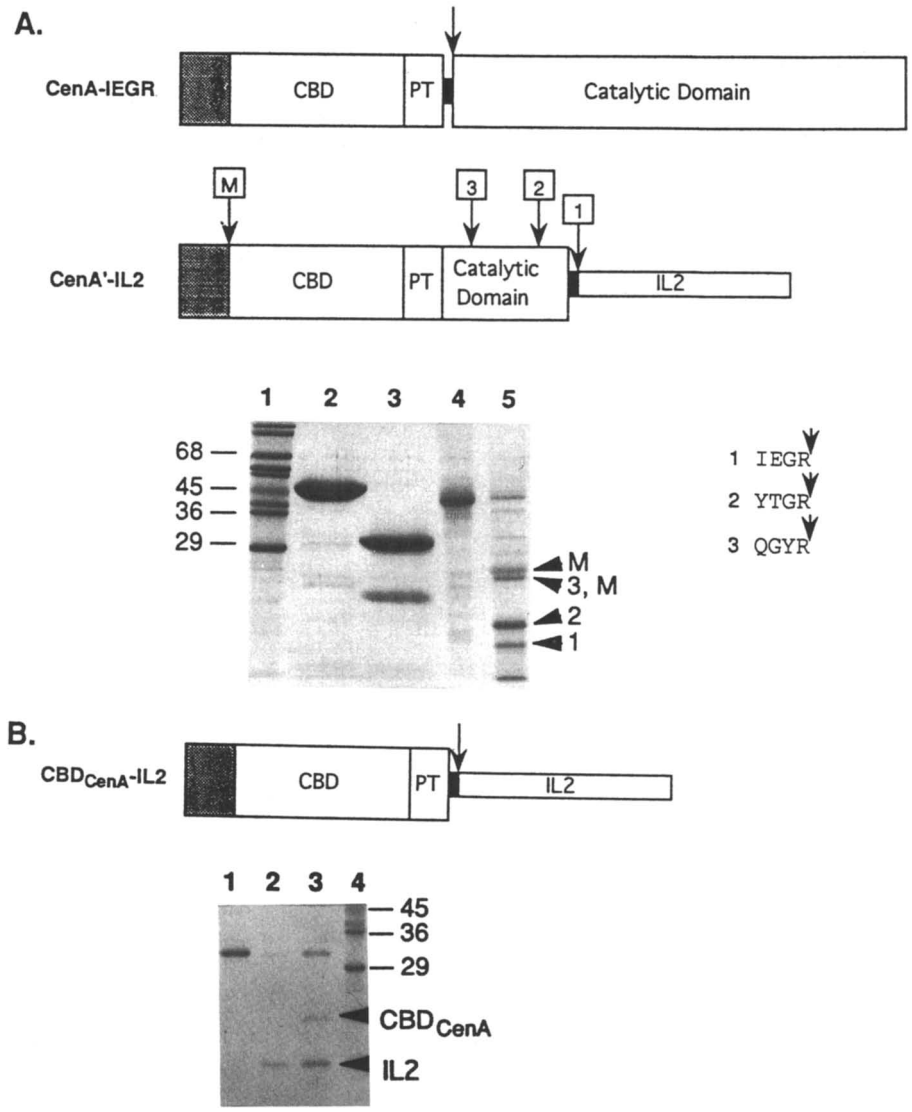


Figure 2. Factor Xa cleavage of CenA-interleukin 2 (IL2) fusion proteins analysed by SDS-PAGE. A: Cleavage of CenA'-IL2 with factor Xa. CenA'-IL2 and control protein CenA-IEGR are shown with the CenA leader peptide and factor Xa cleavage sites indicated in grey and black respectively. The SDS gel shows factor Xa digestion of CenA-IEGR (lane 2—before digestion; lane 3—after digestion) and CenA'-IL2 (lane 4—before digestion; lane 5—after digestion). N-terminal amino acid sequencing of marked bands in lane 5 yielded either the mature N-terminus of CenA (M) and/or three factor Xa cleavage sites (numbered). The locations of these sites in CenA'-IL2 are shown and they are compared with respect to factor Xa recognition sequence. B: Factor Xa cleavage of CBD_{CenA}-IL2 while bound to cellulose. Lane 1—before digestion; lane 2—after digestion, supernatant fraction; lane 3—after digestion, cellulose-bound fraction.

to cellulose (Table I). Binding to cellulose was stable at neutral pH and at ionic strengths from 10 mM to greater than 1 M, which facilitated the removal of non-specifically bound impurities. At neutral pH, the fusion protein could be desorbed from cellulose with distilled water (Figure 3). Alternatively, Abg-CBD_{Cex} could be eluted with 8M guanidinium HCl or by increasing the pH to >8.0. Binding was stable at low pH.

Binding to Cellulose

The adsorption isotherm ($[P]_{ad}$ vs. [P]) of CBD_{Cex} to bacterial microcrystalline cellulose (BMCC) is shown in Figure 4, inset. A Scatchard plot ($[P]_{ad}/[P]$ vs. $[P]_{ad}$) of these data (Figure 4) is non-linear (concave upwards). The cellulose surface can be viewed as a two-dimensional array of overlapping sites (*8*). Therefore, adsorption of cellulases to cellulose cannot be modeled by a simple Langmuir isotherm. Adsorption of DNA-binding proteins on a one-dimensional array has shown that overlapping binding sites inevitably result in a non-linear Scatchard plot (*9*). The presence of more than one type of binding site on cellulose (e.g., amorphous and crystalline) and the negative cooperativity observed at higher concentrations of bound ligand would further increase non-linearity. Detailed modelling of the adsorption process is in progress. A Hanes plot ($[P]/[P]_{ad}$ vs. [P]) (*10*) was used to obtain an estimate of relative affinity of CBD_{Cex} (Table II, legend). The relative affinities of CBD_{Cex} and the fusion protein Abg-CBD_{Cex} are very similar (Table II). The binding affinity of the parent exoglucanase Cex is somewhat less. Comparable values have been found for other CBD fusion proteins. These results confirm that the separate domains of these hybrid proteins function independently and are not influenced by the presence of the adjacent partner.

Adsorption isotherms describing the adsorption of CBD_{Cex} to BMCC at various temperatures (4° to 50°C) and pHs (3 to 11) are shown in Figure 5. Differences in the amounts of total adsorbed protein are apparent at high protein concentrations ($[P]_o > 40$ μM) as saturation is approached. Plots of $[P]/[P]_{ad}$ vs. [P] became non-linear at high protein concentrations ($[P]_o = 30$ to 90 μM) indicating a complex interaction of the CBD_{Cex} with BMCC (data not shown). These results show that CBD_{Cex} binds well to cellulose under a wide range of temperature and pH conditions which may be encountered during operation of an immobilized enzyme column.

Enzyme Immobilization

Cellulose is a convenient matrix for enzyme immobilization. It is cheap, inert and readily available in a variety of pure forms: papers, powders, cotton or membranes. CBD fusion proteins can be immobilized and purified simultaneously by adsorption to cellulose (*11*). Under appropriate conditions binding is virtually irreversible as shown in Figure 6. In this experiment Abg-CBD_{Cex} was bound to a cellulose membrane which was then continuously perfused with a solution of *p*-nitrophenol-ß-D-glucoside (pNPG). Hydrolysis of this chromogenic substrate was monitored by absorbance at 405 nm. No loss in hydrolytic activity was detected over 11 days at 37°C. At 50°C the enzyme lost activity progressively over a period of three days. Examination of the cellulose at the end of the experiment indicated that, although enzyme activity was lost, the protein remained bound to the cellulose.

The thermal stability of immobilized CBD hybrid proteins was further examined using a ß-glucosidase (Cbg) from the thermophile *Caldocellum saccharolyticum* (*12*) linked by its C-terminus to CBD_{Cex}. The resulting fusion protein, Cbg-CBD_{Cex}, was purified from *E. coli* lysates by affinity chromatography on cellulose. The purified fusion protein was then immobilized by adsorption at

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

Table I. Summary of Abg-CBD$_{Cex}$ Purification by Affinity Chromatography on CF1 Cellulose

Purification step	Volume (mL)	Activity (Units.* mL^{-1})	Total activity (Units)	[Protein]** (mg.mL^{-1})	Total [protein] (mg)	Specific activity (Units.mg^{-1})	Yield (%)	Fold-purification
French press cell extract	867	21.9	18987	21.3	18467	1.02	100	1.0
Streptomycin sulfate-treated, filtered cell extract	848	17.4	14755	14.3	12126	1.22	78	1.2
CF1 column flow-through	885	0.12	106	0.78	690			
CF1 column wash	1750	0.98	1719	6.39	11179			
CF1 cellulose eluate (water elution) after Amicon ultrafiltration	22	411.1	9045	3.44	76	119.6	48	117.2

* One ß-glucosidase unit releases one µmol of *p*-nitrophenolate per min from 1.2 mM pNPG in phosphate buffer, pH 7.0 at 37°C.
** As determined by Coomassie blue dye-binding assay.

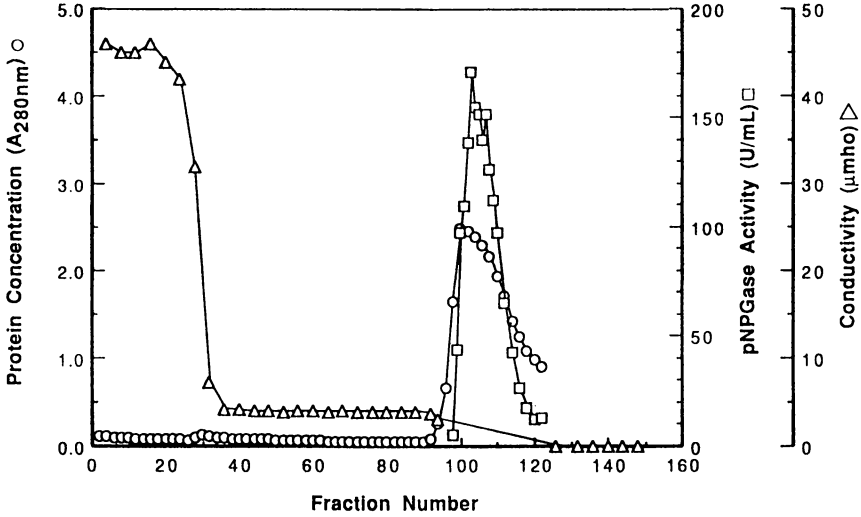


Figure 3. Elution profile of Abg-CBD_{Cex} using affinity chromatography on CF1 (Whatman) cellulose. The fusion protein was eluted using distilled water and 50 mM potassium phosphate buffer, pH 7 in a concave descending gradient.

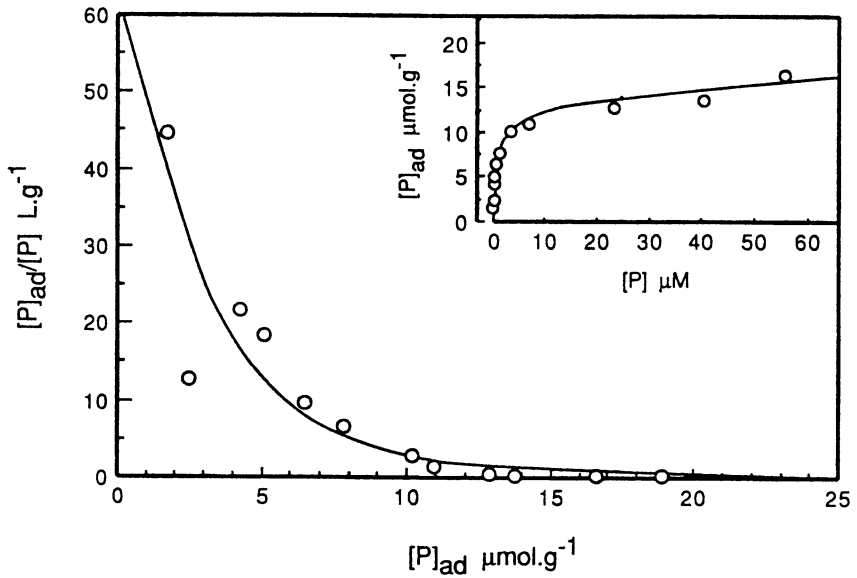


Figure 4. Adsorption of CBD_{Cex} to bacterial microcrystalline cellulose (BMCC). Main panel shows a Scatchard analysis of data for CBD_{Cex} ($[P]_o$ = 0.9 to 90 μM) bound to BMCC (1 mg) at 22°C. Inset shows data plotted as adsorption isotherm. $[P]_o$ refers to initial concentration of CBD_{Cex}. [P] refers to equilibrium concentration of CBD_{Cex} free in solution after adsorption for 24 h. $[P]_{ad}$ refers to concentration of CBD_{Cex} bound to BMCC and is calculated by the difference between $[P]_o$ and [P].

Table II. Relative Affinities and Saturation Levels for CBD_{Cex}, Cex and Abg-CBD_{Cex} for Avicel at 4°C, pH 7*

Protein	Relative Affinity, K_r** ($L.g^{-1}$)	Saturation Level ($\mu mol.g^{-1}$ Avicel)
CBD_{Cex}	7.52	2.70
Cex	1.54	0.85
Abg-CBD_{Cex}	7.81	0.59

*The saturation levels were estimated from plots of $[P]_{ad}$ vs. [P] at high protein concentrations ($[P]_o$ > 40 μM). Relative affinities were estimated from the linear plots of $[P]/[P]_{ad}$ vs. [P] obtained at low protein concentrations ($[P]_o$ (CBD_{Cex}) = 0.9 to 7 μM; $[P]_o$ (Cex) = 1 to 3 μM; $[P]_o$ (Abg-CBD_{Cex}) = 0.15 to 0.88 μM).
**The binding affinity is directly proportional to K_r. $K_r = K_a \cdot [N_o]$ where K_a is the equilibrium association constant and N_o is the total number of binding sites per gram of Avicel.

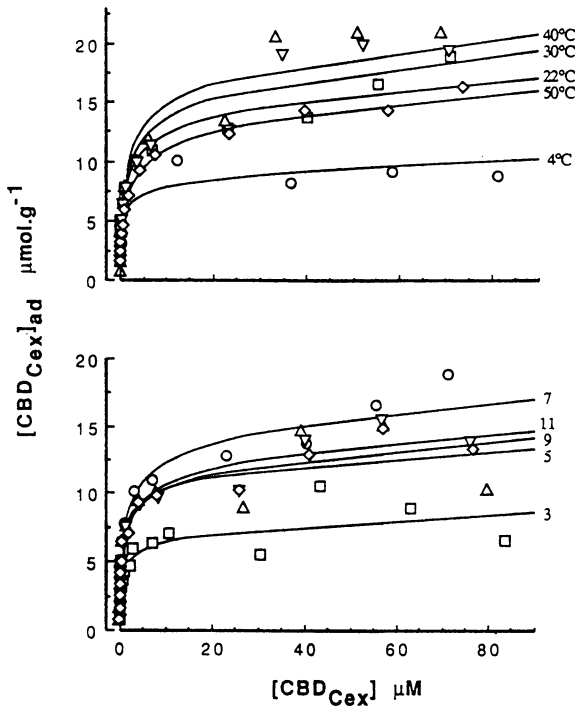


Figure 5. Adsorption of CBD_{Cex} to bacterial microcrystalline cellulose at different temperatures (upper panel) and pHs (lower panel). Details are as in Figure 4. A constant pH of 7 and a constant temperature of 22°C were used for the temperature and pH studies, respectively. Buffers used were 25 mM citrate buffer, pH 3 and 5; 25 mM phosphate buffer, pH 7; and 25 mM carbonate buffer, pH 9 and 11.

room temperature on cellulose acetate membranes. pNPG was perfused continuously through the membranes at 70°C. The activity of the immobilized enzyme was monitored as for Abg-CBD_{Cex}. The immobilized Cbg-CBD_{Cex} was stable at 70°C for more than 70 h (Figure 7). CBD_{Cex} can thus be used for enzyme immobilization at temperature up to at least 70°C.

Our results demonstrate the utility of CBDs for the purification or immobilization of proteins. The stable binding of CBD hybrid proteins to a readily available inexpensive cellulose support provides a convenient generic technology. The CBD can be located at either the N- or C- terminus of the hybrid protein, and, at least for the examples we have tested, can function independently without influencing the biological activity of the fusion partner.

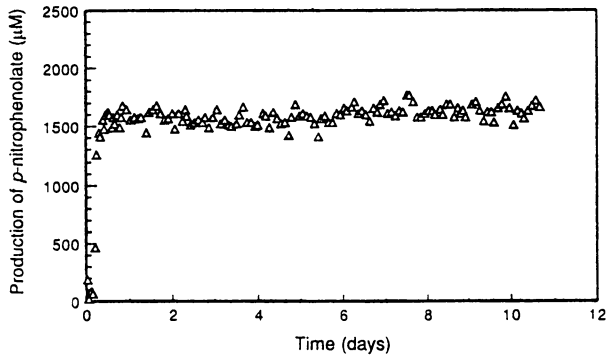


Figure 6. Performance of an Abg-CBD_{Cex} immobilized enzyme column at 37°C. The cellulose used was dewaxed cotton (2 g). Enzyme loading was 1.5 mg protein per g cellulose. Flow rate was 7.8 mL per h. Substrate used was 3 mM pNPG in 50 mM potassium phosphate buffer, pH 7.

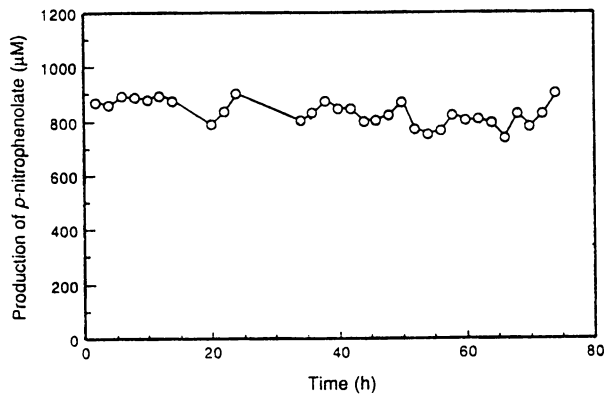


Figure 7. Performance of a Cbg-CBD_{Cex} immobilized enzyme column at 70°C. Five stacked cellulose acetate membranes (25 mm diameter x 1 mm thick) were used at an enzyme loading of 0.25 mg total protein. Flow rate was 9 mL per h. Substrate used was 1.6 mM pNPG in 50 mM phosphate-citrate buffer, pH 6.2.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada. We thank Emily Kwan for technical help.

Literature Cited

1. Wong, W.K.R.; Gerhard, B.; Guo, Z.M.; Kilburn, D.G.; Warren, R.A.J.; Miller, R.C., Jr. *Gene* **1986**, *44*, 315-324.
2. O'Neill, G.P.; Goh, S.H.; Warren, R.A.J.; Kilburn, D.G.; Miller, R.C., Jr. *Gene* **1986**, 44, 325-330.
3. Warren, R.A.J.; Beck, C.F.; Gilkes, N.R.; Kilburn, D.G.; Langsford, M.L.; Miller, R.C., Jr.; O'Neill, G.P.; Scheufens, M.; Wong, W.K.R. *Proteins* **1986**, 1, 335-341.
4. Gilkes, N.R.; Warren, R.A.J.; Miller, R.C., Jr.; Kilburn, D.G. *J. Biol. Chem.* **1988**, 263, 10401-10407.
5. Pilz, I.; Schwarz, E.; Kilburn, D.G.; Miller, R.C., Jr.; Warren, R.A.J.; Gilkes, N.R. *Biochem. J.* **1990**, 271, 277-280.
6. Greenwood, J.M.; Gilkes, N.R.; Kilburn, D.G.; Miller, R.C., Jr.; Warren, R.A.J. *FEBS Lett.* **1989**, 244, 127-131.
7. Ong, E.; Gilkes, N.R.; Warren, R.A.J.; Miller, R.C.; Kilburn, D.G. *Bio/Technology* **1989**, 7, 604-607.
8. Henrissat, B.; Vigny, B.; Buleon, A.; Perez, S. *FEBS Lett.* **1988**, 231, 177-182.
9. McGhee, J.D., Von Hippel, P.H. *J. Mol. Biol.*, **1974**, 86, 469-89.
10. Peitersen, N.; Medeiros, J.; Mandels, M. *Biotechnol. Bioeng.* **1977**, 19, 1091-1094.
11. Ong, E.; Gilkes, N.R.; Miller, R.C., Jr.; Warren, R.A.J.; Kilburn, D.G. *Enzyme Microb. Technol.*, **1991**, 13, 59-65.
12. Love, D.R.; Streiff, M.B. *Bio/Technology* **1987**, 5, 384-387.

RECEIVED March 31, 1992

Chapter 16

Modification of Regulatory Communication in Aspartate Transcarbamoylase

M. E. Wales, C. J. Strang, R. Swanson, and J. R. Wild

Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843–2128

The modification of the T-to-R transition of aspartate transcarbamoylase by the nucleotide end-products of de novo pyrimidine biosynthesis provides a unique opportunity to analyze the structure-function relationships involved in allostery. By combining site-directed substitutions of individual amino acids with larger structural rearrangements between functionally divergent enzymes, it is possible to ascribe functional roles to tertiary and supersecondary components of protein structure. In each of the enzymes examined to date, cytidine-5'-triphosphate (CTP) and uridine-5'-triphosphate (UTP) combine to synergistically inhibit the catalytic efficiency of the various, structurally conserved ATCases. In spite of this common allostery, CTP, UTP, and ATP (adenosine-5'-triphosphate) have different independent effects in various enzymes. The potential for independent pathways for allostery in the enzyme is being examined by protein engineering.

The formation of carbamoylaspartate from aspartic acid and carbamoyl phosphate is the first committed step of pyrimidine nucleotide biosynthesis. The biochemical organization of this step varies from the multi-enzyme complexes of eukaryotic organisms to the independent enzymes found in prokaryotic systems. The allosteric regulation of the prokaryotic enzyme is complex, since the nucleotide end-products differ in their effects on the catalytic activity of the various ATCases. Hybrid enzymes formed with native subunits from different sources demonstrate that the regulatory subunit determines the nature of the allosteric response. Protein engineering studies involving the genetic interchange of selected structural units have resulted in the formation of hybrid enzymes with altered catalytic and regulatory characteristics (*1,2*).

0097–6156/93/0516–0195$06.00/0

ATCases are Oligomeric Enzymes with Regulatory and Catalytic Functions Vested in Different Polypeptides or Within Multifunctional Protein Aggregates.

Aspartate transcarbamoylase (ATCase) possesses tremendous structural and regulatory variety in divergent biological systems (see Table I). For example, the *Escherichia coli* holoenzyme, which is composed of two catalytic trimers (c_3) and three regulatory dimers (r_2), is subject to allosteric regulation by the nucleotide end-products of the pathway (*3,4*). Binding of substrate, or structural analogues, promotes a structural transition (the "T-to-R" transition) in this bacterial enzyme, resulting in homotropic cooperativity. In addition, the binding of heterotropic nucleotide effectors influences the structural transitions of the bacterial dodecamer. The *Bacillus subtilis* ATCase, unlike the *E. coli* enzyme, is composed of catalytic subunits only and has no allosteric controls (*5*). The mammalian enzyme is part of a multifunctional protein aggregate encoding the preceding and subsequent enzymes in the biosynthetic pathway. The ATCase component of this multifunctional enzyme is neither influenced by allosteric responses nor does it experience a cooperative structural transition. Nonetheless, based upon the extent of functional similarities and modeling of predicted structural homologies, the various catalytic subunits of different ATCases appear to share a common ancestry with a very slow evolutionary rate (*6,7*).

Table I. Architecture and Enzymatic Characteristics of Divergent ATCases

Organism	Structure	M_r (x10^{-3})	Kinetics	Allostery
Prokaryotic				
Class A	$(c)_n(r)_n$	340-380	hyperbolic	UTP Inhibition
Class B	$2(c_3):3(r_2)$	275-315	sigmoidal	variable
Class C	(c_3)	100-140	hyperbolic	none
Eukaryotic				
S. cerevisiae	$[CA]_3$	800	hyperbolic	UTP on CPSase
Mammalian	$[CAD]_{3n}$	230-1,200	hyperbolic	UTP on CPSase

SOURCE: Adapted from ref. 2.

Biological Divergence in Regulatory Function Leads to Variation in Allosteric Controls Within the Structurally Complex Class B Oligomers.

Twelve bacterial gene systems of native ATCases have been cloned from divergent bacterial species and overproduced in *E. coli*. Studies with these enzymes have analyzed the basic kinetic and allosteric characteristics. All of the enzymes examined have the same dodecameric organization as the *E. coli*

enzyme ($2c_3$:$3r_2$). These cloned gene systems provide a set of enzymes with functionally unique characteristics which can be used to provide insight into the mechanism of the allosteric response. Several specific differences should be emphasized:

a) Some of the *Yersinia* sp. show no heterotropic response to either CTP or ATP, *E. herbicola* has slight CTP inhibition/no activation and *Y. enterocolitica* has ATP activation/no inhibition (Table II). These enzymes all share the same subunit associations as the *E. coli* enzyme.

Table II. Enzymatic Characteristics of Divergent Class B ATCases

ATCase Type	Bacterial Species	Allosteric Characteristics
ATCase B1 (I)[a]	*E. coli* *S. typhimurium*	CTP, CTP+UTP Inhibition ATP Activation
ATCase B2 (IV)	*Y. intermedia*	CTP, UTP Inhibition ATP Activation
ATCase B3 (V)	*E. carnegiana* *E. herbicola*	sCTP[b], CTP+UTP Inhibition No Activation
ATCase B4 (IV)	*Y. enterocolitica*	No Inhibition ATP Activation
ATCase B5 (IV)	*Y. kristensenii* *Y. frederiksenii*	No Inhibition No Activation
ATCase B6 (II)	*A. hydrophila* *S. marcescens*	No Inhibition CTP, ATP Activation CTP+UTP inhibits Activation
ATCase B7 (III)	*P. vulgaris*	CTP+UTP Inhibition CTP, ATP Activation

SOURCE: Adapted from ref. 2.

[a] Tribal classifications (given in parenthesis) are according to Bergey's Manual of Determinative Bacteriology. B1-B7 indicate subgroups of the ATCase Class B enzymes.

[b] sCTP indicates only slight inhibition by CTP (<20%).

b) The report of UTP/CTP synergistic inhibition of the *E. coli* enzyme (*8*) has been extended to the ATCases from other sources (*P. vulgaris*, *S. marcescens*, and *E. herbicola*). In spite of different allosteric responses to the individual

nucleotides, these enzymes exhibit synergistic inhibition (Table III). The ATCase from *E. herbicola* exhibits a depressed response to CTP and no response to ATP and UTP; nonetheless, CTP and UTP are strong synergistic inhibitors. The enzymes from *S. marcescens* and *P. vulgaris* have activation by both CTP and ATP but have diverged in their response to the combination of pyrimidine biosynthetic end products (CTP + UTP). With the *S. marcescens* enzyme, UTP only counteracts the CTP activation, resulting in activity levels near those without any effectors present. However, in *P. vulgaris* UTP acts as a synergistic inhibitor with CTP to give < 10% of the no effector activity level. This synergistic inhibition of the enzyme by UTP in the presence of CTP provides a regulatory logic for those enzymes which show CTP activation.

Table III. Kinetic and Allosteric Comparison of Purified Native ATCases

Enzyme[a]	$[S]_{0.5}$[b]	n_{app}	CTP[c]	ATP[c]	CTP+UTP[c]
E. coli	5	2.4	40	240	5
S. marcescens	16	2.6	150	400	130
P. vulgaris	28	3.2	150	550	10
E. herbicola	4	2.3	80	100	10

[a] The bacterial source of the ATCase enzyme is identified. In all cases, the holoenzyme (c_6r_6) was studied.
[b] The $[S]_{0.5}$ for aspartate, given in mM.
[c] Percentage relative activity in the presence of 2 mM nucleotide effector at 1/2 of the $[S]_{0.5}$ aspartate of each holoenzyme

In addition to preliminary kinetic characterization, the entire *pyrBI* gene region of four of the cloned gene systems has been sequenced: *E. coli, P. vulgaris, S. marcescens* and *E. herbicola*. Figure 1 presents an alignment of the deduced amino acid sequences of the *pyrI* genes from these strains. This type of comparative sequence analysis allows for the formulation of experimentally testable hypotheses about the function of divergent regions/residues.

In Spite of Regulatory Divergence, ATCases From Various Bacterial Sources Maintain a Highly Conserved Amino Acid Sequence.

The comparison of sequence information provides an essential element in the evaluation of the functional role of individual residues and larger structural components. The alignment of related sequences can provide predictive information that individual sequence examination can not provide. It is

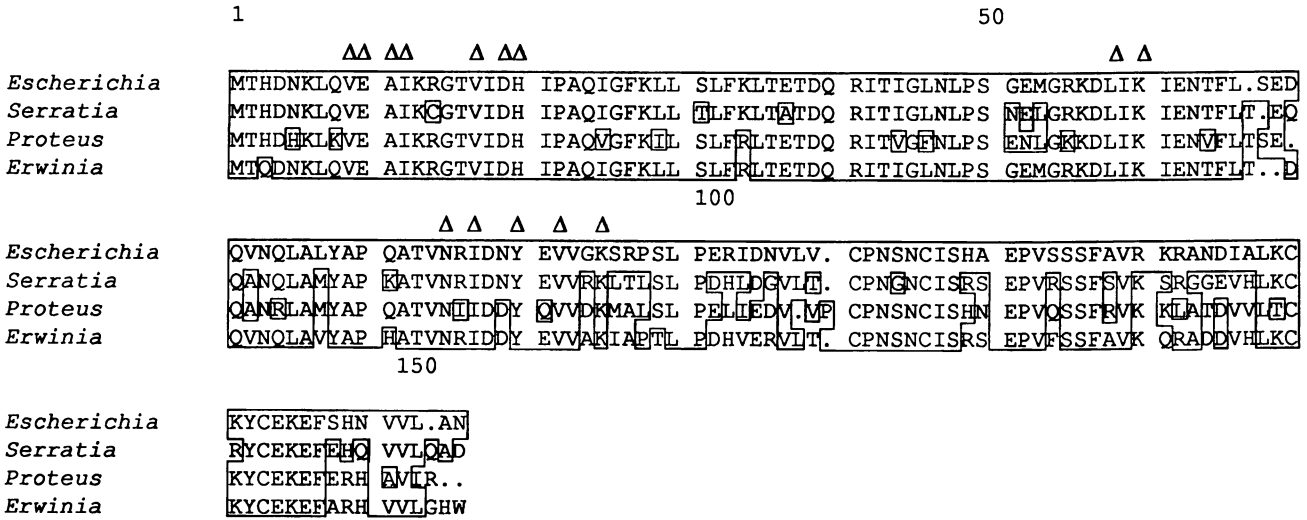


Figure 1. Comparison of the amino acid sequences of the regulatory polypeptide from select members of the *Enterobacteriaceae*. Δ indicates allosteric site residues.

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

possible to identify those residues which are absolutely conserved, those which appear to have constraints (such as charge or size) and to recognize regions in which there appear to be no constraints on substitution. This type of information may reflect the structure-function roles of individual residues and discrete structural regions. The inference is that those residues directly involved in catalysis, substrate binding or folding are constrained, while other residues will have more freedom for substitution during evolution. Table IV summarizes the residues of the allosteric binding site of ATCase which have been implicated in binding either ATP or CTP in the T and R conformations by X-ray crystallography (*9-11*).

Table IV. Residues of the Allosteric Binding Site of ATCase (*9-11*)

T^{ATP}	R^{ATP}	T^{CTP}	R^{CTP}	Chemistry[a]
Val 9			Val 9	ATP, CTP-ribose
Glu 10	Glu 10			ATP-base
Ala 11			Ala 11	ATP, CTP-base
Ileu 12	Ileu 12	Ileu 12	Ileu 12	ATP-base, CTP-base
		Val 17		CTP-3' OH, base
Asp 19	Asp 19	Asp 19	Asp 19	ATP, CTP-ribose
	His 20		His 20	ATP, CTP-γP
Leu 58				ATP-3' OH
Lys 60	Lys 60	Lys 60	Lys 60	ATP, CTP-base, ribose
Asn 84		Asn 84	Asn 84	ATP, CTP-αP
			Ile 86	CTP-base
Tyr 89	Tyr 89	Tyr 89	Tyr 89	ATP, CTP-base
Val 91	Val 91	Val 91	Val 91	ATP,CTP-αP,CTP-ribose
Lys 94	Lys 94	Lys 94	Lys 94	ATP,CTP-triphosphate

[a] Indicates the nucleotides and their components that interact with the amino acid.

The amino acid sequence comparisons of the enteric ATCases were remarkable for the extent of conservation observed among functionally diverged ATCases . When compared with *E. coli* sequences (*12,13*), the residues of the catalytic chain of *S. marcescens* (*14*), *P. vulgaris*, *S. typhimurium* (*15*), and *E. herbicola* have 90%, 78%, 93% and 91% absolute positional identity, respectively. [The issue of positional identity is very important relative to predictive modeling of the structural homologies.] All of the active site residues and most protein interface residues are conserved. The regulatory chains of *S. marcescens*, *P. vulgaris* and *E. herbicola* are slightly less conserved at 77%, 71%, and 80%, respectively, while the *S. typhimurium* sequence had 94% identity. Like the catalytic chain, the allosteric binding site and protein interface

residues of the regulatory chain are extensively conserved (Figure 1). Although all of the directly implicated amino acids forming the structural allosteric site are conserved, there are a number of residues located in secondary positions either very near the allosteric site or very near allosteric site residues that have important consequences (Table V). Natural variation can be used to propose functional relationships within both the catalytic and regulatory polypeptides which have diverged in the ATCases from the various biological sources. For example, the site-directed change of Lys56r to Ala56r resulted in the loss of the activation by ATP and synergistic inhibition by CTP+UTP (*16*). Gly51r in the *E. coli* enzyme is substituted by glutamic acid in *P. vulgaris*, and residue Arg96r in *E. coli* has various polar amino acid substitutions in *S. marcescens*, *P. vulgaris* and *E. herbicola*.

Table V. Site-directed Mutations in the *E. coli* Enzyme Affecting Allostery

Site Substitution[a]	Allosteric Response[b] CTP	ATP	CTP+UTP
Wild-type Enzyme	Inhibition	Activation	Synergism
Allosteric Binding Domain			
Lys 56r - Alanine(*16*)	Inhibition	-0-	-0-
Lys 60r - Alanine(*17*)	-0-	Activation	-0-
Lys 94r - Glutamine(*18*)	Inhibition	-0-	nd
Zinc Binding Domain			
Cys 109 - Histidine	-0-	Activation	Synergism
Asn 111 - Alanine(*19*)	-0-	-0-	-0-
Asn 113 - Alanine(*19*)	Inhibition	Activation	nd
Asn 113 - Glycine(*20*)	Inhibition	Activation	nd
Glu 119 - Aspartate	Inhibition	Activation	Synergism
Arg 130 - Glycine(*21*)	Inhibition	Activation	nd

[a] The position of the substitution is indicated, along with the naturally occurring amino acid and the site-directed alteration (e.g. Lys 56r - Alanine indicates that lysine 56 of the regulatory chain was replaced with an alanine). The reference is indicated in parenthesis after each mutation.

[b] The allosteric response is indicated as either activation, inhibition or synergistic inhibition under the respective allosteric effector. 0 indicates no effect, while nd means no data.

This type of mutational analysis can be enhanced by the use of natural variation, along with the available structural information from X-ray crystallographic refinements.

Refined Structures of Variously Liganded ATCases From *E. coli* Provide Insight into Functional Relationships.

A series of conformations of the aspartate transcarbamoylase of *E. coli* have been elucidated by W.N. Lipscomb and coworkers over the past twenty-five years. The crystal structures of the *E. coli* ATCase have been determined to 2.3 - 2.8 Å resolution in the native unliganded state (*22*), in the presence of the allosteric inhibitor CTP and the allosteric activator ATP (*9-11*), with the bisubstrate analogue N-(phosphonacetyl)-L-aspartate (PALA) (*23-25*), and related substrate analogues (*9,26,27*). The amino terminal "carbamoyl phosphate domain" (CP domain) is involved in binding carbamoylates while the carboxyl terminal "aspartate domain" (ASP domain) is involved in binding aspartate and its analogues. The active sites in each of the catalytic trimers are formed at the interfaces between adjacent catalytic polypeptides, and closure of the two domains of the catalytic polypeptide upon the substrates provides necessary tight binding for catalysis (*28,29*). The binding of purine or pyrimidine nucleotides to the allosteric binding domain (ALLO domain) of the aspartate transcarbamoylases of enteric bacteria initiates either a positive or negative heterotropic effect on the subsequent binding of substrate ligands at the active sites of the enzyme, located over 60 Å away. The coherent transmission of the allosteric signals requires faithful protein:protein communication channels within the regulatory subunit between the allosteric binding domain and the zinc binding domain (ZN domain) and from the zinc binding domain into the catalytic trimers. The nature of this transmission and the mechanism by which allosteric effectors influence the conformational transition of the holoenzyme is still unknown (*30-32*).

In spite of the presence of many, extensively refined structures of aspartate transcarbamoylase, the sheer number of residues in the dodecamer (Mr=310,000) makes it difficult to define the detailed spatial organization of the various components and structural domains. In concert with the construction of hybrid and chimeric varieties of ATCases, a simplified structural description of the holoenzyme has been developed to make it easier to conceptualize the spatial relationships among the domains (Figure 2). This description is based on a central beta sheet structure located in each domain. The holoenzyme is represented in terms of rectangles placed in the position of each beta sheet in the holoenzyme. (The relationships were established by drawing the rectangles to enclose the entire β-sheet of each domain as represented by the molecular graphics program FRODO using the coordinates of the unliganded T-structure. In this way, the visual complexity of the structure was simplified to emphasize the tertiary structure of the domain, while still maintaining the proper spatial relationships.) The darkly shaded rectangles indicate the beta sheet structures of the upper half of the holoenzyme and lighter rectangles represent locations in the lower half. Three catalytic chains form close associations around a central three-fold axis, with the resulting trimers forming the top and bottom of the holoenzyme. The regulatory chains associate as dimers; these dimers are located around the circumference of the

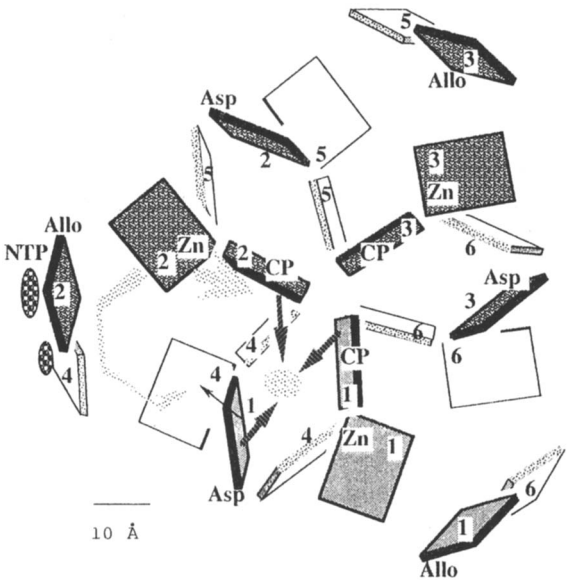


Figure 2. The holoenzyme represented in terms of rectangles in the position of the beta sheets in each domain (see text).

catalytic trimers and bridge the two trimers. In the regulatory polypeptides, the two adjacent allosteric domain beta sheets (e.g. allo 1:allo 6) dimerize into one continuous, 10-stranded sheet. This 10-stranded sheet is the locus of the nucleotide binding sites, one at each end of the sheet, on the outside surface. One pair of nucleotide binding sites is marked by textured ovals in Figure 2 ("NTP").

The beta sheets in the domains of the catalytic trimers are arranged so that they surround each active site on three sides. The wide, textured arrows schematically indicate the contributions from the three different domains to one of the active sites. This conceptualization expresses the way in which the active sites are formed "at the interfaces between adjacent catalytic polypeptides." In the nomenclature established by W.N. Lipscomb and co-workers, this active site is formed from residues contributed by catalytic chains c1 and c2. It is to be noted that each T-state active site is bridged by one regulatory dimer. For example, Zn domain 4 interacts with the 240s loop of active site 1, while Zn domain 2 interacts near the 80s loop of active site 1. These two loops, the 240s loop from c1 and the 80s loop from c2, move in to form the active site in the R state.

A zinc domain beta sheet maintains a relatively fixed relation to its associated carbamyl phosphate beta sheet between the T and R states. The two allosteric beta sheets of the dimer also move as a nearly rigid unit. However, the zinc domain and allosteric domain beta sheets of a single regulatory chain change their relative angle by about 8°, and their separation by about 0.5 Å. Thus, the molecule exhibits rigid links in the two main non-covalent associations made by the regulatory chains, but allows for movement between the covalently joined zinc and allosteric domains.

The Zinc Domains Transfer the Allosteric Signal From the Regulatory Subunits into the Catalytic Subunits.

As described in the previous section, all of the contact residues between the regulatory and catalytic chains are within the Zinc domain. Thus it is the Zinc domain which physically must transmit the regulatory signals. There are four contact regions involving the zinc domain and either the signal source (the allosteric binding domain) or the signal target (the catalytic chains). One of these regions, a "source" contact, is the large hydrophobic zinc:allo interface. This interface mediates the 8° rotation of the allosteric domain relative to the zinc domain during the T→R transition. The other three zinc contact regions involve the catalytic chains, or the "target" contacts.

The most extensive and geometrically unchanging contact with the catalytic chains involves the region of the zinc domain containing three of the four zinc-binding cysteines (residues 113-121 and 137-142). As would be expected for any essential structural feature, the amino acid sequence in this region is extensively conserved. The next most extensive interaction involves the short stretch between residues 109-114 (both of which are zinc-binding cysteines) and the C-terminal helix. This interaction is striking for two reasons:

1) the carboxy-terminus is one of the regions which exhibit maximal sequence disparity among the different ATCases, and 2) this region interacts with the catalytic chain loop (the 240s loop) that moves in to contribute to aspartate binding in the active site. This interaction exist only in the T-state of the enzyme. Finally, a sketchy interaction exists between the 130s loop of the zinc domain and helix 7 (H7) in the aspartate domain. Both of these regions exhibit high sequence variability when compared across the known sequences.

Thus, the zinc domains mediate the response of the holoenzyme to nucleotide effectors in a very interesting way. Each active site integrates input from three different nucleotide binding sites via the different zinc domains it contacts. For example, active site 1 can be influenced by 1) zinc domain 2 interactions near the 80s loop, 2) zinc domain 4 interactions with the 240s loop or 3) a long interaction between zinc domain 1 and H7 of aspartate-binding domain 1. Conversely, each nucleotide effector binding site signals three different active sites indirectly through its associated zinc domain.

Demonstrated Ability to Form Hybrid Oligomeric Enzymes From Native Regulatory and Catalytic Subunits.

An alternative to single residue mutagenesis that can provide insight into the role of protein interfaces is the use of hybrid enzymes. The first such hybrid enzyme was formed in vitro from purified catalytic subunits of *Salmonella typhimurium* and regulatory subunits from *E. coli* (*33*). These two native enzymes are very similar, and the hybrid enzyme results were somewhat inconclusive. Since that time, this approach has been extended to the in vivo construction of hybrid ATCases utilizing the cloned regulatory subunits of one bacterial enzyme and the catalytic subunits of another (*34*). The formation of over twenty hybrid holoenzymes has resulted in the following observations about the nature of the structure-function relationships of the subunits:

1. Stable hybrid enzymes can be formed both in vivo from heterologous subunits expressed from different plasmids or assembled from separated subunits in vitro (*33-36*).
2. The hybrid enzymes demonstrate a pattern of allosteric control that is determined by the nature of the regulatory subunit.
3. Most of the hybrid enzymes formed from native subunits are catalytically efficient as indicated by low $S_{0.5}$ values of 3-8 mM aspartate and the maintenance of maximal velocity.
4. The r:c protein:protein interfaces provide important molecular interactions which affect the enzymatic characteristics expected in the T-R conformational transition of the enzyme.

One of the first hybrid enzymes formed (c_3-*Serratia marcescens*::r_2-*E. coli*) was a stable oligomer of 300 kD which possessed homotropic kinetic responses and was subject to activation by ATP and inhibition by CTP (*36*). In spite of this typical allosteric response, the enzyme appeared catalytically paralyzed,

requiring an increase in aspartate concentration at the $S_{0.5}$ from 5 mM to 125 mM. In most other cases, the hybrids were efficient for catalysis and, independent of the catalytic effects, the regulatory subunits always determined the nature of the allosteric response (*2*). This model of control extends across the widest pattern of hybrid enzyme formation (see Table VI).

Table VI. The Regulatory Chain Dictates the Allosteric Response

ATCase Organization[a]		Response to Effectors (%)[b]		
catalytic subunit	regulatory subunit	CTP	ATP	UTP
E. coli	*E. coli*	40	140	ne
	S. marcescens	125	135	ne
	P. vulgaris	140	170	ne
S. marcescens	*E. coli*	50	130	ne
	S. marcescens	135	150	ne
	P. vulgaris	145	150	ne
P. vulgaris	*E. coli*	25	130	ne
	S. marcescens	115	125	ne
	P. vulgaris	130	190	ne

[a] All enzymes are class B ATCases with catalytic trimers = 100,000 Da and holoenzymes = 300,000 Da.
[b] Percent relative activity in the presence of 2 mM nucleotide. Percentages below 100% represent inhibition (e.g. 5% relative activity indicates 95% inhibition), percentages above 100% are activation (e.g. 180% relative activity indicates 80% activation), ne means no effect.

Interchange of sub-domain modules of divergent ATCases to examine the structural effect of secondary and supra-secondary structures in transmitting the allosteric signals

Since the studies with hybrid enzymes demonstrated that the regulatory subunits impose their unique heterotropic responses on associated catalytic trimers in hybrid holoenzymes, the next step is to genetically reconstruct chimeric enzymes to distinguish more basic structural contributions to the catalytic and regulatory characteristics of the different enzymes. For that reason, a collection of chimeric enzymes was formed from genetically engineered subunits containing domains from the regulatory polypeptides possessing diverse allostery. The constructions currently available are derived from the ATCases of *E. coli*, *S.marcescens* and *P. vulgaris*. Those constructions which have been formed and

the preliminary data available from these chimeric enzymes is presented in Table VII.

Table VII. Kinetic and Allosteric Responses of Chimeric Enzymes

Enzyme[a]	$[S]_{0.5}$[b]	n_{app}	Response to Effectors (%)[c] ATP	CTP	C+U
C_{ec}:$R_{zn\text{-}sm::allo\text{-}ec}$	3.0	1.8	130	ne	40
C_{ec}:$R_{zn\text{-}ec::allo\text{-}sm}$	13.5	1.7	180	ne	30
C_{ec}:$R_{zn\text{-}ec::allo\text{-}pv}$	2.7	1.3	ne	ne	60
C_{ec}:$R_{zn\text{-}pv::allo\text{-}ec}$	4.0	1.2	ne	ne	88
C_{sm}:$R_{zn\text{-}sm::allo\text{-}ec}$	6.0	2.0	150	80	3
C_{sm}:$R_{zn\text{-}ec::allo\text{-}sm}$	140.0	3.3	110	ne	60
C_{sm}:$R_{zn\text{-}pv::allo\text{-}ec}$	4.0	1.3	ne	ne	65
C_{sm}:$R_{zn\text{-}ec::allo\text{-}pv}$	9.0	1.6	120	5	5
C_{pv}:$R_{zn\text{-}pv::allo\text{-}ec}$	3.7	1.0	ne	ne	ne
C_{pv}:$R_{zn\text{-}ec::allo\text{-}pv}$	4.0	1.2	ne	ne	79

[a] C, catalytic subunit; R, regulatory subunit; ec, *E. coli*; sm, *S. marcescens*; pv, *P. vulgaris*; zn, zinc domain; allo, allosteric domain; chimeric notation: C_{ec}:$R_{zn\text{-}ec::allo\text{-}sm}$, *E. coli* catalytic subunit associated with a chimeric regulatory subunit composed of the *E. coli* zinc domain and the *S. marcescens* allosteric domain.
[b] The $[S]_{0.5}$ for aspartate, given in mM.
[c] Percent relative activity in the presence of 2 mM nucleotide (see Table VI).

These studies have shown that chimeric regulatory subunits will stably associate with either *E. coli*, *S. marcescens* or *P. vulgaris* catalytic subunits. However, it appears that neither domain alone is sufficient to promote the characteristic allosteric response seen in the native enzymes and in all hybrids examined so far. In evaluating the native enzymes, the hybrid enzymes and these chimeric enzymes, the one region which is perturbed in these chimeric enzymes that is not altered in either of the other configurations, is the zinc::allosteric interface. As mentioned above, this interface undergoes significant rearrangement during the T→R transition. This is consistent with the observation that six of these enzymes have dramatically reduced Hill coefficients. What is surprising is that, in spite of this reduced cooperativity, most of the enzymes (all but one: C_{sm}:$R_{zn\text{-}ec::allo\text{-}sm}$) exhibit very low aspartate requirements. The singular exception is the C_{sm}:$R_{zn\text{-}ec::asp\text{-}sm}$ enzyme has an $[S]_{0.5}$ of 140 mM aspartate, providing further evidence that the contact between the *E. coli* zinc domain and the *S. marcescens* catalytic subunit residues results in a significantly increased substrate requirements.

Conclusion

From the viewpoint of subunit and domain exchanges of ATCases, the holoenzyme is quite resilient. Its structure-function relationships are complex yet appear to be modular and robust. The hybrid and chimeric enzymes are catalytically active and often retain consistent regulatory characteristics, although the values of various enzymatic parameters shift slightly. However, when site-directed mutations are evaluated, the enzyme appears to be more easily disrupted by a single change. The difference between these two observations is that the shuffling of evolutionarily derived units, despite their multiple amino acid differences, may be intrinsically more compatible than single site-directed alterations. In evaluating the significance of divergence between the enteric ATCases, it should be considered that the nucleotide-free ATCase is not the physiological condition. Under physiological concentrations of nucleotides, the ATCase would be liganded, so that the unliganded-catalytic response has not been tuned by selection. Perhaps this is the reason for the activation-inhibition variability seen among the native enzymes. In spite of this, all the regulated ATCases show the physiologically logical behavior of equal or higher activity with ATP than with CTP, and equal or still lower activity with both CTP and UTP.

Acknowledgments

We acknowledge the National Institute of Health, the National Science Foundation, and the Robert A. Welch Foundation for ongoing support of this research.

Literature Cited

1. Wild, J.R.; Grimsley, J.K.; Kedzie, K.M.; Wales, M.E. In *Chemical Aspects of Enzyme Biotechnology*, Baldwin, T.O., Rauschel, F.M. and Scott, A.I., Ed., Plenum Press: New York, New York, 1991, pp. 95-109.
2. Wild, J.R.; Wales, M.E. *Annu. Rev. Microbiol.* **1990**, *44*, pp. 93-118.
3. Gerhart, J.C. and Pardee, A.B. *J. Biol. Chem.* **1962**, *237*, pp. 891-896.
4. Gerhart, J.C. and Schachman, H.K. *Biochemistry* **1965**, *4*, pp. 1054-1062.
5. Brabson, J.S.; Switzer, R.L. *J. Biol. Chem.* **1975**, *250*, pp. 8664-8669.
6. Major, J.G. PhD dissertation, Texas A&M University, 1989. pp. 1-122.
7. Scully, J.L.; Evans, D.R. *Proteins* **1991**, *9*, pp. 191-206.
8. Wild, J.R.; Loughrey, S.J.; Corder, T.C. *Proc. Nat. Acad. Sci. USA* **1989**, *86*, pp. 52-56.
9. Ke, H.M.; Honzatko, R.B.; Lipscomb, W.N. *Proc. Nat. Acad. Sci. USA* **1984**, *81*, pp. 4037-4040.
10. Gouaux, J.E.; Stevens, R.C.; Lipscomb, W.N. *Biochemistry* **1990**, *29*, pp. 7702-7715.
11. Stevens, R.C.; Gouaux, J.E.; Lipscomb, W.N. *Biochemistry* **1990**, *29*, pp. 7691-7701.

12. Hoover, T.A.; Roof, W.D.; Foltermann, K.F.; O'Donovan, G.A.; Bencini, D.A.; Wild, J.R. *Proc. Nat. Acad. Sci. USA* **1983**, *80*, pp. 2462-2466.
13. Schachman, H.K.; Pauza, C.D.; Navre, M.; Karels, M.J.; Wu, L.; Yang, Y.R. *Proc. Nat. Acad. Sci. USA* **1984**, *81*, pp. 115-119.
14. Beck, D.A.; Kedzie, K.M.; Wild, J.R. *J. Biol. Chem.* **1989**, *264*, pp. 16629-16637.
15. Michaels, G.; Kelln, R.A.; Nargang, F.E. *Eur. J. Biochem.* **1987**, *166*, pp. 55-61.
16. Corder, T.S.; Wild, J.R. *J. Biol. Chem.* **1989**, *264*, pp. 7425-7430.
17. Zhang, Y.; Kantrowitz, E.R. *Biochemistry* **1989**, *28*, pp. 7313-7318.
18. Zhang, Y.; Landjimi, M.M.; Kantrowitz, E.R. *J. Biol. Chem.*, **1988**, *263*, pp. 1320-1324.
19. Eisenstein, E.; Markby, D.W.; Schachman, H.K. *Proc. Nat. Acad. Sci. USA* **1989**, *86*, pp. 3094-3098.
20. Xu, W.; Pitts, M.A.; Middleton, S.A.; Kelleher, K.S.; Kantrowitz, E.R. *Biochemistry* **1988**, *27*, pp. 5507-5515.
21. Stebbins, J.W.; Kantrowitz, E.R. *J. Biol. Chem.* **1989**, *264*, pp. 14860-14864.
22. Kim, K.H.; Pan, Z.; Honzatko, R.B; Ke, H-m; Lipscomb, W.N. *J. Mol. Biol.* **1987**, *196*, pp. 853-875.
23. Ke, H.; Lipscomb, W.N.; Cho, Y.; Honzatko, R.B. *J. Mol. Biol.* **1988**, *204*, pp. 725-747.
24. Krause, K.L.; Volz, K.W.; Lipscomb, W.N. *J. Mol. Biol.* **1987**, *193*, pp. 527-553.
25. Volz, K.W.; Krause, K.L.; Lipscomb, W.N. *Biochem. Biophys. Res. Commun.* **1986**, *136*, pp. 822-
26. Gouaux, J.E.; Lipscomb, W.N. *Proc. Nat. Acad. Sci. USA* **1988**, *85*, pp. 4205-4208.
27. Gouaux, J.E.; Lipscomb, W.N. *Biochemistry* **1990**, *29*, pp. 389-402.
28. Kantrowitz, E.R.; Lipscomb, W.N. *TIBS* **1990**, *15*, pp. 53-59.
29. Kantrowitz, E.R.; Lipscomb, W.N. *Science* **1988**, *241*, pp. 669-74.
Wild, J.R.; Johnson, J.L.; Loughrey, S.J. *J. Bacteriol.* **1988**, *170*, pp. 446-448.
30. Hervé, G. In *Allosteric Enzymes*; Editor, G. Hervé, CRC Press: Boca Raton, Florida, 1989, pp. 61-79.
31. Stevens, R.C.; Lipscomb, W.N. *Biochem. Biophys. Res. Commun.*, **1990**, *171*, pp. 1312-1318.
32. Schachman, H.K. *J. Biol. Chem.*, **1988**, *263*, pp. 18583-19262.
33. O'Donovan, G.; Holoubek, H.; Gerhart, J.C. *Nature New Biology*, **1972**, *238*, pp. 264-266.
34. Foltermann, K.F.; Shanley, M.S.; Wild, J.R. *J. Bacteriol.*, **1983**, *157*, pp. 891-898.
35. Shanley, M.S. **1988**, PhD Dissertation, Texas A&M University, College Station, Texas.
36. Shanley, M.S.; Foltermann, K.F.; O'Donovan, G.A.; Wild, J.R. *J. Biol. Chem.* **1984**, *259*, pp. 12672-12677.

RECEIVED March 31, 1992

Chapter 17

C_1-Tetrahydrofolate Synthase

Dissection of Active Site and Domain Structure by Protein Engineering

Anice E. Thigpen[1], Charles K. Barlowe[2], and Dean R. Appling[3]

Department of Chemistry and Biochemistry, University of Texas, Austin, TX 78712

C_1-tetrahydrofolate (H_4folate) synthase is a trifunctional protein possessing the activities 10-formyl-H_4folate synthetase, 5,10-methenyl-H_4folate cyclohydrolase, and 5,10-methylene-H_4folate dehydrogenase. The current model divides the eukaryotic protein into two functionally independent domains with dehydrogenase/cyclohydrolase activities sharing an overlapping site on the N-terminal domain and synthetase activity associated with the C-terminal domain. In prokaryotes, these activities are generally catalyzed by separate monofunctional enzymes. The genes or cDNAs encoding members of this family of enzymes have been isolated from mammalian, yeast, and bacterial sources and reveal considerable conservation of amino acid sequence. In an effort to elucidate structure/function relationships in these proteins, we have utilized several recombinant DNA methodologies to dissect active site and domain structure in this family of enzymes. Site-directed and random mutagenesis has been used to identify important amino acid residues in the overlapping dehydrogenase/cyclohydrolase active site. Deletion analysis and construction and expression of chimeric forms of the trifunctional enzyme in which the two domains derive from different species have provided insights into domain structure and function.

Folate-mediated one-carbon metabolism plays an essential role in several major cellular processes including nucleic acid biosynthesis, mitochondrial and chloroplast protein biosynthesis, amino acid biosynthesis and interconversions, vitamin metabolism, and

[1]Current address: Department of Molecular Genetics, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75235
[2]Current address: Department of Biochemistry, University of California, Berkeley, CA 94720
[3]Corresponding author

0097–6156/93/0516–0210$06.00/0

methyl group biogenesis. These pathways are summarized in Figure 1. The variety of the 3-carbon of serine, derived from glycolytic intermediates (*1*). The one-carbon unit is transferred to H_4folate in a reaction catalyzed by serine hydroxymethyltransferase (reaction 4, Figure 1), generating 5,10-methylene-H_4folate and glycine. This form of the coenzyme can be used directly in *de novo* dTMP synthesis by thymidylate synthase (reaction 6, Figure 1). 5,10-Methylene-H_4folate may also be reduced to 5-methyl-H_4folate or oxidized to 10-formyl-H_4folate depending on the needs of the cell. In rapidly growing cells, the synthesis of purines is a critical folate-dependent pathway, requiring two moles of 10-formyl-H_4folate per mole of purine ring. 5,10-Methylene-H_4folate is oxidized to 10-formyl-H_4folate via the sequential enzymes 5,10-methylene-H_4folate dehydrogenase and 5,10-methenyl-H_4folate cyclohydrolase (reactions 3 and 2, respectively). Formate represents another potential one-carbon donor (*2*) and is activated in an ATP-dependent reaction catalyzed by 10-formyl-H_4folate synthetase (reaction 1, Figure 1).

In prokaryotes, reactions 1-3 (Figure 1) are catalyzed by three separate monofunctional enzymes, with the known exceptions of *Escherichia coli* and *Clostridium thermoaceticum* in which the cyclohydrolase and dehydrogenase activities are catalyzed by bifunctional proteins. In eukaryotes, these three activities are present on one polypeptide in the form of a trifunctional enzyme. This enzyme, termed C_1-tetrahydrofolate synthase (C_1-THF synthase), is thus responsible for interconversion of the one-carbon unit between the formate and formaldehyde oxidation levels.

All of the known eukaryotic C_1-THF synthases exist as homodimers of approximately 100,000 Da. The current model divides the enzyme into two functionally independent domains with dehydrogenase/cyclohydrolase activities sharing an overlapping active site on the N-terminal domain and synthetase activity associated with the C-terminal domain. This model is supported by several lines of evidence. For example, limited proteolysis results in physical separation of activities, with the synthetase activity associated with a large proteolytic fragment (subunit M_r = 60,000-80,000) and the

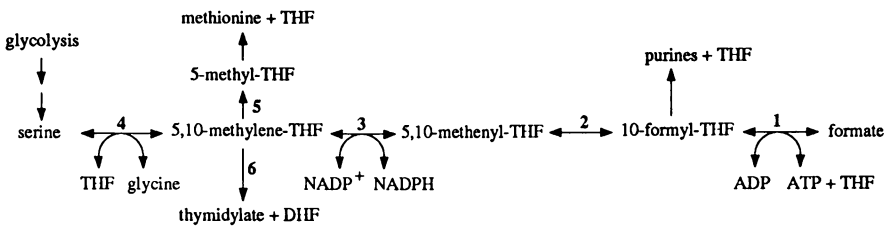


Figure 1. Tetrahydrofolate coenzymes in one-carbon metabolism. Pathways which utilize or generate one or more folate coenzymes are indicated schematically by arrows. Reactions **1**, **2** and **3**; 10-formyl-THF synthetase (EC 6.3.4.3), 5,10-methenyl-THF cyclohydrolase (EC 3.5.4.9) and 5,10-methylene-THF dehydrogenase (EC 1.5.1.5), are catalyzed by C_1-THF synthase. Other reactions shown are **4**, serine hydroxymethyltransferase (EC 2.1.2.1); **5**, 5,10-methylene-THF reductase (EC 1.5.1.20); **6**, thymidylate synthase (EC 2.1.1.45). THF, tetrahydrofolate; DHF, dihydrofolate.

dehydrogenase and cyclohydrolase activities with a small fragment (subunit M_r = 30,000) (*3-5*). Immunochemical experiments with the yeast enzyme (*6*) support the concept of two functionally independent domains. Villar *et al.* (*5*) used differential scanning calorimetry to demonstrate the existence of two domains in the rabbit enzyme. Schirch (*7*) demonstrated coordinate protection by $NADP^+$ of the rabbit liver dehydrogenase/cyclohydrolase activities against heat inactivation. 5,10-Methenyl-H_4folate, a product of the dehydrogenase reaction, does not accumulate in the coupled dehydrogenase/cyclohydrolase reaction (*7-9*). Chemical modification studies with the trifunctional enzyme (*6,10-12*) also support the concept of an overlapping active site for the dehydrogenase/cyclohydrolase activities.

These observations raise two major questions. First, what are the spatial and functional relationships between the two structural domains of C_1-THF synthase? Second, what is the structure of the overlapping dehydrogenase/cyclohydrolase active site? A more complete understanding of the enzymology of these reactions requires new experimental approaches. The amino acid sequences of several species of C_1-THF synthase have been derived from their cloned genes or complementary DNAs (*13-16*). In addition, sequences are available for several monofunctional and bifunctional versions of these enzymes (*17-20*). In the absence of a three dimensional structure, these clones provide the opportunity for a detailed analysis of catalytic mechanisms and active site architecture of this family of enzymes. This chapter describes several approaches utilizing recombinant DNA methods to address these questions.

Site-directed mutagenesis

The structure of the dehydrogenase/cyclohydrolase active site, mechanism of the coupling of these two reactions, and the identity of catalytic residues are unknown. Chemical modification experiments with the yeast cytoplasmic C_1-THF synthase indicate the existence of at least two critical histidyl residues and at least two critical cysteinyl residues at the dehydrogenase/cyclohydrolase active site (*6*). The putative dehydrogenase/cyclohydrolase domain, encompassing the first 300 or so amino acids, contains 11 histidines, but only 3 cysteines (*13*). Oligonucleotide-mediated site-directed mutagenesis was used to individually change each cysteine contained within the dehydrogenase/cyclohydrolase domain (Cys-11,Cys-144 and Cys-257) to serine (*21*). The resulting proteins were over-expressed in yeast and purified for kinetic analysis. Site-specific mutations in the dehydrogenase/cyclohydrolase domain do not affect synthetase activity, consistent with the proposed domain structure. The C144S and C257S mutations result in 7-fold and 2-fold increases, respectively, in the dehydrogenase K_m for $NADP^+$. C144S lowers the dehydrogenase maximal velocity roughly 50% while C257S has a maximal velocity similar to the wild type. Cyclohydrolase catalytic activity is reduced 20-fold by the C144S mutation but is increased 2-fold by the C257S mutation. Conversion of Cys-11 to serine has a negligible effect on dehydrogenase/cyclohydrolase activity. A double mutant, C144S/C257S, results in catalytic properties roughly multiplicative of the individual mutations. The dehydrogenase K_m for the folate substrate is not significantly changed in any of the mutants generated. These results implicate Cys-144 and Cys-257 as important active site residues for both the dehydrogenase and cyclohydrolase reac-

tions, consistent with the overlapping active site model. Cys-144 is conserved in all five NADP$^+$-dependent 5,10-methylene-H_4folate dehydrogenases for which sequence information is known (Figure 2). In contrast, the two known NAD$^+$-dependent 5,10-methylene-H_4folate dehydrogenases, which are approximately 50% identical to their NADP$^+$-dependent counterparts, have an alanine in the equivalent position. Since the major effect of the C144S mutation is on the K_m for NADP$^+$, it is tempting to speculate that Cys-144 is involved in the binding specificity of NADP$^+$ vs. NAD$^+$.

HBF	MLPATPWK	(166-173)
MBF	MLPATPWK	(172-179)
YM	FIPCTPYG	(173-180)
YC	FLPCTPKG	(141-148)
HC	FIPCTPKG	(144-151)
RC	FIPCTPKG	(144-151)
EC	..PCTPRG	(140-145)

Figure 2. Amino acid sequence conservation surrounding active site cysteine (**bold**) in NADP- vs. NAD-dependent 5,10-methylene-H_4folate dehydrogenases. HBF, human bifunctional NAD-dependent dehydrogenase/cyclohydrolase (*20*); MBF, mouse bifunctional NAD-dependent dehydrogenase/cyclohydrolase (*19*); YM, yeast mitochondrial C_1-THF synthase (*14*); YC, yeast cytoplasmic C_1-THF synthase (*13*); HC, human cytoplasmic C_1-THF synthase (*15*); rat cytoplasmic C_1-THF synthase (*16*); EC, *E. coli* bifunctional NADP-dependent dehydrogenase/cyclohydrolase (*33*). All of the C_1-THF synthases contain the NADP-dependent dehydrogenase. Boxed residues indicate identical residues or conservative substitutions.

Random mutagenesis/functional selection

Without a crystal structure of the enzyme or additional chemical modification data that might suggest important residues, picking targets for site-directed mutagenesis becomes a guessing game. In families of related enzymes, stretches of high sequence similarity often correspond to active site regions. However, the C_1-THF synthases are highly conserved across their entire sequence, making it impossible to predict active site regions. We therefore developed a method in which random mutagenesis is coupled to a functional screening to identify those mutations that affect one or more of the activities of the trifunctional enzyme (*22*). The method relies on plasmid-borne expression of genes in strains of *Saccharomyces cerevisiae* that are missing one or more of the activities of C_1-THF synthase. These specially constructed strains allow a very simple screening procedure for the detection of mutants in any of the three reactions. Specific segments of the gene or cDNA are subjected to random mutagenesis *in vitro* before expression and screening. The chemical mutagen used, nitrous acid, deaminates deoxycytosine, deoxyadenosine, and deoxyguanosine, resulting in C to T and A to G transitions (*23*). Plasmids encoding mutant enzymes are easily recovered for sequence analysis and subsequent overexpression of the mutant protein for enzymatic or structural analysis.

The screening procedure is based on the earlier observation (*24*) that the nutritional requirement of *ser1* mutants can be met by either serine, or glycine plus formate, although

growth on glycine plus formate is considerably slower. The phosphorylated pathway to serine at the phosphoserine aminotransferase reaction is blocked in *ser1* mutants. Under the glycine plus formate condition, serine is synthesized via the combined activities of C_1-THF synthase and serine hydroxymethyltransferase (reactions 1-4, Figure 1). Therefore, *ser1* mutants which are also deficient in one or more of the activities in this pathway are unable to grow on glycine plus formate.

Using this method, we have generated several mutant forms of the yeast cytoplasmic C_1-THF synthase that result in deficient synthetase and/or dehydrogenase/cyclohydrolase activities (*22,25,26*). When the synthetase domain was targeted, inactivating mutations were isolated at codons 428, 487, 492, and 565. None of these mutations affected the dehydrogenase/cyclohydrolase activities. Conversely, when the dehydrogenase/cyclohydrolase domain was targeted, inactivating mutations were isolated at codons 138, 145, 177, 188, 267, 271, and 284. The effect on cyclohydrolase activity has not been determined in these mutants. These mutations did not affect the synthetase activity. Some of the mutant enzymes harbor multiple amino acid replacements, so it will be necessary to determine which mutated residues affect activity and which are silent.

Beyond contributing information on active site residues, these redesigned enzymes have been extremely valuable in studying folate-mediated one-carbon metabolism in yeast. We have replaced the normal chromosomal gene of yeast with the mutant forms, resulting in new strains of yeast expressing full-length enzyme, but lacking one, two or all three of the C_1-THF synthase activities. These 'metabolic engineering' experiments have revealed a structural role for C_1-THF synthase in a putative multienzyme complex in *de novo* purine biosynthesis (*25*) and a new folate-dependent enzyme in yeast (*27*).

Deletion analysis

Deletion mutagenesis has been widely used to probe domain structure in multifunctional proteins. For example, many transcriptional activators have been dissected with this technique to delineate DNA binding domains, dimerization domains, and activation domains. We have used deletion mutagenesis to further define the domain structure of C_1-THF synthases from yeast and rat.

Several deletions were constructed in the yeast *ADE3* gene, encoding the cytoplasmic C_1-THF synthase, shown schematically in Figure 3 (*26*). A small internal deletion in the dehydrogenase/cyclohydrolase domain (yΔS/X; deletion of residues 114-144) resulted in a very unstable protein when expressed in yeast, with none of the three activities detectable over background. Deletion of 91 amino acids near the C-terminus of the synthetase domain (yΔCla) completely eliminated synthetase activity. Again, the protein was quite unstable and the dehydrogenase/cyclohydrolase activities were less than 10% of wild-type. Deletion of the entire synthetase domain (residues 335-946; yΔH3) results in a protein with no synthetase activity, with dehydrogenase/cyclohydrolase activity at about 10% of wild-type. Deletion of the entire dehydrogenase/cyclohydrolase domain (residues 2-328; yΔDC) resulted in a stable protein of approximately 70 kDa when expressed in yeast. This deletion completely abolished dehydrogenase/cyclohydrolase activity, as expected, but also reduced synthetase activity to less than 10% of wild-type. It is quite likely that this deletion extended through the connecting peptide into the N-

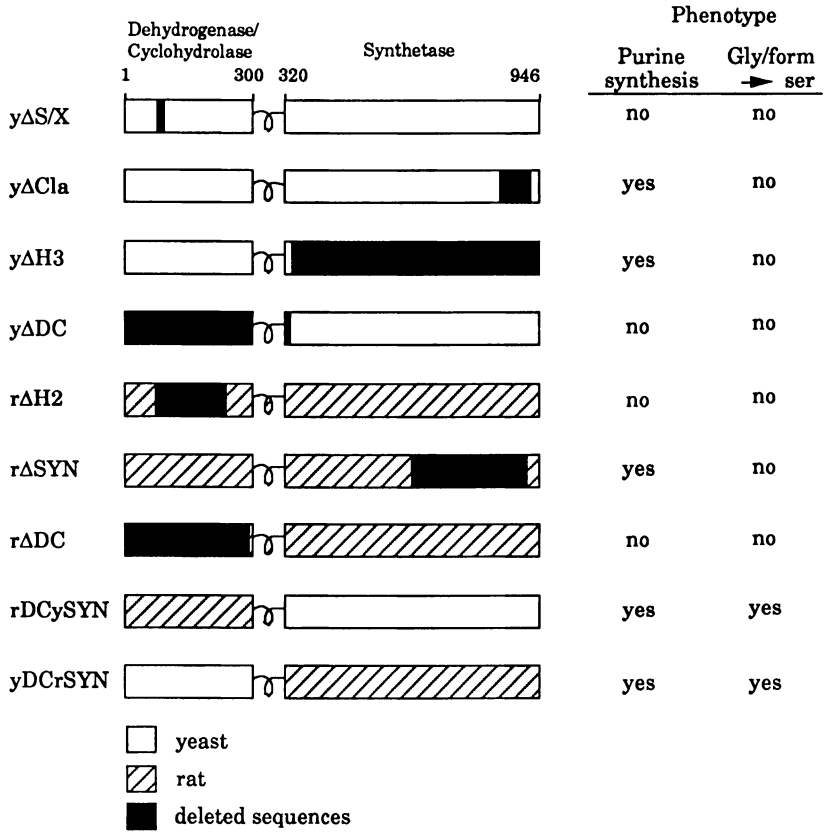


Figure 3. Growth phenotypes of *ade3⁻ ser1⁻* yeast expressing deletion mutants and domain chimeras of rat and yeast cytoplasmic C_1-THF synthases from high copy number plasmids. The two columns on the right indicate whether the construct supports the synthesis of purines or the synthesis of serine from glycine + formate. The wild-type forms of both enzymes are positive for both phenotypes. Lower case *r* and *y* refer to rat and yeast proteins, respectively. Descriptions of each construct can be found in the text.

terminus of the synthetase domain, perhaps deleting critical residues for that activity. Recently, Hum and MacKenzie (*28*) used deletion analysis to define the interdomain region of the human C_1-THF synthase to residues 292-310 (corresponding to residues 297-315 of the yeast enzyme). Furthermore, homology at the N-termini of the two bacterial monofunctional synthetases begins around position 319 of the yeast enzyme. Examination of the amino acid sequence of the yeast enzyme reveals a likely connecting region between amino acids 306-321 that contains 6 proline residues and several potential protease digestion sites.

Three deletions in the rat cytoplasmic C_1-THF synthase have recently been con-

structed. The first (rΔH2) deleted amino acids 133-272 from the dehydrogenase/cyclohydrolase domain, the second (rΔSyn) deleted amino acids 563-921 from the synthetase domain, and the third (rΔDC) deleted the entire dehydrogenase/cyclohydrolase domain (amino acids 2-277) (Figure 3). None of these constructs produced stable protein that could be detected by activity or immunoblot.

Genetic complementation assays provide a sensitive measure of *in vivo* function(s) of a protein. We previously used genetic complementation of *ade3* mutants for functional selection in our random mutagenesis experiments (see above). Complementation of a yeast strain harboring an *ade3* chromosomal deletion with these deletion constructs on high-copy-number plasmids reveals some surprising results. Several of the constructs that did not exhibit enzyme activity in crude extracts nonetheless support growth of the *ade3* deletion strain in media lacking purines. In fact, all the constructs involving synthetase deletions support growth in this strain, albeit at reduced rates. On the other hand, none of the dehydrogenase/cyclohydrolase domain deletions supported growth in the absence of exogenous adenine. We have proposed that cytoplasmic C_1-THF synthase is required as a structural component of a multienzyme complex involved in *de novo* purine biosynthesis, at least in yeast (*25*). This putative structural function can be differentiated from the enzymatic function of C_1-THF synthase by testing for growth on glycine plus formate in place of serine. None of the deletion constructs support serine synthesis from glycine plus formate, since all three activities of C_1-THF synthase are required in this pathway (Figure 1). Thus, the structural role in purine synthesis is independent of any enzymatic function of C_1-THF synthase. The deletion results suggest that the dehydrogenase/cyclohydrolase domain is the structural portion of the protein involved in critical protein-protein interactions with other participants of the putative purine multienzyme complex. The nature and composition of this complex remains to be determined.

Domain Chimeras

One drawback to deletion analysis is that the deletions often have global effects. For example, the folding, stability, or activity of one domain may be dependent on the integrity of the deleted structure. This is apparently the case for some of the constructs described above, since many of the proteins proved to be quite unstable. A more subtle strategy is to substitute whole regions or domains of a protein with sequences from other proteins. We have used this approach to study the modular nature of the C_1-THF synthase family. These enzymes share a great deal of sequence similarity. The rat and human enzymes are greater than 90% identical; the monofunctional synthetases from *Clostridium* share as high as 46% identity with the synthetase domain of the rat enzyme (*16*). It is quite likely then, that these proteins are structurally related as well, and it might be possible to generate stable chimeric enzymes in which the dehydrogenase/cyclohydrolase and synthetase domains derive from different proteins. We have recently constructed two of these domain chimeras. rDCySYN is composed of the rat dehydrogenase/cyclohydrolase domain (amino acids 1-277) fused to the yeast synthetase domain (amino acids 283-946). yDCrSYN is composed of the yeast dehydrogenase/cyclohydrolase domain (amino acids 1-334) fused to the rat synthetase domain (amino acids 331-935).

Both of these proteins have been expressed from high-copy-number plasmids in an *ade3* deletion strain. Both constructs are able to complement the purine auxotrophy of the host strain, and both support the synthesis of serine from glycine plus formate (Figure 3). The enzymatic activity of these chimeric proteins, as measured in crude extracts, is quite low, suggesting that the constructs are unstable and/or poorly expressed. Some of these extracts have been examined by immunoblotting with antisera raised against yeast or rat C_1-THF synthase and only very low levels of the full-length chimeras can be detected. Nevertheless, the *in vivo* complementation results confirm the modular nature of C_1-THF synthase and indicate that the domains derived from different sources function together efficiently enough to support growth of the yeast.

Future Directions

The work presented here represents only our initial efforts at studying the structure and functions of the multifunctional enzyme C_1-THF synthase though protein engineering. We are currently working on replacement of the $NADP^+$-dependent dehydrogenase/cyclohydrolase domain of C_1-THF synthase with a human NAD^+-dependent dehydrogenase/cyclohydrolase protein and replacement of the synthetase domain of the eukaryotic enzyme with a bacterial synthetase structure. These constructs are of interest from both the structural and metabolic points of view, and the yeast expression system provides a valuable window on *in vivo* function. However, several technical problems remain to be solved, the foremost being that of protein instability. Although most of the constructs described clearly function *in vivo*, their relative instability makes it difficult to obtain them in sufficient amounts for structural analysis. Towards this end, we are constructing a series of yeast expression vectors in which the various C_1-THF synthase sequences are fused to the C-terminus of the 76-amino acid human ubiquitin. This strategy has proven successful for the stable expression in yeast of a number of heterologous proteins (*29-32*). An improved expression system should allow the purification of enough of these proteins to carry out detailed structural studies so that we may better understand how these important multifunctional proteins function in the cell.

Acknowledgments

This work was supported by grants to D.R.A. from the National Institutes of Health (DK36913) and from the Foundation for Research.

Literature Cited

(1) Schirch, L. In *Folates and Pterins*; R. L. Blakley and S. J. Benkovic, Ed.; Wiley: New York, 1984; Vol. 1; pp 399-431.
(2) Barlowe, C. K.; Appling, D. R. *Biofactors* **1988**, *1*, 171-176.
(3) Paukert, J. L.; Williams, G. R.; Rabinowitz, J. C. *Biochem. Biophys. Res. Comm.* **1977**, *77*, 147-154.
(4) Tan, L. U. L.; Drury, E. J.; MacKenzie, R. E. *J. Biol. Chem.* **1977**, *252*, 1117-1122.
(5) Villar, E.; Schuster, B.; Peterson, D.; Schirch, V. *J. Biol. Chem.* **1985**, *260*, 2245-2252.
(6) Appling, D. R.; Rabinowitz, J. C. *Biochemistry* **1985**, *24*, 3540-3547.
(7) Schirch, L. *Arch. Biochem. Biophys.* **1978**, *189*, 283-290.

(8) Cohen, L.; MacKenzie, R. E. *Biochim. Biophys. Acta* **1978**, *522*, 311-317.
(9) Wasserman, G. F.; Benkovic, P. A.; Young, M.; Benkovic, S. J. *Biochemistry* **1983**, *22*, 1005-1013.
(10) Schirch, L.; Mooz, E. D.; Peterson, D. In *Chemistry and Biology of Pteridines*; R. L. Kisluik and G. M. Brown, Ed.; Elsevier/North-Holland: Amsterdam, 1979; pp 495-500.
(11) Smith, D. D. S.; MacKenzie, R. E. *Can. J. Biochem. Cell Biol.* **1983**, *61*, 1166-1171.
(12) Smith, D. D. S.; MacKenzie, R. E. *Biochem. Biophys. Res. Comm.* **1985**, *128*, 148-154.
(13) Staben, C.; Rabinowitz, J. C. *J. Biol. Chem.* **1986**, *261*, 4629-4637.
(14) Shannon, K. W.; Rabinowitz, J. C. *J. Biol. Chem.* **1988**, *263*, 7717-7725.
(15) Hum, D. W.; Bell, A. W.; Rozen, R.; MacKenzie, R. E. *J. Biol. Chem.* **1988**, *263*, 15946-15950.
(16) Thigpen, A. E.; West, M. G.; Appling, D. R. *J. Biol. Chem.* **1990**, *265*, 7907-7913.
(17) Whitehead, T. R.; Rabinowitz, J. C. *J. Bacteriol.* **1988**, *170*, 3255-3261.
(18) Lovell, C. R.; Przybyla, A.; Ljungdahl, L. G. *Biochemistry* **1990**, *29*, 5687-5694.
(19) Belanger, C.; MacKenzie, R. E. *J. Biol. Chem.* **1989**, *264*, 4837-4843.
(20) Peri, K. G.; Belanger, C.; MacKenzie, R. E. *Nucleic Acids Res.* **1989**, *17*, 8853.
(21) Barlowe, C. K.; Williams, M. E.; Rabinowitz, J. C.; Appling, D. R. *Biochemistry* **1989**, *28*, 2099-2106.
(22) Barlowe, C. K.; Appling, D. R. *Biofactors* **1989**, *2*, 57-63.
(23) Myers, R. M.; Lerman, L. S.; Maniatis, T. *Science* **1985**, *229*, 242-247.
(24) McKenzie, K. Q.; Jones, E. W. *Genetics* **1977**, *86*, 85-102.
(25) Barlowe, C. K.; Appling, D. R. *Mol. Cell. Biol.* **1990**, *10*, 5679-5687.
(26) Barlowe, C. K. Ph.D. Thesis, The University of Texas at Austin, 1990.
(27) Barlowe, C. K.; Appling, D. R. *Biochemistry* **1990**, *29*, 7089-7094.
(28) Hum, D. W.; MacKenzie, R. E. *Prot. Eng.* **1991**, *4*, 493-500.
(29) Ecker, D. J.; Stadel, J. M.; Butt, T. R.; Marsh, J. A.; Monia, B. P.; Powers, D. A.; Gorman, J. A.; Clark, P. E.; Warren, F.; Shatzman, A.; Crooke, S. T. *J. Biol. Chem.* **1989**, *264*, 7715-7719.
(30) Butt, T. R.; Khan, M. I.; Marsh, J.; Ecker, D. J.; Crooke, S. T. *J. Biol. Chem.* **1988**, *263*, 16364-16371.
(31) Sone, T.; McDonnell, D. P.; O'Malley, B. W.; Pike, J. W. *J. Biol. Chem.* **1990**, *265*, 21997-22003.
(32) Baker, R. T.; Varshavsky, A. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 1090-1094.
(33) D'Ari, L.; Rabinowitz, J. C. *J. Biol. Chem.* **1991**, *266*, 23953-23958.

RECEIVED March 31, 1992

Chapter 18

Properties of Native and Site-Mutagenized Cellobiohydrolase II

C. Barnett[1], L. Sumner[1,4], R. Berka[1,5], S. Shoemaker[1,6], H. Berg[2], M. Gritzali[3], and R. Brown[3]

[1]Genencor International, Inc., 180 Kimball Way, South San Francisco, CA 94080
[2]Biology Department, Memphis State University, Memphis, TN 38152
[3]Food Science and Human Nutrition Department, University of Florida, Gainesville, FL 32611

The cellulase system of *Trichoderma reesei* (also *T. longibrachiatum*) comprises endoglucanases, cellobiohydrolases and β-D-glucosidases which act synergistically to convert cellulose to glucose. Although cellobiohydrolase I (CBH I) is the most abundant component, cellobiohydrolase II (CBH II) is required for optimum rates of conversion of crystalline cellulose. The three dimensional structure of the catalytic core of CBH II recently reported by Swedish and Finnish investigators (*1*) has made it possible to interpret more precisely the results of experiments regarding structure-function relationships. Changes in activity and specificity due to site specific mutagenesis have been used to study the putative active site of CBH II to confirm the essentiality of specific amino acid residues and the effect on the substrate specificity and kinetics of reaction using model substrates. In this study, CBH II genes were expressed in *Aspergillus awamori* and the properties of the resulting enzymes were examined. The results indicate a multi-site enzyme with activity on polymeric and soluble substrates dependent on specific amino acid residues.

The cellulase system of *Trichoderma reesei* has been the subject of intense study over the past twenty years because of its ability to depolymerize cellulose, the most abundant component of plant derived materials and agricultural wastes. Advances in cellulase manufacturing making available lower cost commercial cellulases, as well as increased pressure to reduce the amount of solid waste have heightened interest in

[4]Current address: Tufts Medical School, Tufts University, Boston, MA 02215
[5]Current address: Novo Nordisk Biotech, Inc., Davis, CA 95616
[6]Current address: Department of Food Science and Technology, California Institute of Food and Agricultural Research, University of California, Davis, CA 95616

0097–6156/93/0516–0220$06.00/0

developing cellulase-based conversion processes. One such process is the development and commercialization of a cellulase-based simultaneous-saccharification fermentation process for the production of fuel ethanol. This process is based on mixed feedstocks, including renewable and waste materials, and is generating a clean burning transportation fuel. Ethanol is being used today as an octane enhancer and as a transportation fuel. In the future its use can significantly reduce our dependence on petroleum based fuels.

The cost and efficiency of cellulases remain major considerations in their use in commercial processes. Cellulases are about a hundred times less efficient than amylases and though much of the reason for this is due to differences in the substrates, it is felt that the cellulases can be improved to increase both the rate and extent of depolymerization. Examination of the cellulase system of *T. reesei* shows it to consist of at least five distinct enzymes that act synergistically to depolymerize crystalline forms of cellulose to give glucose and cellobiose as the major products. Among the five enzymes are three categories of hydrolytic activities. The endoglucanases, EG I and EG II are β-glucan glucanohydrolases (EC 3.2.1.4.) and randomly cleave internal β–1,4-glucosidic linkages in cellulose. The exocellobiohydrolases, CBH I and CBH II are β-1,4-glucan cellobiohydrolases (EC 3.2.1.91) and degrade cellulose from the ends of the cellulose polymer chains. CBH I, when acting upon cellulose, releases primarily cellobiose with some glucose; CBH II, however, is unique among the endoglucanases and cellobiohydrolases in that it releases almost entirely cellobiose from the hydrolysis of cellulose. The β-glucosidase or cellobiase (EC 3.2.1.21) converts cellobiose and other cellooligosaccharides to glucose.

The genes encoding the two endoglucanases, *egl1* (*2*) and *egl3* (*3*), and two cellobiohydrolases, *cbh1* (4) and *cbh2* (*5-6*) have been cloned and their nucleotide sequences determined. Furthermore, *cbh1* and *cbh2* have been expressed in *Saccharomyces* (*7*) and *cbh1* and *egl1* have been expressed in *Aspergillus nidulans* (*8*). Although well characterized at the genetic and biochemical level, the active site residues and their respective roles have yet to be determined for any of the cellulase enzymes. Chen et al. (*5*) suggested a model for CBH II based on homology to the catalytic mechanism of T4 lysozyme (*9*) where residues equivalent to the aspartic acid at position 175 and the glutamic acid at position 184 are involved in catalysis. At about the same time, Knowles et al. (*6*) suggested that the glutamic acid at position 244 is implicated in catalysis. More recently, the three-dimensional structure of the catalytic core of CBH II was reported (*1*). This study strongly implicated two aspartyl residues (Asp 175 and Asp 221) in the catalytic mechanism as well as several tryptophanyl residues in an extensive binding site. The structure proposed a "tunnel" into which a β-1,4-glucan chain could enter with the consequent addition of H_2O to alternate glycosidic bonds and the release of α-cellobiose as product molecules.

To test the validity of these hypotheses and to identify the active site residues in CBH II by site-directed mutagenesis, we have developed an expression system combining *Aspergillus awamori*, a fungus with a relatively low background for cellulolytic activity and no cellobiohydrolase, and the expression vector pGPT-*pyrG* (*10*). Using this heterologous expression system, we have identified one residue essential to the catalytic mechanism of CBH II and a second residue which seems to affect the catalytic activity of CBH II. Neither of these alterations of the CBH II

protein resulted in a loss of binding to cellulose in cell walls, thus indicating the significance of the cellulose binding domain (CBD) in the substrate:enzyme association.

Materials and Methods

Fungal Strains. *Aspergillus awamori* strain (*A. niger* var. *awamori*) GC12 was derived from strain UVK143 (a glucoamylase hyper-producing mutant of strain NRRL 3112) by parasexual crossing of the following auxotrophic mutants: *A. awamori* GC5 (*pyrG5*), a uridine requiring auxotroph isolated by selection on 5-fluoro-orotic acid (*11*) following mutagenesis of UVK143f with ultraviolet light (this mutant is deficient in the enzyme orotidine 5′-monophosphate decarboxylase); *A. awamori* GC3 (*argB3*), which is an arginine requiring auxotroph isolated by filtration enrichment (*12*) following nitrosoguanidine mutagenesis of UVK143f (this mutant is specifically deficient in the enzyme ornithine carbamoyl transferase). The resultant double auxotroph (*pyrG*, *argB*) was designated strain GC12.

Bacterial Strains, Cloning Vectors and Plasmids. *Escherichia coli* JM101 (*13*) was used for propagation of all plasmids. The cloning vectors pUC218 and pUC219 are chimeric DNA phage-plasmid molecules derived from pUC18 and pUC19, respectively (*14*), with the insertion of the restriction sites *Xho*I, *Bgl*II, and *Cla*I in the polylinker between the *Bam*HI and the *Xba*I sites. These vectors contain the intergenic region of the bacteriophage M13 (*15*) and when used in conjunction with the helper phage M13/K07, generate single stranded DNA for use as template for site directed oligonucleotide mutagenesis (*16*) and for sequencing.

Construction of pGPT-cbh2. In order to create sites to easily remove the coding region of *cbh2*, oligonucleotide mutagenesis primers 30 base pairs in length were synthesized to insert a *Bgl*II site 24 nucleotides upstream of the methionine at the translation start site and an *Nhe*I site 21 nucleotides downstream of the stop codon in both the intronless and genomic copies of the *cbh2* gene (Figure 1).

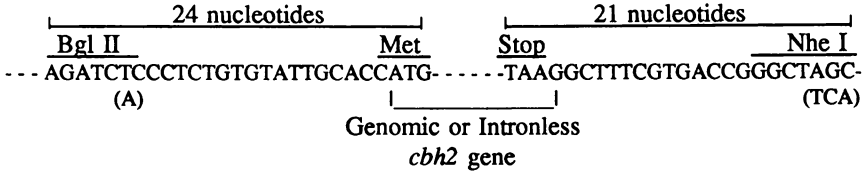


Figure 1
Nucleotide Sequence of 5′ and 3′ *cbh2* Regions

These sites were used to ligate the full length coding regions of both forms of *cbh2* into the *Bgl*II and *Xba*I sites of the expression vector pGPT-*pyrG* (*17*). This vector uses the *A. awamori* glucoamylase (*glaA*) promoter and the *A. niger glaA* terminator

as transcriptional and translational controls for the expression of CBH II. Also included on the vector is the *A. nidulans pyrG*, which complements the uridine auxotrophy of our GC12 strain and serves as a selectable marker for identification of transformants on solid media containing no uridine. The entire promoter-coding region-terminator cassette can be liberated from vector sequences by restriction digestion with *Cla*I to yield a 5.0 kb fragment.

Transformation Procedure for *A. awamori*. *A. awamori* strain GC12 was protoplasted as described (*18*) and transformed by an electroporation technique described by Ward et al. (*19*). The number of transformants ranged from 5-15 per microgram of DNA.

Culture conditions. *A. awamori* CBH II transformants were grown in liquid medium containing 1 g/L Bacto Peptone (Difco), 20 g/L malt extract (Difco), 1 g/L yeast extract (Difco), 6 g/L $NaNO_3$, 0.52 g/L KC1, 34 g/L KH2PO_4, 1.0 g/L $MgSO_4 \cdot 7H_20$, 50 g/L maltose, 1 g/L trace elements (1.0 g/L, $FeSO_4 \cdot 7H_20$, 8.8 g/L $ZnSO_4 \cdot 7H_2O$, 0.4 g/L $CuSO_4 \cdot 5H_20$ 0.15 g/L $MnSO_4 \cdot 4H_20$, 0.1 g/L $Na_2B_40_7 \cdot 10H_20$, 50 mg/L $[NH_4]_6Mo_70_{24} \cdot 4H_20$), 10 mL/L met-bio solution (50 g/L L-methionine, 200 mg/mL d-biotin), 0.1% Tween 80, 50 mg/mL streptomycin. All cultures were inoculated with conidia to a final concentration of $1x10^6$ conidia/mL of culture medium and grown for 4 days at 37°C on a rotary shaker (New Brunswick Scientific Company, Inc.) at 200 rpm. Cultures were filtered through miracloth and filtrates collected for enzyme characterization.

Construction of CBH II Expression Vectors. Native and site-specific variants of CBH II in pJC218 were excised with *Bgl*II and *Nhe*I and ligated into *Bgl*II/*Xba*I cut pGPT-*pyrG* vector. These were used to transform *E. coli* JM101 and selected for on medium containing 50 mg/mL carbenicillin.

Isolation and Characterization of Nucleic Acids. *A. awamori* DNA and RNA were isolated as described previously (*20*). DNA samples from transformants were digested with an appropriate restriction enzyme, fractionated on 0.5% agarose gels and blotted to Nytran (Schleicher & Schuell, Keene, NH) nylon membrane. The membranes were hybridized in 50% formamide, 5X SSPE, 200 mg/mL sheared and denatured (95°C) salmon sperm DNA and with nick translated pGPT-*cbh2* plasmid ($1x10^6$ cpm/mL). After overnight incubation at 42°C, the membranes were washed at 55°C in 2X SSC, as described by Davis et al. (*21*) Total RNA from selected *A. awamori* transformants was fractionated electrophoretically on 1% agarose gels containing 2% formaldehyde and subsequently blotted to Nytran in 20X SSPE. The membranes were hybridized with a nick translated *cbh*2 coding region fragment. Hybridization and washing conditions were the same as described above for DNA hybridization.

Site Specific Mutagenesis. Oligonucleotide-directed site specific mutagenesis was accomplished according to the method described by Carter (*16*). The precise nucleotide sequence of the genetic modifications were confirmed using the Sanger di-deoxy sequencing method (*22*).

Antibody Analysis. Screening transformants for the presence of recombinant CBH II was done using an enzyme linked immunosorbent assay. (Shoemaker and Sumner, unpublished data.) Western blot analysis was carried out as described by Towbin et al. (*23*).

Enzyme Purification and Analysis of rCBH II Enzymes. One liter cultures of the *A. awamori* transformants (cDNA type) which expressed CBH II enzymes were concentrated twenty fold and diafiltered into 5 mM phosphate buffer at pH = 7.8. The rCBH II protein was purified by FPLC (Pharmacia) using a Mono Q anion exchange column equilibrated in 5 mM phosphate buffer at pH = 7.8. Flow-through fractions containing rCBH II were collected and were incubated in 1% phosphoric acid-swollen cellulose (PSC) for 1 hour at 45°C. The samples were centrifuged at 2500×g to remove PSC. Production of reducing sugars, expressed as cellobiose equivalents, was determined according to the method of Nelson and Somogyi (*24-25*). The designation of the CBH II enzymes used in the kinetic and binding studies is given in Table I.

Table I. Recombinant CBH II (rCBH II) Enzymes[a]

Enzyme	Designation
rCBH II (from native cDNA)	rCBH II
rCBH II (Glu 184 to Gln 184)	E184Q
rCBH II (Asp 173, 175 to Asn 173, 175)	D173N/D175N
rCBH II (Asp 173 to Asn 173)	D173N
rCBH II (Asp 175 to Asn 175)	D175N
rCBH II (Glu 244 to Gln 244)	E244Q

[a]rCBH II refers to the *Aspergillus* gene product which is derived from native and genetically modified *Trichoderma* CBH II genes. CBH II purified directly from *T. reesei* cultures is designated as CBH II or native CBH II.

Reaction Product Analysis. Reaction supernatants were analyzed by HPLC using a Bio-Rad HPX65A column at 80°C. Glucose and cellobiose (Sigma) were used as standards.

Kinetic Studies. The enzymatic hydrolysis of Avicel or PSC was carried out in 50 mM pH 5.0 sodium acetate buffer at 40°C after which the reaction was stopped by exposure to boiling water for 5 minutes. Following centrifugation of the incubation mixtures, the supernatants were analyzed for reducing sugar by the Nelson-Somogyi method (*24-25*). The model substrate, methylumbelliferyl-β-D-cellobioside (MUC), was used in studies of specificity and kinetics of the purified native and recombinant forms of CBH II. These studies were carried out at 40°C in 50 mM pH 5.0 sodium acetate buffer and the extent of reaction determined by measuring absorbance at 346 nm (*26*).

Ultrastructure Analysis. The enzyme-linked colloidal gold labelling technique of Berg et al. (*27*) without poststaining was used to compare the binding affinities of native and recombinant CBH II forms (native—purified, recombinant—wild type; D173N/D175N; and E184Q) with identical sections of *Casuarina* branchlet parenchyma cell wall material. The same ratio of protein to gold was used to make each of the probes, which were then diluted similarly to give working solutions according to standard procedures (*27*).

Results and Discussion

Expression of CBH II in *A. awamori*. Intronless and genomic forms of rCBH II and the specific genetic variants D173N/D175N, E184Q, D173N, D175N, and E224Q were introduced into the expression vector pGPT-*pyrG* as outlined in the previous section. These plasmids were used to transform *A. awamori* GC12. All transformants were picked to fresh minimal medium in the absence of uridine twice successively and inoculated into the medium described in the previous section. The addition of maltose as a carbon source was necessary to induce transcription from the *glaA* promoter. Duplicate samples of ten transformants, from both the genomic and intronless constructions were analyzed by an immunochemical method (ELISA). Approximately 70-80% of the transformants tested showed the presence of secreted CBH II. The maximum expression among the genomic transformants was determined to be 25 mg/mL and the intronless transformants reached a maximum expression level of 40-50 mg/mL (data not shown).

Total chromosomal DNA was isolated from selected high producing transformants and from selected low producing transformants. Equal amounts of DNA from each transformant were cut with the restriction endonuclease *Cla*I, fractionated on 0.5% agarose gels, blotted to Nytran and hybridized with the nick translated pGPT-CBH II. The autoradiogram given in Figure 2 shows that in each selected transformant, pGPT-*cbh2* has integrated into the host genome. In this figure the letter C denotes intronless clones and the letter G denotes genomic clones. The end lanes (p18CB and p1G1) are purified controls cut with *Cla*I, a restriction enzyme that separates the GAM-CBH II expression unit from the vector DNA. The lane denoted Uc is *A. awamori* untransformed DNA as a negative control. Five micrograms of genomic DNA isolated from various transformants and a negative control were run on 0.5% agarose, blotted and probed with nick translated genomic pGPT-*cbh2* plasmid.

The high producing intronless transformants, cI and cJ, contain full length tandem integrants as seen by the restriction pattern consistent with the plasmid control. They also show different copy numbers which is consistent with circular plasmid integration into the chromosome. The intronless transformants with undetectable levels of expression of CBH II show a restriction pattern that is markedly different from the control plasmid. This is due to an uneven recombination event resulting in the loss of either the promoter portion of the coding region and resulting in undetectable levels of CBH II expression. The high producing genomic transformants, gC and gW, do not show a restriction pattern that is consistent with a full length tandem integration as the restriction pattern does not contain the same size fragments as the control.

This is the result of uneven recombination into the chromosome while maintaining the integrity of the expression cassette.

Northern blot analysis of the high producing transformants cI, gC and gW, probed with the *BglII-NheI* coding region fragment is given in Figure 3. In this case, total *A. awamori* RNA was subjected to electrophoresis on a 3% agarose/formaldehyde gel (panel A). The gel was transferred to a nylon membrane filter (Nytran R). The membrane was probed with ^{32}P labelled intronless *BglII-NheI* fragment and visualized by autoradiogram (panel B). Lanes Gc, Gw and Cj are individual transformants (G is genomic and C is intronless). Lane Uc is an untransformed control and M is a Bethesda Research Labs (BRL) RNA ladder. Northern analysis shows that ample mRNA is being produced from the *glaA* promoter in both the intronless and the genomic CBH II producing transformants.

Analysis of rCBH II. Analysis of culture supernatants of selected transformants by SDS polyacrylamide gel electrophoresis (SDS-PAGE) with silver staining (Figure 4A) and western blot analysis (Figure 4B) show rCBH II as a broad heterogeneous band and of a lower mobility, compared to native CBH II. The lanes in both gels designated Cqf, Gqb, Cqa, Cg, Gb, Cf and Ga are individual transformants. The lane designated MW are protein molecular weight markers. The CBH II lane is purified CBH II from *T. reesei* and the Uc is untransformed *A. awamori* supernate as a negative control. This is indicative of hyperglycosylation by the host, *A. awamori*. The purified rCBH II gave a similar pattern upon gel analysis (data not shown). Reaction products analyzed by HPLC confirmed that the reaction product is greater than 95% cellobiose (data not shown).

The FPLC procedure allowed the concentration and purification of native and recombinant forms of CBH II. The enzymatic properties of these enzymes were compared using phosphoric acid-swollen cellulose (PSC) as substrate. The reaction products were analyzed by HPLC and confirmed as cellobiose (>95%). A comparison of specific activities using either PSC or Avicel (0.1% w/v) is shown in Table II.

Table II. Specific Activities of CBH II Enzymes

	Specific Activity (mmol glc equiv/min/mg protein)	
CBH II Enzyme[a]	PSC	Avicel
Native CBH II	4.9	0.13
Recombinant CBH II		
rCBH II-1	3.9	0.09
rCBH II-2	5.4	0.14
E184Q	5.9	0.17
D173N/D175N	0.19	<0.01
D173N	5.9	0.15
D175N	<0.05	<0.01
E244Q	8.8[b]	0.13

[a]rCBH IIs were expressed in *Aspergillus* and purified from culture filtrates using immunosorption columns. [b]Not yet confirmed in cultures of the same transformant.

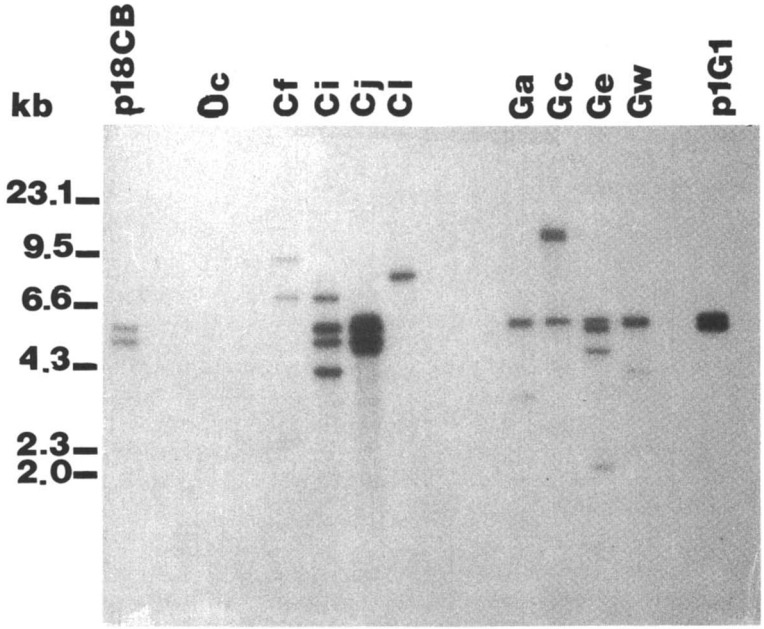


Figure 2. Southern blot analysis of selected CBH II transformants.

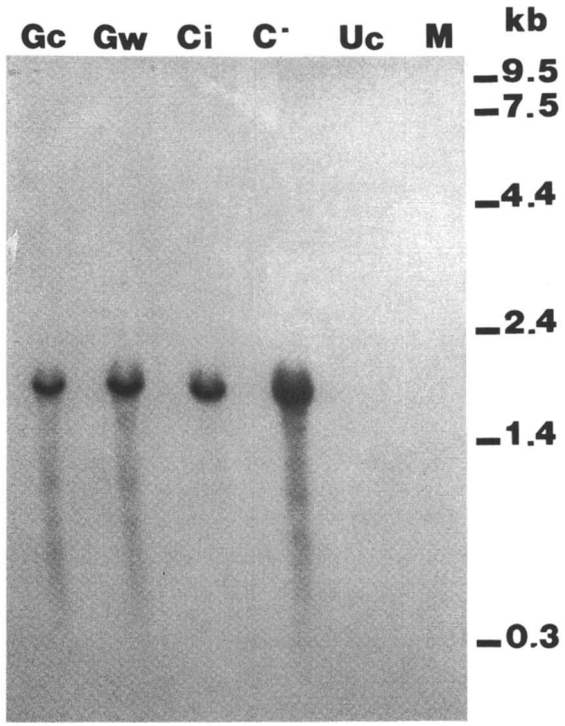


Figure 3. Northern analysis of CBH II transformants.

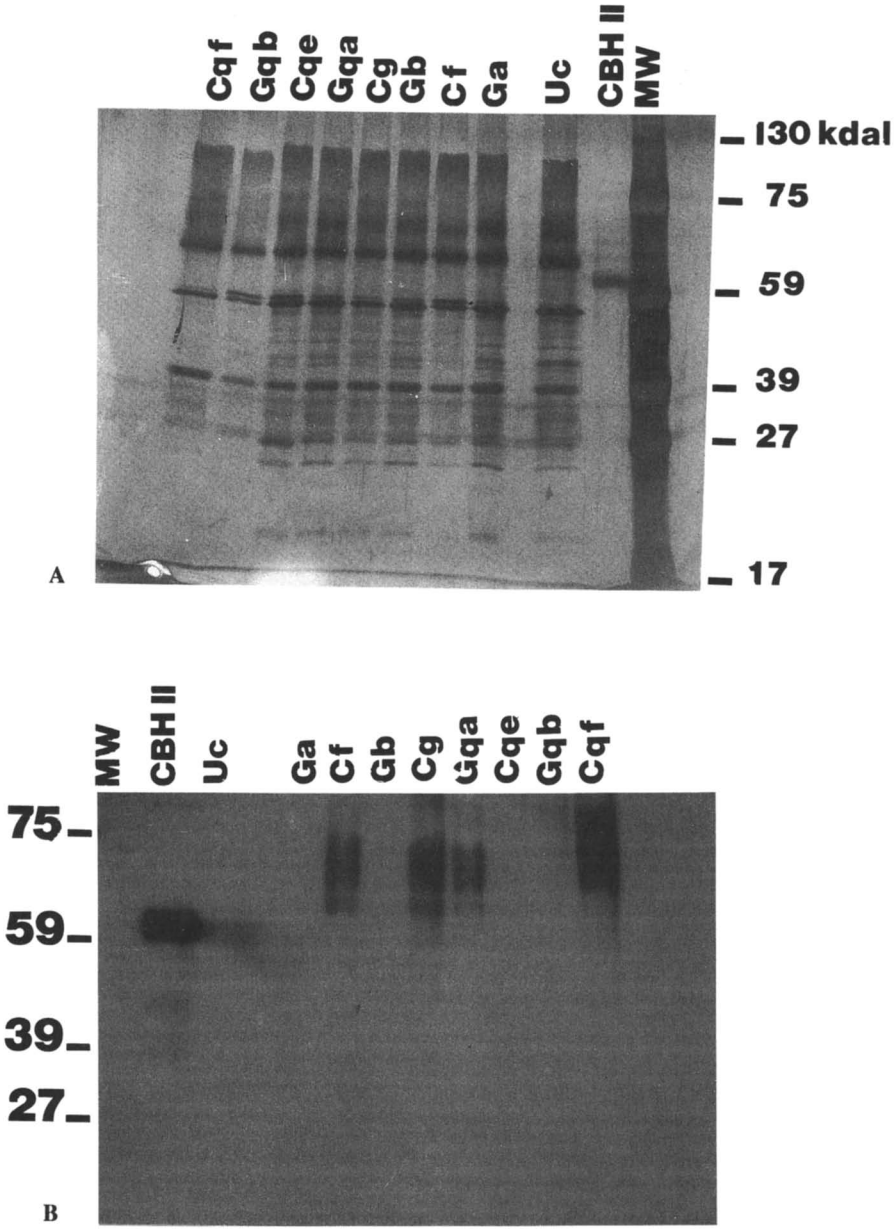


Figure 4. Polyacrylamide gel analysis of CBH II transformants. Panel A shows an SDS-PAGE with silver staining. Panel B shows western blot analysis after blotting gel onto nitrocellulose, incubating with antibody specific for CBH II, and visualizing by ^{125}I protein A.

Catalytic Specificity. Using methylumbelliferyl-β-D-cellobioside (MUC) as a model substrate the rCBH II forms were analyzed for catalytic activity. As shown in Table IIIall recombinant forms, except for those with the sequence for native CBH II, had the capacity to hydrolyze MUC.

Table III. Specific Activity of Purified Enzymes using 0.8 mM Methylumbelliferyl β-D-Cellobioside as Substrate

Enzyme	Specific Activity (mmol/min/mg protein)
Native	
CBH I	0.086
CBH II	<0.001
Recombinant CBH II	
rCBH II	<0.001
D173N	0.011
D175N	0.014
D173N/D175N	0.010
E184Q	<0.007

In experiments with CBH I and the double mutant D173N/D175N, the kinetic parameters for MUC hydrolysis were determined (Lineweaver-Burk data analysis) and compared. Whereas the native CBH II has no activity against MUC, the double mutant of CBH II has an enzyme-concentration-normalized V_{max} comparable to that of CBH I. (V_{max} = 0.04 μmol min^{-1} mg^{-1} for D173N/D175N, compared with V_{max} = 0.09 μmol min^{-1} mg^{-1} for CBH I). It appears that very small perturbations in the structure near the active site of CBH II permit binding and cleavage of MUC. It is noteworthy that the K_m value for the double mutant (2.27 x 10^{-3} M) is 50 times higher than that of CBH I (4.7 x 10^{-5} M); this suggests an intrinsically higher affinity for such substrates in the CBH I active site.

Ultrastructure Studies. The enzyme-gold affinity technique (*27*) was used with *Casuarina* branchlet parenchyma cell walls to examine specific binding of CBH II enzymes (Figure 5). The rCBH II enzymes used in this study were native (panel A), D173N/D175N (panel B), E184Q (panel C) and E244Q (panel D). The observations from many fields and several degrees of magnification, including that shown in Figure 5, reveal no significant differences in the observed density and the distribution of gold particles on the cellulose fibrils of the wall material. The evidence for this observed specificity for the cellulose component, compared to other components of the cell wall, has been described previously (*27*).

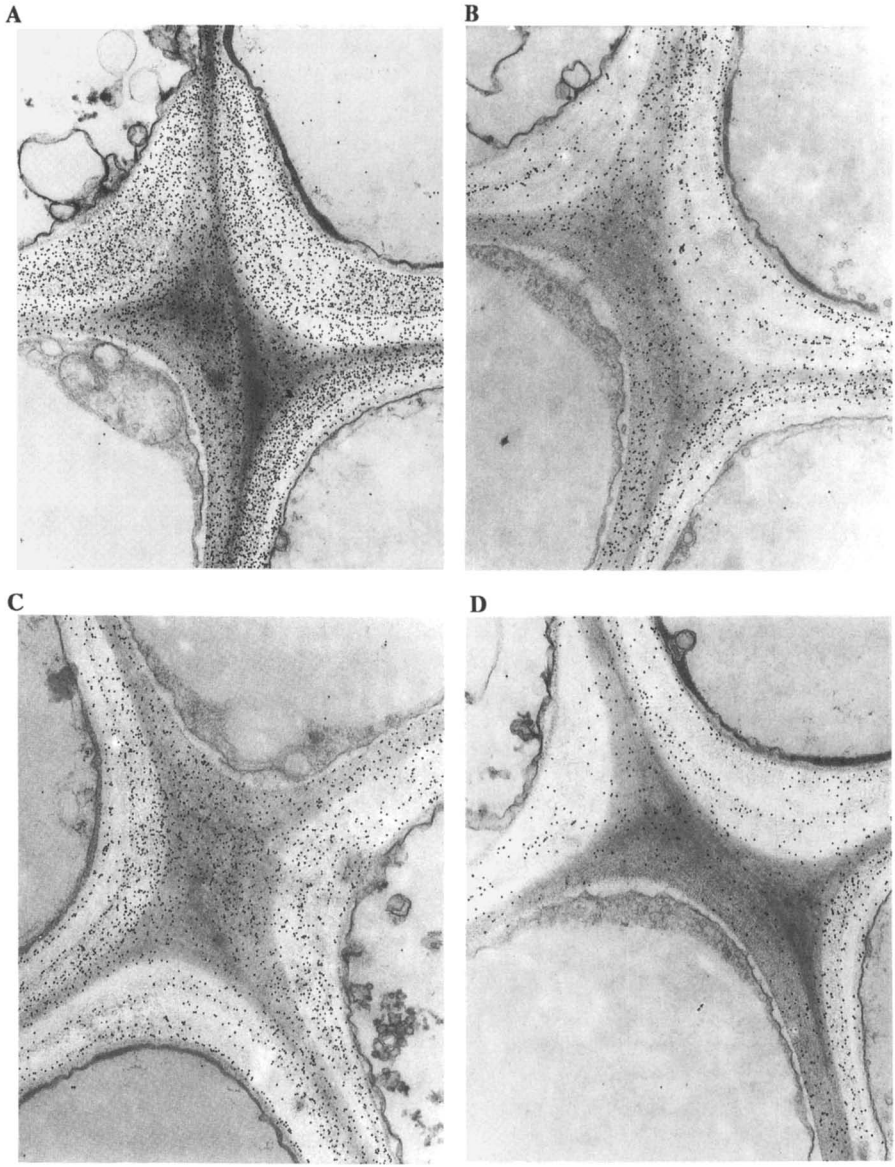


Figure 5. Unstained electron micrograph of CBH II enzyme-linked colloidal gold labeling of *Casuarina* branchlet parenchyma cell wall material. Magnification is 20,000x.

Conclusions

The different effects of the genetic modifications of CBH II enzymes on rate of PSC hydrolysis, as compared to Avicel hydrolysis, reveal that the most significant change is in the rate of Avicel hydrolysis. This seems to indicate a greater dependence on native CBH II structure for the hydrolytic cleavage of the crystalline substrate than for PSC hydrolysis.

In agreement with a previous report from other laboratories (*28*), we observe no activity of native CBH II on methylumbelliferyl-β-D-cellobioside (MUC). With respect to the present studies using MUC and rCBH II enzymes, the D173N, D175N, and D173N/D175N mutants probably allow greater accessibility and greater freedom of substrate binding, transition state formation or product release for small substrates, such as MUC. This in turn may reflect a loss of fidelity with respect to the mechanism of the native depolymerase, which is most clearly exhibited in the dramatic effect of genetic modification on the hydrolytic cleavage of Avicel. The association with the cellulose fibrils, as evidenced by the ultrastructural studies, indicate that the loss of activity is not due to an inability of the cellulose binding domain to provide proximity of enzyme and substrate.

Detailed interpretation of the structural basis for the observed changes in activity and specificity will depend on the resolution of the structure of the active site region either for intact rCBH II forms or their respective catalytic domains.

Acknowledgments

The authors would like to gratefully acknowledge support from The National Renewable Energy Laboratory (NREL; formally SERI; Subcontract HK-7-07122), Genencor International and from the GRI/IFAS co-funded research program. We also greatly appreciate the expert technical assistance of Katalin Vienne in the kinetic studies.

Literature Cited

1. Rouvinen, J.; Bergfors, T.; Teeri, T.; Knowles, J. K. C.; Jones, T. A. *Science* **1990,** *249,* 380-386.
2. Penttila, M.; Lehtovaara, P.; Nevalainen, H.; Bhikhabhai, R.; Knowles, J. *Gene* **1986,** *45,* 253-263.
3. Saloheimo, M.; Lehtovaara, P.; Penttila, M.; Teeri, T. T.; Stahlberg, J.; Johansson, G.; Pettersson, G.; Claeyssens, M.; Tomme, P.; Knowles, J. K. C. *Gene* **1988,** *63,* 11-21.
4. Shoemaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis, M. *Bio/Technol.* **1983,** *1(8),* 687-690.
5. Chen, C. M.; Gritzali, M.; Stafford, D. W. *Bio/Technol.* **1987,** *5,* 274-278.

6. Teeri, T. T.; Lehtovaara, P.; Kauppinen, S.; Salovuori, I.; Knowles, J. *Gene* **1987,** *51,* 43-52.
7. Penttila, M. E.; Andre, L.; Lehtovaara, P.; Bailey, M.; Teeri, T. T.; Knowles, J. K. C. *Gene* **1988,** *63,* 103-112.
8. Barnett, C. C.; Berka, R. M.; Shoemaker, S. P.; Sumner, L. M.; Ward, M.; Wilson, L. J. *Abstract, 14th Fungal Biology Conference.* **1987.**
9. Paice, M. G.; Jurasek, L. *Adv. Chem. Series.* **1979,** *181,* 361-374.
10. Berka, R. M.; Barnett, C. C. *Biotech. Adv.* **1989,** *7,* 127-154.
11. van Hartingsveldt, W.; Mattern, I. K.; van Zeijl, C. M. J.; Pouwels, P. H.; van den Hondel, C. A. M. J. J. *Mol. Gen. Genet.* **1987,** *206,* 71-75.
12. Buxton, F. P.; Gwynne, D. I.; Davis, R. W. *Gene.* **1985,** *37,* 207-214.
13. Messing, J.; Crea, R.; Seeburg, P. H. *Nucleic Acids Res.* **1981,** *9,* 309-321.
14. Yanisch-Peron, C.; Viera, J.; Messing, J. *Gene.* **1985,** *33,* 103-119.
15. Mead, D. A.; Kemper, B. In *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*; Rodriquez, Denhardt, D. T., Eds.; Butterworth Publishers: Stoneham, MA, 1987; pp. 85-102.
16. Carter, P. *Methods in Enzymology* **1987,** *154,* 382-403.
17. Berka, R.; Ward, M.; Thompson, S.; Fong, K.; Wilson, L.; Lamsa, M.; Gray, G. In *Advances in Gene Technology, Protein Engineering and Production*; Brew, K., Ed.; IRL Press: Washington, DC, 1988.
18. Cullen, D.; Gray, G. L.; Wilson, L. J.; Hayenga, K. J.; Lamsa, M. H.; Rey, M. W.; Norton, S.; Berka, R. M. *Bio/Technol.* **1987,** *5,* 369-376.
19. Ward, M; Wilson, L. J.; Carmona, C.; Turner, G. *Current Genetics* **1988,** *13,* 37-42.
20. Timberlake, W. E.; Barnard, E. C. *Cell* **1981,** *26,* 29-37.
21. Davis, L. G.; Dibner, M. D.; Battey, J. F. *Basic Methods in Molecular Biology;* Elsevier: New York, NY, 1986; pp. 84-87.
22. Sanger, F.; Niclen, S.; Coulson, A. R. *Proc. Nat. Acad. Sci. USA.* **1977,** *74,* 5463-5467.
23. Towbin, H.; Staehlin, T.; Gordon, J. *Proc. Nat. Acad. Sci. USA.* **1979,** *76,* 4350-4354.
24. Nelson, N. J. *J. Biol. Chem.* **1944,** *153,* 375-380.
25. Somogyi, M. J. *J. Biol. Chem.* **1952,** *195,* 19-23.
26. Rosenthal, A. L.; Saifer, A. *Anal. Biochem.* **1973,** *55,* 85-92.
27. Berg, R. H.; Erdos, G. W.; Gritzali, M.; Brown, Jr., R. D. *J. Electron Microscopy Tech.* **1988,** *8,* 371-379.
28. Claeyssens, M.; Van Tilbeurgh, H.; Tomme, P. Wood, T. M.; McRae, S. *Biochem. J.* **1989,** 261, 819-825.

RECEIVED September 16, 1992

Chapter 19

Recombinant β-Glucosidase of *Trichoderma reesei*

Tim Fowler

Genencor International, Inc., 180 Kimball Way, South San Francisco, CA 94080

Recent advances in the genetic manipulation of *T. reesei* provide a molecular tool with which the cellulase enzyme composition can be altered to provide new and useful mixtures. In addition, mutagenesis experiments can be designed to answer direct questions regarding cellulase gene regulation and mechanism of enzyme action.

ß-glucosidase is one member of a family of cellulase enzymes that act synergistically to hydrolyze cellulose to glucose. Studies in our laboratory have focused on the role of extracellular ß-glucosidase in the regulation of the other enzymes of the cellulase complex of *T. reesei* and its function in the production of glucose from cellulose. To this end we have cloned the gene encoding the extracellular ß-glucosidase (*bgl1*) from *Trichoderma reesei*. (*1*). Re-introduction of the *bgl1* gene back into the host in extra copies gives increased expression of ß-glucosidase and results in a cellulase complex that has an increased rate of glucose production from cellulosic substrates. Results of *bgl1* gene disruption experiments and proposed site directed mutagenesis of the enzyme are described.

In this chapter I will describe the cellulase complex of *Trichoderma reesei* and in particular the genetic manipulation of this complex to produce novel strains in which the cellulase complex profile and activity has been altered. I will focus on the extracellular ß-glucosidase enzyme encoded by the *bgl1* gene.

Historically, filamentous fungi have been used to produce many enzymes, antibiotics and other biochemical products. However, methods for DNA-mediated transformation of industrially important fungi only became generally available in the mid 80s. This achievement provided a viable alternative to classical genetics for rapid improvement of existing fungal strains which produce salable commodities. The industrially important deuteromycete, *Trichoderma reesei*, is an example of such a fungal species. *T. reesei* secretes large quantities of a hydrolytic mixture of enzymes that act in concert to degrade crystalline cellulose to glucose. These cellulases are used in a variety of applications; examples include the extraction of fruit and vegetable juices, animal feed and silage processing, and malting and brewing. In addition, the

0097–6156/93/0516–0233$06.00/0

introduction of genetic engineering techniques to *Trichoderma* is leading to rapid advances in our understanding and manipulation of the cellulase complex.

The cellulase enzyme complex of *T. reesei* consists of two known cellobiohydrolases, CBHI and CBHII (EC 3.2.1.91), at least two endoglucanases, EGI and EGII (EC 3.2.1.4), and at least one ß-glucosidase (EC 3.2.1.21). The endoglucanases and cellobiohydrolases are believed to act synergistically to hydrolyze crystalline cellulose to small cello-oligosaccharides (mainly cellobiose). The small oligo-dextrins are subsequently hydrolyzed to glucose by ß-glucosidase. The exact biochemistry, regulation and synergy between the different cellulases of *T. reesei* are the subject of a great deal of investigation.

Wild type strains of *T. reesei* produce extracellular cellulase at levels of a few grams per liter and in molar ratios of about CBHI (60) : CBHII (20) : EGI (10) : ß-glucosidase (1) (*5,6*). In the biotechnology industry, selection of *T. reesei* production strains by mutagenesis and screening have resulted in strains capable of secreting in excess of 40 g/L (*38,39*). The cellulase products from these strains can be further modified by blending with other cellulase preparations (sometimes enriched for specific components). In addition, variation in fermentation conditions enables a measure of control over the ratios of secreted enzymes. However, these methods tend to be expensive and/or labor intensive. Methods for genetic manipulation of *T. reesei* became possible with the introduction of both dominant and auxotrophic markers (*2,3,4*) and techniques such as electroporation and protoplast fusion for introducing the foreign DNA into the fungal cells. These techniques are now being employed to manipulate the cellulase complex at the genetic level. Most recently, genetic engineering of *T. reesei* strains enables the complete removal or overproduction of either individual or multiple components of the cellulase complex. This means novel strains with cellulase profiles specifically tailored to defined applications can now be made (See Table I from ref. 7).

The role of the ß-glucosidase component in cellulose hydrolysis and in the regulation of the cellulase complex is currently under investigation at Genencor International. Presumably one function of ß-glucosidase is the breakdown of the cellobiose produced by the cellobiohydrolases and endoglucanases to provide glucose as a carbon source. Another function of ß-glucosidase may be to use glucose (via a transglycosylation reaction) to produce oligosaccharides that have been shown to act as potent inducers of the cellulase complex (*5,41,42*).

Cell fractionation, immunoflourescence and electron microscopic localization indicate that the majority of ß-glucosidase is associated with the outer integuments of *T. reesei* (*37,43,44,45,46*). It therefore makes teleological sense that for the organism to make more efficient use of available cellobiose, the major portion of the detectable ß-glucosidase activity remains associated with the cell wall (*8,9,10*). It is believed that the association of ß-glucosidase with the cell wall is a significant factor in the reduced ability of commercial preparations of cellulase to produce glucose. Improvement of the cellulase preparations by the addition of purified ß-glucosidase (*11,12,13*) or isolation of mutant strains of *T. reesei* that have increased levels of ß-glucosidase are possible solutions to this problem (*40*).

We have begun to investigate the role of the ß-glucosidase enzyme in cellulase system at the molecular level by cloning and sequencing of the extracellular ß-glucosidase gene, *bgl1*, from *T. reesei* (*1*). We demonstrate that transformation of the *bgl1* gene into the *T. reesei* genome in multiple copies can be used to generate strains with significant increase in extracellular ß-glucosidase activity. Finally site directed mutants of the *bgl1* gene have been created to locate residues central to catalytic activity. The long term goal of these experiments is to identify mutant forms of the ß-glucosidase enzyme that possess novel and useful enzymatic properties. The deletion

of the *bgl1* gene from *T. reesei* and the use of this strain as a host for the site directed mutants of ß-glucosidase are also described.

Isolation and Overexpression of the *bgl1* Gene in *T. reesei.*

We have previously described the cloning and sequencing of the gene encoding the extracellular ß-glucosidase gene (*bgl1*) from *T. reesei* (*1*). The primary structure of the deduced extracellular ß-glucosidase protein has several properties of note. The predicted molecular weight of the encoded ß-glucosidase protein is 74,341 daltons. The size and composition is in good agreement with the protein purified by Chirico and Brown (*14*). A 31 amino acid secretion signal peptide precedes the mature amino terminus of ß-glucosidase as deduced from the amino terminal peptide sequence and homologies to the other cellulase genes (*15, 16, 17, 18, 19*). The primary amino acid sequence of ß-glucosidase shows 7 potential N-linked glycosylation sites (*20*). The *bgl1* coding region is interrupted by two putative introns of 70 and 64 base pairs that show homology to the consensus splice signals emerging for *T. reesei* and other filamentous fungi (*21*).

The following sections describe experiments in which the cellulolytic capacity of *T. reesei* strains was altered by transformation with extra copies of the *bgl1* gene, resulting in the overexpression of extracellular ß-glucosidase. Alternatively, transformation with a vector designed to disrupt the *bgl1* gene resulted in a strain with no apparent extracellular ß-glucosidase activity.

The ability to increase expression of interesting and/or commercially valuable proteins by genetic engineering has recently become possible in filamentous fungi. As a general rule increasing the copy number of a fungal gene through transformation usually leads to increased expression of that gene. Recent examples of this phenomenon include glucose oxidase (*goxA*, ref. *31*), glucoamylase (*glaA*, ref. *32*), and prepro-polygalacturonidase II (*pgaII*, ref. *33*) expressed in *A. niger*. We have undertaken a similar approach in obtaining enhanced expression of extracellular ß-glucosidase. Extra copies of a genomic clone of *bgl1* were introduced into the genome of *T. reesei* using the transformation vector called pSASß-glu (Figure 1).

Positive selection of *T. reesei* transformants containing extra copies of *bgl1* was made using the *A. nidulans amdS* gene as a selectable marker (*2*). *Trichoderma* does not contain a functional equivalent of the *amdS* gene and is therefore unable to utilize acetamide as a sole nitrogen source unless the *amdS* gene is stably inherited during transformation. Several stable transformants were seen to contain multiple copies of the *bgl1* gene (T. Fowler, unpublished Southern blot results). The cellulase products from these multicopy *bgl1* gene transformants showed an increase in the rate of glucose release from Avicel. One transformant was chosen for a more detailed analysis and was shown to contain 5-10 additional copies of the *bgl1* gene integrated into the genome, which gave rise to a 4.2 fold increase in *bgl1* mRNA levels (See figure 5 from ref. *1*). In addition, data is given for the action of the cellulase preparation (named C5X) from the transformant on cellobiose, Avicel and phosphoric acid swollen cellulose (PSC). An increase in the rate of production of glucose from the cellulosic substrates was observed for each substrate (See Figure 5 from ref. *1*). These data suggest that integration of additional copies of *bgl1* in the genome of *T. reesei* leads to an increase in specific message levels and corresponding extracellular protein levels. Furthermore the cellulolytic activity of *T. reesei* strains is specifically improved by transformation with the *bgl1* gene.

Increased saccharification observed from relatively pure substrates were an indicator that similar results upon treatment of less well defined cellulosic materials may be obtained. The resulting increase in the rates of hydrolysis may have potential applications for biomass conversion and ethanol production.

Cytolase 123 is a Genencor cellulase preparation from a proprietary *T. reesei* strain. The same strain was used as the parent for transformation with extra copies of *bgl1* resulting in the transformant that produce C5X. A comparison of the action of Cytolase 123 and C5X was made on a cellulosic waste floc from diaper manufacturing (figure 2). The cellulase preparation containing increased levels of ß-glucosidase (C5X) released higher levels of glucose from the diaper waste over time than did an equivalent amount of Cytolase 123.

Pretreatment of a variety of biological waste materials, such as agricultural residues, forestry waste, and municipal solid waste (MSW) gives rise to crude substrates that can subsequently be used in separate hydrolysis fermentation (SHF) or simultaneous saccharification and fermentation reaction (SSF) to produce ethanol (*34*). We were interested in seeing whether the cellulase mixture, C5X, could improve the SSF process by providing more glucose as a substrate to ferment into ethanol. In the SSF process a single vessel is used to mix pretreated MSW, hydrolytic cellulase enzymes and an organism capable of fermenting the released glucose to ethanol (such as the yeast strain *Saccharomyces cerevisiae*). This simplification can be approximated in laboratory shake flask analysis using cellulase enzyme preparations, a cellulosic substrate and a yeast innoculum. Table I shows results from such an experiment on a paper fraction of municipal solid waste using Cytolase 123 and C5X as the cellulase enzyme mixtures.

From this experiment, it appears that the enhanced ß-glucosidase cellulase preparation increases the availability of glucose resulting in increased ethanol production from yeast fermentation. This result is especially pronounced at the lower dosages. One possible explanation for this observation is that higher dosages of either cellulase mixture saturates the system with glucose and the resultant high alcohol levels become rate limiting.

In conclusion, novel cellulase mixtures can be produced from genetic manipulation of a single *T. reesei* strain. In the future these designer strains will result in simplified fermentation protocols and remove the need for supplementation of additional enzymes to existing enzyme preparation. Structural cellulose found in biomass materials is a mixture of crystalline cellulose, hemicellulose and lignin. Further improvement of strains of *T. reesei* for more rapid and complete conversion of cellulosic biomass may include the cloning and overexpression of the genes encoding hemicellulases and ligninolytic enzymes.

Deletion of the *bgl1* gene from *T. reesei*.

We have described how the composition of the cellulase complex of *T. reesei* can be altered by introduction of extra copies of cellulase genes. Conversely, a targeted gene disruption can remove or interrupt cellulase gene coding sequences. Such mutations not only alter the overall composition and mode of action of the cellulase preparation but provide a host into which the disrupted gene can be reintroduced following site directed mutagenesis that is designed to modify the enzyme's activity. In addition, the manipulation of each individual component at the genetic level presents an opportunity to look in new ways at several questions central to the regulation of the cellulase complex and the hydrolysis of cellulose. For example, how is each gene is involved in the induction and regulation of the cellulase complex? What is the contribution of each component in the hydrolysis of cellulosic substrates? What is the exact composition of the *Trichoderma* cellulase complex and multiplicity of the enzyme species?

Mutants of *Trichoderma reesei* lacking the coding sequence for the extracellular ß-glucosidase gene, *bgl1*, were obtained by a targeted gene replacement event (Figure 3). A gene replacement vector was first constructed (illustrated in Figure 3). The vector

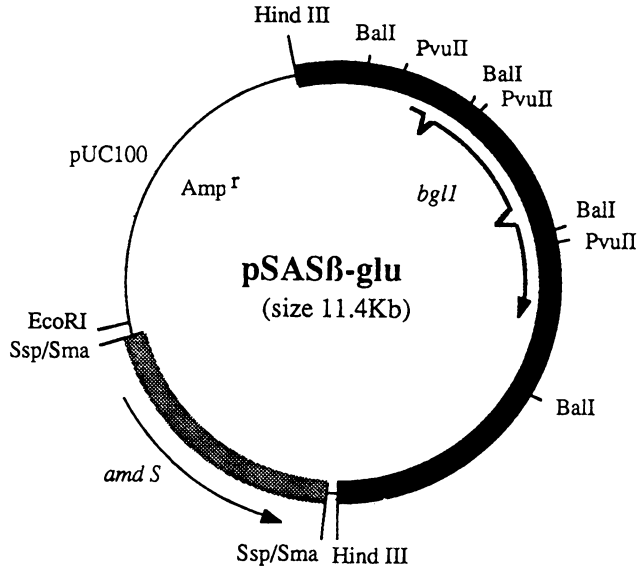


Figure 1. *Trichoderma reesei bgl1* gene overexpression vector pSASß-glu. The genomic *T. reesei bgl1* gene is contained on a 6.0 kb *Hind*III fragment. Figure source: from reference *1*.

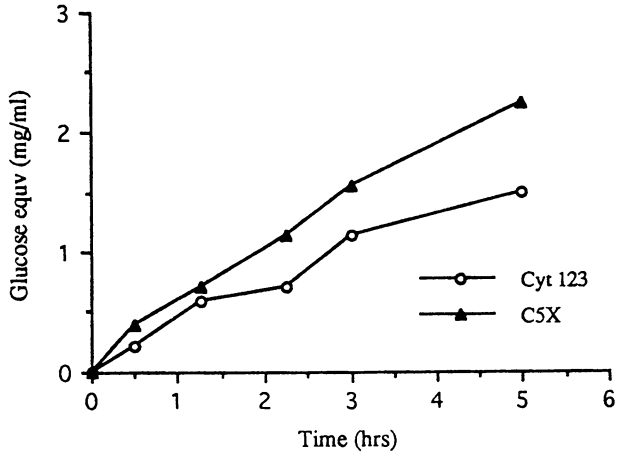


Figure 2. Comparison of Cytolase 123 and C5X action on cellulosic diaper waste. The cellulase mixtures were prepared using the same fermentation conditions. Dosage of the cellulase was 0.4mg enzyme/100mg substrate.

Table I. A Comparison of Cytolase 123 and C5X in the production of Ethanol from Cellulosic Biomass

Cellulase Activity(FPU)/ Gram Cellulose (MSW)	Comparison of Cytolase 123 and C5X[a] Grams EtOH/Liter SSF RXN	
	Cytolase 123	C5X
6	1.3	3.8
12	3.3	4.5
18	4.3	5.5
24	5.0	5.8
30	5.3	5.8
36	5.3	5.8

a) The cellulase activity of a standard batch of Cytolase 123 was calculated and dosed as shown. C5X was then compared to Cytolase 123 by using an equivalent amount of protein.

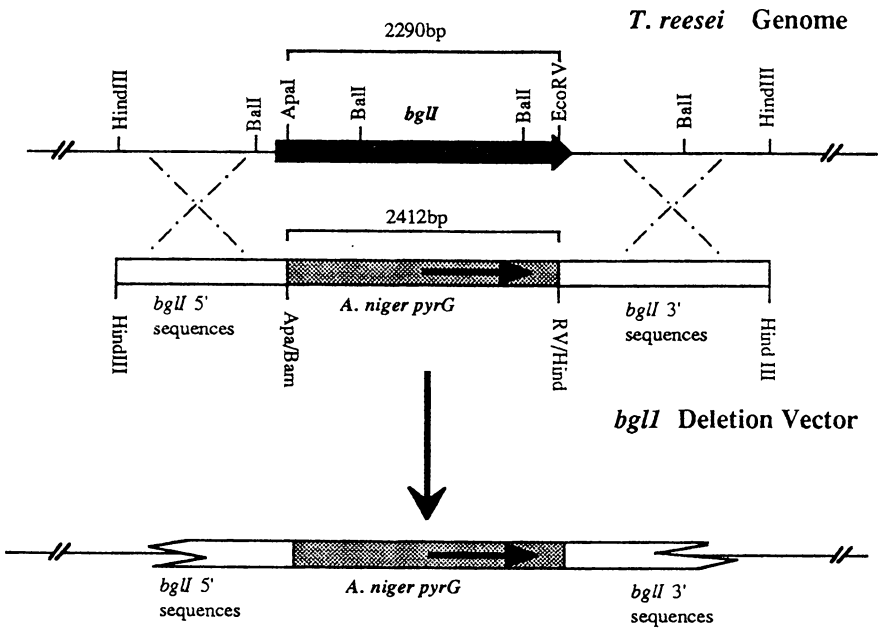


Figure 3. Schematic of the *bgl1* Gene Disruption Strategy.

was digested with *Hind*III to release a linear fragment in which the *bgl1* coding sequences were replaced with the *pyrG* gene from *Aspergillus niger* (*pyrG* encodes an enzyme required in the uridine biosynthetic pathway). Purified linear *Hind*III fragment from the disruption vector was used to transform a *T. reesei* strain P37 *pyrG69*, a uridine requiring auxotroph, to prototrophy. Transformants were screened by dot blot hybridization using *bgl1* coding sequences as a probe. The *bgl1* coding region was shown to be deleted by Southern blot analysis of transformant genomic DNA, as shown by a lack of hybridization signal when compared to the parental strain P37 *pyrG69* (data not shown). To confirm that transcription from the *bgl1* gene had also been disrupted, total RNA was isolated from the deletion transformant and P37 *pyrG69*. Northern blot analysis using the same *bgl1* probe indicated that *bgl1* specific mRNA present in P37 *pyrG69* is absent in the *bgl1* deletion strain (data not shown).

Western blot analysis established that extracellular ß-glucosidase is encoded by the *bgl1* gene. Extracellular ß-glucosidase was purified to near homogeneity by the method of Chirico and Brown (*14*) and used in the production of rabbit polyclonal antibodies. Extracellular protein from the *bgl1* deletion strain and P37, grown under inducing conditions, was separated on a 10-20% SDS polyacrylamide gel, electroblotted onto Nytran and probed using polyclonal antibodies raised against ß-glucosidase. Under conditions of induction, ß-glucosidase protein is observed in the culture broth of strain P37 but not in the *bgl1* deletion strain (data not shown). Interestingly, however, residual ß-glucosidase activity can still be detected using chromogenic substrates although the activity has been reduced by greater than an order of magnitude in comparison to P37 (data not shown). Experiments are currently in progress to determine the nature of this activity.

Other laboratories have attempted to obtain ß-glucosidase deficient strains of *T. reesei* by mutagenesis with either UV light (*35*) or gamma irradiation (*36*) Their inability to obtain non-leaky ß-glucosidase null mutations suggested to the authors that either ß-glucosidase is an essential gene (for example, involved in morphogenesis, Ref. *37*, or as an essential structural component of the cell wall) or there are at least two different ß-glucosidase genes in *T. reesei*. Evidently the extracellular ß-glucosidase encoded by the *bgl1* gene is not essential as demonstrated by the successful selection of a disruption within *bgl1*. Moreover, on the basis of DNA hybridization experiments we have not seen any evidence of sequences showing even weak homology to the *bgl1* gene within the genome of *T. reesei*. If there are other ß-glucosidase enzymes produced by *T reesei* (perhaps responsible for the residual activity on chromogenic substrates described above) they must be genetically distinct from *bgl1*. The lack of signal detection upon Western analysis of the *bgl1* disruption, using antibodies raised against the extracellular ß-glucosidase, argues against other ß-glucosidase enzyme forms arising from post-transcriptional modification of the *bgl1* mRNA or post-translational modification of the ß-glucosidase protein.

Future Directions and Conclusions.

Site Directed Mutagenesis Experiments on *bgl1*.

The deletion of the *bgl1* gene from *T. reesei* provides a host with which to study the role of extracellular ß-glucosidase in the regulation of other enzymes of the cellulase complex and its function in the hydrolysis of cellulose. The effect of the *bgl1* gene deletion on the regulation of the *T. reesei* cellulase genes and the effect on the breakdown of cellulosic materials is currently under investigation. Preliminary studies indicate that in the absence of ß-glucosidase, induction of the other cellulase genes is delayed at the level of transcription and that after a lag the *bgl1* deletion strain is able to

utilize cellulose as a carbon source. It is also interesting to note that the *bgl1* deficient strain is able to utilize cellobiose as a carbon source indicating the presence of alternative uptake and utilization mechanisms.

This section describes the use of the *T. reesei* strain containing the *bgl1* disruption as a host for site directed mutants of ß-glucosidase. Significant homology was observed between the deduced ß-glucosidase protein sequence of *T. reesei* and a catalytic ß-glucosidase peptide fragment from *Aspergillus wentii* (*22*). Furthermore, this region is shared by a number of other ß-glucosidases from cellulolytic fungi, yeasts, and bacteria (Table II). A twin aspartic acid motif appears in the form of a highly conserved S/T-D-W box in the first region followed by a conserved G-L-D-M box in the second region. The twin carboxylic acid motif also appears in the catalytic site of lysozyme, giving rise to the notion that the catalytic mechanism of lysozyme and cellulases may be similar (*23*). A comparable aspartic acid/glutamic acid mechanism has been demonstrated for the ß-glucosidase from *Agrobacterium* (Withers et al, personal communication). In addition carboxyl groups have been proposed to be essential for the function of the ß-glucosidase of *Schizophyllum commune* (*24*).

To investigate further the importance of residues conserved between ß-glucosidase enzymes (See Table II) on the catalytic activity of *T. reesei* ß-glucosidase, a series of targeted changes have been introduced into the sequence of the *bgl1* gene. The long term goal of these experiments is to create ß-glucosidase enzymes with altered catalytic activities. Some useful properties that will be screened for are different pH and temperature optima, increased catalytic turnover rates and decreased product inhibition profiles. We are currently in the process of expressing these mutant enzymes (using essentially the same vector as described in Figure 1) as a component of the whole cellulase complex, or individually in a host that has been deleted for other cellulase activities.

Table II. Homology around the known active site residue of *A. wentii* ß-glucosidase with other sequenced ß-glucosidase genes

		Reference
Butyvibrio fibrisolvens		
758	D E W G F E **G** V V V **S** **D** **W** W G F G E H Y K E V L A **G** **N** **D** I K M G C	*25*
Ruminococcus albus		
685	K Q W G F D **G** F T M **Y** **D** **W** W A N I N D R G C A P D **K** **N** **N** **F** A A M V	*26*
Kluyveromyces fragalis		
214	D E W K W D **G** M L M **S** **D** **W** F G T Y T T A A A I K N **G** **L** **D** I E F P	*27*
Candida pelliculosa		
288	E E L G F Q **G** F V M **T** **D** **W** G A L Y S G I D A A N A **G** **L** **D** **M** D M P	*28*
Saccharomycopsis fibuligera 1		
284	E E L G F Q **G** F V V **S** **D** **W** G A Q L S G V Y S A I S **G** **L** **D** **M** S M P	*29*
Saccharomycopsis fibuligera 2		
288	E E L G F Q **G** F V V **S** **D** **W** A A Q M S G A Y S A I S **G** **L** **D** **M** S M P	*29*
Aspergillus wentii		
?	A Z L G F Z **G** F V M **S** **D** **W** A A H H A G V S G A L A **G** **L** **B** **M** G S M P	*22*
Clostridium thermocellum		
220	N E W M H D **G** F V V **S** **D** **W** G A V N D R V S G L D A **G** **L** **D** **L** E M P	*30*
Trichoderma reesei		
225	D E L G F P **G** Y V M **T** **D** **W** N A E H T T V E S A N S **G** **L** **D** **M** S M P	*1*

The identified active site aspartic acid residue of *A. wentii* is underlined. Conserved residues are shown in bold face type.
Source: adapted from ref. *1*.

Conclusions.

The cellulolytic activity of *T. reesei* strains may be specifically improved by transformation with cloned cellulase genes. Recently transformation methods have been developed for *Trichoderma* that enable the amplification or deletion of the cellulase genes resulting in altered cellulase enzyme profiles. Similar genetic manipulations of *T. reesei* can be performed with the *bgl1* gene encoding the extracellular ß-glucosidase. The extracellular levels of ß-glucosidase can be increased by specific amplification of the *bgl1* gene in the genome of *T. reesei*. Extra copies of *bgl1* lead to increased message levels and a cellulase product that has increased cellulolytic activity. Deletion of the *bgl1* gene from the genome of *T. reesei* enables its role in cellulose hydrolysis to be determined directly. This well defined null-mutation is currently being used in our laboratory to address the questions of ß-glucosidase multiplicity, location, and role in signal transduction. In addition, the *bgl1* disruption strain will be used as a host for mutant ß-glucosidase enzyme forms whose catalytic activity has been altered by site directed mutagenesis.

Acknowledgments.

The previously unpublished data given in Table I and Figure 2 were the result of experiments performed with the help of Sharon Shoemaker, Dan Wendt and Steve Lewis of Genencor International Incorporated. The author would also like to thank Drs. Mikalina Gritzali and Ross D. Brown Jr. (University of Florida, Gainesville) for collaborative efforts with protein purification and biochemical analysis.

Literature Cited.

1. Barnett, C.B.; Berka R.; Fowler T. *Bio/technology* **1991**, *9*, 562-567.
2. Pentitila, M. ; Nevalainen, H.; Ratto, M.; Salminen, E.; Knowles, J. *Gene* **1987**, *61*, 155-164.
3. Smith, J.L.; Bayliss, F. T.; Ward, M. *Cur. Genet.* **1991**, *19*, 27-33.
4. Gruber, F.; Visser, J.; Kubicek, C.P.; de Graaff, L.H. *Curr. Genet.* **1990**, *18*, 71-76.
5. Gritzali, M.; Brown, R.D., Jr. *Advances in Chemistry series.* **1979**, *181*, 237-260.
6. Niku-Paavola, M.L. In: Proceedings of the Soviet Union - Finland seminar on bioconversion of plant raw materials by microorganisms. Tashkent, **1983**, 26-31.
7. Uusitalo, J.M.; Nevalainen, K.M.H.; Harkki, A.M.; Knowles, J.K.C.; Penttila M.E. *J. Biotechnol.* **1991**, *17*, 35-50.
8. Messner, R.; Kubicek, C. P. *Enzyme Microb. Technol.* **1990**, *12*, 685-690.
9. Kubicek, C. P. *Eur. J. Appl. Biotechnol.* **1981**, *13*, 226-231.
10. Messner, R.; Hagspiel, K.; Kubicek, C. P. *Arch. Microbiol.* **1990**, *154*, 150-155.
11. Enari, T.M.; Niku-Paavola, M.L.; Harju ,L.; Lappalainen, A.; Nummi, M. *J. Appl. Biochem.* **1981**, *3*, 157-163.
12. Sternberg, D.; Vijayakumar, P.; Reese, E. T. *Can. J. Microbiol.* **1977**, *23*, 139-147.
13. Kadam, S. K.; Demain, A. L. *Biochem. Biophys. Res. Comm.* **1989**, *161*, 706-711.
14. Chirico, W.J.; Brown, R.D. Jr. *Eur. J. Biochem.* **1987**, *165*, 333-341.
15. Shoemaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis, M. *Bio/Technology* **1983**, *1*, 691-696.

16. Chen, M.C.; Gritzali, M.; Stafford, W.D. *Bio/Technology* **1987**, *5*, 274-278.
17. Teeri, T. T.; Lehtovaara, P.; Kauppinen, S.; Salovuori, I.; Knowles, J.K.C. *Gene* **1987**, *51*, 43-52
18. Penttila, M.; Lehtovaara, P.; Nevalainen, H.; Bhikhabhai, R.; Knowles, J.K.C. *Gene* **1986**, *45*, 253-263.
19. Saloheimo, M.; Lehtovaara, P.; Penttila, M.; Teeri, T.T.; Stahlberg, J.; Pettersson, G.; Claeyssens, M.; Tomme, P.; Knowles J.K.C. *Gene* **1988**, *63*, 11-21.
20. Gavel, Y.; Heijne von, G. *Protein Engineering* **1990**, *3*, 433-442.
21. Gurr, S. J.; Unkles, S. E.; Kinghorn J. R. In: Gene structure in eukaryotic microbes. Kinghorn J. R. (Ed.) IRL Press. **1987**, 93-139.
22. Bause, E.; Legler, G. *Biochim. Biophys. Acta.* **1980**, *626*, 459-465.
23. Paice, M.G.; Jurasek, L. *Adv. Chem. Ser.* **1979**, *181*, 361-374
24. Clarke, A. J. *Biochim. Biophys. Acta.* **1990**, *1040*, 145-152.
25. From a proprietaty Genentech database.
26. Ohmiya, K.; Takano, M.; Shimizu S. *Nucleic Acids Research* **1990**, *18*, 671.
27. Raynal, A.; Gerbaud, C.; Francingues, M.C.; Guerineau, M. *Curr. Genet.* **1987**, *12*, 1987.
28. Kohchi, C.; Toh-e, A. *Nucleic Acids Research* **1985**, *13*, 6273-6282.
29. Machida, M.; Ohtsuki, I.; Fukui, S.; Yamashita, I. *App. Env. Micro.* **1988**, *54*, 3147-3155.
30. Grabnitz, F.; Ruecknagel, K.P.; Seiss, M.; Staudenbauer, W.L. *Mol. Gen. Genet.* **1989**, *217*, 70-76.
31. Whittington, H.; Kerry Williams, S.; Bidgood, K.; Dodsworth, N.; Peberdy, J.; Dobson, M.; Hinchliffe, E.; Ballance, D.J. *Curr. Genet.* **1990**, *18*, 531-536.
32. Fowler, T.; Berka, R.; Ward, M. *Curr. Genet.* **1990**, *18*, 537-545
33. Bussink, H.J.D.; Kester, H.C.M.; Visser, *J. FEMS Lett.* **1990**, *273*, 127-130.
34. Gauss, W.F.; Suzuki, S.; Takagi, M. U.S. Patent #3990944 **1976**.
35. Mishra, S.; Rao, S.; Deb, J.K. *J. Gen. Micro.* **1989**, *135*, 3459-3465.
36. Strauss, J.; Kubicek, C.P. *J. Gen. Micro.* **1990**, *136*, 1321-1326.
37. Jackson, M.A.; Talburt, D.E. *Exp. Micol.* **1988**, *12*, 203-216.
38. Durand, H.; Baron, M.; Calmels, T.; Tiraby, G. In : Biochemistry and Genetics of Cellulose Degradation, Aubert, J.-P.; Beguin, P. and Millet, J. (eds), New York, Academic Press. **1988a**, 135-151.
39. Durand, H.; Clanet, M.; Titalby, G. *Enzyme Microb. Technol.* **1988b**, *10*, 341-345.
40. Kowamori, M.; Ado, Y.; Takasawa, S. *Agr. Biol Chem.* **1986**, *50*, 2477-2482.
41. Valheri, M.P.; Leisola, M.; Kaupinnen, V. *Biotechnol. Lett.* **1979**, *1*, 41-46.
42. Kubicek, C.K. *J. Gen. Microbiol.* **1987**, *133*, 1481-1487.
43. Messner, R.; Hagspiel, K.; Kubicek, C.P. *Arch. Microbiol.* **1990**, *154*, 150-155.
44. Sprey, B. *FEMS Microbiol. Lett.* **1986**, *36*, 287-292.
45. Umile, C.; Kubicek, C.P. *FEMS Microbiol. Lett.* **1986**, *34*, 291-295.
46. Usami, S.; Kirimura, K.; Imura, M.; Morikawa, S. *J. Ferm. Bioeng.* **1990**, *70*, 185-187.

RECEIVED March 31, 1992

Chapter 20

Structure–Function Relationships in Cellulase Genes

David B. Wilson

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853

Cellulases are mechanistically interesting because the hydrolysis of crystalline cellulose requires the action of a set of enzymes acting together, despite the fact that all of the enzymes catalyze the hydrolysis of a β-1,4-linkage between two glucose residues. Genetic engineering techniques are being applied to cellulase genes in order to determine the mechanism of action of cellulases and to try to engineer more active cellulases. This review describes various approaches currently used to improve cellulases and defines new directions for recombinant cellulase technology.

Cellulases are being actively studied because of the strong interest in commercializing a process for the conversion of biomass cellulose to ethanol. A major problem in such a process is the high cost of cellulase, which could be reduced from about 40% of the total cost to 2% if cellulases having a 20-fold higher specific activity could be developed (*1*). While it is easy to visualize how an exocellulase and an endocellulase might cooperate in cellulose hydrolysis, it is difficult to see how exocellulases could be synergistic to each other (i.e., in multi-exocellulase systems). Yet, selected exocellulases display about as much synergism as do an exocellulase and endocellulase (*2*).

There are several reasons for cloning cellulase genes. One is to determine the amino acid sequence of a cellulase by sequencing the DNA of the cloned cellulase gene and then translating the open reading frame. Another is to introduce a cloned gene into an organism that does not produce cellulases and that can express the gene at a high level in order to simplify the purification of the cellulase it encodes or to create useful microorganisms by adding a cellulase gene to an organism that lacks either a specific type of cellulase gene or any cellulase genes. A third is to use a cloned gene to carry out region specific mutagenesis, site directed mutagenesis, or domain shuffling experiments which can help determine the role of specific residues

0097–6156/93/0516–0243$06.00/0

or domains in the function of the cellulase encoded in the cloned gene. A fourth is to study the role of modification, such as glycosylation on enzyme activity. *Thermomonospora fusca* E_2 is an endoglucanase that is lightly glycosylated (~2% sugar). When the *T. fusca* E_2 gene is expressed in *Streptomyces lividans*, the E_2 produced is not glycosylated and yet its enzymatic activity is nearly identical to that of *T. fusca* E_2 (*3,4*).

All of these reasons apply to any gene encoding any protein; but the second and third reasons are particularly important for cellulase genes, since most cellulolytic microorganisms contain a number of cellulase genes, thus making it difficult to isolate mutations in them. Furthermore, because some cellulases display synergism (i.e., the activity of the mixture is greater than the sum of the activities of the enzymes acting alone), traces of an endocellulase in a purified exocellulase can significantly alter its apparent properties. We have seen this with *T. fusca* cellulase E_3. Our earlier preparations contained low levels of activity on carboxymethyl cellulose (CMC) which comigrated with E_3 during polyacrylamide gel electrophoresis. Recently, we found that the CMCase activity could be removed by affinity chromatography, and the new preparation had significantly less activity on filter paper than our original preparation (*2*). However, when the new E_3 preparation was assayed on filter paper with an endocellulase or with the exocellulase, cellobiohydrolase I (CBH I) from *Trichoderma reesei*, the new E_3 preparation gave a slightly higher activity than our original sample. This result shows that the change in activity of E_3 was not caused by denaturation, but probably was due to the removal of an endoglucanase that contaminated our original E_3 preparations. Similar results were reported for cellobiohydrolases I and II purified from *Penicillum pinophilum* (*5*). In this paper, it was shown that the addition of very small amounts of an endoglucanase to highly purified preparations of *P. pinophilum* CBH I and CBH II could dramatically increase the activity of mixtures of these enzymes on either cotton or Avicel. The problems caused by not removing the last traces of endoglucanases from exocellulases can be avoided if noncellulolytic host strains containing only a single cloned exocellulase gene are used as the source of the enzyme.

Certain anaerobic bacteria, such as *Clostridium thermocellum*, assemble their cellulases in large particles (called cellulosomes) on their outer surface (*6,7*). Cloning and expression of cellulase genes often provides the best way to obtain active preparations of an individual cellulase from such organisms. This is due to the difficulty in isolating individual cellulases in an active form from cellulosomes. In one case, a cloned *C. thermocellum* cellulase was not only obtained in large amounts from *E. coli* cells carrying the cloned gene, but the enzyme from the cloned strain was also crystallized (*8*).

Many different cellulase genes have been cloned from a number of species of bacteria, fungi, and plants, and these studies have been recently reviewed (*9*). Over 70 cellulase genes have been sequenced and comparisons of these sequences show that the regions coding for the catalytic domains can be grouped into seven families, with the sequences within each family showing some similarity with all the members of that family (*10*). One family contains eight xylanase genes, as well as two cellulase genes. The fact that xylanases and cellulases can have similar sequences is not surprising, because some hydrolytic enzymes have activity on both cellulose and xylan (*11,12*).

From work with other proteins, it is likely that the three-dimensional structures of all of the members of a cellulase family will be quite similar in their overall polypeptide folding pattern, although there will be significant differences that account for the specific properties of each enzyme. It is also likely that the structures of enzymes in different families will show few similarities. At this time, only two native cellulases have been crystallized and both of them were unable to bind to cellulose, thus indicating the absence of a cellulose binding domain (*8,13*). It has been possible to crystallize the catalytic domains of several cellulases and the three-dimensional structure of *T. reesei* CBH II has been determined to 2.7 Å resolution by X-ray crystallography (*13*). We have crystallized the catalytic subunit of *T. fusca* E_2 (*15*), which belongs to the same cellulase family as *T. reesei* CBH II, even though it is an endocellulase. We have determined the three-dimensional structure of the E_2 catalytic subunit and our 1.8 Å model has the same basic conformation as *T. reesei* CBH II, however, its active site is present in a cleft rather than in a tunnel as it is in CBH II. These results confirm the prediction that cellulases from the same family will have a similar overall three-dimensional structure, and a careful comparison of the two structures should provide detailed information about some of the important structural differences between exocellulases and endocellulases.

One property of cellulases and xylanases that appears to be constant among the members of a given cellulase family is the stereochemistry of cleavage. Some cellulases invert the conformation of the glycoside linkage during cleavage while others retain the β-conformation in the product. Thirteen cellulases and three xylanases have been studied at this time, including members of five different families, and all members of a given family catalyze hydrolysis with the same stereochemistry (*16,20*).

Cellulose Binding Domains

Genetic engineering techniques currently are being used to study the role of cellulose binding domains in cellulase function. These domains are present in many bacterial and fungal cellulases (*21-23*) and they can be present at either the N-terminal or C-terminal end of the catalytic domain. Since cellulases within the same gene family can differ in the location of their cellulose binding domain, the location may not be critical. Most bacterial cellulose binding domains show some homology with each other, even then the catalytic domains belong to different families (*10*). Fungal cellulase binding domains have a very different sequence from the bacterial binding domains and are much smaller (33 amino acids (aa) long rather than 100 aa)(*24*). The three dimensional structure of the *T. reesei* CBH I cellulose binding domain has been determined by two dimensional NMR, but it is not yet clear how it binds to cellulose (*25*). Studies of a *Cellulomonas fimi* cellulose binding domain, that was purified from a strain containing a modified gene that only encoded the binding domain, have shown that the binding domain disrupts the structure of cellulose fibers as seen in the electron microscope (*26*). The cellulose binding domain appears to function like the C_1 activity postulated many years ago (*27*); that is, it disrupts the structure of cellulose without catalyzing its hydrolysis. These results suggest that the binding domain of a cellulase not only binds the enzyme to cellulose, but also helps disrupt the interactions between cellulose molecules that make cellulose so resistant to digestion.

This probably explains why cellulases that do not have binding domains have little activity on crystalline cellulose (*28*).

Genetic Engineering Studies

Genetic engineering has been used to produce proteins that only contain a cellulose binding domain (*26*), or that only contain a catalytic domain (*29*), or to add a cellulose binding domain to enzymes that lack them (*30*). This was first done with *E. coli* alkaline phosphatase and the modified enzyme retained activity and could be immobilized on cellulose (*30*). We have attached a cellulose binding domain to the C-terminus of a *Bacteroides ruminicola* catalytic domain (*31*). The native enzyme did not bind tightly to cellulose, but the modified enzyme did (*31*) Furthermore, the modified enzyme had a specific activity on acid swollen cellulose that was ten times that of the original enzyme and its activity on ball milled cellulose was eight times that of the original enzyme. The modified enzyme still had little activity on Avicel, but, unlike that native enzyme, it was able to synergize with an exocellulase in Avicel digestion (*31*). These experiments confirm the results of previous experiments in which removal of the binding domain from active cellulases reduced their activity on insoluble substrates, but not on soluble ones and shows that at least some CMCase catalytic domains can hydrolyze insoluble cellulose when joined to a cellulose binding domain.

Cloned cellulase genes have been introduced into organisms that ferment glucose to ethanol efficiently, such as yeast or Zymomonas, to try to create organisms that can ferment cellulose to ethanol (*32-34*). At this time none of these modified organisms degrade cellulose fast enough to give a significant rate of ethanol production from cellulose. This is not surprising, since all known cellulolytic organisms produce several cellulases. The *cen* A and *cex* genes of *Cellulomonas fimi* were cloned on the same plasmid, and both were introduced into yeast (*34*). The yeast transformants had some activity on crystalline cellulose, but it was less than 0.1% of the activity produced by *T. reesei*, the most active cellulolytic organism. This low activity is probably due to low expression and the fact that the *cex* enzyme has a much higher activity on xylan than on cellulose. The challenge is to introduce a set of cellulase genes into the organism and get them all expressed and secreted at a high level.

Cloned cellulase genes have also been transformed back into the parent strain in order to alter the composition of the crude cellulase mixture produced by the strain so as to alter its specific activity. This has been done with *T. reesei*, where the gene for endoglucanase I (*egl* I) was introduced into both a while type strain and a strain in which the cellobiohydrolase I gene (*cbh* I) was inactivated (*35*). The *egl* I gene was present in an expression vector which contained both the *cbh* I promoter and terminator regions. Stable transformants were isolated in which the *egl* I gene had inserted at the chromosomal *cbh* I locus in both the wild type and *cbh* I negative strains. The transformed wild type strain made two times more endoglucanase I (EG I) than the parent strain, but the overall activity on filter paper of the crude cellulase was unchanged from that of the parent strain. The *cbh* I negative mutant strain carrying the *egl* I gene made about four times more EG I than the wild type strain, presumably because CBH I makes up 60% of the total protein produced by the wild type strain, thus its absence allowed a higher level of expression of the cloned *egl* I gene.

A similar experiment has been carried out in which the gene for CBH II was transformed into wild type *T. reesei* (*36*). Several stable strains were isolated that produced from two to three times more CBH II than the wild type. These trains also had more filter paper activity than the parent strain, but did not produce cellulase complex with a higher specific activity, because the amount of protein they secreted also increased several fold (*36*). In these studies, there was no correlation between the number of *cbh* II genes present in a given transformant and the amount of CBH II produced where the number of *cbh* II genes present in different strains ranged from 10 to 20 (*36*).

Site directed mutagenesis of potentially important residues in cellulases has been reported even though detailed structural information is only available for a few cellulases. Rouvinen et al. (*14*) identified two Asp residues (Asp^{175} and Asp^{221}) in the active site of *T. reesei* CBH II in their three dimensional structure that were also conserved in most related cellulases. These Asp residues were converted to Ala by site directed mutagenesis. The Asp^{175} to Ala ^{175} mutant retained 20% of the wild type activity, whereas the Asp^{221} to Ala^{221} mutant was inactive. Thus, Asp^{221} appears to be directly involved in catalysis while Asp^{175} may function to alter the pK of Asp^{221}.

Another way to identify important catalytic residues is to look for residues that are conserved in members of a given cellulase family. Baird et al. (*37*) found a sequence Asn-Glu-Pro that was present in 16 members of cellulase family A. They converted the Glu residue in this sequence in both a *Bacillus polymyxa* (Glu^{167}), and a *Bacillus subtilis* endoglucanase, to Gln. In each case, the modified cellulase had less than 5% of its original activity as tested by a CMC overlay assay of the mutant colonies. Activity was restored when the mutant genes were changed back to the original sequence, proving that it was the Glu to Gln change that inactivated the enzymes rather than secondary mutations. These workers changed two other nonconserved Glu residues in the *B. subtilis* enzyme to either Ala, Asp, or Pro, and all of the mutants retained activity. Py et al. (*38*) have studied another member of cellulase family A, *Erwinia chrysanthemi* EG2, by site directed mutagenesis. They found that conversion of either Glu^{133} or His^{98} to Ala gave proteins that did not have activity (less than 0.25%) on either CMC or p-nitrophenyl-β-D-cellobioside. Both mutants produced proteins that bound tightly to Avicel, reacted with antibodies against EG2, and had the correct molecular weight. At this time, no information is available to determine whether the mutations prevented correct folding of the protein or inactivated essential residues. Glu^{133} in EG2 is similar to Glu^{167} in the *B. polymyxa* CMCase.

Chemical Studies

Chemical modification studies also have been used to identify potential active site residues in cellulases. One such study proposed that either Glu^{126}, Asp^{130}, or Asp^{132} of *T. reesei* CBH I was an essential residue (*39*). This proposal was tested by site directed mutagenesis experiments in which Glu^{126} was changed to Gln, Asp^{130} to Asn, and Asp^{132} to Ala (*40*). Each of these mutations retained at least 30% of the wild type activity, showing that none of the residues was essential. Surprisingly, when Glu^{127} in *T. reesei* EG I, which is the equivalent of Glu^{126} in CBH I, was changed to Gln,

the enzyme lost all activity. All the mutant enzymes in this study were produced in yeast which extensively glycosylates the enzymes. Mutant EG I appeared to be more extensively glycosylated than wild type EG I and overglycosylated enzyme from the wild type strain did not have activity. This result suggests that overglycosylation may have caused the loss of activity in the Glu^{127} mutant. These experiments on *T. reesei* CBH I show the danger of overinterpreting chemical modification experiments, which can give incorrect results due to secondary changes in the enzyme. Experiments using site directed mutagenesis also have to be interpreted carefully, especially when the mutations reduce activity, since the mutation can prevent the protein from folding correctly, giving the appearance of effecting an essential catalytic residue. This is probably what has occurred with the EG I Gln^{127} mutant, since it appears unlikely that the conversion of a Glu residue to a Gln residue would lead to a dramatic change in glycosylation, unless it changed the folding of the protein.

A combination of chemical mutagenesis and site directed mutagenesis was used to identify a residue that was important, but not essential, for cellulase function. Chemical modification of a single His residue resulted in a loss of 70%-80% of the activity of *Clostridium thermocellum* endoglucanase D (Cel D) as measured with 2-chloro-4-nitrophenyl-β-D-cellobioside (*41*). Attempts to identify the modified His by digestion and peptide fractionation were inconclusive because several His residues were modified at similar rates. All 12 His residues in Cel D were converted individually to Ala or to Ser by site directed mutagenesis. Six of the mutants retained 80% or more of the original activity, whereas the activity of the other six ranged from 2% to 25%. The residue that was modified in the native enzyme was determined by reacting each of the last six mutants with diethyl pyrocarbonate (DEPC), because serine does not react with DEPC. Only the mutant where His^{516} was changed to Ser was unaffected by DEPC. This result proved that His^{516} was the residue that had been modified in the wild type enzyme. Consistent with this conclusion is the fact that this mutant retained 25% of the wild type activity and His^{516} is the only His residue conserved in members of family E.

The results of these site directed mutagenesis experiments show the power of this approach in determining whether a specific residue in a cellulase is or is not essential for activity. Furthermore, they strongly support the proposal (*42*) that carboxyl side chains play an essential role in cellulase activity. However, many more experiments are needed before we have a clear understanding of the molecular mechanisms by which the different types of cellulases catalyze cellulose hydrolysis.

Literature Cited

1. Lynd, L.R.; Cushman, J.H.; Nichols, R.J.; Wyman, C.E. *Science* **1991**, *251*, 1318-1323.
2. Irwin, D.; Wilson, D.B., manuscript in preparation.
3. Ghangas, G.S.; Wilson, D.B. *Appl. Environ. Microbiol.* **1988**, *54*, 2521-2526.
4. Spezio, M. unpublished.
5. Wood, T.; McCrae, S.; Bhat, K. *Biochem J.* **1989**, *260*, 37-43.
6. Kobayashi, T.; Romaniec, M.P.M.; Fauth, U.; Demain, A.L. *Appl. Environ. Microbiol.* **1990**, *56*, 3040-3046.
7. Lamed, R.; Setter, E.; Bayer, E.A. *J. Bacteriol.* **1986**, *156*, 828-836.

8. Joliff, G.; Beguin, P.; Joy, M.; Millet, J.; Peyter, A.; Poljak, R.; Avbert, J.-P. *Bio/Technology* **1986**, *4*, 896-900.
9. Beguin, P.; Gilkes, N.R.; Kilburn, D.G.; Miller, R.C.; O'Neill, G.P.; Warren, R.A.J. *Crit. Rev. Biotechnol.* **1987**, *6*, 129-162.
10. Gilkes, N.R.; Henrissat, B.; Kilburn, D.G.; Miller, R.C. Jr.; Warren, R.A.J. *Microbiol. Reviews* **1992**, *55*, 303-315.
11. Shoemaker, S.; Watt, K.; Tsitousky, G.; Cox R. *Bio/Technology* **1983**, *1*, 687-690.
12. Woods, J.R.; Hudman, J.F.; Gregg, K.J. *Gen. Microbiol.* **1989**, *135*, 2543-2549.
13. Katsube, Y. personal communication.
14. Rouvian, J.; Bergfors, T.; Teeri, T.; Knowles, J.K.C.; Jones, T.A. *Science* **1990**, *249*, 380-386.
15. Spezio, M.; Wilson, D.B.; Karplus, A.D. manuscript in preparation.
16. Gebler, J.; Gilkes, N.; Claeyssens, M.; Wilson, D.; Beguin, P.; Wakarchuk, W.; Kilburn, D.; Miller, R.; Warren, R.A.; Withers, S. *J. Biol. Chem.* **1992** in press.
17. Withers, S.G.; Dombroski, D.; Berven, L.A.; Kilburn, D.F.; Miller, R.C. Jr.; Warren, R.A.J.; Gilkes, N.R. *Biochem. Biophys. Res. Commun.* **1986**, *139*, 487-494.
18. Meinke, A.; Braun, C.; Gilkes, N.R.; Kilburn, D.G.; Miller, R.C. Jr.; Warren, R.A.J. *J. Bacteriol* **1991**, *173*, 308-314.
19. Knowles, J.K.C.; Lehtovaara, P.; Murray, M.; Sinnott, M.L. *J. Chem. Soc. Chem. Commun.* **1988**, 1401-1402.
20. Claeyssens, M.; van Tilbeurgh, H.; Kamerling, J.P.; Berg, J.; Vrsnska, M.; Biely, P. *Biochem. J.* **1990**, *270*, 251-256.
21. van Tilbeurgh, H.; Tomme, P.; Claeyssens, M.; Bhikhabhai, R.; Pettersson, G. *FEBS Lett.* **1986**, *204*, 223-227.
22. Saloheimo, M.; Lehtovaara, P.; Penttila, M.; Teeri, T.; Stahlberg, J.; Johansson, G.; Pettersson, G.; Claeyssens, M.; Tomme, P.; Knowles, J. *Gene* **1988**, *63*, 11-21.
23. Ong, E.; Greenwood, J.M.; Gilkes, N.R.; Kilburn, D.G.; Miller, R.C. Jr.; Warren, R.A.J. *Trends Biotechnol.* **1989**, *7*, 239-243.
24. Johansson, G.; Stahlberg, J.; Lindeberg, G.; Engstrom, A.; Pettersson, G. *FEBS Lett.* **1989**, *243*, 389-393.
25. Kraulis, P.; Clore, M.; Nilges, M.; Jones, A.; Pettersson, G.; Knowles, J.; Gronenborn, A. *Biochemistry* **1989**, *28*, 7241-7257.
26. Din, N.; Gilkes, N.R.; Tekant, B.; Miller, R.C. Jr.; Anthony, R.; Warren, J.; Kilburn, D.G. *Bio/Technology* **1992**, *9*, 1096-1099.
27. Reese, E.T.; Sui, R.G.H.; Levinson, H.S.S. *Bacteriol.* **1950**, *59*, pp. 485-497.
28. Klyosov, A. *Biochemistry* **1990**, *29*, 10577-10585.
29. Irwin, D.; Wilson, D.B. manuscript in preparation.
30. Greenwood, J.M.; Gilkes, N.R.; Kilburn, D.G.; Miller, R.C. Jr.; Warren, R.A.J. *FEBS Lett.* **1989**, *244*, 127-131.
31. Maglione, G.; Russell, J.B.; Wilson, D.B. manuscript in preparation.
32. Van Arsdell, J.N.; Kwok, S.; Schweickart, V.L.; Ledner, M.B.; Gelford, D.H.; Innis, M.A. *Bio/Technology* **1987**, *5*, 60-64.
33. Brestic-Goachet, N.; Gunasekaran, P.; Cami, B.; Baratti, J.C. *J. Gen. Microbiol.* **1989**, *135*, 893-902.

34. Wong, W.K.R.; Corry, C.; Parekh, R-S.; Parekh, S.R.; Wayman, M.; Davies, R.W.; Kilburn, D.G.; Skipper, N. *Bio/Technology* **1988**, *6*, 713-719.
35. Harkki, A.; Mantyla, A.; Pentilla, M.; Muttilainen, S.; Buhler, R.; Suominen, P.; Knowles, J.; Nevalainen, H. *Enzyme Microb. Technol.* **1991**, *13*, 227-233.
36. Kubicek-Pranz, E.; Gruber, F.; Kubicek, C.J. *Biotechnol.* **1991**, *20*, 83-94.
37. Baird, S.D.; Hefford, M.A.; Johnson, D.A.; Sung, W.L.; Yaguchi, M.; Seligy, V.A. *Biotechmical and Biophys. Res. Commun.* **1990**, *169*, 1035-1039.
38. Py, B.; Bortoli-German, I.; Haiech, J.; Chippaux, M.; Barras, F. *Protein Engin.* **1991**, *4*, 325-333.
39. Tomme, P.; Claeyssens, M. *FEBS Lett.* **1989**, *243*, 239-243.
40. Mitsuishi, Y.; Nitisinprasert, S.; Saloheimo, M.; Biese, I.; Reinikainen, T.; Claeyssens, M.; Keranen, S.; Knowles, J.; Teeri, T. *FEBS Lett.* **1990**, *275*, 135-138.
41. Tomme, P.; Chauvaux, S.; Beguin, P.; Millet, J.; Aubert, J-P.; Claeyssens, M. *J. Biol. Chem.* **1991**, *266*, 10313-10318.
42. Clarke, A.J.; Yaguchi, M. *Eur. J. Biochem.* **1985**, *149*, 233-238.

RECEIVED June 12, 1992

Chapter 21

Clostridium thermocellum Cellulosome

New Mechanistic Concept for Cellulose Degradation

J. H. David Wu

Department of Chemical Engineering, University of Rochester, Rochester, NY 14627–0166

Clostridium thermocellum is an anaerobic and thermophilic bacterium with an optimum growth temperature of 60°C. It produces an extracellular cellulase system shown to be highly active on crystalline cellulose. The Avicelase activity (i.e., activity against Avicel, a microcrystalline cellulose) of this cellulase system resides mainly in an unusually large protein aggregate called the cellulosome. The cellulosome is now known to contain multiple, discrete subunits and comprise a total molecular weight of millions. An apparent ordering to this supramolecular structure, first indicated by organization of the genetic elements, may provide a basis for the cellulase action attributed to *C. thermocellum* and related surface-active cellulase producers.

The *C. thermocellum* cellulosome aggregate has a very stable quaternary structure, resistant to most detergents and other dissociating reagents (*1,2*). Although the whole complex has been isolated and characterized, functions of individual subunits remain unclear. Purification of individual components from the cellulosome appears to be a prohibitive task. This issue has seriously impeded the mechanistic study of this cellulase system.

The first insight into the mechanism of cellulose degradation by the cellulosome, and its organization, was obtained when the two major subunits, CelL (or S_L) and CelS (or S_S), which degrade crystalline cellulose synergistically, were identified (*3*). An "anchor-enzyme model" was proposed to explain the synergism between these two subunits (*4*). In this model, CelL functions as an anchor on the cellulose surface for CelS, the catalytic subunit. This mechanism is clearly different from the well-accepted model based on the fungal cellulase system. Its novelty has prompted interest in cloning the *celL* and *celS* genes. Recently, the *celL* (*5,6*) and *celS* (*7,8*) genes have been successfully cloned. Their DNA sequences reveal surprisingly striking features which shed more light onto how the cellulosome is organized.

0097–6156/93/0516–0251$06.00/0

Furthermore, the group at Institute Pasteur, France, has independently discovered that other catalytic subunits, CelD and XynZ, also complex with the anchor CelL (*9*), as CelS does. The complex formation involves specific ligand-receptor interaction (*10*). These results confirm the original anchor-enzyme model and expand it into a more sophisticated picture, allowing initial construction of the basic structure of cellulosome. Not surprisingly, this structure depicts a novel mechanistic concept for both the supramolecular organization and enzymatic cellulose degradation.

In this article, the recent progress in cellulosome research leading to the unveiling of a proposal for its complex quaternary structure will be reviewed.

Properties of the *C. thermocellum* Cellulase System

A very important characteristic of the *C. thermocellum* cellulase system is its high specific activity toward crystalline cellulose. A comparison based on equal amounts of extracellular protein (*11,12*) demonstrates that the *C. thermocellum* cellulase system is much more active on cotton and Avicel than *Trichoderma reesei* cellulase. In fact, the cellulase system degrades Avicel faster than phosphoric acid-swollen Avicel (*11*), indicating it prefers crystalline cellulose to amorphous substrate. This is very unusual, because most cellulase systems more readily degrade amorphous substrates.

Another unusual property of the *C. thermocellum* cellulase system is its requirement for Ca^{++} for the maximum activity (*12*). Recent evidence indicates that Ca^{++} serves to stabilize the protein at high temperature (*13,14*). Ethylenediaminetetraacetate (EDTA) completely inhibits Avicelase activity (*12*). The cellulase system is also inhibited by certain apolar chelating agents (*15*). This inhibition is reversed by mixture of Fe^{++} and Fe^{+++}, suggesting that iron is involved in its catalytic action.

As expected from its anaerobic and thermophilic nature, the clostridial cellulase has a high thermostability and is sensitive to oxygen inactivation (*12,15*). Similar to the fungal system, it is also subject to feedback inhibition by cellobiose and, to a much lesser extent, glucose (*16*).

The Cellulosome: A Multicomponent and Multifunctional Protein Complex

The preference for a crystalline substrate and the requirement for Ca^{++}, and possibly iron, are unique properties among cellulases. These observations strongly suggest that the clostridial cellulase system may adopt a completely different enzymatic mechanism than those derived from the fungal system. Indeed, it did not take long for researchers to realize that this system exists as an extremely unusual multisubunit protein complex, termed cellulosome (*1*). Isolation and characterization of the cellulosome have been reviewed by Lamed and Bayer (*17,18*). Only a brief historical sketch will be provided here.

The Cellulosome as a Cellulose-Binding Factor. The cellulosome was first purified as the cellulose-binding factor from the YS strain of *C. thermocellum* based on the ability to bind to cellulose (*2*). Under the electron microscope, these molecules look like a group of closely related complexes with complicated quaternary structure. The complexes have diameters of about 18 nm and estimated molecular weights of about 2.1 million Daltons. At least fourteen protein subunits (S1 - S14) with molecular weights ranging from 210 to 48 KDa are associated with the isolated aggregates as revealed by SDS-PAGE. At least eight components of the cellulosome are active against CMC (i.e., determined by a CMC-overlay assay of the components separated by SDS-PAGE).

Among the subunits, S8 (M_r = 75,000) and S1 (M_r = 210,000) are the two most abundant species. The strikingly large size and abundance of S1 suggests that it probably plays an important role in the function of cellulosome. This role is suggested by studies on a mutant (AD2 [adherence-defective]) of the YS strain, which lacks the cell bound S1. The AD2 mutant fails to bind to the cellulose surface (*19*). Furthermore, the same mutant produces less S1 in the culture broth than the wild strain. This phenomenon is associated with an increased amount of lower molecular weight endoglucanases with decreased affinity toward cellulose (*20*). The studies based on the AD2 mutant, therefore, suggest that S1 is a multifunctional protein which is involved in the organization of the cellulosome, serves to anchor the cellulases to the cellulose surface, and mediates cell adhesion.

The cellulosome also occurs on the cell surface, especially during the exponential growth phase, where they form distinctive polycellulosomal protuberances with sizes ranging from 60 to 200 nm (*20,21*). As the culture enters the stationary phase, more cellulosomes are released into culture broth (*19,22*). Upon growth on cellulose, the polycellulosomal protuberances on the cell surface appear to aggregate at the contact site between the cell and the cellulose surface, forming the "contact corridor" bridging the cell and the cellulose surface (*21*). On the other hand, the cellulose surface, when void of cells, is coated by a layer of cellulosomes.

Cellulosome molecules of larger sizes (M_r = 100 x 10^6 and 4.5 x 10^6) have been isolated from the JW20 strain of *C. thermocellum*, which contain 20 to 35 subunits (*22,23*). Many of them have been demonstrated to have cellulase or xylanase activities (*24*). The morphology of these cellulosome molecules has been studied using electron microscopy (*25*). Various forms of cellulosomes, i.e. the "tight cellulosome" with tightly packed subunits and the "loose cellulosome" with loosely packed subunits, have been observed. Both forms may "decompose" into smaller structural elements containing five to eight subunits arranged equidistantly in a parallel and symmetric array along a central groove or axis. These observations indicate that the size and the morphology of cellulosomes may vary depending on the number of the structural elements in a cellulosome molecule and how these elements are arranged. [See discussions of the revised anchor-enzyme model below].

The Cellulosome as a Cellulase Complex. Attempts to purify the cellulase components from the ATCC 27405 strain of *C. thermocellum* revealed that most of the Avicelase and the majority of the CMCase activities are associated with the protein aggregates with molecular weights of a few million Daltons (*3*). This

indicates that the major function of the cellulosome is cellulose degradation. Indeed, at least eight subunits of the YS strain cellulosome are active against CMC (*2*). Furthermore, isolated cellulosome from the YS strain possesses the same properties as the crude enzyme, including the requirement for a reducing agent and Ca^{++} for maximum activity, and the sensitivity to cellobiose inhibition (*26*).

The involvement of a protein complex with complicated quaternary structures clearly distinguishes the *C. thermocellum* cellulase from the fungal system. This raises questions of how many subunits are required and whether the quaternary structure they form is crucial for cellulose degradation. These questions were partially answered when a subcellulosome was isolated (*27*). This subcellulosome contains only six major subunits, yet displays much higher specific activity against CMC and Avicel than the crude enzyme. The subcellulosome is inactivated by sulfhydryl reagents and is inhibited by EDTA and the apolar chelating agent, *o*-phenanthroline. These properties are consistent with those of the crude enzyme. Therefore, at least for the standard laboratory assay, cellulose degradation by cellulosomes seems to involve only a "core unit" of the cellulosome, consisting of a small number of key subunits.

Subunits of the Cellulosome. The structure-function relationship of the cellulosome became the central theme of research since it was established that the cellulosome is the reaction center for cellulose degradation. The first step in this study is to characterize the role of the individual components. Since the cellulosome is very stable and cannot be dissociated without denaturing the protein, study in this respect has been approached mostly through the molecular cloning of cellulase (*cel*) genes. As a result, more than twenty cellulase, xylanase, and ß-D-glucosidase genes have been cloned into *Escherichia coli* (*28-30*). Many of those genes (including *celA* [*31*], *celB* [*32*], *celC* [*33*], *celD* [*34*], *celE* [*35*], *celF* [*36*], *celH* [*37*], *xynZ* [*38*], *bglA* [*39*], *bglB* [*40*]) have been sequenced. Table 1 lists the endoglucanase and xylanase genes that have been cloned and sequenced.

Table 1. Cloned *C. thermocellum* Endoglucanase and Xylanase Genes

Cloned Gene	Size (b.p.)	Predicted MW (Daltons)	Observed MW (Daltons)	ref.
CelA	1,344	52,503	56,000	(*41*)
CelB	1,689	63,857	66,000	(*42*)
CelC	1,032	40,439	38,000	(*43*)
CelD	1,947	72,344	65,000	(*44*)
CelE	2,442	90,211	------	
CelF	2,219	87,409	------	
CelH	2,702	102,301	------	
XynZ	2,511	92,159	90,000	(*38*)

It is clear from these cloned genes that the observed diversity in cellulase molecules produced by this bacterium is not solely due to post-translational modifications or proteolysis. The predicted sizes of most of the proteins encoded fall in the range of those of the cellulosome components (48 - 210 KDa). One striking feature of these genes is that all, except *celC*, code for a conserved, duplicated sequence of 24 amino acid residues. This duplicated sequence is located at the C-terminus of all the Cel proteins except CelE and XynZ, where the sequence is in the middle of the proteins (Figure 1).

The duplicated sequence on the CelD and XynZ has been shown to function as a binding ligand to CelL (M_r = 250,000), the largest subunit of the cellulosome of the ATCC 27405 strain (*9*). Deletion of the duplicated sequence deprive the proteins of the ability to bind to CelL. These results indicate that proteins with this conserved, duplicated sequence bind to CelL and that they are subunits of the cellulosome. In this respect, it is interesting to note that the CelCCA of *C. cellulolyticum* also contains this duplicated sequence (*45*).

CelL, with many subunits binding to it, could serve as the "scaffolding" for the cellulosome. Since many proteins contain the duplicated sequence, they could compete for the same binding site on CelL. Alternatively, a single CelL molecule could provide multiple binding sites. As will be discussed later, the DNA sequence of *celL* indicates that it indeed serves as the supporting structure of the cellulosome and provides multiple binding sites.

The Anchor-Enzyme Model

Although the crude *C. thermocellum* cellulase is very active against crystalline cellulose, none of the cloned genes code for protein with significant activity on the crystalline substrate. It is possible that the gene(s) encoding the essential subunit(s) for such activity has not been cloned. In addition, the complicated quaternary structure may be crucial for such activity. The high stability of the cellulosome and the large number of its subunits have been a serious technical impediment for attempts to unveil the mystery of how the cellulosome degrades crystalline cellulose. However, the first insight was provided when two cellulosome subunits, CelS (or S_S; M_r = 82,000) and CelL (or S_L; M_r = 250,000) were identified.

CelS was first purified by gel filtration chromatography of the cellulosome from the ATCC 27405 strain and partially dissociated with SDS (*3*). Upon dissociation, the Avicelase activity was lost. However, the activity was partially restored by combining CelS with another chromatographic fraction containing predominantly CelL. CelL was later purified by elution from an SDS-gel. The CelL thus purified retained the activity to degrade Avicel synergistically with CelS. Cellulose degradation occurred only when CelL and CelS were combined, indicating their cooperative action.

It would be tempting to explain this cooperative action by the model of synergism between endo- and exoglucanase. However, experimental evidence indicates that a unique mechanism is probably involved. The cellulose degradation probably involves the interactions between CelL, CelS and the water-insoluble cellulosic substrate. An

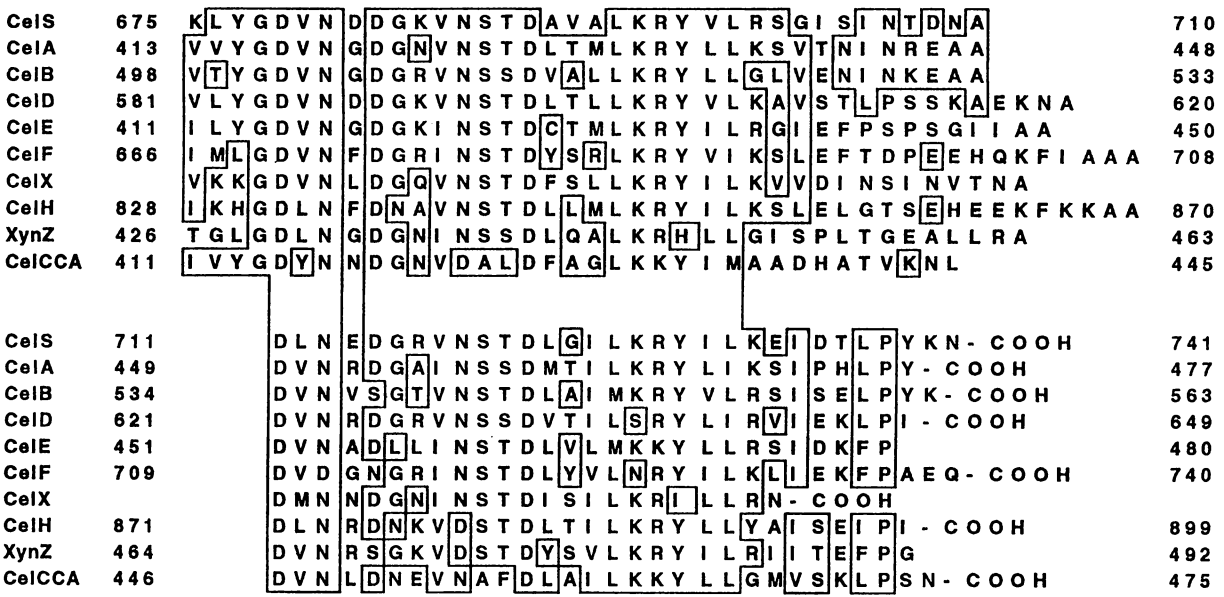


Figure 1. Alignment of the conserved duplicated sequence between CelS, CelA, CelB, CelD, CelE, CelF, CelX, CelH, and XynZ of *C. thermocellum* and CelCCA of *C. cellulolyticum*. Boxed amino acids are identical or have similar chemical properties. Numbers indicate the position, within the sequence of each protein, of the first or the last amino acid shown on a line. Similar residues are: V, L, I, M, F; R, K; D, E; N, Q; Y, F, W; S, T.

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

anchor-enzyme model (Figure 2 [*4*]) has been proposed to explain this synergistic effect. In this model, CelL functions as an anchorage subunit which connects the CelS, a catalytic subunit, to the cellulose surface. This model is based on the following observations: 1) although CelS is not active on crystalline cellulose, it is active on CMC; 2) the CMCase activity of CelS is not significantly enhanced by CelL; 3) adsorption of CelS to the cellulose surface is dependent on CelL; 4) the purified CelL has no enzymatic activity.

It became apparent that CelL and CelS are the two most abundant subunits of the cellulosome. Furthermore, the reconstituted Avicelase (CelL + CelS) is capable of degrading most of the Avicel particles in the assay system (Wu and Demain, unpublished data), indicating that the crystalline portion of the cellulose is attacked. Finally, the reconstituted Avicelase has the same properties as the crude enzyme, such as requiring Ca^{++} for the maximum activity, generating predominantly cellobiose as the hydrolysis product, and the sensitivity to cellobiose inhibition (*4*). The reconstituted Avicelase, therefore, appears to be representative of the Avicelase activity of the crude enzyme.

The Revised Anchor-Enzyme Model

The complex of CelS and CelL represents the simplest subcellulosome active on crystalline cellulose. The anchor-enzyme model provides the first clue to the mechanism and organization of the cellulosome. The intriguing feature of the model and the technical difficulty in purifying CelL and CelS through SDS treatment have triggered interest in targeting the molecular cloning at genes for these two subunits. Recently, both celL and celS genes have been cloned. Not surprisingly, analysis of their gene structures reveals much greater details regarding how the cellulosome is organized in an anchor-enzyme configuration.

CelL gene (or *cipA* for cellulosome integrating protein) was cloned by using antibody against CelL (*5*). Its open-reading-frame codes for a protein of 196,800 KDa (*6*). An extremely striking feature of its DNA sequence is that about 75% of the gene consists of repeated sequence of about 0.5 kbp with two direct repeats in the 5' end (N-terminus of CelL) and seven direct repeats in the 3' end (C-terminus of CelL). In addition, the CelL also has the conserved, duplicated sequence at the C-terminus, which serves as a binding ligand to itself as mentioned above. The DNA sequence, therefore, suggests that CelL binds to itself. The implication of this self-binding is that CelL molecules, by binding to each other, provide a scaffolding for the quaternary structure of the cellulosome.

If CelL is the key building block for the cellulosome, how then may the various subunits interact with this hypothetical CelL scaffolding? Since all Cel proteins with the conserved, duplicated sequence would potentially bind to CelL, it is likely that each repeated domain on the CelL serves as the binding receptor for the duplicated sequence. This interesting possibility has been experimentally verified. Using ^{125}I-labeled CelD protein as a probe to screen an expression genomic library, Fujino et al. (*10*) obtained a truncated *celL* (or *cipA*) gene coding for the C-terminal end of CelL. The encoded CelL fragment contains two full repeated domains (out of nine

in the intact CelL). Gene deletion analysis indicated that CelD is able to bind to only one repeated domain and the same protein with the duplicated sequence deleted fails to bind.

Description of the New Model. These recent results indicate that CelL functions as an anchor not only for CelS but also for other catalytic subunits. A schematic drawing of this new model is shown in Figure 3. In this model, various catalytic subunits bind to CelL through the interaction between the duplicated sequence and the repeated domain to form a "core unit" of the cellulosome. The core units are linked to each other through the CelL's own duplicated sequence. The fact that most of the duplicated sequences lie in the C-terminus probably allows the catalytic domain to "stick out" from the core structure. Some catalytic subunits have their own cellulose binding domain (CBD) which may further enhance the binding to cellulose surface (*36,46*). It is interesting to note that the cellulose binding domain and the catalytic domain of CelE reside on the opposite sides of the duplicated sequence (*35,46*), again allowing the domains to "stick out" from the core structure. At this time, it is not clear whether there is any protein-protein interaction between the "anchored" subunits.

In a cellulosome molecule of about 2.1 million Daltons (*2*), two to three such core units may exist. Theoretically, larger or smaller cellulosome molecules could be formed depending on the degree of the self-association of CelL. Furthermore, the CelL scaffolding could be linear or circular and more than one of the repeated domains could be used for self-association. The geometry and the size of this self-association may depend on whether the cellulosome is in solution, on the cell surface, or on the substrate surface. This explains the observation of cellulosome molecules of different sizes and shapes (*20,25*) and the morphology change of cellulosome, for example, "transformation" upon binding to cellulose surface (*21*) or "structural decomposition" in the later stage of fermentation (*25*).

It is worth noting that the structure of the core unit depicted in Figure 3 is strikingly similar to the structural element revealed by electron microscopy (*25*). The structural element mentioned earlier is formed by "rows of equidistantly spaced polypeptides arranged parallel to the major axis" Based on the structure of this element, Mayer et al. (*25*) and Coughlan and Ljungdahl (*47*) proposed the model of "simultaneous multicutting event" leading to the cellulose degradation. If the CelL represents the major axis and the various catalytic subunits represents "the equidistantly spaced polypeptides arranged parallel to the axis", then this revised anchor-enzyme model is strikingly consistent with the structural element and the multicutting model. Mayer et al. (*25*) even indicated that five to eight identical subunits are present in an element, in close agreement with the number of subunits shown in Figure 3.

CelL: The Key Building Block of the Cellulosome. In the first anchor-enzyme model proposed, CelL functions only as an anchor having no catalytic activity. The lack of catalytic activity has been reported by two separate groups (*2-4*) both using SDS-treated CelL for assay. Contrary to this finding, a truncated, non-denatured recombinant version of CelL was shown to be active on CMC (*5*). This discrepancy

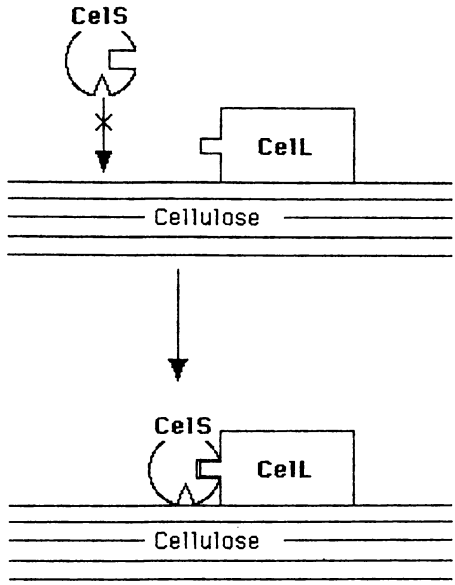


Figure 2. The anchor-enzyme model indicating that CelL functions as an anchor on the cellulose surface for CelS, the catalytic subunit.

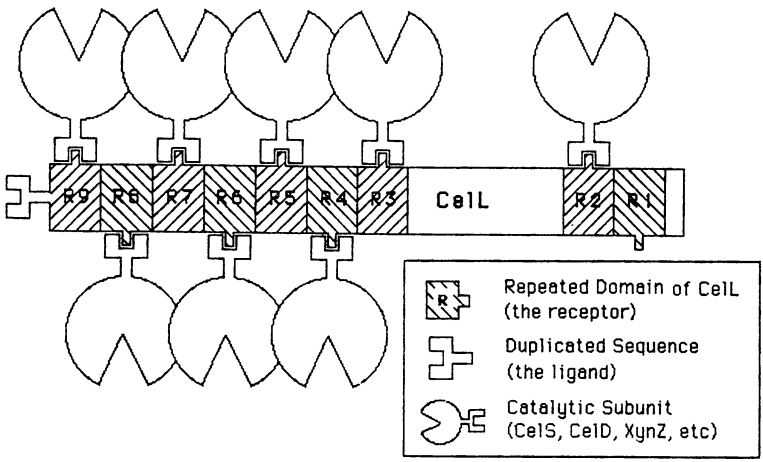


Figure 3. The revised anchor-enzyme model indicating that CelL functions as an anchor for other subunits to form a "core unit" of the cellulosome. One repeated domain (R1) on CelL is left open to indicate the potential self-association site for CelL. This site is randomly chosen in this drawing; however, this site could be specific due to the possible protein-protein interactions beyond the ligand-receptor interaction described in the model.

is likely due to the irreversible denaturation of the catalytic domain by SDS in earlier experiments. If the catalytic site lies on the repeated domain, CelL is not only an anchor, but also provides a hydrolysis site in every domain to synergize with the catalytic activities of the other subunits. As an anchor on the cellulose surface, CelL is likely to have its own cellulose binding domain. It is not clear whether the cellulose binding domain is located in the repeated domains or in the non-repeated domain. It would also be interesting to determine if CelL possesses non-hydrolytic, cellulose fiber-disrupting activity (*48*).

Besides its large size and repeated domains, CelL is also unique in being glycosylated. CelL contains 5-7% carbohydrate (*49*) and a novel carbohydrate-peptide linkage (*50*). The fact that the truncated recombinant CelL is active on CMC (*5*) and capable of binding the catalytic subunits (*10*) indicates that glycosylation is not essential for its catalytic activity or receptor function.

CelS: The Major Cellulosome Subunit. Although the receptor sites in CelL could be occupied by any subunits with the duplicated sequence according to the revised model, the majority of those sites will probably be occupied by CelS since it is the most abundant subunit in the cellulosome. In fact, CelS is the most abundant protein species of the total extracellular protein of the ATCC 27405 strain (Figure 4). Its equivalent in the YS strain, the S8 subunit, is also the most abundant subunit (*1,2*). In the structural element reported by Mayer et al. (*25*), the equidistantly arranged subunits have been reported to be identical; however, smaller subunits are also present in the same cellulosome molecule. Therefore, the receptor sites on CelL are probably occupied mostly by CelS with other subunits interspersed.

The S8 subunit has been shown to be active on CMC using a CMC-overlay assay (*2*). Its mode of action has been determined using its proteolytic fragment (S8-tr) with molecular weight of 68 KDa, which dissociates from the cellulosome after proteolysis and can be easily purified (*13*). It appears that the binding ligand (the duplicated sequence) is cleaved by the protease, resulting in the dissociation of S8-tr from the cellulosome (*9,13*). Although activity on CMC in the overlay assay suggests endoglucanase activity, S8-tr displays typical cellobiohydrolase activity (*13*). More importantly, its activity is stabilized by Ca^{++} and reducing agent and is strongly inhibited by cellobiose. It also produces cellobiose as the major hydrolysis product. These properties are consistent with those found in the crude enzyme, the isolated cellulosome, the subcellulosome, and the reconstituted Avicelase (CelL + CelS). Furthermore, the biosynthesis of S8 appears to be induced by growth on cellulose (*20*). The DNA sequence of *celS* gene shares no homology with other known *cel* genes (*7,8*) except the region coding for the duplicated sequence. The duplicated sequence (Figure 1) confirms its ability to bind to CelL. Finally, it is the only subunit which has been demonstrated to degrade a major portion of crystalline cellulose synergistically with CelL (*3*). CelS therefore plays a critical role in cellulose degradation by the cellulosome.

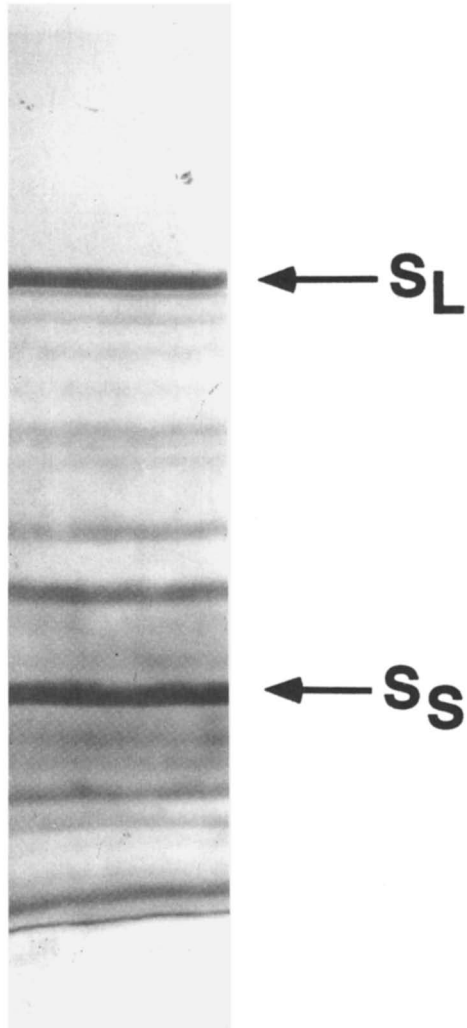


Figure 4. SDS-PAGE pattern of the total extracellular proteins of *C. thermocellum* ATCC 27405, indicating that CelS (S_S) and CelL (S_L) are the most abundant protein species. The crude proteins were obtained by growing the bacterium on cotton.

Conclusion

In spite of the complexity of the cellulosome, it appears that its organization probably follows a very simple rule: attack of the repeated and ordered structure of the substrate with repeated and ordered domains through the quaternary structure of the enzyme molecules. An even simpler rule is used in organizing the quaternary structure by employing a universal set of ligands/receptors. Although many questions remain to be answered, the elegance and the beauty revealed by the model are fascinating. It represents a new concept not only for cellulose degradation, but also for supramolecular protein structure.

The unique properties of the *C. thermocellum* cellulase system apparently are rooted in its unique structure. As the details of this structure are elucidated at the molecular level, it may be possible to redesign the cellulosome following the same simple rules of the game.

Literature Cited

1. Lamed, R.; Setter, E.; Kenig, R.; Bayer, E.A. *Biotechnol. Bioeng. Symp.* **1983**, *13*, 163-181.
2. Lamed, R.; Setter, E.; Bayer, E.A. *J. Bacteriol.* **1983**, *156*, 828-836.
3. Wu, J.H.D.; Orme-Johnson, W.H.; Demain; A.L. *Biochemistry* **1988**, *27*, 1703-1709.
4. Wu, J.H.D.; Demain; A.L. In *Biochemistry and Genetics of Cellulose Degradation*; Aubert; J.-P.; Beguin, P.; Millet, J., Eds.; Academic Press: London, 1988; pp 117-131.
5. Romaniec, M.P.M.; Kobayashi T.; Fauth U.; Gerngross U.T.; Demain, A.L. *Appl. Biochem. Biotechnol.* **1991**, *31*, 119-134.
6. Gerngross U.T.; Romaniec, M.P.M.; Huskisson, N.S.; Demain, A.L. Personal communication.
7. Wang, W.K.; Wu, J.H.D. *Appl. Biochem. Biotechnol.* In press.
8. Wang, W.K.; Wu, J.H.D. 1992. Submitted.
9. Tokatlidis, K.; Salamitou, S.; Beguin, P.; Dhurjati, P.; Aubert, J.-P. *FEBS Lett.* **1991**, *291*, 185-188.
10. Fujino, T.; Beguin, P.; Aubert, J.-P. *FEMS Microbiol. Lett.* **1992**, *94*, 165-170.
11. Johnson, E.A. PhD Thesis, M.I.T., Cambridge, MA, 1983.
12. Johnson, E.A.; Sakajoh, M.; Halliwell, G.; Madia, A.; Demain, A.L. *Appl. Environ. Microbiol.* **1982**, *43*, 1125-1132.
13. Morag, E.; Halevy, I.; Bayer, E.A.; Lamed, R. *J. Bacteriol.* **1991**, *173*, 4155-4162.
14. Chauvaux, S., Beguin, P.; Aubert, J.-P.; Bhat, M.K.; Gow, L.A.; Wood, T.M.; Bairoch, A. *Biochem. J.* **1990**, *265*, 261-265.
15. Johnson, E.A.; Demain, A.L. *Arch. Microbiol.* **1984**, *137*, 135-138.
16. Johnson, E.A.; Reese, E.T.; Demain, A.L. *J. Appl. Biochem.* **1982**, *4*, 64-71.
17. Lamed, R; Bayer, E. A. *Adv. Appl. Microbiol.* **1988**, *33*, 1-46.

18. Lamed, R.; Bayer, E. A. In *Biochemistry and Genetics of Cellulose Degradation*; Aubert; J.-P.; Beguin, P.; Millet, J., Eds.; Academic Press: London, 1988; pp 101-116.
19. Bayer, E.A.; Kenig, R.; Lamed., R. *J. Bacteriol.* **1983**, *156*, 818-827.
20. Bayer, E.A.; Setter, E.; Lamed, R. *J. Bacteriol.* **1985**, *163*, 552-559.
21. Bayer, E.A.; Setter, E.; Lamed, R. *J. Bacteriol.* **1986**, *167*, 828-836.
22. Hon-Nami, K.; Coughlan, M.P.; Hon-nami, H.; Ljungdahl, L.G. *Arch. Microbiol.* **1986**, *145*, 13-19.
23. Coughlan, M.P.; Hon-Nami, K.; Hon-Nami, H.; Ljungdahl, L.G.; Paulin, J.J.; Rigsby, W.E. *Biochem. Biophys. Res. Commun.* **1985**, *130*, 904-909.
24. Kohring, S.; Wiegel, J.; Mayer, F. *Appl. Env. Microbiol.* **1990**, *56*, 3798-3804.
25. Mayer, F.; Coughlan, M.P.; Mori, Y.; Ljungdahl, L.G. *Appl. Env. Microbiol.* **1987**, *53*, 2785-2792.
26. Lamed, R.; Kenig, R.; Setter, E.; Bayer, E.A. *Enzyme Microb. Technol.* **1985**, *7*, 37-41.
27. Kobayashi, T.; Romaniec, M.P.M.; Fauth, U.; Demain, A.L. *Appl. Env. Microbiol.* **1990**, *56*, 3040-3046.
28. Schwarz, W.; Bronnenmeier, K.; Staudenbauer, W.L. *Biotechnol. Lett.* **1985**, *7*, 859-864.
29. Hazlewood, G.P.; Romaniec, M.P.M.; Davidson, K.; Grepinet, O.; Beguin, P.; Millet, J.; Raynaud, O.; Aubert, J.-P. *FEMS Microbiol. Lett.* **1988**, *51*, 267-282.
30. Beguin, P.; Millet, J.; Grepinet, O.; Navarro, A.; Juy, M.; Amit, A.; Poljak, R.; Aubert, J.-P. In *Biochemistry and Genetics of Cellulose Degradation*; Aubert; J.-P.; Beguin, P.; Millet, J., Eds.; Academic Press: London, 1988; pp 11-30.
31. Beguin, P.; Cornet, P.; Aubert, J-P. *J. Bacteriol.* **1985**, *162*, 102-105.
32. Grepinet, O.; Beguin, P. *Nucleic Acids Res.* **1986**, *14*, 1791-1799.
33. Schwarz, W.H.; Schimming, S.; Rucknagel, K.P.; Burgschwaiger, S.; Kreil, G.; Staudenbauer, W. *Gene* **1988**, *63*, 23-30.
34. Joliff, G.; Beguin, P.; Aubert, J.-P. *Nucleic Acids Res.* **1986**, *14*, 8605-8613.
35. Hall, J.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J. *Gene* **1988**, *69*, 29-38.
36. Navarro, A.; Chebrou, M.-C.; Beguin, P.; Aubert, J.-P. *Res. Microbiol.* **1991**, *142*, 927-936.
37. Yague, E.; Beguin, P.; Aubert, J.-P. *Gene* **1990**, *89*, 61-67.
38. Grepinet, O.; Chebrou, M.-C.; Beguin, P. *J. Bacteriol.* **1988**, *170*, 4582-4588.
39. Grabnitz, F.; Seiss, M.; Rucknagel, K.P.; Staudenbauer, W.L. *Eur. J. Biochem.* **1991**, *200*, 301-309.
40. Grabnitz, F.; Rucknagel, K.P.; Seiss, M.; Staudenbauer, W.L. *Mol. Gen. Genet.* **1989**, *217*, 70-76.
41. Schwarz, W.; Grabnitz, F.; Staudenbauer, W.L. *Appl. Env. Microbiol.* **1986**, *51*, 1293-1299.
42. Beguin, P.; Cornet, P.; Millet, J. *Biochimie* **1983**, *65*, 495-500.
43. Petre, D.; Millet, J.; Longin, R.; Beguin, P.; Girard, H.; Aubert, J-P. *Biochimie* **1986**, *68*, 687-695.
44. Joliff, G.; Beguin, P.; Juy, M.; Millet, J.; Ryter, A.; Poljak, R.; Aubert, J.-P. *Bio/Technology* **1986**, *4*, 896-900.
45. Faure, E.; Belaich, A.; Bagnara, C.; Gaudin, C.; Belaich, J.-P. *Gene* **1989**, *84*, 39-46.

46. Durrant, A.J.; Hall, J.; Hazlewood, G.P.; Gilbert, H.J. *Biochem. J.* **1991**, *273*, 289-293.
47. Coughlan, M.P.; Ljungdahl, L.G. In *Biochemistry and Genetics of Cellulose Degradation*; Aubert; J.-P.; Beguin, P.; Millet, J., Eds.; Academic Press: London, 1988; pp 11-30.
48. Din, N.; Gilkes, N.R.; Tekant, B.; Miller, R.C.; Warren, R.A.J.; Kilburn, D. *Bio/technology* **1991**, *9*, 1096-1099.
49. Gerwig, G.J.; Kamerling, J.P.; Vliegenthart, J.F.G.; Morag, E.; Lamed, R.; Bayer, E.A. *Eur. J. Biochem.* **1991**, *196*, 115-122.
50. Gerwig, G.J.; Waard, P.; Kamerling, J.P.; Vliegenthart, J.F.G.; Morgenstern, E.; Lamed, R.; Bayer, E.A. *J. Biol. Chem.* **1989**, *264*, 1027-1035.

RECEIVED August 20, 1992

Chapter 22

Protein Chemical Cross-Linking

Implications for Protein Stabilization

Shan S. Wong[1], Michael Losiewicz[2], and Lee-Jun C. Wong[3,4]

[1]Department of Pathology and Laboratory Medicine, University of Texas Health Science Center, Houston, TX 77030
[2]Department of Chemistry, University of Massachusetts, Lowell, MA 01854
[3]Department of Biological Sciences, University of Massachusetts, Lowell, MA 01854

Chemical cross-linking of proteins is a special application of chemical modification. The cross-linkers are bifunctional compounds containing two reactive functionalities derived from group specific reagents. They may be classified into homobifunctional, heterobifunctional and zero-length cross-linkers. Different physical and chemical properties have been integrated into these cross-linking reagents, e.g., lengths and sizes, hydrophobicity and hydrophilicity, cleavability, iodinability, fluorofors, chromophores and spin labels. Examples of each class of compounds are presented. The use of the bifunctional reagents in the stabilization of proteins are discussed with regard to immobilization onto solid supports, cross-linking to other soluble proteins and intramolecular cross-linking. Examples are presented to demonstrate that the reticulation of proteins or enzymes stabilizes native molecular structures against denaturation by chemical, thermal and mechanical forces. It is suggested that the thermal stability of cross-linked proteins may be evaluated by the Arrhenius equation.

The stability of proteins and enzymes is a major concern in their industrial applications in organic synthesis, isolation and purification of chemicals, in their use as a reagent for chemical analysis, in therapeutics and diagnostics, and in the study of their structures and functions. Many methods have been evolved to preserve the integrity and activity of the native proteins. Based on thermodynamic reasoning, proteins were invariably kept at low temperatures to prolong their longetivity. Although many enzymes and proteins can be stored in this manner for an extended period of time, low temperatures are not preferable in many industrial operations.

[4]Current address: Institute for Molecular Genetics, Baylor College of Medicine, Houston, TX 77030

0097–6156/93/0516–0266$06.00/0

Furthermore, some proteins are cold sensitive and may be more stable at ambient temperatures. In an attempt to mimic the microenvironment of the proteins in the cell, various substances have been added to interact with isolated proteins. Glycerol and ethylene glycol have been used, particularly enabling the proteins to be kept at very low temperatures without being frozen. Other substances such as carbohydrates like sucrose and proteins like albumin have also been used to stabilize proteins of interest. While these methods have preserved the biological activities to various extents, the procedure introduces foreign substances which may not be desirable in many applications. A new method for protein stabilization using chemical modification has evolved. Cross-linking reagents have been used to intramolecularly cross-link and to conjugate protein or enzymes to other molecules including solid supports. This technique has been demonstrated to greatly enhance the stability of proteins. This Chapter will briefly review the characteristics of cross-linking reagents and their use in protein stabilization. Examples of proteins stabilized by chemical cross-linking will be presented.

General Characteristics of Cross-Linking Reagents

Chemical cross-linking of biological components has been applied to the study of membrane components and soluble proteins, the preparation of immunoconjugates and immunotoxins. Various reviews have appeared (*1-6*). A recent monograph covering the principle of cross-linking and its applications to various areas has also been published (*7*). Readers who are interested in detailed treatment of this area are encouraged to consult these publications. A summary review with particular reference to protein stabilization will be presented here.

Cross-linking reagents used for the stabilization of proteins are essentially chemical modification agents. These compounds contain two reactive functionalities which will react with amino acid side-chains of a protein, thus bringing two components together. A schematic presentation of these reagents is shown in Figure 1. The reactive groups are located at the two ends of the molecule connected by a backbone which may be designed to contain specific characteristics. Although any reactive functionalities may be incorporated into the head groups, the most common functional entities are acylating and alkylating agents (*8,9*). This is because the most reactive amino acid side-chains susceptible for modification are nucleophiles, such as sulfhydryl group of cysteine, amino groups of lysine and N-terminal amino acids, carboxyl groups of aspartic and glutamic acids and C-terminal amino acids, imidazolyl group of histidine, and thioether group of methionine. Thus the most common reactions with the two-headed compounds are acylation and alkylation of the protein.

Acylating agents. Acylating agents are compounds that confer acyl groups to nucleophiles. These compounds contain a good leaving group attached to the acyl group so that they are easily replaced by the reacting nucleophile. Since water is a nucleophile and is present in abundance in aqueous media, hydrolysis of the acylating agent may be an important side reaction, consuming considerable amount of the reactant. There are many acylating agents. Of particular importance are isocyanates, isothiocyanates, imidoesters, sulfonyl chlorides, N-hydroxysuccinimidyl

and other activated esters, such as p-nitrophenyl ester. All of these reactive functional groups have been incorporated into cross-linkers (*7*). Some of these reactions are represented in Figure 2.

Alkylating agents. Among the alkylating functionalities that have been incorporated in the two-headed reagents, α-haloacetyl, N-maleimidyl, and halo-aryl derivatives are the most commom. In these reactions the nucleophile attacks the activated carbon displacing a leaving group as shown in Figure 3. When a nucleopile reacts with a halobenzene, arylation takes place.

Group specific reactions. The specificity of the cross-linking reagent for a specific amino acid side-chain depends on the relative reactivity of the nucleophile. Since the nucleophilicity of an amino acid side-chain depends on several factors, such as its electronic structure, its pK_a and its microenvironment, the reactivity of an amino acid side-chain is not specific and several side-chains may react with the same alkylating and acylating functionalities of the bifunctional reagents (*7*). However, the thiolate ion is the most nucleophilic (*10*). Thus the sulfhydryl groups of proteins will react with most of the cross-linkers at alkaline pH. For example, α-haloacetyl compounds and N-maleimido derivatives are generally considered as sulfhydryl selective. A different group of reagents that are thiol specific are disulfide and mercurial compounds. These functionalities makes the bifunctional reagents react with only thiol groups of proteins. Disulfide compounds react through a disulfide interchange reaction as shown in Figure 4 (*11*).

In the absence of any sulfhydryl moiety, the amino group becomes the major target of reaction. Similarly, for proteins with high content of surface amino groups, the competition for the reagent may favor such a group. In addition, thioacyl esters as a result of acylation are susceptible to hydrolysis. Thus acylating agents are in general considered amino group directing.

Other reactions. In addition to amino acids, some proteins contain prosthetic groups which may be used for cross-linking. Carbohydrates of glycoproteins are particularly useful. The sugar moieties can be oxidized with periodate to form dialdehydes which will form Schiff bases with amines. These Schiff bases may be stablized by reducing agents in a process called reductive alkylation as shown in Figure 5 (*12*). Instead of glycoproteins, polysaccharides may be used as cross-linkers coupling two or more proteins through their amino groups (*7*).

Homobifunctional cross-linkers. When two identical functional groups are incorporated at the ends of a cross-linker, a homobifunctional reagent is formed. The diacylating agents include bis-imidoesters, bis-succinimidyl derivatives, di-isocyanates, di-isothiocyantes, di-sulfonyl halides, bis-nitrophenyl esters, dialdehydes and diacylazides (*7*). A few of the common representive compounds are shown in Figure 6. Of the dialdehydes, glutaraldehyde has been extensively used. While it is proposed that the reaction proceeds through a Schiff base formation, the mechanism of the reaction is far from clear (*13*).

Among the dialkylating agents are bismaleimides, bis-haloacetyl derivatives, di-alkyl halides, and bis-oxiranes (*7*). Figure 7 illustrate the structure of some of these

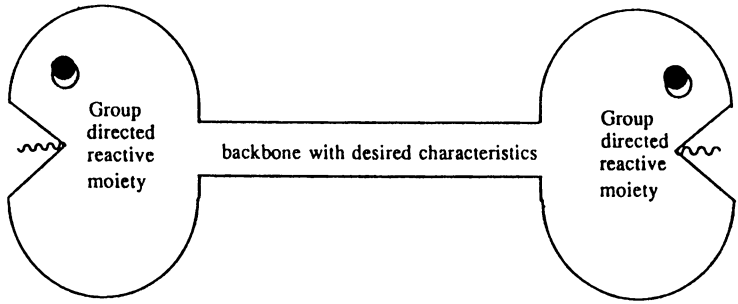


Figure 1. Schematic representation of cross-linking reagents. The reactive functional groups of these two-headed compounds are connected through a backbone. If the functional groups are identical, they are homobifunctional reagents, if different, heterobifunctional. Various structural characteristics may be incorporated into the backbone. (Reproduced with permission from ref. 7. Copyright 1991 CRC Press).

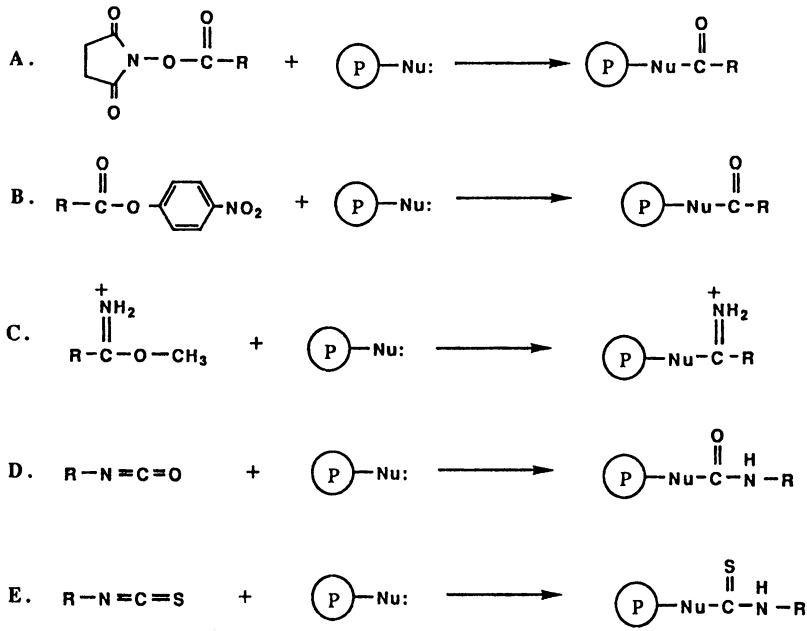


Figure 2. Commonly found acylating groups in cross-linking reagents and their reactions with nucleophiles. A. N-Hydroxysuccinimidyl ester; B. p-Nitrophenyl ester; C. Methyl imidoester; D. Isocyanate; E. Isothiocyanate.

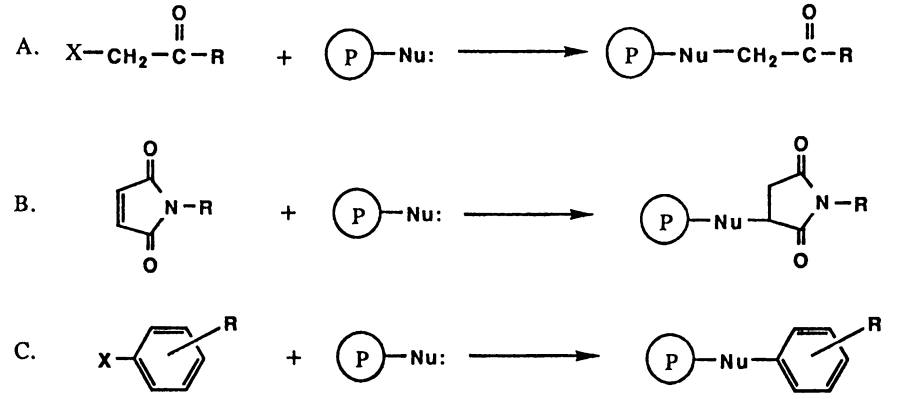


Figure 3. Alkylating groups and their reactions with nucleophiles. A. Haloacetyl group; B. N-Maleimides; C. Aryl halides.

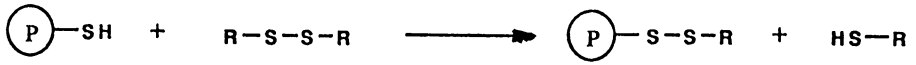


Figure 4. The disulfide interchange reaction between a thiol group of a protein and a disulfide compound.

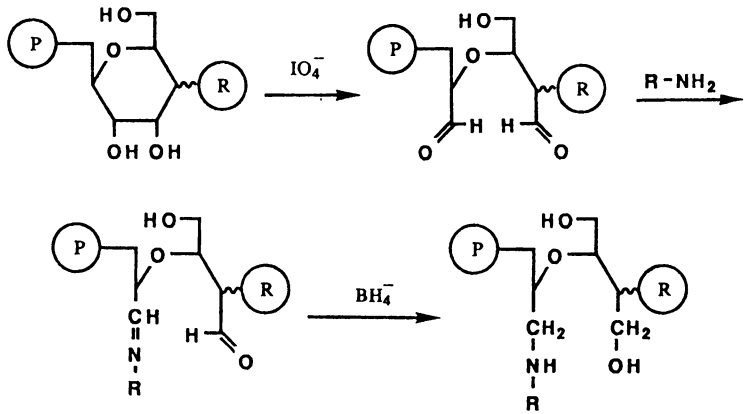


Figure 5. Reductive alkylation of an amino group with an aldehyde derived from oxidation of a sugar moiety of a glycoprotein.

compounds. As stated above, although any of the nucleophilic amino acid side-chains will react with these cross-linkers, the diacylating agents are generally thought to be amino group selective while the dialkylating agents are sulfhydryl specific. Thiol specific bis-disulfide cross-linkers have also been synthesized (*14*).

Heterobifunctional cross-linkers. Heterobifunctional cross-linkers contain two different functional groups. These two reactive functionalities can be any combination of those alkylating and acylating agents mentioned above. For example, one end of the cross-linker may be a disulfide such that it would be sulfhydryl group specific and the other end may be an acylating agent for amino group selectivity. Examples of such combinations are shown in Figure 8. Recently, the extremely reactive nitrene- and carbene-generating moieties have been incorporated to provide nondiscriminatory reactions with relatively inert proteins (*15*) (Figure 8D).

Zero-length cross-linking reagents. Zero-length cross-linkers are a special class of compounds which induce direct joining of two chemical moieties of proteins without the introduction of any extrinsic material. This is different from homo- and hetero-bifunctional reagents where a spacer is incorporated between the two cross-linked groups. Reagents that cause the formation of disulfide bonds, such as cupric di(1,10-phenanthroline), are zero-length cross-linkers. Other reagents include carbodiimides, isoxazolium derivatives, chloroformates and carbonyldiimidazole, which couple carboxyl and amino groups (*7*). These reagents react by activating the carboxyl group to form an active intermediate with which the amino group reacts as shown for carbodiimides in Figure 9.

Cleavable Reagents. In addition to the reactive head groups of the bifunctional cross-linking reagents, various functional groups have been incorporated into the backbone of these compounds to increase their versatility. For example, cleavable bonds have been incorporated to enable the cross-linked species to be separated. These include disulfide bond, vicinal glycol, azo, sulfone, ester and thioester linkages (*16-22*). Reagents including reducing and oxidizing agents, acids and bases may be used to cleave these bonds (Table I).

Hydrophobicity and Hydrophilicity. The degree of water solubility of cross-linkers may be altered by incorporating hydrophilic or hydrophobic entities into these compounds. For example, Staros (*23*) incorporated a sulfonate group onto the succinimide ring of N-hydroxysuccinimide esters to increase its hydrophilicity and decrease its membrane solubility. The inclusion of ether-oxygen, hydroxyl group, ester and amide bonds also increase the water solubility, while an increase in alkyl chain length decreases the hydrophilicity (*24*). These characteristics may be useful in certain conditions.

Size and Length. Bifunctional cross-linkers also differ in size in regard to the bulkiness of the molecule, and in length with respect to the distance between the two functional groups. The molecules can be as simple as formaldehyde (*25*) and as bulky as di[(iodoacetyl)aminomethyl]-fluorescein (*26*). The distance between the

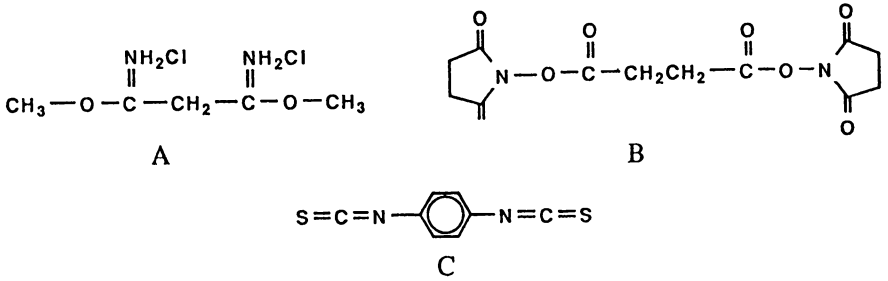


Figure 6. Examples of homobifunctional cross-linkers containing diacylating groups. A. Dimethyl malonimidate dihydrochloride; B. Bis(N-hydroxysucccinimidyl) succinate; C. 1,4-Phenylene di-isothiocyanate.

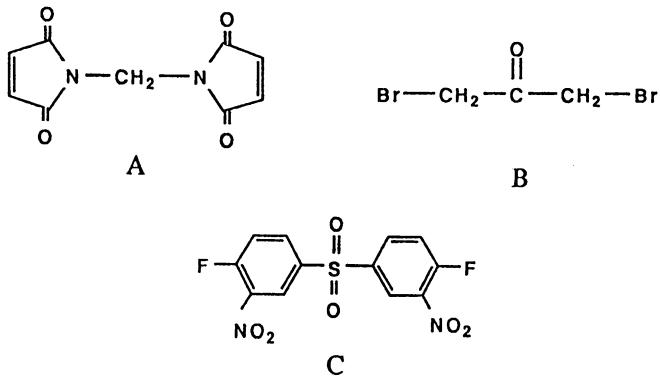


Figure 7. Examples of homobifunctional cross-linkers containing dialkylating groups. A. N, N'-Methylenebismaleimide; B. 1,3-Dibromoacetone; C. Bis(3-nitro-4-fluorophenyl)sulfone.

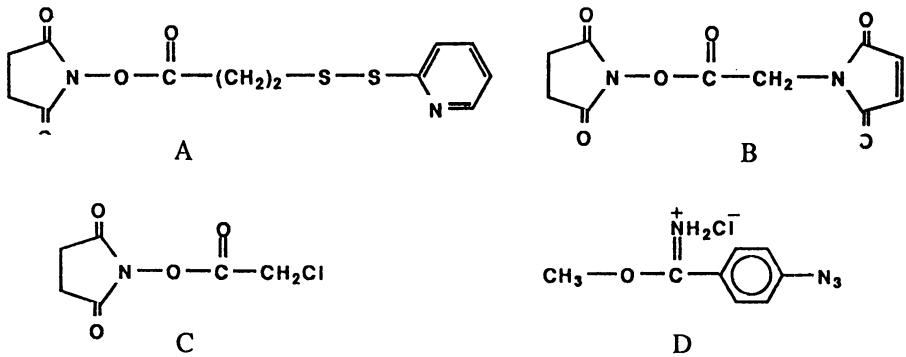


Figure 8. Examples of heterobifunctional cross-linkers. A. N-Succinimidyl 3-(2-pyridyldithio)propionate; B. N-Succinimidyl maleimidoacetate; C. N-Succinimidyliodoacetate; D. Methyl 4-azidobenzimidate.

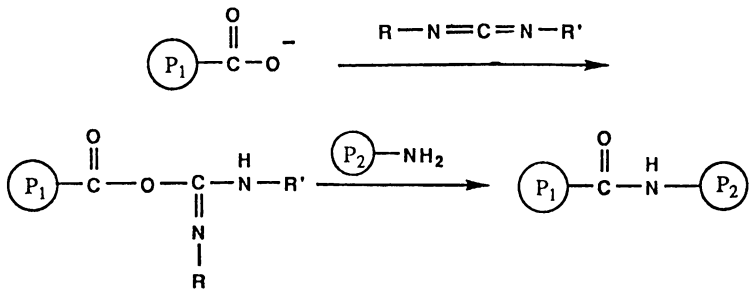


Figure 9. Cross-linking effected by a carbodiimide, a zero-length cross-linking reagent.

TABLE I
EXAMPLES OF CLEAVABLE GROUPS IN BIFUNCTIONAL REAGENTS

CLEAVABLE GROUP	CLEAVAGE CONDITION	CLEAVED PRODUCTS
$R_1-S-S-R_2$	reducing agent e.g. 2-mercaptoethanol	$R_1-SH + R_2-SH$
$R_1-CH(OH)-CH(OH)-R_2$	periodate	$R_1-CHO + R_2-CHO$
$R_1-N=N-R_2$	dithionite	$R_1-NH_2 + R_2-NH_2$
$R_1-S(=O)(=O)-R_2$	base	$R_1-OH + R_2-SO_3^-$

functional groups can be varied as well. For example, alkyl diimidoesters can be synthesized with different methylene groups giving rise to different lengths as depicted in Table II. These molecules are useful for cross-linking different groups on a protein molecule (*27*).

Probe labeled reagents. Various reporter groups have been incorporated into cross-linking reagents. As shown in Figure 10, these include spin labels, fluorescence and absorption probes as well as radioactive moieties (*28-31*). These labeled compounds have been used in the study of biological membranes and protein structures (*7*).

Stability of proteins by cross-linking to other components

Because the forces that contribute to the structural integrity of proteins are multi-factorial, prediction of the shelf-life of a particular protein in its native state is virtually impossible (*32-35*). However the stabilities of proteins and enzymes can be artificially modified. Of the many methods available, chemical cross-linking has recently been used to stabilize proteins. The rational of such an approach is based on the hypothesis that inactivation of a protein is due to unfolding of its structure and that intra- or inter-molecular cross-linking may rigidify the molecule, thus retard conformational changes towards the denatured state caused by external chemical, thermal or mechanical forces (*36-37*). In this context, chemical bifunctional reagents have been used to stabilize proteins by intramolecular cross-linking or by coupling the protein to other components, such as other solid supports or soluble polymers. The following sections will demonstrate some of the characteristics.

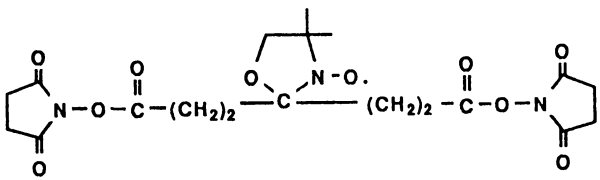
Stabilization by immobilization. Protein immobilization has received great attention because of its important applications in medical and clinical analysis, affinity chromatography, and synthetic chemistry (*38-41*). In addition to the characteristics of solid state chemistry, immobilization of a protein, in many circumstances, has increased the stability of the molecule (*42*). Virtually any cross-linking reagents may be used to couple proteins to solid supports and the various methods have been reviewed extensively (*7,38-43*). We have, for example, coupled galactosyltransferase to CL-Sepharose 4B with 6-aminocaproic acid through N-hydroxysuccinimidyl activation by dicyclohexyl carbodiimide (*44*). As shown in Figure 11, the immobilized enzyme has demonstrated increased stability towards both mechanical and thermal denaturation. The solid-bound enzyme was stable over 10 months at 4°C with the energy of inactivation raised to 30 Kcal/mol from 13 Kcal/mol for the soluble enzyme (*45*). Many other enzymes have demonstrated the same phenomenon of increased stability on immobilization.

Intermolecular cross-linking. Cross-linking of proteins to other soluble components may also provide the reticulating effect, strengtening the molecular structure. This is analogous to immobilization as discussed above, except that the components remain soluble in the aqueous medium. In many occassions, such arrangements also stabilize the protein. Thus, when horseradish peroxidase was cross-linked to immunoglobulin G or Jacalin, a plant lectin, with glutaraldehyde, it becomes more

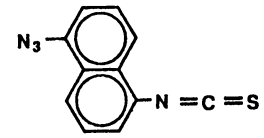
TABLE II

DI-IMIDOESTERS OF DIFFERENT CHAIN LENGTHS

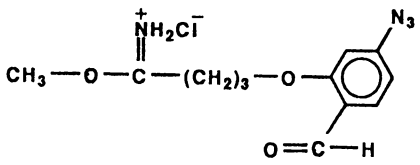
Cross-Linker	Maximum Distance (Å)
$CH_3-O-C(=NH_2Cl)-CH_2-C(=NH_2Cl)-O-CH_3$	5
$CH_3-O-C(=NH_2Cl)-(CH_2)_2-C(=NH_2Cl)-O-CH_3$	6
$CH_3-O-C(=NH_2Cl)-(CH_2)_4-C(=NH_2Cl)-O-CH_3$	9
$CH_3-O-C(=NH_2Cl)-(CH_2)_5-C(=NH_2Cl)-O-CH_3$	10
$CH_3-O-C(=NH_2Cl)-(CH_2)_6-C(=NH_2Cl)-O-CH_3$	11
$CH_3-O-C(=NH_2Cl)-(CH_2)_8-C(=NH_2Cl)-O-CH_3$	14



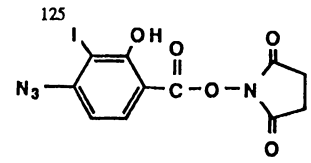
A



B



C



D

Figure 10. Examples of cross-linkers with different probes. A. Spin label; B. Fluorescence; C. Absorption; and D. Radioactive.

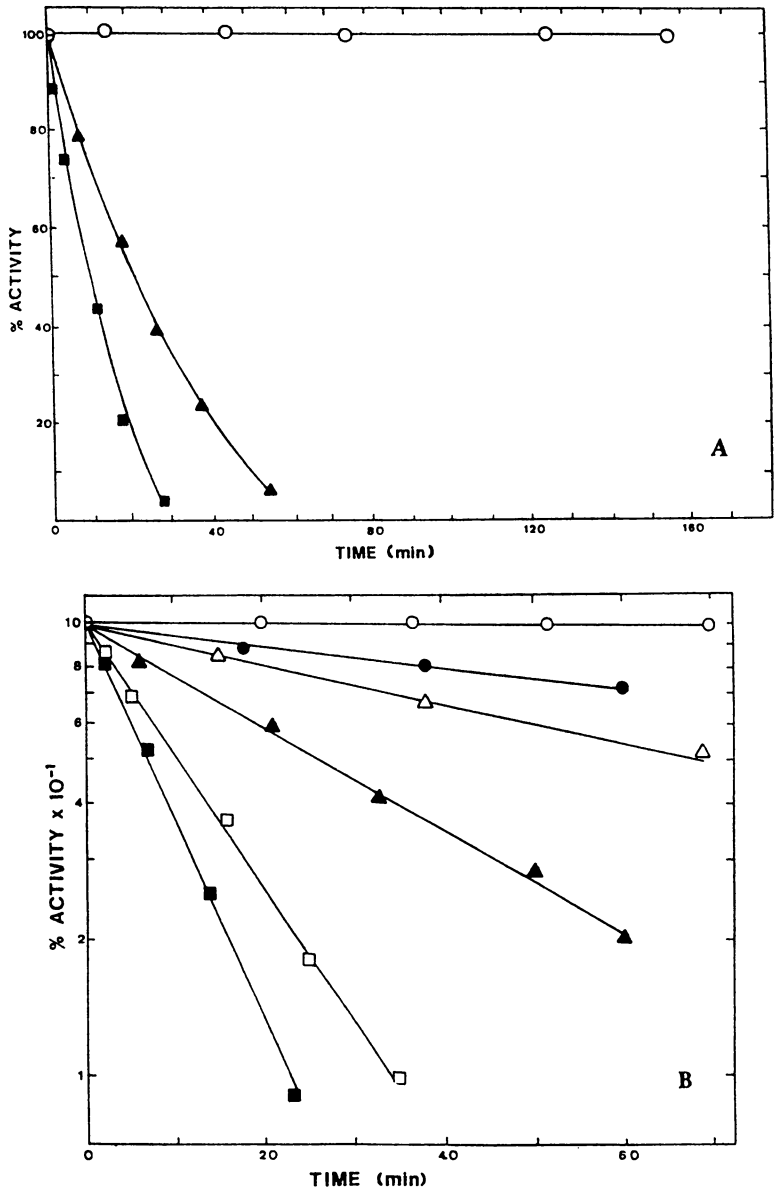


Figure 11. Increased stability of galactosyltransferase immobilized on Sepharose. A. Mechanical stability: O, immobilized enzyme; ▲, free enzyme; ■, free enzyme with Sepharose beads. B. Thermal stability: Free (closed symbols) and immobilized (open symbols) enzyme were inactivated at 75°C (■, □); 60°C, (▲, △); 30°C (●, ○). (Reproduced with permission from ref. 45. Copyright 1985 Academic Press).

stable towards thermal denaturation with an increase in Arrhenius energy of inactivation as shown in Table III.

TABLE III

ENERGY OF INACTIVATION OF HORSERADISH PEROXIDASE CONJUGATES

Conjugates	E_a (Kcal/mol)
Peroxidase alone	35
Peroxidase-IgG	51
Peroxidase-Jacalin	43

The conjugates were prepared with glutaraldehyde. (Reproduced with permission from ref. 7. Copyright 1991 CRC Press)

Intramolecular cross-linking. Intramolecular cross-linking with bi- or poly-functional reagents braces the molecular structure in the native state and has been shown to increase the chemical, thermal, and mechanical stability of many enzymes (*46-50*). For example, Himmel et al. (*50*) used a series of bifunctional reagents of different functional groups and lengths to intramolecularly cross-link amyloglucosidase. Although not all reagents gave the same results, some chemical cross-linkers, for example some of the diimidoesters, enhanced the thermal stability of the enzyme. As shown in Figure 12, the enzyme modified with either dimethyl succinimidate, dimethyl malonimidate, dimethyl suberimidate or dimethyl pimelimidate, retained a much higher activity after incubation for one hour at various temperatures. It was reported that in some cases the half-life of inactivation is more than doubled when incubated at 65°C (*50*).

Prediction of Protein Stability

An increase or a decrease in stability of a protein after cross-linking intramolecularly or intermolecularly to another component may be determined by studying its susceptibility to denaturation by thermal, mechanical or chemical disturbances. However, its half-life may be so prolonged that its determination may not be possible. This is particularly so for thermal denaturation at low temperatures where the half-life may be extended to months or years. In this case, the thermal stability may be inferred from the heat of denaturation obtainable by use of the Arrhenius equation:

$$\ln k = A - E_a/RT$$

When the natural logarithm of the rate constants of inactivation, k, is plotted against the reciprocal of the absolute temperatures at which the inactivation takes place, the negative slope of the line provides the value of $-E_a/R$ where R is the known gas

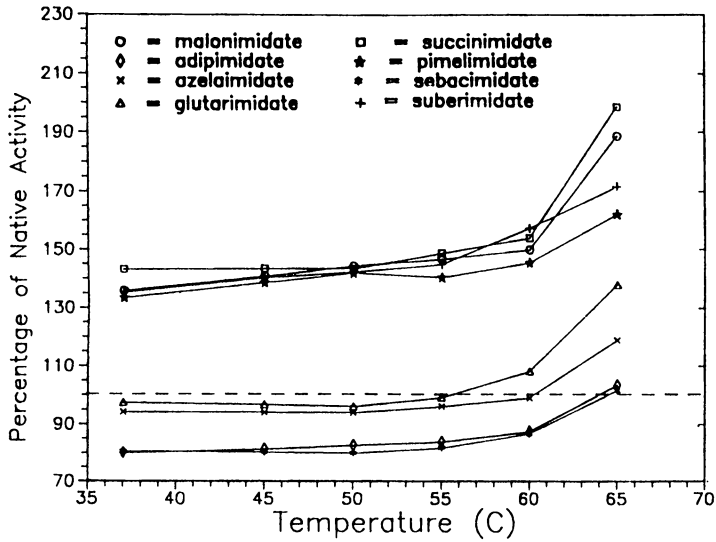


Figure 12. Increased stability of intramolecularly cross-linked glucoamylase. Glucoamylase cross-linked by various indicated reagents was preincubated at the indicated temperatures for one hour before assay. The native enzyme activity was taken as 100%. (Reproduced with permission from ref. 50. Copyright 1989 Humana Press).

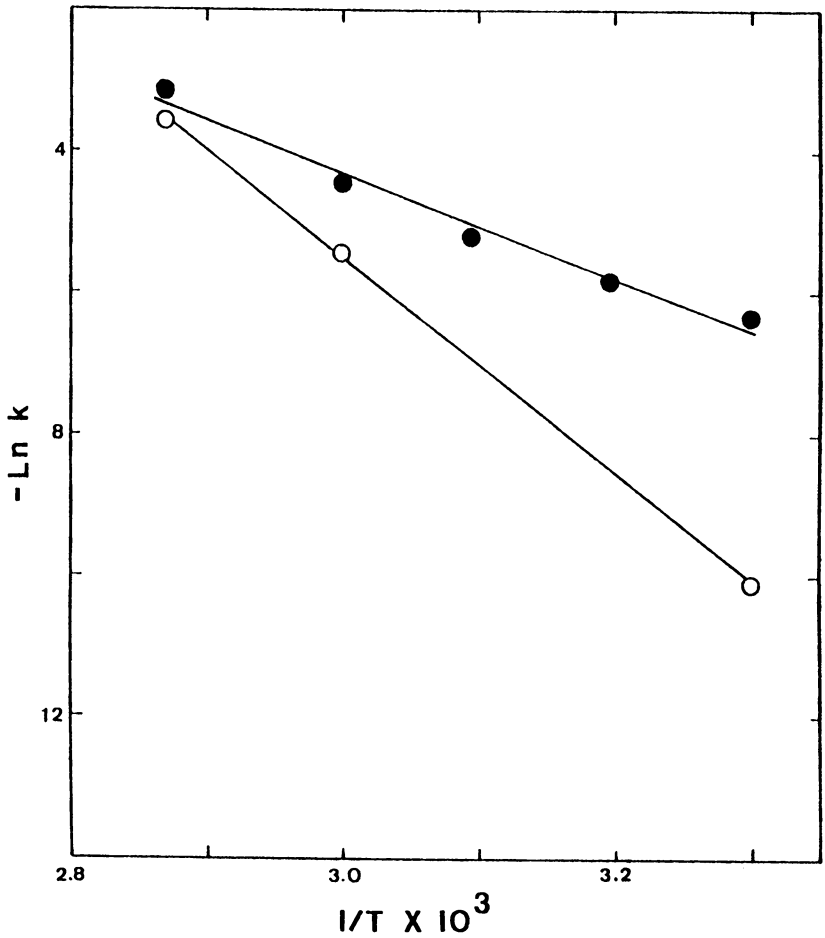


Figure 13. Arrhenius plot of inactivation of galactosyltransferase either free (closed circles) or immobilized on to Sepharose (open circles) (Reproduced with permission from ref. 45. Copyright 1985 Academic Press).

constant. The larger the value of E_a, the more energy is required for inactivation, thus the more stable is the protein. This is examplified by the denaturation of galactosyltransferase shown in Figure 13. The native enzyme has an E_a of 13 kcal/mol, whereas that of the immobilized enzyme is 30 kcal/mol.

The rate constant of inactivation of the protein at certain temperatures may be determined by measuring the rate of inactivation at that temperature. In most cases, the denaturation is a first-order process and the first order rate constant may be obtained by plotting the logarithm of the initial activity, v, against the time of incubation, t, at that particular temperature according to:

$$\ln v = C - kt$$

To evaluate the half-life of crosss-linked protein at a certain temperature, the rate constant of inactivation at that temperature is determined by extrapolation of the Arrhenius plot. The half-life can then be calculated by the following equation:

$$t_{1/2} = 0.693/k$$

Synopsis.

Although not all cross-linked proteins result in stabilization of the biological activity, it has been firmly established that both intra- and inter-molecular cross-linking can confer stability to the molecule. Where no stability or decreased stability is observed, it is possible that the reagents used may be inappropriate. A different cross-linking reagent may give a totally opposite result, as has been demonstrated by Tatsumoto et al. (*50*). Since there are so many cross-linkers available or can be synthesized, it may be safe to prostulate that there is always a cross-linking reagent that can be used to stablize a protein.

Literature Cited.

1. Ji, T. H. *Methods Enzymol.* **1983**, *91*, 508-609.
2. Han, K.-K.; Richard, C.; Delacourte, A. *Int. J. Biochem.* **1984**, *16*, 129-144.
3. Gaffney, B. J. *Biochim. Biophys. Acta* **1985**, *822*, 289-317.
4. Tamura, M.; Tamura, T.; Burnham, D. N.; Uhlinger, D. J.; Lambeth, J. D. *Arch. Biochem. Biophys.* **1989**, *275*, 23-32.
5. Wawrzynczak, E. J.; Thorpe, P. E.; in *Immunoconjugates, Antibody Conjugates in Radioimaging and Therapy of Cancer*; Vogel, E.-W.; Ed.; Oxford University Press: New York, NY, 1987, p. 28.
6. Nonisotopic Immunoassay; Ngo, T. T., Ed.; Plenum Press, New York, N.Y., 1988.
7. Wong, S.S. *Chemistry of protein conjugation and cross-linking*; CRC Press: Boca Raton, FL, 1991.
8. Means, G. E.; Feeney, R. E. *Chemical Modification of Proteins*; Holden-Day, San Francisco, CA, 1971.
9. Lundblad, R. L.; Noyes, C. M.; *Chemical Reagents for Protein Modification*; CRC Press: Boca Raton, FL, 1984; Vols. 1 and 2.

10. Edwards, J. O.; Pearson, R. G. *J. Am. Chem. Soc.* **1962**, *84*, 16-26.
11. Kimura, T.; Matsueda, R.; Nakagawa, Y.; Kaiser, E. T.; *Analyt. Biochem.* **1982**, *122*, 274-282.
12. Cabacungan, J. C.; Ahmed, A. I.; Feeney, R. E.; *Anal. Biochem.* **1982**, *124*, 272-278.
13. Hardy, P. M.; Nicholls, A. C.; Rydon, H. N.; *J. Chem. Soc. Perkin Trans.* **1976**, *1*, 958-962.
14. Bloxham, D. P.; Sharma, R. P. *Biochem. J.* **1979**, *181*, 355-366.
15. Bayley, H.; Knowles, J. R. *Methods Enzymol.* **1977**, *46*, 69-114.
16. Traut, R. R.; Bollen, A.; Sun, T. T.; Hershey, J. W. B.; Sundberg, J.; Pierce, L. R. *Biochemistry* **1973**, *12*, 3266-3273.
17. Webb, J. L. *Enzyme and Metabolic Inhibitors*; Academic Press: New York, N.Y., 1966; Vol. 2, pp. 729.
18. Smith, R. J.; Capaldi, R. A.; Muchmore, D.; Dahlquist, F. *Biochemistry* **1978**, *17*, 3719-3723.
19. Jaffe, C. L.; Lis, H.; Sharon, N. *Biochemistry* **1980**, *19*, 4423-4429.
20. Wold, F. *Methods Enzymol.* **1972**, *25*, 623-651.
21. Sato, S.; Nakao, M. *J. Biochem.* **1981**, *90*, 1177-1185.
22. Friebel, K.; Huth, H.; Jany, K. D.; Trummer, W. E. *Z. Physiol. Chem.* **1981**, *362*, 421-428.
23. Staros, J. V. *Biochemistry* **1982**, *21*, 3950-3955.
24. Fasold, H.; Bäumert, H.; Fink, G. in *Protein Cross-linking: Biochemical and Molecular Aspects*; Friedman, M., Ed.; Plenum Press: New York, NY, 1976; pp. 207.
25. Hopwood, D. *Histochemie* **1969**, *17*, 151-161.
26. Haugland, R. P. *Handbook of Fluorescent Probes and Research Chemicals*, Molecular Probes, Inc.: Eugene, Oregon, 1990, p. 22.
27. Ji, T. H. *J. Biol. Chem.* **1974**, *249*, 7841-7847.
28. Gonzalez-Ros, J. M.; Farach, M. C.; Martinez-Carrion, M. *Biochemistry* **1983**, *22*, 3807-3811.
29. Gaffney, B. J.; Willingham, G. L.; Schepp, R. S. *Biochemistry* **1983**, *22*, 881-892.
30. Ji, T. H.; Ji, I. *Analyt. Biochem.* **1982**, *121*, 286-289.
31. Sigrist, H.; Allegrini, P. R.; Kempf, C.; Schnippering, C.; Zahler, P. *Eur. J. Biochem.* **1982**, *125*, 197-201.
32. Matthew, J. B.; Gurd, F. R. N. *Methods Enzymol.* **1986**, *130*, 437-453.
33. Chothia, C. *Ann. Rev. Biochem.* **1984**, *53*, 537-572.
34. Alber, T. *Ann. Rev. Biochem.* **1989**, *58*, 765-798.
35. Dill, K. A.; Shortle, D. *Ann. Rev. Biochem.* **1991**, *60*, 795-825.
36. Martinek, K.; Klibanov, A. M.; Goldmacher, V. S., Berezin, I. V. *Biochim. Biophys. Acta* **1977**, *485*, 1-12.
37. Torchilin, V. P.; Maksimenko, A. V.; Smirnov, V. N.; Berenzin, I. V.; Klibanov, A. M.; Martinek, K. *Biochim. Biophys. Acta* **1978**, *522*, 277-283.
38. *Medical Applications of Immobilized Enzymes and Proteins*; Chang, T. M. S., Ed.; Plenum Press: New York, NY, 1977; vols. I and II.
39. *Immobilized Enzymes in Food Processing*; Pitcher, W. H., Jr., Ed.; CRC Press: Boca Raton, FL, 1979.

40. Kennedy, J. F.; Cabral, J. M. S. *Immmobilized enzymes in solid phase biochemistry*, Scouten, W. H., Ed., Wiley Interscience: New York, NY, 1983, pp 253.
41. Mosbach, K. *Ciba Found. Symp.* **1985**, *111*, 57-70.
42. Mohr, P.; Pommerening, K. *Affinity Chromatography*, Marcel Dekker: New York, NY, 1985.
43. *Affinity Chromatography*; Dean, P. D. G.; Johnson, W. S.; Middle, F. A., Eds.; IRL Press: Washington, D. C., 1985.
44. Lee, T. K.; Wong, L-J. C.; Wong, S. S. *J. Biol. Chem.* **1983**, *258*, 13166-13171.
45. Demers, A. G.; Wong, S. S. *J. Appl. Biochem.* **1985**, *7*, 122-125.
46. Torchilin, V. P.; Trubetskoy, V. S.; Omel'yanenko, V. G.; Martinek, K. *J. Mol. Catal.* **1983**, *19*, 291-303.
47. Trubetskoy, V. S.; Torchilin, V. P. *Int. J. Biochem.* **1985**, *17*, 661-663.
48. Torchilin, V. P.; Trubetskoy, V. S. *Ann. N. Y. Acad. Sci.* **1984**, *434*, 27-30.
49. Gottschalk, N.; Jaenicke, R. *Biotechnol. Appl. Biochem.* **1987**, *9*, 389-400.
50. Tatsumoto, K.; Oh, K. K.; Baker, J.O., Himmel, M. E. *Appl. Biochem. Biotechnol.* **1989**, *20*, 293-308.

RECEIVED March 31, 1992

Chapter 23

Glutaraldehyde Cross-Linking

Fast and Slow Modes

Timothy J. A. Johnson

Department of Anatomy and Neurobiology, Colorado State University, Fort Collins, CO 80523

Intense activity in molecular cytochemistry has rekindled interest in questions regarding aldehyde chemistry in the fixation of biological specimens. The chemistry of glutaraldehyde crosslinking has been examined numerous times. The reaction of glutaraldehyde with common low molecular weight nucleophiles such as amino acids and sulfhydryl compounds, which are frequently encountered in biological systems, generates a wide range of products. The complexity of the condensation product mixture has thwarted a thorough understanding of crosslinking mechanisms. Despite this, it is useful to consider the condensation products of glutaraldehyde with itself and other low molecular weight precursors as the actual crosslinking structures. With that emphasis, this paper focuses on: 1) the molecular weight range of the condensation products as a function of available precursors, and 2) the kinetics of formation of the products.

Glutaraldehyde crosslinking has been used since the early 1960's to immobilize constituents of cells and tissues for subsequent examination by microscopy (*1*). Glutaraldehyde is often employed for linking affinity ligands to solid state supports for affinity chromatography (*2*). In industry glutaraldehyde can effectively render enzymes stable to heat or other denaturing influences (*3*). Stabilized enzymes are useful for repeated use in bulk processing methods.

Cells add enormous complexity because they contain a diverse biochemical milieu that is reactive with aldehydes. Aldehyde-reactive molecules in cells range from molecular masses of a few hundred daltons to greater than 10^6 daltons. The reactive molecules present themselves in states which range from individual soluble molecules to hydrated "solid state" structures. The concentrations of the reactive structures vary from cell to cell and even from one part of a cell to another.

0097–6156/93/0516–0283$06.00/0

In the preparation of glutaraldehyde stabilized enzymes, the focus is simplified because one can purify the reactive structures (enzymes) and define their concentrations and physical state. The goal of stabilization is to immobilize denaturable structural elements. At the same time, one cannot block access to substrate binding sites by direct reaction with the sites or by burying the sites under excess crosslinking rope. The enzyme assay is an effective monitor of substrate accessibility to its binding site and product egress from the enzyme.

Two condensation polymerization schemes for glutaraldehyde crosslinking are examined in this discussion. The goal will be to give the user some practical knowledge for controlling the polymerization or crosslinking process to produce active and stable enzymes.

One of the crosslinking schemes, aldol condensation, is a slow process. The second crosslinking scheme is rapid by comparison. In the presence of primary amine, glutaraldehyde reacts to form polymers of pyridine. The size of the products of the latter chemistry, and thus the length of potential crosslinkers, can be controlled by the ratio of glutaraldehyde to amine.

Aldehyde-Amine Chemistry

Acid Producing Reactions. Before considering aldehyde-amine crosslinking chemistry, it is useful to attend to the product often overlooked in the reaction of glutaraldehyde with amines. Acid is produced when aldehydes and amines combine rapidly to form complexes which have pK_a values that are far lower than those for the parent amines (Fig. 1) (*4-7*, additional references included in ref. *7*). Proteins, of course, have many primary amino groups. But more important, if one chooses to use the rapid crosslinking mode by adding low molecular weight primary amines to the crosslinking mixture, as outlined below, acid production becomes significant (Fig 1). Thus, adding sufficient buffering capacity to the reaction mixture is suggested (Table 1).

The formation of acid is the primary reason for the use of buffers (*7*). Acid production suggests strategies for the choosing a buffer. First, buffers should not be used if they have primary amine moieties which react with aldehydes (e.g., Tris) (*7*). Tertiary amine based buffers (e.g., MOPS and HEPES) do not combine with aldehydes to produce acid and are excellent possibilities (*7*). Second, since the buffering capacity of a buffer is greatest at the pK_a, the buffer chosen should have a pK_a slightly less (0.2-0.3 pH units) than the pH needed for the experiment. Third, there should be sufficient buffering capacity to minimize the pH decrease in the reaction mixture (Table 1). Buffering equivalents can be added by increasing the concentration and/or volume of the crosslinking buffer. This consideration becomes important if one needs to use a buffer with pKa further from the experimental pH. Possible buffer choices are listed in Table 2.

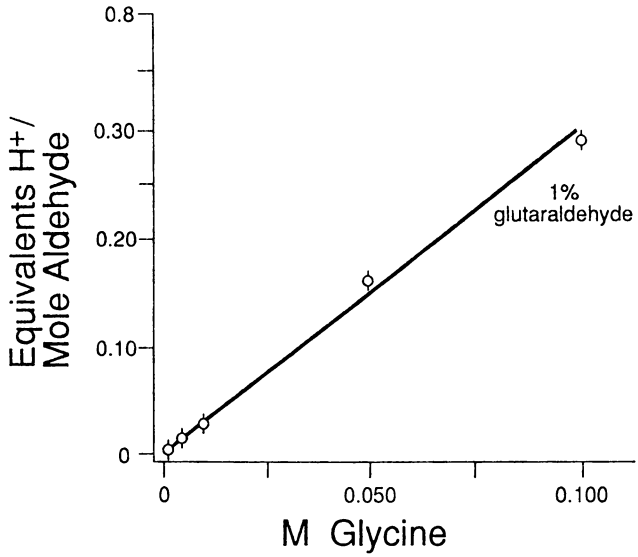


Figure 1. Acid production in glutaraldehyde-glycine reactions.

Table 1
Glutaraldehyde Induced pH Changes in Buffered Glycine[1]

Buffer	Phosphate	MOPS[2]	Cacodylate
0 M	4.50	4.50	4.50
0.05 M	6.90	6.90	6.30
0.20 M	7.20	7.25	6.90

[1]The pH was read 5 minutes after glutaraldehyde was added to buffered 0.05 M glycine. The final glutaraldehyde was 1% (0.1 M). The initial pH was adjusted to 7.40.

[2]3-N-morpholinopropane sulfonic acid.

Table 2
Alternative Buffers[1]

Buffer	pK_a (37°C)
acetate	4.75
MES[2]	5.96
cacodylate	6.19
carbonate	6.37 (pK_{a1})
PIPES[3]	6.66
MOPSO[2]	6.75
BES[3]	6.88
MOPS[2]	7.10
phosphate	7.21 (pK_{a2})
HEPES[3]	7.31
POPSO[3]	7.63
HEPPSO[3]	7.73
HEPPS[3]	7.88
borate	9.23

[1]Alternative buffers for glutaraldehyde crosslinking. The acronyms used for buffers are described in reference 23. The zwitterionic buffers described by Good have pK_a values that are much more sensitive to temperature than those of common buffers like phosphate (*7*, *23*).

[2]Related to group B tertiary amines (*6*). Negligible amounts of acid expected with glutaraldehyde.

[3]Related to group A tertiary amines (*6*). Small amounts of acid are produced with glutaraldehyde.

Aldehyde-amine Crosslinking. Early notions about glutaraldehyde crosslinking of cells and proteins suggested that the dialdehyde formed Schiff bases with two different amino groups. The glutaraldehyde carbon chain thus bridged the amine moieties. Richards and Knowles (*8*) diminished the importance of the double Schiff base concept of crosslinking by pointing out that the irreversible crosslink formed with glutaraldehyde was not compatible with the reversibility of Schiff base formation. Richards and Knowles then noted that self condensation between aldehyde molecules generated α,β-unsaturated aldehydes (Fig. 2). Continued condensation yields products with two or more α,β-unsaturated aldehyde sites. These authors proposed that aldol condensation products of glutaraldehyde provided the "glue" for crosslinking in the sense that nucleophiles add to α,β-unsaturated aldehydes irreversibly (Fig. 2).

The proposal of Richards and Knowles stimulated significant efforts to determine the crosslinking chemistry (*9-13*). Others examined the nature of the crosslink and found that if other precursors are present, namely primary amines, a complex set of pyridine products are formed (*14-17*). The pyridine products provide the structural "glue" for crosslinking.

The present work investigates and discusses glutaraldehyde crosslinking in terms of the precursors and the size of the polymer products generated. The products are regarded as the actual crosslinking molecules. Glutaraldehyde based polymers assemble either by slow or by rapid kinetic processes. Examples are presented of polymer growth in the presence of other precursors in addition to glutaraldehyde.

Slow Crosslinking. The slow crosslinking process is characterized by aldol condensation reactions between glutaraldehyde molecules. The condensation products can be monitored at ~235 nm which is due to the formation of α,β-unsaturated aldol products (*8, 18, 19*). In neutral pH buffer glutaraldehyde condenses with itself, albeit slowly, at temperatures ranging from 0-40°C and pressures of 600-760 mm Hg. Rasmussen explored many parameters of glutaraldehyde self condensation (*19*). He demonstrated the increase in condensation rates with increases in the concentration of aldehyde, temperature, and pH (*18, 19*). Rasmussen did not report the size range of the condensation products.

In the present study, glutaraldehyde self-condensation products in a 0.2 M (2%) glutaraldehyde solution containing 0.05 M MOPS (pH 7.4) were monitored for their presence by absorption at 235 nm and for size by Sephadex G-25 gel filtration. Little change occurs at 235 nm for 60 minutes. After 20 hours at 40°C, a significant increase in adsorption at 235 nm occurs (Fig. 3). The sizes of the condensation products at 20 hours are not significantly larger than glutaraldehyde itself (Fig. 4). That is, the products do not fall within the sieving range of a Sephadex G-25 gel column. The molecular mass range of the products is estimated as between 100 and 1000 daltons.

The increase in adsorption at 235 nm suggests that glutaraldehyde molecules have condensed with each other in 20 hours at 40°C. The gel filtration profile reveals that the self-condensation crosslinking products grow slowly and remain small even after significant time periods. In terms of stabilizing protein molecules by intramolecular crosslinking with glutaraldehyde alone, only amino and sulfhydryl groups, that are proximal in space, are likely to be crosslinked by aldol condensation products.

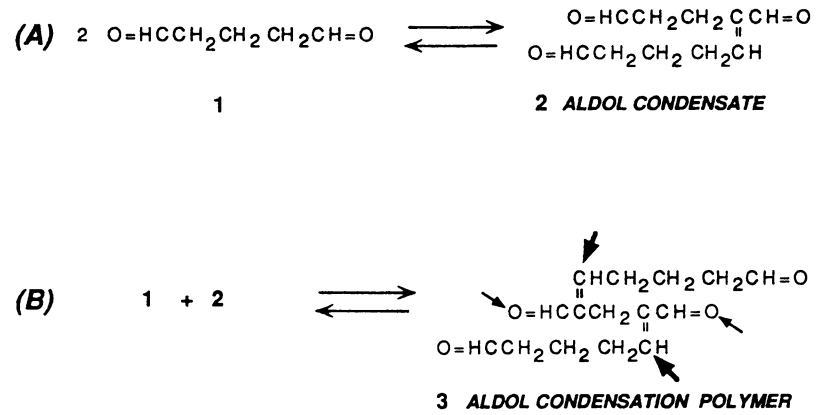


Figure 2. Aldol condensation products of glutaraldehyde. Bimolecular condensation yields the simple condensate 2, and further condensation yields 3 and higher order polymers. Nucleophilic groups such as primary amines react irreversibly with the α,β-unsaturated carbonyl groups forming Michael adducts (large arrowheads) (*8*). Reaction at the carbonyl carbon yield unsaturated Schiff bases (small arrowheads) (*12*). Higher order polymers provide multiple sites for crosslinking.

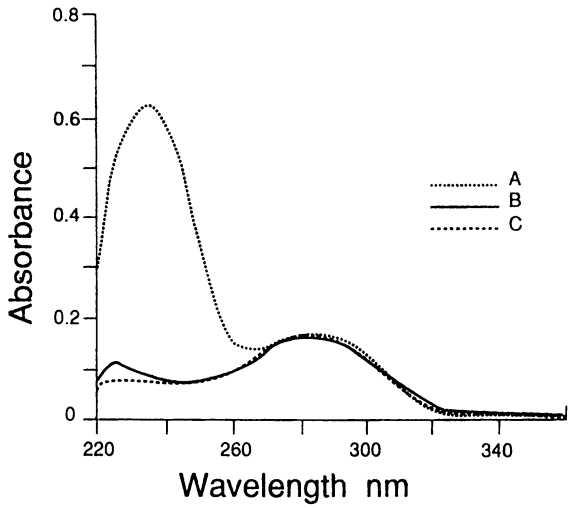


Figure 3. Self condensation products of glutaraldehyde. UV spectra. At 40°C for the time period 0 to 10 minutes there is virtually no change in the UV spectra of 2% glutaraldehyde in 50 mM MOPS (curves represented by A). Curves B and C represent 60 minutes and 20 hours respectively. The spectra were taken on an 8451 Hewlett Packard diode array spectrophotometer in a 1 cm path cell. Samples were diluted with 9 volumes of 50 mM MOPS prior to observing the spectra at 24°C.

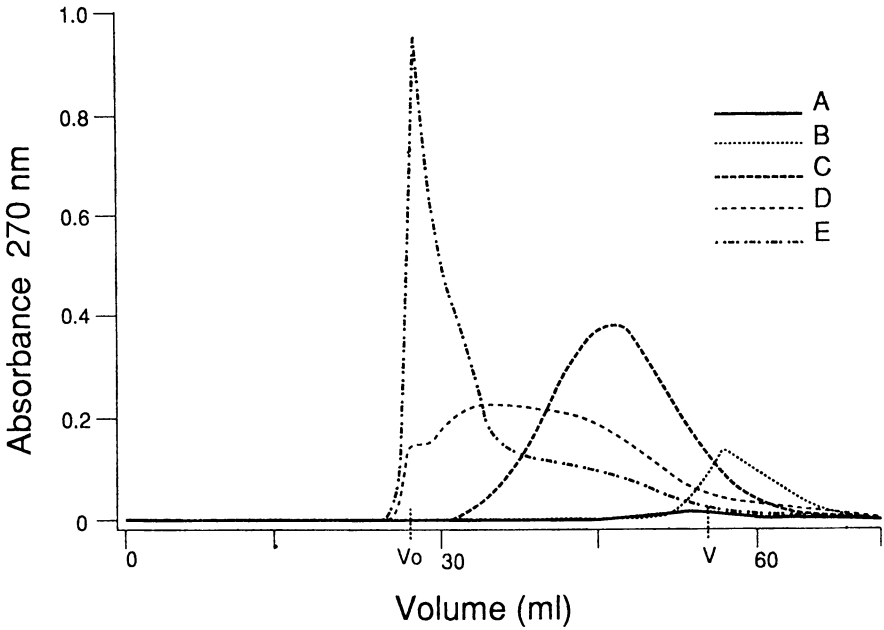


Figure 4. Gel filtration of glutaraldehyde based condensation products on Sephadex G-25. Products of glutaraldehyde self-condensation and condensation with primary amines were chromatographed on a column of Sephadex G-25 (1.41 cm^2 x 48.4 cm) at 30 ml/hr. The column was equilibrated with gel buffer (0.01 M Na acetate, 0.20 M acetic acid, and 0.20 M NaCl, pH 3.4). The column effluent was monitored at 270 nm (5 mm path) with the exception noted using an ISCO V4 absorbance detector. The column liquid volume is marked with V and the excluded volume is marked with V_o. These column parameters were determined with 5' adenosine monophosphate (MW = 347) and bovine serum albumin (MW = 66,000) respectively.

At designated times the reaction products were diluted in 19 volumes of gel buffer except for the products represented by curve B which were diluted with 4 volumes of gel buffer. The experiments represented are as follows: curve A, glutaraldehyde 0 time control; curve B, 0.2 M glutaraldehyde self-condensation (20 hrs, 40°C, see Fig. 3, column effluent monitored at 240 nm); curve C, 0.20 M glutaraldehyde and 0.10 M glycine (5 min., 37°C); curve D, 0.20 M glutaraldehyde, 0.05 M glycine, and 0.025 M lysine (5 min., 37°C); and curve E, 0.20 M glutaraldehyde and 0.05 M lysine (5 min., 37°C).

Rapid Crosslinking. The rapid crosslinking process requires glutaraldehyde and other precursors, namely, low molecular weight primary amines. When the latter are added to glutaraldehyde solutions, substituted pyridine products in a range of sizes are rapidly generated. These products are the crosslinking bridges. One can control the product size by the kind of primary amine added and by the stoichiometric ratio of glutaraldehyde to primary amine (*14, 17*).

The addition of primary amine precursors to glutaraldehyde gives rise to pyridines and polypyridines (Fig. 5) (*14-17*). In a reaction of methylamine with glutaraldehyde, the intermediate cyclic iminium ion has been trapped and recovered in good yield by selective reduction (*20*). The polymer products are 3 dimensional networks because branching is possible at any aldehyde site. Accompanying the process is a rapid increase in absorbance at 265-270 nm (*14-17*). Only a short reaction time (ranging from seconds to minutes) is required to generate structures far larger than glutaraldehyde itself.

Some examples of glutaraldehyde-amine reactions demonstrate the size range of the pyridine products. Glutaraldehyde (0.20 M) and glycine (0.10 M) in 0.20 M MOPS yield products within one minute at 37°C that are large enough that they are partially excluded from the pores of G-25 (data not shown). The profile of 5 minute products is shown in Fig. 4. Products range in size from nearly excluded to smaller structures completely able to enter the gel structure. The apparent molecular masses of the products therefore range from a few hundred daltons to less than 5000. By 5 minutes, the rapid phase of growth is over. There is some growth of products in the period from 5-30 minutes according to the Sephadex G-25 chromatographic profile of the reaction products. After 30 minutes there is virtually no change after many hours at 37°C (data not shown).

Some products of the reaction of 0.20 M glutaraldehyde and 0.05 M lysine in MOPS buffer at 37°C are excluded by one minute, and by 5 minutes, excluded products predominate (Fig. 4). Again, some consolidation into larger structures occurs by 30 minutes and beyond. Most product growth takes place in the first few minutes.

Products intermediate in size range can be obtained by adding glutaraldehyde (0.20 M) to a mixture of glycine (0.05 M) and lysine (0.025 M) again in MOPS buffer. The gel filtration profile of 5 minute reaction products at 37°C shows a wide range of products from excluded products to products barely entering the gel (Fig. 3, curve C). The mixed amine products consolidated substantially between 5-30 minutes. At 30 minutes the larger products predominate.

The formation of pyridines is accompanied by a rapid uptake of oxygen as dihydropyridines are oxidized to pyridines (*21*).

Crosslinking, Enzyme Stabilization, and Cytochemistry. Crosslinking for the purposes of enzyme stabilization or cytochemistry requires a balance between immobilization and access to enzymes and antigens. Some immobilization is necessary for ultrastructure and enzyme stabilization. Excessive immobilization in the form of crosslinking "rope" will restrict access to epitopes by antibodies or access to enzyme catalytic sites by substrates.

The main elements of manipulative control of crosslinking with glutaraldehyde are: 1) concentration of aldehyde, 2) addition of mono- and di- primary amines, and 3) time.

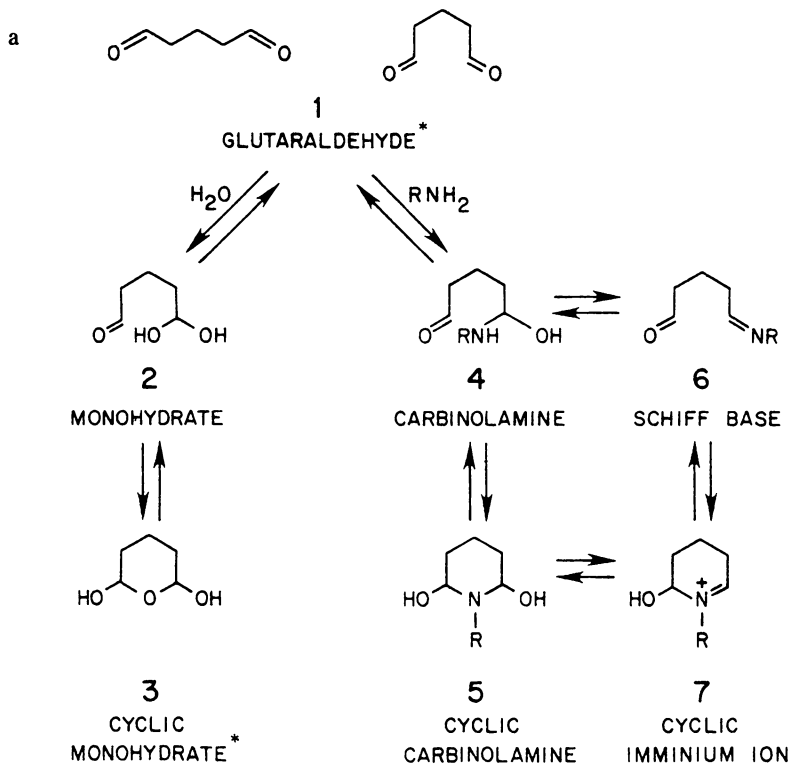
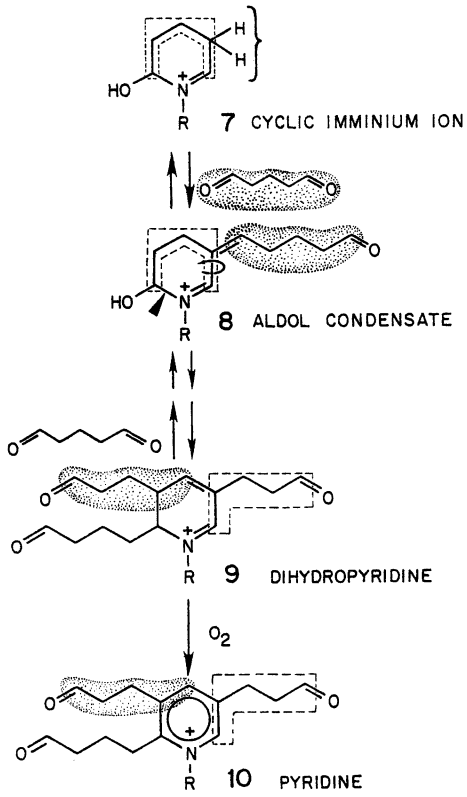


Figure 5. Pathways for obtaining pyridines from glutaraldehyde-amine precursors. (5a) Glutaraldehyde cyclization. In water cyclic structures 3 represent >80% of glutaraldehyde (*20*). Glutaraldehyde cyclizes in the presence of primary amines. The cyclic iminium ion, 7, has been trapped in good yield (*20*, see text). (5b) Low molecular weight pyridines. Abstraction of a proton from 7 facilitates condensation with an additional glutaraldehyde molecule. Dihydropyridine 9 is obtained after another glutaraldehyde molecule condenses. Pyridine 10 is afforded after facile oxidation of 9. (5c) Pyridine polymers. If the last glutaraldehyde molecule in the scheme is replaced by a molecule with a single aldehyde function such as 8, polymers can be generated. This scheme suggests that the ratio of glutaraldehyde/amine is an important determinant of product size (*17*).

Continued on next page

b



c

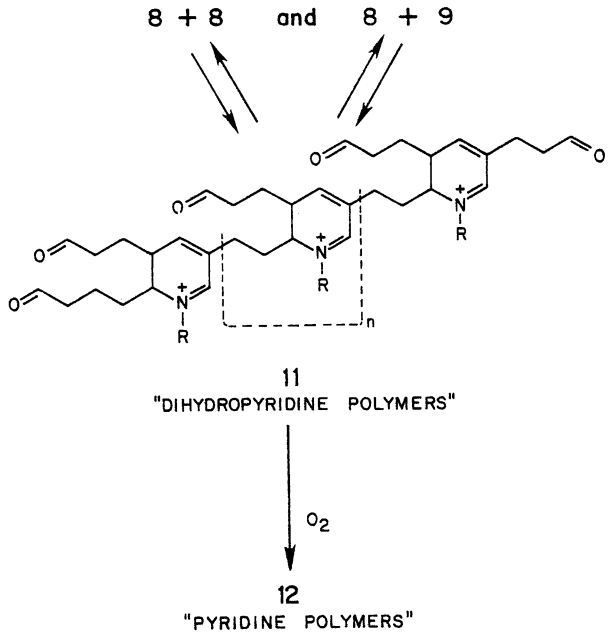


Figure 5. Continued

It is not likely that the largest crosslinking polymers will be helpful for maintaining enzyme activity or retaining antigenicity. Simply put, too much "rope" will sterically hinder access of substrates or antibodies. Or worse yet, excess "rope" will completely block access. It is possible to tailor the crosslink in terms of size by choice of precursors (Table 3). While some aldol condensation does occur between glutaraldehyde molecules themselves, even in the presence of primary amines, the predominant chemistry is the formation of substituted pyridine and polypyridine products (*14-17*). A large excess of glutaraldehyde causes the reaction to revert to the slow default mode of aldol crosslinking.

The time factor should be manipulated because it is easy to do. The goal is to promote the degree of crosslinking needed for stabilization but to permit subsequent biochemistry in terms of enzyme activity or epitope recognition. Product size can be controlled by quenching the polymerization reaction with acid to terminate growth. In the present work the rapid polymerization of glutaraldehyde-amine reactions was terminated by acidifying to pH < 4, i.e., by diluting into 19 volumes of acetate buffer at pH 3.4 (Fig. 4). If acidifying is not compatible with the material to be crosslinked, then rapid dilution or other means of separating small molecules must be employed.

If lysine is chosen to form the crosslinking matrix with glutaraldehyde, then addition of significant concentrations of primary amines such as glycine can be used to restrict the size of the polymer products. In practice, this means adding glycine to one-half of the glutaraldehyde concentration. For 1% glutaraldehyde (0.1 M), add 0.05 volumes of 1 M glycine (≤0.05 M). This step rapidly produces short polypyridines.

In the formation of pyridine products, the ratio of glutaraldehyde to primary amine is as important as the presence of the primary amine. Pyridine polymers are favored by ratios of <2 (*17*). Pyridine monomers are favored by ratios of >3 (*17*).

It is possible to change the chemical nature of the crosslink by choosing different sidechains in the precursor amines. Thus, the negative charge carried by the carboxyl in glycine, for example, can be eliminated and a neutral hydroxyl substituted if ethanolamine is used instead. The worker can select from many amine choices according to experimental needs. Choice of the precursors allows a measure of control over the product size.

Blocking Active Aldehydes. Active aldehyde sites arising from the crosslinking structure are present and usually need to be blocked before subsequent procedures, especially those using other proteins, such as cytochemical labeling. Failure to do so will cause other proteins to accrete to the aldehyde sites generating false positives in labeling. This occurs because monoaldehydes form complexes with primary amines. While the complexes are reversible, the equilibrium lies in favor of complex (*22*). If a protein becomes a polyaldehyde after crosslinking, a polyamine antibody can form multiple aldehyde-amine complexes with the fixed protein. The probability of all complexes reversing simultaneously becomes small and the antibody is nonspecifically and, for practical purposes, irreversibly attached.

The polyaldehyde sites must be blocked with molecules which interfere least with subsequent procedures. Many large molecules have been used to accomplish the blocking. Non-specific serum or IgG are commonly used. Also, solubilized milk powder, gelatin, fish skin gelatin, and bovine serum albumin are used.

Table 3
Glutaraldehyde-based Polymers

Precursors	Product Size (daltons)	Time
Glut	>100 to <1000	5-40 hours
Glut/Glycine	>100 to <5000	5 min
Glut/Glycine/Lysine	>1000 to 5000	5 min
Glut/Lysine	>1000 to >5000	5 min

Another direction for blocking irreversible and non-specific binding is to chemically change the aldehyde sites by reduction to alcohols with $NaBH_4$ (sodium borohydride) (*23*). Sodium borohydride is dissolved in 0.01 M NaOH to a concentration of 100 mM. The borohydride solution is added to fixed preparations to final concentrations of 1-10 mM. After 2-5 minutes additional borohydride is added; a third addition may even be useful. The fixed preparations should be adjusted to a pH of at least ~8 because the borohydride is labile under more acidic conditions (*23*). Borohydride has a half-life of about 1 minute at pH 8, but only 6 sec at pH 7.

Laboratory Safety with Aldehydes. How does one handle glutaraldehyde spills and excess glutaraldehyde? The simplest and fastest method is to have a 1 M solution of glycine on hand. Glycine is a zwitterion and when the amino group reacts with glutaraldehyde, the complex (now charged via the carboxyl group of glycine) is non-volatile. As the precursors progress down the polymerization pathways, the glutaraldehyde precursor is trapped as a non-volatile and minimally reactive polymer. The investigator's eyes, lungs etc. are at little risk from the advanced polymerization products. Sufficient glycine should be added so that there is at least 0.5 moles of glycine/mole glutaraldehyde.

For small spills one should keep a wash bottle with 1 M glycine nearby and spray the spill. One volume of the glycine will react with to 10 volumes of 2% glutaraldehyde (0.2 M). Excess glutaraldehyde waste, for example, is easily detoxified by pouring the waste into a polyethylene bottle (4 liter milk container) into which 700 ml of 1 M glycine was previously added. One can fill the container with up to 4% glutaraldehyde (0.4 M) waste for rapid detoxification. Lower concentrations of glutaraldehyde need proportionately less 1 M glycine.

Acknowledgments

I thank Barbara Johnson for her suggestions and Michael Himmel for his patient efforts to see the work completed. The work was supported by the National Institutes of Health GM 39503-03 to J. Rash and T. Johnson.

Literature Cited

1. Sabatini, D. D.; Bensch, K. J.; Barnett, R. J. *J. Cell Biol.* **1963**, *17*, 19-58.
2. Molin, S.; Nygren, H.; Dolonius, L. *J. Histochem. Cytochem.* **1978**, *26*, 412-414.
3. Tatsumoto, K.; Baker, J. O.; Tucker, M. P.; Oh, K. K.; Mohagheghi, A.; Grohmann, K.; Himmel, M. E. *Appl. Biochem. Biotechnol.* **1988**, *18*, 159-174.
4. Schiff, H. *Ann. Chem.* **1901**, *319*, 59-76.
5. Levy, M. *J. Biol. Chem.* **1933**, *99*, 767-779.
6. Kallen, R. G.; Jencks, W. P. *J. Biol. Chem.* **1966**, *241*, 5864-5878.
7. Johnson, T. J. A. *J. Electron Microsc. Tech.* **1985**, *2*, 129-138.
8. Richards, F. M.; Knowles, J. R. *J. Mol. Biol.* **1968**, *37*, 231-233.
9. Whipple, E. B.; Ruta, M. *J. Org. Chem.* **1974**, *39*, 1666-1668.
10. Hardy, P. M.; Nicholls, A. C.; Rydon, H. N. *J. Chem. Soc. Chem. Comm.* **1969**, 565-566.
11. Hopwood, D. *Histochem. J.* **1972**, *4*, 267-303.
12. Monsan, P.; Puzo, G.; Mazarquil, H. *Biochemie* **1975**, *57*, 1281-1292.
13. Lubig, R.; Kusch, P.; Roper, K.; Zahn, H. *Monatschefte fur Chemie* **1981**, *112*, 1313-1323.
14. Hardy, P. M.; Nicholls, A. C.; Rydon, H. N. *J. Chem. Soc. Perkin I* **1976**, 958-962.
15. Hardy, P. M.; Hughes, G. J.; Rydon, H. N. *J. Chem. Soc. Chem. Comm.* **1976**, 157-158.
16. Hardy, P. M.; Hughes, G. J.; Rydon, H. N. *J. Chem. Soc. Perkin I* **1979**, 2282-2288.
17. Johnson, T. J. A. In The Science of Biological Specimen Preparation 1985; Mueller, M.; Becker, R. P.; Boyde, A.; Wolosewick, J. J., Eds.; SEM Inc.: AMF O'Hare, IL 60666, 1986, pp 51-62.
18. Gillet, R.; Gull, K. *Histochemie* **1972**, *30*, 162-167.
19. Rasmussen, K. E.; Albrechtsen, J. *J. Histochem.* **1974**, *38*, 19-26.
20. Borch, R. F.; Bernstein, M. D.; Durst, H. D. *J. Am. Chem. Soc.* **1971**, *93*, 2897-2904.
21. Johnson, T. J. A. *Eur. J. Cell Biol.* **1987**, *45*, 160-169.
22. Hine, J.; Yeh, C. Y.; Schmalstieg, F. C. *J. Org. Chem.* **1970**, *35*, 340-344.
23. Leppla, S. L.; Bjoraker, B. J.; Bock, R. M. *Methods Enzymol.* **1968**, *XII* Part B, 236-240.
24. Good, N. E.; Winget, G. D.; Winter, W.; Connolly, T. N.; Izawa, S.; Singh, R. M. M. *Biochem. J.* **1966**, *5*, 467-477.

RECEIVED April 29, 1992

Chapter 24

Chemical Modification

Effect on Enzyme Activities and Stabilities

A. Sadana and R. R. Raju

Department of Chemical Engineering, University of Mississippi, University, MS 38677–9740

The chemical modification of enzymes is analyzed using a series-deactivation model. The influence of chemical modification on the specific activity of the initial enzyme state, on the stability and on the residual activity of different enzymes is presented. The chemical modification may change none, one, two, or all of the above variables for an enzyme. The analysis provides fresh physical insights into enzyme deactivation processes, and into the structure-function relationships for enzymes.

A significant amount of effort has been spent in understanding enzyme deactivation mechanisms and what affects the residual activity of enzymes. This type of effort provides significant physical insights into the enzyme structure and function, besides it helps utilize enzyme ability to its fullest extent. Basically, one would like to know how active a particular enzyme state is, and for how long can one maintain a required level of activity. Chemical modification is one means by which one can help an enzyme attain the required characteristics of enzyme activity, residual activity, and stability. Stability generally refers to an inverse measure of the rate constant for transformation of one enzyme state (in this case) to another. Residual activity refers to the activity of an enzyme form (different from the initial enzyme state), which is encountered along the inactivation pathway, and is stable enough under the stress being applied that it can accumulate and persist for a long time. This time period is a long time compared to the time required to transform the native form or some preceding non-native form almost completely into this form. This form then appears to be "permanent." The influence of modification (chemical or otherwise) on initial enzyme activity, stability, and on residual activity is important. More often than not, only one, at best two, of the above three is reported in the literature. In general, initial enzyme activity, residual activity, and stability exhibit opposing tendencies, and

0097–6156/93/0516–0296$06.00/0

thus, it is essential to report the influence of modification of enzymes on initial enzyme activity, residual activity and stability.

In this paper we will attempt to analyze enzyme deactivations based on the influence of chemical modification on the initial enzyme activity, residual activity, and on the stability. This should shed novel physical insights into enzyme deactivation mechanisms, and into the structure and function of enzymes. Surely, it is of interest to develop guiding principles in order that one may predictively attain required (and balanced) levels of initial enzyme activity, stability and residual activity for enzymes. Hopefully, this paper is one step in that direction.

Theory

Since we are interested in both the initial activity and the residual activity, enzyme deactivation data available in the literature was screened for both of these aspects. Generally, the residual activity data is available for enzyme deactivations. Since the data is reported often in terms of normalized activity (enzyme activity at any time, t divided by enzyme activity at time, t = 0), the initial specific activity is generally not available. Also, microheterogeneity of enzymes is often observed (*1*) and this factor should be considered while analyzing the influence of chemical modification on enzyme activity, stability, and residual activity.

The activity-time expression used is (*1*):

$$a = \left(1 + \frac{\beta_1 k_1}{k_2 - k_1} - \frac{\beta_2 k_2}{k_2 - k_1}\right) \exp(-k_1 t) - \left(\frac{k_1}{k_2 - k_1}\right)(\beta_1 - \beta_2) \exp(-k_2 t) + \beta_2 \quad (1)$$

This weighted-average activity expression is obtained for the series-deactivation scheme:

$$E \xrightarrow{k_1} E_1 \xrightarrow{k_2} E_2 \quad (2)$$

Here k_1 and k_2 are first-order deactivation rate coefficients. Let [E], [E_1], and [E_2] be the concentrations of the different enzyme states. Also, let δ, δ_1, and δ_2 be the specific activities of the E, E_1, and E_2 enzyme states, respectively. Let β_1 be the ratio of the specific activities δ_1/δ, and β_2 the ratio of the specific activities δ_2/δ. At time t equal to zero, the normalized activity, a, is equal to one. Very rarely, if at all, does one see units in the activity label in these figures. Figures 1 to 3 show curves fitted by the SAS procedure (*2*) performed on experimental data on immobilized and soluble

enzymes using equation 1. The estimated values of k_1, k_2, β_1, and β_2 are given in a table that follows for the unmodified and the modified enzymes undergoing inactivation.

It may be shown that for the series deactivation scheme:

$$E \xrightarrow{k_1} E_1 \xrightarrow{k_2} E_2 \xrightarrow{k_3} E_3 \xrightarrow{k_4} E_4 \xrightarrow{k_5} \cdots \xrightarrow{k_n} E_n \xrightarrow{k_{n+1}} E_d \qquad (3)$$

the weighted-average activity function is given by:

$$\begin{aligned} a = {} & \left[1 + \frac{\beta_1 k_1}{k_2 - k_1} + \frac{\beta_2 k_1 k_2}{(k_2 - k_1)(k_3 - k_1)} + \frac{\beta_3 k_1 k_2 k_3}{(k_2 - k_1)(k_3 - k_1)(k_4 - k_1)} + \cdots \right. \\ & \left. + \frac{\beta_n k_1 k_2 k_3 \cdots k_n}{(k_2 - k_2)(k_3 - k_1)(k_4 - k_1)\cdots(k_{n+1} - k_1)}\right] \exp(-k_1 t) \\ & + \left[\frac{\beta_1 k_1}{k_1 - k_2} + \frac{\beta_2 k_1 k_2}{(k_1 - k_2)(k_3 - k_2)} + \frac{\beta_3 k_1 k_2 k_3}{(k_1 - k_2)(k_3 - k_2)(k_4 - k_2)} + \cdots \right. \\ & \left. + \frac{\beta_n k_1 k_2 k_3 \cdots k_n}{(k_1 - k_2)(k_3 - k_2)\cdots(k_{n+1} - k_2)}\right] \exp(-k_2 t) + \cdots \\ & + \left[\frac{\beta_n k_1 k_2 k_3 \cdots k_n}{(k_1 - k_{n+1})(k_2 - k_{n+1})(k_3 - k_{n+1})\cdots(k_n - k_{n+1})}\right] \exp(k_{n+1} t) \end{aligned} \qquad (4)$$

Let δ_3, δ_4, δ_5, etc. be the specific activities of the E_3, E_4, E_5, enzyme states. Then $\beta_3 = \delta_3/\delta$, $\beta_4 = \delta_4/\delta$, $\beta_5 = \delta_5/\delta$, etc. Equations (3) and (4) involve a large number of parameters. Nevertheless, a multi-deactivation step model would be more appropriate in some cases. For example, a general equation based on this reaction scheme will be more realistic in predicting whether an enzyme is deactivated to what stage. This would be particularly useful for enzymes that can exist in different conformational states. In spite of the complexity of equation (3) it is apparently inappropriate for cross-linked enzymes, since this procedure generally produces heterogeneous molecules.

Prior to presenting some examples, it is appropriate to present certain "disclaimers" regarding the limited nature of the model that has been simplified enough to be of practical use. (a) The model describes a multi-step process that is indeed useful in describing some 'complex' enzyme inactivations. Note that a large number of inactivations can be modeled quite effectively in terms of simple first-order decays. We have simply chosen to study multi-step processes in this paper. (b) The model describes two sequential, irreversible steps. It is possible for kinetically detectable denaturation intermediates to arise from reversible steps, such as a 'pre-equilibrium' between stable and more labile forms, the latter being transformed by later steps. (c) When "activity remaining" is not measured in the "stress situation" itself, but is measured under other conditions using withdrawn aliquots, partial

reversibility of the inactivation may result in "residual activities" that are not actually present in the enzyme at the point in the stress situation from which the aliquot was withdrawn. (Dilution for assay of some samples denatured in urea is a specific example). (d) Chemical modification of enzymes is extremely likely to introduce heterogeneity in that potential reaction sites may or may not be modified in a given enzyme molecule. Modification at one site may interfere with subsequent modification at a neighboring site, and the effect of modification at different sites may be different even opposing. When a hypothetical chemically modified enzyme, tested under conditions capable of completely inactivating the native enzyme (i.e., with no "residual activity"), shows a "residual activity" equivalent to half the initial activity of the native enzyme, all of the modified enzyme molecules may have been transformed by the stress into a different form that has one-half the activity per molecule. Another possibility is that half the enzyme molecules may be stabilized to such an extent that they retain all of their original activity, and the other half of the enzyme molecules fall prey to competing, non-stabilizing chemical modifications, and, under the testing stress, wind up completely inactivated. Without some means, other than activity, of measuring the state of the protein(s) we are not going to be able to distinguish between the two explanations.

Chemical modification of enzymes may lead to opposing changes in initial specific activity and residual activity. The chemical modification may be by cross-linking agents that make the enzyme more rigid (less flexible). The cross-links act as a clamp to "rigidify" the enzyme in its native conformation. This leads to a decreasing initial specific activity and an increasing residual activity. The modification may also increase the initial specific activity and decrease the residual activity of the enzyme.

Example 1. Effect of modification and presumably cross-linking of penicillinase by toluene 2,4-diisocyanate on the thermostability of the enzyme. The influence of cross-linking by toluene 2,4-diisocyanate on the initial specific activity of penicillinase (*3*) is shown in Figure 1 and in Table I. Toluene 2,4-diisocyanate binds to the ε-amino group of lysine residues in a protein and forms a stable ureido-bond (*4*). Cross-linking increases, in general, the thermostability of enzymes (*5*), including penicillinase (*6*). The cross-linking of penicillinase by toluene 2,4-diisocyanate reduces the initial specific activity by nearly 48 percent. Cross-linking does stabilize the enzyme since the rate constant for inactivation for the first steps, k_1 decreases from 3.4 min^{-1} to 1.0 min^{-1} on modification. The k_2 value is unchanged. The relative residual activity is slightly more than double (β_2 equals 0.19 and 0.41 for the unmodified and modified enzyme forms, respectively). In absolute terms, the specific activities of the final enzyme state for the unmodified and the modified enzyme are 49 and 56 units/μg, respectively. No information was given (*3*) regarding the standard deviation of enzyme activity measurements. In many cases, it is not uncommon to obtain 20 percent variation in enzyme activity measurements. Thus, one might reasonably argue if 49 units/μg is truly different from 56 units/μg?

Example 2. Effect of acetamidination on the thermal stability of pig heart lactate dehydrogenase. Table I also shows the influence of methylacetimidate modification on the thermal denaturation at 60°C of pig heart lactate dehydrogenase (*7*). The figure is not given. Lactate dehydrogenase is a tetrameric enzyme. Out of the 24

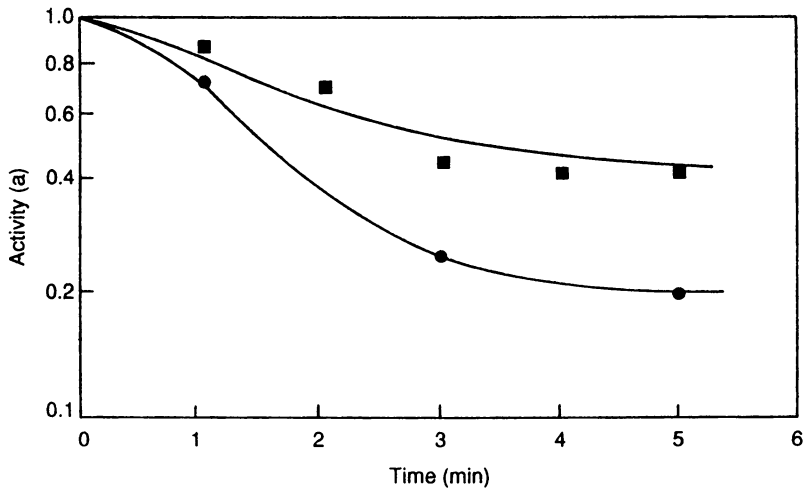


Figure 1. The effect of prior chemical modification with toluene 2,4-diisocynate on the thermal inactivation kinetics of penicillinase measured at 56°C in 0.05 M potassium phosphate buffer containing 0.25% gelatin; ● native; ■ modified by toluene 2,4-diisocyanate in the absence of substrate. (Adapted from ref. 7, Figure 1. Reproduced with permission from ref. 15. Copyright 1985 Wiley.)

Table I. Influence of Chemical Modification on Initial Specific Activity, Residual Activity, and Rate Constants for Inactivation of Enzymes

Enzyme	Chemical Modifier	Initial Specific Activity (units/μg protein)	Residual Activity (units/μg protein)	k_1 (min^{-1})	k_2 (min^{-1})	β_1	β_2	Ref.
penicillinase	None	260	49	3.4	1.1	1.3	0.19	3
penicillinase	toluene 2,4-diisocyanate (0.1 ml/mg enzyme)	136	56	1.0	1.1	1.0	0.41	3
lactate dehydrogenase	None	436	56	170 d^{-1}	180 d^{-1}	0.81	0.43	7
lactate dehydrogenase	methyl acetimidate, 2.8 mg reagent/mg protein	348	320	220 d^{-1}	250 d^{-1}	0.91	0.92	7
penicillinase in urea	None	260	185	0.071	0.094	1.1	0.71	3
penicillinase in urea	toluene 2,4-diisocyanate	136	109	0.078	0.13	1.2	0.80	3

Continued on next page

Table I. Influence of Chemical Modification on Initial Specific Activity, Residual Activity, and Rate Constants for Inactivation of Enzymes (Continued)

Enzyme	Chemical Modifier	Initial Specific Activity (units/μg protein)	Residual Activity (units/μg protein)	k_1 (min^{-1})	k_2 (min^{-1})	β_1	β_2	Ref.
mushroom tyrosinase (immobilized on collagen)	None	0.2 μmol DOPA / h-mg protein	0	14 d^{-1}	19 d^{-1}	1.6	0	10
mushroom tyrosinase (immobilized on collagen)	dimethyl adipimidate	0.3 μmol DOPA / h-mg protein	0	12 d^{-1}	11 d^{-1}	1.2	0	10
E. coli asparaginase	None	100% of native unmodified enzyme	0	0.23	0	0	0	11
E. coli asparaginase	succinylation: fraction of groups acylated equal to 0.26	65% of native unmodified enzyme	0	0.15	0	0	0	11

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.

lysines per subunit, only 6.5 ± 0.5 lysines were left unmodified. On modification, the specific activity of the initial enzyme state decreases by 20 percent and the relative residual activity more than doubles. In absolute terms, the specific activities of the final enzyme state for the unmodified and modified enzyme are 188 and 320 units/μg, respectively. Note that a 20 percent decrease in the initial specific activity by chemical modification increases the specific activity of the final state by 60 percent. Since k_1 and k_2 both increase after treatment with methylacetimidate the cross links in this case destabilize the enzyme. Also, in neither case does the activity-time curve (Figures 2 a,b in reference *7*) appear to have "bottomed out" at the end of the data shown. Both enzyme preparations are losing activity 34 minutes after the start of the incubation at 60°C. Residual activity is the theoretical limiting value obtained by modeling the denaturation data using equation (1).

Chemical modification of enzymes may lead to changes in the initial specific activity and the residual activity in the same direction. These changes could be caused by both external and intrinsic (or non-external) agents. Examples of intrinsic agents could be enzyme or co-enzyme concentrations, which, for example, may convert subactive enzyme forms to more active enzyme forms. This could lead to increases in both initial specific activity and residual activity. Some alcohols stabilize proteins at low concentrations.

Example 3. Effect of cross-linking with toluene 2,4 diisocyanate on the inactivation kinetics of penicillinase in 8 M urea. The influence of cross-linking of penicillinase in urea by toluene 2.4-diisocyanate (*3*) decreases the initial specific activity by nearly 48 percent (see Table I). The relative value of the residual activity of the final enzyme state for the modified enzyme is higher than of the unmodified enzyme (β_2 equals 0.71 and 0.80 for the unmodified and modified forms, respectively). However, in absolute terms, the specific activities of the final enzyme state for the unmodified and the modified enzyme are 185 and 108 units/μg respectively. Since k_1 and k_2 both increase after treatment with toluene 2,4-diisocyanate, the cross links in this case destabilize the enzyme. In this case, the modification of the enzyme reduces both the specific activity of the initial enzyme state and that of the final enzyme state, and also destabilizes the enzyme. (Figure 2)

Example 4. Effect of modification with dimethyl adipimidate on the "reaction inactivation" of mushroom tyrosinase immobilized in a collagen membrane and used in a flow reactor. Table I shows that modification by dimethyl adipimidate of mushroom tyrosinase immobilized on collagen increases the specific activity of the initial enzyme state by 50 percent (*10*). The activity-time figure is not given. The cross links provided by the dimethyl adipimidate suitably change the conformation of the active site to yield a higher catalytic activity. The estimated β_2 values for the unmodified and the modified enzymes are zero, thus, the cross-links do not provide the necessary rigidity to the enzyme for it to exhibit a residual activity. However, the cross-links do stabilize the enzyme, as they slow down the inactivation process. This is because the estimated values of k_1 and k_2 for the modified enzyme are lower than the corresponding values of the unmodified enzyme.

Example 5. Effect of succinylation on the resistance of E. coli asparaginase to inactivation by subtilisin proteolysis. The influence of succinylation on the initial specific activity and resistance of *E. coli* asparaginase to subtilisin proteolysis (*11*) is shown in Figure 3 and Table I. Succinylation was carried out to obtain the desired ratios of succinic anhydride and lysine residues. Native asparaginase is rapidly inactivated by subtilisin at a molar ratio of subtilisin to asparaginase of 1:25. A succinylation (fraction of groups acylated equal to 0.26) of *E. coli* asparaginase (*11*) leads to a decrease in the specific activity of the initial enzyme state. This extent of succinylation decreases the specific activity of the initial enzyme state to 65 percent of the unmodified enzyme value. From Figure 3 it appears that the curves for the unmodified and the 26% modified enzyme inactivation can be modelled adequately by straight lines. For these curves there would appear to be no necessity of postulating an inactivation pathway involving more than one step. The situation for these curves can be accommodated in Equation (1) by assuming $\beta_1 = 0$, and $k_2 = 0$ and/or $\beta_2 = 0$. For 26 percent modification, the observed changes in the inactivation process are changes in the initial activity (the modified enzyme exhibits a decrease compared to the unmodified enzyme) and in the (apparent first-order) rate constant (the modified enzyme exhibits a lower rate constant compared to the unmodified enzyme) for inactivation. Thus, modification in this case leads to a decrease in the specific activity of the initial state, but it also stabilizes the enzyme.

Chemical modification of enzymes may also lead to changes only in the residual activity. The specific activity of the initial enzyme state is unaltered. These types of changes may be observed when chemical modification of an enzyme molecule occurs only at positions distant from the active site. This type of modification can drastically affect residual activity, while leaving its specific activity unaltered.

Also, it is not always necessary that modification must lead to changes in initial specific activity and/or residual activity. In some cases, chemical modification of enzymes leaves both the initial specific activity and the residual activity unaltered. This type of behavior may really be split into two further types. In the first case there is no inactivation of either the unmodified or the modified enzyme. Then, the initial specific activities are equal and so are the residual activities. In the other case there is inactivation but the estimated values of β_2 are the same. Examples of the first type could include the possibility of the introduction of a "defense mechanism" against enzyme inactivation by the introduction of a chemical modifier (*13*). It is perhaps not too unreasonable to assume that examples of the second type would be rare. This is so, since the chemical modification would have to be such that (a) it does not change either the initial specific activity or the residual activity, and (b) other deactivation characteristics such as k_1 and k_2 values may or may not change.

Conclusions

The series-deactivation model is utilized to analyze chemical modification-induced enzyme inactivations. The analysis is based on the influence of modification on the specific activity of the initial enzyme state, the residual activity, and on the inactivation rate constants. The analysis provides an appropriate framework to

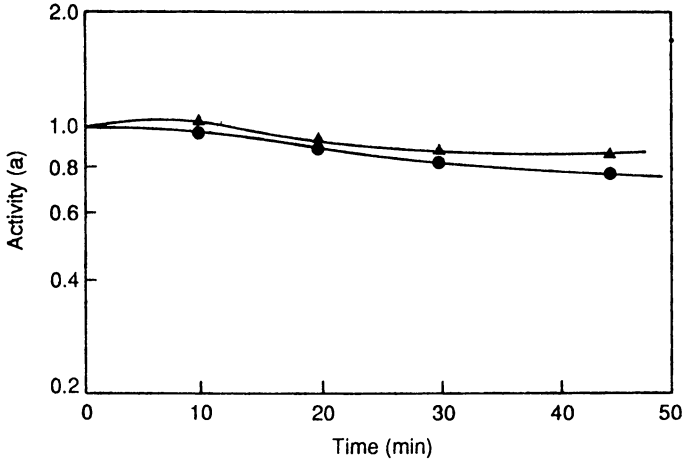


Figure 2. The effect of cross-linking with toluene 2,4-diisocyanate on the inactivation kinetics of penicillinase in 8 M urea, ● no additives present; ▲ toluene 2,4-diisocyanate present. (Adapted from ref. 3, Figures 1 and 3. Reproduced with permission from ref. 15. Copyright 1985 Wiley.)

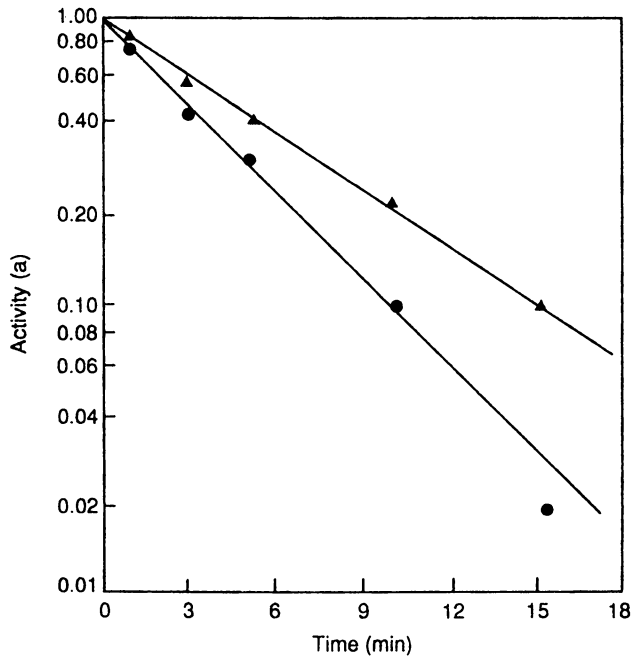


Figure 3. Influence of succinylation on the inactivation kinetics of *E. coli* Asparaginase, ● no modification; ▲ fraction of groups acylated equal to 0.26. (Adapted from ref. 11, Figure 2. Reproduced with permission from ref. 16. Copyright 1986 C. D. Scott.)

compare enzyme deactivation data obtained by different investigators working in a wide variety of areas. By providing judicious examples wherever appropriate the analysis provides an overall perspective of the influence of chemical modification of enzymes undergoing inactivation. Finally, the analysis should provide fresh physical insights into enzyme deactivation processes and stimulate further studies to provide a better understanding of enzymes, in general.

Literature Cited

1. Sadana, A. Biocatalysis: Fundamentals Of Enzyme Deactivation Kinetics, Prentice Hall: New York, NY, 1991, p. 256.
2. SAS User's Guide, Statistics Edition, SAS Institute: Cary, NY, 1982.
3. Klemes, Y.; Citri, N. *Biochem. J.* **1980**, *187*, 529.
4. Schick, A. F.; Singer, S. J. *J. Biol. Chem.* **1961**, *236*, 2477.
5. Zaborsky, O. R., In *Enzyme Engineering*, (E. K. Pye and L. B. Wingard, Jr., eds.), Plenum Press: New York, NY, 1974, Vol. 2, pp. 115-22.
6. Klemes, Y.; Citri, N. *Biochim. Biophys. Acta* **1979**, *567*, 401.
7. Tuengler, P.; Pfleiderer, G. *Biochim. Biophys. Acta* **1977**, *484*, 1.
8. Fornaini, G.; Leoncini, G.; Segni, P.; Calabria, G.A.; Dacha, M. *Eur. J. Biochem.* **1969**, *7*, 214.
9. Yoshida, A. *J. Biol. Chem.* **1966**, *241*, 4966.
10. Letts, D.; Chase, Jr., T. *Adv. Exp. Med. Biol.* **1974**, *42*, 317.
11. Nickle, E. C.; Solomon, R. D.; Torchia, T. E.; and Wriston, Jr., J. C. *Biochim. Biophys. Acta* **1982**, *704*, 345.
12. Keradjopoulos, D.; Holdorf, A. W. *FEMS Microbiol. Lett.* **1977**, *1*, 179.
13. Eaton, J. W.; Boraas, M.; Etkins, N. L. *Adv. Exp. Med. Biol.* **1972**, *28*, 121.
14. McMahon, S.; Stern, A. *Biochim. Biophys. Acta* **1979**, *566*, 253.
15. Sadana, A. Henley, J. P. *Biotechnol. Bioeng.* **1986**, *28*, 256.
16. Sadana, A. Henley, J. P. In *Seventh Symposium on Biotechnology for Fuels and Chemicals;* Scott, C. D., Ed. Wiley: New York, 1985, p. 487.

RECEIVED April 29, 1992

Chapter 25

Cross-Linking Techniques

Applications to Enzyme and Protein Stabilization and Bioconjugate Preparation

Munishwar N. Gupta

Chemistry Department, Indian Institute of Technology, Delhi Hauz Khas, New Delhi–110 016, India

Chemical crosslinking technology was used for (a) enhancing the stability of trypsin, β-galactosidase, and concanavalin A and (b) forming protein-protein conjugates viz. trypsin–chymotrypsin, trypsin–β-galactosidase, and concanavalin A–β-galactosidase. For stabilization of the three proteins, dimethyl adipimidate was found to give the best results. The intramolecularly crosslinked trypsin, with an average of 9.5 groups modified out of 14 free amino groups present, showed a much slower rate of autolysis at 40°C compared to native trypsin. Crosslinked β-galactosidase entrapped in polyacrylamide hydrolyzed 47% of milk lactose in 6 h and at 55°C. Entrapped native enzyme hydrolyzed only 31% substrate under the same conditions. Besides preparing an insoluble aggregate of trypsin–chymotrypsin and β-galactosidase, conjugates of trypsin–chymotrypsin, trypsin–alkaline phosphatase, and Concanavalin A–β-galactosidase were also prepared and evaluated.

There is near unanimity about unfolding of the polypeptide chain as being a primary event in the denaturation of proteins (*1*). For example, in thermal denaturation, this is the initial reversible step (*2*). Thus, it follows that any approach which makes proteins rigid (reduces conformational flexibility) should impart stability. Three distinct approaches attempt to do this by creating additional linkages with parts of the polypeptide chain:

0097–6156/93/0516–0307$06.00/0

(1). Immobilization (*2-4*)

(2). Introduction of disulfide bridges by protein engineering (*5*)

(3). Chemical crosslinking (*6*)

The present chapter would limit itself to the third approach. Apart from protein stabilization, another major application of crosslinking technology has been linking of different biological molecules by intermolecular crosslinking. Such bioconjugates have already found considerable applications in such diverse areas as bioconversion (*7*), medicine (*8*), and bioseparation (*9*). This chapter would also describe some work on heteroenzyme conjugates and other bioconjugate preparations using intermolecular crosslinking.

Crosslinking Methodology

When a crosslinking reagent reacts with a protein molecule, several kinds of reaction products are possible, including intramolecularly crosslinked proteins, oligomers formed due to intermolecular crosslinking, and insoluble protein aggregates. However, it is possible to design a crosslinking experiment so as to obtain the desirable product as the major reaction product.

Some general guidelines mentioned by Wold (*10*) many years ago are quite useful for this purpose:

(1). High protein concentration would favor intermolecular crosslinking over intramolecular crosslinking.

(2). Ph corresponding to minimum net charge on the protein would favor intermolecular crosslinking.

(3). High reagent to protein ratio and prolonged reaction time would favor extensive crosslinking and may result in insoluble protein aggregates.

(4). Successful formation of crosslinks depend upon the availability of suitable reactive groups within the effective range of the reagents. In the context of intramolecular crosslinking, this means that the "span" of the crosslinking reagent is a crucial parameter.

Enzyme Stabilization

Chemical Crosslinking of Trypsin. The property of trypsin to undergo autolysis in solution has resulted in this enzyme being a favorite target of various approaches of enzyme stabilization (*11-13*). Attempts at chemical crosslinking of trypsin with glutaraldehyde and two bisimidoesters [i.e., dimethyl suberimidate (DMS) and dimethyl adipimidate (DMA)] indicated that bisimidoesters are a better choice for obtaining trypsin preparations with decreased autolysis (*14*). Best results were

obtained with DMA (14.3 mg/ml) using trypsin concentration of (0.25 mg/ml) in Tris-HCl buffer (0.2 M, pH 8.3) containing 25 mM Ca^{++} (*13*). Chromatography on CM-Cellulose showed that the crosslinking preparation did not contain any unmodified trypsin. Free amino group analysis showed that an average of 9.5 residues out of 14 present in trypsin were modified. SDS-PAGE analysis showed that no intermolecular crosslinking has occurred. Some characteristics of the crosslinked trypsin are summarized in Table 1 (*15*).

The crosslinked preparation showed much slower rate of autolysis at 40°C as compared to native trypsin (Figure 1). In this particular system, the formation of crosslinks was presumed and not verified. It was also not determined whether the decrease in autolysis was caused by decrease in number of bonds susceptible to tryptic cleavage or crosslinking *per se*.

Lactose Intolerance; Whey Disposal Through Chemical Crosslinking. Another enzyme chosen for chemical crosslinking was β-galactosidase. The choice of this enzyme was based upon its biotechnological usefulness (*16*). This usefulness arises because of lactose intolerance and whey as a biomass (*16*).

Lactose Intolerance. It is a metabolic disorder which is associated with the lack of adequate β-galactosidase activity. There are significant differences in the incidence of lactose intolerance among different ethnic groups. The adult intolerance has so far been observed only in northern Europeans (90%) and in the members of two nomadic pastoral tribes of Africa (80%) (*17*).

Hydrolysis of milk lactose yields low lactose milk. Such a preparation, apart from being low in lactose content, also retains most of the other nutrients present in the milk. Many commercial technologies for production of low lactose milk utilize immobilized β-galactosidase (lactase) (*18*).

Whey as a Biomass. A large amount of milk is converted into whey during the manufacture of cheese. The use of lactose present in whey in food industries has certain associated problems. Whey, as a waste product, cannot be disposed of easily as such because of its high biological oxygen demand (BOD) value. The hydrolysis of whey lactose by β-galactosidase to glucose and galactose solves these problems. These sugars have greater fermentation potentials as compared to lactose. Also, the hydrolyzed whey can also be used in food industries as such (*16*).

Crosslinking of β-Galactosidase. The above considerations prompted us to attempt crosslinking of *E. coli* β-galactosidase—a commercially available and well characterized enzyme. The crosslinking reagents employed in this work (*19*), i.e., glutaraldehyde, DMA and DMS, were specific for free amino groups in proteins. Therefore, they were a safe choice for *E. coli* β-galactosidase since it is reported that the lysine groups in the enzyme are not involved in the catalysis and the enzyme retains 80% of its activity even after all its lysine groups are modified (*20*).

Exploratory experiments indicated that the crosslinked derivative obtained with DMA was most thermostable (*19*). The free amino group analysis showed that 38% free amino groups were modified. The SDS-PAGE analysis showed that neither

Table 1. Properties of Trypsin Crosslinked with Dimethyl Adipimidate

Property	All percentage values are relative to the observed value for native trypsin taken as 100
1. Amidase activity towards benzoyl DL-arginine p-nitroanilide (BAPNA)	113%
2. Esterase activity with p-tosyl L-arginine methyl ester (TAME) as substrate	65%
3. Proteolytic activity towards:	
(a) Casein	27%
(b) Haemoglobin	22%
4. K_m for BAPNA at pH 8.2	1 mM, same as that for native trypsin
5. pH optimum with BAPNA as substrate	Broad range of 7-9, similar to native trypsin

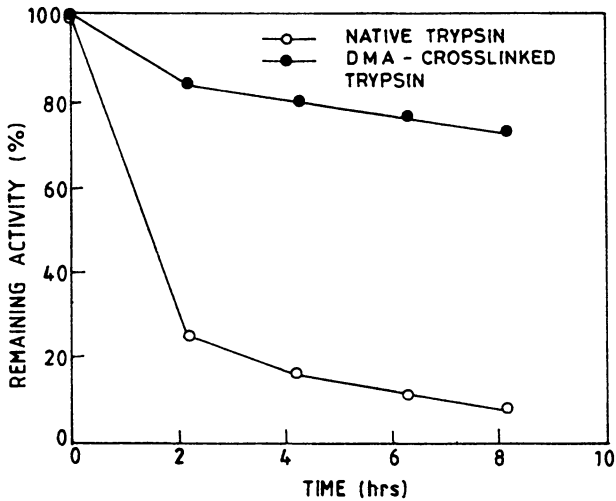


Figure 1. Autolysis of native and DMA-crosslinked trypsin at 40°C. The protein concentration was 125 μg/ml. (Reproduced with permission from ref. 15. Copyright 1988 Butterworth–Heinemann.)

intersubunit nor intermolecular crosslinks had been formed. When the enzyme was treated with ethyl acetimidate (a monofunctional analog of DMA), although extent of free amino group modification was comparable (26%), the o-nitrophenyl β-D-galactopyranoside (ONGP) activity of the modified preparation, both before and after heat treatment (i.e., the enzyme preparations were heated at 55°C in 0.3 M sodium monophosphate buffer pH 8.0 containing 3 mM $MgCl_2$) was quite different as compared to the crosslinked enzyme (Figure 2). While this data did not completely rule out the possibility of some simple chemical modification also having taken place in case of reaction with DMA, it did confirm that the desirable change in the enzyme was the result of formation of intramolecular crosslinks (*19*).

Continuous hydrolysis of milk lactose at 50°C was monitored by using both native and the crosslinked enzyme (Figure 3). Whereas the native enzyme hydrolyzed 40% milk lactose, the crosslinked enzyme hydrolyzed 55% milk lactose in 12 h (*15*). This kind of conversion rate is considered adequate for obtaining low lactose milk based dairy products at the pilot plant level (*21*).

In order to be able to reuse crosslinked enzyme, it was entrapped in polyacrylamide (*22*). The optimization of entrapment conditions was carried out with the native enzyme, and our results show that 50% of the enzyme activity on ONGP was entrapped and the enzyme lost 20% activity during entrapment. The activity of the entrapped enzyme was considerably enhanced when a protective mixture of bovine serum albumin, cysteine, and lactose was present during entrapment. The enzyme crosslinking with DMA and entrapped was also found to be more thermostable as compared to other entrapped enzyme preparations (Figure 4).

The hydrolysis of milk lactose was carried out using native and DMA crosslinked enzymes (Figure 5). The crosslinked preparation entrapped in polyacrylamide hydrolyzed 47% of milk lactose as compared to 31% hydrolysis by entrapped native enzyme in 6 h at 55°C (*22*).

The above system illustrates the usefulness of combining crosslinking with immobilization for obtaining a reusable product with enhanced thermostability.

Insoluble β-Galactosidase Aggregate (*23*). Another way of obtaining a reusable enzyme preparation by crosslinking is to form chemical aggregates. *E. coli* β-galactosidase aggregates prepared by extensive crosslinking with glutaraldehyde retained 63% of the activity on ONGP provided bovine serum albumin and lactose were present during the crosslinking at 4°C. This activity increased to about 70% when the aggregate was homogenized for about 90 s in a mixing blender. All further work discussed below was carried out with the aggregate which had been subjected to this treatment. The K_m value of the aggregate for ONGP was found to be 6×10^{-4} M at pH 7.5 and 25°C as compared to a value of 2.8×10^{-4} M for the native enzyme. The pH optimum was found to remain unchanged at 7.5. The aggregate showed considerable improvement in thermal stability at 55°C (Figure 6) which was reflected in its improved performance during continuous hydrolysis of milk lactose (Figure 7).

One reason why enzyme aggregates have not become popular is that they are difficult to handle and would give poor flow rates in columns. A solution to this

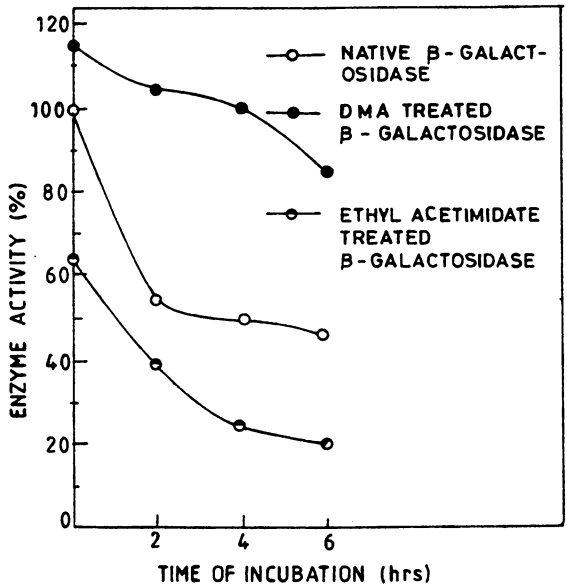


Figure 2. Comparison of the stability of native β-galactosidase with β-galactosidase modified with ethyl acetimidate and DMA respectively. (Reproduced with permission from ref. 23. Copyright 1988.)

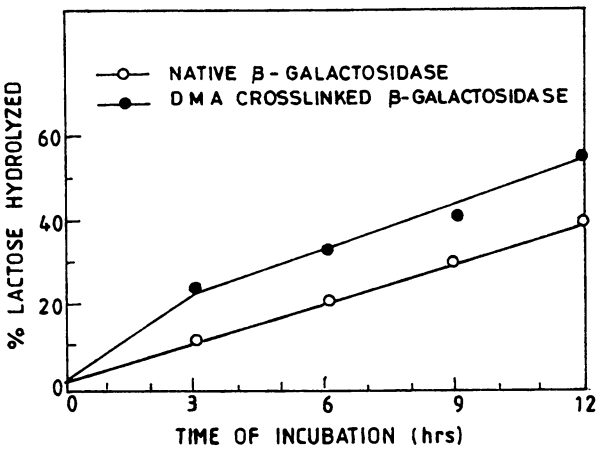


Figure 3. Continuous lactose hydrolysis of milk by DMA modified and native β-galactosidase at 50°C. The enzyme concentration for both native as well as modified preparation was 25 μg in 2.0 ml reaction volume. The initial lactose concentration in the milk was 5%. (Reproduced with permission from ref. 25. Copyright 1988.)

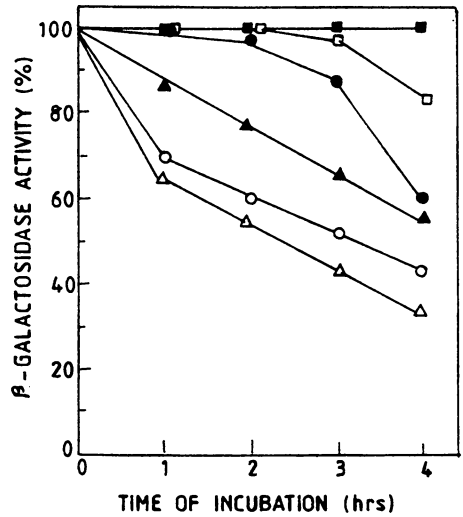


Figure 4. Thermal stability of entrapped β-galactosidase preparations at 55°C in sodium phosphate buffer (20 mM, pH 7.5 containing 0.1 M NaCl and 3 mM $MgCl_2$).
○; Native enzyme entrapped in absence of any protective agent
△; DMS crosslinked enzyme entrapped in absence of any protective agent
□; DMA crosslinked enzyme entrapped in absence of any protective agent
●; Native enzyme entrapped in presence of bovine serum albumin (BSA) + cysteine + lactose
▲; DMS crosslinked enzyme entrapped in presence of BSA + cysteine + lactose
■; DMA crosslinked enzyme entrapped in presence of BSA + cysteine + lactose
(Reproduced with permission from ref. 32. Copyright 1988.)

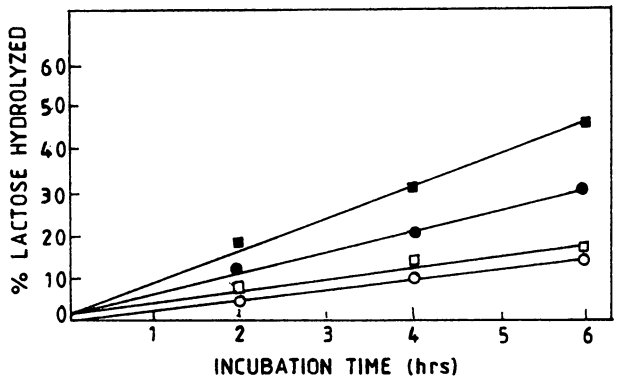


Figure 5. Time course of hydrolysis of milk lactose at 55°C.
○; Native enzyme entrapped in absence of any protective agent
●; Native enzyme entrapped in presence of BSA + cysteine + lactose
□; DMA crosslinked enzyme entrapped in absence of any protective agent
■; DMA crosslinked enzyme entrapped in presence of BSA + cysteine + lactose
(Reproduced with permission from ref. 33. Copyright 1988.)

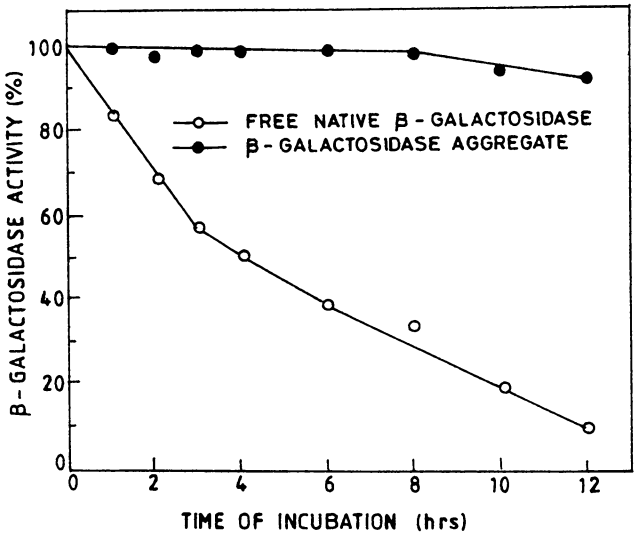


Figure 6. Thermal stability of β-galactosidase preparation at 55°C. The enzyme activity was determined with ONGP as the substrate. (Reproduced with permission from ref. 35. Copyright 1988 Indian Academy of Sciences.)

problem may be to entrap proteins by aggregation within commercially available beads. Chemical aggregation of β-galactosidase by glutaraldehyde inside Sephadex G-200 beads showed that about 17% of the total enzyme activity added at the start was present inside beads (*24*). The K_m towards ONGP (6.3×10^{-4} M) in this case was nearly the same as that of the simple aggregates (*24*). There is obviously need to improve upon these results but they are encouraging.

Thermostabilization of Concanavalin A (*25*). Concanavalin A (Con A), is a lectin of considerable biological interest. Our interest in thermostabilization of Con A stems from the usefulness of this lectin in affinity separations and bioaffinity immobilization. Crosslinking of Con A (1 mg/ml) was carried out at pH 7.5 using DMA (50 mg/ml) in the presence of α-methyl mannoside—a sugar specific for Con A. The crosslinked preparation (112% activity when measured in terms of its ability to precipitate glycogen) was fractionated on Mono S column of FPLC. It resolved into three fractions. The major fraction containing 52% of the total protein was further purified on Mono Q column of FPLC. About 80% of the protein did not bind to the column. This fraction (which did not bind to Mono Q column) showed a single band on SDS-PAGE corresponding to the position of Con A monomer. This, apart from establishing homogeneity, also ruled out inter subunit and intermolecular crosslinking.

The crosslinked derivative had 98% activity as compared to native Con A and free amino group analysis showed that 17% of amino groups were modified in this derivative. This corresponded to about 2 amino residues per Con A subunit. The gel filtration of the derivative on a calibrated column of Fractogel HW-55F showed its molecular weight to be identical with native Con A.

Thermal stability of this crosslinked derivative, native Con A and ethylacetimidate reacted Con A at 70°C is shown in Figure 8. Thus, mere modification of the amino groups in Con A with monofunctional analog did not lead to thermostabilization. These results indicate the formation of one crosslink per subunit in Con A molecule as the cause for thermostabilization. Based upon earlier structural data in the literature, the position of the crosslink can be tentatively assigned between Lys-135 and Lys-138. Becker et al. (*26*) have pointed out that residues 131-168 form part of disordered structure in the molecule. Perhaps, this crosslink introduces an element of order in this region, it reduces conformational flexibility and hence leads to considerable thermostabilization.

Bioconjugate Preparation

One class of bioconjugates are the protein–protein conjugates. Enzyme linked immunoassays (*27*), immunotoxins (*28*), and enzyme targeting (*29*) are some of the important areas where protein–protein conjugates have been used. Con A–ferritin (*30*) is yet another example of a useful protein-protein conjugate. Intermolecular crosslinking has also been employed for the preparation of "neomulti-enzyme complexes" for modelling segments of metabolic pathways (*31*). It was thought that heteroenzyme conjugates combining different enzyme specificities may be useful biochemical reagents e.g. for probing complex biological structures.

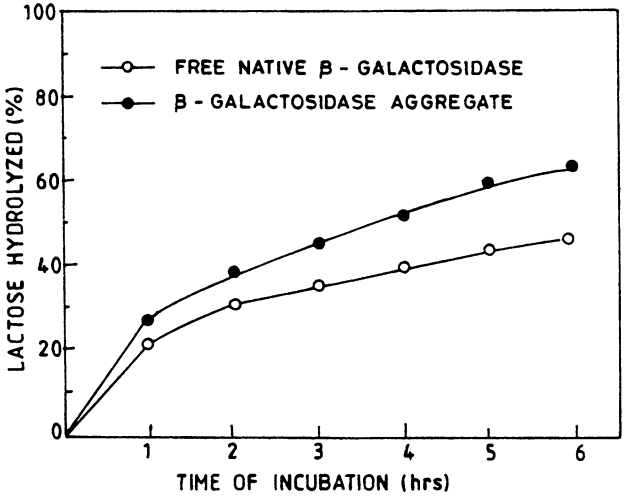


Figure 7. Continuous lactose hydrolysis of milk by β-galactosidase preparation at 55°C. (Reproduced with permission from ref. 35. Copyright 1988 Indian Academy of Sciences.)

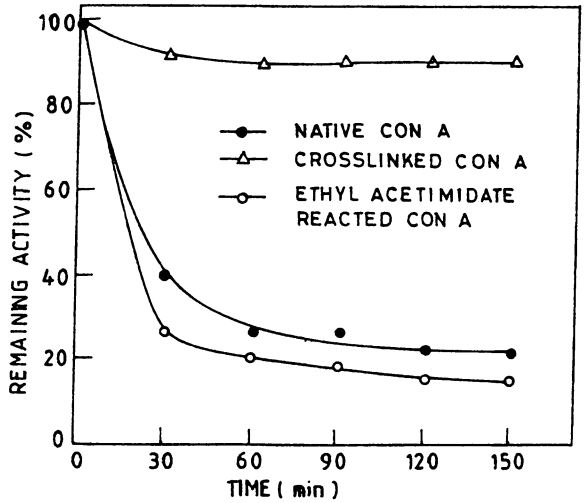


Figure 8. Thermal denaturation of native Con A, crosslinked Con A and ethyl acetimidate reacted Con A at 70°C. The solutions of Con A and its derivatives (1 mg/ml in each case) were made in phosphate buffer (50 mM, pH 7.0) containing 0.5 M NaCl. The lectin activity was assayed by measuring the precipitation of glycogen. (Reproduced with permission from ref. 19. Copyright 1988 Humana Press.)

Trypsin–Chymotrypsin Conjugate (*33*). This conjugate preparation was first attempted by glutaraldehyde. The conjugate formation could be detected by subjecting the reaction product to successive affinity chromatographies on affinity media specific for trypsin and chymotrypsin. The conjugate showed only 4% trypsin activity and 3% chymotrypsin activity assuming starting inputs of activities in each case to be 100.

The conjugate preparation was also attempted by N-succinimidyl pyridyl dithiopropionate (SPDP) (Figure 9). The main advantage of using such crosslinking reagents utilizing multi-step procedure is that undesirable side products (homoconjugates) are not formed. The use of SPDP did not result in greater yield of the desired trypsin-chymotrypsin conjugate. Trypsin component retained 56% activity and chymotrypsin component retained complete activity. The conjugate caseinolytic activity was compared with a mixture of (1:1) trypsin and chymotrypsin. If one remembers that trypsin had lost 46% activity, the conjugate seems to be an improved hybrid protease.

Trypsin–Chymotrypsin Coaggregate (*31*). Preparing insoluble heteroaggregates is yet another way of making enzyme–enzyme conjugates with twin activities and with the added advantage of easy separation and reusability. A coaggregate of trypsin–chymotrypsin prepared by extensive intermolecular crosslinking with glutaraldehyde showed significant reduction in the autolysis of the trypsin component (Figure 10). The aggregate retained 72% of the trypsin activity of native enzyme with benzoyl DL-arginine p-nitroanilide (BAPNA) as substrate, but only 4.8% of the chymotrypsin activity with benzoyl tyrosine ethyl ester BTEE as substrate. Further optimization is required so that greater retention of chymotrypsin is possible.

Trypsin–Alkaline Phosphatase Conjugate (*32*). Preparation of this conjugate was attempted by using glutaraldehyde and (DMA). The strategy for analysis and detection of the conjugate formation was similar to the one outlined above in the case of trypsin–chymotrypsin conjugate. The reaction mixture was chromatographed on benzamidine-agarose (*32*). The bound protein was eluted with benzamidine and assayed for alkaline phosphatase activity. No such activity could be detected in either of the cases. Another effort was made by oxidizing the carbohydrate moiety of alkaline phosphatase by periodate to generate aldehyde groups. This "activated" alkaline phosphatase was coupled to trypsin via Schiff base formation. In this case also, the conjugate formation was analyzed as above. Calculating on the basis of total units of alkaline phosphatase loaded on the benzamidine-agarose, only about 15% was found to bind to the affinity medium. The results did indicate that it is important to try different strategies for preparing bioconjugates. Incidentally, our failure with bifunctional reagents is understandable in view of the reported observation (*34*) that glycoenzymes are relatively "unreactive" towards chemical modification reagents since the protein surface is masked by carbohydrate moieties.

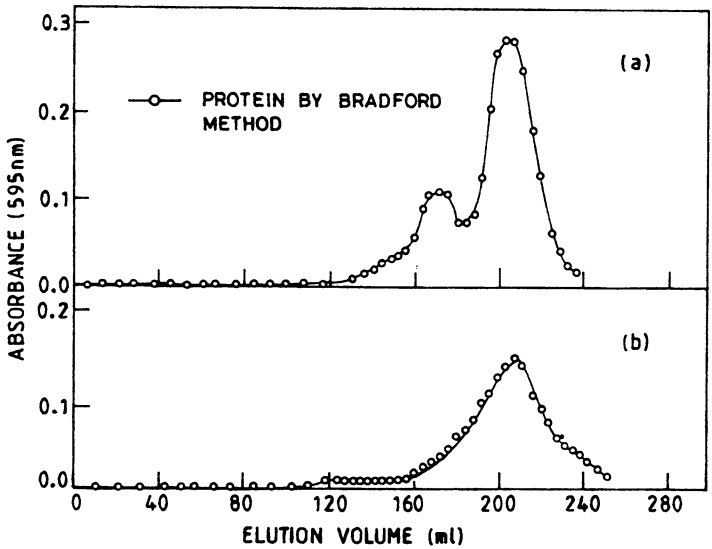


Figure 9. (a) Gel filtration of trypsin–chymotrypsin conjugate prepared by SPDP on Sephadex G-100.
(b) Rechromatography of fractions corresponding to elution volume from 140 to 180 ml after dithiothreitol (DDT) treatment. (Reproduced with permission from ref. 19. Copyright 1988 Humana Press.)

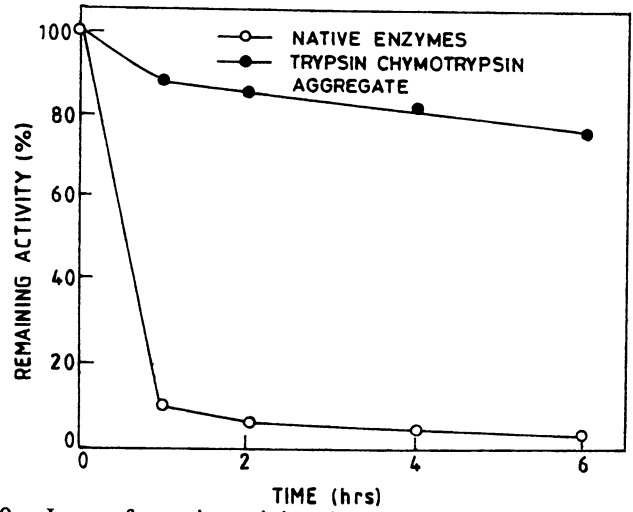


Figure 10. Loss of tryptic activity due to autolysis. The autolysis was carried out in Tris-HCl buffer (0.05 M, pH 8.2) containing 0.025 M ethylenediamine tetraacetate (EDTA) at protein concentration of 102 mg/100 ml at 40°C. Autolysis compared with native enzymes (trypsin and chymotrypsin) mixed in 1:1 ratio on mole basis. Remaining amidase activity was measured with BAPNA as the substrate.
(Reproduced with permission from ref. 22. Copyright 1988.)

Con A–β-Galactosidase Conjugate (*35*). Bioaffinity immobilization (*36*) has lately been considered as an attractive alternate to other methods of immobilization. In view of our interest in the application of β-galactosidase in the production of low lactose milk, a conjugate of Con A with *E. coli* β-galactosidase was prepared by using glutaraldehyde as the crosslinking reagent. The conjugate was expected to bind to Sephadex columns because of the affinity of Con A towards such columns. Recovery of enzyme activity (assayed by using ONGP as substrate) is shown in Table 2. The thermal stability of the Sephadex bound conjugate, measured in terms of survival of the enzyme activity, is shown in Figure 11. The results when the Sephadex bound conjugate was used for lactose hydrolysis at 50°C are shown in Figure 12. These results show that the approach is a promising one and preparation of such lectin-enzyme conjugates may considerably enlarge the scope of affinity immobilization since a large number of lectins and the corresponding affinity media to which they bind are commercially available.

A major problem in preparing protein-protein conjugates is the low yield generally obtained in almost all the cases. Development of more efficient crosslinking techniques and use of genetic engineering to produce fusion proteins are two solutions to this problem (*37, 38*).

Also, some recent successful efforts on developing a computer modeling procedure for assessing the stereochemical suitability of pairs of residues in proteins as potential sites for introduction of cystine bridges (*39*) is a pointer towards futuristic possibilities when it should be possible to choose a proper crosslinking reagent with much less uncertainty for both enzyme stabilization and preparation of protein-protein conjugates.

Conclusion

The results described in this chapter and elsewhere (*6, 40, 41*) show that chemical crosslinking is an attractive and simple approach to enzyme stabilization. Protein engineering (*5*), while most promising, still requires greater efforts. Medium engineering (*42*) for thermostabilization is still in its infancy whereas immobilization (*3*) has an inherent problem of diffusion constraints. However, crosslinking (as well as immobilization) merely prevents unfolding of the polypeptide chain. It does not abolish irreversible covalent changes (*2*) which take place under harsh conditions such as extremes of pH or high temperatures. When end applications require the protein to function under such conditions, protein engineering is perhaps the only solution.

The advent of the protein fusion techniques (*37, 38*) may seem to render intermolecular protein crosslinking obsolete for preparation of protein–protein conjugates. However, until such a time, the fusion techniques become more economical (if ever), chemical crosslinking, as illustrated by our results, would remain a simple and viable alternate.

Table 2. Recovery of Enzyme Activity at Various Stages of Preparation of the Conjugate

Sample	β-Galactosidase activity (%)
Con A–β-galactosidase solution	100
Con A–β-galactosidase solution, 30 min after adding glutaraldehyde	87
Effluent from Sephadex G-50 column eluted with 0.1 M NaCl	70
Effluent from Sephadex G-50 column eluted with 0.2 M glucose	10

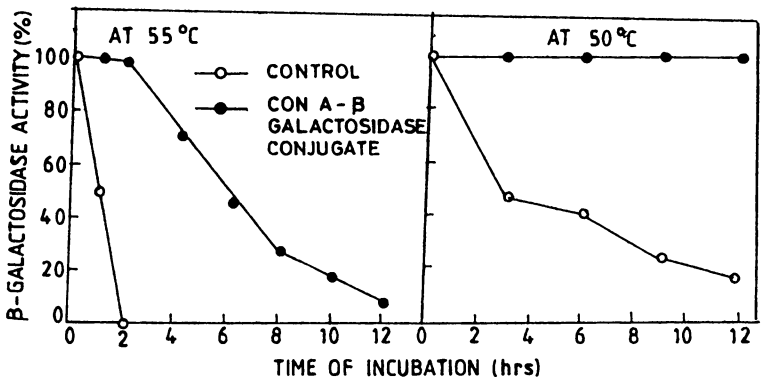


Figure 11. Thermal stability of Spehadex-bound Con A–β-galactosidase. (Reproduced with permission from ref. 22. Copyright 1988.)

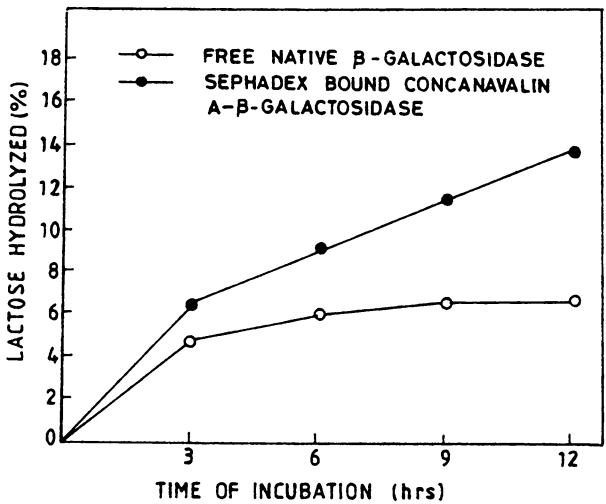


Figure 12. Lactose hydrolysis by Sephadex-bound conjugate at 50°C. The lactose solutions (5%) were prepared in potassium phosphate buffer (0.1 M, pH 7.2, containing 3 mM Mg++). The initial activities (using ONGP) as substrate) of free native enzyme and immobilized enzyme in the samples were identical.

(Reproduced with permission from ref. 23. Copyright 1988.)

Acknowledgement

The work described here was supported by the Department of Science and Technology (Government of India) and Council of Scientific and Industrial Research, India. I wish to thank Dr. Y.S. Rajput, Dr. A. Kamra, and Dr. S.K. Khare, who carried out the work described in this chapter. I also wish to thank current members of my research group particularly Dr. S. Ahmad and Dr. R. Tyagi for their assistance in the preparation of this chapter.

Literature Cited

1. Mozhaev, V.V.; Melik-Nubarov, N.S.; Sergeeva, M.V.; Sikrnis, V.; Martineck, K. *Biocatalysis* **1990**, *3*, 179-187.

2. Gupta, M.N. *Biotechnol. Appl. Biochem.* **1991**, *14*, 1-11.

3. Gupta, M.N.; Mattiasson, B. In *Bioanalytical Applications of Enzymes* Vol. 36; Suelter, C.H., ed., John Wiley: New York, New York, 1992; pp. 1-34.

4. Cabral, J.M.S.; Kennedy, J.F. In *Protein Immobilization: Fundamentals and Applications*; Taylor, R.F., ed., Marcel Dekker: New York, New York, 1991; pp. 73-138.

5. Nosoh, Y; Sekiguchi, T. *Biocatalysis* **1988**, *1*, 257-273.

6. Martinek, K.; Torchilin, V.P. *Methods Enzymol.* **1987**, *137*, 615-624.

7. Taniguchi, M.; Kobayashi, M.; Fujii, M. *Biotechnol. Bioeng.* **1989**, *34*, 1092-1097.

8. Poznansky, M.J. *Methods Enzymol.* **1988**, *137*, 566-574.

9. Senstad, C.; Mattiasson, B. *Biotechnol. Bioeng.* **1989**, *34*, 216-220.

10. Wold, F. *Methods Enzymol.* **1975**, *25*, 623-651.

11. Telefoncu, A. *Biotechnol. Bioeng.* **1983**, *25*, 713-724.

12. Von-Specht, B.U.; Brendel, W. *Biochim. Biophys. Acta* **1977**, *484*, 109-114.

13. Abuchowski, A.; Davis, F.F. *Biochim. Biophys. Acta* **1979**, *578*, 41-46.

14. Rajput, Y.S.; Gupta, M.N. *Enzyme Microb. Technol.* **1987**, *9*, 161-163.

15. Rajput, Y.S.; Gupta, M.N. *Enzyme Microb. Technol.* **1988**, *10*, 143-150.

16. Kosaric, N.; Asher, Y.J. In *Advances in Biochemical Engineering*; Fiechter, A., Ed.; Springer-Verlag: New York, New York, 1985; *19*, pp 25-60.

17. Kretchmer, N. *Sci. Am.* **1972**, *227*, 71-78.

18. Gekas, V.; Lopez-Leiva, M. *Process Biochem.* **1985**, *20*, 2-12.

19. Khare, S.K.; Gupta, M.N. *Appl. Biochem. Biotechnol.* **1988**, *16*, 1-15.

20. Jentoft, N.; Dearborn, D.G. *J. Biol. Chem.* **1979**, *254*, 4359-4365.

21. Pastore, M.; Morisi, F. *Methods Enzymol.* **1976**, *44*, 822-830.

22. Khare, S.K.; Gupta, M.N. *Biotechnol. Bioeng.* **1988**, *31*, 829-833.

23. Khare, S.K.; Gupta, M.N. *Biotechnol. Bioeng.* **1990**, *35*, 94-98.

24. Khare, S.K.; Vaidya, S.; Gupta, M.N. *Appl. Biochem. Biotechnol.* **1991**, *27*, 205-216.

25. Kamra, A.; Gupta, M.N. *Biochim. Biophys. Acta* **1988**, *966*, 181-187.

26. Becker, J.W.; Cunningham, B.A.; Reeke, G.N., Jr.; Wang, J.L.; Edelman, G.M. In *Concanavalin A as a Tool*; Bittiger, H; Schnebli, H., Eds.; John Wiley: New York, New York, 1976; pp 33-54.

27. Walker, J.M. In *Techniques in Molecular Biology*; Walker, J.M.; Gaastra, W., Eds.; Croom Helm: London, 1987; *2*, pp 82-97.

28. Ahmad, A.; Law, K. *Trends Biotechnol.* **1988**, *6*, 246-251.

29. Shier, W.T. *Methods Enzymol.* **1985**, *112*, 248-258.

30. Nicolson, G.L.; Singer, S.J. *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 942-945.

31. Ikura, K.; Okumura, K.; Sasaki, R.; Chiba, H. *Agric. Biol. Chem.* **1984**, *48*, 355-364.

32. Rajput, Y.S.; Gupta, M.N. *Biotechnol. Appl. Biochem.* **1988**, *10*, 424-250.

33. Rajput, Y.S.; Gupta, M.N. *Biotechnol. Bioeng.* **1988**, *31*, 220-223.

34. Royer, G.P. *Methods Enzymol.* **1987**, *135*, 141-146.

35. Khare, S.K.; Gupta, M.N. *J. Biosci.* **1988**, *137*, 47-54.

36. Saleemuddin, M.; Husain, Q. *Enzyme Microb. Technol.* **1991**, *13*, 290-295.

37. Sassenfeld, M.M. *Trends Biotechnol.* **1990**, *8*, 88-99.

38. Ong, E.; Gilkes, N.R.; Miller, R.C.; Warren, A.J.; Kilburn, D.C. *Enzyme Microb. Technol.* **1991**, *13*, 59-65.

39. Sowdhamini, R.; Srinivasan, N.; Shoichet, B.; Santi, D.V.; Ramakrishnan, C.; Balaram, P. *Protein Eng.* **1989**, *3*, 95-103.

40. Baker, J.O.; Oh, K.K.; Grohmann, K.; Himmel, M.E. *Biotechnol. Lett.* **1988**, *10*, 325-330.

41. Tatsumoto, K.; Oh, K.K.; Baker, J.O.; Himmel, M.E. *Appl. Biochem. Biotechnol.* **1989**, *20/21*, 293-308.

42. Gupta, M.N. *Env. J. Biochem.* **1992**, *203*, 25-32.

RECEIVED April 9, 1992

Author Index

Affiliation Index

University of Massachusetts, 265
University of Mississippi, 296
University of Rochester, 251
University of Texas Health Science Center, 167,265
University of Texas, 126,210
University of Virginia Medical School, 68
Yale University, 2

Subject Index

University of Massachusetts, 265
University of Mississippi, 296
University of Rochester, 251
University of Texas Health Science Center, 167,265
University of Texas, 126,210
University of Virginia Medical School, 68
Yale University, 2

Subject Index

In Biocatalyst Design for Stability and Specificity; Himmel, M., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1993.