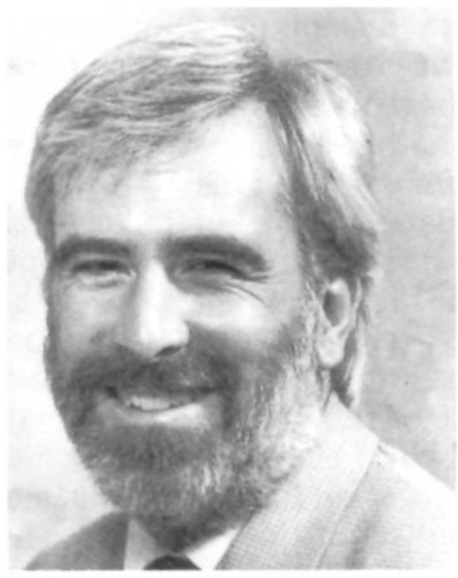
Structure–Function Relationships in Enzymic Catalysis. Can Chimeric Enzymes Contribute?

RICHARD N. ARMSTRONG

Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received February 26, 1990 (Revised Manuscript Received July 25, 1990)

Contents



Dr. Richard N. Armstrong is Professor of Chemistry and Biochemistry at the University of Maryland, College Park. Dr. Armstrong received his B.S. degree in Chemistry from Western Illinois University in 1970. He completed his graduate training in chemistry at Marquette University in the laboratory of Professor Norman E. Hoffman, where he received the Ph.D. degree in 1975. Following postdoctoral study at the University of Chicago with the late Professor Emil Thomas Kaiser, Dr. Armstrong moved to the National Institutes of Health as a Staff Fellow in 1978, where he worked with Dr. Donald M. Jerina. He joined the faculty of the Department of Chemistry and Biochemistry at Maryland in 1980. Dr. Armstrong's research interests are in bioorganic chemistry and mechanistic enzymology with particular emphasis on the enzymes of detoxication. His research group is currently involved in the elucidation of structure–function relationships in the glutathione transferases using physical–organic techniques and the modular assembly of hybrid proteins as well as mechanistic investigations of UDPglucuronyltransferases and epoxide hydrolase.

A. Introduction

1. Rationale for a Modular Approach to Structure–Function Studies

For the past two decades, the investigation of enzyme-catalyzed reactions has been clearly established as a discipline with its roots in physical–organic chemistry. However, until relatively recently the elucidation of relationships between structure and function in enzymic catalysis was largely confined to observation of the effect of reaction conditions and substrate structure on the rate or course of the reaction or the interpretation of three-dimensional structural information. An obvious deficiency in the experimental repertoire of the enzymologist early on was the inability (with the exception of the more or less specific technique of chemical modification of reactive functional groups) to manipulate the structure of the catalyst in a rational fashion. This situation was remedied, at least in part, a bit less than a decade ago with the advent of site-directed mutagenesis.[1,2] The refinement of this technique, particularly with respect to its limitations, has proceeded apace so that it is now a standard and often powerful tool available to the mechanistic enzymologist. Indeed in its most refined version, the site-specific replacement of one or several amino acid side chains in an enzyme guided by X-ray crystallographic data and followed up by a detailed mechanistic analysis can yield unique insight into structure–function relationships in catalysis. It is apparent, however, that an enormous amount of structure–function information is essentially inaccessible through this high-definition methodology. The two fundamental limitations are the slow rate at which sequence space can be sampled and the inherent boundary imposed on sequence space by the naturally occurring amino acids. The latter limitation has attracted considerable attention very recently. The expansion of sequence space by utilization of stop codons[3] or extension of the genetic alphabet[4] to allow incorporation of unnatural amino acids into proteins is possible, in principle, but several technical problems remain to be solved.

0009-2665/90/0790-1309$09.50/0

In contrast, the accelerated exploration of sequence space is currently possible by at least two quite different methods, which are in fact conceptually related to our current understanding of the evolution and refinement of protein function. For example, the creation of large libraries of random mutants, which can be accomplished by several methods,[5-8] coupled with an efficient selection (usually genetic) for a desired catalytic property provides a method to scan sequence space codon by codon. This process mimics the evolutionary refinement of function arising from point mutations in the genetic code. Alternatively, it is possible to scan primary sequence in chunks by modular replacement techniques that result in multiple mutations to a section of the structure thought to be important to a particular function. In this instance the section to be replaced and what it is to be replaced with are more or less rationally derived from other evolutionarily related proteins of somewhat different functional properties. This process crudely simulates the accelerated evolution of protein function by the shuffling of exons, a notion generally attributed to Gilbert.[9,10] The avowed objective of this experimental exercise is to assess whether a particular structural module can confer a defined functional characteristic of the donor to the newly constructed chimeric or hybrid enzyme. It is the manipulation of enzyme structure and function in this modular fashion that is the subject of this review.

Gilbert's hypothesis that a more facile evolution of protein function might be achieved from piecing together disparate exons by recombination within introns prompted Blake[11] to suggest that if exons corresponded to folded units of protein structure, then the probability that exon shuffling would actually result in a stable, useful translation product would be greatly increased. The experimental evidence for such a correlation is very suggestive but certainly not unequivocal.[12] The architecture of proteins can be described by structural motifs of different size and complexity including small compact structural units of 10–30 amino acids, folded functional units of perhaps 20 amino acids or more, and large domains of up to several hundred amino acids. Traut[12] has recently suggested that the average size of internal exons (20–50 codons) is entirely consistent with the normal size of a functional unit of protein structure. The imperfect correspondence of exons with specific structure–function modules is perhaps a vestige of the evolutionary decay of the relationship.[12,13] Although exploration of the relationship between the evolution of enzyme function and modules of protein structure may be sufficient justification for the study of hybrid enzymes, it need not be the only justification.

The replacement of defined structural motifs or regions of primary structure with related sequences is no less reasonable than the site-specific mutation of a single amino acid side chain; the changes are just a bit more complex. The rationale and limitations of the modular exploration of sequence space can be illustrated as follows. Consider, for example, that if one face of an α-helical segment of enzyme A is part of the surface of the active site, then it is not entirely unreasonable to expect that the transfer of that helical segment to a quite similar structural framework of a functionally related enzyme B would impart some of the catalytic character of A to the BAB hybrid. What *is* unreasonable is to expect that changes in the back face of the helix or in mutually compensating interactions with the rest of the active site would be without consequence. This inherent limitation is simply a manifestation of the fact that proteins are highly cooperative structures.

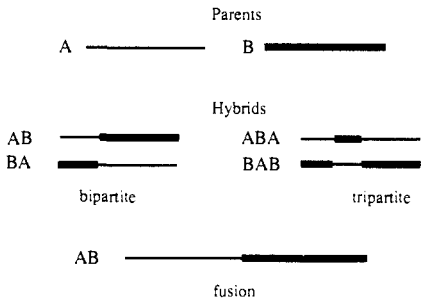


Figure 1. Common types of chimeric enzymes created from two related parents A and B.

2. Definitions, Nomenclature, and Methods

What is a chimeric enzyme? For the purposes of this review a chimeric enzyme is defined as a catalytic protein that is encoded by a gene constructed from at least two genetically distinct, natural or man-made parent genes. Although in the logical extreme site-specific mutant enzymes might be classified as chimera, they are not, by common understanding, considered so. Furthermore, enzymes with simple multiple mutations will not be considered chimeric unless they occur in a specific region of the primary structure and all intervening amino acids are derived from the same parent as the mutations. Not included in this definition are hybrid enzymes held together by noncovalent interactions or by chemical cross-links or hybrids of enzymes with other biopolymers (e.g., DNA, RNA, or carbohydrate).

Before proceeding, it is perhaps instructive to define the several types of hybrid enzymes that can be created and to briefly outline the methods which can be used to construct chimeric genes that encode hybrid catalysts. Usually, though certainly not always, chimeric enzymes are created from two or more related parent proteins as illustrated in the two-dimensional schematic diagrams of Figure 1. The several types of chimeric products can be classified as either bipartite hybrids which consist of the N-terminal section of one protein and the C-terminal section of another or as multipartite (usually tripartite) hybrids in which an internal section(s) of one polypeptide is replaced with the corresponding sequence of another. In these instances the hybrid product is roughly, if not exactly, the same length as either parent and is expected to possess the functional characteristics of one parent or the other. For reasons obviously related to the highly cooperative nature of protein structure, this expectation is often not the case. A different class of hybrids, referred to here as fusion proteins, result from the grafting of a functional domain of one protein onto essentially the entire sequence of another, creating a chimeric product substantially larger than either parent that exhibits the distinct properties of *both* parents.

There is no recommended nomenclature for hybrid enzymes. As a result, most authors create names or abbreviations appropriate to their own particular sys-

tem. Although the nomenclature usually used for site-specific mutants (e.g., E43D staphylococcal nuclease) can be applied to one of the parent proteins, in many cases it can become quite cumbersome as the number of mutations increases. A more general nomenclature that conveys the block construction of the hybrid but no specific sequence information has been adopted by the author. For example, a tripartite hybrid of enzymes A and B of identical length, a situation that is common with closely related isoenzymes, can be designated $A^{15}B^{17}A^{122}$, to indicate that of the 154 amino acids the first 15 are derived from A, the next 17 (residues 16–32) from B, and the last 122 (residues 33–154) again from A. For the more general situation where additions or deletions may appear in the sequence alignment of A and B or in the case of fusion proteins, the actual sequence interval of each parent can be specified in the superscript (e.g., $A^{1-15}B^{16-32}A^{33-154}$). In either situation the actual sequence changes in a hybrid can be readily derived by consulting the primary structure of the relevant segments of the parents.

The preliminary identification of functionally important modules to be targeted for investigation is best accomplished by examination of the three-dimensional structure of either parent enzyme or a homologous isoenzyme. Crystallographic information and model building are also quite useful in anticipating or interpreting the more than occasionally encountered "unusual result". It is of course possible to locate regions of interest in the absence of a three-dimensional structure by comparison of the primary structure of the parent proteins. As a general rule one might seek regions of relatively high homology in enzymes that are related by some distance in evolution. In contrast, hypervariable regions are probably the best targets for the analysis of structure–function relationships in very closely related isoenzymes. If genomic DNA sequences are known, then additional guidance may be obtained from the location of exon–intron interfaces since these junctions may define boundaries of structure–function units in the products of translation (vida supra).

The actual construction of chimeric genes to encode hybrid proteins can be accomplished by several techniques, the details of which will not be discussed here. Although the assembly of chimeric genes can, in principle, be achieved by the application of site-specific mutagenesis, the use of cassette mutagenesis[6,14] is much more efficient for the regional introduction of multiple mutations. Cassettes may be defined by naturally occurring restriction sites in the plasmid DNA, or appropriately located restriction sites may be introduced in the gene by site-specific mutagenesis. The cassettes may be obtained as restriction fragments of a gene encoding a donor protein or as synthetic oligonucleotide duplexes. However, the location of splice sites in the construction of chimeric genes need not be limited to available or engineered restriction sites. In vitro, primer-directed, methods can be used for the deletion or addition of modules to a parent protein.[15] Insertion of large (50–2000 bp) fragments of DNA (derived from gene A) can be inserted into a template gene (gene B) without use of restriction sites by a technique known as "sticky feet" directed mutagenesis.[16] Long primers with sticky feet are generated by amplification of the appropriate segment of gene A with the polymerase chain reaction using primers that bracket the segment of interest and that are tagged with the sticky feet. The long primer is annealed to a single-stranded template of gene B by virtue of the sticky feet sequences, which are complementary to the template.

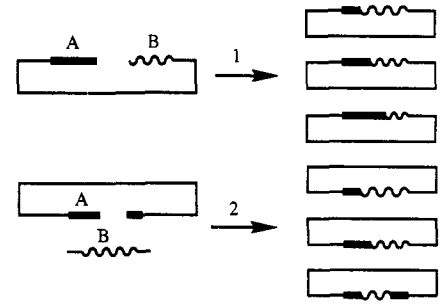


Figure 2. Illustration of the construction of libraries of chimeric genes encoding hybrid enzymes by (1) in vivo intramolecular recombination between partially homologous genes harbored on a single piece of linear DNA and (2) intermolecular recombinational gap repair between two linear gene fragments of partially homologous sequences.

All of the above methods are only useful for generating a single well-defined chimeric construct at a time. Techniques are also available for the assembly of a library of chimeric genes. For example, a library of chimeric genes encoding a family of hybrid proteins can be derived from two partially homologous genes by requiring an in vivo recombination event in the selection for recombinant cDNAs. Bipartite hybrid genes can be efficiently assembled in both *Escherichia coli*[17,18] and yeast[19] by intramolecular recombination within homologous segments of the regions encoding the two parent proteins as is illustrated in Figure 2. Intermolecular recombination in *E. coli* is also possible as demonstrated by Yanofsky and co-workers.[20] Tripartite hybrids can be created in yeast[19,21] by intermolecular recombinational gap repair of plasmid DNA linearized within the region encoding enzyme A by cotransformation of yeast with the linear plasmid encoding A and a partially homologous linear fragment encoding part of enzyme B (Figure 2). Both of these methods have the potential for generating a relatively large number of hybrid enzymes that can be used to map functional regions of the polypeptides. It should be pointed out that such methods do require, as does random mutagenesis, a detailed characterization of the products (e.g., sequencing the DNA).

3. Scope of This Review

Of the literally hundreds of hybrid proteins constructed in the past decade by recombinant DNA technology, only a handful have been created to address issues of structure–function relationships in enzymic catalysis. It is this small number of chimeric enzymes at which this review is primrily directed. Nonetheless the enormous number of other hybrid proteins (many of which are enzymes) and their utility do warrant some general comment by way of introduction. For example, a number of chimeric variants of specific binding proteins have been assembled for the purpose of elucidating structural motifs crucial to binding interactions. Particularly noteworthy examples are the DNA binding proteins between which the binding specificities have been interchanged by swapping DNA binding mo-

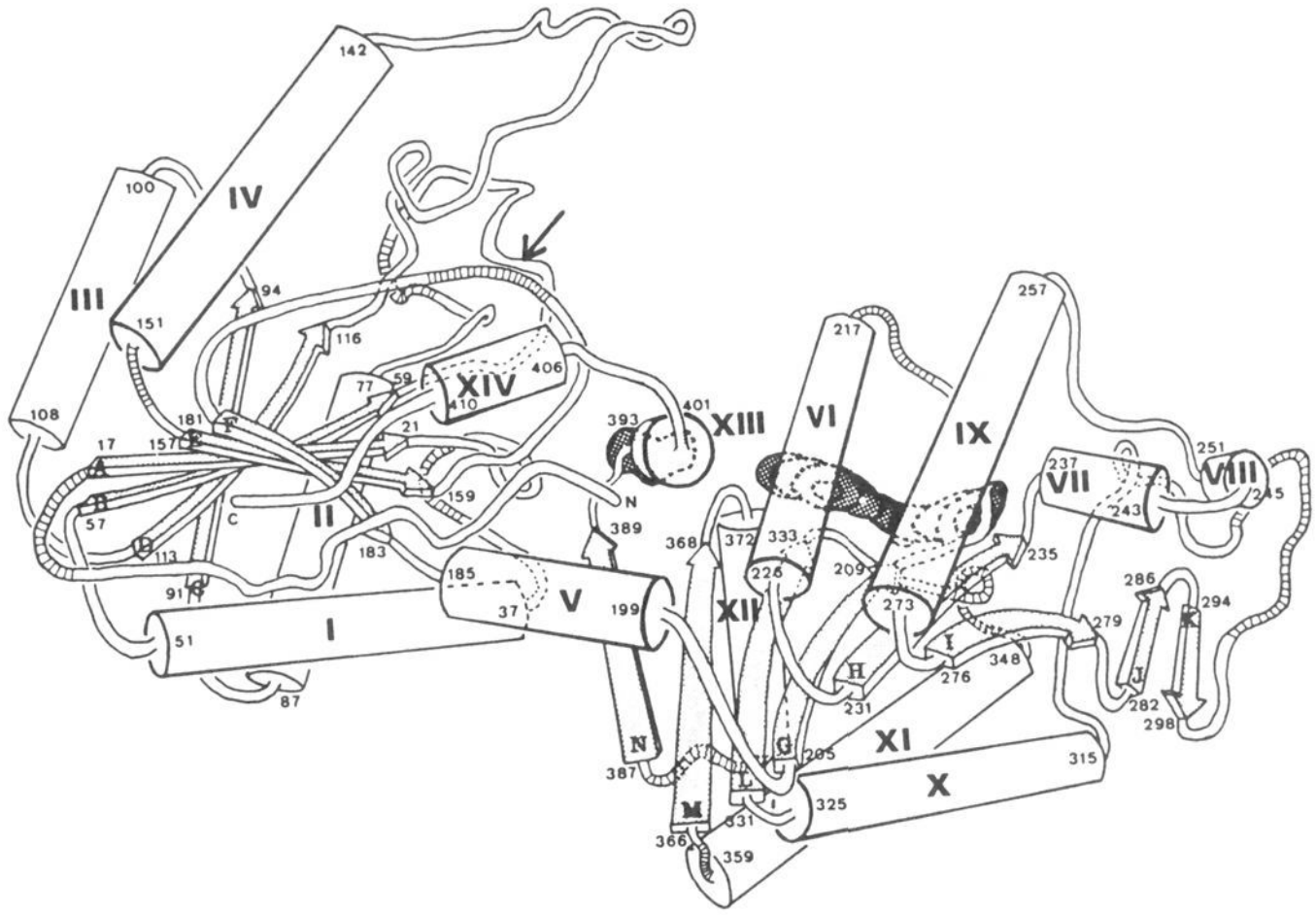


Figure 3. Schematic drawing (Reprinted with permission from ref 42. Copyright 1982 IRL Press.) of the three-dimensional structure of yeast phosphoglycerate kinase without substrates bound (open form, from ref 42). The C-terminal ATP binding domain is the lobe on the right. The arrow indicates the splice site in the N-terminal domain at residue 172 just before the β-strand E that precedes helix V. This helix links the two domains. The binding sites for ATP and 3-phosphoglycerate are shown behind helices VI and IX and at the end of helix XIII, respectively.

tifs.[22-24] Hybrid proteins have also proven to be quite useful for mapping residues or secondary structural regions important to the activity of hormones,[25] cytokines,[26] and regulatory proteins.[27] Fusion and chimeric proteins have also been utilized to examine oncogene activation[28,29] and structure–function relationships in drug receptors.[30] On a more practical note, chimeric exotoxins with altered cell binding properties have been shown to hold tremendous potential as clinically useful medicinal agents.[31]

Fusion proteins with catalytic activity have been particularly useful for the analysis of membrane protein topology.[32] In addition, the processing and translocation of secretory proteins have long been explored through the use of fusion proteins.[33-35] More recently, a systematic analysis of the requirements for the illicit secretion of a cytoplasmic protein (triosephosphate isomerase) led by the signal sequence and proximal sequences of β-lactamase has been performed by using a series of fusion hybrids.[36] Chimeric proteins have also been used to elucidate the requirements for N-terminal methionine excision by methionylaminopeptidase in *E. coli*.[37] In each of these instances the hybrid protein has served as a substrate for translocating or processing machinery with the enzymic activity of the hybrid, when present, serving a marker function.

The coverage of this review will be limited to chimeric enzymes that have been assembled for the express purpose of examining structure–function relationships in catalysis. Inasmuch as the modular approach to elucidating these relationships has really only developed in the past few years, this is clearly a review of an emerging field rather than a mature one. This review includes literature available up to about January 15, 1990.

B. Domain Interchange in Divergent Isofunctional Enzymes

Many proteins, particularly ones with molecular weights in excess of 20000, consist of two or more large, independently folded[38] functional domains. Each domain usually serves a particular, unique function (e.g., binding a substrate or regulatory molecule) while a connecting or hinge region facilitates the appropriate interdomain interactions necessary for regulation or catalysis. Isofunctional enzymes from highly divergent organisms often have quite similar backbone structures, even though their sequence homology may be 50% or less. However, it is not generally known whether selective pressure results in conservation of essential interdomain and hinge region interactions in such enzymes as it does with essential catalytic residues. The alternative possibility is that mutually compensating mutations might occur at the domain interface. Domain interchange between divergent enzymes is one way to examine these ideas.

1. Phosphoglycerate Kinase

The first well-defined domain swapping experiments were performed on 3-phosphoglycerate kinase and reported by Mas et al.[39] in 1986. Phosphoglycerate kinase catalyzes the reversible formation of 1,3-diphosphoglycerate from ATP and 3-phosphoglycerate. The enzyme from yeast exhibits 65% sequence identity with the mammalian proteins from horse and man.[40,41] The crystal structures of both the yeast[42] and horse[43] enzymes reveal that the protein consists of two distinct domains (residues 1–184 and 200–415) of roughly equal size linked by a 15-residue α-helical segment. This structural motif is readily apparent from the ribbon

tracing of the yeast protein illustrated in Figure 3. The two lobes are functionally distinct in that the ATP binding site is located in the C-terminal domain and the binding locus for 3-phosphoglycerate resides in the N-terminal lobe.[42,43] The active site is therefore located at the domain interface. In the absence of substrates the enzyme exhibits the rather open conformation shown in Figure 3 which appears to close upon binding of substrates.[43,44]

Two complementary chimeric phosphoglycerate kinases in which the two functional domains have been switched were constructed by using an existing *Nco*I site in codon 174 of the human structural gene and *Kpn*I and *Hpa*II restriction sites in codons 164 and 183, respectively, in the yeast gene. The gaps were spanned with the appropriate linkers such that the splice site in both hybrid genes was located at equivalent codons: 174 of the human gene and 172 of the yeast sequence. There are two natural deletions in the N-terminal half of the yeast sequence and one at the C-terminal of the human sequence. The two hybrids can be designated $H^{1-174}Y^{173-415}$ and $Y^{1-172}H^{175-416}$ following the nomenclature suggested in the Introduction.

The two parent and chimeric enzymes can be efficiently expressed in yeast.[39] It is quite striking that the two chimeric proteins are almost indistinguishable from one another or the parents with respect to their kinetic constants. Values of k_{cat} and k_{cat}/K_m for both ATP and 3-phosphoglycerate differ by no more than a factor of 2 for all four enzymes. The largest differences recorded were in the K_m for 3-phosphoglycerate, which was 1.5–2-fold higher in the two hybrids. This corresponds to a trivial difference in the binding energies of 0.2–0.4 kcal/mol. It is very obvious in this instance that the structures of the active sites at the domain interface as well as the hinge regions are insensitive to the 145 differences in primary structure between the yeast and human enzymes. The essential features of the hinge, the interdomain surfaces, and the active site have been conserved during evolution as revealed in the properties of the chimeric enzymes.

2. Aspartate Transcarbamoylase

Aspartate transcarbamoylase (ATCase) catalyzes the carbamoylation of aspartate by carbamoyl phosphate to form carbamoyl aspartate and inorganic phosphate, a key reaction in the pyrimidine biosynthetic pathway. Hamster ACTase is part of the CAD trifunctional protein complex which also includes the preceding (carbamoyl phosphate synthetase) and subsequent (dihydroorotase) enzymes in the pathway.[45] The CAD protein is a single polypeptide that is organized into three structurally distinct domains,[46] with the ATCase unit occupying the C-terminal portion of the protein. The CAD ATCase domain has been expressed in *E. coli* from the 3′-end of the hamster CAD cDNA.[47]

ATCase from *E. coli* is an allosterically regulated dodecameric protein that is functionally independent. The oligomeric structure of the *E. coli* enzyme consists of two catalytic trimers and three regulatory dimers, $[(C_3)_2(R_2)_3]$.[48,49] The catalytic trimer can catalyze the reaction in the absence of the regulatory dimers and under such conditions exhibits normal Michaelian kinetic behavior. The active sites are located at the three subunit interfaces of the trimer. In the holoenzyme the catalytic trimers are stacked one on top of the other. Thus, each catalytic subunit can be divided into a polar domain which caps the top and bottom of the hexameric catalytic core and an N-terminal equatorial domain that faces the waist of the holoenzyme.[49] The polar and equatorial domains are primarily involved in the binding of carbamoyl phosphate and aspartate, respectively.[49]

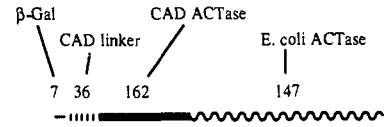


Figure 4. Schematic diagram of the primary structure of the CAD/*E. coli* aspartate transcarbamoylase hybrid described in ref 50. The hybrid consists of four segments of polypeptide, the first of which is a seven amino acid sequence derived from β-galactosidase. The next 198 residues are contributed by the CAD protein, of which 36 residues are from the linker region that bridges the ACTase domain to the rest of the CAD protein followed by the first 162 residues from the CAD ACTase. The 147 residues of the C-terminal portion of the chimera are contributed from the *E. coli* ACTase. Numbers indicate the number of residues encoded by each of the four gene segments.

Alignment of the primary structure of the ATCase domain of the hamster CAD deduced from the cDNA sequence reveals a 44% sequence identity with the *E. coli* catalytic subunit.[50] Wild, Davidson, and co-workers have recently constructed an active chimeric ATCase composed of the polar (C-terminal) domain, residues 162–312, of the *E. coli* catalytic subunit and the N-terminal portion of the CAD ATCase superdomain plus 36 amino acids which normally links the ATCase polypeptide to the rest of the CAD complex and 7 residues from β-galactosidase (a fusion artifact of the expression vector) as shown in Figure 4. Although this protein has not yet been characterized in detail, it appears to be catalytically competent since transformation of a *pyrB* auxotroph of *E. coli* with the plasmid harboring the chimeric gene allows growth on media without a pyrimidine supplement. A preliminary characterization of the enzyme suggests that it is not particularly stable, a property shared with the ATCase superdomain of CAD when expressed in *E. coli*. The formation of an active, hybrid ATCase with the polar domain derived from a bacterial enzyme and the equatorial domain from a highly divergent mammalian protein demonstrates again the very strong selective pressure for the conservation of essential interdomain structure and function.

3. Tryptophan Synthetase α-Subunit

Tryptophan synthetase is a bienzyme complex that catalyzes the two-step synthesis of tryptophan from indole 3-glycerol phosphate and L-serine. The enzyme is a tetramer with a subunit composition $\alpha_2\beta_2$. The α-subunit catalyzes the formation of indole and D-glyceraldehyde 3-phosphate from indole 3-glycerol phosphate. The β-subunit then catalyzes the condensation of indole with L-serine to yield tryptophan and water. The three-dimensional structure of the holoenzyme from *Salmonella typhimurium* reveals an approximately linear subunit arrangement, αββα, with a channel connecting the active sites of the α- and β-subunits.[51] The α-subunit polypeptide is 268 residues in length. The tertiary structural motif of the subunit in the holoenzyme is an α/β-barrel built from eight

repeating supersecondary structural units each containing a β-strand followed by an α-helix. The tertiary structure of the α-subunit appears to constitute a single domain. However, physical evidence for stepwise folding of the isolated α-subunit as well as the proteolytic division of the polypeptide at residue 188 into two folding units, designated α-1 (residues 1–188) and α-2 (residues 189–268), clearly suggests that the α/β-barrel consists of two independent folding domains.[52-54]

Yanofsky and co-workers were the first to construct a chimeric enzyme by the in vivo recombination of the genes encoding the α-subunit of tryptophan synthetase from *E. coli* and *S. typhimurium*.[20] The two α-subunits are 85% identical in primary structure with the majority of the mutations (25 of 40) occurring in the C-terminal domain of the protein as defined by the site of proteolysis. Intermolecular recombination of plasmids harboring parts of the *trypAB* gene from each organism permitted the construction and isolation of five chimeric α-subunits in which the N-terminal segment was derived from *S. typhimurium* and the C-terminal fragment from *E. coli*. The hybrid proteins which, because of the identical length of the parents, can be designated $S^{60}E^{208}$, $S^{98}E^{170}$, $S^{124}E^{144}$, $S^{173}E^{95}$, and $S^{183}E^{85}$ contain 6, 8, 12, 14, and 15 mutations in the N-terminal domain of the *E. coli* protein, respectively. The splice junctions are derived from the sites of recombination.[20] None of the hybrid subunits exhibited any significant difference in catalytic competence in the presence of the β-subunit as judged by the ratio of the β-reaction (indole to tryptophan) to the overall α+β reaction (indole 3-glycerol phosphate to tryptophan).[20] Little if any deterioration is evident in the ability of the chimeric subunits to complex with and activate the β-subunit. It is interesting to note, however, that most (63%) of the sequence differences in the two polypeptides are in the C-terminal domain (strand/helix combinations 6–8)[51] which is distal to the α,β-subunit interface in the holoenzyme.

Although the irreversible heat denaturation of the α-subunit in the presence of a crude extract containing the β-subunit suggests that four of the five hybrids are more heat labile than the parents,[20] a more detailed examination of the reversible denaturation of one of the hybrids, $S^{173}E^{95}$, is more revealing. This chimera is essentially composed of the N-terminal α-1 domain of the *Salmonella* subunit and the α-2 domain of the *E. coli* protein as defined by the proteolytic site at arginine 188.[52] Reversible denaturation of $S^{173}E^{95}$ with guanidinium chloride occurs in a stepwise process as it does with both parents. Furthermore, if it can be assumed that the α-2 domain is the first to unfold in each protein, then the two independent folding domains of the hybrid appear to inherit the stability of the domains of the parents from which they were derived. Moreover, the thermal stability of the hybrid is not significantly different from that of either of the parents. Thus, the hybrid subunit does not appear to be destabilized by any unfavorable interdomain interactions. It would be of interest to know if the complementary hybrid behaves in an analogous manner.

4. Thioredoxin

Thioredoxin is a small (12 000 Da) redox protein which serves as a reducing agent for ribonucleotide reductase and as a protein disulfide reductase in general.[55] It is also involved in the replication of some viruses in *E. coli*.[56] The active site of the protein is a redox active disulfide pair involving Cys32 and Cys35 in the *E. coli* version. The protein is converted to its active form by the enzyme thioredoxin reductase, an NADPH-dependent flavoprotein. The primary structures of the proteins from *E. coli* and *Anabaena* PCC 7119 exhibit 49% amino acid identity, with the *Anabaena* polypeptide having one less residue at both the amino and carboxyl termini.[57] The crystal structure of the oxidized form of thioredoxin from *E. coli* has been determined[58] and is illustrated in ribbon form in Figure 5. The active-site disulfide formed between C32 and C35 is located at the N-terminal end of the second α-helix in a highly conserved sequence WCGPC (residues 31–35). The degree of homology between the two proteins suggests that their three-dimensional structures should be quite similar.[59]

Gleason and co-workers have recently constructed two complementary chimeric thioredoxins from the genes encoding the two homologous proteins using a common restriction site located in the active site between the two codons for the cysteine residues.[59] The

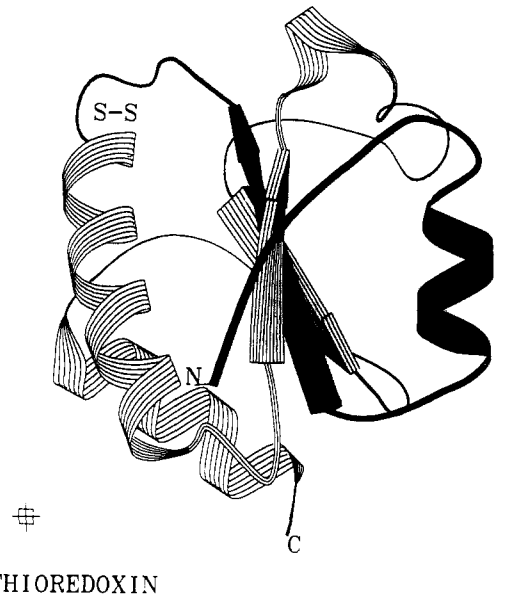


Figure 5. Ribbon representation of the three-dimensional structure[58] of oxidized thioredoxin from *E. coli*. The amino-terminal one-third of the molecule is shown in black. The position of the active-site disulfide is indicated in the upper left-hand corner. The diagram was produced by using the program RIBON, along with the coordinates on deposit in the Brookhaven Protein Data Bank.

TABLE I. Catalytic Efficiency of Reduction of Thioredoxins with *E. coli* Thioredoxin Reductase and Reduction of *E. coli* Ribonucleotide Reductase by Thioredoxins[a]

parent or chimeric thioredoxin	reduction by thioredoxin reductase		reduction of ribonucleotide reductase	
	k_{cat}, s^{-1}	k_{cat}/K_m, M^{-1} s^{-1}	k_{cat}, s^{-1}	k_{cat}/K_m, M^{-1} s^{-1}
E. coli (E^{1-108})	39	1.9×10^7	2.0	1.0×10^6
Anabaena (A^{1-106})	39	2.3×10^6	2.0	1.0×10^6
$A^{1-32}E^{34-108}$	52	5.2×10^7	2.7	2.7×10^6
$E^{1-33}A^{33-106}$	29	2.9×10^5	0.80	2.4×10^5

[a] Data retabulated from ref 59.

two chimeric proteins thus contain the N-terminal one-third of one protein with the remainder derived from the other. Although the protein does not have any obvious domain structure, the net result of this construction is that the first α-helix and large β-strand are transferred from one protein to the other (Figure 5). As is evident from Table I, both oxidized chimeric cofactors are substrates for thioredoxin reductase from *E. coli* and are competent to reduce ribonucleotide reductase from *E. coli*. It is also very interesting that the $A^{1-32}E^{34-108}$ hybrid is somewhat better at both functions than either of the parents. Although it might be concluded from these data that the C-terminal two-thirds of the protein is more important for productive interactions with the *E. coli* thioredoxin reductase and ribonucleotide reductase, it is evident from tryptophan fluorescence spectra of the oxidized and reduced hybrids that there is a signifiant change in the environment of $Trp28^{59,60}$ near the active site or perhaps of Trp31 in the active site. The exact nature of the conformational difference between this hybrid and the parent molecules that contribute to its enhanced electron-transfer efficiency remains to be determined.

C. Modular Replacement of Subdomains

Although it is reasonable to expect that the replacement of an entire domain of an enzyme with one derived from a structurally related isofunctional species, as above, might result in a structurally robust if not functionally viable molecule, the fact remains that there are no general rules for predicting the impact of domain replacement on function. The same may be said for the much more modest venture of the replacement of subdomains or even individual secondary structural elements. Intuition suggests that an enzyme is likely to be much more sensitive to structural alterations made in interior regions or in folding units of relatively high conformational order such as α-helices or β-strands than to changes at the surface or in conformationally flexible loops. The experimental pursuit of guidelines for the modular modification of small local folding units in enzymes is now underway as illustrated by the following examples.

1. α-Helical Segments and Loops in Alkaline Phosphatase

Bacterial alkaline phosphatase, the *pho*A gene product of *E. coli*, has recently been used as a vehicle for ascertaining the sensitivity of an enzyme to the shuffling or replacement of small secondary structural units. The three-dimensional structure of the dimeric enzyme has been determined to a resolution of 2.8 Å (Figure 6).[61] Dubose and Hartl[62] have recently examined the impact of the replacement of several α-helical segments identified by crystallography with other α-helical units or with sequences that are not likely to form helices. Three of the 14 α-helical regions identified by X-ray crystallography,[62] helices 2 (residues 55–66), 4 (residues 103–110), and 7 (residues 171–178), were chosen for modification or replacement. Each segment differs in its location in the protein, as illustrated in Figure 6. For example, helix 7 is on the periphery of the enzyme, helix 4 is just to the carboxyl-terminal side of Ser102 in the active site, and helix 2 is part of the interface between the two subunits. The types of modifications made in

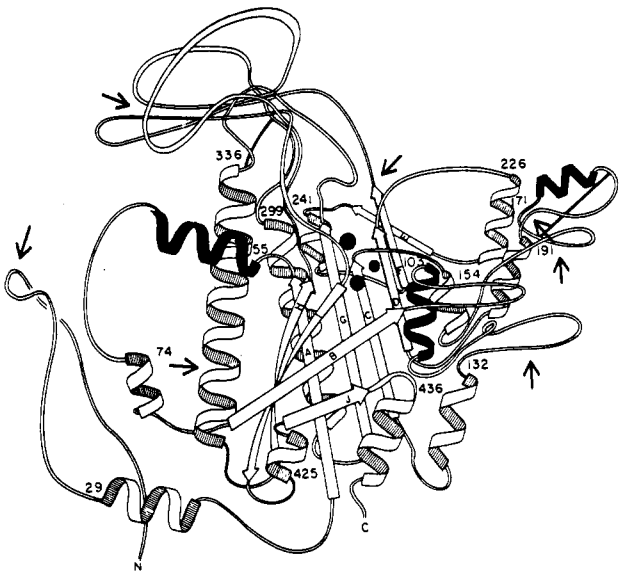


Figure 6. Representation of the three-dimensional structure of the monomer of *E. coli* alkaline phosphatase (Reprinted with permission from ref 61. Copyright Academic Press Limited.). The three helices that were targets for modification described in ref 62 are shown in black. Arrows indicate the approximate sites of insertion of peptide sequences as described in ref 66.

each helix included simple single or double mutations not expected to alter the potential for helix formation, the partial or complete replacement of the helix with a helical segment from another source, and finally the replacement of the segment with a random array of amino acids.

The introduction of conservative single or double mutations in the three helical segments did not significantly alter the catalytic efficiency of the enzyme. To the contrary, replacement of each helix, by frameshift mutations, with a random stretch of amino acids not expected to form an α-helix abolished catalytic activity in all cases. The more conservative replacement of the α-helices with parts of helical units from other sources gave chimeric proteins with catalytic efficiencies ranging from undetectable to greater than that of wild type as shown in Table II.[62] A number of general observations were made. First, not surprisingly, each helix exhibits a different sensitivity to modification which is perhaps an index of their functional role in structure and catalysis. The most interesting finding is that the N-terminal halves of the helices seem to be much more sensitive to modification than the C-terminal portion. This is clearly true for helices 4 and 7 as illustrated by the data of Table II. Although this might be rationalized in the case of helix 4, in which the N terminus of the helix points at the active site, it is not evident why this should be true for helix 7, which is located on the edge of the molecule. Whether this may turn out to be a very general observation remains to be seen. Even though it appears that the disruption of a helical unit is generally disastrous, the converse of merely conserving the helix-forming potential of the sequence is not sufficient to guarantee a happy catalyst. This is apparent from the fact that helix 2, located at the subunit interface, is not tolerant of replacement with a synthetic structural unit known to form an α-helix in solution[63] but that introduces a number of changes in charge into the region. Nevertheless, it is obvious that it is possible to substitute new α-helical

TABLE II. Helix Replacement in Alkaline Phosphatase[a]

helix	composition of native and chimeric proteins	relative k_{cat}/K_m
2	.....AspSerGluIleThrAlaAlaArgAsnTyrAlaGlu.....	1.00
2-1[b]	GluLeu Leu LysLysLeu	0
2-2	Leu GluGlu LeuLys Gly	0
2-3	GluLeu Leu LysLys LeuLeu GluGlu LeuLys Gly	0
4	.....AlaAlaSerAlaThrAlaTrpSer.....	1.00
4-1[c]	Ser Ala Thr Ser	0.02
4-2	Glu Lys	0.97
4-3	Ser Ala Thr Ser Glu Lys	0.01
7	.....ProSerAlaThrSerGluLysCys.....	1.00
7-1[d]	Gln Thr Ala Ala	0.21
7-2	Leu Arg Asp	1.42
7-3	Gln Thr Ala AlaLeu Arg Asp	1.21
7-4	Thr Asn Ser Leu	0.02
7-5	Arg Met Leu	1.23
7-6	Thr Asn Ser Leu Arg Met Leu	0.07

[a] Data and sequence information are retabulated from ref 62. [b] Helical replacement sequences for 1-1–2-3 were derived from a synthetic α-helical peptide.[63] [c] Sequences of chimeras 4-1–4-3 are from the native helix 7 sequence. [d] Replacement sequences for 7-1–7-3 are derived from helix 1 (residues 29–35) of the native protein and those for 7-4–7-6 are from an α-helical segment of bacteriophage T4 lysozyme.[64]

units with nonconservative mutations into alkaline phosphatase without, in the majority of instances, severe impairment of catalysis.

Loops in the tertiary structure of proteins often occur at the surface of the molecule and are relatively flexible with respect to the rest of the protein scaffold. The N and C termini of loops are, in many instances, close to one another with several amino acids looped out into solvent such that the structure resembles the Greek letter Ω. These Ω-loops[65] are often associated with sites of deletion or insertion found in the primary structures of related isofunctional enzymes. The tolerance of loops to variation in length is a property of enzyme structure that is not well understood. Preliminary attempts to address this problem have been recently reported, again with alkaline phosphatase as the test system.[66] The insertion of di- and tripeptide sequences into five different loop regions (see Figure 6) of alkaline phosphatase results in the retention of significant catalytic activity in each instance. Insertions in loops after residues 13, 129, 166, 190, and 383 gave enzymic activities in the periplasm of 3, 24, 84, 102, and 71%, respectively, of wild-type activity per cell. Although the insertion after residue 13 interfered with the processing and transport of the protein into the periplasm, most of the other insertions had only a marginal effect. It was also possible to insert a pentadecapeptide analogue of dynorphin into the sites after residues 129 and 166 without disruption of either transport, processing, or catalytic activity of the enzyme. Quite in contrast, insertion of ProGly after residue 258 in β-strand F or after residue 350 in α-helix 12 results in inactive enzyme. Although none of the chimeras in this study were characterized with respect to their kinetic behavior, the lesson is relatively clear; loops are considerably more tolerant of modification than other secondary structural elements.

2. Ω-Loops in Iso-1-cytochrome *c*

The cytochromes *c* provide a very instructive example of the malleable nature of Ω-loops in proteins as recently pointed out by Sherman and co-workers.[15] The crystal structures of both prokaryotic and eukaryotic cytochromes *c* reveal that the tertiary structures of the proteins are highly conserved despite differences in the primary structures.[21,67-71] Crystal structures are available for the proteins from *Rhodospirillum rubrum*,[67] *Paracoccus denitrificans*,[68] tuna,[69] *Pseudomonas aeruginosa*,[70] and yeast.[71] The basic structural motif is a heme binding pocket constructed of α-helices connected by four Ω-loops. Interspecies comparisons show that the four loops are of variable length but are located in the same positions in the tertiary structure from one species to the next. The first indication of the importance (or lack thereof) of the Ω-loops in iso-1-cytochrome *c* from *Saccharomyces cerevisiae* was recently investigated with respect to both the biosynthesis and function of the molecule.

The four Ω-loops in yeast iso-1-cytochrome *c*, designated loops A–D, encompass residues 23–37, 39–48, 45–59, and 75–89, respectively; loops B and C overlap. Deletions in the central portion of each loop were found to have quite different consequences for the biosynthesis and function of the protein in yeast. For example, yeast harboring plasmids encoding deletions of amino acids 27–33 in loop A or 78–83 in loop D were completely deficient in iso-1-cytochrome *c*. These deletions result in either impaired biosynthesis, increased lability, or a significantly altered structure of the protein. However, deletion of residues 40–45 in loop B or 48–55 in loop C results in a reasonable level of synthesis (40–60% of normal) and a measurable cytochrome *c* activity (10–20% of wild type). Loops B and C are obviously not crucial to either the biosynthesis or electron-transport function of the protein.

Although yeast does not tolerate the partial deletion of loop A, it does, to varying degrees, accommodate replacement of the loop with analogous sequences of the same length from other species (tuna or *R. rubrum*) or of shorter length (e.g., *P. aeruginosa*) and of longer length (*P. denitrificans*). Heterologous substitution of a longer sequence from porcine pancreatic esterase was also tolerated. The levels of biosynthesis of the five chimeric proteins ranged from 30 to 100% of normal. However, the activity of the hybrids varied considerably, from 0 to 90%. The chimera with a replacement sequence three residues shorter than normal (from *P. aeruginosa*) was not functional. Thus, it would appear that yeast will tolerate replacement of loop A with a longer sequence even from an entirely unrelated protein but is quite intolerant of significant deletions in the same region.

Unfortunately, it is not known which steps in the biosynthesis of the protein are sensitive to loop deletion and replacement. A number of steps in the biosynthesis might be affected, including translation, binding of the apoprotein to the mitochondria, heme attachment catalyzed by heme lyase, or mitochondrial import. In addition, for those hybrids that are successfully synthesized, it is not clear how or to exactly what extent the electron-transport properties of the protein are im-

Figure 7. (A) Sequences of ribonuclease A (shown in bold type) and angiogenin (in outline type) and the hybrid angiogenin ARH-I containing a fourth disulfide loop derived from RNase A as described in ref 73. (B) N-Terminal sequences of RNase A (bold type), human angiogenin (outline type), and the hybrid ARH-III.[80]

paired. Although the loops are not always essential to the electron-transport function, the fact that this function appears to be impaired to different extents with different hybrids has been interpreted to indicate that the loops are involved in surface interactions of cytochrome *c* with its physicological partners.[15]

D. Recruitment of Catalytic Specificity

A few attempts have been made to actually recruit a new or enhance an existing catalytic activity of an enzyme by the transplantation of a portion of the active site of one enzyme into a compatible tertiary framework of another evolutionary kin. Several recent papers discussed below tend to indicate that this basic strategy may be generally applicable to the design of new catalysts and the exploration of structure–function relationships in enzymes. In special cases, for example, with enzymes having macromolecular substrates, an additional substrate recognition site can be added distal to the catalytic site by the construction of a fusion protein.

1. Angiogenin/Ribonuclease

Human angiogenin is a ribonucleolytic protein that is related to the secretory and nonsecretory ribonucleases (RNases) but is involved in the induction of human blood vessels.[72] Although angiogenin shares 35% sequence identity with pancreatic ribonuclease including the conservation of several well-characterized catalytic residues,[73] it has a very different ribonucleolytic activity toward conventional substrates; it is considerably less efficient.[74] Furthermore, chemical modification and site-specific mutagenesis studies have shown that active-site residues Lys40, His13, and His114 are essential for both the angiogenic and ribonucleolytic activities.[73,75-77] Conversely, even though these residues are conserved in the active site of RNase, the enzyme does not possess detectable angiogenic activity. By a careful comparison of primary structures and the crystal structure of RNase S[78] and a calculated three-dimensional structure of angiogenin,[77] Vallee and co-workers have identified regions that are unique to angiogenin and therefore may constitute structural features critical to the unique properties of the protein.[73,80]

Residues 55–75 of human angiogenin correspond to residues 62–71 of mammalian pancreatic ribonuclease which, in the latter, forms a loop that constitutes part of the purine binding site.[78] The sequence homology in the central portion of this region (residues 62–71) is very low, with the notable absence of the disulfide bond found between Cys65 and Cys72 that is found in RNase as illustrated in Figure 7A. Furthermore, the loop in angiogenin is two residues shorter than in RNase. Harper and Vallee first constructed the angiogenin–ribonuclease hybrid shown in Figure 7A by cassette or, as these authors prefer, "regional" mutagenesis.[73] The resulting chimeric protein, ARH-1 ($A^{1-57}R^{59-73}A^{71-123}$), in which residues 58–70 of human angiogenin were replaced by the equivalent residues (59–73) of RNase A, was found to have properties intermediate between the two parents. The hybrid folds and corretly forms all disulfide bonds including the one introduced from RNase A. However, it has significantly less angiogenic potency but much higher ribonucleolytic activity toward conventional substrates than does native angiogenin.[73] The introduction of the loop and disulfide bond apparently impairs the angiogenic activity and enhances the enzymic activity of the molecule.

The enhanced ribonucleolytic activity has been investigated in some detail.[73] The activity of the hybrid toward tRNA was about 300-fold greater than that of angiogenin and approximately 200-fold less than that of RNase A under the same conditions. Similar results were obtained with wheat germ RNA, 18S and 28S rRNA, and the homopolymers poly(C) and poly(U).

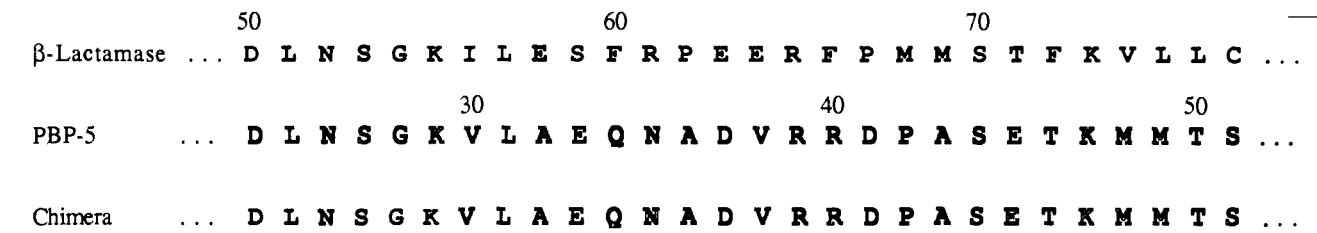


Figure 8. Amino acid sequences of the RTEM-1 β-lactamase (bold type), the corresponding sequence of the PBP-5 enzyme (outline type), and the chimera as described in ref 84. The active-site serine is residue 70 in the β-lactamase sequence.

With the latter substrates the chimera was roughly 200-fold more active than native angiogenin but less active than RNase A by a factor of 10^3. Kinetic results obtained with dinucleoside 3′,5′-phosphate substrates reveal that modest changes in the catalytic specificity accompany the enhanced efficiency. For example, the relative order of k_{cat}/K_m for angiogenin toward four such substrates is CpA > CpG > UpA > UpG, but with the hybrid, which is up to 2 orders of magnitude more efficient, it is CpA ≫ UpA > CpG > UpG. This order is qualitatively similar to that for RNase A, although the k_{cat}/K_m values of RNase A are 10^3–10^4 greater than those of the hybrid.

The residues 8–21 region of human angiogenin is a highly conserved region of the molecule found in angiogenins from four mammalian species. The region differs substantially (Figure 7B) from the corresponding region of RNase A which, as it happens, is poorly conserved in the pancreatic ribonucleases and is not required for RNase A activity. Bond and Vallee reasoned that this highly conserved region in mammalian angiogenins, which is completely nonessential to RNase A, may be one of the structural keys to the unique catalytic and biological properties of the angiogenin molecule.[80] To test this hypothesis, another angiogenin–ribonuclease hybrid (ARH-III) was constructed as illustrated in Figure 7B. The hybrid protein, here designated $A^{1-7}R^{7-21}A^{23-123}$, was assembled by cassette mutagenesis such that residues 7–21 of RNase A were substituted for the angiogenin sequence containing residues 8–22.

Substitution of the RNase sequence in the amino-terminal region of angiogenin has virtually no effect on the ribonucleolytic activity of the protein toward dinucleotide substrates or naked 18S and 23S rRNA. In contrast, this hybrid was 20–30-fold less effective than angiogenin in inhibition of protein synthesis in a rabbit reticulocyte assay. This latter activity is the result of a very specific cleavage of 18S rRNA in the 40S subunit of the ribosome. Interestingly, the hybrid elicits its full angiogenic activity at a dose that is roughly 10-fold lower than that required with human angiogenin. The full activity is about 80% of the human protein. Finally, the placental ribonuclease inhibitor binds 10 and 600 times more tightly ($K_i \leq 7 \times 10^{-17}$ M) to the hybrid than it does to angiogenin and RNase A, respectively. Together these results suggest that the general ribonucleolytic activity of angiogenin is not coupled to either the angiogenic or protein synthesis inhibitory activities. That the hybrid has altered properties with respect to the latter activities is a strong indication that the highly conserved region near the N terminus is responsible, to some extent, for the unique biological properties of this protein.

Even though the precise structural details of changes in the substrate or physiological specificity of angiogenin introduced by the transfer of domains from a distantly related enzyme are not known, this work remains an elegant demonstration of the power of scanning sequence space in chunks rather than residue by residue. Complementary experiments designed to enhance the angiogenic activity of RNase are desirable to further test the generality of domain transfer in sorting out structure–function relationships between these related classes of proteins. It should be noted that some experiments complementary to these have been conducted by Benner and Allemann in which a nine amino acid region of the angiogenin sequence has been substituted for the corresponding sequence in a secretory RNase to yield a molecule with altered ribonucleolytic activity.[81] The details of these studies should appear shortly.

2. β-Lactamase/Penicillin-Binding Proteins

The penicillin-binding proteins (PBPs) catalyze the cross-linking of the peptidoglycan in bacterial cell wall synthesis by cleavage of a D-Ala–D-Ala peptide bond at the C terminus of one polypeptide chain and transferring the resulting N-terminal segment to the N terminus of another polypeptide. These proteins bear a remarkable resemblance to the class A β-lactamases which catalyze the hydrolysis of β-lactam antibiotics, structural analogues of the D-Ala–D-Ala dipeptide. The substantial similarity of the primary and secondary structures suggests a strong evolutionary relationship between these two classes of enzymes.[82] For example, the RTEM-1 β-lactamase and the PBP-5 carboxypeptidase from *E. coli* share a 27% sequence identity. Despite the similarities in amino acid sequence, particularly in the active-site region, the β-lactamases do not normally catalyze the hydrolysis of acyclic peptide substrates, nor do the PBPs catalyze the efficient hydrolysis of β-lactam antibiotics. Instead, the PBPs usually form a stable acyl enzyme intermediate in their attempt to hydrolyze β-lactams. This is one mechanism by which the β-lactams exert their antibiotic action.[83]

Richards and co-workers have very elegantly demonstrated that the carboxypeptidase activity of PBP-5 can be recruited into the RTEM β-lactamase structure by the transplantation of a 28 amino acid module from the active site of PBP-5 into the active-site region of the β-lactamase.[14,84] The resulting chimera shown in Figure 8 was found to have substantially lower β-lactamase activity toward benzylpenicillin compared to that of the parent RTEM β-lactamase; k_{cat}/K_m is reduced by a factor of about 10^6. More interesting, however, is the fact that the hybrid protein has a detectable D,D-carboxypeptidase activity which is about 0.5–1% (comparing k_{cat}/K_m for two synthetic substrates) of that of the native PBP-5. This is quite remarkable considering that the parent RTEM β-lactamase has no de-

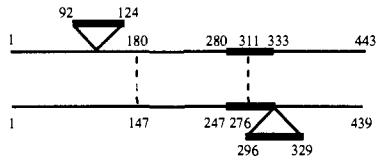


Figure 9. Linear representation of the amino acid sequence alignments of the Φ3T and SPR DNA methyltransferases.[86] The bold lines indicate the variable regions of sequence in the two proteins. The dashed lines show the positions, relative to the protein sequence, of common restriction sites in the genes used to construct the various chimeric enzymes. Amino acid residue numbers at the important splice sites for each protein sequence are indicated.

tectable carboxypeptidase activity and that the chimera retains 93% of the sequence of the β-lactamase polypeptide. However, the chimera does not catalyze the transpeptidation reaction, which is a general property of the PBPs.

An interesting and important point concerning this particular chimeric enzyme is the apparent necessity for the presence of a β-lactam to assist in the proper initial folding of the hybrid polypeptide into an active enzyme. The chimera is labile to heat inactivation at relatively low (37 °C) temperature in the absence of a β-lactam and exhibits a pronounced induction period for its β-lactamase activity. These facts suggest that the hybrid protein has a defect in its folding and conformational stability, observations which are likely to be common with chimeric enzymes. Although it is not yet clear which residue changes are most important for the recruitment of the carboxypeptidase activity, the availability of three-dimensional structures of both β-lactamases and PBPs should aid in the future refinement of chimeric constructs.

3. Chimeric DNA Methyltransferases

In principle it is possible to locate regions in the primary structure of an enzyme crucial for substrate recognition by the construction of chimeric enzymes of two homologous parents possessing different catalytic properties. The successful recruitment of the catalytic specificity of one enzyme into another by the transfer of a specific region of sequence can be taken as evidence for the participation of that region in catalysis. As an example, chimeric DNA methyltransferases have been used to partially define amino acid sequences of the enzymes that are involved in substrate recognition.

DNA methyltransferases (Mtases) catalyze the sequence-specific transfer of the methyl group of *S*-adenosylmethionine to DNA. The Mtases from the *Bacillus subtilis* phages SPR and Φ3T catalyze the specific methylation of the sequence GGCC but differ in their ability to methylate other related sequences.[85] Thus, the SPR Mtase characteristically methylates the sequences CCGG and CC(A/T)GG, while the Φ3T enzyme is specific for the sequence GCNGC. Although the two enzymes share a high amount of sequence homology, each contains one unique region of primary structure (residues 92–124 in the Φ3T enzyme and residues 296–329 in the SPR Mtase) and a common variable region (residues 280–333 and 247–295 in the Φ3T and SPR proteins, respectively) as illustrated in Figure 9.[86]

To test the reasonable hypothesis that the structural differences which distinguish the sequence specificity

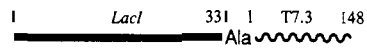


Figure 10. Schematic representation of the fusion hybrid of the *lac I* protein and the T7.3 single-strand-specific endonuclease.[87] The numbers delineate the sequences contributed by each parent polypeptide. The alanine residue joining the two segments is a result of the splicing strategy.

of the two enzymes reside in the variable regions of the primary structure, Trautner and co-workers[85] constructed a series of chimeric plasmids encoding hybrid Mtases and followed changes in Mtase specificity by determining the resistance of the plasmid to various restriction enzymes that are sensitive to methylation of the restriction site. The two chimeric constructs, $\Phi3T^{1-180}SPR^{148-439}$ and $SPR^{1-147}\Phi3T^{181-443}$, encoded by DNA joined at the common *Sac*I restriction site (Figure 9) exhibit the sequence specificity of the parent from which the C-terminal two-thirds of the polypeptide was derived. This result clearly suggests that the variable region in the N-terminal domain of the Φ3T enzyme is not necessary for expression of either the GGCC recognition common to both proteins or the differential specificity toward the other sequences. Furthermore, hybrids in which the splice site is at the *Stu*I restriction site in the middle of the variable region found in the C-terminal region of the two parents show sequence specificity patterns that are slightly different from that of either parent. Thus, the chimeric Mtases $\Phi3T^{1-311}SPR^{277-439}$ and $SPR^{1-276}\Phi3T^{312-443}$ both methylate GGCC sequences as do the parents, but the former does not recognize GCNGC or CC(A/T)GG sites and the latter methylates no other sites. In all cases the hybrid enzymes have relative catalytic activities not terribly different from that the parents. Finally, the tripartite hybrid, $SPR^{1-147}\Phi3T^{181-311}SPR^{277-439}$, was found to be slightly more efficient at methylation and to have a relaxed specificity. The results, taken together, strongly suggest that the variable regions of the C termini are involved in the recognition of the alternate methylation sites. Moreover, the results support the notion of the modular organization of the DNA methyltransferases.

4. Fusion Hybrid of T7 Endonuclease

In certain special instances an additional substrate specificity can be introduced into an enzyme, particularly those with macromolecular substrates, by the fusion of the enzyme with a specific recognition protein. One very clear demonstration of this concept was recently reported[87,88] in which the fusion of the gene encoding the phage T7 endonuclease, T7.3, to the 3′-end of the *lac* repressor gene (*lac I*) results in a T7 endonuclease specifically targeted to sites near the *lac* operator site. A representation of the primary structure of the repressor/nuclease (R/N) hybrid is shown in Figure 10. The hybrid consists of the first 331 (of a total of 360) amino acids of the *lac I* protein, an alanine residue linker, and then the entire 148-residue sequence of T7.3.

The T7.3 endonuclease is specific for the cleavage of single-stranded DNA, a property that has been used to detect native cruciform structures in supercoiled DNA.[89] The enzyme also cleaves duplex DNA nonspecifically and much less efficiently. The cleavage of duplex DNA probably occurs in regions of transiently melted struc-

ture.[90] Although the hybrid enzyme maintains its nuclease activity toward the major cruciform structure in supercoiled DNA, it exhibits, unlike its catalytic parent, specificity for a sequence near the *lac* operator site in linear DNA. Most interesting is the finding that the enzyme cleaves at two specific sites located upstream, 155 and 190 base pairs from the center of symmetry of the operator. The specific cleavage is maintained even if the orientation of the operator is reversed. The inversion of the orientation of the operator displaces the center of the operator an additional 40 base pairs from the two cleavage sites now located at 195 and 230 base pairs from the center. The limits as to how close or far the cleavage can be from the recognition site to observe the selectivity is not known. Given the relative size of the hybrid and the fact that it is probably a tetramer, it is clear that the DNA must loop back toward the hybrid protein such that the active site of the nuclease can encounter the cleavage site irrespective of the orientation of the recognition site to the cleavage site. The binding of the *lac I* domain of the hybrid enzyme to the *lac O* sequence presumably increases the residence time of the nuclease near the sensitive site, which results in a specific cleavage of the linear DNA. Consistent with this is the observation that the specificity of the hybrid nuclease toward linear DNA containing *lac O* is significantly decreased in the presence of isopropyl 1-thio-β-D-galactopyranoside, which is known to decrease the affinity of the *lac* repressor for the operator by increasing the off-rate of *lac I*.[90] The foregoing is a very nice demonstration of the potential of simple fusion hybrids in the engineering of more specific catalysts.

E. Mosaic Isoenzymes of Detoxication

Many enzymes involved in the metabolism of xenobiotic compounds exist in families of isoenzymes possessing slightly different catalytic specificity. Such a diversification of substrate specificity allows an organism much more flexibility in its metabolic response to a host of unexpected chemical insults. Biochemists interested in these catalysts have spent a good deal of time cataloging isoenzymes, their substrate preferences, and, more recently, their primary structures as well as, where possible, their three-dimensional structures. Comparisons of the primary structures of the isoenzymes within a given class of catalyst reveal that the proteins share a good deal of sequence homology with one another. This obviously should not be a surprise for related isofunctional enzymes. Perhaps more importantly, such comparisons show that many of the differences in sequence appear to be clustered so that the primary structure of the proteins may be divided into highly conserved and highly variable regions. It is therefore tempting, if not always correct, to assume that the unique catalytic character of various isoenzymes derives in large part from differences in the structure of the variable regions. Furthermore, a scenario for the evolutionary diversification of the catalytic properties of detoxication enzymes by the recruitment of exons encoding parts of these variable sequences can be easily imagined. There is in fact some suggestive evidence that gene conversion, a nonreciprocal recombination event, has played a role in the evolution of both the cytochromes P-450[91] and the glutathione *S*-transferases.[92] For these reasons the enzymes of detoxication are obvious targets for a modular analysis of structure–function relationships.

1. Cytochromes P-450

Of all the detoxication enzymes the cytochromes P-450 have received the most attention with respect to both mechanistic investigation and the elucidation of structure–function relationships. The enzymes catalyze an extraordinarily diverse number of monooxygenation reactions including hydroxylation, epoxidation, heteroatom oxidation, oxygen and nitrogen dealkylations, deaminations, heteroatom release, and reductions.[93] The substrate specificities of the numerous isoenzymes vary quite remarkably with respect to kinetics and the regiochemistry and stereochemistry of the reactions. A number of attempts have been made to track down segments of primary structure important to the substrate specificity of the various isoenzymes obtained from mammalian sources through the construction and evaluation of chimeric enzymes. Inasmuch as the mammalian enzymes are membrane-bound proteins, the construction and expression of parent and chimeric enzymes have been accomplished primarily in yeast. The availability of a crystal structure of the bacterial cytochrome P-450CIA1[94] (formerly P-450$_{cam}$)[95] is of considerable help in designing and interpreting this sort of experiment.

The earliest reports of chimeric cytochromes P-450 were not particularly encouraging from the standpoint of understanding structure–function relationships. The first construction of chimeric cytochromes P-450 was reported by Sakaki et al.[96] with two rat liver isoenzymes P-450IA1 and P-450IA2 having an overall homology of 68% and quite different substrate specificities. Three chimeric expression vectors were assembled encoding the hybrid rat proteins IA1$^{1-185}$IA2$^{183-512}$, IA1$^{1-372}$IA2$^{367-512}$, and IA1$^{1-185}$IA2$^{183-366}$IA1$^{373-523}$. Although all three proteins could be expressed in yeast, the tripartite hybrid was found not to bind heme. The observation that microsomes containing the two bipatite hybrids exhibited somewhat different catalytic activities toward three different substrates (acetanilide, 7-ethoxycoumarin, and benzo[*a*]pyrene) such that the chimera composed of the C-terminal two-thirds of IA2 tended to have the specificity of the IA2 parent and the one composed of the C-terminal one-third of IA2 mimicked the IA1 parent was interpreted to suggest that all or part of the central one-third of the protein was important to the catalytic specificity of the enzyme. Given the absence of complementary results from the reciprocal hybrids and the fact that the tripartite hybrid did not bind heme and therefore was not active, the above interpretation must be considered a bit liberal.

The catalytic profile of other early chimeric cytochromes P-450 appeared to be consistent with the conclusion that a region of the C-terminal two-thirds of the polypeptide sequence was crucial (or conversely that the N-terminal one-third was not) to the architecture of the substrate binding site of the enzyme. Pompon[97] reported a chimeric enzyme, IA2$^{1-143}$IA1$^{144-518}$, assembled from rabbit P-450IA1 and P-450IA2 (formerly LM6 and LM4) that retained the general catalytic character of the P-450IA1 parent. A similar finding has been reported for the mouse IA1 and IA2

isoenzymes.[98] However, such a limited group of results may simply be concidental and potentially misleading.

To obtain a higher resolution picture of the situation obviously necessitates the construction of additional chimeric enzymes that focus on more limited regions of the primary structure. Identification of hot spots in the sequence may be done in a rational fashion by examining variable regions of sequence and sniffing around experimentally until one locates and defines a responsive segment. Alternatively, a library of chimeras can be generated and screened with no preconceived notion of the possible location of critical sequence modules. Both of these techniques have been used very recently to generate new, informative hybrid cytochromes P-450.

Pompon co-workers[19] have engineered of a new group of interspecies hybrid enzymes in yeast by in vivo recombination within partially homologous sequences of the mouse P-450IA1 and rabbit P-450IA2 genes. The intramolecular assembly of two bipartite chimeric genes was achieved by forcing recircularization of a linearized plasmid containing all or parts of both genes through recombination of homologous segments. In addition, two tripartite chimeric genes were prepared by gap repair of one gene through intermolecular recombination with a piece of the other. The relative activities of the four hybrid enzymes encoded by the recombinant plasmids toward three substrates appear to suggest that a particular region (residues 203–238) of P-450IA1 is crucial to the catalytic efficiency of this isoenzyme toward polycyclic aromatic hydrocarbons since the one chimera that does not contain this sequence has a very low activity toward benzo[*a*]pyrene, a characteristic of rabbit P-450IA2. Whether this module is in itself sufficient to impart specificity toward benzo[*a*]pyrene is not known but could, in principle, be determined by substitution of the region into the rabbit P-450IA2. Part of the sequence segment in question (residues 203–238) appears to correspond to the E helix (residues 150–170) in the crystal structure of the bacterial enzyme[94] and may indeed be involved in substrate recognition.

A similar region has been tentatively identified as being important for the binding of laurate to rabbit P-450IIC2 through characterization chimeric enzymes that are not catalytically active, not an uncommon finding in the construction of hybrid enzymes. For example, rabbit liver cytochromes P-450IIC14 and P-450IIC2 (formerly pHP3 and pHP2-1, respectively) are closely related proteins sharing 81% amino acid sequence identity.[99] The P-450IIC2 isoenzyme is a laurate ω-1 hydroxylase; P-450IIC14 is not. Inasmuch as most of the sequence differences appear to be clustered in three variable regions, residues 91–110, 213–248, and 472–490, it is reasonable to expect that one or more of these sequence modules in P-450IIC2 are largely responsible for this specific catalytic activity. Of three hybrids derived from the parents, namely, $IIC14^{1-43}IIC2^{44-90}$, $IIC14^{1-210}IIC2^{211-490}$, and $IIC14^{1-261}IIC2^{262-490}$, only the first had significant fatty acid hydroxylase activity. Both $IIC14^{1-43}IIC2^{44-490}$ and $IIC14^{1-210}IIC2^{211-490}$ appear competent at binding fatty acid substrates as judged by spectral titrations, but the latter has no hydroxylase activity. This in addition to the fact that $IIC14^{1-261}IIC2^{262-490}$ does not bind substrate

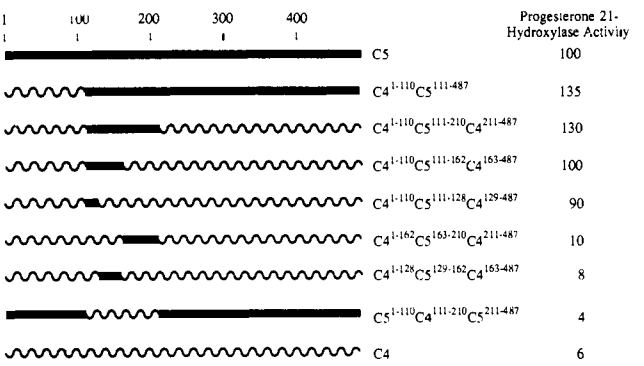


Figure 11. Schematic illustration of the primary structures and progesterone 21-hydroxylase activities of the chimeric cytochromes P-450 constructed from the P-450IIC4 and P-450IIC5 isoenzymes by Kronbach et al.[100] The scale at the top indicates the residue position of the 487 amino acid polypeptides. The relative catalytic activities tabulated on the right were determined at subsaturating concentrations of progesterone so that the relative activities would be roughly comparable to differences in V/K.

was taken as evidence that the region between residues is necessary for binding fatty acid substrates but not sufficient for productive binding or catalysis. This region does encompass the variable region bounded by residues 213 and 248. Although this interpretation must be regarded as very preliminary and subject to further experimental verification, it does point out the importance in cataloging "failures" of hybrid proteins at either binding or catalysis.

The feasibility of experimental identification of a small segment of primary structure that may impart a defined catalytic characteristic was recently demonstrated by Kronbach and co-workers[100] with the highly homologous cytochromes P-450IIC4 and P-450IIC5. The two isoenzymes share 95% amino acid sequence identity but have quite different catalytic efficiencies at least toward progesterone. P-450IIC5 is unique in that it catalyzes the hydroxylation of progesterone with high efficiency at the 21-position in the formation of deoxycorticosterone. Although the actual catalytic function of the IIC4 isoenzyme has not been identified, it does catalyze the hydroxylation of progesterone albeit with a relative V/K about 25-fold lower than that of P-450IIC5. Most of the difference is manifest in K_m, which is 1.6 μM with P-450IIC5 and 25 μM with P-450IIC4. Four restriction sites were used to splice the genes encoding the two isoenzymes at codons 110, 128, 162, and 210, which permitted tracking a relatively small stretch of sequence in P-450IIC5 apparently responsible for the high V/K toward progesterone. The seven chimeras constructed and expressed in COS-1 cells are illustrated in Figure 11 along with their relative progesterone 21-hydroxylase activities. It is evident that only the four hybrid isoenzymes that contain the region between residues 111 and 128 derived from the IIC5 isoenzyme catalyze the 21-hydroxylation of progesterone efficiently. It is quite striking that the mutation of just three amino acids in the 111–128 segment of P-450IIC4 enhances the specificity of the enzyme toward progesterone. It is also interesting that this region of the mammalian enzyme corresponds to a region between the B and C helices of the bacterial protein which contains Tyr96 that hydrogen bonds to the carbonyl of the camphor substrate.[94] It would seem likely that this same region in the mammalian enzyme

is involved in substrate binding.

Others have also found that changing only a few amino acids can have an effect on the substrate selectivity of a cytochrome P-450. The phenobarbitol-inducible cytochromes P-450IIB1 and P-450IIB2 are highly homologous enzymes differing in only 14 amino acids in the C-terminal half of the protein. Both enzymes catalyze the oxidation of testosterone with essentially the same regiochemistry and stereochemistry. The major products of the oxidation are the 16α- and 16β-hydroxy, 17-keto products and the doubly oxidized 16β-hydroxy, 17-keto metabolite. Aoyama et al.[101] have recently isolated a variant cDNA encoding another IIB2 enzyme. The variant IIB2 which contains three mutations, L58F, I114F and E322V, exhibits catalytic properties that are different from those of the IIB1 and IIB2 isoenzymes. For instance, the catalytic turnover of the IIB2 variant was very low and, perhaps more interestingly, seemed not to produce significant amounts of the 16β-hydroxy product. This observation prompted the construction of a chimeric IIB1 isoenzyme in which the sequence between and inclusive of L58 and I114 was replaced with that derived from the variant IIB2, thus producing the double mutant L58F/I114F P-45oIIB1. The chimeric IIB1 with the two phenylalanine residues had about 30% of the catalytic activity of the parent P-450IIB1 toward testosterone as well as a different product distribution. Specifically, the chimera gave very little of the 16β-hydroxy and no detectable 17-keto products. The hybrid protein thus appears unable to catalyze the efficient hydroxylation of the 16β-position in either the starting material or the 17-keto product. A similar observation was made with androstenedione as substrate.[101]

Site-specific mutants in which only one of the positions was changed to phenylalanine resulted either in a protein (I114F P450IIB1) that resembled the parent with respect to product distribution or in an unstable protein product (L58F P-450IIB1). Taken together the data suggest that two amino acids in the N-terminal one-third of the protein are particularly important in the orientation of the substrate at the active site. It is notable that mutations at both positions, which are separated by over 50 residues in sequence space, are necessary to produce a stable enzyme with altered catalytic properties. One of these residues is in a region of sequence that corresponds to part of the substrate binding site of the bacterial enzyme (vide supra).[94]

2. UDPglucuronosyltransferase

The UDPglucuronosyltransferases are a group of microsomal detoxication enzymes that catalyze the addition of a glucuronosyl group to both endogeneous and exogeneous substrates that contain nucleophilic functional groups.[102] As is typical of the enzymes of detoxication, the catalysts exhibit rather broad substrate specificity but do show distinct substrate preferences. Inasmuch as these enzymes are integral membrane-bound proteins the possibility of a solution of the three-dimensional crystal structure of one of the isoenzymes in the near future is quite remote. For this reason the role of chimeric constructs is likely to be one of the few tools available for delineating which portions of the polypeptide influence the substrate specificity of the enzymes. Some preliminary but quite impressive work in this regard has been recently reported by Mackenzie.[103]

Two isoenzymes of UDPglucuronosyltransferase with quite distinct catalytic properties have been expressed in COS cells. One form, UDPGTr-3, is very active toward testosterone, whereas the other, UDPGTr-4, prefers to glucuronidate etiocholanolone.[104,105] The two enzymes exhibit an 85% sequence identity. Furthermore, comparison of their sequences to other isoenzymes reveals that the C-terminal ends of the proteins are of somewhat higher homology than the amino-terminal portions, perhaps suggesting that the N-terminal domain is involved in recognition of the aglycon substrates. Construction and expression of two complementary chimeric isoenzymes in which the N-terminal domain of one isoenzyme was substituted for the other have now provided the first real evidence that at least part of the architecture of the aglycon binding site is derived from the N-terminal domain of the protein.[103] The chimera UDPGTr-3.4 ($3^{1-298}4^{299-530}$) consists of the first 298 amino acid of the testosterone-specific enzyme, UDPGTr-3, and the last 232 residues of UDPGTr-4, which is specific for etiocholanolone. Although the kinetic properties of the chimera have not been investigated in detail, it is clear that this hybrid construct prefers testosterone as the aglycon substrate. Similarly, the reciprocal hybrid UDPGTr-4.3 ($4^{1-298}3^{299-530}$), which is built from the N-terminal half of the etiocholanolone-specific enzyme and the C-terminal domain of the other, efficiently catalyzes the glucuronidation of etiocholanolone and not testosterone. The most impressive result of this investigation is the fact that the two reciprocal hybrids in fact have, to a first approximation, reciprocal substrate specificities. Strictly speaking, this is a necessary condition for demonstrating that a particular domain of an isoenzyme does impart a unique catalytic property to a given protein framework.

3. Glutathione *S*-Transferases

The glutathione *S*-transferases are a group of proteins that catalyze the nucleophilic addition of the tripeptide glutathione (GSH) to a wide variety of lipophilic molecules having electrophilic functional groups.[106,107] This general reaction represents a major detoxication pathway for reactive electrophiles including many generated by the cytochromes P-450 discussed above. The soluble or cytosolic isoenzymes are dimeric proteins composed of subunits (molecular weights of approximately 25 000) derived from at least three gene families. Homodimeric and heterodimeric enzymes occur within gene families, but intergene family heterodimers are not known. Amino acid sequence homologies are high (>70%) within a gene family and significantly lower (ca. 40%) between different gene families.[106,107] Furthermore, sequence differences appear to be clustered into variable regions in a given gene family.[108] In addition, it is well established that the substrate specificities for isoenzymes even within a family are often quite different. There is no evidence for active monomeric species, nor do the catalytic properties of a given subunit appear to be influenced by the identity of its partner in the holoenzyme. The GSH transferases are therefore good candidates for the modular analysis of structure–function relationships in catalysis and per-

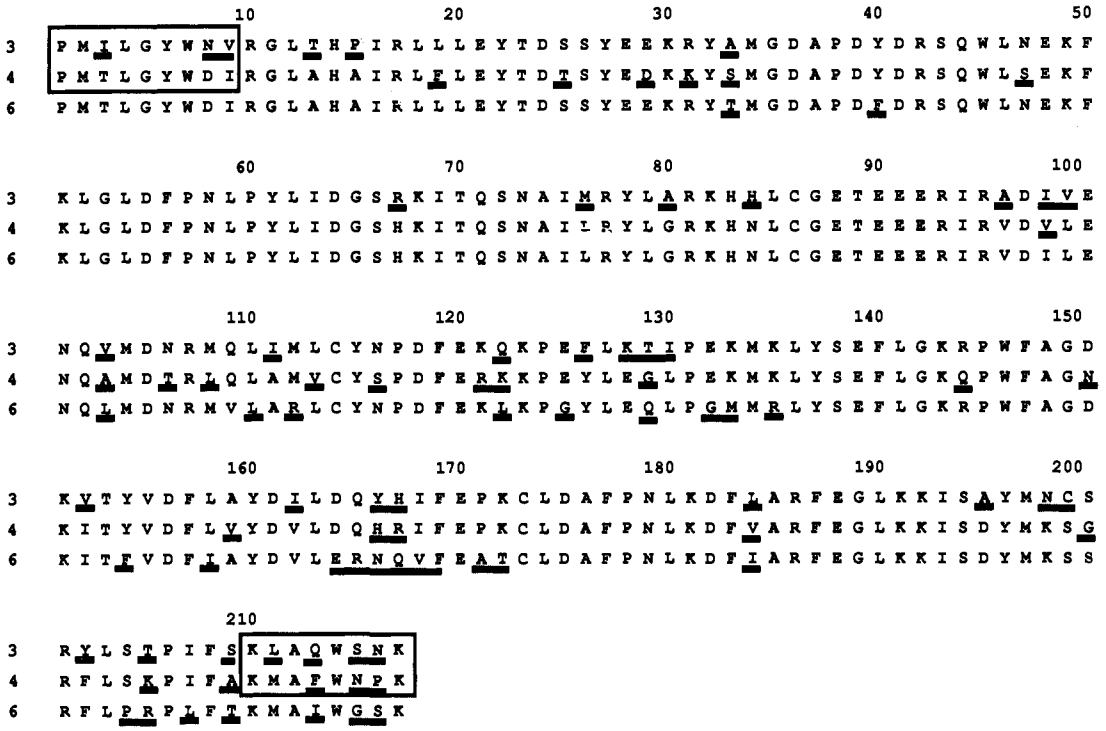


Figure 12. Amino acid sequences of three of the *mu*-gene-class subunits of GSH transferases from rat. Nonconserved residues are underlined. The boxed regions at the termini of polypeptides 3 and 4 show the regions that were changed in the construction of the first-generation chimeric isoenzymes[114] listed in Table III.

TABLE III. Catalytic Properties of Native and Hybrid GSH Transferases toward Phenanthrene 9,10-Oxide[a]

isoenzyme or chimeric enzyme	K_d^{GSH}, μM	k_c, s^{-1}	k_c/K_m, $M^{-1}\ s^{-1}$	mole fraction of (*S*,*S*) product	$\delta\Delta G^{\ddagger}$,[b] kcal/mol
3-3	21	0.42	1.2×10^4	0.43	−0.2
4-4	24	0.79	6.6×10^5	1.00	>2.7
$(4^{93^{208}})_2$	42	0.46	1.9×10^3	0.85	1.0
$(3^{209}4^8)_2$	180	0.30	1.4×10^4	0.29	−0.5
$(4^{93^{200}}4^8)_2$	20	1.8	3.2×10^3	0.74	0.6

[a] Data taken from ref 114. [b] Calculated from RT ln ([*S*,*S*-product]/[*R*,*R*-product]).

haps eventually with respect to subunit–subunit recognition. Unfortunately, no three-dimensional structure of a GSH transferase has been reported, although diffraction-quality single crystals have been prepared from an isoenzyme from each gene class.[109-111] The design of hybrid enzymes must be guided by the primary structure alone.

The subunit polypeptides of the *mu* gene class of GSH transferases, designated subunits 3, 4, and 6, are 217 residues in length and share 77% sequence identity with one another.[108] Most of the sequence variations are located in four clusters bounded by residues 1–33, 98–135, 150–172, and 198–217 as illustrated in Figure 12 and more schematically in Figure 13. In spite of the rather high sequence homology, the isoenzymes of the *mu* gene class have quite distinct catalytic properties. For instance, isoenzyme 4-4 is very efficient and stereospecific in the addition of GSH to arene oxide substrates, whereas isoenzyme 3-3 is not.[112] Conversely, isoenzyme 3-3 is very good at catalyzing nucleophilic aromatic substitution reactions.[113]

Some small initial steps have been taken in the fabrication of chimeric GSH transferases to assess the influence of variable regions at the N and C termini on the catalytic specificity of these two isoenzymes.[114,115] An expression plasmid for isoenzyme 3-3 of rat liver GSH transferases was constructed and manipulated to

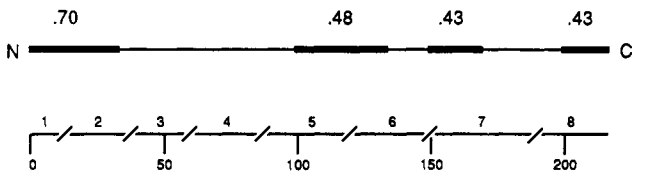
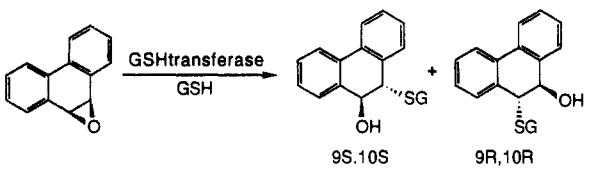


Figure 13. Relationship of the variable regions of the class *mu* glutathione transferases from rat to the exon–intron interfaces found in the genomic DNA. The four variable regions in the protein sequence are indicated by the bold lines. The fractions of positions in each region that are conserved in subunits 3, 4, and 6 are indicated at the top. The three intervening sequences shown as narrow lines exhibit ≥90% sequence identity among the three isoenzymes. The bottom line indicates the relationship of the exon–intron interfaces to the residue number of the protein. Exons are numbered 1–8, and the bottom scale is the amino acid residue number.

encode chimeric type 3 subunits in which parts of both the N- and C-terminal variable domains from subunit 4 were substituted into the analogous regions of subunit 3. The hybrid polypeptides, which, because of the identical length of the parents, are designated $4^{93^{208}}$, $3^{209}4^8$, and $4^{93^{200}}4^8$, contain triple and quadruple mutations in the N and C termini, respectively, as illustrated in the boxed regions of Figure 12. Active dimeric hybrid enzymes were obtained in each instance. The catalytic properties of the chimeric enzymes are all distinct from each other and the parents when exam-

SCHEME I



ined with three substrates including 1-chloro-2,4-dinitrobenzene, 4-phenyl-3-buten-2-one, and phenanthrene 9,10-oxide. The kinetic constants and stereoselectivity of the parent and hybrid isoenzymes with the latter substrate are given in Table III. The most striking observations from these data are the changes observed in the stereoselectivity of the hybrids toward phenanthrene 9,10-oxide (Scheme I) and in the dissociation constant for GSH. The three mutations introduced in the N-terminal module, which, by the way, corresponds to exon 1 (see Figure 13) of the *mu*-class genomic DNAs,[92] has a significant effect on the stereoselectivity of enzyme changing the $\delta\Delta G^{\ddagger}$ for the two diastereomeric transition states by about 1.2 kcal/mol, even though the K_m for the substrate is increased. The quadruple mutation in the C-terminal variable region has a much more modest effect on the stereoselectivity that is in the opposite direction. However, the dissociation constant, K_d, for GSH is about 10-fold higher than that for either native enzyme. Interestingly, this defect in GSH binding introduced in $(3^{209}4^{8})_2$ is apparently corrected in the tripartite hybrid $(4^{9}3^{200}4^{8})_2$. This complementation of a change in properties introduced through the C terminus by comodification of the N terminus clearly implicates both domains as important to the architecture of the active site. Furthermore, it is an experimental indication that the N and C termini of the protein, which are maximally separated in sequence space, are quite close in three dimensions. This is not an uncommon occurrence inasmuch as about one-third of all proteins whose three-dimensional structures are known have N and C termini that interact.[116] Finally, a recent report of photoaffinity labeling of the C-terminal domain of isoenzymes 1-1 and 2-2 fully supports the contention that this variable region is part of the active-site ensemble.[117]

Preliminary results from the construction of an interspecies class *alpha* chimera assembled from the N-terminal two-thirds of the human isoenzyme ϵ and rat subunit 1 suggest that such hybrid proteins are also feasible.[118] Although this chimera exhibited properties more closely related to the human isoenzyme, it would be optimistic, in light of the results mentioned above, to conclude that the catalytic properties of the class *alpha* enzymes were determined by the C-terminal region of the primary structure. The situation is obviously more complex than this and will be much clearer after the crystal structure is determined.

F. Summary

It should be obvious that the modular approach to evaluating structure–function relationships in enzymic catalysis is still in its infancy. The great advantage of the technique is the potential for rapidly scanning sequence space in search of functional hot spots. Moreover, it expands the repertoire of the enzymologist for the exploration of structure–function relationships in instances where the availability of a three-dimensional structure is unlikely, e.g., membrane-bound enzymes. Of course there are drawbacks to this approach, not the least of which is the much lower certainty with which the actual structural changes in the catalyst can be predicted, even in the presence of native crystal structure. As with most experimental designs, the value of chimeric enzymes must be viewed in the context of the questions being asked. The inherent uncertainties in the interpretation of results should not dissuade us from the use of this approach where appropriate. The technique can and will certainly be used unwisely. However, when combined with a thorough evaluation of catalytic properties of the chimeric product and, where possible, three-dimensional structural information, the judicious construction of hybrid enzymes should stimulate new insight into the nature of enzymic catalysis and perhaps the rules by which catalytic function might have evolved. It is obvious from several of the studies described above that the construction and analysis of chimeric enzymes can enhance our understanding of catalysis in a way not possible by other approaches. The answer to the question posed in the title is clearly, then, yes.

G. Acknowledgment

I thank the NIH (GM 30910 and ES 00133) and the NSF (DMB8716307) for financial support of the work performed at the University of Maryland and Prof. Herman Ammon for his assistance in creating the ribbon diagram of Figure 5.

H. References

(1) Knowles, J. R. *Science* **1987**, *236*, 1252.
(2) Gerlt, J. A. *Chem. Rev.* **1987**, *87*, 1079.
(3) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182.
(4) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature (London)* **1990**, *343*, 33.
(5) Myers, R. M.; Lerman, L. S.; Maniatis, T. *Science* **1985**, *229*, 242.
(6) Wells, J. A.; Vasser, M.; Powers, D. B. *Gene* **1985**, *34*, 315.
(7) Hermes, J. D.; Knowles, J. R. Proceedings of the Robert A. Welch Foundation, XXXI. Design of Enzymes and Enzyme Models, 1987, p 207.
(8) Lehtovaara, P. M.; Koivula, A. K.; Bamford, J.; Knowles, J. K. C. *Protein Eng.* **1988**, *2*, 63.
(9) Gilbert, W. *Nature (London)* **1978**, *271*, 501.
(10) Gilbert, W. *Science* **1985**, *228*, 823.
(11) Blake, C. C. F. *Nature (London)* **1978**, *271*, 267.
(12) Traut, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2944.
(13) Darnell, J. E.; Doolittle, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1271.
(14) Richards, J. H. *Nature (London)* **1986**, *323*, 187.
(15) Fetrow, J. S.; Cardillo, T. S.; Sherman, F. *Proteins: Struct., Funct. Genet.* **1989**, *6*, 372.
(16) Clackson, G.; Winter, G. *Nucleic Acids Res.* **1989**, *17*, 10163.
(17) Weber, H.; Weissmann, C. *Nucleic Acids Res.* **1983**, *11*, 5661.
(18) Abastado, J. P.; Darche, S.; Godeau, F.; Cami, B.; Kourilsky, P. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6496.
(19) Pompon, D.; Nicolas, A. *Gene* **1989**, *83*, 15.
(20) Schneider, W. P.; Nichols, B. P.; Yanofsky, C. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2169.
(21) Orr-Weaver, T. L.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4417.
(22) Warton, R. P.; Brown, E. L.; Ptashne, M. *Cell* **1984**, *38*, 361.
(23) Schleif, R. *Science* **1988**, *241*, 1182.
(24) Sellers, J. W.; Struhl, K. *Nature (London)* **1989**, *341*, 74.
(25) Kaiser, E. T.; Kezdy, F. J. *Science* **1984**, *223*, 249.
(26) Kaushansky, K.; Shoemaker, S. G.; Alfaro, S.; Brown, C. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1213.
(27) Woon, C. W.; Soparkar, S.; Heasley, L.; Johnson, G. L. *J. Biol. Chem.* **1989**, *264*, 5687.
(28) Eilers, M.; Picard, D.; Yamamoto, K.; Bishop, J. M. *Nature (London)* **1989**, *340*, 66.
(29) Lowe, D. G.; Ricketts, M.; Levison, A. D.; Goeddel, D. V. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1015.

(30) Lefkowitz, R. J.; Kobilka, B. K.; Caron, M. G. *Biochem. Pharmacol.* **1989**, *38*, 2941.
(31) Pastan, I.; FitzGerald, D. *J. Biol. Chem.* **1989**, *264*, 15157.
(32) Manoil, C.; Beckwith, J. *Science* **1986**, *233*, 1403.
(33) Talmadge, K.; Kaufman, J.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3988.
(34) Moreno, F.; Fowler, A. V.; Hall, M.; Silhavy, T. J.; Zabin, I.; Schartz, M. *Nature (London)* **1980**, *286*, 356.
(35) Bashford, P. J.; Silhavy, T. J.; Beckwith, J. R. *J. Bacteriol.* **1979**, *139*, 19.
(36) Summers, R. G.; Knowles, J. R. *J. Biol. Chem.* **1989**, *264*, 20074.
(37) Hirel, Ph.-H.; Schmitter, J.-M.; Dessen, P.; Fayat, G.; Blanquet, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8247.
(38) Wetlaufer, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 697.
(39) Mas, M. T.; Chen, C. Y.; Hitzeman, R. A.; Riggs, A. D. *Science* **1986**, *233*, 788.
(40) Hitzeman, R. A.; Hagie, F. E.; Hayflick, J. S.; Chen, C. Y.; Seeburg, P. H.; Derynck, R. *Nucleic Acids Res.* **1982**, *10*, 7791.
(41) Michelson, A. M.; Markham, A. F.; Orkin, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 472.
(42) Watson, P. L.; Walker, N. P. C.; Shaw, P. J.; Bryant, T. N.; Wendell, P. L.; Fothergill, L. A.; Perkins, R. E.; Conroy, S. C.; Dobson, M. J.; Tuite, M. F.; Kingsman, A. J.; Kingsman, S. M. *EMBO J.* **1982**, *1*, 1635.
(43) Banks, R. D.; Blake, C. C. F.; Evans, P. R.; Haser, R.; Rice, D. W.; Hardy, G. W.; Merrett, M.; Phillips, A. W. *Nature (London)* **1979**, *279*, 773.
(44) Roustan, C.; Fattoum, A.; Jeanneau, R.; Pradel, L.-A. *Biochemistry* **1980**, *19*, 5168.
(45) Coleman, P. F.; Suttle, D. P.; Stark, G. R. *J. Biol. Chem.* **1977**, *252*, 6379.
(46) Mally, M. I.; Grayson, D. R.; Evans, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 6647.
(47) Maley, J. A.; Davidson J. N. *MGG, Mol. Gen. Genet.* **1988**, *213*, 278.
(48) Gerhart, J. C.; Schachman, H. K. *Biochemistry* **1965**, *4*, 1054.
(49) Honzatko, R. B.; Crawford, J. L.; Monaco, H. L.; Ladner, J. E.; Edwards, B. F. P.; Evan, D. R.; Warren, S. G.; Wiley, D. C.; Ladner, R. C.; Lipscomb, W. N. *J. Mol. Biol.* **1982**, *160*, 219.
(50) Major, J. G.; Wales, M. E.; Houghton, J. E.; Maley, J. A.; Davidson, J. N.; Wild, J. R. *J. Mol. Evol.* **1989**, *28*, 442.
(51) Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R. *J. Biol. Chem.* **1988**, *263*, 17857.
(52) Miles, E. W.; Yutani, K.; Ogasahara, K. *Biochemistry* **1982**, *21*, 2586.
(53) Beasty, A. M.; Matthews, C. R. *Biochemistry* **1985**, *24*, 3547.
(54) Yutani, K.; Sato, T.; Ogasahara, K.; Miles, E. W. *Arch. Biochem. Biophys.* **1984**, *229*, 448.
(55) Holmgren, A. *Annu. Rev. Biochem.* **1985**, *54*, 237.
(56) Modrich, P.; Richardson, C. C. *J. Biol. Chem.* **1975**, *250*, 5515.
(57) Gleason, F. K.; Whittaker, M. M.; Holmgrem, A.; Jornvall, H. *J. Biol. Chem.* **1985**, *260*, 9567.
(58) Holmgren, A.; Soderberg, B.-O.; Eklund, H.; Branden, C.-I. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2305.
(59) Lim, C.-J.; Gleason, F. K.; Jacobson, B. A.; Fuchs, J. A. *Biochemistry* **1988**, *27*, 1401.
(60) Holmgren, A. *J. Biol. Chem.* **1972**, *247*, 1992.
(61) Sowadski, J. M.; Handschumacher, M. D.; Krishna Murthy, H. M.; Foster, B. A.; Wyckoff H. W. *J. Mol. Biol.* **1985**, *186*, 417.
(62) Dubose, R. F.; Hartl, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9966.
(63) Regan, L.; DeGrado, W. F. *Science* **1988**, *241*, 976.
(64) Matthews, B. W.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 4178.
(65) Leszczynski, J. F.; Rose, G. D. *Science* **1986**, *234*, 849.
(66) Freimuth, P. I.; Taylor, J. W.; Kaiser, E. T. *J. Biol. Chem.* **1990**, *265*, 896.
(67) Salemme, F. R.; Freer, S. T.; Xuong, Ng. H.; Alden, R. A.; Kraut, J. *J. Biol. Chem.* **1973**, *248*, 3910.
(68) Timkovich, R.; Dickerson, R. E. *J. Biol. Chem.* **1976**, *251*, 4033.
(69) Takano, T.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6371.
(70) Matsuura, Y.; Takano, T.; Dickerson, R. E. *J. Mol. Biol.* **1982**, *156*, 389.
(71) Louie, G. V.; Hutcheon, W. L. B.; Brayer, G. D. *J. Mol. Biol.* **1988**, *199*, 295.
(72) Fett, J. W.; Strydom, D. J.; Lobb, R. R.; Alderman, E. M.; Bethune, J. L.; Riordan, J. F.; Vallee, B. L. *Biochemistry* **1985**, *24*, 5480.
(73) Harper, J. W.; Vallee, B. L. *Biochemistry* **1989**, *28*, 1875.
(74) Strydom, D. J.; Fett, J. W.; Lobb, R. R.; Alderman, E. M.; Bethune, J. L.; Riordan, J. F.; Vallee, B. L. *Biochemistry* **1985**, *24*, 5486.
(75) Shapiro, R.; Fox, E. A.; Riordan, J. F. *Biochemistry* **1989**, *28*, 1726.
(76) Shapiro, R.; Riordan, J. F.; Vallee, B. L. *Biochemistry* **1986**, *25*, 3527.
(77) Shapiro, R.; Strydom, D. J.; Weremowicz, S.; Vallee, B. L. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 530.
(78) Wodak, S. Y.; Liu, M. Y.; Wyckoff, H. W. *J. Mol. Biol.* **1977**, *116*, 855.
(79) Palmer, K. A.; Scheraga, H. A.; Riordan, J. F.; Vallee, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1965.
(80) Bond, M. D.; Vallee, B. L. *Biochemistry* **1990**, *29*, 3341.
(81) Benner, S. A. *Trends Biochem. Sci.* **1989**, *14*, 396.
(82) Joris, B.; Ghuysen, J. M.; Dive, G.; Renard, A.; Dideberg, O.; Charlier, P.; Frere, J. M.; Kelly, J. A.; Boyington, C.; Moews, P. C.; Knox, J. R. *Biochem. J.* **1988**, *250*, 313.
(83) Waxman, D. J.; Strominger, J. L. *Annu. Rev. Biochem.* **1983**, *52*, 825.
(84) Chang, Y.-H.; Labgold, M. R.; Richards, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2823.
(85) Balganesh, T. S.; Reiners, L.; Lauster, R.; Noyer-Weidner, M.; Wilke, K.; Trautner, T. A. *EMBO J.* **1987**, *6*, 3543.
(86) Behrens, B.; Noyer-Weidner, M.; Pawlek, B.; Lauster, R.; Balganesh, T. S.; Trautner, T. A. *EMBO J.* **1987**, *6*, 1137.
(87) Panayotatos, N.; Fontaine, A.; Backman, S. *J. Biol. Chem.* **1989**, *264*, 15066.
(88) Panayotatos, N.; Backman, S. *J. Biol. Chem.* **1989**, *264*, 15070.
(89) Panayotatos, N.; Fontaine, A. *J. Biol. Chem.* **1987**, *262*, 11364.
(90) Barkley, M. D.; Riggs, A. D.; Jobe, A.; Bourgeois, S. *Biochemistry* **1975**, *14*, 1700.
(91) Atchison, M.; Adesnik, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2300.
(92) Lai, H.-C. J.; Qian, B.; Grove, G.; Tu, C.-P. D. *J. Biol. Chem.* **1988**, *263*, 11389.
(93) Guengerich, F. P.; Macdonald, T. L. *Acc. Chem. Res.* **1984**, *17*, 9.
(94) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *J. Mol. Biol.* **1987**, *195*, 687.
(95) The designation of isoenzymes used here follows the recently proposed uniform system of nomenclature. Nebert, D. W.; Nelson, D. R.; Adesnik, M.; Coon, M. J.; Estabrook, R. W.; Gonzalez, F. J.; Guengerich, F. P.; Gunsalus, I. C.; Johnson, E. J.; Kemper, B.; Levin, W.; Phillips, I. R.; Sato, R.; Waterman, M. R. *DNA* **1989**, *8*, 1.
(96) Sakaki, T.; Shibata, M.; Yabusaki, Y.; Ohkawa, H. *DNA* **1987**, *6*, 31.
(97) Pompon, D. *Eur. J. Biochem.* **1988**, *177*, 285.
(98) Cullin, C.; Pompon, D. *Gene* **1988**, *65*, 203.
(99) Imai, Y. *J. Biochem.* **1988**, *103*, 143.
(100) Kronbach, T.; Larabee, T. M.; Johnson, E. F. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8262.
(101) Aoyama, T.; Korzekwa, K.; Nagata, T.; Adesnik, M.; Reiss, A.; Lapenson, D. P.; Gillette, J.; Gelboin, H. V.; Waxman, D. J.; Gonzalez, F. J. *J. Biol. Chem.* **1989**, *264*, 21327.
(102) Armstrong, R. N. *CRC Crit. Rev. Biochem.* **1987**, *22*, 39.
(103) Mackenzie, P. I. *J. Biol. Chem.* **1990**, *265*, 3432.
(104) Mackenzie, P. I. *J. Biol. Chem.* **1986**, *261*, 14112.
(105) Mackenzie, P. I. *J. Biol. Chem.* **1987**, *262*, 9744.
(106) Pickett, C. B.; Lu, A. Y. H. *Annu. Rev. Biochem.* **1989**, *57*, 357.
(107) Mannervik, B.; Danielson, U. H. *CRC Crit. Rev. Biochem.* **1988**, *23*, 283.
(108) Abramovitz, M.; Listowski, I. *J. Biol. Chem.* **1987**, *262*, 7770.
(109) Sesay, M. A.; Ammon, H. L.; Armstrong, R. N. *J. Mol. Biol.* **1987**, *197*, 377.
(110) Schaffer, J.; Gallay, O.; Ladenstein, R. *J. Biol. Chem.* **1988**, *263*, 17404.
(111) Cowan, S. W.; Bergfors, T.; Jones, T. A.; Tibbelin, G.; Olin, B.; Board, P. G.; Mannervik, B. *J. Mol. Biol.* **1989**, *208*, 369.
(112) Boehlert, C. C.; Armstrong, R. N. *Biochem. Biophys. Res. Commun.* **1984**, *121*, 980.
(113) Graminski, G. F.; Kubo, Y.; Armstrong, R. N. *Biochemistry* **1989**, *28*, 3562.
(114) Zhang, P.; Armstrong, R. N. *Biopolymers* **1990**, *29*, 159.
(115) Armstrong, R. N.; Ammon, H. L.; Graminski, G. F.; Zhang, P.; Sesay, M. A.; Kubo, Y.; Dickert, L. A. In *Glutathione S-Transferases and Drug Resistance*; Hayes, J. D., Pickett, C. B., Mantle, T. J., Eds.; Taylor & Francis: London, 1990; pp 65–74.
(116) Richardson, J. *Adv. Protein Chem.* **1981**, *34*, 167.
(117) Hoesch, R. M.; Boyer, T. D. *J. Biol. Chem.* **1989**, *263*, 17712.
(118) Mannervik, B.; Board, P. G.; Berhane, K.; Bjornestedt, R.; Castro, V. M.; Danielson, U. H.; Hao, X. Y.; Kolm, R.; Olin, B.; Principato, G. B.; Ridderstrom, M.; Stenberg, G.; Widersten, M. In *Glutathione S-Transferases and Drug Resistance*; Hayes, J. D., Pickett, C. B., Mantle, T. J., Eds.; Taylor & Francis: London, 1990; pp 35–46.