Combining Chemical and Biological Techniques to Produce Modified Proteins

Roger S. Goody,* Kirill Alexandrov, and Martin Engelhard[a]

KEYWORDS:

biosynthesis · ligation · peptides · protein modifications · synthesis design

Introduction

The ability to produce proteins in the laboratory and to change their structures, and therefore their properties, in a controlled fashion is of crucial importance in basic biological research, in biotechnology, and increasingly in medical applications. Recent developments have led to a substantial expansion of the spectrum of methods available for the production of proteins and have extended the semiclassical approaches of overexpression and peptide synthesis so that the limitations of these methods no longer dictate the availability of desired protein (or protein analogue) structures. Here we review standard and newer techniques, with an emphasis on a comparison of the advantages and disadvantages of the different approaches.

Biosynthetic Methods

Protein expression

Under this heading we understand the production of proteins by expression in cells that, in general, are not the cells in which they occur naturally (although their homologues may occur). This is referred to as heterologous expression. Homologous overexpression is, of course, also a useful technique. The range of cell types available for this approach has increased over the years, and it is very likely that further systems will become available. The most convenient organism for protein expression is the bacterium *Escherichia coli* and a very large number of proteins from a wide variety of sources have been successfully produced in these cells.[1] These proteins have been, and continue to be, used in basic research (structural and mechanistic studies), in biotechnological applications (for example, synthesis with the help of enzymes as catalysts), and for therapeutic purposes (for example, human insulin). Ease of handling, rapid growth, and profound genetic and molecular biological knowledge of *E. coli* are the major advantages of this species. Its main disadvantage is its lack of the post-translational modification machinery common and essential to eukaryotic organisms, so that proteins which require this modification cannot generally be expressed in *E. coli* or other prokaryotes. Eukaryotic systems which have been used effectively for expression of proteins that cannot be expressed in prokaryotes include yeast,[2] *Dictyostelium discoidum*,[3, 4] and insect cells infected with baculovirus.[5] The latter system is often chosen initially after failure to achieve expression of the functional polypeptide in *E. coli*, and a large number of eukaryotic proteins have been expressed in quite high yields with its help. However, this system is not the solution to all expression problems and has the disadvantages of slow growth, a long development cycle under inconvenient conditions, and a relatively high cost. There is a clear need for development of alternative systems capable of expressing eukaryotic proteins in their correctly modified form.

Once a reasonable expression system has been found, production of mutant proteins by site-directed mutagenesis is a well-established technique that allows, in principle, substitution of each amino acid by any other naturally occurring amino acid.[6] This technique is often used to answer questions on the role of specific amino acids in protein properties or to change properties in a specific manner. Such experiments normally involve relatively detailed knowledge of the structure and mechanism of the protein. A potentially powerful approach to alteration of protein properties in a desired manner even without accurate knowledge of the protein structure and mechanism, involves random or quasi-random mutagenesis techniques,[7] or DNA-shuffling techniques, coupled to efficient methods to select mutants with the desired specificity.[8]

The methods outlined above constitute a powerful arsenal of techniques for the production of proteins with desired structures and properties. They are, however, ultimately limited by the properties of the naturally occurring amino acids and the peptide bond. The methods outlined below offer the possibility of overcoming this limitation.

Introduction of unnatural amino acids by in vitro suppression

In this approach, an unnatural amino acid is incorporated into a protein in a cell-free transcription/translation system.[9, 10] In order to achieve this incorporation, the codon for the amino acid to be replaced is substituted by one of the three naturally occurring stop codons, which must be different from the codon used for termination of the translation. A corresponding

[a] *Prof. R. S. Goody, Dr. K. Alexandrov, Dr. M. Engelhard*
Max-Planck Institute for Molecular Physiology
Department of Physical Biochemistry
Otto-Hahn-Strasse 11, 44227 Dortmund (Germany)
Fax: (+49) 231-133-2399
E-mail: goody@mpi-dortmund.mpg.de

suppressor tRNA is synthesized which is charged with the desired unnatural or labeled amino acid. This synthetic aminoacyl tRNA is then used to suppress termination by the chosen stop codon in a cell-free system. A large number of unnatural amino acids have been incorporated in this manner. Examples are fluorinated tyrosine,[11] spin-labeled[9], and ^{13}C-labeled[12] amino acids. Such modified proteins can, in principle, be used for a number of interesting studies on protein structure – function relationships. Their use is somewhat limited at present by the low yields obtained from cell-free in vitro transcription/translation systems, a problem which is further aggravated by suboptimal suppressor efficiency.

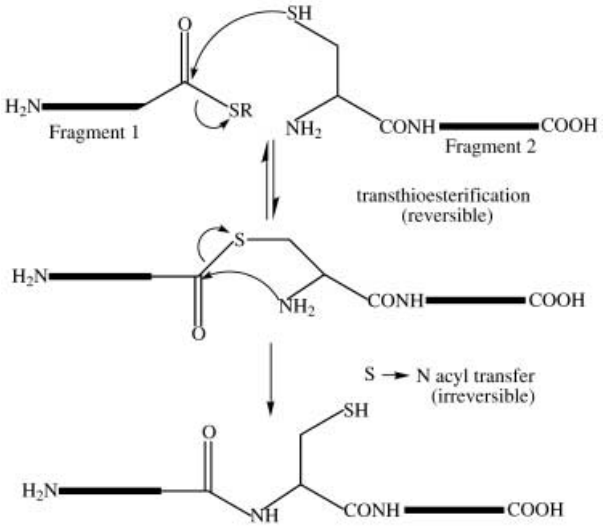
Chemical Methods

Peptide synthesis

The seminal development in this area was the introduction of solid-phase synthetic methods.[13] Together with the development of appropriate protecting group chemistry, this technique has now progressed to the point at which peptides with approximately 60—70 amino acid residues can be synthesized relatively routinely, and in some cases longer fragments have been prepared. The major advantage of this approach is that there is, in principle, complete freedom to incorporate amino acid analogues, which can differ from the naturally occurring amino acids in the nature of their side chain, in their sterochemistry at the α-carbon atom (that is, D- instead of L-amino acids), or even in their backbone chemistry (in other words, this is not limited to the peptide bond). The yields obtainable are normally considerably higher than those from in vitro suppression. There are significant possibilities to introduce of post-translational modifications, which include phosphorylation, glycosylation, and lipidation, although this can sometimes present formidable synthetic problems. The two disadvantagesof solid-phase synthesis compared to cellular expression and cell-free synthesis concern the length of fragments which can be produced in single syntheses (most proteins of interest are significantly longer than 60 – 70 amino acids) and the fact that peptides are initially produced in a denatured form. In many cases, the latter problem has not proved to be serious, and progress is continually being made in the development of methods to renature proteins.

Generation of larger peptides and proteins by chemical ligation of fragments

The size of peptides and proteins which can be synthesized chemically has been increased dramatically by the introduction of methods for ligation of smaller fragments. While some success has been achieved by using an enzymatic method,[14] chemical ligation methods appear to be more versatile. A number of approaches have been developed[15–17] and the method known as native chemical ligation has been particularly successful.[15, 18, 19] If the strategy for production of the full-length protein involves two fragments, the first (that is, N-terminal) peptide is synthesized on a thioester resin, so that after removal from the support, a C-terminal thioester is generated. The second (C-terminal) fragment is synthesized in the standard manner with an N-terminal cysteine residue. The C terminus of the first fragment is activated by transesterification with thiophenol or another thiol reagent, after which it is allowed to react with the C-terminal cysteine of the second fragment in a second transesterification step. A subsequent rearrangement leads to generation of a native peptide bond linking the two fragments (see Scheme 1 for the basic chemical mechanism). This procedure has now been used in several impressive examples to produce proteins which could not have been made in such relatively large amounts (tens of milligrams) by any of the other methods described.[19] As well as its use for the introduction of unnatural amino acids, as described below, this method has been used, for example, to produce covalently linked dimers, which could not have been achieved by recombinant methods.[20, 21]



Scheme 1. *Chemical mechanism of native chemical ligation.*

Recent examples of the use of the native chemical ligation method to connect two or more synthetic peptides include studies on an effector molecule which interacts with the protein product of the *ras* oncogene. The Ras protein is a guanine triphosphatase (GTPase) which in the GTP-bound state interacts with effector molecules as part of a signal transduction cascade. One of the effectors of Ras action is the protein kinase cRaf-1 and the fragment which interacts with Ras is referred to as the Ras binding domain (RBD). The known structure of the complex formed by this interaction was used to define a position in the RBD for the introduction of a fluorescent residue[22] in order to generate an assay principle for biophysical and cell biology studies, and also as a potential diagnostic technique. This decision was based on the principle that the modification should not interfere with the interaction, but should be near enough to the binding interface to be affected in the complex. The RBD fragment was synthesized as two peptides, one of which

harbored a tryptophan derivative, *N*-1-methyl-7-azatryptophan, at position 91 instead of the naturally occurring leucine. After ligation of the two fragments, the fluorescent derivative of the RBD was obtained and could be easily folded by stirring in buffer containing salt and magnesium chloride. It was shown that changes in the fluorescence intensity of the synthetic protein could be used to monitor its interaction with Ras and that the kinetics of this interaction were essentially the same as those of the wild-type protein expressed in *E. coli*. It was also shown that the RBD produced by in vitro transcription/translation showed identical binding behavior[23] so that in this particular case, the conclusion could be reached that the final state of the protein is identical independent of whether it is produced by cellular expression, in vitro transcription/translation in a cell-free system, or chemical synthesis. While this result is both the desired and expected one based on our understanding of protein folding, it is important to demonstrate that these different methods of protein production, which have relative advantages and disadvantages, can nevertheless lead to identical products in cases where the comparison can be made. In most recent work, it has been shown that the partner protein of the cRaf-1 protein, Ras, can also be prepared by complete chemical synthesis and ligation of three fragments. Again, it was shown that the properties of this completely synthetic molecule are indistinguishable from those of the cellularly expressed protein.[24]

Further work on the Ras – Raf interacting system has led to the generation of an RBD with a more useful fluorescent amino acid incorporated at position 91.[25] The side chain of this amino acid contains the nitrobenz-2-oxa-1,3-diazole group and, as in the case of the tryptophan analogue, the fluorescence of this moiety can be used to monitor the interaction of the RBD with activated Ras. The fluorescent RBD also contained a C-terminal histidine tag and this was used to immobilize the protein on a solid surface. The immobilized and labeled RBD could be used as a sensor to detect activated Ras molecules in the submicromolar concentration range.

In the examples given above, it was possible to take advantage of naturally occurring cysteines in the sequences of the proteins prepared. Clearly, this will not always be possible. In such cases, the only possible way to use the standard ligation chemistry is to replace certain nonessential amino acids by cystein. There is in fact a large amount of literature available on introduction of cysteine residues at strategic positions in labeling procedures and it is known that such introduction is generally tolerated at well-chosen positions. However, it is probable that this will not always be an optimal strategy and in this sense, an important new development is the recent extension of the native ligation method to coupling of fragments without the need for an N-terminal cysteine residue in the C-terminal polypeptide.[26] In this technique, a thiol-containing auxiliary group (1-phenyl-2-mercaptoethyl) is added to the α-amino group at the N terminus of the C-terminal polypeptide. After ligation by a mechanism essentially identical to that shown in Scheme 1, this group can be removed by treatment with anhydrous hydrogen fluoride. The technique has been used to synthesize cytochrome b562, which contains no cysteine residues.[27]

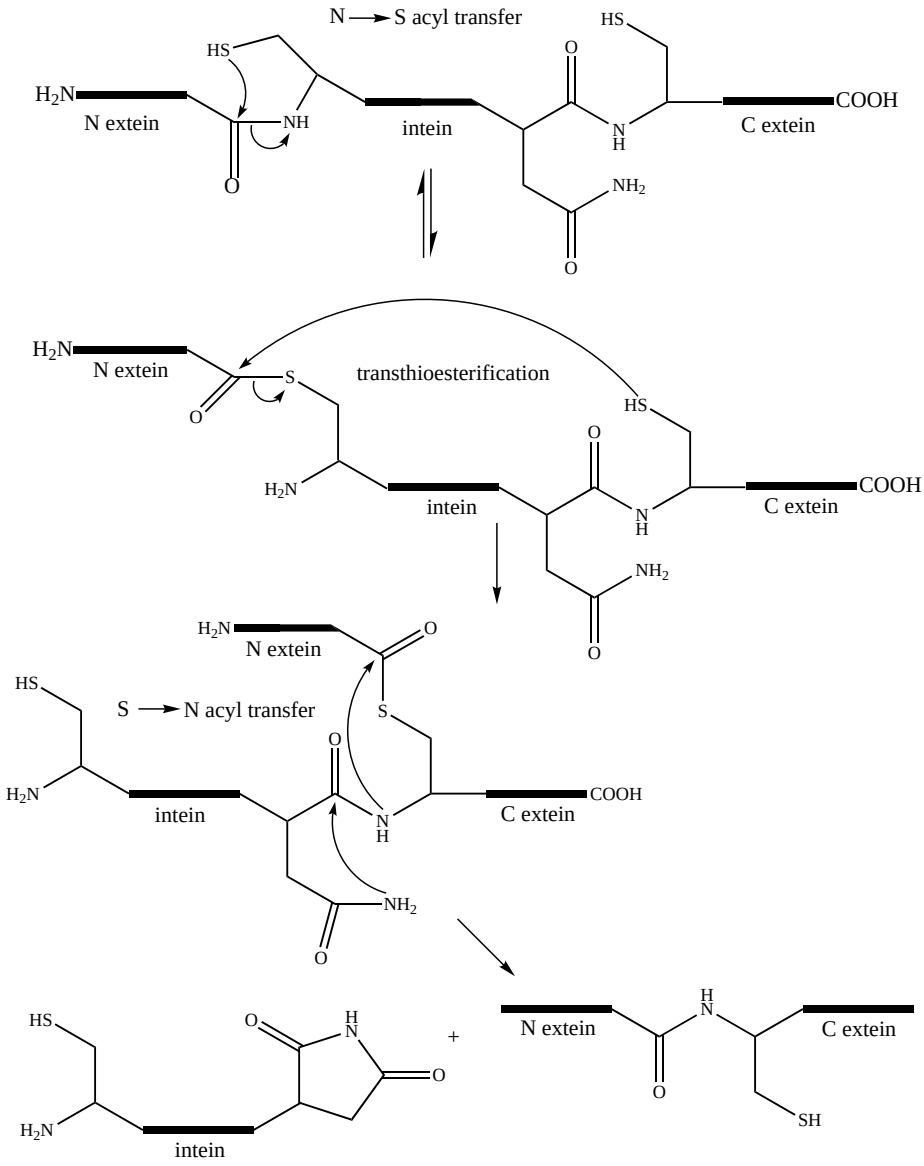
The combination of peptide synthesis and ligation techniques appears to be a very useful approach to the total synthesis of proteins of a size up to about 200 amino acids. Ligation of more than three fragments, while technically feasible, is not likely to be generally attractive because of the relatively low yields to be expected from such a technically complex process. A combined synthetic and biosynthetic approach is more appropriate for larger proteins that contain a fragment which has been produced by chemical synthesis.

Combination of Synthetic and Biosynthetic Methods

The technical difficulties with ligation of several synthetic polyeptides have led to emphasis in recent years on the use of a combination of chemical synthesis of part of the target molecule and cellular expression of the other part or parts, since the biosynthetic route does not have any principle limitation in terms of size. Application of the native ligation method requires production of an N-terminal peptide with a C-terminal thioester and a C-terminal fragment with an N-terminal cysteine, as described above. One way to do this is to synthesize the N-terminal peptide chemically, and express the remainder of the protein, for example in bacteria, in such a manner that an N-terminal cysteine is produced or can be generated (for example, by proteolytic removal of a fusion peptide). This is the appropriate approach for production of proteins with targeted modifications in the N-terminal region of the protein and allows, in its simplest form, such manipulations in approximately the first 60—70 residues. There are as yet not many examples of the application of this approach.

If the part of the protein which is to be modified, and should therefore be synthesized chemically, is the C-terminal region, a different strategy is needed. The N-terminal region must be produced biosynthetically if it is too large for solid-phase synthesis, but convenient chemical procedures for the conversion of the C-terminal carboxy group of an expressed protein into a thioester are not available. Fortunately, nature provides a way to overcome this problem in the form of protein splicing.[28, 29]

Proteins which undergo this intramolecular splicing typically have an N-extein – intein – C-extein structure and splicing involves excision of the intein part and ligation of the N extein and the C extein (Scheme 2). This process involves the following steps: 1) N $\rightarrow$S acyl transfer at a cysteine residue at the first N-extein/intein junction; 2) transthioesterification with a cysteine residue at the intein/C-extein junction, which results in transfer of the N extein to the N terminus of the C extein through a thioester link; 3) S $\rightarrow$N acyl transfer with the C-terminal residue of the intein, which results in release of the intein with a succinimide C terminus and generation of a native peptide bond between the two exteins. Mutant intein proteins have been made in which only the first of these steps occurs.[30] The thioester-linked product of this step can be cleaved by thiol reagents. Thus, if the protein of interest is present as an N-terminal fusion protein with the intein (that is, in place of the N extein), it can be cleaved by thiol reagents and is generated with



Scheme 2. *Chemical mechanism of protein splicing.*

a thioester at its C terminus, which in turn can be used for native chemical ligation.

This general approach has been termed expressed protein ligation and there are a number of examples of its application.[31–35] The procedure has been made convenient by inclusion of an affinity tag in the intein fragment that allows immobilization on a solid support before thiol-induced cleavage (see Figure 1). In the example shown, the method was used to prepare a derivative of the Raslike GTPase Rab7, a protein involved in vesicular protein transport in the cell. This example illustrates some unique advantages of the approach.[36] The full-length Rab7 protein has two reactive cysteines near the C terminus, which inevitably react with reagents used for fluorescent labeling of proteins. While this reaction can be advantageous in examinations of the interaction of the GTPase with other proteins,[37, 38] these are the cysteines which become prenylated on interaction with geranylgeranyl transferase so it is obviously not desirable to label them in studies in which prenylation should occur. Thus, in the C-terminal peptide used for the ligation reaction, a peptide with the sequence C–K(dansyl)–S–C–S–C was used. The result was a fluorescent Rab7 analogue that showed significant changes on interaction with partner proteins and could also be post-translationally modified (geranylgeranylated) at the two C-terminal cysteines.

Summary

In this brief review of methods currently available for the production of modified proteins, we have compared the advantages and disadvantages of the techniques available. These arguments are summarized in Table 1. In addition, Figure 2 illustrates the possibilities offered by the presently available combination of chemical and molecular biological methods for the production of proteins with tailor-made properties. This makes it clear that the goal of completely controlled manipulation of protein structures is becoming realistic, opening up a large number of potential uses in basic and applied research and production.

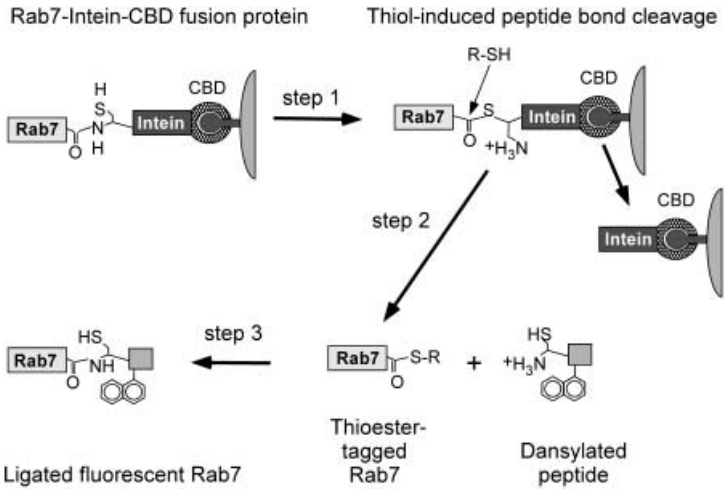


Figure 1. *Use of expressed protein ligation to generate Rab7 labeled at its C terminus with the fluorescent dansyl group. CBD = chitin binding domain.*[36]

Table 1. *Advantages and disadvantages of methods for the production of modified proteins.*

Method	Advantages	Disadvantages	Comments
cellular expression	potentially high yields; convenient once system established; no molecular weight limit; rapid generation of mutants; generation of libraries possible	often not useful for eukaryotic proteins; only natural amino acids can be used	Potential for development of convenient, efficient eukaryotic expression systems
cell-free synthesis including in vitro suppression	no molecular weight limit; generation of libraries possible, unnatural amino acids possible	low yields; demanding chemistry involved	improved in vitro transcription/translation technology could lead to better yields
chemical synthesis	unnatural amino acid and backbone chemistry possible; generation of libraries possible	limited to less than about 100 amino acids	size limitation not likely to change significantly in the near future
chemical synthesis plus native ligation	as for chemical synthesis	larger proteins possible, but ligation of several fragments technically demanding	further strategic developments possible
combination of synthetic and biosynthetic routes including intein chemistry	for synthetic part as for chemical synthesis; large proteins possible	in general only N- and C-terminal regions amenable to unnatural modifications	complex strategies allow a synthetic peptide to be incorporated in the middle of a protein

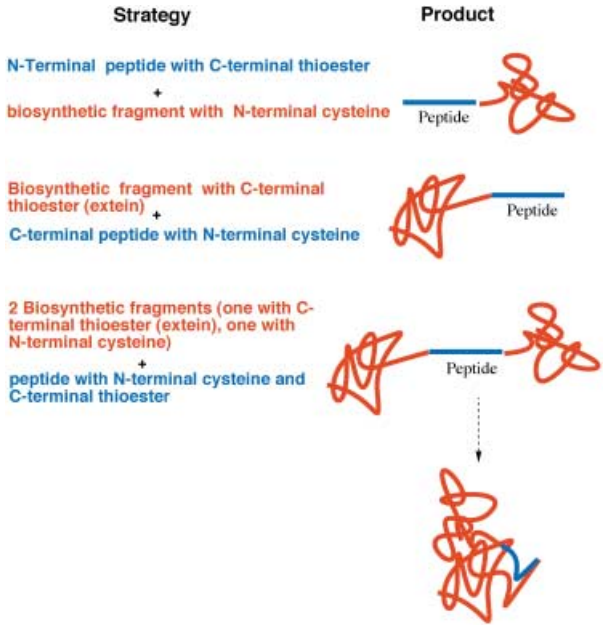


Figure 2. *Illustration of the potential for incorporation of a synthetic peptide, which can contain modified amino acids or other residues, into different locations in a protein molecule. Note that the assumption is made here that the native ligation reaction is carried out at a cysteine residue but, as discussed in the text, this is also possible at other positions.[26] The arrow indicates that folding of the full-length protein can occur so that the synthetic peptide or peptide analogue is part of the highly structured native conformation.*

[1] F. Baneyx, *Curr. Opin. Biotechnol.* **1999**, *10*, 411 – 421.
[2] J. L. Cereghino, J. M. Cregg, *FEMS Microbiol. Rev.* **2000**, *24*, 45 – 66.
[3] D. J. Manstein, K. M. Ruppel, L. Kubalek, J. A. Spudich, *J. Cell Sci. Suppl.* **1991**, *14*, 63 – 65.
[4] M. B. Slade, K. R. Emslie, K. L. Williams, *Biotechnol. Genet. Eng. Rev.* **1997**, *14*, 1 – 35.
[5] T. A. Kost, J. P. Condreay, *Curr. Opin. Biotechnol.* **1999**, *10*, 428 – 433.
[6] G. L. Costa, J. C. Bauer, B. McGowan, M. Angert, M. P. Weiner, *Methods Mol. Biol.* **1996**, *57*, 239 – 248.
[7] J. Light, R. A. Lerner, *Bioorg. Med. Chem.* **1995**, *3*, 955 – 967.
[8] J. A. Kolkman, W. P. Stemmer, *Nat. Biotechnol.* **2001**, *19*, 423 – 428.
[9] J. A. Ellman, D. Mendel, S. Anthony-Cahill, C. J. Noren, P. G. Schultz, *Methods Enzymol.* **1991**, *202*, 301 – 336.
[10] J. A. Ellman, B. F. Volkman, D. Mendel, P. G. Schultz, D. E. Wemmer, *J. Am. Chem. Soc.* **1992**, *114*, 7959 – 7961.
[11] J. S. Thorson, E. Chapman, P. G. Schultz, *J. Am. Chem. Soc.* **1995**, *117*, 9361 – 9362.
[12] V. W. Cornish, D. R. Benson, C. A. Altenbach, K. Hideg, W. L. Hubbell, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2910 – 2914.
[13] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149 – 2154.
[14] D. Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom, J. A. Wells, *Science* **1994**, *266*, 243 – 247.
[15] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776 – 779.
[16] C. F. Liu, J. P. Tam, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6584 – 6588.
[17] J. P. Tam, Q. Yu, Z. Miao, *Biopolymers* **1999**, *51*, 311 – 332.
[18] T. W. Muir, P. E. Dawson, S. B. Kent, *Methods Enzymol.* **1997**, *289*, 266 – 298.
[19] P. E. Dawson, S. B. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923 – 960.
[20] M. Baca, T. W. Muir, M. Schnölzer, S. B. Kent, *J. Am. Chem. Soc.* **1995**, *117*, 1881 – 1887.
[21] L. E. Canne, A. R. Ferredemare, S. K. Burley, S. B. Kent, *J. Am. Chem. Soc.* **1995**, *117*, 2998 – 3007.
[22] J. R. Sydor, C. Herrmann, S. B. Kent, R. S. Goody, M. Engelhard, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 7865 – 7870.
[23] J. R. Sydor, R. P. Seidel, R. S. Goody, M. Engelhard, *FEBS Lett.* **1999**, *452*, 375 – 378.
[24] C. F. Becker, C. L. Hunter, R. P. Seidel, S. B. Kent, R. S. Goody, M. Engelhard, *Proc. Natl. Acad. Sci. USA*, submitted.
[25] C. F. Becker, C. L. Hunter, R. P. Seidel, S. B. Kent, R. S. Goody, M. Engelhard, *Chem. Biol.* **2001**, *8*, 243 – 252.
[26] P. Botti, M. R. Carrasco, S. B. Kent, *Tetrahedron Lett.* **2001**, *42*, 1831 – 1833.
[27] D. W. Low, M. G. Hill, M. R. Carrasco, S. B. Kent, P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554 – 6559.
[28] H. Paulus, *Chem. Soc. Rev.* **1998**, *27*, 375 – 386.
[29] H. Paulus, *Annu. Rev. Biochem.* **2000**, *69*, 447 – 496.
[30] S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus, M. Q. Xu, *Gene* **1997**, *192*, 271 – 281.
[31] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705 – 6710.
[32] G. J. Cotton, T. W. Muir, *Chem. Biol.* **2000**, *7*, 253 – 261.
[33] M. Huse, T. W. Muir, L. Xu, Y. G. Chen, J. Kuriyan, J. Massague, *Mol. Cell* **2001**, *8*, 671 – 682.
[34] D. Cowburn, T. W. Muir, *Methods Enzymol.* **2001**, *339*, 41 – 54.
[35] J. A. Camarero, D. Fushman, D. Cowburn, T. W. Muir, *Bioorg. Med. Chem.* **2001**, *9*, 2479 – 2484.
[36] A. Iakovenko, E. Rostkova, E. Merzlyak, A. M. Hillebrand, N. H. Thoma, R. S. Goody, K. Alexandrov, *FEBS Lett.* **2000**, *468*, 155 – 158.
[37] K. Alexandrov, I. Simon, V. Yurchenko, A. Iakovenko, E. Rostkova, A. J. Scheidig, R. S. Goody, *Eur. J. Biochem.* **1999**, *265*, 160 – 170.
[38] K. Alexandrov, A. J. Scheidig, R. S. Goody, *Methods Enzymol.* **2000**, *329*, 25 – 39.

Received: August 20, 2001 [M288]