Journal of Peptide Science
J. Peptide Sci. **9**: 574–593 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.475

JOURNAL OF PEPTIDE SCIENCE

Total Chemical Synthesis of Enzymes‡

STEPHEN KENT*

Institute for Biophysical Dynamics, Departments of Chemistry, Biochemistry & Molecular Biology, The University of Chicago

Received 26 March 2003
Accepted 26 March 2003

Abstract: The total synthesis, at will, of a wide variety of protein and enzyme molecules is made feasible by modern chemical ligation methods. As Emil Fischer intuitively understood, synthetic access to the enzyme molecule enables the power of chemical science to be applied to elucidating the molecular basis of catalytic function in unprecedented detail. Copyright © 2003 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: chemical protein synthesis; enzymes; chemical ligation; native chemical ligation; peptide synthesis; peptide thioester; solid phase peptide synthesis; solid phase chemical ligation; backbone engineering; HIV-1 protease

OBJECTIVE – ENZYMES

In this issue, we celebrate the centennial of the first chemical synthesis of a dipeptide [1] by the great German chemist, Emil Fischer. When Fischer turned his attention to the world of proteins, he had already carried out the body of research in sugar and purine chemistry for which he would receive the Nobel Prize in Chemistry [2]. And, he had already promulgated the Fischer 'lock and key' theory of enzyme action [3] for which he is best remembered to this day. Thus, at the very outset, Fischer had realized the central importance of chiral specificity in biochemical transformations, and had a clear conceptual picture of the origin of this specificity in the shape complementarity of enzymes and the chiral substrate chemicals on which they act.

For the next phase of his career, Fischer very deliberately set out to undertake the total chemical synthesis of enzyme molecules, in order to be able to define the molecular basis of enzyme action.

> My entire yearning is directed toward the first synthetic enzyme. If its preparation falls into my lap with the synthesis of a natural protein material, I will consider my mission fulfilled.
>
> E Fischer
> (letter to Adolf Baeyer, 5 Dec 1905†)

In 1902, Fischer and Franz Hofmeister had separately put forward the 'peptide theory' of protein structure: that proteins are made up of α-amino acids linked head-to-tail by anhydride (i.e. amide) bonds [4,5]. Fischer understood that enzymes were protein in nature, and that the total synthesis of an enzyme would entail the chemical preparation of a protein. To that end, he developed ingenious methods for the chemical synthesis of peptide chains of defined amino acid sequence, including chiral amino acids [6]. Within a few years, these methods led to the total synthesis of an 18 amino acid residue peptide, of defined chemical composition according to the analytical methods of that era [7]. However, Fischer never succeeded in his original goal of the total synthesis of an enzyme molecule.

* Correspondence to: Stephen Kent, The University of Chicago, Cummings Life Sciences Center #325, 920 East 58th Street, Chicago, IL 60637, USA; e-mail: skent@uchicago.edu

‡ A tribute to Hermann Emil Fischer on the 100th anniversary of the first chemical synthesis of a peptide.

† This letter is in the Fischer Collection, Bancroft Library, U.C. Berkeley [Quoted in 'A Skeptical Biochemist', Joseph S. Fruton, Harvard University Press, 1992].

Today Fischer's dream has been realized, and total synthesis is used to apply the science of chemistry to the study of the molecular basis of enzyme action [8]. An example of the power of modern synthetic methods for the elucidation of fundamental aspects of enzyme action can be seen in the case of the 'HIV-1 protease', an enzyme essential to the replication of the human immunodeficiency virus.

Case Study: HIV-1 Protease

Crystal structures. The original structures of the HIV-1 protease molecule, both unliganded ('empty') and as three separate complexes with substrate-derived inhibitors, were obtained by x-ray diffraction studies performed on crystalline enzyme prepared by total chemical synthesis [9–12] (Figure 1). These data were made freely available to the research community and formed the basis of the successful worldwide programmes in structure-based drug design that resulted in the development of many of the 'protease inhibitor' class of AIDS therapeutics [13].

Site specific labelling – NMR probe nuclei. More recently, total chemical synthesis was used to determine the chemical properties and ionization states of the catalytic aspartic acid side chain carboxyl functionalities in the active site of the HIV-1 protease [14]. An autolysis-resistant form of the enzyme was prepared that was specifically labelled with 98% ^{13}C enrichment at a single atom site, the 4 position (i.e. the β-carboxyl carbon atom) of Asp^{25}, in each subunit of the homodimeric enzyme. Studies of the chemical shift behaviour of the catalytic carboxyl functionalities were carried out by high field ^{13}C NMR. In the presence of the inhibitor pepstatin, two signals were observed that remained invariant as a function of pH from pH 4 to pH 8. Measurements in D_2O showed that the signal at 178.8 ppm originated from the un-ionized –COOH, while, surprisingly, the signal at 172.4 ppm arose from an ionized $-COO^-$ carboxylate. The anomalous chemical shift of the ionized carboxylate indicates that this functionality is found in a region of very low effective dielectric constant.

Examination of the crystal structure of the corresponding (recombinant) HIV-1 protease-pepstatin complex [15] showed that one Asp^{25} side chain carboxyl is strongly H-bonded to the —OH of the pseudo-peptide bond 'isostere' (i.e. that part of the inhibitor thought to mimic the tetrahedral intermediate gem-diol in the productive enzyme-substrate complex). Significantly, the other Asp^{25} side chain carboxyl is in a hydrophobic pocket with at least 3.5 Å to the same –OH; i.e. it is *not* H-bonded. This is consistent with our ^{13}C NMR observations, using the assignments shown in Figure 2.

These observations immediately suggested a molecular basis for the enhanced nucleophilicity, observed in complexes of this type, of just one

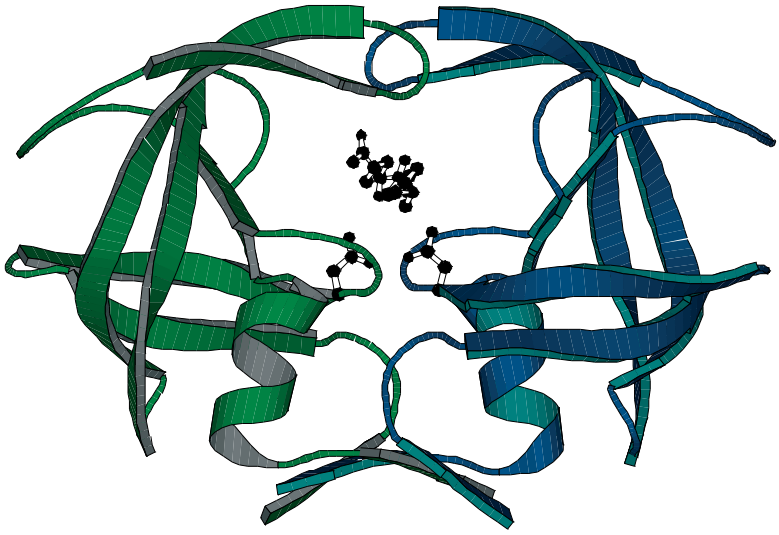


Figure 1 HIV-1 PR-MVT101 structure. Ribbon representation of the crystal structure of the chemically synthesized enzyme HIV-1 protease, complexed with the reduced peptide isostere inhibitor MVT101. This was the original structure determined for this important enzyme molecule [9,10], and formed the basis of the structure-based drug design programs that led to the development of the 'protease inhibitor' class of AIDS therapeutics [13].

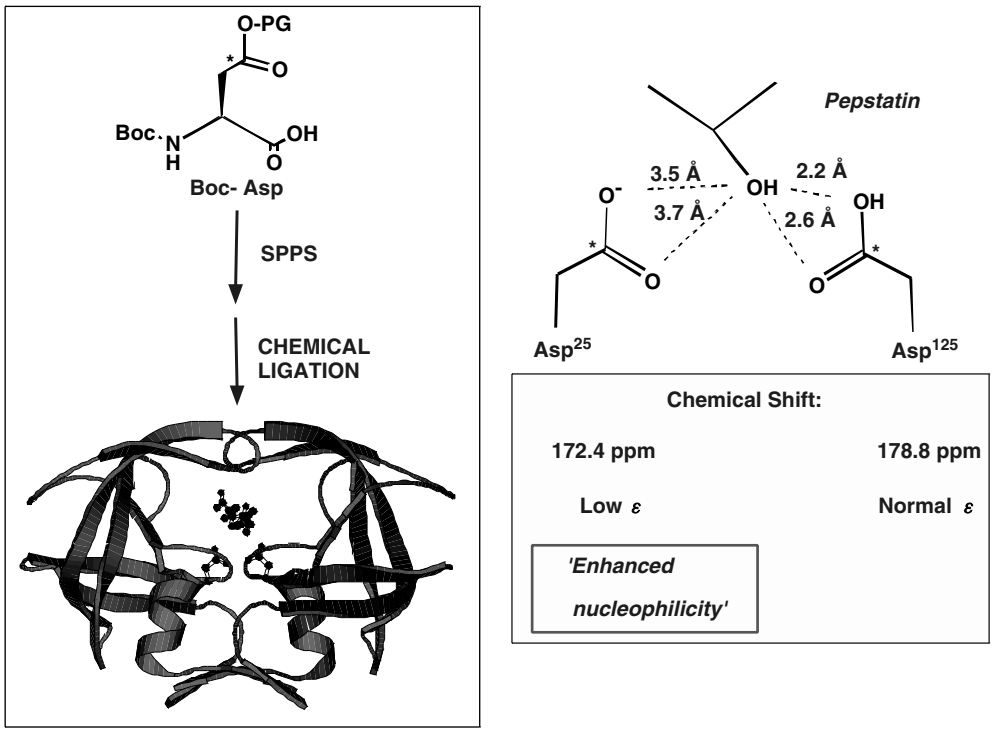


Figure 2 Site-specific labeling of the catalytic apparatus of an enzyme molecule. Total chemical synthesis was used to prepare ([4-^{13}C]Asp^{25,125})-labeled HIV-1 protease [14]. The ^{13}C nmr chemical shifts were measured for the active site side chain carboxyls in the presence of the inhibitor 'pepstatin'. The chemical shifts together with the protonation states, defined by D_2O exchange, showed that one of the two side chains existed as a carboxylate (i.e. in ionized form), yet was in a region of very low effective dielectric constant and not H-bonded. It is this side chain that shows the enhanced nucleophilicity that is the defining characteristic of the aspartyl proteinase enzymes [16].

of the two active site aspartic acid side chain carboxyls — a chemical property that defines the 'aspartic acid proteinase' class of enzymes [16]. An ionized carboxylate that is 'naked' (i.e. desolvated) in a non-polar pocket formed by the enzyme–ligand complex can be expected to show greatly enhanced nucleophilic reactivity. Total chemical synthesis has thus enabled single atom site-specific labelling with NMR probe nuclei, and in combination with NMR studies of the chemically synthesized enzyme has provided an insight into the chemical mechanism underlying the defining characteristic of the 'aspartic acid proteinases', an important class of enzymes in biology and medicine.

'Backbone Engineering' – role of the flaps. A unique feature of the HIV-1 protease is the pair of 'flap' structures that close in a near-symmetric fashion over substrate-derived inhibitors, as observed in numerous crystal structures [9–12,15]. Each flap consists of two anti-parallel β-strands connected by a reverse turn; the flaps close down in an edge-on fashion, and the Ile^{50} amide –NH– of each flap form hydrogen-bonds to a water molecule that is coordinated between the flaps and the carbonyl on each side of the 'scissile' bond in the substrate-derived inhibitor. These water-mediated hydrogen bonds are believed to be catalytically relevant, effecting electrophilic catalysis by imposing geometric distortion of the scissile amide bond. This tetrahedrally coordinated water is *not* observed in the crystal structures of the cell-encoded aspartyl proteinases, and has therefore served as the basis for design of inhibitors specific to the retroviral HIV-1 protease.

Previously, we had used total chemical synthesis to knock-out *both* of these Ile^{50} backbone –CO–NH– structures, replacing them with –CO–S– isosteres not capable of donating H-bonds [17]. Gratifyingly, we observed a ~3500-fold reduction in k_{cat} for this backbone-engineered HIV-1 protease compared with the control enzyme. This corresponds to a ~5 kcal/mole increase in activation energy for the rate limiting step in the enzyme–substrate complex, and was interpreted as being due to the deletion of

the two H-bonds. Because the {two-domain, ~300 residue polypeptide chain} cell-encoded aspartyl proteinases have only a single flap, we asked: what would be the effect of knocking out only one of the two Ile^{50} amide –NH– moieties in the HIV-1 protease? If both flap H-bonds contributed equally, then we would expect a ~2.5 kcal/mole effect. If on the other hand, one Ile^{50} amide –NH– is sufficient for catalysis (as must be the case for the single flap cell-encoded enzymes), then we would expect *full catalytic efficiency* in an enzyme with only one Ile^{50} amide –NH– knocked out.[‡]

In order to construct such a 'single flap' enzyme, it was necessary to establish synthetic access to a 'tethered dimer' form of the HIV-1 protease enzyme molecule [18–20].[§] A uniquely chemical approach was taken, employing a convergent synthetic strategy that enabled us to make a 202-amino acid residue protein with full enzymatic activity [21] [Figure 3]. Using this synthetic route, a construct was prepared with just *one* Ile^{50} amide –CO–NH– replaced by a –CO–O– moiety [22]. This enzyme, backbone engineered in a single flap, showed a reduction of k_{cat} of 2.2-fold. That is, correcting for loss of symmetry, the enzyme with only one Ile^{50} amide –CO–NH– knocked out had *full intrinsic enzymatic activity*.

[‡] Superimposed on these effects will be a factor of 2-fold reduction in k_{cat} due to the loss of symmetry in an enzyme with only one functional flap: i.e. whereas in the homodimeric enzyme with native backbone, the substrate can bind productively in either orientation, in the enzyme with only one native backbone flap it may be that only one orientation can be productive.

[§] The native HIV-1 protease is a homodimer of identical 99 residue polypeptide chains that make up a single functional enzyme molecule with a single active site at the interface between the two subunits (domains). It had previously been pointed out that, to effect change in only one subunit, it was necessary to make a *single polypeptide form* of the enzyme comprising 200+ amino acids; several such 'tethered dimers' had been made by recombinant means.

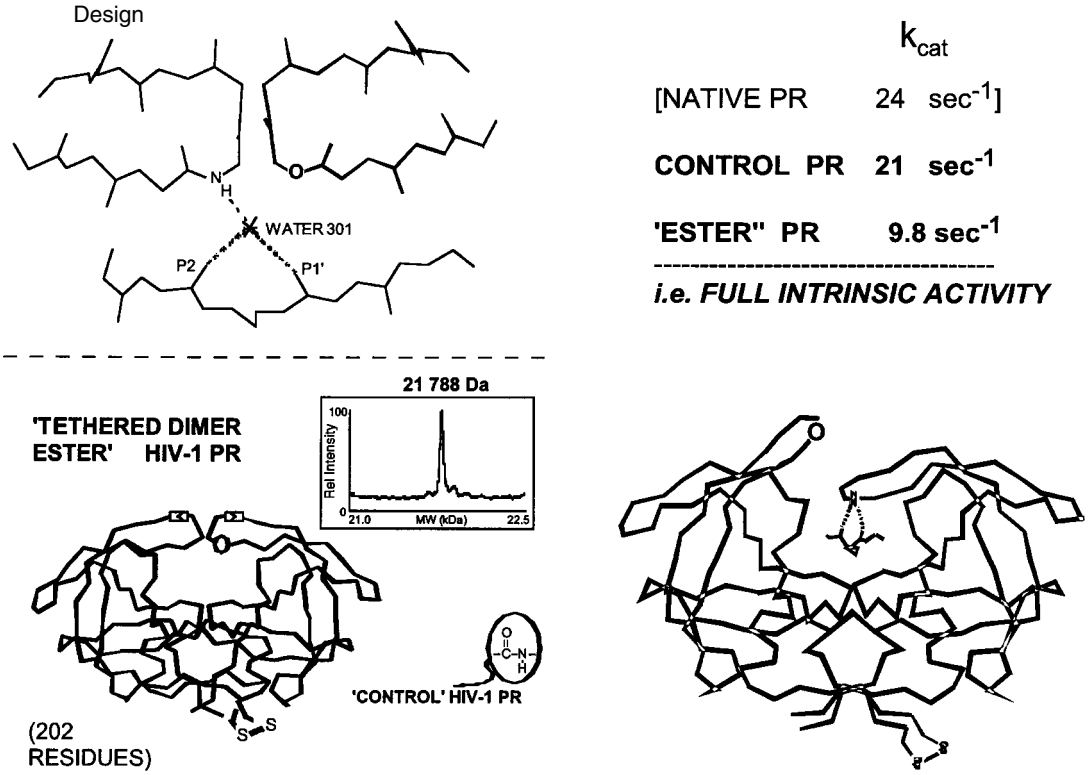


Figure 3 HIV-1 protease 'backbone engineered' in a single flap retains full intrinsic catalytic activity. Total chemical synthesis was used to prepare a 202 residue, 21kiloDalton tethered dimer form of the normally homodimeric enzyme molecule [21]. The control enzyme that contained *both* essential flap amide bonds showed full catalytic activity. The essential backbone amide –NH– [that hydrogen bonds the water molecule tetrahedrally coordinated between substrate and flaps] was deleted in *one* of the two flaps, by replacement with an ester moiety. This backbone-engineered molecule showed full intrinsic activity, even though deleting *both* essential flap amide bonds [17] reduced catalytic activity by a factor of 3500. Thus, the enzyme HIV-1 protease requires only a single flap for catalytic activity.

From this it is concluded that the HIV-1 protease can use just a single flap in the catalytically relevant enzyme–substrate complex. In our view, the tetrahedrally coordinated water molecule observed in many (but not all) structures of the HIV-1 protease complexed with substrate-derived inhibitors is most probably an artefact of the high C2 symmetry of the homodimeric retroviral enzyme molecule and is not catalytically productive. This has important implications for the design of improved HIV-1 protease inhibitors.

These two case studies illustrate the power of modern total chemical synthesis of enzymes in the elucidation of the molecular/chemical basis of enzyme catalysis. The next section describes some of the advances in synthetic chemistry that have made this type of research possible.

MODERN SYNTHETIC METHODS

Enzymes are protein molecules characterized by extraordinary efficiency for catalysing chemical reactions under physiological conditions. These protein catalysts are responsible for biochemical metabolism, and their action is characterized by very high rate enhancements and exquisite specificity of reaction. These functions originate in the precise three-dimensional folded structure of the protein molecule's polypeptide chain, i.e. from the 'tertiary' structure of the protein. Enzymes are typically protein molecules containing ~300 amino acids, corresponding to two 'domains'. The smallest enzymes (e.g. ribonculease, lysozyme) are polypeptide chains of ~125 amino acids.

Thus, in order to effect the total chemical synthesis of an enzyme, it is necessary to construct polypeptide chains of more than 100 amino acids, of defined covalent structure and in high purity. Over much of the twentieth century, the total synthesis of proteins was a major objective of the most talented synthetic organic chemistry research groups throughout the world. Sophisticated and ingenious synthetic chemistries arose from these efforts, particularly in Japan and Germany [23–28]. However, the successful total synthesis of proteins[¶] generally eluded the synthetic community. The chief reasons for this were: (i) the poor solubility of protected peptide intermediates formed in classical solution synthesis; and, (ii) the statistical accumulation of byproducts in stepwise solid phase methods. Added to this was the extreme degree of chemical sophistication of the most advanced synthetic methods, so that they were accessible to only a handful of extremely able chemists.

CHEMICAL LIGATION

Some 10 years ago, we introduced a new concept for the total chemical synthesis of proteins — the 'chemical ligation' method [29] — that has made the routine, reproducible total synthesis of proteins a practical reality [Figure 4].

Chemical ligation is based on two principles. First, the method makes use of *unprotected peptides*. The skill sets and chemistries needed to make unprotected peptides are widespread and accessible. Furthermore, we know how to keep unprotected peptides in solution and how to handle them; we know how to purify unprotected peptides by highly efficient yet standard methods such as reverse phase HPLC; and, we know how to characterize them with great precision using the powerful modern biological mass spectrometry techniques. This renders accessible high purity

[¶] We use an operational criterion for successful total synthesis: the ability to *reproducibly* undertake the preparation *at will* of protein molecules of high purity in multiple tens of milligram amounts for structure–function studies.

Figure 4 Chemical synthesis — a retrosynthetic analysis. Total chemical synthesis of an enzyme molecule consists essentially of three steps: i. chemical synthesis of peptide segments spanning the target sequence; ii. ligation of the peptide segments to give the full length polypeptide; and, iii. folding the linear polypeptide chain to form the active tertiary structure of the protein molecule. Use of the chemical ligation principle to stitch together *unprotected* peptide segments enables the use of building blocks that are easily made and that can be kept in solution at high concentrations, and thus reacted rapidly and in good yield. Absence of protecting groups also enables the direct use of electrospray mass spectrometry to characterize the starting materials, the intermediate products, the full length polypeptide, and the folded protein molecule with respect to purity and covalent structure.

Figure 5 Principles of chemical ligation [29]. Unprotected peptides are covalently joined by the chemoselective reaction of unique, mutually reactive functional groups to give a single product of defined covalent structure.

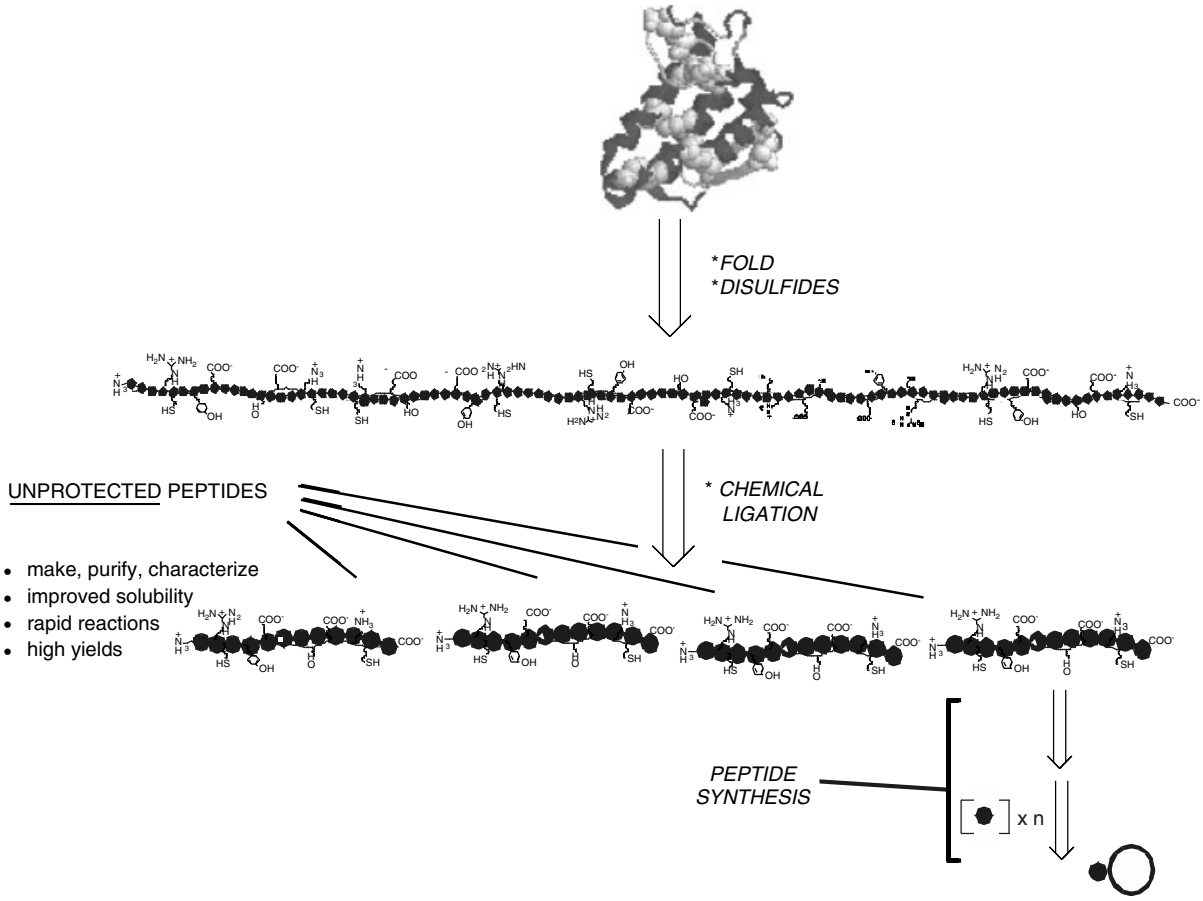


Figure 4

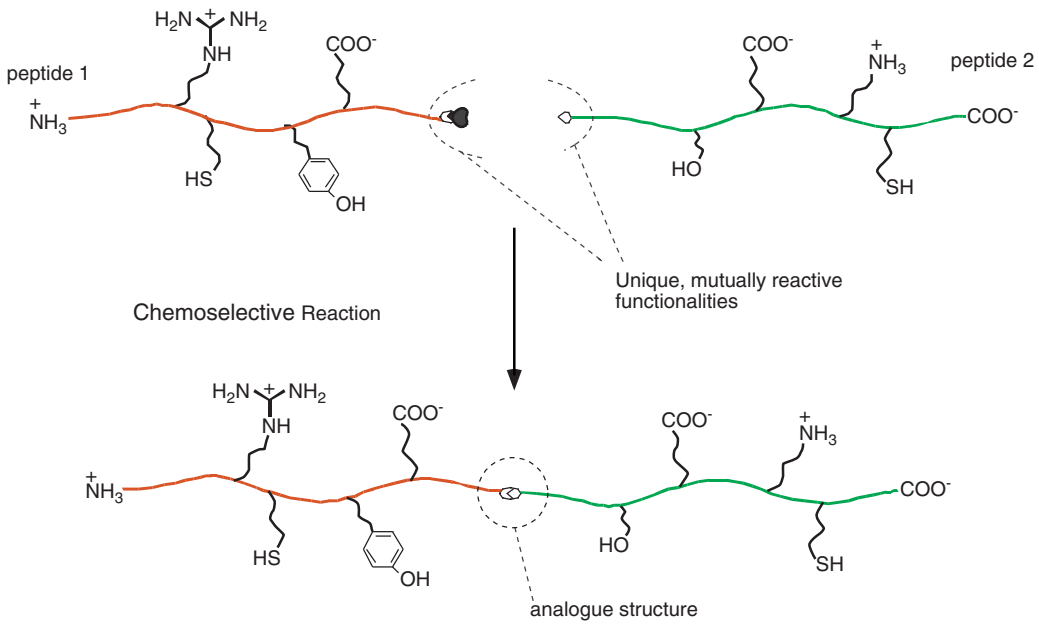


Figure 5

synthetic peptides of defined covalent structure. Second, we introduced the use of *chemoselective reaction* to join together, unambiguously and in high yield, unprotected peptide segments as building blocks for the preparation of larger polypeptide chains [29] [Figure 5].

By taking advantage of the ability to make and handle unprotected peptides, and the ability to join them together using simple chemistries, we have been able to successfully undertake the total chemical synthesis of hundreds of protein molecules, including enzymes [8]. Thanks to such chemical access to biologically active proteins, we have been able to make unique analogues that have provided important insights into the chemical and molecular principles underlying enzyme catalysis. A striking example of the power of chemical protein synthesis is the preparation of mirror-image enzymes.

Case Study: Mirror-Image Enzymes[‖]

D-HIV-1 protease. In order to demonstrate the unique capabilities of chemical protein synthesis[††], we undertook the total synthesis of the enantiomeric forms of the HIV-1 protease enzyme molecule, as shown in Figure 6. Using our original

[‖] The idea of mirror image enzymes has been around since the chiral specificity of enzyme catalysis was first understood. It has been a staple of science fiction writers for many decades. I remember in the early 1970s as a graduate student speculating with some of my fellow students — over a few beers on the Northside of Berkeley campus — about making mirror image enzyme molecules (we were going to use mirror image recombinant DNA methods — fantasy might as well go all the way). Imagine my surprise to find, upon joining Bruce Merrifield as a post-doc, that the preparation of mirror image forms of RNase A was a major objective of the research programme in the Merrifield laboratory. Despite years of effort on the parts of several of us, the synthetic methods and chemistries in use at the time were simply not up to the task of producing an enzyme molecule of this size

[††] To this day, it is technically difficult to simultaneously introduce more than one non-coded amino acid into a protein molecule by recombinant DNA methods. It is not possible at all to introduce D-amino acids into a protein using recombinant methods. The synthesis of the D-HIV-1 protease described here simultaneously

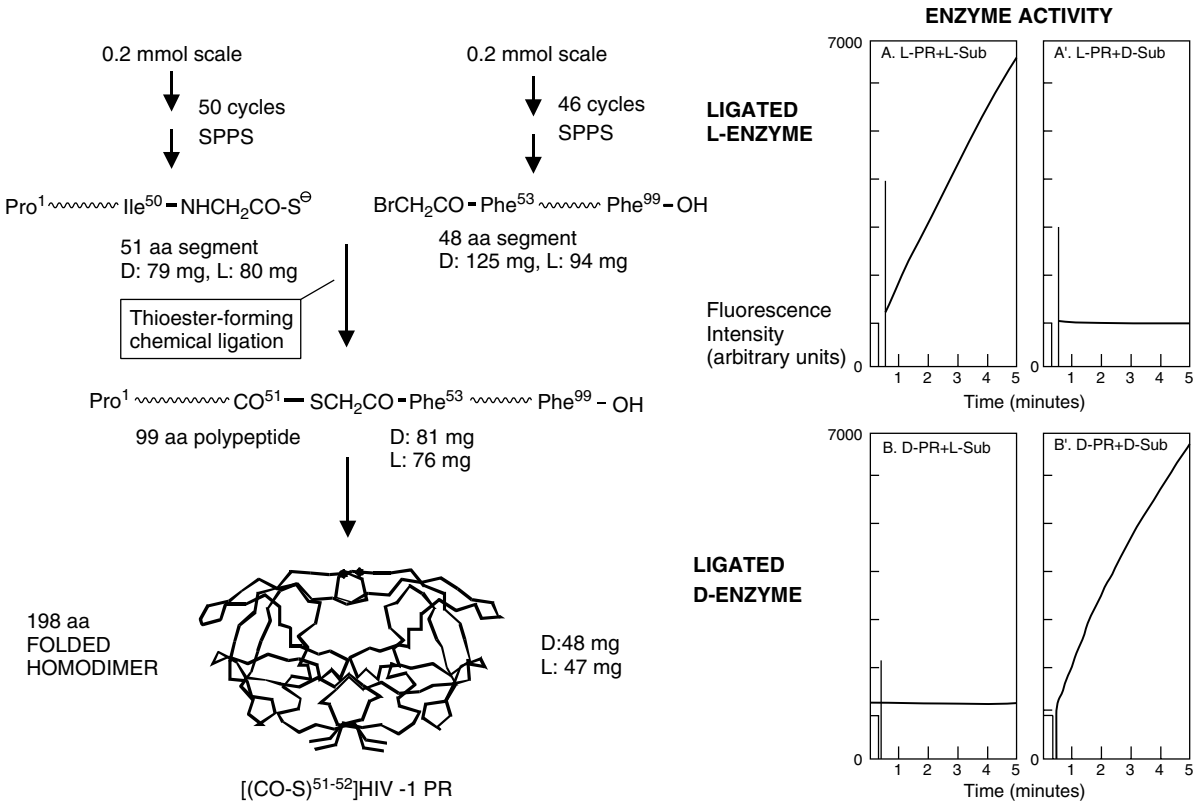


Figure 6 Total chemical synthesis and reciprocal chiral specificity of the mirror image enzymes D- and L-HIV-1 protease [30]. Thioester-forming chemical ligation [29] of two ~50 residue synthetic polypeptide segments was used to make ~80 milligrams each of the D-polypeptide and L-polypeptide chains. These were folded in good yield to give ~50 milligrams each of the high purity mirror image enzyme molecules. The L-enzyme acted only on the L-substrate, while the D-enzyme showed equivalent activity on the D-substrate, but was inactive on the L-substrate.

thioester-forming nucleophilic ligation chemistry, we were able to produce ~50 mg amounts of high purity preparations of each protein enantiomer [30]. The two proteins had full activity, and exhibited reciprocal chiral specificity; i.e. the L-protease cut the L-peptide substrate but did not act on the D-peptide substrate; conversely, the D-protease did not act on the L-peptide substrate, but did cut the D-peptide substrate.

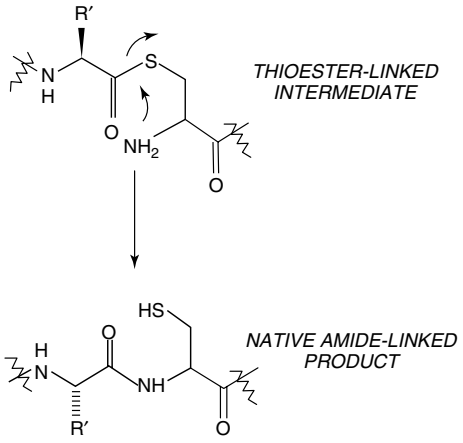
Maria Miller and Mohan Rao at the NCI-Frederick used the ligated D-HIV-1 protease to elucidate the high-resolution crystal structure of the mirror image form of the enzyme complexed to the mirror image form of a substrate-derived inhibitor [31]. As expected, the protein was in all essential respects the mirror image of the original synthetic {L-protease + L-inhibitor} HIV-1 protease structure. Of particular interest, were the mirrored relationships of secondary structural elements such as right- and left-handed alpha-helices, and the right- and left-handed twists of the anti-parallel beta-sheets. These experimentally determined independent structures of mirror image forms of a protein molecule illustrate in a clear and compelling fashion many of the general features of the folded tertiary structure of proteins, and serve as a vivid demonstration of the power of total chemical protein synthesis.

PEPTIDE-BOND FORMING CHEMICAL LIGATION

Our original thioester-forming ligation chemistry involved a simple chemoselective nucleophilic reaction that gave products containing a thioester moiety[‡‡] at the site of ligation (see Figure 6). This and similar ligation chemistries (oxime-forming; thiazolidine-forming; thioether-forming) have been used to make a variety of protein molecules with full biological activity, and to make a wide variety of protein analogues for structure–function studies [8]. Despite the utility of these original ligation chemistries, there was a continuing demand for a ligation method that would give a native amide bond at the site of ligation.

Native Chemical Ligation

To achieve the formation of a native amide bond at the site of ligation of two unprotected peptide segments, we envisioned the intramolecular nucleophilic attack [32] of an α-amine on an initial thioester-containing ligation product.[§§] This reaction would result in an S- to N-acyl shift [33], by analogy with the well-known O- to N-acyl shift observed under a variety of circumstances in peptide chemistry. The key intermediate is:



Initially, we sought to make the thioester-linked intermediate by attack of a peptide α-thiocarboxylate on β-bromoAla-peptide[¶¶]; however, this resulted in aziridine formation and ambiguous reaction pathways. My colleague Phil Dawson, who was at that time a PhD student, came up with the idea of using thiol-thioester exchange, in the presence of a thiol catalyst, which had the essential feature of promoting reversibility in the initial step of thioester-forming ligation. In this way, we developed the 'native chemical ligation' method for thioester-mediated chemical ligation of unprotected peptides at Cys residues [34], as shown in Figure 7.

Native chemical ligation relies on the reaction of a peptide-$^{\alpha}$thioester with a Cys-peptide to give an amide-linked product. A typical ligation reaction mixture is shown in Figure 8.

introduced 2×86 non-coded, D-amino acids into a single protein molecule (i.e. every amino acid in the 99 residue monomer sequence, less the 13 Gly residues). This dramatically illustrates that, once synthetic access has been established to a protein, there is effectively no limit to the modifications that can be made.

[‡‡] Thioesters are widely believed to be very labile to hydrolysis in aqueous solution, under neutral or slightly basic conditions. This is an erroneous belief. Careful studies have shown that under such conditions thioesters, –CO–SR, are slightly more stable towards hydrolysis than the corresponding oxoester, –CO–OR. Under standard conditions for handling synthetic peptides and proteins, typically water–acetonitrile mixtures containing 0.1% trifuoroacetic acid, peptide-thioesters are completely stable.

[§§] Anomalous susceptibility of thioesters to amine nucleophiles: much more reactive than corresponding oxoesters. Real, but not yet convincingly explained [32].

[¶¶] Unpublished observations. M. Baca, T. Walters, Summer, 1993.

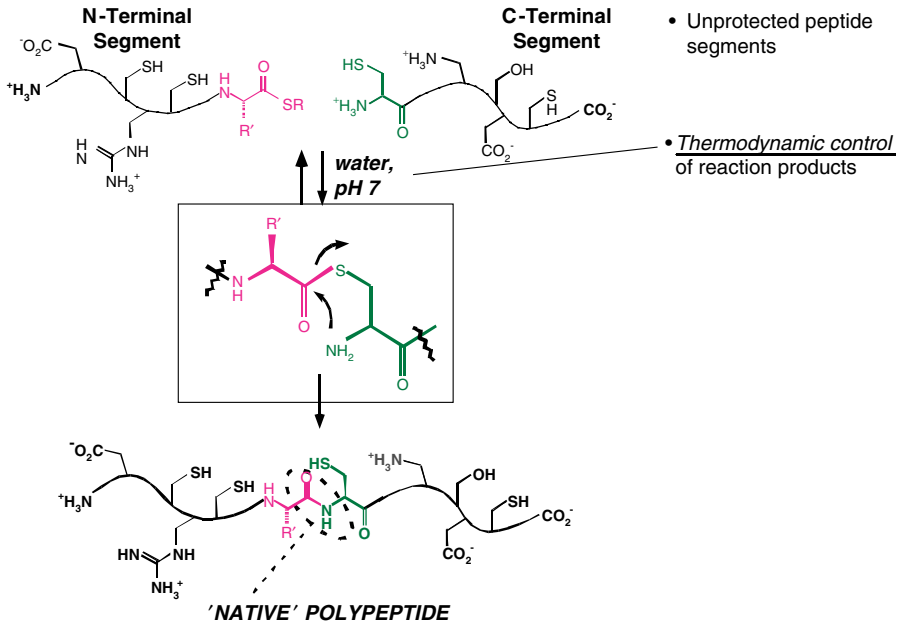


Figure 7 Native chemical ligation: *reversible* transthioesterification, followed by amide formation. Intramolecular nucleophilic attack of an α-amino group on the initial thioester product occurs only when the thiol is on the side chain of an *N*-terminal Cys, thus regenerating the thiol functional group of the Cys side chain and giving a final ligation product containing a native peptide bond at the site of ligation. The initial thiol-thioester exchange step is fully reversible, while the second amide-forming step is irreversible under the reaction conditions used; because of this, eventually only the desired amide-containing product is formed, even in the presence of internal Cys residues in either peptide segment. It is this reversible-irreversible two step reaction mechanism that is the *essence* of the native chemical ligation method.

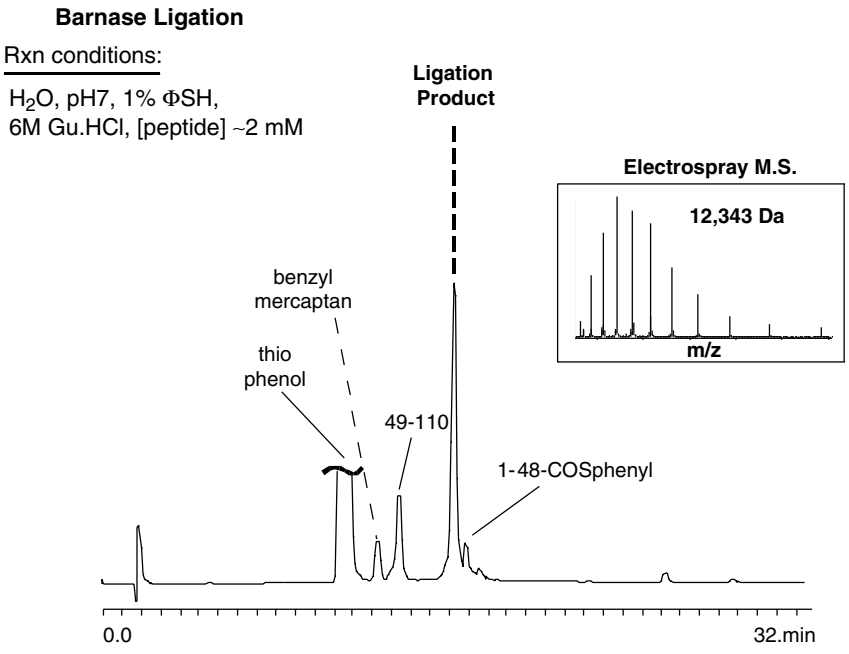


Figure 8 Native chemical ligation to give the 110 residue barnase molecule [37]. The two unprotected peptide segments 1-48$^{\alpha}$thioester and Cys^{49}-110 were reacted in aqueous 6M guanidine. HCl at pH7 in the presence of a thiol catalyst to give a near-quantitative yield of the full length 110 residue polypeptide chain. Analytical reverse phase HPLC of the reaction after 7 hours is shown, together with the electrospray mass spectrum of the full length product.

Reaction occurs in a matter of hours, and gives near-quantitative yields of the desired ligation product, even in the presence of other Cys residues in either or both peptide segments (see legend to Figure 7).

Case Study: Human Secretory PLA2

Human secretory phospholipase A2 is a small enzyme with a polypeptide chain of 124 amino acids; the mature folded protein contains 14 Cys residues as seven disulfide bonds. The total chemical synthesis of the Group II enzyme was undertaken by native chemical ligation of two large synthetic peptides (see Figure 9A,B) [35]. Early in the reaction, several thioester-linked species were observed. On more prolonged reaction, these intermediates converged on the single amide-linked target product polypeptide chain. This is consistent with our understanding of the mechanism of thioester-mediated native chemical ligation (see below). The linear polypeptide chain was readily purified by reverse phase HPLC, and folded in good yield to give a protein of extraordinary purity with full enzymatic activity, as shown in Figure 9C.

This synthesis also illustrates the power of modern biological electrospray mass spectrometry [36] for the characterization of the unprotected peptide building blocks, the linear polypeptide product and the final folded protein molecule. For example, from the precise mass difference of 15 ± 2 Da between the linear and folded molecules, it can be readily seen that seven disulfides, each resulting in a mass decrease of 2.0 Da, have been formed from the 14 Cys residues in the target sequence. The change in distribution of charge states between the linear polypeptide and the final protein product is also characteristic of the formation of a native tertiary structure. Finally, the absence — to great sensitivity — of products of lower or higher mass is rigorous proof of the purity and homogeneity of the final protein product and is a testament to the efficacy of the synthetic methods used.

CHEMICAL PROTEIN SYNTHESIS

The covalent coupling of unprotected peptide segments, by native ligation or otherwise, has been proven to be an effective route to the reproducible total synthesis of proteins. Enough experience has

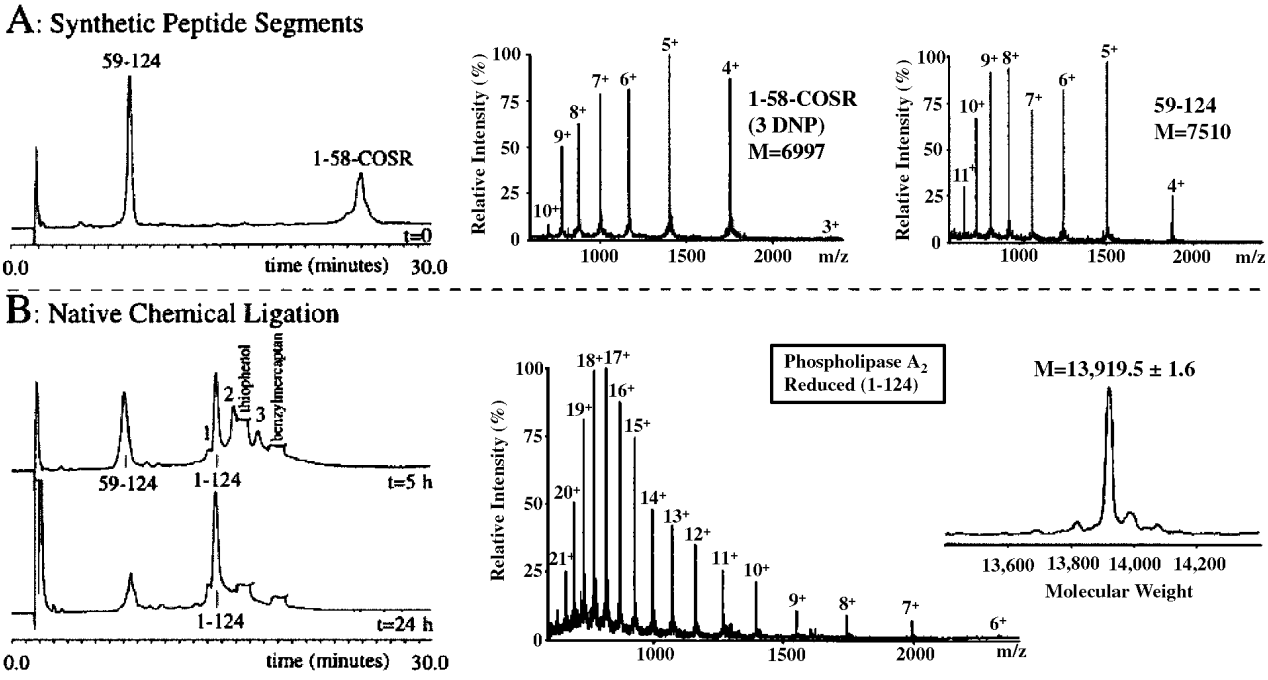


Figure 9 Total chemical synthesis of the enzyme human secretory PLA2 (Group 2) [35]. The target 124 amino acid residue polypeptide chain contains 14 Cys residues. Panel A: ligation reaction, t = 0. Panel B: ligation reaction, t = 5 h, 24 h. Panel C: folding the purified synthetic polypeptide to give the native enzyme molecule. The final product contained seven disulfides and had a measured mass of 13905 ± 1 Da, in agreement with the calculated mass.

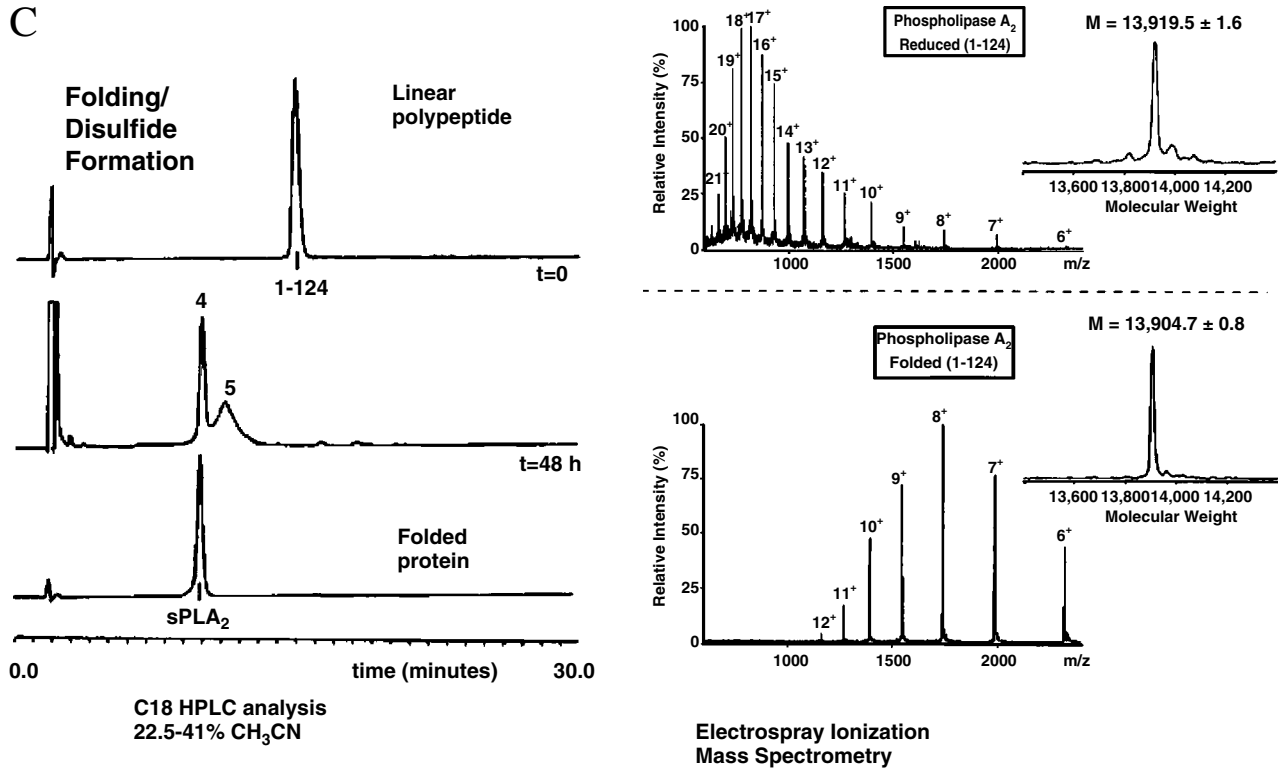


Figure 9 (*Continued*).

accumulated for a reliable understanding of many aspects of the method.

Mechanism and Kinetics

Amide-forming thioester mediated chemical ligation of unprotected peptides at Cys residues is the most practical chemistry for the total chemical synthesis of proteins. The reaction mechanism of thioester-mediated amide-forming chemical ligation at cysteine has been the subject of detailed investigation [37]. Based on these insights, simple adjustment of reaction conditions has led to routine quantitative yields of desired products. The kinetics of native chemical ligation have also been studied, and the effect of the identity of the *C*-terminal amino acid in the peptide-thioester has been quantified [38]. Reaction is slower for β-substituted amino acids at the *C*-terminal position of the peptide-thioester reactant, but is generally useful for most amino acids at that position.

Scope and Limitations

While it is possible to find many proteins whose polypeptide chains, in native form, may not be susceptible to the direct application of the native chemical ligation method, because of an absence of suitably placed Cys residues, there are an almost unlimited number of proteins (and whole *families* of important proteins) to which total synthesis by chemical ligation can be applied to good effect. For example, disulfide containing secretory proteins and zinc-finger proteins are two extensive families — each comprising thousands of proteins — that are directly amenable to total synthesis by native chemical ligation.

Simple stratagems can render virtually any protein amenable to total synthesis by the native chemical ligation of unprotected peptide segments at Cys residues. For example, where the target polypeptide does not contain suitably disposed natural Cys residues for native chemical ligation, it is straightforward to simply insert a Cys residue at the desired ligation site(s). It has been shown in a number of cases that a Cys can be introduced without compromising the folding and biological activity of the product protein [39]. In some cases, for ligation a Cys has been introduced where there is a Glu residue in the native sequence and then modified, post ligation, by alkylation with

bromoacetic acid to give a 'pseudo-Glu' residue, maintaining the negative charge of the native side chain [40].

New Ligation Chemistries

The principle of thioester-mediated amide forming intramolecular rearrangement has been used with *N*-linked thiol-containing cleavable auxiliaries to extend native chemical ligation to sites other than Cys residues, with some success [41–44]. For reasons not yet fully understood, where an *N*-linked thiol auxiliary is used the successful rearrangement of the thioester-linked intermediates to give the desired amide bond is restricted to ligation sites that contain a Gly residue.

A 'traceless' amide forming ligation chemistry has been prototyped, using an ingenious modification by Bertozzi and coworkers of the Staudinger reaction of tertiary phosphines with organic azides [45]. The Bertozzi group and the laboratory of Ron Raines at the University of Wisconsin [46] have developed the reaction with the aim of effecting the chemical ligation of peptide-$^{\alpha}$phosphinothioesters with azido-peptides to give native amide linked ligation products. This chemistry is still under development and holds great promise as an important addition to the repertoire of amide forming ligations of unprotected peptides.

Solubility

The use of unprotected peptides in conjunction with chaotropes such as guanidine.HCl has overcome the solubility problems associated with the maximal protection of functional groups in classic solution approaches, and has enabled the routine use of high concentrations of peptides and rapid, high yield ligation reactions [8]. This provides a simple and effective solution to the principal obstacle to classic solution chemical synthesis of proteins.

Racemization

The mild thioester activation, non-basic reaction conditions (ligation is carried out at neutral pH), and the intramolecular mechanism of amide formation result in no detectable racemization at the site of native chemical ligation [47].

Analytical Control

Because the peptide building blocks and the intermediate and final ligation products are all unprotected, they are readily purified by high resolution techniques such as reverse phase HPLC and/or charge based methods. Furthermore, modern techniques such as HPLC-ion trap electrospray mass spectrometry can be directly applied to analytical control of the synthesis, giving high precision information on the identity and purity of reaction products, both before and after purification.

Folding

Finally, and perhaps most gratifyingly, the synthetic polypeptide products have been shown to invariably fold *in vitro* in high yield to form protein molecules of defined tertiary structure [8]. The products of chemical synthesis by ligation methods are among the purest and best defined protein preparations ever produced, and are routinely used for multidimensional NMR and for high resolution x-ray crystallography to determine the folded structure of biologically active proteins.

Scale of Synthesis

In the research laboratory, ~50 mg amounts of each peptide building block are used, and the folded protein is obtained in amounts of several tens-of-milligrams after final purification (typically, 20–30 mg). This large quantity of product enables the routine application of modern methods for biophysical characterization of the synthetic protein products: e.g. biomolecular NMR; x-ray crystallography; differential scanning calorimetry; and CD spectroscopy. It also provides ample sample for the evaluation of the biological and biochemical activities of the protein in multiple systems.

Types of Proteins

Proteins from a wide variety of types have been prepared by total chemical synthesis (see Table 1), including enzymes, enzyme inhibitor proteins, secretory proteins, adaptor domains from intracellular signalling pathways, transcription factors and electron transfer proteins, among others.

Size of Proteins

Chemical ligation methods have enabled the synthesis of much larger protein molecules than was previously possible (Figure 10). Thus far, routine synthetic access has been established to single domain[III]

[III] Polypeptide 'domains' are the building blocks of the protein world. Typically, domains are 130 ± 40 amino acid residues in size.

Table 1 Chemical Protein Synthesis (TO DATE: >300 PROTEINS: >20 FAMILIES)

SIZE 4 kDa to >50 kDa	Secretory proteins	Enzymes
	EGF	Asp proteases
	AGRP	PLA2s
NOVEL	Chemokines	4OT/MIF
Crystal structures	Cytokines	Ras
NMR structures	TGF-β	Adaptor proteins
Biology	Ser-PR inhibitors	SH2
Precision labelling	OMTKY3	SH3
	Eglin c	Electron transport proteins
	BPTI	Cyt b562
	Transcription proteins	Rubredoxin
	Zn fingers	
	Myc/Max/Mad	

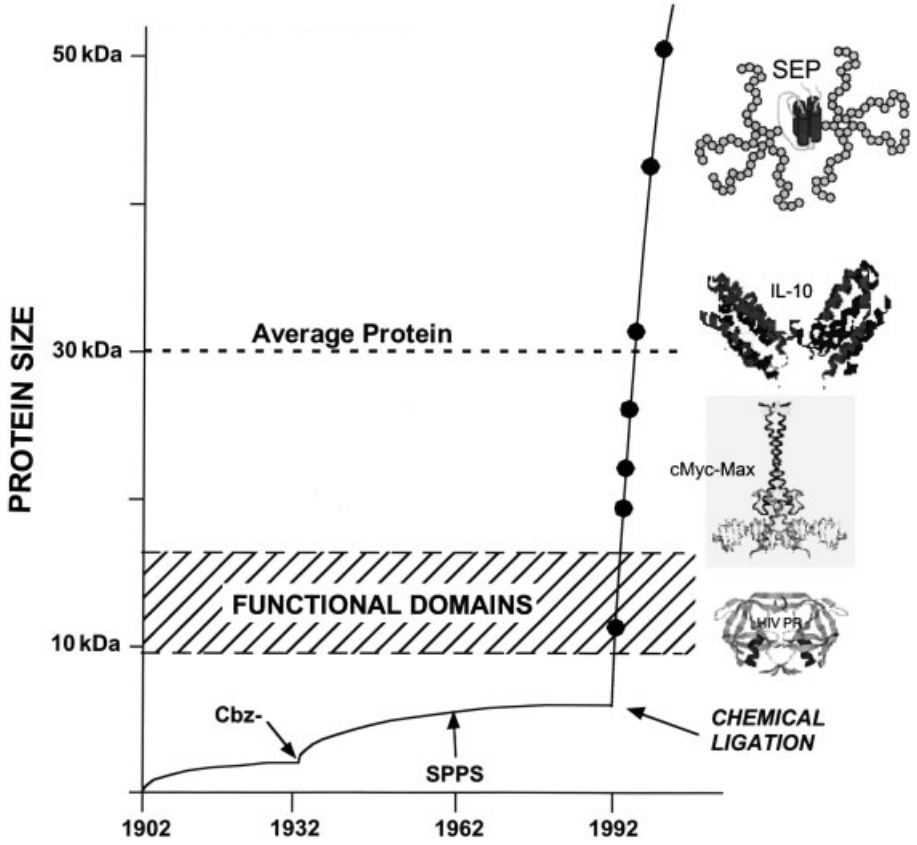


Figure 10 Size of protein molecules accessible by total chemical synthesis. The use of of chemical ligation and unprotected peptide segments as building blocks has enabled the modular construction of ever-larger protein molecules. Representative different size proteins prepared by total chemical synthesis are shown as cartoons.

proteins consisting of ~150 amino acids in size (i.e. up to ~18 kDa) [8]. A number of larger protein constructs have been prepared by total chemical synthesis. These include the ~22 kDa tethered dimers of the HIV-1 protease [21,22] (discussed above), and a ~21 kDa covalently linked cMyc-max transcription factor related protein [48]. More recently, a polymer-modified analogue of erythropoietin of 50,825 Da has been prepared by total chemical synthesis [40]. This glycoprotein mimetic was prepared by chemical

ligation of six unprotected building blocks. Each molecule contained a polypeptide of 166 amino acids and two copies of a branched {polyethyleneoxide-polyamide} precision polymer construct each of which contained four negative charges.

Data from genome sequencing projects shows that the typical protein consists of two domains and contains about 300 amino acids. Despite the successes of chemical protein synthesis, improved methods are still needed for the routine total synthesis of such protein molecules.

FUTURE DEVELOPMENTS

At the start of the 21st century, just 100 years after Emil Fischer's original chemical synthesis of a peptide, we find ourselves in an incongruous situation: in terms of the effort and time required, it is now easier to prepare *proteins* by total chemical synthesis than it is to prepare the necessary *peptide building blocks!*

How has it come about that chemical *protein* synthesis is more straightforward than chemical *peptide* synthesis?

The techniques used in chemical protein synthesis from unprotected peptide thioester building blocks are simple and effective. The chemoselective ligation principle [29] and chemistries such as thioester-mediated native chemical ligation for joining unprotected peptides has made these manipulations simple and independent of molecular size[†††]. Similarly, modern biological mass spectrometry has rendered straightforward the high precision analytical control of the total synthesis of large polypeptides by chemical ligation methods. Folding the synthetic polypeptide, which ought to have been the problem step in total synthesis, is dictated by the primary amino acid sequence and is invariably efficient and unambiguous under *in vitro* conditions.

By contrast, methods for the preparation by chemical synthesis of unprotected peptide-thioesters for use as building blocks are still technically challenging and time consuming. Relative times spent on peptide synthesis and chemical protein synthesis in the preparation of typical size proteins are shown in Figure 11.

[†††] This is key. There is now abundant empirical evidence that chemical ligation of large polypeptide chains proceeds with exactly the same rates and yields as for small peptide segments. This is true over a wide range of molecular sizes involving peptide building blocks from fewer than ten amino acids, to those with hundreds of amino acid residues.

Preparation of the Unprotected Peptide Building Blocks

The key starting materials for chemical protein synthesis are unprotected peptide-$^{\alpha}$thioesters of 30–40 amino acids in length.

Solid phase peptide synthesis. The synthesis of unprotected peptide-$^{\alpha}$thioesters of 30–50 amino acids is straightforward [38]. Modern, highly optimized Boc chemistry stepwise SPPS [49] is about 80% reliable in producing high purity peptide-thioesters of this size, when used in conjunction with reverse phase HPLC purification and analytical control by means of LC-ion trap mass spectrometry. A repeat synthesis of problem peptides invariably gives the desired peptide-$^{\alpha}$thioester product. The scale of synthesis in the research laboratory is typically ~200 μmol, yielding >100 mg of high purity peptide-$^{\alpha}$thioester for use in chemical protein synthesis.

The time required for such a synthesis, including chain assembly, cleavage and deprotection, lyophilization, preparative HPLC/analytical control, and final lyophilization, is about *5 days for each peptide product*. Including the 80% success rate, this means that if the syntheses are performed consecutively the preparation of four peptide-$^{\alpha}$thioesters will require 25 days, and the preparation of eight peptide-$^{\alpha}$thioesters will require *more than 50 days*! (See Figure 11). Thus, the preparation of the peptide building blocks takes about twice as much effort as the chemical synthesis of the protein molecule. If solid phase chemical ligation [50,51] is used to facilitate chemical protein synthesis, there is an even greater disparity between the time taken for chemical peptide synthesis and the time required for chemical protein synthesis (Figure 11).

The conclusion is clear: *it is the preparation of these peptide building blocks that is by far the rate limiting step in chemical protein synthesis.*

What is needed are rapid, parallel methods for the chemical synthesis and purification of unprotected peptide thioesters in multiple tens-of-milligram amounts, so that the preparation of a set of 4 to 8 peptide-$^{\alpha}$thioester building blocks can be accomplished in no more than 5 days. The dramatic impact that such synthetic peptide chemistry would have on the preparation of a 300 residue protein by chemical ligation of ten unprotected peptide segments is shown in Figure 11.

Expressed protein ligation. An ingenious method for the preparation of peptide-$^{\alpha}$thioesters by recombinant techniques was developed by Muir and

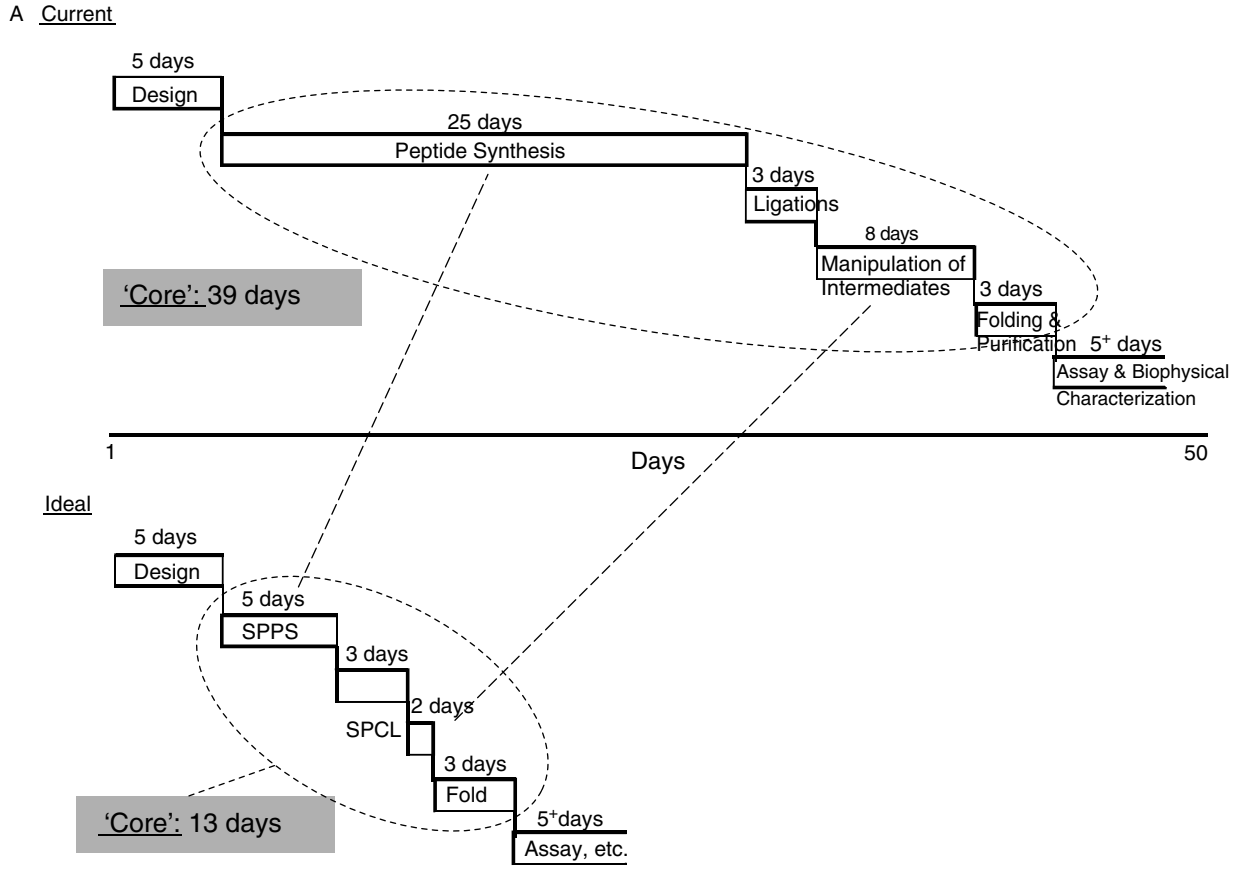


Figure 11 Chemical synthesis of the unprotected peptide building blocks is the principal bottleneck in chemical protein synthesis. There is also room for significant improvement in the handling and purification of intermediate ligation products. Parallel preparation of the necessary sets of peptide building blocks together with the use of solid phase chemical ligation [41,42] results in a 5-fold reduction of the time for 'core' synthetic processes, as illustrated for the preparation of a ~150 residue protein (A), and for a ~300 residue protein (B).

coworkers at The Rockefeller University [52]. This method makes use of a commercially available recombinant protein expression kit to generate a polypeptide$^{\alpha}$thioester that can then be used in native chemical ligation with a Cys-peptide prepared by chemical synthesis. In this way, larger polypeptide$^{\alpha}$thioesters can be obtained, and chemical modifications can be introduced into selected region of protein molecules without limitation as to size. For certain applications, this is a very effective way of overcoming the peptide synthesis bottleneck in chemical protein synthesis.

Chemical Ligations and the Manipulation of Intermediate Products

The most arduous aspect of chemical protein synthesis from unprotected peptide building blocks is the repeated preparative HPLC steps used for the isolation of the intermediate products at each stage of the synthesis. Together with the analytical evaluation of fractions and the lyophilization steps that accompany each purification, these manipulations can take 2–3 days for each intermediate product. The cumulative effect is shown in Figure 11, for four segment and eight segment syntheses. In the next section, we discuss a way to simplify and speed up these manipulations of intermediate products in chemical protein synthesis.

Solid phase chemical ligation. Rapid isolation of intermediates in a repetitive series of reactions is a problem that is directly amenable to the principles of polymer-supported chemical synthesis, pioneered by Bruce Merrifield in his development of solid phase peptide synthesis [53]. Prototype methods for the *solid phase chemical ligation* of unprotected peptides have been reported [50,51]. It can be estimated

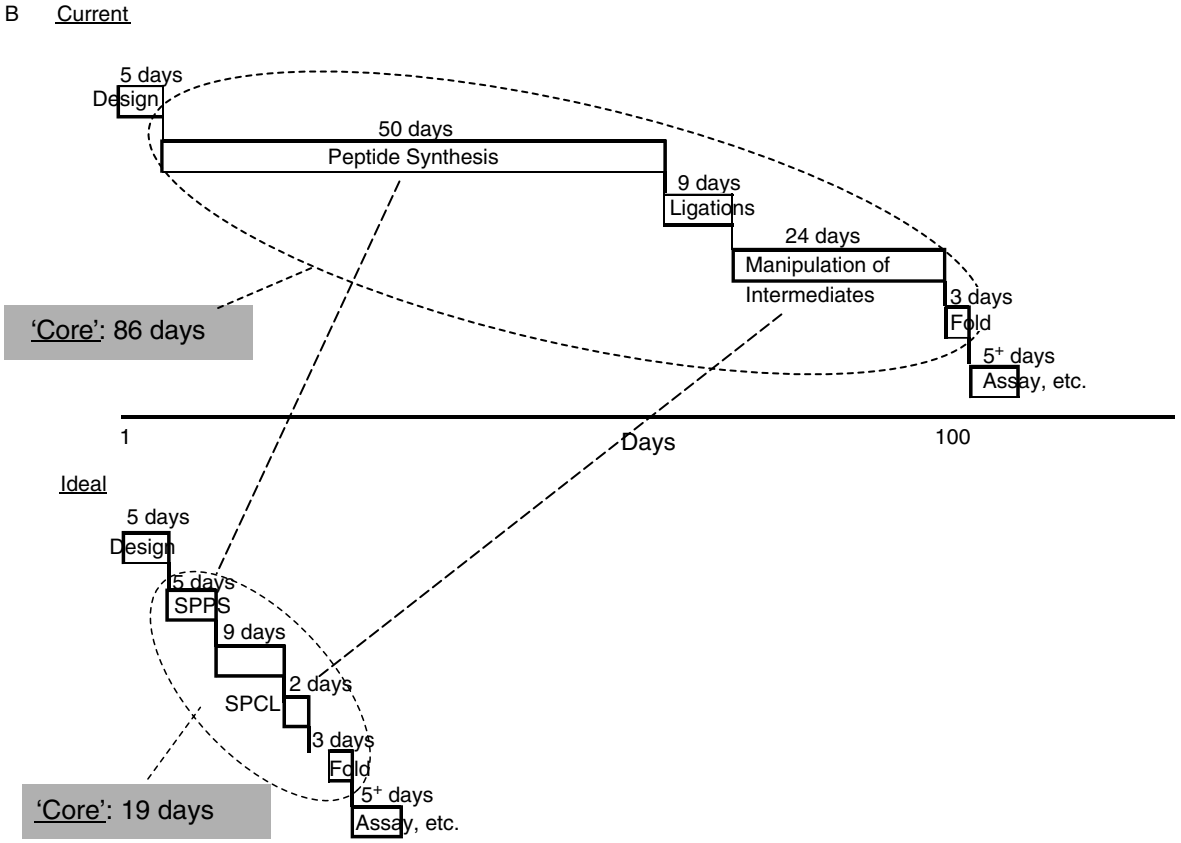


Figure 11 (*Continued*).

that the consequent ease of intermediate chemical manipulations and the rapid isolation of ligation products will reduce by 60–70% the time taken for the assembly of a long polypeptide from 4 or more unprotected peptide building blocks (see Figure 11).

It is worth remembering that, as with solid phase peptide synthesis, there is a price to pay for such synthetic convenience: the analytical control of synthesis becomes much more intractable, and the accumulation of polymer-bound byproducts leads to final products of reduced homogeneity. Nonetheless, with sufficient careful development of solid phase ligation chemistries, it is reasonable to suppose that the same level of efficacy will be achieved in the assembly of up to 10 peptide segments by solid phase chemical ligation as has been achieved for the synthesis of 10 residue peptides by addition of individual amino acids in stepwise solid phase peptide synthesis. This will revolutionize chemical protein synthesis.

A summary of recent innovations in chemical protein synthesis is shown in Table 2.

Table 2 Chemical Protein Synthesis — New Methods

- Expressed protein ligation
  Muir *et al.*, Rockefeller U [52]
- Solid phase chemical ligation
  Canne *et al.*, Gryphon Sciences [50]
  Dawson *et al.*, TSRI [51]
- Native ligation at non-Cys residues
  Canne *et al.*, TSRI [41]
  Botti *et al.*, Gryphon Sciences [43]
  Dawson *et al.*, TSRI [42]
- 'Traceless' ligation (Staudinger chemistry)
  Bertozzi *et al.*, U.C. Berkeley [45]
  Raines *et al.*, U. Wisconsin [46]

APPLICATIONS OF CHEMICAL PROTEIN SYNTHESIS

Using the chemical ligation approach, the science of chemistry can now be applied without restriction to the study of the protein molecule. Chemical

synthesis enables the application of all the ingenuity of the modern chemical methods to be applied to the study of the molecular basis of protein function. Applications range from the straightforward replacement of individual amino acid building blocks to much more elaborate and ingenious chemical schemes to engineer new forms of the protein molecule:

- Non-coded amino acids can be incorporated without limitation as to kind, position within the polypeptide chain and number of substitutions[‡‡‡]. Non-amino acid building blocks can also be used. For example, a bicyclic β-turn mimetic of fixed geometry was introduced into the HIV-1 protease molecule [54].
- Post-translational modifications: glycoproteins [55] and glycoprotein mimetics [40].
- Chemical synthesis can be used to introduce NMR probe nuclei at specific single atom sites in a protein molecule [14], in any desired number and combination. This can be invaluable for sorting out residue assignments in overlapping regions of the spectra[§§§]. Using expressed protein ligation, it is readily possible to mix and match biosynthetically isotope-enriched domains with unlabelled domains in order to simplify the interpretation of NMR spectra of larger proteins [56].
- Reporter moieties for physical techniques such as EPR or fluorescence spectroscopy can be introduced at will at any desired location within the protein molecule being studied [8].
- Radical re-engineering of the protein molecule has included: building in chemical cleavage sites to unzip the peptide chain at will for protein foot printing [57]; the preparation of proteins containing cyclic polypeptide chains [58]; the construction of topological analogues of proteins (e.g. two *N*-terminals, no *C*-terminus [48]; interpenetrating cyclic polypeptide chains [59]).

Physical Organic Chemistry of Enzyme Action

Perhaps the best reason to apply chemistry to the protein molecule is to study enzyme catalysed chemical reactions. The active site of the enzyme protein molecule provides a defined environment like no other. In principle, and to an increasing degree in reality, every aspect of the 'reaction medium' can be tailored for experimental purposes: geometry, electrostatics, dielectric constant, etc. This enables the systematic dissection of factors contributing to the reaction rate [14,17].

Integral Membrane Enzymes

Integral membrane proteins represent a very important class of understudied molecules with numerous critical biological functions[¶¶¶]. Recently, it has become clear that there exist integral membrane *enzymes* that work by mechanisms analogous to globular protein enzymes, but whose catalytic function is exerted *within* the lipid bilayer [60].

Total synthesis by means of chemical ligation of unprotected peptides has been shown to be an effective way of preparing ion channel proteins that are refractory to recombinant expression [61]. We are applying chemical protein synthesis to the study of integral membrane aspartyl proteinases and to integral membrane serine proteinases, with the objective of obtaining high resolution structural information and understanding the molecular basis of function in these unusual enzymes.

CONCLUSIONS

A mere century after Emil Fischer set the goal, the total chemical synthesis of enzyme molecules has been achieved. Modern chemical ligation methods enable the routine synthesis, at will, of a wide variety of protein and enzyme molecules.[‖‖‖‖] As

[‡‡‡] Despite considerable effort and scientific ingenuity, the use of recombinant methods to engineer proteins is thus far limited to the simultaneous introduction of no more than two different non-coded amino acids. The range of non-coded amino acids compatible with ribosomal synthesis is also limited.

[§§§] This will have particular application to polytopic helical integral membrane proteins; these molecules contain large numbers of identical hydrophobic amino acids in similar chemical environments. Labelling subsets of these residues with NMR probe nuclei will be essential to interpretation of high resolution magic angle spinning NMR spectra of membrane protein preparations.

[¶¶¶] It has been estimated that 20%–30% of all open reading frames across a wide variety of sequenced genomes code for integral membrane proteins with two or more membrane spanning ('transmembrane') regions. For example, the human genome contains a predicted ~1000 open reading frames coding for G-protein coupled receptors. At this date, there are no high resolution structures of any human GPCR.

[‖‖‖‖] It is instructive to compare the achievements of the past decade and the current state-of-the-art in the chemical synthesis of peptides and proteins with predictions of at the beginning of the 1990s (John Jones, in *The Chemical Synthesis of Peptides*, Clarendon Press: Oxford, 1991; 201). Jones foresaw the development of chemical synthesis based on ligation methods (the 'template method'), in general terms if not in terms of the

Fischer intuitively understood, synthetic access to the enzyme molecule enables the power of chemical science to be applied to elucidating in unprecedented detail the molecular basis of catalytic function of these protein molecules.

Acknowledgements

I wish to acknowledge the essential contributions of my coworkers and collaborators over the 10 years that it has taken to develop and apply the chemical ligation method of protein synthesis. I have endeavoured to appropriately cite their work throughout. This perspective on chemical protein synthesis is focused on the work performed in the author's laboratory. It is not possible in this forum to do justice to the large amount of outstanding work performed by other research groups on the chemical synthesis of proteins. Many research groups have made important and original contributions to the development of chemical ligation methods for use in protein synthesis. Numerous laboratories around the world have made use of these and other chemical synthesis methods in elegant studies of a wide variety of biologically significant protein systems.

REFERENCES

1. Fischer E, Fourneau E. Über einige Derivate des Glykokolls. *Ber. dtsch. chem. Ges.* 1901; **34**: 2868–2879.
2. Fischer Nobel Prize: http://www.nobel.se/chemistry/laureates/1902/
3. Fischer E. Einfluss der Configuration auf die Wirkung den Enzyme. *Ber. Dtsch. Chem. Ges.* 1894; **27**: 2985–2993.
4. Hofmeister F. Ueber Bau und Gruppierung der Eiweisskörper. *Ergeb. Physiol.* 1902; **1**: 759–802.
5. Fischer E. Ueber die Hydrolyse der Proteinstoffe. *Chemiker Zeitung* 1902; **26**: 939–940. (Hofmeister had presented his famous plenary lecture on protein structure in the morning of the 74th meeting of the Society of German Scientists and Physicians (1902, Karlsbad). That same afternoon, Fischer read his paper on the isolation of amino acids and peptides from protein hydrosylates, and suggested the coupling of amino acids in proteins). Adapted from: http://www.pepnet.com/fischerbio.html
6. Fischer E. Synthese von Polypeptiden, IV. Derivate des Phenylalanins. *Ber. dtsch. chem. Ges.* 1904; **37**: 3062–3071.
7. Fischer E. Synthese von Polypeptiden, XVII. *Ber. dtsch. chem. Ges.* 1907; **40**: 1754–1767.
8. Dawson PE, Kent SBH. Synthesis of native proteins by chemical ligation. *Ann. Rev. Biochem.* 2000; **69**: 925–962.
9. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SBH. Crystal structure of synthetic HIV-1 protease: conserved fold in retroviral proteases. *Science* 1989; **245**: 616–621.
10. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SBH, Wlodawer A. Structure of a complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* 1989; **246**: 1149–1152.
11. Swain AL, Miller M, Green J, Rich DH, Schneider J, Kent SBH, Wlodawer A. X-ray crystallographic structure of a complex between synthetic HIV-1 protease and a substrate-based hydroxyethylamine inhibitor. *Proc. Natl Acad. Sci. USA* 1990; **87**: 8805–8809.
12. Jaskolski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrikson RL, Schneider J, Kent SBH, Wlodawer A. Structure at 2.5 Å resolution of chemically synthesized HIV-1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* 1991; **30**: 1600–1609.
13. Kent S, Marshall GR, Wlodawer A, Perlmutter RM. Determining the 3D Structure of HIV-1 Protease. *Science* 2000; **288**: 1590.
14. Smith R, Brereton IM, Chai RY, Kent SBH. Ionization states of the catalytic residues in HIV-1 Protease. *Nature Structural Biol.* 1996; **3**: 946–950.
15. Fitzgerald PMD, Mckeever BM, Vanmiddlesworth JF, Springer JP, Heimbach JC, Leu C-T, Herber WK, Dixon RAF, Darke PL. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution. *J. Biol. Chem.* 1990; **265**: 14 209.
16. Hartsuck JA, Tang J. The carboxylate ion in the active center of pepsin. *J. Biol. Chem.* 1972; **247**: 2575–2580.
17. Baca M, Kent SBH. Catalytic contribution of flap-substrate hydrogen bonds in HIV-1 protease explored by chemical synthesis. *Proc. Natl Acad. Sci. US* 1993; **90**: 11 638–11 642.
18. DiIanni CL, Davis LJ, Holloway MK, Herber WK, Darke PL, Kohl NE, Dixon RA. Characterization of an active single polypeptide form of the human immunodeficiency virus type 1 protease. *J. Biol. Chem.* 1990,; **265**: 17 348–17 354.
19. Cheng YE, Yin FH, Foundling SF, Blomstrom S, Kettner CA. Stability and Activity of Human Immunodeficiency Virus Protease: Comparison of the

detailed chemistries. He also foresaw chemical protein synthesis, and the synthesis of glycoproteins. His prediction that 'AIDS will be curbed with a substantial contribution from peptide chemistry' has also come about in the form of the 'protease inhibitor' class of therapeutics [13].

Natural Dimer with a Homologous, Single-Chain Tethered Dimer. *Proc. National Academy Sciences USA* 1990; **87**: 9660–9664.
20. Tozer J, Yin FH, Cheng YS, Bagossi P, Weber IT, Harrison RW, Oroszlan S. Activity of tethered human immunodeficiency virus 1 protease containing mutations in the flap region of one subunit. *Eur. J. Biochem.* 1997; **244**: 235–241.
21. Baca M, Muir TW, Schnölzer M, Kent SBH. Chemical ligation of cysteine-containing peptides: synthesis of a 22 kDa tethered dimer of HIV-1 protease. *J. Am. Chem. Soc.* 1995; **117**: 1881–1887.
22. Baca M, Kent SBH. Protein backbone engineering through total chemical synthesis: new insight into the mechanism of HIV-1 protease catalysis. *Tetrahedron* 2000; **56**: 9503–9513.
23. Sieber P, Eisler K, Kamber B, Riniker B, Rittel W, *et al.* Synthesis and biological activity of two disulphide bond isomers of human insulin: [A7-A11,A6-B7-cystine]- and [A6-A7,A11-B7-cystine]insulin (human). *Hoppe Selyler's Z. Physiol. Chem.* 1978; **359**: 113–123.
24. Akaji K, Fujii N, Yajima H, Hayashi K, Mizuta K, *et al.* Studies on peptides. CXXVII. Synthesis of a tripentacontapeptide with epidermal growth factor activity. *Chem. Pharm. Bull.* 1985; **33**: 184–201.
25. Zawadzke LE, Berg JM. The structure of a centrosymmetric protein crystal. *Proteins* 1993; **16**: 301–305.
26. Inui T, Bodi J, Kubo S, Hishio H, Kimura T, Kojima S, Maruta H, Muramatsu T, Sakakibara S. Solution synthesis of human midkine, a novel heparin-binding neurotrophic factor consisting of 121 amino acid residues with five disulphide bonds. *J. Peptide Sci.* 1996; **2**: 28–39.
27. Hojo H, Kwon Y, Kakuta Y, Tsuda S, Tanaka I, Hikichi K, Aimoto S. Development of a linker with an enhanced stability for the preparation of peptide thioesters and its application to the synthesis of a stable-isotope-labeled hu-type DNA-binding protein. *Bull. Chem. Soc. Jpn.* 1993; **66**: 2700–2706.
28. Woo DD-L, Clarke-Lewis I, Chait BT, Kent SBH. Chemical synthesis in protein engineering: total synthesis, purification, and covalent structural characterization of a mitogenic protein, human transforming growth factor-α. *Protein Eng* 1989; **3**: 29–37.
29. Schnölzer M, Kent S. Constructing proteins by dovetailing unprotected synthetic peptides: backbone engineered HIV protease. *Science* 1992; **256**: 221–225.
30. Milton R. deL, Milton S, Schnölzer M, Kent SBH. Synthesis of proteins by chemical ligation of unprotected peptide segments: mirror-image enzyme molecules, D- & L-HIV protease analogues. In *Techniques in Protein Chemistry IV*, Angeletti R (ed.). Academic Press: New York, 1993; 257–267.
31. Miller M, Baca M, Rao JKM, Kent S. Probing the structural basis of the catalytic activity of HIV-1 protease through total chemical protein synthesis. *J. Molec. Structure (THEOCHEM)* 1998; **423**: 137–152.
32. Yang W, Drueckhammer DG. Computational studies of the aminolysis of oxoesters and thioesters in aqueous solution. *Organic Lett.* 2000; **2**: 4133–4136.
33. Wieland Th, *et al.* Uber peptidsynthesen. 8. Bildung von s-haltigen peptiden durch intramolekulare wanderung von aminoacylresten. *Liebig's Annalen der Chemie* 1953; **583**: 129–149.
34. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Synthesis of proteins by native chemical ligation. *Science* 1994; **266**: 776–779.
35. Hackeng TM, Mounier CM, Bon C, Dawson PE, Griffin JH, Kent SBH. Total chemical synthesis of enzymatically active human type II secretory phospholipase A2. *Proc. Natl Acad. Sci. USA* 1997; **94**: 7845–7850.
36. Chait BT, Kent SBH. Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science* 1992; **257**: 1885–1894.
37. Dawson PE, Churchill M, Ghadiri MR, Kent SBH. Modulation of reactivity in native chemical ligation through the use of thiol additives. *J. Am. Chem. Soc.* 1997; **119**: 4325–4329.
38. Hackeng TM, Griffin JH, Dawson PE. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl Acad. Sci. USA* 1999; **96**: 10 068–10 073.
39. Lu W, Randal M, Kossiakoff A, Kent SBH. Probing intermolecular backbone H-bonding in serine-proteinase protein-inhibitor complexes. *Chem. Biol.* 1999; **6**: 419–427.
40. Gerd GK, Chen S-Y, Mao F, Cressman S, Traviglia S, Shao H, Hunter C, Low D, Cagle N, Carnevali M, Gueriguian V, Keogh P, Porter H, Stratton SM, Con Wiedeke M, Wilken J, Tang J, Levy JJ, Miranda LP, Crnogorac M, Kalbag S, Botti P, Schindler-Horvath J, Savatski L, Adamson JW, Kung A, Kent SBH, Bradburne JA. Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* 2003; **299**: 884–887.
41. Canne LE Bark SJ, Kent SBH. Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* 1996; **118**: 5891–5896.
42. Offer J, Dawson PE. Nalpha-2-mercaptobenzylamine-assisted chemical ligation. *Org Lett.* 2000; **2**: 23–26.
43. Botti P, Carrasco MR, Kent SBH. Native chemical ligation using removable N^{α}-(1-phenyl-2-mercaptoethyl) auxiliaries. *Tetrahedron Lett.* 2001; **42**: 1831–1833.
44. Low DW, Hill MG, Carrasco MR, Botti P, Kent SBH. Total synthesis of cytochrome b562 by native chemical ligation using a removable auxiliary. *Proc. Natl Acad. Sci. USA* 2001; **98**: 6554–6559.
45. Saxon E, Armstrong JI, Bertozzi CR. A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* 2000; **2**: 2141–2143.

46. Nilsson BL, Kiessling LL, Raines RT. High-yielding Staudinger ligation of a phosphinothioester and azide to form a peptide. *Org. Lett.* 2001; **3**: 9–12.
47. Lu W, Qasim MA, Laskowski M Jr., Kent SBH. Probing intermolecular main chain H-bonding in serine proteinase-protein inhibitor complexes: chemical synthesis of a backbone-engineered turkey ovomucoid third domain. *Biochemistry* 1997; **36**: 673–679.
48. Canne LE, Ferré-D'Amaré AR, Burley SK, Kent SBH. Total chemical synthesis of a unique transcription factor-related protein: *cMyc-Max*. *J. Am. Chem. Soc.* 1995; **117**: 2998–3007.
49. Schnolzer M, Alewood P, Alewood D, Kent SBH. "In situ" neutralization protocols in Boc-chemistry solid phase peptide synthesis: rapid, high yield assembly of difficult sequences. *Int. J. Pept. Prot. Res.* 1992; **40**: 180–193.
50. Canne LE, Botti P, Simon RJ, Yijun Chen Y, Dennis EA, Kent SBH. Chemical protein synthesis by solid phase ligation of unprotected peptide segments. *J. Am. Chem. Soc.* 1999; **121**: 8720–8727.
51. Brik A, Keinan E, Dawson PE. Protein synthesis by solid-phase chemical ligation using a safety catch linker. *J. Org. Chem.* 2000; **65**: 3829–3835.
52. Muir TW. Semisynthesis of proteins by expressed protein ligation. Tom W. Muir, *Annu. Rev. Biochem.* (2003) **72**: 249–289.
53. Merrifield RB. Solid phase peptide synthesis. 1. Synthesis of a tetrapeptide. *J. Am. Chem. Soc.* 1963; **85**: 2149.
54. Baca M, Alewood P, Kent SBH. Structural engineering of HIV-1 protease with a β-turn mimic of fixed geometry. *Protein Sci.* 1993; **2**: 1085–1091.
55. Marcaurelle LA, Mizoue LS, Wilken J, Oldham L, Kent SB, Handel TM, Bertozzi CR. Chemical synthesis of lymphotactin: a glycosylated chemokine with a C-terminal mucin-like domain. *Chemistry* 2001; **7**: 1129–1132.
56. Cowburn D, Muir TW. Segmental isotopic labeling using expressed protein ligation. *Methods Enzymol* 2001; **339**: 41–54.
57. Muir TW, Dawson PE, Fitzgerald MC, Kent SBH. Probing the chemical basis of binding activity in an SH3 domain by protein signature analysis. *Chem. Biol.* 1996; **3**: 817–825.
58. Craik DJ, Simonsen S, Daly NL. The cyclotides: Novel macrocyclic peptides as scaffolds in drug design. *Curr. Opin. Drug Discov. Devel.* 2003; **5**: 251–260.
59. Blankenship JW, Dawson PE. Thermodynamics of a designed protein catenane. *J. Mol. Biol.* 2003; **327**: 537–548.
60. Urban S, Freeman M. Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Molecular Cell* 2003; **11**: 1425–1434.
61. Salom D, Lear JD, Kent SBH, DeGrado WF. Total chemical synthesis of the integral membrane protein influenza A virus M2 proton channel: role of its cytoplasmic domain for pore assembly. *Biochemistry* 1999; **38**: 11 905–11 913.

