Peptide ligation and its application to protein engineering

Graham J Cotton and Tom W Muir

The ability to assemble a target protein from a series of peptide fragments, either synthetic or biosynthetic in origin, enables the covalent structure of a protein to be modified in an unprecedented fashion. The present technologies available for performing such peptide ligations are discussed, with an emphasis on how these methodologies have been utilized in protein engineering to investigate biological processes.

Address: Laboratory of synthetic protein chemistry, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

Correspondence: Tom W Muir
E-mail: muirt@rockvax.rockefeller.edu

Chemistry & Biology September 1999, **6**:R247–R256
http://biomednet.com/elecref/10745521006R0247

1074-5521/99/$ – see front matter

Introduction

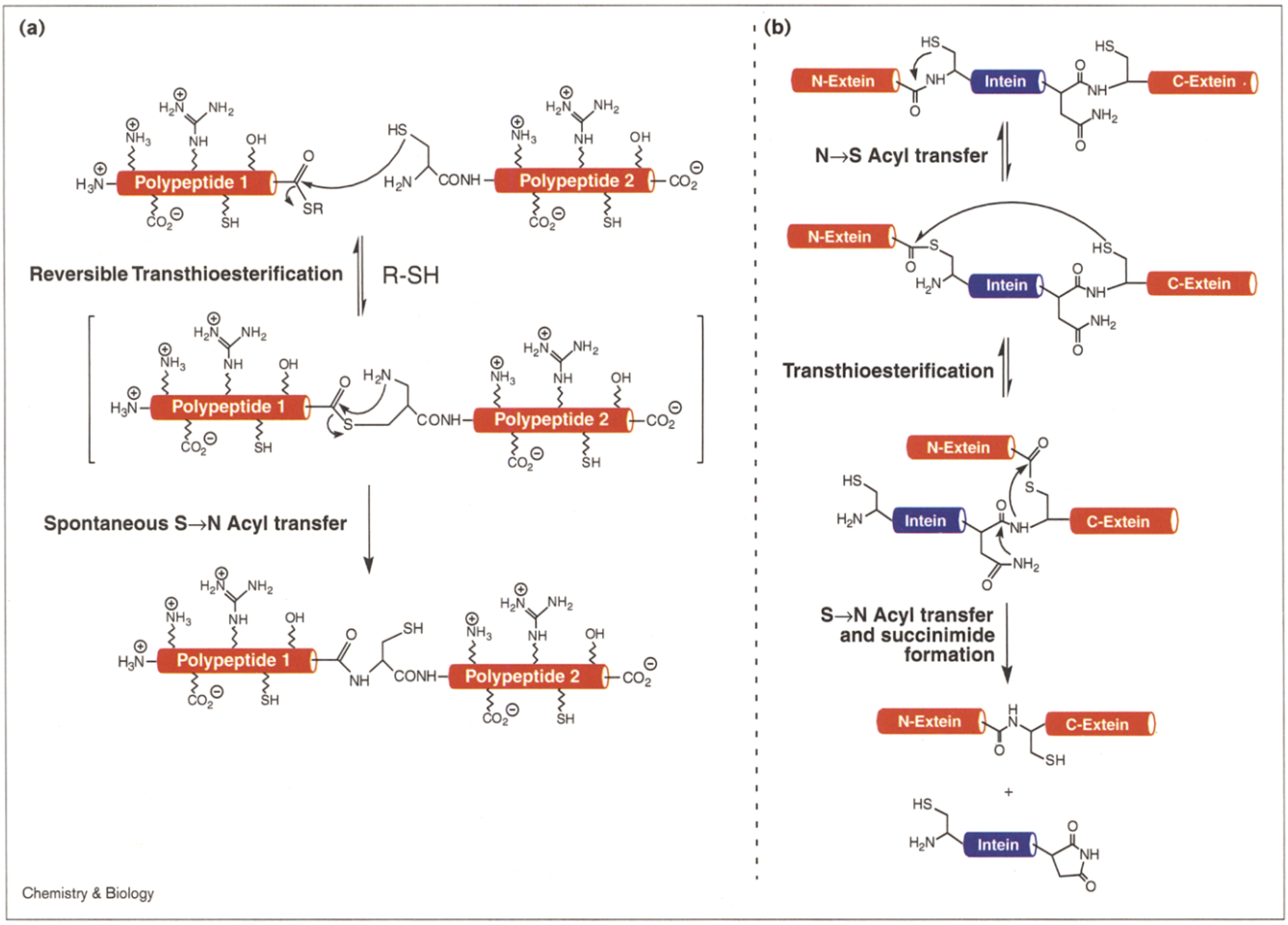
With the genetic blueprint of several organisms already available and the completion of the human genome project apparently well within sight, we are now facing a challenge worthy of a new millennium: characterization of many tens of thousands of novel gene products at the chemical and biological levels. Often referred to as the 'proteome project', this is an enormously daunting undertaking, particularly given the emerging picture of complexity in biological processes. It seems likely that both established and novel technologies will have to be brought to bear on the problem, the latter probably requiring the input of researchers from many branches of science and engineering. Chemistry, and in particular organic chemistry, is well poised to make a huge impact in this area, both through the use of small-molecule probes of biological processes [1] and, as will be the focus of this review, through the direct chemical manipulation of protein molecules. Precisely altering the covalent structure of a protein using synthetic chemistry provides a freedom in protein engineering unattainable by standard site-directed mutagenesis techniques. Accordingly, the development of approaches for the site-specific incorporation of unnatural amino acids into proteins has received considerable attention in recent years, and several biosynthetic and synthetic strategies are now available for this purpose [2–6].

This review will focus on just one of these technologies, namely the 'peptide ligation' approach [4,6]. Peptide ligation actually refers to a suite of approaches, all of which allow fully unprotected polypeptide building blocks to be regioselectively joined together in aqueous solution to create a target protein molecule. Ligation can be achieved either chemically [4], by incorporating unique functionalities at the amino and carboxyl termini of the peptide segments, or enzymatically using sophisticated reverse-proteolysis techniques [3,5]. The attraction of these modular ligation approaches is that they allow a selected region of a long protein sequence to be chemically manipulated in a manner analogous to a synthetic peptide, enabling proteins to be site-specifically modified in an hitherto unprecedented fashion. In this article, we will briefly review some recent technological developments in the peptide-ligation field and then attempt to illustrate the extraordinary potential of this area in protein engineering through a series of examples.

Bridging peptide chemistry and protein biotechnology

There are several different chemical and enzymatic ligation approaches available for the assembly of proteins

Figure 1



Comparison of the mechanisms of **(a)** native chemical ligation and **(b)** protein splicing.

from unprotected peptides (these have been reviewed extensively [3,4,6]). Indeed, small proteins and protein domains are now readily accessible to total synthesis, which is emerging as the method of choice for the rapid generation of these molecules [6]. Difficulties associated with the solid-phase peptide synthesis (SPPS) of peptides longer than ~50 amino acids do, however, render the assembly of large protein targets (> 100 residues) a significant challenge using only two synthetic peptide segments. One solution to this problem has been to link together three or more synthetic peptides either using a combination of orthogonal ligation approaches [7–9] or through the sequential use of the same methodology [9–12]. Only a handful of proteins, however, have been prepared using these strategies to date, primarily because they are technically demanding to perform. Perhaps as a consequence of this, recent developments in the field have centered on an alternative approach in which a combination of synthetic peptides and recombinantly derived polypeptides are used as the ligation building blocks. In principle, this semisynthetic strategy should allow synthetic access to extremely large protein systems by ligating short synthetic peptides (containing the chemical probe of interest) to much larger recombinant polypeptides. Pioneering work in this area includes the use of conformationally assisted condensation reactions [3], oxime and hydrazone bond forming ligation chemistries [13–15] and engineered peptide ligases [5].

The well-established 'native chemical ligation' method [16] has proved pivotal for many of the recent developments in protein semisynthesis. As illustrated in Figure 1a, the first step in native chemical ligation involves the chemoselective reaction that occurs at physiological pH between a peptide fragment containing an amino-terminal cysteine residue and a second peptide fragment containing an $^{\alpha}$thioester group. This initial transthioesterification reaction is then followed by a rapid intramolecular S→N acyl shift to generate an amide bond at the ligation junction. Note, additional cysteine residues

are permitted in one or both peptide segments because of the reversible nature of the initial transthioesterification step. This reaction has been widely used for ligating two synthetic peptide fragments together [6], where the necessary reactive functionalities can be incorporated into the fragments during SPPS. Significantly, the necessary reactive groups can now be incorporated in a general fashion into recombinant proteins by utilizing new directives in protein engineering.

Techniques for producing recombinant amino-terminal cysteine proteins, for use in native chemical ligations, have been established for some time [17]. It is only recently, however, that the requisite technologies have been developed that enable recombinant $^{\alpha}$thioester proteins to be generated. This breakthrough was achieved by manipulating a naturally occurring biological phenomenon known as protein splicing [18], a process that bears remarkable similarities to native chemical ligation (Figure 1). Protein splicing is a post-translational process in which a precursor protein undergoes a series of intramolecular rearrangements that result in precise removal of an internal region, referred to as an intein, and ligation of the two flanking sequences, termed exteins (Figure 1b). Although there are, with one exception (see below), no sequence requirements in either of the exteins, inteins are characterized by several conserved sequence motifs and approximately one hundred members of this protein domain family have now been identified (for a comprehensive listing, see www.neb.com/neb/Frame_tech.html).

The first step in protein splicing involves an N→S (or N→O) acyl shift in which the N-extein unit is transferred to the sidechain SH or OH group of a conserved Cys/Ser/Thr residue, always located at the immediate amino terminus of the intein. Note, it has been speculated that the intein structure provides the driving force for this thermodynamically unfavorable rearrangement by twisting the scissile amide-bond into a higher energy conformation [19]. The entire N-extein unit is then transferred to a second conserved Cys/Ser/Thr residue at the intein-C-extein boundary (+1 position) in a transesterification step. The resulting branched intermediate is then resolved through a cyclization reaction involving a conserved asparagine residue at the carboxyl terminus of the intein. The intein is therefore excised as a carboxy-terminal succinimide derivative. In the final step, an amide bond is formed between the two exteins following an S→N (or O→N) acyl shift, a step reminiscent of native chemical ligation. These mechanistic insights have led to the design of a number of mutant inteins that can only promote the first step of protein splicing [20–23]. Proteins expressed as in-frame amino-terminal fusions to one of these engineered inteins can be cleaved by thiols via an intermolecular transthioesterification reaction to generate the recombinant protein $^{\alpha}$thioester derivative [21]. Peptide sequences containing an amino-terminal cysteine residue can then be specifically ligated to the carboxyl termini of these recombinant $^{\alpha}$thioester proteins [24–26], in a procedure termed expressed protein ligation (EPL) or intein-mediated protein ligation (IPL). It is important to note inteins are not the only protein domains that can be used to generate recombinant protein $^{\alpha}$thioesters. Indeed, Beachy and coworkers [27] have demonstrated that the autoprocessing domain of the protein Hedgehog (which is structurally related to inteins) can be used to generate semisynthetic proteins in a manner analogous to EPL.

Methods for generating recombinant amino-terminal cysteine proteins rely on cleavage of an appropriate precursor protein. In the enzymatic method developed by Verdine and coworkers [17], the recognition sequence for the protease factor Xa is introduced immediately in front of the cryptic amino-terminal cysteine in the protein. Because factor Xa cleaves directly after its recognition site, the desired amino-terminal cysteine protein is simply liberated by mild enzymatic treatment. Other proteases that cleave after their recognition site, such as enterokinase or ubiquitin carboxy-terminal hydrolase, should also be compatible with such an approach. An alternative strategy has recently been described that negates the need for this proteolytic step [22,23,28]. This clever approach was again developed from studies on protein splicing, and utilizes inteins that have been engineered to promote direct cleavage at their carboxyl terminus, even with a cysteine residue at the +1 position. The protein of interest is simply expressed as carboxy-terminal fusion to one of these engineered inteins and the desired material is liberated through spontaneous cleavage at the intein-C-extein junction.

Benzyl $^{\alpha}$thioester containing peptides have been shown to be excellent substrates for the peptide ligase subtiligase [29], a double mutant version of the serine protease subtilisin [29,30]. Welker and Scheraga [31] have recently shown that benzyl $^{\alpha}$thioester derivatives of recombinant proteins can be prepared using intein technology, a method that the authors suggest will allow synthetic peptides to be enzymatically ligated to the carboxyl terminus of recombinant proteins. Although this remains to be experimentally confirmed, this strategy would be a powerful complement to the earlier work of Wells and coworkers [32] who demonstrated that subtiligase can be used to ligate synthetic peptides to the amino terminus of recombinant proteins.

The observation that protein splicing can be triggered by reconstituting inactive amino- and carboxy-terminal fragments of an intein [33–37] provides yet another application of this versatile process in protein engineering. Usually referred to as *trans*-splicing, this phenomenon allows two recombinant proteins to be joined together

in vitro. Each protein is expressed as a fusion with one of a pair of complementary intein fragments; simply combining the two fusion proteins together under appropriate conditions results in a noncovalent association of the intein fragments, activation of protein splicing and so generation of the desired recombinant protein chimera. Most *trans*-splicing studies have used artificially engineered split intein systems [33–35], although a naturally occurring split intein system has recently been observed [36] — a remarkable finding that adds an additional layer of complexity to the post-translational control of protein function. Finally, a semisynthetic *trans*-splicing system has also been reported in which the carboxy-terminal component is a synthetic polypeptide, thereby providing another route to the preparation of semisynthetic proteins [37].

With the availability of the technologies outlined above, all possible permutations for joining recombinant and synthetic polypeptides together are, in principle, amenable through chemical/enzymatic ligation. As illustrated in the following sections, this opens up many exciting opportunities for the application of synthetic chemistry to protein engineering.

Protein–ligand interactions

The ability to perform 'unnatural amino acid mutagenesis' makes the peptide ligation approach well suited for investigating protein–ligand interactions. In a recent example, the interaction between the σ^{70} protein subunit of *Escherichia coli* RNA polymerase and the T4 anti-sigma protein (AsiA) was mapped using expressed protein ligation [25]. A truncated version of the 600 amino acid σ^{70} subunit lacking residues 567–600 was generated as the $^{\alpha}$thioester protein and found to lack AsiA binding activity. A synthetic peptide, corresponding to the missing residues, was then ligated to the carboxyl terminus of the truncated protein, which restored AsiA binding activity, thus elucidating the binding region. This functionally important carboxy-terminal region of σ^{70} can now be readily modified with all manner of biophysical probes and unnatural amino acids through ligation of the appropriate synthetic peptide.

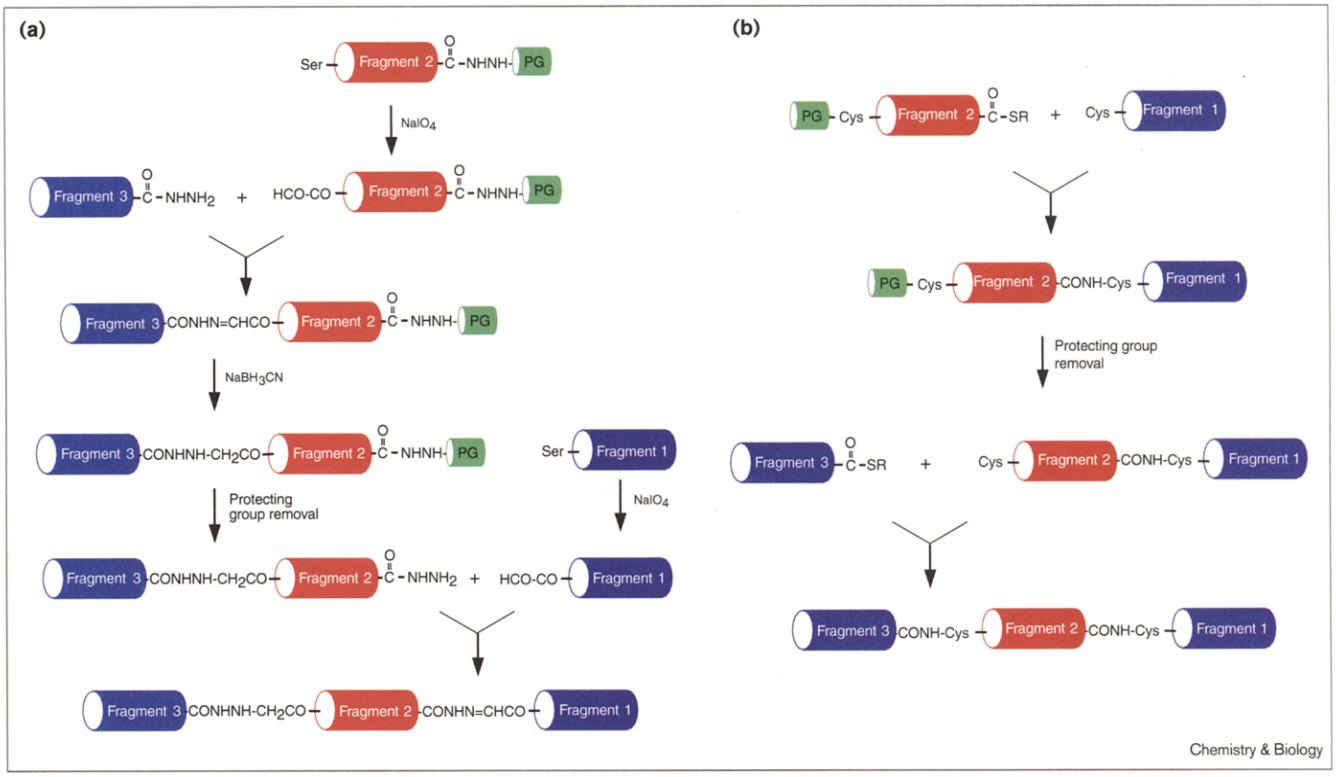
An extension of native chemical ligation has been used to study the interaction between the AP-1 transcription activator complex and its DNA recognition sequence [17,38]. An affinity cleavage reagent was prepared by chemically ligating a synthetic thioester derivative of EDTA to a series of recombinant AP-1 complexes (composed of c-Fos and c-Jun subunits) each possessing an amino-terminal cysteine residue generated by proteolytic cleavage [17]. These semisynthetic proteins have been used in an elegant series of studies to elucidate the structural determinants of orientation within an NFAT•AP-1•DNA ternary complex [17], and to study how this orientation affects transcription activation [38].

An alternative ligation chemistry has been used to specifically attach biophysical probes to recombinant proteins containing an amino-terminal serine residue, which can be converted into a glyoxyl group by mild periodate oxidation [13,39]. This aldehyde functionality can then be chemoselectively reacted with compounds containing either hydrazide, aminooxy or thiosemicarbazide groups to produce the corresponding hydrazone, oxime or thiosemicarbazole derivatives. This approach has been used to generate fluorescein and biotin derivatives of SH3 domains for studying their ligand-binding properties [40,41]. In a seminal series of studies, Proudfoot and coworkers [42] have used these peptide ligation techniques to identify potent inhibitors of HIV-1 infection based on the β-chemokine RANTES. A natural ligand for the HIV-1 coreceptor CCR5, RANTES is a competitive inhibitor of HIV infection *in vitro*, but its pro-inflammatory properties render it of little value as an antiviral agent. Using an oxime ligation approach, these researchers were able to generate an amino-terminally modified version of the protein, AOP-RANTES, which blocks viral infection in a variety of immune cells, but importantly has no pro-inflammatory effects [42]. In a follow up series of studies, native chemical ligation has been used to prepare even more potent RANTES analogs and to facilitate high-resolution structural studies on these molecules [43,44].

The groups of Offord and Rose [13–15] have pioneered the use of the hydrazone- and oxime-forming reactions for chemically ligating synthetic and recombinant peptide fragments together. In one original application, ligation through hydrazone formation was used to investigate the biological activity of a series of granulocyte colony stimulating factor (G-CSF) analogs [13,14]. Recombinant fragments of the protein were first generated through specific proteolysis at in-built Lys–Ser sequences. Amino-terminal aldehyde groups or carboxy-terminal hydrazides were then incorporated into the appropriate fragments through periodate oxidation or reverse proteolysis, respectively. Backbone engineered derivatives of the 174-residue protein were therefore obtained by ligating the fragments back together, and importantly these were shown to retain full biological activity [13]. There followed a sophisticated extension of this work in which a central fragment of G-CSF was replaced by a synthetic peptide, essentially a kind of synthetic cassette mutagenesis [14]. The appropriate reactive functionalities were introduced into the synthetic peptide insert, which then allowed it to be regioselectively reacted with the flanking recombinant pieces of the protein (Figure 2a).

Site-specific modification of peptides and proteins with fluorescent probes offers an extremely powerful way of studying biological processes. In a recent example, a fluorescent analog of the Abelson (Abl) protein tyrosine kinase was generated in which a synthetic dansylated peptide

Figure 2



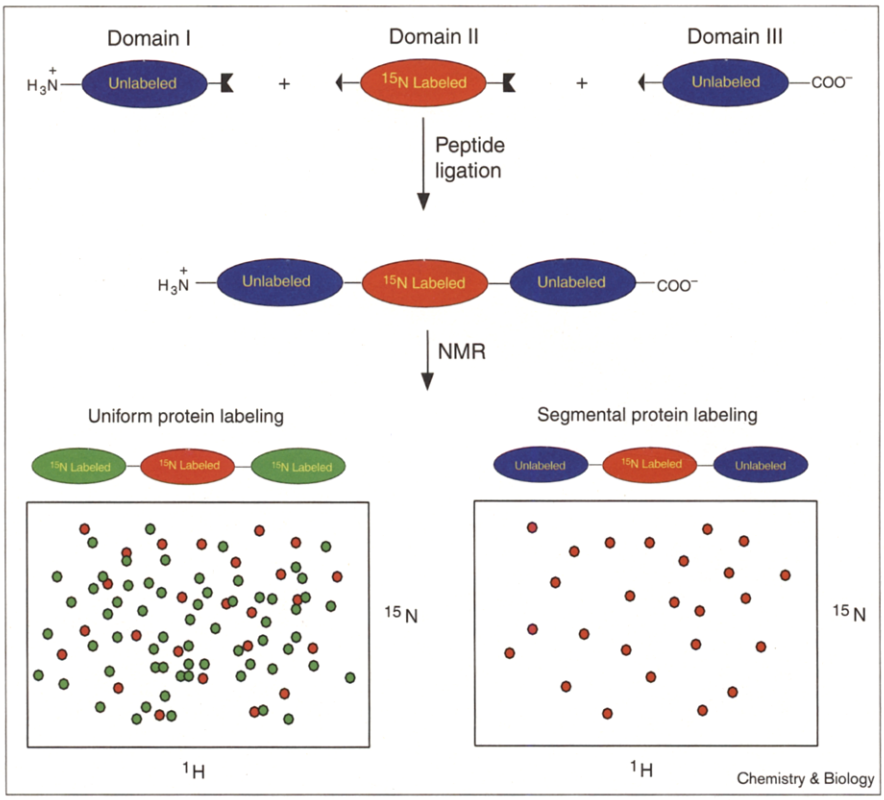
Strategies for the sequential chemical ligation of recombinant and/or synthetic protein fragments utilizing **(a)** the hydrazone bond forming reaction or **(b)** the native chemical ligation reaction. PG refers to a protecting group that can either be chemically or enzymatically removed.

was inserted between the recombinantly derived Src homology 3 (SH3) and SH2 domains of the protein [45]. To perform this synthetic cassette mutagenesis, a sequential native chemical ligation strategy was developed in which all steps were performed at physiological pH (Figure 2b). In this method, the synthetic insert contained a carboxy-terminal thioester and a cryptic amino-terminal cysteine residue masked by a factor Xa cleavable pro-sequence. This reversible cysteine protection was necessary to prevent the insert reacting with itself in either an intermolecular or intramolecular fashion [11,46]. In the first reaction, recombinant amino-terminal cysteine Abl–SH2 (generated by factor Xa treatment of the appropriate precursor) was ligated to the carboxyl terminus of the insert. The pro-sequence was then removed from the ligation product by treatment with factor Xa, to reveal the requisite amino-terminal cysteine for the second ligation reaction. This was facilitated by the addition of recombinant Abl–SH3 $^{\alpha}$thioester (produced using intein technology) to yield the desired insertion product. A series of fluorescence studies indicated that this semisynthetic protein is a specific biosensor for high-affinity bidentate interactions involving the regulatory region of Abl [45].

This significant extension to EPL renders any region of a protein accessible to synthetic manipulation, at least in principle. Moreover, as the factor Xa cleavable pro-sequence can be incorporated at the DNA level, the insert in this strategy can itself be a recombinant protein. Indeed, recombinant sequences that contain both an amino-terminal cysteine and an $^{\alpha}$thioester have been used to produce circular and polymeric proteins [47,48], something previously accomplished using synthetic peptides [11,46]. Two different approaches have been developed, both of which utilize inteins with amino-terminal cleavage activity to produce the $^{\alpha}$thioester moiety, but differ in the way the amino-terminal cysteine is produced. In one report, the amino-terminal cysteine originates from factor Xa proteolysis [47], a strategy that allowed a circular version of an SH3 domain from the protein c-Crk to be efficiently generated and studied. In the so-called TWIN (TWo INtein) approach developed by Xu and coworkers [48], the requisite amino-terminal cysteine was generated using carboxy-terminal intein cleavage, a system that allowed a number of circular and polymeric peptides and proteins to be prepared.

In an indication that peptide ligation could be a useful tool in cell biology, Tam and coworkers [49] used a thiazolidine ligation approach to attach a membrane-permeable peptide sequence (MPS) to a series of bioactive peptides. These researchers ligated a carboxy-terminal aldehyde

Figure 3



Principle of segmental isotopic labeling for structural analysis of proteins using NMR spectroscopy. In the hypothetical example shown, an internal domain of a multidomain protein is selectively labeled with ^{15}N, allowing this region to be selectively analyzed using heteronuclear NMR spectroscopy.

derivative of the MPS to an amino-terminal cysteine peptide and demonstrated that the resulting chimera was imported into eukaryotic cells [49]. The significance of this approach is that it should allow import of both peptide and nonpeptide bioactive molecules into cells.

Structural studies

Peptide ligation is beginning to make important contributions to the field of structural biology, both by providing access to interesting molecules for high-resolution studies and, as will be the primary focus here, through the development of novel strategies that actually aid the structure determination process in nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography [32,50–52].

Although developments over the last several years now allow proteins of up to ~35 kDa to be studied using NMR spectroscopy, larger systems remain inaccessible because of the loss of spectral resolution that occurs as a result of increased line widths and increased numbers of signals with similar chemical shifts. The former of these problems has to some extent been addressed with the development of new methods for spectral observation (reviewed in [53]). Peptide ligation appears to offer a solution to the latter by allowing selected segments of a protein to be isotopically labeled with NMR sensitive nuclei (Figure 3). In principle, segmental isotopic labeling of a large protein will lead to simplified NMR spectra; unlabeled regions of the protein can be filtered out using suitable heteronuclear correlation experiments leaving only signals from the labeled part of the protein. Consequently, segmental labeling should allow NMR structure analysis of discrete regions of very large proteins (Figure 3).

The feasibility of segmental isotopic labeling has recently been demonstrated using two different peptide ligation strategies—*trans*-splicing [50] and expressed protein ligation [51]. Yamazaki and coworkers used a *trans*-splicing system based on the PI-*Pfu1* intein to selectively ^{15}N label the carboxy-terminal domain of the *E. coli* RNA polymerase α subunit [50]. The carboxy-terminal domain of α (αC) was expressed as two fragments, each attached to one half of the split intein system. Reassembly of the intein module induced protein splicing, resulting in the generation of the intact αC domain. Each half of the protein was selectively labeled by simply expressing the corresponding split intein fragment in ^{15}N-enriched medium. Heteronuclear NMR experiments on these samples clearly revealed the improvement in spectral resolution that segmental labeling provides [50]. In a complementary study, expressed protein ligation was used to ^{15}N label a single domain within a Src-homology domain pair

derived from the Abl protein tyrosine kinase [51]. An ethyl $^{\alpha}$thioester derivative of recombinant Abl–SH3 was generated from the corresponding intein fusion and then chemically ligated, under physiological conditions, to ^{15}N-labeled recombinant Abl–SH2 possessing a factor Xa generated amino-terminal cysteine. Comparison between the ^{1}H{^{15}N} NMR spectra of fully labeled and segmental labeled Abl–SH(32) again illustrated the power of segmental isotopic labeling for studying large proteins using NMR spectroscopy.

Peptide ligation also provides a route to protein derivatives for use in X-ray structure determination. In an early example, Wells and coworkers [32] used subtiligase-mediated peptide ligation to attach a mercurated-cysteine-containing peptide to human growth hormone. The ability to incorporate heavy-atom-laden sequences into proteins using ligation is likely to prove useful for solving the relative phases in protein crystal structures using isomorphous replacement techniques. In a similar vein, peptide ligation also provides an alternative route to selenomethionine-containing proteins, which are now commonly used for solving the phase problem in conjunction with multiwavelength anomalous diffraction (MAD) [54]. The feasibility of this strategy was recently demonstrated by Lolis and coworkers [52] who used native chemical ligation to prepare a specific selenomethionine derivative of the chemokine MIP-II protein for subsequent crystallographic studies.

Enzymology

A number of enzyme-catalyzed reactions have been probed using peptide-ligation approaches. Fine details of enzyme mechanism can be directly studied through the incorporation of unnatural amino acids that have unique chemical properties into an enzyme active site. This approach is exemplified by a series of studies on the HIV-1 aspartyl-protease, a 99-residue polypeptide that is active as the homodimer. In a pioneering peptide-ligation study, the fully active protease was produced by chemical ligation of two unprotected synthetic fragments, each of ~50 amino acids in length [55]. Synthetic access has allowed the covalent structure of the enzyme to be manipulated in ways possible only through chemical synthesis, by modifying the sidechains or backbone of the protein [56,57], or incorporating NMR probe nuclei in the active site [58], which has allowed fundamental mechanistic questions to be asked. Native chemical ligation has been used to study intermolecular interactions between various serine proteinases and the protein inhibitors OMKTY3 [59] and eglin c [60]. Synthetic access to these protein inhibitors has allowed a systematic and quantitative analysis of the contribution of intermolecular backbone–backbone hydrogen bonding to the stability of an enzyme–inhibitor complex. In the recent study on eglin c, a series of five protein analogs were prepared, each containing a unique mainchain -NHCO- to -COO- modification and designed to delete a particular hydrogen-bonding interaction with the proteinase [60]. Using these synthetic proteins, the authors were able to show that the contribution of hydrogen-bonding to stability is highly context dependent and that there is a complex reciprocity between rigidity and adaptability in determining inhibition.

In a beautiful piece of protein engineering, the 124-residue enzyme ribonuclease A was assembled from six synthetic fragments using the subtiligase technology described earlier [30]. The catalytic histidine residues at positions 12 and 119 were then replaced by 4-fluorohistidine by modifying the corresponding synthetic peptide. The difference in pKa between this fluoro derivative and the natural histidine residue was used, among other things, to fully elucidate the general acid/general base mechanism employed by the enzyme.

Intein technology is also beginning to see application in enzymology. In one recent example, the role of the carboxy-terminal carboxyl group of T7 RNA polymerase in catalysis was investigated [61]. The protein was expressed as an amino-terminal fusion to an intein and then cleaved with either cysteine, cysteine methyl ester or 2-mercaptoethylamine to generate the corresponding carboxy-terminal cysteine, cysteine methyl ester or decarboxy-cysteine derivatives. Study of these semisynthetic analogs revealed a direct role for the carboxy-terminal carboxylate group in Mg^{2+}-dependent catalytic activity. In a second example, Begley and coworkers [62] demonstrated that recombinant protein $^{\alpha}$thiocarboxylates can be prepared from the corresponding intein fusions—a process that was developed to study the enzymology of thiamin biosynthesis, which involves a protein thiocarboxylate intermediate.

Post-translational modifications

The difficulty in producing proteins that have defined and specific patterns of post-translational modification has hindered studies on their functional role. This is an area in which peptide ligation could have a major impact, considering these alterations are central to the molecular events controlling cellular signaling networks and nature's method for conferring functional diversity onto the same translated sequence. A typical example is the regulation of the Src family of protein tyrosine kinases through tyrosine phosphorylation of their carboxy-terminal tails. This post-translational event induces an intramolecular association between the tail region and an SH2 domain, which acts to down regulate kinase activity. This tyrosine phosphorylation reaction is catalyzed by the kinase Csk, which is highly homologous to Src family members but, interestingly, lacks this regulatory tail region. Peptide ligation was used to test the hypothesis that addition of a phosphotyrosine tail to Csk would result in down-regulation of its kinase activity, in a manner analogous to Src kinases [24]. Full length Csk (450 amino

acids) and a short synthetic phosphopeptide were hooked together using EPL to give the ~53 kDa phosphoprotein. The semisynthetic Csk protein did indeed form the expected intramolecular phosphotyrosine–SH2 association but surprisingly showed an increase in the catalytic activity relative to the wild-type protein.

Systematic studies on glycoproteins have proved troublesome because of difficulties in producing homogeneous samples of defined oligosaccharide structure. Consequently, a number of synthetic methods have been developed to address this problem, including chemical and enzymatic ligation strategies (reviewed in [63]). The general approaches developed by Bertozzi and coworkers make use of the ability to incorporate unique ketone or aldehyde groups into peptides [64,65] and glycopeptides [66], respectively. Such groups can then be elaborated in a truly chemoselective fashion through oxime/hydrazone/thiosemicarbazole chemistries to give the desired neoglycopeptides [64–66]. A related strategy has been developed in which free reducing sugars are chemoselectively reacted with peptides bearing aminooxy groups [67–69]. Indeed, Mutter and coworkers [69] demonstrated that the use of *N,O*-disubstituted aminooxy groups allowed synthetic peptides to be glycosylated with high anomeric stereoselectivity.

Chemical methods are now available for the preparation of phosphorylated, glycosylated and prenylated versions of small synthetic peptides. This, coupled with the availability of the peptide-ligation technologies discussed above, suggests that it should be possible to assemble large proteins that have defined post-translational modifications. Indeed, a chemoenzymatic ligation approach has recently been used to prepare unique glycoforms of the enzyme RNAse B from constituent peptide fragments [70]. It would therefore appear that synthetic chemistry is in an excellent position to drive fundamental advances in this important area of protein biology.

Future outlook

The peptide-ligation field has evolved rapidly over the last several years. This has brought us to the point where peptide building blocks can be efficiently ligated together using any one of a number of technologies. Moreover, as we have attempted to illustrate in this review, synthetic and recombinant polypeptides can now be freely intermixed in these ligation strategies, allowing extremely large semisynthetic proteins to be manipulated using the tools of organic chemistry. There are still several technical problems, however, that need to be addressed, particularly in the native chemical ligation strategy, which is arguably the most powerful peptide-ligation approach. One important challenge for the future will be to develop 'chemical tricks' that negate the need for an amino-terminal cysteine residue in one of the peptide segments. Possible solutions to this problem might involve the use of transient auxiliary groups [71] or the use of conformationally-assisted ligations [72]. The field would also benefit greatly from the development of reliable routes to peptide $^{\alpha}$thioesters using the Fmoc–SPPS strategy — the method of choice for synthesizing phosphopeptides. Recent developments in resin-linker technology suggest this problem will be short lived [73–75].

The ability to perform sequential ligation reactions is extremely exciting because it provides synthetic access to very large proteins and, when interfaced with expressed protein ligation, allows the segmental isotopic labeling of internal regions of a protein. As pointed out earlier, these sequential ligation strategies are technically demanding to perform, and therefore the routine application of this approach will require new developments. Perhaps the most likely solution to this problem will involve performing stepwise ligations on a solid support (akin to SPPS), something that has already been formally achieved using synthetic peptides [10,11,76] and, in our own group, using a combination of synthetic and recombinant peptides (G.J.C. and T.W.M. unpublished observations).

This review began by pointing out the enormous opportunities that the proteome project presents to chemists. The field of peptide ligation is sure to play an expanding role in this endeavor by providing rapid access to novel sequences through total synthesis and by facilitating new lines of investigation in the areas ranging from structural biology and biochemistry to basic cell biology.

Acknowledgements

We thank all members of the Muir group for many useful discussions. This work was supported by the N.I.H. (GM55843, T.W.M.), the Pew Charitable Trusts (T.W.M.) and the Burroughs-Wellcome Fund (T.W.M. and G.J.C.).

References

1. Hinterding, K., Alonso-Diaz, D. & Waldmann, H. (1998). Organic synthesis and biological signal transduction. *Angew. Chem. Int. Ed. Engl.* **37**, 688-749.
2. Cornish, V.W., Mendel, D. & Schultz, P.G. (1995). Probing protein structure and function with an expanded genetic code. *Angew. Chem. Int. Ed. Engl.* **34**, 621-633.
3. Wallace, C.J.A. (1995). Peptide ligation and semisynthesis. *Curr. Opin. Biotech.* **6**, 403-410.
4. Muir, T.W. (1995). A chemical approach to the construction of multimeric protein assemblies. *Structure* **3**, 649-652.
5. Braisted, A.C., Judice, J.K. & Wells, J.A. (1997). Synthesis of proteins by subtiligase. *Methods Enzymol.* **289**, 298-313.
6. Wilken, J. & Kent, S.B.H. (1998). Chemical protein synthesis. *Curr. Opin. Biotech.* **9**, 412-426.
7. Canne, L.E., Ferre-D'Amare, A.R., Burley, S.K. & Kent, S.B.H. (1995). Total chemical synthesis of a unique transcription factor-related protein: cMyc-Max. *J. Am. Chem. Soc.* **117**, 2998-3007.
8. Mutter, M., *et al.*, & Tuchscherer, G. (1996). Template assembled synthetic proteins (TASP) as functional mimetics of proteins. *Angew. Chem. Int. Ed. Engl.* **35**, 1482-1485.
9. Muir, T.W., Dawson, P.E. & Kent, S.B.H. (1997). Protein synthesis by chemical ligation of unprotected peptides in aqueous solution. *Methods Enzymol.* **289**, 266-298.
10. Canne, L.E., Thompson, D.A., Simon, R.J. & Kent, S.B.H. (1997). Solid phase protein synthesis by chemical ligation of unprotected peptide segments in aqueous solution. In *Peptides, Frontiers in Peptide Science; Proceedings of the 15th American Peptide Symposium.* (Tam, J.P. & Kaumaya, P.T.P., eds), pp. 301-302, Nashville, USA.

11. Camarero, J.A., Cotton, G.J., Adeva, A. & Muir, T.W. (1998). Chemical ligation of unprotected peptides directly from a solid support. *J. Pept. Res.* **51**, 303-316.
12. Hennard, C. & Tam, J.P. (1997). Sequential orthogonal ligation for the synthesis of β defensin. In *Peptides, Frontiers in Peptide Science; Proceedings of the 15th American Peptide Symposium.* (Tam, J.P. & Kaumaya, P.T.P., eds) pp. 247-248, Nashville, USA.
13. Gaertner, H.F., Rose, K., Cotton, R., Timms, D., Camble, R. & Offord, R.E. (1992). Construction of protein analogues by site-specific condensation of unprotected peptides. *Bioconj. Chem.* **3**, 262-268.
14. Gaertner, H.F., Offord, R.E., Cotton, R., Timms, D., Camble, R. & Rose, K. (1994). Chemo-enzymic backbone engineering of proteins. *J. Biol. Chem.* **269**, 7224-7230.
15. Rose, K. (1994). Facile synthesis of homogeneous artificial proteins. *J. Am. Chem. Soc.* **116**, 30-33.
16. Dawson, P.E., Muir, T.W., Clark-Lewis, I. & Kent, S.B.H. (1994). Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779.
17. Erlanson, D.A., Chytil, M. & Verdine, G.L. (1996). The leucine zipper domain controls the orientation of AP-1 in the NFAT·AP-1·DNA complex. *Chem. Biol.* **3**, 981-991.
18. Paulus, H. (1998). The chemical basis of protein splicing. *Chem. Soc. Rev.* **27**, 375-386.
19. Klabunde, T., Sharma, S., Telenti, A., Jacobs Jr, W.R. & Sacchettini, J.C. (1998). Crystal structure of GyrA intein from the *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat. Struct. Biol.* **5**, 31-36.
20. Xu, M.-Q. & Perler, F.B. (1996). The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **15**, 5146-5153.
21. Chong, S., *et al.,* & Xu, M.-Q. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271-281.
22. Evans, T.C., Benner, J. & Xu, M.Q. (1999). The *in vitro* ligation of bacterially expressed proteins using an intein from *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.* **274**, 3923-3926.
23. Mathys, S., *et al.,* & Xu, M.-Q. (1999). Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* **231**, 1-13.
24. Muir, T.W., Sondhi, D. & Cole, P.A. (1998). Expressed protein ligation: A general method for protein engineering. *Proc. Natl Acad. Sci. USA* **95**, 6705-6710.
25. Severinov, K. & Muir, T.W. (1998). Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J. Biol. Chem.* **273**, 16205-16209.
26. Evans Jr., T.C., Benner, J. & Xu, M.-Q. (1998). Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Prot. Sci.* **7**, 2256-2264.
27. Porter, J.A., *et al.,* & Beachy, P.A. (1996). Hedgehog patterning activity: Role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21-34.
28. Southworth, M.W., Amaya, K., Evans, T.C., Xu, M.-Q. & Perler, F.B. (1999). Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* Gyrase A intein. *Biotechniques* **27**, 110-120.
29. Abrahamsen, L., Tom, J., Burnier, J., Butcher, K.A., Kossiakoff, A. & Wells, J.A. (1991). Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **30**, 4151-4159.
30. Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J. & Wells, J.A. (1994). A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **266**, 243-247.
31. Welker, E. & Scheraga, H.A. (1999). Use of benzyl mercaptan for direct preparation of long polypeptide benzylthio esters as substrates of subtiligase. *Biochem. Biophys. Res. Commun.* **254**, 147-151.
32. Chang, T.K., Jackson, D.Y., Burnier, J.P. & Wells, J.A. (1994). Subtiligase: A tool for semisynthesis of proteins. *Proc. Natl Acad. Sci. USA* **91**, 12544-12548.
33. Mills, K.V., Lew, B.M., Jiang, S.-Q. & Paulus, H. (1998). Protein splicing in *trans* by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proc. Natl Acad. Sci. USA* **95**, 3543-3548.
34. Wu, H., Xu, M.-Q. & Liu, X.-Q. (1998). Protein *trans*-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim. Biophys. Acta* **1387**, 422-432.
35. Southworth, M.W., Adam, E., Panne, D., Byer, R., Kautz, R. & Perler, F.B. (1998). Control of protein splicing by intein fragment reassembly. *EMBO J.* **17**, 918-926.
36. Wu, H., Hu, Z. & Liu, X.-Q. (1998). Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl Acad. Sci. USA* **95**, 9226-9231.
37. Lew, B.M., Mills, K.V. & Paulus, H. (1998). Protein splicing *in vitro* with a semisynthetic two-component minimal intein. *J. Biol. Chem.* **273**, 15887-15890.
38. Chytil, M., Peterson, B.R., Erlanson, D.A. & Verdine, G.L. (1998). The orientation of the AP-1 heterodimer on DNA strongly affects transcriptional potency. *Proc. Natl Acad. Sci. USA* **95**, 14076-14081.
39. Geoghegan, K.F. & Stroh, J.G. (1992). Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconj. Chem.* **3**, 138-146.
40. Kapoor, T.M., Andreotti, A.H. & Schreiber, S.L. (1998). Exploring the specificity pockets of two homologous SH3 domains using structure-based, split-pool synthesis and affinity-based selection. *J. Am. Chem Soc.* **120**, 23-29.
41. Chen, J.K., Lane, W.S., Brauer, A.W., Tanaka, A. & Schreiber, S.L. (1993). Biased combinatorial libraries: Novel ligands for the SH3 domain of phosphatidylinositol 3-kinase. *J. Am. Chem. Soc.* **1993**, 12591-12592.
42. Simmons, G., *et al.,* & Proudfoot, A.E.I. (1997). Potent inhibition of HIV-1 infectivity in macrophages and lymphocytes by a novel CCR5 antagonist. *Science* **276**, 276-279.
43. Mosier, D.E., *et al.,* & Wilken, J. (1999). Highly potent RANTES analogues either prevent CCR-5-using human immunodeficiency virus type 1 infection in vivo or rapidly select for CXCR4-using variants. *J. Virol.* **73**, 3544-3550.
44. Wilken, J., *et al.,* & Kent, S.B.H. (1998). Total chemical synthesis and high-resolution crystal structure of the potent anti-HIV protein AOP-RANTES. *Chem. Biol.* **6**, 43-51.
45. Cotton, G., Ayers, B., Xu, R. & Muir, T.W. (1999). Insertion of a synthetic peptide into a recombinant protein framework: A protein biosensor. *J. Am. Chem. Soc.* **121**, 1100-1101.
46. Zhang, L. & Tam, J.P. (1997). Synthesis and application of unprotected cyclic peptides as building blocks for peptide dendrimers. *J. Am. Chem. Soc.* **119**, 2363-2370.
47. Camarero, J.A. & Muir, T.W. (1999). Biosynthesis of a head to tail cyclized protein with improved biological activity. *J. Am. Chem. Soc.* **121**, 5597-5598.
48. Evans, T.C., Jr, Benner, J. & Xu, M.-Q. (1999). The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J. Biol. Chem.* **274**, 18359-18363.
49. Zhang, L., *et al.,* & Tam, J.P. (1998). Preparation of functionally active cell-permeable peptides by single-step ligation of two peptide molecules. *Proc. Natl Acad. Sci. USA* **95**, 9184-9189.
50. Yamazaki, T., *et al.* & Nakamura, H. (1998). Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* **120**, 3391-3392.
51. Xu, R., Ayers, B., Cowburn, D. & Muir, T.W. (1999). Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies. *Proc. Natl Acad. Sci. USA* **96**, 388-393.
52. Shao, W., *et al.,* & Schweitzer, B.I. (1998). Accessibility of selenomethione proteins by total chemical synthesis: structural studies of human herpesvirus-8 MIP-II. *FEBS Lett.* **441**, 77-82.
53. Dötsch, V. & Wagner, G. (1998). New approaches to structure determination by NMR spectroscopy. *Curr. Opin. Struct. Biol.* **8**, 619-623.
54. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.
55. Schnölzer, M. & Kent, S.B.H. (1992). Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. *Science* **256**, 221-225.
56. Baca, M. & Kent, S.B.H. (1993). Catalytic contribution of flap-substrate hydrogen bonds in HIV-1 protease explored by chemical synthesis. *Proc. Natl Acad. Sci. USA* **90**, 11638-11642.
57. Miller, M., Baca, M., Rao, J.K.M. & Kent, S.B.H. (1998). Probing the structural basis of the catalytic activity of HIV-1 protease through total chemical protein synthesis. *J. Mol. Struct.* **423**, 137-152.
58. Smith, R., Brereton, I.M., Chai, R.Y. & Kent, S.B.H. (1996). Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.* **3**, 946-950.
59. Lu, W., Qasim, M.A., Laskowski, M. & Kent, S.B.H. (1997). Probing intermolecular main chain hydrogen bonding in serine proteinase-protein inhibitor complexes: Chemical synthesis of backbone engineered turkey ovomucoid third domain. *Biochemistry* **36**, 673-679.

60. Lu, W., Randal, M., Kossiakoff, A. & Kent, S.B.H. (1999). Probing intermolecular backbone H-bonding in serine proteinase-protein inhibitor complexes. *Chem. Biol.* **6**, 419-427.
61. Lykke-Andersen, J. & Christiansen, J. (1998). The C-terminal carboxy group of T7 RNA polymerase ensures efficient magnesium ion-dependent catalysis. *Nucl. Acids Res.* **26**, 5631-5635.
62. Kinsland, C., Taylor, S.V., Kelleher, N.L., McLafferty, F.W. & Begley, T.P. (1998). Overexpression of recombinant proteins with a C-terminal thiocarboxylate: Implications for protein biosynthesis. *Prot. Sci.* **7**, 1839-1842.
63. Lemieux, G.A. & Bertozzi, C.R. (1998). Chemoselective ligation reactions with proteins, oligosaccharides and cells. *Trends Biotech.* **16**, 506-513.
64. Rodriguez, E.C., Marcaurelle, L.A. & Bertozzi, C.R. (1998). Aminooxy-, hydrazide-, and thiosemicarbazide-functionalized saccharides: Versatile reagents for glycoconjugate synthesis. *J. Org. Chem.* **63**, 7134-7135.
65. Marcaurelle, L.A., Rodriguez, E.C. & Bertozzi, C.R. (1998). Synthesis of an oxime-linked neoglycopeptide with glycosylation-dependent activity similar to its native counterpart. *Tetrahedron Lett.* **39**, 8417-8420.
66. Rodriguez, E.C., Winans, K.A., King, D.S. & Bertozzi, C.R. (1997). A strategy for the chemoselective synthesis of *O*-linked glycopeptides with native sugar-peptide linkages. *J. Am. Chem. Soc.* **119**, 9905-9906.
67. Zhao, Y., Kent, S.B.H. & Chait, B.T. (1997). Rapid, sensitive structure analysis of oligosaccharides. *Proc. Natl Acad. Sci. USA* **94**, 1629-1633.
68. Cervigni, S., Dumy, P. & Mutter, M. (1996). Synthesis of glycopeptides and lipopeptides by chemoselective ligation. *Angew. Chem. Int. Ed. Engl.* **35**, 1230-1232.
69. Peri, F., Dumy, P. & Mutter, M. (1998). Chemo- and stereoselective glycosylation of hydroxylamino derivatives: a versatile approach to glycoconjugates. *Tetrahedron* **54**, 12269-12278.
70. Witte, K., Sears, P., Martin, R. & Wong, C.H. (1997). Enzymatic glycoprotein synthesis: Preparation of ribonuclease glycoforms via enzymatic glycopeptide condensation and glycosylation. *J. Am. Chem. Soc.* **119**, 2114-2118.
71. Canne, L., Bark, S. & Kent, S.B.H. (1996). Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* **118**, 5891-5896.
72. Beligere, G.S. & Dawson, P.E. (1999). Conformationally assisted protein ligation using C-terminal thioester peptides. *J. Am. Chem. Soc.* **121**, 6332-6333.
73. Li, X., Kawakami, T. & Aimoto, S. (1998). Direct preparation of peptide thioesters using an Fmoc solid-phase method. *Tetrahedron Lett.* **39**, 8669-8672.
74. Jensen, K.J., Alsina, J., Songster, M.F., Vagner, J., Albericio, F. & Barany, G. (1998). Backbone amide linker (BAL) strategy for solid-phase peptide synthesis of C-terminal-modified and cyclic peptides. *J. Am. Chem. Soc.* **120**, 5441-5452.
75. Backes, B.J., Virgilio, A.A. & Ellman, J.A. (1996). Activation method to prepare a highly reactive acylsulfonamide "safety-catch" linker for solid-phase synthesis. *J. Am. Chem. Soc.* **118**, 3055-3056.
76. Canne, L., Kent, S.B.H. & Simon, R. (1998). Solid phase native chemical ligation of unprotected or N-terminal cysteine protected peptides in aqueous solution. International Patent Number WO 98/56807.