TUTORIAL REVIEW

CSR
www.rsc.org/csr

The synthesis of peptides and proteins containing non-natural amino acids

David R. W. Hodgson* and John M. Sanderson*
Department of Chemistry, South Road, Durham, UK DH1 3LE.
E-mail: j.m.sanderson@durham.ac.uk; d.r.w.hodgson@durham.ac.uk;
Fax: +44 (0)191 384 4737; Tel: +44 (0)191 334 2100

Received 16th February 2004
First published as an Advance Article on the web 13th August 2004

Methods for the incorporation of non-natural amino acids into proteins have advanced significantly over recent years and in this *tutorial review* we aim to give a general overview of the area. These techniques offer the possibility of modulating the structures and functions of proteins and thus permit the generation of novel designed systems for both biocatalytic and mechanistic studies. Four complementary approaches are discussed in detail along with examples of their application. The advantages and disadvantages of each technique are also discussed.

Introduction

The wish to generate designer proteins that rival the properties of their natural counterparts has been a long sought after goal. The complexities involved in the *de novo* design of proteins are overwhelming and current computational methods do not permit any great degree of freehand protein design. Therefore many researchers apply protein engineering techniques to existing protein scaffolds in order to derive new systems that retain the desirable traits of the original system in addition to the newly designed properties. However standard protein engineering techniques limit the user to the standard twenty commonly encountered amino acids. Thus the possibilities in terms of producing modified proteins with new properties or functions using standard engineering methods are severely limited. To overcome this obstacle, several strategies for the incorporation of non-natural amino acids into naturally occurring protein and peptide molecules have been developed. This review will focus on four strategies that have been widely exploited to date, and are still undergoing continual advancement and optimisation.

Bioconjugation

Bioconjugation is the simplest and longest standing method for the introduction of non-natural amino acids into proteins. The approach relies on the existing functionality of a protein, and through the use of appropriate chemical reagents, chosen amino acid side chains can be modified. All side chains displaying reactive functionality can be modified, and, depending on the functionality in question and the choice of modifying reagent, highly selective, quantitative modifications can take place. Clearly quantitative, site-selective bioconjugations are the most desirable outcome as these lead to homogeneous populations of modified products that are suitable for quantitative experimental studies and analysis. This section of the review will discuss some of the most widely used bioconjugation systems in order to give the reader a flavour of the types of reagents in use and also the end uses of the remolded protein products. In addition the advantages and disadvantages of the approach will be discussed. Several other authors have reviewed the area comprehensively and the reader is referred to these articles and references therein for a fuller picture of the subject.[1,2] The reader is also referred to an excellent textbook that gives detailed experimental protocols for a wide variety of bioconjugation experiments.[3]

Cysteine modification

The most widely used bioconjugation strategy exploits the latent nucleophilicity of the thiol side chain of cysteine.

Dr David Hodgson graduated from the University of Cambridge in 1995 with a B.A. in Natural Sciences. He then undertook a PhD on small molecule enzyme model systems, again at Cambridge, with Professor Tony Kirby. In 1999 he moved to the University of Toronto to work on the chemical modification of subtilisins with Professor J. Bryan Jones. The summer of 2000 saw a move to SUNY at Buffalo to work on ribozyme selection with Hiroaki Suga with support from a Royal Society-Fulbright post-doctoral fellowship. In January 2003 he took up a lectureship at the University of Durham.

David Hodgson

Dr John Sanderson obtained his PhD from the University of Leeds in 1996. He worked as a postdoctoral researcher at the University of Sheffield for one year, and then in the group of Professor J.-M. Lehn at the Collège de France, Paris, until September 1999. Upon returning to the UK, he took up a lectureship at the University of Sussex. Since 2001 he has been a lecturer in the Department of Chemistry at the University of Durham, where his research interests include the design and synthesis of peptides and the study of peptide–lipid interactions.

John Sanderson

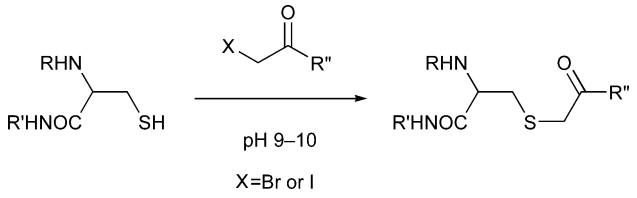
DOI: 10.1039/b312953p

 This journal is © The Royal Society of Chemistry 2004

Through the use of appropriate electrophiles thiol groups can be rapidly, selectively and quantitatively modified. Cysteine modifications fall into two major categories: alkylation and mixed disulfide formation. Within these categories further subdivisions can be made according to the choice of modifying reagent employed.

Cysteine alkylation

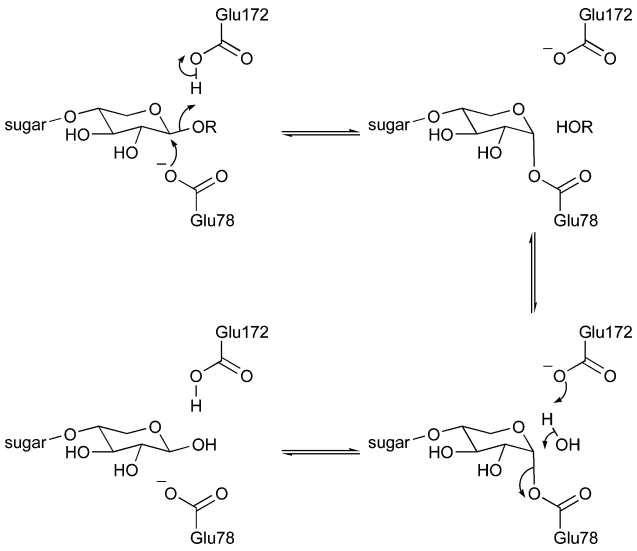
Probably the most widely used approach is alkylation using α-halo carbonyl compounds in order to generate thioether products (Scheme 1). Under mildly basic (pH 9–10) conditions



Scheme 1 Thiol alkylation using α-halo carbonyl compounds.

the acidic thiol group can be partially deprotonated and reveal the more nucleophilic thiolate anion. The soft character of this nucleophile complements the electrophilic characteristics of α-halo carbonyl compounds and thioether formation occurs rapidly between these components. Furthermore α-halo carbonyl compounds show little reactivity towards other proteinogenic functional groups; thus cross-reactivity is limited, and the products of the modification process tend to be highly homogeneous.

An illustrative example of this approach was used by Withers in order to gain insight into the mechanisms of glycosyl hydrolases.[4] Glycosyl hydrolases usually contain two catalytic active site carboxylate residues, and the relative positioning of these two carboxyl groups is known to be a critical parameter in determining whether the glycosyl hydrolase in question is either an inverting or retaining hydrolase. One carboxylate is deprotonated and acts as a nucleophile in order to trap the forming oxocarbenium ion, whereas the other is protonated and acts as a general acid catalyst in order to facilitate the departure of the alkoxy leaving group of the aglycon (Scheme 2).



Scheme 2 Mechanism of a retaining glycosyl hydrolase.

Withers' team chose to study the effects of perturbing the carboxylate–carboxylate distance in the xylanase of *B. circulans*. The active site of the wild-type (WT) xylanase contains Glu78 and Glu172, which lie ~5.5 Å apart and function as the active site nucleophile and the general acid catalyst respectively. In previous work a Glu78Asp mutant, where the carboxylate–carboxylate distance is larger than in the wild-type enzyme, was studied, but was found to have little catalytic activity. Therefore a new study centred on *reducing* the carboxylate–carboxylate distance by introducing a lengthened carboxylic acid analogue of Glu78 was undertaken. Unfortunately, unlike the case for the chain-shortened Glu78Asp mutant, the standard twenty amino acid repertoire does not contain a lengthened-chain analogue of glutamic acid; therefore a combined site-directed mutagenesis (SDM)–chemical modification strategy was employed in order to overcome this limitation. First a cysteine mutant Glu78Cys of the xylanase was produced through SDM. The use of SDM and *in vivo* expression systems allows relatively large quantities of the mutant protein to be produced in a conventional manner. However, *in vivo* expression systems limit the user to the twenty naturally occurring amino acids, therefore if the wild-type protein already contains a cysteine residue multiple labelling can occur. Fortunately, wild-type *B. circulans* xylanase does not contain any cysteine residues, thus the Glu78Cys mutation introduces a unique thiol functional group into the xylanase which can then be modified in a highly selective manner. Iodoacetic acid was chosen as the electrophilic modifying reagent with the resultant protein displaying an added methylene-carboxy group extending from the added cysteine residue (Fig. 1).

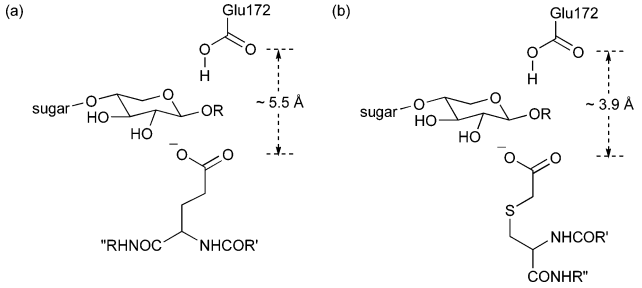


Fig. 1 (a) Wild-type and (b) modified xylanase active sites.

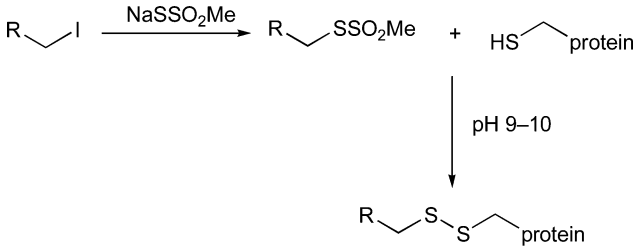
The extent of thiol modification was then confirmed using Ellman's Reagent[5] and electrospray mass spectrometry (ESMS) was used in order to confirm that the expected modification had indeed taken place. Together this information was used to conclude that the modified protein was homogenous and suitable for detailed kinetic analyses. The Glu78Cys–CH_2–CO_2^- mutant protein retained a significant level of activity with k_{cat}/K_M values only 16–100-fold lower than the WT enzyme. The authors concluded that the modified active site nucleophile was still able to participate in nucleophilic catalysis whilst only causing moderate perturbations in the active site conformation owing to its increased size compared to the WT Glu residue.

The example cited above illustrates the use of an α-halo compound in the production of a modified enzyme for mechanistic studies on its existing glycosyl transferase activity. However bioconjugated proteins have also been employed in order to ascertain structural information upon the active sites of proteins using a variety of thiol-reactive agents including fluorophores and the structural probe Fe-BABE.[6]

Disulfide formation

Disulfides offer two particular advantages over and above other thiol-selective bioconjugates: first, disulfide formation is readily reversible through the use of standard reducing agents such as dithiothreitol (DTT) or β-mercaptoethanol.

Second, the linker length between the thiol that is modified and the added moiety is limited to the bond lengths of the S–S–C portion of the bioconjugate. This is a particular advantage as the space requirements of the disulfide within the confines of a protein active site are modest, plus the disulfide does not bristle with functionality that is prone to hydrogen bonding or other interactions that may distort the active site or override the intended effects of the bioconjugate. Several types of chemical reagent have been employed for the production of protein–S–S–label bioconjugates with the thiopyridyls and methanethiosulfonates (MTS) being the most widely utilised. Probably the most simple to use are the methanethiosulfonates which can be synthesised readily from organo-bromide and -iodide compounds through nucleophilic displacement with sodium methanethiosulfonate (Scheme 3).



Scheme 3 Synthesis and use of methanethiosulfonate reagents.

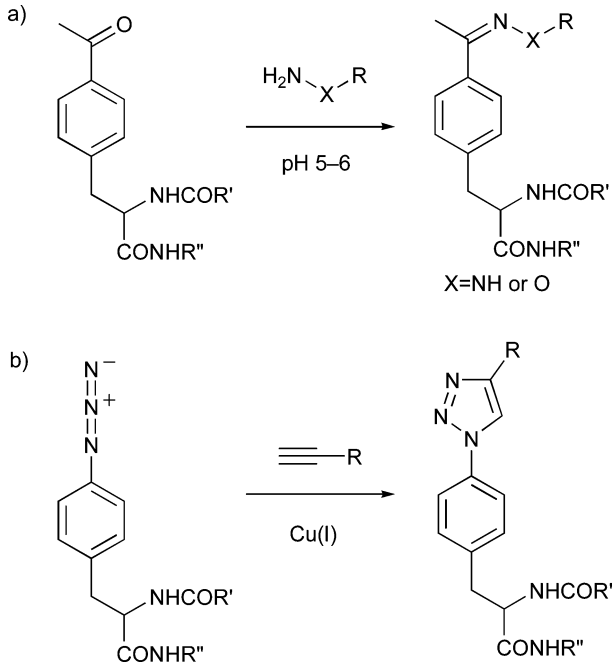
The straightforward synthesis of MTS reagents from organo-halides, which are themselves readily accessible or available commercially, allows the user to generate diverse arrays of MTS reagents. These in turn can be used in order to generate large numbers of modified proteins where the modifications can incorporate many different structural and functional features. This approach has been elegantly exploited in Jones' systematic engineering of subtilisin mutants, which has also been reviewed comprehensively.[7]

In one example chemically modified subtilisin mutants were generated and screened for their ability to function as catalysts for peptide ligation—an example of using an engineered protein in order to provide a tool to facilitate the generation of other mutant peptides.[8] The approach relied on generating a combinatorial array of mutant proteins using a range of different MTS reagents. The array was then screened for amidase and esterase activities using a microtiter plate-based assay. The chemical modification approach is most apt for this parallel screening approach as only one batch of cysteine-containing protein needs to be produced. This protein can then be modified in parallel using different MTS reagents that are available within the laboratory or commercially using a bioconjugation process that can be completed within one or two days.

A further example of the use of MTS reagents also serves to illustrate the diversity of roles that proteins containing non-natural amino acids can fulfil. Under the banner of "Catalytic Antagonists" a series of subtilisin mutants bearing a range of small-molecule protein recognition elements including enzyme inhibitors, protein affinity labels and antigens were prepared.[9] These newly engineered proteases were then assessed for their ability to selectively recognise a specific protein target within a mixture of proteins and preferentially subject this target protein to proteolytic degradation.

Future prospects

Although the bioconjugation approach towards the introduction of non-natural amino acids into peptides and proteins is well established, recent developments in the use of suppressor tRNA techniques offer great opportunities for further expansion. As discussed in a later part of this review, the use of suppressor tRNA methods by Schultz and co-workers has shown that it is possible to use the ribosomal protein synthesis machinery in order to incorporate a non-natural amino acid into a protein using both *in vitro* and *in vivo* expression systems. However, rather than just introducing the desired end-product amino acid, Schultz has instead incorporated a unique amino acid side chain that possesses an "orthogonal" chemical functional group that can be bioconjugated without interfering with the chemistries of the natural proteinogenic amino acid side chains. Two examples are keto-carbonyl containing amino acids and amino acids that contain an azido functional group. These non-standard amino acids can be employed in order to form hydrazone and oxime derivatives in the case of the keto amino acid[10] or triazole derivatives in the case of the azido functionalised amino acid, through the use of copper(I) catalysed [3+2] cycloadditions to terminal alkynes[11] (Scheme 4).



Scheme 4 Bioconjugations to (a) ketone and (b) azide containing non-natural amino acid side chains.

This combination of methods has led to the facile generation of glycoprotein mimetics which mirror the type of glycosyl modifications that frequently occur post-translationally *in vivo* and are often difficult to reproduce in convenient bacterial expression systems. Furthermore this system capitalises on the advantages of both bioconjugation and suppression methods whilst minimising the limitations of each. The Staudinger ligation, which again uses azide-based chemistry, has also received recent attention both in simple bioconjugation applications[12] and in concert with other tRNA engineering strategies (see tRNA Engineering Methods section).

Non-ribosomal peptide synthesis

In eukaryotes, the synthesis of cellular peptides and proteins is conducted by the ribosomal machinery, in which the genetic information encoded in mRNA is translated into specific peptide sequences by a protein/RNA complex that moves progressively along each mRNA strand. In prokaryotes, a number of biologically important peptides are synthesised by a different mechanism, in which the entire sequence of a peptide is dictated

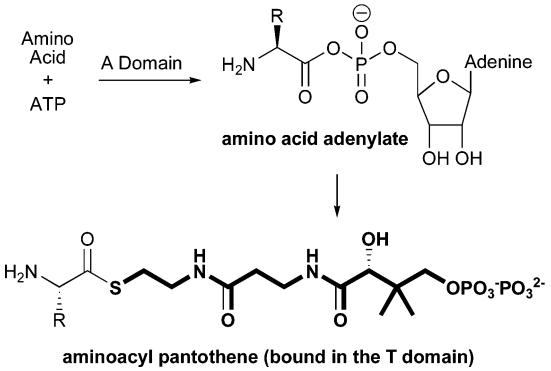
by the structure of the enzyme that synthesizes it. This process, non-ribosomal peptide synthesis (NRPS),[13,14] has much in common with polyketide synthesis[15] and is directed by large modular enzymes termed non-ribosomal peptide synthases. Some synthases are multimeric complexes, while others are single, massive proteins. Each module is composed of about 1000–1200 amino acids, giving the complete enzymes a mass in the region of 2 MDa. NRPS is responsible for the synthesis of a diverse set of peptides, including several antimicrobial agents, and peptide fragments that are subsequently incorporated into other secondary metabolites, such as the tripeptide δ-(L-α-aminoadipoyl)-L-cysteinyl-D-valine, which is a precursor for the synthesis of the penicillins and cephalosporins.

Peptide diversity

A unique feature of NRPS is the diversity of the peptides that can be synthesised. Whereas ribosomal peptide synthesis produces linear peptides composed of all L-amino acids containing only minor post-translational modifications, NRPS produces an array of chemically diverse peptides that frequently contain highly esoteric modifications, including cyclic structures, high D-amino acid contents, *N*-methylation, unusual heteroaromatic groups, and an array of N- and C-terminal modifications. Some examples of these peptides are presented in Fig. 2. The synthesis of each of these peptides is catalysed by a unique non-ribosomal peptide synthase. Each module of the synthase contains the requisite components for the activation and reaction of a single amino acid. Diversity is achieved by the recruitment of different components in each module according to the amino acid to be incorporated.

Module structure

In recent years, many of the structures of the components of NRPS modules have been solved to high resolution, enabling the mechanisms of the reactions they catalyse to be elucidated. The basic module consists of an activation (A) domain, a thiolation domain (T domain, sometimes called a peptidyl carrier protein, PCP) and a condensation (C) domain. In the first step of the elongation cycle, catalysed by the A domain, an amino acid adenylate is formed from an amino acid and ATP (Scheme 5).



Scheme 5 Activation of amino acids by the A and T domains of non-ribosomal peptide synthases. The pantothene group is shown in bold

The activated acyl group is transferred from the A domain to the thiol of a pantothene moiety covalently bound to the T domain, forming a thioester as the active acylating agent. This formally corresponds to the activation of acyl groups as thioesters of Coenzyme A in polyketide synthesis, and utilises the capability of thioesters to act as relatively active acylating agents whilst retaining a degree of stability against hydrolysis. The aminoacyl pantothenate of the T domain then undergoes nucleophilic attack by the α-amino group of the amino acid bound to the T domain of the next module of the synthase. In this process, a peptide bond is formed, and the growing peptide chain is transferred to the T domain of the next module, a process that is mediated by the C domain of the next module.[16] This process is shown schematically in Scheme 6. There are some notable exceptions to this cycle: in the first module, there is no C domain (as there is no amino acyl group to transfer); in the final step, the thioester is hydrolysed to release the free peptide by a thioesterase (TE) domain.[17,18] Some TE domains are additionally able to direct aminolysis of the thioester by the N-terminal amino group to form a cyclic peptide. In some non-ribosomal peptide synthases, such as those of the gramicidins and tyrocidins, certain modules contain epimerisation domains that catalyse the epimerisation of amino acids bound to the preceding T domain, leading to the incorporation of a D-amino acid into the nascent peptide. Other domains exist that are capable of performing *N*-methylation and heterocycle

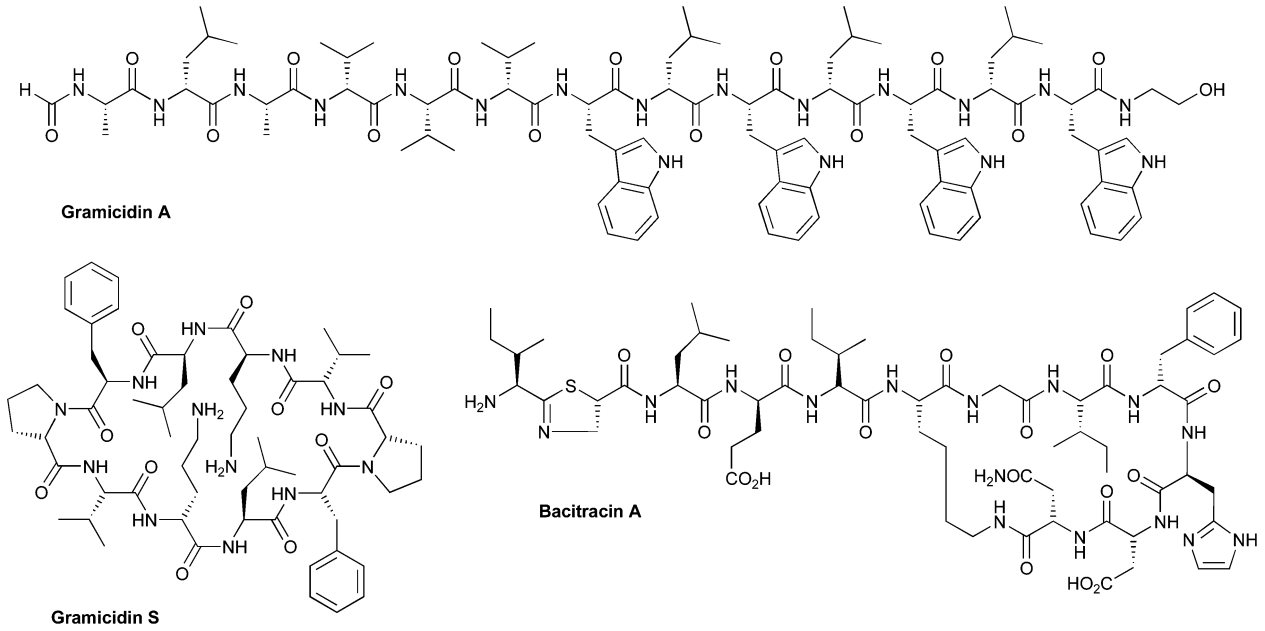
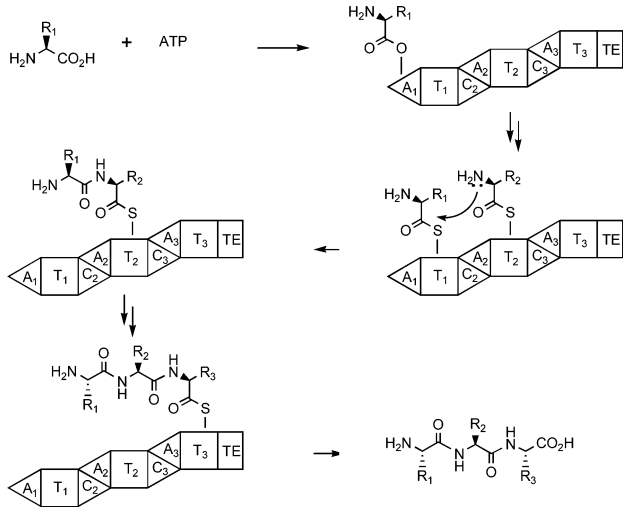


Fig. 2 Examples of peptides produced by non-ribosomal peptide synthesis.



Scheme 6 Schematic representation of the synthesis of a tripeptide by non-ribosomal peptide synthesis. Each module is composed of a C, A and T domain, with the exception of the first module, which has no C domain.

formation. A complete peptide is synthesised by sequential progression of a growing peptide chain from one module to the next before release by the TE domain.

The sequence of the peptide that is synthesised is determined by the A and C domains. Each A domain is selective for the amino acid that it activates, which is the principal determinant of the amino acid subsequently incorporated. Some A domains exhibit only modest selectivity however, with the result that families of related peptides are produced (*cf.* the gramicidins). The C domains also display a level of selectivity in the amino acids that they are able to transfer to the A domain, and provide a proof-reading mechanism for the condensation process, preventing elongation if the A domain is activated with the wrong amino acid.[19]

Future prospects

A number of aspects of these enzyme systems make them appealing for applications in biotechnology:

i. They catalyse peptide bond formation in water.

ii. They use unprotected amino acids as substrates.

iii. With the use of epimerisation domains, they are capable of selective incorporation of D-amino acids into a growing chain.

iv. They are able to direct the synthesis of cyclic peptides and other esoteric peptides.

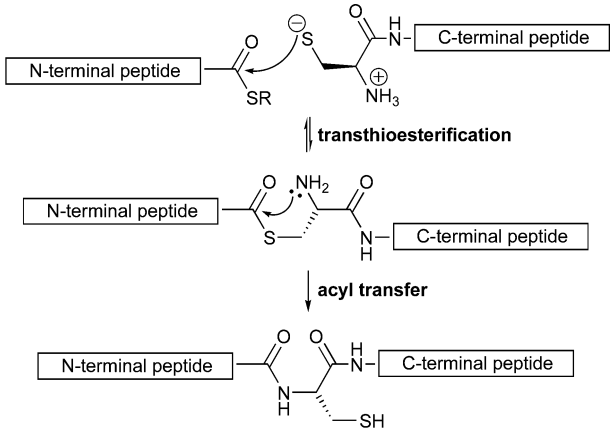
Current *in vitro* methods of peptide synthesis are becoming increasingly sophisticated, but as these methods become more advanced and complex peptide structures become accessible, the cost of their synthesis increases accordingly. Part of this increased cost lies in the requirement for orthogonally protected amino acids, and the extensive use of protecting groups produces processes with very low atom efficiency. Non-ribosomal peptide synthases offer a potentially rewarding route to diverse peptide structures that is highly atom efficient and environmentally benign. In order to be of practical 'off the shelf' use however, there are a number of features of these modular systems that need to be addressed. The greatest concern is the effort that will be required to generate a synthase specific for the peptide(s) that one desires to prepare. There are however, some encouraging precedents for engineering these systems in the literature: some A domains have been 'forced' into activating the wrong amino acid *in vitro* by raising the concentration of the 'wrong' amino acid;[20] two separated modules are capable of functioning independently (*i.e.* they can be loaded with amino acid), and form dipeptides when mixed together;[16] A domains retain their activity independent of the rest of the protein (*i.e.* when expressed without the other domains), and maintain their selectivity in such cases.[21] Although some synthases have been engineered to produce peptides with a different sequence to the parent synthase, engineering has not yet been routinely applied to the synthesis of peptides containing highly unusual or non-natural chemical functionality, although TE domains have been used to mediate peptide cyclisation and release to produce peptides containing non-natural functionality.[22,23] The most promising example of this has been the release of peptides from solid supports using TE domains, opening the door to mixed chemical/enzymatic approaches to novel peptide libraries.[24]

Peptide ligation

An attractive methodology for the preparation of novel proteins involves the manufacture of bite-sized peptides by any of an array of suitable methodologies, such as those described above, followed by their assembly into a complete macromolecule through the intervention of selective covalent bond-forming reactions. A number of selective reagents suitable for the covalent modification of proteins have been known for many years, but general methodologies for the ligation of peptide fragments through the formation of natural backbone amides have only recently become widely available. These methods have enormous potential for the preparation of macromolecules containing novel peptide or protein domains, and have proven particularly useful for the synthesis of proteins that are traditionally difficult to obtain by *in vivo* molecular biology-based approaches, such as membrane proteins.

Thioester-based methods

The most widely applied ligation method is the reaction of an N-terminal cysteine of a peptide or protein with a C-terminal thioester of another peptide (Scheme 7).[25,26] Following a

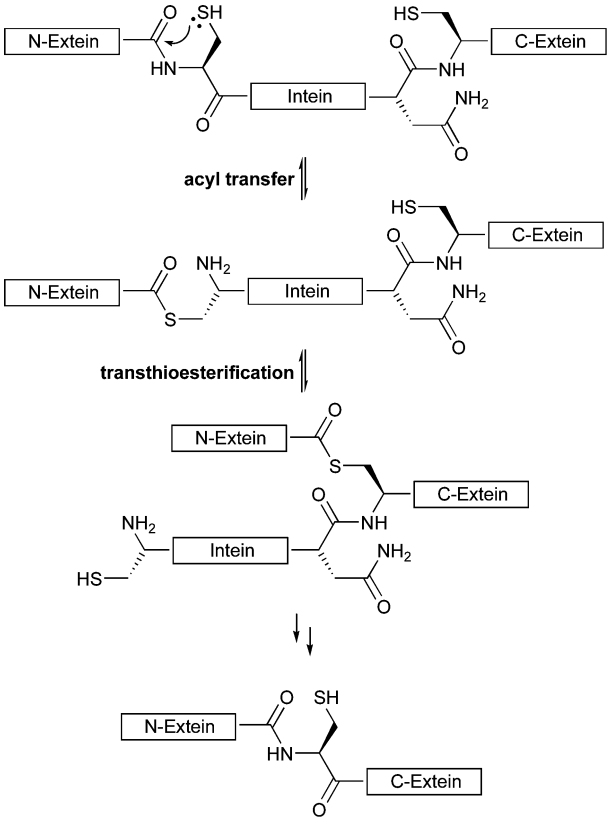


Scheme 7 Thioester-mediated (native) peptide-ligation.

reversible intramolecular transthioesterification to form a thioester bridge between the two peptides, intramolecular attack of the α-amino group of the C-terminal peptide furnishes the desired peptide bond. This approach, often termed 'native chemical ligation', requires that one of the peptides to be fused has a cysteine at the N-terminus, and produces a ligated product containing at least one cysteine. In favourable circumstances, *i.e.* where no functionally important cysteines are present elsewhere in the protein, the ligated peptide can be desulfurised using Pd/Al_2O_3 to form an alanine residue.[27] N-terminal cysteines are readily prepared by a

number of conventional *in vitro* and *in vivo* methods. The major challenge with this approach is the synthesis of a peptide or protein with a C-terminal thioester. Two excellent methods for achieving this are available.

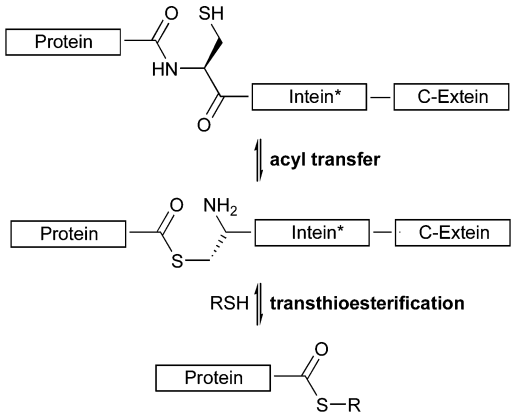
Inteins. Some proteins undergo a process termed *self-splicing*, in which two protein domains termed *exteins* are ligated with the concomitant excision of the protein fragment (*intein*) between them.[28,29] The splicing reaction is catalysed by the intein itself, and is in many respects the reverse of protein ligation. The process commences with reversible N to S transfer of an acyl group that lies at the C-terminal end of the N-extein, followed by reversible transthioesterification to form an intermediate in which the two exteins are linked *via* a thioester (Scheme 8). In the final steps of the splicing process, the intein



Scheme 8 Protein self-splicing mediated by inteins.

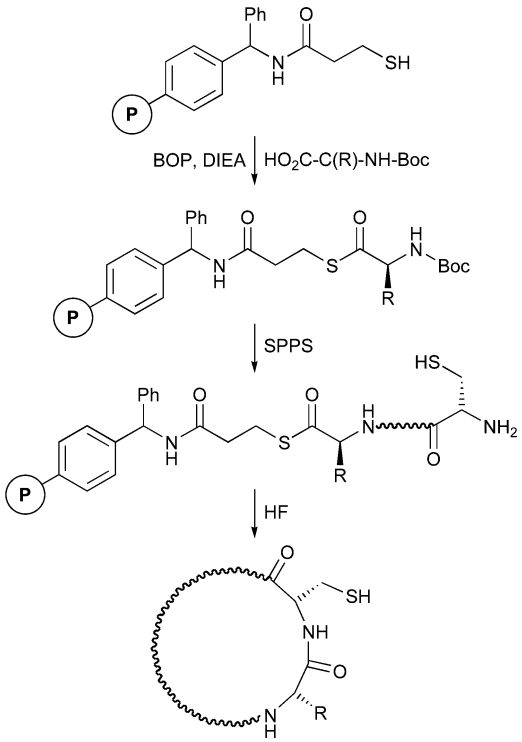
is excised *via* aspartimide formation at its C-terminus, and the *S*-acyl group transferred to the liberated α-amino group. Inteins from a variety of sources exhibit significant sequence homology, and are capable of splicing together whichever proteins are found as C- and N-terminal exteins. This reaction has been exploited to enable a facile preparation of C-terminal thioesters (Scheme 9). Using molecular biology approaches, a recombinant protein is prepared consisting of the peptide or protein of interest fused at the C-terminus to a modified intein domain. Through modification of the C-extein, the splicing reaction is prevented from proceeding past the initial acyl transfer reaction. The resulting thioester then becomes susceptible to intermolecular transthioesterification by an added thiol to release the intein and form an isolable thioester. Potentially, any thioester may be produced by this method, although generally the range of thioesters prepared is restricted to simple alkyl esters or more reactive phenyl thioesters.

Solid-phase methods. As solid-phase peptide synthesis usually proceeds in a C→N direction, with the C-terminus of



Scheme 9 Preparation of protein thioesters.

the peptide attached to the resin, there are two potential strategies for the preparation of C-terminal thioesters. The first and simplest is to attach the peptide to the solid support *via* a thioester linkage that remains following cleavage from the resin. Due to the extensive use of amines Fmoc-SPPS, Boc chemistry is generally favoured for the preparation of thioesters. This has been put to good use in the preparation of cyclic peptides by thioester ligation (Scheme 10).[30] Synthetic



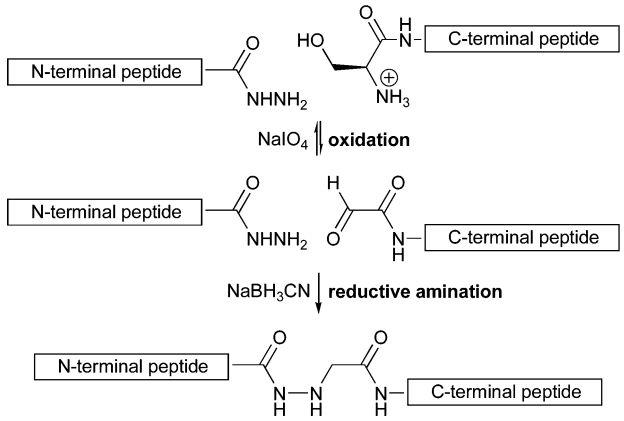
Scheme 10 Synthesis of cyclic peptides using native chemical ligation.[30]

peptides are also useful building blocks for semisynthetic approaches to the preparation of membrane proteins, in which the soluble parts of the protein (intracellular and extracellular domains) are prepared by molecular biology approaches, and the less soluble parts (transmembrane domains) prepared using solid-phase methods. Controlled thioester ligation of the peptide and protein fragments subsequently yields the complete protein.[31] The second approach involves orthogonal protection of the C-terminal carboxylate and immobilisation of the peptide to the solid support through the side chain of the C-terminal amino acid. Unmasking of the C-terminal carboxylate at a late stage in the synthesis then allows formation of the

thioester prior to deprotection from the solid support. This approach is restricted to peptides with suitable residues at the C-terminus for immobilisation, such as lysine, aspartate or glutamate.

Other methods

A variety of other ligation methods have been described for peptide ligation, including examples of both enzymatic[32] and selective chemical[33] approaches. In the former approach, unprotected peptides have been esterified with short aliphatic diols using the serine protease subtilisin[34] and subsequently ligated by protease-mediated aminolysis.[8,35] This approach is limited however, by the substrate specificity of subtilisin. The latter approach requires an N-terminal serine. Oxidation of the serine with sodium periodate furnishes a glyoxyl group, to which another peptide with a C-terminal hydrazide can be ligated by reductive amination (Scheme 11).



Scheme 11 Hydrazide-mediated chemical ligation of peptides.

Mixed approaches

The applicability of combined approaches using ligation and bioconjugation has been demonstrated to great effect in the production of a synthetic erythropoiesis protein–polymer system.[36] Using both oxime formation and native chemical ligation, yields in excess of 100 mg of a 51 kD construct were prepared in high purity. Peptide fragments for ligation were prepared by solid phase methods, permitting ready incorporation of non-natural amino acids into the protein.

tRNA engineering methods

Over the course of the last twenty-five years several groups have worked towards harnessing the power of Nature's coded protein synthesis machinery in order to produce proteins that contain non-natural amino acids. During coded protein synthesis, RNA polymerases transcribe genomic DNA in order to provide coded messages for protein synthesis in the form of mRNAs. Each mRNA message is made up of series of "codons" that code for a single amino acid in the series of amino acid residues that make up the desired protein. Each codon is made up of three bases that are then recognised by a corresponding transfer RNA molecule (tRNA) through Watson–Crick base-pairing to three complementary bases within each tRNA molecule known as the "anti-codon". Each different tRNA molecule, however, is "charged" with, that is chemically bonded to, a specific amino acid (Fig. 3) that corresponds directly to the specific anticodon present in that tRNA molecule (Fig. 4).

Thus by attaching a non-natural amino acid to a tRNA molecule it is possible to insert that non-natural amino acid in

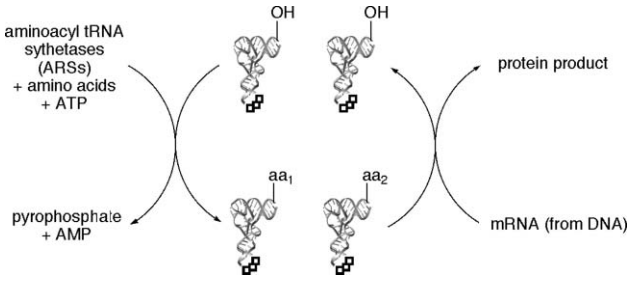


Fig. 3 Basic pathway of translation and the role of tRNA and ARSs.

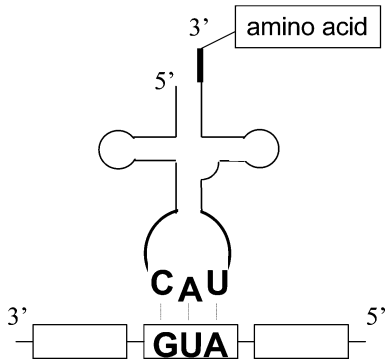


Fig. 4 The codon reading event of coded protein synthesis.

place of the originally intended amino acid. However, in so doing a naturally occurring amino acid would be replaced by a non-natural amino acid in all instances, and the number of possible amino acids within a protein would still be limited to twenty. In addition to the codons that correspond to each of the natural amino acids, the genetic code also contains three codons that are used to send "stop" signals to the protein synthesis machinery and thus terminate translation when required. Clearly only one "stop" codon is actually required whilst two could be left for other uses.

The development of "suppressor" tRNAs has allowed this to occur. By using a tRNA molecule that possesses an anti-codon that is complementary to one of the stop codon signals it is possible to "suppress" the stop signal and in addition, insert an amino acid that is attached to the suppressor tRNA in place of the stop signal. There have been two main obstacles on this pathway: first, the chemical or enzymatic generation of tRNA molecules that are charged with non-natural amino acids can be challenging. Second, the existing tRNA aminoacylation system must not charge the suppressor tRNA with any other amino acid—the suppressor tRNA must be "orthogonal" to the existing aminoacylation system. This is critical in order to maintain the fidelity in terms of the amino acid composition of translated proteins.

Chemical aminoacylation of tRNAs

Pioneering work in this field was undertaken by Hecht and co-workers, who through the use of extensive chemical synthesis combined with an enzymatic ligation process were able to generate tRNA molecules bearing non-natural amino acids.[37] The process involved the preparation of an aminoacylated dinucleotide pCpA-aa that mirrors the 3′-terminus of a tRNA molecule. This aminoacylated oligonucleotide molecule was then attached to a tRNA molecule that lacked the two 3-terminal nucleotides using T4 RNA ligase. Although effective, the process was extremely laborious, requiring purifications at most stages. Furthermore, the use of this chemical aminoacylation system usually requires the use of *in vitro*

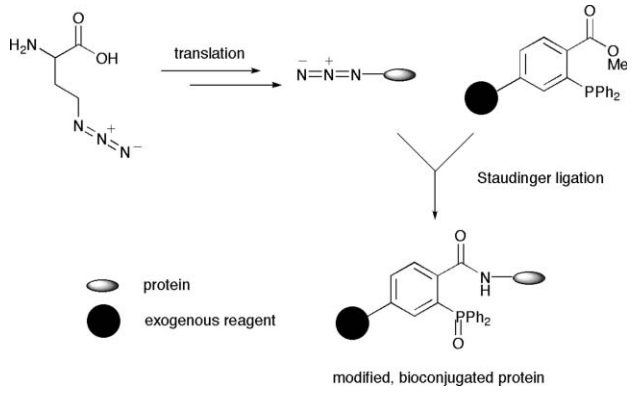
translation, which until recently has only afforded modest yields of translated protein product.

The chemical aminoacylation method was taken a step further by Schultz' group, who developed cyanomethyl esters as aminoacylating agents.[38] In the presence of base cyanomethyl esters selectively aminoacylate either the 2′ or 3′ hydroxyl group of an unprotected pCpA dinucleotide and can even be used in a mixed aqueous–organic solvent system. Recent advances exploiting micelles have further enhanced the use of cyanomethyl esters in aqueous mixtures.[39]

Ideally one would like to employ an *in vivo* system that is capable of promoting aminoacylation of tRNAs using non-natural amino acids whilst still retaining the high yields associated with *in vivo* expression systems. The problem with this approach is the intrinsically high fidelity of enzymatic aminoacylation *in vivo*. Aminoacylation *in vivo* is performed by aminoacyl tRNA synthetases (ARSs) which are able to select a specific amino acid, selectively activate this amino acid to enable esterification with a tRNA molecule then select and aminoacylate the correct tRNA molecule from within the mixture of tRNAs present in the cell. Several ARSs also possess editing domains which are able to selectively degrade misacylated tRNAs. This property is necessary in some cases in order to differentiate between amino acids that are closely related in structure.

Enzymatic aminoacylation of tRNAs

One approach taken by Tirrell is to exploit the inherent promiscuity of certain synthetases. The approach relies on using auxotrophic strains of *E. coli* that are unable to produce one of the naturally occurring amino acids. Thus in order to survive these strains must either be fed the naturally occurring amino acid or a close analogue that can be used in place of the absent amino acid. Thus by screening a range of structural analogues using the methionine auxotroph, it was possible to identify several analogues of methionine that could function in translation and allow cell growth.[40] Subsequent work based on this approach has centred on introducing fluorinated isoleucine analogues into proteins.[41] This approach allows easy incorporation of a non-natural amino acid in place of one of the natural amino acids and thus does not expand the number of amino acids that may be used in protein synthesis. However, the ability to introduce analogues of natural amino acids in a straightforward *in vivo* expression system lends this approach to applications in the large scale production of proteins that can be used for extensive mechanistic or crystallographic studies. A recent extension of the methionine analogue approach has allowed the incorporation of of azidohomoalanine (Scheme 12) into proteins, which could then be used as a handle for bioconjugation *via* the Staudinger ligation.[42]



Scheme 12 Staudinger ligation of azidohomoalanine.

Orthogonal suppressor tRNA aminoacylation *in vitro* and *in vivo*

Over the course of many years Schultz' group has developed re-engineered ARS proteins that are able to aminoacylate tRNA molecules with non-natural amino acids. With the ultimate aim of generating organisms that are able to support translation using more than the standard twenty proteinogenic amino acids, this approach has presented challenges on several levels. The first challenge is generating a suitable mutant ARS protein that is able to charge a suppressor tRNA molecule with the desired non-natural amino acid. The second challenge is to ensure that the endogenous ARS proteins of the organism cannot aminoacylate the suppressor tRNA with a natural amino acid and thus insert a natural amino acid rather than the required non-natural amino acid. Such tRNAs are termed "orthogonal". In order to generate suppressor tRNAs that are orthogonal to the aminoacylation systems in *E. coli*, attention was turned to the tRNAs present in other organisms. In particular, it has been found that when appropriately modified, certain tRNAs derived from *Methanococcus jannaschii* are orthogonal to the ARSs present in *E. coli* and other common expression systems, and can also be engineered to function as efficient stop codon suppressors. Furthermore, by performing ingenious directed evolution experiments on *M. jannaschii* ARS proteins new ARS proteins could be isolated that were able to charge the derived orthogonal suppressor tRNA with non-natural amino acids.[43] These newly derived ARSs were selected in a manner that prevented charging of endogenous tRNA species with the non-natural amino acid, and thus maintained a high level of amino acid fidelity during protein synthesis. With a viable ARS selection system in hand an explosion in the number of non-natural amino acid specific ARS systems is set to occur.

Ribosome engineering

Another strategy for the introduction of non-natural amino acids is the re-engineering of the ribosome in order to permit more radical changes in amino acid structure whilst still allowing translation to proceed. This approach has been exemplified by the use of D-amino acids in ribosomal protein synthesis by Hecht.[44] Mutant ribosomes were prepared with alterations in the sequence of the peptidyl transferase region of the 23S rRNA. Protein synthesis was then performed using an *in vitro* assay which showed that a model mRNA (dihydrofolate reductase) could be translated to yield a product protein containing a D-amino acid. The D-amino acid was introduced *via* suppression of a stop codon using a suppressor tRNA charged with either D-phenylalanine or D-methionine using chemical ligation methods. With the mutant ribosome suppression efficiency was increased up to four-fold over the the wild-type ribosome, with the efficiency reaching nearly 50% of the L-amino acid value in some cases.

Ribozyme catalysed tRNA aminoacylation

Thus far we have discussed the use of suppressor tRNAs that are able to suppress stop codons alone, however systems that are able to suppress frame shifts by reading a codon made up of four or even five bases have been developed.[45] These multibase coding systems offer the possibility of inserting more than one non-natural amino acid into a single protein molecule and recent work has realised this possibility. Using a ribozyme-powered aminoacylation system, Murakami *et al.* has generated aminoacylated suppressor tRNAs that have been used in order to incorporate two different phenylalanine analogues into a protein at the same time.[46] The approach relied on using both amber suppression and frame shift suppression in order to permit the simultaneous use of two different non-natural

amino acids. The resin-immobilised system known as "flexiresin" uses cyanomethyl esters as the source of non-natural amino acids. Owing to its broad specificity the same flexiresin system was used to aminoacylate *both* suppressor tRNAs used in this *in vitro* experiment, each with a different amino acid. Thus a single "off the shelf" catalyst system could potentially be used to generate aminoacyl-tRNAs for a broad range of purposes. In addition to the use of multibase codons, the genetic code can also be expanded by cutting the degeneracy of the genetic code as illustrated by Kwon *et al.*[47] or indeed expanding the number of nucleic acids that can participate in the coding event.[48]

Future prospects

The use of coded protein synthesis for the production of proteins containing non-natural amino acids opens up an immense number of possibilities. Applications of this type of approach have included the production of mutant enzymes for structural studies *via* fluorescence[49] and the generation of mutant nicotinic acetylcholine receptors for detailed receptor–agonist and receptor–antagonist studies.[50] However, current work on the use of *in vivo* systems will allow existing fermentors to be used with new strains in order to generate large quantities of engineered protein for a multitude of commercial and research end uses. In combination with bioconjugation,[51] NRPS and ligation approaches the tRNA engineering approach should allow the wide scale exploitation of combinatorial methods for the screening of proteins with novel properties and functions. A recent example gives a taste of the types of designer systems we may come to expect.[52]

References

1 B. G. Davis, *Curr. Opin. Biotechnol.*, 2003, **14**, 379.
2 D. F. Qi, C. M. Tann, D. Haring and M. D. Distefano, *Chem. Rev.*, 2001, **101**, 3081.
3 G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, San Diego, California, USA, 1996.
4 S. L. Lawson, W. W. Wakarchuk and S. G. Withers, *Biochemistry*, 1996, **35**, 10110.
5 E. L. Ellman, *Arch. Biochem. Biophys.*, 1959, **82**, 70.
6 D. P. Greiner, R. Miyake, J. K. Moran, A. D. Jones, T. Negishi, A. Ishihama and C. F. Meares, *Bioconj. Chem.*, 1997, **8**, 44.
7 J. B. Jones and G. Desantis, *Acc. Chem. Res.*, 1999, **32**, 99.
8 K. Khumtaveeporn, A. Ullmann, K. Matsumoto, B. G. Davis and J. B. Jones, *Tetrahedron: Asymmetry*, 2001, **12**, 249.
9 B. G. Davis, R. F. Sala, D. R. W. Hodgson, A. Ullman, K. Khumtaveeporn, D. A. Estell, K. Sanford, R. R. Bott and J. B. Jones, *ChemBioChem*, 2003, **4**, 533.
10 H. Liu, L. Wang, A. Brock, C.-H. Wong and P. G. Schultz, *J. Am. Chem. Soc.*, 2003, **125**, 1702.
11 A. Deiters, A. Cropp, M. Mukherji, J. W. Chin, C. Anderson and P. G. Schultz, *J. Am. Chem. Soc.*, 2003, **125**, 11782.
12 M. B. Soellner, K. A. Dickson, B. L. Nilsson and R. T. Raines, *J. Am. Chem. Soc.*, 2003, **125**, 11790.
13 M. A. Marahiel, T. Stachelhaus and H. D. Mootz, *Chem. Rev.*, 1997, **97**, 2651.
14 H. von Döhren, U. Keller, J. Vater and R. Zocher, *Chem. Rev.*, 1997, **97**, 2675.
15 C. R. Hutchinson, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3010.
16 V. Bergendahl, U. Linne and M. A. Marahiel, *Eur. J. Biochem.*, 2002, **269**, 620.
17 T. A. Keating, D. E. Ehmann, R. M. Kohli, C. G. Marshall, J. W. Trauger and C. T. Walsh, *ChemBiochem*, 2001, **2**, 99.
18 D. Schwarzer, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 2001, **8**, 997.
19 U. Linne and M. A. Marahiel, *Biochemistry*, 2000, **39**, 10439.
20 A. Lawen and R. Traber, *J. Biol. Chem.*, 1993, **268**, 20452.
21 H. D. Mootz and M. A. Marahiel, *J. Bacteriol.*, 1997, **179**, 6843.
22 R. M. Kohli, M. D. Burke, J. Tao and C. T. Walsh, *J. Am. Chem. Soc.*, 2003, **125**, 7160.
23 T. L. Schneider, C. T. Walsh and S. E. O'Connor, *J. Am. Chem. Soc.*, 2002, **124**, 11272.
24 R. M. Kohli, C. T. Walsh and M. D. Burkart, *Nature*, 2002, **418**, 658.
25 G. J. Cotton and T. W. Muir, *Chem. Biol.*, 1999, **6**, R247.
26 J. P. Tam, Q. Yu and Y. A. Lu, *Biologicals*, 2001, **29**, 189.
27 L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526.
28 E. C. Schwartz, T. W. Muir and A. B. Tyszkiewicz, *Chem. Commun.*, 2003, 2087.
29 H. D. Mootz, E. S. Blum, A. B. Tyszkiewicz and T. W. Muir, *J. Am. Chem. Soc.*, 2003, **125**, 10561.
30 Y. Sun, G. Lu and J. P. Tam, *Org. Lett.*, 2001, **3**, 1681.
31 F. I. Valiyaveetil, R. MacKinnon and T. W. Muir, *J. Am. Chem. Soc.*, 2002, **124**, 9113.
32 J. S. Fruton, *Adv. Enzymol.*, 1982, **53**, 239.
33 J. P. Tam, J. X. Xu and K. D. Eom, *Biopolymers*, 2001, **60**, 194.
34 C. F. Liu and J. P. Tam, *Org. Lett.*, 2001, **3**, 4157.
35 D. Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom and J. A. Wells, *Science*, 1994, **266**, 243.
36 G. G. Kochendoerfer, S. Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Y. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent and J. A. Bradburne, *Science*, 2003, **299**(5608), 884.
37 T. G. Heckler, L.-H. Chang, Y. Zama, T. Naka, M. S. Chorghade and S. M. Hecht, *Biochemistry*, 1984, **23**, 1468.
38 S. A. Robertson, J. A. Ellman and P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 2722.
39 K. Ninomiya, T. Kurita, T. Hohsaka and M. Sisido, *Chem. Commun.*, 2003, 2242.
40 J. C. M. van Hest, K. L. Kiick and D. A. Tirrell, *J. Am. Chem. Soc.*, 2000, **122**, 1282.
41 P. Wang, Y. Tang and D. A. Tirrell, *J. Am. Chem. Soc.*, 2003, **125**, 6900.
42 K. L. Kiick, E. Saxon, D. A. Tirrell and C. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 19.
43 S. W. Santoro, L. Wang, B. Herberich, D. S. King and P. G. Schultz, *Nat. Biotechnol.*, 2002, **20**, 1044.
44 L. M. Dedkova, N. E. Fahmi, S. Y. Golovine and S. M. Hecht, *J. Am. Chem. Soc.*, 2003, **125**, 6616.
45 T. Hohsaka, Y. Ashizuka, H. Taira, H. Murakami and M. Sisido, *Biochemistry*, 2001, **40**, 11060.
46 H. Murakami, D. Kourouklis and H. Suga, *Chem. Biol.*, 2003, **10**, 1077.
47 I. Kwon, K. Kirschenbaum and D. A. Tirrell, *J. Am. Chem. Soc.*, 2003, **125**, 7512.
48 E. Lee Tae, Y. Wu, G. Xia, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 2001, **123**, 7439.
49 L. E. Steward, C. S. Collins, M. A. Gilmore, J. E. Carlson, J. B. A. Ross and A. R. Chamberlin, *J. Am. Chem. Soc.*, 1997, **119**, 6.
50 E. J. Petersson, A. Choi, D. S. Dahan, H. A. Lester and D. A. Dougherty, *J. Am. Chem. Soc.*, 2002, **124**, 12662.
51 J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. W. Zhang and P. G. Schultz, *Science*, 2003, **301**(5635), 964.
52 L. Alfonta, Z. Zhang, S. Uryu, J. A. Loo and P. G. Schultz, *J. Am. Chem. Soc.*, 2003, **125**, 14662.