

## 1

**Introduction**

*The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.*

[W. L. Bragg]

## 1.1

**Classical Approaches to Protein Modification**

Chemical modification of solvent-accessible reactive side-chains has a long history in protein science and a number of group-specific modifying agents are well known. The N-terminal  $\alpha$ -amino groups in protein sequences and  $\epsilon$ -amino groups of lysine side-chains are common targets of chemical modifications. Chemical conjugations still play an important and general role in protein and cellular chemistry along with biotechnology. For example, human Annexin-V modified by conjugation of its lysine side-chains with the fluorescent dye fluoroisothiocyanate is commonly used as a marker to study cell apoptosis [1]. The modification of proteins and peptides with polyethylene glycol was, and continues to be, one of the most frequent chemical modifications used in the manufacture of biopharmaceuticals [2]. Moreover, a widely used protein modification method called “crystal soaking”, where a protein crystal is “soaked” in a solution with certain heavy metals, is an essential tool for protein X-ray crystallography (“structural genomics”) [3]. Chemical modifications were also extensively used for the introduction of different labels (fluorescence markers, spin labels, etc.) and crosslinking with various photolinkers, fluorophores and cages in order to study protein topology or protein–protein interactions [4].

In the era of classical enzymology, chemical modifications (acylations, amidations, reductive alkylations, cage reagents, etc.) of functional groups (lysine, cysteine, tyrosine, histidine, methionine, arginine and tryptophan side-chains) represented the only available approach for studies of structure–function relationships. At that time, a typical application of these approaches was to identify the residue involved in catalysis or binding (e.g. carboxyl groups of aspartic or glutamic acids, imidazole groups of histidine side-chains, hydroxy groups of tyrosine, serine and

threonine, etc.) by a substance with specific modifying capacity. Cysteine side-chains were always an attractive target for site-specific modification in proteins. This is largely driven by the relative ease of specific modification of cysteine in proteins without concomitant modification of other nucleophilic sites such as lysine and histidine. As a result, a large number of reagents are available for the modification of cysteine. For example, Bender and Koschland used this approach to chemically modify (“chemical mutation”) the active site of serine to cysteine in subtilisin, which resulted in the loss of the peptidase activity [5, 6]. These early protein structure–function studies also resulted in the mapping of the receptor-binding regions in insulin. The chemical modification of the N-terminal glycine of chain A provided a model to help understand the prerequisites of productive binding. However, it was not possible to resolve the influence of the steric bulk imposed upon modification from the effects of the positive charge neutralization at the N-terminal amino group [7].

The lack of specificity and unpredictability of the reaction accompanied by a distinct reduction or even loss of activity of the protein under study are, among others, the major reasons why protein modifications are no longer popular today. Site-directed mutagenesis offers a much more elegant and precise tool to address such issues. For example, in the absence of a precise three-dimensional structure, alanine (or glycine)-scanning mutagenesis [8] is suitable for systematic substitutions of all residues in (or around) the putative active center of the protein under study and the subsequent identification of the “essential” functional group.

## 1.2

### Peptide Synthesis, Semisynthesis and Chemistry of Total Protein Synthesis

The “peptide theory” put forward in 1902 by Emil Fisher and Franz Hofmeister [9] correctly postulated that proteins are made up of  $\alpha$ -amino acids that are linked head-to-tail by amide bonds. At that time chemists were mainly interested in the total chemical synthesis of protein by using the techniques of classical organic chemistry. Emil Fisher succeeded in synthesizing an 18-residue peptide composed of glycine and L-leucine in which amino acids were combined to yield small peptides that could be coupled together to produce larger peptides [10]. The amino acids could then be linked via the peptide (amide) in a stepwise manner [11]. Alternatively, in convergent fragment condensation, small peptides can be coupled together to give a larger peptide. Fragment synthesis of insulin, the first protein molecule to have been sequenced, was accomplished by three groups working independently at approximately the same time. The most recent achievement of solution synthesis was the preparation of a linear 238-residue protein, a precursor of green fluorescent protein (GFP) from *Aequorea victoria* [12]. Convergent condensation in combination with peptide bond synthesis in a stepwise manner was successfully applied in the solid phase, resulting in the first total synthesis of the enzyme ribonuclease S [13, 14]. In this method, the C-terminal residue of the peptide which is covalently anchored to an insoluble support is used to assemble the remaining amino acids in a stepwise manner and, finally, to cleave the synthesized

product from the solid phase. These initial results have been followed by a series of successful syntheses of a variety of enzymes.

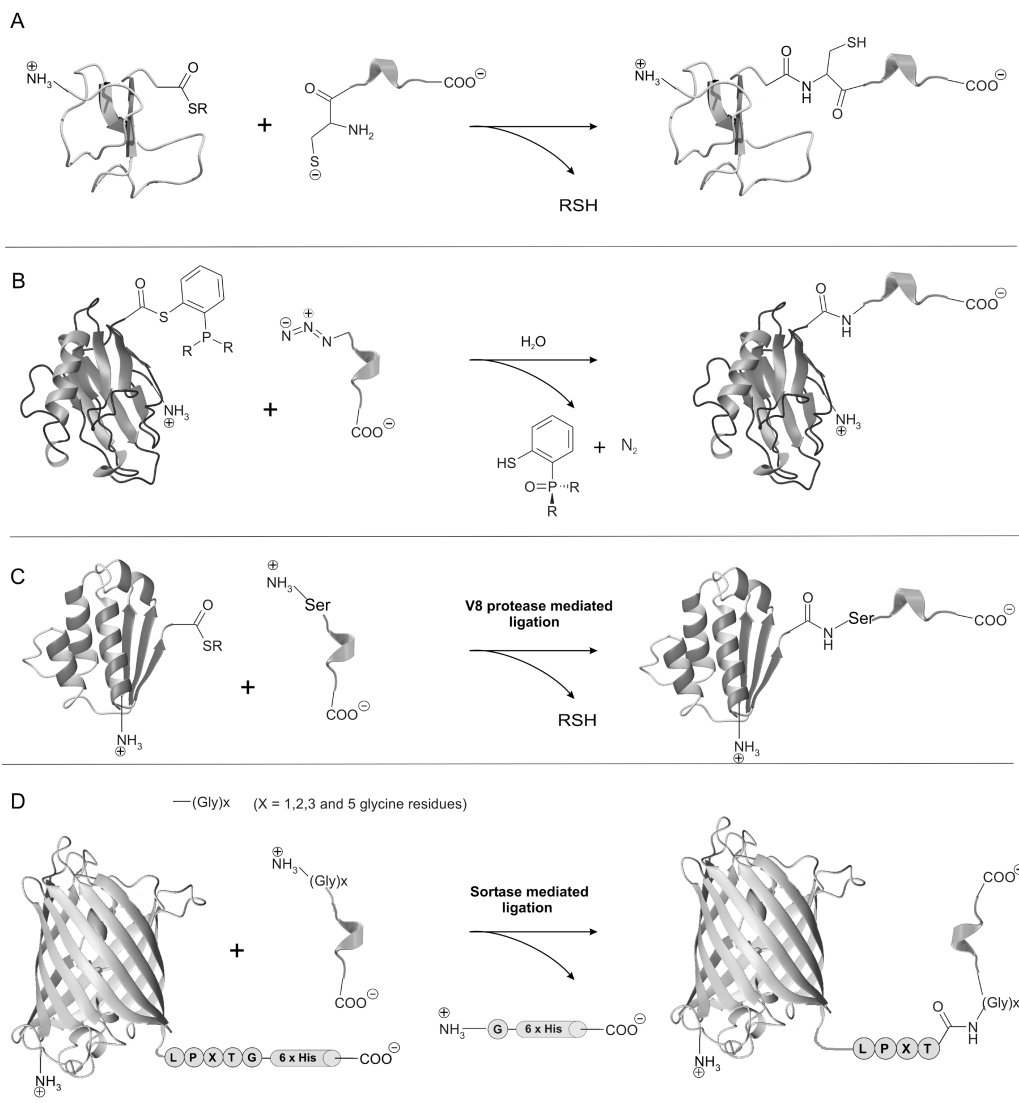
These methods require a high degree of chemical sophistication, advanced synthetic methods and skillful experimentation. These are some of the chief reasons why they are still not widespread among the community of synthetic chemists. These approaches have their own drawbacks, such as the poor solubility of the protected peptide intermediates formed by solution synthesis along with the accumulation of byproducts that block reactions in a stepwise solid-phase synthesis. For all these reasons, the total chemical synthesis of a homogeneous protein longer than 100 residues still presents a formidable challenge.

The strategy of convergent assembly (i.e. condensation) of synthetic and natural peptide fragments is termed “protein semisynthesis”. The basic requirements for semisynthesis are: (i) the synthetic donor peptide has to be protected and activated, and (ii) an acceptor protein fragment that has to be prepared by enzymatic or chemical fragmentation of the parent protein should be available and properly protected. Offord and Rose pioneered the use of hydrazone- and oxime-forming reactions for chemically ligating such fragments [15]. Although these chemistries are selective, they were in practice often hampered by the insolubility of the large protected peptide building blocks. Most recently, the rediscovery of the Staudinger ligation has represented an additional breakthrough in this field (see Fig. 1.1) [16]. From a purely chemical perspective, it is an excellent tool for protein/peptide ligation that allows different protein/peptide fragments to be coupled at any desired position, without the requirement for a particular sequence composition [17].

In order to chemically create a native amide bond between interacting fragments, Kent and coworkers successfully performed a thiol–thioester exchange between unprotected fragments in aqueous solution – a technique they called “native chemical ligation” (NCL) [18]. The N-terminal fragment contains a C-terminal electrophilic  $\alpha$ -thioester which can be conjugated via a thiol–thioester exchange reaction to the N-terminal thiol-harboring fragment. Such developments are based on the early discovery by Wieland and co-workers [19] that thioesters of amino acids and peptides form peptide bonds with N-terminal cysteine in neutral aqueous solution via a spontaneous  $S \rightarrow N$ -acyl shift, along with the procedures of Blake [20] and Yamashiro [21] for thioester preparation by solid-phase synthesis. See Fig. 1.1.

Some of the most striking examples that demonstrate the unique capacity of NCL are total protein synthesis of the D-chiral form of HIV-1 protease (100 residues) [28] and the preparation of the posttranslationally modified artificial variant of erythropoietin [29] (polymer modified; 166 residues). The D-enantiomer of HIV-1 protease is completely active and exhibits reciprocal chiral specificity just as much as the L-enantiomer, i.e. it is capable of cutting only the D-peptide substrates. Its three-dimensional structure is in all respects the mirror image of the “natural” L-protease. Indeed, these examples dramatically illustrate the considerable potential of NCL as a complementary approach for protein engineering based on templated ribosome-mediated protein synthesis.

Since peptide synthesis is generally limited by the size of polypeptide-chain (around 10 kDa) and genetically encoded protein modification (*vide infra*) still suf-



**Fig. 1.1.** Basic approaches for chemoselective ligations [22, 23]. (A) NCL between unprotected peptide fragments that contain the requisite reactive groups (N-terminal Cys and C-terminal thioester). (B) Staudinger ligation forms an amide bond from an azide and a specifically fictionalized phosphine. This method allows independent amino acid sequences to be coupled at any desired Xxx–Yyy bond and, in this respect, is potentially almost universal [24]. (C) EEL [25] uses specific thioesters (generated by “classical” intein-mediated approaches) as a substrate mimetic

for a *Staphylococcus aureus* V8 serine protease in order to bypass the requirement for the cysteine at the ligation site. (D) Sortase from *S. aureus*, a membrane-anchored trans-peptidase, cleaves any polypeptide provided with a C-terminal sorting signal (LPXTG motif). This strategy when applied to tagged green fluorescent protein (Nt-GFP-LPXTG-6His-Ct) enables its successful conjugation with various donor molecules containing two or more N-terminal glycines [26] (D- and L-peptides, nonpeptide fragments and even other GFPs). (From Budisa [27]; © Wiley 2004.)

fers from limited scope, NCL has no current rival in the various types of experimental designs that include sequential isotopic labeling, preparation of circular proteins, insertion of the non-native polypeptide fragments or nonpeptide molecules at predefined sites. The most commonly used ligation techniques for chemical transformation of proteins are thioester ligation, Staudinger ligation, oxime or hydrazone ligation and disulfide ligation.

### 1.3

#### Chemoselective Ligations Combined with Biochemical Methods

In parallel with these purely chemical approaches, a number of biochemical methods have also entered this arena. For example, protein splicing can now be coupled with native chemical ligation and to give rise to expressed protein ligation (EPL) [30]. This technique takes advantage of recombinant DNA technology to generate protein fragments of any size via ribosomal synthesis. In this way large proteins become accessible for chemoligation. An interesting alternative to these chemoligation procedures is enzyme-catalyzed condensation [31] which was demonstrated as early as 1938 by Max Bergman [32]. The studies of Kaiser and co-workers [33] on subtilisin in the 1960s and 1970s provided a solid base for the success of Wells and associates [34] in the engineering of an active site for this enzyme. They generated an enzyme (“subtiligase”) capable of efficiently catalyzing the ligation of peptide fragments. Subtiligase exhibits a largely reduced proteolytic activity and is functionally active as an acyl transferase. This property was exploited for enzymatic condensation of six peptide fragments of ribonuclease A (each 12–30 residues, one of them containing the noncanonical amino acid 4-fluorohistidine) [35].

Recent research in this field yielded expressed enzymatic ligation (EEL) [25], which combines the advantages of EPL and the substrate mimetic strategy of protease-mediated ligation. However, genome and proteome mapping among different biological kingdoms might offer attractive tools for such purposes. For example, sortase-catalyzed proteolysis entered the arena of enzyme-mediated native protein ligation very recently. Sortases are bacterial (*Staphylococcus aureus*) enzymes that are responsible for the covalent attachment of specific proteins to the cell wall of Gram-positive bacteria in a two-step transpeptidation reaction either *in vivo* or *in vitro* [36]. This strategy has now been “borrowed” from nature and was shown to be suitable for protein–peptide and protein–protein ligations [26]. It is reasonable to expect that the recruitment of chemical strategies that living organisms have optimized and developed during their evolution and their application in chemistry in the future will be crucial for the development of novel technologies.

### 1.4

#### Methods and Approaches of Classical Protein Engineering

The use of the polymerase chain reaction (PCR) – originally developed by Kary Mullis for efficient multiplication of specific DNA sequences [37], their sequencing

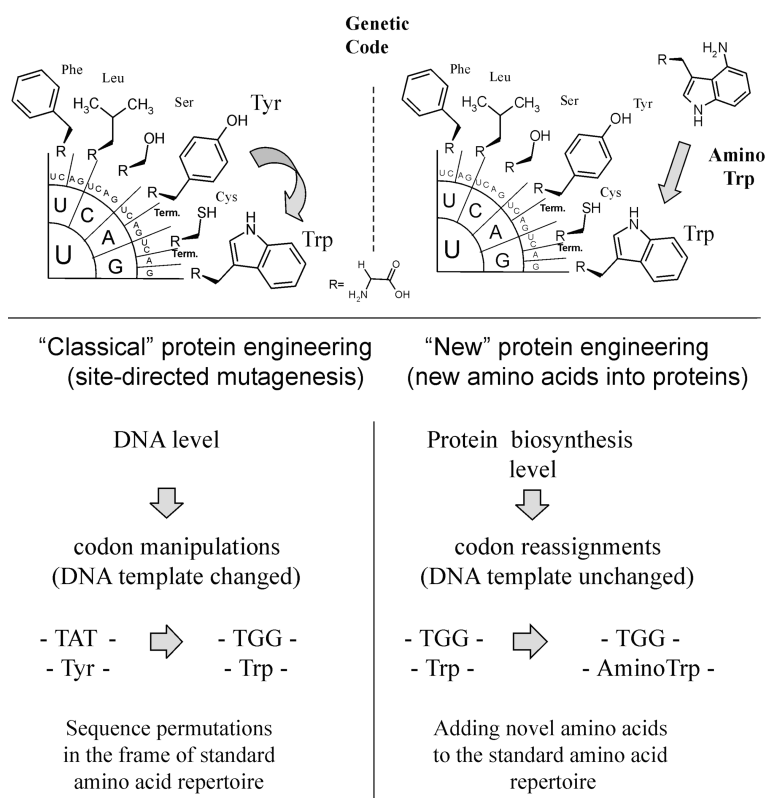
and cloning – has revolutionized the possibilities for protein engineering. The PCR greatly simplified experimental procedures to tailor new genes *in vitro* via fragment deletion or insertion or nucleotide substitutions. Indeed, it is only with the approach of oligonucleotide-mediated site-directed substitutions of particular amino acids in a target sequence [38] that the term “protein engineering” entered the vocabulary of protein science [39]. The design and identification of proteins with novel functions and properties was dramatically powered by methods that mimic Darwinian evolutionary processes, i.e. natural evolution produces a large number of variants by mutation and subsequently selects the “fittest” among them. Routine molecular biology methods of mutation/recombination and screening/selection in the test tube allow for rapid and direct isolation of biomolecules based on their functional properties. This collection of methods has been termed directed evolution [40] and provides a powerful tool for the development of biocatalysts with novel properties without requiring an understanding of their complicated structure–function relationships, or knowledge of enzyme structures or catalytic mechanisms.

However, the major limitation of these methods and approaches (usually neglected in the current literature) is that the changes introduced are limited to the repertoire of the canonical 20 amino acids. The above-mentioned methods, in combination with the experimental extension of the amino acid repertoire of the genetic code through its engineering, will open a new era for designing not only protein structure and function, but also the design of novel cell types. Thus, traditional (“classical”) methods for protein engineering and design can be supplemented or even fully replaced by these novel approaches (Fig. 1.2).

### 1.5

#### Genetically Encoded Protein Modifications – Reprogramming Protein Translation

The ability and capacity of living cells to synthesize functional proteins is unrivaled. From the synthetic chemist’s point of view, the basic features which demonstrate the power and versatility of ribosome-mediated protein synthesis over synthetic approaches are the structural homogeneity of the synthesized polypeptide and the possibility for the precise control of the (stereo)chemical composition of the desired sequence. On the other hand, the same mechanisms that ensure high fidelity of such templated protein synthesis limit the diversity of the amino acid basic building blocks in this process. Therefore, breaking these limits either *in vitro* or *in vivo* (eubacteria and eukaryotic cells) should offer the possibility to add novel amino acids into the existing repertoire of the genetic code. This would be possible only if the protein translational apparatus is reprogrammed, and subsequently the scope and utility of the protein engineering is greatly expanded. This means that genetically encoded (i.e. templated) protein modifications in combination with genetics, physiology and metabolic manipulation of living cells should have great advantages over classical chemical modification, peptide synthesis, chemoligation and even routinely used site-directed mutagenesis. This would make



**Fig. 1.2.** Classical versus new protein engineering. The discovery of site-directed mutagenesis [38] allowed for permutation of any existing gene-encoded protein sequence by codon manipulation such as TAT (Tyr) → TGG (Trp), i.e. replacement of one standard and conserved amino acid with another. The delivery of novel amino acids into the existing amino acid repertoire as prescribed by the genetic code is a novel form of protein engineering [41]. This is exemplified above by the experimental reassignment of the UGG (TGG) coding unit for Trp to aminotryptophan (AminoTrp). In this way, the interpretation of the genetic code is changed, i.e. canonical →

noncanonical amino acid replacement at the level of the target protein sequence is fully achieved. Related proteins can be defined as mutants and variants. *Mutant* denotes a protein in which the wild-type sequence is changed by site-directed mutagenesis (codon manipulation at the DNA level) in the frame of the standard amino acid repertoire. *Variant* denotes a protein in which single or multiple canonical amino acids from a wild-type or mutant sequence are replaced with non-canonical ones (expanded amino acid repertoire, codon reassignment at the protein translation level).

not only novel side-chain and backbone chemistries accessible, but also open a general perspective for novel chemistry to be performed in a controlled manner exclusively inside the living cells.

Methods for the expansion of the number of amino acids that can serve as basic building blocks in ribosome-mediated protein synthesis are described in the cur-

rent literature under different names, such as “expanded scope of protein synthesis” [42], “expanded amino acid repertoire” [43], “expanded genetic code” [44], “tRNA-mediated protein engineering” (TRAMPE) [45], “site-directed non-native amino acid replacement” (SNAAR) [46], etc. Their common feature is experimental re-coding, read-through or changes in meaning of coding triplets in the frame of the existing universal genetic code. The experimental read-through can be achieved either by reassignment of evolutionarily assigned coding triplets (i.e. sense codons), suppression of termination triplets (UGA, UGG and UGU) or non-triplet coding units. In the context of protein expression, this can be achieved by controlling environmental factors (i.e. selective pressure) of the intact, but genetically modified, cells, by their supplementation with redesigned translation components or by a combination of both. At the level of the universal genetic code these experiments lead to an increase in its coding capacity by expanding its amino acid repertoire. The term “engineering of the genetic code” covers all these aspects. It strictly refers to experiments aimed at changing the interpretation of the universal genetic code or changing the structure of the code by the introduction of novel coding units.

## 1.6

### Basic Definitions and Taxonomy

A precise terminology is usually hampered in the early stages of development of any novel research field because not enough is known to permit accurate definitions. A pragmatic strategy applied in this book is to accept provisional, rough terminological characterizations which can leverage the field’s early developmental stages with the taxonomic refinements emerging as the surrounding facts become clearer. For example, a great deal of knowledge gained from code engineering experiments is actually related to the chemistry of *Escherichia coli* which, to date, remains to the main cell type for protein expression experiments. There should be no doubt that genetic code engineering in eukaryotic cells will provide additional, novel and exciting facts, insights and concepts, and subsequently lead to revisions of the existing ones. By taking into account the engineering of the genetic code’s integration in several disciplines, it is obvious how difficult it is to gain a comprehensive overview in this field. Molecular and systems biology, genetics, metabolic research, research on posttranslational modifications or pharmacological properties are also bringing their own rather complex terminology in this field. For example, functional genomics or proteomics cover gene actions and interactions on a genome- or proteome-wide scale, and include at least four levels of complexity: genes (the genome), messenger RNA (the transcriptome), proteins (the proteome) and metabolism (the metabolome).

In the early studies [47], the basic criterion for one substance to be regarded as an “analog” was the requirement to have a shape and size similar to a naturally occurring molecule without any dramatic differences in biophysical properties. Such amino acid analogs that are sterically almost identical to the canonical



ones, e.g. Met/SeMet or Arg/canavanine, are named isosteres. Modifications that include addition/deletion of one or more side-chain methylene, e.g. Met/ethionine, or other groups resulted in amino acids termed homologs. Relatively simple terminology, already proposed in the current literature [43], that classifies  $\alpha$ -amino acids into two general groups (canonical and noncanonical) is used throughout this book. Other terms that will occasionally be used in the text are cognate/noncognate, coded/noncoded, proteinogenic/nonproteinogenic, standard/non-standard, natural/unnatural/non-natural, special canonical amino acids (generated via cotranslational modifications) and biogenic amino acids (generated via posttranslational modifications). Indeed, such  $\alpha$ -amino acids, e.g. selenocysteine or diiodo-tyrosine, are also “natural”, “proteinogenic” or “common” and even “canonical” in the context of their physiological appearance. All these aspects are summarized in Tab. 1.1.

Nowadays, the true meaning of the nature of the universal genetic code is often confused. For example, it is often reported in the media that the human, yeast or mouse “genetic code” has been “cracked”, thereby causing unnecessary confusion. In fact, this relates to the complete genome sequencing (i.e. nucleotide sequence mapping) of the human, yeast or mouse genome. The term “genetic code” specifically refers to the correspondence between nucleotide sequence and amino acid

**Tab. 1.1.** Taxonomy of canonical and noncanonical amino acids (these amino acids that participate in ribosome-mediated protein synthesis are classified according to their assignments/reassignments in the code table, metabolic activity and origin; posttranslational modifications are strictly separated from coding process).

Feature	$\alpha$ -amino acids <sup>a</sup>	
Mode of translation in the polypeptide		
sense codon-dependent incorporation	canonical	noncanonical
mostly nonsense coding unit reassignment in a context-dependent manner	special canonical <sup>b</sup>	special noncanonical <sup>c</sup>
posttranslational modifications (strictly separated from basic coding)	special biogenic	–
Other aspects		
position in the genetic code table <sup>d</sup>	naturally assigned	experimentally assigned
<i>in vivo</i> effects on cellular viability	vital	mostly toxic
origin	mostly metabolism	mostly anthropogenic

<sup>a</sup>Most canonical or noncanonical amino acids can indeed be built into peptides and proteins by peptide synthesis or total chemical synthesis protocols as discussed in previous sections.

<sup>b</sup>“Special canonical amino acid” refers to formyl-methionine, selenocysteine and pyrrolysine (see Section 3.10).

<sup>c</sup>Site-directed introduction of special noncanonical amino acids is extensively discussed in Chapter 5.

<sup>d</sup>This does not apply for special biogenic amino acids since they enter the protein structure after translation.

sequence. Canonical amino acids were defined by a three-letter code, e.g. methionine (Met), leucine (Leu), phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp), proline (Pro), etc. Noncanonical amino acids are given comparable names often denoted by a three letter code: norleucine (Nle), ethionine (Eth) or with prefixes that denote chemical functionality, e.g. 7-azatryptophan [(7-Aza)Trp]. They are all defined upon their first appearance in the text. The term “analog” defines strict isosteric exchange of canonical/noncanonical amino acids (e.g. methionine/selenomethionine) while the term “surrogate” defines nonisosteric changes (e.g. methionine/ethionine, tryptophan/thienopyrrolylalanine). Their mode of translation in proteins can be position-specific (directed by reassignments of rare codons and by suppression of termination or frameshifted coding units) or multiple-site (usually directed by reassignments of common coding triplets or codon families in the target sequence).

## References

- 1 KOOPMAN, G., REUTELINGSPERGER, C., KUIJTEN, G., KEEHNEN, R., PALS, S. and VAN OERS, M. (1994). Annexin-V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood* **84**, 1415–1420.
- 2 KATRE, N. V., KNAUF, M. J. and LAIRD, W. J. (1987). Chemical modification of recombinant interleukin 2 by polyethylene glycol increases its potency in the murine Meth-A sarcoma model. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 1487–1491.
- 3 BLUNDELL, T. L. and JOHNSON, L. N. (1976). Protein Crystallography. Academic Press, New York.
- 4 MEANS, G. E. and FEENEY, R. E. (1990). Chemical modifications of proteins: history and applications. *Bioconjugate Chemistry* **1**, 2–12.
- 5 NEET, K. E. and KOSHLAND, D. E. (1966). The conversion of serine at the active site of subtilisin to cysteine: a chemical mutation. *Proceedings of the National Academy of Sciences of the United States of America* **56**, 1606–1616.
- 6 POLGAR, L. and BENDER, M. L. (1969). The nature of general base-general acid catalysis in serine proteases. *Proceedings of the National Academy of Sciences of the United States of America* **64**, 1335–1342.
- 7 PULLEN, R. A., LINDSAY, D. G., WOOD, S. P., TICKLE, I. J., BLUNDELL, T. L., WOLLMER, A., KRAIL, G., BRANDENBURG, D., ZAHN, H., GLIEMANN, J. and GAMMELTOFT, S. (1976). Receptor-binding region of insulin. *Nature* **259**, 369–373.
- 8 LEFÈVRE, F., RÉMY, M. H. and MASSON, J. M. (1997). Alanine-stretch scanning mutagenesis: a simple and efficient method to probe protein structure and function. *Nucleic Acids Research* **25**, 447–448.
- 9 HOFMEISTER, F. (1902). Über Bau und Gruppierung der Eiweißkörper. *Ergebnisse der Physiologie* **1**, 759–802.
- 10 FISCHER, E. (1906). Untersuchungen über Aminosäuren, Polypeptide und Proteine. *Berichte der deutschen chemischen Gesellschaft* **39**, 530–610.
- 11 MERRIFIELD, R. B. (1963). Solid phase synthesis. *Journal of the American Chemical Society* **85**, 21–49.
- 12 NISHIUCHI, Y., INUI, T., NISHIO, H., BODI, J., KIMURA, T., TSUJI, F. I. and SAKAKIBARA, S. (1998). Chemical synthesis of the precursor molecule of the Aequorea green fluorescent protein, subsequent folding, and development of fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1151–1156.

- Sciences of the United States of America* **95**, 13549–13554.
- 13 GUTTE, B. and MERRIFIELD, R. B. (1969). Total synthesis of an enzyme with ribonuclease A activity. *Journal of the American Chemical Society* **91**, 501–507.
  - 14 DENKEWALTER, R. G., WEBER, D. F., HOLLY, F. W. and HIRSCHMANN, R. (1969). Studies on total synthesis of an enzyme. I. Objective and strategy. *Journal of the American Chemical Society* **91**, 502–507.
  - 15 GAERTNER, H. F., ROSE, K., COTTON, R., TIMMS, D., CAMBLE, R. and OFFORD, R. E. (1992). Construction of protein analogs by site-specific condensation of unprotected fragments. *Bioconjugate Chemistry* **3**, 262–268.
  - 16 STAUDINGER, H. and MEYER, J. (1919). Phosphinethylene derivatives and phosphinimines. *Helvetica Chimica Acta* **2**, 635–646.
  - 17 KÖHN, M. and BREINBAUER, R. (2004). The Staudinger ligation – a gift to chemical biology. *Angewandte Chemie* **43**, 3106–3116.
  - 18 DAWSON, P. E., MUIR, T. W., CLARKE LEWIS, I. and KENT, S. B. H. (1994). Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779.
  - 19 WIELAND, T. and SCHAFER, W. (1951). Synthese Von Oligopeptiden Unter Zellmöglichen Bedingungen. *Angewandte Chemie* **63**, 146–147.
  - 20 BLAKE, J. (1981). Peptide segment coupling in aqueous-medium – silver ion activation of the thiolcarboxyl group. *International Journal of Peptide and Protein Research* **17**, 273–274.
  - 21 YAMASHIRO, D. and LI, C. H. (1988). New segment synthesis of alpha-inhibin-92 by the acyl disulfide method. *International Journal of Peptide and Protein Research* **31**, 322–334.
  - 22 CASI, G. and HILVERT, D. (2003). Convergent protein synthesis. *Current Opinion in Structural Biology* **13**, 589–594.
  - 23 MUIR, T. W. (2003). Semisynthesis of proteins by expressed protein ligation. *Annual Review of Biochemistry* **72**, 249–289.
  - 24 NILSSON, B. L., HONDAL, R. J., SOELLNER, M. B. and RAINES, R. T. (2003). Protein assembly by orthogonal chemical ligation methods. *Journal of the American Chemical Society* **125**, 5268–5269.
  - 25 MACHOVA, Z., VON EGGELKRAUT-GOTTANKA, R., WEHOFSKY, N., BORDUSA, F. and BECK-SICKINGER, A. G. (2003). Expressed enzymatic ligation for the semisynthesis of chemically modified proteins. *Angewandte Chemie-International Edition* **42**, 4916–4918.
  - 26 MAO, H. Y., HART, S. A., SCHINK, A. and POLLOK, B. A. (2004). Sortase-mediated protein ligation: a new method for protein engineering. *Journal of the American Chemical Society* **126**, 2670–2671.
  - 27 BUDISA, N. (2004). Adding new tools to the arsenal of expressed protein ligation. *ChemBioChem* **5**, 1176–1179.
  - 28 SCHNOIZER, M. and KENT, S. B. H. (1992). Constructing proteins by dovetailing unprotected synthetic peptides – backbone-engineered HIV protease. *Science* **256**, 221–225.
  - 29 KOCHENDOERFER, G. G., CHEN, S. Y., MAO, F., CRESSMAN, S., TRAVIGLIA, S., SHAO, H. Y., HUNTER, C. L., LOW, D. W., CAGLE, E. N., CARNEVALI, M., GUERIGUIAN, V., KEOGH, P. J., PORTER, H., STRATTON, S. M., WIEDEKE, M. C., WILKEN, J., TANG, J., LEVY, J. J., MIRANDA, L. P., CRNOGORAC, M. M., KALBAG, S., BOTTI, P., SCHINDLER-HORVAT, J., SAVATSKI, L., ADAMSON, J. W., KUNG, A., KENT, S. B. H. and BRADBURN, J. A. (2003). Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* **299**, 884–887.
  - 30 MUIR, T. W., SONDEHI, D. and COLE, P. A. (1998). Expressed protein ligation: a general method for protein engineering. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6705–6710.

- 31 BORDUSA, F. (2002). Proteases in organic synthesis. *Chemical Reviews* **102**, 4817–4867.
- 32 BERGMANN, M. and FRAENKEL-CONRAT, H. (1938). The enzymatic synthesis of peptide bonds. *Journal of Biological Chemistry* **124**, 1–6.
- 33 NAKATSUKA, T., SASAKI, T. and KAISER, E. T. (1987). Peptide segment coupling catalyzed by the semisynthetic enzyme thiolsubtilisin. *Journal of the American Chemical Society* **109**, 3808–3810.
- 34 JACKSON, D. Y., BURNIER, J., QUAN, C., STANLEY, M., TOM, J. and WELLS, J. A. (1994). A designed peptide ligase for total synthesis of ribonuclease-A with unnatural catalytic residues. *Science* **266**, 243–247.
- 35 CHANG, T. K., JACKSON, D. Y., BURNIER, J. P. and WELLS, J. A. (1994). Subtiligase – a tool for semisynthesis of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12544–12548.
- 36 MAZMANIAN, S. K., LIU, G., HUNG, T. T. and SCHNEEWIND, O. (1999). *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763.
- 37 MULLIS, K. (1987). Process for amplifying nucleic acid sequences. *US Patent* 4683202.
- 38 HUTCHISON, C. A., PHILLIPS, S., EDGELL, M. H., GILLAM, S., JAHNKE, P. and SMITH, M. (1978). Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry* **253**, 6551–6560.
- 39 FERSHT, A. R. (1985). *Enzyme Structure and Mechanism*. Freeman, San Francisco.
- 40 ARNOLD, F. H. (1997). Design by directed evolution. *Accounts of Chemical Research* **31**, 125–131.
- 41 MINKS, C., ALEFELDER, S., MORODER, L., HUBER, R. and BUDISA, N. (2000). Towards new protein engineering: *in vivo* building and folding of protein shuttles for drug delivery and targeting by the selective pressure incorporation (SPI) method. *Tetrahedron* **56**, 9431–9442.
- 42 KIRCK, K. L. and TIRRELL, D. A. (2000). Protein engineering by *in vivo* incorporation of non-natural amino acids: control of incorporation of methionine analogues by methionyl-tRNA synthetase. *Tetrahedron* **56**, 9487–9493.
- 43 BUDISA, N., MINKS, C., ALEFELDER, S., WENGER, W., DONG, F. M., MORODER, L. and HUBER, R. (1999). Toward the experimental codon reassignment *in vivo*: protein building with an expanded amino acid repertoire. *FASEB Journal* **13**, 41–51.
- 44 NOREN, C. J., ANTHONYCAHILL, S. J., GRIFFITH, M. C. and SCHULTZ, P. G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188.
- 45 ROTHSCHILD, K. J. and GITE, S. (1999). tRNA-mediated protein engineering. *Current Opinion in Biotechnology* **10**, 64–70.
- 46 IBBA, M. and HENNECKE, H. (1994). Towards engineering proteins by site-directed incorporation *in vivo* of nonnatural amino acids. *Biotechnology* **12**, 678–682.
- 47 RICHMOND, M. H. (1962). Effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriological Reviews* **26**, 398–420.