PROTEIN ASSEMBLY TO MINE THE HUMAN GENOME

BRADLEY L. NILSSON, MATTHEW B. SOELLNER,
RONALD T. RAINES*
Department of Chemistry and Department of Biochemistry
University of Wisconsin–Madison
Madison, Wisconsin 53706, USA
E-mail: raines@biochem.wisc.edu

1. Introduction

The human genome contains 30,000 or so genes [1,2]. Scientists from a broad range of disciplines are now working to reveal the structure and function of the proteins encoded by these genes. Their findings could lead to the solution of a multitude of problems in biology and medicine. In addition to structure–function analyses of extant proteins, chemical biologists are working to create new proteins with desirable properties, either by *de novo* design or by altering natural frameworks.

The study of natural proteins and the creation of nonnatural ones requires the ability to access and manipulate proteins. The isolation of proteins from their natural source is often tedious, idiosyncratic, and impractical. In contrast, the production of proteins by using recombinant DNA (rDNA) technology, either in a heterologous host or *in vitro*, can provide access to large quantities of protein and allows for the exchange of one of the 20 natural amino acid residues for another. Yet, aggregation often limits the yield of properly folded proteins produced by recombinant proteins. Moreover, the restrictions of the genetic code severely limit the possible modifications.

The chemical synthesis and semisynthesis of proteins harbors the potential to overcome many of the disadvantages of current protein production methods. Chemical synthesis using established solid-phase techniques could be rapid and easily automated, and facilitate purification. Folding synthetic proteins while they remain attached to a solid support could prevent aggregation. In addition, chemical synthesis enables the facile incorporation of nonnatural amino acids or modules.

2. Native Chemical Ligation

New methods are facilitating the total chemical synthesis of proteins [3]. In particular, the ligation of synthetic peptides provides a convergent route to proteins by chemical synthesis. Currently, the most common ligation method is "native chemical ligation". This ligation method was discovered in 1953, when the reaction of ValSPh and CysOH in aqueous buffer was shown to yield the dipeptide: ValCysOH [4]. In the 1990's, this seminal discovery was developed into a practical method to ligate large peptide fragments [5]. In native chemical ligation, the thiolate of an *N*-terminal cysteine residue of one peptide attacks the *C*-terminal

359

thioester of another peptide. An amide linkage forms after a rapid S- to N-acyl shift (Scheme 1). "Expressed protein ligation" is an extension of native chemical ligation in which the *C*-terminal thioester is produced by rDNA technology rather than chemical synthesis [6,7].
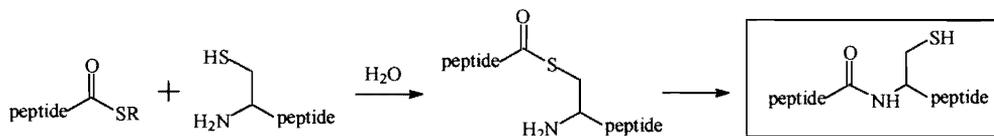
Scheme 1

2.1. PROTEIN PROSTHESIS: SELENOCYSTEINE

We have expanded the use of native chemical ligation and expressed protein ligation to include ligation at selenocysteine (Sec, U) rather than cysteine [8,9]. In natural proteins, selenocysteine is incorporated translationally and has its own $tRNA^{Sec}$ and codon, UGA, which is also the *opal* stop codon. Decoding a UGA codon for selenocysteine requires a unique structure in the 3′ untranslated region of the *m*RNA called a *se*lenocysteine *in*sertion *s*equence (SECIS) element. The production of eukaryotic selenocysteine-containing proteins in prokaryotes is problematic because eukaryotic and prokaryotic cells use a different SECIS element. We reasoned that selenocysteine, like cysteine, could effect native chemical ligation, and thereby provide a semisynthetic means to incorporate selenocysteine into proteins [8,10].

Through the use of a model thioester, we were able to demonstrate the feasibility of using selenocysteine in native chemical ligation [8]. In addition, we generated a pH–rate profile that demonstrates the increased rate of selenocysteine ligations compared to cysteine ligations at lower pH. For example, reaction with selenocysteine is 10^3-fold faster than with cysteine at pH 5.0. We next explored the use of selenocysteine in expressed protein ligation using ribonuclease (RNase) A as a model protein [8,9]. We used rDNA technology to prepare a fragment corresponding to residues 1–109 of RNase A with a *C*-terminal thioester. We used standard solid-phase methods to synthesize a peptide corresponding to residues 110–124 with selenocysteine at residue 110. We ligated the thioester fragment and the peptide fragment, and then folded and purified the protein product. The identity of the desired C110U RNase A protein was verified by mass spectrometry. The wild-type catalytic activity of this enzyme is indicative not only of an intact protein, but also of one that is folded properly and has a selenosulfide (Se–S) bond between Sec110 and Cys56.

Thus, the isomorphous replacement of sulfur with selenium can be effected with expressed protein ligation [8,9,11]. This exchange could enhance the conformational stability of proteins. A selenosulfide bond has a lower reduction potential than does a disulfide bond. Accordingly, semisynthetic proteins with selenosulfide bonds should be more resistant to denaturation in a reducing environment, such as the cytosol.

2.2. PROTEIN PROSTHESIS: β-TURN MIMIC

In another type of protein prosthesis, we have used expressed protein ligation to incorporate nonnatural β-turn mimics into the RNase A scaffold [12]. In native RNase A, residues

Gly112–Asn113–Pro114–Tyr115 form a type VI reverse turn within a hairpin. We suspected that residues Asn113–Pro114 could be replaced by a β-turn mimic composed of two β-amino acid residues, *R*-nipecotic acid–*S*-nipecotic acid (*R*-Nip–*S*-Nip), which was known to promote β-hairpin formation.

We used expressed protein ligation to replace Asn113–Pro114 with *R*-Nip–*S*-Nip (Scheme 2). RNase A fragment 1–94 was produced as a *C*-terminal thioester and fragment 95–124 containing *R*-Nip–*S*-Nip was synthesized by solid-phase peptide synthesis. The two fragments were ligated, and the protein product was folded and purified. The catalytic activity of the variant containing *R*-Nip–*S*-Nip was indistinguishable from that of wild-type RNase A. Moreover, the conformational stability of the *R*-Nip–*S*-Nip variant was enhanced slightly over that of the wild-type protein, with $\Delta T_m = (1.2 \pm 0.3)$ °C. The diastereomeric *R*-Nip–*R*-Nip module cannot adopt a turn conformation. When incorporated in place of Asn113–Pro114, *R*-Nip– *R*-Nip resulted in a protein that did not fold into an active conformation. Thus, the nonnatural segment does not serve merely as a passive linker. Instead, the specific conformational propensity of the prosthetic segment is critical for successful replacement of a natural β-turn.

Gly112-Asn113-Pro114-Tyr115   Gly112-*R*-Nip-*S*-Nip-Tyr115
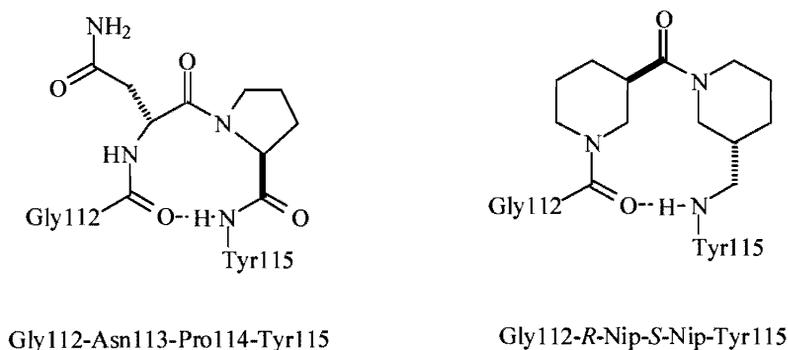
Scheme 2

This work demonstrates the power of semisynthesis to create proteins with desirable attributes. The incorporation of nonnatural modules such as *R*-Nip–*S*-Nip into proteins would be impractical or impossible by other means. In addition to endowing proteins with enhanced stability, semisynthesis can be used to incorporate specific probes for the study of protein structure–function relationships in ways that the traditional swapping of the 20 common amino acids cannot.
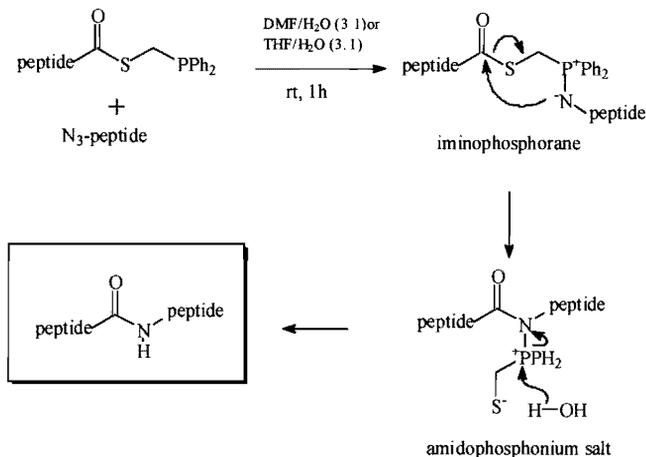
3. Staudinger Ligation

A limitation of native chemical ligation is its intrinsic reliance on having a cysteine or selenocysteine residue at the ligation junction. Cysteine is uncommon, comprising <2% of all residues in proteins. Selenocysteine is even more rare. Modern peptide synthesis is typically limited to peptides of ≤50 residues. Hence, the vast majority of proteins cannot be prepared by any method that allows for peptides to be coupled only at cysteine or selenocysteine residues.

The removal of the cysteine limitation by the development of a more general ligation

technology would expand greatly the utility of total protein synthesis. We believe that the Staudinger reaction provides the basis for such an alternative. In the Staudinger reaction, a phosphine is used to reduce an azide to an amine: $PR_3 + N_3R' + H_2O \rightarrow O{=}PR_3 + H_2NR' + N_2(g)$ [13,14]. This reaction occurs via a stable intermediate, an iminophosphorane ($R_3P^+{-}^-NR'$), which has a nucleophilic nitrogen. This nitrogen can be acylated, both in intermolecular and intramolecular reactions [15,16]. Hydrolysis of the resulting amidophosphonium salt gives an amide and phosphine oxide.

To apply the Staudinger reaction to peptide synthesis, we use a phosphinothiol to unite a thioester and azide. A putative mechanism for our version of the "Staudinger ligation" is shown in Scheme 3 [17,18]. The ligation begins by transthioesterification with the phosphinothiol. Coupling of the resulting phosphinothioester with a peptide azide leads to the formation of the reactive iminophosphorane. Intramolecular attack of the iminophosphorane nitrogen on the thioester leads to an amidophosphonium salt. Hydrolysis of the amidophosphonium salt produces the desired amide. Significantly, no atoms from the phosphinothiol remain in the amide product.

Scheme 3

## 3.1. CHOICE OF PHOSPHINOTHIOL

We have analyzed a series of phosphinothiol compounds for their ability to mediate the Staudinger ligation (Scheme 4). First, we synthesized phosphinothiol 1, and coupled it to several *N*-acetyl amino acids as a thioester. Upon incubation with N3GlyNHBn, these thioesters reacted to form a dipeptide, albeit in low isolated yields (≤35%) [17]. Still, phosphinothiol 1 provided us with proof-of-principle for the Staudinger ligation of peptides.

Phosphinothiol 2 was our second generation phosphinothiol [18]. Staudinger ligation with phosphinothiol 1 occurs through an intermediate with a six-membered ring. We reasoned that reducing the size of this ring to five could improve the ligation yield. To access a transition state with a smaller ring, we replaced the *o*-phenyl group of phosphinothiol 1 with a single methylene group. Thioesters of 2 derived from AcOH, AcGlyOH, and AcPheOH were prepared either by transthioesterification or by coupling with

dicyclohexylcarbodiimide. These amino acid thioesters with phosphinothiol **2** formed dipeptides with N₃GlyNHBn in isolated yields of >90% [18] (Table I).

Phosphinothiols **3** and **4** were also screened for their ability to effect the Staudinger ligation when incorporated into amino acid thioesters. We found thioesters of **3** gave dipeptide yields with N₃GlyNHBn that were comparable to thioesters of **2**, although the reaction rates were substantially slower. Phosphinothiol **4** and its thioesters are extremely difficult to handle to due its high propensity towards oxidation in air.

To date, phosphinothiol **2** is the most efficient reagent for effecting the Staudinger ligation. The couplings that we have performed using this reagent are summarized in Table I. All of these ligation reactions occur in nearly quantitative yield.

1	2	3	4
low yield	high yield	slow rate	unstable
	fast rate		(easily oxidized)
	stable		
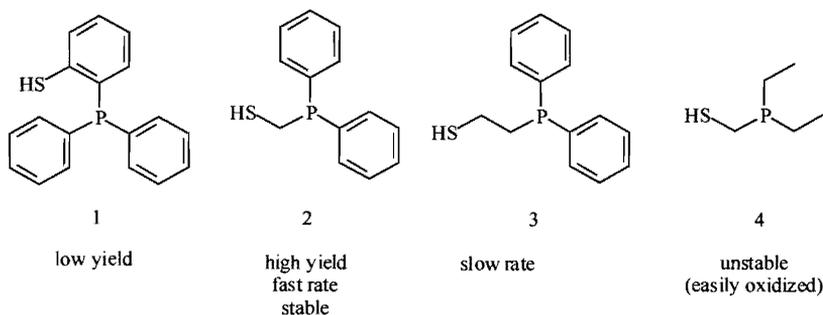
Scheme 4

3.2. RETENTION OF STEREOCHEMISTRY

All natural α-amino acids except glycine have a stereogenic center at their α-carbon. To be an effective tool for the total chemical synthesis of proteins, a peptide ligation reaction must proceed without epimerization. The coupling of thioesters in native chemical ligation, which like the Staudinger ligation involves transthioesterification followed by an S- to N-acyl shift [5], is known to proceed without detectable racemization [19]. It therefore remained important to assess the propensity for epimerization of the α-azido acid during the Staudinger ligation. This issue is of concern because an α-carbanion of both an azide and an iminophosphorane can be stabilized by both inductive and resonance effects.

To make this assessment, we used the Staudinger ligation to couple D- and L-α-azido acids [20]. The chirality of the product dipeptides was analyzed by HPLC using a D-phenylglycine chiral column. The chromatographic conditions enabled the baseline resolution of the two possible enantiomeric products. The reaction of the D epimer produced no evidence of product containing the L epimer, and *vice versa*. Thus, the Staudinger ligation proceeds without detectable epimerization of the α-carbon of the azido acid.

TABLE 1. Staudinger ligation of thioesters of phosphinothiol 2 with α-azido acids [18,20]

Phosphinothioester	Azide	Peptide	Isolated Yield (%)	
			L	D
AcS⁀PPh₂	N₃⁀C(O)⁀N(H)⁀Ph	AcGlyNHBn	91	
AcGlyS⁀PPh₂	N₃⁀C(O)⁀N(H)⁀Ph	AcGlyGlyNHBn	80 a	
AcPheS⁀PPh₂	N₃⁀C(O)⁀N(H)⁀Ph	AcPheGlyNHBn	92	
AcGlyS⁀PPh₂	Ph CH₂ N₃⁀C(O)⁀N(H)⁀Ph (L and D)	AcGlyPheNHBn	90	93
AcGlyS⁀PPh₂	MeO₂C CH₂ N₃⁀C(O)⁀N(H)⁀Ph (L and D)	AcGlyAsp(OMe)NHBn	91	95
AcGlyS⁀PPh₂	PhCH₂O CH₂ N₃⁀C(O)⁀N(H)⁀Ph (L and D)	AcGlySer(Bn)NHBn	92	99

a Isolated by recrystallization. Other peptides were isolated by chromatography on silica gel.

3.3. SYNTHESIS OF PHOSPHINOTHIOL 2

The original synthetic route to phosphinothiol 2, is shown in Scheme 5 [18]. Phenylmagnesium bromide is added to chloromethylphosphonic dichloride (5), and the resulting Grignard reaction refluxed for 12 h to give a diphenyl phosphine oxide. A mixture of this diphenyl phosphine oxide with thioacetic acid and triethylamine in dry tetrahydrofuran (THF) is heated at reflux for 12 h. After purification by flash chromatography and treatment with decolorizing charcoal, thiophosphine oxide 6 is isolated in a 54% combined yield for the two steps. An excess of trichlorosilane in chloroform for 72 h is used to reduce thiophosphine oxide 6 to phosphinothioester 7, which is isolated by flash chromatography in nearly quantitative yield. Hydrolysis of the phosphinothioester 7 with sodium hydroxide in methanol for 2 h gives phosphinothiol 2. During this reaction, Ar(g) is bubbled through the reaction mixture to prevent oxidation of the resultant thiol. Phosphinothiol 2 is purified by chromatography on alumina and isolated in 94% yield. The overall yield for the route in Scheme 5 is 50%.
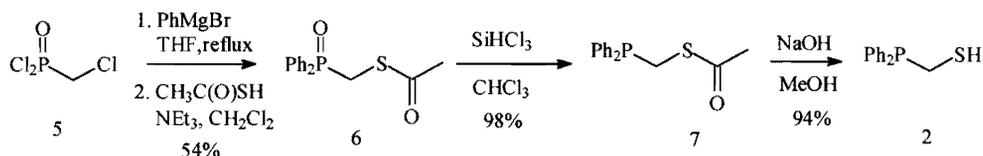
Scheme 5

We have since developed an improved synthetic route that uses air-stable borane protection of the phosphine (Scheme 6) [20]. This route is based on the easily prepared alkylating agent **8** [21] and the commercially available borane–diphenylphosphine complex. Borane diphenylphosphine is deprotonated by sodium hydride in dimethyl formamide (DMF) followed by addition of **8** to give borane-complex **9** (86% yield). Complex **9** is stable to air and moisture, and can be stored at room temperature under air for months without any sign of oxidation or other decomposition. The borane complex is disrupted by mild heating with 1,4-diazabicyclo[2.2.2]octane (DABCO) in toluene for 4 h (95% yield). The protecting group of the resulting acyl phosphinothiol **7** is removed as described above to give phosphinothiol **2** (94% yield). The overall yield for the route in Scheme 6 is 74%.
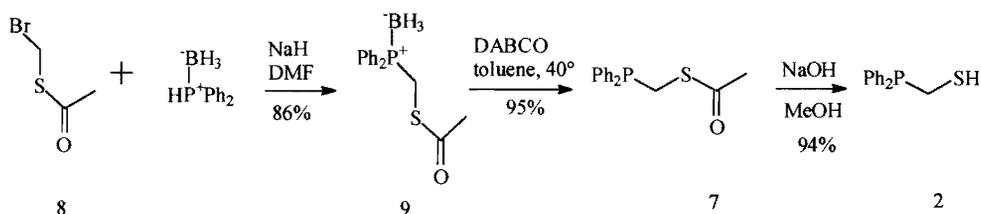
Scheme 6

3.4. THERMODYNAMICS AND KINETICS

The thermodynamics and kinetics of the Staudinger ligation highlight its exceptional promise for effecting amide bond formation between peptide fragments. Bond energies reveal that $\Delta H < -100$ kcal/mol for the overall reaction, due largely to the great strength of the N≡N and O=P bonds in the products. This decrease in enthalpy is much greater than that for amide bond formation by more typical acyl transfer reactions.

We have used ^{13}C NMR spectroscopy to observe the kinetics of the Staudinger ligation. These experiments involve the reaction of N_3GlyNHBn labeled at the α-carbon with ^{13}C and AcGlySCH$_2$PPh$_2$ in DMF-d_7. Significantly, intermediates do not accumulate, indicating that the rate-limiting step is the association of the phospinothioester with the azido acid. The starting material disappears and the product appears with a half-life of approximately 7 minutes at room temperature. This rapid rate makes the Staudinger ligation of peptides practical for the iterative assembly of proteins from component peptides.

3.5. PROTEIN SEMISYNTHESIS WITH THE STAUDINGER LIGATION

Having demonstrated the efficient use of HSCH$_2$PPh$_2$ (**2**) in mediating the Staudinger ligation, we next sought to exploit this new synthetic methodology. The Staudinger ligation

of protected peptide fragments on a solid support represents an orthogonal amide-bond forming technique to existing methods. We once again chose RNase A as a model system for semisynthesis to demonstrate a potential advantage of this orthogonality.

TABLE 2. Coupling of peptides by orthoganol methods to form ribonuclease A

Bond	Coupling Method
1–109	mRNA Translation by Ribosome (sequential)
109–110	Expressed Protein Ligation (convergent)
110–111	Solid-Phase Peptide Synthesis (sequential)
111–112	Staudinger Ligation (convergent)
112–124	Solid-Phase Peptide Synthesis (sequential)

We envisioned a semisynthesis of RNase A that involved assembly of the protein by a variety of amide-bond forming reactions. These reactions included mRNA translation by the ribosome, expressed protein ligation, solid-phase peptide synthesis, and the Staudinger ligation (Table II). Specifically, fragment 1–109 was prepared as a C-terminal thioester by rDNA techniques. The amide bonds in this segment were formed by mRNA translation by the ribosome. Fragment 110–124 was prepared by two methods. The amide bonds between residues 110–111 and 112–124 were formed by standard solid-phase peptide synthesis using activation with HATU. The protected fragment 110–111 was elaborated as a phosphinothioester of phosphinothiol 2. The protected fragment 112–124 was elaborated as an N-terminal azide while immobilized to resin. Soluble fragment 110–111 was coupled with immobilized fragment 112–124 by using the Staudinger ligation directly on the solid support. Finally, the resulting fragment 110–124 was liberated from the resin and deprotected. Fragment 110–124, in which residue 110 is a cysteine, was coupled to fragment 1–109 by using expressed protein ligation to give semisynthetic RNase A. The value of k_{cat}/K_m for this semisynthetic RNase A was 0.94 X 10^7 $M^{-1}s^{-1}$, which is nearly identical to the value of 1.1 X 10^7 $M^{-1}s^{-1}$ obtained for biosynthetic RNase A. The identity of the two enzymes was also verified by mass spectrometry. We believe that a judicious combination of the methods in Table II holds great promise for obtaining access to semisynthetic proteins.

3.6. PROTEIN ASSEMBLY WITH THE STAUDINGER LIGATION

Another promising application for the Staudinger ligation is the total chemical synthesis of a protein on the solid phase. In our semisynthesis of RNase A, we demonstrated the ability to couple protected peptide fragments on a solid support by using the Staudinger ligation. Likewise, the reaction can be used to assemble entire proteins (Scheme 7). One convergent strategy is to divide a target protein into fragments of 20–30 residues. The ultimate C-terminal fragment, while still immobilized to a solid support, would be capped with an α-azido acid. The remaining fragments would be synthesized as protected peptides and elaborated as C-terminal phosphinothioesters. These peptidyl thioesters would be coupled to the growing peptide chain on the solid support. After each ligation, the nascent peptide would be capped with an α-azido acid to prepare for the next ligation. When the protein

synthesis was complete, the protein would be deprotected and folded while still attached to the resin to avoid aggregation. The folded protein could be left attached to the resin by its *C*-terminus for high-throughput assays or liberated for detailed structure–function analyses. This entire process would be amenable to automation.
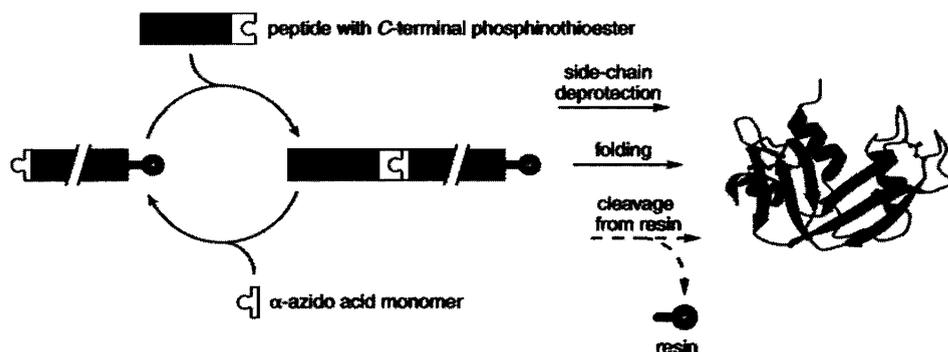
Scheme 7

4. Envoi

On December 12, 1902, Emil Fischer delivered his Nobel Prize lecture in Stockholm, Sweden, saying in part [22]:

Of the chemical aids in the living organism the ferments—mostly referred to nowadays as enzymes—are so pre-eminent that they may justifiably be claimed to be involved in most of the chemical transformations in the living cell. The examination of the synthetic glucosides has shown that the action of the enzymes depends to a large extent on the geometrical structure of the molecule to be attacked, that the two must match like lock and key. Consequently, with their aid, the organism is capable of performing highly specific chemical transformations which can never be accomplished with the customary agents. To equal Nature here, the same means have to be applied, and I therefore foresee the day when physiological chemistry will not only make extensive use of the natural enzymes as agents, but when it will also prepare synthetic ferments for its purposes.

A century later, Fischer's vision is becoming reality. Enzymes and other proteins are not only accessible targets for synthetic chemistry, but are poised to become the dominant targets of the 21st century. We believe that the Staudinger ligation has the potential to facilitate this effort (Scheme 7).

In addition to playing a major role in protein synthesis and semisynthesis, the Staudinger ligation could be used in conjunction with natural thioesters. For example, both the biosynthesis of polyketides and the nonribosomal biosynthesis of peptides proceed by the elaboration of thioester intermediates [23]. Interception of these intermediates with a phosphinothiol would allow for Staudinger ligation to an azide. The ligation of a

biosynthetic library of thioesters with a chemical library of azides could be a facile means to access molecular diversity.

Finally, the Staudinger ligation could also be useful for the immobilization of proteins to a solid support. Chemical synthesis or semisynthesis could be used to produce a protein that contains a single phosphinothioester or azide. The Staudinger ligation of such a protein to a surface could result in the display of proteins in a uniform manner for a variety biotechnical applications, including the creation of protein chips.

5. References

1. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.
2. Venter, C.J., *et al.* (2001) The sequence of the human genome, *Science* **291**, 1304–1351.
3. Dawson, P.E. and Kent, S.B.H. (2000) Synthesis of native proteins by chemical ligation, *Annu. Rev. Biochem.* **69**, 923–960.
4. Wieland, T., Bokelmann, E., Bauer, L., Lang, H.U., Lau, H. and Schafer, W. (1953) Polypeptide syntheses. VIII. Formation of sulfur containing peptides by the intramolecular migration of aminoacyl groups, *Liebigs Ann. Chem.* **583**, 129–149.
5. Dawson, P.E., Muir, T.W., Clark-Lewis, I. and Kent, S.B.H. (1994) Synthesis of proteins by native chemical ligation, *Science* **266**, 776–779.
6. Muir, T.W., Sondhi, D. and Cole, P.A. (1998) Expressed protein ligation: A general method for protein engineering, *Proc. Natl. Acad. Sci., USA* **95**, 6705–6710.
7. Evans, T.C., Benner, J. and Xu, M.Q. (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element, *Protein Sci.* **7**, 2256–2264.
8. Hondal, R.J., Nilsson, B.L. and Raines, R.T. (2001) Selenocysteine in native chemical ligation and expressed protein ligation, *J. Am. Chem. Soc.* **123**, 5140–5141.
9. Hondal, R.J. and Raines, R.T. (2002) Semisynthesis of proteins containing selenocysteine, *Methods Enzymol.* **347**, 70–83.
10. Gieselman, M.D., Xie, L. and van der Donk, W.A. (2001) Synthesis of a selenocysteine-containing peptide by native chemical ligation, *Org. Lett.* **3**, 1331–1334.
11. Quaderer, R., Sewing, A. and Hilvert, D. (2001) Selenocysteine-mediated native chemical ligation, *Helvetica Chim. Acta* **84**, 1197–1206.
12. Arnold, U., Hinderaker, M.P., Nilsson, B.L., Huck, B.R., Gellman, S.H. and Raines, R.T. (2002) Protein prosthesis: A semisynthetic enzyme with a □-peptide reverse turn, *J. Am. Chem. Soc.* **124**, 8522–8523.
13. Staudinger, H. and Meyer, J. (1919) New organic compounds of phosphorus. III. Phosphinemethylene derivatives and phosphinimines, *Helvetica Chim. Acta* **2**, 635–46.
14. Staudinger, H. (1961) *Arbeitserinnerungen*, Dr. Alfred Hüthig Verlag GmbH., Heidelberg, Germany, pp. 30–35.
15. Bosch, I., Romea, P., Urpí, F. and Vilarrasa, J. (1993) Alternative procedures for the macrolactamisation of □-azido acids, *Tetrahedron Lett.* **34**, 4671–4674.
16. Saxon, E. and Bertozzi, C.R. (2000) Designing chemical reactions for a cellular environment: Cell surface engineering via a modified Staudinger reaction, *Science* **287**, 2007–2010.
17. Nilsson, B.L., Kiessling, L.L. and Raines, R.T. (2000) Staudinger ligation: A peptide from a thioester and azide, *Org. Lett.* **2**, 1939–1941

18. Nilsson, B.L., Kiessling, L.L. and Raines, R.T. (2001) High-yielding Staudinger ligation of a phosphinothioester and azide to form a peptide, *Org. Lett.* **3**, 9–12.

19. Lu, W.Y., Qasim, M.A. and Kent, S.B.H. (1996) Comparative total syntheses of turkey ovomucoid third domain by both stepwise solid phase peptide synthesis and native chemical ligation, *J. Am. Chem. Soc.* **118**, 8518–8523.

20. Soellner, M.B., Nilsson, B.L. and Raines, R.T. (2002) Staudinger ligation of α-azido acids retains stereochemistry, *J. Org. Chem.* **67**, 4993–4996.

21. Farrington, G.K., Kumar, A. and Wedler, F.C. (1989) A convenient synthesis of diethyl (mercaptomethyl)phosphonate, *Org. Prep. Proced. Int.* **21**, 390–392.

22. http://www.nobel.se

23. Cane, D.E., Walsh, C.T. and Khosla, C. (1998) Harnessing the biosynthetic code: Combinations, permutations, and mutations, *Science* **282**, 63–68.