

# SPLICING OF UNNATURAL AMINO ACIDS INTO PROTEINS: A PEPTIDE MODEL STUDY

Shoufa Han<sup>†</sup> and Ronald E. Viola\*

University of Toledo, Department of Chemistry, Toledo, Ohio 43606, USA

<sup>†</sup> Present address: Scripps Research Institute, Department of Molecular Biology, La Jolla, CA 92037

**Abstract:** *S*-Ethyl 2-azidohexanethioate ( $N_3$ -*Hex*-SEt), an unnatural amino acid analog of leucine, is coupled with L-cysteine ethyl ester ( $NH_2$ -Cys-OEt) to obtain  $N_3$ -*Hex*-Cys-OEt by native chemical ligation. Coupling of this dipeptide with *N*-*t*-butoxycarbonyl-2-diphenylphosphinoethanethioglycinate produces the tripeptide, *t*-Boc-Gly-*Hex*-Cys-OEt, in high yield. These reactions suggest an approach for the incorporation of unnatural amino acids into proteins by successive native chemical ligation and Staudinger ligation.

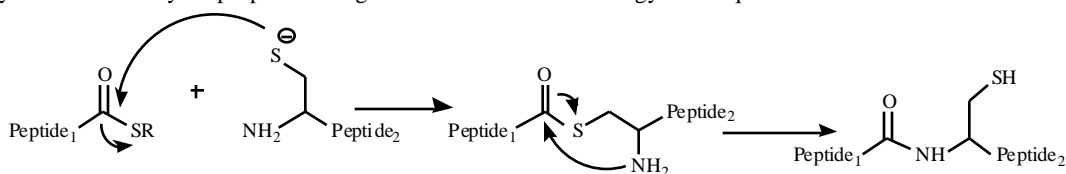
**Key Word:** phosphinothiol, peptide coupling, unnatural amino acid incorporation, Staudinger ligation, native chemical ligation

## INTRODUCTION

Incorporation of unnatural amino acids into proteins can greatly expand the scope of site-directed mutagenesis, and could provide a method for systematic structure-function relationship studies. Among the approaches that have been used to introduce unnatural amino acids into proteins, the nonsense suppression methodology is potentially the most powerful. This method involves conversion of the codon of a target amino acid into the amber codon, TAG [1], that is recognized by an engineered suppressor tRNA [2]. This suppressor tRNA is selectively acylated with the target unnatural amino acid either chemically or by an engineered tRNA-synthetase [3-6]. This method has recently culminated in the generation of a bacterium with a 21 amino acid genetic code [7], providing a powerful tool for studying proteins in their native environment. The major shortcoming of this approach is the complex sequence of steps required to introduce a single unnatural amino acid. To incorporate another unnatural amino acid by this *in vivo* method the corresponding biosynthetic pathway of the unnatural amino acid and a different orthogonal

tRNA and tRNA-synthetase pair must be engineered and incorporated into the bacterium. Considering the diversity of unnatural amino acids that are potentially available and the complexity of the expanded genetic code method, additional means are needed to incorporate unnatural amino acids into proteins using a simple and straightforward methodology.

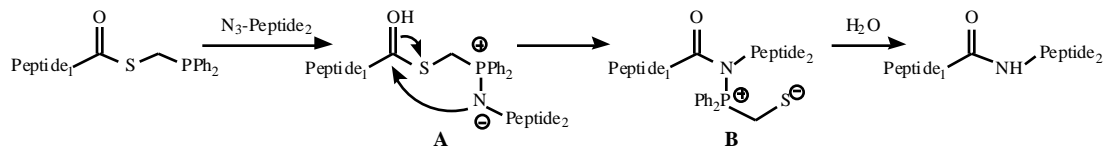
Native chemical ligation can selectively couple a unprotected peptide containing a C-terminal thioester with another unprotected peptide that has an N-terminal cysteine in aqueous solution [8,9]. The thiolate of the N-terminal cysteine residue attacks the carbonyl carbon of the C-terminal thioester, followed by a rapid intramolecular S–N acyl transfer to produce an amide bond between the two peptides (Scheme 1). Recombinantly expressed proteins or protein fragments with either a C-terminal thioester or an N-terminal cysteine can easily be prepared using available molecular biology techniques.



**Scheme 1.** Peptide coupling by native chemical ligation.

Expressed protein ligation was developed as an extension of native chemical ligation to couple synthetic peptides with expressed protein segments [10-14], taking advantage of the virtually unlimited size of these biosynthesized peptides, thereby greatly facilitating the convergent chemical synthesis of proteins [13,14].

The Staudinger reaction is an alternative method for amide bond formation that involves the reduction of an azide with a phosphine *via* a stable aza-ylide intermediate containing a nucleophilic nitrogen [15-17]. Intramolecular acylation of this nucleophile affords an amido-phosphonium salt, which can be hydrolyzed to produce an amide bond and release phosphine oxide [18-21]. Recently, the Staudinger reaction has been applied to the synthesis of peptides through the coupling of phosphinothioesters and an azide [20,21]. Attack of the phosphine on the azide generates an aza-ylide intermediate (Scheme 2, compound A), followed by internal attack of the nucleophilic nitrogen on the carbonyl carbon of the thioester. Hydrolysis of the rearranged adduct (Scheme 2, compound B) in aqueous solution liberates phosphine oxide and leaves the peptides coupled by an amide bond.



**Scheme 2.** Peptide ligation by the Staudinger reaction.

We have begun to explore a combination of these two ligation chemistries with the goal of selectively splicing any synthetic amino acid or peptide analogs into a defined position between two

expressed protein fragments. Here we report the successive application of these ligation reactions in a peptide model system.

## MATERIALS AND METHODS

$^1\text{H}$ - and  $^{13}\text{C}$ -NMR spectra were obtained for all synthetic intermediates and products on a Varian VXR400. Mass spectra (ESI) were obtained on Bruker Esquire-LC. 2-Chloroethyldiphenylphosphine was obtained from Organometallics. All other reagents were purchased from either Aldrich or Lancaster. Synthetic reactions were conducted in oven-dried glassware under argon and all solvents were distilled prior to use. Flash column chromatography was performed on 230-400 mesh silica gel from Lancaster under positive argon pressure. Thin layer chromatography (TLC) was performed on Kodak 13181 plates with fluorescence indicator. Compounds were visualized by iodine chamber or UV light.

### ***S*-Ethyl 2-azidohexanethioate (I).**

A mixture of 2-bromohexanoic acid (25.0 g, 0.13 mmol), sodium azide (20.0 g, 0.31 mmol) and 18-crown-6 (1.0 g, 3.8 mmol) in DMF (100 mL) was stirred at room temperature for 24 h, and then filtered. The organic solution, acidified to pH 2, was extracted extensively with ethyl ether and water. The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and then concentrated to afford 2-azidohexanoic acid as a brownish liquid (20.0 g, 0.10 mol) in 78 % yield. A mixture of ethanethiol (5.0 g, 0.83 mol), 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (7.5g, 39.3 mmol), the 2-azidohexanoic acid (6.1g, 38.9 mmol), and triethylamine (2 mL) in THF (150 mL) was stirred overnight, and then filtered. The organic solution was concentrated and the residue was dissolved in ethyl ether (200 mL). The organic solution was washed with acid (1.0 M HCl), base (1.0 M NaOH), water, and brine. The organic layer was dried over  $\text{Na}_2\text{SO}_4$ , and then concentrated to give a pale yellow solid (6.5 g, 32.7 mmol, 84 % yield).

### ***N*<sub>3</sub>-Hex-Cys-OEt (II).**

L-Cysteine ethyl ester (2.92 g, 15.8 mmol) and *S*-ethyl 2-azidohexanethioate (I) (3.16 g, 15.8 mmol) were dissolved in THF and water (3:1, 40 mL), pH 8.0, and stirred at room temperature overnight. THF was removed under reduced pressure and the water layer was extracted with ethyl ether (150 mL). The ethyl ether solution was dried over  $\text{Na}_2\text{SO}_4$  and concentrated to afford a pale yellow solid (5.0 g, 15.4 mmol) in 97% yield.

### **2-(Diphenylphosphino)ethanethiol.**

To a solution of 2-chloroethyldiphenylphosphine (3.0 g, 13.0 mmol) in THF (40 mL) at  $-78\text{ }^\circ\text{C}$  was added hexamethyldisilathiane (3.0 g, 17.0 mmol), followed by dropwise addition 15 mL of 1.0 M tetrabutylammonium fluoride (TBAF) in THF. The reaction mixture was allowed to warm to room temperature and stir overnight. The solution was concentrated and the residue was dissolved in ethyl ether (200 mL) and then washed with saturated aqueous ammonium chloride, water, and brine. Evaporation of the solvent affords a pale yellow liquid. Further purification by column chromatography (10:1 hexanes:ethyl acetate,  $R_f = 0.67$ ) gives the desired product (2.8 g, 12 mmol) as a colorless oil in 90% yield.

***N*-*t*-Butoxycarbonyl-*S*-(2-diphenylphosphinoethane)thioglycinate (III).**

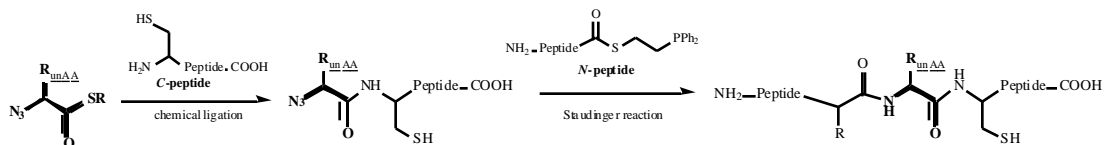
A mixture of 2-(diphenyl-phosphino)ethanethiol (1.5 g, 6.5 mmol), 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide (EDC) hydrochloride (1.3 g, 6.8 mmol), *N*-*t*-butoxycarbonylglycine (1.2g, 6.5 mmol), and triethylamine (0.2 mL) in ethyl ether (150 mL) was stirred overnight. The mixture was washed successively with acid (1.0 M HCl), base (1.0 M NaOH), water, and brine. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, concentrated, and purified by chromatography (10:1 petroleum ether:ethyl acetate, R<sub>f</sub> = 0.15) to give the desired product as a white solid (2.2 g, 5.2 mmol, 80% yield).

**Boc-Gly-Hex-Cys-OEt (IV).**

The *N*-*t*-butoxycarbonyl-*S*-(2-diphenylphosphinoethane)thioglycinate (III) (0.30 g, 0.74 mmol) and the dipeptide (II) (0.25g, 0.77 mmol) were dissolved in THF/H<sub>2</sub>O (9:1, 10 mL) and stirred at room temperature for 6 hr. The coupling reaction was completed in quantitative yield as monitored by TLC, followed by removal of the solvent under reduced pressure. Production of the tripeptide was verified by mass spectral analysis.

**RESULTS AND DISCUSSION**

The combination of the Staudinger ligation and expressed protein ligation has the potential to provide a simple and selective method to replace a target amino acid at any position in a protein with a series of unnatural amino acid analogs. The implementation of this strategy involves preparation of three components: an *N*-terminal protein fragment expressed from the amino terminus to the X-1 position with a carboxyl phosphinothioester, a *C*-terminal protein fragment expressed from the X+1 position to the carboxyl terminus with conversion of the X+1 amino acid into cysteine, and target unnatural amino acid precursors synthesized with a  $\alpha$ -azide and a  $\beta$ -carboxyl thioester. Native chemical ligation of the unnatural amino acid precursor with the cysteine-containing protein fragment results in peptides each with the unnatural amino acid coupled to the cysteine. The azido group of the unnatural amino acid is then used to couple with the carboxyl phosphinothioester of the *N*-terminal protein fragment *via* Staudinger ligation. The product of each coupling reaction is the complete protein with an unnatural amino acid spliced into the desired position (Scheme 3).



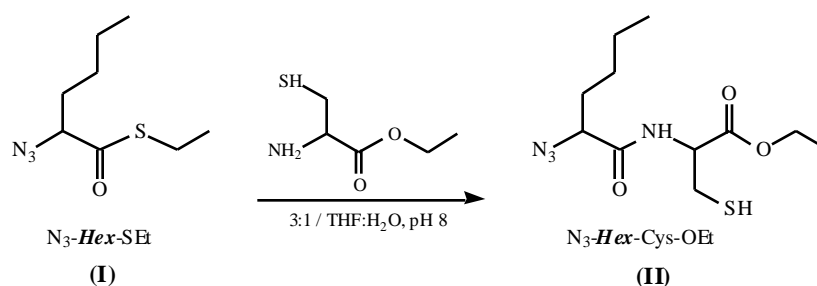
**Scheme 3.** Proposed splicing of unnatural amino acids (R<sub>unAA</sub>) by chemical ligation and Staudinger ligation.

We have examined a model system to test the chemistry required for the splicing of unnatural amino acids between protein fragments by successive native chemical ligation and Staudinger ligation. To test these coupling reactions *S*-ethyl 2-azidohexanethioate was synthesized as an activated straight chain

analog of leucine. This amino acid analog is first coupled with NH<sub>2</sub>-Cys-OEt, a mimic of a C-terminal protein fragment containing a terminal cysteine. The resulting dipeptide is then coupled by Staudinger ligation with an activated mimic of an N-terminal protein fragment containing a terminal thioester.

#### Coupling by Native Chemical Ligation.

To splice unnatural amino acids into proteins by these orthogonal ligations, the first step is to ligate the unnatural amino acid analogs to a protein fragment containing an N-terminal cysteine (the C-peptide) by native chemical ligation. S-Ethyl 2-azidohexanethioate (N<sub>3</sub>-Hex-SEt, **I**) was used as the unnatural amino acid precursor, and is smoothly coupled with NH<sub>2</sub>-Cys-OEt by native chemical ligation affording the coupled product, N<sub>3</sub>-Hex-Cys-OEt (**II**), in quantitative yield (Scheme 4).



**Scheme 4:** Synthesis of a dipeptide by native chemical ligation.

This chemical ligation attaches the unnatural amino acid analog to the cysteine by formation of an amide bond, thus extending this C-peptide model by one amino acid, and introducing an N-terminal azide group that is stable enough to survive this coupling step and can then be used for the subsequent Staudinger reaction.

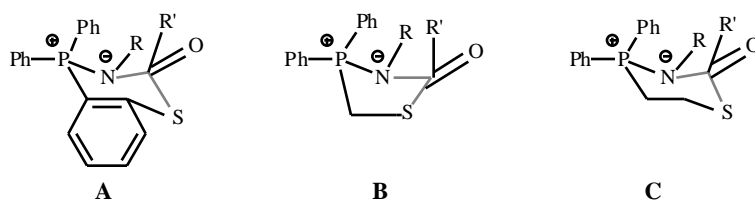
#### Coupling by Staudinger ligation.

The carboxyl group of the N-peptide model must be activated through attachment of a phosphinothiol to provide a site for initial coupling with the azide *via* P-N bond formation. This aza-ylide intermediate (Scheme 2, compound A) of Staudinger ligation can be either hydrolyzed to cleave the P-N bond and release an amine, or proceed through an intramolecular acylation (Scheme 2, compound B) to yield an amide bond upon hydrolysis. To achieve peptide coupling in high-yield in aqueous medium the nucleophilic nitrogen of aza-ylide intermediate must attack the intramolecular carbonyl carbon in a kinetically favorable manner to overwhelm the undesirable hydrolysis pathway. Two types of phosphinoesters have been reported that promote this rearrangement in peptide ligation [20,21], but the reported synthetic procedures for these phosphinothioesters require multiple steps, resulting in low overall yields [20-23]. The use of 2-(diphenylphosphino)benzenethioesters (Fig. 1A) in these reactions give yields that are too low for practical application [20], while coupling of diphenylphosphinomethylthioesters (Fig. 1B) with azides affords the peptide products in very high yield [21]. The higher coupling yields promoted

by phosphinomethanethiols have been ascribed to the reduced ring size of its five-membered ring transition state that can bring the nucleophilic imide nitrogen proximal to the electrophilic carbonyl carbon, relative to its position in the 6-membered ring transition state of phosphinobenzenethiols (Fig. 1) [21]. However, the low yields with benzenethiols in peptide synthesis may also be a consequence of the fusion of the benzene ring to the six-membered transition state, leading to higher energy barrier for its formation. If this proposal is correct then removal of benzene ring with retention of the six-membered transition state (Fig. 1C) should result in greatly enhanced yields.

#### A modified phosphinothiol for Staudinger ligation.

To test our hypothesis, a slightly modified phosphinothiol, 2-(diphenylphosphino)ethanethiol, was prepared and examined for its ability to promote facile synthesis and high peptide coupling yields.



**Figure 1.** Proposed transition states derived from phosphinothioesters used in peptide coupling: 2-(diphenylphosphino)benzenethiol (A), diphenylphosphinomethanethiol (B), 2-(diphenylphosphino)ethanethiol (C).

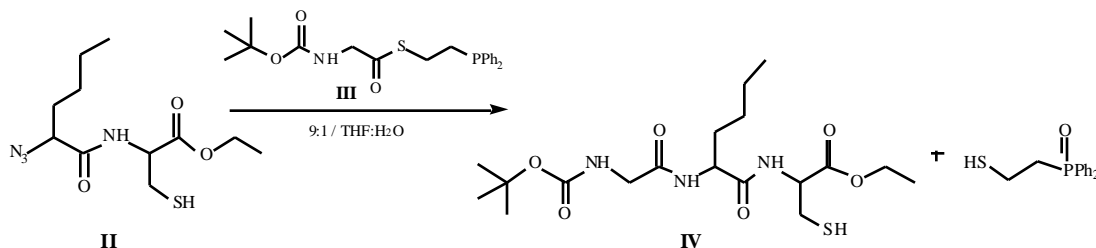
2-(Diphenylphosphino)ethanethiol is available commercially, but is fairly expensive. An economic, high-yield, one-pot synthesis of this phosphinothiol was achieved from commercially available 2-chloroethyldiphenylphosphine by a trimethylsilylthioxy-dehalogenation reaction [24]. *t*-Butylammonium trimethylsilylthiolate ( $\text{Me}_3\text{SiS}^-\text{Bu}_4\text{N}^+$ ), generated *in situ* by the addition of tetrabutylammonium fluoride (TBAF) to hexamethyldisilane, acts as the nucleophile to react with alkyl chloride to yield the masked thiol functionality. Upon aqueous workup the desired product is obtained in 90% yield. The synthetic reaction conditions are quite mild, thereby minimizing possible air-oxidation of the starting material or the product.

The utility of this phosphinothiol reagent in Staudinger peptide ligation was examined by coupling *t*-butyl azidoacetate with *N*-*t*-butoxycarbonylglycine that has been activated as the diphenylphosphinoethane thioester (**III**). This test reaction proceeded quite rapidly and coupling was achieved in quantitative isolated yield of the protected gly-gly dipeptide. This facile peptide coupling with the phosphinoethanethiol reagent supports the proposal that removal of the benzene ring from the phosphinothiol reagent enhances the peptide coupling yield due to the ease of formation of this six-membered transition state.

#### Completion of the Splicing Process

The utility of this phosphinothiol reagent in Staudinger peptide ligation was examined by the coupling of  $\text{N}_3$ -*Hex*-Cys-OEt (**II**) with *N*-*t*-Butoxycarbonyl-*S*-(2-diphenylphosphinoethane)thioglycinate

(III). The azide functional group of **II** couples smoothly with **III**, a mimic of an *N*-terminal expressed protein fragment to yield the tripeptide, Boc-Gly-*Hex*-Cys-OEt (**IV**) (Scheme 5). The coupling reaction proceeds rapidly and the starting material is completely consumed during the reaction as judged by TLC. Formation of the tripeptide product was verified by MS analysis.



**Scheme 5.** Staudinger ligation to produce a tripeptide.

The Staudinger reaction was recently used to couple a synthetic dipeptide with an azido-containing peptide (representing amino acids 112-124 of ribonuclease A) by solid phase synthesis [25]. Following cleavage from the resin, side-chain deprotection, and HPLC purification, the coupled peptide was attached to the biosynthesized RNase A fragment (1-109) by native chemical ligation to produce a fully functional ribonuclease. Our model reaction shows that Staudinger coupling can be performed in the presence of an unprotected cysteine thiol, and that these reactions can be carried out in aqueous solution with absolute regioselectivity and chemoselectivity to produce the coupled peptides in high yield. These results clearly support the idea that unnatural amino acid analogs can be incorporated into proteins by the sequential application of native chemical ligation followed by Staudinger ligation. By using a series of unnatural amino acid precursors with systematically altered side chain properties in the coupling with protein fragments, a series of proteins could be obtained with systematically altered unnatural functionalities at any targeted position.

## ACKNOWLEDGEMENTS

This work was supported by a grant (MCB0196103) from the National Science Foundation.

## REFERENCE

- [1] Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., and Schultz, P.G. (1989) *Science*, 244, 182-188.
- [2] Bruce, A.G., Atkins, J.F., Wills, N., Uhlenbeck, O.C., and Gesteland, R.F. (1982) *Proc. Natl. Acad. Sci. USA*, 79, 7127-7131.
- [3] Ellman, J.; Mendel, D.; Anthony-Cahill, S.J.; Noren, C.J.; Schultz, P.G. (1991) in *Methods in Enzymology* (Langone, J.J., Ed). pp. 301-336. Academic Press, San Diego.
- [4] Ellman, J.A., Mendel, D., and Schultz, P.G. (1992) *Science*, 255, 197-200.
- [5] Mendel, D., Ellman, J.A., Chang, Z., Veenstra, D.L., Kollman, P.A., and Schultz, P.G. (1992) *Science*, 256, 1798-1802.

- [6] Judice, J.K., Gamble, T.R., Murphy, E.C., De Vos, A.M., and Schultz, P.G. (1993) *Science*, 261, 1578-1581.
- [7] Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., and Schultz, P.G. (2003) *J. Am. Chem. Soc.*, 125, 935-939.
- [8] Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B. (1994) *Science*, 266, 776-779.
- [9] Dawson, P.E. and Kent, S.B. (2000) *Annu. Rev. Biochem.*, 69, 923-960.
- [10] Hofmann, R.M. and Muir, T.W. (2002) *Curr. Opin. Biotechnol.*, 13, 297-303.
- [11] Cotton, G.J. and Muir, T.W. (1999) *Chemistry & Biology*, 6, R247-R256.
- [12] Muir, T.W. (2001) *Synlett* 733-740.
- [13] Valiyaveetil, F.I., MacKinnon, R., and Muir, T.W. (2002) *J. Am. Chem. Soc.*, 124, 9113-9120.
- [14] Cotton, G.J. and Muir, T.W. (2000) *Chemistry & Biology*, 7, 253-261.
- [15] Staudinger, H. and Meyer, J. (1919) *Helv. Chim. Acta*, 2, 635-646.
- [16] Gololobov, Y.G., Zhmurova, I.N., and Kasukhin, L.F. (1981) *Tetrahedron*, 37, 437-472.
- [17] Gololobov, Y.G. and Kasukhin, L.F. (1992) *Tetrahedron*, 48, 1353-1406.
- [18] Saxon, E. and Bertozzi, C.R. (2000) *Science*, 287, 2007-2010.
- [19] Saxon, E., Armstrong, J.I., and Bertozzi, C.R. (2000) *Organic Letters*, 2, 2141-2143.
- [20] Nilsson, B.L., Kiessling, L.L., and Raines, R.T. (2000) *Organic Letters*, 2, 1939-1941.
- [21] Nilsson, B.L., Kiessling, L.L., and Raines, R.T. (2001) *Organic Letters*, 3, 9-12.
- [22] Soellner, M.B., Nilsson, B.L., and Raines, R.T. (2002) *J. Org. Chem.*, 67, 4993-4996.
- [23] Block, E., Ofori-Okai, G., and Zubieta, J. (1989) *J. Am. Chem. Soc.*, 111, 2327-2329.
- [24] Hu, J. and Fox, M.A. (1999) *J. Org. Chem.*, 64, 4959-4961.
- [25] Nilsson, B.L., Hondal, R.J., Soellner, M.B., and Raines, R.T. (2003) *J. Am. Chem. Soc.*, 125, 5268-5269.

Received on \_\_\_\_\_, 2003, accepted on \_\_\_\_\_, 2003.