Extending the Applicability of Native Chemical Ligation

Lynne E. Canne,* Steven J. Bark, and Stephen B. H. Kent

Contribution from The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

Received February 7, 1996[⊗]

Abstract: A more general approach to native (amide-forming) chemical ligation of unprotected peptide segments is described that extends the technique beyond the previously reported X-Cys ligation site to now include X-Gly and Gly-X ligation sites. A peptide, $[\text{peptide}_1]^{\alpha}$COSR, is reacted with a second peptide, $HSCH_2CH_2(O)$-$N^{\alpha}[\text{peptide}_2]$, under conditions promoting thioester exchange. The intermediate thioester-linked product rearranges to form a ligation product linked by an N-substituted amide bond. If desired, the -oxyalkyl substitution on the amide bond can be removed by facile treatment with Zn in acidic medium to give a native peptide bond at the ligation site. The techniques described have been employed to ligate small model peptide segments to yield peptides with native or modified backbones, proving the feasibility of this approach.

Introduction

The application of chemical ligation to peptide synthesis is becoming increasingly popular.[1−8] Typically, this method involves the chemoselective reaction of unprotected peptide segments to give a product with an *unnatural* backbone structure at the site of ligation.[1,4,5] This revolutionary concept makes possible the synthesis of peptides of greater size than attainable by standard methods of peptide synthesis[8,9] as well as the synthesis of peptides of unusual structure and topology.[2,4,6,8,9] The chemical ligation method has allowed us to harness the power of stepwise solid phase peptide synthesis[10,11] to routinely make unprotected peptides of up to 60 amino acid residues in good yield and purity,[12] and to use this for the purpose of routine total chemical synthesis of proteins.

Recently, a chemical ligation technique was reported for the preparation of proteins with native backbone structure.[13] In this technique, an unprotected synthetic peptide bearing a C-terminal α-thioester is reacted in a chemoselective manner with an unprotected peptide containing an N-terminal Cys residue. Thiol exchange yields an initial thioester-linked intermediate which spontaneously rearranges to give a native amide bond at the ligation site joining the two peptide segments, in the process regenerating the Cys side chain thiol. This original version of native ligation uses chemistry first described by Wieland[14] for reacting amino acids. As originally described,[13] native ligation is restricted to joining peptide segments at an X−Cys bond.

Here we report important advances which extend this procedure, alleviating the necessity for an N-terminal Cys residue, and thus expanding the range of sites amenable to native chemical ligation to X-Gly and Gly-X as well as X-Cys. The method allows for the synthesis of peptides and proteins with either native or modified backbone structures. Scheme 1 outlines the procedure. The peptide-α-carboxythioester (**1**) reacts, via thiol exchange, with either an N^{α}(ethanethiol) peptide (**2a**) or an N^{α}(oxyethanethiol) peptide (**2b**) to produce the ligation product **3**. This thioester-linked intermediate rearranges through a favorable geometric arrangement involving a five- or six-membered ring to give the amide-linked product **4**, containing a secondary amide in **4a** or the analogous *N*-oxyalkyl compound in **4b**. The resulting N^{α}(substituted)amides also have potentially favorable solubility properties;[15] the N^{α}(oxyalkyl)-amide ligation product has the further advantage of being stable to HF cleavage conditions, yet is simply removed under mild conditions. In the case of **4b**, zinc dust can be added directly to the reversed phase HPLC-purified peptide in the acidic eluant to reduce the N−O bond of the *O*-alkylhydroxamate and give the native backbone structure of ligation product **5b**. Model studies were performed to explore the scope and limitations of this approach.

Results

All peptide segments were synthesized in stepwise fashion by established solid phase methods using *in situ* neutralization/2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) activation protocols for *tert*-butoxycarbonyl (Boc) chemistry, purified by preparative reversed phase HPLC, and characterized by electrospray mass spectrometry (ESMS).[12] Peptide-α-thioesters (**1**) were generated from the corresponding peptide-α-thioacids, which in turn, were synthesized on a thioester resin as previously described.[16] The peptide-α-thioacids were converted to either the corresponding 3-car

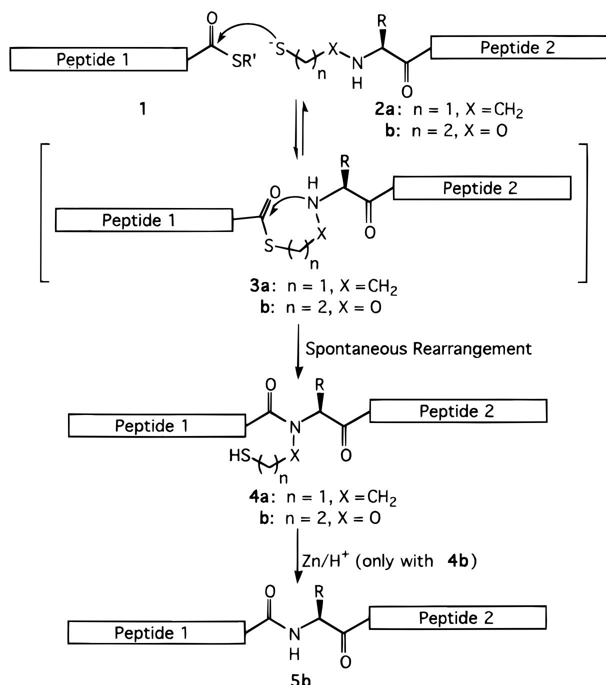
* To whom correspondence should be addressed. Present address: Gryphon Sciences, 250 E. Grand Ave., Suite 90, South San Francisco, CA 94080. Phone: (415) 952-8518. FAX: (415) 952-3055. E-mail: lynne@gryphonsci.com.

⊗ Abstract published in *Advance ACS Abstracts,* June 1, 1996.

(1) Schnölzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221−225.
(2) Dawson, P. E.; Kent, S. B. H. *J. Am. Chem. Soc*. **1993**, *115*, 7263−7266.
(3) Baca, M.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A*. **1993**, *90*, 11638−11642.
(4) Rose, K. *J. Am. Chem. Soc*. **1994**, *116*, 30−34.
(5) Liu, C. F.; Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A*. **1994**, *91*, 6584−6588.
(6) Muir, T. W.; Williams, M. J.; Ginsberg, M. H.; Kent, S. B. H. *Biochemistry* **1994**, *33*, 7701−7708.
(7) Williams, M. J.; Muir, T. W.; Ginsberg, M. H.; Kent, S. B. H. *J. Am. Chem. Soc*. **1994**, *116*, 10797−10798.
(8) Canne, L. E.; Ferré-D'Amaré, A. R.; Burley, S. K.; Kent, S. B. H. *J. Am. Chem. Soc*. **1995**, *117*, 2998−3007.
(9) Baca, M.; Muir, T. W.; Schnölzer, M.; Kent, S. B. H. *J. Am. Chem. Soc*. **1995**, *117*, 1881−1887.
(10) Merrifield, R. B. *J. Am. Chem. Soc*. **1963**, *85*, 2149−2154.
(11) Kent, S. B. H. *Annu. Rev. Biochem*. **1988**, *57*, 957−989.
(12) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Peptide Protein Res*. **1992**, *40*, 180−193.
(13) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779.
(14) Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H. *Liebigs Ann. Chem*. **1953**, *583*, 129−149.
(15) Quibell, M.; Packman, L. C.; Johnson, T. *J. Am. Chem. Soc*. **1995**, *117*, 11656−11668.
(16) Canne, L. E.; Walker, S. M.; Kent, S. B. H. *Tetrahedron Lett*. **1995**, *36*, 1217−1220.

S0002-7863(96)00398-8 CCC: $12.00 © 1996 American Chemical Society

Scheme 1. Generalized Native Chemical Ligation of Unprotected Peptide Segments



boxy-4-nitrophenyl thioesters in 6 M guanidine·HCl, 0.1 M sodium acetate, pH 5.0−6.5, by reaction with 1.5 equiv of 5,5′-dithiobis(2-nitrobenzoic acid)[17] or the corresponding benzyl ester in 6 M guanidine·HCl, 0.1 M sodium acetate, pH 4.0, using 10 equiv of benzyl bromide.[13] Both these thioesters provide satisfactory leaving groups for the ligation reactions, with the 3-carboxy-4-nitrophenyl thioesters demonstrating a somewhat faster reaction rate than the corresponding benzyl thioesters.[13]

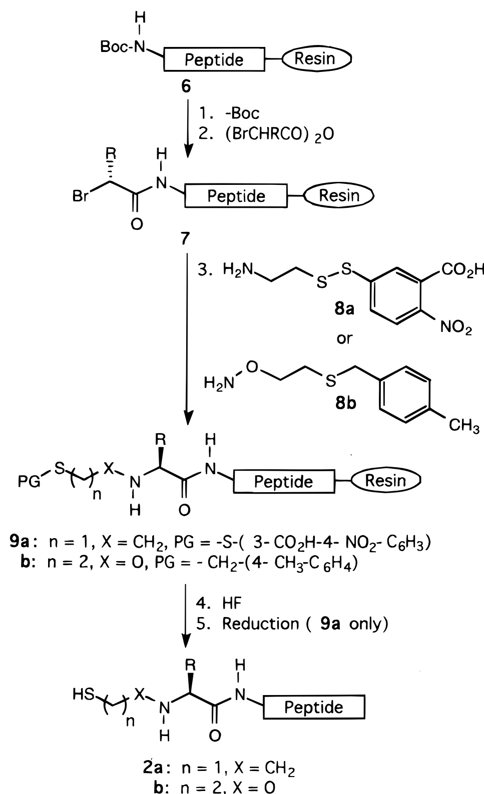
The syntheses of N^{α}(ethanethiol) (**2a**) and N^{α}(oxyethanethiol) (**2b**) peptides are outlined in Scheme 2. The appropriate α-bromocarboxylic acid, activated as the symmetric anhydride (0.5 equiv of 1,3-diisopropylcarbodiimide (DIC) in dichloromethane),[18] was coupled to the deprotected N-terminal amino acid of peptide-resin **6** to give bromoacyl-peptide-resin **7**. The bromide was then displaced, with inversion of stereochemistry, by the amine function of either structure **8a** or **8b** in DMSO to give peptide-resin **9**. Deprotection and cleavage from the resin in anhydrous HF gave **2b** directly or **2a** still in the form of the disulfide which was reduced to the free thiol.

The aminoethanethiol derivative **8a** was synthesized in one step from the reaction of 2-aminoethanethiol and 5,5′-dithiobis-(2-nitrobenzoic acid) in 80% acetonitrile in water. The synthesis of the (aminooxy)ethanethiol derivative **8b** was more involved and is outlined in Scheme 3. Bromide **11** was produced from the reaction of *N*-hydroxyphthalimide (**10**) with a large excess of 1,2-dibromoethane.[19] Bromide **11** was then converted to the protected (aminooxy)ethanethiol derivative **12** with 4-methylbenzyl mercaptan in the presence of the base 1,8-diazabicyclo-[5.4.0]undec-7-ene (DBU). The phthalimide group of **12** was removed in a two-step process[20] involving reduction with $NaBH_4$ followed by treatment with acetic acid to give the desired (aminooxy)ethanethiol derivative **8b**.

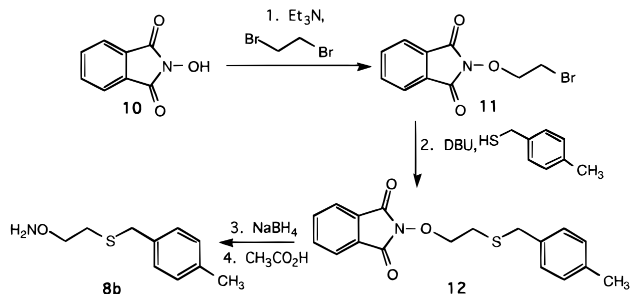
Model Ligations. The results of the model ligations are summarized in Table 1. All ligations were run at concentrations

(17) Muir, T. W.; Dawson, P. E.; Kent, S. B. H. *Tetrahedron Lett.*, submitted for publication.
(18) Robey, F. A.; Fields, R. L. *Anal. Biochem.* **1989**, *177*, 373−377.
(19) Bauer, L.; Suresh, K. S. *J. Org. Chem.* **1963**, *28*, 1604−1608.
(20) Osby, J. O.; Martin, M. G.; Ganem, B. *Tetrahedron Lett.* **1984**, *25*, 2093−2096.

Scheme 2. Synthesis of N^{α}(substituted) Peptide Segments



Scheme 3. Synthesis of Compound **8b**



ranging from 4 to 8 mg/mL of each peptide in either 8 M urea, 0.1 M Na_2HPO_4, pH 7.0, or 6 M guanidine·HCl, 0.1 M Na_2HPO_4, pH 7.5. Immediately after solvation of the peptide segments, 2−5% benzyl mercaptan (model ligation no. 1) or thiophenol (model ligation nos. 2−5), by volume of ligation buffer, was added to keep thiol functions in the reduced form. This had the added consequence of exchanging a significant amount of the original peptide-α-thioester (**1**) to a peptide-α-benzyl thioester or peptide-α-phenyl thioester, both of which are still capable of reacting in the desired fashion with peptide **2**. The ligation reactions were followed by analytical reversed phase HPLC and the products identified by ESMS. For clarity, in the following text the amino acid residues involved in the ligation are underlined.

Model Ligation No. 1. The peptide-α-(3-carboxy-4-nitro)-phenyl thioester **1** had the sequence LYRAG-$^{\alpha}COSC_6H_3$(3-CO_2H-4-NO_2) (observed mass 760 Da, calcd 760 Da). The N^{α}(ethanethiol) peptide **2a**, protected as the 5-thio-2-nitrobenzoic acid disulfide, had the sequence [(3-CO_2H-4-NO_2)-C_6H_3-S-SCH_2CH_2]-GAGPAGD-$^{\alpha}CONH_2$ (observed mass 800 Da, calcd 800 Da) which was reduced to [$HSCH_2CH_2$]-GAGPAGD-$^{\alpha}CONH_2$ (observed mass 603 Da, calcd 603 Da) in the ligation mixture. After 1 h, the product of the ligation, **4a**, [LYRAG-G((N^{α}-CH_2CH_2SH)AGPAGD-$^{\alpha}CONH_2$ (observed mass 1163 ±

Table 1. Summary of Model Ligations

model ligation	peptide **1**[a]	peptide **2**[b]	pH	temp (°C)	reaction time (h)	approximate yield[c] (%) unrearranged product (**3**)	rearranged product (**4**)
1	LYRAG-SNB	N^{α}(etsh)-GAGPAGD	7.0	25	4	not detected	90
2	LYRAG-SNB	N^{α}(oetsh)-GRNTATIMMQRGNFR	7.5	25	16	not detected	75
3	LYRAF-Sbzl	N^{α}(oetsh)-GRNTATIMMQRGNFR	[7.5	37	11.5	30	(35)]
			4.5[d]	37	10	0	64
4	LYRAG-Sbzl	N^{α}(oetsh)-AARHTVHQRHLHG	[7.5	37	17.5	39	(52)]
			4.5[d]	37	6.5	20	69
5	LYRAF-Sbzl	N^{α}(oetsh)-AARHTVHQRHLHG	7.5	37	19	58	not detected
			4.5[d]	37	22	52	not detected

[a] Thioester peptides where SNB = 3 carboxy-4-nitrophenyl thioester and Sbzl = benzyl thioester. [b] N^{α}(substituted) peptides where etsh = N^{α}(ethanethiol) and oetsh = N^{α}(oxyethanethiol). [c] Based on peptide **2**, estimated from analytical reversed phase HPLC (peak areas) and ESMS. [d] Adjusted to pH 4.5 after the indicated time at pH 7.5.

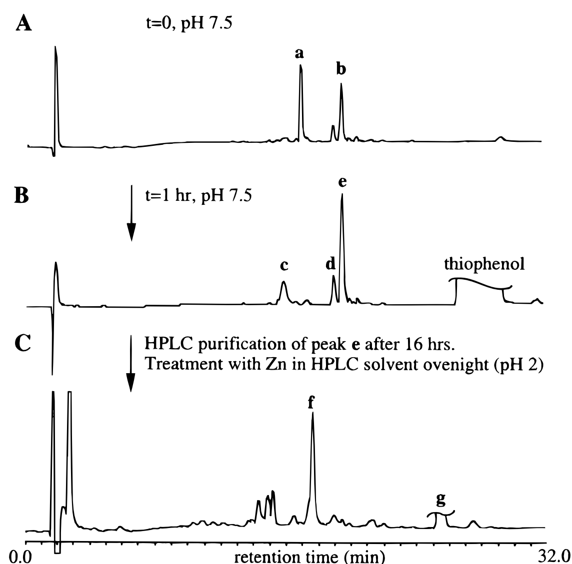


Figure 1. Model ligation no. 2 to form amide -Gly(N^{α}-OCH_2CH_2-SH)Gly- compound **4b**. Carried out in 6 M guanidine·HCl, 0.1 M Na_2HPO_4, pH 7.5. (A) Analytical HPLC (10−50% B over 30 min) at $t = 0$; peak a, thioester peptide **1**, LYRAG-$^{\alpha}COSC_6H_3$(3-CO_2H-4-NO_2); peak b, N^{α}(oxyethanethiol) peptide **2b**, [$HSCH_2CH_2O$-GRNTATIM-MQRGNFR-$^{\alpha}CONH_2$. (B) Analytical HPLC (10−50% B over 30 min) at $t = 1$ h; peak c, non-peptide impurity; peak d, LYRAG-$^{\alpha}COSC_6H_5$, resulting from transthioesterification of **1** (peak a) with thiophenol; peak e, intermediate ligation product **4b**, LYRAGG(N^{α}-OCH_2CH_2SHR-NTATIMMQRGNFR-$^{\alpha}CONH_2$ and a minor amount of unreacted [$HSCH_2CH_2O$-GRNTATIMMQRGNFR-$^{\alpha}CONH_2$ (**2b**), as determined by electrospray ionization MS. (C) Analytical HPLC (15−40% B over 30 min) of peak e after HPLC purification and treatment with Zn in acidic HPLC solvent overnight; peak f, final ligation product **5b**, LYRAGGRNTATIMMQRGNFR-$^{\alpha}CONH_2$; peak g, non-peptide impurity.

1 Da, calcd 1163 Da)] had already formed to a significant degree. After purification of the ligation product by reversed phase HPLC, it was treated with 5,5′-dithiobis(2-nitrobenzoic acid) for 1 h in 6 M urea, 0.1 M phosphate, pH 6.0, to yield the expected *S*-nitrobenzoic acid disulfide product LYRAGG[N^{α}-CH_2CH_2S-SC_6H_3(3-CO_2H-4-NO_2)]AGPAGD-$^{\alpha}CONH_2$ (observed mass 1360 ± 1 Da, calcd 1360 Da). This confirmed the rearrangement of **3a** to form **4a**, since **3a** would have been unreactive toward 5,5′-dithiobis(2-nitrobenzoic acid).

Model Ligation No. 2. Figure 1 shows the course of model ligation no. 2. The peptide-α-thioester segment **1** (peak a) had the sequence LYRAG-$^{\alpha}COSC_6H_3$(3-CO_2H,4-NO_2) (observed mass 760 Da, calcd 760 Da). The N^{α}(oxyethanethiol) peptide **2b** (peak b) consisted of the sequence [$HSCH_2CH_2O$]-GRN-TATIMMQRGNFR-$^{\alpha}CONH_2$ (observed mass 1827 ± 1 Da, calcd 1828 Da). After 1 h of reaction at room temperature in the presence of added thiophenol, the ligation product LYRAG-G(N^{α}-OCH_2CH_2SH)RNTATIMMQRGNFR-$^{\alpha}CONH_2$ (**4b**; observed mass 2388 ± 1 Da, calcd 2389 Da) had formed to a significant degree, coeluting with unreacted N^{α}(oxyethanethiol) peptide **2b** (peak e). After further reaction overnight at room temperature, the ligation product was purified by HPLC, and zinc dust was added directly to the collected peptide in HPLC eluant and stirred overnight at room temperature. Under these conditions, the zinc effectively reduced the N−O bond. Reductions of this type are possible through a variety of reagents.[21] The resulting peptide (peak f) gave a mass consistent with reduction of the N−O bond to produce LYRAGGRNTATIM-MQRGNFR-$^{\alpha}CONH_2$ (**5b**; observed mass 2313 ± 1 Da, calcd 2313 Da). This confirmed the rearrangement of **3b** to **4b** as shown in Scheme 1, since cleavage of the N−O bond in unrearranged **3b** would have resulted in the formation of two distinct peptides of significantly lower masses.

Model Ligation No. 3. Figure 2 shows the course of model ligation no. 3. The peptide-α-thioester segment **1** (peak d) consisted of a peptide with a C-terminal phenylalanine thioester [LYRAF-$^{\alpha}COSCH_2C_6H_5$ (observed mass 775 Da, calcd 775 Da)], thus providing a more sterically hindered model than the previous cases (model ligations nos. 1 and 2) which involved C-terminal glycine thioesters. The N^{α}(oxyethanethiol) peptide **2b** (peak a) consisted of the sequence [$HSCH_2CH_2O$]-GRN-TATIMMQRGNFR-$^{\alpha}CONH_2$ (observed mass 1827 ± 1 Da, calcd 1828 Da). The presence of the more sterically hindered thioester slowed the reaction relative to the unhindered models. However, heating at 37 °C was found to accelerate the rate of initial ligation. Figure 2A shows the ligation reaction after 11.5 h at 37 °C. Surprisingly, the rate of rearrrangement of **3b** to **4b** was slowed enough to observe the unrearranged product **3b**, LYRAF-[$^{\alpha}COSCH_2CH_2O$]-GRNTATIMMQRGNFR-$^{\alpha}$C-ONH_2 (peak b; observed mass 2478 ± 1 Da, calcd 2478 Da), eluting slightly before the rearranged product **4b**, LYRAFG(N^{α}-OCH_2CH_2SH)RNTATIMMQRGNFR-$^{\alpha}CONH_2$ (peak c; observed mass 2478 ± 1 Da, calcd 2478 Da). Having identical masses, the unrearranged and rearranged products **3b** and **4b** were identified by their zinc reduction products. Intermediate **3b**, peak b, gave two peptides upon zinc reduction with masses of 669 and 1753 Da corresponding to the peptide sequences LYRAF-$^{\alpha}$COOH (thioester hydrolysis) and GRNTATIMM-QRGNFR-$^{\alpha}CONH_2$, respectively. It was subsequently determined that lowering the pH to 4.5 after initial ligation by diluting the crude reaction mixture (to 5 times the volume) with 6 M guanidine·HCl, 0.1 M sodium acetate, pH 4.0, accelerated the rate of rearrangement. Figure 2B shows that rearrangement of the initial ligation product **3b** to **4b** (peak c) was complete after 10 h at pH 4.5 at 37 °C. HPLC purification of this peak and subsequent zinc reduction gave a peptide (peak f) of the expected

(21) Keck, G. E.; McHardy, S. F.; Wager, T. T. *Tetrahedron Lett*. **1995**, *36*, 7419−7922.

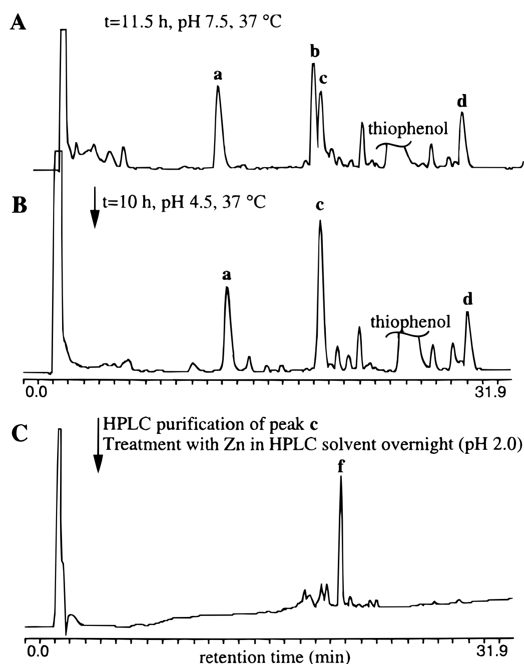


Figure 2. Model ligation no. 3 to form a -Phe(N$^{\alpha}$-OCH_2CH_2SH)Gly-linked product **4b**. Ligation and rearrangement carried out at 37 °C. (A) Analytical HPLC (20−40% B over 30 min) after 11.5 h in 6 M guanidine·HCl, 0.1 M Na_2HPO_4, pH 7.5; peak a, N$^{\alpha}$(oxyethanethiol) peptide **2b**, [$HSCH_2CH_2O$]-GRNTATIMMQRGNFR-$^{\alpha}CONH_2$; peak b, unrearranged intermediate ligation product **3b**, LYRAF-[$^{\alpha}COSCH_2CH_2O$]-GRNTATIMMQRGNFR-$^{\alpha}CONH_2$; peak c, rearranged intermediate ligation product **4b**, LYRAFG(N$^{\alpha}$-OCH_2CH_2SH)RNTATIMMQRGNFR-$^{\alpha}CONH_2$; peak d, thioester peptide **1**, LYRAF-$^{\alpha}COSCH_2C_6H_5$. (B) Analytical HPLC (20−40% B over 30 min) after 10 h in 6 M guanidine·HCl, 0.1 M $NaCH_3CO_2$, pH 4.5. (C) Analytical HPLC (0−67% B over 30 min) of HPLC-purified peak c after treatment with Zn in acidic HPLC solvent overnight; peak f, final ligation product **5b**, LYRAFGRNTATIMMQRGNFR-$^{\alpha}CONH_2$.

mass, LYRAFGRNTATIMMQRGNFR-$^{\alpha}CONH_2$ (**5b**; observed mass 2403 ± 1 Da, calcd 2404 Da).

Model Ligation No. 4. LYRAG-$^{\alpha}COSCH_2C_6H_5$ **1** (observed mass 685 Da, calcd 685 Da) was ligated to [$HSCH_2CH_2O$]-AARHTVHQRHLHG-$^{\alpha}COOH$ **2b** (observed mass 1595 ± 1 Da, calcd 1596 Da) at pH 7.5. This model provided steric hindrance, in the form of an Ala residue, on the opposite side of the ligation site from model ligation no. 3. The rate of reaction was similar to that of the previous model, the rate being significantly slower than those of the unhindered models (model ligation nos. 1 and 2), but enhanced with heating at 37 °C. Both unrearranged (**3b**, LYRAG-$^{\alpha}$[$COSCH_2CH_2O$]-AARHTVHQRHL-HG-$^{\alpha}COOH$) and rearranged (**4b**, LYRAGA(N$^{\alpha}$-OCH_2C-H_2SH)ARHTVHQRHLHG-$^{\alpha}COOH$) ligation products were formed, observed mass (both unrearranged and rearranged) 2156 ± 1 Da, calcd 2157 Da. However, unlike model ligation no. 3, the rate of rearrangement was *not* enhanced by lowering the pH to 4.5 after initial ligation, though rearrangement was nearly complete after 2 days at 37 °C regardless of pH as shown by successful reduction to the amide. It should also be noted that the presence of a significant number of histidines in the final ligation product **5b** (LYRAGAARHTVHQRHLHG-$^{\alpha}COOH$; observed mass 2080 ± 1 Da, calcd 2080 Da) resulted in binding of the peptide to the zinc. EDTA had to be added to the HPLC buffer/Zn mixture to free the peptide from the zinc after reduction of the N−O bond.

Model Ligation No. 5. A final model ligation was performed which provided steric bulk on both sides of the ligation site. LYRAF-$^{\alpha}COSCH_2C_6H_5$ **1** (observed mass 775 Da, calcd 775 Da) was ligated to [$HSCH_2CH_2O$]-AARHTVHQRHLHG-$^{\alpha}COOH$ **2b** (observed mass 1595 ± 1 Da, calcd 1596 Da) at pH 7.5. Though the initial (i.e., unrearranged) ligation product **3b** was observed, LYRAF-$^{\alpha}$[$COSCH_2CH_2O$]-AARHTVHQRHL-HG-$^{\alpha}COOH$ (observed mass 2246 ± 1 Da, calcd 2247 Da), there was no evidence of rearrangement over time, even at lower pH. The presence of side chains on both sides of the ligation site apparently provided too much steric hindrance for the rearrangement to occur via a six-membered ring intermediate under the conditions used.

Discussion

The observations reported here serve to emphasize the extraordinary facility of our original native ligation chemistry,[13] which recently was independently repeated in essentially identical form in model ligation studies.[22] We have used this original native peptide bond-forming ligation reaction in the chemical synthesis of a number of proteins with full biological activity, including the chemokine IL-8,[13] the enzymes HIV-1 protease and barnase, the serine proteinase inhibitors turkey ovomucoid third domain and eglin C, and a b/HLH transcription factor. This ligation reaction was based on principles enunciated by Max Brenner[23] and made use of chemistry first described by Wieland.[14]

Compared with the original native chemical ligation at X−Cys sites,[13] the most notable feature of the ligation chemistry reported here is the slower rearrangement of the initial thioester ligation product. Clearly, intramolecular attack by the α-NH-(oxyalkyl) group via a six-membered ring intermediate in the current instance is considerably less favored[24] than the facile five-membered ring mediated attack of the unsubstituted α-NH_2 group in the X−Cys ligation.[13] Except for the Gly−Gly ligation, in the present study we were able to isolate the thioester-linked initial ligation product (Table 1), which rearranged only slowly. Remarkably, for C$^{\alpha}$-substituted amino acids on both sides of the ligation site, we were unable to observe subsequent rearrangement to an amide bond: the thioester-linked intermediate was indefinitely stable under the conditions of reaction. These slow rearrangements are similar to those observed in an (N-substituted)amide-forming ligation chemistry previously described.[5]

The use in the current study of a temporary auxiliary functional group in amide-forming ligation is reminiscent of the "thiol capture" strategy proposed by Kemp.[25] Both native chemical ligation and the thiol capture method have as their stated goal the use of unprotected peptides in a segment condensation strategy to achieve the synthesis of long polypeptide chains. The key aspect of the native chemical ligation approach is reaction under conditions that promote exchange of the thiol moiety of the initial thioester-linked intermediate products, to give regioselective ligation at the N-terminal Cys even in the presence of other Cys residues in both segments.[13] By contrast, in the thiol capture strategy and related methods,[26] regiospecific protection of Cys side chain functionalities prior to the ligation reaction is a necessity,[27] thus frustrating the original intent.

(22) Tam, J. P.; Lu, Y. A.; Liu, C. F.; Shao, J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12485−12489.
(23) Brenner, M. In *Peptides. Proceedings of the Eighth European Peptide Symposium*; Beyerman, H. C., Ed., North-Holland: Amsterdam, 1967; pp 1−7.
(24) Mandolini, L. *J. Am. Chem. Soc.* **1978**, *100*, 550−554.
(25) Fotouhi, N.; Galackatos, N. G.; Kemp, D. S. *J. Org. Chem.* **1989**, *54*, 2806−2817.
(26) Liu, C. F.; Rao, C.; Tam, J. P. *Tetrahedron Lett.* **1996**, *37*, 933−936.

Similarly, the use in the current study of an activated α-thiocarboxyl function (i.e., an α-COSR) is reminiscent of chemistry previously used in peptide segment condensation.[28−30] However, these syntheses were based on conventional non-chemoselective attack by the α-amine nucleophile of the second segment on a Ag^+-activated peptide-$^{\alpha}$COSR, and thus necessitated regiospecific (re)protection of all other α- and ε-amine functional groups in both segments.[28−30]

By contrast, thioester-mediated amide-forming ligation chemistry is compatible with the use of completely unprotected peptide segments with the full range of side chain functionalities found in nature, including thiols.[9,13] For this reason, native chemical ligation is simple and practical, and is a general approach to the total chemical synthesis of proteins provided they contain appropriate ligation sites.

The work reported here increases the utility of the native chemical ligation method reported by Dawson *et al.*[13] by extending the number of dipeptide sequences that can be used as ligation sites. In addition to the X−Cys ligation site of that study, we can now make use of X−Gly and Gly−X ligation sites. Our previous studies have shown that X can be any amino acid, including *β*-branched amino acids such as Val.[13,31] Our results potentially extend the number of suitable sites for native chemical ligation by a factor of 3 to more than 50 of the 400 dipeptide sequences found in proteins. Because there is considerable latitude in choosing a ligation site in a target sequence, this extended applicability will render most polypeptides accessible by native chemical ligation. Continuing studies will establish the feasibility of a further extension of the applicability of ligation chemistry based on the same principles to an even wider range of dipeptide sequences.

Native chemical ligation in its original form[13] or in the form described here provides direct synthetic access to polypeptide chains the size of typical protein domains. Other ligation chemistries[1,4,5] can be used to join synthetic domains in a modular fashion to produce large (i.e., >20 kDa), fully functional synthetic proteins.[8,9] In its most general form, incorporating all suitable chemistries,[1,4,5,13] the chemical ligation approach represents the next stage in the evolution of methods for the chemical synthesis of polypeptides and, for the first time, provides for reproducible, practical total chemical synthesis of proteins.

Experimental Section

Materials and Methods. Machine-assisted solid phase peptide syntheses were carried out on a custom-modified Applied Biosystems 430A peptide synthesizer.[12] Reversed phase HPLC was performed on a Rainin HPLC system with 214-nm UV detection, using Vydac C-18 analytical (5 *μ*m, 0.46 × 15 cm) and semiprepatative (10 *μ*m, 1.0 × 25 cm) columns. Chromatographic separations were achieved using linear gradients of buffer B in A (A = 0.1% TFA in water, B = 90% CH_3CN/10% water containing 0.09% TFA) over 30−60 min at 1 mL/min (analytical) or 3 mL/min (semipreparative). Mass spectra of all peptide segments were obtained with a Sciex API-III electrospray quadrupole mass spectrometer; observed masses were derived from the experimental *m*/*z* values for all observed protonation states of a molecular species, using the program MacSpec (Sciex). Calculated masses were based on average isotope composition and were derived using the program MacProMass (Terry Lee and Sunil Vemuri, Beckman Research Institute, Duarte, CA). All other mass spectrometry was performed at The Scripps Research Institute Mass Spectrometry Facility. ^{1}H NMR spectra were recorded on a Bruker 250-MHz spectrophotometer, and chemical shifts are reported in parts per million downfield from Me_4Si. Microanalyses were performed at The Scripps Research Institute X-ray Crystallographic Facility and agreed with calculated values ±0.4%. Boc-L-amino acids and HBTU were purchased from Novabiochem (La Jolla, CA). [[4-(Hydroxymethyl)phenyl]acetamido]-methyl (PAM) resins and diisopropylethylamine (DIEA) were obtained from Applied Biosystems (Foster City, CA), and methylbenzhydrylamine (MBHA) resin was obtained from Peninsula Laboratories, Inc. (San Carlos, CA). Synthesis grade dimethylformamide (DMF) was obtained from Baker, and AR grade CH_2Cl_2 and HPLC grade CH_3CN were obtained from Fisher. Trifluoroacetic acid (TFA) was obtained from Halocarbon (New Jersey). HF was purchased from Matheson Gas. 4-Methylbenzyl mercaptan was obtained from Lancaster. All other reagents were AR grade or better and were obtained from Aldrich Chemical or from Fisher.

Peptide Segment Synthesis. Chain Assembly. Peptides were synthesized in stepwise fashion by established machine-assisted or manual solid phase methods using *in situ* neutralization/HBTU activation protocols for Boc chemistry.[12] Side chain protection was as follows: Boc-Arg(*p*-toluenesulfonyl)-OH, Boc-Asn(xanthyl)-OH, Boc-Asp(*O*-cyclohexyl)-OH, Boc-His(dinitrophenyl)-OH, Boc-Thr(benzyl)-OH, and Boc-Tyr(2-bromobenzyloxycarbonyl)-OH. Boc-Gln-OH and Boc-Met-OH were used without side chain protection. Coupling reactions were monitored by quantitative ninhydrin assay[32] and were typically >99%. After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin by treatment with HF containing 5% *p*-cresol for 1 h at 0 °C to give the peptide-$^{\alpha}$COSH, -$^{\alpha}CONH_2$, or -$^{\alpha}CO_2H$. After removal of the HF under reduced pressure, the crude peptide was precipitated in anhydrous Et_2O, dissolved in HPLC buffer (40−50% B), and lyophilized.

Peptide-α-thioesters (1). Thioacid peptides were synthesized on the appropriate Boc-aminoacyl-*S*-resins, made by coupling [4-[α-(*N*-*t*-Boc-aminoacyl-*S*)benzyl]phenoxy]acetic acid, DCHA salt[16] (2.0 equiv), and aminomethyl-resin (1 equiv, washed with 10% DIEA in DMF) with HBTU (1.6 equiv) added as an activating agent and DIEA (1 equiv), in DMF. Peptide-$^{\alpha}COSC_6H_3(3\text{-}CO_2H\text{-}4\text{-}NO_2)$ thioesters were generated by dissolving the crude peptide-$^{\alpha}$COSH (15−20 mg) in 6 M guanidine·HCl, 0.1 M sodium Acetate, pH 5.0−6.5, to which was added 1.5 equiv of 5,5′-dithiobis(2-nitrobenzoic acid).[13,17] The mixture was vortexed briefly and purified after 10 min. The identity of the peptide-$^{\alpha}COSC_6H_3(3\text{-}CO_2H\text{-}4\text{-}NO_2)$ ester was unambiguously confirmed by precise electrospray mass measurements, in contrast with Liu.[26] Peptide-$^{\alpha}COSCH_2C_6H_5$ thioesters were generated by dissolving the crude peptide-$^{\alpha}$COSH (15−20 mg) in 6 M guanidine·HCl, 0.1 M sodium acetate, pH 4.0, to which was added 10 equiv of benzyl bromide.[13] The mixture was vortexed briefly and purified after 1 h. LYRAG-$^{\alpha}COSC_6H_3(3\text{-}CO_2H\text{-}4\text{-}NO_2)$ (observed mass 760 Da, calcd 760 Da) was purified by semipreparative HPLC (20−60% B over 40 min) to give 20−30% yield. LYRAG-$^{\alpha}COSCH_2C_6H_5$ (observed mass 685 Da, calcd 685 Da) was purified by semipreparative HPLC (15−45% B over 60 min) to give 25−30% yield. LYRAF-$^{\alpha}COSCH_2C_6H_5$ (observed mass 775 Da, calcd 775 Da) was purified by semipreparative HPLC (30−60% B over 60 min) to give 25−30% yield. Most of the losses in yield arose simply from HPLC recoveries.

N$^{\alpha}$(ethanethiol) and N$^{\alpha}$(oxyethanethiol) Peptides (2). These peptides were synthesized on either MBHA or the appropriate Boc-aminoacyl-OCH_2-PAM-resins. After chain assembly was complete and the N$^{\alpha}$Boc group removed with neat TFA (two 1-min treatments) and neutralized with 10% DIEA in DMF (two 1-min treatments), the peptide was bromoacetylated by the method of Robey.[18] Bromoacetic acid (2.0 mmol) was dissolved in CH_2Cl_2 (2 mL), to which was added DIC (1 mmol). After activation for 10−15 min to form the symmetric anhydride, the mixture was diluted with DMF (2 mL), added to the

(27) Kemp, D. S.; Carey, R. I. In *Peptides: Proceedings of the 11th American Peptide Symposium*; Rivier, J. E., Marshall, G. R., Eds.; ESCOM: Lieden, 1990; pp 920−922.

(28) Blake, J.; Li, C. H. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4055−4058.

(29) Yamashiro, D.; Li, C. H. *Int. J. Peptide Protein Res.* **1988**, *31*, 322−334.

(30) Hojo, H.; Yoshimura, S.; Aimoto, S. In *Peptide Chemistry*; Okada, Y., Ed.; Protein Research Foundation: Osaka, 1994; pp 9−12.

(31) Simon, R. J. Personal communication. Ligations involving *β*-branched amino acids at the ligation site proceed at slower rates than ligations involving other amino acids.

(32) Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147−157.

peptide-resin, and coupled for 30 min. The resin was then washed with DMSO, and **8a** (2 M in DMSO) or **8b** (1 M in DMSO) was added and allowed to react with the bromoacetylated peptide-resin for 8−23 h. The peptides were purified without further modification after cleavage from the resin. [(3-CO_2H-4-NO_2)-C_6H_3-S-SCH_2CH_2]-$\underline{A}$G-PAGD-$^{\alpha}CONH_2$ (**2a**) (observed mass 800 Da, calcd 800 Da) was purified by semipreparative HPLC (0−67% B over 60 min) to give ∼24% yield. [$HSCH_2CH_2O$]-$\underline{G}$RNTATIMMQRGNFR-$^{\alpha}CONH_2$ (**2b**) (observed mass 1827 ± 1 Da, calcd 1828 Da) was purified by semipreparative HPLC (15−40% B over 60 min) to give ∼25% yield. [$HSCH_2CH_2O$]-$\underline{A}$ARHTVHQRHLHG-$^{\alpha}COOH$ (**2b**) was synthesized by the above method using the racemate of 2-bromopropionic acid instead of bromoacetic acid. The resulting crude lyophilized product was a mixture of the desired peptide and the peptide that results from elimination of HBr from the bromopropionylated peptide (CH_2=CHCO-ARHTVHQRHLHG-$^{\alpha}COOH$; observed mass 1502 ± 1 Da, calcd 1502 Da). The mixture was purified by semipreparative HPLC (10−40% B over 60 min) to give [$HSCH_2CH_2O$]-$\underline{A}$ARHTVHQRHLHG-αCOOH (observed mass 1595 ± 1 Da, calcd 1596 Da) in ∼10% yield.

***S*-[(3-Carboxy-4-nitrophenyl)thio]-2-aminoethanethiol (8a).** 2-Aminoethanethiol (1.0 g, 13.0 mmol) and 5,5′-dithiobis(2-nitrobenzoic acid) (1.75 g, 4.4 mmol) were combined in acetonitrile (100 mL) and water (25 mL) and stirred at room temperature for 14 h. The reaction mixture was diluted with water (450 mL) and purified by preparative reversed phase HPLC on a Waters Delta Prep 4000 with a Vydac 5.0 × 2.5 cm preparative C-18 column to give **8a** (1.0 g, 40%). ^{1}H NMR (D_2O): δ 7.95 (d, 1H, J = 8.7 Hz), 7.61 (dd, 1H, J = 8.7, 2.2 Hz), 7.58 (d, 1H, J = 2.2 Hz), 3.18 (t, 2H, J = 6.5 Hz), 2.92 (t, 2H, J = 6.7 Hz).

***N*-(2-Bromoethoxy)phthalimide (11).** *N*-(2-Bromoethoxy)phthalimide was synthesized by a modification of the procedure of Bauer and Suresh.[19] *N*-Hydroxyphthalimide (16.0 g, 98 mmol), triethylamine (30 mL, 215 mmol), and 1,2-dibromoethane (35 mL, 406 mmol) were combined in DMF (115 mL) and stirred at room temperature overnight. Solids were filtered and washed with DMF. The filtrate was diluted with water (800 mL) and the resulting precipitate filtered, dissolved in EtOAc (200 mL), and washed with 1 N HCl (2 × 50 mL), water (1 × 50 mL), and saturated NaCl (1 × 50 mL), and dried over $MgSO_4$. Volatiles were removed *in vacuo*, and the resulting solid was recrystallized from 95% EtOH to give **11** as a white solid (16.6 g, 63%). ^{1}H NMR ($CDCl_3$): δ 7.82 (m, 4H), 4.49 (t, 2H, J = 6.9 Hz), 3.65 (t, 2H, J = 6.9 Hz). FAB MS (sodium ion): calcd for [$C_{10}H_8BrNO_3$, H^+] 291.9585, found 291.9579.

***N*-[2-[*S*-(4-Methylbenzyl)]thio]ethoxy]phthalimide (12).** Bromide **11** (16.6, 62 mmol), 4-methylbenzyl mercaptan (8.5 mL, 63 mmol) and DBU (9.5 mL, 64 mmol) were combined in benzene (150 mL) and stirred at room temperature for 8 h. Solids were filtered and washed with benzene, and the filtrate was washed with 1 N HCl (2 × 35 mL), water (1 × 35 mL), and saturated NaCl (1 × 35 mL) and dried over $MgSO_4$. Volatiles were removed *in vacuo* and the resulting solid recrystallized from EtOAc/hexane to yield **12** as a white solid (14.8 g, 74%). ^{1}H NMR ($CDCl_3$): δ 7.80 (m, 4H), 7.18 (d, 2H, J = 8.0 Hz), 7.04 (d, 2H, J = 8.0 Hz), 4.22 (t, 2H, J = 7.4 Hz), 3.75 (s, 2H), 2.79 (t, 2H, J = 7.4 Hz), 2.27 (s, 3H). FAB MS: calcd for [$C_{18}H_{17}NO_3S$, H^+] 328.1007, found 328.1016. Anal. Calcd for $C_{18}H_{17}NO_3S$: C, 66.03; H, 5.23; N, 4.28; S, 9.79. Found: C, 66.04; H, 4.95; N, 4.30; S, 9.58.

***S*-(4-Methylbenzyl)-2-(aminooxy)ethanethiol (8b).** *S*-(4-methylbenzyl)-2-(aminooxy)ethanethiol was synthesized by the method of Osby.[20] The N-substituted phthalimide **12** (7.4 g, 23 mmol) was suspended in 2-propanol (203 mL) and water (35 mL), to which was added $NaBH_4$ (3.5 g, 92 mmol). The mixture was stirred at room temperature overnight. Acetic acid (25 mL) was slowly added till bubbling ceased and the flask stoppered and heated to 50 °C for 2−3 h. Volatiles were removed *in vacuo*, and the resulting solution was diluted with 1 N NaOH and extracted with EtOAc (4 × 50 mL). The combined EtOAc extractions were washed with saturated NaCl (1 × 50 mL) and dried over $MgSO_4$. Volatiles were removed *in vacuo*, and the resulting oil was purified by flash chromatography (1:1 hexane/EtOAc) to yield **8b** as a clear, colorless oil (3.2 g, 72%). ^{1}H NMR ($CDCl_3$): δ 7.21 (d, 2H, J = 8.0 Hz), 7.12 (d, 2H, J = 8.0 Hz), 5.40 (br s, 2H, D_2O exchanged), 3.77 (t, 2H, J = 6.5 Hz), 3.71 (s, 2H), 2.64 (t, 2H, J = 6.5 Hz), 2.33 (s, 3H). FAB MS: calcd for [$C_{10}H_{15}NOS$, H^+] 198.0953, found 198.0958. Anal. Calcd for $C_{10}H_{15}NOS$: C, 60.88; H, 7.66; N, 7.10; S, 16.25. Found: C, 60.79; H, 7.88; N, 7.03; S, 16.11.

Acknowledgment. This work was supported by funds from the NIH (Grants PO1 HL31950 and RO1GM43397, S.B.H.K.). We gratefully acknowledge the assistance of Dr. Gary Suizdak and Dr. Raj K. Chandra, The Scripps Research Institute, for FAB MS measurements and for elemental analyses, respectively, and our other colleagues at The Scripps Research Institute for numerous helpful discussions.

JA960398S