Orthogonal Ligation of Unprotected Peptide Segments through Pseudoproline Formation for the Synthesis of HIV-1 Protease Analogs[†,‡]

Chuan-Fa Liu, Chang Rao, and James P. Tam*

Contribution from the Department of Microbiology and Immunology, A-5119 MCN, Vanderbilt University, 1161 21st Avenue South, Nashville, Tennessee 37232-2363

Received August 14, 1995[⊗]

Abstract: We describe the total synthesis of three active HIV-1 protease analogs by an orthogonal ligation through thiaproline formation of two large, unprotected peptide segments. The central element of this strategy in achieving orthogonality of peptide bond formation is through proximity effect, and the key reaction is a side chain thiol initiated aldehyde capture to overcome the entropic problem of coupling between two large molecular weight peptide segments. The capture step also leads to specific entropic activation of the acyl segment because the respective termini are brought to close proximity, resulting in a spontaneous acyl rearrangement to form the amide bond. A general method using a thioester for introducing an aldehyde moiety to the C-terminus of an unprotected peptide segment was also developed. Three active analogs of HIV-1 protease were obtained in excellent yield by ligating two segments of 38 and 61 residues. Two analogs contained a thioproline residue at position 39, and the third contained a 38−39 non-peptide backbone. Efficient ligation at pH 4 was attained at peptide segment concentrations as low as 50 μM, a concentration which is not feasible with conventional convergent methods using protected peptide segments.

Introduction

Conventional approaches for segment condensation in peptide synthesis focus on an α-amine-directed amide bond formation with an enthalpically activated α-carboxylic group.[1,2] These approaches require a full or partial protection scheme for other functionalities to prevent random acylation, and suffer the limitations of solubility and purification difficulties attendant to protected peptide segments.[1] Moreover, the intrinsic entropic barrier imposed by the high molecular weights of large protected peptide segments reduces the efficiency of intermolecular peptide bond formation.

Considerable efforts to overcome these limitations have resulted in the development of various approaches using minimally protected or totally unprotected peptide segments.[2−6] Those employing the principle of entropy-driven intramolecular acyl transfer for the formation of peptide bonds hold particular promise.[4−6] Kemp *et al.*[4] utilized a capture step that brings together the respective C- and N-termini of the two unprotected peptides through a tricyclic template to facilitate the subsequent amide bond formation through intramolecular acyl rearrangement. On the basis of the same rationale, we have recently designed an orthogonal ligation without a template using a side chain initiated approach for coupling unprotected peptide segments in which the focus of the first bond formation is not directed to the α-amine, but rather between an N^{α}-side chain of the amino segment and an acyl segment bearing a glycolaldehyde ester moiety.[5] In this scheme, the capture step is an orthogonal coupling reaction and forms a non-amide, thiazolidine (Thz) ring. The orthogonality of this reaction is derived from the specific thiazolidine formation between the aldehyde and the 1,2-amino thiol of these segments at acidic pH in the presence of other unprotected and reactive functional groups. The thiazolidine formation also brings the α-amine close to the α-carboxyl moiety of the acyl segment to increase their effective molar concentrations. The reactive carboxyl and amino termini are positioned in close proximity for a spontaneous peptide bond formation through an entropy-driven intramolecular *O,N*-acyl transfer. Moreover, a proline-like residue (5-(hydroxymethyl)-γ-thioproline, abbreviated as SPro) is formed at the ligation site and can be used as a proline surrogate. This strategy obviates the need for protecting groups and overcomes the entropy problem of conventional peptide bond formation between large peptide segments.

A crucial step in this approach is the introduction of the glycol aldehyde ester onto the C-terminus of the unprotected acyl peptide segment. In this paper, we describe a general method involving the use of a thioester and Ag^{+} as an activation agent

* To whom all correspondence should be addressed.

[†] An abstract of this work was presented at the Third International Chinese Peptide Symposium in Beijing, China, on June 16, 1994.

[‡] Abbreviations: Abu, L-α-aminobutyric acid; AcOH, acetic acid; Boc, *tert*-butoxycarbonyl; BOP, (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate; DMF, dimethylformamide; DMSO, dimethyl sulfoxide; DTT, 1,4-dithiothreitol; EDTA, ethylenediaminetetraacetic acid; HIV-1 PR, human immunodeficiency virus-1 protease; HOSu, *N*-hydroxysuccinimide; MALDMS, matrix-assisted laser desorption mass spectrometry; RP HPLC, reversed-phase high-performance liquid chromatography; SPro, thioproline; TFA, trifluoroacetic acid; Thz, thiazolidine.

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1995.

(1) (a) Hirschmann, R.; Nutt, R. F.; Veber, D. F.; Vitali, R. A.; Varga, S. L.; Jacob, T. A.; Holly, F. W.; Denkewalter, R. C. *J. Am. Chem. Soc.* **1969**, *91*, 507−508. (b) Kiyama, S.; Fujii, N.; Yajima, H.; Moriga, M.; Takagi, A. *Int. J. Pept. Protein Res.* **1984,** *23*, 174−186. (c) Kuroda, H.; Chen, Y.-N.; Kimura, T.; Sakakibara, S. *Int. J. Pept. Protein Res.* **1992,** *40*, 294−299.

(2) (a) Blake, J.; Li, C. H. *Proc. Natl. Acad. Sci. U.S.A.* **1981,** *78*, 4055−4058. (b) Yamashiro, D.; Li, C. H. *Int. J. Pept. Protein Res.* **1988**, *31*, 322−334. (c) Hojo, H.; Aimoto S. *Bull. Chem. Soc. Jpn.* **1992,** *64*, 111.

(3) (a) King, T. P.; Zhao, S. W.; Lam, T. *Biochemistry* **1986**, *25*, 5774−5779. (b) Gaertner, H. F.; Rose, K.; Cotton, R.; Timms, D.; Camble, R.; Offord, R. E. *Bioconjugate Chem.* **1992**, *3*, 262−268. (c) Schnolzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221−225.

(4) (a) Kemp, D. S. In *Peptides: Synthesis, Structure and Function*; Rich, D., Gross, E., Eds.; Pierce Chemical Co.: Rockford, IL, 1981; pp 73−79. (b) Kemp, D. S. *Biopolymers* **1981**, *20*, 1793. (c) Fotouhi, N.; Galakatos, N. G.; Kemp, D. S. *J. Org. Chem.* **1989**, *54*, 2803. (d) Kemp, D. S.; Carey, T. I. *J. Org. Chem.* **1993,** *58*, 2216.

(5) (a) Liu, C.-F.; Tam, J. P. *J. Am. Chem. Soc.* **1994,** *116*, 4149. (b) Liu, C.-F.; Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994,** *91*, 6584−6588.

(6) Dawson, P. E., Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994,** *266*, 776−779.

0002-7863/96/1518-0307$12.00/0

for this purpose in the synthesis of three analogs of the aspartic protease of the human immunodeficiency virus-1, HIV-1 PR. The syntheses of these active HIV-1 PR analogs provide a stringent test for this entropy-driven orthogonal ligation strategy and demonstrate the suitability of a thioproline as a replacement for Pro residue in protein synthesis.

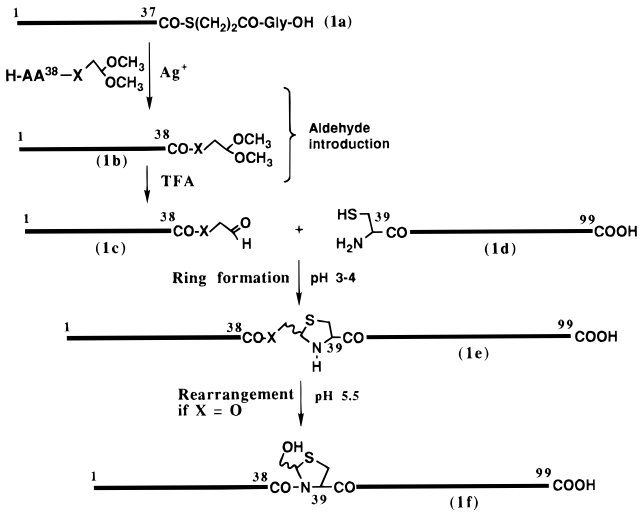
Results

Overall Synthetic Scheme. HIV-1 PR, a member of the aspartic protease family, is a homodimer in which each of the 99-residue monomers contributes one of the two aspartic acid residues to constitute the active site of the enzyme.[7,8] Total chemical synthesis of this protease was first achieved by stepwise solid phase peptide synthesis.[9ab] In addition, several analogs of HIV-1 PR have also been prepared by chemoselective ligation via formation of thioester bonds as peptide bond replacement.[3c,9c,d] In our strategy, two unprotected peptides are ligated through formation of an Xaa−SPro amide bond, which is more stable than the thioester and other non-amide linkages, such as hydrazone and oxime. Two proline residues in the HIV-1 PR sequence are suitable for our target ligating sites: Pro[39] at the broad loop (sequence 36−42) prior to the flap region and Pro[44] located within the first β strand of the flap. In this study, we chose Pro[39] as the ligating site to accommodate the pseudoproline structure. Because the Cys[67,95] are not involved in disulfide bridge, we used the strategy of Wlodawer *et al.*[8c] and replaced these Cys residues with α-aminobutyric acid (Abu) to give [SPro[39],Abu[67,95]]HIV-1 PR. Two other analogs, Leu[38]→Ala, [Ala[38],SPro[39],Abu[67,95]]HIV-1 PR and a Leu[38]-amidomethyl-Thz[39] analog with a non-peptide backbone, [Leu-NHCH$_2$-Thz[38−39],Abu[67,95]]HIV-1 PR, were also prepared (Scheme 1). The latter is an isosteric but stable analog of the intermediate ester product prior to intramolecular acyl rearrangement and provides an opportunity to compare the rearranged with the unrearranged product.

The overall strategy for the syntheses of these HIV-1 protease analogs consists of ligating two unprotected segments, the (1−38)-acyl and the (39−99)-amino segments (Scheme 1). While the synthesis of the amino segment was relatively straightforward in accordance with conventional solid phase methods and was achieved with satisfactory results (Figure 1B), the synthesis of the acyl segment containing a glycol aldehyde ester (acyloxyacetaldehyde) requires a new methodology.

Preparation of Acyl Segment Bearing a Glycol Aldehyde Ester. A general method for the preparation of the acyl segment with a glycol aldehyde ester was developed using an $n + 1$ strategy by adding a single amino acid derivative, a leucinyl[38] glycol aldehyde dimethyl acetal (Leu-oxyacetaldehyde dimethyl acetal), onto the (1−37)-segment containing a C$^{\alpha}$-thioester, taking the synthesis of [SPro[39],Abu[67,95]]HIV-1 PR as example.

Scheme 1. Overall Strategy for the Synthesis of HIV-1 PR Analogs[a]



[a] The aldehyde introduction step consists of coupling of the thioester, **1a**, activated by silver ion with an α-amino(acyloxy(amido))acetaldehyde masked as methyl acetal in large excess (≥100-fold) and deblocking the acetal by aqueous TFA to yield the desired (1−38)-acyl segment **1c**. Ring formation between **1c** and **1d** upon aldehyde capture of the 1,2-amino thiol of the N-Cys of **1d** gave the thiazolidine product, **1e**, which underwent a spontaneous *O*,*N*-acyl transfer to yield the final product with a thioproline (SPro) at position 39, **1f**. AA = amino acid residue; X = O or NH. If X = O, [SPro[39],Abu[67,95]]PR and [Ala[38],SPro[39],Abu[67,95]]PR were obtained with AA[38] = Leu or Ala, respectively. If X = NH, the synthetic scheme ended at stage **1e** and [Leu-NHCH$_2$-Thz[38−39],Abu[67,95]]PR was obtained with AA[38] = Leu. Amino acid sequence of HIV-1 protease: H-**P[1]QITLWQRPL VTIRIGGQLKEALL*DTG*ADDTVLEEMNLP[39]GKW**<u>**KPKMI-GGIGGFIKVRQYDQIP VEICGHKAIGTVLVGPTPVNIIGRN-LLTQ IGCTLNF**</u>**[99]**-OH. Amino acid residues are represented by the standard single letter code. The ligation site was at Pro[39]. The underlined sequence represents the flap region; the highly conserved active site sequence is in italics. The two Cys[67,95] residues were replaced with L-α-aminobutyric acid (Abu) in our synthetic analogs.

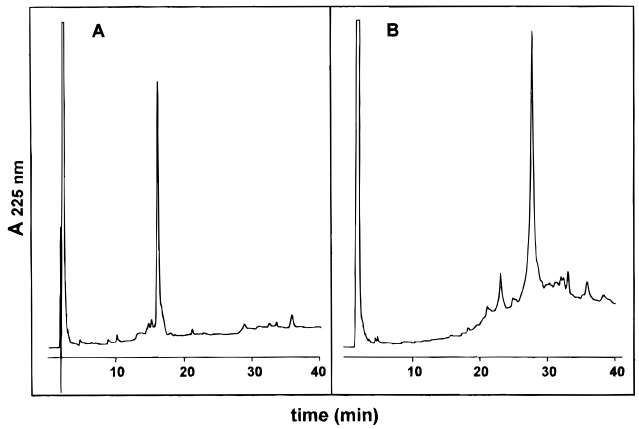


Figure 1. C$_{18}$ RP HPLC elution profile of the crude products of HIV-1 PR(1−37)-S(CH$_2$)$_2$CO-Gly-OH and [Cys[39],Abu[67,95]]HIV-1 PR(39−99) after HF cleavage and dialysis. (A) HIV-1 PR(1−37)-S(CH$_2$)$_2$CO-Gly-OH; (B) [Cys[39],Abu[67, 95]]HIV-1 PR(39−99). HPLC linear gradient: 40−80% buffer B at 1%/min in buffer A.

The synthesis of the (1−37)-segment thioester was achieved by the stepwise solid phase synthesis using Boc chemistry on a Boc-Asn-SCH$_2$CH$_2$CO-Gly-OCH$_2$-Pam resin using the method developed by Aimoto *et al.*[2c] HF cleavage of the protected resin-bound (1−37)-thioester segment yielded the desired unprotected peptide in excellent yield (Figure 1A). Activation of thioester by Ag$^+$ in the presence of a large excess of the masked Leu[38]-oxyacetaldehyde efficiently yielded the (1−38)-

(7) (a) Debouck, C.; Gorniak, J. G.; Strickler, J. E.; Meek, T. D.; Metcalf, B. W.; Rosenberg, M. *Proc. Natl. Acad. Sci. U.S.A.* **1987,** *84*, 8903−8906. (b) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U.S.A.* **1988,** *85*, 4686−4690.

(8) (a) Pearl, L. H.; Taylor, W. R. *Nature (London)* **1987,** *329*, 351−354. (b) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. *Nature (London)* **1989,** *337*, 615−620. (c) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Calwson, L.; Schneider, J.; Kent, S. B. H. *Science* **1989**, *245*, 616−621.

(9) (a) Schneider, J.; Kent, S. B. H. *Cell* **1988**, *54*, 363−368. (b) Darke, P. L.; Nutt, R. F.; Brady, S. F.; Garsky, V. M.; Ciccarone, T. M.; Leu, C.-T.; Lumma, P. K.; Freidinger, R. M.; Veber, D. F.; Sigal, I. S. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 297−303. (c) Baca, M.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90*, 11638−11642. (d) Baca, M.; Muir, T. W.; Schnolzer, M.; Kent, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 1881−1887.

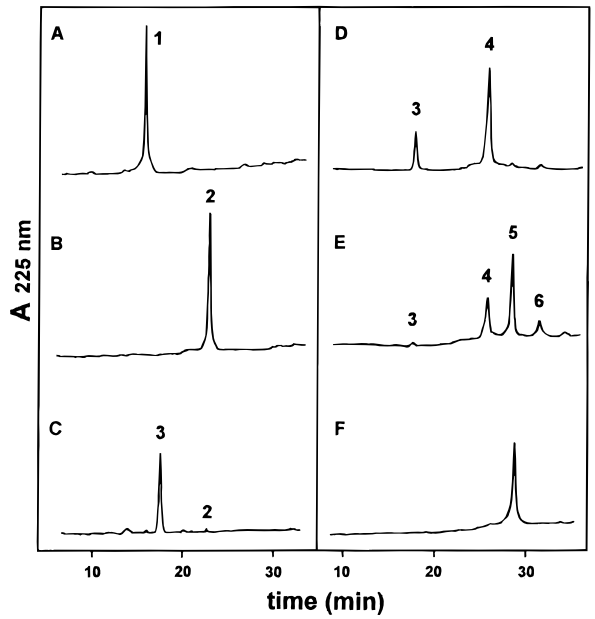


Figure 2. C_{18} RP HPLC monitoring of the progress of the synthesis of [SPro^{39},Abu^{67,95}]PR: (A) purified thioester segment, HIV-1 PR(1−37)-S$(CH_2)_2$CO-Gly-OH (peak 1); (B) purified HIV-1 PR(1−38)-OCH$_2$-CH$(OCH_3)_2$ (peak 2) after coupling of HIV-1 PR(1−37)-S$(CH_2)_2$CO-Gly-OH activated by silver nitrate with H-Leu^{38}-OCH$_2$CH$(OCH_3)_2$ (large excess) in DMSO; (C) demasking of the acetal by TFA (5% H_2O) to give HIV-1 PR(1−38)-OCH$_2$CHO (peak 3) in 10 min at 0 °C; (D) ligation of HIV-1 PR(1−38)-OCH$_2$CHO (Peak 3) with [Cys^{39},Abu^{67,95}]HIV-1 PR(39−99) (peak 4), t = O; (E) after 5 h; desired ligation product (peak 5) and peak 6 is the disulfide dimer of peak 4; (F) purified product, the *N*-acyl product was indisdinguishable in HPLC with the *O*-acyl product. HPLC was run with a linear gradient of 40−80% buffer B at 1%/min in buffer A.

acyl segment, **1b** (Figure 2B). No side reactions with other amines and nucleophiles on the (1−37)-acyl segment itself were detectable because of their relatively low molar ratio when compared to the large excess of the small Leu^{38} component. A concern on the stability of the (1−37)-thioester was the presence of Asn at the C-terminus which might be susceptible to aspartimide formation due to the amide side chain participation during the activation by Ag^+. However, no aspartimide formation was found in model experiments with Boc-Asn-S$(CH_2)_2$COOH. The likelihood of aspartimide formation during the reaction of (1−37)-thioester was further diminished by the large excess of the small amino component.

The masked amino acid glycol aldehyde ester was introduced enzymatically in previously published methods.[5] However, the chemical method via Ag^+ described here is more versatile because it overcomes the limitation imposed by the substrate specificity of an enzymatic reaction. Blake and Yamashiro as well as Aimoto *et al.* have used the silver ion activated thiocarboxylic acid and thioester for the condensation of partially protected peptide segments.[2] The use of totally unprotected fragments has not been attempted because of the concern of random acylation on other amine nucleophiles present in the naked amine and acyl components. Our $n + 1$ coupling makes this feasible because it involves formation of a peptide bond between a large unprotected peptide and a small functionalized amino acid. The selectivity and efficiency of this coupling was driven by the silver ion mediated activation which distinguished the C^{α}-thioester from other side chain unprotected carboxylic groups and by an overwhelming excess of the small amino acid derivative of the masked glycol aldehyde which largely reduced the probability of undesired random acylation. In addition, *N*-hydroxysuccinimide was also used in large excess as an additive to minimize the possible danger of racemization. In combination with the orthogonal ligation, the $n + 1$ method of Ag^+-mediated thioester activation also provided a convenient approach for the systematic preparation of protein analogs, including those containing small organic functionalities. For example, we have introduced a Boc-protected hydrazide onto the C-terminus of the acyl (1−37)-segment by reaction with Boc-NHNH$_2$ in large excess (unpublished data). Such a derivative was then be used for conjugation, after removal of Boc, with a second segment bearing an N-terminal aldehyde function to form an HIV-1 PR analog containing an unusual backbone (unpublished data).

Acetal was chosen for aldehyde protection because it is easily removed. Treatment of the acetal, **1b**, by TFA at low temperature (0 °C) quickly (10 min) demasked the aldehyde, **1c** (Figure 2C). The TFA treatment gave side reactions in HIV-1 PR segments as revealed by HPLC when longer reaction time and/or higher temperature was applied (data not shown).

Similarly, two other analogs were also prepared by the $n + 1$ method. Ala^{38} analog was prepared using the masked Ala-oxyacetaldehyde and the Leu^{38}-amidomethyl-Thz^{39} analog by the masked Leu-amidoacetaldehyde.

Ligation of the Amine and Acyl Segments. Ligation between the unmasked acyl segment (1−38) containing a glycol aldehyde, **1c**, and [Cys^{39}]amine segment **1d** gave the HIV-1 PR thiazolidine product **1e,** in this case, [SPro^{39},Abu^{67,95}]HIV-1 PR. This ring formation was efficient and was 80% completed within 5 h (Figure 2E) even when the concentrations of both segments were low (0.05−0.2 mM). Such high reactivity provided orthogonality of our ligation scheme as well as the driving force to overcome the entropy barrier of the conventional intermolecular reaction between large peptide molecules. This efficiency could be attributed to the high reactivity of an aldehyde with 1,2-amino thiol and possibly to the complementarity between the two structurally folded unprotected peptides of the HIV-1 PR. Such complementarity is usually not possible for protected peptide segments and points to the unique advantage of using unprotected peptide segments as building blocks in our strategy. This ring forming reaction has also found its application in the preparation of large peptide dendrimers with high density.[10]

Unlike the conventional method, our strategy generates an active product because the *O*,*N*-acyl transfer reaction also occurs at acidic condition and is the slow step in our scheme.[5] This greatly simplifies the synthetic scheme by eliminating the harsh deprotection and the necessity of the refolding steps. In our synthesis, autoproteolytic cleavages[11] of synthetic HIV proteases and degradation of synthetic segments occurred almost immediately. To prevent enzymatic cleavages, an aspartic protease inhibitor, acetylpepstatin was added to the ligation solution. The product obtained after HPLC purification was further incubated in 6 M guanidine hydrochloride (pH 5.5)/glycerol solution for 3−4 days to assure the completion of the *O*,*N*-acyl transfer. Because of the large size of HIV-1 protease, the rearranged *N*-acyl product was indistinguishable from the unarranged *O*-acyl product by HPLC. However, studies on small model compounds and peptides have determined that the *O*,*N*-acyl transfer reaction occurred at an optimal pH of around 5 with a $t_{1/2}$ of about 20 h.[5] Furthermore, the formation of the amide bond was confirmed by its stability under saponification

(10) (a) Rao, C.; Tam, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 6975−6976. (b) Jun, S.; Tam, J. P. *J. Am. Chem. Soc.* **1995**, *117*, 3893−3899.

(11) Mildner, A. M.; Rothrock, D. J.; Leone, J. W.; Bannow, C. A.; Lull, J. M.; Reardon, I. M.; Sarcich, J. L.; Howe, W. J.; Tomich, C.-S. C.; Smith, C. W.; Heinrikson, R. L.; Tomasselli, A. G. *Biochemistry* **1994,** *33*, 9405−9413.

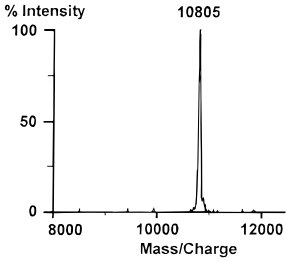


Figure 3. MS analysis of [SPro^{39},Abu^{67,95}]HIV-1 PR by MALDMS.

conditions, since the *O*-acyl intermediate was found susceptible to hydrolysis by 0.05−0.1 N NaOH solution in 5−10 min while the *N*-acyl product was resistant to the same treatment. Because of the unactivated feature of the α-carboxyl group, this *O*,*N*-acyl transfer proceeded smoothly without the side reactions commonly seen in conventional peptide synthesis methods. All three analogs were prepared similarly except for Leu^{38}-amidomethyl-Thz^{39}, which did not involve an acyl rearrangement. All three protease analogs were analyzed by MALDMS, which gave the expected molecular weight. For example, the [SPro^{39},Abu^{67,95}]-HIV-1 PR gave a mass unit of 10 805 that agreed closely with the calculated mean unit of 10 803 (Figure3).

Enzymatic Activity of the Synthetic HIV-1 PR Analogs. Three different peptide substrates were used to determine the activity of these synthetic protease analogs. As shown in HPLC analysis (Figure 4), these HIV-1 PR analogs were active in catalyzing the hydrolysis of these substrates. Detailed kinetic studies (Table 1) with one of these substrates, the *Gag* p24/p15 synthetic peptide (substrate IV, Figure 4C), showed that all synthetic protease analogs exhibited binding affinity (K_m) similar to that of the native enzyme, but showed different catalytic activity (V_{max}): ∼100% with [SPro^{39},Abu^{67,95}]HIV-1 PR, ∼70% with Ala^{38} and ∼30% with the Leu^{38}-amidomethyl-Thz^{39} analog. The catalytic activity (V_{max} or converted k_{cat}) obtained for the native enzyme as well as for our synthetic analogs in this study are, however, lower than those determined by continuous spectrophotometry in the literature.[12] This may be due to different assay methods and conditions. For example, the assay temperature (22 °C) in our study was lower than that (37 °C) used by others. Another reason may be the method for determining the enzyme concentration. Instead of using active-site titration for the folded enzyme solution, we measured the UV absorption of the denatured solution and assumed that all the material would be refolded to the active form when diluted into the assay buffer, which usually leads to overestimation of the actual active concentration. Nevertheless, we used identical assay procedures and conditions for all enzyme preparations including the native HIV-1 PR to provide a consistent basis for comparison.

It is not surprising that [SPro^{39},Abu^{67,95}]HIV-1 PR retained full activity when compared to the native enzyme, because the peptide backbone structures of the wild-type protease and this analog are identical and the (hydroxymethyl)thioproline is structurally a close approximation of the Pro residue. The epimeric hydroxymethyl group is on the side chain Pro ring and would not be expected to change the overall conformation of a proline-containing peptide chain, since Pro often occurs at the reverse turn and is solvent exposed. In the Ala^{38} analog, the change of the isobutyl side chain to a methyl may result in loss of hydrophobic interaction. This loss leads to an increase in the activation energy for the catalytic hydrolysis and accounts for the decrease in catalytic activity. In the case of Leu^{38}-amidomethyl-Thz^{39} analog, there are three atoms inserted between Leu^{38} C═O and Pro^{39} >NH, and this nonpeptide backbone may cause conformational distortion. Surprisingly, this analog still exhibited about 30% activity and comparable binding affinity toward substrate IV as the native enzyme. The relatively flexible structure of the broad loop[8] apparently can tolerate this structural modification.

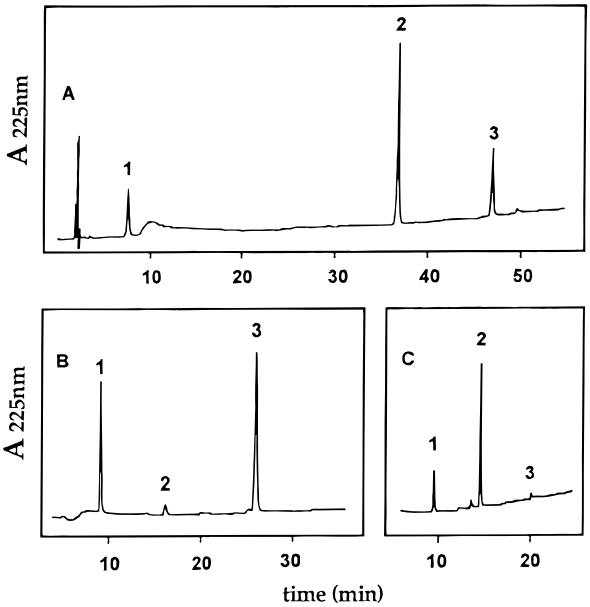


Figure 4. C_{18} RP HPLC analysis of the enzymatic hydrolysis of three different synthetic substrates by the protease analogs. (A) Hydrolysis of 100 μL of 0.2 mM 2-(aminobenzoyl)-Thr-Ile-Nle-Phe(*p*-NO_2)-Gln-Arg-NH_2 in 0.1 M MES (pH 5.5) buffer catalyzed by [SPro^{39},Abu^{67,95}]-HIV-1 PR (∼0.5 μg) at 22 °C for 15 min. Peak 1: H-Phe(p-NO_2)-Gln-Arg-NH_2. Peak 2: 2-(aminobenzoyl)-Thr-Ile-Nle-OH. Peak 3: remaining substrate. HPLC gradient: 0−55% of buffer B in buffer A over 55 min. (B) Hydrolysis of 100 μL of 0.1 mM H-Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val-OH catalyzed by ∼0.35 μg of [Ala^{38},SPro^{39},Abu^{67,95}]-HIV-1 PR in 50 mM acetate buffer (pH 4.7) containing 1 M NaCl at 22 °C for 20 min. Peak 1: H-Val-Ser-Gln-Asn-Tyr-OH. Peak 2: H-Pro-Ile-Val-OH. Peak 3: remaining substrate. HPLC gradient: 0−40% buffer B in buffer A in 40 min. (C) Hydrolysis of substrate IV (100 μL, 30 μM), H-Lys-Ala-Arg-Val-Nle-(*p*-NO_2)-Phe-Glu-Ala-Nle-NH_2 by [Abu^{67,95},SPro^{39}]HIV-1 PR (∼0.2 μg) in 50 mM NaOAc buffer, pH 4.7, with 1 M NaCl and 10% glycerol for 15 min. Peak 1: H-Lys-Ala-Arg-Val-Nle-OH. Peak 2: H-(*p*-NO_2)-Phe-Glu-Ala-Nle-NH_2. Peak 3: remaining substrate. HPLC gradient: 5−65% buffer B in buffer A in 30 min.

These results show that protease analogs with a pseudoproline structure in the sequence are able to fold, dimerize correctly into their active conformation, and retain nearly full catalytic activity. It is also interesting to note that our HIV-1 protease analogs already possess partial enzymatic activity at their *O*-acyl ester stage as concluded from results obtained from the Leu^{38}-$NHCH_2$-Thz^{39} analog. However, the *O*,*N*-acyl transfer was necessary to render these analogs fully active.

Discussion

A central element in achieving orthogonality in peptide bond formation between unprotected segments is the exploitation of proximity effect. Thus, our strategy focuses on side chain nucleophile capture for the initial bond formation, an important contributing factor to achieve orthogonality which leads to the subsequent entropy-driven amide bond formation by intramolecular acyl transfer. Interestingly, such a transfer reaction is naturally observed in the posttranslational protein splicing[13] in which the ultimate excision of an intervening sequence involves

(12) Tummino, P. J.; Ferguson, D.; Hupe, L.; Hupe, D. *Biochem. Biophys. Res. Commun.* **1994,** *200,* 1658−1664.

Table 1. Kinetic Parameters of the Synthetic HIV-1 PR Analogs on the Hydrolysis of Substrate IV in Comparison with Recombinant HIV-1 PR

Protease	38-39 linkage	K_m (μM)	V_{max} (μmol/min.mg)	V_{max}/K_m (μmol/min.mg.M)
rHIV-1 PR	-Leu^{38}-N(39)-C(=O)-	10.1	3.43	3.40×10^5
[SPro^{39}, Abu^{67,95}]-HIV-1 PR	-Leu^{38}-N(OH, S, 39)-C(=O)-	11.9	3.96	3.33×10^5
[Ala^{38}, SPro^{39}, Abu^{67,95}]-HIV-1 PR	-Ala^{38}-N(OH, S, 39)-C(=O)-	8.2	2.26	2.76×10^5
[Leu-NHCH$_2$-Thz$^{38-39}$, Abu^{67,95}]-HIV-1 PR	-Leu^{38}-NH-(S, 39, HN)-C(=O)-	11.4	1.14	1.00×10^5

Kinetic studies were performed at 22 °C under steady state condotions and substrate hydrolysis was analyzed by RP-HPLC. See Experimental Section for details. Enzyme concentrations were measured by UV spectrophotometry on denatured enzyme preparations in 6M guanidine hydrochloride.

an *N*,*O*(*S*)-acyl transfer between the amide and ester linkages of Ser, Thr, or Cys residue. Similar *N*,*O*-acyl interconversions are observed in peptides with Gly-Ser/Thr sequences in strongly acidic and mildly basic conditions.[14] Thus, the ester formation of an acyl component with the side chains of Ser, Thr, or Cys in the amine component leads to amide bond formation through the proximity-driven *O*(*S*),*N*-acyl transfer. This principle was first utilized for peptide synthesis by Kemp *et al.*[4] and recently adopted by us[5] and Dawson *et al.*[6]

At present, there are three strategies that utilize the proximity-driven principle besides those using conformation-assisted couplings.[15] First, Kemp used a tricyclic aromatic template containing an acyl ester aryl thiol to capture an activated cysteinyl side chain through a disulfide exchange to place both the α-amine and acyl component in close proximity to effect the amide bond formation. The tricyclic template is then removed by reduction.[4] Second, as shown in this work, our strategy uses a side chain directed strategy to capture an acyl segment containing an aldehyde group.[5] The third strategy uses either thioester or disulfide exchange without template to facilitate the proximity-driven *S*,*N*-acyl transfer. Dawson *et al.* have developed a thioester exchange between an acyl thioester and a cysteinyl segment,[6] while we have developed similar schemes through a phosphine-assisted reaction as well as thioesterification with a *β*-bromoalanine peptide segment.[16] We have also developed a facile disulfide exchange strategy through an acyl thiocarboxylic acid and an activated unsymmetrical disulfide of the N-Cys amino segment.[17] A common theme in these strategies is the involvement of the N-Cys side chain to form a covalent bond between the acyl and the side chain of the amine segment and then an entropy-driven amide bond formation to produce a cysteine residue at the ligation site. Our strategy produces a proline-like backbone that may have the advantage of being more accessible because of the high frequency of proline residues occurring at reverse turns in protein sequences. Furthermore, by varying the aldehyde moiety, other nongenetically coded amino acids could be generated. The thiazolidines are also reversible to give cysteinyl bonds. These chemical approaches, together with the recently developed enzymatic method,[18] represent important advances in the total synthesis of proteins.

With large unprotected peptide segments and protein domains readily accessible either by solid phase synthesis or recombinant techniques, a chemical ligation method to form an amide backbone between these high molecular weight building blocks would greatly expand our ability to synthesize and modify proteins with noncoded amino acids as well as to design artificial proteins with unusual architectures.[19,20] The syntheses of HIV-1 protease analogs clearly demonstrate the ease and efficiency of our chemical ligation approach. Moreover, our method demonstrates that a pseudoproline structure can be used to replace a proline residue in the sequence of a protein without altering its biological activity. Most importantly, our strategy employs the principle of orthogonality, utilizing the side chain participation in the initial formation of a covalent complex and the subsequent intramolecular acyl transfer to overcome the problems of conventional segment condensation method. In the future, it should open new avenues for other chemical approaches in forming peptide bonds between very large peptide segments.

Experimental Section

General. Analytical C_{18} HPLC was run on a Shimadzu system with a Vydac column (0.46 cm × 25 cm, C_{18} reversed phase) at a flow rate of 1.5 mL/min with a linear gradient of buffer B (60% acetonitrile in H_2O/0.04% TFA) in buffer A (5% acetonitrile in H_2O/0.045% TFA), with UV detection at 225 nm. Mass spectra were obtained with the Matrix-assisted laser desorption mass spectrometry (MALDMS) with

(13) (a) Kane, P. M.; Yamashiro, C. T.; Wolczyk, D. F.; Neff, N.; Goebl, M.; Stevens, T. H. *Science* **1990,** *250*, 651−657. (b) Davis, E. O.; Jenner, P. J.; Brooks, P. C.; Colston, M. J.; Sedgwick, S. G. *Cell* **1992,** *71*, 201−210. (c) Hodges R. A.; Perler, F. B.; Noren, C. J.; Jack, W. E. *Nucleic Acids Res.* **1992,** *20*, 6153−6157.

(14) Sakakibara, S.; Shin, K. H.; Schneider, W.; Hess, G. P. *J. Am. Chem. Soc.* **1962,** *84*, 4921−4928.

(15) (a) Homandberg, G. A.; Chaiken, I. M. *J. Biol. Chem.* **1980**, *255*, 4903. (b) Wallace, C. J. A.; Corthasy, B. E. *Protein Eng.* **1986,** *1*, 23. (c) Proudfoot, A. E. I.; Rose, K.; Wallace, C. J. A. *J. Biol. Chem.* **1989,** *264*, 8764−8770.

(16) (a)Tam, J. P.; Lu, Y.-A.; Liu, C. F.; Shao, J. *Proc. Natl. Acad. Sci. U.S.A.*, in press. (b) Tam, J. P.; et al. In *Proceedings of the 14th American Peptide Symposium*, in press.

(17) (a) Liu, C. F.; Rao, C.; Tam, J. P. In *Proceedings of the 14th American Peptide Symposium*, in press. (b) Liu, C. F.; Rao, C.; Tam, J. P. *Tetrahedron Lett.* In press.

(18) Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. *Science* **1994,** *266*, 243.

(19) Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5409−5413.

(20) Mutter, M.; Vuilleumier, S. *Angew. Chem., Int. Ed. Engl.* **1989,** *28*, 535.

an accuracy of ±0.05%. The calculated mass units were given as average values of the isotopic composition.

Recombinant HIV-1 PR was purchased in its highest grade (affinity-purified, crystallizable grade) from BACHEM Bioscience Inc. (Bubendorf, Switzerland). This material gave a $V_{max} \approx 19.7$ μmol·min^{-1}·mg^{-1} toward substrate IV at 37 °C. Acetylpepstatin was also from the same company. All HIV-1 PR substrates were purchased from BACHEM (Torrance, CA).

Synthesis of HIV-1 PR(1−37)-S(CH$_2$)$_2$CO-Gly-OH (1a) and [Cys^{39},Abu^{67,95}]-HIV-1 PR (39−99) (1d). Both the thioester segment, **1a**, and amine segment, **1d**, were prepared by solid phase synthesis[21] using the conventional Boc-benzyl type protection strategy and the BOP coupling protocol.[22] **1a** was synthesized on an Asn^{37}-S(CH$_2$)$_2$CO-Gly-OCH$_2$-Pam resin which was obtained through coupling of Boc-Asn-S(CH$_2$)$_2$COOH to H-Gly-OCH$_2$-Pam resin as reported.[2c] **1d** was synthesized on a Phe^{99}-OCH$_2$-Pam resin.[23] The final cleavage was performed using the low−high HF cleavage procedure.[24] After HF cleavage, the crude products were extracted with 50% AcOH/H$_2$O, dialyzed against decreasing concentrations of AcOH to 10%, and then purified by preparative HPLC with a Vydac C$_{18}$ column. The purified products gave a single peak in analytical RP HPLC (Figure 2A,D, peaks 1 and 4 respectively): MS [M + H]$^+$ *m*/*z* (**1a**) 4281.8 (calcd), 4282.0 (found); *m*/*z* (**1d**) 6529.7 (calcd), 6530.0 (found).

Preparation of α-amino(acyloxy)- and Amidoacetaldyde Dimethyl Acetal. Aldehyde Introduction. α-amino(acyloxy)acetaldyde dimethyl acetal was prepared by esterification of the Z-protected α-amino acid cesium salt with bromoacetaldehyde dimethyl acetal in DMF followed by catalytic hydrogenolysis as reported previously.[5] α-Amino(acylamido)acetaldehyde was prepared by coupling of Z-amino acid to aminoacetaldehyde dimethyl acetal followed by catalytic hydrogenolysis. The thioester segment, **1a**, was dissolved in DMSO (5−10 mM) containing 0.5−1.0 M of the amino acid derivative, α-(aminoacyloxy(amido))acetaldehyde dimethyl acetal, HOSu (0.5−1.0 M), and dimethyl sulfide (∼50 mM). $AgNO_3$ predissolved in DMSO (2 equiv to the thioester segment) was added. In general, the reaction was completed in 20−60 min with >80% yield as determined by analytical HPLC: MS [M + H]$^+$, HIV-1 PR(1−38)-OCH$_2$CH(CH$_3$)$_2$, *m*/*z* 4337.9 (calcd), 4338.0 (found); [Ala^{38}]HIV-1 PR(1−38)-OCH$_2$-CH(CH$_3$)$_2$, *m*/*z* 4296.8 (cald.), 4296 (found); HIV-1 PR (1−38)-NHCH$_2$-CH(CH$_3$)$_2$, *m*/*z* 4336.9 (calcd), 4338.0 (found).

The acetal deprotection was performed using 95% trifluoroacetic acid in H$_2$O at 0 °C for 5−10 min to give the desired glycol aldehyde acyl segment, **1c**, as shown in Figure 2C for the synthesis of [SPro^{39},Abu^{67,95}]HIV-1 PR. The reaction mixture was immediately diluted with 10-fold 30% acetonitrile/H$_2$O and applied to an HPLC column for purification. The collected fractions were concentrated by a centrifuge vacuum and used immediately for the next step.

Ring Formation and Acyl Transfer. The amino segment, [Cys^{39},Abu^{67,95}]HIV-1 PR (39−99), **1d**, was dissolved in a minimal amount of 60% acetonitrile/H$_2$O (0.1% TFA) and mixed with the acyl segment containing a glycol aldehyde ester, **1c**. The solution was diluted with H$_2$O (0.1% TFA) to the desired concentrations of the two segments (∼50 μM for **1c** and ∼100 μM for **1d**). Acetylpepstatin (0.5 mM) and EDTA (0.5 mM) were then added. The pH was adjusted by addition of solid sodium acetate to 3.0−4.0 (by pH paper). The thiazolidine ring product, **1e**, was 60−80% formed as determined by HPLC in 5−10 h based on the acyl segment of glycol aldehyde ester (the amino segment was used in 2−3 equiv). The ring formation would be faster at higher pH as stated in previous studies,[5] but in the case of HIV-1 protease, pH >4 led to precipitation of both segments (especially the amino segment). Some precipitation (mostly of the ligation product) was indeed observed during the reaction which was dissolved by guanidine hydrochloride before HPLC purification. The ligation product was isolated by HPLC, lyophilized, and redissolved in the denaturing condition of 6 M guanidine hydrochloride (pH 5.5)/40% glycerol (v/v) solution to complete the *O*,*N*-acyl transfer reaction. The protein concentration was determined by UV spectroscopy in its 6 M guanidine hydrochloride solution prior to the addition of glycerol using a $\epsilon_{280} = 25\,000$ M^{-1}·cm^{-1}.[9c] This concentration then served as basis for determination of enzymatic activities and for calculation of kinetic parameters such as V_{max}, assuming that 100% of the material would refold and remained active in enzymatic assays. The guanidine/glycerol solution served as a storage buffer for the protease and to preserve enzymatic activity. A >90% yield of the rearranged *N*-acyl product, **1f**, was obtained after incubating for 3−4 days in the above buffer. These enzyme solutions were stored at −70 °C. The enzyme sample was active immediately when diluted into the assay buffer as shown by the hydrolysis of the synthetic peptides derived from the HIV *Gag* protein sequences (Figure 4): MS [M + H]$^+$, [SPro^{39},Abu^{67,95}]HIV-1 PR (monomer), *m*/*z* 10 803 (calcd), 10 805 (found); [Ala^{38},SPro^{39},Abu^{67,95}]-HIV-1 PR, *m*/*z* 10 761 (calcd), 10 764 (found); [Leu-NHCH$_2$-Thz$^{38-39}$,Abu^{67,95}]HIV-1 PR, *m*/*z* 10 802 (calcd), 10 806 (found).

HPLC Analysis of the Enzymatic Hydrolysis of HIV-1 PR Substrates by the Synthetic Protease Analogs. Three different synthetic substrates derived from the *Gag* p17/p24 and p24/p15 cleavage sites were used for qualitative HPLC assays: H-Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val-OH (substrate VIII) from the natural p17/p24 cleavage site, H-Lys-Ala-Arg-Val-Nle-Phe(*p*-NO$_2$)-Glu-Ala-Nle-NH$_2$ (substrate IV) and 2-aminobenzoyl-Thr-Ile-Nle-Phe(*p*-NO$_2$)-Gln-Arg-NH$_2$[25a] derived from the two p24/p15 cleavage sites. Synthetic protease analogs efficiently catalyzed the hydrolysis of these substrates in a weakly acidic buffer. Typically, for HPLC assay, an aliquot of protease analog solution in 6 M guanidine hydrochloride/40% glycerol was directly added to an appropriate assay buffer with a substrate for an indicated time period at 22 °C and stopped with 20% TFA/H$_2$O before applying to HPLC column (Figure 4). The hydrolytic products were collected from analytic HPLC and confirmed by MS analysis.

Kinetic Studies. Substrate IV, derived from the first cleavage site of the *Gag* p24/p15,[25b] was used for all kinetic studies due to its excellent solubility in aqueous buffer. Assay buffer was 50 mM sodium acetate, pH 4.7, containing 5 mM DTT and 2 mM EDTA, with an ionic strength equal to 1 M NaCl and with glycerol added to give a 10% (v/v). DTT and EDTA were added to prevent the oxidation of the free Cys^{67,95} residues in the recombinant HIV-1 protease. This buffer was used for comparison in all kinetic studies. The enzyme sample was activated by simple dilution into the assay buffer. No further refolding/dialysis procedure was employed. The enzyme concentrations typically used for kinetic assays were 0.02−0.05 μM, and the substrates were between 5 and 60 μM. The reaction proceeded for 3−6 min during which less than 25% of the substrate was hydrolyzed and was stopped with 20% TFA in H$_2$O. Initial rates were calculated on the basis of the UV absorption (area of HPLC peaks) at 225 nm of the substrate and hydrolyzed products. The ratio of UV absorption of the substrate to the hydrolyzed C-terminal half, H-Phe(*p*-N$_2$O)-Glu-Ala-Nle-NH$_2$ was determined to be 1:0.68, calibrated in the presence of internal standards. Typically, five to six data points were used for each experiment. Kinetic parameters were obtained by fitting data into the Michaelis−Menten equation using the Lineweaver−Burk plotting method. All kinetic studies were performed at 22 °C.

Acknowledgment. This work was in part supported by grants from USPHS, AI 28701 and CA 36544.

JA952790W

(21) Merrifield, R. B. *J. Am. Chem. Soc.* **1963,** *85*, 2149−2154.
(22) Castro, B.; Dormoy, J. R. G.; Evin, G.; Selve, C. *Tetrahedron Lett.* **1975**, 1219−1222.
(23) Mitchell, A. R.; Kent, S. B. H.; Engelhard, M.; Merrifield, R. B. *J. Org. Chem.* **1978,** *43*, 2845−2852.
(24) Tam, J. P.; Heath, W. F.; Merrifield, R. B. *J. Am. Chem. Soc.* **1983,** *105*, 6442−6455.
(25) (a) Toth, M. V.; Marshall, G. R. *Int. J. Pept. Protein Res.* **1990,** *36*, 544−550. (b) Richards, A. D.; Phylip, L. H.; Farmerie, W. G.; Scarborough, P. E.; Alvarez, A.; Dunn, B. M.; Hire, P.-H.; Konvalinka, J.; Strop, P.; Pavlickova, L.; Kostka, V.; Kay, J. *J. Biol. Chem.* **1990,** *265*, 7733−7736.