Chemical Ligation of Cysteine-Containing Peptides: Synthesis of a 22 kDa Tethered Dimer of HIV-1 Protease

Manuel Baca, Tom W. Muir, Martina Schnölzer,† and Stephen B. H. Kent*

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

Received September 26, 1994[⊗]

Abstract: Thioester-forming chemoselective reaction of unprotected peptide fragments containing cysteine residues has been investigated. This work shows that free sulfhydryl groups are compatible with the reactive components of thioester-forming ligation chemistry. This allows conjugation by chemical ligation of cysteine or other thiol-containing peptides, followed by postligation disulfide bond formation to form folded protein domains, or large multisubunit synthetic proteins. Under acidic conditions, peptides bearing bromoacetyl or α-thiocarboxylate groups did not undergo intermolecular reaction with the sulfhydryl group of cysteine. Intramolecular reaction also did not occur, provided a sufficient number of intervening residues separated the functionalities. The results of these studies have been used in the design and synthesis of a 22 kDa tethered dimer HIV-1 protease analogue, prepared by the convergent chemical ligation of four unprotected peptide segments. Two pairs of ~50 residue peptides were ligated via formation of thioester bonds to form the individual monomer polypeptide chains. The ligated monomers each possessed a two residue extension, either at the N-terminal or at the C-terminal, containing an unprotected sulfhydryl group. These were subsequently linked via directed formation of a disulfide bond. The resulting enzyme analogue retained full catalytic activity, showing that the placement of the backbone thioesters and the disulfide bond were in functionally unimportant parts of the molecule.

Introduction

Recently, we introduced a novel route to the total synthesis of proteins—the chemical ligation of large unprotected peptide segments.[1] Chemical ligation employs unique functionalities, one on each of two unprotected peptides, with reactivities tailored to one another. These mutually reactive moieties undergo chemoselective reaction but do not react with the other functional groups found in peptides. Hence there is no need for terminal or side-chain protection of the peptide segments. In the original ligation chemistry,[1] two peptides were joined to one another by nucleophilic reaction of a C-terminal α-thiocarboxylate on one peptide with an N-terminal α-bromoacyl moiety of a second peptide. A thioester bond was formed at the ligation site. The peptide segments can be readily prepared by solid phase peptide synthesis (SPPS),[2] and the absence of protecting groups ensures that these precursor segments are soluble and that they can be purified and characterized by conventional means.

The range of protein targets which have now been prepared by thioester-forming chemical ligation includes analogues of the HIV-1 protease,[1,3] the fibronectin 10F3 module,[4] and a four-helix bundle template-assembled synthetic protein.[5] At the acidic pH's used in the thioester-forming chemical ligation reaction, the common functional groups (α-NH_3^+, ϵ-NH_3^+, -OH, Im, α-COO^-, β,γ-COO^-, etc.) found in protein-derived peptide segments are non-nucleophilic. Thus far, however, none of these ligated protein analogues has contained cysteine residues. The side-chain sulfhydryl group of cysteine is nucleophilic, and a potential complication of thioester-forming chemical ligation is the undesired attack of this side chain upon the α-bromoacyl group, resulting in formation of a thioether bond. Indeed the reaction of a thiol with an α-bromoacyl group has been used to generate cyclic peptides,[6] to couple antigenic peptides to carrier proteins,[7] and to ligate two peptides in the synthesis of an artificial neoprotein mimic of the cytoplasmic domain of a dimeric integrin receptor.[8]

It was important to establish whether or not the use of unprotected cysteine is compatible with the thioester-forming chemical ligation approach to the synthesis of proteins. The ability to incorporate unprotected cysteine (or unnatural sulfhydryl-bearing residues) in a chemoselective ligation strategy for total synthesis of proteins would render the method more general for proteins regardless of sequence. Furthermore, free sulfhydryl groups could subsequently be oxidized to form intra- or intermolecular disulfide bonds, or alternatively serve as "handles" for postsynthetic chemical modification of the target protein.

In this paper, we show that in general the presence of free sulfhydryl groups is compatible with the thioester-forming ligation chemistry. This was explored by studies on model peptides and applied to the preparation of a 22 kDa HIV-1 protease (HIV-1 PR) analogue. The 11 kDa monomers containing terminal sulfhydryl groups were prepared by thioester-forming ligation. Subsequent oxidation gave a disulfide-linked tethered dimer enzyme analogue with catalytic properties indistinguishable from the native homodimeric molecule.

† Current address: Division of Cell Biology, German Cancer Research Center, D-69120 Heidelberg, Germany.

⊗ Abstract published in *Advance ACS Abstracts*, February 1, 1995.

(1) Schnölzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221−225.
(2) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180−193.
(3) Baca, M.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11638−11642.
(4) Williams, M. J.; Muir, T. W.; Ginsberg, M. H.; Kent, S. B. H. *J. Am. Chem. Soc.* **1994**, *116*, 10797.
(5) Dawson, P. E.; Kent, S. B. H. *J. Am. Chem. Soc.* **1993**, *115*, 7263−7266.
(6) Robey, F. A.; Fields, R. L. *Anal. Biochem.* **1989**, *177*, 373−377.
(7) Kolodny, N.; Robey, F. A. *Anal. Biochem.* **1990**, *187*, 136−140.
(8) Muir, T. W.; Williams, M. J.; Ginsberg, M. H.; Kent, S. B. H. *Biochemistry* **1994**, *33*, 7701−7708.

0002-7863/95/1517-1881$09.00/0

Table 1. Synthetic Peptide Segments Prepared, and Their Characterization by ESMS

no.	sequence	calculated mass (Da) [isotope comp] mono. av.	experimental mass (Da)
	Model Peptides		
1	BrAc-Arg-Ala-Cys-Ala-Arg-Gly[a]	752.2; 753.7	752.5 ± 0.7
2	Arg-Ala-Cys-Ala-Arg-GlyCOSH	648.3; 648.7	648.5 ± 0.7
3	BrAc-Thr-Pro-Gly-Ser-Arg-Asn-Leu-Cys-Asn [Lysozyme BrAc(69−77)]	1080.4; 1082.0	1081.0 ± 1.1
	HIV-1 Fragments		
4	(1−51)α-COSH	5612.0; 5615.6	5614.3 ± 1.5
5	BrAc(53−99)Gly-$NHCH_2CH_2SH$	5347.9; 5352.1	5351.5 ± 0.9
6	$HSCH_2CO$-Gly(1−51)α-COSH	5743.0; 5746.8	5747.3 ± 1.3
7	BrAc(53−99)	5231.8; 5236.0	5234.4 ± 0.6

[a] Abbreviation: BrAc = bromoacetyl.

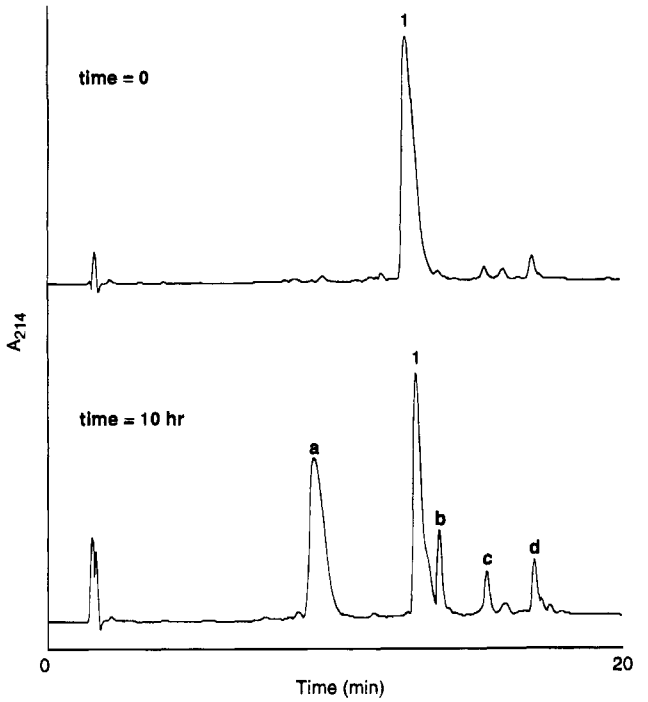

Figure 1. Breakdown of BrAc-Arg-Ala-Cys-Ala-Arg-Gly (**1**) in 100 mM sodium phosphate, pH 3.2, at 25 °C. Peptide was dissolved at 5 mg/mL, and the time-dependent decomposition was analyzed by HPLC, using a 5−30% solvent B gradient over 20 min. (Upper) HPLC chromatogram of **1** immediately upon dissolution. (Lower) HPLC chromatogram after 10 h. Each peak was collected and further analyzed by ESMS. On the basis of the masses obtained and referencing these to the mass (M) of **1**, the identity of each breakdown product was determined. While **a** was found to be monomeric, **b**−**d** were dimeric species. Formation of each product was rationalized as follows: **a** (average mass = M − 80), intramolecular thioether formation; **b** (average mass = 2M − 160), bis(intermolecular) thioether formation; **c** (average mass = 2M − 80), intermolecular thioether formation; **d** (mass = 2M − 2), intermolecular disulfide.

Results

Stability of Bromoacetyl and α-Thioacid Peptides Containing Cysteine. In order to facilitate study of ligation reactions, a series of model peptides were synthesized (Table 1), each designed to be highly soluble under the moderately acidic conditions used during ligation. The bromoacetylated peptide **1** (BrAc-Arg-Ala-Cys-Ala-Arg-Gly) was synthesized and purified by standard optimized protocols,[2] and its stability in 0.1 M sodium phosphate, pH 3.2, was determined using reverse phase HPLC and electrospray mass spectrometry (ESMS). Peptide **1** was observed to be unstable under these conditions, as indicated by the appearance of four new peaks on HPLC (Figure 1). By far, the predominant of these (peak **a**) corresponded to a monomeric cyclic species, produced via the *intramolecular* attack of the cysteine sulfhydryl group on the bromoacetyl function with displacement of HBr. Other minor breakdown products were also observed, corresponding to formation of dimeric species with concomitant loss of two molecules of HBr (peak **b**), one molecule of HBr (peak **c**), or two H^+ (disulfide linked dimer, peak **d**). After 10 h in solution at pH 3.2, approximately 65% of peptide **1** had been converted into these combined breakdown products.

In contrast to **1**, peptide **3** (BrAc-Thr-Pro-Gly-Ser-Arg-Asn-Leu-Cys-Asn) in which the bromoacetyl group and cysteine are separated by seven intervening residues, did not undergo any detectable intramolecular cyclization after 24 h in solution at pH 3.2, 25 °C. Additionally, no dimeric breakdown products corresponding to intermolecular thioether formation were observed, suggesting that the minor appearance of these in solutions of **1** may be due to the small size (higher concentration) of that peptide.

The α-thioacid-containing peptide **2** (Arg-Ala-Cys-Ala-Arg-GlyCOSH) was found to have undergone only minor side reactions after 24 h in solution at pH 3.2, 25 °C. Small amounts of hydrolysis to the α-carboxylic acid (5%) and disulfide-linked dimerization (9%) were observed over a 24 h period. These minor side reactions are not atypical for peptides containing either α-thioacid or cysteine residues. Also observed was the appearance of an HPLC peak corresponding to material of mass 34 Da less than the parent peptide **2**. Again this component comprised only a small amount (8%) of the final mixture. Loss of a single molecule of H_2S would account for such a mass, suggesting that an intramolecular thioesterification had occurred between the cysteine sulfhydryl and the thioacid function, thus yielding a cyclic peptide thioester.

Chemoselective Ligation of Cysteine-Containing Peptides. To demonstrate that an unambiguous chemical ligation can be performed when both peptide segments contain unprotected cysteine sulfhydryls, peptides **2** and **3** were ligated in aqueous phosphate buffer at pH 3.2 (Figure 2). This ligation reaction proceeded so quickly that even immediate HPLC analysis upon initiating the reaction showed considerable product formation. Subsequent analysis of the reaction mixture after 1 h showed almost quantitative formation of the ligation product. The ligation rate was sufficiently rapid that none of the previously observed minor breakdown products of peptide **2** accumulated prior to its near total consumption in the reaction. There was no evidence to suggest that an alternate ligation had occurred by reaction of the cysteine sulfhydryl with bromoacetyl moiety as only one reaction product was observed by HPLC. Similarly, the cysteine residues in the product peptide did not react with either the excess bromoacetyl peptide or the thioester bond created upon ligation, even after 7 h.

Design and Synthesis of the Tethered Dimer HIV-1 PR. A practical application of the compatibility of free sulfhydryl groups with the thioester ligation chemistry involved the design

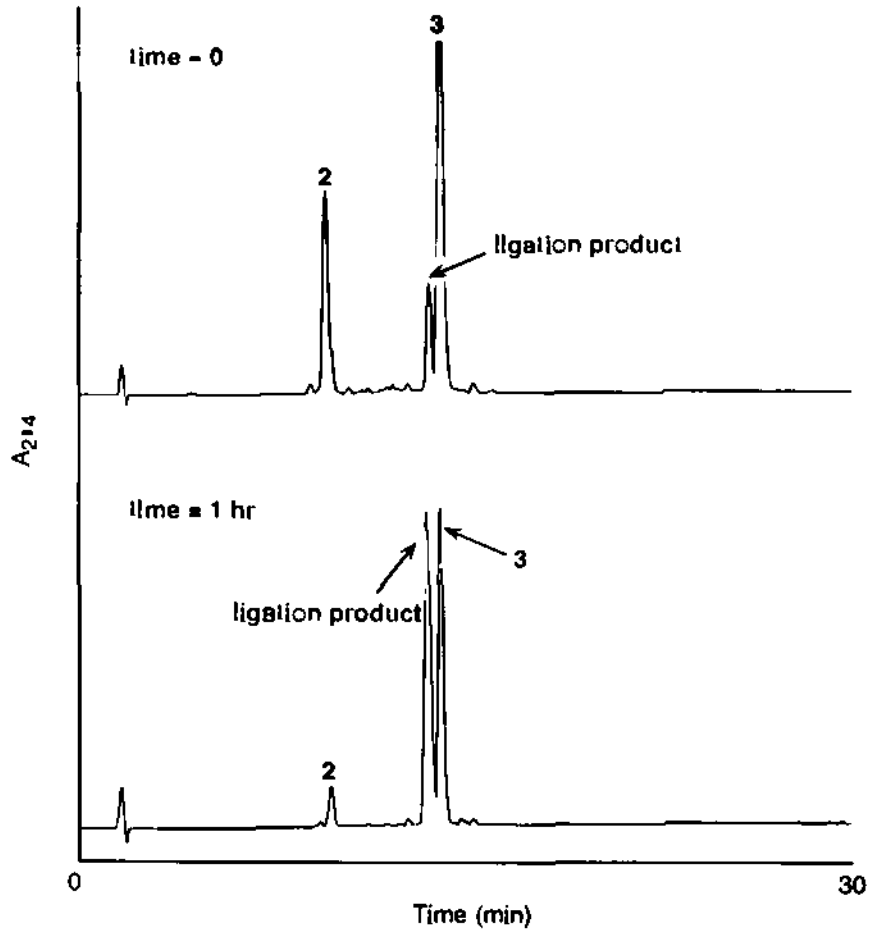

Figure 2. Chemoselective ligation of Arg-Ala-Cys-Ala-Arg-GlyCOSH (2) and BrAc-Thr-Pro-Gly-Ser-Arg-Asn-Leu-Cys-Asn (3). Peptides were dissolved together in 100 mM sodium phosphate, pH 3.2, and stirred at room temperature. The ligation reaction was monitored by HPLC using a 0−67% solvent B gradient over 60 min. Peaks were collected and analyzed by ESMS. (Upper) HPLC chromatogram of the ligation mixture immediately upon dissolving the combined peptide mixture. Even in the few seconds between initiating the ligation mixture and injecting a sample onto the HPLC, considerable ligation can be seen to have taken place. (Lower) HPLC chromatogram after 1 h reaction time. At this time the reaction was essentially complete. Only one ligation product was formed, indicating exclusive reaction between the bromoacetyl and α-thioacid groups. Excess of bromoacetylated peptide remained after the reaction. No new peaks were observed at longer time (7 h), indicating no reaction between sulfhydryl groups and the excess bromoacetyl peptide or the newly formed thioester bond.

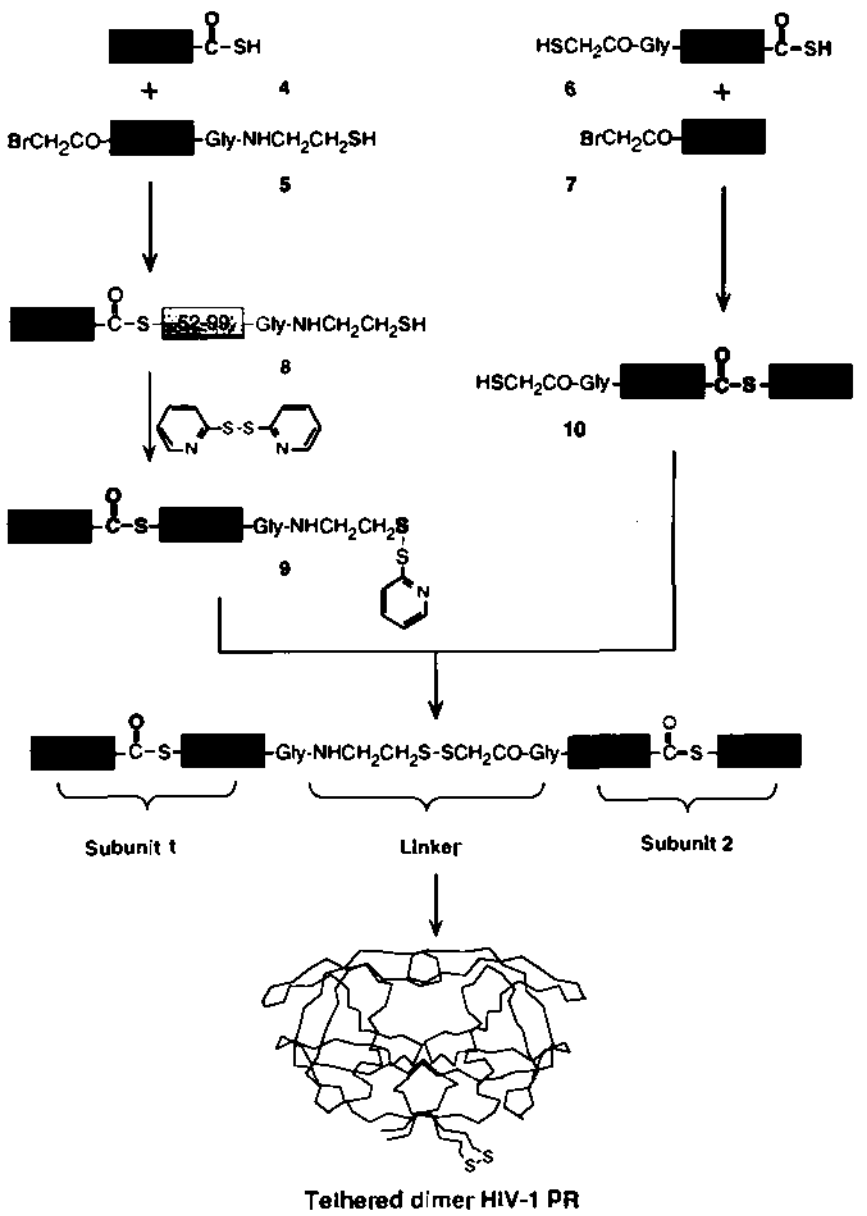

Figure 3. Schematic representation of the total chemical synthesis of the 22 kDa tethered dimer HIV-1 PR. Unprotected peptide segments were ligated via nucleophilic attack of the deprotonated α-thioacid group on the bromoacetyl moiety to form the two monomers each with a short N- or C-terminal extension containing an unprotected sulfhydryl group. After derivatization of the cysteamine-containing monomer with 2,2′-dipyridyl disulfide, the desired disulfide-linked heterodimer was formed by thiolysis of the *S*-(2-pyridylsulfenyl)cysteamine derivative. The tethered dimer comprised the two 99 amino acid subunits of HIV-1 PR (residues 1−99 and 1′−99′), covalently linked by the four residue intersubunit link -Gly-$NHCH_2CH_2SSCH_2C(O)$-Gly-. Each subunit contained a single backbone thioester bond between residues 51 and 52. The synthetic tethered dimer was folded to give a fully active HIV-1 PR analogue, with catalytic properties indistinguishable from those of the wild-type enzyme.

and synthesis of a fully functional tethered dimer HIV-1 PR analogue. The 202 residue polypeptide chain was constructed via a convergent segment chemical ligation approach (Figure 3) in which the individual 101 residue monomer polypeptide chains were first assembled by thioester-forming chemical ligation of two ~50 residue peptides. The nonequivalent subunits were then covalently linked via directed formation of an intersubunit disulfide bond.

The HIV-1 PR segments BrAc(53−99)Gly-$NHCH_2CH_2SH$ and $HSCH_2CO$-Gly(1−51)α-COSH each contained free sulfhydryl groups in addition to bromoacetyl or α-thioacid moieties. They were both completely stable to the acidic conditions used for HF cleavage and workup, followed by HPLC purification.[2] The purified and lyophilized peptides did not show any sign of breakdown, even after storage for 1 month at −20 °C. This again emphasizes the stability of thiol-containing bromoacetyl and α-thioacid peptides, provided these groups are not in close proximity within the sequence.

The chemoselective ligation of bromoacetyl and α-thioacid peptides to form the individual subunit polypeptide chains occurred without any competing side reaction from the terminal thiol groups, as judged by formation of only one ligation product when analyzed by HPLC (Figure 4A). The cysteamine-containing monomer **8** was treated with 2,2′-dipyridyl disulfide[9] directly in the crude ligation mixture, having first quenched the unreacted α-thioacid groups still present in excess **4** by addition of bromoacetic acid. This obviated the need for prior isolation of the free cysteamine peptide and gave a superior yield of the *S*-(2-pyridylsulfenyl)cysteamine derivative. No side reactions of the 2,2′-dipyridyl disulfide reagent or the *S*-(2-pyridylsulfenyl)cysteamine group with the thioester bond were observed. The resulting thiol- or disulfide-containing thioester-ligated monomers were stable to the conditions of reverse phase HPLC purification, and the lyophilized products showed no breakdown after several weeks at −20 °C.

In the final covalent linking of the monomer subunits (Figure 4B), thiolysis of the *S*-(2-pyridylsulfenyl)cysteamine peptide **9** by the mercaptoacetyl functionality on the second monomer **10** led to the directed formation of a disulfide bond, thereby generating the final 22 kDa tethered dimer with the desired mainchain topology. The product was then purified and characterized (Figure 4C). To prevent the potential autolytic breakdown of this product, the HPLC fraction obtained after purification was diluted with 6 M guanidinium hydrochloride (GuHCl), concentrated, and dialyzed against 5 M GuHCl, pH 3.8. In the denatured state, HIV-1 PR is completely stable to

(9) Grassetti, D. R.; Murray, J. F. *Arch. Biochem. Biophys.* **1967**, *119*, 41−49.

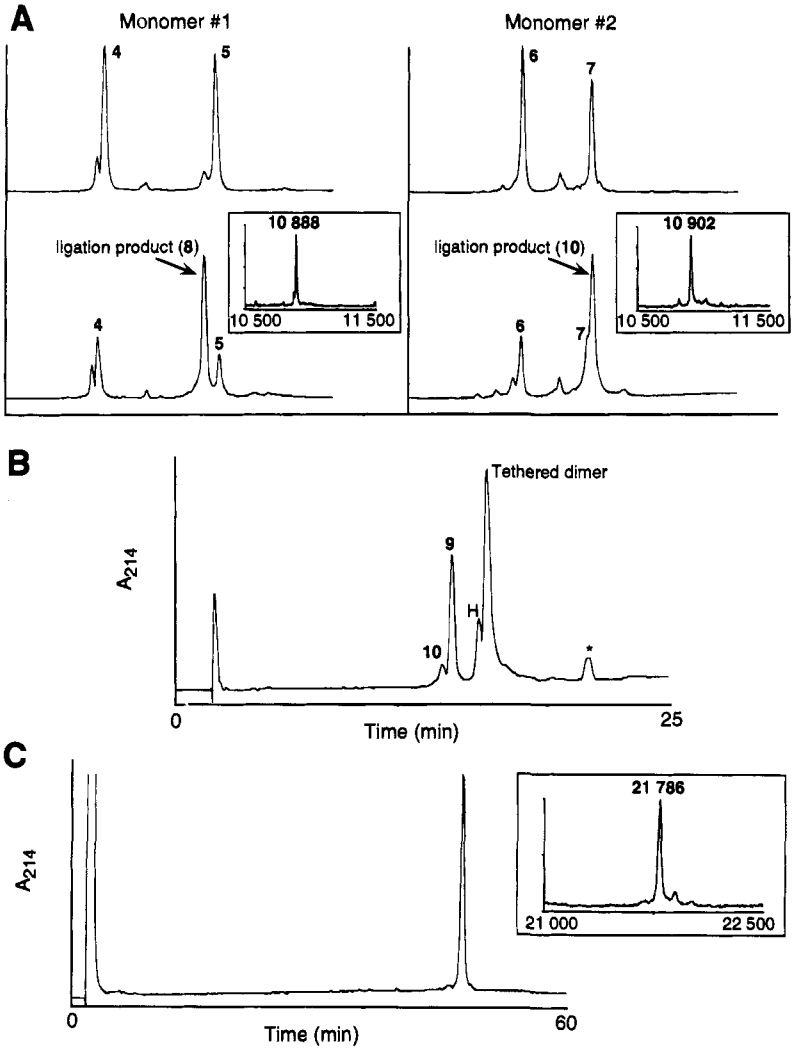

Figure 4. Synthesis and characterization of tethered dimer HIV-1 PR. (**A**) Synthesis of HIV-1 PR monomers by thioester-forming chemical ligation. (Left) The upper HPLC trace represents the combination of peptides **4** [(1−51)αCOSH] and **5** [BrAc(53−99)Gly-NHCH$_2$CH$_2$SH] at zero time. The lower trace shows the formation of the HIV-1 PR monomer segment (1−51)-[COS]-(52−99)Gly-NHCH$_2$CH$_2$SH (**8**) after overnight reaction of peptides **4** and **5** at 4 °C, pH 3.8. The inset on the lower trace corresponds to the mass of peak **8** as analyzed by ESMS. (Right) As for the left pair of traces, except representing the synthesis of the monomer segment HSCH$_2$CO-Gly(1−51)-[COS]-(52−99) (**10**) from peptides **6** [HSCH$_2$CO-Gly(1−51)αCOSH] and **7** [BrAc(53−99)]. (**B**) Intersubunit ligation of HIV-1 PR monomers by directed disulfide formation. HPLC trace shows formation of tethered dimer HIV-1 PR 4 h after combination of the purified monomer segments **9** [where **9** is the *S*-(2-pyridylsulfenyl)-cysteamine derivative of **8**] and **10**. The peak labeled "H" represents minor formation of the homodimeric disulfide of peptide **8**, while the peak labeled with an asterisk is a column artifact. (**C**) Characterization of the synthetic tethered dimer HIV-1 PR. An aliquot of the purified enzyme stored in 5 M GuHCl, pH 3.8, was analyzed by reverse phase HPLC using a linear gradient of 0−67% solvent B over 60 min. The presence of only a single sharp peak is an indication of the purity and stability of the synthetic protein when stored under these conditions. (Inset) Covalent characterization of tethered dimer HIV-1 PR by ESMS. The protein peak was collected from the HPLC in a single fraction and analyzed by ESMS. The reconstructed plot of intensity versus molecular mass was derived from the raw mass spectral data. The experimentally determined mass of tethered dimer HIV-1 PR (21 786 ± 3 Da) is in good agreement with the calculated mass of 21 787 Da (average isotope composition).

autolysis.[10] After storage in 5 M GuHCl, pH 3.8 for 1 month at 4 °C, no chemical or autolytic breakdown products could be detected by HPLC.

Enzyme Folding and Activity. Initial attempts to fold the synthetic enzyme using protocols previously described for thioester-containing analogues of HIV-1 PR[3,10] led to almost total loss of the protein by precpitation and adsorption. This result is in contrast to the high yields (typically 70%) obtained for folding of dilute solutions of the thioester-containing 99 residue polypeptide [(COS)51-52]HIV-1 PR to form the homodimeric folded enzyme molecule.[1,3,10] To reduce the level of protein loss occurring during the folding process, particularly from adsorption, 0.5 mg/mL of BSA was added to the denatured enzyme prior to folding by dialysis. Although considerable material was still lost, a modest yield of folded enzyme (~20%) was obtained which was sufficient for characterization of the enzymatic properties. Solubilization of the precipitant from this folding in 6 M GuHCl, followed by HPLC and ESMS analysis, showed that this material had not undergone hydrolysis at the thioester bond or any other chemical modification. This was consistent with the known stability of thioester-containing peptides below pH 7.[1,10] Thus, the low yield obtained from folding was not a consequence of any chemical instability in the synthetic tethered dimer. Clearly, the lowered folding yield is due to the presence of the intermonomer tether since otherwise identical thioester-containing monomers fold correctly and in high yield.[10]

Enzymatic properties of the folded tethered dimer HIV-1 PR were defined by kinetic studies based on cleavage of a fluorogenic substrate,[11] as well as by the specific cleavage of synthetic peptide analogues of the matrix/capsid and capsid/

(10) deLisle Milton, R. C.; Milton, S. C. F.; Schnölzer, M.; Kent, S. B. H. In *Techniques in Protein Chemistry IV*; Angeletti, R. H., Ed.; Academic Press: New York, 1992; pp 257−267.

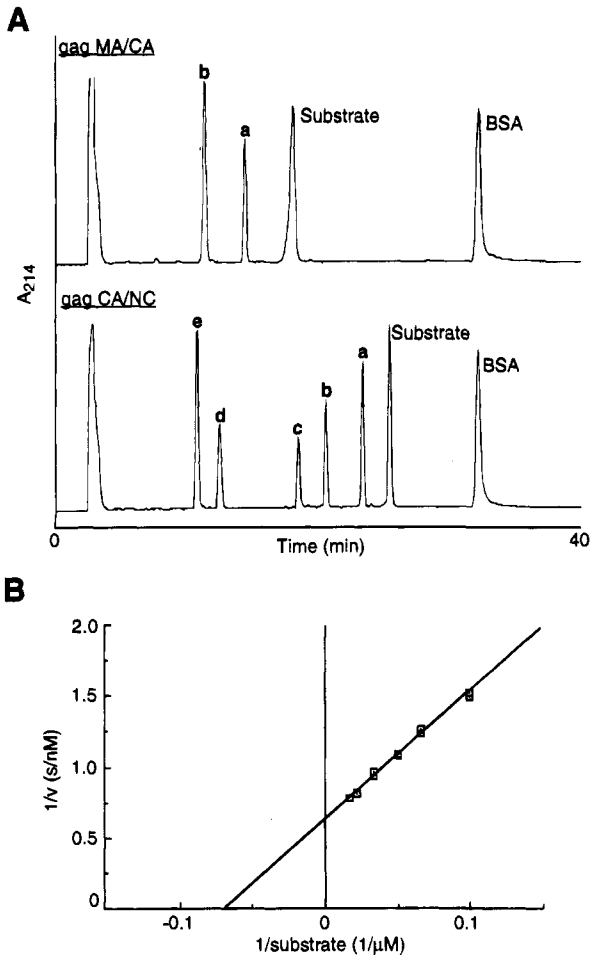

Figure 5. Enzymatic activity of synthetic tethered dimer HIV-1 PR. (**A**) Specificity. Cleavage of synthetic peptide analogues of the matrix/capsid (MA/CA) and capsid/nucleocapsid (CA/NC) processing sites in the HIV-1 gag translation product. (Upper) The MA/CA analogue RRSNQVSQNY−PIVQNIQGRR (500 μM) was incubated with tethered dimer HIV-1 PR (7 nM) at pH 5.5, 37 °C. High-pressure liquid chromatography analysis of the reaction mixture after 21 h showed significant cleavage of substrate. Reaction products were identified by electrospray mass spectrometry and as expected result from specific cleavage at the Tyr−Pro bond to give PIVQNIQGRR (peak **a**) and RRSNQVSQNY (peak **b**). The late-eluting BSA, a component in the assay buffer, was not cleaved under these conditions. (Lower) Similarly, the CA/NC processing site analogue GHKARVL−AEAM-SQVTNSATIM−MQRGNFRNQRK (500 μM) was incubated with tethered dimer HIV-1 PR (7 nM) at pH 5.5, 37 °C. After 21 h, HPLC and ESMS analysis of the reaction mixture showed five products corresponding to the expected specific cleavage at the Met−Met and/or Leu−Ala bonds. Products were characterized as GHKARVLAE-AMSQVTNSATIM (peak **a**), AEAMSQVTNSATIMMQRGNFRN-QRK (peak **b**), AEAMSQVTNSATIM (peak **c**), GHKARVL (peak **d**) and MQRGNFRNQRK (peak **e**). (**B**) Kinetics. Determination of steady state kinetic parameters for cleavage of the fluorogenic substrate Abz-Thr-Ile-Nle−Phe(NO_2)-Gln-Arg-amide (Abz = anthranilic acid). Initial velocities were measured as a function of substrate concentration, and the data are presented graphically in a double-reciprocal plot. Kinetic parameters were calculated by nonlinear regression analysis as $k_{cat} = 21 \pm 1\ s^{-1}$ and $K_m = 15 \pm 1\ \mu M$, values typical of those obtained with a wild-type enzyme.[15]

nucleocapsid junctions of the viral gag-pol polyprotein.[12] The gag-pol-derived peptides were cleaved at the expected Tyr-Pro, Met-Met, and Leu-Ala positions (Figure 5A) corresponding to the natural processing sites cleaved during virus maturation.[13]

Kinetic parameters were determined for cleavage of the fluorogenic substrate, and these data are presented graphically in Figure 5B. Derived values for the turnover number and Michaelis constant were $k_{cat} = 21 \pm 1\ s^{-1}$; $K_m = 15 \pm 1\ \mu M$. These values are indistinguishable from those of native backbone, homodimeric HIV-1 PR from chemical synthesis[14] or from recombinant microorganisms.[15]

Discussion

Our experimental results demonstrate that, in general, unprotected sulfhydryl groups are compatible with the reactive elements used in thioester-forming chemical ligation of peptides. That is, bromoacetyl and α-thioacid functionalities do not normally react with free thiol groups under the conditions used for ligation. This selective reaction can be attributed to the relative pK_a's of sulfhydryl versus α-thioacid groups. A nucleophile with a negative charge is a more powerful nucleophile than its conjugate acid. Thus thiocarboxylate and thiolate ions ($RCOS^-$ and RS^-, respectively) are far more potent nucleophiles than neutral RCOSH and RSH. The chemoselective ligation reaction proceeds via an S_N2 mechanism, so the rate of reaction is directly proportional to the strength of the attacking nucleophile. Under the acidic conditions typically used for the thioester ligation chemistry (pH 3.0−5.5), the α-thioacid group ($pK_a \approx 3$[16]) is predominantly deprotonated while thiol groups are protonated ($pK_a \approx 9$ for Cys side chain[17]), thus reaction occurs almost exclusively between the thiocarboxylate and bromoacetyl group. In the absence of a thiocarboxylate functionality, no ligation can be brought about between bromoacetyl- and cysteine-containing peptides at low pH.[8]

In the course of this study, we have identified one potential limitation to the general thioester-forming chemical ligation of cysteine-containing peptides. In those instances where the cysteine residue is in close intramolecular proximity to either the reactive bromoacetyl or α-thioacid groups, some level of side reaction is observed. Proximity to a bromoacetyl group leads to intramolecular thioether formation, while nearness to an α-thioacid can cause internal thioester formation. The former side reaction was particularly severe in the case of peptide **1**. Apparently, intramolecular cyclization with resulting formation of a 12 membered thioether ring system is highly favored, even at acidic pH. In the case of peptide **2**, a small amount of the cyclic thioester was formed by attack of the cysteine side chain on the terminal α-thioacid group. Occurrence of these side reactions is limited however to peptides with sulfhydryl groups proximal to the bromoacetyl or α-thioacid moieties. For instance, more distal placement of the cysteine residue in the lysozyme fragment BrAc-Thr-Pro-Gly-Ser-Arg-Asn-Leu-Cys-Asn completely prevented formation of an internal thioether bond, while no intramolecular thioester formation was observed for the HIV-1 PR fragment $HSCH_2CO$-Gly(1-51)αCOSH.

When working with cysteine-containing peptides, some level of oxidation to the disulfide will occur. In general, this was minimized in our experiments by the acidic pH of the reaction medium and also by slow stirring of the ligation mixtures.

(11) Toth, M. V.; Marshall, G. R. *Int. J. Pept. Protein Res.* **1990**, *36*, 544−550.
(12) Baca, M.; Alewood, P. F.; Kent, S. B. H. *Protein Sci.* **1993**, *2*, 1085−1091.
(13) Oroszlan, S. In *Viral Proteinases as Targets For Chemotherapy*; Kräusslich, H., Oroszlan, S., Wimmer, E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989; pp 87−100.
(14) Kent, S. B. H.; Schneider, J.; Clawson, L.; Selk, L.; Delahunty, C.; Chen, Q. In *Viral Proteinases as Targets For Chemotherapy*; Kräusslich, H., Oroszlan, S., Wimmer, E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989; pp 223−230.
(15) Richards, A. D.; Phylip, L. H.; Farmerie, W. G.; Scarborough, P. E.; Alvarez, A.; Dunn, B. M.; Hirel, P. H.; Konvalinka, J.; Strop, P.; Pavlickova, L. *J. Biol. Chem.* **1990**, *265*, 7733−7736.
(16) Janssen, M. J. In *The Chemistry of Carboxylic Acids and Esters*; Patai, S., Ed.; Interscience: London, 1969; pp 705−764.
(17) Tanford, C. *Adv. Protein Chem.* **1962**, *17*, 69−165.

Consequently, this side reaction rarely exceeded 10%. We did however explore the use of reducing agents to reverse oxidation of the sulfhydryl group (data not shown). Use of excess dithiothreitol (DTT) was found to promote some transthioesterification of the thioester ligation product, resulting in its scission at the ligation site to give a thioester-linked DTT adduct and mercaptoacetyl peptides. By contrast, use of the non-nucleophilic, water-soluble phosphine tris(2-carboxyethyl)phosphine[18] gave quantitative reduction of disulfides when added to ligation mixtures, without any observable side reactions involving the reactant or product peptides. Thus, when necessary, use of this reducing agent is compatible with components of the thioester ligation chemistry.

The linking of domains by formation of a disulfide bond has often been used to covalently join protein subunits that otherwise associate by noncovalent interactions, for instance to form tethered dimers of DNA-binding proteins.[19-21] In the work described in this report, we have used a similar strategy to generate a 22 kDa tethered dimer analogue of HIV-1 PR by total chemical synthesis. The compatibility of free sulfhydryl groups with the thioester ligation chemistry permitted assembly of the individual HIV-1 PR monomers, each containing a terminal residue with an unprotected thiol moiety. These then formed the basis for covalent dimerization of the two monomers by formation of an intersubunit disulfide bond. Exercising our total synthetic control over the composition of these peptides, we opted for cysteamine and mercaptoacetic acid as the sulfhydryl-bearing residues instead of cysteine. This was to avoid any potential structural perturbation of the folded HIV-1 PR domains by branched chiral elements in the intersubunit connecting loop. Because the monomer subunits were non-identical, directed formation of the disulfide bond was necessary in order to ensure that only the desired heterodisulfide of correct main-chain topology was formed. Successful use of 2,2'-dipyridyl disulfide shows that reaction of sulfhydryl groups with this reagent does not lead to side reactions with the ligation-derived thioester bond. Where regiospecific disulfide formation is not required, we have found that conventional oxidation methods, such as by DMSO/aqueous buffer,[22] can also be applied to the oxidation of free sulfhydryl groups in thioester-containing peptides (data not shown).

The tethered dimer HIV-1 PR prepared by the methods described here possessed full enzymatic activity. Kinetic parameters obtained for cleavage of a fluorogenic substrate are typical of values obtained for cleavage of other -Nle-Phe(p-NO_2)- substrates by wild-type HIV-1 PR.[15] This indicates that neither the single thioester bond in each monomer nor the linking disulfide-containing loop had any deleterious effect on enzymatic activity. This comes as no surprise, as the positioning of the thioester ligation site within each subunit was identical to that used in previous syntheses of fully active homodimeric HIV-1 PR,[1] while the wide variety of short, glycine-rich loops successfully used by others as intersubunit covalent links[23-25] indicates that the connecting loop need only be flexible and two or more residues in length. In support of this, a recent crystallographic structure of a recombinant tethered dimer HIV-1 PR showed that a -Gly-Gly-Ser-Ser-Gly- loop connecting the two domains was mobile and did not perturb the folded structure of the protein.[26]

The synthesis of a tethered dimer HIV-1 PR is significant. First, this represents the largest functional protein ever prepared in homogeneous form by total chemical synthesis. This demonstrates that the chemical ligation approach can be applied to the preparation of protein targets in excess of 200 amino acids. Second, chemical access to tethered dimer HIV-1 PR will allow preparation of analogues in which unnatural elements of structure are assymetrically incorporated into only one subunit of this enzyme. Interest in such analogues arises from a desire to understand the similarities and differences between the homodimeric two-chain retroviral aspartyl proteases and the related two domain, single-chain pepsin-like proteases. We are actively pursuing the study of such HIV-1 PR analogues.

This study has verified the applicability of thioester-forming chemical ligation to the synthesis of proteins of any sequence. With minor exceptions, we have shown that peptides can be unambiguously ligated without side reactions from sulfhydryl groups and that following ligation these sulfhydryls can be oxidized to disulfide bonds. Finally, the full activity of a synthetic tethered dimer HIV-1 PR analogue demonstrates that by judicious placement of the non-native backbone linkages introduced by chemical ligation of peptide segments, the functional properties of target proteins need not be compromised.

Experimental Section

Materials and Methods. Boc-amino acids and HBTU were obtained from Novabiochem (San Diego, CA). Boc-aminoacyl-OCH_2-Pam-resins and diisopropylethylamine were purchased from Applied Biosystems (Foster City, CA). *N,N*-Dimethylformamide was obtained from Mallinckrodt Chemical Co. (Paris, KY). HPLC grade acetonitrile was purchased from EM Science (Gibbstown, NJ), and all other reagents were reagent grade from Aldrich Chemical Co.

Preparative HPLC was performed on a Waters Delta-prep 4000 HPLC system. Analytical and semipreparative HPLC were performed either on a Rainin HPXL dual pump system with detection on a Dynamax UV detector or on an integrated Hewlett-Packard 1050 system. Preparative HPLC was run on a Vydac C18 column (15 μm, 50 × 250 mm) at a flow rate of 30 mL/min; semipreparative HPLC was run on a Vydac C18 column (10 μm, 10 × 250 mm) at 3 mL/min; and analytical HPLC was run on a Whatman C18 (5 μm, 4.0 × 140 mm) or Vydac C4 column (5 μm, 4.6 × 150 mm) at 1 mL/min. All runs used linear gradients of 0.1% aqueous TFA (solvent A) versus 90% acetonitrile, 0.1% TFA (solvent B). Mass analysis of all synthetic peptides and ligation products was performed on a Sciex API-III triple quadrupole electrospray mass spectrometer as previously described.[27] Calculated masses were obtained using the program MacProMass (Sunil Vemuri and Terry Lee, City of Hope, Duarte, CA).

Solid-Phase Peptide Synthesis. All peptides were synthesized according to the *in situ* neutralization/HBTU activation protocol for Boc solid phase chemistry as previously described.[2] Synthesis of short model peptides was performed by manual methods, while assembly of the HIV-1 PR fragments was achieved by machine-assisted synthesis on an Applied Biosystems 430A peptide synthesizer. The HIV-1 PR fragment BrAc(53-99)Gly-$NHCH_2CH_2CH$ (**5**) was prepared on a N^{α}-Boc-cysteamine-*S*-benzyl-4-(oxyacetamidomethyl)-resin (P. Alewood, personal communication). Peptide α-thiocarboxylic acids (α-COSH) were constructed on a Gly-thioester support,[28-29] yielding peptides with

(18) Burns, J. A.; Butler, J. C.; Moran, J.; Whitesides, G. M. *J. Org. Chem.* **1991**, *56*, 2648-2650.
(19) Sauer, R. T.; Hehir, K.; Stearman, R. S.; Weiss, M. A.; Jeitler-Nilsson, A.; Suchanek, E. G.; Pabo, C. O. *Biochemistry* **1986**, *25*, 5992-5998.
(20) Hubbard, A. J.; Bracco, L. P.; Eisenbeis, S. J.; Gayle, R. B.; Beaton, G.; Caruthers, M. H. *Biochemistry* **1990**, *29*, 9241-9249.
(21) Talanian, R. V.; McKnight, C. J.; Kim, P. S. *Science* **1990**, *249*, 769-771.
(22) Tam, J. P.; Wu, C. R.; Liu, W.; Zhang, J. W. *J. Am. Chem. Soc.* **1991**, *113*, 6657-6662.
(23) Cheng, Y. S.; Yin, F. H.; Foundling, S.; Blomstrom, D.; Kettner, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9660-9664.
(24) DiIanni, C. L.; Davis, L. J.; Holloway, M. K.; Herber, W. K.; Darke, P. L.; Kohl, N. E.; Dixon, R. A. *J. Biol. Chem.* **1990**, *265*, 17348-17354.
(25) Kräusslich, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 3213-3217.
(26) Bhat, T. N.; Baldwin, E. T.; Liu, B.; Cheng, Y. E.; Erickson, J. W. *Nature Struct. Biol.* **1994**, *1*, 552-556.
(27) Schnölzer, M.; Jones, A.; Alewood, P. F.; Kent, S. B. H. *Anal. Biochem.* **1992**, *204*, 335-343.
(28) Blake, J. *Int. J. Pept. Protein Res.* **1981**, *17*, 273-274.
(29) Yamashiro, D.; Li, C. H. *Int. J. Pept. Protein Res.* **1988**, *31*, 322-334.

C-terminal Gly-αCOSH after cleavage. All other peptides were synthesized on appropriate Boc-aminoacyl-OCH_2Pam-resins. Where required, the bromoacetyl group was introduced at the N-terminal of peptides by coupling the preformed symmetric anhydride.[6] The terminal mercaptoacetyl residue on the HIV-1 PR fragment $HSCH_2CO$-Gly(1−51)α-COSH (**6**) was coupled as the preformed symmetric anhydride of the *S*-(4-methylbenzyl) derivative. Where necessary, the N^{α}Boc group was first removed, then side-chain protecting groups were removed and the peptides were cleaved from the resin by treatment with liquid HF containing 4% anisole for 1 h at 0 °C. After evaporation of the HF, crude peptide products were precipitated, washed with diethyl ether, dissolved in aqueous acetic acid (10−50%), and lyophilized.

Purification and Characterization of Peptide Segments. Crude lyophilized peptides were dissolved in either acidic aqueous buffers or 50% aqueous acetic acid and purified by preparative or semipreparative reverse phase HPLC. Purified peptides were stored as lyophilized powders at −20 °C. The purified peptide segments were characterized by ESMS, and all had observed masses within experimental error of the calculated masses. Table 1 summarizes synthetic peptides prepared for this study and includes the calculated and experimentally derived masses.

Model Peptide Stability and Ligation. (1) Stability of Peptides 1, 2, and 3. Peptides were dissolved in 100 mM sodium phosphate, pH 3.2, at a concentration of 5 mg/mL. The resulting solutions were monitored by reverse phase HPLC using a gradient of 5−30% solvent B over 20 min. Peaks were collected on the basis of their UV absorbance (214 nm) and analyzed by ESMS.

(2) Synthesis of Arg-Ala-Cys-Ala-Arg-Gly-[COS]-Gly-Thr-Pro-Gly-Ser-Arg-Asn-Leu-Cys-Asn. Ligation was carried out by combining peptides **2** (0.92 mg, 1.4 μmol) and **3** (1.68 mg, 1.6 μmol) in 150 μL of 100 mM sodium phosphate, pH 3.2, at 25 °C. The ligation reaction was monitored by reverse phase HPLC using a gradient of 0−67% solvent B over 30 min. Peaks were collected and analyzed by ESMS. The ligation reaction was observed to be complete after 1 h, and the product was characterized by ESMS [calcd 1648.7 Da (monoisotopic), 1649.9 Da (average); found 1649.5 ± 0.7 Da].

Synthesis of Tethered Dimer HIV-1 PR. (1) Synthesis of HIV-1 PR (1−51)-[COS]-(52−99)Gly-$NHCH_2CH_2SH$ (8). Peptides **4** (10.27 mg, 1.83 μmol) and **5** (10.16 mg, 1.90 μmol), prepared by stepwise SPPS, were dissolved in 6 M GuHCl, 50 mM sodium phosphate, pH 3.8 (1.8 mL), and stirred at 4 °C for 20 h and room temperature for 7 h. Formation of the target ligation product **8** was followed by analytical HPLC (35−60% solvent B gradient), and it appeared as a new peak with retention time intermediate between the two precursor fragments. The identity of this peak was confirmed by ESMS as being the target ligation product [calcd mass 10 887 Da (average isotopes); found 10 888 ± 3 Da].

(2) Conversion of 8 to the *S*-(2-Pyridylsulfenyl)cysteamine Derivative (9). Without isolating peptide **8** from the ligation reaction, a solution of 0.5 M bromoacetic acid in 50% acetonitrile/water (40 μL) was slowly added to the crude mixture, and the solution was stirred for 1 h (to quench residual thiocarboxylate group). This solution was then slowly added to excess 2,2′-dipyridyl disulfide (PySSPy, 12 mg), dissolved in 90% acetonitrile/0.1% TFA (400 μL), and stirred for 3 h at room temperature. Reaction of the terminal cysteamine thiol group was essentially complete after this time, as judged by a shift in the HPLC retention time of the ligated product. The derivatized product was purified by semipreparative reverse phase HPLC, and lyophilized to give HIV-1 PR (1−51)-[COS]-(52−99)Gly-$NHCH_2CH_2SSPy$ (**9**), 7.70 mg (0.70 μmol, 38%). **9** was characterized by ESMS [calcd mass 10 996 Da (average isotopes); found 10 997 ± 3 Da].

(3) Synthesis of HIV-1 PR $HSCH_2CO$-Gly(1−51)-[COS]-(52−99) (10). Peptides **6** (4.23 mg, 0.74 μmol) and **7** (3.74 mg, 0.71 μmol), prepared by stepwise SPPS, were dissolved in 6 M GuHCl, 50 mM sodium phosphate, pH 3.8 (600 μL), and stirred at 4 °C for 20 h and then 12 h at room temperature. Formation of the target ligation product **10** was followed by analytical HPLC (35−60% solvent B), and it appeared as a new late-eluting peak. The ligation product was then purified by semipreparative reverse phase HPLC and lyophilized to give $HSCH_2CO$-Gly(1−51)-[COS]-(52−99) (**10**), 4.60 mg (0.42 μmol, 59%). **10** was characterized by ESMS [calcd mass 10 902 Da (average isotopes); found 10 902 ± 1 Da].

(4) Ligation of the Two Monomers 9 and 10. HIV-1 PR monomers **9** (3.0 mg, 0.27 μmol) and **10** (4.60 mg, 0.42 μmol) were dissolved in 6 M GuHCl, 100 mM NaOAc, pH 5.0 (600 μL), and stirred at 4 °C. Thiolysis of the *S*-(2-pyridylsulfenyl) derivative was monitored by analytical HPLC (35−60% solvent B), and the disulfide-linked tethered dimer product appeared as a new peak at longer retention time than either of the two monomer fragments. After 4 h the yield of dimer was ~70%, with no further reaction occurring. The heterodimer was purified by HPLC, and the product-containing fractions were pooled. Then 6 M GuHCl, 100 mM NaOAc, pH 3.8, was added to the combined fractions, such that the final GuHCl concentration was 4 M. Acetonitrile was removed *in vacuo*, and the sample was ultrafiltered in a stirred Amicon cell, using a 10 kDa cutoff membrane. After concentrating, the sample was dialyzed twice against 5 M GuHCl, 50 mM NaOAc, pH 3.8, and stored at 4 °C. The concentration of protein in this sample was determined by absorbance at 280 nm, using a molar extinction coefficient of 25 500 M^{-1} cm^{-1}.[30] Synthetic tethered dimer HIV-1 PR was characterized by ESMS, after first removing all salts by HPLC [calcd mass 21 787 Da (average isotopes); found 21 786 ± 3 Da]. Under the storage conditions used, no breakdown of the synthetic enzyme was observed after a 1 month period.

Enzyme Folding. Tethered dimer HIV-1 PR was folded by adding BSA and Ac-Thr-Ile-Nle-ψ[CH_2NH]Nle-Gln-Arg·amide[31] (MVT-101) to a diluted sample of the denatured protein in 5 M GuHCl (final concentrations: 0.5 mg/mL BSA, 1 μM MVT-101, 1.2 μM enzyme) and dialyzing against 50 mM NaOAc, pH 5.5, containing 10% glycerol and 1 μM MVT-101. The competitive inhibitor MVT-101 was added to prevent autolytic breakdown during folding. Following dialysis, the sample was centrifuged to remove any precipitate and stored at 4 °C. The concentration of folded enzyme was determined by measuring the peak area obtained in an analytical HPLC run and comparing to the area obtained from a sample of denatured enzyme of known concentration. The yield of folded protein obtained was ~20%.

Enzyme Assays. Enzymatic activity and substrate specificity of tethered dimer HIV-1 PR was assessed by the cleavage of two synthetic peptides spanning the matrix/capsid and capsid/nucleocapsid cleavage sites in the viral gag-pol polyprotein.[12] The matrix/capsid peptide (500 μM) was incubated with tethered dimer HIV-1 PR (7 nM) at 37 °C in a pH 5.5 assay buffer containing 50 mM HOAc, 50 mM Mes, 100 mM Tris, 10% glycerol, and BSA at 0.5 mg/mL. After 21 h, formation of cleavage products was determined by HPLC. The identity of these products was confirmed by ESMS. Cleavage of the capsid/nucleocapsid peptide (500 μM) by tethered dimer HIV-1 PR was similarly assessed.

Kinetic parameters were determined for cleavage of a fluorogenic substrate[11] Abz-Thr-Ile-Nle−Phe(*p*-NO_2)-Gln-Arg·amide (Abz = anthranilic acid; cleaved at the Nle−Phe bond) in assay buffer. Assays were run at 37 °C using a chromatographic protocol described previously,[3] with an enzyme concentration of 78 pM and substrate concentration varied between 10−60 μM. Residual MVT-101 carried over into these assays was <1 nM, well below its inhibition constant.[31] Initial rates were determined by single time points of reactions in which consumption of substrate was less than 10% of the starting concentration (15 min reaction time). Kinetic parameters were obtained by fitting data to the Michaelis−Menten equation using a nonlinear regression computer program.[32]

Acknowledgment. We thank Lynne Canne and Paul Alewood for helpful advice and discussions. T.W.M. was supported by an Amgen Postdoctoral Fellowship; M.S. was supported by an AIDS scholarship from the Bundesministerium für Forschung und Technologie, Germany. This research was supported by funds from NIH Grants PO1 GM48870 and RO1 GM48897.

JA9431925

(30) Gill, S. C.; von Hippel, P. H. *Anal. Biochem.* **1989**, *182*, 319−326.

(31) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. *Science* **1989**, *246*, 1149−1152.

(32) Duggleby, R. G. *Comput. Biol. Med.* **1984**, *14*, 447−455.