J|A|C|S
ARTICLES
Published on Web 08/18/2007

Modular Total Chemical Synthesis of a Human Immunodeficiency Virus Type 1 Protease

Erik C. B. Johnson,[†] Enrico Malito,[‡] Yuequan Shen,[‡,§] Dan Rich,[||] Wei-Jen Tang,[‡] and Stephen B. H. Kent*[,†,⊥]

Contribution from the Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, Ben-May Department for Cancer Research, and [D]Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, and School of Pharmacy, The University of Wisconsin—Madison, Madison, Wisconsin 53705

Received April 24, 2007; E-mail: skent@uchicago.edu

Abstract: As part of our ongoing studies of the human immunodeficiency virus type 1 (HIV-1) protease enzyme, we set out to develop a modular chemical synthesis of the protein from multiple peptide segments. Our initial attempts were frustrated by the insolubility of intermediate peptide products. To overcome this problem, we designed a synthetic strategy combining the solubility-enhancing properties of C-terminal $(Arg)_n$ tags and the biological phenomenon of autoprocessing of the Gag−Pol polyprotein that occurs during maturation of the HIV-1 virus in vivo. Synthesis of a 119-residue peptide chain containing 10 residues of the reverse transcriptase (RT) open reading frame plus an $(Arg)_{10}$ tag at the C-terminus was straightforward by native chemical ligation followed by conversion of the Cys residues to Ala by Raney nickel desulfurization. The product polypeptide itself completed the final synthetic step by removing the C-terminal modification under folding conditions, to give the mature 99-residue polypeptide. High-purity homodimeric HIV-1 protease protein was obtained in excellent yield and had full enzymatic activity; the structure of the synthetic enzyme was confirmed by X-ray crystallography to a resolution of 1.07 Å. This efficient modular synthesis by a biomimetic autoprocessing strategy will enable the facile synthesis of unique chemical analogues of the HIV-1 protease to further elucidate the molecular basis of enzyme catalysis.

Introduction

The HIV-1 protease (HIV-1 PR) is a virally encoded 21.4 kDa aspartyl protease made up of two identical 99-residue polypeptide chains that together form an enzyme molecule with a single active site.[1] Inhibition of the HIV-1 PR results in noninfectious viral particles, and this fact has made the protease a focus of antiviral therapy through development of selective inhibitors of the enzyme.[2] During HIV-1 replication, the HIV-1 PR enzyme is required for processing of the Gag and Gag−Pol polyprotein precursors into the individual protein components required for maturation of the virus.[3] Because the HIV-1 PR 99-residue polypeptide monomer sequence is itself contained within the Gag−Pol polyprotein, a conundrum arises: where does the first molecule of the homodimeric HIV-1 protease come from?[4] Experiments with recombinant expression of Gag−Pol polyprotein constructs in bacteria and reticulocyte lysate have suggested that the protease has the ability to fold into an enzymatically active form while still contained within the Gag−Pol polyprotein, which allows it to cleave itself from the polyprotein by an autoproteolytic mechanism.[4,5]

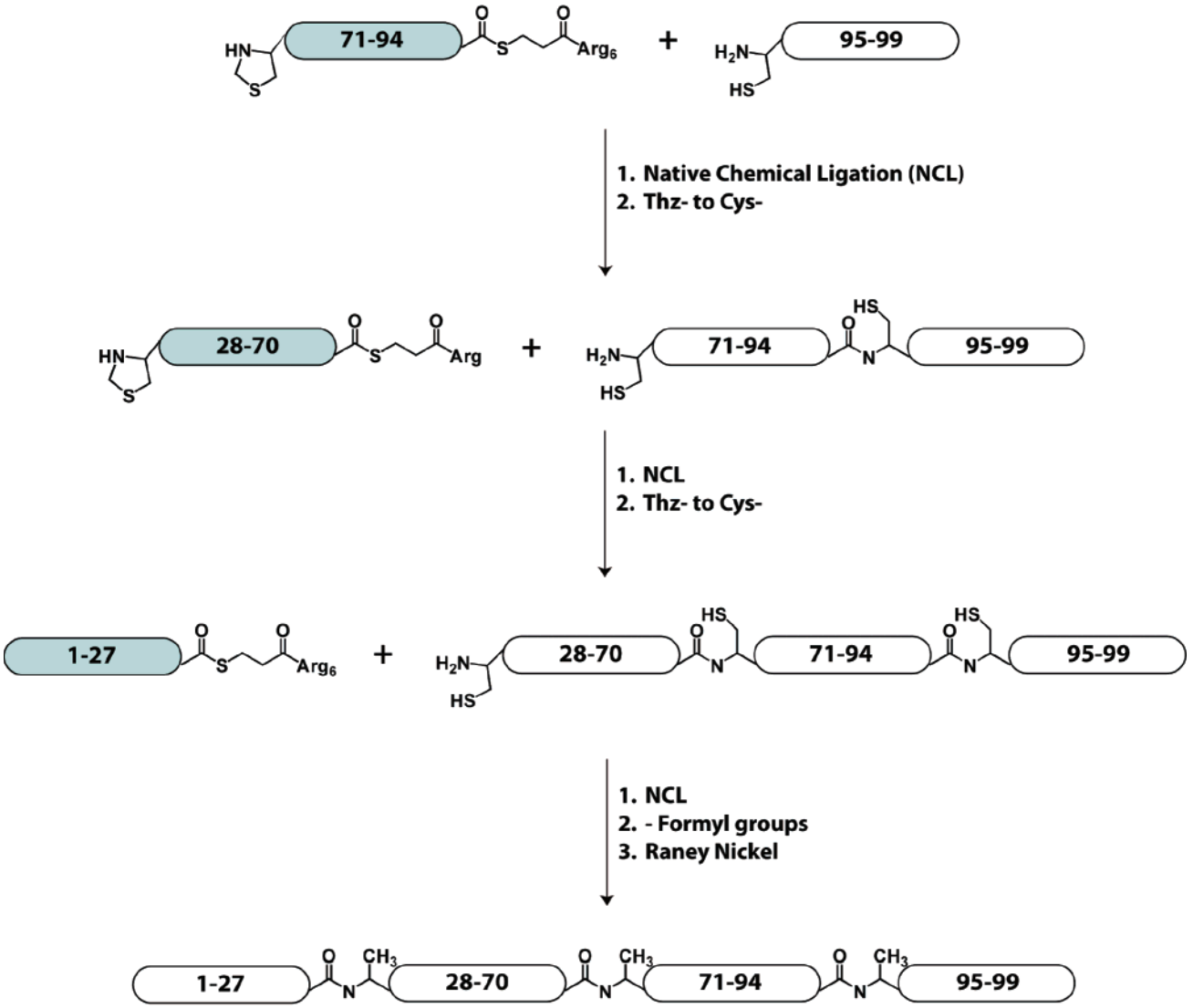
Although the HIV-1 PR has been well-studied,[6] details of its catalytic mechanism remain to be experimentally determined. For example, hydrolysis of the substrate peptide bond is thought to occur by nucleophilic attack of a water molecule activated by a catalytic aspartate residue located at position 25 in each monomer.[7] Curiously, this water nucleophile has never been observed in enzyme−inhibitor crystal structures. To test the validity of this aspect of the mechanism, it would be useful to systematically vary the electronic properties of the catalytic aspartate side chains and observe the effects on catalysis.[8] Other aspects of HIV-1 PR catalysis that require further study include the role of the highly mobile "flaps" that enclose the substrate within the binding groove of the enzyme.[9−12]

[†] Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, The University of Chicago.
[‡] Ben-May Department for Cancer Research, The University of Chicago.
[§] Present address: College of Life Science, Nankai University, Tianjin, China 300071.
[||] The University of Wisconsin—Madison.
[⊥] Department of Chemistry, The University of Chicago.

(1) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B.; Wlodawer, A. *Science* **1989**, *246*, 1149−1152.
(2) Swanstrom, R.; Erona, J. *Pharmacol. Ther.* **2000**, *86*, 145−170.
(3) Frankel, A. D.; Young, J. A. *Annu. Rev. Biochem.* **1998**, *67*, 1−25.
(4) 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.
(5) Pettit, S. C.; Everitt, L. E.; Choudhury, S.; Dunn, B. M.; Kaplan, A. H. *J. Virol.* **2004**, *78*, 8477−8485.
(6) Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249−284.
(7) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; Freeman: New York, 1999.
(8) Short, G. F., 3rd; Laikhter, A. L.; Lodder, M.; Shayo, Y.; Arslan, T.; Hecht, S. M. *Biochemistry* **2000**, *39*, 8768−8781.
(9) Baca, M.; Kent, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11638−11642.
(10) Baca, M.; Kent, S. B. H. *Tetrahedron* **2000**, *56*, 9503−9513.

 10.1021/ja072870n CCC: $37.00

Scheme 1. Initial Synthetic Strategy for the HIV-1 PR 1−99 Polypeptide[a]



[a] Four peptide segments are joined sequentially by native chemical ligation[14] (NCL), with conversion of the N-terminal (4*R*)-1,3-thiazolidine (Thz) into cysteine after each NCL reaction in order to proceed with the next ligation.[16,17] After the third ligation is complete, formyl protection is removed from tryptophan, and the cysteines are converted into alanine by Raney nickel desulfurization.[18]

Total synthesis enables precise atom-by-atom control over the covalent structure of a protein molecule and is thus very useful for dissecting essential aspects of the molecular basis of enzyme catalysis.[13] Modern chemical protein synthesis is based on the chemoselective reaction of peptides by native chemical ligation. In native chemical ligation, an unprotected peptide thioester reacts with an unprotected Cys peptide in aqueous solution at neutral pH to give a single product with a native peptide bond at the ligation site.[14] The peptide thioester building blocks are readily made by optimized Boc chemistry stepwise solid-phase peptide synthesis (SPPS).[15] Because of the statistical accumulation of resin-bound byproducts in SPPS, yields and purity are generally better for shorter peptides, and it is therefore desirable to synthesize proteins in a modular fashion through ligation of multiple peptide segments. This is of particular importance when one aims to incorporate nonnatural amino acids or biophysical probes into the protein molecule, and is also useful if a large series of analogue proteins is to be synthesized.

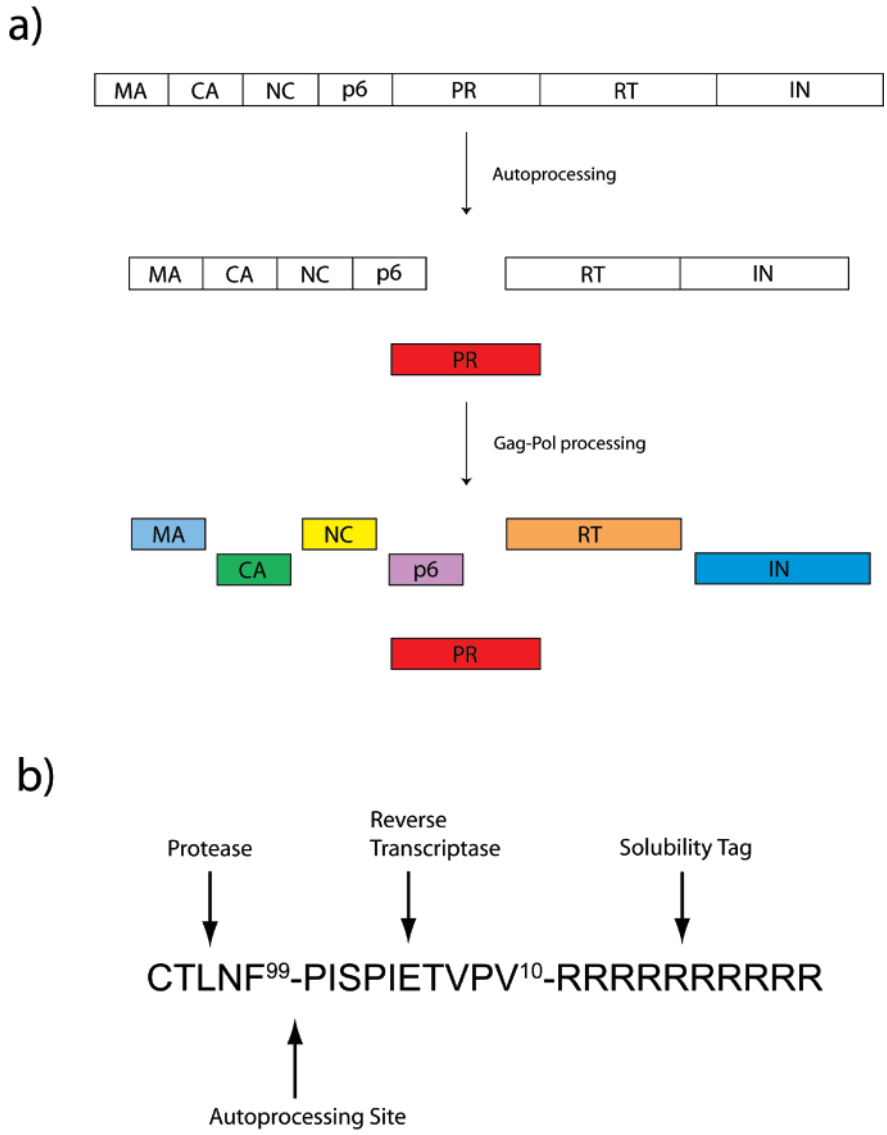
We set out to develop a robust, modular synthetic route to the HIV-1 PR for our studies of the enzyme. Our initial efforts to synthesize the HIV-1 PR using multiple peptide segments were frustrated by the insolubility of intermediate peptide products. In this paper, we describe how we overcame these difficulties using state-of-the-art chemical synthesis based on native chemical ligation (NCL) and Raney nickel desulfurization, in a biomimetic synthetic approach inspired by the natural autoproteolytic maturation of the HIV-1 PR in vivo.

Results

Initial Synthetic Design. Our initial synthetic design is shown in Scheme 1. The strategy was based on ligation of four peptides, consisting of residues 1−27 (1−27$^{\alpha}$COSR), 28−70 (Thz28−

(11) Trylska, J.; Tozzini, V.; Chang, C. E.; McCammon, J. A. *Biophys. J.* **2007**, *92*, 4179−4187.
(12) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915−920.
(13) Kent, S. *J. Pept. Sci.* **2003**, *9*, 574−593.
(14) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776−779.
(15) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. *Int. J. Pept. Protein Res.* **1992**, *40*, 180−193.

Scheme 2. (a) Simplified Scheme of Maturation of HIV-1 Gag−Pol Polyprotein[a] and (b) Modification of C-Terminal 95−99 Peptide[b]



[a] The protease (PR) catalyzes its own removal from within the polyprotein and then processes the rest of the polyprotein into mature matrix (MA), capsid (CA), nucleocapsid (NC), p6, reverse transcriptase (RT), and integrase (IN) viral proteins. [b] To increase the solubility of Cys71−99, 10 residues from the N-terminus of the reverse transcriptase protein sequence adjacent in the Gag−Pol polyprotein, plus 10 arginine residues, are added to the Cys95−99 peptide.

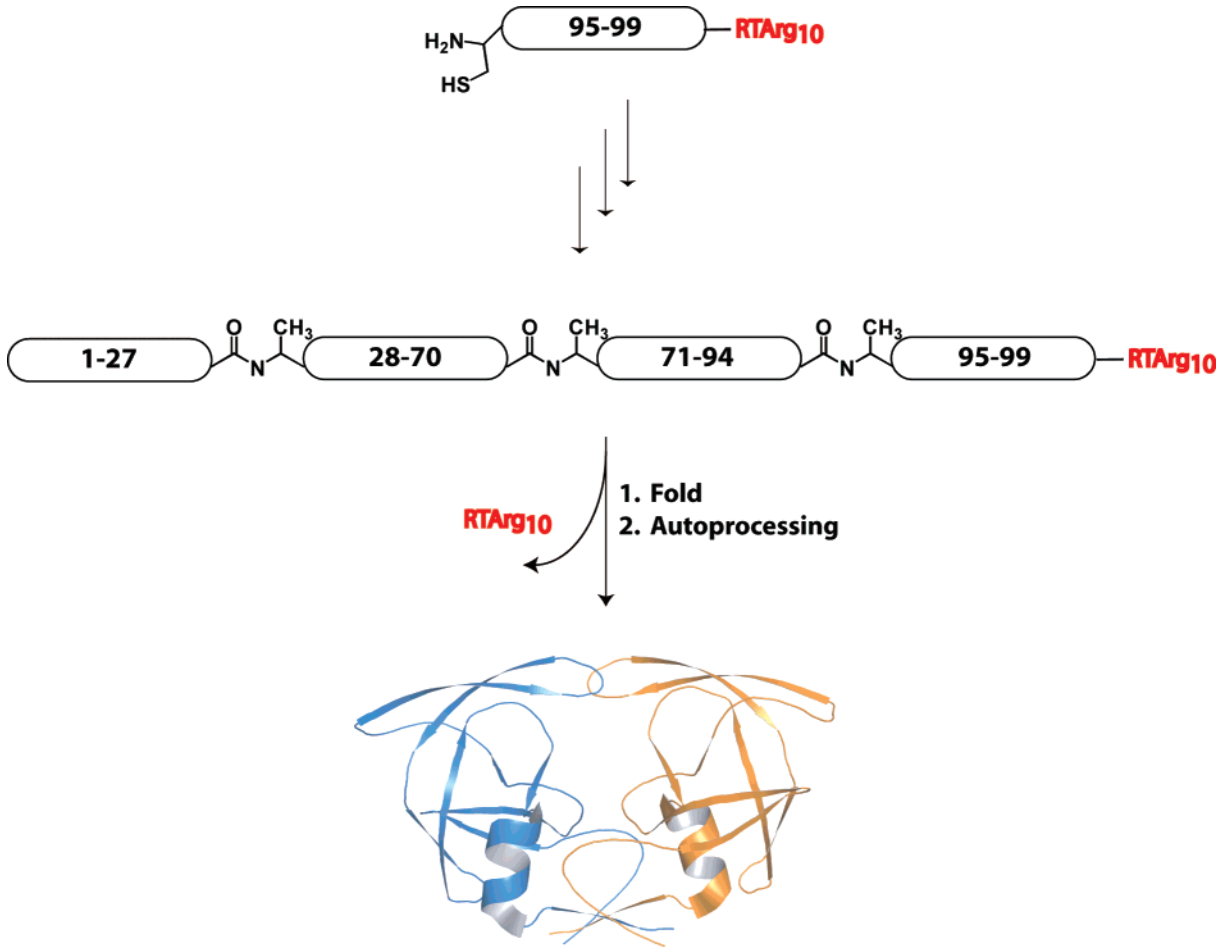
70$^{\alpha}$COSR), 71−94 (Thz71−94$^{\alpha}$COSR), and 95−99 (Cys95−99) of the protease (for peptide sequences, see Table S1, Supporting Information). The design involved ligation of the peptides in sequential fashion, starting with the C-terminal Cys95−99 peptide.

Ligation of Cys95−99 to Thz71−94$^{\alpha}$COSR was rapid and was complete within 1 h (data not shown). However, the solution became cloudy as the ligation progressed, suggesting that the product Thz71−99 peptide was aggregating and precipitating. In addition to the insolubility of the Thz71−99 peptide, the synthesis was further complicated by the relative instability of the Thz28−70$^{\alpha}$COSR peptide thioester during the subsequent ligation (see Figure S2, Supporting Information). Because of the poor solubility of the Cys71−99 peptide and the poor stability of the Thz28−70$^{\alpha}$COSR peptide thioester, it was reasoned that ligation between these two peptides could not be optimized to give satisfactory yields without mutating residues in Thz28−70$^{\alpha}$COSR to stabilize the thioester. In order to simplify the synthesis, increasing the solubility of Cys71−99 was imperative.

Revised Synthetic Strategy. Our idea was to add a solubilizing sequence of Arg residues at the C-terminus of the first peptide segment in order to keep all the intermediate peptide products in solution. The challenge was to devise a way to remove this Arg$_n$ tag after assembly of the full-length polypeptide chain, to give the mature 99-residue product. Our revised synthetic design relied on the autoproteolytic biological function of the protease during HIV-1 maturation (Scheme 2a). On the basis of this proposed autoprocessing mechanism of HIV-1 PR maturation, we added 10 residues from the reverse transcriptase (RT) protein adjacent to the HIV-1 PR sequence in the Gag−Pol polyprotein to the C-terminus of the 95−99 peptide, followed by 10 arginine residues to act as a solubilizing "tag" to keep the polypeptide intermediates in solution (Scheme 2b).[19]

(16) Villain, M.; Vizzavona, J.; Rose, K. *Chem. Biol.* **2001**, *8*, 673−679.
(17) Bang, D.; Kent, S. B. *Angew. Chem., Int. Ed.* **2004**, *43*, 2534−2538.
(18) Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123*, 526−533.

Scheme 3. Biomimetic Synthetic Strategy for HIV-1 PR[a]



[a] The protease is synthesized from four peptide segments. The product 1−99*RTArg*$_{10}$ polypeptide contains the 1−99 protease polypeptide in addition to 10 residues from the N-terminus of the reverse transcriptase polypeptide adjacent in the Gag−Pol polyprotein and 10 arginine residues to increase the solubility of intermediate products during the synthesis. After synthesis of the 119-residue polypeptide chain is complete, the polypeptide is dialyzed into conditions where protease activity is expected, and the protease removes the C-terminal modification through autoprocessing to yield the 2 × 99 residue homodimeric enzyme.

After assembly of the 119-residue polypeptide was complete, it was anticipated that the protease would remove the C-terminal modification after folding through an autoprocessing mechanism similar to what occurs in vivo, yielding the 2 × 99 residue homodimeric enzyme. The final biomimetic synthetic strategy is shown in Scheme 3.

To test this synthetic strategy, the C-terminal peptide was resynthesized to include the aforementioned modifications, and the first ligation was performed as before. Satisfyingly, the Cys71−99RTArg$_{10}$ peptide product was much more soluble in ligation buffer than the Cys71−99 peptide, and Thz deprotection proceeded to completion overnight (Figure 1).

Methoxylamine was removed from the solution by solid-phase extraction (SPE), and after lyophilization the peptides were redissolved in ligation buffer. Thz28−70$^{\alpha}$COSR was then added in order to perform the next NCL reaction, and the progress of the ligation was monitored by liquid chromatography−mass spectrometry (LC-MS) (Figure 2). Ligation of Cys71−99RTArg$_{10}$ to Thz28−70$^{\alpha}$COSR went to completion without any of the complications previously observed.

The Thz group was then removed as before, the methoxylamine was removed by SPE, and the final ligation was performed with 1−27$^{\alpha}$COSR (Figure 3). Ligation was allowed to proceed for 2 h, after which the ligation solution was made 20% in piperidine to remove formyl protection on tryptophan. The nonpeptidic materials were removed by SPE, and the peptide products were treated with Raney nickel to convert Cys^{28} and Cys^{71} to the native alanine residues and Cys^{95} to the mutant Ala^{95} (Figure 3). Desulfurization at neutral pH required approximately 40 h. After desulfurization was complete, the crude mixture was purified by gradient reverse-phase high-performance liquid chromatography (RP-HPLC). The final isolated yield of 1−99RTArg$_{10}$, based on the amount of Cys95−99RTArg$_{10}$ starting peptide, was 26%.

The purified 119-residue 1−99RTArg$_{10}$ polypeptide containing the reverse transcriptase plus arginine tag extension at the C-terminus was then dissolved in 6 M guanidine hydrochloride phosphate buffer (0.2 M, pH 7.4) and folded by dialysis against pH 5.6 acetate buffer—conditions expected to result in folding of the polypeptide chain to form enzymatically active protein. Analysis by LC-MS and matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS of the folding reaction after overnight incubation at 4 °C showed that quantitative cleavage of 1−99RTArg$_{10}$ had occurred between residues 99 and 100 to yield the mature 1−99 polypeptide chain (Figure 4

(19) Johnson, E. C.; Kent, S. B. *Tetrahedron Lett.* **2007**, *48*, 1795−1799.

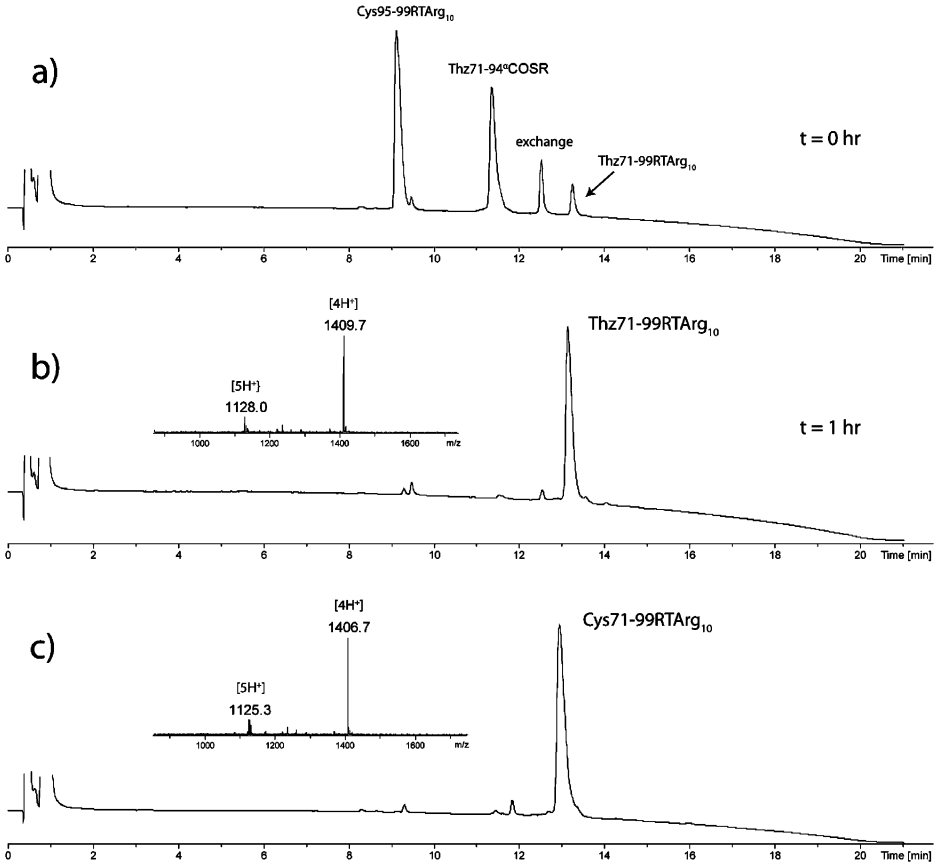


Figure 1. Ligation of Thz71−94$^{\alpha}$COSR to Cys95−99RTArg$_{10}$ and Thz deprotection. (a) RP-HPLC analysis of the ligation reaction after addition of Thz71−94$^{\alpha}$COSR (t < 1 min). Product has begun to appear before a reaction sample could be quenched with acid and subsequently analyzed. (b) Ligation is essentially complete within 1 h (expected mass 5633.7 Da, observed mass 5634.9 Da). (c) Thz deprotection after 12 h (expected mass 5621.7, observed mass 5622.2 Da). Reactions were analyzed by analytical RP-HPLC with online ESI-MS detection, using a 5−65% gradient of solvent B (CH_3CN + 0.08% TFA) vs solvent A (H_2O + 0.1% TFA). UV detection was at 214 nm. Masses are an average over the principal UV peak during LC-MS analysis, with an experimental uncertainty of ∼1 part in 5000.

and Figure S3, Supporting Information). Folding yield was 45%, for an overall synthetic yield of 12% based on the limiting peptide segment.

Characterization of Synthetic HIV-1 PR. The mature enzyme's polypeptide had an observed mass of 10 704.4 ± 5.4 Da (calculated mass 10 703.5 Da, average isotopes). To verify that the synthetic HIV-1 PR possessed full activity identical to the virally encoded enzyme, the synthetic enzyme was assayed directly from folding buffer by incubation with synthetic substrate peptides corresponding to the p17/24 (data not shown) and to the p24/15 Gag−Pol junctions (Figure 5a). The synthetic enzyme showed the expected substrate specificity by catalyzing peptide bond hydrolysis only at the expected cleavage sites. Further kinetic characterization of the synthetic enzyme by use of a fluorogenic peptide substrate gave steady-state kinetic parameters k_{cat} and K_m of 26.0 s^{-1} and 25 μM, respectively, in agreement with previously reported values for the protease under similar assay conditions (Figure 5b and Table S2, Supporting Information).[10,20,21]

The structure of the chemically synthesized HIV-1 PR was characterized by X-ray crystallography. Crystals were grown from protein material concentrated directly from folding buffer, after addition of the inhibitor JG-365, a hydroxyethylamine substrate isostere.[22] The HIV1PR·JG365 crystals diffracted to 1.07 Å (for data collection and refinement statistics, see Table S3, Supporting Information). Main-chain electron density was observed to unambiguously end at residue Phe^{99}, showing that the protease had indeed cut itself from the 1−99RTArg$_{10}$ sequence at residue 99 (Figure 5c). The atomic resolution of the electron density map also allowed us to unambiguously determine maintenance of L stereochemistry at the desulfurization sites. The structure with the JG-365 inhibitor showed the hallmark features present in previously determined structures of HIV-1 PR inhibitor complexes, including the coplanar catalytic aspartate side-chain carboxyls, and the enzyme flaps mediating contact with the inhibitor through a water molecule.

Discussion

When nonnatural amino acids or biophysical probes are incorporated into a protein by total chemical synthesis, it is desirable to construct the protein in a modular fashion from shorter peptide segments in order to minimize the difficulty and expense of synthesizing the peptide that contains the modification of interest. However, this approach can occasionally lead to problems with intermediate polypeptide solubilities during

(20) Szeltner, Z.; Polgar, L. *J. Biol. Chem.* **1996**, *271*, 32180−32184.
(21) Bergman, D. A.; Alewood, D.; Alewood, P. F.; Andrews, J. L.; Brinkworth, R. I.; Englebretsen, D. R.; Kent, S. B. H. *Lett. Pept. Sci.* **1995**, *2*, 99−107.
(22) Rich, D. H.; Green, J.; Toth, M. V.; Marshall, G. R.; Kent, S. B. H. *J. Med. Chem.* **1990**, *33*, 1285−1288.

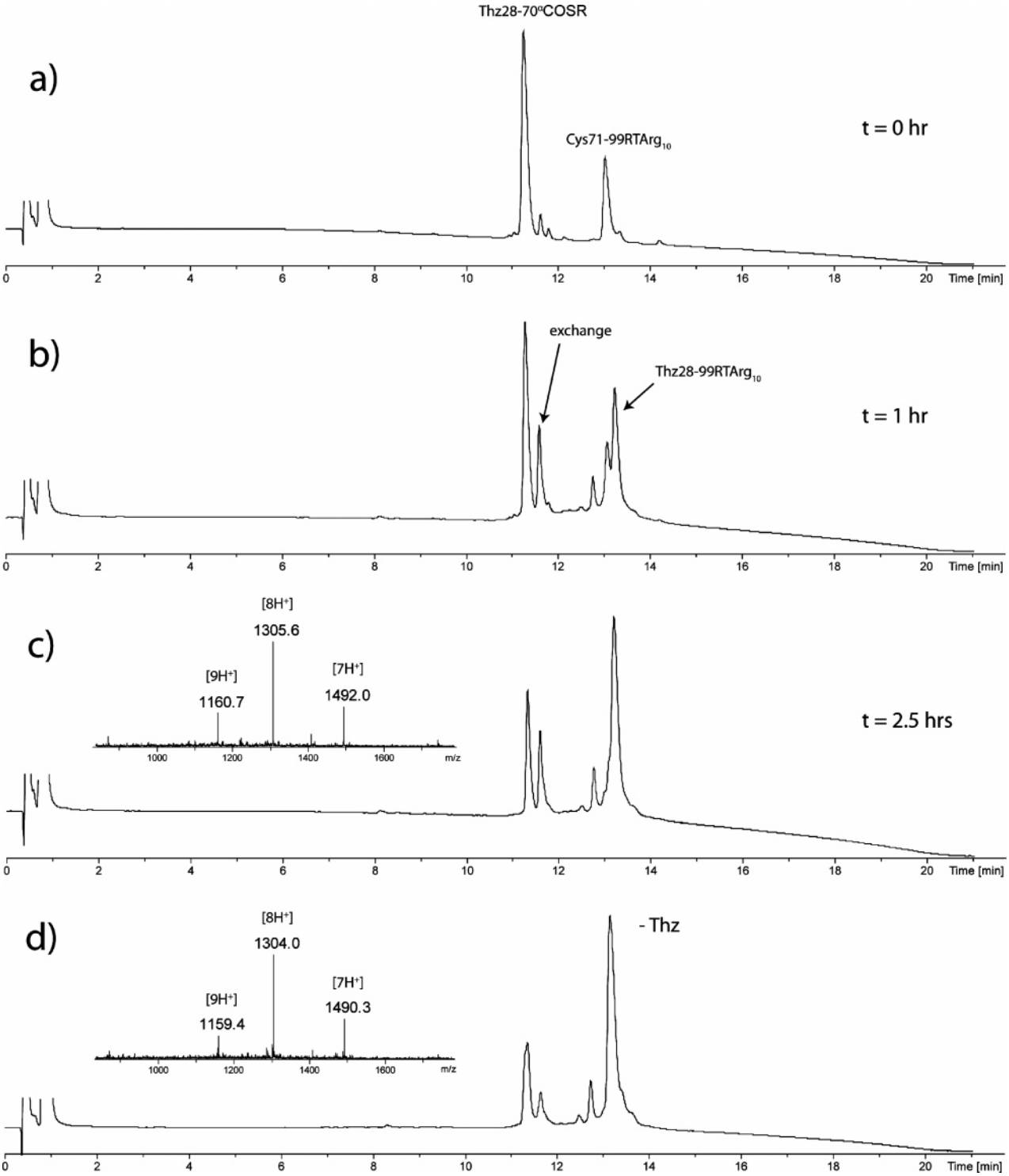


Figure 2. Ligation of Thz28−70αCOSR to Cys71−99RTArg$_{10}$ and Thz deprotection. RP-HPLC analysis of the ligation reaction is shown (a) at $t = 0$, (b) after 1 h, and (c) after 2.5 h. Most of the starting Cys71−99RTArg$_{10}$ peptide has reacted with Thz28−70αCOSR, which is present in slight molar excess (Thz28−99RTArg$_{10}$ expected mass 10 434.2 Da, observed mass 10 437.0). The reaction was allowed to proceed for 5.5 h, at which time methoxylamine hydrochloride was added to remove Thz protection. (d) Thz deprotection after 15 h (expected mass 10 422.2 Da, observed mass 10 424.8 Da). Reactions were analyzed as described for Figure 1.

the ligation reactions, even when they are performed in aqueous denaturant solution at neutral pH. We encountered this difficulty during our initial attempts at developing a modular synthesis of the HIV-1 PR, which contains a β-sheet-rich region at the C-terminus of the protein that has a tendency to form intermolecular H-bonded aggregates, creating problems with chain assembly during SPPS[23] and, in our case, solubility problems during peptide ligations. While this ordinarily would not be an insurmountable obstacle to a protein synthesis, as one could simply extend the reaction time to increase the yield of ligated product, we were presented with the compounding problem of reactant instability in the Thz28−70αCOSR peptide thioester.

(23) Henkel, B.; Bayer, E. *J. Pept. Sci.* **1998**, *4*, 461−470.

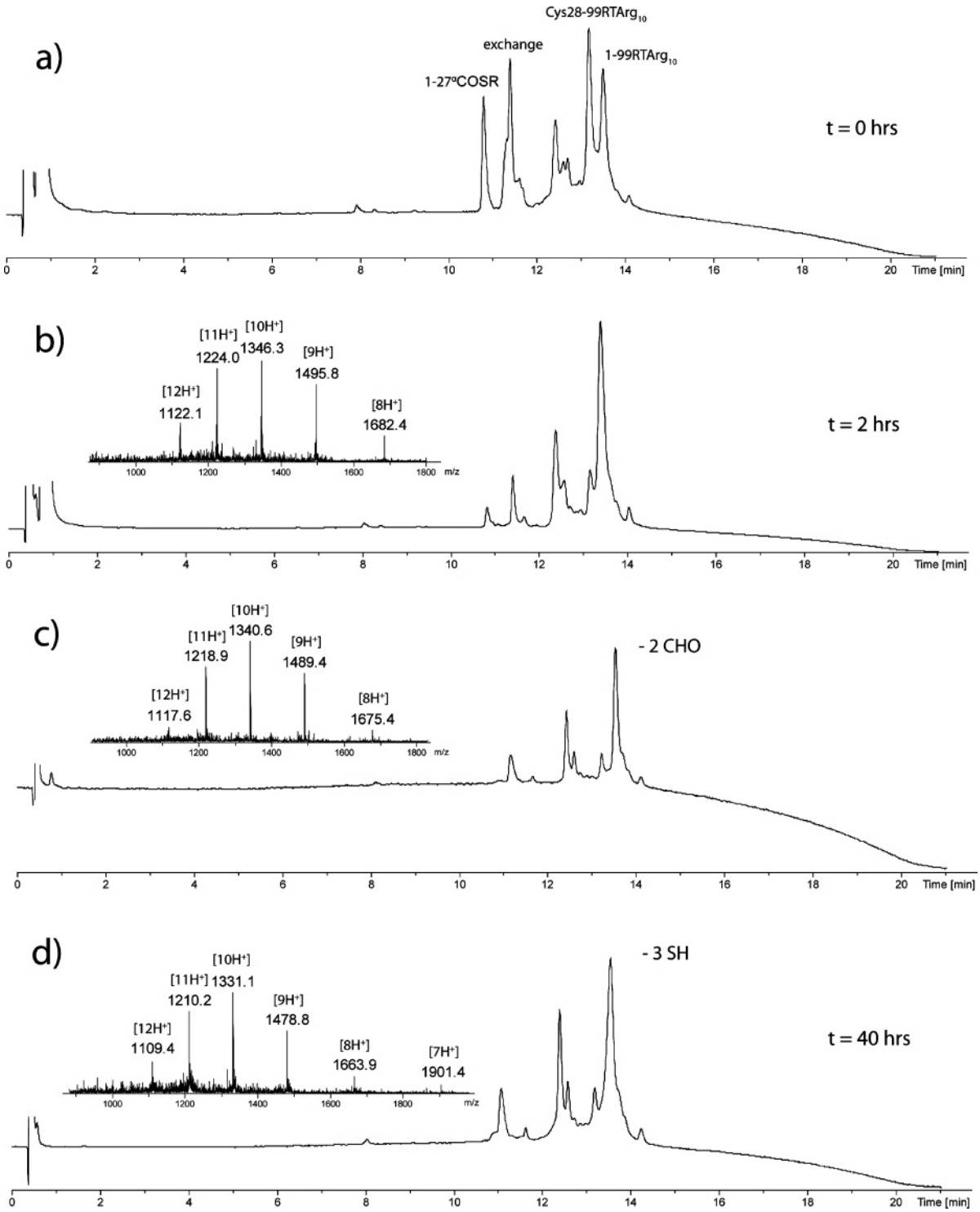


Figure 3. Ligation of 1−27$^{\alpha}$COSR to Cys28−99RTArg$_{10}$, formyl deprotection, and desulfurization. (a) RP-HPLC analysis of the ligation reaction at t = 0 after the addition of 1−27$^{\alpha}$COSR. A significant amount of ligation has occurred before the reaction could be quenched for analysis. (b) Analysis of the ligation reaction at 2 h. The ligation was determined to be complete (1−99RTArg$_{10}$ expected mass 13 449.8 Da, observed mass 13 452.6 Da). (c) RP-HPLC analysis of the crude products after formyl deprotection, SPE, lyophilization, and redissolution in ligation buffer before addition of Raney nickel (expected mass 13 393.7 Da, observed mass 13 396.5 Da). (d) Analysis of the desulfurization reaction after 40 h (expected mass 13 296.5 Da, observed mass 13 301.5 Da). Reactions were analyzed as described previously.

Unexpectedly, the Thz28−70$^{\alpha}$COSR peptide thioester was found to be highly labile to hydrolysis and methoxylaminolysis during the one-pot ligation procedure.[17] This reactant thioester lability, coupled with the poor solubility of reactant cysteine peptide Cys71−99, created an insurmountable practical barrier to the attainment of acceptable ligation yields for this reaction.

To overcome this problem, we redesigned the synthesis of HIV-1 PR, taking inspiration from the natural function of the protease in vivo. Molecular biology experiments have suggested that the protease is generated from the Gag−Pol polyprotein by means of an autoprocessing mechanism, the details of which are still not completely understood. The protease then functions

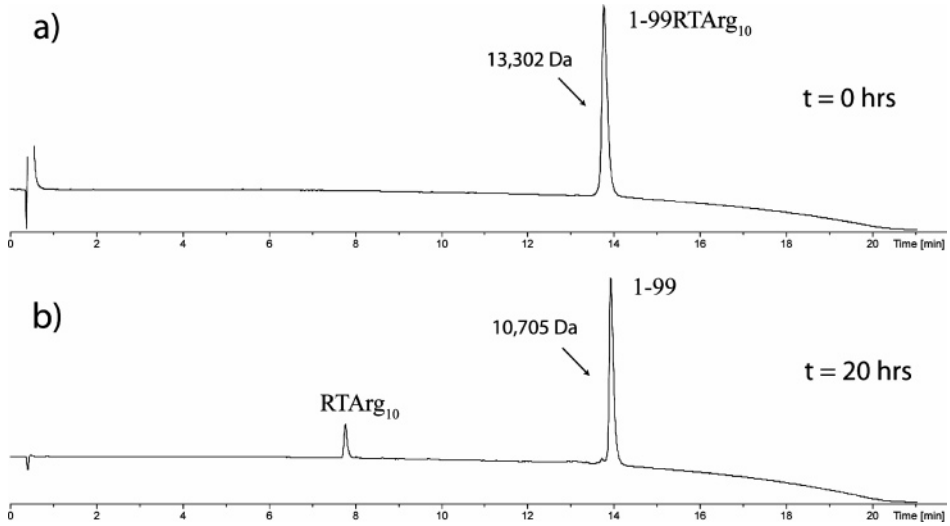


Figure 4. Folding of 1−99RTArg$_{10}$. (a) RP-HPLC analysis of purified 1−99RTArg$_{10}$ at $t = 0$ in 6 M guanidine hydrochloride phosphate buffer (0.2 M, pH 7.4). (b) Analysis of folding solution after 20 h. The principal peak has shifted to a slightly later elution time, and a peak corresponding in mass to the RTArg$_{10}$ peptide has appeared (for details of mass analyses, see Figure S3, Supporting Information). Analytical RP-HPLC was performed as described previously.

to process the viral polyproteins into their individual protein components, including the HIV-1 PR itself, that are required for viral replication and propagation. We took advantage of this autoprocessing mechanism in our revised synthetic design by adding a sequence of the reverse transcriptase polypeptide adjacent to the protease polypeptide in the Gag−Pol polyprotein to the C-terminus of the protease, in addition to 10 arginine residues to increase the solubility of the product Thz71−99RTArg$_{10}$ peptide after ligation of Cys95−99RTArg$_{10}$ to Thz71−94$^{\alpha}$COSR. After all ligations were complete, we reasoned that the protease would recapitulate in vitro its function in vivo and remove the RTArg$_{10}$ modification from its own C-terminus to give the mature 99-residue polypeptide. To our satisfaction, not only did the RTArg$_{10}$ modification increase the solubility of the Cys71−99RTArg$_{10}$ peptide to such an extent where ligation was favored over thioester hydrolysis, but also the product synthetic protein was able to remove the additional residues from its own C-terminus upon folding of the 119-residue polypeptide to yield the desired enzyme molecule, a homodimer of identical 99-residue polypeptide chains. As demonstrated by multiple experimental techniques, the final synthetic protein material was of very high purity and correct structure, with full catalytic activity.

Efficient autoprocessing of synthetic 1−99RTArg$_{10}$ provides strong additional support for the proposed nature of the enzyme maturation mechanism in vivo: in principle, only one molecule of HIV-1 PR or another cell-encoded protease would be required to initiate the autoproteolysis leading to complete polyprotein maturation. This level of experimental control is difficult to achieve by the techniques of traditional molecular biology; it is impossible to definitively rule out a first proteolytic event by a cell-encoded enzyme. Total chemical protein synthesis is inherently free from contamination by cellular enzymes and, in the case of autoproteolytic enzymes, provides the ability to separate synthesis of the polypeptide from activation of the enzyme.

This is the first time to our knowledge that a synthetic product itself completes a synthesis through cleavage of a covalent bond. This type of "recursive" synthetic design may find other uses in the synthesis of other enzymes, when one wishes to remove a functionality useful during the synthesis but not part of the desired final synthetic product.

Conclusion

We have developed a novel biomimetic approach to the total chemical synthesis of the HIV-1 protease. A modular synthesis using native chemical ligation, Raney nickel desulfurization, and autoprocessing gave good yields of high-purity synthetic protein, with full enzymatic activity and the correct fold as determined by atomic-resolution X-ray crystallography. As exemplified here by the incorporation of Nle[36,46] and Abu[67], such a modular synthetic strategy enables the efficient incorporation of nonnatural amino acids into the protein in a straightforward and economical manner. Future structure−function studies on the HIV-1 protease can now take advantage of this modular synthetic route to elucidate the details of enzyme catalysis.

Experimental Procedures

Materials. 2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), *S*-tritylmercaptopropionic acid, *p*-methylbenzhydrylamine (MBHA) resin, Boc-Lys(2ClZ)-OCH$_2$-phenylacetamidomethyl (Pam) resin, and *tert*-butoxycarbonyl- (Boc-) amino acids were obtained from Peptides International. Boc-Arg(Tos)-OCH$_2$-Pam resin and Boc-Phe-OCH$_2$-Pam resin were purchased from Applied Biosystems, Foster City, CA. Side-chain protecting groups used were Arg (Tos), Cys (4MeBzl), Asp (cHex), Asn (Xan), Glu (cHex), His (Bom), Lys (2ClZ), Ser (Bzl), Thr (Bzl), Trp (CHO), and Tyr (BrZ). *N*,*N*-Diisopropylethylamine (DIEA) was obtained from Applied Biosystems. *N*,*N*-Dimethylformamide (DMF), dichloromethane, diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were purchased from Fisher. Trifluoroacetic acid (TFA) was obtained from Halocarbon Products. HF was purchased from Matheson. All other reagents were purchased from Sigma−Aldrich. All glassware was thoroughly washed with chromic acid cleaning solution before use.

Peptide Synthesis. The Thz28−70 peptide-aCOSCH$_2$CH$_2$COArg thioester was synthesized as described previously.[24] Peptide-$^{\alpha}$COSCH$_2$-CH$_2$COArg$_6$ thioesters were synthesized by coupling five arginine

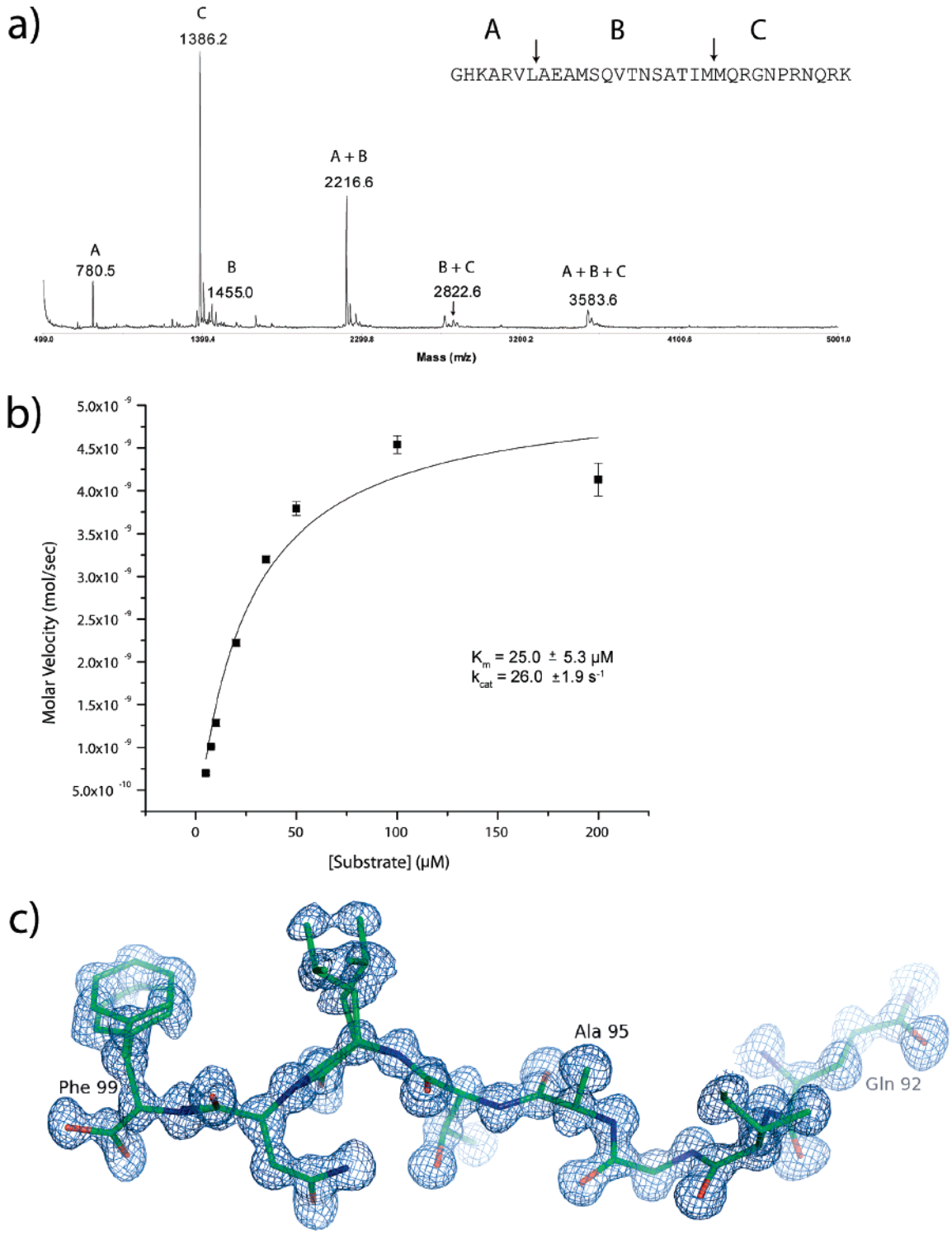


Figure 5. (a) Cleavage of the p24/p15 substrate peptide by synthetic HIV-1 PR. Substrate peptide (1.8 mg, 0.5 mM) was dissolved at 0.5 mM in acetate buffer (50 mM, pH 5.6) and incubated with 13 pmol of HIV-1 PR at 37 °C for 9 h. The reaction was then analyzed by MALDI-TOF MS, with 2,5-dihydroxybenzoic acid as the matrix. All possible peptide cleavage products can be accounted for (p24/p15 expected mass 3582.1 Da, observed mass 3582.6 ± 1.8 Da; GHKARVL expected mass 779.9 Da, observed mass 779.5 ± 0.4 Da; AEAMSQVTNSATIM expected mass 1453.6 Da, observed mass 1454.0 ± 0.7 Da; MQRGNPRNQRK expected mass 1384.6 Da, observed mass 1385.2 ± 0.7 Da; GHKARVLAEAMSQVTNSATIM expected mass 2215.5 Da, observed mass 2215.6 ± 1.1 Da; AEAMSQVTNSATIMMQRGNPRNQRK expected mass 2820.2 Da, observed mass 2821.6 ± 1.4 Da). The small peaks to the right of the principal peaks correspond to sodium adducts. (b) Steady-state kinetics of synthetic HIV-1 PR. The fluorogenic substrate Abz-TI(Nle)F-(NO_2)QR-CO-amide, where Abz = 2-aminobenzoyl, Nle = norleucine, and F(NO_2) = *p*-nitrophenylalanine, was incubated at various concentrations with the enzyme (100 nM), and initial velocities were determined by the rate of increase in fluorescence upon substrate cleavage. k_{cat} and K_m were determined after fitting of the data points to the Michaelis−Menten equation by use of a nonlinear least-squares fitting program in Origin 7.0. (c) Final σ_A-weighted $2F_o - F_c$ electron density map, contoured at a level of 1σ shown in blue wire mesh, for the C-terminal region of synthetic HIV-1 PR, superimposed on a stick representation of residues 92−99. Carbon, oxygen, and nitrogen atoms are shown in green, red, and blue, respectively. The map clearly shows the successful desulfurization of Cys[95] to L-Ala,[95] as well as the termination of main-chain density at residue Phe.[99]

residues to the preloaded Boc-Arg(Tos)-OCH_2-Pam resin before coupling with β-mercaptopropionic acid and continuing with chain assembly.[19] The fluorogenic substrate and Cys95−99RTArg$_{10}$ peptides were synthesized on MBHA resin; the p24/p15 substrate peptide was synthesized on Boc-Lys(2ClZ)-OCH_2-Pam resin. Peptides were synthesized on a 0.4 mmol scale by the manual Boc chemistry in situ neutralization/HBTU protocol described previously,[15] with slight modifications. Briefly, N^{α}-Boc removal was achieved with neat TFA, via one rapid wash followed by a 2 min batch treatment. The N^{α}-deprotected peptide−resin was subject to a single flow wash with DMF. Amino acids (2.2 mmol) were dissolved in 4 mL of 0.5M HBTU (i.e., 2.0 mmol), were activated by addition of DIEA (1 mL, 6.6 mmol) for 1 min, and then added to the peptide−resin; coupling was carried out for 12 min, followed by a single DMF flow wash to remove excess activated amino acid and soluble coproducts. After chain assembly was complete, the N^{α}-Boc group was removed as described above, and the peptide−resin was washed with dichloromethane and dried by aspiration. The peptide was cleaved from the resin and the side-chain protecting groups were simultaneously removed by treatment with anhydrous HF at 0 °C for 1 h, with 10% *p*-cresol added as a scavenger. After removal of HF by evaporation, the peptide was precipitated, washed with cold diethyl ether, dissolved in aqueous acetonitrile + 0.1% TFA, and then lyophilized. Isolated yields after HPLC purification, based on the loading of the resin, were 14.1% for 1−27$^{\alpha}$COSR, 13.3% for Thz28−70$^{\alpha}$COSR, 23.7% for Thz71−94$^{\alpha}$COSR, and 46% for Cys95−99.

HPLC. Analytical reversed-phase HPLC was performed on an Agilent 1100 system with Microsorb C-4 (5 μm, 2.1 × 50 mm) silica columns packed in-house. Peptides were eluted from the column with a gradient of 5−65% acetonitrile/0.08% TFA versus water/0.1% TFA at a flow rate of 0.5 mL/min. Preparative HPLC of crude peptides after SPPS was performed on a Vydac C-18 (50 × 250 mm) column, and preparative HPLC of 1−99RTArg$_{10}$ after desulfurization was performed on an Agilent C-3 (22 × 250 mm) column. Peptides were eluted from the column with an appropriate shallow gradient of acetonitrile/0.08% TFA versus water/0.1% TFA. Fractions containing the desired purified peptide product were identified by analytical LC-MS, combined, and lyophilized.

MS Analysis. Peptide masses were obtained on an Agilent 1100 LC-MS with on-line electrospray ion trap detection. MALDI-TOF MS analysis of the folding reaction was performed by mixing a sample of the reaction mixture and a sample of sinapinic acid (saturated in aqueous acetonitrile + 0.1% TFA) in a 1:1 mixture. MALDI spectra were acquired on a Perseptive Biosystems Voyager-DE in linear mode.

Peptide Ligations. Ligations were performed on a 5 μmol scale (2.5 mM) in 6 M guanidine hydrochloride phosphate buffer (0.2 M) at pH 7.0, with 200 mM 4-mercaptophenylacetic acid (MPAA) added as a catalyst and 20 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) added as a reducing agent. MPAA was repurifed from the commercial source by HPLC before use. After each ligation, the Thz was converted into cysteine with 0.2 M methoxylamine HCl at pH 4.0. After Thz conversion was complete, the crude peptide products were subjected to solid-phase extraction (SPE) on C-4 silica SPE columns and recovered in 50% aqueous acetonitrile and lyophilized. After the final ligation was complete, cysteine (1 M) was added and the solution was made 20% in piperidine for 20 min to remove formyl protection on tryptophan. The solution was then acidified with HCl, purified by SPE to remove all nonpeptidic material, and then lyophilized.

Desulfurization. Lyophilized crude peptide 1−99RTArg$_{10}$ (67 mg) was dissolved in 10 mL of 6 M guanidine hydrochloride phosphate buffer (0.2 M, pH 7.4) with 20 mM TCEP. Raney nickel was prepared by dissolving 3.6 g of nickel acetate in 24 mL of H_2O and slowly adding 600 mg of $NaBH_4$. The charged nickel was then washed in batch 3× with 50 mL of water and added to the peptide solution. The reaction was allowed to proceed under argon until complete as judged by LC-MS of aliquots, after which the reaction solution was centrifuged and the supernatant was decanted. The catalyst was washed 3× with 5 mL of ligation buffer, with the final wash including 1 M imidazole to displace any peptide adsorbed to the nickel. The supernatants were pooled and the combined solution was acidified to pH 2 and then purified by RP-HPLC as described above. The overall yield of purified 119 residue polypeptide for all ligations and other synthetic manipulations was 26% based on the starting peptide segment Cys95−99RTArg$_{10}$.

Folding. 1−99RTArg$_{10}$ polypeptide (15 mg) was dissolved in 12 mL of 6 M guanidine hydrochloride phosphate buffer (0.2 M, pH 7.4) and folded by a three-step dialysis procedure against sodium acetate buffer (pH 5.6) by use of a 3500 MW cutoff membrane. The first step was against 2 L of 50 mM sodium acetate buffer (pH 5.6) for 3 h. The second step was against 2 L of 25 mM sodium acetate buffer (pH 5.6) for 2 h. The third step was against 2 L of 10 mM sodium acetate buffer (pH 5.6) overnight at 4 °C. Final conditions were 10 mM sodium acetate, pH 5.6. The final concentration of folded protein was determined by absorbance at 280 nm with a calculated dimer extinction coefficient of 25 120 M^{-1} cm^{-1}.

Enzyme Kinetics. The fluorogenic substrate Abz-TI(Nle)F(NO_2)-QR-CONH, where Abz = 2-aminobenzoyl, Nle = norleucine, and F(NO_2) = *p*-nitrophenylalanine, was incubated at various concentrations with the enzyme (100 nM), and initial velocities were determined by the initial rate of increase in fluorescence after substrate cleavage.[25] Final assay conditions were 50 mM sodium acetate buffer, pH 5.6, and 1% dimethyl sulfoxide (DMSO), 37 °C. Assays were performed in triplicate for each substrate concentration. k_{cat} and K_m were determined after fitting of the data points to the Michaelis−Menten equation by use of a nonlinear least-squares fitting program in Origin 7.0.

Crystallization. Crystals were grown at room temperature by the hanging drop vapor diffusion method from a well solution consisting of 0.1 M citrate, 0.2 M sodium phosphate, 30% saturated ammonium sulfate, and 10% DMSO, pH 6.0. Protein solution (aliquot of concentrated folding solution) (∼0.5 mM enzyme) was preincubated with a 20-fold molar excess of (*S*)-JG-365 and was then mixed in a 2:1 (v/v) ratio with well solution. Crystals grew within 5 days and were frozen in liquid nitrogen with a cryoprotectant of 30% glycerol.

Structure Determination and Refinement. Diffraction data were collected at 19-BM station in the Advanced Photon Source (APS) at Argonne National Laboratory and processed with HKL2000.[26] PHASER[27] was used to obtain the initial phases, by use of the model of HIV-1 protease in complex with JG-365 (PDB code 7HVP). The starting model was initially refined by the CNS program[28] with data between 50 and 1.5 Å. After free *R*-factor convergence,[29] the model was fed into SHELX97[30] for full anisotropic temperature factor refinement, with data to the highest resolution shell and with low-resolution cutoff at 10 Å. Iterative model building and refinement were performed with COOT[31] and SHELXL[30] until the final R_{cryst}/R_{free} values converged to 14.36/18.37%. Coordinates for the refined model have been deposited in the Protein Data Bank with accession code 2JE4. Model analysis

(24) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10068−10073.
(25) Toth, M. V.; Marshall, G. R. *Int. J. Pept. Protein Res.* **1990**, *36*, 544−550.
(26) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307−326.
(27) Read, R. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 1373−1382.
(28) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905−921.
(29) Brunger, A. T. *Nature* **1992**, *355*, 472−475.
(30) Sheldrick, G. M.; Schneider, T. R. *Methods Enzymol.* **1997**, *277*, 319−343.
(31) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.

was carried out with programs of the CCP4 package.[32] Graphics were generated by Pymol.[33]

Acknowledgment. We thank Brad Pentelute for helpful discussions regarding the desulfurization reaction and Hookang Im for assistance with X-ray diffraction data collection. This work was supported by the National Institutes of Health (GM075993-01 to S.B.H.K., GM62548 to W.-J.T.) and by the Department of Energy GTL-Genomics Program (DE-FG02-04ER63786 to S.B.H.K.). Y.S. is funded by an AHA fellowship (0520123Z), and E.C.B.J. is funded by the Graduate Training Program in Growth and Development (NIH T32 HD007009).

Supporting Information Available: Synthetic peptide sequences, LC-MS of crude polypeptides, ligation data using initial synthetic strategy, mass analysis of the folding reaction, kinetic parameters of chemically synthesized HIV-1 PR (PDF); X-ray crystallography of synthetic HIV-1 PR (CIF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA072870N

(32) Bailey, S. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760−763.
(33) DeLano, W. L. DeLano Scientific LLC, South San Francisco, CA, 2005; http://www.pymol.org.