Conformational Selection of Inhibitors and Substrates by Proteolytic Enzymes: Implications for Drug Design and Polypeptide Processing

David P. Fairlie,*,† Joel D. A. Tyndall,† Robert C. Reid,† Allan K. Wong,† Giovanni Abbenante,† Martin J. Scanlon,† Darren R. March,† Douglas A. Bergman,† Christina L. L. Chai,‡ and Brendan A. Burkett‡

Centre for Drug Design and Development, University of Queensland, Brisbane, Queensland 4072, Australia, and Research School of Chemistry, The Australian National University, Canberra, Australian Capital Territory 2600, Australia

Received June 21, 1999

Inhibitors of proteolytic enzymes (proteases) are emerging as prospective treatments for diseases such as AIDS and viral infections, cancers, inflammatory disorders, and Alzheimer's disease. Generic approaches to the design of protease inhibitors are limited by the unpredictability of interactions between, and structural changes to, inhibitor and protease during binding. A computer analysis of superimposed crystal structures for 266 small molecule inhibitors bound to 48 proteases (16 aspartic, 17 serine, 8 cysteine, and 7 metallo) provides the first conclusive proof that inhibitors, including substrate analogues, commonly bind in an extended β-strand conformation at the active sites of all these proteases. Representative superimposed structures are shown for (a) multiple inhibitors bound to a protease of each class, (b) single inhibitors each bound to multiple proteases, and (c) conformationally constrained inhibitors bound to proteases. Thus inhibitor/substrate *conformation,* rather than sequence/composition alone, influences protease recognition, and this has profound implications for inhibitor design. This conclusion is supported by NMR, CD, and binding studies for HIV-1 protease inhibitors/substrates which, when preorganized in an extended conformation, have significantly higher protease affinity. Recognition is dependent upon conformational equilibria since helical and turn peptide conformations are not processed by proteases. *Conformational selection* explains the resistance of folded/structured regions of proteins to proteolytic degradation, the susceptibility of denatured proteins to processing, and the higher affinity of conformationally constrained 'extended' inhibitors/substrates for proteases. Other approaches to extended inhibitor conformations should similarly lead to high-affinity binding to a protease.

Introduction

Proteolytic enzymes of the four major classes (aspartic, cysteine, metallo, serine) catalyze both intracellular and extracellular peptide cleavages during numerous physiological processes, including digestion, fertilization, growth, differentiation, cell signaling and migration, immunological defense, wound healing, apoptosis, protein turnover, and general 'housekeeping' functions. Proteases are also vital in the propagation of most disease processes. Thus selective inhibition of proteases[1] is increasingly being targeted for the treatment of cancers;[2] parasitic, fungal, and viral infections (e.g. schistosomiasis,[3] malaria,[4] *C. albicans*,[5] HIV,[6] hepatitis C,[7] herpes,[8] common cold[9]); inflammatory, immunological, and respiratory conditions;[10] and cardiovascular[11] and degenerative disorders including Alzheimer's disease.[12] There are now many designed potent and selective protease inhibitors that slow or halt disease progression,[13] HIV-1 protease inhibitors being notable for their speedy emergence from clinical trials.[6]

Progress toward the design of small molecule inhibitors of proteases is significantly handicapped by the unpredictable conformational changes that can occur upon substrate/inhibitor binding to proteases, and by the lack of structural information until recently for protease−inhibitor complexes. Even now there is no structural information available for most known proteases, let alone their inhibitor-bound complexes. Most approaches to protease inhibitors have begun with some knowledge of which peptide substrates are processed by the target protease. Proteases tend to recognize short (<10 residue) peptide sequences of their polypeptide substrates, the amino acid side chains being crucial for determining selectivity and susceptibility to cleavage.[1] However when isolated, these short peptides are most commonly found in rapidly interconverting random conformations, whereas native polypeptide substrates can fold via different intramolecular hydrogen bonds into well-defined fixed conformations such as α-helices, β- or γ- turns, and β-sheets. Very little is presently known about whether substrate conformation influences protease recognition or cleavage. Also, the most abundant substrate/inhibitor conformation in solution may not necessarily be the one recognized by a protease, and conformational equilibria may significantly influence rates of peptide processing.

We now pool together some compelling structural evidence that aspartic, serine, metallo, and cysteine proteases all share a common conformational requirement for recognition, namely an extended (β-strand) conformation for their active-site-directed inhibitors and substrate analogues. It has previously been noted for individual crystal structures of protease−inhibitor com-

* To whom correspondence should be addressed. Tel: +61 7 336 51271. Fax: +61 7 336 51990. E-mail: d.fairlie@mailbox.uq.edu.au.
† University of Queensland.
‡ The Australian National University.

10.1021/jm990315t CCC: $19.00
Published on Web 03/11/2000

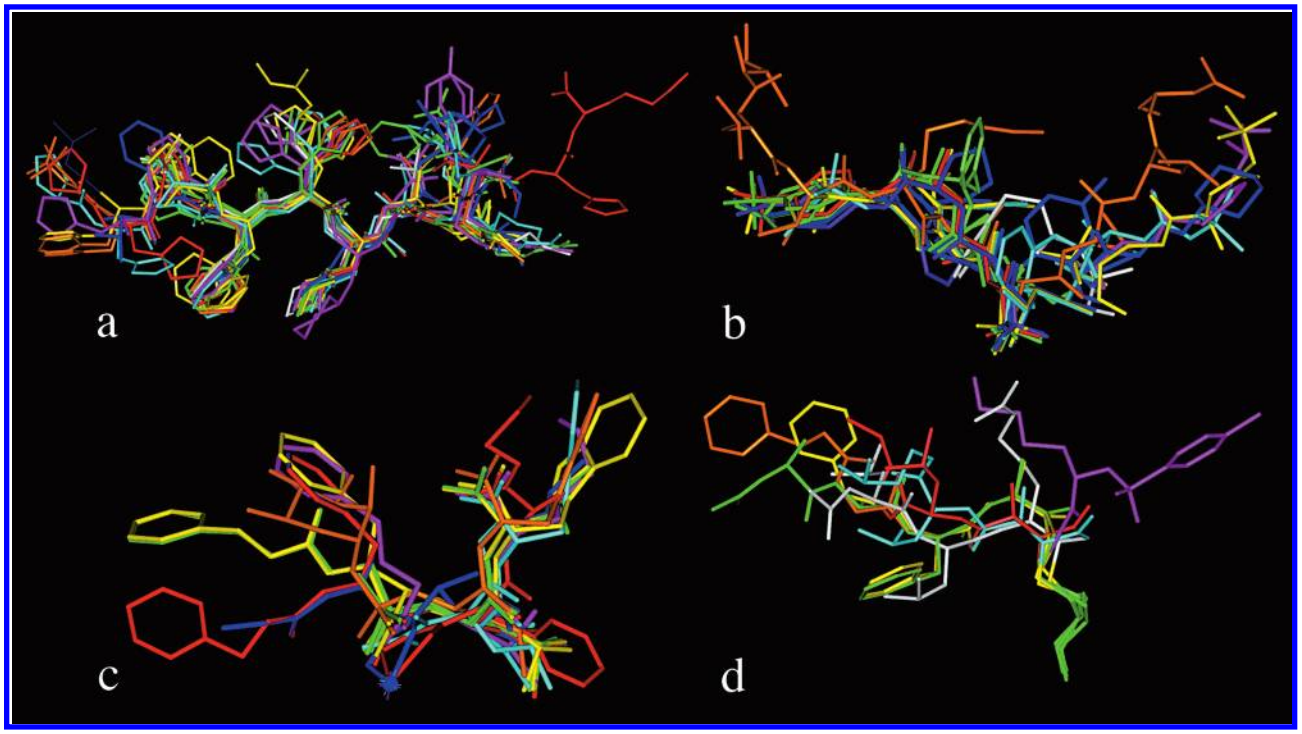


Figure 1. Superimposed protease−inhibitor X-ray crystal structures (PDB codes), proteases omitted: (a) endothiapepsin-binding conformations of 21 inhibitors (1eed, 1ent, 1epl, 1epm, 1epn, 1epo, 1epp, 1epq, 1epr, 1er8, 2er0, 2er6, 2er7, 2er9, 3er3, 3er5, 4er1, 4er2, 4er4, 5er1, 5er2); (b) porcine pancreatic elastase-binding conformations of 16 inhibitors (1eal, 1eas, 1eat, 1eau, 1eld, 1ele, 1elf, 1esb, 1bma, 1fle, 1inc, 1jim, 1nes, 2est, 5est, 7est); (c) thermolysin-binding conformations of 13 inhibitors (1thl, 1tlp, 1tmn, 2tmn, 3tmn, 4tln, 4tmn, 5tln, 5tmn, 6tmn, 7tln, 7tmn, 8tln), metal shown as white sphere; (d) papain-binding conformations of 8 inhibitors (1pad, 1pe6, 1pip, 1pop, 1ppp, 4pad, 5pad, 6pad), cysteine in green with sulfur shown as yellow sphere. Rmsd values were 0.30 ± 0.19 Å (a), 0.30 ± 0.06 Å (b), 0.17 ± 0.04 Å (c), and 0.37 ± 0.30 Å (d).

plexes (e.g. separate reports for each PDB-filed structure in Table 1, Supporting Information) that the extended strand conformation is important for recognition,[14] but recent publication of large numbers of three-dimensional structures for protease−inhibitor complexes now permits this more extensive analysis of inhibitor conformational preferences for all four classes of proteases.

A strategy that fixes the extended conformation of an inhibitor is also examined below, for several proteases from different classes, and found to enhance inhibitor affinity for a protease. For one protease (HIV-1 protease), additional structural (CD, NMR), binding, and limited mechanistic observations for substrates and substrate analogues support the hypothesis that they too are recognized and processed in only an extended conformation. Proteases appear most likely to select for, rather than structurally reorganize inhibitors/substrates to, a single (extended) conformation. This *conformational selection* phenomenon appears to be common for proteases and has profound implications for protein folding and stability, as well as for design of protease inhibitors as potential drugs.

Results

Proteases with Multiple Inhibitors. Inhibitors of the aspartic[15a] protease, endothiapepsin,[15b] adopt an extended strand conformation when bound to the protease as exemplified in Figure 1a for 21 superimposed PDB-filed crystal structures. All proteases were superimposed, but only protease-bound conformations of inhibitors are displayed. We also examined 15 other aspartic proteases for which there are PDB coordinates for cystal structures with bound inhibitors (Table 1) including human immunodeficiency virus protease, HIV-1PR (50 inhibitors) and related viral proteases (HIV-2 (13 inhibitors), SIV (5), FIV (1)), cathepsin D (1), renin (3), rennin/chymosin (1), penicillopepsin (7), secreted aspartic protease (2), pepsin (2), mucoropepsin (1), retropepsin (1), saccharopepsin (1), rhizopuspepsin (4), and plasmepsin II (1). All 114 structures (Table 1) show a common *extended* (*β*-strand) conformation for inhibitors bound to aspartic proteases. Since many of these inhibitors are substrate analogues, containing identical peptide sequences to substrates but with the cleavable amide bond replaced by uncleavable transition-state analogues, substrates are also likely to be recognized in their strand conformations.

Serine proteases[16] are classified by their substrate specificity as either trypsin-like (positively charged residues (Lys/Arg) preferred at P1), chymotrypsin-like (large hydrophobic resiudes (Phe/Tyr/Leu) at P1), or elastase-like (small hydrophobic residues (Ala, Val) at P1). All three types were compared. Figure 1b shows superimposed conformations of 20 inhibitors complexed to porcine pancreatic elastase. Other PDB-filed crystal structures examined for inhibitor−protease complexes of serine proteases (Table 1) include trypsin (10 inhibitors), α-chymotrypsin (1), *γ*-chymotrypsin (19), human neutrophil elastase (3), α-lytic protease (9), thrombin (10), subtilisin (7), proteinase A (4), achromobacter (1), human cathepsin G (1), glutamic acid-specific protease (1), carboxypeptidase (3), blood coagulation factor VIIa (1), porcine factor IXA (1), mesentericopeptidase (1), and thermitase (1). All 95 structures (Table 1) show a common extended inhibitor conformation for these 17 serine proteases.

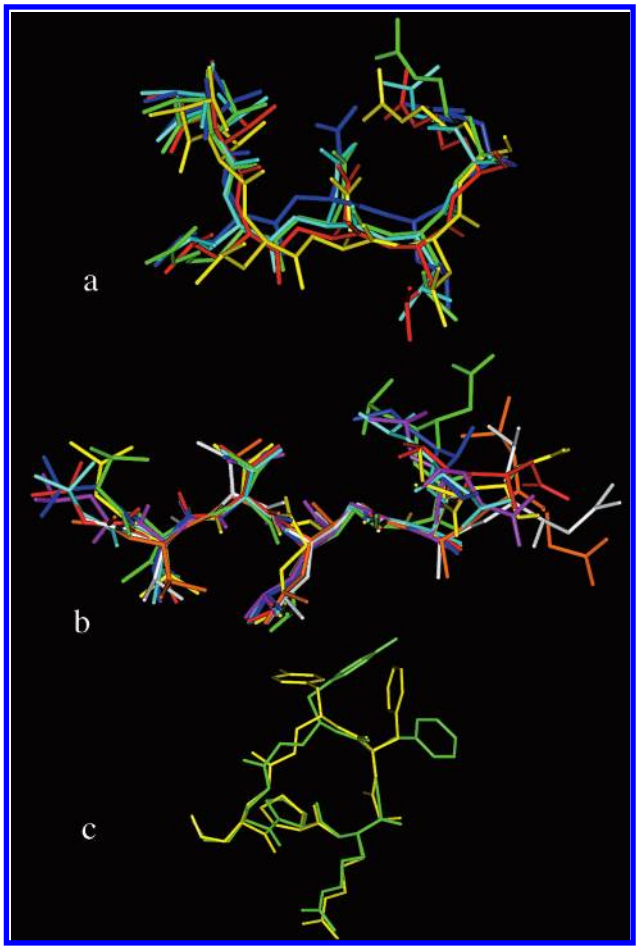


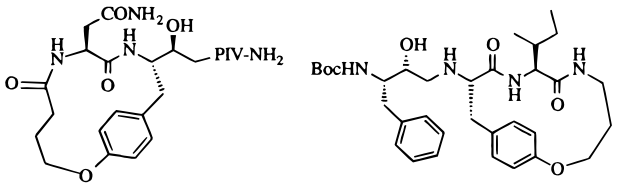
Figure 2. Superimposed protease−inhibitor X-ray crystal structures (PDB codes), proteases omitted: (a) binding domain of Eglin C complexed with the serine proteases α-chymotrypsin (dark blue, 1acb), subtilisin carlsberg (red, 1cse), mesentericopeptidase (yellow, 1mee), subtilisin novo (green, 1sbn), and thermitase (light blue, 1tec); (b) binding domain of pepstatin complexed with the aspartic proteases cathepsin D (pink, 1lyb), pepsin A (red, 1pso), plasmepsin II (yellow, 1sme), mucoropepsin (green, 2rmp), endothiapepsin (light blue, 4er2), rhizopuspepsin (dark blue, 6apr), HIV-1 (acetylpepstatin, orange, 5hvp), and HIV-2 (acetylpepstatin, white, 1phv); (c) binding domain of cyclotheonamide A (**4**) complexed with the serine proteases thrombin (green, 1tmb) and trypsin (yellow, 1tyn).

Metalloproteases[17] for which inhibitor−protease crystal structures could be accessed (Table 1) were thermolysin (13 inhibitors), matrilysin (3), neutrophil collagenase (6), interstitial collagenase (2), atrolysin C (1), stromelysin (4), and carboxypeptidase A (9). These 38 structures establish that all 7 metalloproteases bind to their inhibitors in extended strand conformations as depicted in Figure 1c for thermolysin-bound inhibitors. Cysteine proteases[18] exist in three structurally different classes typified as papain-like (e.g. cathepsins), picornaviral proteases (similar to serine proteases with cysteine replacing serine), and ICE-like proteases (caspases). Crystal structures were examined for inhibitors bound to papain (8 inhibitors; Figure 1d), and for inhibitors complexed with cathepsin B (2), ICE (2), apopain or CPP32 (2), actinidin (1), glycyl endopeptidase (1), cruzain (2), and calpain (1). In all 19 structures (Table 1) the protease-bound inhibitors adopt extended strand conformations (e.g. Figure 1d).

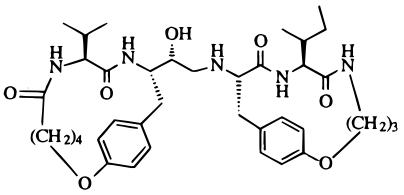
These results demonstrate that all four classes of proteases bind inhibitors in a common extended β-strand conformation forming enzyme−inhibitor−enzyme antiparallel β-sheet arrangements.

Inhibitors with Multiple Proteases. Figure 2 shows conformations of single inhibitors each bound to multiple proteases, an arbitrary structure being chosen for superimposition of the others. The serine protease-binding domain of Eglin C, a proteinacious inhibitor from a leech,[19] adopts an extended conformation between P4−P1 when bound to five different serine proteases (Figure 2a). The classic aspartyl protease inhibitor, pepstatin (or acetylpepstatin), is a hexapeptide inhibitor which adopts the same extended conformation between P3−P3′ when bound to at least eight different aspartic proteases (Figure 2b, rmsd 0.1−0.5 Å versus PDB: 4er2). The natural product cyclotheonamide binds in an extended conformation between P3−P1 (Figure 2c) to both the digestive enzyme trypsin and the trypsin-like thrombin involved in clotting (rmsd 0.43 Å for superimposed backbone atoms). These examples strengthen the notion of conformational homogeneity in protease recognition.

Cyclic Protease Inhibitors. To test the proposition that proteases preferentially recognize an extended conformation, we and others have devised methods to stabilize the β-strand inhibitor conformation.[20,21] Our strategy was to fix inhibitors in protease-binding conformations by linking together side chains from $i,i+2$ amino acids to form macrocycles (e.g. **1**−**3**). We have



1 (K_i 12 nM) **2** (K_i 4 nM)

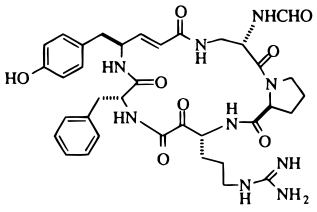


3 (K_i 3 nM)

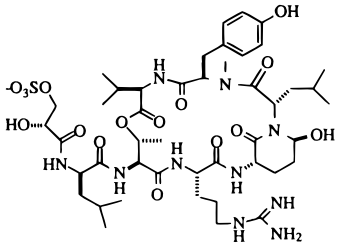
earlier reported the syntheses of **1**−**3** which are potent inhibitors of HIV-1 protease.[22] We now show (Figure 3a) that the structures of **1** and **2** bound to HIV-1 protease (PDB codes: 1mtr, 1cpi)[23] superimpose extremely well upon the structures of 20 other acyclic inhibitors of HIV-1 protease, not all of which are peptides (Figure 3a). This structural mimicry is achieved through conformational restrictions jointly imposed by cyclization and constraints in the cycle (two planar amides and aromatic ring). The hydroxyethylamine isostere attached to the cycle adopts slightly different positions in the enzyme, due to the influence of the charged secondary amine of **2** which also hydrogen bonds to one of the catalytic aspartate residues in the enzyme.[22b,23] This also has an effect on the location of the benzyl substituent at P1, and these changes reflect some

flexibility in the capacity of the enzyme to bind cooperatively to the inhibitor.

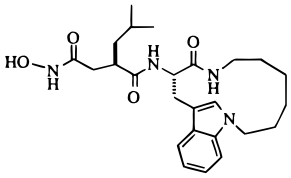
Unlike most known inhibitors of HIV-1 protease,[6] **1**−**3** are restrained in an extended conformation (Figure 3a) that is essentially preorganized for receptor binding.[20,22,23] Due to reduced conformational entropy for protease-binding, these cyclic inhibitors of HIV-1 protease are 10−100 times more potent than acyclic analogues of comparable size.[20c,24] Cyclization also protects amide bonds from proteolytic degradation, a major problem with bioavailability of acyclic peptides. This mimetic strategy can be extended to other proteases since, for example, we find that protease-bound cyclic natural products **4** (Figure 2c, PDB: 1tyn) and **5** (Figure 3B, PDB: 1tps) and the synthetic inhibitor **6** (PDB: 1mmp) also exist in an extended conformation. These compounds are known to potently inhibit trypsin, thrombin, and matrilysin, respectively.



4 (K_i 10 nM)



5 (K_i 1 nM)



6 (K_i 1 nM)

Cyclic peptides in general are usually associated with mimicking turn conformations of peptides.[20a,b] It appears though that if the macrocycle is small enough, some of its components present their peptide backbone or equivalent residues in an extended β-strand conformation.[20b] For example, Figure 4 shows a Ramachandran plot of Φ,Ψ or pseudo Φ,Ψ angle pairs taken from available crystal structures for the peptide backbone of 10 cyclic inhibitors bound to the proteases: HIV-1 protease (3),[22,23] rhizopuspepsin (1),[25] trypsin (2),[26,27] matrilysin (1),[28] and penicillopepsin (3).[29,30] Most of the coordinates lie in the upper left quadrant of the plot (Φ −56.9° to −169.5°, Ψ 35.2° to 161.7°) and typify[31] the extended β-strand peptide conformation. The two exceptions in the upper right quadrant are for compounds **2** and **3** but are due to the transition-state isostere replacement of an amide bond. The two other exceptions in the lower left quadrant result from the piperidine ring of the trypsin inhibitor **5** (lower point) and the turn region of the rhizopuspepsin cyclic inhibitor (upper point). These data establish that the 10 cyclic inhibitors are effective structural mimics of the extended β-strand peptide conformation. Other cyclic protease inhibitors are also known,[20,32] and all adopt an extended peptide or pseudo-peptide conformation.

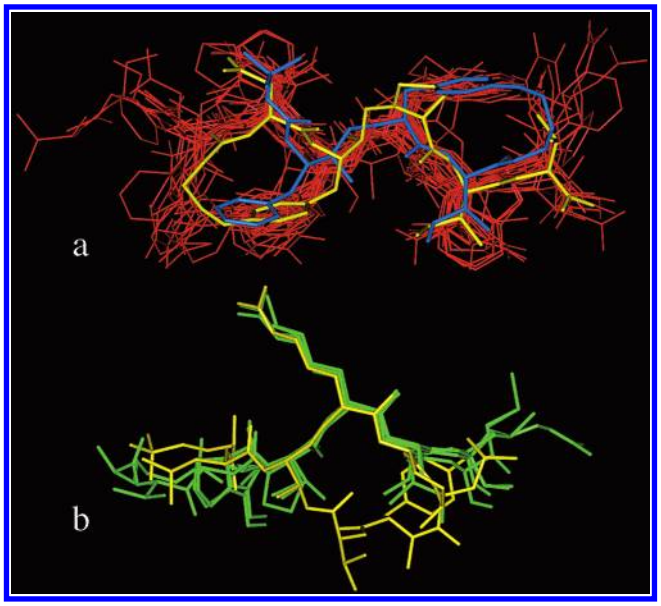


Figure 3. Superimposed protease−inhibitor X-ray crystal structures (PDB codes), proteases omitted: (a) HIV-1 protease-binding conformations of cyclic inhibitors **1** (yellow, 1cpi) and **2** (blue, 1mtr) and 20 acyclic inhibitors (red, 1aaq, 1gnm, 1gnn, 1gno, 1hbv, 1hef, 1heg, 1hih, 1hiv, 1hps, 1hsg, 1hvi, 1hxb, 1sbg, 4hvp, 4phv, 5hvp, 7hvp, 8hvp, 9hvp); (b) trypsin-binding conformation of cyclic inhibitor **5** (yellow, 1tps) and four acyclic inhibitors (green, 1brb, 1brc, 1ppe, 1smf). Rmsd values were 0.57 ± 0.15 Å (a) and 0.43 ± 0.07 Å (b).

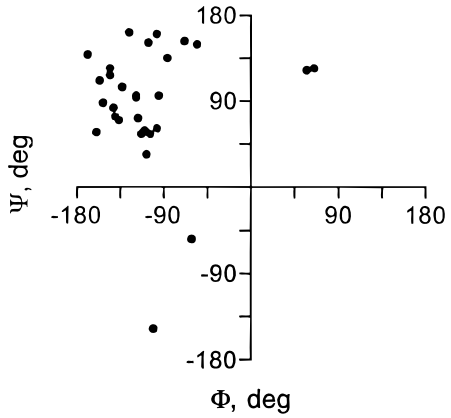
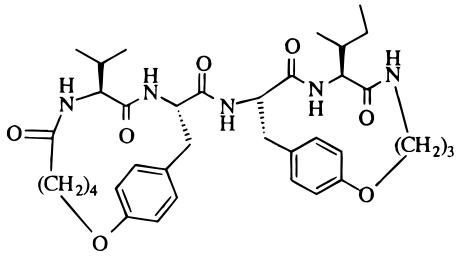


Figure 4. Ramachandran plot of 30 angle pairs (Φ,Ψ° or pseudo Φ,Ψ°) for 10 cyclic inhibitors of HIV-1 protease (3), rhizopuspepsin (1), penicillopepsin (3), trypsin (2), and matrilysin (1). PDB codes: 1cpi,[22a] 1mtr,[22b] 1b6p,[23] 5apr,[25] 1tps,[26] 1tyn,[27] 1mmp,[28] 1bxo,[29] 2wea, 2wed.[30]

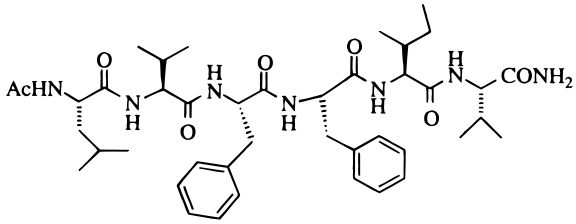
Cyclic Protease Substrates. Since cyclization preorganizes *inhibitors* for binding to a protease (e.g. **1**−**6**), we decided to examine this approach for optimizing protease-binding of *substrates*. Scheme 1 summarizes the synthesis of a novel bicyclic substrate **7**, an analogue of **3** but with the N- and C-terminal cyclic peptides linked together by an amide bond instead of a hydroxyethylamine transition-state isostere. The two chiral centers in each cycle of **7** are simply derived from L-amino acids. The N-terminal cyclic acid was prepared (Scheme 1) by bromoacylating the N-terminus of the dipeptide, H-Val-Tyr-OBz, followed by ring closure via the *p*-hydroxy substituent of Tyr.

Each cycle in **7** is constrained by the presence of only 15 or 16 ring atoms, by two trans amide bonds, and by a para-substituted aromatic ring. Evidence for strain

in each cycle is demonstrated in the ^{1}H NMR data for **7** (see Experimental Section) by the appearance of four separate aromatic proton resonances per cycle, the inequivalence of the aromatic protons being indicative of restricted rotation of each aromatic ring.



7 (Ki 0.8 μM)



8 (Ki 60 μM)

We were interested in comparing the relative affinities for HIV-1 protease of bicycle **7** and its more flexible linear hexapeptide analogue **8**. Traditionally substrates such as **7** and **8** are examined for their turnover kinetics, measuring K_m and k_{cat}. The latter parameter was difficult to obtain in water because these small molecules are not very water soluble and because they are only processed slowly. Also while K_m ($=(k_{-1} + k_2)/k_1$) does reflect the affinity of the substrate for the enzyme, it is complicated by also including the rate of decomposition of the enzyme−substrate complex:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

A much more effective method of obtaining meaningful information about the binding affinity of **7** versus **8** for a protease is to consider these compounds as competitive inhibitors, since:

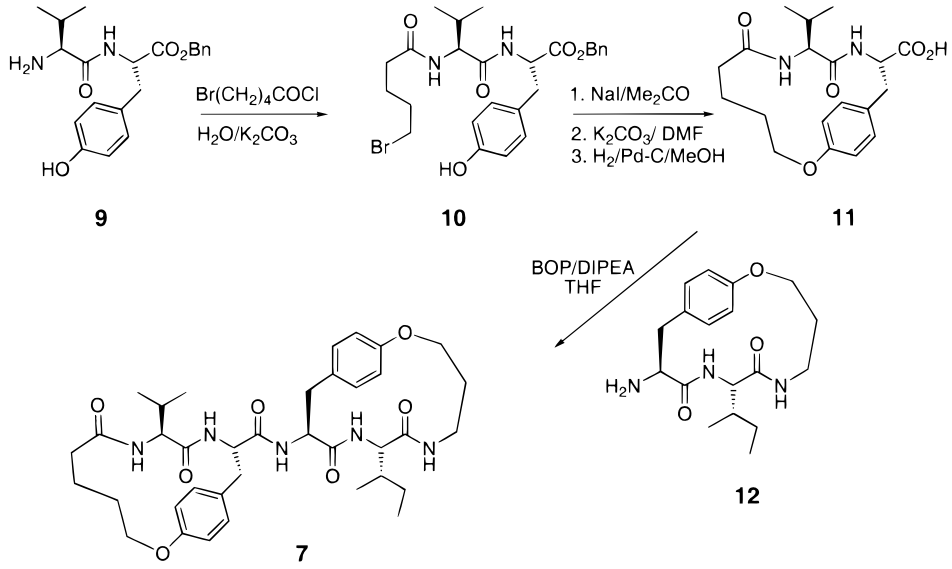
$$E + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EI$$

We therefore chose to measure substrates **7** versus **8** as competitive inhibitors of the processing of another substrate, the fluorogenic substrate 2-Abz-Thr-Ile-Nle-Phe(*p*-NO_2)-Gln-Arg-NH_2 (Abz-NF*-6; $K_m = 26\ \mu M$), by HIV-1 protease at pH 6.5 ($I = 0.1$ M, 37 °C). Competition for the protease by **7** and **8**, which are not directly monitored, is detectable only as reduced proteolytic processing of Abz-NF*-6. We could thus directly obtain protease affinities for substrates by determining K_i ($=k_1/k_{-1}$) for **7** (0.8 μM) and **8** (60 μM) in these experiments. K_i is the equilibrium constant which defines the affinity of the inhibitor (I) for the enzyme (E) and is directly related to the free energy of binding ($\Delta G = -RT \ln K$). Thus bicycle **7** was found to have about 80-fold higher affinity for HIV-1 protease than linear **8**, corresponding to about 11 kJ/mol difference in binding energy.

We also directly measured the processing of substrates **7** and **8** using HPLC to monitor peptide cleavage products. Interestingly, **7** was a poorer substrate (K_m = 110 μM) than **8** (K_m = 20 μM) as defined by the Michaelis−Menton equilibrium constant, yet **7** is more constrained to an extended conformation and has the higher affinity for the protease as determined above using competition experiments with a fluorogenic substrate. This difference can be accounted for if the enzyme−substrate complex (ES) decomposes more slowly for **7** than for **8**, not a surprising result since a good substrate should not bind so tightly to enzyme as to impede turnover, but rather there needs to be a compromise between enzyme affinity and product formation and dissociation. Substrate **7** consists of the same two cyclic components as in **1** and **2**, which each constrain the peptide backbone to the preferred extended conformation (Figure 3a), and its N-terminal cycle is predicted to align in the protease active site in a similar location as does the cycle in **1**.

Acyclic Protease Substrates. *Conformational selection* of substrates has not previously been demon-

Scheme 1



9 **10** **11** **12** **7**

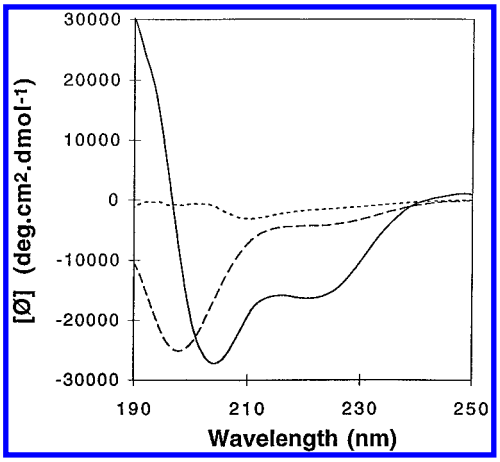


Figure 5. CD spectra of peptides **13** (Ac-K**ARVL**↓**AEA**MSQVTNP-NH_2, – – –), **14** (EQADKLAKEAA**RKVL**↓**FLN**EAAAKAA, —), and **15** (Ac-PSYKGR**PGNF**↓**LQS** RP-NH_2, - - -) in water at 22 °C. Arrow indicates cleavage site for HIV-1 protease. Protease-binding domain is shown in bold.

strated for any protease, although many of the inhibitors crystallized with proteases are substrate analogues and do adopt an extended conformation. If fixing the extended conformation of a substrate leads to higher affinity for a protease (e.g. **7** vs **8**), then the reverse action of stabilizing substrates in alternative conformations (e.g. α-helices or *β*-turns) should reduce their affinity for proteases. To test this proposition, we synthesized three peptide substrates for HIV-1 protease (**13**−**15**), each containing seven sequential amino acids (P4−P3′) from polypeptide substrates that are known to bind in the active site of the protease and are known to be cleaved by HIV-1 protease (Figure 4, bold). For example, peptide **13** contains the p24 cleavage site, and **14** has the RT/IN cleavage site, while **15** contains the p7/p6 cleavage junction. The solution conformations of these peptides are predicted[6e,33] to exist in different solution conformations (**13**, random coil; **14**, α-helix; **15**, turn) when incorporated into longer peptides (Figure 5), and these qualitative structure predictions are confirmed by their CD spectra in Figure 5.

Peptide **14** has the known protease-binding heptapeptide component of polypeptide substrates incorporated into a conformationally biased unnatural sequence designed to further stabilize the α-helix. This compound appears to be helical in water by CD (Figure 5), and ^{1}H NMR spectroscopy confirmed that **14** and **15** retain significant structure in aqueous solution (Figure 6). Sequential NOE connectivities, coupling constants, and chemical shift indices (Figure 6) indicate that **14** is helical along its entire length, while **15** has a small number of medium range NOEs consistent with turns around residues PSYK and RPGN.

Using rpHPLC to monitor substrate cleavages over time, we found that **13**−**15** are all processed by HIV-1 protease but only very slowly, far too slowly to measure K_i in the presence of the fluorogenic substrate Abz-NF*-6. We therefore resorted to measuring K_m values for these substrates. The random coil substrate **13** (K_m = 140 μM), the helical substrate **14** (K_m = 300 μM), and the turn-containing substrate **15** ($K_m = 10^6$ μM) are very poor substrates for HIV-1 protease, requiring days to completely turn over thus preventing us from obtaining reliable values for k_{cat} for these poor substrates. While

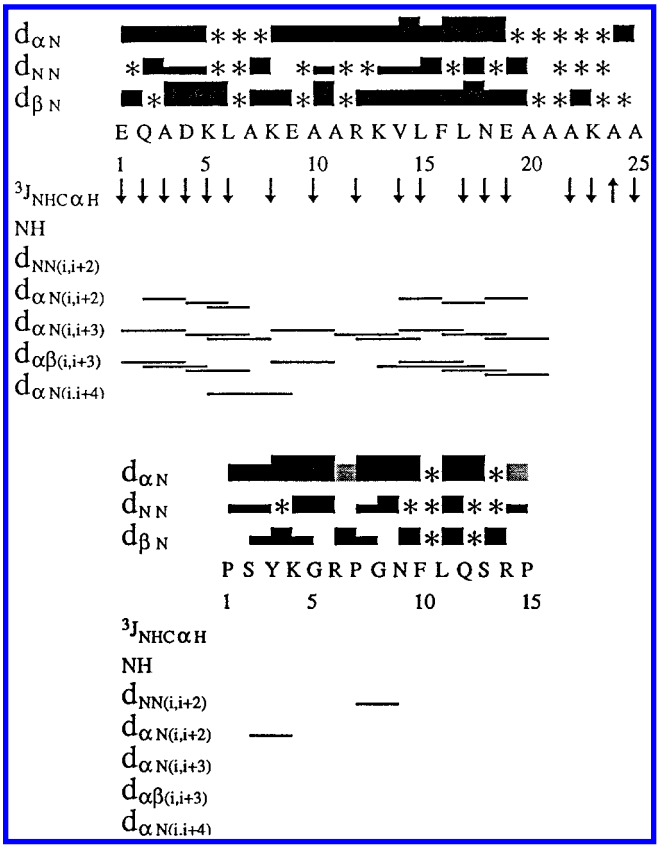


Figure 6. ^{1}H NMR data for sequence assignment and identification of secondary structure in **14** (top) and **15** (bottom). Filled bars indicate sequential NOE connectivities, bar height indicating their strength; $d(i,i+1)$ connectivities not detected due to overlap are asterisked. Lightly shaded bars correspond to sequential connectivities to Pro Hδ's. Values of $^3J_{HNH\alpha}$ are indicated by ↓ if <5 Hz, ↑ if >8 Hz.

the experimental difficulties and different substrate sequences make it difficult for direct quantitative comparisons of protease affinities and substrate-processing rates to be made between **13**−**15**, it is clear from their K_m values at least that **14** and **15** do have low affinities for the protease. These results offer some qualitative support for the idea that helical and turn conformations are not preferred by the protease and that the protease binds only extended conformations of substrates (Figures 1−3). The low rates of substrate processing may reflect the unfavorable equilibrium between the recognized extended conformation and the unrecognized helical/turn conformation (Figure 7). Further work is progressing toward stabilizing a single peptide sequence in three different conformations in order to less ambiguously confirm conformational recognition of substrates by a protease.

Conformational Selection. The requirement of extended conformations for protease substrates is logical for several reasons. First, energy that is expended in folding peptides into helices, turns, or sheets will be conserved if these structures are resistant to proteolytic degradation but would be wasted if peptide cleavage was conformationally indiscriminant. Second, it is intuitively obvious that the stretching of the substrate amide bond toward the transition state, which is largely dictated by interactions between enzyme and 'pocket-filling' inhibitor/substrate substituents, is most efficient when these forces are directionally opposed as enforced by the

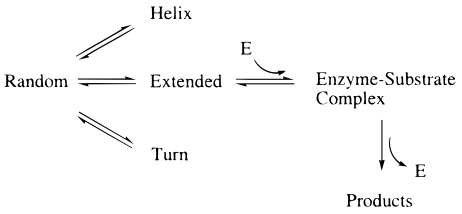


Figure 7. Conformational equilibria for substrates and inhibitors and the proposed conformational selection by proteases.

extended form due to the trans amide bond and the L-conformation of each flanking amino acid. Third, linear extension of a peptide exposes the main chain amide atoms, otherwise shielded by side chains in intramolecularly hydrogen-bonded secondary structures, and promotes intermolecular hydrogen bonding with an enzyme. Fourth, proteins that are substrates for HIV-1 protease are processed *between* elements of secondary structure, rather than within helical, turn, or sheet regions, and rates of processing tend to be greater after denaturation of protein substrates.[34] Fifth, and importantly in the context of the substrates **13**−**15**, the diameter of the inhibitor-bound conformation of the active site of HIV-1 protease (∼6 Å) is far too small to accommodate substrates in either helical or turn conformations so the latter forms would not be expected to be recognized prior to rearrangement to an extended form. This latter observation is not consistent with a model[35] in which the peptide is recognized in its most stable solution conformation (e.g. helix, turn) and then rearranged into the extended strand on protease binding.

These factors all support a conformational selection hypothesis, where different conformers interconvert (Figure 7) but only the extended conformer is recognized and processed by proteases. This could be construed as a controversial hypothesis. For example, it has been suggested elsewhere[35] that entropic considerations do not support the idea that protein−ligand interactions in general proceed via conformational selection, but rather are more likely to proceed via a two/multistep 'zipper' process. We agree that such an induced fit model of protein−ligand interaction is more attractive than a rigid lock-and-key model for protease−ligand interaction, but the truth may well be somewhere in between. Conformational restrictions that preorganize the ligand structure close to its final protease-bound conformation can be expected to reduce the entropic barrier en route to a protease−ligand complex. We have only presented data in this paper that suggest that proteases *ultimately* bind their ligands in an extended ligand conformation and cannot definitively conclude from the crystal structural data alone that the *initial* interaction between protease and ligand requires the ligand to be in an extended conformation. However it does appear from our preliminary modeling on the size of the active site, and the structural and binding data for just one protease, HIV-1 protease, that there is support for the idea that helical and turn conformations are not recognized by this particular protease. For other protease active sites which are larger or more open and accessible to helices, turns, and sheets, these may well be accommodated in the active site though such examples are not presently known.

It is conceivable that molecules that are more closely related to the ultimate protease-binding shape, the extended conformation, still bind to the protease via a multistep templating process[35a] in which part of the extended ligand surface draws together the protease surfaces. Whether proteases in general initially select for an extended conformation remains to be confirmed. If it is true in general, then the rate of substrate processing should be influenced by conformational equilibria and disfavored by high populations of conformations that need to rearrange prior to protease binding. This was not possible to fully test here with the limited range of (poor) substrates available for the one protease examined but is the subject of a more detailed kinetic study for HIV-1 protease using a single substrate sequence locked into different highly stable conformations that do not readily interconvert (unpublished data).

Implications for Drug Design. Biological properties of peptides and proteins are frequently interpreted on the basis of their observed solution or solid-state structures. However the most abundant solution structure of a molecule is not necessarily the one recognized by an enzyme. The above results clearly demonstrate that inhibitors and substrate analogues at least end up bound in an extended conformation to the active site of a protease. There are profound consequences for drug design of binding to proteases by their substrates, substrate analogues, and inhibitors in extended β-strand conformations. First, it suggests the underutilized strategy of conformationally fixing protease *inhibitors* in an extended β-strand in order to optimize their potency and selectivity, an approach highlighted here by conformationally constrained macrocycles, but there are also other ways of stabilizing extended β-strands. Second, the design of protease inhibitors should be greatly simplified since new proteases, for which there is no or limited structural information, can still be effectively targeted based only upon knowledge of the class of protease and its substrate specificity. Third, proteolytic degradation of other bioactive peptides (agonists, antagonists, etc.) might be minimized by the reverse strategy of stabilizing their turn or helix conformations so that the extended strand conformation is less accessible. Conformational selection should now be examined for other proteases, and attempts should be made to see if there are correlations between conformational equilibria and processing rates of polypeptide substrates and conformationally biased analogues.

Conclusions

This work provides the most comprehensive analysis to date for the importance of inhibitor conformation in inhibitor−protease interactions. On the basis of superpositions of over 250 protease−inhibitor crystal structures, the first strong evidence has been pooled together to show that aspartic, serine, cysteine, and metallo proteases *commonly* bind to the extended β-strand conformation of their inhibitors, including substrate analogues and nonpeptidic inhibitors. Cyclic inhibitors and a cyclic substrate have been used to show that the basic strategy of conformationally constraining inhibitors/substrates to the protease-binding structure (preorganization) is an effective way of increasing affinity

for a protease. This suggests that general methods which can successfully preorganize molecules in the protease-binding conformation will similarly increase protease affinity, and some of these methods are already being described for specific proteases.[20b,21] Furthermore, for proteases where the structure is unknown, but substrate specificity and protease type are known, approaches to fixing an extended conformation for inhibitors may be useful for producing lead molecules that bind the protease tightly. Finally, some preliminary data is presented which is consistent with the idea that peptide conformations other than the extended strand may be more resistant to protease cleavage, suggesting that protein folding protects peptide amide bonds and thus that equilibria between peptide conformations could be important determinants of substrate activity. We are currently examining peptide substrates in more detail for kinetic and structural evidence of a conformational preference in substrate processing by proteases.

Experimental Section

All materials were reagent grade; amino acids were purchased from Novabiochem. Gradient HPLC was performed on Waters C-18 analytical (15 μm, 8 mm × 100 mm) and semipreparative (15 μm, 25 mm × 100 mm) columns. Analytical runs were 100% A for 2 min, then 0−60% B gradient over 60 min at 2 mL/min, where buffer A is 0.1% TFA in H_2O and buffer B is 0.1% TFA in 90% CH_3CN/H_2O. Mass spectra were obtained on a triple quadrupole mass spectrometer (PE SCIEX API III) equipped with an ionspray (pneumatically assisted electrospray) atmospheric pressure ionization source (ISMS). Solutions of compounds in 9:1 acetonitrile/0.1% aqueous trifluoroacetic acid were injected by syringe into the spectrometer. Molecular ions, $\{[M + nH]^{n+}\}/n$, were generated by the ion evaporation process and focused into the analyzer of the mass spectrometer through a 100-mm sampling orifice. Full scan data was acquired by scanning quadrupole-1 from *m/z* 100−900 with a scan step of 0.1 Da and a dwell time of 2 ms. Mass spectra were processed using MacSpec version 3.2 (PE-Sciex Instruments). Accurate mass determinations were performed on a KRATOS MS25 mass spectrometer using electron impact ionization.

Computer Superimpositions. Protease−inhibitor crystal structures were identified in the Protein Data Bank (http://www.rcsb.org/pdb) and the PDB codes used are listed in Table 1 (Supporting Information). These represented the majority of structures available at the time of analysis. The protease−inhibitor structures were displayed using the program InsightII 95.0 (Molecular Simulations Inc., San Diego, CA) and the α-carbon atoms of a selection of proteases were superimposed using the homology module and superimposition function. Only the inhibitor/substrate conformations are displayed in Figures 1−3. The reported rmsd values for Figures 1 and 3 only reflect structural deviations among the protein conformations. In the cases of multiple proteases bound to a common inhibitor (e.g. pepstatin A), only the inhibitor conformations were superimposed (e.g. Figure 2) and the rmsd values indicate the extent of conformational match for the same inhibitor bound to different proteases.

Synthesis. The synthesis of inhibitors **1**−**3** has been described previously.[22] The bicyclic substrate **7** was made by some modifications to this method involving solution-phase coupling of the N-terminal cyclic carboxylate **11** with the C-terminal cyclic amine **12** using BOP/DIPEA. It was characterized by NMR (^{1}H, ^{13}C) and MS data. Peptides **8** and **13**−**15** were synthesized by standard solid-phase synthesis protocols using Boc or Fmoc chemistry, purified by rpHPLC and characterized by MS and NMR spectra.

Peptides 8 and 14. Amino acids were Boc-protected and assembled by manual stepwise solid-phase peptide synthesis using HBTU activation and DIPEA in situ neutralization. This involved washing the resin in DMF, cleaving the N-terminal Boc protecting group by shaking in 100% TFA (2 × 1 min) and flow washing with DMF (1 min) to remove the TFA. Each coupling involved activating the amino acid (4 equiv) by mixing with 0.5 M HBTU (4 equiv) and DIPEA (5 equiv), and then adding this mixture to the resin (1 equiv) and shaking 10 min. Couplings were monitored by the quantitative ninhydrin test for remaining free amine. The resin was then flow washed for 1 min and the procedure of deprotection and coupling was repeated for the other amino acids. After the final amino acid was added the resin was flow washed with DMF and DCM and then dried under a flow of nitrogen gas. Peptide **14** was synthesized on MBHA resin (Peptide Institute) with substitution value of 0.57 mmol/g. Peptide **8** was synthesized synthesized on a Boc-Val-Pam resin (Peptide Institute) with substitution value of 0.70 mmol/g and then methylated (HCl/MeOH). Side-chain-protecting groups were *N*-tosyl for Arg, xanthyl (Xan, temporary protection) for Asn and Gln, benzyloxy (BzlO) for Asp and Glu, and *N*-2-chlorobenzyloxycarbonyl (ClZ) for Lys. Prior to cleavage of the peptide from the resin the N-terminal Boc protecting group was removed, the resin was flow washed with DMF and DCM and then dried under a flow of nitrogen. Cleavage of ∼300−400 mg resin was performed in an all Teflon apparatus specially designed for handling HF. Treatment with liquid HF (10 mL) and *p*-cresol (1 mL) at −5 °C for 1−2 h was followed by treatment with deoxygenated ether to precipitate the cleaved peptide, which was filtered and dissolved in 50% acetonitrile/water and then freeze-dried. Purification by rpHPLC used a gradient of 100% A (2 min), 100% A/0% B to 20% A/80% B (40 min), 20% A/80% B to 100% B (10 min), 100% B (5 min). Ac-LVFFIV-OMe (**8**): t_R = 56.0 min; ES-MS calcd 793.5, found 793.2. Suc-EQADKLAKEAA-RKVLFLNEAAAKAA-NH_2 (**14**): t_R = 24.5 min; ES-MS calcd 2755.5, found 2755.3.

Peptide 15. This was synthesized by Fmoc chemistry on Ramage resin (Novabiochem) with a substitution value of 0.55 mmol/g. Amino acids were assembled by manual stepwise solid-phase peptide synthesis using HBTU/DIPEA activation. The procedure involved washing the resin in DMF, cleavage of the N-terminal Fmoc protecting group by shaking in 50% piperidine/DMF (2 × 1 min) and flow washing with DMF (1 min). Each coupling involved activating the amino acid (4 equiv) by mixing with 0.5 M HBTU (4 equiv) and DIPEA (4 equiv), and then adding this mixture to the resin (1 equiv) and shaking 10 min. Couplings were monitored by the quantitative ninhydrin test. The resin was then flow washed for 1 min and the procedure of deprotection and coupling was repeated for the other amino acids. After the final amino acid was added the resin was flow washed with DMF and DCM and then dried under a flow of nitrogen gas. Side-chain-protecting groups were 2,2,5,7,8-pentamethylchroman-6-sulfonyl (Pmc) for Arg, trityl (Trt) for Asn and Gln, *tert*-butyloxycarbonyl (Boc) for Lys, and *tert*-butyl (tBu) for Ser and Tyr. The peptide was cleaved from resin (200−300 mg) by stirring with 3 mL of a solution of TFA:TIPS:H_2O (95:2.5:2.5) for 2−3 h for Arg(Pmc)-protected peptides and for 5−6 h for Arg(Mtr)-protected peptides. The TFA was removed on a rotary evaporator at 25 °C, ether was used to precipitate the cleaved peptide which was filtered and dissolved in 50% acetonitrile/water and then freeze-dried. Purification by rpHPLC used a gradient of 100% A (2 min), 100% A/0% B to 20% A/80% B (40 min). Ac-PSYKGRPGN-FLQSRP-NH_2 (**15**): t_R = 18.54 min; ES-MS calcd 1744.9, found 1744.8.

5-Bromopentanoyl-L-valinyl-L-tyrosine Benzyl Ester (10). A solution of L-valinyl-L-tyrosine benzyl ester *p*-toluenesulfonate (**9**; 10 mmol) in THF (30 mL) and water (30 mL) was stirred at room temperature while solutions of K_2CO_3 (2.83 g) in water (10 mL) and 5-bromopentanoyl chloride in THF (5 mL) were added over 10 min. After a further 30 min the solution was diluted with EtOAc and washed with 2 M HCl, 5% $NaHCO_3$ and brine and dried over $MgSO_4$. Removal of solvent gave a white powder **10** (5.06 g, 95%): R_f 0.55 (75% EtOAc/hexane); ^{13}C NMR ($CDCl_3$) δ 172.9, 171.1, 171.0, 155.4,

135.0, 130.4, 128.6, 126.7, 115.6, 67.4, 58.3, 53.3, 36.9, 35.4, 33.1, 32.0, 31.2, 24.2, 19.1, 18.2.

(9*S*,12*S*)-12-Carbobenzyloxy-7,10-dioxo-9-isopropyl-2-oxa-8,11-diazabicyclo[12.2.2]octadeca-14,16,17-triene (11). 5-Bromopentanoyl-L-valinyl-L-tyrosine benzyl ester (935 mg, 1.75 mmol) was converted to the corresponding iodide by refluxing with NaI (395 mg) in acetone (20 mL) for 2 h, then the precipitated NaCl was filtered off and the solvent was evaporated. The residual iodide was dissolved in dry DMF (50 mL) and anhydrous K_2CO_3 (500 mg) was added. The mixture was stirred at room temperature overnight then evaporated in vacuo. The residue was partitioned between water and 50% EtOAc/$CHCl_3$ and the organic layer was filtered to remove solid polymer, then washed with dilute sodium thiosulfate solution, 2 M HCl, and $NaHCO_3$, and dried over $MgSO_4$. Removal of solvent gave a solid which was purified by flash chromatography ($CHCl_3$−50% EtOAc/$CHCl_3$) giving the benzyl ester of **11** as a white powder (290 mg, 37%): R_f 0.28 (75% EtOAc/hexane); ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.45−7.35 (m, 5H, ArH), 7.21 (dd, J = 8.4, 2.2 Hz, 1H, ArH), 6.92−6.84 (m, 2H, ArH), 6.78 (dd, J = 8.3, 2.7 Hz, 1H, ArH), 5.83 (d, J = 9.9 Hz, 1H, tyr NH), 5.69 (d, J = 8.8 Hz, 1H, val NH), AB system: δ_A 5.26, δ_B 5.20 J_{AB} = 13.1 Hz, 2H, OCH_2Ph), 5.06 (ddd, J = 14.9, 10.0, 4.8 Hz, 1H, tyr αH), 4.29−4.12 (m, 2H, CH_2OAr), 3.95 (dd, J = 8.8, 7.1 Hz, 1H, val αH), 3.43 (dd, J = 13.6, 4.8 Hz, 1H, tyr βH), 2.53 (dd, J = 13.6, 12.1 Hz, 1H, tyr βH), 2.26−2.17 (m, 1H) and 2.09−1.98 (m, 1H, CH_2CO), 1.98−1.70 (m, 2H, CH_2), 1.83 (m, 1H, val βH), 1.55−1.31 (m, 2H, CH_2), 0.84 (d, J = 6.8 Hz, 6H, $(CH_3)_2$); ^{13}C NMR ($CDCl_3$) δ 172.07, 171.25, 170.02, 155.71, 135.04, 131.12, 129.99, 128.72, 128.69, 128.46, 128.34, 118.34, 116.52, 67.81, 67.49, 58.16, 52.48, 38.46, 36.12, 31.78, 25.98, 21.66, 18.77, 18.40; HRMS *m/e* 452.2312 (M^+), calcd for $C_{26}H_{32}N_2O_5$ 452.2311.

A solution of this benzyl ester of **11** (3.0 g, 6.6 mmol) in MeOH (75 mL) was hydrogenated over 10% Pd−C, 2 atm, room temperature for 3 h. The catalyst was filtered off and the solvent was removed in vacuo giving the carboxylic acid **11** as a white powder (2.4 g, 100%): ^{1}H NMR (300 MHz, CD_3OD) δ 8.35 (d, J = 9.9 Hz, 1H tyr NH), 7.47 (d, J = 9.3 Hz, 1H, Val-NH) 7.19 (dd, J = 8.5, 2.2 Hz, 1H, ArH), 7.00 (dd, J = 8.2, 2.2 Hz, 1H, ArH), 6.85 (dd, J = 8.4, 2.6 Hz, 1H, ArH), 6.76 (dd, J = 8.2, 2.6 Hz, 1H, ArH), 4.87−4.76 (m, 1H, tyr αH), 4.29−4.19 (m, 1H, CH_2OAr), 4.14−4.02 (m, 1H, CH_2OAr), 4.00 (d, J = 8.4 Hz, 1H, val αH (NH exchanged)), 3.35 (dd, J = 13.5, 4.3 Hz, 1H, tyr βH), 2.62 (dd; J = 13.5, 12.7 Hz, 1H, tyr βH), 2.20−2.07 (m, 2H, CH_2CO), 1.88−1.80 (m, 2H, val βH and CH_2), 1.60−1.25 (m, 3H, CH_2), 0.90 (d, J = 6.8 Hz, 3H, CH_3), 0.84 (d, J = 6.7 Hz, 3H, CH_3); ^{13}C NMR (CD_3OD) δ 174.93, 174.35, 172.48, 156.66, 132.56, 131.46, 130.56, 119.63, 119.40, 69.36, 59.54, 54.18, 38.50, 36.46, 32.89, 27.24, 22.84, 19.54, 19.04; HRMS *m/e* 362.1841, calcd for $C_{19}H_{26}N_2O_5$ 362.1842.

(8*S*,11*S*)-11-Amino-7,10-dioxo-8-(2-butyl)-2-oxa-6,9-diazabicyclo[11.2.2]heptadeca-13,15,16-triene (12). The *N*-Boc-protected form[22b] of macrocycle **12** (1.00 g, 2.31 mmol) was dissolved in TFA (10 mL), then after a further 5 min, the TFA was evaporated in vacuo. The residue was dissolved in water and purified by preparative reverse-phase HPLC (linear gradient, 100% water increasing MeCN at 1% per min + 0.1% TFA), room temperature 15 min. Lyophilization gave the macrocyclic amine **12** as a white powder (525 mg): ^{1}H NMR (300 MHz, CD_3OH) δ 7.79 (broad m, 1H, N*H*CH_2), 7.22 (d, J = 7.8 Hz, 1H, Ile-NH), 7.23 (dd, J = 8.4, 2.1 Hz, 1H, ArH), 6.97−6.79 (m, 3H, ArH), 4.41−4.31 (m, 1H, H-3), 4.28−4.16 (m, 1H, H-3), 4.08 (dd, J = 10.7, 7.0 Hz, 1H, Tyr-αCH), 3.60−3.45 (m, 1H, H-5), 3.47 (t, J = 7.6 Hz, 1H, Ile-αCH), 3.28 (dd, J = 12.4, 7.0 Hz, 1H, Tyr-βCH), 2.86−2.75 (m, 1H, H-5), 2.70 (dd, J = 12.4, 10.7 Hz, 1H, Tyr-βCH), 2.30−2.09 (m, 1H, H-4), 1.83−1.69 (m, 1H, H-4), 1.60−1.39 (m, 2H, Ile-βCH and Ile-γCH_2), 1.01−0.88 (m, 1H, Ile-γCH_2), 0.83 (t, J = 7.2 Hz, 3H, Ile-δCH_3), 0.75 (d, J = 6.8 Hz, 3H, Ile-γCH_3); ^{13}C NMR (CD_3-OH) δ 171.5, 168.6, 160.0, 132.2, 130.4, 128.2, 118.8, 118.7, 68.6, 59.8, 56.0, 40.0, 38.0, 37.5, 27.6, 26.2, 14.9, 11.7; IS-MS *m/z* 334 (MH^+).

Bicyclic Substrate 7. A solution of the macrocyclic acid **11** (155 mg, 0.428 mmol), the macrocyclic amine **12** (109 mg, 0.244 mmol) and BOP reagent (190 mg, 0.43 mmol) was stirred with THF (2 mL) at 0 °C, then diisopropylethylamine (160 μL) was added. The mixture was stirred at 0 °C for 60 min then concentrated in vacuo. The residue was dissloved in the minimum volume of warm DMF then diluted with about 2 volumes of 50% MeCN/H_2O and immediately purified by reverse-phase HPLC before the product aggregated from solution (linear gradient 30% MeCN/70% H_2O/0.1% TFA to 90% MeCN/10% H_2O/0.1% TFA over 20 min, retention time 17 min). Lyophilization gave a white powder **7** (91 mg, 55%): ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.09 (d, J = 9.8 Hz, 1H, Phe-NH), 7.90 (d, J = 7.4 Hz, 1H, Phe-NH), 7.33 (d, J = 9.3 Hz, 1H, Val-NH), 7.28 (m, 1H, CH_2NH), 7.18−7.13 (m, 2H, Ar−H), 7.01 (m, 2H, Ile-NH and Ar−H), 6.84 (m, 2H, Ar−H), 6.81−6.74 (m, 2H, Ar−H), 6.67 (dd, J = 8.2, 2.6 Hz, 1H, Ar−H), 4.58 (m, 1H, Phe-αH), 4.53 (m, 1H, Phe-αH), 4.30 (m, 1H, H-3′), 4.11 (m, 1H, H-3′), 4.09 (m, 1H, H-3), 4.04 (m, 1H, H-3), 3.99 (t, J = 8.5 Hz, 1H, Val-αCH), 3.49 (t, J = 7.9 Hz, 1H, Ile-αCH), 3.30 (m, 1H, H-5), 3.10 (m, 1H, Phe-βCH_2), 3.08 (m, 1H, Phe-βCH_2), 2.66 (m, 1H, H-5′), 2.48 (m, 1H, Phe-βCH_2), 2.43 (m, 1H, Phe-βCH_2), 2.13 (m, 1H, H-6), 1.98 (m, 1H, H-4′), 1.83 (m, 1H, H-6), 1.68 (m, 1H, H-4′),1.66 (m, 1H, Val-βCH), 1.54 (m, 1H, H-5), 1.46 (m, 1H, H-4), 1.41 (m, 1H, Ile-βCH), 1.30 (m, 1H, H-5), 1.28 (m, 1H, Val-γCH_2), 1.17 (m, 1H, H-4), 0.83 (m, 1H, Ile-γCH_2), 0.75 (d, J = 6.8 Hz, Ile-γCH_2), 0.76−0.70 (m, 9H, Val-γCH_3 and Ile-γCH_3), 0.63 (d, J = 6.8 Hz, 3H, Ile-δCH_3); ^{13}C NMR (DMSO-d_6) δ 171.3, 170.7, 170.3, 169.8, 169.8, 157.6, 155.1, 131.4, 131.1, 130.5, 129.6, 129.5, 129.3, 118.5, 118.4, 118.0, 117.5, 68.2, 68.1, 57.2, 57.1, 55.2, 54.1, 38.3, 38.2, 37.7, 35.7, 35.0, 31.4, 26.7, 25.9, 24.6, 21.7, 19.3, 18.8, 15.0, 11.5; HRMS *m/z* 700.3666 (MNa^+), calcd for $C_{37}H_{51}N_5O_7Na$ 700.3686.

Enzyme Inhibition and Processing. HIV-1 protease assays[36] were carried out using synthetic [Aba 67,95,167,195]-HIV-1 protease (SF2 isolate) with Cys residues replaced by α-aminobutyric acid (Aba). K_i values were calculated from IC_{50} values determined for substrates **7**, **8**, and **13**−**15** or inhibitors **1**−**3** incubated at pH 6.5, I = 0.1 M, 37 °C, 50 mM substrate 2-Abz-Thr-Ile-Nle-Phe(p-NO_2)-Gln-Arg-NH_2 (Abz-NF*-6) using a continuous fluorimetric assay[37] on a Perkin-Elmer LS50B luminescence spectrometer and assuming competitive inhibition. K_m values were determined at pH 5.5, I = 0.1 M and 37 °C using fixed point assays for substrates incubated over 6−24 h and the products were identified and analyzed using rpHPLC and electrospray mass spectrometry.

CD Spectra. CD data were recorded on a JASCO 710 spectropolarimeter. Peptides were dissolved in 10 mM phosphate buffer (pH 7) and measurements were made at 20 °C in a 3-mL cuvette with a path length of 1 cm, scanning from 250 to 190 nm every 0.10 nm, with bandwidth of 1 nm. Wavelength scans were corrected for buffer scans taken at 20 °C. Peptide concentrations were determined by amino acid analysis. Percent helicity was calculated[38] based on a predicted mean residue ellipticity ([ø]Hn) using [ø]Hn = [ø]H∞ (1−2.5/n), where [ø]H∞ = −37400 deg·cm^2·dmol^{-1} at a wavelength of 222 nm for a helix of infinite length, n is the number of residues in the helix, and the value of 2.5 is a wavelength-dependent constant for 222 nm.

NMR Spectroscopy. Proton NMR spectra were recorded for peptides (1−3 mM; 90% H_2O/10% ^2H_2O) on a Bruker ARX-500 at 280 K as described,[39] processed using XWIN NMR (Bruker), and analyzed using AURELIA (Bruker). Resonances were assigned using standard sequential procedures. The pattern of NOEs was used to identify secondary structure. Sequence specific assignments for **14** were based on sequential NOE connectivities (Figure 6a). Despite overlap in the spectra, medium range NOEs ($d\alpha N(i,i+2)$, $d\alpha N(i,i+3)$, $d\alpha N(i,i+4)$, $d\alpha\beta(i,i+3)$) were observed for a number of residues, consistent with a helical conformation. For the majority of residues where these NOEs have not been identified, the presence of other overlapping cross-peaks in the spectra prevents their observation. Also consistent with a helical conformation for the

peptide, a large number of residues have coupling constants $^3J_{HNH\alpha} < 5$ Hz and the Hα chemical shifts of the majority of the residues are shifted significantly upfield compared with their random coil values. The pattern of NOE connectivity observed, $^3J_{HNH\alpha}$ coupling constants, and chemical shift indices (Figure 6a) all strongly suggest that the peptide is helical along its entire length. For peptide **15**, sequence-specific assignments were made from sequential NOE connectivities (Figure 6b). Strong sequential $d\alpha N(i,i+1)$ connectivities and coupling constants $^3J_{HNH\alpha} = 5-8$ Hz suggest that most residues are in an extended conformation, but a small number of medium range NOEs are observed consistent with turn conformations at residues P_1SYK_4 and R_6PGN_9.

Acknowledgment. We thank the Australian National Health and Medical Research Council and the Australian Research Council for some financial support.

Supporting Information Available: Table 1 containing PDB codes for 266 protease−inhibitor structures, as mentioned in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Leung, D.; Abbenante, G.; Fairlie, D. P. Protease Inhibitors: Current Status and Future Prospects. *J. Med. Chem.* **2000**, *43*, 305−41. (b) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein−Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97*, 1359−472. (c) Craik, C. S.; Debouck, C. Proteases as Therapeutic Targets. In *Perspectives in Drug Discovery and Design*; McKerrow, J. H., James, M. N. G., Eds.; ESCOM: Leiden, 1995; Vol. 2, pp 1−125. (d) Van Noorden, J. F. IBC's 2nd International Conference on Protease Inhibitors: Novel Therapeutic Applications and Development, Washington D.C., Feb 24−26, 1997; *Acta Histochem.* **1997**, *99*, 249−55.

(2) (a) Beckett, R. P.; Davidson, A. H.; Drummond, A. H.; Huxley, P.; Whittaker, M. Recent Advances in matrix metalloprotease inhibitor research. *DDT* **1996**, *1*, 16. (b) Johnson, L. L.; Dyer, R.; Hupe, D. J. Matrix metalloproteinases. *Curr. Opin. Chem. Biol.* **1998**, *2*, 466−71. (c) Yan, S.; Sameni, M.; Sloane, B. F. Cathepsin B and human tumor progression. *J. Biol Chem.* **1998**, *379*, 113−23.

(3) (a) Becker, M. M.; Harrop, S. A.; Dalton, J. P.; Kalinna, B. H.; McManus, D. P.; Brindley, P. J. Cloning and characterization of the *Schistosoma japonicum* aspartic proteinase involved in hemoglobin degradation. *J. Biol. Chem.* **1995**, *270*, 24496−501; erratum: *J. Biol. Chem.* **1997**, *272*, 17246. (b) Brindley, P. J.; Kalinna, B. H.; Dalton, J. P.; Day, S. R.; Wong, J. W.; Smythe, M. L.; McManus, D. P. Proteolytic degradation of host hemoglobin by schistosomes. *Mol. Biochem. Parasitol.* **1997**, *89*, 1−9.

(4) (a) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Maier, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from Plasmodium falciparum. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034−9. (b) Li, Z.; Chen, X.; Davidson, E.; Zwang, O.; Mendis, C.; Ring, C. S.; Roush, W. R.; Fegley, G.; Li, R.; Rosenthal, P. J.; et al. Anti-malarial drug development using models of enzyme structure. *Chem. Biol.* **1994**, *1*, 31−7.

(5) (a) Abad-Zapatero, C.; Goldman, R. W.; Muchmore, S. W.; Hutchins, C.; Stewart, K.; Navaza, J.; Payne, C. D.; Ray, T. L. Structure of a secreted aspartic protease from C. Albicans complexed with a potent inhibitor: implications for the design of antifungal agents. *Protein Sci.* **1996**, *5*, 640−52. (b) Abad-Zapatero, C.; Goldman, R.; Muchmore, S. W.; Hutchins, C.; Oie, T.; Stewart, K.; Cutfield, S. M.; Cutfield, J. F.; Foundling, S. I.; Ray, T. L. Structure of secreted aspartic proteinases from Candida. Implications for the design of antifungal agents. *Adv. Exp. Med. Biol.* **1998**, *436*, 297−313.

(6) (a) Wlodawer, A.; Erickson, J. W. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **1993** *62*, 543−85. (b) Darke, P. L.; Huff, J. R. HIV protease as an inhibitor target for the treatment of AIDS. *Adv. Pharmacol.* **1994**, *25*, 399−454. (c) West, M. L.; Fairlie, D. P. Targeting HIV-1 protease: a test of drug-design methodologies. *Trends Pharmacol. Sci.* **1995**, *16*, 67−75. (d) Kempf, D. J.; Sham, H. L. HIV protease inhibitors. *Curr. Pharm. Des.* **1996**, *2*, 225−46. (e) March, D. R.; Fairlie, D. P. *Designing new antiviral drugs for AIDS: HIV-1 Protease and its inhibitors*; R. G. Landes Co.: Austin, TX, 1996.

(7) (a) Kim, J. L.; Morgenstern, K. A.; Lin, C.; Fox, T.; Dwyer, M. D.; Landro, J. A.; Chambers, S. P.; Markland, W.; Lepre, C. A.; O'Malley, E. T.; Harbeson, S. L.; Rice, C. M.; Murcko, M. A.; Caron, P. R.; Thomson, J. A. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **1996**, *87*, 343−55; erratum: *Cell* **1997,** *8*, 159. (b) Love, R. A.; Parge, H. E.; Wickersham, J. A.; Hostomsky, Z.; Habuka, N.; Moomaw, E. W.; Adachi, T.; Hostomska, Z. The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **1996**, *87*, 331−42.

(8) (a) Gibson, W.; Hall, M. R. Assemblin, an essential herpesvirus proteinase. *Drug Des. Discov.* **1997**, *15*, 39−47. (b) Shieh, H. S.; Kurumbail, R. G.; Stevens, A. M.; Stegeman, R. A.; Sturman, E. J.; Pak, J. Y.; Wittwer, A. J.; Palmier, M. O.; Wiegand, R. C.; Holwerda, B. C.; Stallings, W. C. Three-dimensional structure of human cytomegalovirus protease. *Nature* **1996**, *383*, 279−82; erratum: *Nature* **1996**, *384*, 288.

(9) Boycott, R.; Klenk, H. D.; Ohuchi, M. Cell tropism of influenza virus mediated by hemagglutinin activation at the stage of virus entry. *Virology* **1994**, *203*, 313−9.

(10) (a) Bernstein, P. R.; Edwards, P. D.; Williams, J. C. Inhibitors of human leukocyte elastase. *Prog. Med. Chem.* **1994**, *31*, 59−120. (b) Hugli, T. E. Protease inhibitors: novel therapeutic application and development. *Trends Biotechnol.* **1996**, *14*, 409−12. (c) Fath, M. A.; Wu, X.; Hileman, R. E.; Linhardt, R. J.; Kashem, M. A.; Nelson, R. M.; Wright, C. D.; Abraham, W. M. Interaction of secretory leukocyte protease inhibitor with heparin inhibits proteases involved in asthma. *J. Biol. Chem.* **1998**, *273*, 13563−9. (d) Tanaka, R. D.; Clark, J. M.; Warne, R. L.; Abraham, W. M.; Moore, W. R. Mast cell tryptase: a new target for therapeutic intervention in asthma. *Int. Arch Allergy Immunol.* **1995**, *107*, 408−9.

(11) Stubbs, M. T.; Bode, W. A player of many parts: the spotlight falls on thrombin's structure. *Thromb Res.* **1993**, *69*, 1−58.

(12) (a) Selkoe, D. J. Alzheimer's disease: genotypes, phenotypes, and treatments. *Science* **1997**, *275*, 630−1. Vassar, R.; Bennett, B. D.; Babu-Khan, S.; Khan, S.; et al. Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **1999**, *286*, 735−741.

(13) Seife, C. Blunting nature's Swiss army knife. *Science* **1997**, *277*, 1602−3.

(14) (a) See references to all PDB-filed structures in Table 1. (b) Rich, D. H. Peptidase Inhibitors. In *Comprehensive Medicinal Chemistry;* Samuels, P. J., Ed.; Pergamon Press: Oxford, 1990; Vol. 2, pp 391−441 and references therein.

(15) (a) James, M. N. G.; Sielecki, A. R. Aspartic proteases and their Catalytic Pathway inhibitors. In *Biological Macromolecules and Assemblies*: *Active sites of enzymes;* Jurnak, F. A., McPherson, A., Eds.; John Wiley & Sons: New York, 1987; p 413. (b) Bailey, D.; Cooper, J. B. A structural comparison of 21 inhibitor complexes of the aspartic proteinase from Endothia parasitica. *Protein Sci.* **1994**, *3*, 2129−43.

(16) Powers, J. C.; Harper, J. W. Inhibitors of serine proteases. In *Protease Inhibitor*s; Barret, A. J. S., Salverson, G., Eds.; Elsevier: Amsterdam, 1986; pp 55−152.

(17) Powers, J. C.; Harper, J. W. Inhibitors of metalloproteases. In *Protease Inhibitors*; Barret, A. J. S., Salverson, G., Eds.; Elsevier: Amsterdam, 1986; pp 219−98.

(18) (a) Hans-Hartwig, O.; Schirmeister, T. Cysteine proteases and their inhibitors. *Chem. Rev.* **1997**, *97*, 133−71. (b) Rasnick, D. Small synthetic inhibitors of cysteine protease. In *Perspectives in Drug Discovery and Design*; McKerrow, J. H., James, M. N. G., Eds.; ESCOM: Leiden, 1996; pp 47−63.

(19) Dauter, Z.; Betzel, C.; Genov, N.; Pipon, N.; Wilson, K. S. Complex between the subtilisin from a mesophilic bacterium and the leech inhibitor eglin-C. *Acta Crystallogr. B* **1991**, *47*, 707−30.

(20) (a) Fairlie, D. P.; Abbenante, G.; March, D. Macrocyclic Peptidomimetics: Forcing Peptides into Bioactive Conformations. *Curr. Med. Chem.* **1995**, *2*, 672−705. (b) Fairlie, D. P.; West, M. L.; Wong, A. K. Towards protein surface mimetics. *Curr. Med. Chem.* **1998**, *5*, 29−62. (c) Reid, R. C.; Fairlie, D. P. Mimicking extended conformations of protease substrates: designing cyclic peptidomimetics to inhibit HIV-1 Protease. *Advances in Amino Acid Mimetics and peptidomimetics*; JAI Press Inc.: Greenwich, CT, 1997; pp 77−107.

(21) (a) Smith III, A. B.; Keenan, T. P.; Holcomb, R. C.; Sprengeler, P. A.; Guzman, M. C.; Wood, J. L.; Carol, P. J.; Hirschman, R. Design, synthesis and crystal structure of a pyrrolinone-based peptidomimetic possessing the conformation of a *β*-strand: Potential application to the design of novel inhibitors of proteolytic enzymes. *J. Am. Chem. Soc.* **1992**, *114*, 10672−74. (b) Smith III, A. B.; Hirschmann, R.; Pasternack, A.; Akaishi, R.; Guzman, M. C.; Jones, D. R.; Keenan, T. P.; Sprengeler, P. A.; Darke, P. L.; Emini, E. A.; Holloway, M. K.; Schleif, W. A. Design and synthesis of peptidomimetic inhibitors of HIV-1 proteases and renin. Evidence for improved transport. *J. Med. Chem.* **1994**, *37*, 215−8. (c) Szewczuk, Z.; Rebholz, K. L.; Rich, D. H. Synthesis and biological activity of new conformationally restricted analogues of pepstatin. *Int. J. Pept. Protein Res.* **1992**, *40*, 233−42.

(22) (a) Abbenante, G.; March, D.; Bergman, D.; Hunt, P. A.; Garnham, B.; Dancer, R. J.; Martin, J. L.; Fairlie, D. P. Regioselective Structural and Functional Mimicry Of Peptides: Design Of Hydrolytically Stable Cyclic Peptidomimetic Inhibitors Of Hiv 1 Protease. *J. Am. Chem. Soc.* **1995**, *117*, 10220−6. (b) March, D.; Abbenante, G.; Bergman, D.; Brinkworth, R. I.; Wickramasinghe, W.; Begun, J.; Martin, J. L.; Fairlie, D. P. Substrate Based Cyclic Peptidomimetics Of Phe Ile Val That Inhibit Hiv 1 Protease Using a Novel Enzyme Binding Mode. *J. Am. Chem. Soc.* **1996**, *118*, 3375−9. (c) Reid, R. C.; March, D. R.; Dooley, M. J.; Bergman, D. A.; Abbenante, G.; Fairlie, D. P. A novel bicyclic enzyme inhibitor as a consensus peptidomimetic for the receptor-bound conformations of 12 peptidic inhibitors of HIV-1 protease. *J. Am. Chem. Soc.* **1996**, *118*, 8511−7.

(23) Martin, J. L.; Begun, J.; Schindeler, A.; Wickramasinghe, W. A.; Alewood, D.; Alewood, P. F.; Bergman, D. A.; Brinkworth, R. I.; Abbenante, G.; March, D.; Reid, R. C.; Hunt, P. A.; Fairlie, D. P. Molecular Recognition of Macrocyclic Peptidomimetic inhibitors by HIV-1 Protease. *Biochemistry* **1999**, *38*, 7978−88.

(24) Abbenante, G.; Bergman, D. A.; March, D. R.; Reid, R. C.; Hunt, P. A.; James, I. W.; Dancer, R. J.; Garnham, B.; Stoermer, M. L.; Fairlie, D. P. Structure−activity relationships for macrocyclic peptidomimetic inhibitors of HIV-1 protease. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2531−6.

(25) Suguna, K.; Padlan, E. A.; Bott, R.; Boger, J.; Parris, K. D.; Davies, D. R. Structures of complexes of rhizopuspepsin with pepstatin and other statine-containing inhibitors. *Proteins* **1992**, *13*, 195−205.

(26) Lee, A. Y.; Smitka, T. A.; Bonjouklian, R.; Clardy, J. Atomic structure of the trypsin-a90720a complex: a unified approach to structure and function. *Chem. Biol.* **1994**, *1*, 113.

(27) Lee, A. Y.; Hagihara, M.; Karmacharya, R.; W.albers, M. W.; Schreiber, S. L.; Clardy, J. Atomic structure of the Trypsin-Cyclotheonamide a complex: lessons for the design of serine protease inhibitors. *J. Am. Chem. Soc.* **1993**, *115*, 12619.

(28) Browner, M. F.; Smith, W. W.; Castelhano, A. L. Matrilysin-inhibitor complexes: common themes among metalloproteases. *Biochemistry* **1995**, *34*, 6602−10.

(29) Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; James, M. N. Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* **1998**, *37*, 16839−45.

(30) Ding, J. H.; Fraser, M. E.; Meyer, J. H.; Bartlett, P. A.; James, M. N. G. Macrocyclic inhibitors of penicillopepsin. 2. X-ray crystallographic analyses of penicillopepsin complexed with a P3−P1 macrocyclic peptidyl inhibitor and with its two acyclic analogues. *J. Am. Chem. Soc.* **1998**, *120*, 4610−21.

(31) (a) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−637. (b) Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **1981**, *34*, 167−339.

(32) (a) Ksander, G. M.; De Jesus, R.; Yuan, A.; Ghai, R. D.; McMartin, C.; Bohacek, R. Meta-substituted benzofused macrocyclic lactams as zinc metalloprotease inhibitors. *J. Med. Chem.* **1997**, *40*, 506−14. (b) Janetka, J. W.; Raman, P.; Satyshur, K.; Flentka, G.; Rich, D. H. Novel cyclic biphenyl ether β-strand mimetics and HIVProtease Inhibitors. *J. Am. Chem. Soc.* **1997**, *119*, 441−2. (c) Xue, C.-B.; He, X.; Roderick, J.; DeGrado, W. F.; Cherney, R. J.; Hardman, K. D.; Nelson, D. J.; Copeland, R. A.; Jaffe, B. D.; Decicco, C. P. Design and sunthesis of cyclic inhibitors of matrix metalloproteinases and TNF-α production. *J. Med. Chem.* **1998**, *41*, 1745−8. (d) Cherney, R. J.; Wang, L.; Meyer, D. T.; Xue, C.-B.; Wasserman, R.; Hardman, K. D.; Welch, P. K.; Covington, M. B.; Copeland, R. A.; Arner, E. C.; DeGrado, W. F.; Decicco, C. P. Macrocyclic amino carboxylates as selective MMP-8 inhibitors. *J. Med. Chem.* **1998**, *41*, 1749−51.

(33) Ptitsyn, O. B.; Finkelstein, A. V. Theory of protein secondary structure and algorithm of its prediction. *Biopolymers* **1983**, *22*, 15−25.

(34) (a) Poorman, R. A.; Tomasselli, A. G.; Heinrikson, R. L.; Kezdy, F. J. A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate database. *J. Biol. Chem.* **1991**, *266*, 14554−61. (b) Tomasseli, A. G.; Heinrikson, R. L. Specificity of retroviral proteases: an analysis of viral and nonviral protein substrates. *Methods Enzymol.* **1994**, *241*, 279−301.

(35) (a) Burgen, A. S. V.; Roberts, G. C. K.; Feeney, J. Binding of flexible molecules to macromolecules. *Nature* **1975**, *253*, 753−5. (b) Jencks, W. P. *Catal. Chem. Enzymol.* **1987**, 717−9.

(36) Bergman, D. A.; Alewood, D.; Alewood, P. F.; Andrews, J. L.; Brinkworth, R. I.; Englebretsen, D. R.; K. S. B. H. Kinetic-properties of HIV-1 protease produced by total chemical synthesis with cysteine residues replaced by isosteric l-alpha-amino-N-butyric acid. *Lett. Pept. Sci.* **1995**, *2*, 99−107.

(37) Toth, M. V.; Marshall, G. R. A simple, continuous fluorometric assay for HIV protease. *Int. J. Pept. Protein Res.* **1990**, *36*, 544−50.

(38) Chen, Y. H.; Yang, J. T.; Chau, K. H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **1974**, *13*, 3350−9.

(39) Scanlon, M. J.; Fairlie, D. P.; Craik, D. J.; Englebretsen, D. R.; West, M. L. NMR solution structure of the RNA-binding peptide from human immunodeficiency virus (type 1) Rev. *Biochemistry* **1995**, *34*, 8242−9.

JM990315T