

Prime Site Binding Inhibitors of a Serine Protease: NS3/4A of Hepatitis C Virus

Paolo Ingallinella, Daniela Fattori,[‡] Sergio Altamura, Christian Steinkühler, Uwe Koch, Daniel Cicero, Renzo Bazzo, Riccardo Cortese, Elisabetta Bianchi, and Antonello Pessi*

IRBM P. Angeletti, Via Pontina Km 30.600, 00040 Pomezia (Rome), Italy

Received January 29, 2002; Revised Manuscript Received March 4, 2002

ABSTRACT: Serine proteases are the most studied class of proteolytic enzymes and a primary target for drug discovery. Despite the large number of inhibitors developed so far, very few make contact with the prime site of the enzyme, which constitutes an almost untapped opportunity for drug design. In the course of our studies on the serine protease NS3/4A of hepatitis C virus (HCV), we found that this enzyme is an excellent example of both the opportunities and the challenges of such design. We had previously reported on two classes of peptide inhibitors of the enzyme: (a) product inhibitors, which include the P₆–P₁ region of the substrate and derive much of their binding energy from binding of their C-terminal carboxylate in the active site, and (b) decapeptide inhibitors, which span the S₆–S₄' subsites of the enzyme, whose P₂'–P₄' tripeptide fragment crucially contributes to potency. Here we report on further work, which combined the key binding elements of the two series and led to the development of inhibitors binding *exclusively* to the prime site of NS3/4A. We prepared a small combinatorial library of tripeptides, capped with a variety of constrained and unconstrained diacids. The SAR was derived from multiple analogues of the initial micromolar lead. Binding of the inhibitor(s) to the enzyme was further characterized by circular dichroism, site-directed mutagenesis, a probe displacement assay, and NMR to unequivocally prove that, according to our design, the bound inhibitor(s) occupies (occupy) the S' subsite and the active site of the protease. In addition, on the basis of the information collected, the tripeptide series was evolved toward reduced peptide character, reduced molecular weight, and higher potency. Beyond their interest as HCV antivirals, these compounds represent the first example of prime site inhibitors of a serine protease. We further suggest that the design of an inhibitor with an analogous binding mode may be possible for other serine proteases.

Serine proteases are perhaps the most studied class of proteolytic enzymes and a primary target for drug discovery. Despite the large number of inhibitors developed so far, very few make contact with the prime site of the enzyme.¹ Given that the starting point for inhibitor design is generally the substrate, this is not surprising, since for this class of enzymes the key determinants of substrate specificity reside in the P-region (1, 2–5). Therefore, developing inhibitors of serine proteases, which specifically target the prime site, is an almost untapped opportunity for drug design. In the course of our studies on the serine protease NS3/4A of hepatitis C virus (HCV), we found that this enzyme is an excellent example of both the opportunities and the challenges of such design.

The target of our work is a key virally encoded enzyme, whose inhibition holds much promise for an effective anti-

HCV therapy. Infection by hepatitis C virus (HCV) is widely recognized today as a huge public health concern, with more than 170 million people infected worldwide, most of them unknowingly. HCV infection is associated with a slowly progressing liver disease that over 2 decades can lead to liver cirrhosis or liver cancer. Unfortunately, neither a generally effective treatment nor a preventive vaccine is available (6, 7). The chymotrypsin-like serine proteinase domain, which is strictly required for polyprotein maturation, is contained within the N-terminal 180 amino acids of the NS3 protein (8–13). Although NS3 has proteolytic activity of its own, interaction with a second viral protein, NS4A, is essential for efficient processing of all the NS3-dependent polyprotein cleavage sites (8, 9, 14), and it is now common to refer to this proteinase as the NS3/4A protease.

As for serine proteases in general, for NS3/4A the main determinants of ground-state substrate binding (reflected in the K_m) reside in the P region (2–5). Accordingly, we (2, 15) as well as others (16, 17) were able to develop potent peptide inhibitors derived from the nonprime region of the substrate. By contrast, the prime region of the substrate, while being important for catalysis, contributes little to ground-state binding (3–5). This notwithstanding, inspection of the NS3/4A structure (18–22) shows binding pockets in the S' region, which can be exploited for active site directed inhibition. We reasoned that a ligand, taking advantage of S' binding, could display a range of interactions with the enzyme different from the ones used by the substrate and

* Address correspondence to this author at the Biopolymers Laboratory, Department of Molecular & Cell Biology, IRBM P. Angeletti, Via Pontina Km 30.600, 00040 Pomezia (Rome), Italy. Telephone: +39-06-91093445. Fax: +39-06-91093482. E-mail: antonello_pessi@merck.com.

[‡] Present address: Department of Chemistry, Menarini Ricerche Sud, Via Tito Speri 10, 00040 Pomezia (Rome), Italy.

¹ We follow the nomenclature of Schechter and Berger [Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* 27, 157–162] in designating the cleavage sites as P₆–P₅–P₄–P₃–P₂–P₁...P₁'–P₂'–P₃'–P₄' etc., with the scissile bond between P₁ and P₁' and the C-terminus of the substrate on the prime site. The binding sites on the enzyme corresponding to residues P₆–P₅–P₄–P₃–P₂–P₁...P₁'–P₂'–P₃'–P₄' are indicated as S₆–S₅–S₄–S₃–S₂–S₁...S₁'–S₂'–S₃'–S₄' etc.

represent an entry point for a novel class of NS3/4A inhibitors. Following on this, we optimized prime site binding in the context of noncleavable decapeptides spanning P₆–P₄' and were able to develop subnanomolar inhibitors (23). In addition to a previously optimized P₆–P₁ sequence (2), these inhibitors include a P₂'–P₄' tripeptide fragment, which contributes >2000-fold gain in potency. While the binding mode of the nonprime fragment was found to be similar to the natural substrates, novel interactions were indeed present in the prime site of the decapeptide. This was most clearly demonstrated by the finding that a substrate, whose prime site derived from the optimized sequence, was cleaved by the protease with much *reduced* turnover (23).

While proving that prime site binding can be a component of a high-affinity interaction, our inhibitors did not bind to the prime site *in the absence* of contacts with the nonprime region of the enzyme. In this respect, they were thus similar to previously reported examples of serine protease inhibitors spanning both enzyme subsites (4, 24–31). Here we report on further work, which led to the development of peptidic and nonpeptidic inhibitors that bind *exclusively* to the prime site of NS3/4A. Beyond their interest as antivirals, these compounds represent the first such case for a serine protease.

MATERIALS AND METHODS

Enzyme Preparations and Site-Directed Mutagenesis. *Escherichia coli* BL21(DE3) cells were transformed with plasmids containing the cDNA coding for the serine protease domain of the HCV J strain NS3 protein (amino acids 1027–1206, followed by the sequence ASKSKK) under the control of the bacteriophage T7 gene 10 promoter in a pT7-7 vector. Point mutations were introduced by PCR-directed mutagenesis using appropriate mutagenic oligonucleotides. Clones were fully sequenced on both strands to exclude the introduction of additional mutations by PCR. All enzymes were purified as previously described (15) and found to be >95% pure as judged from reversed-phase HPLC performed using a 4.6 × 250 mm Vydac C4 column. Enzyme preparations were routinely checked by mass spectrometry done on HPLC-purified samples using a Perkin-Elmer API 100 instrument and by N-terminal sequence analysis carried out using Edman degradation on an Applied Biosystems model 470A gas-phase sequencer. Enzyme stocks were quantified by amino acid analysis, shock-frozen in liquid nitrogen, and kept in aliquots at –80 °C until use.

Protease Activity Assays. The HPLC assay was performed as previously described (2, 15). Briefly, we used 0.2–1 nM NS3 protease in 50 mM HEPES,² pH 7.5, 1 mM DTT, 15% glycerol, 150 mM NaCl, and 1% CHAPS, containing 80 μM Pep4AK. Pep4AK (KKKGSVVIVGRILSGR-NH₂) is the protease cofactor, a peptide spanning the central hydrophobic core of the NS4A protein, with a three-lysine tag at the N-terminus to increase solubility (32). Reactions were started by addition of 3 μL of substrate peptide, whose DMSO stock solution concentration was determined by quantitative amino

acid analysis on HCl-hydrolyzed samples. Six duplicate data points at different substrate concentrations were used to calculate kinetic parameters. Incubation times were chosen in order to obtain <7% substrate conversion, and reactions were stopped by addition of 40 μL of 1% TFA. Kinetic parameters were calculated from nonlinear least-squares fit of initial rates as a function of substrate concentration with the help of Kaleidagraph software, assuming Michaelis–Menten kinetics.

K_i values of peptide inhibitors were calculated from substrate titration experiments performed in the presence of increasing amounts of inhibitor. Experimental data sets were simultaneously fitted using a multicurve fit macro (SigmaPlot software) with the equation $V = (V_{\max}S)/(K_m(1 + K_i/I) + S)$. Alternatively, K_i values were derived from IC₅₀ values, calculated using a four-parameter logistic function, according to the equation $IC_{50} = (1 + S/K_m)K_i$. Adaptation of the assay to the microplate is reported in detail in ref 33.

Circular Dichroism Spectroscopy. Circular dichroism (CD) measurements were performed using a Jasco J-710 spectropolarimeter equipped with a cell holder thermostatically controlled by a circulating water bath. Measurements were recorded at 15 °C, with an 8 s time constant and 5 nm/min and averaged for two acquisitions. Protein concentration was 1.2 mg/mL, and spectra were collected with rectangular quartz cells of 1 cm path length in the near-UV region (320–250 nm). The concentration of both protein and inhibitor stock solutions was determined by quantitative amino acid analysis. Results are reported as mean residue ellipticity [Θ] having units of deg cm² dmol⁻¹ and were calculated using a mean residue mass of 106 Da for NS3. Titration experiments of NS3/Pep4AK and the inhibitor were performed in sodium phosphate buffer, 50 mM, pH 7.5, 15% glycerol, 2% CHAPS, and 3 mM DTT by incubating the protein with a saturating amount of Pep4AK and increasing amounts of inhibitor for 10 min at 15 °C. Spectra were recorded before and after the addition of Pep4AK and inhibitor and routinely corrected for background signal and dilution effects. The titration curve was followed by monitoring ellipticity at 288 nm as a function of inhibitor concentration. For each point of the titration curve the observed ellipticity Θ can be expressed by the equation:

$$\Theta = \alpha\Theta_{\text{bound}} + (1 - \alpha)\Theta_{\text{free}}$$

where α is the percentage of bound enzyme, Θ_{bound} is the ellipticity of the complexed enzyme, and Θ_{free} is the ellipticity of the uncomplexed enzyme. From this equation it is possible to derive the percentage of bound enzyme:

$$\alpha = (\Theta - \Theta_{\text{free}})/(\Theta_{\text{bound}} - \Theta_{\text{free}})$$

The bound inhibitor at each point of the titration curve is equal to α[E]_{total}, where [E]_{total} is the total enzyme concentration. By plotting the concentration of the bound inhibitor versus the concentration of the free inhibitor for each titration point, it was possible to derive the dissociation constant K_d.

Probe Displacement Assay. The assay was run as described (34). Briefly, fluorescence emission spectra were recorded on a Perkin-Elmer LS50B instrument with a cuvette holder thermostated at 23 °C. Protease (100 nM) was incubated in 2.5 mL of 50 mM HEPES, pH 7.5, 1 mM DTT, 15%

² Abbreviations: ACE, angiotensin-converting enzyme; Cha, β-cylohexylalanine; CHAPS, 3-(3-cholamidopropyl)dimethylammonio-1-propanesulfonate; Dap, diaminopropionic acid; Dif, 3,3-diphenylalanine; DMSO, dimethyl sulfoxide; DTT, dithiothreitol; HEPES, 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid; Phg, phenylglycine; TFA, trifluoroacetic acid; SAR, structure–activity relationship.

glycerol, and 1% CHAPS containing 80 μ M Pep4AK until attainment of thermal equilibrium. Active site probe (500 nM), having the sequence Ac-Asp-Glu-Dap(*N*- β -dansyl)-Glu-Cha-Cys-OH (Dap, diaminopropionic acid; Cha, β -cyclohexylalanine), was added, and emission spectra were recorded between 400 and 600 nm at a scan speed of 60 nm/min upon excitation at 280 nm. Emission and excitation slits were 2.5 and 5 nm, respectively. Spectra were routinely corrected for the background signal of buffer and for dilution effects. Increasing amounts of competing, nonfluorescent ligand were added, spectra were recorded as described above, and probe displacement was monitored as a decrease in the fluorescence peak intensity at 510 nm. IC₅₀ values for probe displacement were calculated by fitting a two-parameter logistic equation to the fluorescence intensity data.

NMR. NMR spectra were recorded on a Bruker AMX-500 spectrometer, equipped for multichannel operation. Two-dimensional ¹H–¹⁵N HSQC experiments to monitor complex formation by NH chemical shift changes were acquired at 25 °C and analyzed as previously described (35).

Modeling. Energy minimization and molecular dynamics were used for modeling. All calculations were carried out with the program BatchMin and the molecular modeling package Insight II/Discover (Biosym Technologies Inc., San Diego, CA). Hydrogen atoms were included, and the potential energy of the complex was expressed by the force field MMFF, as implemented in the MacroModel V5.0 distribution of the simulation program BatchMin.

Synthesis. Detailed procedures and analytical data for the solid-phase and solution synthesis of the libraries and the single compounds of the Results section are described in the Supporting Information.

RESULTS

The work described here evolves from our previous work, which exploited binding to the *S'* region of NS3/4A to develop inhibitors of the enzyme (23). Since the prime region of all the natural substrates, in the form of synthetic peptides spanning residues P₁'–P₁₀', does not inhibit NS3/4A to any significant extent (15, 40), we had explored its binding potential using noncleavable substrate analogues, which included both P and P' residues. The optimization process led to an increase in binding of more than 3 orders of magnitude, with the best decapeptide, Ac-Asp-Glu-Dif-Ile-Cha-Cys-Pro-Cha-Asp-Leu-NH₂ (Dif, 3,3-diphenylalanine), showing IC₅₀ <200 pM. Prime site optimization (residues P₂'–P₄') had improved potency >2000-fold (23), to be compared with the 20-fold gain of the equivalent region (residues P₂–P₄) on the P-side (2). This was because for P-region binding inhibitors (“product inhibitors”) a substantial portion of the binding energy is due to interactions of the C-terminal carboxylate into the active site (2, 15), a feature which was absent in the P–P' decapeptides. In the absence of the carboxylate contribution, inhibitors binding mainly to the S or *S'* subsites of NS3/4A showed comparable potency (23). More importantly for drug design, prime site binding derived from interactions not used by the substrate, as shown by the finding that, in the context of a cleavable substrate, the optimized prime site sequence *decreased* turnover by the protease (23).

Combinatorial Library of Prime Site Inhibitors. The above data suggested to us that a viable strategy to develop

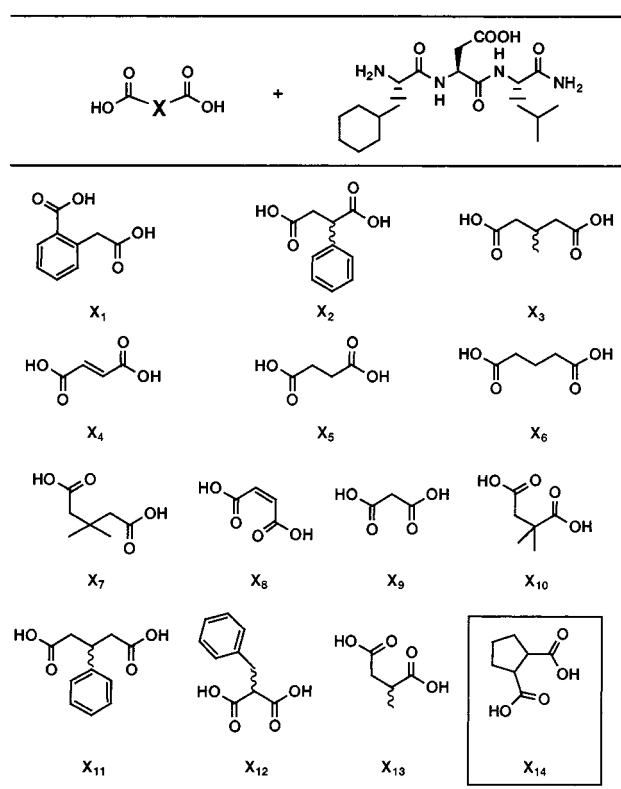


FIGURE 1: Diacid fragments used in the prime site inhibitor library. Due to the presence of isomers, a total of 26 compounds were produced. The fragment (X₁₄) yielding the active mixture is highlighted.

inhibitors, which would bind exclusively to the prime site of NS3/4A, should combine two elements: (a) the optimized prime site binding tripeptide (Cha-Asp-Leu-NH₂) and (b) an *N*-terminal carboxylic acid, suitably positioned in the active site to engage in interactions, similar to those previously established for the *C*-terminal carboxylate of peptide product inhibitors (2, 15). The key to a successful design was therefore how to devise the proper linkage between these two elements. Two elements argued against the design of a specific molecule, using computer modeling: (i) no structural data were available to us of NS3/4A with the prime site occupied by an inhibitor; (ii) NS3/4A shows an induced-fit mechanism of binding (37, 38), with the prime site in particular forming different tertiary structures with different decapeptide inhibitors (39). Therefore, we decided to probe our hypothesis by the synthesis of a small combinatorial library.

The chosen structure for the library was very simple, as shown in Figure 1; the P₂'–P₄' tripeptide fragment was capped by a diacid moiety. By including fragments with a flexible or a more rigid tether between the acids, with or without an intervening side chain, we collected a small representative set of 14, commercially available, diacids. Because of the presence of isomers, the library contained 26 compounds, not all of which were actually produced, because of spontaneous decarboxylation of the inhibitors containing some of the fragments (see Materials and Methods). To reassure ourselves that any inhibitor selected from the library would exploit binding of the optimized P₂'–P₄' fragment, we also prepared a control library, based on the nonoptimized -Ser-His-Leu-NH₂ fragment of the NS4A/

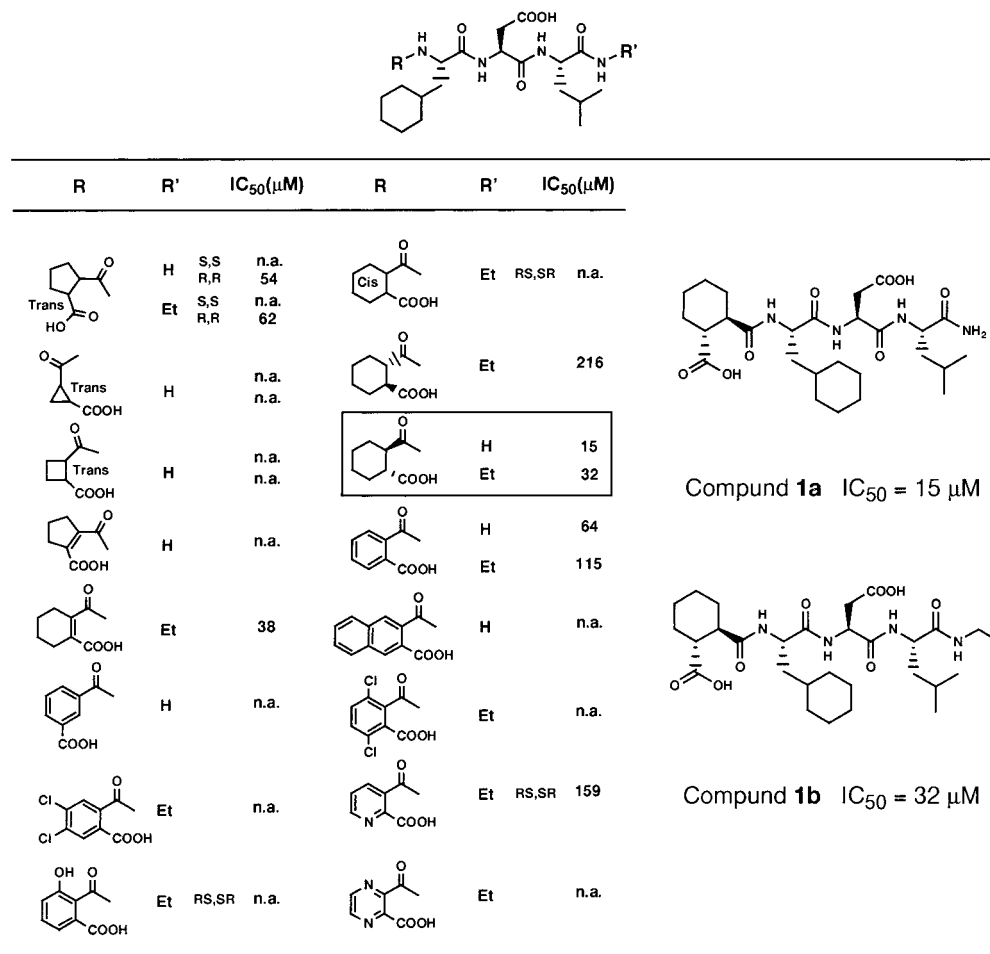


FIGURE 2: SAR of the cyclic diacid moiety of the capped tripeptide inhibitors. The complete structure (and activity) of the two more active compounds is given on the left.

4B cleavage site (3). As expected, no inhibitor was selected from the control library. By contrast, the combination of the optimized -Cha-Asp-Leu-NH₂ tripeptide with the cyclic diacid fragment X₁₄ yielded an active compound. It is worth noting that the active fragment (*trans*-1,2-cyclopentanedicarboxylic) is the diacid analogue of proline, the residue which was used as the P₁' residue during the selection of the P-P' spanning decapeptides.

The *S,S* and *R,R* isomers present in the active mixture of the derivatives with *trans*-1,2-cyclopentanedicarboxylic acid were chromatographically separated and retested, and only one compound was found active (IC₅₀ = 54 μ M). Kinetic analysis on the active compound showed competitive behavior, with K_i = 32 μ M.

Structure-Activity Relationship (SAR) of the Capped Tripeptides. We prepared multiple analogues of the initial lead, which specifically addressed the various elements of the molecule, i.e., the N-terminal cyclic diacid, the side chains in positions P₂' and P₃', and the C-terminal carboxamide.

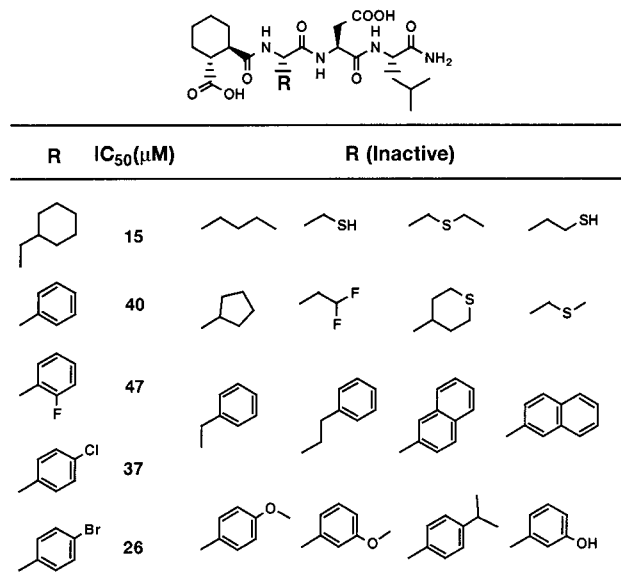
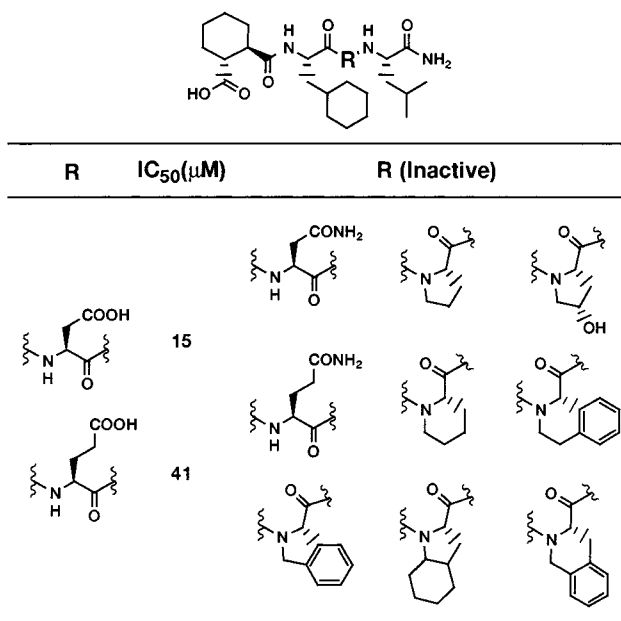
(A) **The Capping Diacid.** The SAR around the cyclic diacid moiety was very tight, with only a few fragments resulting in active compounds (Figure 2). The best capping groups were *trans*-cyclohexanedicarboxylic acid (compound **1**, IC₅₀ = 15 μ M), cyclohex-1-enedicarboxylic acid (IC₅₀ = 38 μ M), and phthalic acid (compound **2**, R' = H, IC₅₀ = 64 μ M). The cyclohexene derivative, despite the good affinity and lack of chiral centers, was abandoned because of its intrinsic

instability, with easy loss of the dicarboxylic moiety, a phenomenon previously observed for analogous molecules (40, 41). This tendency was present, to a lower extent, also in the phthalic acid derivatives. Moreover, substitutions in the ring lowered the potency of the compound. The *trans*-cyclohexanedicarboxylic acid (compound **1a**, R' = H, IC₅₀ = 15 μ M; compound **1b**, R' = CH₂-CH₃, IC₅₀ = 32 μ M; Figure 2) was therefore selected for all our subsequent studies. The commercial availability of the pure (*R,R*) enantiomer also allowed us to establish the absolute configuration of the active compound, which was found to be (*R,R*) in agreement with modeling predictions (see Figure 2).

Within the capping group, the presence of a free carboxylate was absolutely required, as this could not be deleted, nor transformed into an amide, without complete loss of activity (data not shown). By contrast, the C-terminal amide could be either primary or secondary: -CO-NH₂ and -CO-NH-CH₂-CH₃ had similar activity (Figure 2).

(B) **P₂' and P₃' Positions.** We next turned our attention to the tripeptide moiety and prepared a series of analogues aimed at comparing the SAR in the current series with that previously established for the same prime binding tripeptide in the context of a decapeptide (23). Results for the P₂' residue are shown in Figure 3 and those for the P₃' residue in Figure 4.

Probably due to the much reduced potency of the tripeptides with respect to the decapeptides, many P₂' residues,

FIGURE 3: SAR of the capped tripeptide inhibitors: the P₂' residue.FIGURE 4: SAR of the capped tripeptide inhibitors: the P₃' residue.

which were accepted in the latter (23), gave rise to inactive compounds in the former: for example, Nle and homophenylalanine. However, the optimal P₂' amino acids remained the same, i.e., Cha followed by phenylglycine (Phg), with or without substituents in the aromatic ring.

Similarity between the two series was also observed for the P₃' residues (Figure 4), where a negative charge was required, and Asp was preferred over Glu (compare ref 23).

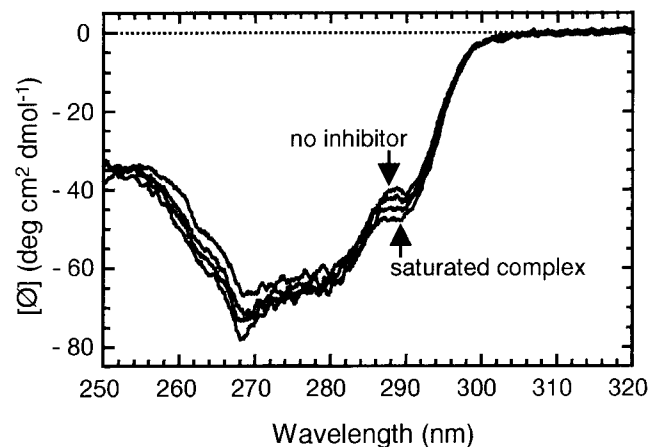
Site-directed mutagenesis of the two potential basic contacts for this acid, Lys¹³⁶ and Arg¹⁰⁹ (Table 1) indicated that the latter was responsible for the P₃' preference, as previously observed for the decapeptide series (23).

(C) *N-Methyl Analogues*. The importance of the backbone primary amide for inhibitor binding was probed through a series of *N*-methyl analogues. *N*-Methylation at both P₁' and P₃' markedly decreased (≥ 10 -fold) inhibitory potency, while the P₂' position was almost unaffected: we found a 2-fold and 3-fold increase in the IC₅₀ in the series based on

Table 1: Site-Directed Mutagenesis of the NS3 Residues in Proximity of the P₃' Negative Charge

enzyme	IC ₅₀ of compound 1a ^a
WT	32
Lys ¹³⁶ Arg	17
Lys ¹³⁶ Met	27
Lys ¹³⁶ Ser	50
Arg ¹⁰⁹ Gln	>300

^a The IC₅₀ is shown of compound **1a** on the mutant enzymes (WT = wild-type, i.e., no mutation).

FIGURE 5: Binding of compound **1a** to NS3/4A studied by circular dichroism. Near-UV region of the CD spectra of the protease/cofactor with increasing concentrations of inhibitor. The dissociation constant K_d was calculated by plotting bound versus free inhibitor concentration (see text for more explanation).

compounds **1** and **2**, respectively. The latter finding formed the basis for the design of a new series (see below). Sensitivity of P₁' to *N*-methylation was anticipated by modeling, which predicted the existence of a hydrogen bond with Ser⁴² of the enzyme (see below).

Overall, all of the above data were highly suggestive that, according to our design, the capped tripeptides bound to the S' region of the enzyme in a manner very similar to the decapeptides from which they derived, with their N-terminal carboxylate positioned in the active site, as observed for the product inhibitors (2, 15).

Detailed Studies of Compounds 1a and 1b. In addition to the clear, but indirect evidence from SAR, we sought further proof of the binding mode of the inhibitors by a series of experiments, carried out on compounds **1a** and **1b** (Figure 2). Kinetic analysis (not shown) confirmed that the inhibitor is competitive with the substrate: compound **1a**, $K_i = 13 \mu\text{M}$; compound **1b**, $K_i = 22 \mu\text{M}$.

(A) *Binding Studies by CD*. Binding of compound **1a** to NS3/4A was then studied by circular dichroism (CD) spectroscopy in the near-UV region (Figure 5). CD of a protein in the near-UV range is contributed by the aromatic side chains and is characteristic of the protein tertiary structure. The addition of compound **1a** to the NS3/4A complex causes a change in the near-UV spectrum of the protease, with a small but consistent increase of intensity between 252 and 292 nm. Therefore, a tertiary structure rearrangement of the protease occurs upon binding of **1a**, suggesting that an induced-fit mechanism is associated with the inhibition process. By monitoring the variation of the ellipticity at 288 nm upon addition of increasing concentra-

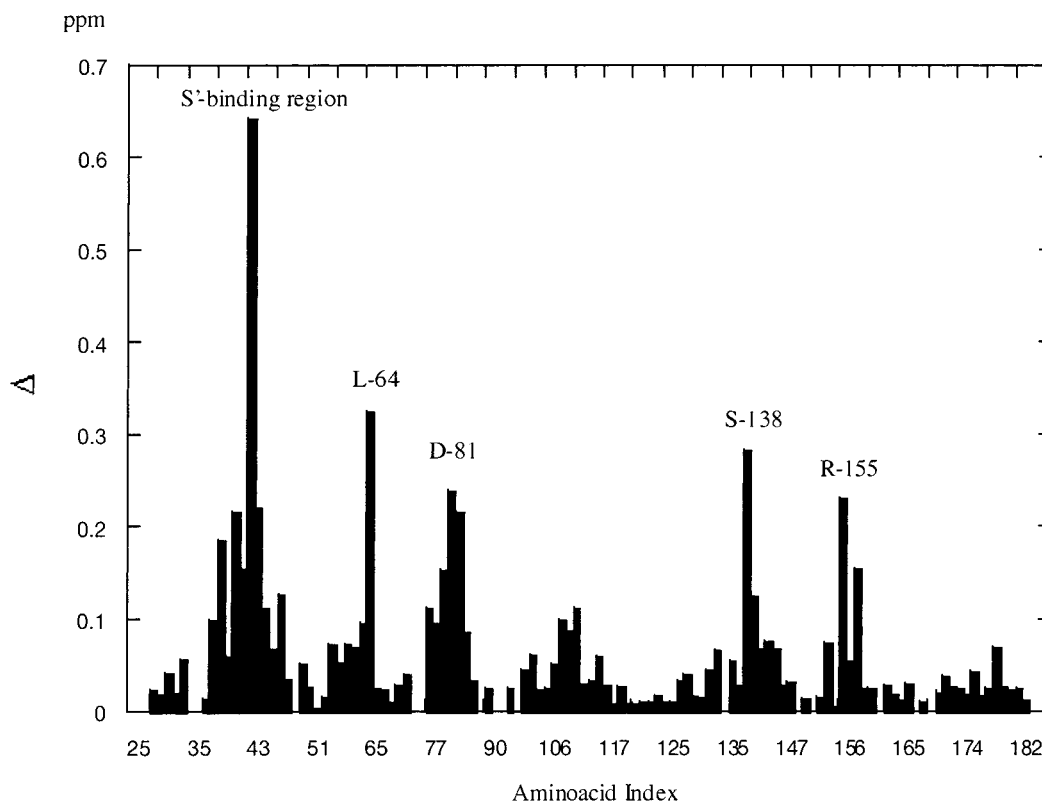


FIGURE 6: Chemical shift perturbation experiment with compound **1b**. For details see text.

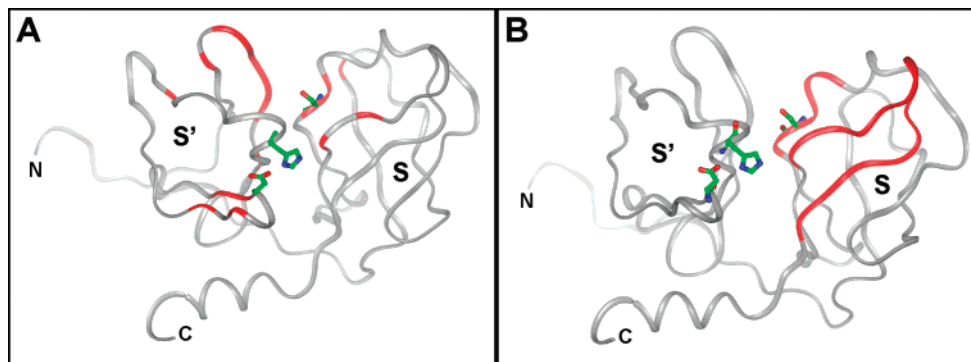


FIGURE 7: Comparison between the chemical shift perturbations observed with (A) compound **1b** and (B) a representative product inhibitor (see text). The ligand-binding region in the backbone of the enzyme, as determined in the titration experiments, is colored in red. Residues of the catalytic triad are in ball-and-stick representation.

tions of inhibitor, we could indeed determine a K_d of 10 μM for the inhibitor–enzyme complex formation, in good agreement with the inhibitory constant derived from kinetic experiments ($K_i = 13 \mu\text{M}$).

(B) *Binding Studies by NMR*. The binding mode of the inhibitors was further investigated by NMR, using the chemical shift perturbation method. Titration of the ^{15}N -labeled enzyme with compound **1b** (Figure 6) produced chemical shift variations of both the ^1H and ^{15}N nuclei of the backbone amide group. These shifts were used to map the interaction surface between the enzyme and the inhibitor, characterizing the region of the protein involved in compound recognition. This study was performed in the absence of the cofactor, since the complex between NS3 and Pep4AK peptide (2, 3, 15, 23) is not stable enough in solution to allow characterization by NMR (21).

The titration is shown in Figure 6. The quantity Δ (42) is a weighted average of the perturbations of the chemical shifts

of ^1H and ^{15}N and is the same previously used to characterize binding of the product inhibitors to NS3 (35). The most important perturbations were observed around Ser⁴² (the region from Ser³⁷ to Thr⁴⁶) and around two of the members of the catalytic triad (Asp⁸¹ and Ser¹³⁹). The shifts observed for these residues are similar to those previously measured (35), highlighting that both the old (2, 15, 35) and the new inhibitors occupy the active site of the enzyme. By contrast, no significant shift was observed in the previously defined S region of the enzyme (residues Ala¹⁵⁶ to Val¹⁷⁰), in striking contrast with the binding of inhibitors based on the P region of the substrate (35). New, large shifts were instead measured for Leu⁶⁴ and Arg¹⁵⁵.

The difference in the binding mode of the two classes of inhibitors is illustrated in Figure 7. Figure 7B shows a previously characterized product inhibitor derived from the P region of the substrate, Ac-Asp-Glu-Dif-Glu-Cha-Cys-OH (2, 15, 35), that binds to the S region of the enzyme and to

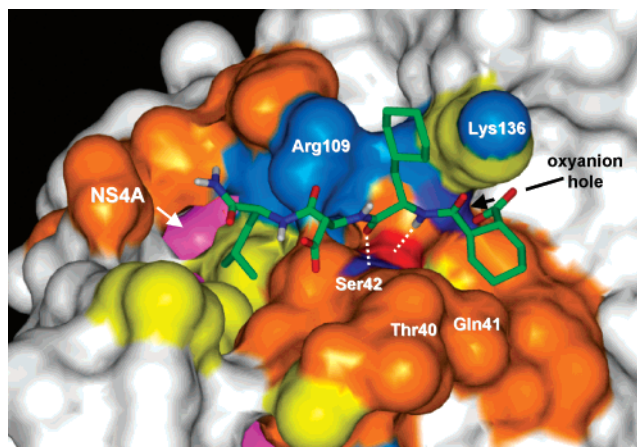


FIGURE 8: Modeling of compound **1a** in the S' region of NS3/4A protease.

the active site. By contrast, Figure 7A shows the capped tripeptide inhibitor, compound **1b**, which extends from the active site into the S' region.

(C) *Probe Displacement Assay.* Figure 7 highlights that, according to our design, the only region in common between the capped tripeptide series and the P-side derived product inhibitors is the active site, where the carboxylate binds. Since both series derive a major part of their binding energy from this interaction, they should be unable to simultaneously bind to NS3. We tested this hypothesis by using a previously developed displacement assay (34) based on a dansylated hexapeptide probe, Ac-Asp-Glu-Dap(*N*- β -dansyl)-Glu-Cha-Cys-OH. The probe, which derives from an optimized product inhibitor (2), binds to the S region of the enzyme and positions its C-terminal carboxylate into the active site (15, 34). Active site occupancy by this compound can be monitored by following fluorescence energy transfer between the dansyl fluorophore and tryptophan residues of NS3/4A. The probe was previously used to unambiguously assess active site binding of NS3/4A inhibitors (34). Compound **1b** is able to displace the dansylated probe, with an IC_{50} of 30 μ M, in excellent agreement with the one established by kinetic analysis (32 μ M).

(D) *Modeling of the Complex between Compound 1a and NS3/4A.* All of the above knowledge, direct and indirect, was used to model the compound into the S' region of the NS3/4A protease (Figure 8). The P_1' interaction site is a narrow cleft formed by Lys¹³⁶ on one side and Gln⁴¹ and Thr⁴⁰ on the other side. Together, these residues severely limit the space available to the capping diacid group. This observation provides a structural explanation for the tight SAR observed in this position. The carboxylate of the capping group contributes to binding electrostatically by forming a number of hydrogen bonds to the residues of the oxyanion hole and to His⁵⁷, similarly to the carboxylate of hexapeptide product inhibitors (35). The apolar parts of the capping group contribute to binding by lipophilic interactions with the aliphatic part of the Lys¹³⁶ side chain. The S_2' is formed by the side chains of Lys¹³⁶ and Arg¹⁰⁹, which create a narrow interaction site for lipophilic residues. The side chains of Cha and Phg are well placed to interact with the lipophilic parts of the side chains of these two residues, in a stacking fashion which maximizes the lipophilic contact area.

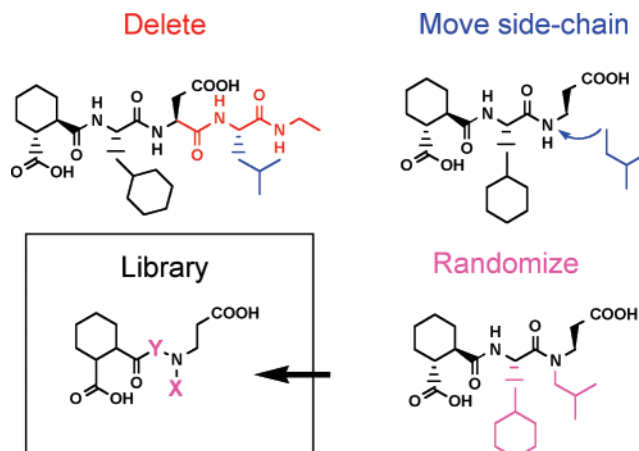


FIGURE 9: Design of a second generation NS3/4A prime site inhibitor series. For explanation, see text.

Modeling and experimental data also indicate that the P_2' residue forms backbone hydrogen bonds to Ser⁴², as highlighted in Figure 8. These backbone interactions help in stabilizing the binding conformation of the inhibitor. In the P_3' position the vicinity of Arg¹⁰⁹ favors acidic residues, and aspartate in particular is especially suited for short-range interaction with Arg¹⁰⁹. NS4A and its binding region provide a lipophilic surface patch, which is exploited by hydrophobic residues in position P_4' . Overall, the interactions contributing to the binding of this class of inhibitors are distributed over the whole S' region. Moreover, many of these interactions are lipophilic in nature, thus offering an attractive opportunity for further improvement of potency and pharmacological properties.

Evolution of the Lead: Reduction in Peptide Character. Having established the validity of our initial design, we used the information collected to design a new series of inhibitors with reduced peptide character. The design of the new series used the same logic of the first one (Figure 9).

We designed a generic structure and looked for a specific compound by preparing a small combinatorial library. From modeling, it appeared that (i) the P_4' residue was contributing to binding only with its side chain; very little, if any, contact was made by the C-terminal carboxamide, and no hydrogen bond was formed with the enzyme; (ii) likewise, the amide NH of the P_2' residue was pointing outward, toward the solvent; (iii) accordingly, alkylation of the P_2' NH was well tolerated; (iv) the acidic side chain in P_3' was critical for activity and was to be preserved. Our design therefore deleted residue P_4' (in red in Figure 9) and moved its side chain (in blue in Figure 9) onto the amide nitrogen of residue P_2' , creating a "peptoid-like" (43) center bearing simultaneously the P_3' and the P_4' side chains. One of these side chains was held fixed as the side chain of an aspartic acid, while the other was randomized (in violet in Figure 9). The expectation was that the binding mode of the compounds would be different from the capped tripeptide series and therefore the specific side chain filling the S_4' cavity would be better selected than designed. For the same reason, we used three different amino acids in position P_2' (in violet in Figure 11), including a glycine to allow for maximal flexibility in positioning of the randomized side chain.

The building blocks used in the two variable elements of the library (X and Y in Figure 12) are shown in Figure 10.

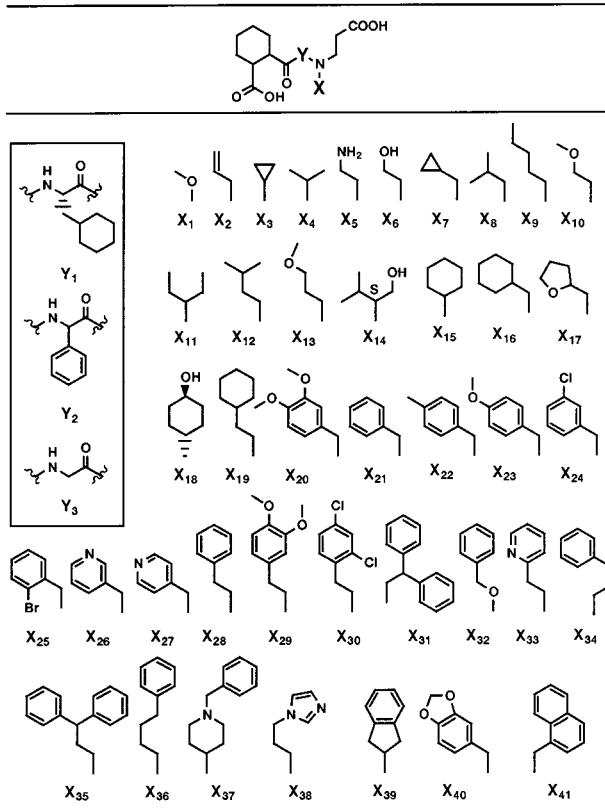


FIGURE 10: Building blocks used for the second combinatorial library of NS3/4A prime site inhibitors.

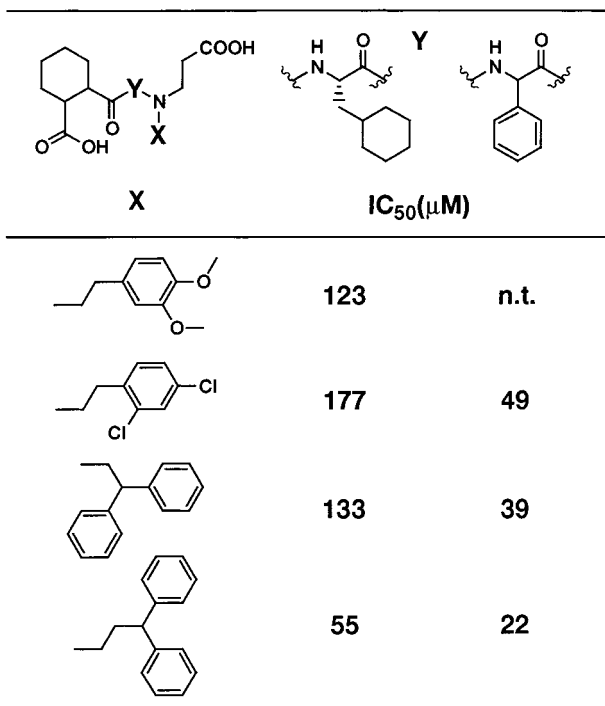


FIGURE 11: Active compounds from the combinatorial library illustrated in Figure 10.

Overall, the library was composed of $41 \times 3 = 123$ compounds. In Figures 10–12 the phenylglycine is indicated with no defined chirality, because we realized during the synthesis that complete racemization was occurring at the phenylglycine C α during the coupling step (for more details, see Materials and Methods). In Figure 11, the activity values

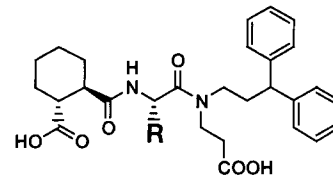


FIGURE 12: Effect of ring substituents on the activity of prime site inhibitors.

refer to the most active isomer, after HPLC separation. Although we assume, on the basis of the SAR and modeling, that the active isomer contains (L) phenylglycine, the absolute configuration has not yet been formally assigned.

Four different X fragments gave active compounds, in combination with Y = Phg or Cha, while all of the library members with Y = Gly were inactive (Figure 11). The comparison between compounds having the same X fragment, but different Y, highlighted that consistently the phenylglycine side chain was superior to the β -cyclohexyl-alanine side chain. In the capped tripeptides, the opposite was true, highlighting some difference in the binding mode of the two series.

(A) *Compounds with Phenylglycine Derivatives.* This finding offered the opportunity to introduce substituents into the phenyl ring, which readily lead to improved potency (Figure 12), and indeed the 4-Br substituent yielded a compound with 7-fold lower IC₅₀ (2 μ M). This inhibitor, with a molecular weight below 600, is a promising lead for a new iteration of design.

DISCUSSION

The evidence presented in this paper supports the notion that we have successfully designed prime site inhibitors of NS3/4A. The correctness of our design was confirmed through several lines of evidence: First, kinetic analysis and CD showed that inhibition was due to occupation of the substrate-binding site. Second, SAR around the lead compound **1** showed that in the capped tripeptides all of the elements were maintained, which were key to binding of the prime fragment of the decapeptides, and in addition that the N-terminal acid was absolutely required for activity. Third, NMR showed that the regions of NS3, which make

contact with the inhibitor, are indeed the prime site and the active site. Although this result was acquired in the absence of the NS4A cofactor, it was confirmed in its presence by displacement, by compound **1**, of a product inhibitor-derived fluorescent probe.

We then used modeling, site-directed mutagenesis, and selective N-methylation to derive a binding hypothesis, which suggested a way to evolve the capped tripeptide series toward reduction in size and peptide character, with concomitant improvement in potency. Notably, the SAR of the new series suggested similarities, but also differences, with respect to the previous one, holding promise for a new iteration of the design cycle.

Overall, the above work on NS3/4A shows that targeting the prime site of a serine protease can be a successful strategy for drug design. Prior to our work on NS3/4A (23), other authors had already shown that prime site binding could be an effective component of a potent inhibitor: for example, Imperiali and Abeles reported that in their difluoromethyl ketone inhibitors of chymotrypsin, spanning P–P', prime site residues could improve potency >1000-fold (25). However, binding of all the previously described inhibitors depended on contacts in the nonprime region of the enzyme. Therefore, the current work represents the first example of purely prime site inhibitors of a serine protease.

The closest analogy with our inhibitors is found in a different class of proteolytic enzymes, namely, metalloproteases. Our design is indeed strongly reminiscent of Byers and Wolfenden's concept of "collected product" (biproduct) inhibitor (44, 45). In rationalizing the unusual high potency of L-benzylsuccinic acid as an inhibitor of carboxypeptidase A, these authors noted that its structure resembled the collected products of peptide hydrolysis and was thus bound with an affinity resembling their combined affinity (45). The concept was then successfully exploited by Patchett and co-workers to develop potent inhibitors of angiotensin-converting enzyme (ACE; reviewed in ref 46) and thermolysin (47, 48). In the X-ray structure of a collected product inhibitor bound to the latter enzyme (48), and in the putative binding interaction, deduced from SAR, of enalaprilat with the former enzyme (46), the carboxylate is in bidentate coordination with the catalytic Zn²⁺ ion, having displaced a water molecule from the metal. The oxygen atoms of the carboxylate are located precisely where the oxygens of the hydrated scissile peptide bond of a substrate would be. In addition to binding into the active site, ACE inhibitors extended into the S₁' and S₂' subsites. Occupancy also of the S₁ subsite, although not strictly required for activity, greatly improved potency (46).

In analogy with the collected product design, our capped tripeptides collect in a single molecule features of the N-terminal (the carboxylate) and the C-terminal (the optimized P₂'–P₄' sequence) cleavage products, and both elements are strictly required for activity. It is worth noting that the potency of our best NS3/4A inhibitor (2 μM) is equivalent to the potency (IC₅₀ = 4.9 and 2.4 μM, respectively) of N-carboxyalkyl or N-carboxymethyl dipeptide inhibitors of ACE lacking the P₁ side chain (46). Introduction of the proper P₁ side chain in the collected product leads to enalaprilat, and this is also the current direction of our design for NS3/4A.

We believe it conceivable that the design of an inhibitor with an analogous binding mode may be possible for other

serine proteases. In the case of NS3/4A, we found that the oxygen atoms of the P₁ α-carboxylate are the recipients of extensive hydrogen-bonded interactions, in particular with the imidazole ring of the catalytic histidine and with the main chain amido groups of the residues of the oxyanion hole (15, 35). This binding motif is not limited to NS3 but has been found in the crystal structure of the complex of other serine proteases with their N-terminal cleavage products: Sindbis virus core protein (49), *Streptomyces griseus* protease A (50), and bovine thrombin (51). Further interactions in the active site are more peculiar to each enzyme; for example, NS3/4A interacts with the carboxylate also with the side chain of Lys¹³⁶ (15, 35). When an acidic moiety is not a viable option for active site binding, a classical "serine trap" might be more indicated, especially for those proteases displaying a "lock-and-key" catalytic machinery, rather than an induced-fit mechanism like NS3/4A (41, 42). The active site binding element might be combined with an optimized prime site binding fragment, generated by the same course of action used for NS3 (23). The needed starting point, a noncleavable substrate analogue, can be easily obtained by simply substituting the P₁' residue with proline or a D-amino acid (52). The critical choice of the proper active site binding fragment will of course depend on the particular enzyme under study. As in our case, a combinatorial search of the best fragment could help to compensate for the inevitable uncertainties of the specific design.

ACKNOWLEDGMENT

The authors thank Stefano Acali for help with peptide synthesis, Francesca Naimo and Fabio Bonelli for mass spectrometry, Silvia Pesci for analytical NMR, and Sergio Serafini, Nadia Gennari, Mauro Cerretani, Gabriella Biasiol, and Mirko Brunetti for the protease assays. We are especially grateful to Raffaele De Francesco for helpful discussions throughout the work and to Victor G. Matassa for drawing our attention to the collected product analogy.

SUPPORTING INFORMATION AVAILABLE

Detailed procedures and analytical data for the solid-phase and solution synthesis of the libraries and the single compounds of the Results section. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

1. Perona, J. J., and Craik, C. S. (1995) *Protein Sci.* 4, 337–360.
2. Ingallinella, P., Altamura, S., Bianchi, E., Taliani, M., Ingenito, R., Cortese, R., De Francesco, R., Steinkühler, C., and Pessi, A. (1998) *Biochemistry* 37, 8906–8914.
3. Urbani, A., Bianchi, E., Narjes, F., Tramontano, A., De Francesco, R., Steinkühler, C., and Pessi, A. (1997) *J. Biol. Chem.* 272, 9204–9209.
4. Landro, J. A., Raybuck, S. A., Luong, Y. P. C., O'Malley, E. T., Harbeson, S. L., Morgenstern, K. A., Rao, G., and Livingston, D. J. (1997) *Biochemistry* 36, 9340–9348.
5. Zhang, R., Durkin, J., Windsor, W. T., McNemar, C., Ramanathan, L., and Le, H. V. (1997) *J. Virol.* 71, 6208–6213.
6. Houghton, M. (1996) Hepatitis C Viruses, in *Fields' Virology* (Fields, B. N., Knipe, D. M., and Howley, P. M., Eds.) 3rd ed., pp 1035–1058, Lippincott-Raven, Philadelphia and New York.
7. Hoofnagle, J. H. (1997) *Hepatology* 26 (Suppl. 1), 15S–20S.

8. Bartenschlager, R., Ahlborn Laake, L., Mous, J., and Jacobsen, H. (1994) *J. Virol.* 68, 5045–5055.
9. Lin, C., Pragai, B. M., Grakoui, A., Xu, J., and Rice, C. M. (1994) *J. Virol.* 68, 8147–8157.
10. Tanji, Y., Hijikata, M., Hirowatari, Y., and Shimotohno, K. (1994) *Gene* 145, 215–219.
11. Failla, C., Tomei, L., and De Francesco, R. (1995) *J. Virol.* 69, 1769–1777.
12. Hahm, B., Han, D. S., Back, S. H., Song, O. K., Cho, M. J., Kim, C. J., Shimotohno, K., and Jang, S. K. (1995) *J. Virol.* 69, 2534–2539.
13. Han, D. S., Hahm, B., Rho, H. M., and Jang, S. K. (1995) *J. Gen. Virol.* 76, 985–993.
14. Failla, C., Tomei, L., and De Francesco, R. (1994) *J. Virol.* 68, 3753–3760.
15. Steinkühler, C., Biasiol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A., and De Francesco, R. (1998) *Biochemistry* 37, 8899–8905.
16. Llinas-Brunet, M., Bailey, M., Fazal, G., Goulet, S., Halmos, T., Laplante, S., Maurice, R., Poirier, M., Poupert, M. A., Thibeault, D., Wernic, D., and Lamarre, D. (1998) *Bioorg. Med. Chem. Lett.* 8, 1713–1718.
17. Llinas-Brunet, M., Bailey, M., Déziel, R., Fazal, G., Gorys, V., Goulet, S., Halmos, T., Maurice, R., Poirier, M., Poupert, M. A., Rancourt, J., Thibeault, D., Wernic, D., and Lamarre, D. (1998) *Bioorg. Med. Chem. Lett.* 8, 2719–2724.
18. 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., and Thomson, J. A. (1996) *Cell* 87, 343–355.
19. Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., and Hostomska, Z. (1996) *Cell* 87, 331–342.
20. Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J., Sardana, M., Steinkühler, C., Tomei, L., De Francesco, R., Kuo, L., and Chen, Z. (1998) *Protein Sci.* 7, 837–847.
21. Barbato, G., Cicero, D. O., Nardi, M. C., Steinkühler, C., Cortese, R., De Francesco, R., and Bazzo, R. (1999) *J. Mol. Biol.* 289, 370–384.
22. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W., and Weber, P. (1999) *Structure* 7, 1353–1363.
23. Ingallinella, P., Bianchi, E., Ingenito, R., Koch, U., Steinkühler, C., Altamura, S., and Pessi, A. (2000) *Biochemistry* 39, 12898–12906.
24. Eichler, J., and Houghten, R. A. (1993) *Biochemistry* 32, 11035–11041.
25. Imperiali, B., and Abeles, R. H. (1987) *Biochemistry* 26, 4474–4477.
26. Betzel, C., Singh, T. P., Visanji, M., Peters, K., Fittkau, S., Saenger, W., and Wilson, K. S. (1993) *J. Biol. Chem.* 268, 15854–15858.
27. Bastos, M., Maeji, N. J., and Abeles, R. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 6738–6742.
28. Saxena, A. K., Singh, T. P., Peters, K., Fittkau, S., and Betzel, C. (1996) *Protein Sci.* 5, 2453–2458.
29. Tong, L., Qian, C., Massariol, M.-J., Déziel, R., Yoakim, C., and Lagacé, L. (1998) *Nat. Struct. Biol.* 5, 819–826.
30. Abato, P., Conroy, J. L., and Seto, C. T. (1999) *J. Med. Chem.* 42, 4001–1009.
31. Speri, S., Jacob, U., de Prada, N. U., Stürzenbecher, J., Wilhelm, O. G., Bode, W., Magdolen, V., Huber, R., and Moroder, L. (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97, 5113–5118.
32. Bianchi, E., Urbani, A., Biasol, G., Brunetti, M., Pessi, A., De Francesco, R., and Steinkühler, C. (1997) *Biochemistry* 36, 7890–7897.
33. Cerretani, M., Di Renzo, L., Serafini, S., Vitelli, L., Gennari, N., Bianchi, E., Pessi, A., Urbani, A., Colloca, S., De Francesco, R., Steinkühler, C., and Altamura, S. (1999) *Anal. Biochem.* 266, 192–197.
34. Fattori, D., Urbani, A., Brunetti, M., Ingenito, R., Pessi, A., Prendergast, K., Narjes, F., Matassa, V. G., De Francesco, R., and Steinkühler, C. (2000) *J. Biol. Chem.* 275, 15106–15113.
35. Cicero, D. O., Barbato, G., Koch, U., Ingallinella, P., Bianchi, E., Nardi, C., Steinkühler, C., Cortese, R., Matassa, V., De Francesco, R., Pessi, A., and Bazzo, R. (1999) *J. Mol. Biol.* 289, 385–396.
36. La Plante, S. R., Cameron, D. R., Aubry, N., Lefebvre, S., Kukolj, G., Maurice, R., Thibeault, D., Lamarre, D., and Llinas-Brunet, M. (1999) *J. Biol. Chem.* 274, 18618–18624.
37. Bianchi, E., Orru', S., Dal Piaz, F., Ingenito, R., Casbarra, A., Biasiol, G., Koch, U., Pucci, P., and Pessi, A. (1999) *Biochemistry* 38, 13844–13852.
38. Pessi, A. (2001) *J. Pept. Sci.* 7, 2–14.
39. Ingallinella, P., Pucci, P., Dal Piaz, F., Pessi, A., and Bianchi, E. (2001) in *Peptides for the New Millennium, Proceedings of the 17th American Peptide Symposium* (Houghten, R. H., and Lelbl, M., Eds.) pp 545–546.
40. Glusenkamp, K.-H., Mengede, C., Drosdzioł, W., Jahde, E., and Rajewsky, M. F. (1998) *Bioorg. Med. Chem. Lett.* 8, 285–288.
41. Kirby, A. J. (1980) *Adv. Phys. Org. Chem.* 17, 183–278.
42. Grzesiek, S., Stahl, S. J., Wingfield, P. T., and Bax, A. (1996) *Biochemistry* 35, 10256–10261.
43. Simon, R. J., Kania, R. S., Zuckerman, R. N., Huebner, V. D., Jewell, D. A., Banville, S., Ng, S., Wang, L., Rosenberg, S., Marlowe, C. K., Spellmeyer, D. C., Tan, R., Frankel, A. D., Santi, D. V., Cohen, F. E., and Bartlett, P. A. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89, 9367–9371.
44. Byers, L. D., and Wolfenden, R. (1972) *J. Biol. Chem.* 247, 606.
45. Byers, L. D., and Wolfenden, R. (1973) *Biochemistry* 12, 2070–2078.
46. Wyvratt, M. J., and Patchett, A. A. (1985) *Med. Chem. Rev.* 5, 483–531.
47. Maycock, A. L., DeSousa, D. M., Payne, L. G., Den Broeke, J., Wu, M. T., and Patchett, A. A. (1981) *Biochem. Biophys. Res. Commun.* 102, 963.
48. Monzingo, A. F., and Matthews, B. W. (1984) *Biochemistry* 23, 5724.
49. Tong, L., Wengler, G., and Rossmann, M. G. (1993) *J. Mol. Biol.* 230, 228–247.
50. James, M. N. G., Sielecki, A. R., Brayer, G. D., and Delbaere, L. T. J. (1980) *J. Mol. Biol.* 144, 43–88.
51. Martin, P. D., Robertson, W., Turk, D., Huber, R., Bode, W., and Edwards, B. F. P. (1992) *J. Biol. Chem.* 267, 7911–7920.
52. Meldal, M., Svendsen, I., Juliano, L., Juliano, M. A., Del Nery, E., and Scharfstein, J. (1998) *J. Pept. Sci.* 4, 83–91.

BI025603X