Pergamon

Bioorganic & Medicinal Chemistry, Vol. 5, No. 4, pp 707–714, 1997

PII: S0968-0896(97)00016-3
0968-0896/97 $17.00 + 0.00

Synthesis of *N*-Glyoxylyl Peptides and Their In Vitro Evaluation as HIV-1 Protease Inhibitors

Driss Qasmi,[a] Eve de Rosny,[b] Loïc René,[a] Bernard Badet,[a,*] Isabelle Vergely,[b] Nicole Boggetto,[b] and Michèle Reboud-Ravaux[b]

[a]*Institut de Chimie des Substances Naturelles-CNRS, 91198 Gif-sur-Yvette Cedex, France*
[b]*Laboratoire d'Enzymologie Moléculaire et Fonctionnelle, Département de Biologie Supramoléculaire et Cellulaire, Institut Jacques Monod & Université Paris VII, 2 place Jussieu 75251, Paris Cedex 05, France*

Abstract—A series of novel synthetic peptides containing an *N*-terminal glyoxylyl function (CHOCO-) have been tested as inhibitors of HIV-1 protease. The *N*-glyoxylyl peptide CHOCO-Pro-Ile-Val-NH_2, which fulfills the specificity requirements of the MA/CA protease cleavage site together with the criteria of transition state analogue of the catalyzed reaction, was found to be a moderate competitive inhibitor although favorable interactions were visualized between its hydrated form and the catalytic aspartates using molecular modeling. Increasing the length of the peptide sequence led to compounds acting only as substrates. © 1997 Elsevier Science Ltd.

Introduction

Human immunodeficiency virus (HIV) is the causative agent for acquired immunodeficiency syndrome (AIDS). The molecular organization of the HIV genome is known to consist of *gag*, *pol* and *env* genes, which are necessary for viral replication.[1] The virally encoded protease (HIV protease) is responsible for the release of the protease itself, reverse transcriptase, integrase, and other proteins from the *gag–pol* fusion polyprotein.[2] Since inactivation of HIV-1 protease, by either chemical or genetic means, leads to the production of immature, noninfectious viral particles,[3] design methods to inhibit this enzyme have been recognized as a highly promising strategy to combat the growing threat of AIDS. In fact, a few HIV protease inhibitors have recently been approved to be used in combination with reverse transcriptase inhibitors for AIDS treatment.[4]

The HIV-1 protease is a homodimeric aspartyl protease with a C_2 axis of symmetry.[5] Its catalytic mechanism has been extensively studied and several X-ray structures of native and inhibitor-complexed forms have been determined.[5,6] The knowledge of enzyme structure and mechanism as well as the experience gained from the studies of other aspartic proteases such as renin, have facilitated the development of a significant number of inhibitors. Many of them are mimics of the transition-state structure generated during the hydrolysis of the peptide substrate.[7] In our attempt to find novel inhibitors for HIV-1 protease, we have designed and synthesized a new *N*-glyoxylyl-tripeptide (**5a**). The basic skeleton of **5a** is the tripeptide (Pro-Ile-Val-NH_2) derived from the *N*-terminal sequence of the capsid protein p24 generated via the processing of the precursors $Pr55^{gag}$ and $Pr180^{gag-pol}$ by the HIV-1 protease. Therefore, this compound is expected to occupy the S'_1–S'_3 subsites on binding with the protease.[8] The basis of our inhibitor design relied on the fact that the terminal aldehyde of glyoxylate exists exclusively in the hydrated form,[9] and thus the hydrated *N*-glyoxylyl-tripeptide **5a** might be considered as a structural mimic of the tetrahedral transition state of the hydrolytic process. Furthermore, molecular modeling revealed that the hydrated form of **5a** can readily fit in the binding pocket of the HIV-1 protease, and the *gem*-diol moiety of the hydrated **5a** is within hydrogen bond distance to both active site aspartate residues of

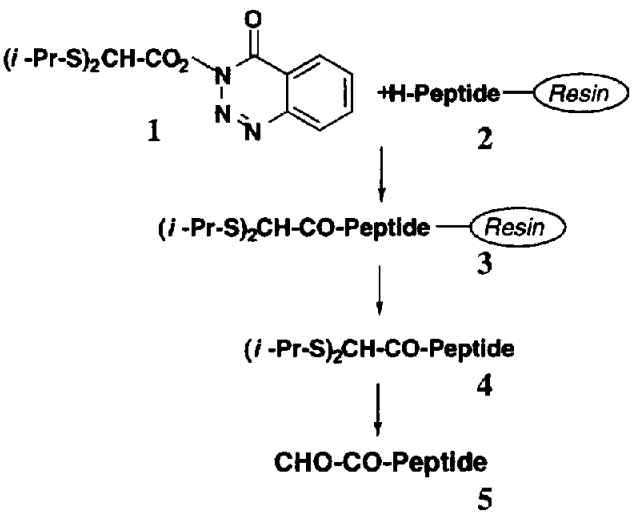

Figure 1. Synthetic access to CHO-CO-peptides.

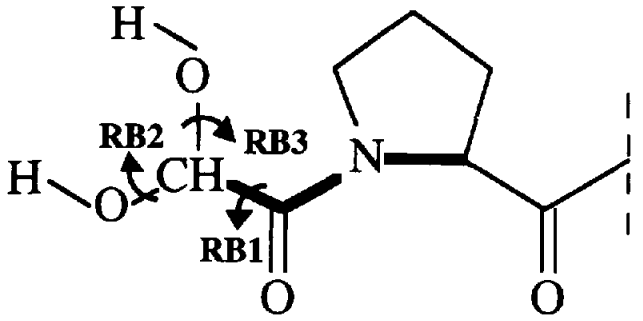

Figure 2. Systematic conformational search performed with **5a**^{hyd}-*trans* (value of 180° for the torsion angle ω defined by the thick bonds). RB1, RB2, and RB3 were the torsion angles used in reference for the rotamers shown with arrows.

the enzyme. Hence, **5a** is expected to act as a transition state analogue for HIV-1 protease, and its binding may cause inhibition of the enzyme activity. A few *N*-glyoxylyl-oligopeptides (**5b–f**) of varied chain lengths, which fulfilled the structural requirements for binding to the enzyme S_n–S'_3 subsites,[8] were also prepared to verify the role of the hydrated glyoxylyl group in the inhibition.

Results

Molecular modeling

The interaction between *N*-glyoxylyl-tripeptide **5a** and the HIV-1 protease was modeled using the known crystal structure of the complex formed between the protease and the inhibitor JG365 (Ac-Ser-Leu-Asn-Pheψ[CH(OH)CH_2N]-Pro-Ile-Val-OMe) used as a template.[10] Compound **5a** could exist either in its keto form (**5a**, CHO-CO-Pro-Ile-Val-NH_2) or its hydrated form (**5a**^{hyd}, $CH(OH)_2$-CO-Pro-Ile-Val-NH_2). In addition, two conformations of the amide bond involving the proline nitrogen [*cis* (ω = 0°) and *trans* (ω = 180°)] may exist. Thus, the four structures, **5a-*cis***, **5a-*trans***, **5a**^{hyd}-***cis***, and **5a**^{hyd}-***trans*** were examined by exploring the potential hydrogen bonding within the active site, especially with the catalytic aspartates 25 (A or B). Molecular modeling for the analysis of the interaction of **5a-*cis*** and **5a-*trans*** with the protease consisted in exploring the hydrogen bond possibilities between the glyoxylyl group of the inhibitor and the protease. Whereas no hydrogen bond was found between the *N*-glyoxylyl function of **5a-*trans*** and the catalytic aspartates of the protease, one was observed between the aldehydic oxygen of **5a-*cis*** and the protonated B/Asp25 residue. For the hydrated form of the inhibitor, a systematic conformational search was performed to study the interactions of **5a**^{hyd}-***cis*** or **5a**^{hyd}-***trans*** with the protease (Fig. 2). A unique family of 18 conformations was obtained for **5a**^{hyd}-***trans***. None of them showed favorable interaction between the *gem*-diol group of the inhibitor and the catalytic aspartates. The analysis with **5a**^{hyd}-***cis*** led to 96 conformations arranged in two families, f1 (78 conformers) and f2 (18 conformers), exhibiting RB1 angle values (Fig. 2) of approximately 0 and 180°, respectively. For 70% of the conformations, hydrogen bonds may exist between the *gem*-diol group of **5a**^{hyd}-***trans*** and one or the two Asp25 of the protease. The lowest energy conformer of each family was selected and its energy further minimized by a conjugate gradient method. The minimized conformation of f1 family is shown in Figure 3.

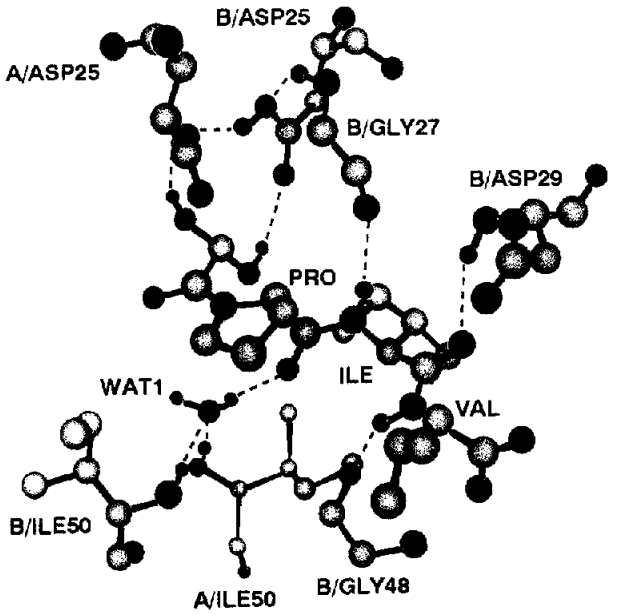

Figure 3. Hydrogen bond network between **5a**^{hyd}-***cis***, HIV-1 protease, and the structural water molecule Wat1 (minimized representative conformation of the structural family f1 generated after a systematic conformational search). Atoms are colored by atom type and hydrogen bonds are displayed as dashes. Hydrogen atoms not involved in hydrogen bonds are omitted.

Chemical synthesis of *N*-glyoxylyl peptides

In addition to the basic compound **5a**, five additional oligopeptides of varied chain length, whose amino acid sequences matched that of the S_n–S'_3 sites of the protease were also prepared. The protected glyoxylate, diisopropylthioacetic acid 3,4-dihydro-4-oxo-1,2,3-benzotriazin-3-yl ester (**1**, Fig. 1), was selected as the synthon for its versatility in solid-phase peptide synthesis. This activated ester was obtained from diisopropylthioacetic acid in good yield using classic activation procedures.[11] The precursor, diisopropylthioacetic acid, was readily prepared by coupling the commercially available glyoxylic acid (in aqueous solution) with 2-propanethiol.[12] As shown in Figure 1, the glyoxylate equivalent was introduced in the last step of solid-phase peptide synthesis operating in the Fmoc mode. The dithioacetal-peptide amides were purified and fully characterized at this stage. Since the CHO-CO moiety could not survive the drastic acidic conditions required for resin cleavage and side-chain deprotection, demasking the aldehyde function was performed after cleavage

Table 1. Inhibition of HIV-1 protease by compounds **5a–5f**

Compound	CHOCO-sequence	$K_i \times 10^5$(M) (inhibition)	$K_i \times 10^5$ (M)[a]
5a	-Pro-Ile-ValNH$_2$	52	–
5b	-Phe-Pro-Ile-ValNH$_2$	70[b]	–
5c	-Asn-Phe-Pro-Ile-ValNH$_2$	–	95[b]
5d	-Gln-Asn-Phe-Pro-Ile-ValNH$_2$	–	9
5e	-Ser-Gln-Asn-Phe-Pro-Ile-ValNH$_2$	–	52
5f	-Val-Ser-Gln-Asn-Phe-Pro-Ile-ValNH$_2$	–	12

[a]Equals to K_m.[17]
[b]Calculated using the equation of Cheng and Prussov.[21]
The standard deviations of the K_i values are below 20%.

of the peptide from its support. The deprotection was accomplished using *N*-bromosuccinimide in acetonitrile/H_2O or acetic acid/H_2O for longer peptides. The high polarity of these CHO-CO-peptides compared to their dithioacetal precursors allowed them to be readily purified by reversed-phase HPLC. The purified products **5a–f**, isolated in 40–60% yields, all exhibited a single peak by HPLC analysis.

Biochemical evaluation

A new chromogenic peptide Val-Ser-Gln-Asn-*p*-NO_2Phe-Pro-Ile-Val (V-S-Q-N-FNO_2-P-I-V) was synthesized and used as the substrate for the HIV-1 protease. This octapeptide, which sequence derived from that of the cleavage site at the MA–CA junction of $Pr55^{gag}$, carries a *p*-NO_2Phe (FNO_2) at the P_1 site. After incubation of the HIV-1 protease with this substrate (t_R = 36.9 min), two smaller peptides could be detected by HPLC with no detectable secondary degradation products. These two peptides were identified as V-S-Q-N-FNO_2 (t_R = 25.2 min) and P-I-V (t_R = 24.4 min) based on their retention times and spectral characteristics. The absorption change at 307 nm, as a result of the bond cleavage between *p*-NO_2Phe and Pro, allowed the reaction to be monitored continuously. Spectrophotometrical scanning (250–400 nm) of the incubation mixture at different time points also revealed the existence of two isobestic points at 284 and 392 nm. The initial rate of the reaction was measured (pH 4.5, 30 °C) by following the continuous release of V-S-Q-N-FNO_2 at 307 nm. The kinetic parameters for the enzymic hydrolysis were determined from the initial rates of hydrolysis measured at 307 nm, pH 4.5 and 30 °C: $k_{cat} = 1.6 (\pm 0.1)\ s^{-1}$, $K_m = 0.47 (\pm 0.5) \times 10^{-3}$ M. The kinetic parameters for inhibitors were determined from double reciprocal plots constructed from data for at least five inhibitor concentrations.

Compounds **5a** and **5b** behaved as competitive inhibitors (Table 1) of HIV-1 protease. Compounds **5c–f** behaved as substrates as demonstrated by HPLC analysis of their reaction mixture with the enzyme. Their affinities for the protease were estimated by competitive inhibition[7] versus the chromogenic substrate used in our activity assay.

Discussion

Peptide-aldehydes are well-known inhibitors for serine and cysteine proteases. They react with the catalytically active residue of these enzymes to yield the corresponding hemiacetal or hemithioacetal, which mimic the structure of the transition state of the enzymatic reaction.[13] The tight binding of the resulting tetrahedral adduct leads to enzyme inactivation. Although the catalytic mechanisms of serine/cysteine proteases and aspartic proteases are distinct, these compounds have also been employed as inhibitors for aspartic proteases, such as renin[14] and HIV protease.[15] For example, carbobenzoxydipeptide-aldehydes Cbz-P_2-P_1-H had been reported to inhibit HIV protease with an IC_{50} in the range of 1 μM.[15] The inhibition was attributed to the hydration of the *C*-terminal aldehyde of these peptides making them look like a reaction intermediate. However, the hydration is not expected to be complete for a simple aldehyde, and a significant portion of the peptide-aldehyde is left in its ineffective carbonyl form. One possible solution to overcome such a problem is to enhance the hydrophilicity of the aldehyde function. Hence, we have incorporated in our inhibitor design a glyoxylyl group, which has a much greater propensity for hydration than a simple aldehyde.[9,16] Since all peptide-aldehyde inhibitors reported so far for different proteases have the aldehyde function at the *C*-terminus, we thought it would be interesting to investigate the effect of an aldehyde group linked to the *N*-terminus of the peptides interacting with the S'_1, ... S'_n subsites of the protease.[8] Thus, the *N*-glyoxylyl-tripeptide **5a** was designed to present the hydroxyl groups of the hydrated aldehyde moiety to the catalytic aspartates of the HIV-1 protease. Our design is supported by the results of molecular modeling which clearly showed that **5a**, in its hydrated form, could interact with both active site aspartate residues via hydrogen bonding.

The protected *N*-glyoxylyl-tripeptide **4a** was prepared by solid-phase peptide synthesis and the product was isolated and fully characterized. The use of NBS in aqueous medium provided a mild and clean procedure to remove the diisopropylthio protecting group from **4a** to give **5a** in satisfactory yield. In order to assess the effect of a glyoxylyl group at the *N*-terminus of a peptide substrate, *N*-glyoxylyl-oligopeptides (**5b–f**) were synthesized using a similar sequence of reactions. ^1H

NMR analyses of three of these peptides performed in dimethylsulfoxide (DMSO) showed that the integration of the aldehydic proton appearing as a singlet at 9.3 ppm was well below unity (0.5) and even lacking in **5a**. A singlet at ~4.6 ppm characteristic for a hydrated aldehyde group with the complementary integration was, however, present in the three spectra arguing for the presence of the hydrated form in all cases. This result was confirmed from the mass spectroscopy analysis (fast atom bombardment) giving systematically the *m/z* value of the hydrated form of peptide **5**. Finally, the fact that no signal around 9 ppm was detected for **5b** when the ^{1}H NMR was performed in the buffer attests of complete hydratation in the biological conditions.

As expected, compound **5a** behaved as a competitive inhibitor of HIV-1 protease with a K_i value of 0.52 mM (Table 1). The inability of its precursor PIVNH$_2$ to interfere with the enzyme clearly illustrates the necessity of an unmasked aldehyde for the activity. The *N*-glyoxylyl-tetrapeptide **5b** was also an inhibitor for HIV-1 protease, albeit with reduced potency (K_i = 0.70 mM). The longer-chain analogues **5c–f** behaved as substrates and were cleaved at the Phe–Pro junction. These results demonstrate the importance of having the *N*-glyoxylyl group at the P$_1$–P$'_1$ cleavage site of the peptide substrate to be effective as a transition state mimic.

It must be noticed that the affinity of **5a** for the protease (K_i = 0.52 mM) is identical to that of the heptapeptide **5e** (K_i = 0.52 mM). Assuming there is no significant interaction between the *N*-glyoxylyl appendix of **5e** and the protease, the binding contribution of the *N*-glyoxylyl group of **5a** as a transition state structural mimic is equivalent to that of the four amino acid extension. This notion seems to be supported by the K_i of 0.95 mM for **5c**, which peptide chain is two residues longer than **5a**, but two amino acids shorter than **5e**. As anticipated, further chain extension in the case of **5f** led to an increase in affinity (K_i = 0.12 mM). However, it is not obvious why the hexapeptide **5d** should have higher affinity (K_i = 0.09 mM) than **5f**.

While compound **5a** is an inhibitor for HIV-1 protease, its activity is clearly not impressive. It is possible that **5a** is not completely hydrated in the enzyme's active site, thus its hydrogen bonding network is not fully established. Furthermore, if **5a** is not completely hydrated, its α-ketoamide moiety may adopt an orthogonal conformation.[18] The resulting twisted proline amide bond is certainly not a favorable conformation for binding with protease. Another likely scenario is that the enzyme-bound hydrated **5a** exists mainly in the *trans* conformation (**5a**^{hyd}-***trans***), which leads to a weaker interaction with the protease. Further investigations to answer these questions will be pursued after identification of more promising inhibitors of this class.

Although the activity of the designed inhibitor **5a** against the HIV-1 protease is not practically useful, these compounds listed in Table 1 represent the first series of a new class of peptides carrying an *N*-glyoxylyl group. The potential of incorporating this functional group into inhibitor design for other enzymes remains to be explored. It is worth mentioning that a similar strategy had been applied by Walker et al. in their design of peptide-glyoxals as inhibitors for serine and cysteine proteases.[19] Interestingly, the K_i value for chymotrypsin was found to be 10-fold lower than that reported for the corresponding peptide–aldehyde inhibitor.

Experimental

Chemistry

Melting points were determined with an electrothermal melting point apparatus and are uncorrected. ^{1}H NMR and ^{13}C NMR spectra were recorded on a Bruker AM300 spectrometer. Chemical shifts are reported on the δ scale relative to internal standard (tetramethylsilane or appropriate solvent peaks) with coupling constants given in Hertz. The chemical shift, multiplicity (s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; dd, doublet of doublets; and br, broad), relative integral, coupling constant (if applicable), and assignment are given in this order for each discernible NMR signal. MS spectra were recorded on a AEI MS9 spectrometer for chemical ionization (CI), and a Kratos MS80 spectrometer for fast atom bombardment (FAB) analysis. Microanalyses were performed at ICSN, Gif-sur-Yvette, and the accuracy of the data are within 0.3%. High-performance liquid chromatography purification were conducted with a Waters 600E instrument equipped with a Waters 490E UV detector operating at 214 and 254 nm. Flash column chromatography was performed on columns of various diameters with Merck (230–400 mesh) silica gel by elution with the solvents reported. Analytical thin-layer chromatography (TLC) was carried out on Merck silica gel 60 G-254 plates (25 mm). TLC spots were visualized either with UV light or by dipping into the staining solutions of vanillin:ethanol:H_2SO_4 (1:98:1) or phosphomolybdic acid (7% EtOH solution) and then heating. The drying reagent used in the routine work up was anhydrous magnesium sulfate. Solvents, unless otherwise specified, were of analytical reagent grade or the highest quality commercially available.

Diisopropylthioacetic acid. A mixture of glyoxylic acid (148 g, 1 mol, 50% aqueous solution), 2-propanethiol (122 mL, 228 g, 3 mol) and *p*-toluenesulfonic acid (0.5 g) in toluene (500 mL) was refluxed with a Dean–Stark trap until water formation ceased. Toluene was evaporated and the residue was distilled (bp 85 °C/0.2 torr) to afford 181 g (87% yield) of diisopropylthioacetic acid; ^{1}H NMR ($CDCl_3$) δ 10.38 (brs, 1H, COOH), 4.34 (s, 1H, SCHS), 3.18 (m, 2H, *i*Pr, J = 6.7 Hz), 1.33–1.16 (2d, 12H, CH_3, J = 6.7 Hz); ^{13}C NMR ($CDCl_3$) δ 176.35, 48.06, 36.14, 23.34, 23.14. MS (CI) *m/z* 209 (M+1). Anal. calcd for $C_8H_{16}O_2S_2$: C, 46.12; H, 7.74; O,

15.36; S, 30.78. Found: C, 46.11; H, 7.48; O, 15.21; S, 30.57.

Diisopropylthioacetic acid 3,4-dihydro-4-oxo-1,2,3-benzotriazin-3-yl ester (1). Diisopropylthio acetic acid (10.4 g, 50 mmol) was dissoved in THF (300 mL) and the solution cooled to −15 °C. Dicyclohexylcarbodiimide (10.2 g, 50 mmol) was added and the solution was stirred at 15 °C for 5 min. Solid 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine (8.15 g, 50 mmol) was added and the mixture stirred at −10 °C for 30 min and then at 0 °C for 4 h. After standing overnight at 4 °C, precipitated dicyclohexylurea was removed by filtration and the filtrate concentrated in vacuo. The resulting syrup was crystallized from ethylacetate:pentane (15.5 g, 88%) mp 76 °C; ^{1}H NMR ($CDCl_3$) δ 8.40 (d, 1HAr, J = 7.8 Hz), 8.26 (d, 1HAr, J = 8 Hz), 8.02 (t, 1HAr, J = 7.3 Hz), 7.86 (t, 1HAr, J = 7.6 Hz), 4.79 (s, 1H, SCHS), 3.4 (m, 2H *i*Pr, J = 6.8 Hz), 1.42–1.39 (2d, 12H, CH_3, J = 6.8 Hz); ^{13}C NMR ($CDCl_3$) δ 167.50, 150.00, 145.30, 135.59, 132.94, 129.17, 125.85, 122.00, 45.06, 36.77, 23.46, 23.28. MS (CI) *m/z* 354 (M+1). Anal. calcd for $C_{15}H_{19}N_3O_3S_2$: C, 50.97; H, 5.42; N, 11.89; O, 13.58; S, 18.14. Found: C, 50.97; H, 5.14; N, 11.91; O, 13.40; S, 18.42.

Diisopropylthioacetyl-peptide amides (general procedure for the preparation of compounds 4a–f). Standard procedure of solid-phase peptide synthesis using Fmoc strategy were used starting with 4-(α-Fmoc-amino-2′,4′-dimethoxybenzyl)phenoxy acetic acid as a linker.[11] Last acylation was performed using four equivalents of compound **1**. After cleavage from support by TFA:H_2O:ethanedithiol (95:2.5:2.5) these peptides were purified by C18 reversed-phase HPLC.

Diisopropylthioacetyl-Pro-Ile-Val-amide (4a). ^{1}H NMR ($CDCl_3$) δ 8.2 (bs, 1H, NH_2), 7.2 (m, 2H, NH), 6.8 (bs, 1H, NH_2), 4.56 (s, 1H, SCHS), 4.5 (t, 1H, CHα Ile), 4.3 (m, 1H, CHα Val), 4.2 (t, 1H, CHα Pro), 3.8 (m, 2H, CH_2-N Pro), 3.2 (m, 2H, CH *i*Pr), 2.25 (m, 1H, CHβ Val), 2.2–1.9 (m, 5H, CHβ Ile, $2CH_2$ Pro), 1.45 (m, 1H, CH_2 Ile), 1.25 (m, 13H, CH_3 *i*Pr, CH_2 Ile), 0.9 (m, 12H, CH_3). MS (FAB, thioglycerol) *m/z* 539 (M + Na), 517 (M + H).

Diisopropylthioacetyl-Phe-Pro-Ile-Val-amide (4b). ^{1}H NMR (DMSO-d_6) δ 8.14 (d, 1H, NH), 8.05 (d, 1H, NH), 7.7 (d, 1H, NH), 7.3–7.5 (m, 5H, Ph), 7.47 (s, 1H, NH_2), 7.18 (s, 1H, NH_2), 5.44 (s, 1H, SCHS), 4.8 (m, 1H, CHα Phe), 4.53 (m, 1H, CHα Pro), 4.3–4.1 (m, 2H, CHα Val + Ile), 3.7 (m, 2H, CH_2-N Pro), 3.2–2.7 (m, 4H, CH_2 Phe, CH *i*Pr), 2.2–1.8 (m, 6H, CHβ Val, CHβ Ile, $2CH_2$ Pro), 1.55 (m, 1H, CH_2 Ile), 1.25 (m, 13H, CH_3 *i*Pr, CH_2 Ile), 0.9 (m, 12H, CH_3). MS (FAB, thioglycerol) *m/z* 686 (M + Na), 664 (M + H).

Diisopropylthioacetyl-Asn-Phe-Pro-Ile-Val-amide (4c). ^{1}H NMR (DMSO-d_6) δ 8.2 (d, 1H, CONH), 8.05 (d, 1H, CONH), 8.0 (d, 1H, CONH), 7.7 (d, 1H, CONH), 7.5–7.2 (m, 7H, $CONH_2$ Asn+Ph), 7.4 (s 1H, $CONH_2$), 7.18 (s 1H, $CONH_2$), 5.8 (s, 1H, SCHS), 4.8–4.6 (m, 2H, αCH Phe + Asn), 4.53 (m, 1H, αCH Pro), 4.6–4.4 (m, 2H, αCH Val + Ile), 3.7 (m, 2H, CH_2-N Pro), 3.2–2.8 (m, 4H, CH_2 Phe, CH *i*Pr), 2.5 (m, 2H, CH_2 Asn), 2.2–1.8 (m, 6H, βCH Val, βCH Ile, $2CH_2$ Pro), 1.6 (m, 1H, CH_2 Ile), 1.35 (m, 13H, CH_3 *i*Pr, CH_2 Ile), 0.9 (m, 12H, CH_3). MS (FAB, thioglycerol) *m/z* 800 (M + Na), 778 (M + H).

Diisopropylthioacetyl-Gln-Asn-Phe-Pro-Ile-Val-amide (4d). ^{1}H NMR (DMSO) δ 8.3 (d, 1H, CONH), 8.2 (d, 1H, CONH), 8.1 (d, 1H, CONH), 8 (d, 1H, CONH), 7.7 (d, 1H, CONH), 7.5–6.8 (m, 11H, Ph + 3 $CONH_2$), 5.52 (s, 1H, SCHS), 4.8–4.6 (m, 3H, αCH Asn + Phe + Pro), 4.45 (m, 1H, αCH Gln), 4.3–4.2 (m, 2H, αCH Val + Ile), 3.7 (m, 2H, CH_2-N Pro), 3.2–2.8 (m, 4H, CH_2 Phe, CH *i*Pr), 2.5 (m, 2H, CH_2 Asn), 2.3–1.7 (m, 10H, βCH Val, βCH Ile, $2CH_2$ Pro, $2CH_2$ Gln), 1.6 (m, 1H, CH_2 Ile), 1.35 (m, 13H, CH_3 *i*Pr, CH2 Ile), 0.9 (m, 12H, CH_3). MS (FAB, thioglycerol) *m/z* 928 (M + Na), 906 (M + H).

Diisopropylthioacetyl-Ser-Gln-Asn-Phe-Pro-Ile-Val-amide (4e). ^{1}H NMR (DMSO-d_6) δ 8.4 (d, 1H, CONH), 8.3 (d, 1H, CONH), 8.2 (d, 1H, CONH), 8.1 (d, 1H, CONH), 8 (d, 1H, CONH), 7.85 (d, 1H, CONH), 7.6–6.9 (m, 11H, Ph + 3 $CONH_2$), 5.8 (s, 1H, SCHS), 5.25 (m, 1H, OH Ser), 4.7–4.4 (m, 4H, αCH Asn + Phe + Pro + Ser), 4.4–4.2 (m, 3H, αCH Val + Ile + Gln), 3.8–3.5 (m, 4H,CH_2-N Pro + CH_2-O Ser), 3.2–2.8 (m, 4H, CH_2 Phe, CH *i*Pr), 2.6–2.4 (m, 2H, CH_2 Asn), 2.35–1.8 (m, 10H, βCH Val, βCH Ile, $2CH_2$ Pro, $2CH_2$ Gln), 1.6 (m, 1H, CH_2 Ile), 1.35 (m, 13H, CH_3 *i*Pr, CH_2 Ile), 0.9 (m, 12H, CH_3). MS (FAB, thioglycerol) *m/z* 1015 (M+Na), 993 (M+H)

Diisopropylthioacetyl-Val-Ser-Gln-Asn-Phe-Pro-Ile-Val-amide (4f). ^{1}H NMR (DMSO-d_6) δ 8.4 (d, 1H, CONH), 8.3 (d, 1H, CONH), 8.2 (d, 1H, CONH), 8.1 (d, 1H, CONH), 8,0 (d, 1H, CONH), 7.9 (m, 2H, CONH), 7.5–6.8 (m, 11H, Ph + 3 $CONH_2$), 5.6 (s, 1H, SCHS), 4.8–4.2 (m, 8H, αCH Asn + Phe + Pro + Ser + 2Val + Ile + Gln), 3.9–3.3 (m, 5H, CH_2-N Pro + CH_2-O Ser + OH Ser), 3.2–2.8 (m, 4H, CH_2 Phe, CH *i*Pr), 2.6–2.4 (m, 2H, CH_2 Asn), 2.3–1.8 (m, 11H, 2βCH Val, βCH Ile, $2CH_2$ Pro, $2CH_2$ Gln), 1.6 (m, 1H, CH_2 Ile), 1.25 (m, 13H, CH_3 *i*Pr, CH_2 Ile), 0.9 (m, 18H, CH_3). MS (FAB, thioglycerol) *m/z* 1114 (M + Na), 1092 (M + H).

Glyoxylyl-peptides (general procedure for the preparation of compounds 5a–f). A solution of diisopropylthioacetyl peptide (0.1 g) in acetonitrile:water (8:2) or acetic acid:water (8:2) for compounds **4e** and **4f** was treated at 0 °C with solid *N*-bromosuccinimide (1.1 equiv) in one portion. After 1 h at 0 °C solvent was evaporated and glyoxylyl-peptides were purified by reversed-phase HPLC (delta pack C18 column, 2.5 × 30 cm, water:acetonitrile: 0.05% TFA, 20 mL min^{-1}) to afford **5** in 40–60% yield. The purity of each product was checked by re-injecting each sample into an analytical HPLC column. All compounds exhibited a single peak when analyzed under isocratic conditions

(delta pack C18 column, 0.8 × 10 cm, water:acetonitrile: 0.05% TFA, 2 mL min $^{-1}$).

Glyoxylyl-Pro-Ile-Val-amide (5a). ^{1}H NMR (DMSO-d_6) δ 8.4 (d, 1H, CONH), 8.2 (m, 3H, CONH, $CONH_2$), 4.6 (s, 1H, $CH(OH)_2$), 4.42 (t, 1H, CHα-Ile), 4.2 (d, 1H, CHα-Val), 4.0 (t, 1H, αCH Pro), 3.8 (m, 2H, CH_2-N Pro), 2.2 (m, 2H, CH_2CH Pro), 2.06–1.8 (m, 4H, βCH Val, βCH Ile, γCH_2 Pro), 1.6–1.4 (m, 1H, CH_2 Ile), 1.3–1.1 (m, 1H, CH_2 Ile), 1.1–0.8 (m, 12H, CH_3). MS (CI, isobutane), *m/z* 383 (M + H), 365 (M-NH_3), MS (FAB, thioglycerol), *m/z* 660 (M + Na+thioglycerol-H_2O), 638 (MH + thioglycerol-H_2O), 423 (M + Na, $CH(OH)_2$), 405 (M + Na, CHO), 401 (M + H, $CH(OH)_2$).

Glyoxylyl-Phe-Pro-Ile-Val-amide (5b). ^{1}H NMR (DMSO-d_6) δ 9.3 (s, 0.5H, CHO), 890 (d, 0.5H, CONH Phe), 8.14 (d, 1H, CONH), 7.9 (d, 0.5H, CONH Phe), 7.7 (d, 1H, CONH), 7.4–7.2 (m, 7H, Ph, $CONH_2$), 4.8 (m, 1H, αCH Phe), 4.53 (m, 1H, αCH Pro), 4.5 (s, 0.5H, $CH(OH)_2$), 4.4–4.2 (m, 2H, αCH Val + Ile), 3.9–3.6 (m, 2H, CH_2-N Pro), 3.2–2.7 (m, 2H, CH_2 Phe), 2.2–1.8 (m, 6H, βCH Val, βCH Ile, $2CH_2$ Pro), 1.6 (m, 1H, CH_2 Ile), 1.2 (m, 1H, CH_2 Ile), 0.9 (m, 12H, CH_3). MS (FAB, thioglycerol), *m/z* 660 (M + Na + thioglycerol-H_2O), 638 (MH + thioglycerol-H_2O) 570 (M + Na, $CH(OH)_2$), 548 (M + H, $CH(OH)_2$).

Glyoxylyl-Asn-Phe-Pro-Ile-Val-amide (5c). ^{1}H NMR (DMSO-d_6) δ 9.2 (s, 0.5H, CHO), 8.2–8 (2d, 2H, CONH), 8.05–7.7 (2d, 2H, CONH), 7.5–7.2 (m, 7H, $CONH_2$ Asn + Ph), 7.4 and 7.18 (2s 2H, $CONH_2$), 5.1 (s, 0.5H, $CH(OH)_2$), 4.8–4.6 (m, 2H, αCH Phe + Asn), 4.53 (m, 1H, αCH Pro), 4.6–4.4 (m, 2H, αCH Val + Ile), 3.8–3.4 (m, 2H, CH_2-N Pro), 3.2–2.8 (m, 2H, CH_2 Phe), 2.7–2.3 (m, 2H, CH_2 Asn), 2.2–1.8 (m, 6H, βCH Val, βCH Ile, $2CH_2$ Pro), 1.7–1.5 (m, 1H, CH_2 Ile), 1.4–1.1 (m, 1H, CH_2 Ile), 1.1–0.8 (m, 12H). MS (FAB, thioglycerol) *m/z* 684 (M + Na, $CH(OH)_2$), 662 (M + H, $CH(OH)_2$).

Glyoxylyl-Gln-Asn-Phe-Pro-Ile-Val-amide (5d). MS (FAB, thioglycerol) *m/z* 812 (M + Na, $CH(CHOH)_2$).

Glyoxylyl-Ser-Gln-Asn-Phe-Pro-Ile-Val-amide (5e). MS (FAB, thioglycerol) *m/z* 899 (M + Na, $CH(CHOH)_2$).

Glyoxylyl-Val-Ser-Gln-Asn-Phe-Pro-Ile-Val-amide (5f). MS (FAB, thioglycerol) *m/z* 958 (M + H, CHO), 976 (M + H, $CH(OH)_2$), 998 (M + Na, $CH(OH)_2$).

Biochemistry

Recombinant HIV protease from *Escherichia coli*[20] was used. The protein concentration of the enzyme solution was measured by total amino acid analysis on a PicoTag apparatus (Waters Millipore, Milford, MA). Kinetic measurements were performed with an Uvikon 941 spectrophotometer equipped with a thermostated cell holder. All chromatographic analyses were conducted with a Waters 600 MS–HPLC apparatus.

HIV protease activity towards V-S-Q-N-FNO_2-P-I-V was measured spectrophotometrically at 307 nm and 30 °C. The chromogenic substrate was first dissolved in 1 M acetic acid. After sonication, water was added (acetic acid final concentration: 0.17 M). In order to determine the kinetic parameters of the enzymic hydrolysis, the concentration of the chromogenic substrate varied from 31.2 to 500 μM in 0.1 M acetate, 1 M NaCl, 5 mM EDTA (final pH 4.5). The reaction was initiated by the addition of a freshly prepared enzyme solution (1 mM sodium phosphate containing 1 mg mL^{-1} bovine serum albumin at pH 8.0). The experimental data for initial rates v were fitted to the Michaelis equation to determine the kinetic parameters (nonlinear regression analysis using the software package Kaleidagraph 2.1.3, Abelbeck Software, Reading, PA). The concentration of the chromogenic product liberated during the enzymatic hydrolysis was deduced from the absorbance at 307 nm of V-S-Q-N-FNO_2 and V-S-Q-N-FNO_2-P-I-V in the incubation mixture using the molar absorption coefficient of 1212 M^{-1} cm^{-1}.

The compounds being tested were dissolved either in DMSO (**5a–c**) or in DMF (**5d–f**). For the determination of K_i constants, the experimental conditions were: [protease] = 0.24 μM; [V-S-Q-N-FNO_2-P-I-V] = 31.2–500 μM; [**5a**] = 0.164–1.32 mM; [**5d**] = 13.2–200 μM; [**5e**] = 0.13–1 mM; [**5f**] = 50–200 μM in 0.1 M acetate buffer containing 1 M NaCl and 5 mM EDTA (final pH 4.5) at 30 °C. The double reciprocal plots of rate vs substrate concentration gave straight lines with a slope of $(K_m/V_m) \times (1+[I]_0/K_i)$. The constant ($K_i$) for each inhibitor was obtained as the *x* intercept from the linear plot of the slopes vs inhibitor concentration. The IC_{50} determinations were performed in duplicate using: [protease] = 0.24 μM; [V-S-Q-N-FNO_2-P-I-V] = 280 μM; [**5b**] = [**5c**] = 0.5–2 mM. An approximate value of K_i was estimated from the above IC_{50} using the equation of Cheng and Prussov.[21]

The sites of the enzyme-induced cleavages of *N*-glyoxylyl peptides **5** (0.15–1 mM) and of V-S-Q-N-FNO_2-P-I-V (280 μM) were determined after 1 h incubation with protease (0.24 μM) at 25 °C, followed by 8 h incubation at 4 °C. After filtration on microspins (CML), the enzyme digest was analyzed by reversed-phase HPLC (C8 column, 5 μm, 250 × 4.6 mm, Interchrom) using a linear gradient from 0 to 60% of solvent B (acetonitrile:0.07% TFA) in solvent A (water:0.1% TFA) in 60 min. The flow rate was 0.75 mL min^{-1} and the effluent was scanned in the range of 200–400 nm. The data were analyzed using the Millenium software. Only two peaks were detected in each run. The two fragments resulting from cleavage of compounds **5** were identified by comparison with authentic samples of V-S-Q-N-FNO_2 and PIV and their identity confirmed by mass spectroscopy. The *N*-glyoxylyl peptides and the chromogenic substrate were checked under the same experimental conditions.

Modeling studies

All the computations were performed on a Silicon Graphics station Iris4D-35 and graphics were displayed with an Evans and Sutherland PS390 using SYBYL software.[22] Two accessory programs were written: one to arrange conformations into families after a systematic conformational analysis and another to give information on hydrogen bonds; they were created with Microsoft Excel[25] and SYBYL languages, respectively.

The atomic coordinates of the crystal structure of synthetic HIV-1 protease (SF2 isolate sequence) complexed with the inhibitor Ac-Ser-Leu-Asn-Phe ψ[CH(OH)CH_2N]-Pro-Ile-Val-OMe called JG365[10,24] were retrieved from the Protein Data Bank[25] (file 7hvp). Residues were numbered from 1 to 99 in each subunit of the protease and were also labeled with the prefix A and B by the crystallographers to indicate the first and second monomer, respectively. The residues 67 and 95, which were Aba (L-α-amino-*n*-butyric acid) in the X-ray protein structure, were replaced by Cys, the amino acid in the wild-type protein. Water molecules were omitted from the starting model except the structural water molecule (Wat1) involved in hydrogen bonds with JG365 and the enzyme residues A/Ile50 and B/Ile50 of the flaps. All hydrogen atoms were added. The coordinates were refined using the AMBER all-atom force field, which takes all explicit hydrogens into account.[26] For nonstandard residues in the inhibitor (Pheψ[CH(OH)CH_2N]-Pro) and one of the essential aspartic acids in the catalytic diad, which was protonated in all simulations[27-29] (protonation state B/Asp25.Oδ1), we used the electrostatic charges calculated by Ferguson et al.[27] A cut-off of 8 Å was used for nonbonded interactions and a constant dielectric of 4 was applied. The energy of the system was minimized without constraints by a conjugate gradient method to a minimum energy change of 0.05 kcal mol^{-1}.

CHOCO-Pro and $CH(OH)_2CO$-Pro structures were built with the standard parameters of SYBYL and were used as models for the *N*-terminal structures of **5a** and **5a^{hyd}**, respectively. The glyoxal function of **5a** was modeled under its most stable configuration, that is, *trans* (O=C–C=O torsion angle value was 180°). Missing force-field parameters were chosen by analogy with existing ones in AMBER or other force fields. Atomic charges of the models were calculated with MOPAC[30] using the AM1 method and the keyword MMOK.

5a and **5a^{hyd}** were modeled in the active site of HIV-1 protease starting from the structure of JG365. The first step consisted of the replacement of the Ac-Ser-Leu-Asn-Pheψ[CH(OH)CH_2] moiety of JG365 by the CHOCO or $CH(OH)_2CO$ groups for compounds **5a** or **5a^{hyd}**, respectively. The second step was the replacement of the methoxy *C*-terminal group of JG365 by NH_2. Four possible inhibitor structures (**5a-*cis***, **5a-*trans***, **5a^{hyd}-*cis***, and **5a^{hyd}-*trans***) were attempted since the configuration of the amide bond formed between the *N*-terminal group of **5a** or **5a^{hyd}**, and the proline residue may be *cis* (ω = 0°) or *trans* (ω = 180°).

Systematic conformational analyses on the complex of **5a^{hyd}-cis** or **5a^{hyd}-*trans*** with the HIV-1 protease were performed with a 360° rotation of three rotatable bonds using a 60° increment (Fig. 2), and a van der Waals radii scaling factor of 0.63. For each conformation, the possibilities of creation of hydrogen bonds involving the *gem*-diol group of **5a^{hyd}** were explored and the energy was calculated. Conformations were further classified into families using the criteria of continuity for rotamer angle values with a 60° resolution corresponding to the increment value used in the search. For example, a newly generated conformation was added to family *n* if, for each rotamer, the absolute difference of its angle value with that of at least one conformation of family *n* is equal to zero or 60°. In each family, the lowest energy conformer was extracted and minimized with a calculation protocol similar to that used for the complex formed between JG365 and the protease.

Acknowledgments

We are grateful to the Fondation pour la Recherche Médicale (programme Sidaction) for a research grant-in-aid and for financial support to I.V. We thank Hoechst for a generous gift of glyoxylic acid solution

References

1. Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K. J.; Starcich, B.; Josephs, S. F.; Doran, E. R.; Rafalski, J. A.; Whitehorn, E. A.; Baumeister, K.; Ivanoff, L.; Petteway, Jr S. R.; Pearson, M. L.; Lautenberger, L. A.; Papas, T.; Ghrayeb, J.; Chang, N. T.; Gallo, R. C.; Wong-Staal, F. *Nature (London)* **1985**, *313*, 277.

2. Mous, J.; Heimer, E. P.; Le Grice, S. F. *J. Virology*, **1988**, *62*, 1433.

3. (a) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686. (b) McQuade, T. K.; Tomasselli, A. G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T. K.; Heinrikson, R. L.; Tarpley, W. G. *Science* **1990**, *247*, 454.

4. (a) Kempf, D.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentge, C. A.; Green, B. G.; Fino, L.; Park, C. H.; Kong, X.-P.; Wideburg, N. E.; Saldivar, A.; Ruiz, L.; Kati, W. M.; Sham, H. L.; Robins, T.; Stewart, K. D.; Hsu, A.; Plattner, J. J.; Leonard, J. M.; Norbeck, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2484. (b) Vacca, J. P.; Dorsey-Schlief, W. A.; Levin, R. B.; McDaniel, S. L.; Darke, P. L.; Zugay, J.; Quintero, J. C.; Blahy, O. M.; Roth, E.; Sardana, V. V.; Schlabach, A. J.; Graham, P. I.; Condra, J. H.; Gotlib, L.; Holloway, M. K.; Lin, J.; Chen, I.-W.; Ostovic, D.; Anderson, P. S.; Emini, E. A.; Huff, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4096. (c) Kitchen, V. S.; Skinner, C.; Ariyoshi, K.; Lane, E. A.; Duncan, I. B.; Burckhardt, J.; Burger, H. U.; Bragman, K.; Pinching, A. J.; Weber, J. N. *Lancet* **1995**, *345*, 952.

5. Wlodawer, A.; Erickson, J. W. *Ann. Rev. Biochem.* **1993**, *62*, 543.

6. Rodriguez, E. J.; Angeles, T. S.; Meek, T. D. *Biochemistry* **1993**, *32*, 12380.

7. Darke, P. L.; Huff, J. R. *Adv. Pharmacol.* **1994**, *25*, 399.

8. Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157.

9. Chastrette, F.; Bracoud, C.; Chastrette, M.; Mattioda, G.; Christidis, Y. *Bull. Soc. Chim. Fr.* **1985**, *II*, 66.

10. Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805.

11. Atherton, E.; Sheppard, R. C. In *Solid Phase Peptide Synthesis*; IRL: 1989.

12. Qasmi, D.; René, L.; Badet, B. *Tetrahedron Lett.* **1994**, *35*, 4343.

13. Wolfenden, R. *Annu. Rev. Biophys. Bioeng.* **1976**, 271. Abeles, R. H. *Drug Develop. Res.* **1987**, *10*, 221.

14. Fehrentz, J.-A.; Heitz, A.; Castro, B.; Cazaubon, C.; Nisato, D. *FEBS Lett.* **1984**, *167*, 273.

15. Sarubbi, E.; Seneci, P. F.; Angelastro, M. R.; Peet, N. P.; Denaro, M.; Islam, K. *FEBS Lett.* **1993**, *319*, 253.

16. Ocain, T.; Rich, D. H. *J. Med. Chem.* **1992**, *35*, 451.

17. Cornish-Bowden, A. In *Fundamentals of Enzyme Kinetics*; Portland, 1995.

18. (a) Liu, J.; Albers, M. W.; Chen, C. M.; Schreiber, S. L.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2304. (b) Rosen, M. K.; Standaert, R. F.; Galat, A.; Nakatsuka, M.; Schreiber, S. L. *Science* **1990**, *248*, 863.

19. Walker, B.; McCarthy, N.; Healy, A.; Ye, T.; McKervey, M. A. *Biochem. J.* **1993**, *293*, 321.

20. Hirel, P. H.; Parker, F.; Boiziau, J.; Jung, G.; Outerovitch, D.; Dugué, A.; Peltiers, C.; Giuliacci, C.; Boulay, R.; Lelièvre, Y.; Cambou, B.; Mayaux, J.-F.; Cartwright, T. *Antiviral Chem. Chemother.* **1990**, *1*, 9.

21. Cheng, Y. C.; Prussov, W. H. *Biochem. Pharmacol.* **1973**, *22*, 3099.

22. SYBYL 5.5, Tripos Associates, St Louis, Missouri.

23. Microsoft Excel 3.0, Microsoft Corporation.

24. Rich, D. H.; Green, J.; Toth, M. V.; Marshall, G. R.; Kent, S. B. H. *J. Med. Chem.* **1990**, *33*, 1285.

25. Bernstein, F. C.; Koetzel, T. F.; Williams, G. J. G.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.*, **1977**, *112*, 535.

26. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.

27. Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. *J. Med. Chem.* **1991**, *34*, 2654.

28. Trospha, A.; Hermans, J. *Protein Eng.* **1992**, *5*, 29.

29. Chen, X.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 42.

30. Stewart, J. J. P. *J. Comput. Aided Mol. Des.* **1990**, *4*, 1.

(*Received in U.S.A.* 14 *November* 1996; *accepted* 31 *December* 1996)