PROTEINS: Structure, Function, and Genetics 10:325–339 (1991)

Comparative Analysis of the Sequences and Structures of HIV-1 and HIV-2 Proteases

Alla Gustchina and Irene T. Weber
Macromolecular Structure Laboratory, NCI Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, Maryland 21702

ABSTRACT **The different isolates available for HIV-1 and HIV-2 were compared for the region of the protease (PR) sequence, and the variations in amino acids were analyzed with respect to the crystal structure of HIV-1 PR with inhibitor. Based on the extensive homology (39 identical out of 99 residues), models were built of the HIV-2 PR complexed with two different aspartic protease inhibitors, acetylpepstatin and a renin inhibitor, H-261. Comparison of the HIV-1 PR crystal structure and the HIV-2 PR model structure and the analysis of the changes found in different isolates showed that correlated substitutions occur in the hydrophobic interior of the molecule and at surface residues involved in ionic or hydrogen bond interactions. The substrate binding residues of HIV-1 and HIV-2 PRs show conservative substitutions of four residues. The difference in affinity of HIV-1 and HIV-2 PRs for the two inhibitors appears to be due in part to the change of Val 32 in HIV-1 PR to Ile in HIV-2 PR.**

Key words: retroviral proteases, aspartic proteases, HIV, protease inhibitors, sequence analysis

INTRODUCTION

The acquired immunodeficiency syndrome (AIDS) is caused by human immunodeficiency virus type 1 (HIV-1), which has been the subject of very extensive studies. More recently, two similar viruses infecting primates and humans were discovered by a serological study of a healthy West African population.[1] These are Simian immunodeficiency virus (SIV) and human immunodeficiency virus type 2, (HIV-2) and are structurally and genetically related to HIV-1. HIV types 1 and 2 share a similar genomic organization, indicating a common evolutionary origin, but differ significantly in terms of nucleotide and amino-acid sequence: the more conserved *gag* and *pol* genes, respectively, display only 56 and 60% nucleotide sequence homology and less than 60% amino-acid identity.[2] HIV-2 is more closely related to the SIV than it is to HIV-1.[3] The *gag*, *pol*, and *env* proteins of SIV and HIV-2 are antigenically cross-reactive, whereas their cross-reactivity to HIV-1 is restricted to some *gag* and *pol* antigens.[4] In order to develop drugs active against both HIV-1 and HIV-2, it is important to analyze the similarities and differences between their proteins.

One virally encoded enzyme that is essential for viral maturation is the retroviral aspartyl protease (PR).[5] The three-dimensional structures are now available for the native HIV-1 PR[6–8] and for several complexes with inhibitors.[9–11] The monomer of HIV-1 PR contains the sequence, Asp-Thr-Gly, which is characteristic of the active sites of nonviral aspartic proteases, and the structural core resembles one of the two domains of the pepsinlike enzymes. The active site of the retroviral protease is formed by residues from both subunits in the PR dimer, whereas in pepsin the active site lies between two domains in the monomeric enzyme. This is in agreement with the results of biochemical studies of the behavior of active forms of avian myeloblastosis virus (AMV) and HIV proteases in gel filtration.[12,13]

A number of different isolates of both HIV-1 and HIV-2 are currently available. They show variations in amino acid sequences for the HIV proteins coded by their genes that are presumably due to the reduced fidelity of transcription by the reverse transcriptase. The crystal structures of HIV-1 PR provide a guide to interpreting these variations. Here we have examined the substitutions in the amino acid sequences of HIV-1 and -2 PR from different isolates, modeled the structure of HIV-2 PR with inhibitor, and analyzed the changes in amino acid residues in terms of the conserved three-dimensional structure.

EXPERIMENTAL PROCEDURES

The HIV PR sequences corresponding to both the HIV-1 and HIV-2 isolates were obtained from the

Received August 23, 1990; revision accepted January 16, 1991.

Address reprint requests to Irene T. Weber at her new address: Department of Pharmacology, Life Sciences Bldg., Thomas Jefferson University, Philadelphia, PA 19107.

A. Gustchina is on leave from the V.A. Engelhardt Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow, USSR.

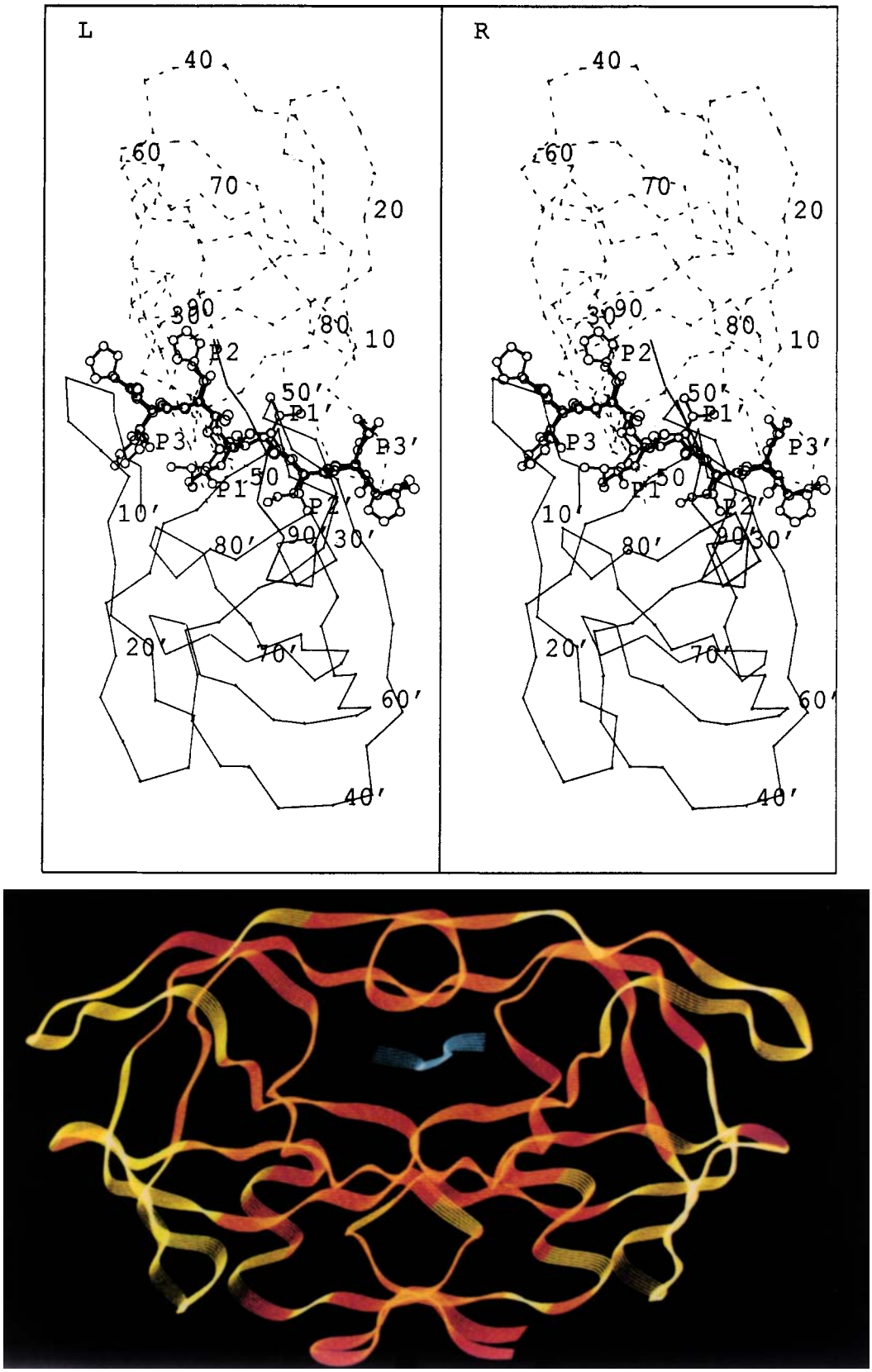


Fig. 1. Legend appears on page 327.

Los Alamos HIV Sequence Database.[14] Duplicate sequences were removed and the unique sequences were aligned with a program written by Hein.[15]

The atomic coordinates of the crystal structure of the complex of HIV-1 PR with inhibitor JG-365 (Ac-Ser-Leu-Asn-Phe-ψ[CH(OH)CH_2N]-Pro-Ile-Val-OMe)[10] were taken as a basis for obtaining the atomic coordinates of HIV-2 PR with the sequence of *rod* isolate. The program FRODO[16] run on an Evans and Sutherland PS390 molecular graphics system was used for the modeling. Amino acid side chains that were different in HIV-1 and -2 PR were built in similar conformations to those in the crystal structure and positioned to avoid close contacts with neighboring residues. During the modeling process it was observed that a change from a small amino acid side chain to a larger one was generally accommodated by nearby changes to smaller residues. The conformation of the main chain was changed only for positions where Pro residues, or Gly residues with positive ϕ torsion angles, were replaced by another type of residue. No insertions or deletions were necessary.

In the crystal structure of HIV-1 PR with inhibitor JG365,[10] the two subunits show deviations from twofold symmetry. However, there is a crystal structure of HIV-1 protease with symmetric inhibitor with almost exact twofold symmetry.[11] Our model for HIV-2 PR is a symmetrical dimer of two subunits related by a rotation of about 178°. The atomic coordinates of the second monomer in the dimer were obtained from the first one by a superposition on the C_α atoms of the second monomer in the crystal structure of HIV PR with inhibitor. The first subunit consists of residues 1 to 99, and the second is numbered 1′ to 99′. The root-mean-square (rms) deviation of all 99 C_α atoms of the first subunit compared to the crystal structure is 0.20 Å, whereas the second subunit, generated by symmetry, has an rms deviation of 0.59 Å. The second monomer in this symmetrical dimer was compared with that in the crystal structure of HIV-1 PR with inhibitor.[10] There were deviations from symmetry for residues that were involved in inhibitor-enzyme or crystal packing contacts. The largest deviations were observed in the residues 14–16, which are involved in crystal packing contacts, and around residue 81, which is near the inhibitor. The symmetrical positions were kept in almost all cases, since there were no undesirable contacts. One adjustment from symmetry was made in the model dimer. The side chain of Ile 50′ on the flap was moved slightly, and the carbonyl oxygen was positioned to form a hydrogen bond interaction with NH of Gly 51, exactly as observed in the crystal structures of HIV-1 PR with inhibitors.

The HIV-2 PR was modeled with two different inhibitors, acetyl-pepstatin (Ac-Val-Val-Sta-Ala-Sta, where Sta is the rare amino acid, statine, (4S,3S)-4-amino-3-hydroxyl-6 methylheptanoic acid), and the renin inhibitor, H-261 (tBoc-His-Pro-Phe-His-Leuψ[CHOH-CH_2]-Val-Ile-His) (Fig. 1a). Since the conformations of the inhibitors are very similar in all the available HIV-1 PR crystal structures, the inhibitors were modeled by altering residues in the cocrystal structures.[9,10] Small rotations of side chains were made when necessary to relieve unfavorably close contacts between atoms. The side chains were positioned to overlap with those in the different cocrystal structures of HIV-1 PR and inhibitors. Since only small alterations were made, no energy minimization was performed and only a qualitative comparison has been made.

RESULTS AND DISCUSSION

Comparison of HIV PR Sequences From Different Isolates

The amino acid sequences of PR from different isolates of HIV-1 and -2 are aligned in Table I, and the variability in amino acid sequences is summarized in Table II and Figure 1b. The level of similarity within each type of HIV is extremely high—if the conservative substitutions are included it corresponds to 91 similar residues within HIV-1 isolates and to 82 in HIV-2 isolates, out of a total of 99 residues. The similarity between the two types of virus approaches 62% if conservative changes are taken into account. The HIV-1 isolates show 10 conserved basic and 8 conserved acidic residues, whereas the HIV-2 PR sequences have 9 conserved basic and 8 conserved acidic residues. Five of the basic residues are conserved between types 1 and 2. These are Arg 8, Lys 45, Lys/Arg 57, Lys/Arg 70, and Arg 87. Four of the acidic residues are conserved between the isolates of HIV-1 and -2 PR. These are Glu 21, Asp 25, Asp 29, and Asp 30. Asp 25 and 29 interact with the inhibitor, and Arg 8′, Asp 29, and Arg 87 form a set of ionic interactions that is also present in the Rous sarcoma virus (RSV) PR dimer.[17] Other such conserved interactions are described later.

To provide a basis for the discussion of the similarity and variability in the sequences of the different isolates and their correlations with the structure of the enzyme, the major structural elements are summarized below. HIV-1 PR shares a structural fold with all other known aspartic proteases. The

Fig. 1. **(a) Top:** Stereo view of the modeled structure of the HIV-2 PR dimer with two modeled inhibitors. The C_α atoms of the two subunits are shown in dashed and continuous lines, and labeled every 10 residues. The inhibitors are in a ball and stick representation. The renin inhibitor, H-261, is in thin lines, and the acetyl pepstatin is in thicker lines. The inhibitor residues P3 to P3′ are indicated. **(b) Bottom:** A chain tracing of the dimer of HIV-2 PR in a different orientation to Figure 1a. Red indicates residues that are conserved among all isolates of HIV-1 and HIV-2, and yellow indicates variable residues. The peptide inhibitor, shown in blue, lies below the two flaps and above catalytic Asp 25.

TABLE I. Variations in Protease Sequences: HIV-1 PR and HIV2-PR*

```
hivl: .............K......................S...R.....................LI...................................  bh5
      ...........A.K.................................................I....Q..............................  eli
      .............K.....................I..........................LI....Y..M...........................  pondk
      ..V..........K......................S...R.....................LI...................................  hxb2
      .........I...K..........................R.....................LI............................L......  oyi
      .........I...K...................G.....RR.....................TIG...........................L......  mn
      .............K................................................LI...................................  z2z6
      .........I..VK................................................LI...................................  rf
      .........V..V.V....................I..........................LI....K.....I.............M..........  mal
      .............K..........................R...............G.....LI...................................  n143
      PQITLWQRPLVTIRIGGQLKEALLDTGADDTVLEEMNLPGKWKPKMIGGIGGFIKVRQYDQIPVEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF  sf2
      **..**.**.**..*.**..*.********.......*.....**..*******....*.....*....*...*...*.**.**.***.**..*..**.
hiv2: PQFSLWKRPVVTAYIEGQPVEVLLDTGADDSIVAGIELGNNYSPKIVGGIGGFINTKEYKNVEIEVLNKKVRATIMTGDTPINIFGRNILTALGMSLNL  rod
      ................D......................S....................D...R..................................  isy
      .............H............R.N..........S...........................G.R.K................V..........  nihz, fg
      ................D................T.....D..T.....................K....R.............................  ben
      .............F..D......................D..T.....................K....R....................AT.......  d194
      ......R....K.C....S........V...........S..T.................D.....VG.R....................NT...T..F  d205
      .............HV....................V...S........................R....R.............................  st
      ........10........20........30........40........50........60........70........80........90.........
      | N |  |     A1     ||   B1   |         |     flap     ||     A2      ||    B2    ||α-helix||  C |
```

*Protease sequences from different isolates of HIV-1 are shown on the top half, and from HIV-2 isolates on the lower half of the table. The single letter amino acid code is used, and only differences from the sequence of the HIV-1 PR crystal structure (isolate *sf2*) or the modeled HIV-2 PR (isolate *rod*), shown in the center, are indicated. Asterisks indicate identical residues between the *sf2* isolate of HIV-1 and the *rod* isolate of HIV-2. Structural elements are indicated at the bottom. N: N-terminal β strand; C: C-terminal strand; A1 and A2: antiparallel β strands forming loops A1 and A2 of the A sheet; B1 and B2 : antiparallel β strands forming loops B1 and B2 of the B sheet. The flap and helix are indicated. Residue R27 in HIV-2 *nihz, fg* was considered suspicious due to the presence of G at this position in all other aspartic proteases and was omitted from the analysis in Table II.

TABLE II. Summary of Variability of PR Sequences in HIV Isolates*

		Substitutions in sequences	
Virus type	Identical residues	Conservative	Variable
HIV-1 isolates	79	12	8
	39	23	37
HIV-2 isolates	73	9	17

*PR sequences for the isolates shown in Table 1 are compared to give the numbers of identical, similar, or variable amino acid residues both within the two types, and between HIV-1 and HIV-2. The amino acid types considered to be conservative substitutions are: V/L/I/A/M, F/Y/W, R/K, D/E, N/Q, T/S.

structure of one subunit of HIV-1 PR is shown in Figure 2a. The central part of the HIV PR monomer is formed from two layers of β-sheets with approximately orthogonal directions of the β-strands, and the distribution of hydrophobic residues in the hydrophobic nucleus is in accordance with the structural template for aspartic proteases.[18] The first β-sheet, called by Andreeva[19] the A-layer, contains two antiparallel β-hairpins, which include residues 9–22 and 59–75 (Fig. 2b). The second, B-layer, is composed of two wide loops that cross to form an antiparallel system of two pairs of parallel strands (residues 23–33 and 76–86). This is equivalent to β-strands b-c, b′-c′ (the A-sheet), and the strands c-d, c′-d′ (the B-sheet) in the terminology of Blundell.[20] This conserved core structure resembles one domain of the pepsinlike enzymes. Each B-layer in the dimer provides one catalytic Asp 25, and forms part of the substrate binding pocket. The two loops forming the A-sheet are not involved in contacts to the inhibitor. These loops can have different lengths in other PRs, for example, they are extended by 5 to 10 residues in the structure of RSV PR compared to HIV-1 PR.

Another important structural element consists of the N- and C-termini (residues 1–5 and 95–99).

Fig. 2. **(a) Top:** The C_α backbone of one subunit from the crystal structure of HIV-1 PR is shown in a stereo view. Every tenth residue is numbered. Residues 9 to 22 and 59 to 75 form the A-layer, and residues 23 to 33 and 76 to 86 form the B-layer. The flap consists of residues 42 to 57. Other structural elements are indicated in Table I. **(b) Bottom:** The major structural elements in one subunit of HIV-1 PR are shown in different colors. The two β-sheets consist of the A-layer in yellow (loops A1 and A2) and the B-layer in dark blue (wide loops B1 and B2). The flap is in purple, the short helix in green, and the catalytic Asp25 is red.

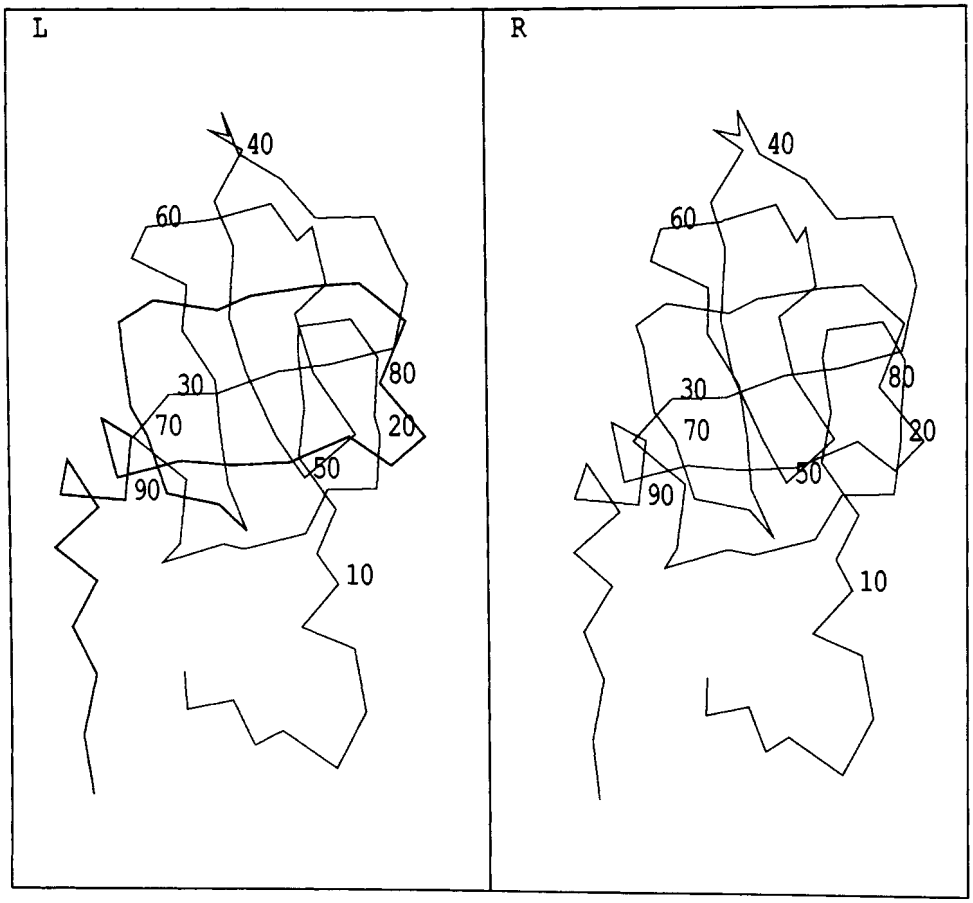
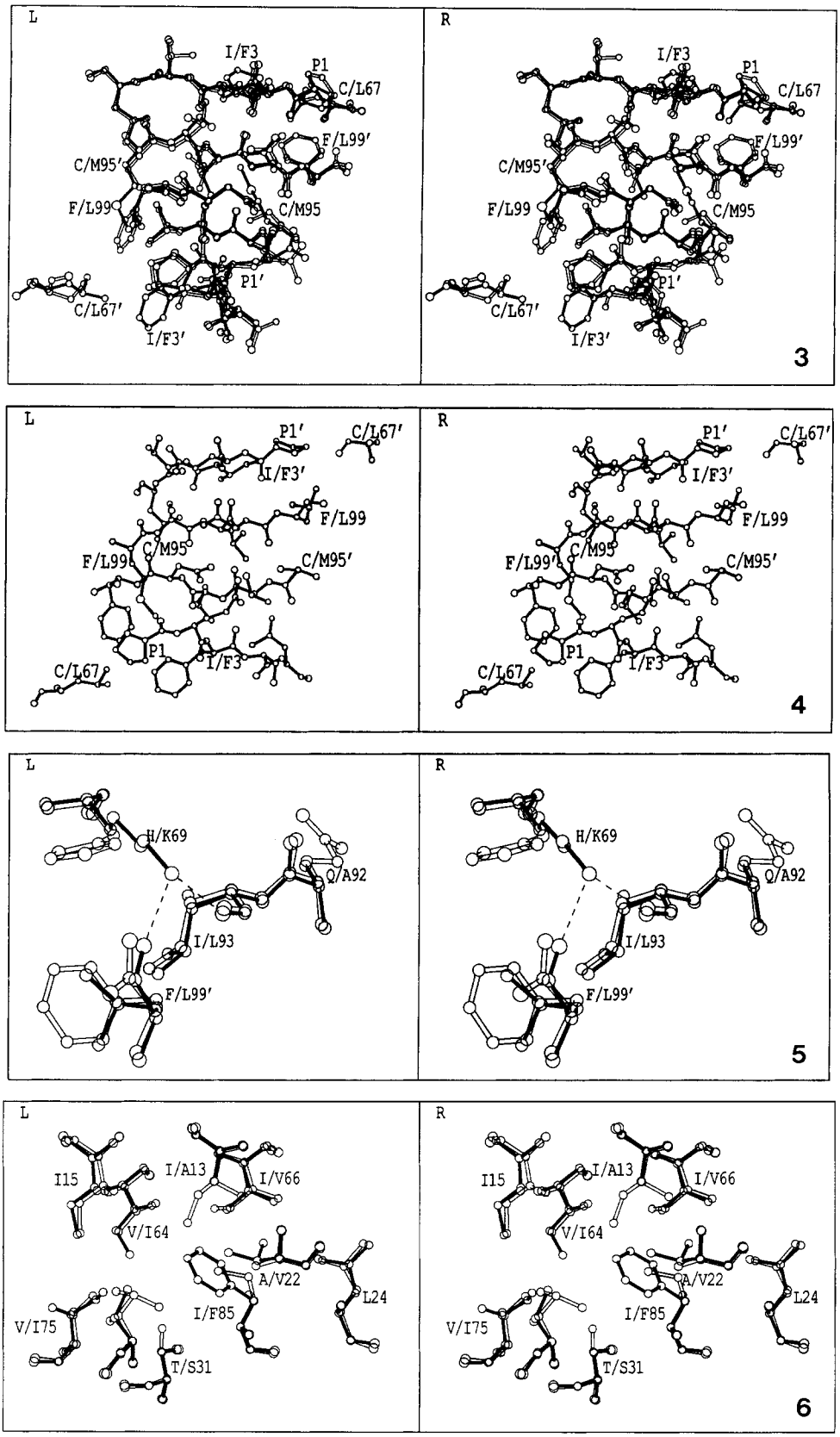


Fig. 2. Legend appears on page 328.



Figs. 3–6. Legends appear on page 331.

These form a four-stranded antiparallel β-sheet, which provides many intersubunit interactions in the dimer[17] and is adjacent to the active site. The flexible surface β-hairpin called the "flap" (residues 42–58) has a structural and functional analog in the nonviral aspartic proteases (residues 70–83 in pepsin numbering). Therefore, the major structural elements of the PR structure consist of the double layer of A- and B-sheets, the flap, the C-terminal α-helix and the 4-stranded β-sheet formed by the termini.

All the important structural elements have highly conserved sequences (Table I, Fig. 1b). The most conserved region is a segment containing the Asp-Thr-Gly sequence that is common to all aspartic proteases. This is part of the B1-loop and forms the active site by dimerization of two subunits (Fig. 2b). The B2-loop and the adjacent helical region, residues 86–93, are also highly conserved, due to restrictions imposed by the three-dimensional structure. The residues in these regions provide numerous intra- and intermolecular interactions in the active dimer and also enzyme-substrate interactions. Residues 87–89, Gly-Arg-Asn/Asp, at the N-terminus of the helix are conserved in the sequences of all known retroviral PRs and are involved in hydrogen bond and ionic interactions with a region near the catalytic triplet. The helix dipole of both helices in the PR dimer lies with the positively charged end near the two catalytic aspartic acids and this may help stabilize the aspartic acids during catalysis.

The other highly conserved sequence in different isolates of HIV-1 and HIV-2 PRs occurs in the flap region (Fig. 2b). The flap is a flexible structural unit in all aspartic proteases, and the residues at the tip of the flap are involved in enzyme-substrate interactions. All isolates of HIV-1 and HIV-2 PR have a conserved sequence at the tip of the flap (Table I). This sequence, -GGIGGFI-, involves four Gly residues that are very important for providing flexibility. In comparison, the single flap in nonviral aspartic proteases contains only three conserved residues at its tip: Tyr 75, Gly 76 (which is replaced by Ser only in human kidney renin), and Gly 78 (in pepsin numbering). A difference in position of the flap of about 2 Å has been observed on comparing the structures of pepsinlike enzymes with and without inhibitors.[21–23] The dimeric molecule of HIV-1 PR differs from the cellular proteases due to the presence of two flaps, as opposed to the single flap in pepsinlike enzymes. The tips of the two flaps in HIV-1 PR almost overlap, thus closing the active site so that both the binding of a substrate or inhibitor to the enzyme and the release of products require a substantial movement of both flaps.[24]

Residues from the N-terminus and the C-terminus are conserved within the two types of virus due to the formation of the four-stranded antiparallel β-sheet from both subunits in the PR dimer. The A1 and A2 loops have more variation in amino acid residues since they are on the protein surface and do not appear to be involved in contacts with the inhibitor (Fig. 1b, 2b). A2, in particular, shows much variation in amino acid sequence. The internal residues show more conservative substitutions since they form part of the hydrophobic core.

Model Structure of HIV-2 PR

The fact that the important structural elements are highly conserved in sequence and that most substitutions occur in the surface loops was the basis for modeling the structure of HIV-2 PR using the three-dimensional structure of HIV-1 PR. Homologous modeling was used previously to build a structure of HIV-1 PR from the crystal structure of RSV PR that has 30% identical residues.[25] This model was reasonably accurate, except for the region of the flaps.[26] The structures of HIV-1 and HIV-2 PRs are expected to be much more similar, since the sequences have the same length and show 39% identical residues. Therefore, the model of HIV-2 PR was built with only minor alterations from the experimental structure of HIV-1 PR (Fig. 1a).

The active site region is the most conserved both in the amino acid sequence, and in the three-dimensional structure for all aspartic proteases. No insertions, deletions, or large changes in the conformation of the main chain are required to model the HIV-2 PR structure. Four of the glycines in HIV-1 PR with positive ϕ torsion angles are substituted by other types of amino acids in HIV-2 PR, so their ϕ torsion angles were adjusted to the more energetically favorable negative values. Gly 16, at the hairpin turn of loop A1, is Glu in all HIV-2 isolates. The hairpin turn in loop A2 also has a change from Gly

Fig. 3. The 4-stranded antiparallel β-sheet formed by the termini is shown in a stereo view for residues 1 to 4 and 95 to 99 from each subunit in the PR dimer. Residue 67 is also shown. Open bonds indicate the HIV-1 PR crystal structure, and solid bonds indicate the HIV-2 PR modeled structure. The residues are labeled in the single letter code for amino acid type, with the type in HIV-1 PR followed by the type in HIV-2 PR (HIV-1 PR residue/HIV-2 PR residue), and the residue number.

Fig. 4. Stereo view of a putative chimera between HIV-1 and HIV-2 PR showing the residues 1 to 4 and 95 to 99 that form the intersubunit β-sheet and residue 67 in the same representation as in Figure 3. Residues 1–4, 67, and 95–99 are from HIV-2 PR, and residues 1′–4′, 67′, and 95′–99′ are from HIV-1 PR. Larger residues are grouped on the left-hand side.

Fig. 5. Stereo view of the interactions of the C-terminus with His 69 in HIV-1 PR, or with Lys 69 in HIV-2 PR, in a representation similar to that of Figure 3. Both residues at 69 can form an ionic interaction with the carboxylate group of residue 99.

Fig. 6. Stereo view of the hydrophobic core of the PR. Residues from the crystal structure of HIV-1 PR are shown in open bonds, and residues from the model of HIV-2 PR are indicated by solid bonds. The labeling is similar to Figure 3.

68 with a positive ϕ torsion angle to Asn. Both Glu and Asn are residues that are frequently found at surface turns in protein structures.[27] Gly 40 is substituted by Asn in HIV-2 PR *rod* isolate in the wide loop with many variable residues. A small adjustment was made in the ϕ torsion angle for Gly 73, which becomes Ala in HIV-2 PR. Although it is possible for homologous loops to have divergent conformations,[28] comparison of the HIV-2 PR model and the HIV-1 PR crystal structures showed that the internal packing of hydrophobic residues and the external polar interactions were simlar and more consistent with similar loop conformations.

The energies of the model and of the crystal structures[10] were analyzed using XPLOR.[29] The overall energy is very similar, and there are only five bad contacts (2.3–2.5 Å) between hydrogen bonded pairs of atoms in both cases. In the crystal structure 0.2% carbon-carbon contacts are below 3.3 Å, compared with 0.35% in the model structure. The root mean square deviation from ideal stereochemistry was 0.032 Å for the model, and was 0.024 Å for the crystal structure.

Intersubunit β Sheet

Only those substitutions in the sequence that modify the interactions in the three-dimensional structure or produce new interactions in HIV-2 compared to HIV-1 PR are described. The first variable residue in HIV-2 PR is Phe 3, which is Leu in HIV-1 PR. It is found at the N-terminus of both monomers and, together with the C-termini, participates in the formation of the four-stranded β-sheet at the intersubunit interface. There are three important changes in residues in this region between HIV-1 and HIV-2 PR: Ile 3 to Phe, Cys 95 to Met, and Phe 99 to Leu (Phe in HIV-2 *d205* isolate). All of these residues are involved in internal hydrophobic interactions (Fig. 3), which are expected to stabilize the dimer. The presence of the larger hydrophobic residues, Phe 3 and Met 95, may enhance these interactions in HIV-2 PR. This may counteract the substitution of Phe 99 for Leu, which is expected to weaken the hydrophobic interaction with Pro 1′ (Fig. 3). Cys 67, which is close to Phe 99′ is substituted by the larger residue, Leu, whereas the bulky Phe changes to the smaller Leu at 99 in HIV-2 PR. While Phe occupies position 99 in the *d205* isolate, a concerted change to a smaller residue (Val 67) also occurs. The importance of the Phe 99 in HIV-1 PR has been tested by deleting this residue, which inactivates the PR.[30]

The other residues involved in these hydrophobic interactions are conserved in all isolates in both types (Pro 9 and 9′, Leu 23 and 23′, Leu 24 and 24′, Leu 90 and 90′). The hydrophobic side chain of Leu 97 lies near Leu 24′ and Thr 26′ near the active site, and mutation of any of these three residues reduces the activity of HIV-1 PR.[31] In HIV-2 PR, the side chains of Gln 2, Thr or Ser 96, and Asn 98 can produce the same hydrogen bonds on the exterior of this intersubunit β-sheet, as shown in[17] for HIV-1 PR.

The distribution of variable residues in the four β-strands can affect the interactions across the sheet. In HIV-1, for example, hydrophobic contacts of Pro 1 with Phe 99 enhance the interactions between the N- and C-termini in the sheet. The substitutions of Phe 99 to Leu and Cys 95 to Met provide additional hydrophobic intractions between both C-termini in the sheet. These observations might be useful in designing inhibitors of dimer formation.

The heterodimer consisting of an HIV-1 PR monomer and an HIV-2 PR monomer (Fig. 4), is less stable than either of the parent homodimers.[32] In the case of a heterodimer, the hydrophobic interactions would become asymmetrical. There would be fewer hydrophobic moieties on one side where all three variable residues are smaller (Ile(Phe) 3, Cys(Met) 95, Phe(Leu) 99, where residues in parentheses are from the HIV-2 PR monomer), and in contrast, more hydrophobic moieties would be found on the other side where only the larger residues would be present.

One interesting difference between HIV-1 and -2 PRs occurs at residue 69. His 69 forms an ion pair with the negatively charged C-terminus of the HIV-1 PR dimer (Fig. 5). Modeling of the substitution of His to the Lys that is present in all isolates of HIV-2 PR and in the *mal* isolate of HIV-1 showed that, in these cases, the ionic interaction can be conserved. Two of the HIV-1 isolates are different, with residue 69 being either Tyr or Gln, so that no ion pair could be formed with the C-terminus. Hydrogen bond interactions between residue 69 and the C-terminus, however, are still possible.

Hydrophobic Interior Residues

Numerous substitutions occur among the residues forming the interior hydrophobic core of the PR subunit. As shown in Figure 6, the residues involved in these interactions are 13, 15, 22, 31, 33, 64, 66, 75, and 85. Substitution of Ile 85 in the center of the hydrophobic core by the bulky Phe residue in HIV-2 PR is accompanied by a decrease in the sizes of neighboring residues: Ile 13 to Ala, Leu 33 to Val, Thr 31 to Ser, and Ile 66 to Val. Simultaneously, Ala 22 and Val 75 are altered to the larger Val and Ile. The result is that all these internal hydrophobic residues have a favorable steric fit in the model structure of HIV-2 PR.

Tyr 59 is absolutely conserved in all isolates of both types of HIV PR, even though it occurs in a region (residues 55–77) that shows many differences in sequence between HIV-1 and HIV-2 PRs (Table I). Tyr 59 makes hydrophobic contacts with both the structural core of the monomer, and the outer wide loop of less regular structure (residues 39–43) (Figs. 7,8) and thus stabilizes the conformation of this sur-

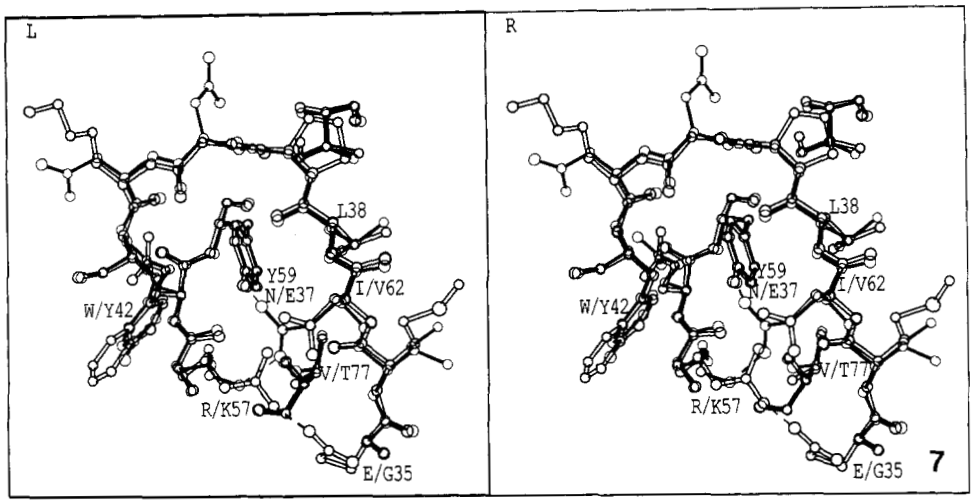


Fig. 7. Stereo view of Tyr 59 in HIV-1 PR (open bonds) and in HIV-2 PR (solid bonds). The labeling is similar to Figure 3. Glu 37 interacts with Tyr 59 and Lys 57 in HIV-2 PR. In HIV-1 PR, Arg 57 forms a salt bridge with Glu 35.

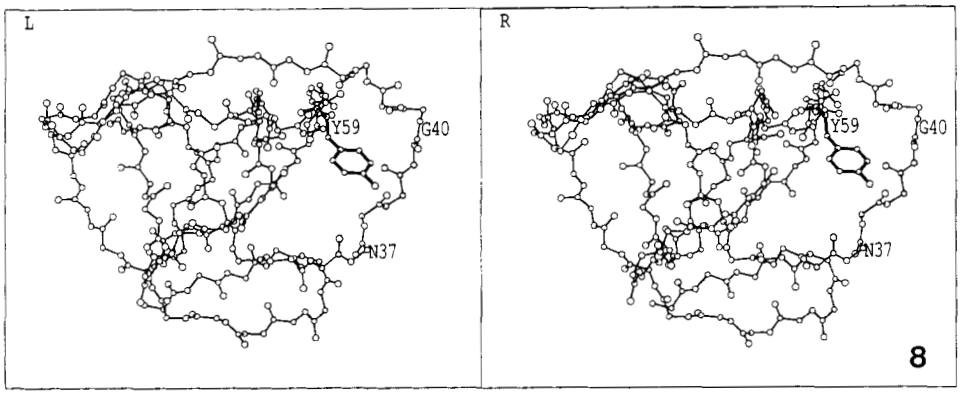


Fig. 8. Stereo view of Tyr 59 and the main chain atoms of one subunit of HIV PR. Tyr 59 interacts with residues of the wide loop 33–41.

face loop. The structural analog of this residue in RSV PR is Trp 50, which shows similar hydrophobic contacts (Fig. 9). Trp 54 plays a similar role in the RSV PR structure. It stabilizes the A1 loop, which is much longer than in the HIV PR structure. Bulky residues that provide numerous hydrophobic contacts between a structural core and surface elements can be used to maintain the globular conformation of proteins. The related destabilizing effect of substituting smaller residues for bulky hydrophobic residues has been measured for barnase.[33] In addition, the ionic interactions of side chains may be used to connect surface structural elements as previously suggested by Barlow and Thornton,[34] for example, residues 23 and 78 in the RSV PR structure, or residues 14 and 65 in HIV PR, as described later.

The wide loop with the residues 33–41 is a variable structural element in all isolates of both types of HIV PR. The residues in this loop participate in the formation of a hydrophobic surface pocket in HIV-2 PR structure (Fig. 10). This region differs in the three-dimensional structures of HIV-1 and -2 PR (Fig. 10), due to the variability of the residues that form it. The conformation of the polypeptide backbone in this region also differs greatly in the three-dimensional structures of other retroviral PRs, although most of the core structure is very similar. For example, in the structures of RSV PR and the almost identical avian myeloblastosis associated virus (MAV) PR, there is a helical turn in this region that is not present in HIV-1 PR. Mous et al.[35] reported that peptide antibodies were raised against the HIV-1 PR sequence in the range of residues 18–40. It may be suggested that such different surface structural features in different retroviral PRs can serve as an epitope for interactions with antibodies.

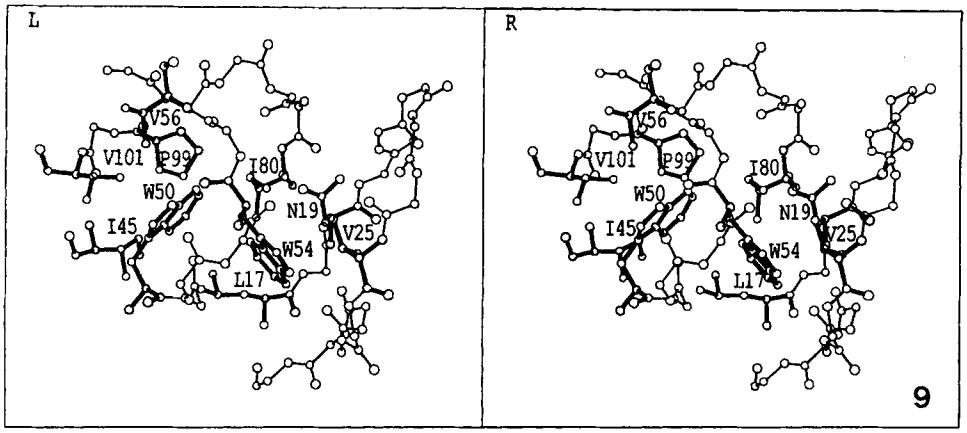


Fig. 9. Stereo view of the environment of Trp 50 and Trp 54 in RSV PR. The residues forming hydrophobic interactions are shown in thicker lines, the main chain is in thinner lines. Trp 54 interacts with residues of the A1 hairpin loop (residues 10 to 34) in RSV PR.

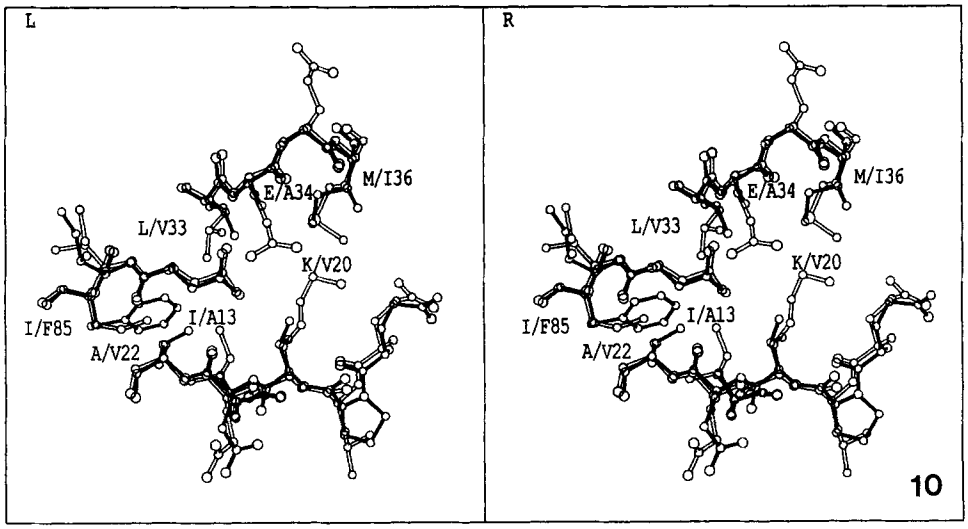


Fig. 10. Stereo view of residues from part of the wide loop and the neighboring residues. Residues 13, 18–22, 33–36, and 83–85 are shown for HIV-1 (open bonds) and HIV-2 (solid bonds) PRs, with labeling as described in Figure 3.

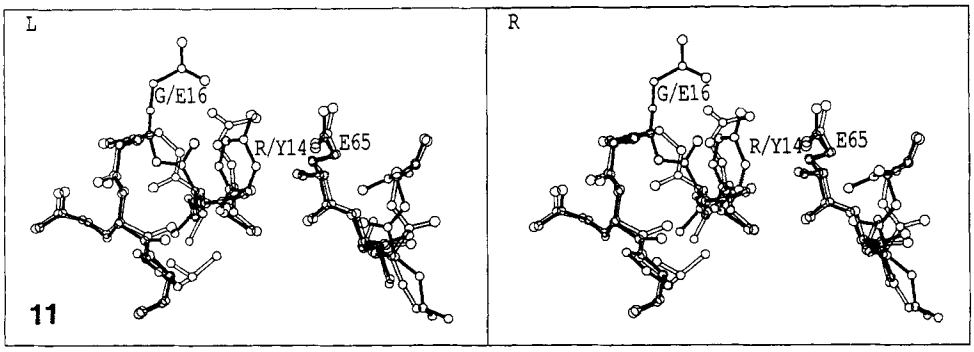


Fig. 11. Stereo view of the interactions between residues 14 and 65 in HIV-1 (open bonds) and HIV-2 (solid bonds) PRs with labeling as in Figure 3. In HIV-1 PR, Arg 14 forms an ionic interaction with Glu 65. In HIV-2 PR, the hydroxyl of Tyr 14 may form hydrogen bond interactions with the side chains of residues 65 and 16.

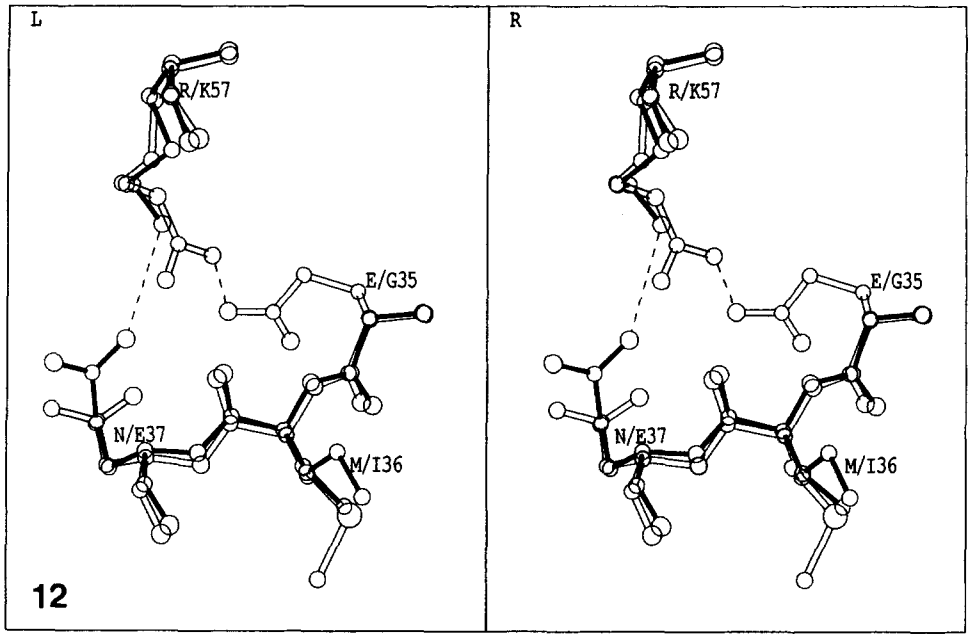


Fig. 12. Stereo view of ionic interactions between residue 57 and 35 or 37. In HIV-1 PR (open bonds), Arg 57 forms an ion pair with Glu 35. In HIV-2 PR (solid bonds), Lys 57 may interact with Glu 37. The labeling is as described in Figure 3.

This might be useful for immunological approaches to distinguishing between HIV-1 and -2 or for creating possible drugs.

Ionic Interactions

When the crystal structure of HIV-1 PR is compared to the modeled structure of HIV-2 PR, it can be seen that several regions are connected by similar ionic or polar interactions involving different types of residues in the two types of HIV PR. Here we describe several of these.

In the crystal structure of HIV-1 PR, Arg 14 forms an ion pair with Glu 65, which stabilizes the position of both the A1 and A2 surface hairpin turns (Fig. 11). Arg or Lys occurs at residue 14, and Glu 65 is present in all but one of the HIV-1 isolates. In the HIV-2 isolates there are four substitutions observed at residue 14, Tyr, His, Phe, and Cys. Modeling shows that in some cases it is possible to obtain similar side chain interactions (Fig. 11). If Glu is found at position 65 and His or Tyr at 14, they can interact easily. When position 14 is occupied by an aromatic residue, the side chain of residue 65 may interact with the π-electron cloud of the aromatic ring.[36] In the HIV-2 d205 isolate residue 14 is substituted by Cys, and only in this isolate does the highly conserved Thr 12 change to Lys. Modeling shows that the small size of Cys side chain does not prevent the formation of an ion pair between Lys 12 and Glu 65. The variability of the residues in the surface loop 12–17 in different isolates of HIV-2 PR makes it possible to utilize that region as a target for production of monoclonal antibodies. If such antibodies could be obtained, they might distinguish not only the HIV-1 PR from HIV-2 PR, but also among different isolates.

In the known crystal structures of retroviral proteases, there is a conserved ion pair between Glu 35 and Arg 57 of HIV-1 PR (which correspond to Glu 47 and Arg 74 in RSV PR[37,38] and MAV PR[39]). Arg 57 is at the entry to the flap, and Glu 35 belongs to the variable surface loop. In HIV-2 PR, Glu 35 is replaced by Gly, but there is a correlated change of Asn 37 to Glu. Modeling shows that it is possible to form an ion pair between Lys 57 and Glu 37 in HIV-2 PR (Fig. 12), so that the flap and the variable loop are connected by similar, but not identical, ionic interactions in HIV-1 and -2 PRs. The exact position of this ion pair may have a subtle effect on the interaction of the flap with substrates or inhibitors in HIV-1 or HIV-2 PRs.

Lys 43 and Glu 58 interact by formation of a hydrogen bond and are conserved in all isolates of HIV-1 PR. In HIV-2 PR Lys 43 is changed to Ser or Thr, whereas Gln 58 is changed to Glu and Asp 60 is changed to Lys. Modeling shows that it is possible to form ionic interactions between Glu 58 and Lys 60 in HIV-2 PR. Therefore, residue 58 may have different interactions in HIV-1 and -2 PRs.

Inhibitor Binding in HIV-1 and HIV-2 PRs

Inhibition constants for HIV-1[40] and HIV-2[41] PR were measured for two characteristic inhibitors of aspartic proteases, acetyl-pepstatin and a renin in-

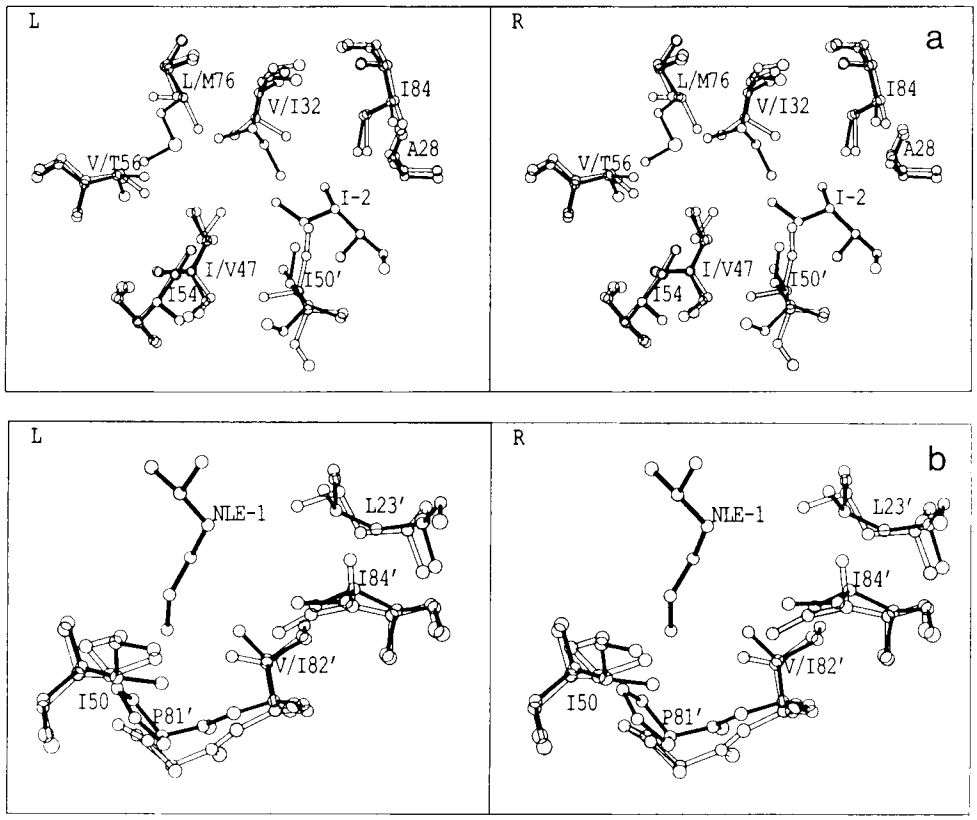


Fig. 13. (**a**) Stereo view of the amino acid substitutions around the S2 subsites of HIV-1 (open bonds) and HIV-2 (solid bonds) PRs. The labeling is as described in Figure 3. The substitutions Val 32 to Ile, Ile 47 to Val, and Leu 76 to Met are shown near Ile 2 of the inhibitor from the cocrystal structure.[9] (**b**) Stereo view similar to (a) of the subsites S1 showing the Val 82 to Ile change near inhibitor residue, Nle-1, which is Norleucine from the cocrystal structure.[9]

hibitor, H-261. The acetyl-pepstatin has a measured K_i of 20 nM for HIV-1 PR, and K_i of 5 nM for HIV-2 PR, whereas H-261 has a K_i of 5 nM for HIV-1 PR and K_i of 35 nM for HIV-2 PR. Acetyl-pepstatin was more effective toward HIV-2 PR than the H-261, which is the opposite of what was observed previously for the HIV-1 PR. We have compared the models of these two inhibitors bound to the HIV-1 and -2 PR structures in order to attempt to explain the differences in their relative affinities. After submission of this manuscript, the crystal structure of HIV-1 PR with acetyl-pepstatin was publsihed.[42] Our model structure appears to agree with the experimental data.

The substrate binding pocket is formed by residues 8, 23, 25, 27 to 30, 32, 47 to 50, 53, 76, 80 to 82, and 84. Only four of these residues involved in enzyme-substrate interactions show differences between HIV-1 and -2 PRs. These are all conservative substitutions: Val 32 to Ile (part of the S2′ and S2 subsites), Ile 47 to Val (involved in the S2 and S4 subsites), Leu 76 to Met (involved in S2 and S2′), and Val 82 to Ile (in the S1′ subsite) (Fig. 13). It is quite clear that these substitutions can only change the hydrophobicity of the respective subsites by the loss or gain of one methyl group. Within the different isolates of HIV-1 and -2 there are no changes in the residues forming the PR substrate binding pocket.

Modeling of the acetyl-pepstatin and reinin inhibitors in the binding sites of both types of HIV PR showed that at least one of these four residues, which differ and are involved in enzyme-ligand interactions, contributes to the difference in relative inhibition constants. This is residue 32, which is Val in HIV-1 and Ile in HIV-2 PR and forms part of the S2 and S2′ subsites. The arrangement of residues in the S2 and S2′ subsites of HIV-1 and -2 PRs is shown in Figure 14. The renin inhibitor has a larger residue in both P2 and P2′ compared to acetyl-pepstatin: His compared to Val at P2, and Ile compared to Ala at P2′. Ile 32 in HIV-2 PR in some conformations appears to have unfavorably close contacts with P2 and P2′ of the renin inhibitor, whereas Val 32 in HIV-1 PR does not have close contacts. In the case of acetyl-pepstatin, on the contrary, the presence of smaller residues in P2 and P2′ make its binding in HIV-2 PR more effective, since there are better hydrophobic interactions with Ile 32. Leu/Met 76 also forms part of these subsites and may contribute to

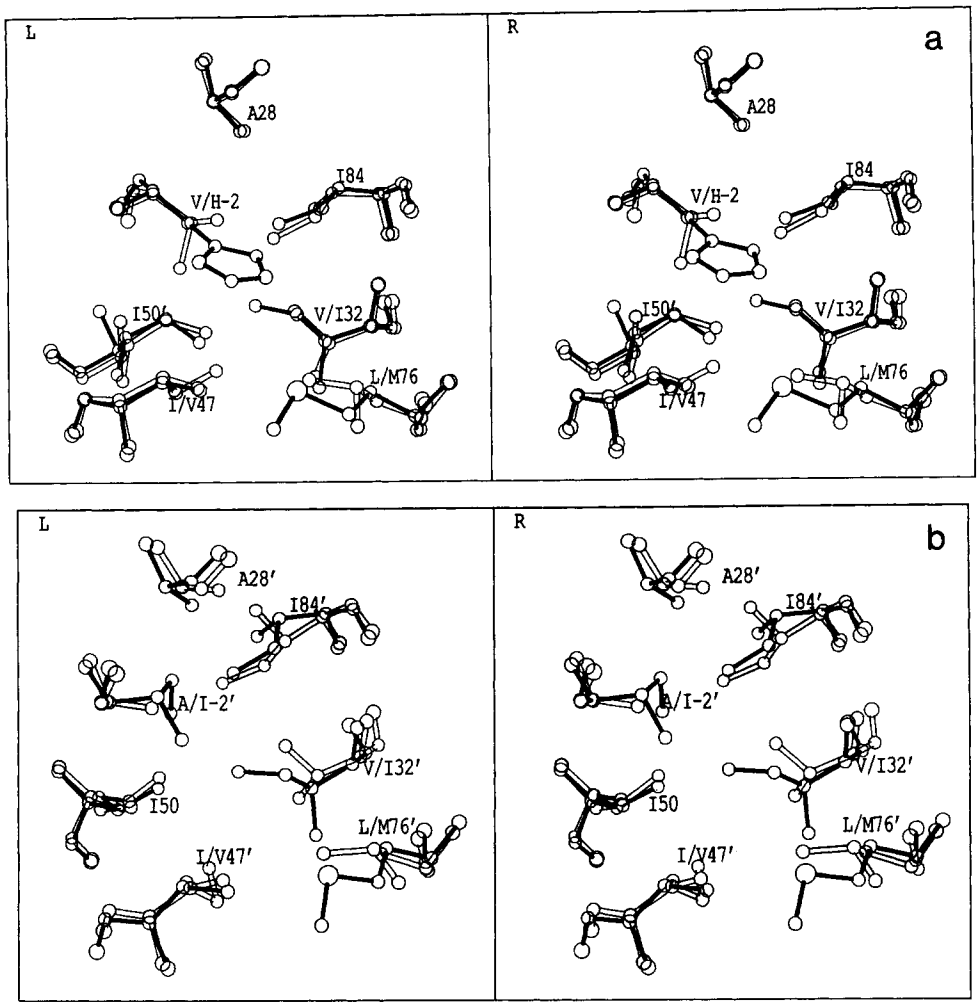


Fig. 14. (**a**) Stereo view of the interactions of the P2 inhibitor residues in the S2 subsites of HIV-1 (open bonds), or HIV-2 (solid bonds) PRs. The labeling is as described in Figure 3. (**b**) Stereo view of the interactions of the P2′ inhibitor residues in the S2′ subsites of HIV-1 and -2 PRs, as shown in (a).

the hydrophobic interaction with inhibitor. The role of the substitutions of the other two residues, 47 and 82, with respect to the binding of these inhibitors is less obvious.

CONCLUSIONS

The model of HIV-2 PR described here has been constructed on the basis of the homologous structure of HIV-1 PR and as such may not be correct in all details. However, the similarity of the sequences of these proteins (39% identity) suggests that the main structural characteristics of the model should be sufficiently accurate for qualitative analysis. Examination of the PR structures with two different inhibitors has identified one residue in the substrate binding site where the size of amino acid present in HIV-1 or -2 PR is correlated with the relative binding constants of the inhibitors.

Comparison of the crystallographic and model structures of HIV-1 and HIV-2 PRs shows that all substitutions of residues occur such way that the important structural interactions can be maintained. Modeling of the substitutions in different isolates of both types of HIV PR leads to the same conclusion. These results can serve as a basis for a general approach to reveal the structural features of proteins with unknown three-dimensional structures. The comparison of the amino acid sequences of homologous enzymes and their isozymes (when available) can provide some important conclusions about the possible structural role, not only for conserved residues that obviously have an important function, but for those residues that show correlated substitutions. Residues showing such correlated substitutions may be involved in the same set of local interactions (hydrophobic or polar) in the three-dimensional structure, and such predictions could be tested by site-directed mutagenesis.

Another practical result can be obtained by locating the fragments in the sequences of homologous proteins and their isolates where there are frequent substitutions. This study shows that those variable regions can form different surface structural elements and therefore can serve as an epitope for monoclonal antibodies. This serves as a possible guide to an immunological approach for separation and purification of different proteins.

ACKNOWLEDGMENTS

We thank Drs. Alexander Wlodawer, Dominique Housset, Joseph Mack, Stephen Foundling, Mariusz Jaskólski, and Robert Harrison for helpful discussions, and Dr. Amy Swain for providing us with the coordinates of the protease complexed with the JG-365 inhibitor prior to publication. This research was sponsored by the National Cancer Institute, DHHS, under contract No. N01-C01-74101 with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

REFERENCES

1. Kanki, P.J., Barin, F., M'Boup, S., Allan, J.S., Romet-Lemonne, J.L., Marlink, R., McLane, M.F., Lee, T.-H., Arbeille, B., Denis, F., Essex, M. New human T-lymphotropic retrovirus related to simian T-lymphotropic virus type III (STLV-IIIagm). Science 232:238–243, 1986.
2. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., Alizon, M. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature 326:662–669, 1987.
3. Zaguru, J.F., Francini, G., Reitz, M., Collalti, E., Starcich, B., Hall, L., Fargnoli, K., Jagodzinski, L., Guo, H.-G., Laure, F., Arya, S.K., Josephs, S., Zaguru, D., Wong-Staal, F., Gallo, R.C. Genetic variability between isolates of human immunodeficiency virus (HIV) type 2 is comparable to the variability among HIV type 1. PNAS 85:5941–5945, 1988.
4. Brun-Vezinet, F., Katlama, C., Roulot, Lenoble, L., Alizon, M., Madjari, J.J., Rey, M.A., Girard, P.M., Yeni, P., Clavel, F., Gadelle, S., Harzic, M. Lymphadenopathy-associated virus type 2 in AIDS and AIDS-related complex. Lancet I:128–132, 1987.
5. Kohl, N.E., Emini, E.A., Scheif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Skolnick, E.M., Sigal, I.S. Active human immunodeficiency virus protease is required for viral activity. PNAS 85:4686–4690, 1988.
6. Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337:615–620, 1989.
7. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B.H. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 245:616–621, 1989.
8. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danley, D.E., Geoghegan, K.F., Hawrylik, S.J., Lee, S.E., Scheld, K.G., Hobart, P.M. X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. Nature 342:299–302, 1989.
9. Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H., Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 246:1149–1152, 1989.
10. Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Kent, S.B.H., Wlodawer, A. X-ray crystallographic structure of a complex between HIV-1 protease and a hydroxyethylamine inhibitor. Proc. Natl. Acad. Sci. U.S.A. 87:8805–8809, 1990.
11. Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Poul, D.A., Knigge, M. Design, activity, and 2.8 Å crystal structure of C2 symmetric inhibitor complexed to HIV-1 protease. Science 249:527–533, 1990.
12. Hansen, J., Billich, S., Schulze, T., Sukrow, S., Moelling, K. Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J 7:1785–1791, 1988.
13. Kotler, M., Danho, W., Katz, R., Leis, J., Skalka, A.-M. Avian retroviral protease and cellular aspartic proteases are distinguished by activities on peptide substrates. J. Biol. Chem. 264:3428–3435, 1989.
14. Myers, G., Josephs, S.F., Wong-Staal, F., Rabson, A.B., Smith, T.F., Berzofsky, J.A. (eds.). "Human Retroviruses and AIDS 1989: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequence," Los Alamos, NM: Los Alamos National Laboratory, 1989.
15. Hein, J. Unified approach to alignment and phylogenies. Methods Enzymol. 183:626–645, 1989.
16. Jones, A.T. A graphics model building and refinement system for macromolecules. J. Appl. Crystallogr. 11:268–272, 1978.
17. Weber, I. Comparison of the crystal structures and intersubunit interactions of Human Immunodeficiency and Rous Sarcoma Virus proteases. J. Biol. Chem. 265:10492–10496, 1990.
18. Andreeva, N.S. Structural template of aspartic proteases. In: Biosynthesis and Structure of Proteins. Pushchino 3: 97–115, 1988.
19. Andreeva, N.S., Zdanov, A.S., Gustchina, A.E., Fedorov, A.A. Structure of ethanol-inhibited porcine pepsin at 2 Å resolution and binding of the methyl ester of phenylalanyl-diiodotyrosine to the enzyme. J. Biol. Chem. 259:11353–11366, 1984.
20. Blundell, T., Jenkins, J., Pearl, L., Sewell, T., Pedersen, V. The high resolution structure of Endothiapepsin. In Kostka, V., (ed.): "Aspartic Proteinases and Their Inhibitors." New York: Walter de Gruyter, 1985:151–161.
21. Bott, R.R., Subramanian, E., Davies, D.R. Three-dimensional structure of the complex of the *Rhizopus chinensis* carboxyl proteinase and pepstatin at 2.5 Å resolution. Biochemistry 21:6956–6962, 1982.
22. James, M.N.G., Sielecki, A.R. Structure and refinement of penicillopepsin at 1.8 Å resolution. J. Mol. Biol. 163:299–361, 1983.
23. Foundling, S.I., Cooper, J. Watson, F.E., Cleasby, A., Pearl, L.H., Sibanda, B.L., Hemmings, A., Wood, S.P., Blundell, T.L., Valler, M.J., Norey, C.G., Kay, J., Boger, J., Dunn, B.M., Leckie, B.J., Jones, D.M., Atrash, B., Hallett, A., Szelke, M. High resolution X-ray analyses of renin inhibitor-aspartic proteinase complexes. Nature 327: 349–352, 1987.
24. Gustchina, A., Weber, I. Comparison of inhibitor binding in HIV-1 protease and in nonviral aspartic proteases: the role of the flap. FEBS Lett. 269:269–272, 1990.
25. Weber, I., Miller, M., Jaskólski, M., Leis, J., Skalka, A.M., Wlodawer, A. Molecular modelling of the HIV-1 protease and its substrate binding site. Science 243:928–931, 1989.
26. Weber, I.T. Evaluation of homology modeling of HIV protease. Proteins 7:172–184, 1990.
27. Wilmot, C.M., Thornton, J.M. Analysis and prediction of the different types of β-turns in proteins. J. Mol. Biol. 203: 221–232, 1988.
28. Greer, J. Comparative modeling methods: application to the family of mammalian serine proteases, Proteins 7:317–334, 1990.
29. Brünger, A.T., Kuriyan, J., Karplus, M. Crystallographic R-factor refinement by molecular dynamics. Science 235: 458–460, 1987.
30. Hostomsky, Z., Appelt, K., Ogden, R.C. High-level expres-

sion of self-processed HIV-1 protease in Escherichia coli using a synthetic gene. Biochem. Biophys. Res. Commun., 161:1056–1063, 1989.
31. Loeb, D.D., Hutchison III, C.A., Edgel, M.H., Farmerie, W.G., Swanstrom, R. Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. J. Virol., 63:111–121, 1989.
32. Babé, L.M., Pichuantes, S., Craik, C.S. Inhibition of HIV protease activity by heterodimer formation, Biochemistry, 30:106–111, 1991.
33. Kellis, J.T., Nyberg, K., Sali, D., Fersht, A.R. Contribution of hydrophobic interactions to protein stability, Nature 333:784–786, 1988.
34. Barlow, D.J., Thornton, J.M. Ion pairs in proteins. J. Mol. Biol. 168:867–885, 1983.
35. Mous, J., Heimer, E.P., Le Grice, S.F.J. Processing protease and reverse transcriptase from Human Immunodeficiency Virus Type I polyprotein in Escherichia coli. J. Virol. 62:1433–1436, 1988.
36. Burley, S.K., Petsko, G.A. Amino-aromatic interactions in proteins. FEBS Lett. 203:139–143, 1986.
37. Miller, M., Jaskólski, M., Mohana Rao, J.K., Leis, J., Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature 337:576–579, 1989.
38. Jaskólski, M., Miller, M., Mohana Rao, J.K., Leis, J., Wlodawer, A. Structure of the aspartic protease from Rous Sarcoma Retrovirus refined at 2 Å resolution. Biochemistry 29:5889–5898, 1990.
39. Foundling, S.J., Salemme, F.R., Korant, B., Wendoloski, J.J., Weber, P.C., Treharne, A.C., Schadt, M.C., Jaskólski, M., Miller, M., Wlodawer, A., Strop, P., Kostka, V., Sedlacek, J., Ohlendorf, D.H. Crystal structure of a retroviral proteinase from Avian Myeloblastosis associated Virus. In: Viral Proteinases as Targets for Chemotherapy. New York: Cold Spring Harbor, 1989:181–190.
40. Richards, A.D., Roberts, R., Dunn, B.M., Graves, M.C., Kay, J. Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases. FEBS Lett. 247:113–117, 1989.
41. Richards, A.D., Broadhurst, A.V., Ritchie, A.J., Dunn, B.M., Kay, J. Inhibition of the aspartic proteinase from HIV-2. FEBS Lett 253:214–216, 1989.
42. Fitzgerald, P.M.D., McKeever, B.M., VanMiddlesworth, J.F., Springer, J.P., Heimbach, J.C., Leu, C.-T., Herber, W.K., Dixon, R.A.F., Darke, P.L. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0 Å resolution. J. Biol. Chem. 265:14209–14219, 1990.