23 Ferris, T, *Sky Telesc*, 1987, **73**, 486, *Nature*, 1992, **356**, 657, *Science*, 1992, **257**, 1208, Lynden-Bell, D *et al*, *Astrophys J*, 1988, **326**, 19, Willick, J, *Astrophys J*, 1990, **351**, L5, Dressler, A and Faber, S, *Astrophys. J*, 1990, **354**, L45, Mathewson, D S *et al*, *Astrophys J*, 1992, **389**, L5

ACKNOWLEDGMENTS I thank N D Hari Dass and S Uma Sankar for arranging the One-Day Meeting and Institute of Mathematical Sciences for support and use of their facilities K R Anantharamaiah's comments helped improve the presentation

Received 25 May 1993; accepted 14 July 1993

X-ray crystal structure and computer modelling studies of HIV protease and its inhibitor complexes

Bangalore K. Sathyanarayana and Alexander Wlodawer

Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, BRP-ABL, P O Box B, Frederick, MD 21702, USA

HIV-1 protease is essential for the replication of HIV or the acquired immunodeficiency (AIDS) virus and is considered as an attractive target for the design of specific inhibitors. In order to design drugs which inhibit the action of HIV protease, it is essential to obtain the 3-D structures of these proteases. The native HIV protease and the very first inhibitor complex of the protease were studied at our laboratory and in this article we summarize the X-ray structure analysis and molecular modelling studies of the HIV-1 protease both in its native as well as with different inhibitor complexes studied both at our laboratory as well as laboratories elsewhere.

THIS article summarizes the results of X-ray crystallographic and computer modelling studies of the human immunodeficiency virus protease (HIV-PR) and its inhibitor complexes. These studies are quite promising and may lead to the design of drugs that could be therapeutically beneficial in cases of acquired immunodeficiency syndrome (AIDS). HIV-PR is an aspartic protease (PR), that is encoded by the human immunodeficiency virus (HIV). This enzyme is essential for proper maturation of the HIV virons, and if it is inactivated by either mutation or chemical inhibition, the assembled viral particles are not infectious. The techniques of rational drug design are being employed in many industrial and academic laboratories all over the world, in search of compounds that could accomplish such inactivation in the human body. However, for a rational approach to drug design to succeed, it is important to characterize the structures of the target enzyme and of enzyme–inhibitor complexes. Observations based on structural data, when used with other sources of information such as the results of investigations of binding constants and kinetics parameters, can lead to a better understanding of the enzyme–inhibitor interactions. Currently, extensive efforts have been devoted to the structure determination of HIV protease at various laboratories, and in this article, we summarize the work originating from our laboratory, as well as the results from other laboratories throughout the world.

HIV has been classified as a member of the lentivirus family[1] and from the array of processes required to sustain the viral life cycle, it may be possible that there are numerous points that could be exploited in development of drugs for AIDS therapy[2]. Figure 1 shows the HIV life cycle and the places of possible attack by the drugs at several stages. The cycle begins when HIV binds to the outside of the host cell and injects its core, which includes two identical strands of RNA and structural proteins and enzymes, all of which are needed at the later stages of the virus life cycle. Since HIV is a retrovirus with RNA as the source of its genetic information, instead of DNA as in most other viruses, it needs the help of two specific enzymes, namely polymerase and ribonuclease. These two enzymes form reverses transcriptase whose role is to make a second DNA copy using the first one as a template and to destroy the original RNA of the virus. RT actually causes the genes in the cell to make the proteins that the virus needs to reproduce. This is the first stage wherein the function of the RT can be halted by the intervention of any suitable drug. The genetic information of the virus in the regular DNA form is now carried on to the cell nucleus, where another viral enzyme called integrase (IN) will then splice the HIV genome into the host cell's DNA. This may well be considered as the second potential major stage of drug intervention against the virus. At the final stage, the

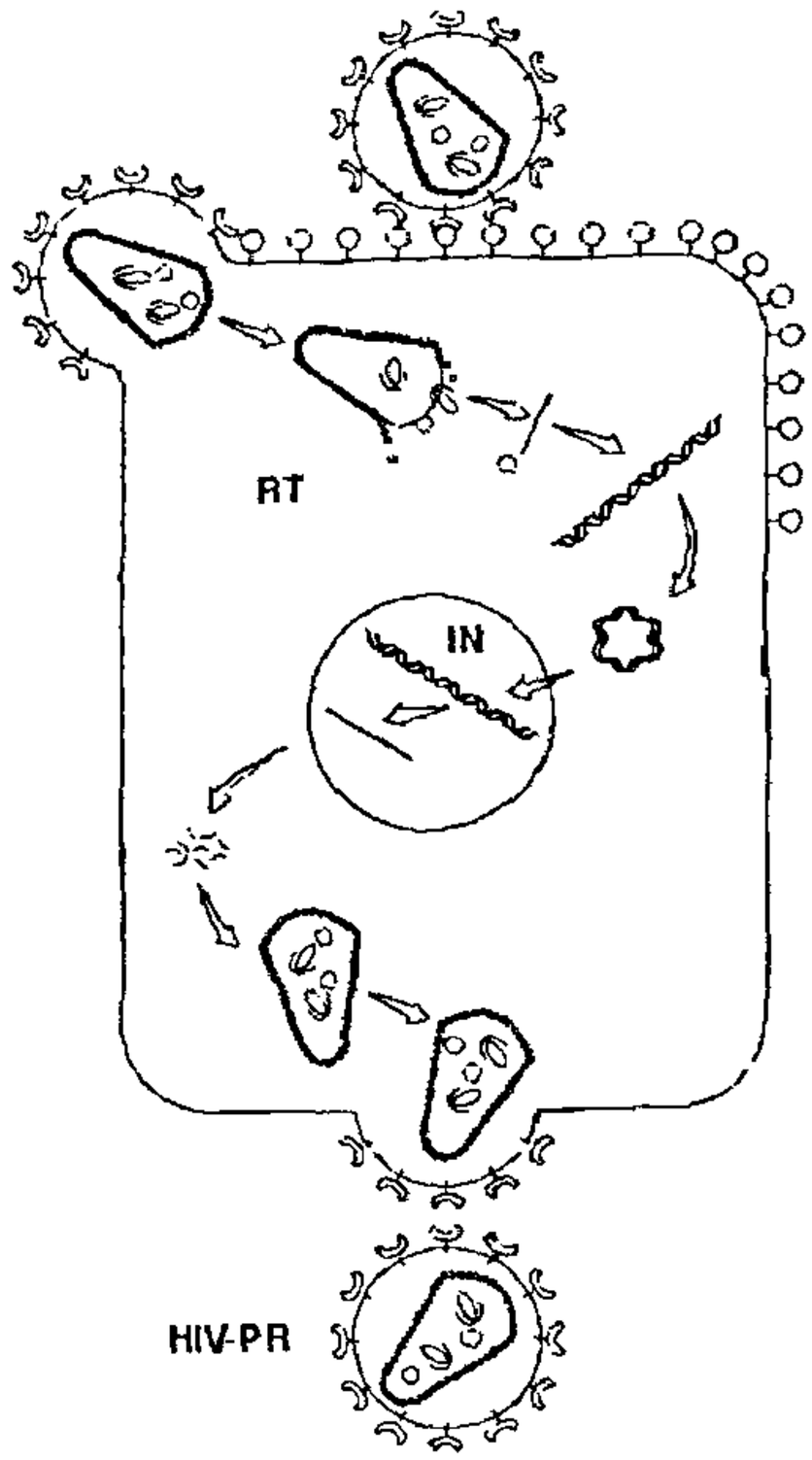

Figure 1. Life cycle of human immunodeficiency virus and the three major stages of the virus wherein it can be intervened by drug to stop the replication of the virus

virus undergoes modifications with the help of the protease (HIV PR). In this article we discuss ways to inhibit this protease using peptide and non-peptide inhibitors. A full knowledge of the three-dimensional structures of all the HIV components, in particular, the reverse transcriptase, integrase and protease is essential for understanding the mechanism of this viral cycle. Currently, the only such component of HIV with structures available to a high resolution is the HIV protease. The X-ray structure of the reverse transcriptase has recently been solved at low or medium resolution by two different groups[3,4] and RNase H by Hostomska *et al*[5] The need to understand the structures of other HIV components is critical and work is underway in this direction at various laboratories including ours

Crystal growth and data collection

The enzyme that we used for X-ray study of the native HIV-1 PR without the inhibitor was prepared by total chemical synthesis[6]. Single crystals of the PR were grown by modifying the method described by McKeever *et al*[7] and were isomorphous with those obtained by them (tetragonal space group $P4_12_12$, unit cell parameters $a = b = 50.24$ Å, $c = 106.56$ Å) Crystals of HIV-1 PR with various inhibitor complexes at various laboratories were grown at almost similar but slightly different conditions. For example, the HIV-1 PR with the hydroxyethylamine inhibitor were grown by vapour diffusion using the hanging drop technique. Conditions in the reservoir were 55–60% (vol/vol) saturated ammonium sulphate/0.1 M sodium accetate, pH 5.4. The protein concentration was 5 mg/ml. The inhibitor was dissolved in dimethyl sulphoxide and diluted into the protein sample to give a final inhibitor concentration of 9 mg/ml [9% (vol/vol) dimethyl sulphoxide] The 6-µl drops consisted of 50% of this mixture and 50% of the protein-inhibitor sample. Crystals grew at 23°C in 3 days. In the case of MVT-101 inhibitor, the enzyme (stored in phosphate buffer, pH 7, in 20% glycerol in –20°C) was concentrated to 6 mg/ml using Centricon microconcentraters, while simultaneously exchanging the buffer for 20 mM sodium acetate, pH 5.4. Five mg of MVT-101 was dissolved in 100-µl DMSO and mixed with protein to yield 10-fold molar excess of the inhibitor. Crystals grew at room temperature in hanging drops from 60% ammonium sulphate. They appeared within a few days as thin rods with maximum dimensions of 0.3 × 0.12 × 0.06 mm.

The crystals used for data collection normally measured up to 0.06 × 0.1 × 0.4 mm and were mounted in quartz capillaries. Data were collected at our laboratory using a Siemens area detector mounted on a three-axis camera. Table 1 shows the resolution of the data collected in various experiments ranging from 2.0 to 3.0 Å. All the inhibitor complexes reported so far crystallize in either the orthorhombic or the hexagonal form. The chemical formulae of the 15 inhibitors studied both in this laboratory and elsewhere are shown in Table 1. Crystallographic information for some of the inhibitor complexes is summarized in Table 2.

X-ray structure of HIV-1 PR

Scientists from Merck Sharp and Dohme were the first to report the X-ray structure of the native HIV-1 PR, which was almost immediately verified independently by other laboratories including ours[6,8]. The crystal structure was solved at our laboratory by molecular replacement and confirmed by the multiple isomorphous replacement (MIR) method. The molecular replacement method used the HIV-1 protease model constructed by Weber[9] based on the known structure of the RSV PR which was solved by Miller *et al*[10] (Figure 2). The program package MERLOT[11] was used for molecular replacement and CORELS program[12] was used for rigid-body refinement, which was then confined (refined?) using the slow-cooling technique coupled with molecular dynamics of

Table 1.

Inhibitor	P5	P4	P3	P2	P1	Linkage	P1'	P2'	P3'	P4'	P5'	Resol. Å	R
MVT-101			Ac-Thr	Ile	Nle	CH_2NH	Nle	Gln	Arg			2.0	0.154
JG-365		Ac-Ser	Leu	Asn	Phe	$CH(OH)CH_2NH$	Pro	Ile	Val-OMe			2.4	0.146
U-85548e	Val	Ser	Gln	Asn	Leu	$CH(OH)CH_2$	Val	Ile	Val			2.5	0.138
AcPepstatin			Ac-Val	Val	Sta	$CH(OH)CH_2$	(CO–NH)	Ala	Sta			2.5	0.167
U-75875			Noa	His	Cha	$CH(OH)CH(OH)$	Val	Ile	Amp			2.0	0.169
A-74704			Cbz	Val	Phe	CH(OH)	Phe'	Val'	Cbz'			2.8	0.182
L-700,417				Ahe'	Phe'	CH(OH)	Phe	Ahi				2.1	0.180
A-77003			Pyr	Val	Phe	$CH(OH)CH(OH)$	Phe'	Val'	Pyr'			1.8	0.167
PS-1	Pro	Pro	Gln	Val	Psta	$CH(OH)CH_2$	(CO–NH)	Ala	Gln	Pro	Pro	2.0	0.162
Ro-31-8558				Boc	Cha	$CH(OH)CH_2$	Val	Ile	Epy			2.3	0.173
L-689,502				Boc	Phe	$CH(OH)CH_2$	Emt	Ahi				2.25	0.173
Pepstatin			Iva	Val	Sta	$CH(OH)CH_2$	(CO–NH)	Ala	Sta			2.8	0.147
AG-2		Ser	Phe	Asn	Sta	$CH(OH)CH_2$	(CO–NH)	Gln	Ile			2.5	0.158
AG-4		Ser	Gln	Asn	Sta	$CH(OH)CH_2$	(CO–NH)	Ile	Val	Gln		2.5	0.148
Lilly-765			Cbz	Val	Phe	$CH(OH)CH(OH)$	Phe	Val	Cbz			2.6	0.145

Abbreviations used: Nle, norleucine; Ac, acetyl; OMe, methoxy; Boc, t-butoxy; Cha, cyclohexylalanine; Epy, ethylpyridine; Cbz, carbobenzyloxy; Iva, isovaleryl; Sta, statine; Psta, phenyl statine; Noa, 1-naphthoxyacetyl; Amp, 2-pyridylmethylamine; Pyr, 2-pyridil-N-methylurea; Ahi, 1-amino-2-hydroxyindan; Emt, O-ethylmorpholinyltyrosine. Prime indicates reversal of amino acid linkage.

Table 2. Reported crystal forms of HIV-1 PR, both as apoenzyme and as inhibitor complexes. The first reference to each crystal form is listed.

Inhibitor	Space group	*a* α	*b* β	*c* γ	Crystal morphology	First reference
	$P4_12_12$	50.3 90.0	50.3 90.0	106.8 90.0	monomer	McKeever *et al.*[8]
MVT-101	$P2_12_12_1$	51.7 90.0	59.2 90.0	62.45 90.0	dimer	Miller *et al.*[10]
Acetyl-Pepstatin	$P2_12_12$	58.39 90.0	86.70 90.0	46.27 90.0	dimer	Fitzgerald *et al.*[24]
A-74704	$P6_1$	63.3 90.0	63.3 90.0	83.6 120.0	dimer	Erickson *et al.*[23]
JG-365	$P6_122$	63.4 90.0	63.4 90.0	83.6 120.0	monomer	Dreyer *et al.*[27]

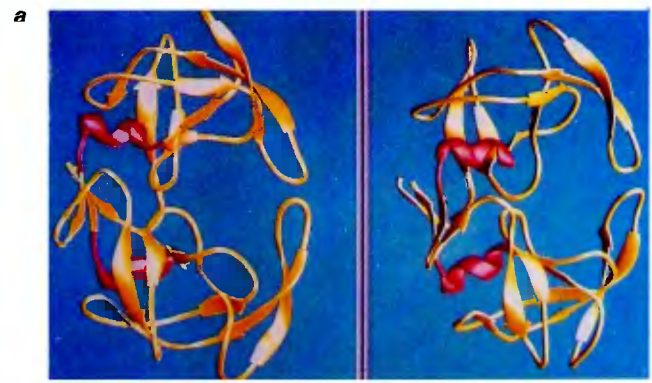

Figure 2. 3D structure from X-ray studies of RSV PR is shown on the left along with and HIV PR on the right. The monomer of RSV has 126 amino acids compared to 99 in HIV and the flaps on RSV PR do not exist in the X-ray structure, but have been added here merely for comparison.

Figure 3. Ribbon stereo diagram of the X-ray structure of (*a*) native HIV-1 PR and (*b*) HIV-1 PR complexed with MVT 101 inhibitor. Note that the flaps are opened in the native which closes upon the inhibition.

XPLOR program[13] followed by restrained least squares, using PROLSQ program[14], and manual rebuilding into 2Fo-Fc Fourier maps using the program FRODO[15]. While these studies initially disagreed about certain details of the structure, they all confirmed that the molecule is a homodimer, and that its active site closely resembles the active sites of other aspartic proteases such as pepsin, chymosin, and rhizopuspepsin, among others. The general topology of the HIV-1 PR molecule is similar to that of a single domain in pepsin-like aspartic protease, with the main difference being that the dimer interface in the HIV-1 PR is made up of four much shorter strands, rather than six longer strands present in pepsin. The overall structure is characterized by beta sheet with an extended four-stranded sheet forming the dimer interface (Figure 3 *a*). The active site triad (Asp25, Thr26, Gly27) located in a loop participates in a network of hydrogen bonds similar to that found in the other enzymes of the family (Figure 4). The carboxylic groups of Asp25 from both chains are nearly coplanar and show close contacts involving OD1 atoms. The network is quite rigid due to the interaction called 'fireman's grip', in which each Thr26 accepts a hydrogen bond from the Thr amide group in the other loop and donates a hydrogen bond to the carbonyl oxygen atom in the residue preceding the catalytic triad on the other stand. A difference between these related enzymes is that only one flap is present in pepsins, while two flaps are present in HIV-1 PR. The flap is a β hairpin that covers the active site and participates in the binding of inhibitors and/or substrates. All in all, the knowledge of the structures of many cellular aspartic proteases that has accumulated in the last 15 years greatly assisted in achieving rapid progress of the investigations of the structure and function of HIV-1 PR.

HIV-1 PR-inhibitor complexes

Since the structures of retroviral and cellular aspartic proteases share many features, inhibitors of cellular aspartic protease were tested early on for their ability to inhibit HIV-1 PR. Subsequently, a large number of complexes of HIV-1 PR with new substrate-based inhibitors have been prepared, and many crystal structures of these complexes have been solved. At least ten different structures of HIV-1 PR-inhibitors have already been published and other structures have been discussed at scientific meetings, with atomic coordinates becoming available either through the Protein Data Bank or by direct distribution from the investigators.

HIV-1 PR has two 99 residue peptide chains whose sequence is shown in Figure 5 and the monomer one is numbered 1 to 99 and monomer two is numbered 101 to 199. A typical Ramachandran plot (shown in Figure 6 for the native HIV-1 PR) for any HIV-1 PR indicates that the main chain torsion angles are all well within the normally allowed region. Figure 7 shows the temperature factors for all the residues of the PR and a PR-inhibitor complex. The B's for most of the main chains are well within 30 Å^2 which is normal for a well-refined structure at 3.0 Å or higher resolution. The flaps which are supposed to be involved in a sort of hinge motion, particularly when they have to accommodate the inhibitor, did not surprisingly have higher B factors. Residues 15–18, 37–40 and 68–70, in particular, which make up loops on the protease surface, are known to be involved in intermolecular interactions in the crystal structures[16, 17]. These residues also have the largest B factors. The mean B-factor of main chain atoms, averaged over protein domains, can give a quantitative estimate of chain flexibility[18].

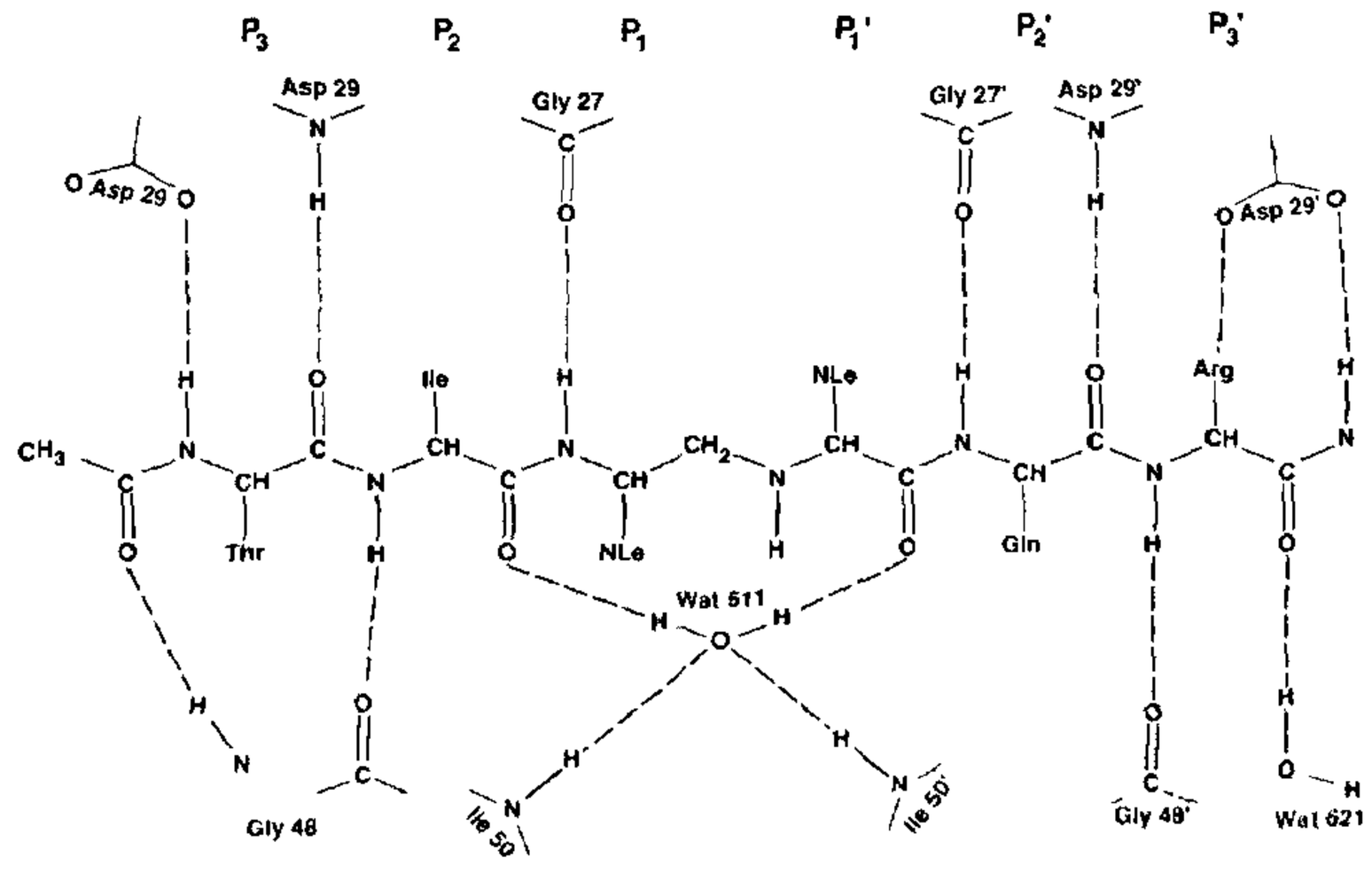

Figure 4. Hydrogen bonding scheme of HIV-1 PR with MVT-101 inhibitor complex. Major hydrogen bonds are indicated by broken lines. This basic H-bond scheme is maintained more or less the same as in other similar peptide inhibitors.

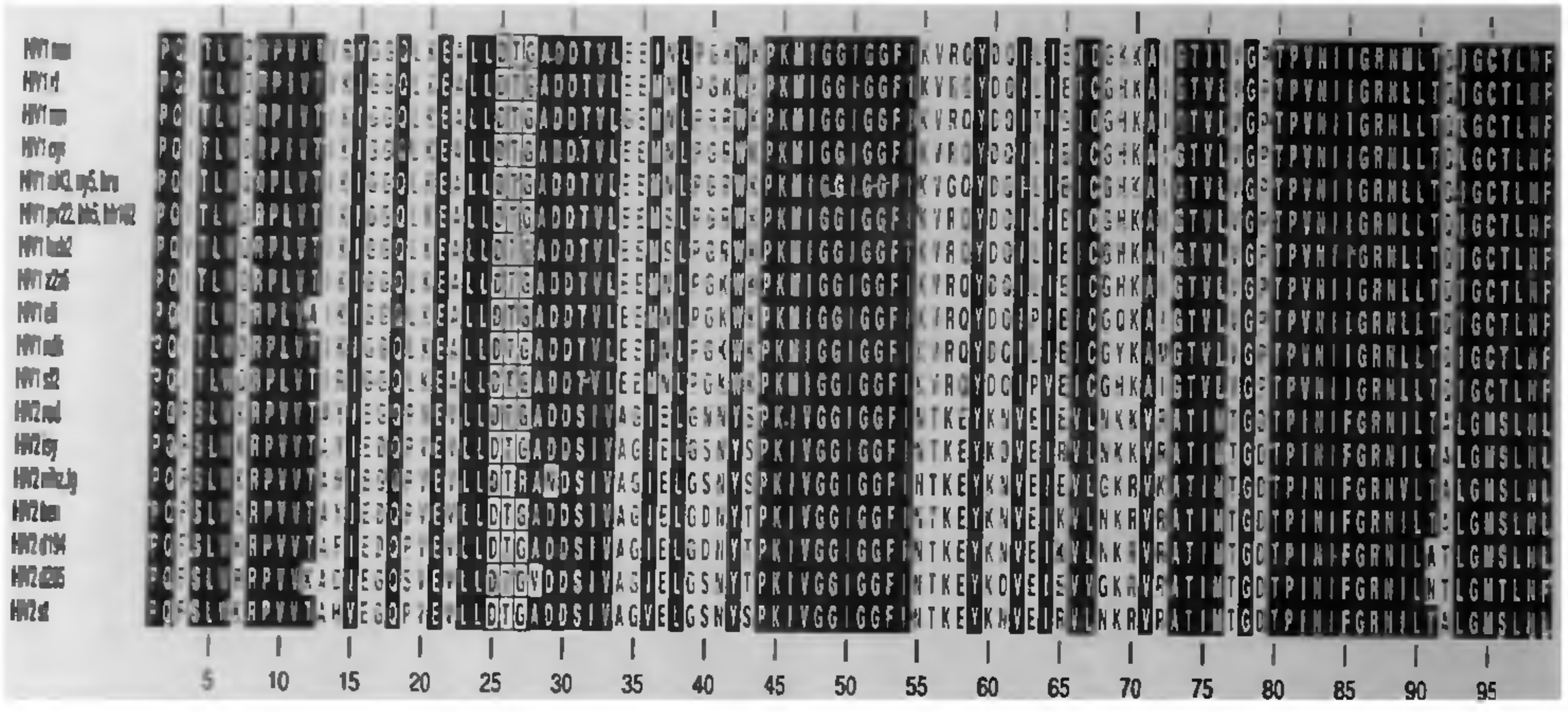

Figure 5. Comparison of the sequences of HIV-1 PR with HIV-2 PR. Residues conserved in proteases from both sources are shaded black, those conserved in each of the proteases are shaded grey, and residues which are not conserved are grey on white background. Those residues which appear to violate conservation are black on white.

Figure 8 shows the electron density map for MVT-101 inhibitor. Coordinate sets for ten PR-inhibitor complex structures available both from our laboratory as well as those solved in other laboratories are used as the basis of most of the comparisons described here. The sequence of inhibitors whose complexes were studied in our laboratory, as well as relevant crystallographic data, is given in Table 1, along with information on those structures received from other laboratories. These inhibitors are peptide analogues with different sequences and lengths, except for the UCSF8. Some of the inhibitors are nonspecific and able to interact with almost all aspartic proteases (pepstatin, acetyl-pepstatin), while others were designed on the basis of the sequences of known substrates of HIV-1 PR. They all bind to the protease in the same general conformation, making similar contacts with the enzyme. Different nonscissile groups replace a central peptide bond. These groups include an unusual amino acid, statine, as well as linkages involving reduced peptide bond, hydroxyethylamine, and hydroxyethylene moieties.

The methods used in our laboratory to solve the structure of the native HIV-1 PR as of the inhibitor complexes, have been published[10, 17, 19, 22]. A detailed description of the methods used to design a quasi-symmetric inhibitor, A-74704, and to solve its structure was published by Erickson *et al*[23]. The unofficial count on the number of HIV PR-inhibitor complexes studied so far at various laboratories within these couple of years, amounts to over 160, but many of them for various reasons have not been published. In Table 1, the three sets of coordinates (AG-2, AG-4, and Lilly-765) were provided by Krzysztof Appelt from Agouron Pharmaceuticals, La Jolla, California. While the structure of a complex of HIV-1 PR with acetyl-pepstatin was published by Fitzgerald *et al*[24], the coordinates used here were from an analogous structure, solved subsequently in our laboratory.

The resolution of the structures solved so far varies from 2.0 to 2.8 Å. Many of these structures were solved by molecular replacement, and therefore some features of the protease structure itself have been carried through by these techniques, rather than uniquely observed. In our experience, in almost all the cases of native as well as inhibitor complexes of HIV PR, the fit of the structure to the electron density map was always excellent for the final refined structure. There was also one common feature of all the HIV PR-inhibitor complexes so far studied, which is a strong peak of electron density, located between flaps and the inhibitor which has been designated as a water molecule. This unique water molecule is tetrahedrally coordinated by the two carbonyl oxygen atoms adjacent to the scissile bond analogue of the inhibitor and by two amide nitrogen atoms in the flaps. This water molecule mediates contacts between the tips of the flaps and the inhibitor.

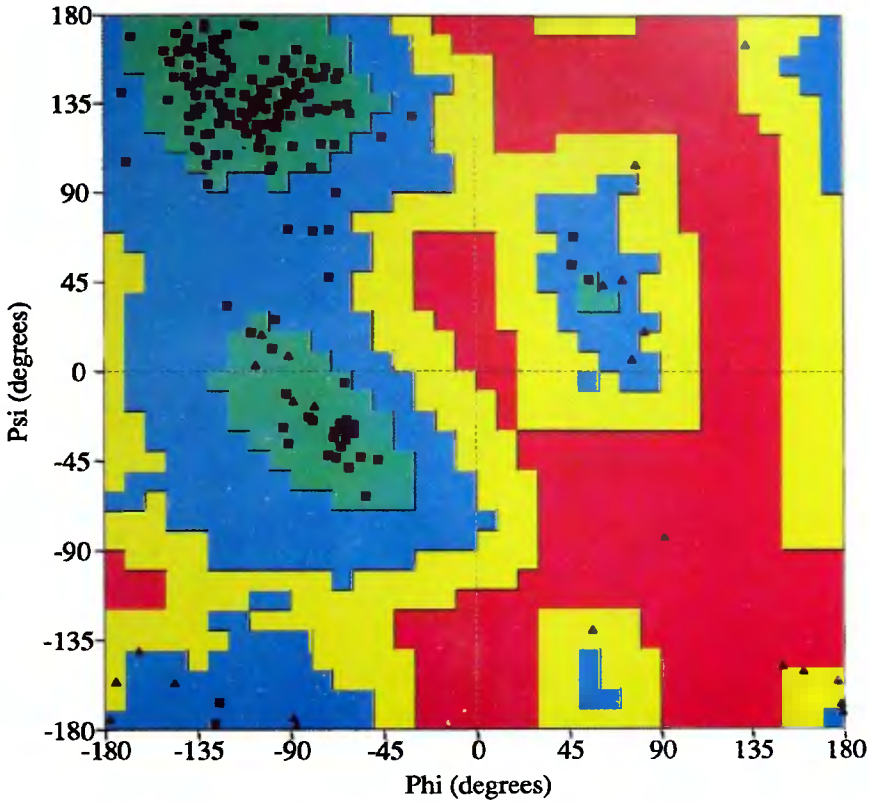

Figure 6. The Ramachandran plot for HIV-1 PR. The red region is the fully disallowed region, the yellow and blue partially allowed region and the green the fully allowed region and the dotted squares represent the position of the amino acids on the Φ–Ψ map

The complex of synthetic HIV-1 PR with a hexapeptide inhibitor with sequence N-acetyl–Thr–Ile–Nle–Ψ[CH_2–NH]–Nle–Gln–Arg–amide (referred to as MVT-101) was the very first HIV-1 PR inhibitor structure that was solved[10]. This asymmetric inhibitor lies in a single orientation and makes extensive interactions at the interface between the two subunits of the homodimeric protein. Later two more HIV PR-inhibitor complexes were studied at our laboratory[16,17]. These three inhibitors were hexa, hepta and octapeptide respectively and are designated as MVT-101, JG-365 and U85548e. The hexapeptide MVT-101 has a reduced bond in place of the normal scissile bond. The heptapeptide, JG-365, has a hydroxyethylene linkage and the octapeptide, U-85548e, has a hydroxymethylene linkage. Originally it was thought that MVT-101 has only one orientation but a new refinement with 2.0 Å resolution indicated conformational disorder, whereas such a disorder has not been observed for the JG-365 and U-85548e inhibitors. Similar disorder has been reported for the 2 Å structure of acetyl-pepstatin which was crystallized with HIV PR in a different crystal form[24].

Another HIV-1 PR complex studied recently in this laboratory is with the inhibitor CHa Ψ[CH(OH)CH-(OH)]-Val-Ile-Amp or designated[22] as U-75875. This has a dihydroxyethylene insert which mimics the transition state of substrate for the enzyme. Also a molecular modelling study was already done on this complex and

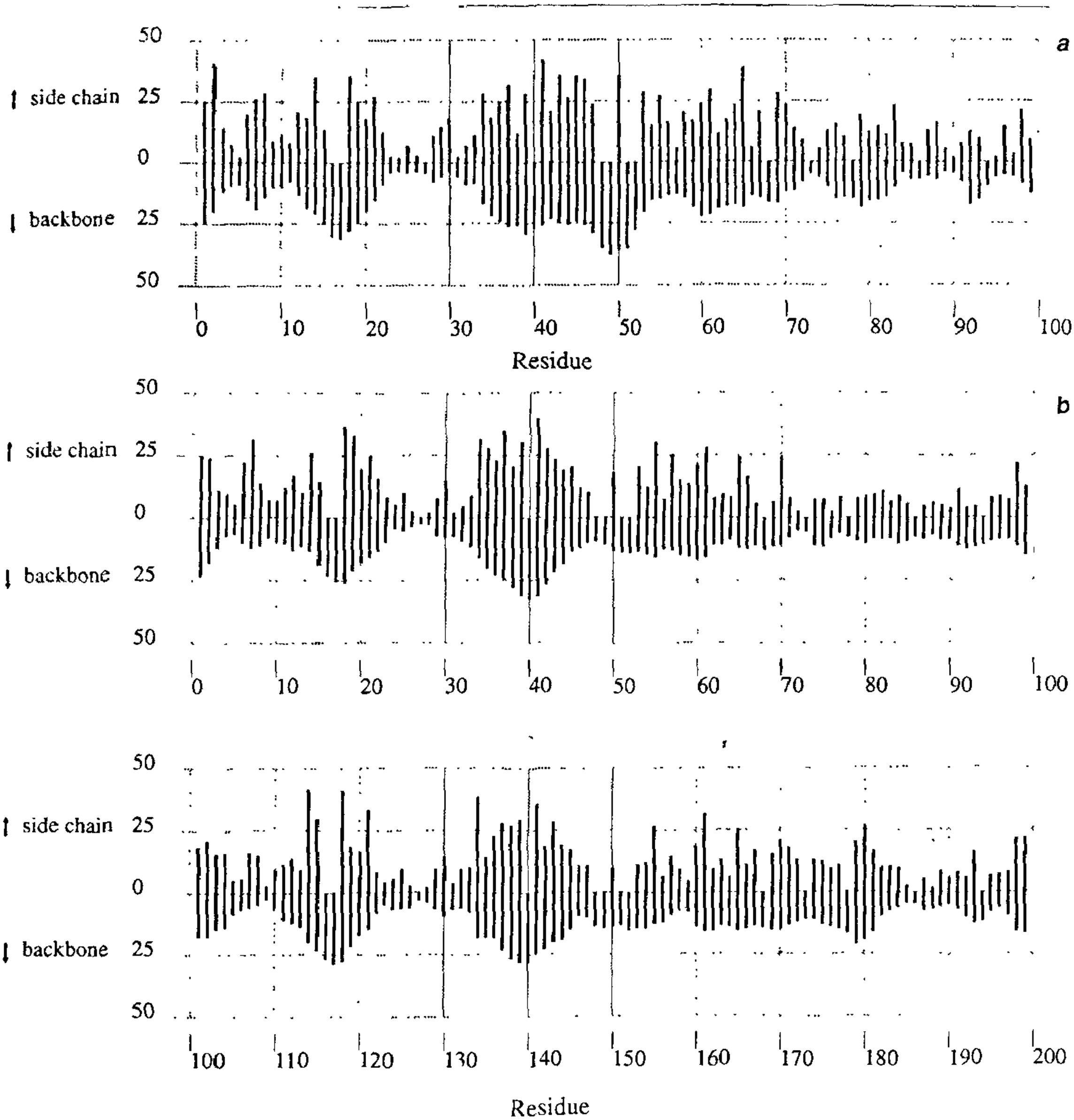

Figure 7. Plot of the isotropic temperature factor vs the residues for HIV-1 PR (*a*) native and (*b*) complexed with the inhibitor

the crystal structure analysis was the perfect way to test the correctness of that study. This inhibitor which is a diastereomer has the R configuration at both the hydroxyl chiral carbon atoms. One of the diol hydroxyl groups is positioned such that it forms hydrogen bonds with the active site aspartate, while the other interacts with only one of them. Like the other PR inhibitor complexes these interactions are also found in an extended chain in a well-defined and extensive active site cleft. Unlike the MVT-101 complex, the U-75875 does not appear to have another conformation other than the single conformation that was found.

All the initially studied inhibitor complexes crystallized in space group $P2_12_12_1$. The first deviation from this was the 2-fold (C_2) symmetric inhibitor of the HIV PR designed based on the 3-D symmetry of the enzyme active site[23]. This inhibitor (A-74704) binds to the enzyme in a highly symmetric fashion. The major difference is that the inhibitor possesses almost exact C_2 symmetry. The space group was $P6_1$. The water

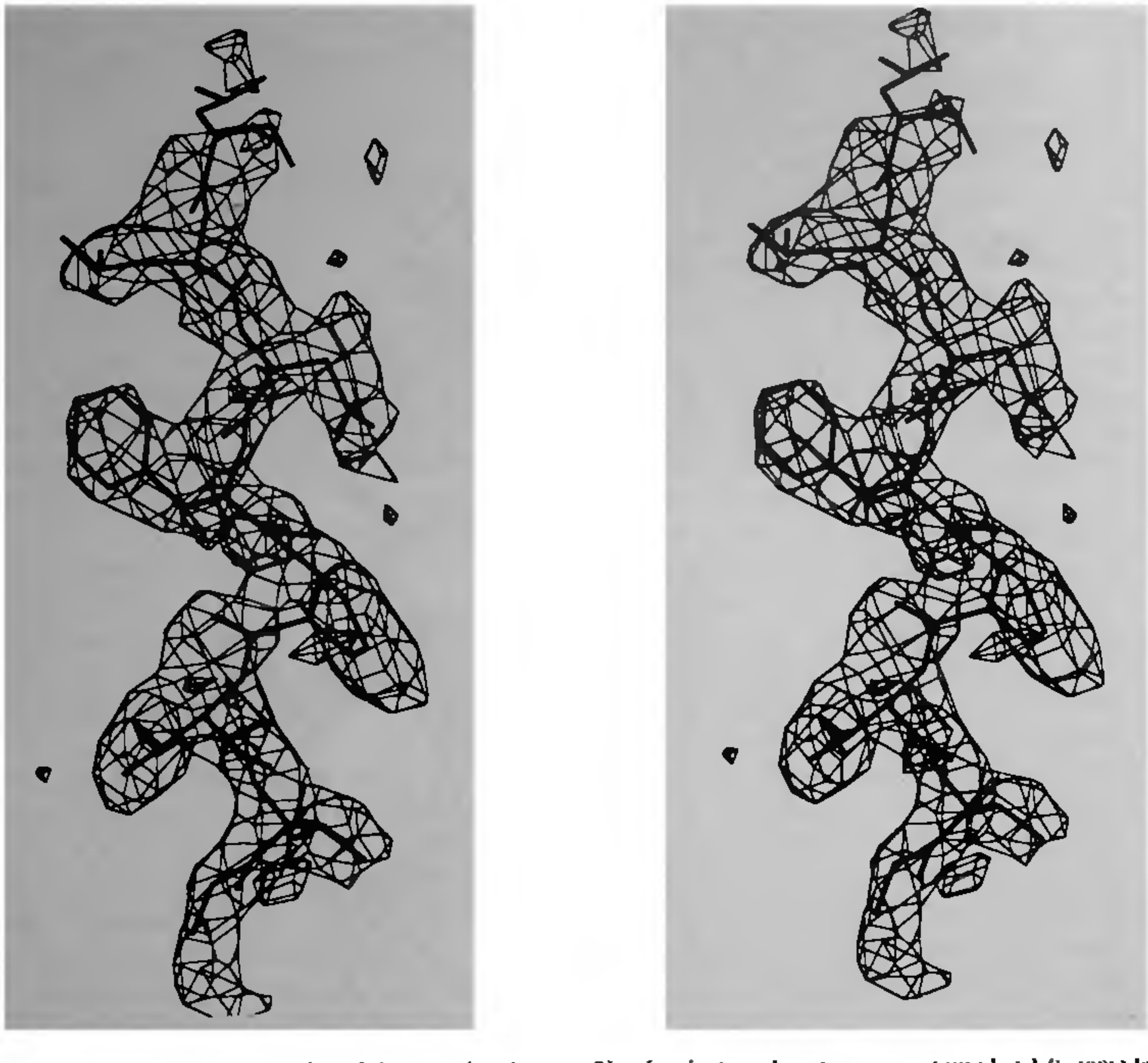

Figure 8. Stereo view of the inhibitor molecule in its 2I o–Fc electron density map contoured at 1 0 sigma level

WAT301 which bridges the inhibitor and the enzyme lies in this structure within about 0.2 of the molecular two-fold axis and is tetrahedrally coordinated to the amide NH atoms of ILE50 and ILE50′ of the flaps and to the carbonyl oxygens of the P2 valyl groups of A-74704 All in all it appears that HIV-1 PR attempts to enforce symmetric binding even with a structurally asymmetric inhibitor. Their study of this particular enzyme complex bad two different goals One was to provide symmetric interactions between the inhibitor and enzyme, which would make the C2 axes coincide. Secondly, the inhibitor should fill the enzyme subsite's S1 and S1′ which usually interact with the side chains if the inhibitor is asymmetric. This inhibitor is C_2 symmetric except for the secondary OH groups on the central atom The initial design resulted in an inhibitor A-74702, which was considered earlier having weaker activity *in vitro* This led the investigators to the A-74704 inhibitor which had a carbobenzoxy valine (Cbz Val) attached to both ends of the core This inhibitor in that sense was different from MVT-101 which is an asymmetric, reduced peptide inhibitor Another difference between these studies was that in the MVT-101 structure the protein was synthetic, whereas this one was recombinant and different from the synthetic protease at six positions. Even with these differences, surprisingly, this inhibitor A-74704 fitted into the active site in the same fashion as MVT-101, with almost the same hydrogen bonding scheme. A total of 29 residues of the PR interacts with the inhibitor compared to 21 residues interacting with the MVT-101 inhibitor. Overall, it suggests that the PR which itself is asymmetric forces a symmetric binding even with the asymmetric inhibitor

The X-ray structure of HIV-1 PR complexed with another inhibitor, acetyl-pepstatin studied by Fitzgerald *et al* [21], proposed two symmetric orientations for the inhibitor, which are related to one another by pseudo-2-fold axis of symmetry of the subunits of the dimeric enzyme Assuming equal temperature factors of both of the inhibitor positions, the occupancy was 0 58 and 0 42 respectively Although the authors refer to some major changes in the structure of the HIV-1 PR itself compared to that of the MVT-101 complex, it appears that these changes do not affect the way the PR interacts with the inhibitor. This was mainly because the amino

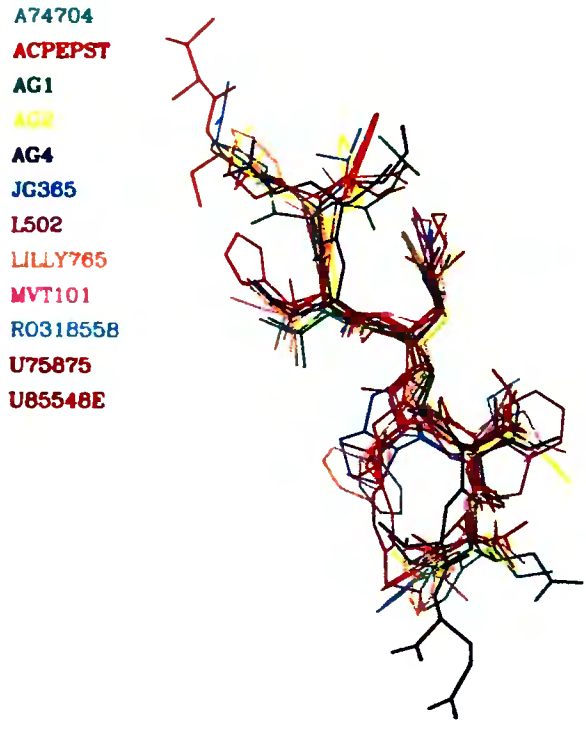

Figure 9. Overlay of 12 inhibitors as obtained from X-ray studies.

acids in the active site in both the structures remained in almost the same conformation.

After the first C_2 symmetric inhibitor was studied, Bone and coworkers[25] designed a new set of pseudo-symmetric inhibitors, using similar strategy. The structures are shown in Figure 9. Once again, it was observed that the inhibitor can be fit into the difference density in two orientations. These orientations differed by rotation of the inhibitor in the active site by 180° about an approximate 2-fold axis which relates the 2 HIV PR subunits. It was also concluded that both these conformations are equally possible for these inhibitors in this complex. These types of studies indicate that the direct hydrogen bonding and indirect hydrogen bonding through water, of the inhibitor with the HIV PR can be improved by suitable modifications of the inhibitor. Such studies lead to better understanding and easier design of the drug that inhibits the activity of HIV inhibitor.

Only one non-peptidic inhibitor of HIV PR has been described so far. This is a derivative of a compound discovered through computational screening of the Cambridge structure database using a shape complimentary algorithm. The main inference between this and the peptide inhibitors is that this inhibitor binds within the enzyme active site but does not overlap the binding site observed for peptide-based inhibitor. A computer search was conducted of the database for 3D compounds based on shape and also to some extent on possible chemical complementarity with the active site of HIV PR. It is interesting to note that the UCSF8 inhibitor binds in a different orientation than predicted by the DOCK program for a closely similar molecule bromperidol. Also, the conformation of UCSF8 in the complex is somewhat different than the structure of bromperidol in the Cambridge structural database. This indicates the strong complex nature of the force that is encoded in the active site of the HIV-1 protease.

Another aspect to note in the case of UCSF8 was that in the crystal structure, the B value for the inhibitor was about 70 Å^2, while the expected B values for protein will be around 30 Å^2.

Differences and similarities among the HIV-1 protease molecules

The structure of the protease itself is very similar among the various protease–inhibitor complexes, despite differences in the crystal parameters. Among all the structures of HIV PR inhibitor studied, more similarities have been observed than differences. As the structures are visually compared using computer graphics, the notable differences among them are not only in side chain positions at the surface of the protein, as expected, but also in some rearrangements in torsion angles of part of the main chain. These differences are at the outer edge of the protease, in the areas involved in crystal lattice contacts, but are distant from the active site, so they have no apparent bearing on inhibitor binding.

In most of the structures, the peptide bond at the tip of the flap between residues 50 and 51 is turned 180° in molecule 1 relative to its position in molecule 2. With the exception of AG-4, the torsion angles identifying monomer 1 are Φ in the range of –20° to –45° and Ψ between –60° and –95°. In monomer 2, Φ is in the range of 120° to 160° and Ψ is between 75° and 120° These are details that may not accurately reflect in the lower-resolution structures. Eventually, high-resolution structures of PR–inhibitor complexes may provide final proof of whether this asymmetry is necessary for binding all inhibitors.

Upon binding of inhibitors, the structure of HIV-1 PR changes considerably, compared to native protease. In the structures discussed here, the inhibitor enzyme has undergone a transition compared to the apoenzyme, which is particularly visible in the flap region (Figure 3 *b*). However, the changes are not limited to the flaps only, and altogether, almost a quarter of α-carbon atoms shift by more than 1 Å each upon inhibitor binding. However, as mentioned above, the structure of the protein part of the complexes is quite similar for all inhibitors studied.

Conformation of the inhibitors of HIV-1 PR

Almost all the inhibitors so far reported by various groups are bound in the protease active site in an extended conformation, so that when they are superimposed upon one another, the functional elements align quite similarly (Figure 9). The contacts between the main chain of the inhibitor and the protease are very similar for all the complexes, with the nonhydrolyzable scissile bond analogue of each inhibitor aligned with the aspartate carboxyl groups (Asp 25/125) of the active site. In all cases except for MVT-101 (which has a reduced peptide bond as a nonscissile moiety), the hydroxyl group at that P1–P1′ junction is positioned between the aspartate carboxyl groups in the protease, within hydrogen bonding distance of each carboxylate oxygen. The configuration of the tetrahedral carbon of the inhibitors is S, with the notable exception of inhibitor U-75875, although the R configuration of this inhibitor results from the naming convention and its linkage is actually isostructural with the linkages involving S carbons in the other inhibitors.

HIV-1 PR is a symmetric enzyme, but loses its symmetry when an inhibitor is bound. This is induced by the N > C polarity of the inhibitor peptide. The differences between monomers 1 and 2 are so subtle that a two-fold disorder with respect to the inhibitor is frequently observed. The comparison of HIV-1 PR complex with the native enzyme, indicates that large conformational changes accompany ligand binding. In the HIV-1 PR complex with MVT-101, the entire flap region has moved considerably relative to the position observed in the native HIV-1 PR (Figures 3 *a* and *b*[6, 20]). This movement for all the C_α atoms of residues 35 through 57 is between 1 Å and 7 Å. This may be visualized as an overall hinge motion, wherein beta-strands in the flap region rearrange themselves after opening to permit substrate or inhibitor binding. Also because of the inhibitor/substrate binding, the overall size of the active site cavity is reduced, due to the movement of the residues 77 to 82, whereas the c_α atoms shift by about 2 Å. These changes explain the wide range of substrates that can be accommodated in the active region. This is the most important mechanism that must be understood for rational drug design.

Molecular modelling and HIV-2 protease

Apart from a large number of X-ray studies of HIV-1 PR inhibitor that are being undertaken at various laboratories, another area of study that is rapidly progressing is the molecular modelling of this protease with suitable inhibitors. Although X-ray crystallographic investigations of biological molecules unequivocally give the 3D structure of the molecules, nevertheless, in the actual biological system, the enzymes may exist in conformations somewhat different. Also, due to the difficulty and time-consuming process of growing these crystals and solving the structures, molecular modelling is the best alternative to actual X-ray crystal structure determination of macromolecules. In fact the starting set of coordinates for the determination of the crystal structure of the native HIV-1 PR was from the modelled set of coordinates of HIV-1 PR based on the crystal structure of the RSV protease.

In the modelling studies of the U-75875 inhibitor, the inhibitor was constructed using the MVT-101 structure, using graphics and energy minimization techniques. The resulting structure kept the hydrogen bonding scheme between the PR and inhibitor as a constraint and led to an asymmetric position of the two hydroxyls. However, this and most other inhibitors were asymmetric, while the apoenzyme was fully symmetric. This observation led to design of another PR-inhibitor, namely the A-74704 from the Abbott laboratory. Three-dimensional fingerprint analysis using the program ALADDIN, was also used to discover novel, nonpeptidic inhibitors of HIV-1 protease from existing database of small molecules. Approximately 40 compounds were selected from the hit list and tested for HIV-1 protease inhibition. The water molecule WAT301, which is so far a common feature of all the HIV-1 PR inhibitor complexes studied, must play an important role in the inhibitor binding because of its maximum hydrogen bonding capability. Another area to explore is to find a covalently bound group to replace this water and test for the activity of the protease. The current goal of all these investigations is to find a potential (?), either peptidic or nonpeptidic, which will inhibit the action of HIV-1 PR. Scientists all over the world are putting their effort to find an answer to this important problem and it would appear that the discovery of such an important drug is not far away.

Although AIDS is caused mainly by the type 1 (HIV-1) virus, the disease is sometimes caused by the second variant, HIV-2. Recently there have been reports of some unusual strains of the AIDS virus which may be mutants of the actual HIV-1 virus. Although these two viruses have a common evolutionary origin, they differ in their nucleotide and amino-acid sequences, with 60% overall amino acid homology between these two. A brief description of the methodology used in modelling HIV-2 PR based on the existing HIV-1 PR is given here. The sequence for the HIV-1 PR was extracted from the Los Alamos HIV sequence database and the atomic coordinates of the HIV-1 PR with inhibitor JG-365 was used for mutating the residues. After replacing, deleting and inserting the residues of HIV-2 PR on to the HIV-1 PR structure, the bad contacts were relieved manually on the graphics. Basically the original HIV-1 PR conformation was maintained for the HIV-2 PR. The atomic coordinates of the second monomer in the dimer were obtained from the first one by superposition on the c_α atoms of the second monomer in the crystal structure of HIV-1 PR. In addition, the inhibitors were also modelled binding into the active site area of the protease. All the substitutions of residues occur in such a way that the important structural interactions are maintained. The examination of the protease structures with two different inhibitors identified one residue in the substrate binding site where the size of the amino acid present in HIV-1 PR correlated with the relative binding constants of the inhibitors.

Recently the X-ray structure of native HIV-2 PR native as well as complexed with two different peptidomeric inhibitor has been determined[26] and an rms deviation of 1.5 Å was found between the main chain atoms and the conserved side chain atoms of the HIV-2 and HIV-1 PR structure. The complexes crystallized in the space group $P2_12_12_1$ with cell dimensions a = 33.28 Å, b = 45.35 Å and c = 135 84 Å. The major differences were only at residues 16–20, 33–44 and 66–72 in each monomer which occur at the external flap segments. Also, most of the hydrogen bonds were maintained similar to those found in the HIV-1 inhibitor structure. The single most important difference of HIV-2 PR from the HIV-1 PR structure was that the flaps are closed in the former case whereas they are open in the HIV-1 PR. The flaps do not have that hinge motion in the case of HIV-1 PR when its active site is occupied by the inhibitor. In the X-ray structure of HIV-2 PR inhibitor complex with U-75875 the inhibitor and PR contacts were almost the same as those found in the case of HIV-1 PR and U75875 complex.

Currently, 12 million people all over the world have been infected by the HIV virus and an estimated 2 million people are developing the disease by the end of this century. Although several drugs like AZT and ddI have been partially successful in controlling this disease, the search for more effective drugs is still going on. HIV PR is an obvious choice for a target of such drugs, and hence a lot of money and efforts are being spent on their search. X-ray crystallography, NMR and molecular modelling are being extensively used for studying the HIV PR inhibitor complexes. Due to the speed with which these studies are going on at various laboratories, the hope for a rationally designed drug for AIDS therapy may not be very far. Also the recent structure of the RT molecule will definitely help find a proper cure for this deadly disease of the humankind.

Summary

As stated earlier, there are more than 160 different inhibitor complexes of HIV-1 PR (and a few of HIV-2 complexes recently solved) that have been studied so far by X-ray crystallographic methods at various laboratories. This is the first time that the structure of one single enzyme has been studied by such a large number of independent investigators, in a variety of crystalline modifications. This gives us the unique opportunity of studying various aspects of protein–ligand binding, hydrogen bonding scheme, nonbonded interactions and a full insight into the structure of enzymes and its interactions with a variety of inhibitors that hitherto was not possible. Although these studies have been made in such a short time, we still do not have a rationally designed drug which inhibits the protease activity of the HIV-1 PR in the real system without complications of

side effects. There are still several unanswered questions. Possibly, with the understanding of the structure of the other two important vital components namely the RT and IN, we may go one step further in this endeavour. Structural information will also benefit the design of vaccines for AIDS.

Due to the extensive data being collected on this complex, Paula Fitzgerald[11] presented at the recent ACA Meeting at Pittsburgh, a poster wherein she outlined some guidelines for publications of the HIV PR inhibitor complexes. We refer to some of them along with some additional comments of our own: (i) prepare tables that describe the quality and completeness of the data and the geometry of the final refinement model; (ii) prepare tables describing hydrogen bonding and nonbonded contacts between the protein and the inhibitor; (iii) prepare the standard set of stereo drawing for each inhibitor structure as well as a close-up view of the active site; (iv) prepare tables comparing the geometry of the PR and inhibitor portion of the family of structures. In addition (v) the final temperature factors for each inhibitor should be published, which gives us some idea of how much the inhibitor position is distorted (vi) The final PDB coordinates should be made available as soon as the structure is published through depositing at the Protein Data Bank. (vii) As indicated above only 10% of the HIV PR inhibitor complexes have so far been published and many of the rest may not be published at all. This is mainly due to the fact that some companies may need to try these inhibitors for clinical test and would like to modify, if needed until a suitable inhibitor is found. Still in the interest of the scientific community working in this area, these data should be made public for a free flow of information among scientists.

1 Narayan, D and Clements, J E, *J Gen Virol*, 1989, **70**, 1639–1717
2 Mitsuya, H and Broder, S, *Nature*, 1987, **325**, 649–773
3 Kohlstaedt, L A, Wang, J, Friedman, J M, Rice, P A and Steitz, T A, *Science*, 1992, **256**, 1783–1790
4 Arnold, E *et al*, *Nature*, 1992, **357**, 85–89
5 Hostomska, Z, Matthews, D A, Davies, J F, Nodes, B R and Hostomsky, Z, *J Biol Chem*, 1991, **266**, 697–702
6 Wlodawer, A, Miller, Jaskoloski, M, Sathyanarayana, B K, Baldwin, E, Weber, I T, Selk, L M, Clawson, L, Schneider, J and Kent, S B, **II**, *Science*, 1989, **245**, 616–621
7 McKeever, B M, Navia, M A, Fitzgerald, P M D, Springer, J P, Leu, C T, Heimbach, J C, Herber, W K, Sigal, I S and Darke, P L, *J Biol Chem*, 1989, **264**, 1919–1921
8 Navia, M. A, Fitzgerard, P M D, McKeever, B M, Leu, C T, Heimbach, J C, Herber, W K, Sigal, I S, Darke, P. L. and Springer, J P, *Nature*, 1989, **337**, 617–620
9 Weber, I T, *Gene*, 1989, **85**, 567–569
10 Miller, M, Schneider, J, Sathyanarayana, B K, Toth, M V, Marshall, G R, Clawson, L, Selk, L, Kent, S B H and Wlodawer, A, *Science*, 1989, **246**, 1149–1152
11 Fitzgerald, P M D, Abstract American Crystallographic Association 50th Annual Meeting, Pittsburgh Diffraction Conference, 9–14 August 1992, University of Pittsburgh, Article No PB31
12 Sussman, J L, *Methods Enzymol*, 1985, **115**, 271–276
13 Brunger, A, Kurian, J and Karplus, M, *Science*, 1987, **235**, 458–460
14 Hendrickson, W A, *Methods Enzymol*, 1985, **115**, 252–270.
15 Jones, T A, *J Appl Crystallogr*, 1978, **11**, 268–272
16 Swain, A, Miller, M M, Green, J, Rich, D H, Schneider, J, Kent, S B H and Wlodawer, A, *Proc Natl Acad Sci USA*, 1990, **87**, 8805–8809
17 Jaskoloski, M, Tomasselli, A G, Sawyer, T K, Staples, D G, Heinrikson, R L, Schneider, J, Kent, S B H and Wlodawer, A, *Biochemistry*, 1991, **30**, 1600–1609
18 Westof, E, Altschuh, D, Moras, D, Bloomer, A C, Mondragan, A, Klug, A and Regenmortel, M H V, *Nature*, 1984, **311**, 123–126
19 Wlodawer, A, Miller, M, Swain, A L and Jaskolski, M, *Methods in Protein Sequence Analysis* (eds Hoog and Gustacsson) 1991, pp 215–221
20 Miller, M, Swain, A L, Jaskolski, M, Sathyanarayana, B K, Marshall, G R, Rich, D, Kent, S B H and Wlodawer, A, in *Retroviral Proteases* (ed Pearl, L H), Stockton Press, NY, 1990, pp 93–106
21 Swain, A L, Gustchina, A and Wlodawer, A, *Structure and Function of the Aspartic Proteinases*, 1991, pp 433–441
22 Thanki, N, Rao, J K M, Foundling, S I, Howe, W J, Moon, J B, Hui, J O, Tomasselli, A G, Heinrikson, R L, Thaisrivongs, S and Wlodawer, A, *Protein Sci*, 1992, **1**, 1601–1625
23 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, Paul, D A and Knigge, M, *Science*, 1990, **249**, 527–533
24 Fitzgerald, P M D, McKeever, B M, VanMiddlesworth, J F, Springer, J P, Heimbach, J C, Leu, C, Herber, W K, Dixon, R. A F and Darke, P L, *J Biol Chem*, 1990, **24**, 1309
25 Bone, R, Vacca, J P., Anderson, P. S and Holloway, M K., *J Am Chem Soc*, 1991, **113**, 9382–9384
26 Mulichak, A M and Watenpaugh, K D, Abstract American Cystallographic Association 50th Annual Meeting, Pittsburgh Diffraction Conference, 9–14 August 1992, University of Pittsburgh, Article No WO3
27. Dreyer, G B, Lambert, D M, Meek, T D, Carr, T J, Tomaszek, T A Jr, Fernandez, A V, Bartus, H, Cacciavillani, E, Hasselt, A M, Minnich, M, Petteway, S R Jr and Metcalf, B W, *Biochemistry*, 1992, **31**, 6646–6659

ACKNOWLEDGEMENT The work presented in this article is the result of a collaborative work of many colleagues Particular mention should be made of M Miller, A Swain, J K M Rao, N Thanki, I Baldwin, M Jaskolski, I Weber, A Gustchina, C Sansom and S Foundling The research was sponsored by the National Cancer Institute, DHHS under contract no NO1-CO-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 it mention trade names, commercial products, or organizations that imply endorsement by the U S Government

Received 20 April 1993, accepted 5 May 1993