Volume 247, number 1, 118–122 FEB 06999 April 1989

Possible role of some groups in the structure and function of HIV-1 protease as revealed by molecular modeling studies

I.V. Pechik, A.E. Gustchina, N.S. Andreeva and A.A. Fedorov

V.A. Engelhardt Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow, USSR

Received 13 February 1989

Retroviral proteases belong to the class of aspartic proteases. A molecular model of HIV-1 protease has been built on the basis of the consensus template specific for the domains of these enzymes. The template region comprises more than a half of the HIV-1 protease monomer structure, it includes the active site, formed at the junction of the two monomers, binding pockets of the enzyme, and some other molecular segments. These regions can be more conveniently described than other parts of the structure. Some properties of the HIV-1 protease molecule are discussed, as well as of probable inhibitors. The properties of the model structure are in good agreement with the recent results of crystallographic studies of Rous sarcoma virus protease.

Retroviral protease; HIV-1; Three-dimensional structure

Retroviral proteases belong to the class of aspartic proteases [1–3]. A molecule of ordinary aspartic protease consists of one polypeptide chain forming two topologically similar domains related by an approximate 2-fold symmetry; its active site is created at the junction of the domains and includes the segments containing the Asp-Thr-Gly-Ser(Thr) sequence in each domain. As retroviral proteases have a molecular mass comparable with the mass of one domain of ordinary enzymes, it has been realized that they can be active as dimers [4], i.e. their molecules must dimerize to create an active site similar to that of other aspartic proteases. A hypothetical scheme for the dimeric structure of retroviral protease has been described by Pearl and Taylor [4].

We have built an atomic model for HIV-1 protease (HIV-1 PR) [5] on the basis of the structural template for aspartic proteases [3]. Atomic coordinates of penicillo-, rhizopus- and endothiapepsin from the Protein Data Bank [6] have been used for these studies, the N-terminal domain has been taken as a basis for the monomer atomic coordinates of HIV-1 PR, while the atomic coordinates of the second monomer in the dimer were obtained from the first one with the use of symmetrical operations around the 2-fold axis.

Fig.1 shows the alignment of amino acid sequences of HIV-1 PR pepsin and penicillopepsin how it follows from the structural template [3], and corresponding double-layer sheet regions of these proteases. It includes the central intradomain double-layer sheet, containing the two A- and the two B-loops [7] with the hydrophobic nucleus between the layers (fig.2). Structural elements which are not included in the template region (the first D-loop of the domain, the small C-helix [7], the N- and the C-terminal strands as well as some connecting segments) have been built in accordance with their position in the three-dimensional structure of ordinary aspartic proteases.

The first outer A-layer of this sheet contains two antiparallel hairpins which include residues 11–21 and 61–73. In accordance with the template the first and the third positions before the turn (in terms of Sibanda and Thornton [8]) are occupied in these hairpins with hydrophobic residues: Ile-13, Ile-15 in the first loop and Ile-64, Ile-66 in the

Correspondence address: I.V. Pechik, V.A. Engelhardt Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow, USSR

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

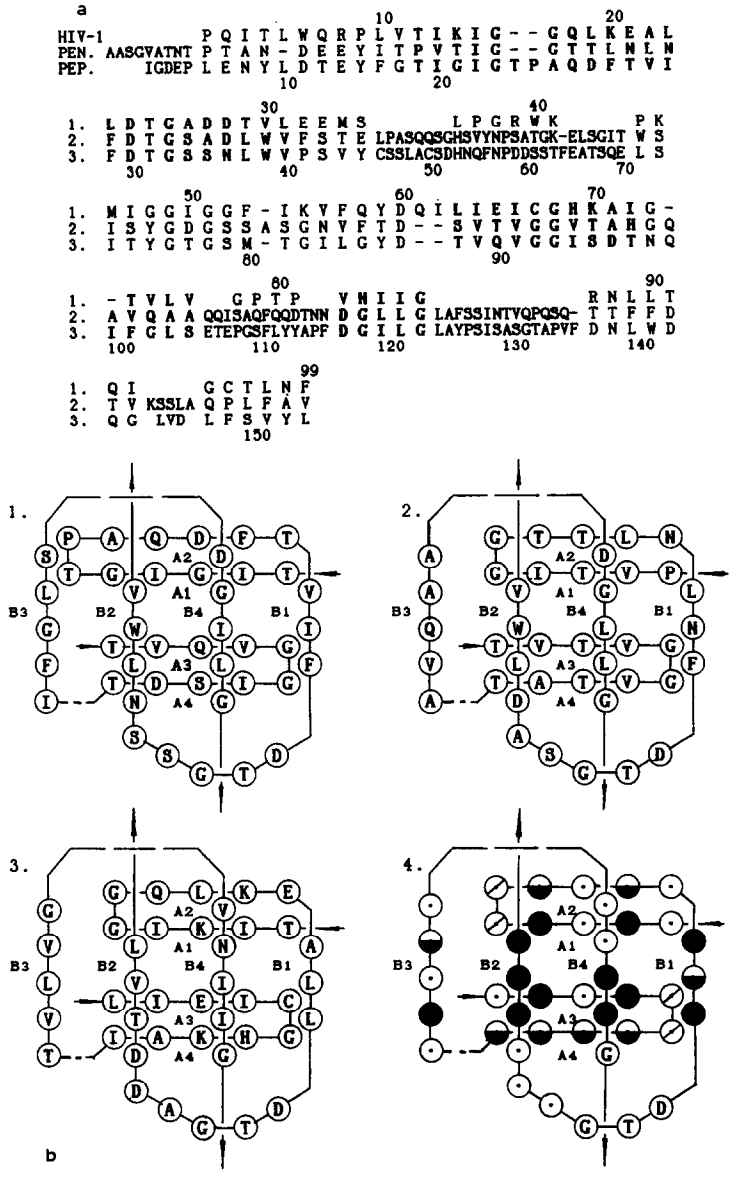

Fig.1. (a) Alignment of HIV-1 PR amino acid sequence with amino acid sequences of penicillopepsin (PEN) and pepsin (PEP). The template regions are indicated with bold letters. (b) Amino acid sequences of chain segments comprising the intradomain double-layer β-sheet in porcine pepsin (1) and penicillopepsin (2) in comparison to the probable location of the same sheet in polypeptide segments of HIV-1 proteinase (3). The scheme representing the consensus template for the pepsin fold [3]: ●, hydrophobic residues; ⊙, hydrophilic residues; ◒, the position without strict requirements; ∅, residues located usually at the turns of β-hairpins (4). The approximate mutual arrangement of the A- and B-strands in the structure is shown.

second. No requirements are imposed on residues of the C-terminal strands of these loops. The second B-layer is composed of two crossing wide loops which form an antiparallel system of two pairs of parallel strands with a 'wedding ring topology' [7]. In accordance with the template the first B-loop (residues with numbers 22–33) contains hydrophobic residues in the first and third positions before the turn (Leu-24, Ala-22), whereas the second B-loop (residues 74–84) has a hydrophobic residue only at the third position before the turn (residue Val-75). After the turn, the

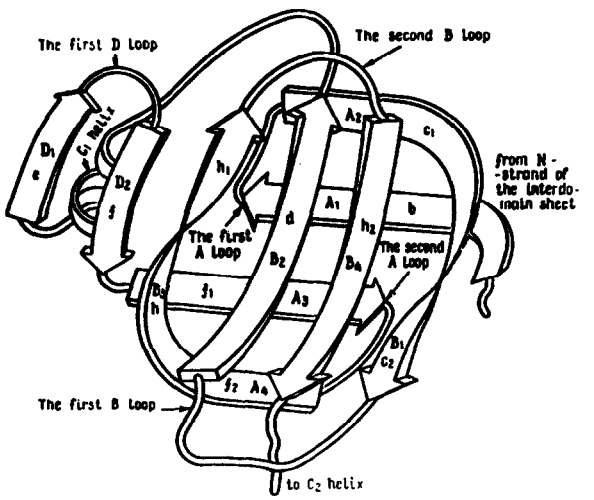

Fig.2. The scheme of the arrangement of the secondary structure elements in domains of aspartic proteinases [7]. The subscripts 1 and 2 correspond to the elements of the first half of the domains, subscripts 3 and 4 to the symmetrically related second half. The orientation and the position of the first D-loop is different in the N- and C-terminal domain. The lower-case letters correspond to the notation of segments used in [9].

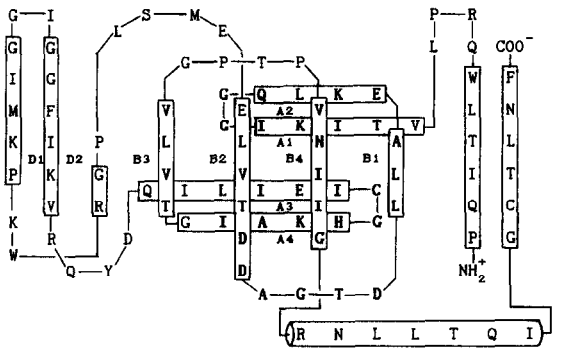

Fig.3. The scheme of the secondary structure HIV-1 PR from [5]. The segments in frames are the β-strands, helical region is shown as cylinder. Template regions are indicated with bold letters.

first B-loop must have hydrophobic residues in the third and the fifth positions. However, the C-terminal strand of the first B-loop HIV-1 PR has a hydrophobic group only in the fifth position (Leu-33) whereas the third one is occupied with Thr-31. This seems quite permissible as the turn of the first B-loop in retroviral proteases is partly exposed to solvent. The two hydrophobic residues followed with glycine located at the third, fourth and fifth positions after the turn of the second B-loop are also presented in HIV-PR. They are Ile-84, Ile-85 and Gly-86.

The aforementioned hydrophobic residues of the A- and the B-layers form a hydrophobic nucleus standard for the domains of aspartic proteases. Some groups make an additional contribution to this nucleus, such as Val-11, Ile-62, Ile-72, Leu-89, Leu-90.

If, during model building, one uses atomic coordinates of the N-terminal domain of pepsins, the first D-loops, which are out of the template region, overlap in the dimeric structure. It has made us change their configuration slightly. The scheme of the secondary structure of HIV-1 PR is presented in fig.3.

The region of the contacts between monomers has no equivalent in the structure of ordinary aspartic proteases. They have the interdomain six-stranded antiparallel sheet. In our model the inter-monomer sheet contains only four strands.

The active site of the retroviral proteases as well as the primary and the secondary substrate-binding pockets are formed from the chain segments which belong almost completely to the template region. Therefore the description of these important fragments of a molecule can be done with more convenience than that of the other regions of the molecular structure. The active site of these enzymes is characterized with a complete symmetry, as it is formed from similar segments of two identical monomers, arranged around the two-fold axis. All hydrogen bonds specific for the active site of ordinary aspartic proteases are presented in this enzyme except the bonds between the hydroxyl groups of Ser(Thr) at the fourth position of the active loop turn and the active aspartic acid residues (Asp-32...Ser-35 and Asp-215...Thr-218 hydrogen bonds in pepsin) as this position is occupied with Ala in HIV-1 PR.

On the basis of this model the first approximate analysis of the structural grounds for the enzyme specificity can be performed. If a molecule of the retroviral protease is completely symmetric, then any one of the two symmetrical positions for the substrate binding is equally probable in relation to the identical monomers P and Q. Therefore the notation of binding pockets as S_n and as S_n' is rather formal. It is of interest that the position of the binding site for a cation or a water molecule

equidistant to the active carboxyls found in other aspartic proteinases [9] is very close to the axis of the molecular symmetry in this case.

The positions of residues forming binding pockets of HIV-1 PR have been found in accordance with the standard positions of residues in substrate-binding sites of ordinary aspartic proteases. The S_1-binding site is formed with the hydrophobic residues Leu-23 and Ile-84. The D-loop residues also contribute to the S_1-binding pocket of ordinary enzymes, in the case of retroviral enzymes the role of this 'flap' cannot be understood until their three-dimensional structure is solved; however, it is probable that the hydrophobic residue Ile-50 at the turn of the 'flap' takes part in the formation of the S_1' pocket providing proline fixation. The presence of Asp-30 and Asp-29 residues at the B_3-strand is an important feature of HIV-1 PR, relevant to the specificity. These residues contribute to the formation of S_2- and S_3-pockets. The S_2-pocket also contains hydrophobic residue Ala-28; the contribution of the connecting segment between the N-terminal strand and the A_1-strand to S_3-binding pocket seems probable. It includes Trp-6, Gln-7, Arg-8 and Pro-9. The S_4-binding pocket can locate only on the surface of a molecule, and the position of this site is not quite definite. Table 1 contains amino acid sequences of gag-pol HIV-1 polyprotein segments which are split with HIV-1 PR. It can be seen that substrates containing a proline residue in the P_1'-position are characterized by the specificity for their N-terminal part: the P_1 position is occupied with bulky hydrophobic residues, there is the Asn residue in the P_2-position, the P_3-position is occupied with residues containing a long hydrophobic part, and in the P_4 position Ser or Thr are dominating. The specificity for the C-terminal part is less obvious.

The preference of hydrophobic residues in the primary binding sites clearly follows from the model. A very interesting structural explanation exists for the presence of the proline residue as P_1' side chain, but not as P_1. The critical hydrogen bonds fix the scissile bond of the substrate in the correct position in relation to the active carboxyls of aspartic proteases: one of them is the bond between the carbonyl oxygen of Gly-217 in pepsin (the corresponding residue is Gly-27 of monomer Q in HIV-1 PR) with the NH group of the

Table 1

Sites of gag-pol polyprotein cleavage with HIV-1 protease [10,11]

	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'
gag	Ser	Gln	Asn	Tyr	* Pro	Ile	Val	Gln
pol	Ser	Phe	Asn	Phe	* Pro	Gln	Ile	Thr
pol	Thr	Leu	Asn	Phe	* Pro	Ile	Ser	Pro
gag	Lys	Glu	Leu	Tyr	* Pro	Leu	Thr	Ser
gag	Ala	Arg	Val	Leu	* Ala	Glu	Ala	Met
gag	Ala	Thr	Ile	Met	* Met	Gln	Arg	Gly
gag	Pro	Gly	Asn	Phe	* Leu	Gln	Ser	Arg
pol	Arg	Lys	Ile	Leu	* Phe	Leu	Asp	Gly
gag	Arg	Gln	Ala	Asn	* Phe	Leu	Gly	Lys
gag	Ala	Glu	Ala	Met	* Ser	Gln	Val	Thr
gag	Phe	Arg	Ser	Gly	* Val	Glu	Thr	Thr

P_1-residue of the substrate; it means that in any aspartic protease normal substrate binding is impossible if the first position before the scissile bond is occupied with an imino acid residue, i.e. proline.

The implications of the specificity of the S_2 binding site seem too preliminary. However, it is possible to explain the binding of Asn or Gln residues at this site as being due to the formation of a hydrogen bond between Asp-30 carboxyl and their side chains. Various hydrophobic groups are also permissible for this site. Such groups dominate in substrates which do not contain proline at the P_1'-position. As follows from table 1 the P_3'-position can be occupied both with charged and hydrophobic residues. As pointed out, the Asp-29 residue contributes to the S_3-binding pocket, the residue which can be close to it is Arg-8 from the mentioned connecting segment. It should be mentioned that the presence of two polar groups with the opposite charges in this region is the common feature of other known retroviral proteases [4]. This feature explains the ability of retroviral proteases to bind in the S_3-site residues with any charge. Hydrophobic groups of the mentioned connecting segment provide the binding of $P_3(P_3')$-hydrophobic residues.

On the basis of the model one can describe the main features of the probable HIV-1 PR inhibitors. As $S_2(S_2')$ and $S_3(S_3')$ binding pockets contain charged residues it seems reasonable to test the binding of residues with the opposite charges. To enhance the specificity it is possible to include a proline residue in the P_1' position. As the S_n and S_n'

sites are identical, one can suppose that the most effective binding can be achieved for the symmetrical substrates and inhibitors. This symmetry will also enhance the specificity of the inhibitors for the HIV-1 PR. The modification of the scissile peptide group can provide the inhibitor properties of the compound to be tightly bound.

Another important problem to be considered is the formation of the dimeric molecule. Although the intermonomer sheet is out of the template region one feature of this sheet seems to be very much probable. All retroviral proteases have the conservative residue Arg-87 which follows after the B_4-strand of the monomer. This residue does not seem to contribute to the activity of the enzyme, like the aspartic acid residue located at the same place in ordinary aspartic proteases. It does not contribute anything to the substrate-binding pockets, as different retroviral enzymes have different specificity. One can think that this conservative arginine is involved in stabilization of the dimeric structure of these enzymes. The ionic pair which can be formed between this arginine residue and the carboxyl terminus of the molecule is the only one equally possible for all retroviral enzymes. If such interaction exists it should be duplicated because of the symmetry. At the same time it looks rather improbable if the conservative residue is involved in different net of interactions in different enzymes.

For the protein engineering work the modification of this conservative Arg-87 seems very interesting. The role of the Asp-29 and the Asp-30 residues in the specificity of HIV-1 PR can also be revealed by protein engineering experiments.

A proper explanation of HIV-1 proteinase specificity is one of the problems to be solved on the way to find effective drugs against the disease induced by this retrovirus.

After our model has been completed and described in [5] we have been informed about results of X-ray crystallographic studies for Rous sarcoma virus proteinase [12], which are in agreement with the main results of our investigations.

REFERENCES

[1] Katoh, I., Yasunaga, J., Ikava, Y. and Yoshinaka, Y. (1987) Nature 329, 654–656.
[2] Le Grice, S.F.J., Milles, J. and Mous, J. (1988) EMBO J. 7, 2547–2553.
[3] Andreeva, N.S. (1988) in: Biosynthesis and Structure of Proteins, Pushchino 3, 97–115.
[4] Pearl, L.H. and Taylor, W.R. (1987) Nature 329, 351–354.
[5] Pechik, I.V., Gustchina, A.E., Andreeva, N.S. and Fedorov, A.A. (1988) in: Biosynthesis and Structure of Proteins, Pushchino 3, 87–96.
[6] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rogers, J.R., Kennard, O. and Shimanouchi, T. (1977) J. Mol. Biol. 112, 535–542.
[7] Andreeva, N.S., Zdanov, A.S., Gustchina, A.E. and Fedorov, A.A. (1984) J. Biol. Chem. 259, 11353–11366.
[8] Sibanda, B.L. and Thornton, J.A. (1985) Nature 318, 170–174.
[9] James, M.N.G. and Sielecki, A.R. (1983) J. Mol. Biol. 163, 299–361.
[10] Darke, P.L., Nutt, R.F., Brady, S.F., Garsky, V.M., Ciccarone, T.M., Leu, C., Lumma, P.K., Freidinger, R.M., Veber, D.F. and Sigal, I.S. (1988) Biochem. Biophys. Res. Commun. 156, 297–303.
[11] Henderson, L.E., Benveniste, R.E., Sowder, R., Copeland, T.D., Schultz, A.M. and Oroslan, S. (1988) J. Virol. 62, 2587–2595.
[12] Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J. and Wlodawer, A. (1989) Nature, in press.