Volume 260, number 2, 201-205 FEB 08044 January 1990

Is the pseudo-dyad in retroviral proteinase monomers structural or evolutionary?

J.K. Mohana Rao and Alexander Wlodawer

Crystallography Laboratory, NCI-FCRF, BRI-Basic Research Program, PO Box B, Frederick, MD 21701, USA

Received 24 October 1989; revised version received 29 November 1989

A pseudo-dyad was found to exist in the monomers of the crystal structures of the proteinases from Rous sarcoma virus and the human immunodeficiency virus. This dyad, also discovered earlier in pepsin-like aspartic proteinases and considered to be of probable evolutionary origin, has been shown to arise as a result of the topology and the folding of the proteinase monomers and may not therefore have much evolutionary significance.

Retroviral proteinase; Rous sarcoma virus; Human immunodeficiency virus-1; Structural alignment; Pseudo-dyad

1. INTRODUCTION

The polyproteins necessary for the replication of retroviruses are post-translationally cleaved by proteinases (PR), which are coded in the *gag-pol* regions of the retroviral genes. The three-dimensional structures of two proteinases, from Rous sarcoma virus (RSV) [1] and from human immunodeficiency virus (HIV-1) [2–4], were recently elucidated by X-ray diffraction methods. These structural reports and several other biochemcal studies [5–7] revealed that the retroviral proteinase from dimers with Asp-Thr/Ser-Gly triplets at the active site. Their structures bear resemblance to the mammalian ([8–9], Gililand, Winbome, Nachman and Wlodawer, to be published) and fungal aspartic proteinase [10–11]. As a result of a comparison of the primary [13] and tertiary structures (Erickson, Rao, Wlodawer and Abad-Zapatero - to be published) of the retroviral and aspartic proteinases, suggestions have been made that duplication of the ancestral retroviral gene followed by a gene fusion gave rise to the present aspartic proteinases. Besides the exact (as in HIV-1 PR) or approximate (as in RSV PR) two-fold symmetry that exists in a dimer of the retroviral proteinase, a two-fold symmetry is present within the proteinase monomer itself. The consequences of this symmetry and its significance will be discussed here.

2. MATERIALS AND METHODS

The atomic coordinates used in this study were those deposited in the Protein Data Bank. The RSV PR coordinates are the result of a refinement with 2.0 Å data to an R factor of 0.144 (Jaskolski, Miller, Rao, Leis and Wlodawer, to be published), while those of HIV-1 PR resulted from a refinement at 2.8 Å to an R factor of 0.184 [3]. A program for structural alignments between two proteins starting from their atomic coordinates developed by one of the authors (J.K.M.R.) was used for aligning proteins or protein segments. The method is similar to that of Remington and Mathews [14]. All possible segments of the first protein were slid over all possible segments of the second. The segment length varied from 15 to 25 residues. The approximate matrix obtained from this data base was employed to refine the Eulerian rotation angles to within 0.5°. Using the matrix derived from this refinement, all the distances to within a certain cut-off value between equivalent atoms in the two proteins (or protein segments) were calculated. The whole cycle of calculations was repeated until convergence was achieved in terms of a minimum value for root mean square (RMS) deviation for distances between the atoms that were superimposed.

In the present analysis, the C_α coordinates of RSV PR and HIV-1 PR were compared against themselves. The cut-off distance for obtaining the initial orientation was 3.0 Å. However, atoms separated by slightly larger distances were considered aligned if the secondary structures to which they belonged ran parallel and had the same relative orientation. In order to confirm that the alignment and the superposition were not just restricted to alpha carbons, comparisons were also performed with all equivalent N, C, C_α, C_β, O atoms for residues to be compared. The RSV PR and HIV-1 PR structures were aligned between themselves for the purpose of standardization.

3. RESULTS AND DISCUSSION

The details of the various comparisons are presented in table 1. Table 2 gives the secondary structure alignments as a consequence of the two-fold symmetry in both RSV and HIV PR monomers in terms of the amino acid residues. Fig.1 is a stereo representation of the superposition of RSV PR and HIV-1 PR structures. Fig.2 with similar views shows the differences in the superpositions of RSV PR and HIV-1 PR and their mates arising because of the pseudo-dyad.

Correspondence address: J.K.M. Rao, Crystallography Laboratory, NCI-FCRF, BRI-Basic Research Program, PO Box B, Frederick, MD 21701, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

Table 1

Statistics for structural superpositions of retroviral proteinases

Proteins superimposed	No. of atoms superimposed	Root mean square Deviation (Å)
HIV-1 PR	23	
	C_α	1.51
HIV-1 PR	110	
	C_α,C_β,C,N,O	1.64
RSV1 PR	30	
	C_α	1.45
RSV1 PR	146	
	C_α,C_β,C,N,O	1.61
RSV2 PR	30	
	C_α	1.49
RSV2 PR	146	
	C_α,C_β,C,N,O	1.62
HIV-1 PR	77	
RSV1 PR	C_α	1.30

RSV1 PR and RSV2 PR are the two monomers of the RSV PR dimer related by an approximate dyad

For the purpose of discussion, the structure of the retroviral proteinase is described briefly using the nomenclature of Blundell [4,15]. The retroviral PR monomer consists of a four-stranded moiety (*a,b,c,d*) followed by a helical turn *h* (present only in RSV PR, but not in HIV-1 PR). A similar motif (*a',b',c',d',h'*) follows *a,b,c,d,h*. Between the strands *a'* and *b'* is present the 'flap', which could not be seen in the RSV proteinase due to disorder imposed by the crystal packing. However, this region, unambiguously observed in HIV-1 PR, plays an important role in the substrate binding as seen in the crystal structure of the inhibitor complex of HIV-1 PR (Miller, Sathyanarayana, Toth, Marshall, Clawson, Selk, Schneider, Kent and Wlodawer, to be published). The strand *q* at the carboxy terminal of the retroviral proteinases is involved in the formation of an intermonomer β sheet, and, like the flap, it has no intramonomer dyad-related counterpart. In the case of RSV PR monomers, the angle of rotation for the two pseudo-dyads located in the present analysis is 173° and is oriented at 55° with respect to the two-fold axis of the dimer. In the case of HIV-1 PR, the pseudo-dyad (angle of rotation is 180°) makes once again an angle of 55° with the perfect dyad of the dimer. It can be seen that even though a larger number of atoms could be superimposed in the case of RSV PR than in HIV-1 PR, the two-fold symmetry for the pseudo-dyad is only aproximate for RSV PR, whereas it is almost exact for HIV-1 PR. This is probably due to the smaller number of atoms involved in the case of HIV-1 PR and its more compact structure. The pseudo-dyad in both the RSV and HIV-1 PR monomers relates the two moieties *a,b,c,d* and *a',b',c',d'*. A helical turn *h* in RSV PR corresponds tot the helix *h'*, but curiously HIV-1 PR has no such counterpart in its N-terminal half of the structure. The monomer dyad passes between the mutually perpendicular four-stranded double layers comprising of *c,d',d,c'* and *c,b,b',c'* in a diagonal fashion. The strands *a* and *a'* that occur at the start of the respective motifs are not a part of these layers.

Table 2

Residues related by the intramonomer dyad in RSV PR and HIV PR

Secondary structural pair	HIVC	HIVN	RSVN	RSVC	HIVE
aa'	38 L	1 P	1 L	54 W	38 L
	39 P/G	2 Q	2 A	55 P	39 P/G
	40 G/N	3I/F	3 M	56 V	40 G/N
	41 K/N	4 T/S	4 T	57 M	41 K/N
bb'	60 D/K	9 P	11 P	74 R	57 R/K
	61 Q/N	10 L/V	12 L	75 K	58 Q/E
	62 I/V	11 V	13 V	76 S	59 Y
	63 L/E	12 T	14 R	79 M	61 Q/N
	64 I	13 I/A	15 V	80 I	62 I/V
	65 E	14 K/H	16 I	81 E	63 L/E
	66 I/V	15 I	17 L	82 L	64 I
	67 C/L	16 G/E	18 T	83 G	65 E
cc'			28 R	93 R	
	68 G/N	17 G	29 S	94 P	70 K
	69 H/K	18 Q	30 V	95 L	71 A/V
	70 K	19 L/P	31 Y	96 L	72 I/R
	71 A/V	20 K/V	32 I	97 L	73 G/A
	72 I/R	21 E	33 T	98 F	74 T
	73 G/A	22 A/V	34 A	99 P	75 V/I
	74 T	23 L	35 L	100 A	76 L/M
	75 V/I	24 L	36 L	101 V	77 V/T
	76 L/M	25 D	37 D	102 A	78 G
dd'		27 G	39 G	103 M	79 P/D
		28 A	40 A	104 V	80 T
		29 D	41 D	105 R	81 P
		30 D	42 I	106 G	82 V/I
	85 I/F	31 T/S	43 T	107 S	83 N
	86 G	32 V/I	44 I	108 I	84 I
		33 L/V	45 I	109 L	85 I/F
	34 E/A	46 S	110 G	86 G	
hh'		35 E/G	47 E	111 R	87 R
			48 E	112 D	88 N
			49 D	113 C	89 L/I
			50 W	114 L	90 L

Column 1 gives information about the secondary structural elements (eg. *aa'*) related by the intramonomer dyad symmetry. The residues in columns 2 (HIVC, C for C-terminal) and 3 (HIVN, N for N-terminal) are related by the intramonomer pseudo-dyad for HIV PR. Columns 4 (RSVN) and 5 (RSVC) are similar for RSV PR. Structural correspondence at the N-terminal half exists for residues in columns 3 (for HIV PR) and 4 (for RSV PR). Columns 5 (for RSV PR) and 6 (for HIV PR, HIVE; E for equivalent) show similar correspondence at the C-terminal half. These structural equivalences were obtained from the superposition of RSV PR and HIV-1 PR. Note that the entries in columns 2 and 6 are not always the same. The numbers correspond to residue numbers and the letters to the amino acid residues in the single letter code. Even though the superpositions have been carried out only for the synthetic HIV-1 PR, the amino acid residues for both HIV-1 (BH-10 isolate) and HIV-2 PR (ROD isolate) are mentioned (separated by a '/' where different)

Blundell et al. [16] in their analysis of the aspartic

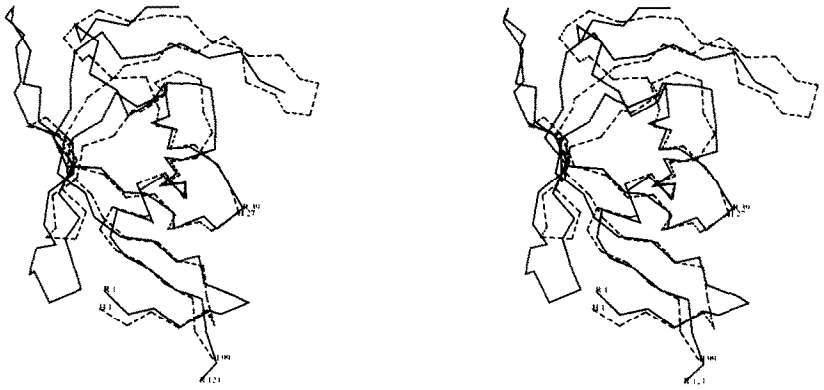

Fig.1. Stereo view of the C_α superposition of RSV PR (solid lines) and HIV-1 PR (broken lines) molecules. For easy chain tracing, the first (R 1 and H 1; R for RSV and H for HIV) and the last residues (R 124 and H 99) together with the Gly residues, R 39 and H 27, two residues past the Asp at the active site marked. The pseudo-dyad is approximately horizontal.

proteinase, endothiapepsin, discuss the possibility that the bilobal proteinase is a tetramer of four subdomains of the type *a,b,c,d,h* as the building blocks. A close examination of table 2 reveals some interesting facts. Out of 35 intramolecular dyad-related residues for RSV PR, only three amino acid residues are identical (columns 4 and 5 of table 2). Out of 23 such pairs of equivalent residues for HIV PR (columns 2 and 3 of table 2), three are identical for HIV-1 PR and two for HIV-2 PR. The mean values of the mininum base change per codon (MBC/C) for the sequences of the aligned segments related by the pseudo-dyad are 1.29 (1.3σ below the random mean), 1.26 (1.8σ below the random mean), 1.57 (0.3σ above the random mean), respectively, for RSV PR, HIV-1 PR and HIV-2 PR, not significantly less than the random value of about 1.45 – 1.50 [17]. The average values in the case of Dayhoff's MDM exchange scores [18], are 10.3 (2.2σ above the random mean), 9.9 (0.9σ above the random mean), 9.5 (0.1σ above the random mean), respectively, for RSV PR, HIV-1 PR and HIV-2 PR, closer to or less than the value of 10, the threshold for homology. Not one of the above scores is beyond three standard deviations from the random mean, a necessary condition for sequence similarity [19]. The standard deviations mentioned above were estimated from the scores obtained for one hundred randomly shuffled sequences of the aligned segments. Thus, even when the structural motifs could be superimposed, there is practically no sequence relationship for the superimposed parts. In the case of *P. aerogenase* ferredoxin, which has a four-stranded antiparallel β sheet, there exists an intramolecular two-fold axis. Even though this protein is relatively small (54 residues only), the regions structurally equivalent due to the intramolecular dyad bear considerable sequence homology [20]. Both sequence and structural homology were also noticed in the case of the eye lens protein γ crystallin [21], which has repeating motifs of strands in its crystal structure. However, for the proteinases of both RSV and HIV-1, more than half of the residues related by the dyad are of the conserved variety, showing that the nature of the amino acid residues (hydrophobic, charged, etc.) at particular positions on the strands are of paramount importance [22]. The dyad-related parts in the strand pairs *bb'* and *cc'* are two residues out of step for HIV PR in comparison with RSV PR, when one takes into account the structural alignment between the two retroviral proteinases themselves. No part of the flap is related by the intramolecular dyad and prima facie. This region could be considered to be the linker between the two motifs. The appearance of the flap between the strands *a'* and *b'* instead of between the moieties *a,b,c,d,h* and *a',b',c',d',h'* gives less credence to the assumption that these intramolecular motifs fused to form the proteinase monomer, unless one proposes selective duplication of the *b',c',d',h'* motif alone with *a'* being part of the flap region. In pepsin-like aspartic proteinases, both domains that are related by the interdomain two-fold axis contain the Asp-Thr-Gly triplets where the aspartic acid residues are involved in a water-mediated hydrogen bond. The interdomain two-fold rotational symmetry is almost perfect in this part of the proteinase structure. On the other hand, the intramolecular dyad symmetry in retroviral proteinases begins to break down precisely at this juncture as the crucial Asp residue is approached (see table 2). Thus, in view of anomalies such as the lack of discernible sequence homology between the monomer dyad-related segments, breakdown of the intramolecular symmetry near the active site, lack of sequence agreement between the dyad-related segments when one takes into account the structural alignment of RSV PR and HIV-1 PR, and the absence of the flap between *a,b,c,d,h* and *a',b',c',d',h'* motifs, one must be extremely cautious when proposing that a gene corresponding to the primordial subdomain consisting of a motif such as *a,b,c,d,h* which duplicated, fused and mutated selectively to produce the retroviral proteinase monomer (note that the helix *h* is not present in HIV PR).

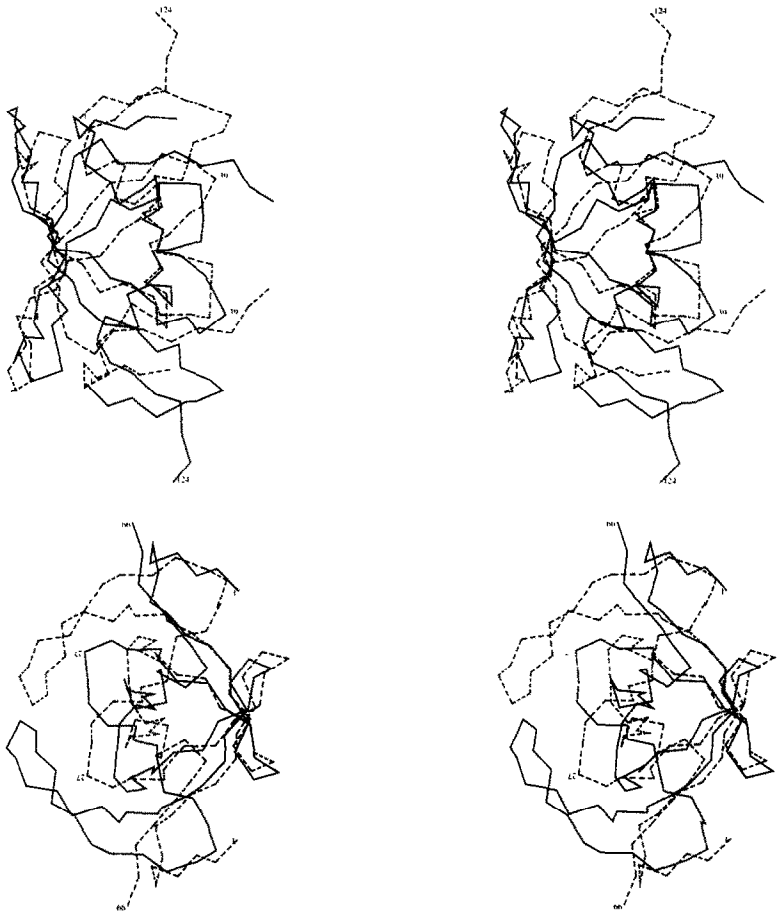

Fig.2. Stereo views of the RSV PR and HIV-1 PR molecules (C_α atoms only) in the presence of the intramonomer pseudo-dyad. The solid lines indicate the original molecules and the broken lines the molecules related by the pseudo-dyad. (a) One of the monomers of the RSV PR, called RSV1 PR; (b) HIV-1 PR. The first and the last residues together with the Gly of the Asp-Thr/Ser-Gly triplet are marked. The primed numbers correspond to the pseudo-dyad-related molecule.

On the other hand, is there an alternate explanation for the existence of the intramolecular dyad in retroviral proteinases? Many proteins require a secondary structural core, generally hydrophobic, around which other structural elements are built to form a compact globular entity. The structural unit of parallel layers of strands with orthogonal strand directions, found in retroviral and aspartic proteinases, happens to be such a preferred core. Although the ψ loop or the 'wedding ring' topology present in both retroviral and aspartic proteinases is quite unique, parallel layers of crossed strands, devoid of the directionality property of the strands in proteinases, are found in several other proteins such as catalase [23], H protein of the photosynthetic reaction center complex [24], and retinol binding protein [25]. The only common characteristic for these apparently unrelated proteins seems to be the aforementioned hydrophobic core. The consequent manifestation of an approximate two-fold symmetry in retroviral and aspartic proteinases seems to be just coincidental with limited or no evolutionary implications and the dyad symmetry is more pronounced in smaller proteins or protein domains.

Acknowledgements: The research work is sponsored by the National Cancer Institute, DHHS, under contract no. NO1-CO-74101 with BRI. 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 US Government.

REFERENCES

[1] Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J. and Wlodawer, A. (1989) Nature 337, 576–579.
[2] 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. and Springer, J.P. (1989) 337, 615–620.
[3] Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E.T., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. and Kent, S.B.H. (1989) Science 245, 616–621.
[4] Lapatto, R., Blundell, T.L., 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. and Hobart, P.M., Nature, in press.
[5] Loeb, D.D., Hutchinson, C.A. iii, Edgell, M.H., Farmerie, W.G. and Swanstrom, R. (1989) J. Virol. 63, 111–121.
[6] Meek, T.D., Dayton, B.D., Metcalf, B.W., Dreyer, G.B., Strickler, J.E., Gomiak, J.G., Rosenberg, M., Moore, M.L., Magaard, V.W. and Debouck, C. (1989) Proc. Natl. Acad. Sci. USA 86, 1841–1845.

[7] Katoh, I., Ikawa, Y. and Yoshinaka, Y. (1989) J. Virol. 63, 2226–2232.
[8] Andreeva, N.S., Zdanov, A.S., Gustchina, A.E. and Fedorov, A.A. (1984) J. Biol. Chem. 259, 11353–11365.
[9] Sielecki, A.R., Hayakawa, K. Fujinaga, M., Murphy, M.E.P., Fraser, M., Muir, A.K., Carilli, C.T., Lewicki, J.A., Baxter, J.D. and James, M.N.G. (1989) Science 243, 1346–1351.
[10] Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H. and Davies, D.R. (1987) J. Mol. Biol. 196, 877–900.
[11] Pearl, L. and Blundell, T.L. (1984) FEBS Lett. 174, 96–101.
[12] James, M.N.G. and Sielecki, A.R. (1983) J. Mol. Biol. 163, 299–361.
[13] Doolittle, R.F., Feng, D.-F., Johnson, M.S. and McClure, M.A. (1989) Quart. Rev. Biol. 64, 1–30.
[14] Remington, S.J. and Matthews, B.W. (1980) J. Mol. Biol. 140, 77–99.
[15] Blundell, T.L., Jenkins, J., Pearl, L., Sewell, T. and Pedersen, V. (1985) in: Aspartic Proteinases and their Inhibitors (Kostka, V. ed.) pp. 151–161, Walter De Gruyter, Berlin.
[16] Blundell, T.L., Sewell, B.T. and McLachlan, A.D. (1979) Biochim. Biophys. Acta 580, 24–31.
[17] Keim, P., Heinrikson, R.L. and Fitch, W.M. (1981) J. Mol. Biol. 151, 179–197.
[18] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure, vol. 5, suppl. 3, pp. 345–352, Natl. Biomed. Res. Found, Washington, DC.
[19] Feng, D.F., Johnson, M.S. and Doolittle, R.F. (1985) J. Mol. Evol. 21, 112–125.
[20] Jensen, L.H. (1974) Annu. Rev. Biochem. 43, 461–474.
[21] Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. and Blundell, T. (1983) J. Mol. Biol. 170, 175–202.
[22] Pechik, I.V., Gustchina, A.E., Andreeva, N.S. and Federov, A.A. (1989) FEBS Lett. 247, 118–122.
[23] Murthy, M.R.N., Reid, T.J. iii, Sicignano, A., Tanaka, N. and Rossmann, M.G. (1981) J. Mol. Biol. 152, 465–499.
[24] Deisenhofer, J. and Michel, H. (1989) EMBO J. 8, 2149–2170.
[25] Newcomer, M.E., Jones, T.A., Åqvist, J., Sundelin, J., Eriksson, U., Rask, I. and Peterson, P.A. (1984) EMBO J. 3, 1451–1454.