Structural and Evolutionary Relationships between Retroviral and Eucaryotic Aspartic Proteinases†

J. K. Mohana Rao,*,‡ John W. Erickson,§,∥ and Alexander Wlodawer‡

Macromolecular Structure Laboratory, National Cancer Institute—Frederick Cancer Research and Development Center, Advanced BioScience Laboratories, Inc.—Basic Research Program, P.O. Box B, Frederick, Maryland 21702, and Protein Crystallography Laboratory, D-47E, Abbott Laboratories, Abbott Park, Illinois 60064

Received November 1, 1990; Revised Manuscript Received February 12, 1991

ABSTRACT: Three-dimensional crystal structures of the homologous retroviral proteinases from Rous sarcoma virus (RSV PR) and from human immunodeficiency virus (HIV-1 PR) are to a large extent similar and bear close resemblance to the six known structures of the bilobal fungal and mammalian aspartic proteinases. Systematic three-dimensional structural superpositions were carried out between the retroviral and the eucaryotic aspartic proteinases. Both retroviral enzymes were found to be similarly related to their fungal and mammalian counterparts. The most strongly conserved parts correspond to those regions in the N- and C-domains of the eucaryotic enzymes that are related by the interdomain dyad and consist of a combination of secondary structural elements that form the ψ loop–α helix motif at the active sites of the aspartic proteinases. The retroviral proteinase monomer exhibits nearly the same degree of structural equivalence to the N- and C-domains of the eucaryotic enzymes. In light of the deduced structural relationships between HIV-1 PR and RSV PR, sequence alignments were performed for a number of retroviral proteinases from different subfamilies. There are three highly conserved amino acid sequence stretches, of which two belong to the ψ loop–α helix motif, that bear moderate sequence similarity with the eucaryotic enzymes. The third conserved sequence stretch among the retroviral proteinases belongs to the flap and bears no resemblance to the flap sequences in the eucaryotic enzymes. The interdomain antiparallel β sheet in the cellular enzymes differs from the intersubunit β sheet in the number, arrangement, and directionality of strands, suggesting the possibility of convergent evolution. However, a more detailed topological analysis of the two sheets revealed that they indeed have a direct structural relationship. This represents an unusual evolutionary example of how a combination of strand deletion and strand interchange events can combine to preserve the functionality of a critical secondary structural element. Overall, the structural comparisons argue strongly for the evolutionary relatedness of the eucaryotic and retroviral aspartic proteinases to a common ancestral domain or subunit. Since no aspartic proteinase motif has yet been discovered in procaryotes, we speculate that the retroviral enzyme evolved from a cellular gene by one or more deletion events, similar to the case for certain viral oncogenes that are deletion products of cellular protooncogenes.

Retroviruses, so named because of their ability to copy their RNA genome into a double-stranded DNA form, encode a proteinase that is responsible for the maturation of newly assembled virions into infectious virus particles. Maturation proceeds via the specific cleavage of the viral polyprotein products of the *gag* and *pol* genes by the viral proteinase. The retroviral proteinases are all characterized by a conserved amino acid triplet, Asp-Thr/Ser-Gly (Toh et al., 1985; Power et al., 1986). Such triplets, occurring in pairs, also exist in the well-studied aspartic proteinase family, for which at least 24 amino acid sequences are known. This led Pearl and Taylor (1987) to propose a structural model for retroviral proteinases based on the crystal structure of the fungal aspartic proteinase endothiapepsin.

While the detailed architectures of only two retroviral proteinases are known to date, from Rous sarcoma virus (RSV PR) (Miller et al., 1989a; Jaskólski et al., 1990) and from HIV-1 (HIV-1 PR) (Navia et al., 1989; Wlodawer et al., 1989; Lapatto et al., 1989), three-dimensional crystal structures for six cell-derived proteinases have been determined. These include three fungal aspartic proteinases—endothiapepsin (Blundell et al., 1990), penicillopepsin (James & Sielecki, 1983), and rhizopuspepsin (Suguna et al., 1987)—and three mammalian enzymes—porcine pepsin (Sielecki et al., 1990; Abad-Zapatero et al., 1990; Cooper et al., 1990), human renin (Sielecki et al., 1989), and bovine chymosin (Gilliland et al., 1990). The cellular aspartic proteinases, with about 325 residues, are all more than twice as long as their retroviral counterparts, which only have 99–125 amino acids each. All aspartic proteinase structures contain two domains possessing more or less similar topological features. The N- and the C-terminal domains, although structurally related by a 2-fold axis, have only limited sequence homology except in the vicinity of the active site. These facts suggested that the enzymes evolved by duplication of an ancestral gene encoding a primordial domain (Tang et al., 1978). In contrast, retroviral proteinases are dimers of identical subunits that are related to each other by a 2-fold axis that is perfect in the case of HIV-PR or that is approximate (angle of rotation about 178°) in the case of RSV PR. It has also been shown that both of the domains of the eucaryotic aspartic proteinases possess pseudodyads, and these were postulated to be of probable evolutionary origin (Blundell et al., 1979). Rao and Wlodawer (1990) demonstrated not only that the monomer subunits of the retroviral proteinases exhibit internal symmetry but also

† Research supported by NIH Grant U01 AI27220-01 (J.W.E.) and by the National Cancer Institute, DHHS, under Control No. N01-CO-74101 with ABL (J.K.M.R. and A.W.).

* To whom correspondence should be addressed.

‡ NCI—Frederick Cancer Research and Development Center.

§ Abbott Laboratories.

∥ Present address: Program Resources Inc.—DynCorp., NCI—Frederick Cancer Research and Development Center, P.O. Box B, Frederick, MD 21702.

0006-2960/91/0430-4663$02.50/0 © 1991 American Chemical Society

Table I: Protein Data Bank Designations of Aspartic Proteinases Used in the Superposition Studies

protein designation	name of protein	reference
1CMS	chymosin from cow	Gilliland et al. (1990)
2APP	penicillopepsin from *P. janthinellum*	James and Sielecki (1983)
2APR	rhizopuspepsin from *R. chinesis*	Suguna et al. (1987)
2RSP	Rous sarcoma virus proteinase	Jaskolski et al. (1990)
3HVP	HIV-1 proteinase (synthetic form)	Wlodawer et al. (1989)
3PEP	pepsin from pig	Abad-Zapatero et al. (1990)
4APE	endothiapepsin from *Endothiaparasitica*	Blundell et al. (1990)

Table II: Statistics for Structural Superpositions of Retroviral and Cellular Aspartic Proteinases[a]

fixed protein	rotated protein	Eulerian angles of rotn (deg)			no. of atoms superimposed	rms devn (Å)
		α	β	γ		
RSV PR	HIV-1 PR	17.00	37.75	146.00	139 C_α	1.28
					675 all[b]	1.51
HIV-1 PR	chymosin	−73.25	163.00	−29.00	104 C_α	1.60
					501 all	1.80
HIV-1 PR	endothiapepsin	−8.75	112.00	−42.00	98 C_α	1.50
					473 all	1.68
HIV-1 PR	penicillopepsin	−90.75	68.50	12.50	99 C_α	1.45
					479 all	1.63
HIV-1 PR	pepsin	95.50	78.00	−119.00	107 C_α	1.58
					512 all	1.80
HIV-1 PR	rhizopuspepsin	172.75	158.00	75.50	107 C_α	1.55
					514 all	1.78
RSV PR	chymosin	169.50	144.00	58.00	96 C_α	1.65
					465 all	1.90
RSV PR	endothiapepsin	158.50	83.00	342.00	85 C_α	1.66
					410 all	1.83
RSV PR	penicillopepsin	65.00	94.00	42.00	95 C_α	1.66
					459 all	1.89
RSV PR	pepsin	−89.00	63.50	−157.50	96 C_α	1.61
					462 all	1.83
RSV PR	rhizopuspepsin	−127.00	153.00	−29.00	94 C_α	1.67
					452 all	1.92

[a] The convention for Eulerian angles is as in Crowther (1972). [b] All = all equivalent C_α, C_β, C, N, and O atoms.

that the intradomain symmetry is due more to topological constraints than to evolutionary causes.

In order to obtain a fuller understanding of the similarities and differences between the eucaryotic and retroviral aspartic proteinases, as well as of their possible evolutionary relationships, we have undertaken to perform a detailed comparison between the structures of these two enzyme families. A preliminary abstract of this effort was presented elsewhere (Erickson et al., 1989).

Materials and Methods

The atomic coordinates used in this study were those deposited in the Protein Data Bank (see Table I for their designations). The program package UPAMA, developed by one of us (J. K. M. Rao, unpublished work) for structural alignments between two proteins starting from their atomic coordinates, was used for aligning proteins or protein segments (Rao & Wlodawar, 1990). The method is similar to that of Remington and Matthews (1980). 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°. By using the matrix derived from this refinement, all the distances to within a certain cutoff 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 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 each of five of the six aspartic proteinases with known three-dimensional structures. (The atomic coordinates for human renin are still unavailable.) The retroviral proteinase structures of RSV and HIV-1 were also compared with each other. The cutoff distance for obtaining the initial orientation was 3.0 Å. In order to confirm that the alignment and the superposition are not just restricted to α carbons, comparisons were also performed with all equivalent N, C, C_α, C_β, and O atoms for residues to be compared. The differences in the conformational angles, ϕ and ψ, for the aligned regions were also calculated.

In addition to the C_α superposition based on the crystal structure data, the amino acid sequences of 13 retroviral proteinases were aligned with the program ALIGN, based on the method of Needleman and Wunsch (1970) and Dayhoff's mutation data matrix (Dayhoff et al., 1978; Schwartz & Dayhoff, 1978). All the sequences were extracted from the PIR protein sequence data base. The alignment scores for all pairwise combinations ranged from 7 to 11 standard deviations above the mean taken for 100 random sequence alignments with sequences of the same length and base composition used as the test cases. Since the alignment score is an overall measure of sequence similarity, the regions of locally poor homology may have very low scores and may be poorly aligned, resulting in the misplacement of gaps. Therefore, gaps in the aligned sequences were placed, wherever possible, by using the results from the structural superposition of HIV-1 PR and RSV PR.

Results and Discussion

Table II displays the quantitative results obtained from various superpositions between the aspartic and retroviral proteinases. Figure 1 is a stereo representation of the C_α

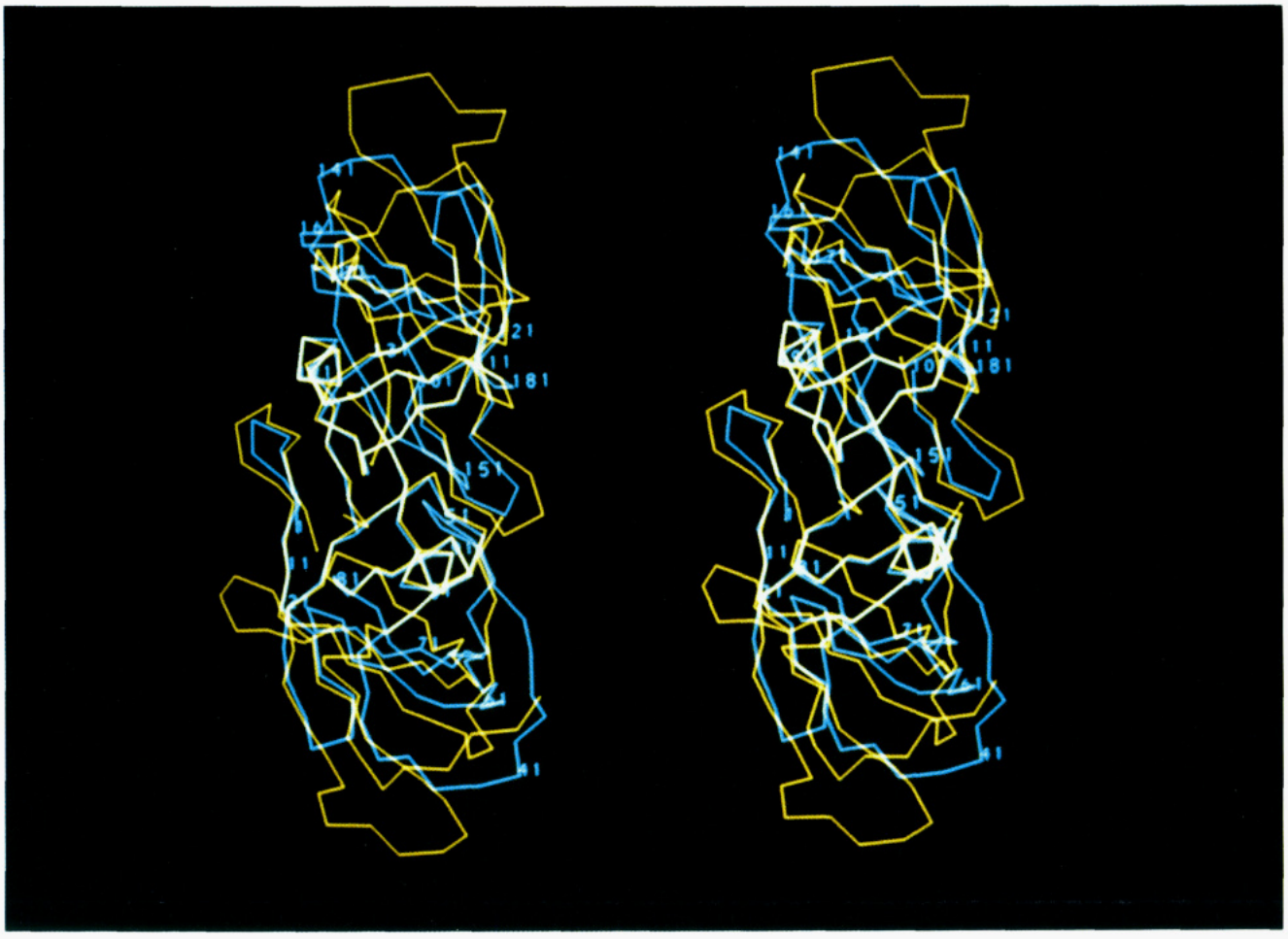


FIGURE 1: Stereoview of the C_α superposition of HIV-1 PR (in blue) and RSV PR (in yellow). The superposition matrix was obtained by the rotation of 14 segments of HIV-1 PR (1–6, 7–16, 17–34, 62–67, 68–69, 70–77, 82–99, 101–106, 107–116, 117–134, 162–167, 168–169, 170–177, and 179–199) onto 14 segments of RSV PR (1–6, 9–18, 29–46, 80–85, 91–92, 94–101, 106–123, 201–206, 209–218, 229–246, 280–285, 291–292, 294–301, and 303–323).

superposition of the two retroviral proteinases of HIV-1 and RSV. Figure 2 shows the aligned sequences of the retroviral proteinases based upon the structural alignments of HIV-1 PR and RSV PR. Figures 3 and 4 are the respective pictorial representations of the superpositions of RSV PR on pepsin as well as HIV-1 PR on rhizopuspepsin. Figure 5 shows the amino acid sequence alignment between the retroviral and aspartic proteinases based on the structural alignments. The nomenclature proposed by Blundell (Blundell et al., 1985) is used for describing the secondary structural elements of aspartic and retroviral proteinases.

Structural Relationships between RSV PR and HIV-1 PR. From Figure 1 and Table II, it is evident that even though the two retroviral proteinases in themselves possess a high degree of structural similarity, there are regions of considerable difference and divergence. First of all, the flaps in HIV-1 PR (residues 44–58) that completely wrap around an inhibitor molecule as observed in several crystal structure studies of various HIV-1 PR–inhibitor complexes (Miller et al., 1989b; Swain et al., 1990; Erickson et al., 1990; Fitzgerald et al., 1990) were not seen in the crystal structure of RSV PR. There seems to be considerable conformational disorder for the flaps in RSV PR but not in HIV-1 PR. The apparent conformational disorder for the flaps in the RSV PR enzyme is probably not due to crystal packing since HIV-1 PR dimers can be packed well in the trigonal RSV PR unit cell without overlap of the flaps. The flap of HIV- PR has a structural analogue in the N-terminal (but not in the C-terminus) domain of the mammalian and fungal aspartic proteinases, although the β bend regions or tips of the flaps superimpose rather poorly (see Figures 4 and 5). These regions are clearly discernible in the crystal structures of all native and inhibitor–enzyme complexes of the eucaryotic aspartic proteinases even though they crystallize in unit cells belonging to different space groups. Thus, it was a surprise not to have observed the flaps in the crystal structure of RSV PR. Paradoxically, the flap contains one of three highly conserved stretches in the aligned sequences of the retroviral proteinases (see Figure 2) and yet it bears little sequence homology to the analogous flaps of the cellular enzymes. The existence of two flaps in the retroviral proteinase dimers versus a single flap for their cellular counterparts results in a more intimate embracing of the inhibitor by the former. This is evidenced by the manner in which the flaps wrap around the inhibitors, which is distinctly different for the proteinases of the two families as observed in the crystal structures of inhibitor-bound enzymes. Moreover, the flaps from HIV-1 PR move by more than 7.0 Å upon binding to a reduced peptide inhibitor (Miller et al., 1989b). In contrast, relatively minor shifts in the position of the N-domain flap have been observed for the cellular enzymes upon inhibitor binding (Davies, 1990). The large number of conserved glycines in the flaps of the retroviral proteinases probably reflects a requirement for substantial conformational flexibility of this structural element. Consequently, this region of the structure may be more prone to conformational disorder.

There are several minor structural differences between the

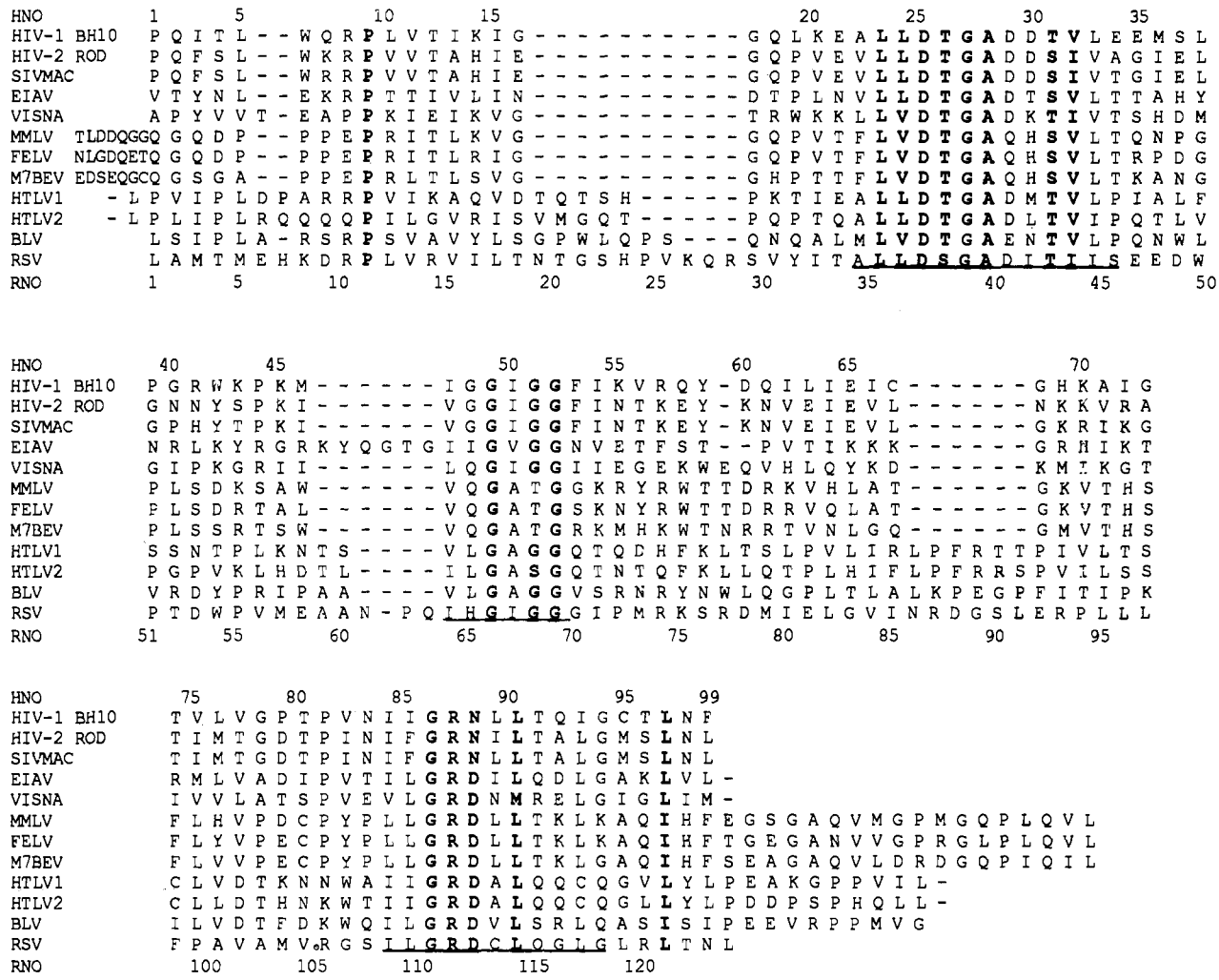


FIGURE 2: Sequence alignments for retroviral proteinases. Conserved residues are in bold letters. Sequences were taken from the PIR sequence data base and the abbreviations used have the following meanings: HNO, HIV numbering; HIV1, human immunodeficiency virus type 1 (BH-10 isolate); HIV2, human immunodeficiency virus type 2 (ROD isolate); SIV, simian immunodeficiency virus (MAC isolate); EIAV, equine infectious anemia virus; VISNA, visna lentivirus (strain 1514); MMLV, Moloney murine leukemia virus; FELV, feline leukemia virus; M7BEV, baboon endogenous virus; HTLV1, human T-cell leukemia virus type I; HTLV2, human T-cell leukemia virus type II; BLV, bovine leukemia virus; RSV, Rous sarcoma virus (Prague C strain); RNO, RSV numbering. Undetermined N- and C-termini are designated by a dash at the end. Owing to the disordered nature of the flap in RSV PR and to the absence of sequence homology in the region immediately preceding the flap, the alignment for residues 50–79 (RSV PR numbering) is uncertain.

proteinases of RSV and HIV-1. The helical turn that is present in the N-terminal region of RSV PR (residues 46–49) is absent in HIV-1 PR. In addition, the RSV PR is 25 residues longer than HIV-1 PR, although there is more internal symmetry in the monomer of the former than in the latter (Rao & Wlodawer, 1990). This is rather surprising since, as proteins become longer, internal symmetry usually becomes less pronounced due to insertions. Thirdly, there is also considerable dissimilarity in the manner in which the backbones of the two retroviral proteinases run just before the a′ strand. Crystallographic elucidation of more retroviral proteinase structures is required in order to resolve these apparent inconsistencies and to establish the more common topological features of the retroviral proteinases.

It should be noted that when this investigation was started over a year ago, the only structure for the recombinant HIV-proteinase was published by Navia et al. (1989). When the RSV PR structure (Miller et al., 1989a) was compared with that of HIV-1 PR, it was found that the helix at the C-terminal end of RSV PR did not have a helical counterpart in HIV-1 PR, but an extended chain was present instead (Erickson et al., 1989). Moreover, the helix in question is contained within a structural motif whose sequence is highly conserved among all the aspartic proteinases (see Figures 2 and 5). However, later independent crystallographic studies by two different groups (Wlodawer et al., 1989; Lapatto et al., 1989) confirmed the presence of such a helix in HIV-1 PR. Recently the authors of the initial structural report revised their chain tracing (Fitzgerald et al., 1990), and presently all groups are in complete agreement on the tertiary structure of HIV-1 PR.

Superpositions of Retroviral and Eucaryotic Aspartic Proteinases. The general structural similarities between the retroviral and eucaryotic aspartic proteinases are easily seen when the C_α backbone structures of RSV PR and porcine pepsin are viewed down their respective intersubunit or interdomain dyads after superposition (Figure 3). The distal portions of these structures contain the interdomain or intersubunit β sheets. The topology of these sheets differs for the two structures with respect to the orientation, number, and apparent polarity of the strands that form the respective sheets. The middle section contains the active-site ψ loops, and this is where most of the structural equivalences between the retroviral and the eucaryotic enzymes are found. The proximal section contains various surface loops and turns and is highly variable between the two structures.

It was possible to superimpose nearly 100 C_α pairs between

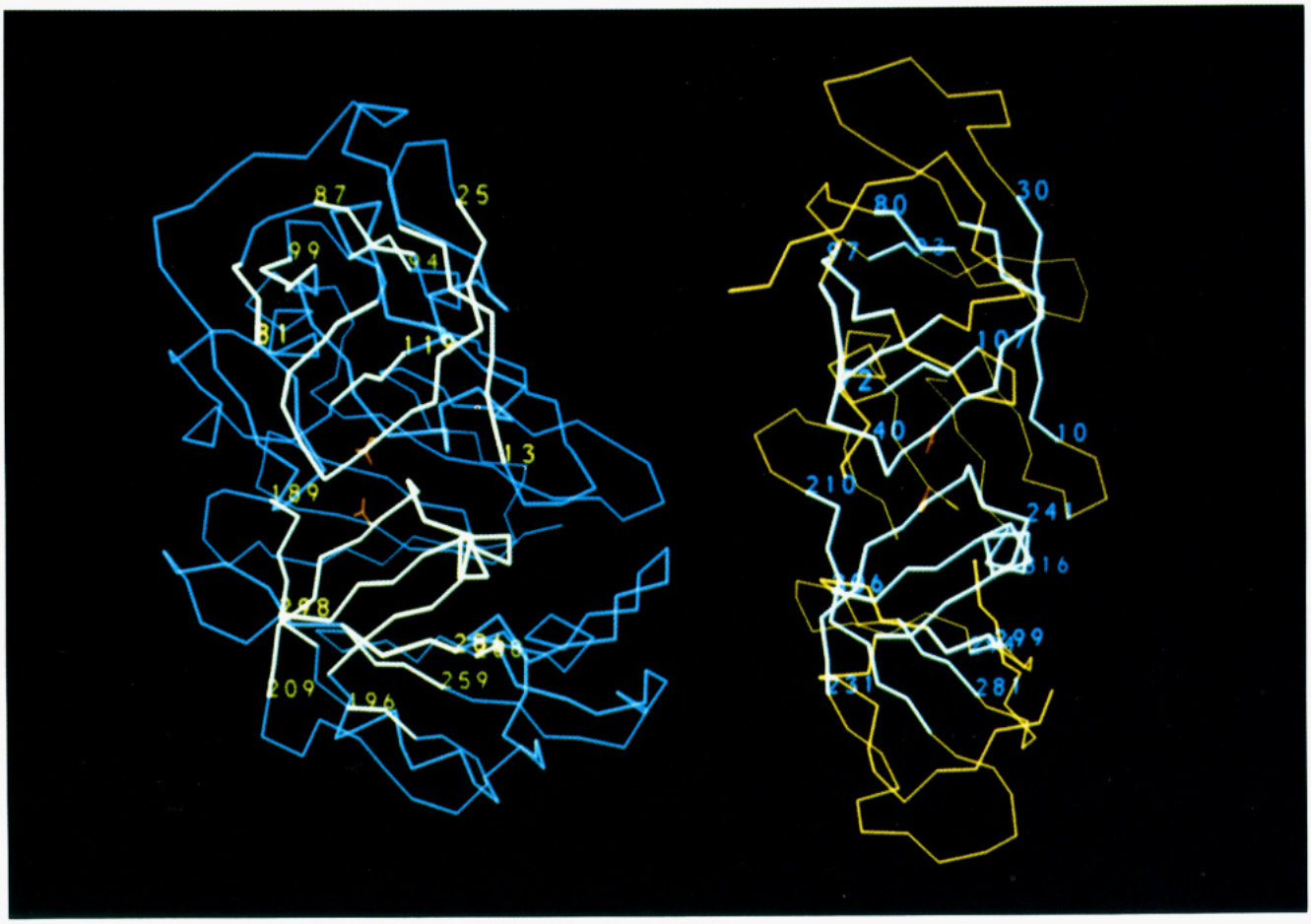


FIGURE 3: C_α superposition of porcine pepsin (in blue) and RSV PR (in yellow). Structurally equivalent segments are highlighted. The 2-fold-related ψ loops are seen as the diagonal three stranded β sheets in the center of each molecule. The active-site Asp residues in red emanate from the innermost strand of each ψ loop. The superposition matrix was obtained by the rotation of 14 segments of pepsin (15–21, 25–40, 41–43, 86–91, 93–96, 98–102, 119–123, 135–136, 189–194, 209–223, 260–263, 267–269, 286–292, and 298–308) on to 14 segments of RSV PR (12–18, 30–45, 49–51, 79–84, 92–95, 96–100, 107–111, 112–113, 210–215, 231–245, 282–285, 293–295, 299–305, and 306–316).

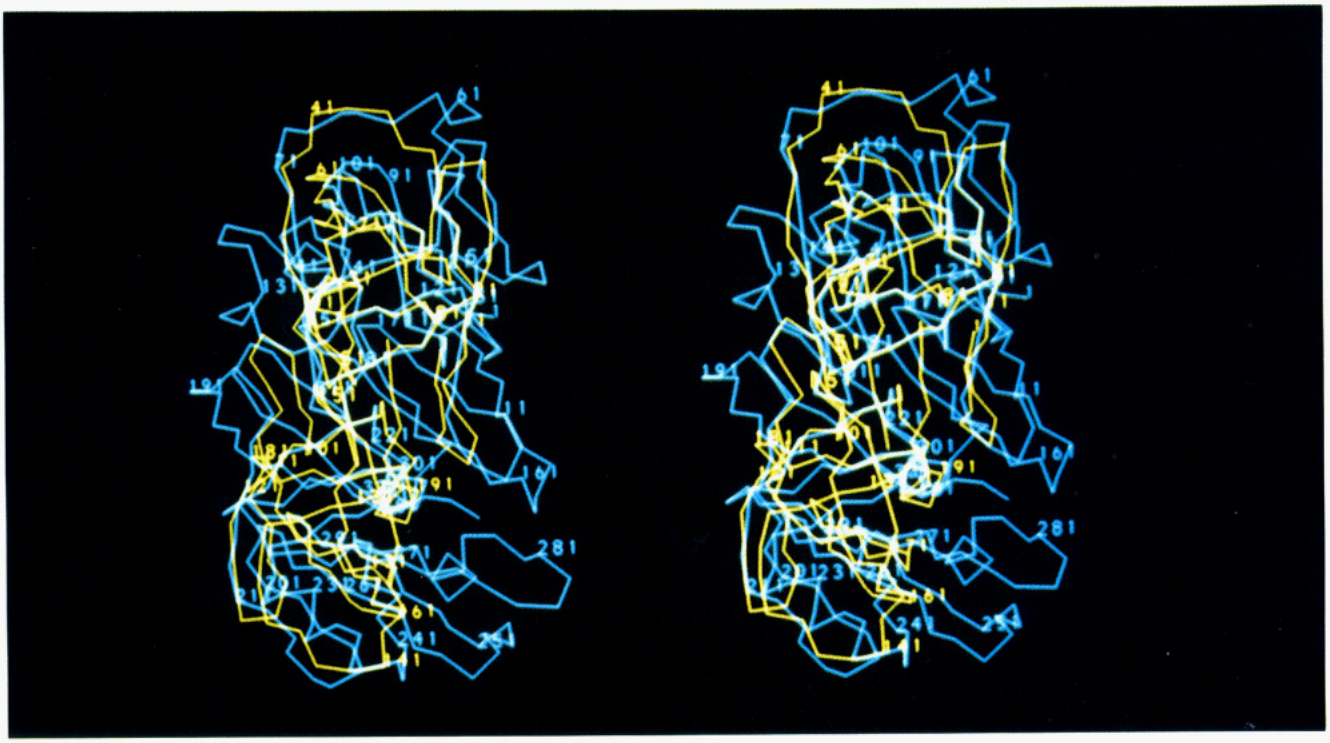


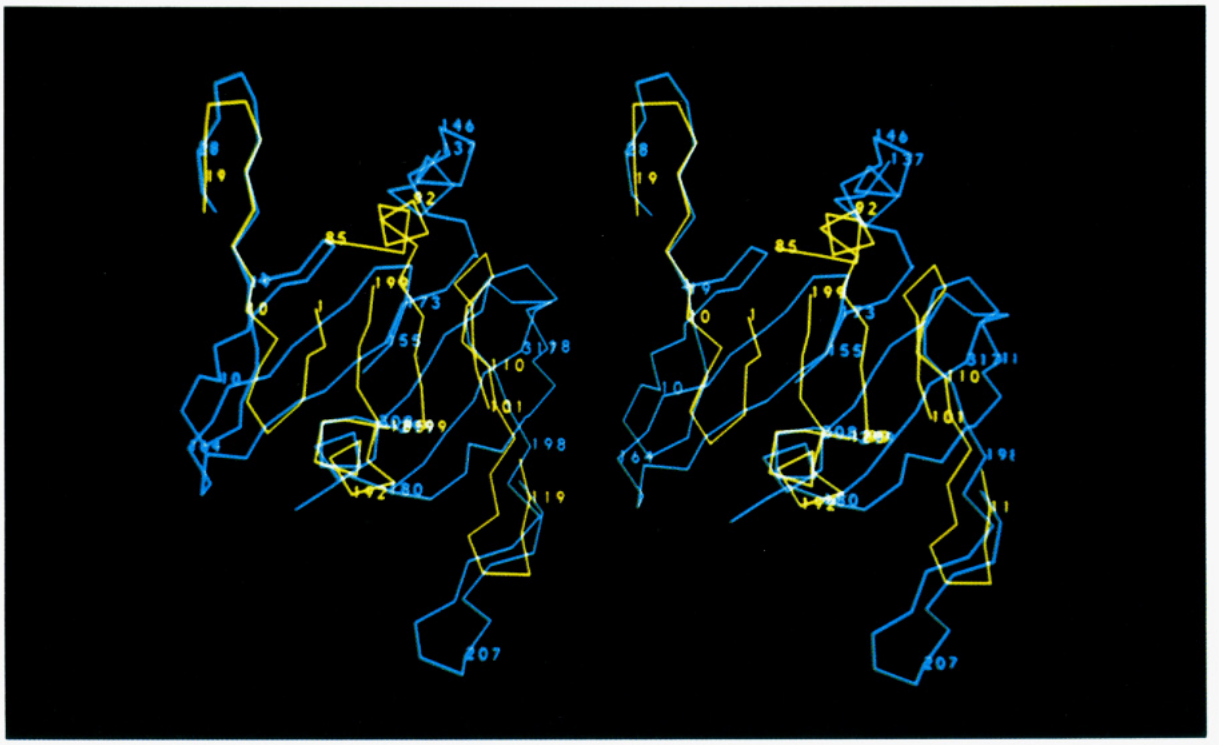
FIGURE 4: Stereoview of the C_α superposition of rhizopuspepsin (in blue) and HIV-1 PR (in yellow). The superposition matrix was obtained by the rotation of 14 segments of rhizopuspepsin (18–24, 27–45, 68–76, 84–88, 90–99, 102–107, 121–125, 138–139, 194–199, 212–229, 241–243, 289–293, 297–306, and 307–308) onto 14 segments of HIV-1 PR (10–16, 17–35, 40–48, 55–59, 62–71, 73–78, 83–87, 88–89, 108–113, 119–136, 142–144, 175–179, 183–192, and 194–195).

FIGURE 5: (Top) Structural alignment of HIV-1 PR and five eucaryotic aspartic proteinases. The sequence of HIV-1 PR corresponds to the BH-10 isolate, and the numbers are on the top. The designations of the secondary structural elements are also given. The correspondences for RSV PR are shown below the HIV-1 PR sequence. The abbreviations for the cellular aspartic proteinases have the following meanings: CHY, bovine chymosin, END, endothiapepsin, PEN, penicillopepsin, PEP, porcine pesin, RIZ, rhizopuspepsin. The letters N and C following these abbreviations represent the N- and C-terminal domains of the respective enzymes. (Bottom) Structural alignment of RSV PR and five eucaryotic aspartic proteinases. The correspondences for HIV-1 PR are shown below the RSV PR sequence. Other symbols have similar meaning as described above.

the dimers of HIV-1 PR and the five eucaryotic enzymes with rms deviations ranging between 1.5 and 1.7 Å (Table II and Figure 5). For the superposition of RSV PR and the cellular aspartic proteinases, the number of superimposed C_α pairs ranged from 85 to 96 and the rms deviations were between 1.6 and 1.7 Å. The number of superimposed C_α pairs was smaller in the latter case due to the fact that the flap was disordered in the crystal structure of RSV PR. When the monomers of the retroviral proteinases were aligned separately with the N- and C-terminal domains, more than half the above number of C_α pairs could be aligned with rms values between 1.1 and 1.5 Å (data not shown). The helix in the N-terminal domain of the eucaryotic enzymes that does not superimpose well with that in the retroviral enzymes superimposes more closely in the domain–monomer comparison (in fact, they run parallel to each other when the bilobal cellular enzymes were structurally aligned with the retroviral dimers; see Figures 3 and 4). This is because the retroviral dimers are more or less perfectly symmetric whereas, for the cellular enzymes, the 2-fold symmetry becomes less pronounced as one moves away from the active site. Similar results were achieved when the RSV PR subunits were switched with one another, this simply reflects the fact that the two monomers are nearly indistinguishable from each other (113 equivalent C_α; rms deviation = 0.4 Å) even though the 2-fold symmetry of the dimer is not expressed crystallographically, nor was it enforced during the structure refinement (Jaskólski et al., 1990).

The mean difference in the conformational ϕ and ψ angles for the aligned regions between the retroviral and eucaryotic aspartic proteinases was uniformly around 30° and the estimated standard deviation was also about the same value. In fact, in the case of porcine pepsin, for which three independent crystal structure data are available—two monoclinic structures (Abad-Zapatero et al., 1990; Silecki et al., 1990) and one hexagonal structure (Cooper et al., 1990)—the difference in the conformational angles was approximately 20° (about one esd). This shows that the alignments of the retroviral and cellular aspartic proteinases display approximately the same degree of precision as independent crystal structure determinations with respect to the main-chain conformations.

Conservation of the Active-Site ψ Loop. The core of the retroviral protein monomers and aspartic proteinase domains is formed by double layers of antiparallel β sheets with the strands in each sheet being approximately orthogonal to those in the other. The topology in this region, which also contains the active-site triplet Asp-Thr/Ser-Gly, goes by different names such as the ψ loop or the "wedding ring". Actually the ψ loop itself is quite unusual since it is formed as a three-stranded, mixed β sheet. It is here that the similarities in the structural motifs, as evidenced by a high degree of overlap of both the retroviral proteinases with the mammalian and fungal aspartic proteinases, are easily noticeable. The sequences in this region are well conserved within and across the two families of aspartic proteinases. It is here also that two of the three stretches of conserved sequences among the retroviral proteinases occur (Figure 2). However, it must be noted that, away from the active site, the amino acid sequence similarity across the members of the two families falls below levels of significance. On the other hand, the hydrophobic and hydrophilic characters of residues are reasonably well preserved (see Figure 5) due to the requirement to form β layers of antiparallel strands (Pechik et al., 1989). The low sequence homology of the retroviral proteinases with the eucaryotic enzymes, in spite of the substantial structural and functional homology, is reminiscent of the situation with the icosahedral RNA virus coat proteins, in which very similar β barrel folds



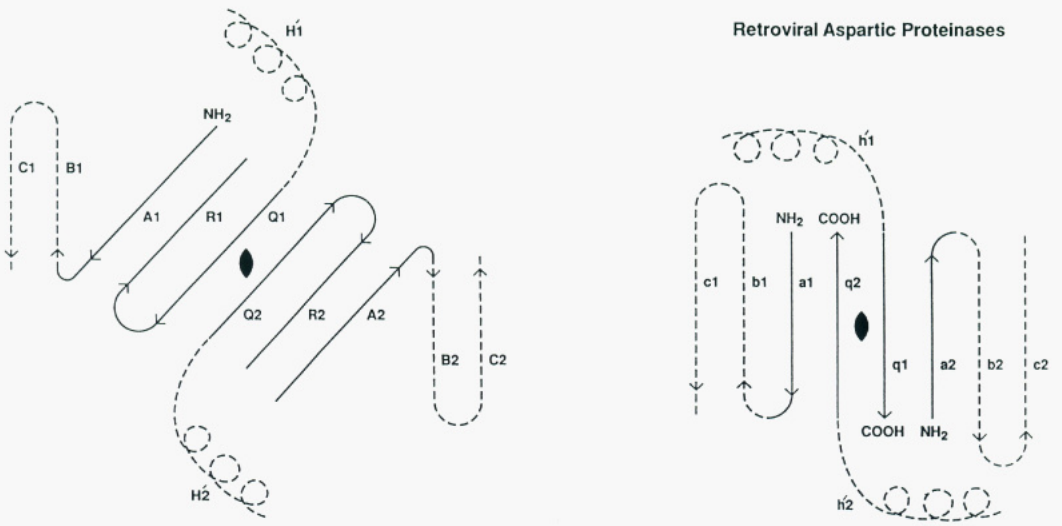


FIGURE 6: (Top) Stereo diagram of the C_α superposition of the interdomain β sheet for rhizopuspepsin (in blue) and the intersubunit β sheet for HIV-1 PR (in yellow). The orientation of the two sheets differs by approximately 40° relative to the core of the two enzyme structures. (Bottom) Cartoon of the interdomain β sheet for eucaryotic aspartic proteinases (e.g., pepsin) and the intersubunit β sheet for retroviral proteinases (e.g., RSV PR). The β strands are represented by arrows (which also indicate the polarity of the strands) and the α helices by coils. The dyads that relate the domains or the subunits are indicated by ♦. Dashed lines represent structural elements behind the β sheet planes.

are observed in the absence of significant sequence homology (Rossmann & Erickson, 1985). The high degree of sequence divergence in both instances may generally be ascribed to the high rate of mutation of RNA viruses (Holland et al., 1982).

Interdomain and Intersubunit β Sheets. There is an antiparallel β sheet formed by strands from both domains (for fungal and mammalian aspartic proteinases) or from both monomers (as in the case of retroviral proteinases). However, this β sheet is formed by six strands in the former and by only four strands in the latter. Initial inspection shows that not only the number of strands but also the strand connectivity and strand directions within the two sheets seem to differ. Besides, the relative orientation and the placement of the strands in the two sheets are significantly different in the four- and the six-stranded β sheets (Figure 6). These facts might argue for structural convergence of the interdomain and intersubunit sheets in the two enzyme classes. However, a detailed examination of the strand topologies in the two sheets reveals a one-to-one structural relationship between the two sets of β strands (Figure 6). The outermost strands of both sheets are formed by the N-termini of the two domains (strands A1 and A2; uppercase letters for the cellular proteinase structural elements with the numbers corresponding to the N- and C-domains) or subunits (strands a1 and a2; lowercase letters for the structural elements of retroviral proteinases with the numbers corresponding to the two subunits). The innermost pair of strands arise from the strands that emanate directly from the conserved α helices (H1′ and H2′). These are the penultimate strands (Q1 and Q2) in the N- and C-domains but are the *C*-terminal strands (q2 and q1) in the retroviral

enzymes. In the cellular enzymes, the order of strands is A1R1Q1 ♦ Q2R2A2 where ♦ represents the pseudodyad; in the retroviral enzymes, the order is a1q2 ♦ q1a2. The inversion in the ordering of the q strands is a consequence of the necessity to preserve the antiparallel topology of the intersubunit β sheet and obviates the need for the C-terminal strands R1 and R2 of the cellular proteinase domains to be present in the retroviral proteinases. Deletion of the terminal strands R1 and R2 without inverting the order of the q1 and q2 strands would have resulted in a mixed parallel/antiparallel four-stranded β sheet instead of a pure antiparallel sheet. Thus, the antiparallel nature of the interdomain sheet of the eucaryotic enzymes has been well preserved in the intersubunit sheet of the retroviral proteinases by the unusual and simultaneous occurrence of strand deletion and strand exchange. The former is a genetic event and the latter is a manifestation of protein folding and assembly; the ultimate result is a stable dimer (Paul Darke, unpublished results) that is held together by interdigitating β strands from the two subunits. Similar interpenetration of subunits resulting in the formation of a stable tetramer has been observed in the crystal structure of the catalase from *Penicillium vitale* (Vainshtein et al., 1986).

Possible Evolutionary Relationship. The acid proteinases belonging to the retroviral and eucaryotic families have one fundamental structural feature in common that underlies a common catalytic mechanism, viz., two aspartic acid residues facing each other across a narrow cleft and bridged by a water-mediated hydrogen bond. Otherwise, they are all involved in diverse activities and functions such as digestion, intracellular protein degradation, and specific processing of precursor proteins. The characteristic signature of the active site in all these enzymes is the pair of Asp-Thr/Ser-Gly triplets. Thus, it is not surprising that they all retain the same basic skeletal architecture enriched with additions and deletions to carry out the specific task for which they are intended. It has been suggested that the cellular enzymes evolved by gene duplication of an ancestral protein that had a motif analogous to one of the two domains of a currently existing proteinase (Tang et al., 1978). Subsequent fusion and independent mutational events in the N- and C-terminal domains gave rise to the present-day eucaryotic enzymes. Whether the retroviral proteinases also evolved independently from such an ancestral gene, whether they represent a direct precursor to the cellular enzymes, or whether they are directly derived from the latter by gene deletion events (similar to the relationship that exists between certain cellular protooncogenes and viral oncogenes (Bishop, 1983) is still unclear. We tend to favor the last hypothesis since the aspartic proteinase motif has not been observed in procaryotes. Dimerization of the basic core motif is the key to the activity of this class of enzymes as evidenced by the experiments of Bianchi et al. (1990), who by autolysis of porcine pepsin were able to show proteolytic activity in a tightly associated molecule consisting of two identical 135 amino acid long fragments derived from the N-terminal domain. This is consistent with the results we obtained in our comparisons of the retroviral proteinase monomer structures with each domain of the cellular aspartic proteinases in which more residues could be superimposed in each domain with a better rms deviation.

A number of intriguing questions remain still unanswered by the present study. Does the subunit structure of the retroviral proteinases confer any functional advantage to the virus during, say, assembly, where structural precursor proteins must certainly dimerize? If not, might one expect, given the unusually high rate of evolution among these enzymes, the possibility of obtaining in a natural form a retroviral proteinase enzyme with fused monomers? Or is it the parsimony in encoding that is of primary importance in the genetic events of the retroviruses? Conversely, since stable dimeric aspartic proteinases can exist, why do the eucaryotic enzymes not utilize this strategy to conserve the genetic material? Is there any advantage to the cell of having a larger gene that would facilitate independent evolution of the two halves of the enzyme? This is indeed a fertile ground for exploration by protein engineering, which can attempt to "evolve" retroviral proteinases into a single fused entity (DiIanni et al., 1990) as the eucaryotic enzymes and vice versa.

ACKNOWLEDGMENTS

We thank Dr. Celerino Abad-Zapatero for useful discussions. We appreciate sincerely the efforts put in by Mr. Mark Kaner and Mr. Frederick Richardson in the preparation of some figures. 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

Abad-Zapatero, C., Rydel, T. J., & Erickson, J. W. (1990) *Proteins: Struct., Funct., Genet. 8*, 62–81.

Bianchi, M., Boigegrain, R. A., Castro, B., & Coletti-Previero, M.-A. (1990) *Biochem. Biophys. Res. Commun. 167*, 339–344.

Bishop, J. M. (1983) *Annu. Rev. Biochem. 52*, 301–354.

Blundell, T. L., Sewell, B. T., & McLachlan, A. D. (1979) *Biochim. Biophys. Acta 580*, 24–31.

Blundell, T. L., Jenkins, J., Pearl, L., Sewell, T., & Pedersen, V. (1985) in *Aspartic Proteinases and their Inhibitors* (Kostka, V., Ed.) pp 155–161, Walter De Gruyter, Berlin.

Blundell, T. L., Jenkins, J. A., Sewell, B. T., Pearl, L. H., Cooper, J. B., Tickle, I. J., Veerapandian, B., & Wood, S. P. (1990) *J. Mol. Biol. 211*, 919–941.

Cooper, J. B., Khan, G., Tickle, I. J., & Blundell, T. L. (1990) *J. Mol. Biol. 214*, 199–222.

Crowther, R. A. (1972) in *The Molecular Replacement Method* (Rossmann, M. G., Ed.) pp 173–178, Gordon & Breach, New York.

Davies, D. R. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 189–215.

Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, pp 345–352, National Biomedical Research Foundation, Washington, DC.

DiIanni, C. L., Davis, L. J., Holloway, M. K., Herber, W. K., Darke, P. L., Kohl, N. E., & Dixon, R. A. F. (1990) *J. Biol. Chem. 28*, 17348–17354.

Erickson, J. W., Rao, J. K. M., Wlodawer, A., & Abad-Zapatero, C. (1989) in *Viral Proteinases as Targets for Chemotherapy* (Krausslich, H., Oroszlan, S., & Wimmer, E. Eds.) pp 191–201, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Erickson, J. W., Neidhart, D. J., VanDrie, J., Kempf, D. J., Wang, X. C., Norbeck, D., Plattner, J. J., Rittenhouse, J., Turon, M., Wideburg, N., Kohlbrenner, W. E., Simmer, R., Helfrich, R., Paul, D., & Knigge, M. (1990) *Science 249*, 527–533.

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. (1990) *J. Biol. Chem. 265*, 14209–14219.

Gilliland, G. L., Winborne, E. L., Nachman, J., & Wlodawer, A. (1990) *Proteins: Struct., Funct., Genet. 8*, 82-93.
Holland, J., Spindler, K., Horodyski, F., Grabov, E., Nichol, S., & Vande Pol, S. (1982) *Science 215*, 1577-1585.
James, M. N. G., & Sielecki, A. R. (1983) *J. Mol. Biol. 163*, 299-361.
Jaskólski, M., Miller, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1990) *Biochemistry 29*, 5889-5898.
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., & Hobart, P. M. (1989) *Nature 342*, 299-302.
Miller, M., Jaskólski, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1989a) *Nature 337*, 576-579.
Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., & Wlodawer, A. (1989b) *Science 246*, 1149-1152.
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. (1989) *Nature 337*, 615-620.
Needleman, S. B., & Wunsch, C. D. (1970) *J. Mol. Biol. 48*, 443-453.
Pearl, L. H., & Taylor, W. R. (1987) *Nature 329*, 351-354.
Pechik, I. V., Gustchina, A. E., Andreeva, N. S., & Fedorov, A. A. (1989) *FEBS Lett. 247*, 118-122.
Power, M. D., Marx, P. A., Bryant, M. L., Gardner, M. B., Barr, P. J., & Luciw, P. (1986) *Science 231*, 1567-1572.
Rao, J. K. M., & Wlodawer, A. (1990) *FEBS Lett. 260*, 201-205.
Remington, S. J., & Matthews, B. J. (1980) *J. Mol. Biol. 140*, 77-99.
Rossmann, M. G., & Erickson, J. W. (1985) in *Virus Structure and Assembly* (Casjens, S., Ed.) pp 29-73, Jones and Bartlett Publishers Inc., Boston, MA.
Schwartz, R. M., & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, pp 353-358, National Biomedical Research Foundation, Washington, DC.
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., & James, M. N. G. (1989) *Science 243*, 1346-1351.
Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., & James, M. N. G. (1990) *J. Mol. Biol. 214*, 143-170.
Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. E., & Davies, D. R. (1987) *J. Mol. Biol. 196*, 877-900.
Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B. H., & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805-8809.
Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A., & Blundell, T. L. (1978) *Nature 271*, 618-621.
Toh, H. H., Ono, M., Saigo, K., Miyata, T., Mellor, J., Fulton, S. M., Dobson, M. J., Wilson, W., Kingsman, S. M., & Kingsman, A. J. (1985) *Nature 315*, 691-692.
Vainshtein, B. K., Melik-Adamyan, W. R., Barynin, V. V., Vagin, A. A., Grebenko, A. I., Borisov, V. V., Bartels, K. S., Fita, I., & Rossmann, M. G. (1986) *J. Mol. Biol. 188*, 49-61.
Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E. T., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616-621.

Hemolytic and Antimicrobial Activities of the Twenty-Four Individual Omission Analogues of Melittin†

Sylvie E. Blondelle and Richard A. Houghten*
Torrey Pines Institute for Molecular Studies, 3550 General Atomics Court, San Diego, California 92121
Received October 10, 1990; Revised Manuscript Received February 20, 1991

ABSTRACT: Although melittin's hemolytic activity has been extensively studied, the orientation of membrane-bound melittin remains uncertain. We have investigated the effect of individually omitted amino acid residues on melittin's activity and related these results to the existing models of melittin–membrane interaction. The extent of hemolysis of the omission analogues closely followed the four known conformational regions of melittin: omission of any of the residues making up the two α-helical regions decreased the hemolytic activity relative to melittin, while omission of any of the residues making up the "hinge" or the C-terminal regions had little or no effect. Our results correlate best with a proposed model in which melittin initially forms "holes" in the membrane, resulting in an initial rapid loss of hemoglobin; the membrane-bound melittin is then internalized into the membrane, resulting in a later slow phase of hemoglobin loss. It was also found that induced structural effects caused by peptide–lipid interactions could be studied by using RP-HPLC, with an excellent correlation found between the retention times of the individual omission analogues and their hemolytic activities.

Melittin, the predominant peptide isolated from honey bee venom (*Apis mellifera*), is known for its marked cytolytic activity (Habermann & Jentsch, 1967). This peptide, consisting of 26 amino acid residues, exhibits strong amphipathic surface activity, although no simple relationship has been found between its surface activity and lytic properties (Fennell et al., 1968; Stocker & Traynor, 1986; Dorman & Markley, 1971; Schröder et al., 1971; Habermann, 1972; Dawson et al., 1978). In our continuing studies of lipid-induced changes in peptide conformations (Houghten & DeGraw, 1987; Houghten & Ostresh, 1987; Büttner et al., 1990a,b; Blondelle et al., 1991;

† Supported in part by National Science Foundation Grant DIR8713707 to R.A.H.

0006-2960/91/0430-4671$02.50/0 © 1991 American Chemical Society