Tapprich, W. E., & Hill, W. E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 556–560.
Tapprich, W. E., Goss, D., & Dahlberg, A. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4927–4931.
Thanaraj, T. A., & Pandit, M. W. (1989) *Nucleic Acids Res. 17*, 2973–2985.
Wakao, H., Romby, P., Westhof, E., Laalami, S., Grunberg-Manago, M., Ebel, J. P., Ehresmann, C., & Ehresmann, B. (1989) *J. Biol. Chem. 264*, 20363–20371.
Watanabe, K., Kuchino, Y., Yamaizumi, Z., Kato, M., Oshima, T., & Nishimura, S. (1979) *J. Biochem. (Tokyo) 86*, 893–905.
Weiel, J., & Hershey, J. W. B. (1981) *Biochemistry 20*, 5859–5865.
Weiel, J., & Hershey, J. W. B. (1982) *J. Biol. Chem. 257*, 1215–1220.
Zucker, F. H., & Hershey, J. W. B. (1986) *Biochemistry 25*, 3682–3690.

Articles

Structure of the Aspartic Protease from Rous Sarcoma Retrovirus Refined at 2-Å Resolution†

Mariusz Jaskólski,‡,§ Maria Miller,‡ J. K. Mohana Rao,‡ Jonathan Leis,‖ and Alexander Wlodawer*,‡
Crystallography Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, P.O. Box B, Frederick, Maryland 21701, and Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106
Received December 21, 1989; Revised Manuscript Received March 20, 1990

ABSTRACT: The structure of Rous sarcoma virus protease has been solved by multiple isomorphous replacement in the crystal form belonging to space group $P3_121$, with unit-cell parameters a = 88.95 Å and c = 78.90 Å. The enzyme belongs to the family of aspartic proteases with two identical subunits composing the active homodimer. The noncrystallographic dyad relating these two subunits was identified after preliminary tracing in the MIR map and was used for phase improvement by electron-density averaging. Structure refinement resulted in a model that included 1772 protein atoms and 252 water molecules, with an R factor of 0.144 for data extending to 2 Å. The secondary structure of a retroviral protease molecule closely resembles that of a single domain in pepsin-like aspartic proteases and consists of several β-strands and of one well-defined and one distorted α-helix. The dimer interface is composed of the N- and C-terminal chains from both subunits which are intertwined to form a well-ordered four-stranded antiparallel β-sheet. In each monomer, the catalytic triad (Asp-Ser-Gly) is located in a loop that forms a part of the ψ-structure characteristic to all aspartic proteases. The position of a water molecule between the active-site aspartate residues and the general scheme of H bonding within the active site bear close resemblance to those in pepsin-like aspartic proteases and therefore suggest a similar enzymatic mechanism. The binding cleft over the active site is covered by two flap arms, one from each monomer, which are partially disordered. The retroviral protease dimer has been compared with several enzymes of cellular origin, with chains aligning to an rms deviation of 1.90 Å or better.

Avian retroviruses contain an RNA genome encapsidated in a symmetric icosahedral core that is surrounded by a lipoprotein envelope (Dickson et al., 1984). The viral genes *gag*, *pol*, and *env* are translated as precursor polypeptides which are cleaved to produce the mature form of the proteins found in virions (Dickson et al., 1984). Processing of the *gag* and *pol* gene protein precursors of avian sarcoma/leukosis virus (ASLV), murine leukemia viruses (MuLV), and human immunodeficiency virus (HIV) has been shown to be directed by a viral-encoded protease (Dickson et al., 1984; Katoh et al., 1985; Kramer et al., 1986).

In avian retrovirus-infected cells, the viral-specific protease (PR) is synthesized as part of two large precursor polypeptides, $Pr76^{gag}$ and $Pr180^{gag\text{-}pol}$. The former includes the PR at its C-terminus, and the latter contains all of the *gag* and the *pol* sequences (Dickson et al., 1984) and is produced at $^1/_{20}$ the amount of $Pr76^{gag}$. The ASLV PR gene encodes a 13.5-kDa protein comprised of a single contiguous polypeptide chain of 124 amino acids. The enzyme purified from virions, however, is a dimer that exhibits a substrate specificity for its homologous *gag* and *pol* polyproteins although there are reports of heterologous substrates being cleaved by the PR (Skalka, 1989). Synthetic peptides that represent homologous *gag* and *pol* precursor cleavage sites are specific substrates for the PR and, in contrast to precursor polypeptides, are rapidly cleaved (Kotler et al., 1988, 1989).

The polyprotein cleavage sites have been deduced by alignment of the amino acid sequences at the C- and N-termini of the mature *gag* and *pol* proteins to sequences found in the precursor. Comparison of the sequence around 27 cleavage sites from several retroviruses indicates that each is cleaved at a unique sequence. The only common feature is that the

† Research sponsored in part by the National Cancer Institute, DHHS, under Contract NO1-CO-74101 with ABL, and in part by NCI Grant CA 38046 to J.L.
* To whom correspondence should be addressed.
‡ ABL-Basic Research Program.
§ On leave from Faculty of Chemistry, A. Mickiewicz University, Poznań, Poland.
‖ Case Western Reserve University.

0006-2960/90/0429-5889$02.50/0

Table I: Completeness of Data as a Function of Resolution for RSV PR[a]

no.	resol	N_{tot}	N_{obs}	1[b]	2[b]	3[b]	4[b]	$\langle F/\sigma \rangle$
1	5.46	1374	1347	98.03	98.03	92.45	94.30	133.97
2	4.29	1375	1353	98.40	98.22	93.12	94.47	124.74
3	3.73	1373	1352	98.47	98.30	93.43	94.61	95.67
4	3.37	1374	1339	97.45	98.09	91.78	94.50	76.42
5	3.12	1375	1324	96.29	97.73	90.75	94.45	58.93
6	2.93	1375	1291	93.89	97.09	88.00	94.33	42.24
7	2.78	1373	1262	91.92	96.35	85.50	94.14	34.64
8	2.65	1374	1214	88.36	95.35	82.31	94.01	28.59
9	2.55	1375	1199	87.20	94.45	80.90	93.87	24.05
10	2.45	1375	1133	82.40	93.24	75.84	93.69	21.09
11	2.38	1374	1102	80.20	92.06	73.13	93.45	19.05
12	2.30	1373	1070	77.93	90.88	71.67	93.33	16.81
13	2.24	1375	1065	77.45	89.85	71.96	93.29	15.00
14	2.18	1373	963	70.14	88.44	64.54	93.20	11.41
15	2.06	1375	821	59.71	86.52	24.50	85.92	9.68

[a] Input data consisted of 20613 reflections of which 17835 were observed. [b] 1, percent of observed as fraction of measured (zone); 2, percent of observed as fraction of measured (cumulative); 3, percent of observed as fraction of theoretical (zone); 4, percent of measured as fraction of theoretical (cumulative).

amino acids at or near the cleavage site are hydrophobic. Thus, the basis for specificity is not fully understood at present.

The viral PRs are similar to the cellular aspartic proteases in that each contains a conserved amino acid sequence, Asp-Thr/Ser-Gly, in the active site of the enzyme (Tang, 1979; Blundell et al., 1983; Pearl & Taylor, 1987). This relationship has been firmly established by the preliminary three-dimensional structure of the Rous sarcoma virus (RSV) PR reported at 3.0-Å resolution (Miller et al., 1989b) and by the structure of recombinant and synthetic HIV-1 PR (Navia et al., 1989; Wlodawer et al., 1989, Lapatto et al., 1989). The structure of the RSV PR has now been further refined to a resolution of 2.0 Å and is the subject of the present discussion.

Materials and Methods

Isolation and Purification of the Protein. RSV (Prague C) was grown and purified on a sucrose density gradient as previously described (Smith et al., 1977). The banded virus (1 g) was suspended in 100 mL of 0.58% NaCl with a Teflon homogenizer. To this suspension, 210 mL of a chloroform–methanol mixture (2:1 v/v) was added at room temperature. The phases were stirred for 5 min and separated in an International centrifuge at 1500 rpm for 60 min at 4 °C. The upper aqueous phase, which contains the solubilized PR as well as the p10 protein, was collected, dialyzed for 3–4 h against 4 L of 10 mM β-mercaptoethanol at 4 °C, and then lyophilized. The lyophilized protein was dissolved in 10 mL of 10 mM Hepes, pH 8.0, 0.1 mM EDTA, and 10 mM β-mercaptoethanol and passed through a Sephadex G-75 column (5 × 30 cm) equilibrated with the same buffer containing 1 M NaCl. The PR, which eluted between 42 and 52% of the column volume, was collected, dialyzed against 2 L of 10 mM β-mercaptoethanol overnight at 4 °C, and lyophilized. The PR was taken up in water prior to crystal growth.

Crystallization and Data Collection. Prismatic crystals of RSV PR were grown by the vapor diffusion method (Miller et al., 1988), using ammonium sulfate as the precipitant, under conditions in which the protease is active (pH 5.4). The crystals are trigonal (space group $P3_121$) with unit-cell parameters a = 88.95 Å and c = 78.90 Å and with two protein molecules in the asymmetric unit (V_M = 3.3 Å^3/dalton). Several heavy-atom derivatives were obtained by soaking the crystals in uranyl acetate, mercury acetate, mersalyl, and methylmercury acetate. Diffraction data for the native crystals and for the derivatives were collected on a Siemens electronic area detector mounted on a Rigaku RU-200 rotating-anode generator, operated at 50 kV and 100 mA. The best native data set consisted of 95106 individual measurements of 20613 unique reflections, which merged with a weighted R_{sym} on intensities of 6.9%. This data set was 86% complete to 2-Å resolution, and 87% of the measurements were significant [$I > 1.5\sigma(I)$]. The fraction of observed reflections in the outermost 2.2–2.05-Å shell was 60% (Table I). Only the observed data as defined above were used in structure solution and refinement.

Structure Solution. The structure was solved by the multiple isomorphous replacement method (MIR). A difference Patterson map for the uranyl acetate derivative revealed a single, highly occupied site that later was identified as being located in the active site of the dimeric enzyme. Three mercury derivatives were solved by difference Fourier and Patterson maps. A combined 3-Å lack-of-closure refinement of the four derivatives, including anomalous scattering for the uranyl derivative, converged with a figure of merit of 0.67 (Table II).

The choice of $P3_121$ as the correct space group enantiomorph was made after calculation of difference Fourier maps for the mercury derivatives using uranyl phases in both enantiomorphic space groups and including anomalous scattering (the heavy-atom peaks were ca. 2.5 times higher for the correct assignment of Friedel pairs). It was confirmed by the correct stereochemistry of the model, which included right-handed helices. Also, the final statistics of the solvent flattening procedure of Wang (1985) indicated $P3_121$ as the correct space group: figure of merit 0.87 (0.85); map inversion R factor 0.184 (0.245); F_o/F_c correlation coefficient 0.977 (0.961) (values in parentheses are for $P3_221$).

The initial 3-Å electron-density map based on MIR phases from two derivatives (uranyl acetate and methylmercury acetate) revealed a number of stretches of continuous density. Polypeptide chain tracing in this map resulted in an initial model consisting of several polyalanine chains with a total of 214 residues (Miller et al., 1989a). This model was used to locate noncrystallographic symmetry, which could not be found with rotation function. Also, the lack of knowledge about the location of noncrystallographic symmetry relating the two molecules made correlation with amino acid sequence impossible. A novel strategy to isolate the chain fragments corresponding to the two independent molecules was developed in which the C_α positions of the initial fit were used to find the orientational relationships among the chain fragments. The search was done by structural comparisons with a package of programs developed by J.K.M.R. The method is similar to that presented by Remington and Matthews (1980). Instead

Table II: Heavy-Atom Parameters after Lack-of-Closure Refinement at 3 Å and Peaks in Difference Fourier Maps at 4 Å Calculated with the Phases from the Refined Protein Model[a]

derivative	site	x	y	z	B (Å^2)	R_c[b]	$O(1)$[c]	$O(2)$[c]	location
uranyl acetate	1	−0.080	0.457	−0.011	31.5	0.49	2.8	2.8	active site
mercury acetate	1	0.498	0.460	0.127	36.2	0.61	1.3	1.3	Cys313[d]
	2	0.755	0.456	0.072	32.7		0.7	0.6	Cys113[d]
	3	−0.082	0.457	−0.013	26.1		1.1	0.8	active site
	4	0.798	0.516	0.034	36.3		0.5	0.6	Cys113′
mersalyl	1	0.498	0.460	0.127	30.5	0.69	1.2	1.1	Cys313
	2	0.753	0.456	0.071	32.8		1.2	1.2	Cys113
	3	−0.080	0.457	−0.013	16.6		0.5	0.4	active site
	4	0.571	0.079	0.084	38.8		0.5	0.5	His7
methylmercury acetate (4 Å)	1	0.502	0.462	0.127	26.7	0.68	0.8	0.6	Cys313
	2	0.756	0.457	0.071	26.3		1.1	1.0	Cys113
	3	0.472	0.495	0.152	37.8		0.6	0.7	Cys313′

[a] The resolution limit of the Fourier maps was chosen so that similar number of structure factors could be used for each derivtive. [b] Cullis R factor: $R_c = (\sum|F_{h(obs)} - F_{h(calc)}|)/\sum F_{h(obs)}$. [c] Occupancies: $O(1)$ in arbitrary units, results of phase refinement; $O(2)$ in 4-Å difference Fourier map, scaled as in $O(1)$. [d] Cysteines 113 and 313 have the thiol groups disordered, with alternate orientations denoted as 113′ and 313′, respectively.

of two different proteins being compared for structural similarities, parts of the same protein were compared. All possible C_α segments of the model were slid over all other segments in both the forward and reverse directions. The comparisons had to be performed with the model running in both directions as the initial fit did not guarantee correct polarity. The segment length varied from 15 to 25 residues. The lower limit could not be increased as the shortest fragment was only 15 residues long. When the model was compared against itself in the forward direction, three pairs of mutually exclusive fragments had very similar Eulerian rotation angles. A different pair of fragments yielded a similar relationship when the segments were running in opposite directions. The rotational and translational parameters thus obtained were refined by least-squares minimization, and a few more residues were assigned that completed the structure superposition. The RSV PR monomers isolated this way were initially not related by a 2-fold rotation, but a dimer with approximate 2-fold symmetry was obtained by making use of the space group symmetry.

The 3-Å electron density map was later calculated on the basis of all four derivatives and the anomalous scattering of the uranyl, and it was significantly improved with the solvent flattening procedure of Wang (1985). The molecular boundary of the dimer and several characteristic features, including the active site and the electron density corresponding to the amino acid sequence Trp-Pro-Thr-Asp-Trp-Pro, could be easily recognized. At that stage, once the two monomers comprising the noncrystallographic dimer were identified, the electron-density map could be averaged. In the present case, because of the still unresolved problem of the exact location of molecular boundaries, a procedure called the "double-averaging electron density correlation technique" was adopted. To begin with, approximate limits to the boundaries of a parallelepiped that probably encloses each subunit were established. Mere knowledge of the C_α coordinates is enough for this purpose. It does not matter if portions of another subunit, besides the solvent, are in this box. In the case of RSV PR, there are two monomers in the asymmetric part, denoted I and II. Each integral grid point in box I is transformed to a point, usually nonintegral, in box II by application of the noncrystallographic symmetry operation that relates monomers I and II. The electron density at the nonintegral grid point is obtained by linear interpolation. The original density at the integral grid point and the interpolated electron density at the nonintegral grid point were averaged and stored at the corresponding grid point in box I. The same procedure was used for integral points in box II, and the nonintegral grid points in box I were obtained by application of the inverse symmetry operation. After averaging, all points were referred to a chosen asymmetric unit. Care is taken to count the points in the asymmetric part that were obtained as a consequence of overlapping regions of the two boxes only once. In this way, even if the molecular masks are not exact and overlapping occurs between them, no problems arise. This procedure is therefore quite useful in the preliminary stages of electron-density averaging in the presence of noncrystallographic symmetry.

The overall correlation coefficient between electron densities in subunits related by noncrystallographic symmetry is sensitive to the rotation and translation parameters that relate noncrystallographically equivalent subunits. In fact, the correlation coefficient between subunits could be improved by as much as 0.05–0.10 by slight changes in translation (less than 1 Å) and Eulerian rotation angles (less than 1°). Thus, the initial rotation and translation parameters that relate the monomers are amenable for easy refinement. Since the density averaging is performed twice (box I to box II as well as box II to box I) and the criterion for improvement of the averaging procedure is the correlation coefficient between the electron densities in the regions related by the noncrystallographic symmetry, the name "double-averaging" (or multiple averaging if there are more than two subunits in the crystallographic asymmetric unit) electron density correlation technique is perhaps apt for this procedure. The averaged density is then Fourier-transformed and the new phases are applied to the observed structure factors, a map is calculated from the new phases, and the entire procedure is repeated. By this method, it was possible to increase the mean density correlation coefficient for the two subunits from 0.35 to 0.60. The quality of the map increased considerably as the averaging procedure was repeated. The other molecule in the dimer was obtained by rotating molecule I around the noncrystallographic dyad. With the exception of 10 amino acids in the flap region (see below), all other residues could be identified.

Structure Refinement and Rebuilding. Two programs were utilized in structure refinement. Conventional restrained refinement was accomplished with the program PROLSQ (Hendrickson, 1985), in an implementation including fast Fourier transforms (Finzel, 1987) and restraining intermolecular contacts (Sheriff, 1987). Refinement that included molecular dynamics was performed with the program X-PLOR (Brünger et al., 1987). Difference Fourier maps were calculated with the PROTEIN package (Steigemann, 1974) and were displayed on an Evans and Sutherland PS390 graphics system running FRODO (Jones, 1985).

The initial, fairly restrained, structure-factor least-squares refinement (using PROLSQ) of the starting model utilized 10–

Table III: Final Statistics of the RSV PR Model

R factor = $\sum\|\|F_o\| - \|F_c\|\|/\sum\|F_o\|$	0.144
weights	$w = \sigma_F^{-2}$
with	$\sigma_F \approx 0.5\langle\|\|F_o\| - \|F_c\|\|\rangle$
resolution (Å)	10–2
no. of reflections	17 609
no. of atoms	2022
rms deviations from ideality (target restraints in parentheses)	
distance restraints (Å)	
bond distance	0.022 (0.020)
angle distance	0.062 (0.040)
planar 1–4 distance	0.076 (0.050)
plane restraints (Å)	0.020 (0.018)
chiral center restraints (Å^3)	0.225 (0.150)
nonbonded restraints (Å)	
single-torsion contact	0.202 (0.300)
multiple-torsion contact	0.214 (0.300)
possible (X···Y) H-bond	0.191 (0.300)
conformational torsion angles (deg)	
planar	3.4 (3.0)
staggered	16.9 (10.0)
orthonormal	39.6 (20.0)
B_{iso} restraints (Å^2)	
main-chain bond	2.218 (2.000)
main-chain angle	3.144 (2.500)
side-chain bond	5.444 (3.500)
side-chain angle	7.612 (4.500)
H-bond	14.346 (16.000)

3-Å data but was quickly expanded to include the reflections in the 10–2.5-Å resolution range (R = 0.307). At this stage the model was passed through one round of molecular-dynamics energy minimization with the crystallographic R factor as an energy term, X-PLOR being used. This round consisted of a heating stage (4000 K for 2 ps) followed by slow cooling (from 4000 to 300 K in 4000 steps, 1 fs each, velocities rescaled every 25 fs) and conjugate-gradient crystallographic refinement. The molecular dynamics refinement converged with an R factor of 0.267. All subsequent refinement was done according to the conventional restrained structure-factor least-squares procedure with individual isotropic temperature factors. In each round of this refinement the model was first idealized with tight geometrical restraints, then refined with loose restraints, and, finally, refined with tight restraints. After each round of refinement the model was carefully examined and manually adjusted in $|2F_o - F_c|\alpha_c$ maps. In the round following the molecular-dynamics minimization, the data set was expanded to include all 17 609 reflections with $I > 1.5\sigma(I)$ in the 10–2-Å resolution range, and the refinement converged at R = 0.239. Water molecules were located and verified during the manual intervention sessions between the rounds of refinement that followed. The final modeling consisted of six rounds of refinement using 10–2-Å data and six manual rebuilding sessions. The structure-factor weighting scheme used during the final refinement was of the form $w_F = \sigma_F^{-2}$ with $\sigma_F = [35.0 - 150.0(\sin\theta/\lambda - 1/6)]$, which corresponded to $\sigma_F \approx 0.5\ \langle\|F_o| - |F_c\|\rangle$ (Table III). The rms positional shift in the last cycle of refinement was 0.007(5) Å, and the final model is characterized by an R factor of 0.144 and by an rms deviation between ideal and observed bond distances of 0.022 Å (Table III). The error in the atomic coordinates estimated by the method of Luzzati (1952) is 0.15 Å. The coordinates and structure factors have been deposited with the Brookhaven Protein Data Bank (Files 2RSP and R2RSPSF, respectively).

RESULTS

Atomic Model. The refined model of RSV PR consists of 1772 protein atoms with an average temperature factor of 20.9 (4) Å^2 and of 252 water molecules. In the following discussion,

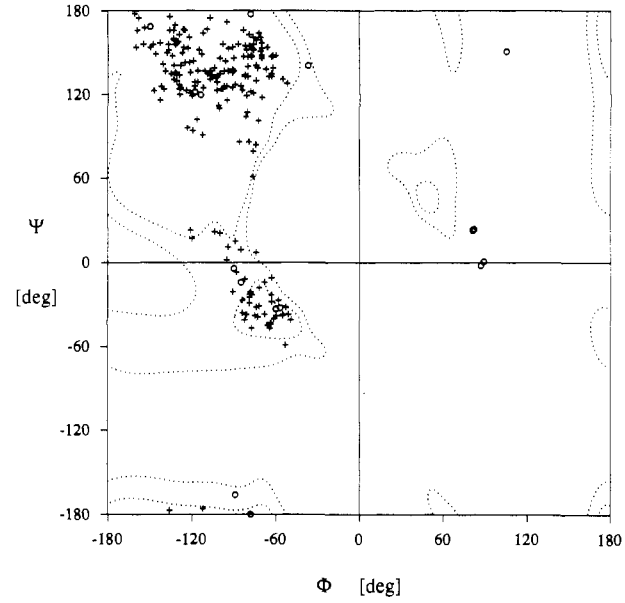

FIGURE 1: Ramachandran (ϕ–ψ) plot for RSV PR. Circles (O) mark glycine and crosses (+) all other residues. The isoenergy contours correspond to 4 and 8 kcal/mol levels in the conformational map of alanine dipeptide (Peters & Peters, 1981).

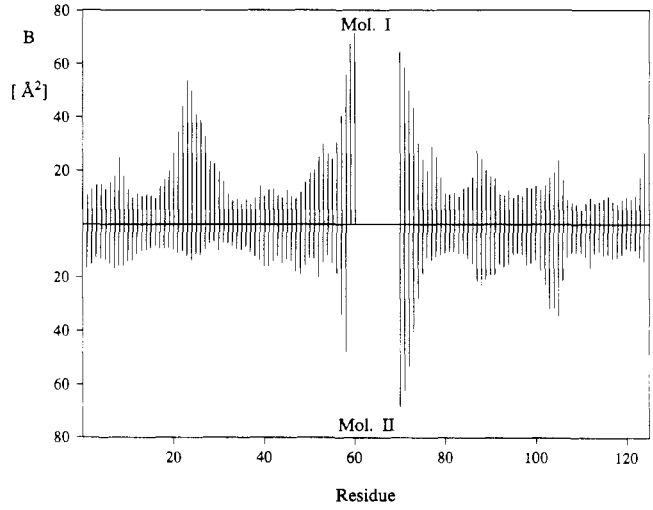

FIGURE 2: Isotropic temperature factors averaged over the main-chain atoms at each residue in molecule I (top) and in molecule II (bottom).

the residues in molecule I (Mol. I) are numbered 1–124 while the corresponding residues in molecule II (Mol. II) have numbers incremented by 200 (201–324). Nine residues in Mol. I and 11 in Mol. II (all in flap regions) are disordered and are not included in the model. The side chains of the two cysteine residues in the structure (one in each monomer) assume two alternate trans–gauche conformations in each molecule in agreement with heavy-atom data (Table II). The quality of the model is illustrated in Figure 1, which plots the main-chain ϕ/ψ angles in the form of a Ramachandran plot. All non-glycine residues are clustered in the allowed regions, consistent with the conformatinal map of the alanine dipeptide calculated by Peters and Peters (1981). The few outliers are all Gly residues, which are known to have much larger conformational freedom and which belong to a region where glycines have been frequently observed.

Figure 2 illustrates the isotropic temperature factors calculated in the final refinement. The B_{iso} values for the main-chain atoms in each residue have been averaged and plotted for both molecules in the noncrystallographic dimer. The alternating low B–high B pattern distinctly seen along the

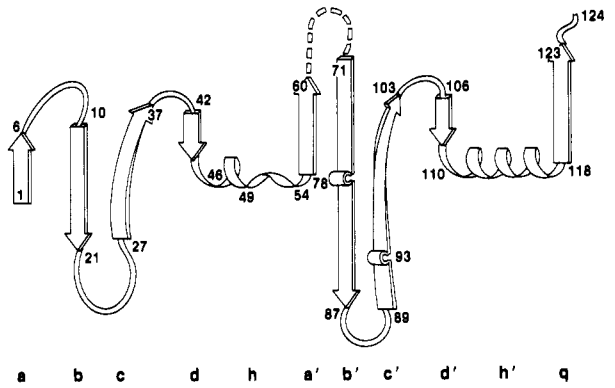

FIGURE 3: Schematic representation of the secondary structure elements in the RSV PR molecule. Designation of these elements is discussed in the text.

polypeptide chain correlates very well with the secondary structure of the monomer discussed below. The higher temperature factors correspond to surface turns and loops around residues 8, 24, 65, 88, 105 (Figures 3 and 4), which connect the β-sheet/helical segments of the chain characterized by low temperature factors. In particular, the highest thermal motion is observed in the flap loops, the tips of which could not be traced in the electron-density maps. Also remarkable is the agreement of the temperature-factor pattern in the two molecules of the dimer. The only exception is the loop 221–227 in Mol. II, since the thermal motion of residues 21–27 in Mol. I is much higher. This phenomenon is, however, not surprising in view of the different crystal-packing contacts in which the 21–27 loops of the two monomers are involved. The 221–227 loop is buried in a densely packed region where it is surrounded by the 221–227 loop from a symmetry related molecule and by the amino end of the c′ β-strand, the top of the h′ helix, and the dimer interface, all from another dimer in the crystal lattice (see Figure 5 and below). In contrast, the 21–27 loop in Mol. I is exposed to an empty channel into which only the flap arms project and has the 206–210 loop from another molecule as the only near neighbor. The crystal packing of RSV PR is shown in Figure 5.

Folding of the Protease Monomer. In principle, the topology of the RSV PR molecule is a simplification of that of a single domain in pepsin-like aspartic proteases and can be described with the conventions of Blundell et al. (1985). A schematic representation of the secondary structure elements in the RSV PR molecule is shown in Figure 3, which is based on the main-chain H-bonding pattern illustrated in Figure 4. In Figure 6 the three-dimensional structure of the RSV PR dimer is shown with the arrows representing individual strands in β-sheets and coils representing α-helices. The N-terminal a β-strand (residues 1–5) forms the outer part of the interface β-sheet (see Architecture of the RSV PR Dimer). The b β-strand (11–20) continues through a broad surface loop (21–27) into the c β-chain, which terminates at the active-site triplet (37–39). At the active-site Asp37 the chain changes its direction and, having completed a full turn at residue 42, continues as chain d (43–47). In pepsin-like proteases chain d is followed by the h helix, while in the RSV PR molecule this helix is limited to one turn (47–49) but is immediately followed by a very characteristic Trp-Pro-Thr-Asp-Trp-Pro loop (50–55), which could be considered a distorted continuation of h. The rest of the moleucle (primed chains in Blundell's notation) has a topology related to that described above by an approximate intramolecular 2-fold symmetry. The intradomain symmetry is illustrated in Figure 7, which represents the C_α backbone of Mol. I. Residues 55–57 form the

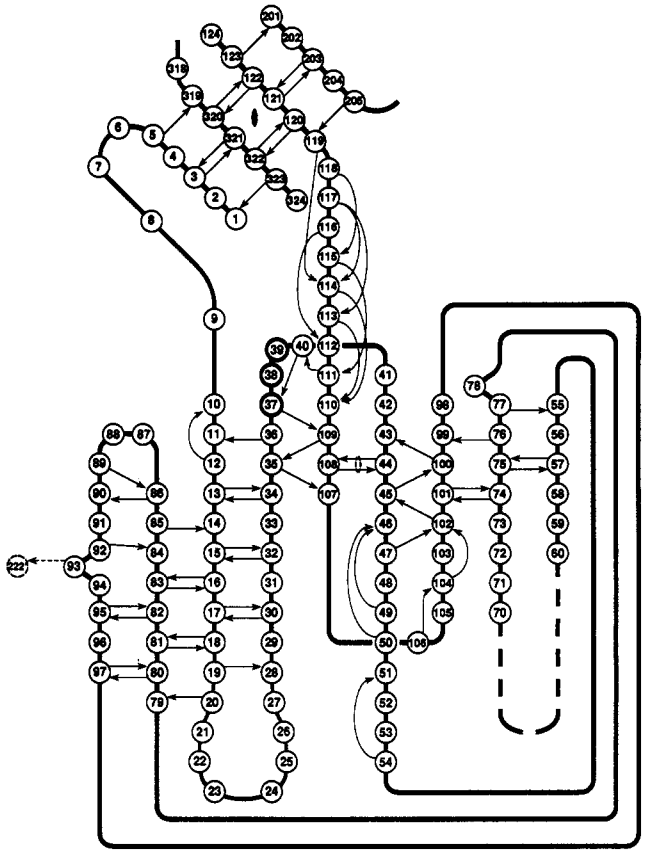

FIGURE 4: Diagram of the secondary structure of the RSV protease with main-chain H-bonds indicated by arrows (→). The active-site triplet (Asp-Ser-Gly) is marked by thicker circles, and the broken circle represents a residue from another dimer in the unit cell. The diagram shows only one molecule (residues 1–124) of the dimer and the terminal strands from the second domain (residues 201–324) in the interface region. The complete dimer is generated by the noncrystallographic dyad indicated by the solid lens, between residues 121 and 321. The H-bond pattern in the second domain is identical with that in the first domain except for the intermolecular 93–222′ and 106–104 H-bonds, which are not present there. The flap residues missing from the model have been omitted from the diagram and are indicated by a broken-line loop. The broken-line lens between residues 44 and 108 represents the approximate 2-fold symmetry present in the single domain of this protease.

a′ β-strand and, as in pepsin-like proteases, lead to the flexible flap loop, which comprises residues 61–69. This loop is disordered in the crystal as there is no interpretable electron density in that region in either of the two molecules. Following the flap loop is the b′ chain (74–86), which after a turn becomes the c′ β-strand (90–103). Another turn at 104–106 changes the chain into d′ (107–110), which leads directly to a well-defined h′ α-helix (110–118). This helix is followed by a straight C-terminal β-strand (119–123), which can be designated as q. In constrast, pepsin-like aspartic proteases have a double-stranded β-sheet here. This C-terminal region forms the inner part of the interdomain (pepsin-like proteases) or intradimer (retroviral protease) interface. The numerous β-strands in the molecular core are organized into a sandwich of two four-stranded β-sheets with perpendicular chain directions (Figure 6). One of these sheets has antiparallel β-chains, c, b, b′, and c′, whereas the other sheet is composed of two ψ-shaped structures superimposed in such a way that the central strand in each makes an outer one in the other (Figure 4). Only the central chains in this sheet are antiparallel; the outer ones are parallel to their neighbors. One of the ψ-structures comprises chains c (35–37), d, and d′ and the other is made up of c′ (100–102), d′, and d.

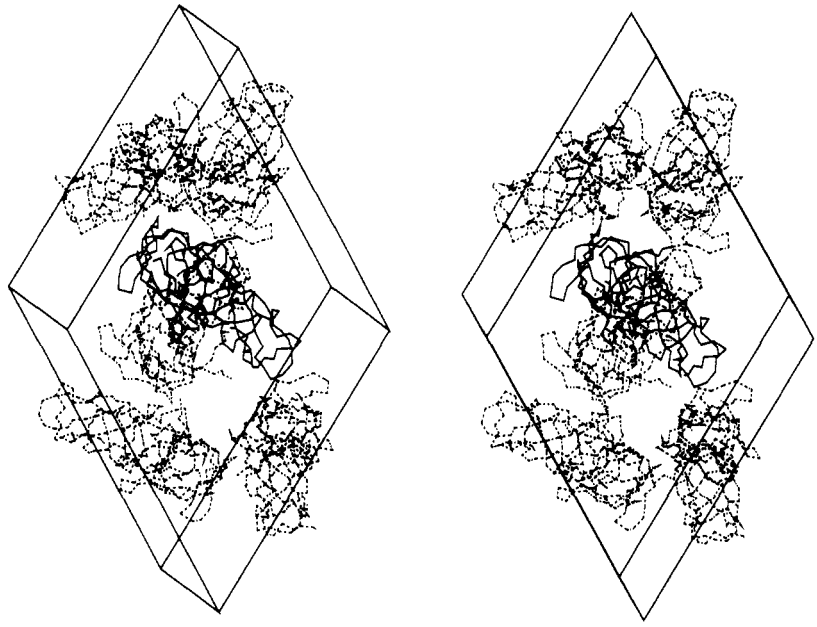

FIGURE 5: Stereo diagram of the unit cell showing the packing of the six symmetry-related RSV PR dimers in the unit cell.

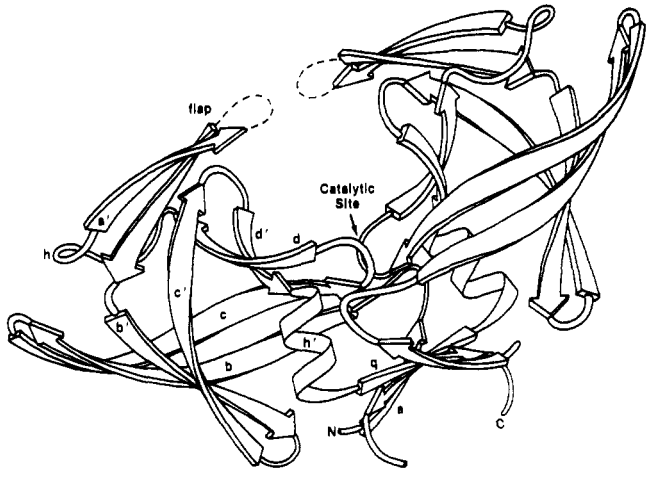

FIGURE 6: Schematic representation of the three-dimensional structure of the RSV PR dimer. Elements of secondary structures are marked for one subunit only.

Architecture of the RSV PR Dimer. Assembling of the two RSV PR monomers into a pepsin-like structure is achieved by using their terminal strands to form a common β-sheet—referred to as the dimer interface (Figures 4 and 6). Even though the detailed connectivity in the interface region is different in these two types of proteases, the overall topology is similar and can be described as an antiparallel β-sheet consisting of six strands in pepsin-like proteases compared to four in the retroviral protease. This sheet forms the bottom of the dimer on which the active-site loops are situated (Figure 6). The two flap arms project over the active site guarding the entrance to a large cleft which can be recognized as the substrate-binding pocket. The two molecules in the RSV PR dimer are related by a nearly perfect 2-fold axis. They superimpose at a 178° rotation with an rms deviation for their C_α atoms of 0.4 Å. The deviations (Δ) between the corresponding atoms of superimposed molecules I and II averaged at each residue over the main-chain and side-chain atoms are plotted in Figure 8. It should be noted that the pattern of alternating small Δ–large Δ observed in this plot follows closely the low*B*–high *B* pattern of the temperature-factor distribution in Figure 2. As expected, the deviations in the side chains are larger than those in the backbone. In particular, the largest deviations between side-chain atoms (e.g., at 105) correspond to residues with long side chains (Arg, Lys, etc.) located in surface loops exposed to the solvent.

The noncrystallographic 2-fold axis of the principal dimer in the unit cell is almost coincident with the crystallographic

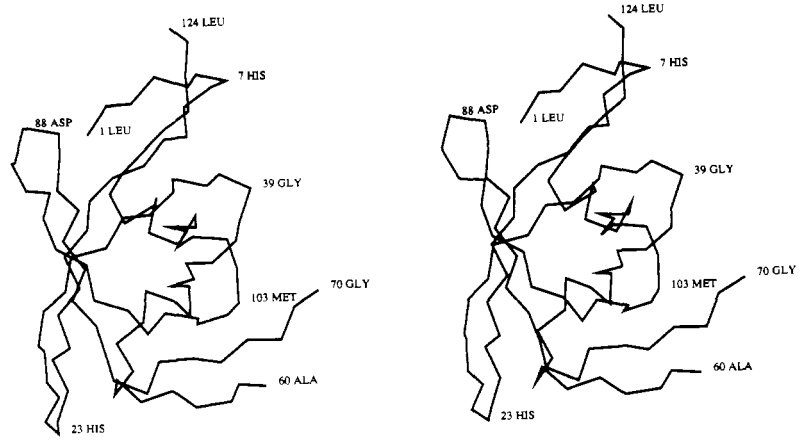

FIGURE 7: Tracing of the C_α atoms in the monomer of RSV PR. In this orientation the approximate 2-fold symmetry of the molecule can be easily seen, the corresponding axis being nearly horizontal and in the plane of the figure.

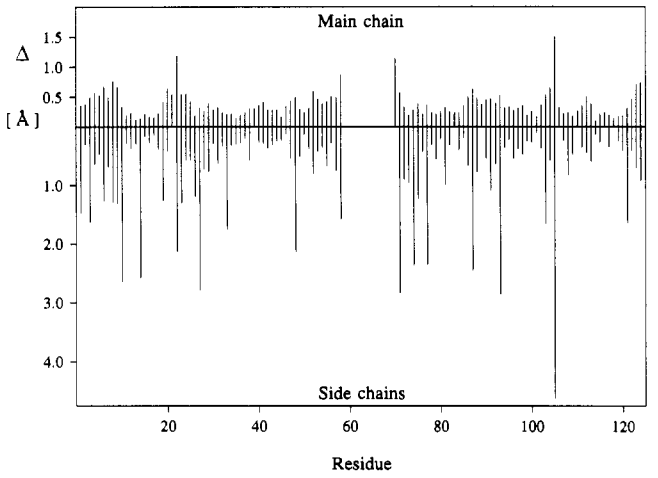

FIGURE 8: Average deviations between corresponding atoms of the two monomers after least-squares alignment.

[21$\bar{3}$] direction. In projection along c, the dimer 2-fold axis is coincident with the diagonal [21] direction on the trigonal (a,b) plane, but it makes an angle of 147° with the c axis (33° with $-c$). The noncrystallographic dyad passes between the flaps, then through a water molecule between the two active-site Asp carboxylic groups, and finally through the interface β-sheet, between the two C-terminal strands (Figure 6). These three loci are the regions where the two monomers come into intimate contact although in the case of the flaps it is only an assumption as their tips are missing from the model. Analysis of crystal packing (Figure 5) suggests that the flaps from dimers related by the crystallographic 2-fold axis project into a common solvent channel where they can assume different positions and become disordered, possibly to avoid clashes that would result if they were to follow the trends of their visible stems.

The RSV PR dimer has been compared with several fungal and mammalian enzymes, for which structures are available in the Protein Data Bank. In all cases, the chains could be aligned to an rms deviation of 1.90 Å or better for nearly half of the C_α atoms in the RSV PR dimer (Table IV). In those structural alignments the C-terminal h′ helix of the RSV PR dimer superimposed quite well with the equivalent C-domain helix in aspartic proteases and ran parallel but displaced relative to the corresponding N-domain helix. This region of the structure also exhibits sequence similarity for the compared proteins. Such sequence similarity has been observed for all known retroviral proteases as well. In particular, the recently

Table IV: Statistics for Structural Superpositions of RSV PR with Cellular Aspartic Proteases[a]

protein	no. of atoms superimposed	rms devn (Å)
porcine pepsin	112	1.88
chymosin	102	1.81
endothiapepsin	98	1.90
rhizopuspepsin	96	1.89
penicillopepsin	105	1.83

[a] The results shown were obtained with a program written by J. K. M. Rao on the basis of a method described by Remington and Matthews (1980). The cutoff value for distance between superimposed C_α atoms that are structurally equivalent was chosen as 3.0 Å.

Table V: Statistics for Structural Superpositions of the Active-Site Atoms of RSV PR with Cellular Aspartic Proteases[a]

protein	no. of atoms superimposed	rms devn (Å)
pepsinogen (porcine)	79	0.46
pepsin (porcine)	72	0.53
chymosin (bovine)	79	0.52
endothiapepsin	81	0.49
rhizopuspepsin		
native	80	0.45
inhibitor complex	83	0.44
penicillopepsin	85	0.48

[a] Eleven residues consisting of the Asp-Ser/Thr-Gly triplet and four residues on either side of this triplet and 11 other residues related by the pseudodyad form the basic motif for the comparison. All equivalent atoms (C_α with C_α, C_β with C_β, and so on) between the superimposed residues were used in the comparison with the same procedure as in Table IV. The cutoff value for distance between superimposed atoms that are structurally equivalent was chosen as 1.0 Å.

published structures of HIV-1 PR (Wlodawer et al., 1989; Lapatto et al. 1989) have the h′ α-helix in perfect agreement in the corresponding region in RSV PR in contrast to the previous report by Navia et al. (1989).

The Active Site. The best superimposing regions between the viral and cellular aspartic proteases, the active-site loops, have been analyzed in a greater detail. For this purpose, all common atoms in the compared proteins were used which belonged to a sequence of 11 residues centered on the active-site triplet in each loop. Comparisons with different fungal and mammalian enzymes gave essentially the same results (Table V). For example, 80 atoms in this active-site region in RSV PR were within 1 Å from their counterparts in rhizopuspepsin PR, and they superimposed at an rms deviation of 0.45 Å. Interestingly, the active sites superimpose equally well for both the native and inhibitor-bound aspartic proteases. In the case of a rhizopuspepsin-inhibitor complex the agreement is at an rms deviation of 0.44 Å for 83 atoms. This

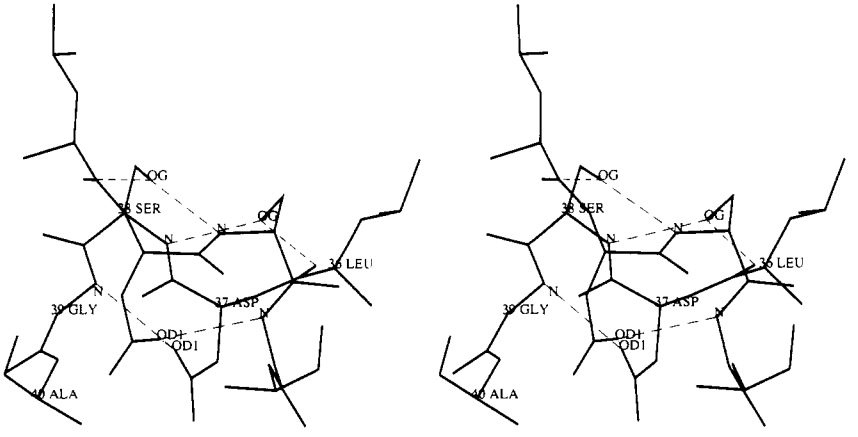

FIGURE 9: Stereo diagram of the region of the active site of RSV PR. The triad Asp37-Ser38-Gly39 and an additional residue on each side are shown, with hydrogen bonds indicted by broken lines. Solvent molecules are not shown.

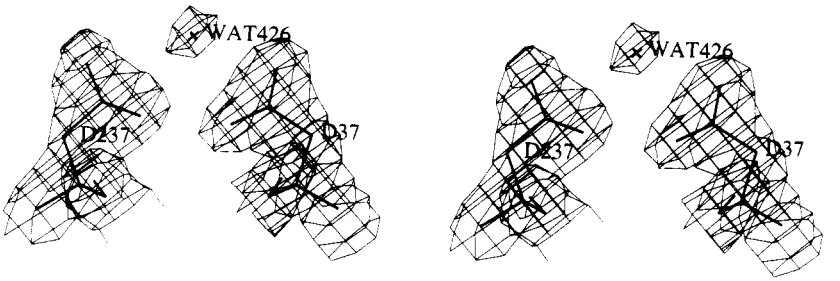

FIGURE 10: Stereo representation of the two aspartates and the water molecule found in the active site of RSV PR. The $2F_o - F_c$ map is contoured on a 1σ level.

Table VI: H-Bonding Distance (Å) within the Active Site

Asp37 OD1···Asp237 OD1	2.81	Ser38 OG···Ser238 N	2.96
		Ser238 OG···Ser38 N	3.11
Wat426 O···Asp37 OD2	2.56	Ser38 OG···Leu236 O	2.77
Wat426 O···Asp37 OD1	2.69	Ser238 OG···Leu36 O	2.52
Wat426 O···Asp237 OD1	2.58		
Wat426 O···Asp237 OD2	3.19	Asp237 OD2···Wat549 O	3.05
Asp37 OD1···Gly39 N	2.86		
Asp237 OD1···Gly239 N	2.94		

byproduct of our comparisons indicates that the active site of a cellular protease has a very rigid structure which does not change even upon inhibitor binding.

The rigidity of the catalytic-site architecture is most likely due to the intricate network of hydrogen bonds which are formed in this region and are very highly conserved in all aspartic proteases (Figure 9 and Table VI). Apart from the short hydrogen-bond interaction between the two Asp carboxylic groups described below, the main link between the two active-site loops comes from the "fireman's grip" in which each serine (or threonine in other proteases), i.e., the middle residue in each active-site triplet, accepts a bond from the amide group of the serine in the other loop and donates a hydrogen bond to the carbonyl O atom located in the residue preceding the catalytic triplet on the other strand. The structure of each individual loop is reinforced by a hydrogen bond between the inner carboxylic O atom of the aspartate and the N atom of the glycine, the last residue in the triplet. In all pepsin-like enzymes the residue immediately following the active-site triplet has a hydroxyl group (the only exception is one domain of human renin, which has alanine in this position), which usually hydrogen bonds to the outer O atom of the catalytic Asp side chain on the same strand. This is in variance with the situation in retroviral proteases, which always have an alanine in this position and where no such hydrogen bonding can occur. The water molecule Wat426 located in the center of the RSV PR active site (Figure 10) has unit occupancy, has a temperature factor of 21.8 Å^2, and is involved in hydrogen bonding with the active-site aspartates in the same manner as in uncomplexed pepsin-like proteases. The geometry of those close contacts (Table VI) suggests that it forms hydrogen bonds with both carboxylic O atoms of Asp37 and one bond with the inner O atom of Asp237. Water Wat549, also found in the active-site region, is within H-bond distance from the outer O atom of the Asp237 carboxylic group. If the previous assumption about the hydrogen bonding around Wat426 is correct, then this extra bond between Wat549 O and Asp237 OD2 would lead to a more balanced saturation of H-bonding properties at the two active-site Asp carboxylic groups. An attempt to locate an analogous water molecule bound to the outer carboxylic O atom of Asp37 was unsuccessful. Although the final $F_o - F_c$ maps had some positive density in the active-site region, there was no meaningful peak in the vicinity of Asp37 OD2, which could be interpreted as the "outer" water molecule at Asp37. This finding is somewhat surprising since in all well-refined structures of pepsin-like aspartic proteases the two active-site aspartates bind outer water molecules in a relatively symmetrical manner. The nearly perfect symmetry of the RSV PR dimer suggested that this should also be true in the present structure, contrary to the above observation.

The least-squares planes through the carboxylic groups of Asp37 and Asp237 make an angle of only 18°. The catalytic water molecule is nearly coplanar with the two aspartates and deviates by only 0.09 Å from the best plane fitted to the two carboxylic groups (Figure 10). The distance of 2.81 Å between the inner O atoms of the aspartates indicates the presence of an acidic proton in this bridge. This is consistent with the low pH at which the protein was crystallized and at which it has its peak enzymatic activity.

Discussion

Two crystallographically independent subunits of RSV PR interact to form a symmetric dimer with an active site that is very similar to the highly conserved active sites of monomeric cellular proteases, such as pepsin (Andreeva et al., 1984), penicillopepsin (James & Sielecki, 1983), endothiapepsin (Blundell et al., 1985), and rhizopuspepsin (Suguna et al., 1987a,b). The present 2.0-Å model of the RSV PR has provided us with a detailed map of the side chains and with the position of 252 water molecules in the crystal structure. In particular, the position of the water molecule between the Asp37 and Asp237 residues (one Asp from each subunit) in the catalytic site, as well as the general scheme of H-bonding within the active site, suggests that the viral and cellular enzymes share a similar enzymatic mechanism. The most detailed mechanism proposed for an aspartic protease is that for rhizopuspepsin (Suguna et al., 1987b). The key features of the catalysis are as follows: (1) a substrate binding into an extended cleft of the enzyme which has a total of at least six subsites, three subsites interacting with the peptide substrate on each side of the scissile bond; (2) an induced rotation of the scissile peptide bond and pyramidalization of the N atom toward the geometry of a tetrahedral transition-state intermediate; (3) a closing of the "flap" over the binding cleft with the addition of both hydrogen-bonding and van der Waals interactions from flap residues with the P2, P1, and P1′ peptide sites; (4) hydrogen bonding between the aspartate residues and the water molecule that will engage in the nucleophilic attack on the carbon of the scissile bond and hydrogen bonding of the Asp with a nearby Gly residue to anchor and orient the former; (5) potential hydrogen bonding between the carbonyl oxygen of the scissile bond with Asp and Ser residues that further polarize the C=O bond making it more susceptible to nucleophilic attack.

This mechanism serves as a model for the viral PR with the exception that an additional enzyme subsite (S4) is required

on the basis of biochemical data (Kotler et al., 1988) and the viral enzyme has two symmetrical flaps. There are also subtle differences in the hydrogen bonding of the catalytic Asp residues as noted in Figure 9 due to the presence of Ala rather than Ser or Thr three residues downstream. Sielecki et al. (1989) suggested that some cellular aspartic proteases, such as renin, have higher pH optima, because they contain alanine in one domain at this site. The finding that the substitution of serine for Ala40 in the RSV PR results in a mutant enzyme with a pH optimum 1 unit lower than in the wild type is consistent with this hypothesis (Leis et al., 1989).

There is evidence from in vivo studies with mutant murine retroviruses (Witte & Baltimore, 1978) that the viral protease becomes active only at or during budding of virus particles from the outer cell membrane. In addition, cytoplasmic extracts of virus-infected avian cells show little processed protein (Dickson et al., 1984). Thus, PR activation appears to be a late event in virus replication. This late activation may be a consequence of a requirement for high concentrations of polyproteins to form active dimers, for acidic pH, and/or for other factors. The optimal pH for protease activity is significantly lower than that found in the cytoplasm (Kotler et al., 1989).

Experiments utilizing recombinant fusion proteins indicate that the PR is capable of autocatalytic activity when expressed in heterologous systems (Kotler et al., 1988; Kramer et al., 1986; Madisen et al., 1987). Examination of the RSV PR structure, however, indicates that the two termini of the free enzyme dimers are distant from the active site, thus raising the question of how the processing occurs. There are at least two possibilities: (1) the processing proceeds via a trans cleavage mechanism involving attack of one precursor dimer on another, or (2) the PR in the precursor polypeptide is in a different conformation, which brings the NC–PR junction sequences into the active-site cleft. This latter model would be comparable to that for activation of pepsinogen to pepsin in which a pH-induced structural change is an important element (James & Sielecki, 1986; J. Erickson, personal communication). Structural studies of precursor polypeptides may distinguish between these two possibilities. The observed structure of the dimer interface, in which both the carboxy and amino termini of the two processed active PR molecules are intertwined, makes the intramolecular cleavage of precursors with similar structures (Navia et al., 1989) very unlikely.

There are now two viral PR structures available for comparison. The HIV-1 PR structure solved with 2.8-Å X-ray data (Wlodawer et al. 1989) and confirmed at 2.7 Å (Lapatto et al. 1989) is very similar to the RSV PR despite differences in molecular size and sequence of the two enzymes. The structures of both proteins are supported by functional studies using site-directed mutagenesis (Loeb et al., 1989a,b; Leis et al., 1989), which have shown that residues predicted to have either catalytic or structural functions in the respective active sites are very sensitive to amino acid substitutions.

The two similar subunits of the RSV PR dimer superimpose onto the crystallographic HIV-1 PR monomer with rms deviations of 1.45 and 1.55 Å for 86 and 88 common C_α positions, respectively. Both enzymes have a conserved structural core consisting of seven β-strands forming two layers and an α-helix in each monomer and of a four-stranded β-sheet in the dimer interface. The enzymes also share common subunit interaction features. These include (1) the active site residues being in an almost identical conformation with 0.5-Å rms deviation for 44 common atoms in one subunit, (2) a network of interactions linking Arg210, Asp41, and Arg111 in the RSV PR corresponding to Arg8′, Asp29, and Arg87 in the HIV-1 PR, and (3) the respective amino and carboxyl termini being intertwined as a four-stranded, antiparallel β-sheet that forms the dimer interface.

A structural alignment of the RSV and HIV enzymes results in 33 identical residues and nine additional conservative amino acid substitutions (Weber et al., 1989). Since the HIV-1 PR has only 99 residues, compared to 124 in the RSV PR, the structural alignment accommodates the size difference at three surface turns. In each case, the HIV-1 PR loops between c and b, between b′ and c′, and in the flap tips are shorter. The presence of longer surface loops in the RSV PR may serve a structural role in virion core assembly since, unlike HIV PR, the RSV PR is synthesized as part of both *gag* and *gag–pol* precursors and is thus present in the virion in concentrations equal to those of the capsid structural proteins.

Other differences can also be observed between the two enzymes. Those found in the active-site cleft may be very important determinants, accounting for the differences noted between the two PRs in turnover numbers and specificity for substrates (Darke et al., 1989; Kotler et al., 1988). An understanding of these differences, especially with respect to selection of a substrate, is very important for the rational design of inhibitors directed toward the PR. Such inhibitors may prove to be effective in blocking cellular replication of HIV, the presumed pathogen for AIDS. On the basis of the present structural and biochemical data, heptapeptides containing nonhydrolyzable peptide bonds or various amino acid substitutions would be potential inhibitors of the enzyme. Other inhibitors could represent half-substrate molecules or contain reactive groups that would covalently link with residues in the active site such as Asp37 and Asp237. The recently solved crystal structure of an HIV-1 PR inhibitor complex (Miller et al., 1989c) revealed that the degree of influence of inhibitor binding on the protease structure is much higher than that reported previously for cellular aspartic proteases (Suguna et al., 1987b; Sali et al., 1989). It would be of importance to determine if RSV PR behaves similarly to HIV-1 PR and to follow the differences in inhibitor binding of the two retroviral proteases. Our initial experiments in which RSV PR crystals were soaked in solutions of either nonspecific aspartic protease inhibitors, such as pepstatin, or HIV PR inhibitors have been unsuccessful (unpublished). Strop et al. (1989) reported the synthesis of specific ASLV inhibitors with binding constants approaching that measured for MVT-101, an HIV-1 PR inhibitor studied by us (Miller et al., 1989c). These inhibitors were recently made available to us, and cocrystallization attempts are underway.

An alternative inhibitor strategy could depend upon disrupting the PR monomer–dimer equilibrium. Since the active site is formed by the combination of two monomers, preventing dimerization should result in an inactive enzyme. Previously, Dutia et al. (1986) and Cohen et al. (1986) reported that small polypeptides representing different sequences of the herpes virus ribonucleotide reductase specifically inhibit the enzyme by disrupting its monomer–dimer equilibrium. If such a strategy is applicable to retroviral proteases as well, then the most likely target to affect the monomer–dimer equilibrium is the interface β-sheet of the two monomers.

ACKNOWLEDGMENTS

We thank Dr. John Erickson for help in molecular dynamics refinement and Dr. Jane Richardson for drawing Figure 6. We also thank Dr. Anna Skalka for discussion of the manuscript. The Advanced Scientific Computing Laboratory,

FCRF, provided a substantial allocation of time on their CRAY X-MP supercomputer. 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

Andreeva, N., Zdanov, A., Gustchina, A., & Fedorov, A. (1984) *J. Biol. Chem. 259*, 11353–11365.

Blundell, T., Sibanda, B., & Pearl, L. (1983) *Nature 304*, 273–275.

Blundell, T. L., Jenkins, J., Pearl, L., Sewell, T., & Pedersen, V. (1985) in *Aspartic Proteinases and Their Inhibitors* (Kostka, V., Ed.) pp 151–161, Walter de Gruyter, Berlin.

Brünger, A., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Cohen, E., Gardreau, P., Brazeau, P., & Langelier, Y. (1986) *Nature 321*, 441–443.

Darke, P., Leu, C.-T., Davis, L., Heimback, J., Diehl, R., Hill, W., Dixon, R., & Sigal, T. (1989) *J. Biol. Chem. 264*, 2307–2312.

Dickson, C., Eisenman, R., Fan, H., Hunter, E., & Teich, N. (1984) in *RNA Tumor Viruses* (Weiss, R., Teich, N., Varmus, H., & Coffin, J., Eds.) Vol. 1, pp 513–648, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Dutia, B., Frame, M., Subak-Sharpe, J., Clark, W., & Marsden, H. (1986) *Nature 321*, 439–441.

Finzel, B. C. (1987) *J. Appl. Crystallogr. 20*, 53–55.

Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.

James, M. N. G., & Sielecki, A. R. (1983) *J. Mol. Biol. 163*, 299–361.

James, M. N. G., & Sielecki, A. R. (1986) *Nature 319*, 33–38.

Jones, A. (1985) *Methods Enzymol. 115*, 157–171.

Katoh, I., Yoshiyuki, Y., Rein, A., Shibuya, M., Odaka, T., & Oroszlan, S. (1985) *Virology 145*, 280–292.

Khan, A., & Stephensen, J. (1979) *J. Virol. 29*, 649–656.

Kotler, M., Katz, R., Danho, W., Leis, J., & Skalka, A. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4185–4189.

Kotler, M., Danho, W., Katz, R. A., Leis, J., & Skalka, A. M. (1989) *J. Biol. Chem. 264*, 3428–3435.

Kramer, R., Schaber, M., Skalka, A., Ganguly, K., Wong-Staal, F., & Reddy, E. (1986) *Science 231*, 1580–1584.

Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Havrylik, S. J., Lee, S. E., Scheld, K. G., & Hobart, P. M. (1989) *Nature 342*, 299–302.

Leis, J., Bizub, D., Weber, I., Cameron, C., Katz, R., Wlodawer, A., & Skalka, A. (1989) in *Viral Proteinases as Targets for Chemotherapy* (Krausslich, H., Oroszlan, S., & Wimmer, E., Eds.) pp 235–243, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Loeb, D., Hutchinson, C., Edgell, M., Farmerie, W., & Swanstrom, R. (1989a) *J. Virol. 63*, 111–121.

Loeb, D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S., & Hutchinson, C. (1989b) *Nature 340*, 397–400.

Luzzati, V. (1952) *Acta Crystallogr. 5*, 802–810.

Madisen, L., Travis, B., Hu, S., & Purchio, A. (1987) *Virology 158*, 248–250.

Miller, M., Leis, J., & Wlodawer, A. (1988) *J. Mol. Biol. 204*, 211–212.

Miller, M., Jaskólski, M., Rao, M., Wlodawer, A., & Leis, J. (1989a) in *Proteases of Retroviruses* (Kostka, V., Ed.) pp 165–174, Walter de Gruyter, Berlin.

Miller, M., Jaskólski, M., Rao, M., Leis, J., & Wlodawer, A. (1989b) *Nature 337*, 576–579.

Miller, M., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Schneider, J., Kent, S. B. H., & Wlodawer, A. (1989c) *Science 246*, 1149–1152.

Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P. (1989) *Nature 337*, 615–620.

Pearl, L., & Taylor, W. (1987) *Nature 329*, 351–354.

Pepinsky, R. (1983) *J. Biol. Chem. 258*, 11229–11235.

Peters, D., & Peters, J. (1981) *J. Mol. Struct. 85*, 107–123.

Remington, S. J., & Matthews, B. W. (1980) *J. Mol. Biol. 140*, 77–99.

Sali, A., Veerapandian, B., Cooper, J. B., Foundling, S. I., Hoover, D. J., & Blundell, T. L. (1989) *EMBO J. 8*, 2179–2188.

Sheriff, S. (1987) *J. Appl. Crystallogr. 20*, 55–57.

Sielecki, A., Hayakawa, K., Fujinaga, M., Murphy, M., Fraser, M., Muir, A., Carilli, C., Lewicki, J., Baxter, J., & James, M. (1989) *Science 243*, 1346–1351.

Skalka, A. (1989) *Cell 56*, 911–913.

Smith, R., Nebes, S., & Leis, J. (1977) *Anal. Biochem. 77*, 226–234.

Steigemann, W. (1974) Ph.D. Thesis, Technische Universität, München.

Strop, P., Konvalinka, J., Pavlickova, L., Blaha, I., Soucek, M., Urban, J., Velek, J., Stys, D., Kostka, V., & Sedlacek, J. (1989) in *Viral Proteinases as Targets for Chemotherapy* (Krausslich, H., Oroszlan, S., & Wimmer, E., Eds.) pp 259–267, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D., & Davies, D. R. (1987a) *J. Mol. Biol. 196*, 877–900.

Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D., & Davies, D. R. (1987b) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7009–7013.

Tang, J. (1979) *Mol. Cell. Biochem. 26*, 93–109.

Wang, B. C. (1985) *Methods Enzymol. 115*, 90–112.

Weber, I. T., Miller, M., Jaskólski, M., Leis, J., Skalka, A. M., & Wlodawer, A. (1989) *Science 243*, 928–931.

Witte, O., & Baltimore, D. (1978) *J. Virol. 26*, 750–761.

Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I., Selk, L., Clawson, L., Schneider, J., & Kent, S. (1989) *Science 245*, 616–621.

Yoshinaka, Y., & Luftig, R. (1981) *Virology 111*, 239–250.