PROTEINS: Structure, Function, and Genetics 14:382–391 (1992)

Structural Studies of the Retroviral Proteinase From Avian Myeloblastosis Associated Virus

Douglas H. Ohlendorf,[1] Steven I. Foundling,[1] John J. Wendoloski,[1] Juraj Sedlacek,[2] Peter Strop,[2] and F. Raymond Salemme[1]

[1]Experimental Station, The Du Pont Merck Pharmaceutical Company, Wilmington, Delaware 19880-0228, and [2]Institutes of Organic Chemistry and Biochemistry, and Molecular Genetics, Czechoslovak Academy of Science, 16610 Prague 6, Czechoslovakia

ABSTRACT **The structure of the retroviral proteinase from avian myeloblastosis associated virus (MAV) has been determined and refined at 2.2 Å resolution. This structure is compared with those of homologous proteinases from Rous sarcoma virus (RSV) and human immunodeficiency type 1 virus (HIV). Through comparison with the structure of a proteinase-inhibitor complex from HIV, a model of a complex between MAV proteinase and a peptide substrate has been generated. Examination of this model suggests structural basis for the diverse specifications of viral proteinases.**

Key words: protein crystallography, retroviral aspartate proteinases, avian myeloblastosis associated virus, enzyme-substrate interactions, macromolecular modelling

INTRODUCTION

Retroviruses, picornaviruses, and most other single-strand RNA viruses produce a polycistronic mRNA whose polyprotein product must be specifically cleaved to generate the mature proteins needed by the virus for its growth.[1,2] In many viruses a viral-encoded seryl, cysteinyl, or aspartyl proteinase carries out this cleavage.[3] In retroviruses such as HIV, MAV, and RSV, the viral proteinase is an aspartyl proteinase containing 99–130 residue subunits that dimerize to form the active enzyme. The active sites of these proteinases lie at the dimer interface with one member of the required pair of aspartate residues being supplied by each subunit.

The structures of two retroviral proteinases have been determined using X-ray crystallography. These are HIV proteinase[4–6] and RSV proteinase.[7,8] We report the determination and refinement of the structure of the MAV proteinase and compare it with those previously determined. Examination of the structure of a complex of HIV proteinase with an inhibitor[9] has allowed generation of a model of a MAV proteinase-substrate complex that provides insight on the substrate specificities observed in viral proteinases.

MATERIALS AND METHODS

Protein Isolation

Protein was produced using an *E. coli* recombinant expression system in which an abbreviated segment of the gene for the p76 gag polyprotein was fused to lacZ gene under control of the tac promoter. The portion of viral genome incorporated into this construct contained sequences corresponding to the carboxyl terminal segment of the nucleocapsid protein, the viral proteinase, and the amino terminal segment of reverse transcriptase.[10] Since naturally occurring transcripts of the p76 gag polyprotein stop after the viral proteinase and there is the normal proteolytic cleavage site between the nucleocapsid protein and the proteinase, high levels of mature MAV proteinase could be produced from this construct. Dissolution of resulting inclusion bodies in urea, followed by refolding and purification using standard chromatographic methods, yielded pure MAV proteinase.[10]

Crystallization

Prior to all crystallization experiments, the purity of the proteinase was verified using SDS-PAGE and isoelectric focusing. Pure MAV proteinase was dialyzed and concentrated by ultrafiltration. Crystallization conditions were systematically investigated using an automated hanging-drop protocol.[11] Crystals suitable for diffraction analysis were obtained from droplets containing 4 μl of 0.1 mM MAV proteinase and 4 μl of reservoir solution (0.05 M citrate/phosphate pH 5.6, 8–12% saturated ammonium sulfate, 5 % v/v dimethyl formamide, 0.05% β-mer-

Received February 26, 1991; revision accepted February 25, 1992.

Address reprint requests to Dr. Douglas H. Ohlendorf, who is now at the Department of Biochemistry, University of Minnesota Medical School, 435 Delaware Street S.E., 4-225 Millard Hall, Minneapolis, MN 55455.

S.I. Foundling is now at the Oklahoma Medical Research Foundation, 825 N.E. 13th Street, Oklahoma City, OK 73104.

J.J. Wendoloski and F.R. Salemme are now at Sterling Research Group, Biophysical and Computational Chemistry, 9 Great Valley Parkway, Malvern, PA 19355.

captoethanol). Crystals grew in a week at room temperature to hexagonal rods from 0.1 to 0.4 mm in length and from 0.1 to 0.2 mm in diameter. Screening with other organic additives indicated that isopropanol and acetone altered crystal habit and yielded crystals unsuitable for further study.

Data Collection and Reduction

Diffraction data were collected with a 3-axis Siemens Area Detector system. Unoriented crystals were mounted within sealed capillaries and data were collected with the detector 16 cm from the crystal at an offset angle of 20°. Approximately 720 data frames were collected in 0.25° steps in each of 2 or 4 orientations differing by 180° or 90° in ϕ, respectively, and were processed using the XENGEN package of programs.[12] Space group symmetry was analyzed using the JUPDISP program.[13]

Refinement and Model Building

Models of the MAV proteinase were generated and evaluated using FRODO[14] and INSIGHT (Biosym Inc., San Diego, CA). Refinement of structural models was accomplished with PROLSQ[15] modified by B.C. Finzel to run on a STAR-100 array processor. Solvent parameters were refined by alternating cycles of occupancy and temperature factor refinement in which the occupancies were constrained to increments of 0.2. Refinement by simulated annealing was carried out using XREF[16]—an amalgam of AMBER[17] and PROLSQ.[15]

RESULTS AND DISCUSSION

Structure Determination

The data set used for analysis of MAV proteinase was collected over 12 days during which time no appreciable radiation damage was observed. It consisted of 2,845 frames containing 196,542 observations of 16,968 unique reflections to 2.2 Å resolution (4 reflections missing). At 2.37 Å the mean intensity was 2.01 σ above background, whereas at 2.2 Å it was 1.6 σ. The unweighted R_{sym} on intensity was 8.16% for all data to 2.2 Å.

The space group and unit cell parameters for MAV proteinase are $P3_121$ with $\mathbf{a} = \mathbf{b} = 88.92$ Å, $c = 78.79$ Å. Calculations using the molecular weight of the proteinase were consistent with a dimer in the crystallographic asymmetric unit. A contemporaneous structural study of RSV proteinase yielded an essentially identical lattice ($a = b = 88.95$ Å, $c = 78.90$ Å)[7] for a protein that was reported to differ from MAV proteinase by only 3 residues.[18,19] Because of this apparent isomorphism, a partially refined model of RSV proteinase[20] was used to produce a set of calculated structure factors. The mean fractional discrepancy of these structure factors compared with the 10-2.2 Å data from MAV proteinase was 0.36 Å. A similar comparison with data from RSV proteinase produced a value of 0.28.[20] The probable origin of this difference was nonisomorphism between the RSV and MAV proteinase crystals. The program BRUTE[21] was used to perform multidimensional searches of a model derived from RSV proteinase against MAV proteinase data. Searches using 5-4 Å data indicated a peak correlation coefficient of 0.74 for a model rotated to 0.6° and translated 0.1 Å. This small repositioning gave an R value of 0.35 for the 10-2.2 Å data with $F > 2\sigma$. refinement with PROLSQ reduced the R value to 0.32. At this point data to higher resolution were included and temperature factor refinement was begun. After several additional cycles, the R value for the 5-2.2 Å data with $F > 2\sigma$ had dropped to 0.28 and the MAV proteinase model had an RMS (ΔB) between bonded atoms of only 0.068 $Å^2$.

To eliminate bias from the probe RSV structure, the model was rebuilt using fragments from highly refined structures in the Protein Data Bank[22] for the backbone conformations[23] and commonly observed rotamers[24] for the side chain conformations. The initial R value for this model was 0.32 for the 5-3.2 Å data. Refinement using simulated annealing protocols similar to that of Brunger et al.[25] lowered the R value to 0.27. PROLSQ refinement was resumed with periodic reexaminations of electron density maps at increasing resolutions. During refinement, omit maps were used to check the placement of the amino termini. In these calculations, selected segments of the model were removed prior to several cycles of PROLSQ refinement to minimize the phase bias of the initial structure. Water molecules were added after the R value had dropped below 0.25 for a model with good geometry. Additional rounds of refinement and map examination resulted in the structure summarized in Table I.

TABLE I. Final Refinement Statistics for MAV Proteinase

Space group	$P3_121$
Cell	$\mathbf{a} = \mathbf{b} = 88.92$ Å, $c = 78.79$ Å
Resolution limits	5-2.2 Å
Reflections	13547 with $F > .1\sigma$
Crystallographic R value	0.168
Protein atoms	1711
Water molecules	133
RMS deviations from ideality	
Bond distances	0.018 Å
Bond angles	1.1°
Planes	0.015 Å
RMS chiral volume	0.206 A^3
RMS difference in thermal parameters between	
Main chain bonded atoms	0.83 A^2
Side chain bonded atoms	1.6 A^2
Hydrogen bonded atoms	3.6 A^2

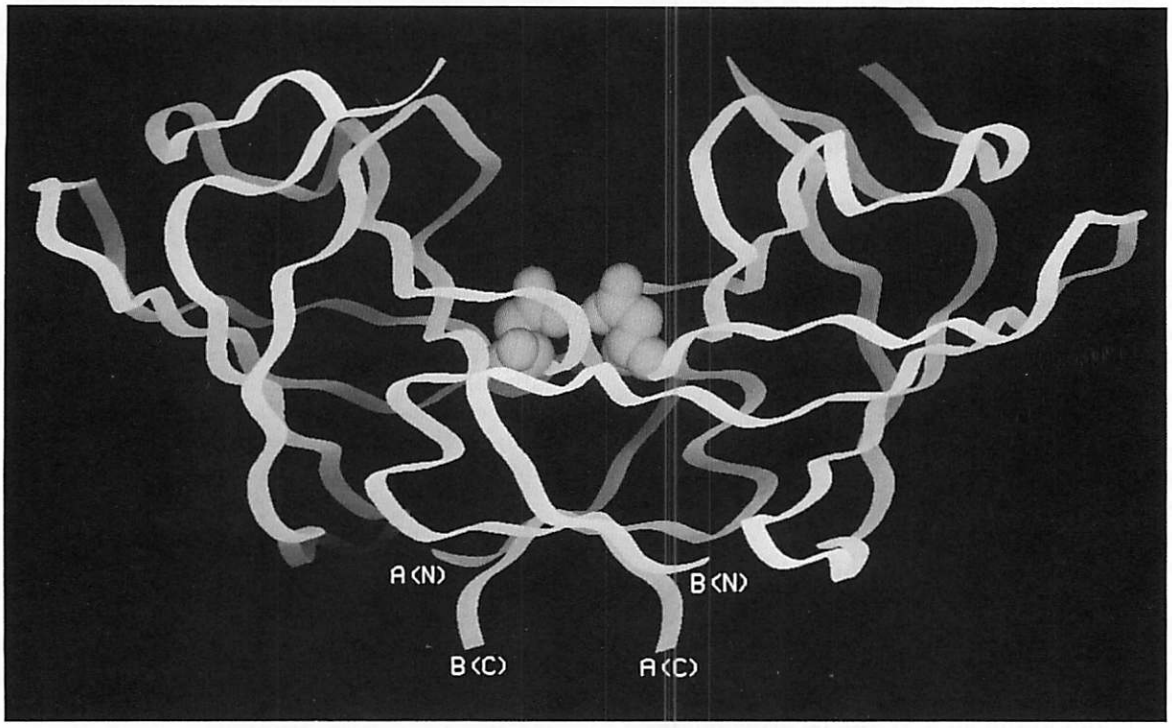

Fig. 1. Ribbon drawing of MAV proteinase backbone. The local symmetry axis relating monomers is vertical. A(N) and A(C) indicate the amino and carboxyl termini of monomer A. B(N) and B(C) indicate the amino and carboxyl termini of monomer B. The active site aspartates are shown as CPK spheres.

MAV Monomer Structure

Figure 1 displays the course of the peptide backbone for each monomer in the dimeric proteinase. Each subunit has a barrellike domain structure. The walls of the barrel are composed of two twisted β-sheets and short lengths of helices. The residues comprising the regions of regular secondary structure are indicated in Figure 2. The strands are labeled **a b c d a′ b′ c′ d′** in accord with convention.[26] There is an additional strand **q** at the end of each monomer. Helical regions are labeled **h** and **h′.**

In sheet 1 of the monomer barrel β-strands are in mixed orientations, whereas in sheet 2 they are antiparallel. A total of six strands—**b, c, d, b′, c′,** and **d′**—form the barrel. As implied by the nomenclature, these strands are related by an approximate twofold symmetry axis associating 28 pairs of C_α's with an RMS error of 1.87 Å (see Fig. 3). This degree of symmetry is comparable to that seen in other aspartyl proteinases.[26] Figure 3 shows that although short helices follow strands **d** and **d′**, the **h** helix is just a helical turn, whereas the **h′** helix consists of two full turns.

Sheet 1 forms top half of the barrel as presented in Figure 3 and includes strands **c d′ d** and **c′.** Descending strand **c** turns, makes a wide loop, and ascends to strand **d.** Descending strand **d′** crosses beneath this wide loop resulting in a structural motif that resembles the Greek letter ψ (the ψ loop). Strand **c′** then forms the outermost strand of this sheet. The topology of sheet 1 is mixed since the innermost strands are antiparallel, whereas the outermost strands strands are parallel with their neighbors. This mixed sheet is conserved in all aspartyl proteinases with known structure.[27]

Sheet 2, the bottom sheet as presented in Figure 3 is formed by strands **c′ b′ b** and **c.** β-strands in this sheet are antiparallel and orthogonal to those in sheet 1. Strands **c** and **c′** participate in both sheets because, as the longest strands, they can bridge the sheets. Also, they contain the active site aspartates and form part of the monomer-monomer interface at the base of the active site. The primary structure associated with strand **c** and this wide loop displays a conserved sequence among retroviral proteinases:

. . . **Hϕ Hϕ Asp (Ser,Thr) Gly Ala Asp** . . .

The first aspartate residue is the essential catalytic residue. The following serine or threonine residue stabilizes the base of the substrate binding cleft by interacting with the other monomer as described below. The glycine is conserved because a side chain would interfere with the active site aspartates. Similarly, the presence of a large side chain on the next residue would reduce accessibility to the active site. The final residue is an aspartic acid that forms a salt linkage with the side chain of an arginine at the amino terminus of helix **d′.** Another component of

```
      1                    10          20          30          40          50
MAV   LAMTME      HK       DRPLVRVILTNTGSHPVKQRSVYITALLDSGADITIISEEDWPADW
      --a--                   -----b----      -----c-----    -d-|-h-|
RSV   LAMTME      HK       DRPLVRVILTNTGSHPVKQRSVYITALLDSGADITIISEEDWPTDW
HIV-1 PQITLW               QRPLVTIKI   G       GQLKEALLDTGADDTVLEEMSLPGKWK
HIV-2 PQFSLW               KRPVVTAYI   E       GQPVEVLLDTGADDSIVAGIELGNNYS
      1                       10                  20          30          40

          60          70          80          90          100         110         120
MAV   PVMEAANPQIHGIGGGIPMRKSRDMIEVGVINRDGSLERPLLLFPAVA  MVRGSILGRDCLQGLGLRLTNL
      -a'-              ------b'------          ---c'---     -d'|---h'--|--q---
RSV   PVMEAANPQIHGIGGGIPMRKSRDMIELGVINRDGSLERPLLLFPAVA  MVRGSILGRDCLQGLGLRLTNL
HIV-1 PKM       IGGIGGFIKVRQY DQILIEIC      GHKAIGTVLVG   PTPVNIIGRNLLTQIGCTLNF
HIV-2 PKI       VGGIGGFINTKEY KNVEIEVL      NKKVRATIMTG   DTPINIFGRNILTALGMSLNL
                  50          60          70          80          90
```

Fig. 2. Amino acid sequence alignment of MAV, RSV, HIV-1, and HIV-2 proteinases. Numbering of residues above each line refers to MAV proteinase; that below each line to HIV proteinase. HIV proteinase sequences are aligned against those of MAV by considering structurally equivalent residues. The two lines of sequences have been aligned to be consistent with the approximate dyad symmetry within each monomer. The locations of the β strands and helixes of MAV proteinase are indicated.

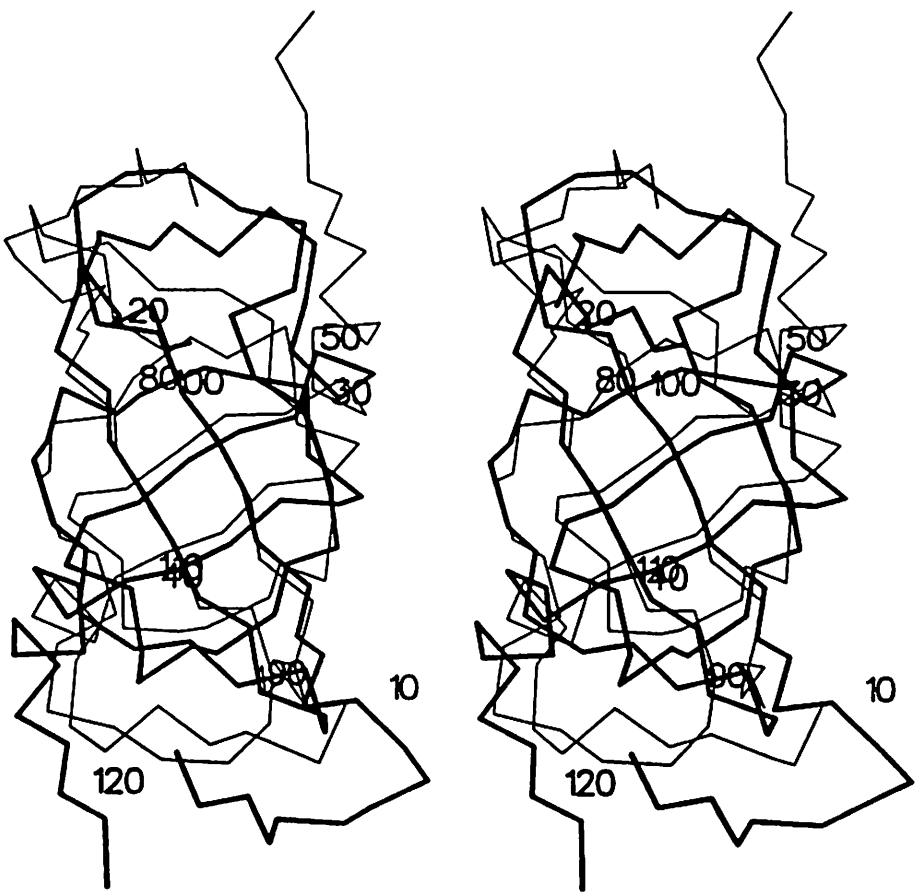

Fig. 3. C_α trace of MAV proteinase monomer (thick lines and residue numbers) rotated onto itself (thin lines). View is down twofold symmetry axis within a monomer.

this interaction is the dipole generated by helix **d'**, which places a partial positive charge directly under this aspartic acid residue. Figure 4 shows the locations of these residues.

Dimer Interactions

As expected from the unit cell volume, a dimer of MAV proteinase is present in the crystallographic

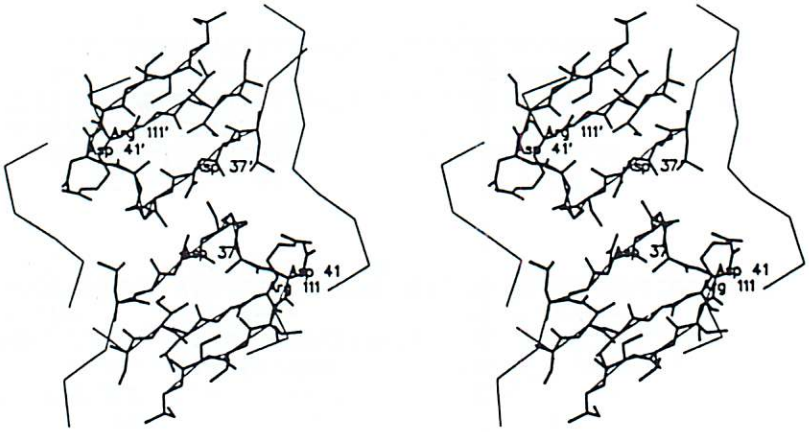

Fig. 4. Drawing of the MAV proteinase active site. Residues in the two ψ loops are shown using thick lines. C_α trace is shown using thin lines. View is down twofold symmetry axis relating monomers.

asymmetric unit. The twofold axis relating the monomers is vertical in Figure 1. In the crystal this dyad is nearly perpendicular to the crystallographic **b** axis and 53° from the crystallographic **c** axis. Taken together the barrellike domains in each monomer are paired to form a bilobal flattened ellipsoid as seen in the other aspartyl proteinases.[28–30]

The twofold axis relating the monomers runs through the center of the active site (Fig. 1). The pair of active site aspartates present in all aspartyl proteinases are supplied by two polypeptide chains related by dyad symmetry. The RMS difference between symmetry-related C_α's is 0.57 Å. During refinement the twofold symmetry was never enforced by either averaging the electron density or constraining the structures of the monomers. The origin of this level of difference between the monomers appears to be their different crystallographic environments.

Two layers of interactions hold the monomers together. The layer nearer the active site contains the hydrogen bonds between the ψ loops from each monomer. This interaction has been described as the "fireman's grip."[31] The side chain hydroxyl of the conserved residue following the active site aspartate is positioned under the corresponding aspartate from the other subunit where it forms hydrogen bonds to the preceding carbonyl oxygen and to the subsequent amide nitrogen. Under this layer is an antiparallel sheet (sheet 3) composed of two strands from each monomer. The strand organization of sheet 3 is [**a Q q A**] where capital letters denote strands from the other monomer. Initial structural reports of HIV proteinase[4–6] differed in the organization of this sheet. Studies using omit maps and maps after forced reinterpretation of sheet topology confirm the interleaving of strands from proteinase monomers.

There are two regions in which significant deviations from twofold symmetry are observed. The first region consists of residues 22–24, which form a loop between strands **b** and **c** farthest from the local dyad. In one monomer this region makes contact with a dimer in a neighboring crystallographic asymmetric unit and the electron density map is clear. In the other monomer this region is exposed to solvent and the electron density map, although discontinuous, suggests a 1.3 Å shift in the main chain. The second region of difference is around Arg 105, which is at the apex of a loop connecting strands **c'** and **d'**. Density for one arginine side chain that makes an intermolecular contact is clear, whereas that for the other residue is weaker and indicates a 2 Å movement from the rest of monomer toward the local dyad.

Symmetry relating MAV monomers extends to side chain orientations and to positions of water molecules. The RMS difference between all symmetry related protein atoms is 1.3 Å. Of 133 water molecules in the final model, 70 are related by the local twofold symmetry with an RMS error of 0.78 Å. The side chains of Cys 113 in both monomers are statically disordered in the same two conformations. Additionally, residues 58–71 in both monomers cannot be seen in the final 5-2.2 Å electron density map. These residues form flaps that fold over the substrate as seen in the HIV proteinase-inhibitor (HIVI) complex.[9] In MAV proteinase the flaps are not involved in any intermolecular interactions that could fix their positions. This is in contrast to HIV proteinase where they make hydrogen bonds to a neighboring molecule.[5]

Comparison of MAV and RSV Proteinases

As mentioned above, both MAV and RSV proteinases crystallize in the same space group with essentially identical unit cell parameters. Initially, their amino acid sequences were believed to differ at 3 locations.[18,19] Subsequent work with the strain

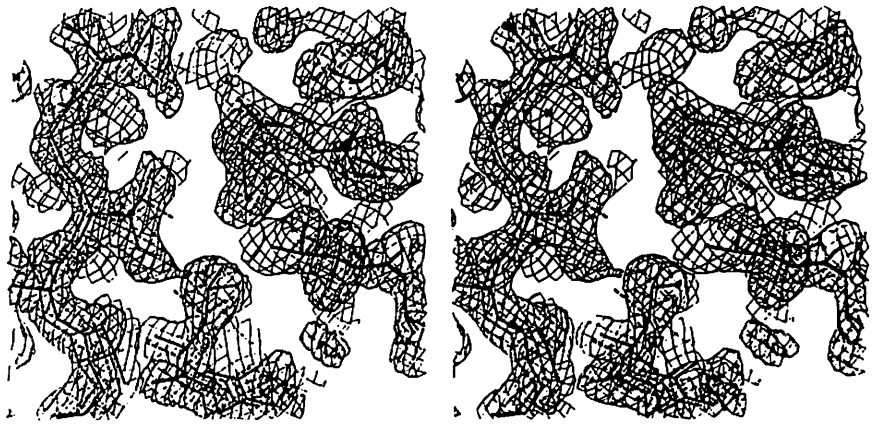

Fig. 5. $2F_o$-F_c electron density map calculated without Val 82 side chain. Thick solid lines show MAV proteinase. Thick dashed lines show RSV proteinase[8] after superimposition onto MAV proteinase.

used here has shown only residue 82 is different, with MAV proteinase having a valine and RSV proteinase having a leucine.[32] Maps calculated after omitting the side chain of Val 82 are consistent with this difference (see Fig. 5).

The RSV proteinase model containing 1,772 protein atoms and 252 water molecules has been refined to a R value of 0.144 using 10-2.0 Å data.[8] The model of MAV proteinase (as summarized in Table I) has been refined to a slightly higher R value at a lower resolution. The geometric and temperature factor constraints applied in the refinement of MAV proteinase were more stringent that those used in the refinement of RSV as evidenced by RMS bond deviations of 0.022 Å and RMS differences in thermal parameters between covalently and hydrogen bonded atoms of 2.2 $Å^2$ and 14.3 $Å^2$, respectively. In addition the model of RSV proteinase has nearly twice as many water molecules. Considering these differences both models are comparable in quality.

Superposition of the RSV and MAV proteinase models yields an RMS difference of 0.22 Å between 215 of 233 C_α's and 0.38 Å between 1610 of 1,721 protein atoms. The C_α differences are localized to flexible regions mentioned above that have altered conformations in the two monomers. The side chain differences are primarily due to degenerate orientations of such residues as aspartic acid, glutamine, and phenylalanine. Both models have Cys 113 in each monomer modeled in the same two conformations. Figure 5 displays a portion of the RSV proteinase structure superimposed on that of MAV. The figure shows that removal of a methylene carbon in residue 82 has a minimal effect on neighboring atoms.

An additional comment needs to be made regarding the better agreement observed between RSV and MAV models than between monomers of MAV. As previously mentioned the RMS agreement of C_α's of MAV monomers is 0.57 Å. Since both proteinases crystallize isomorphously, they are subject to identical interactions with neighboring molecules. These interactions result in similar distortions from dyad symmetry for both molecules producing the improved agreement.

Examination of the refined models of both proteinases verified the initial observation of a small orientational shift of 0.78° and 0.101 Å. Visual examination shows this difference does affect interactions between neighboring molecules.

Comparison of MAV and HIV Proteinases

Figure 6 shows a superposition of the HIV proteinase C_α trace[7] upon that of MAV proteinase. The structures are very similar, with the major differences arising from the 25 additional residues present in MAV proteinase (Fig. 2). The first insert consists of two basic residues in the loop connecting the **a** and **b** strands. This loop occurs at the outer edge of the active site and nearly completely buries Arg 111 of MAV proteinase as compared with Arg 87 of HIV proteinase (Fig. 7).

A 10-residue insert is found in the loop between strands **b** and **c** at the outer edge of the dimer, quite distant from the active site. The position of this loop restricts the conformations of residues 46–57 and 100–105 and compresses the substrate binding cleft as compared with HIV proteinase.[7] As shown in Figure 6, this repositioning is similar to that observed when HIV proteinase binds substrate.[9]

A 6-residue insert occurs in the flaps that cover the active site. Sequence alignments of the retroviral proteinases show the sequence **Gly Ile Gly Gly Ile** to be highly conserved.[33] The importance of these residues is evidenced by a reduction in enzymatic activity caused by their mutations.[34] As shown in the structures of HIV proteinase[7] and the HIVI complex,[9] these conserved residues form the hairpin bend and the strand of β duplex closer to the active site aspartates. The additional residues in

Fig. 6. Superimposed C_α traces from MAV proteinase (thick lines), HIV proteinase[7] (thin solid lines) and HIV proteinase-substrate complex[9] (thin dashed lines). Residue numbers refer to MAV proteinase.

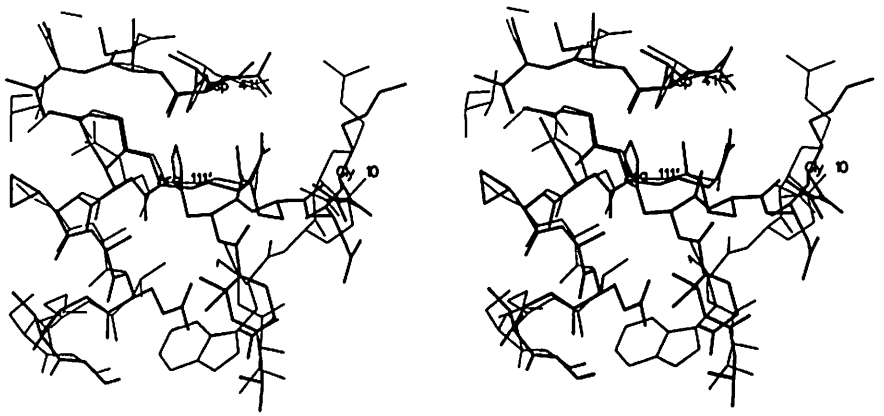

Fig. 7. Drawing of region around Arg 111′ of MAV proteinase (thick lines). The thin lines show corresponding region of HIV proteinase.[7] Residue labels refer to MAV proteinase.

MAV proteinase are thus found in the external β strand that attaches the flap to the rest of the protein.

A 7-residue insert occurs between strands **b′** and **c′** at the bottom of the dimer, away from the active site. These extra residues partially cover the terminal residues of the dimer. As mentioned above, the termini of the monomers interleave to form a β sheet that plays an important role in dimer formation. Covering of the two terminal residues can prevent fraying of this sheet and thus stabilize the dimeric form of the enzyme.

Modeling of MAV Proteinase-Substrate Complex

A characteristic of retroviral proteinases is the wide range of substrates that are specifically cleaved. To understand the basis for this promiscuity, a model of a MAV proteinase-substrate complex was made. The substrate chosen corresponds to the most studied site in vitro, the pol 63/32 site. The sequence of the substrate is:

Thr Phe Gln Ala Tyr | Pro Leu Arg Glu Ala
P5 P4 P3 P2 P1 | P1′ P2′ P3′ P4′ P5′.

Residues in the substrate are labeled P5 through P5′ with the cleavage occurring between residues P1 and P1′. The binding pockets for these residues are similarly labeled S5 through S5′. The numbering scheme is according to convention and emphasizes the symmetric nature of the proteinase.

The modeling process was guided by the structure of the HIVI complex.[9] The first step was to superimpose this structure onto MAV proteinase using 48 homologous pairs of C_α's in the ψ loops of the each dimer. The RMS agreement for these atoms was 0.53 Å. Once oriented, the backbone of the central 6 substrate residues was extracted and a planar peptide bond with a carbonyl oxygen was added between residues P1 and P1′. An automated search of a set of well-refined protein structures for fragments that best fit these α carbons selected residues 24–27 of cytochrome $P450_{CAM}$.[35] This backbone was used to model the central portion of the substrate.

The backbone for the flaps was modeled in two segments. Since residues 64–71 are well conserved in all retroviral proteinases,[33] their conformation in the HIVI complex was used. Connection of residue 71 to the rest of MAV proteinase required only a small rotation of residues 72–75. Automated searches for fragments corresponding to residues 58–63 yielded conformations that either collided with the rest of the molecule or were inconsistent with a proline at position 62. Only incorporation of a

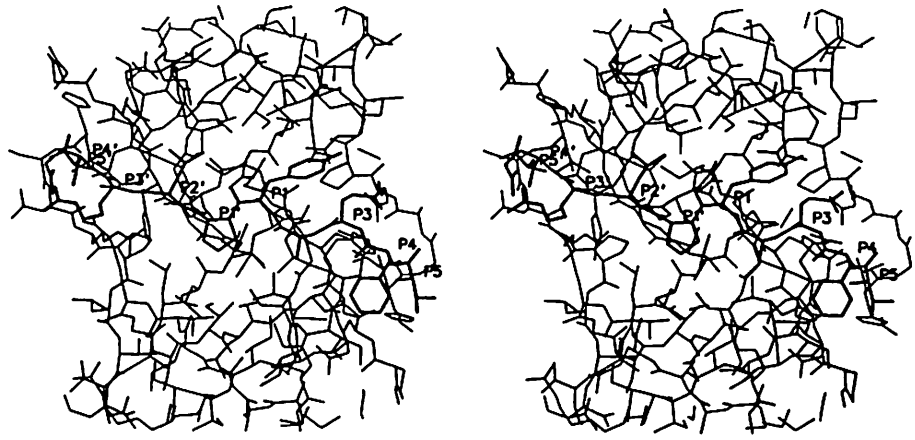

Fig. 8. Active site region of MAV proteinase-substrate model.

cis-peptide linkage to Pro 62 allowed the gap to be bridged without steric conflicts.

At this point side chains were added to the modeled backbone for the flaps and substrate. Inspection of the most common side chain rotamers was used to select allowed side chain conformations. The singular exception to this was Arg P2'. Since HIVI complex also has an arginine in this position, its side chain conformation was adopted.

The final step was to add residues P5, P4, P4', and P5'. Possible paths for the backbone at each end of the substrate were found using automated searches and inspection of the sterically acceptable conformations. The backbone of residues A46 through A48 of tryptophan synthetase[36] provided the most hydrogen bonds between the proteinase and substrate residues P5–P3. With this conformation hydrogen bonds could form between Thr P5 N and His 65 $N_{\delta 1}$, between Thr P5 O_γ and Gln 63 O, and between Phe P4 O and His 65 $N_{\delta 1}$. In addition, the hydrophobic side chain of Phe P4 could insert into a hydrophobic cleft between the flap and rest of the proteinase. For residues P3'-P5', the backbone of residues L24 through L26 of FabKol[37] allowed hydrogen bonds between Glu P4' O and His 65' $N_{\delta 1}$, between Ala P5' N and Gln 63' O and between Ala P5' O and Gln 63' N. Although this conformation brought Arg P3' near to Arg 105 and Arg 10', it was adopted since the Glu P4' side chain could form a neutralizing charge pair with Arg P3'.

On completion of the computer graphics assisted modeling, the structure was minimized using the program AMBER[17] with a 4R dielectric. A molecular dynamics simulation was then run for 125 ps, and the average structure for the last 100 ps was then accumulated and reminimized. The resulting model is presented in Figure 8.

Properties of MAV Proteinase-Substrate Complex

The substrate binding site of MAV proteinase is rather generalized in that it is essentially a long tube lined with hydrophobic side chains that envelope the central six substrate residues. These residues are in an extended conformation resulting in the side chains of residues P2, P1', and P3' being on one side of the substrate and those of residues P3, P1, and P2' on the other.[9] Using frequently observed side chain conformations for the residues lining the active site generated over 10^4 sterically allowed sets. This conformational degeneracy can readily accommodate a large number of substrate sequences.

Dividing the active site cleft longitudinally are short ridges of extended backbone provided by residues 39–41 of the ψ loops and by residues 65–67 of the flaps. In particular, it is these latter residues that form a number of hydrogen bonds with the backbone of the substrate.[9] The clearest lateral boundary is formed by the active site aspartates (Asp 37/Asp 23*), the isoleucines at the ends of the flaps (Ile 67/Ile 50) and Ile 108/Ile 84. This boundary separates the S1 pocket from S2' and, symmetrically, the S1' pocket from S2. Hydrogen bonds hold the active site aspartyl side chains relatively fixed. The isoleucinyl side chains form a more fluid boundary and serve as a hydrophobic anchor for the flaps.

Table II lists the residues of MAV and HIV proteinase that are within 5 Å of substrate side chains and define the binding pockets. Residues that form pockets S1'–S3' can be obtained by switching the primed and unprimed residues. The general impression given by Table II is the extent of homology present in the residues lining the active site. There are, however, some important differences that bear on the character of substrate-proteinase interactions.

The loop encompassing Arg 105'/Pro 81' and Gly 106'/Val 82' separates S1 and S3. This loop is known to be flexible because it is positioned differently in

*The notation Asp 37/Asp 23 refers to homologous positions in the MAV and HIV proteinases. The first residue number is for the MAV proteinase; the second is for HIV proteinase.

TABLE II. Proteinase Residues That Form Substrate Binding Pockets S1 Through S3

S1		S2		S3	
MAV	HIV	MAV	HIV	MAV	HIV
Gly 39	Gly 27	Asp 37	Asp 25	Gly 39	Gly 27
Ile 67	Ile 50	Asp 41	Asp 29	Asp 41	Asp 29
Gly 66	Gly 49	Ala 40	Ala 28	His 65	Gly 48
Leu 35′	Leu 23′	Ile 42	Asp 30	Arg 10′	Arg 8′
Asp 37′	Asp 25′	Ile 44	Val 32	Leu 35′	Leu 23′
Val 104′	Thr 80′	Ile 64	Ile 47	Arg 105′	Pro 81′
Arg 105′	Pro 81′	BB* 65	BB* 48	Gly 106′	Val 82′
Gly 106′	Val 82′	Met 73	Val 56		
Ile 108′	Ile 84′	Ile 108	Ile 84		
		Ile 67′	Ile 50′		

*Only backbone atoms involved in interaction.

the two MAV proteinase monomers and because it moves toward the local dyad upon inhibitor binding by HIV proteinase.[7,9] The change of Pro 81′ in HIV proteinase to Arg 105′ in MAV proteinase is a striking change. However, the additional charge occurs at the end of a long alkyl chain that can easily reach the solvent surrounding the molecule allowing its hydrophobic portion to help separate S1 and S3. The juncture of the S1 and S3 pockets means that residues P1 and P3 must be considered jointly. The total side-chain volume allowed for these two residues is limited and two bulky side chains would not be preferred. The sequences of naturally-occurring cleavage sites are generally consistent with this observation.[2]

Examination of the P2 and P2′ positions of the peptides cleaved by MAV and HIV proteinases shows that the latter prefers small to moderate hydrophobic side chains or neutral hydrophilic side chains while the former prefers only small hydrophobic residues.[2] As shown in Table II, HIV proteinase has Asp 30 in place of Ile 42 on the outer boundary of the S2 pocket. The aspartate side chain can be solvent-exposed or rotated to accept a hydrogen bond from an amide nitrogen or a hydroxyl in the side chain of the P2 residue. Since MAV proteinase has an isoleucine in this position no such hydrogen bond can be made. The preference of MAV proteinase for smaller residues may be due to changing Val 56 of HIV proteinase to Met 73. The larger methionine significantly reduces the amount of space available for the P2 side chain. This is consistent with the observation that MAV proteinase more readily cleaves a mutated substrate with alanines in both P2 and P2′ positions, whereas the sequence with leucines in both positions is not cleaved.[38] In HIV proteinase, the opposite effect has been observed where replacing an isoleucine in the P2′ position by alanine reduces the catalytic efficiency about 30-fold.[39]

The outer residues, P5, P4, P4′, and P5′, do not show strong sequence preferences although their absence reduces cleavage rates in vitro.[40] There is a small preference for side chains with some hydrophobic character in the P4 and P4′ positions[38] that is mirrored by the hydrophobic character of residues in the S4 pocket. Residues in the P5 and P5′ positions are generally neutral and hydrophilic.[38] These side chains can form hydrogen bonds with atoms in the flaps. Altogether, the results of the modeling study suggest that the MAV binding site achieves its diversity in specificity by incorporating some flexible elements that can accommodate a wide variety of substrate sequences.

Coordinates of MAV proteinase and of the modeled MAV proteinase-substrate complex have been deposited in the Brookhaven Data Bank.

ACKNOWLEDGMENT

We thank Maria Miller, Mariusz Jaskolski, and Alex Wlodawer of the National Cancer Institute, Frederick, Maryland, for use of a partially refined model of RSV proteinase prior to publication. We thank Pat Weber and Bruce Korant for many helpful discussions.

REFERENCES

1. Krausslich, H.G., Wimmer, E. Viral proteinases. Ann. Rev. Biochem. 57:701–754, 1988.
2. Hellen, C.U.T., Krausslich, H.-G., Wimmer, E. Proteolytic processing of polyproteins in the replication of RNA viruses. Biochemistry 28:9881–9890, 1989.
3. Kay, J., Dunn, B.M. Viral proteinases: weakness in strength. Biochim. Biophys. Acta 1048:1–18, 1990.
4. 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. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337:615–620, 1989.
5. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B.H. Conserved folding in retroviral proteases. Crystal structure of a synthetic HIV-1 protease. Science 245:616–621, 1989.
6. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danley, D.E., Georghegan, K.F., Hawrylik, S.J., Lee, S.E.,

Scheld, K.G., Hobart, P.M. X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. Nature 342:299–302, 1987.
7. Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J., Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature 337:576–579, 1989.
8. Jaskolski, M., Miller, M., Rao, J.K.M., Leis, J., Wlodawer, A. Structure of the aspartic protease from Rous sarcoma retrovirus refined at 2 Å solution. Biochemistry 29:5889–5898, 1990.
9. Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H., Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 246:1149–1151, 1989.
10. Sedlacek, J., Strop, P., Kapralek, F., Pecenka, V., Kostka, V., Travnicek, M. and Riman, J. Processed enzymatically active protease (P15gag) of avian retrovirus obtained in an *E. coli* system expressing a recombinant precursor (pr251av-Δgag). FEBS Lett. 237:187–190, 1988.
11. Cox, M.J., Weber, P.C. Automation of protein crystallization. J. Appl. Cryst. 20:366–373, 1987.
12. Howard, A.J., Gilliland, G.L., Finzel, B.C., Poulos, T.L., Ohlendorf, D.H., Salemme, F.R. Use of an imaging proportional counter in macromolecular crystallography. J. Appl. Cryst. 20:383–387, 1987.
13. Ohlendorf, D.H. Visualization software for x-ray diffraction data from area detectors. J. Appl. Cryst. 2:4:64–66, 1991.
14. Jones, T.A. A graphics model building and refinement system for macromolecules. J. Appl. Cryst. 11:268–272, 1978.
15. Hendrickson, W.A., Konnert, J.H. Stereochemically restrained crystallographic least-squares refinement of macromolecule structures. In: "Biomolecular Structure, Function, Conformation and Evolution," Vol. 1 (R. Srinvasan, ed.). Pergamon, Oxford, 1980:43–47.
16. Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J., Salemme, F.R. Structural origins of high affinity biotin binding to streptavidin. Science 243:85–88, 1989.
17. Weiner, P., Kollman, P.A. AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interactions. J. Comput. Chem. 2:287–303, 1981.
18. Sauer, R.T., Allen, D.W., Niall, H.D. Amino acid sequence of P15 from avian myeloblastosis virus complex. Biochemistry 20:3784–3791, 1981.
19. Schwartz, D.E., Tizard, R., Gilbert, W. Nucleotide sequence of Rous sarcoma virus. Cell 32:853–869, 1983.
20. Wlodawer, A., personal communication, 1989.
21. Fujinaga, M., Read, R.J. Experiences with a new translation-function program. J. Appl. Cryst. 20:517–521, 1987.
22. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. (1977) The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
23. Finzel, B.C., Kimatian, S., Ohlendorf, D.H., Wendoloski, J.J., Levitt, M., Salemme, F.R. Molecular modeling with substructure libraries derived from known protein structures. In: "Crystallographic and Modeling Methods in Molecular Design" (S. Ealick and C. Bugg, eds.). Springer-Verlag, New York, 1990:175–189.
24. Ponder, J.W., Richards, F.M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol. 193:775–791, 1987.
25. Brunger, A.T., Karplus, M., Petsko, G.A. Crystallographic refinement by simulated annealing: Application to crambin. Acta Cryst. A45:50–61, 1989.
26. Blundell, T.L., Sewell, B.T., McLachlan, A.D. Four-fold structural repeat in the acid proteases. Biochim. Biophys. Acta 580:24–31, 1979.
27. Blundell, T.L., Jenkins, J., Pearl, L., Sewell, T., Pederson, V. The high resolution structure of endothiapepsin. In: "Aspartic Proteinases and Their Inhibitors" (Kostka, V., ed.). Berlin: Walter de Gruyter, 1985:151–161.
28. Blundell, T.L., Jenkins, J.A., Sewell, B.T., Pearl, L.H., Cooper, J.B., Tickle, I.J., Veerapadian, B., Wood, S.P. X-ray analysis of aspartic proteinases: The three-dimensional structure at 2.1 Å resolution of endothiapepsin. J. Mol. Biol. 211:919–941, 1990.
29. Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H., Davies, D.R. Structure and refinement at 1.8 Å resolution of the aspartic proteinase from *Rhizopus chinensis*. J. Mol. Biol. 196:877–900, 1987.
30. James, M.N.G., Sielecki, A.R. Structure and refinement of penicillopepsin at 1.8 A resolution. J. Mol. Biol. 163:299–361, 1983.
31. Pearl, L.H., Blundell, T.L. The active site of aspartic proteinases. FEBS Letters 174:96–101, 1984.
32. Sedlacek, J., unpublished observation, 1989.
33. Erickson, J., Rao, J.K.M., Abad-Zapatero, C., Wlodawer, A. Structural comparisons of retroviral and eukaryotic aspartyl proteinases. In: "Viral Proteinases as Targets for Chemotherapy" (Karusslick, H.-G., Oroszlan, S., Wimmer, E., eds.). Cold Spring Harbor, New York: Cold Spring Harbor Press, 1989:191–201.
34. Loeb, D.D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S.E., Hutchinson, C.A. III. Complete mutagenesis of the HIV-1 protease. Nature 340:397–400.
35. Poulos, T.L., Finzel, B.C. and Howard, A.J. High-resolution crystal structure of cytochrome P450CAM. J. Mol. Biol. 195:687–700, 1987.
36. Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W., Davies, D.R. Three-dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium*. J. Biol. Chem. 263:17857–17871, 1988.
37. Marquart, M., Deisenhofer, J., Huber, R., Palm. W. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. J. Mol. Biol. 141: 369–391, 1980.
38. Margolin, N., Heath, W., Osborne, E., Lai, M., Viahos, C. Substitutions at the P2′ site of GAG P17–P24 affect cleavage efficiency by HIV-1 protease. Biochem. Biophys. Res. Commun. 167:554–560, 1990.
39. Strop, P., Konvalinka, J., Pavlickova, L., Blaha, I., Soucek, M., Urban, J., Velek, J., Stys, D., Kostka, V., and Sedlacek, J. Specificity and inhibition studies on MAV proteinase. In: "Viral Proteinases as Targets for Chemotherapy" (Karusslick, H.-G., Oroszlan, S., and Wimmer, E., eds.). Cold Spring Harbor, NY: Cold Spring Harbor Press, 1989: 259–267.
40. Kotler, M., Katz, R.A., Danho, W., Leis, J., Skalka, A.M. Synthetic peptides as substrates and inhibitors of a retroviral protease. Proc. Natl. Acad. Sci. USA 85:4185–4189, 1988.