Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7872–7876, November 1988
Biochemistry

Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: Structural and functional implications

(picornavirus/protein structure prediction/sequence alignment)

J. FERNANDO BAZAN*† AND ROBERT J. FLETTERICK†

*Department of Biophysics, University of California, Berkeley, CA 94720; and †Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143

Communicated by William J. Rutter, June 28, 1988 (received for review April 8, 1988)

ABSTRACT **Proteases that are encoded by animal picornaviruses and plant como- and potyviruses form a related group of cysteine-active-center enzymes that are essential for virus maturation. We show that these proteins are homologous to the family of trypsin-like serine proteases. In our model, the active-site nucleophile of the trypsin catalytic triad, Ser-195, is changed to a Cys residue in these viral proteases. The other two residues of the triad, His-57 and Asp-102, are otherwise absolutely conserved in all the viral protease sequences. Secondary structure analysis of aligned sequences suggests the location of the component strands of the twin β-barrel trypsin fold in the viral proteases. Unexpectedly, the 2a and 3c subclasses of viral cysteine proteases are, respectively, homologous to the small and large structural subclasses of trypsin-like serine proteases. This classification allows the molecular mapping of residues from viral sequences onto related tertiary structures; we precisely identify amino acids that are strong determinants of specificity for both small and large viral cysteine proteases.**

The naturally occurring proteases can be grouped into four classes according to the prominent functional groups at their active sites: Ser, Cys, Asp, and Zn (1). The Ser protease trypsin and subtilisin subclasses are distinguished by an identical spatial arrangement of catalytic His, Asp, and Ser residues in radically different β/β (trypsin) and α/β (subtilisin) protein scaffolds, an example of locally convergent evolution (1, 2). Within the trypsin subclass of molecules, divergent proteins exhibit widely varying degrees of sequence similarity; analysis of tertiary structures, however, reveals a high level of structural conservation (2, 3). This precedence of tertiary form over primary sequence is observed in other structurally characterized families of proteins such as the immunoglobulins (4) and globins (5).

Cys-active-center viral proteases have been identified in the sequenced genomes of four genera of the picornavirus family: the rhinoviruses [human rhinovirus strains 2–14 (HRV2–14)] (6, 7), the enteroviruses [human poliovirus (HPV1), echovirus strain 9 (EV9), coxsackievirus (CXV), bovine enterovirus (BEV), and hepatitis A virus (HAV)] (8–12), cardioviruses [encephalomyocarditis virus (EMCV) and Theiler's murine encephalomyelitis virus (TMEV)] (13, 14), and aphthoviruses [foot-and-mouth disease virus (FMDV)] (15). Two different classes of plant viruses also encode Cys proteases that are homologous to the picornaviral proteases (16–18): a comovirus with a bipartite genome, cowpea mosaic virus (CPMV) (19), and two potyviruses with monopartite genomes, tobacco-etch virus (TEV) (20) and tobacco-vein-mottling virus (TVMV) (21).

The animal and plant viruses discussed above have in common a positive-strand RNA genome that is translated into a single large polyprotein (two in the case of the segmented CPMV). The precursor is proteolytically processed at preferred Gln-Gly and Tyr-Gly sites to release a number of mature proteins needed for virus replication, structure, and assembly (22). The 3c gene segment in picornaviruses encodes a Cys protease (on average 185 amino acids) that is specific for the Gln-Gly cleavages (22, 23). A second, N-terminal, 2a gene segment present only in rhino- and enterovirus members of the picornavirus family [excluding HAV (12)] encodes a smaller (average 150 amino acids) Cys protease that is similar in sequence to the 3c proteases in the vicinity of the putative active site Cys, but is specific for the Tyr-Gly sites (22, 24). The three plant viruses carry gene regions (termed 24K in CPMV and NIa in TEV–TVMV) (19–22, 25, 26) that encode Cys proteases similar to the picornavirus 3c class.

The catalytic role of a C-terminal Cys in both 2a and 3c proteases has been inferred by the high degree of local primary sequence conservation (16–18, 22) and by site-specific mutagenesis experiments, which showed that replacing the Cys inactivates the enzyme (27). Inactivation by the classical Cys protease inhibitors iodoacetamide, *N*-ethylmaleimide, and *para*-chloromercuribenzoate (28) and the protein cystatin (29) confirms this assignment. These results have prompted classification of these enzymes as another Cys protease structural class, distinct from the papain family (with which there is no discernable sequence similarity) (16–18, 22). Our studies demonstrate that the tertiary fold of these viral Cys proteases is similar to the bilobal β-barrel motif of the trypsin family of Ser proteases (2), a structure that is different from that of papain (1). Conserved His, Asp, and Cys residues equivalent to the trypsin catalytic triad (2) are found at structurally similar positions in the picorna- and plant viral proteases. Other conserved residues in an alignment of viral and cellular proteases fulfill essential structural roles or contribute directly to the homologous catalytic and substrate-binding properties of the Ser proteases.

METHODS

The viral protease sequences were selected from the published literature and the National Biomedical Research Foundation (NBRF) data base (ref. 30, release 15.0). The cellular

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HRV2–14, human rhinovirus, strains 2–14; HPV1, human poliovirus, strain Sabin vaccine P3/Leon/37; EV9, echovirus strain 9; CXV, coxsackievirus, B3 strain Nancy; BEV, bovine enterovirus; HAV, hepatitis-A virus; EMCV, encephalomyocarditis virus; TMEV, Theiler's murine encephalomyelitis virus; FMDV, foot-and-mouth disease virus, strain O[1]K; TVMV, tobacco-vein-mottling virus; TEV, tobacco etch virus; CPMV, cowpea mosaic virus, B genome segment; SAP, *Staphylococcus aureus* (strain V8) protease); SGT, *Streptomyces griseus* trypsin; TRP, trypsin; CHT, chymotrypsin; ELA, elastase; SGPA and SGPB, *Streptomyces griseus* protease A and B; ALP, *Lysobacter enzymogenes* α-lytic protease; NBRF, National Biomedical Research Foundation.

Ser protease sequences [trypsin (TRP), chymotrypsin (CHT), and elastase (ELA)] are found in the alignments of Craik *et al.* (3). Several bacterial enzymes are considered: *Streptomyces griseus* trypsin (SGT) (31), *Staphylococcus aureus* protease (SAP) (32), *Streptomyces griseus* proteases A and B (SGPA and SGPB) (33), and *Lysobacter enzymogenes* α-lytic protease (ALP) (33). Crystallographic coordinates for representative members of the various cellular protease families are available from the Protein Data Bank (34). Tertiary structure modeling of the 2a–3c viral protease folds based on the homologous known structures of small-large trypsin-like Ser proteases was carried out on an Evans and Sutherland PS330 graphics work station (linked to a VAX 8650 computer) with the Biosym INSIGHT software package.

The University of Wisconsin Genetic Computer Group (ref. 35; release 5.2) program PROFILE (36) was used in compiling the alignments of distantly related proteins. Sequence templates that incorporated the disparate and conserved amino acid patterns of both viral and cellular proteases were used to search the (approximately) 6800 sequences of the NBRF data base release 15.0 (30). The discriminatory capability of a template was refined until it could consistently retrieve all known viral Cys protease and/or trypsin-like Ser protease sequences. GENALIGN (37) produced statistical alignment scores of both pairwise and multiple matchings based on the method of Needleman and Wunsch (38). Structural constraints were incorporated into the alignments by allowing gaps to be located only between elements of known or predicted secondary structure. The latter predictions combined the Garnier–Robson algorithm (39) with amphipathic β-strand search techniques (40) to best locate the twelve β-strands of the trypsin-like Ser protease fold in the viral proteases and the bacterial SAP. Accurate location of turns was facilitated by the MATCH algorithm (41).

RESULTS

The viral Cys protease sequences were aligned in an effort to extend the similarities that were noted in the C-terminal third of the proteins (16–18). The picorna- and plant virus sequences were aligned in order of their closer similarity until all sequences had been incorporated into the profile (36). Two absolutely conserved residues in the N-terminal half of the viral 3c proteases, His-40 and Asp-85, accompanied the conserved C-terminal residues Cys-147 and His-161 (Fig. 1). A parallel profile of the smaller 2a proteases showed absolute conservation of His-20, Asp-38, and Cys-109 residues (Fig. 1). Merging of the 3c and 2a alignments (introducing a number of gaps in the shorter 2a sequences) showed that the three conserved 2a residues were equivalent to the conserved His-40, Asp-85, and Cys-147 of the 3c proteases.

Profile searches of the NBRF data bank (ref. 30; release 15.0) utilizing the 2a, 3c, and merged profiles directed attention to the trypsin-like Ser protease family. Notably, the 2a profile retrieved small bacterial protease sequences, while the 3c and merged profiles matched several large trypsin-like Ser proteases, the bacterial SAP (32) scoring highest. Five of the 3c proteases (HRV2, HPV1, EV9, CXV, and BEV) (6–11) and the five available 2a sequences (6–8, 10, 11) were then used to generate pairwise alignment statistics with GENALIGN (37) in matches between themselves, a test group of trypsin-like Ser proteases (SAP, SGT, TRP, CHT, and ELA) (3, 31, 32) and respective randomized sequences. The 3c/2a matches produced comparable identity scores (13 on average) to the large/small Ser protease pairings (average of 14) (33). Viral-2a/small Ser protease alignments scored an average of 13 identities, while the typical viral-3c/large Ser protease alignments were segregated into matches with SAP (scoring ≈15 identities) and all others (4 identities). This low overall identity between viral and cellular proteases must be

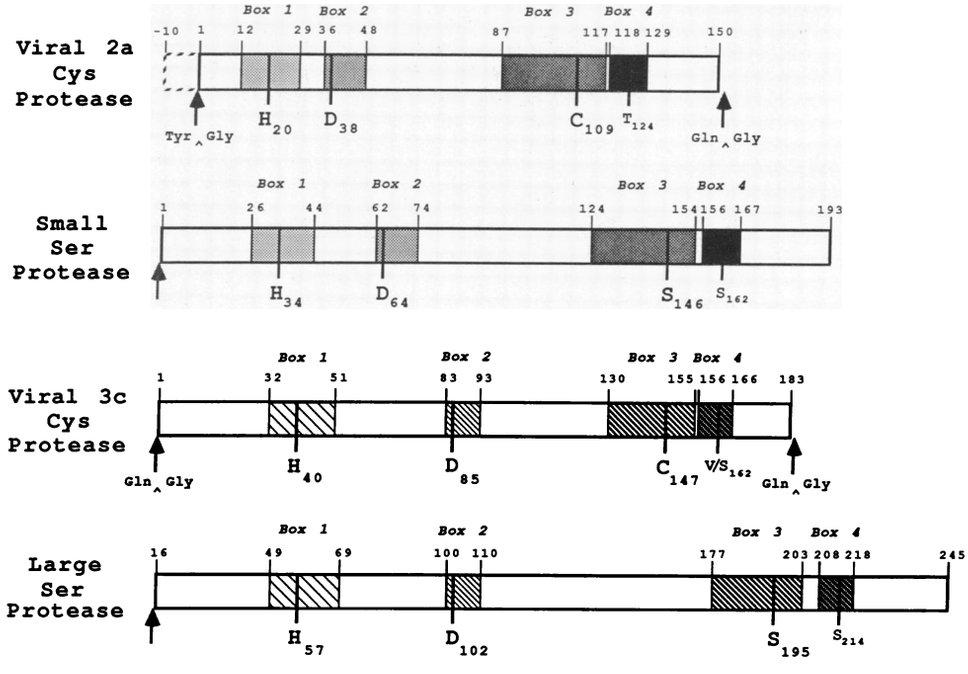

FIG. 1. Schematic comparison of the polypeptide chains of typical viral Cys and cellular Ser proteases that are aligned in Fig. 2. Boxes 1–4 correspond to sequence-similar regions of the chains that contribute to formation of the catalytic site or substrate binding pocket in determined tertiary structures. The different box-shading schemes distinguish sequence patterns of the two known structural subclasses of cellular trypsin-like Ser proteases (labeled small and large) homologous to the viral 2a and 3c Cys protease subclasses. Proteolytic processing sites are identified at the N and C termini of the viral proteases by vertical arrows. Cellular proteases have analogous N-terminal sites, as the enzymes are processed from longer, inactive precursors. The broken-outline box (numbered −10 to 1) at the N terminus of the 2a protease represents sequences N-terminal to the presumed Tyr-Gly cleavage site that display significant similarity to mature protein sequences in 3c and cellular proteases. The numbering of the aligned viral 2a–3c and small Ser protease sequences is absolute; the large Ser protease sequence is numbered according to the chymotrypsinogen scheme (3). H_{20}, His-20; D_{38}, Asp-38, etc.

qualified by the realization that the few conserved residues are known to play essential structural roles in catalysis and substrate binding for trypsin-like Ser proteases. In particular, the three absolutely conserved residues in the viral 2a/3c merged alignment (His-20/40, Asp-38/85, Cys-109/147; see Fig. 1) superimposed on His-57, Asp-102, Ser-195 of the trypsin-like Ser protease catalytic triad (Fig. 1) (2, 3). No significant alignment was possible with the active site triad (Asp-32, His-64, Ser-220) of subtilisin-like Ser proteases (1).

The lead of Taylor (4) and Bashford *et al.* (5) is followed in considering sequence/structure relationships in widely divergent families of proteins. We sought to distinguish whether the low similarity that is observed in viral/cellular sequence comparisons was indeed indicative of structural similarity and evolutionary relationship or suggestive merely of analogous structure. To this effect, available structural information [from Ser protease crystal structures (2) and viral protease secondary structure predictions (39–41)] was incorporated into the viral/cellular protease alignments in an effort to refine the profiles and increase their ability to retrieve

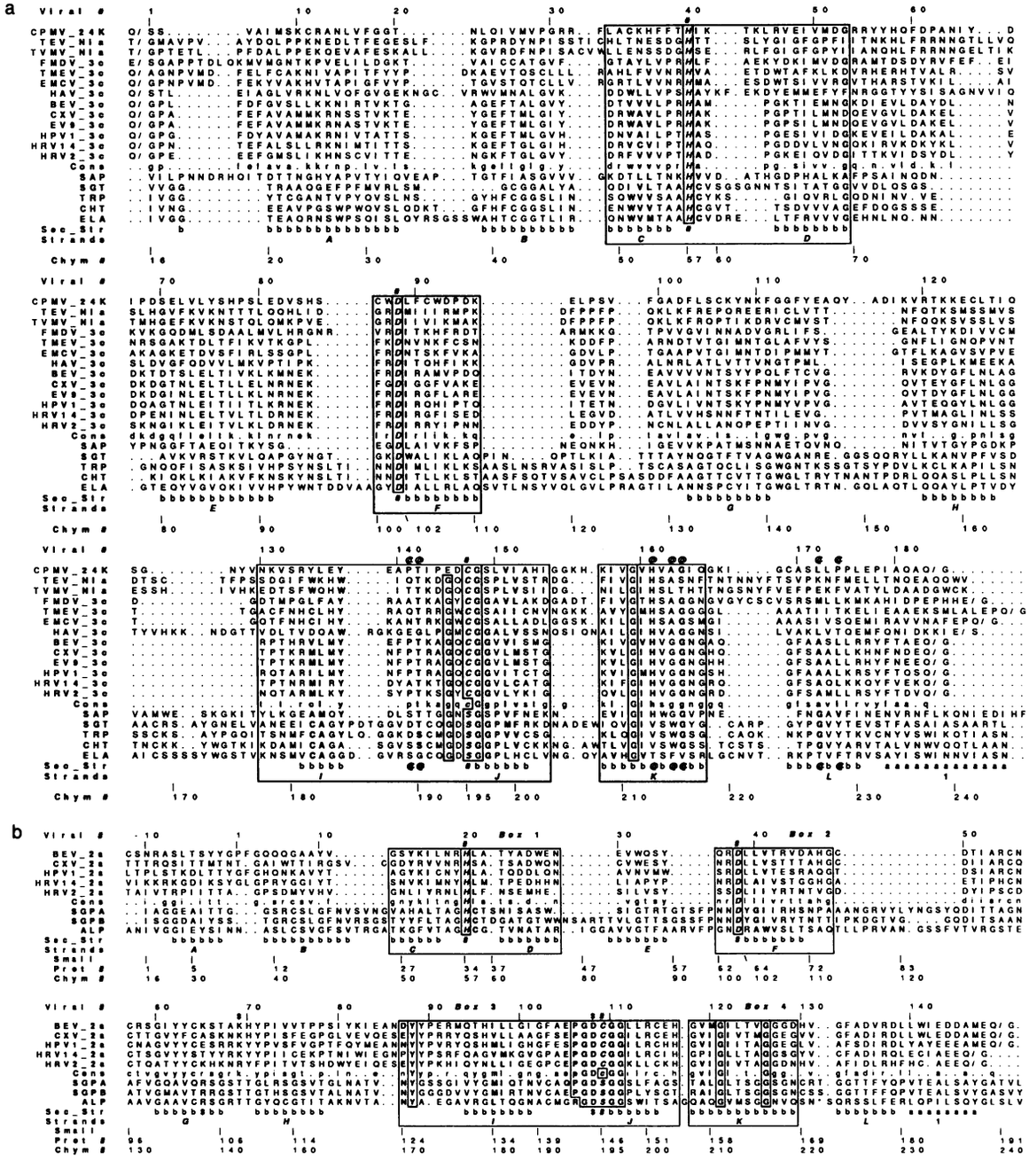

FIG. 2. Alignment of the viral Cys protease sequences with representative trypsin-like Ser protease sequences from mammals and bacteria. The numbering schemes of Fig. 1 are followed. (*a*) Viral 3c/large Ser protease alignment. (*b*) Picornavirus 2a/small bacterial Ser protease alignment. Consensus (Cons) sequences highlight identical (uppercase) and most frequently occurring (lowercase) residues at any given position. The 12 core β-strands (labeled *A–L*) and short C-terminal α-helix (labeled 1) characteristic of all trypsin-like Ser proteases are located under the profiles. Gaps in the alignments map to length- and sequence-variable surface loops. The similarity boxes (labeled 1–4) of Fig. 1 encase the "conserved" catalytic triads of His, Asp, and Ser/Cys residues (#) as well as the few other identical matches (boxed and in **bold**). Other chemically conserved residues are just in **bold** lettering. Notably, residues 189–190, 213–216, and 226 (labeled by @) are important in forming the substrate binding pocket in large Ser proteases. Residue 213 in the 2a proteases is not a conserved His as was the case in the 3c and SAP enzymes. The * in the ALP sequence marks a small insertion (3).

homologous sequences from the data bank. The chymotrypsinogen numbering scheme (3) was adopted for all cellular and viral sequence referrals (Figs. 1 and 2).

The alignment in Fig. 2*a* illustrates the evolutionary relationship between the 3c viral proteases and the large-trypsin family of proteins. While the overall degree of amino acid identity is poor, blocks of significant similarity surround the three active site residues (boxes 1–4). These regions form either the hydrophobic core of the β-barrels or the loops that define the active site and binding pocket of the enzyme. Gaps that must be introduced to align the shorter 3c proteases map to sequence and length-variable surface loops (3) connecting the 12 β-strands of trypsin. One loop stands out regarding both conservation in length and acidic amino acid composition in both viral and cellular enzymes: residues 70–79 have been shown to bind calcium in eukaryotic trypsin molecules (2). The proposed D–E loop in viral proteases may also bind Ca^{2+}.

On the basis of known three-dimensional structures, we can identify the residues that will form the specificity (S1) pocket in the 3c proteases. In particular, residues 189–190 of the large-trypsin class are located at the bottom of the S1 pocket and are a major determinant of specificity (2, 3). These positions are typically Ala/Pro-Thr in the viral 3c proteases, similar to the Ser-Thr pairing of SAP (32). Residues 216 and 226 (which lie on opposite sides of the pocket) are well conserved in character between the viral and cellular enzymes. The striking differences between the viral 3c and trypsin-like enzymes (besides the active site nucleophile) map to residues 213 and 215, respectively positioned at the side and top of the specificity pocket (2). Residue 213 is typically a small hydrophobic amino acid and residue 215 a large aromatic amino acid in the sequenced trypsin homologs (3). The 3c proteases have instead a conserved His-213 and Ala/Gly at position 215. Molecular modeling of this pair of residues in the pocket of a trypsin–inhibitor complex structure (2) reveals possible hydrogen-bonding interactions between viral His-213, Thr-190, and the S1-bound Gln side chain (not shown). Accurate positioning of the His-213 ring may be made possible by the Tyr-228 → Leu viral mutation. We postulate that residues in this collection are the primary determinants of the Gln-Gly cleavage specificity of the 3c proteases. These amino acids are different in all known trypsin-like enzymes (3) with the exception of SAP (32), which has a homologous Thr-190/His-213/Gly-215/Val-228 complement of residues. This bacterial Ser protease recognizes a Glu in the S1 pocket (32).

The 2a picornaviral proteases are aligned with three small trypsin-like Ser proteases of bacterial origin in Fig. 2*b*. Tertiary structures have been determined by x-ray crystallography for SGPA-B and ALP (33). These bacterial enzymes show clear core homology with the large trypsin-like class, differing significantly only in the economy and surface packing of the loops connecting the β-strands (42). The poor sequence identity (typically ≈15%) compounded with the differences in chain length made alignments between bacterial and pancreatic enzymes difficult prior to structural determinations (33, 42). Fig. 2*b* shows that there is a significant degree of sequence identity in the active site residue boxes between bacterial and viral 2a proteases (10 identities in the alignment of 8 sequences). Identical residues map to the interface between β-barrel domains in a small bacterial structure and are catalytically important (such as His-57 and Asp-102 accompanying Cys-195), structurally vital glycines (such as the tightly buried Gly-211, or Gly-216 with stressed dihedral angles), or particularly functional, as in the case of the buried Tyr-171 that hydrogen bonds to residue 214 (a bacterial Ser, viral Thr, that interacts with Asp-102) (2). These residues are highlighted on the α-carbon frame of the SGPA crystal structure (43) in Fig. 3. Gaps in the 2a protease alignment can be accommodated by shortening loops in the three bacterial structures. We can identify specificity-determining residues in the 2a proteases by comparison to the homologous SGPA, SGPB, and ALP structures. The substrate-binding pocket in these small bacterial proteases is not well developed and can be better described as a groove on the surface of the enzyme (42). SGPA and SGPB are characterized as having a chymotryptic specificity for aromatic residues (42). The 2a proteases are similar to SGPA and B in their specificity for Tyr at the S1 site (22).

DISCUSSION

Inspection of the blocks of sequence similarity in Figs. 1 and 2 suggests that the 2a and 3c viral proteases are structurally similar, from which we infer that they diverged from a common ancestor. Their separate alignment with small bacterial and large trypsin-like proteases, respectively, serves to emphasize the parallel evolutionary relationship with the two homologous classes of trypsin-like molecules. A correct alignment (as regards structure) between the 2a and 3c classes will be best attained by structural superposition of spatially equivalent residues that will require large gaps in the smaller 2a sequences. In lieu of the actual crystal structures, this structural alignment is made possible by the separate grouping of the 3c and 2a classes of viral proteases with the large and small trypsin-like molecules (Figs. 1 and 2). We observe that the 2a/3c viral proteases are both 35–45 amino acids shorter than their corresponding small/large cellular homologs (Fig. 1). By conventional reasoning (42), the viral molecules could be judged to antedate the cellular molecules in evolutionary history. However, the greater economy of

FIG. 3. Illustration of important residues in an alignment of small viral/cellular proteases mapped onto the α-carbon framework of SGPA (43). Residues of the catalytic triad (residues 57, 102, and 195), and conserved amino acids at positions 211, 214, 216, and 171 are far apart in sequence but contiguous in space. Residue 142 is a conserved Lys/Arg in the 2a proteases, and is located in a position to interact with the conserved Asp-102 in a manner analogous to the bacterial ion pair of Arg-138·Asp-102 (42). Loop residue 66 in SGPA is the site of a seven-amino acid insertion in SGPB. The corresponding D–E loop in viral 2a proteases is similar to that in SGPA. The E–F loop is abbreviated in the viral proteases in a manner not seen in any bacterial protein and requires the joining of marked residues 91–94. Residues 122–144 denote an abbreviated viral F–G loop (the hinge between β-barrel domains), hypervariable in sequence and length in known Ser enzymes. Both of these loops are spatially distant from the conserved active site in the viral 2a structural model.

viral proteases may just be a reflection of the tightly budgeted viral genome.

Mechanistically the His-57/Asp-102/Cys-195 active site geometry of the viral proteases may function homologously to that of trypsin and analogously to that of papain. The active site geometry of the catalytic His-159/Asn-175/Cys-25 triad of papain is surprisingly similar to that of the trypsin/subtilisin groups (44) but displays no other structural similarity in the rest of the protein (1, 44). We emphasize that the trypsin, subtilisin, and papain families of enzymes have independently evolved a similar active-site constellation of functional residues but share no common ancestor (1, 2). The viral 3c protease model is further inconsistent with a subtilisin or papain secondary and tertiary structure.

Experimental simulations of the papain triad in a trypsin-like Ser protease have used direct chemical modification (45) or mutagenic (46) means to change the active site Ser to a Cys. Higaki *et al.* (46) expressed a mutant trypsin with a Ser-195 → Cys change, producing an enzyme less active than native trypsin by a factor of $\approx 10^5$. This mutant theoretically mimics the 3c viral Cys protease active site. However, the Cys-195 trypsin mutant may be a poorer enzyme because the electronic charge distribution of the native active site is evolutionarily designed for Ser-195 as nucleophile. The observed viral modification of the nucleophilic Ser → Cys may be accompanied by other changes that contribute to a new active site environment.

This study proposes a compelling model for the tertiary fold and active-site geometry of the viral Cys proteases that clearly differentiates them from the papain-like class of cellular Cys proteases. Previous analyses (16–19, 22, 24) have focused on two other conserved viral residues (besides the putative active-site Cys) as important to catalysis. We have argued that His-213 plays an important specificity-determining role in the 3c viral proteases and bacterial SAP. His-203 (conserved in the 2a class) is predicted to be in a surface loop connecting strands J and K (Fig. 3). Others (47) have postulated that (β/β-type trypsin) Ser enzymes and ($\alpha+\beta$-type papain) Cys enzymes are ancestrally related because of a limited sequence similarity in the vicinity of both Cys-25 (papain) and Ser-195 (trypsin) to some members of the 3c viral class of proteases. A comparison of the determined crystal structures of papain and trypsin enzymes (1, 44) does not support this proposal.

Asp active-center proteases from retroviruses have recently been shown to be evolutionarily related to the cellular Asp protease family (48). Molecular modeling of the human immunodeficiency virus protease by using the x-ray coordinates of bacterial pepsin structures has resulted in a useful structure for the design of therapeutic inhibitors (48). Analogously, the twin "small" and "large" trypsin-based structural models proposed for the 2a and 3c classes of proteases are sufficiently detailed to allow testing prior to actual structural determination. We expect that analysis of these viral "Cys trypsins" will provide valuable information on the evolution of catalytic function and protein structure in the trypsin superfamily.

We thank Drs. C. Craik, T. Jukes, J. Higaki, and C.-B. Stewart for stimulating discussions and critical readings of this manuscript. J.F.B. acknowledges the lead of L. Evnin in setting in motion experiments to test the viral Cys protease structural model. The support of Tina S. Bazan (to J.F.B.) is gratefully recognized.

1. Neurath, H. (1984) *Science* **224,** 350–357.
2. Kraut, J. (1977) *Annu. Rev. Biochem.* **46,** 331–358.
3. Craik, C. S., Rutter, W. J. & Fletterick, R. J. (1983) *Science* **220,** 1125–1129.
4. Taylor, W. R. (1986) *J. Mol. Biol.* **188,** 233–258.
5. Bashford, D., Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196,** 199–216.
6. Skern, T., Sommergruber, W., Blaas, D., Gruendler, P., Fraundorfer, F., Pieler, C., Fogy, I. & Kuechler, E. (1985) *Nucleic Acids Res.* **13,** 2111–2126.
7. Stanway, G., Hughes, P. J., Mountford, R. C., Minor, P. D. & Almond, J. W. (1984) *Nucleic Acids Res.* **12,** 7859–7875.
8. Stanway, G., Hughes, P. J., Mountford, R. C., Reeve, P., Minor, P. D., Schild, G. C. & Almond, J. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1539–1543.
9. Werner, G., Rosenwirth, B., Bauer, E., Seifert, J.-M., Werner, F.-J. & Besemer, J. (1986) *J. Virol.* **57,** 1084–1093.
10. Lindberg, A. M., Stalhandske, P. O. K. & Pettersson, U. (1987) *Virology* **156,** 50–63.
11. Earle, J. A. P., Skuce, R. A., Fleming, C. S., Hoey, E. M. & Martin, S. J. (1988) *J. Gen. Virol.* **69,** 253–263.
12. Najarian, R., Caput, D., Gee, W., Potter, S. J., Renard, A., Merryweather, J., Van Nest, G. & Dina, D. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2627–2631.
13. Palmenberg, A. C., Kirby, E. M., Janda, M. R., Drake, N. I., Potratz, K. F. & Collett, M. C. (1984) *Nucleic Acids Res.* **12,** 2969–2985.
14. Pevear, D. C., Calenoff, M., Rohzon, E. & Lipton, H. L. (1987) *J. Virol.* **61,** 1507–1516.
15. Forss, S., Strebel, K., Beck, E. & Schaller, H. (1984) *Nucleic Acids Res.* **12,** 6587–6601.
16. Argos, P., Kamer, P., Nicklin, M. J. H. & Wimmer, E. (1984) *Nucleic Acids Res.* **12,** 7251–7267.
17. Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G. & Zimmern, D. (1984) *EMBO J.* **3,** 855–861.
18. Domier, L. L., Shaw, J. G. & Rhoads, R. E. (1987) *Virology* **158,** 20–27.
19. Lomonossoff, G. P. & Shanks, M. (1983) *EMBO J.* **2,** 2253–2258.
20. Domier, L. L., Franklin, K. M., Shahabuddin, M., Hellmann, G. M., Overmeyer, J. H., Hiremath, S. T., Siaw, M. F. E., Lomonossoff, G. P., Shaw, J. G. & Rhoads, R. E. (1986) *Nucleic Acids Res.* **14,** 5417–5430.
21. Allison, R., Johnston, R. E. & Dougherty, W. G. (1986) *Virology* **154,** 9–20.
22. Wellink, J. & van Kammen, A. (1988) *Arch. Virol.* **98,** 1–26.
23. Hanecak, R., Semler, B. L., Ariga, H., Anderson, C. W. & Wimmer, E. (1984) *Cell* **37,** 1063–1073.
24. Toyoda, H., Nicklin, M. J. H., Murray, M. G., Anderson, C. W., Dunn, J. J., Studier, F. W. & Wimmer, E. (1986) *Cell* **45,** 761–770.
25. Garcia, J. A., Schrijvers, L., Tan, A., Vos, P., Wellink, J. & Goldbach, R. W. (1987) *Virology* **159,** 67–75.
26. Carrington, J. C. & Dougherty, W. C. (1987) *J. Virol.* **61,** 2540–2548.
27. Ivanoff, L. A., Towatari, T., Ray, J., Korant, B. D. & Petteway, S. R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5392–5396.
28. Konig, H. & Rosenwirth, B. (1988) *J. Virol.* **62,** 1243–1250.
29. Korant, B. D., Brzin, J. & Turk, V. (1985) *Biochem. Biophys. Res. Commun.* **127,** 1072–1076.
30. Sidman, K. E., George, D. G., Barker, W. C. & Hunt, L. T. (1988) *Nucleic Acids Res.* **16,** 1869–1871.
31. Read, R. J., Brayer, G. D., Jurasek, L. & James, M. N. G. (1984) *Biochemistry* **23,** 6570–6575.
32. Drapeau, G. R. (1978) *Can. J. Biochem.* **56,** 534–544.
33. Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1979) *Can. J. Biochem.* **57,** 135–144.
34. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
35. Devereaux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
36. Gribskov, M., Homyak, M., Edenfield, J. & Eisenberg, D. (1988) *Comp. Appl. Biol. Sci.* **4,** 61–66.
37. Martinez, H. (1988) *Nucleic Acids Res.* **16,** 1683–1691.
38. Needleman, S. & Wunsch, C. (1970) *J. Mol. Biol.* **48,** 443–453.
39. Garnier, J., Osguthorpe, J. D. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
40. Finer-Moore, J. & Stroud, R. M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 155–159.
41. Cohen, F. E., Abarbanel, R. M., Kuntz, I. D. & Fletterick, R. J. (1986) *Biochemistry* **25,** 266–275.
42. James, M. N. G. (1976) in *Proteolysis and Physiological Regulation*, eds. Ribbons, D. W. & Brew, K. (Academic, New York), pp. 125–142.
43. Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1979) *J. Mol. Biol.* **134,** 781–793.
44. Argos, P., Garavito, R. M., Eventoff, W., Rossmann, M. G. & Branden, C. I. (1978) *J. Mol. Biol.* **126,** 141–158.
45. Yokosawa, H., Ojima, S. & Ishii, S. (1977) *J. Biochem.* **82,** 869–876.
46. Higaki, J. N., Gibson, B. W. & Craik, C. S. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 615–621.
47. Gorbalenya, A. E., Blinov, V. M. & Donchenko, A. P. (1986) *FEBS Lett.* **194,** 253–257.
48. Pearl, L. H. & Taylor, W. R. (1987) *Nature (London)* **329,** 351–354.