Viral Proteases: Evolution of Diverse Structural Motifs to Optimize Function

2 AUTHORS, INCLUDING:

Charles S Craik
University of California, San Francisco

336 PUBLICATIONS **15,125** CITATIONS

Viral Proteases: Evolution of Diverse Structural Motifs to Optimize Function

Minireview

Lilia M. Babé* and Charles S. Craik†
*Arris Pharmaceutical Corporation
South San Francisco, California 94080
†Departments of Pharmaceutical Chemistry,
  Molecular and Cellular Pharmacology, and
  Biochemistry and Biophysics
University of California, San Francisco
San Francisco, California 94143

There is more than one way to cleave a peptide bond, and viral proteases are revealing many new strategies for accomplishing the feat. As information on the three-dimensional structures and biological functions of viral proteases is obtained, unexpected protein folds and unique control mechanisms for proteolysis are being realized. These enzymes are teaching us a great deal about the various posttranslational modifications that occur during viral replication, and also about dramatic variations on the theme of proteolysis.

Much like a classical Greek play, virus assembly is restricted by three unifying rules: time, place, and action. Each viral family may take a different approach and choose an alternate path to arrive at the final infectious product, but all are restricted by a set of overriding rules: time—there is a chronology of events during the viral life cycle; place—the proteolytic activity is localized to a specific cellular micro-environment; and action—the site of substrate cleavage is highly specific.

Not all viruses encode a protease (PR), but those that do ensure that proteolytic action occurs at the proper time and place, independent of the host for this essential function. In general, these viruses express polyproteins that are precisely cleaved by the viral protease(s) during replication and particle assembly. This mode of posttranslational modification serves several purposes: tethered proteins can travel together to the proper assembly site, enzyme activation is delayed until folding and assembly begins, and regulation of proteolysis can control the concentration of key viral proteins. Equally important is the role of the protease in mediating virus assembly/disassembly by converting uncleaved, assembled polyproteins into mature capsids capable of disassembling upon virus entry into a newly infected cell. All of these events may serve to optimize infectious virus production by orchestrating the processes of replication and viral maturation.

Proteases have been identified in a wide range of viruses, with no correlation to capsid complexity, presence of lipid envelope, or nature of their genomes (Kräusslich and Wimmer, 1988). They can be found in nonenveloped ssRNA viruses (picornaviruses), enveloped ssRNA viruses (flaviviruses, retroviruses), nonenveloped dsDNA viruses (adenoviruses), and also in enveloped dsDNA viruses (herpesviruses). Since numerous human pathogens belong to these virus families, interest in these enzymes is understandable. Proof that virally encoded proteases are essential for viral replication and infectivity has led to their identification as key targets

in the design of antiviral agents. This focus has resulted in a wealth of new information on the structure and function of these enzymes.

Several viral proteases will be described here that exemplify the diversity of structures and biological functions displayed by these enzymes (Table 1). Viral and host proteases will be compared and contrasted to highlight the recent discovery of new protein folds, novel constellations of active-site residues, and unique allosteric mechanisms. Though not comprehensive, the following are some of the more striking examples of how viral proteases achieve and regulate their function during the multistep process of viral replication and maturation.

CMV PR: A New Protein Fold and Active Site

The capsid shell of herpesviruses such as human cytomegalovirus (CMV) assembles in a fashion reminiscent of bacteriophage, where scaffold proteins assemble within the capsid core prior to viral DNA packaging and are later extruded from the mature virus. The scaffolding proteins are encoded by two overlapping 3' coterminal genes, and their proteolytic processing is performed by a protease encoded within the N terminus of the larger gene. Confirmation that herpesviruses code for serine proteases came from studies in both herpes simplex and CMV, where active site labeling reagents and mutagenesis experiments identified a conserved serine. Although extensive sequence homology exists among proteases from all herpesviruses, none can be discerned when compared to known serine proteases. The recent elucidation of the CMV PR structure revealed a previously unseen fold and indicated the use of a novel active site (Qiu et al., 1996). The CMV PR monomer consists of a single β-barrel structure instead of the usual two β-barrel fold of proteases in the chymotrypsin family (Figure 1A). The seven-stranded β barrel is surrounded by seven α helices and a signature consensus sequence flanking the active site serine is not found. The structure reveals a novel catalytic triad of His/His/Ser instead of the well-characterized Asp/His/Ser of known serine proteases. Superposition of the chymotrypsin and CMV PR structures reveals a similar positioning of the side chains of the active site residues, strongly suggesting an independent evolutionary pathway converging to yield similar enzymatic function.

Kinetic studies using oligopeptide substrates show a very poor turnover rate for the CMV PR, on the order of 10^4 lower than the digestive serine proteases (Table 1). The low catalytic efficiency observed for CMV PR and other herpesvirus proteases may reflect the altered catalytic triad. His replacement of Asp as the third member of the triad serves to position the catalytic His but does not increase the nucleophilicity of the eponymous Ser. The low turnover rate of CMV PR may be beneficial to the virus assembly process since scaffold maturation is a relatively slow process. Viral DNA encapsidation and extrusion of the scaffold proteins occurs only if proteolytic processing occurs. If processing were too rapid, the coordination of these events might not be optimal, jeopardizing the production of infectious particles. Therefore, the virus may not require a highly efficient enzyme,

Table 1. Chemical, Functional, and Structural Characteristics of Selected Viral Proteases

Protease	Catalytic Efficiency	Site of Action	Viral Substrates Cleavage Sites	Catalytic Residues	Protein Folds	Macromolecular Interactions
CMV PR	low (100–1000)	nucleus	scaffold proteins A↓S	157 63 132	single β-barrel (novel)	allosteric homodimer
AVP	?	nucleus	core proteins GX/G↓G/X	71 54 122	α-helix/β-sheet sandwich (novel)	peptide cofactor binds viral DNA
HRV 3C	medium (~100,000)	cytoplasm	nonstructural proteins Q↓G	71 40 146	double β-barrel (chymotrypsin-like)	tethered to RNA pol binds viral RNA
SCP	?	endoplasmic reticulum	self (core protein) W↓S	163 141 215	double β-barrel (chymotrypsin-like)	binds viral RNA self-inactivates
HCV NS3	medium low (~10,000)	membranes	nonstructural proteins C/T↓S	81 57 139	double β-barrel (chymotrypsin-like)	peptide cofactor tethered to helicase
HIV PR	medium (~100,000)	plasma membrane budding particles	capsid proteins and enzymes F/Y↓P	25 25'	acid protease β-barrel	obligate homodimer

Catalytic efficiencies are expressed as averages of k_{cat}/K_m values ($M^{-1}s^{-1}$) obtained using peptide substrates and are compared to an efficient digestive enzyme (k_{cat}/K_m of $10^7 M^{-1}s^{-1}$). The preferred cleavage site on natural substrates is shown with arrows representing the scissile bond. The amino acid number for each catalytic residue is shown above the depicted side chain.

opting instead for slower turnover and emphasizing affinity for the proper substrate sequence. The enzymatically active CMV PR appears to be a dimer, as shown by preliminary biochemical and structural (Figure 1A) analyses, suggesting yet another use of a structural motif to regulate the timing of proteolysis.

AVP: A Cysteine Protease Active Site in a Novel Protein Fold

In the case of the adenovirus protease (AVP), a cysteine protease catalytic triad is found in a completely novel protein fold, which also incorporates a peptide cofactor

(Ding et al., 1996). The active site of AVP contains a Glu/His/Cys catalytic triad in a similar arrangement to cysteine proteases like papain (Asn/His/Cys). The architecture of seven α helices and a single five-stranded β sheet in the central core forming a "β-sheet sandwich" is unique (Figure 1B). Only the central core resembles some of the motifs seen in the active sites of serine or cysteine proteases, and these similarities suggest a notable convergence in favor of enzyme function.

The major role of AVP is to process the core proteins within an assembled virion in the nucleus of an infected
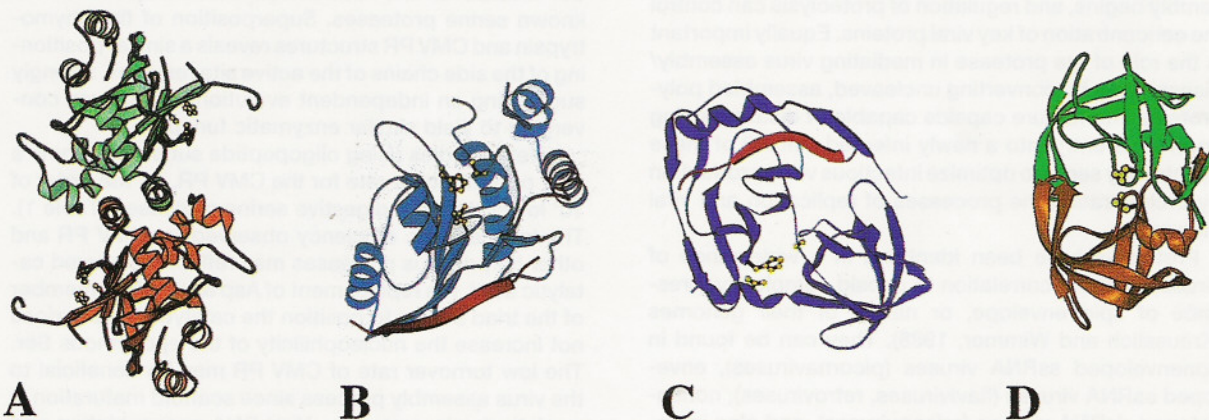

Figure 1. Structures of Selected Viral Proteases

Active site residues are shown in yellow, and unique macromolecular features are distinguished by different colors. (A) CMV PR. Monomers differ in color; coordinates courtesy of W. Stallings. (B) AVP. Peptide cofactor (pVIc) is shown in red; coordinates courtesy of W. Mangel. (C) HCV NS3 PR. A portion of the 54 amino acid cofactor is colored red; coordinates courtesy of K. Wilson and J. Thomson. HRV3C PR and SCP also adopt this fold. (D) HIV PR. Monomers differ in color. Figure created at UCSF by J. Harris and C. S. Craik.

cell. The apoenzyme is relatively inactive, unlikely to prematurely cleave virion precursors in the cytoplasm. An eleven amino acid peptide, identified biochemically and structurally as a cofactor, is a product of the protease activity and can raise k_{cat} by at least 300-fold. In the structure, this pVIc peptide appears as a sixth β strand adjacent to the core β sheet, participating in one disulfide and numerous hydrogen bonds. It does not directly contact any catalytic residues, but it could alter the active site conformation based on its positioning. Incubation of the enzyme–cofactor complex with viral DNA boosts enzyme activity by at least another 20-fold, supporting the idea that internal virion environment serves to optimize proteolytic activity. The use of a peptide cofactor and DNA to enhance activity can be viewed as regulatory steps, such that protease activity is not augmented until required inside the virion.

HRV 3C: A Serine Protease Fold with a Cys Nucleophile

The replication of picornaviruses requires processing of a single polyprotein encoded by the RNA genome. The proteolytic steps fall into three categories: cotranslational processing of capsid precursor by the 2A PR, protease-independent cleavage during final core maturation, and liberation of nonstructural proteins by 3C PR activity. Mutagenesis and biochemical analyses indicate that both the 2A and the 3C proteases have essential cysteine residues. Based on sequence alignments and secondary structure predictions, it was proposed that 3C PR folds similarly to the chymotrypsin family serine proteases, even though it shares less than 10% sequence identity with members of that clan (Bazán and Fletterick, 1988). The human rhinovirus (HRV) 3C structure (Matthews et al., 1994) confirms that prediction and shows that a catalytic triad of Glu/His/Cys is positioned similarly to the Asp/His/Ser of serine proteases. This is a Cys-active-site protease that folds into two equivalent β barrels, much like chymotrypsin, but with differences in some connecting loops, in the precise orientation of catalytic residues, and in regions required for transition state stabilization. The enzyme did not evolve from a simple replacement of a Ser by a Cys nucleophile, but as a concerted structure that supports the coordination of a novel catalytic mechanism (mutagenesis of the active site Cys for a Ser yields an enzyme with a severely reduced activity). The 3C proteases are not highly active enzymes and they may utilize a weaker Cys/His catalytic dyad, although caspases are efficient catalysts that use a Cys/His dyad. The third member of the triad (Glu or Asp) is not strictly conserved among all 3C proteases, and it is pointing away from the active site His, and thus is unlikely to assist in catalysis. However, sequestration of picornavirus replication and assembly to smooth membranes may compensate for reduced 3C activity by increasing the enzyme:substrate ratio.

The 3C PR has been shown to directly bind viral RNA and to be important for the initiation of RNA replication. RNA binding occurs at a region distal from the active site, and generation of this binding domain imposes structural restrictions on this enzyme not found in other proteases. Interestingly, 3C is found fused to the 3D RNA polymerase for a large portion of the viral life cycle. Both the autoprocessed 3C and the fused 3CD have

proteolytic activity, while 3D must be fully processed before it is active. The proteolytic release of an active polymerase may thereby influence virus replication. This multifunctional activity within a single protein structure is not observed in eukaryotic proteases and exemplifies the efficient use of a small viral genome.

Sindbis Core Protein: Both a Protease and a Structural Protein

Sindbis virus genomic RNA codes for two polyproteins that must be posttranslationally cleaved by viral or host proteases. The p130 polyprotein consists of the Sindbis core protein (SCP) at the N terminus, followed by three glycoproteins. As the first step in maturation, SCP performs a single autocatalytic cis cleavage to release itself from this polyprotein, and then has no further catalytic activity, leaving subsequent steps to other proteases. SCP self-inactivates by playing a structural trick: inserting its C-terminal tryptophan into the P1 substrate binding site, thus rendering itself inactive. The inactive structure then assembles into the icosahedrally symmetric core that surrounds the genomic RNA.

Sequence scanning of SCP identified the highly conserved sequence surrounding the active site Ser in the chymotrypsin family. Studies with temperature-sensitive mutants suggested a catalytic triad of Asp/His/Ser for SCP that was later confirmed by crystallographic studies (Tong et al., 1993). The structure revealed a fold very similar to chymotrypsin consisting of two β-barrel domains flanking the substrate binding site. The compact structure with short turns between the β strands could facilitate assembly of the core. Another unique feature of the structure is a very basic N-terminal segment responsible for association with the viral RNA and essential for encapsidation. SCP exemplifies a protein with multiple functions that must be incorporated into a single structure, again showing how small viruses maximize their coding capacity.

HCV NS3: A Tethered Enzyme with a Peptide Cofactor

The emergence of hepatitis C virus (HCV) as a major human pathogen has propelled the virally encoded NS3 PR into the limelight as a potential antiviral target. HCV translates its RNA genome into a single polyprotein precursor. The structural proteins at the N terminus are processed by host proteases, while nonstructural proteins at the C terminus are processed by the two virally encoded proteases. NS2-3 performs a single cis cleavage, while NS3 cleaves at four sites to release the nonstructural viral proteins. Although it has no sequence homology to any known protease, the NS3 protein was predicted to be a serine protease based on active-site labeling and mutagenesis studies. The predicted NS3 PR domain is physically linked at its C terminus to the viral RNA helicase activity.

The recently elucidated crystal structures of NS3 PR confirm the catalytic triad residues Asp/His/Ser and, like HRV 3C and SCP, it also adopts the somewhat "universal fold" of a chymotrypsin, dual β barrel (Figure 1C). One structure also confirms the presence of a peptide cofactor that intercalates within the enzyme core completing a β barrel and stabilizing the N-terminal α-helix (Kim et al., 1996). This 54 amino acid cofactor is the product of the only cis cleavage performed by NS3. Although quite

distant from the catalytic triad, NS4A may enhance activity by optimizing and rigidifying the extended substrate binding site. The complex can aid viral assembly by substantially increasing catalytic efficiency when processing of the remaining nonstructural proteins is required. Another role for the NS4A cofactor is the delivery of NS3 PR to cellular membranes. Membrane localization is a requirement for HCV capsid assembly and maturation; thus, coordination of full protease activation and membrane localization may be a regulatory step for the virus. In vitro studies mixing NS3 and NS4A suggest that complex formation is not absolutely required for protease activity, giving the membrane targeting role of NS4A added importance in vivo. The physical link between two unrelated enzymes as exemplified by the NS3 protease-helicase has no precedent in the realm of nonviral proteases. The peptide cofactor may serve to deliver these two enzymes to their requisite cellular location. Again we see the requirements for viral replication dictating new structural motifs in viral enzymes.

HIV PR: Regulation of Protease Activity by Dimerization

Retroviruses such as the human immunodeficiency virus (HIV) encode a single protease. The HIV PR is responsible for the cleavage of gag and gag-pol polyproteins to yield all capsid proteins and viral enzymes. Alignment of the viral protease region with known sequences lead to the unorthodox suggestion that it could fold as a single domain of an aspartic protease such as pepsin (Pearl and Taylor, 1987). Thus, the virus generates an active aspartic protease only when two identical 99 amino acid domains join to create a homodimer (Figure 1D). Determination of the HIV PR structure confirmed these predictions (Wlodawer et al., 1989). It has been suggested that retroviral proteases preceded the single-chain, nonviral aspartic proteases and that dimerization arose as a strategy to maximize the coding capacity. Another explanation is that dimerization serves a regulatory role. Expression of a single-chain, tethered HIV PR dimer results in premature cytoplasmic maturation of viral polyproteins and dramatically reduces production of infectious virions. This shows the need for the virus to suppress proteolytic maturation until the virions are assembling at the plasma membrane. Only then are polyproteins in sufficient proximity to allow protease dimerization, initiating the cleavage events that yield infectious virions. This supports the proposal that HIV PR dimerization controls proteolytic activity, unifying the principles of viral assembly.

Conclusions

The various viral proteases discussed here serve as examples of the diversity in the structural motifs so far discovered (Figure 1). These structures encompass the fundamental aspects of a protease: a substrate binding cleft, a reaction center, and a mechanism for transition state stabilization to permit catalysis (Perona and Craik, 1995). The many variations on the protein folds and catalytic residues displayed by these enzymes (Table 1) are not surprising when their origins are examined. Assuming that evolution will explore the limits of the allowed protein structures that are compatible with retention or acquisition of function, then viruses are particularly capable of achieving this task. Their replication

occurs on a scale much accelerated when compared to eukaryotes, especially RNA viruses that lack proofreading mechanisms and can thus introduce many mutations. Selection for tolerable mutations is built-in since only viable viruses survive, expediting the search for structures that achieve the two major constraints placed on enzymes: preservation of catalytic competence and structural stability.

Viral proteases are optimized to regulate and coordinate viral replication and assembly. Unlike digestive enzymes, they are highly selective catalysts performing limited proteolysis. The evolved protease sequence may not be the most catalytically robust enzyme, but one capable of performing proteolysis as well as other roles in viral replication. The ability of a viral protease to associate with the viral genome, become a core protein, or accommodate a cofactor may result in structural compromises leading to less efficient enzymes. But colocalization of enzyme and substrate within a virion may serve to compensate for reduced activity and may also impede random hydrolysis of host proteins. Perhaps these unique folds and structural complexities (Figure 1) may lead to novel antiviral designs that extend beyond active site–directed, small molecules, selectively inhibiting these enzymes while not affecting host functions (Babé et al., 1995). The uniqueness of viral proteases could become the ultimate "Achilles' heel" of the virus, as seen recently with effective HIV PR antiviral therapies.

Selected Reading

Babé, L.M., Rosé, J., and Craik, C.S. (1995). Proc. Natl. Acad. Sci. USA 92, 10069–10073.

Bazán, J.F., and Fletterick, R.J. (1988). Proc. Natl. Acad. Sci. USA 85, 7872–7876.

Ding, J., McGrath, W.J., Sweet, R.M., and Mangel, W.F. (1996). EMBO J. 15, 1778–1783.

Kim, J.L., Morgenstern, K.A., Lin, C., Fox, T., Dwyer, M.D., Landro, J.A., Chambers, S.P., Markland, W., Lepre, C.A., O'Malley, E.T., et al. (1996). Cell 87, 343–355.

Kräusslich, H.-G., and Wimmer, E. (1998). Annu. Rev. Biochem. 57, 701–754.

Matthews, D.A., Smith, W.W., Ferre, R.A., Condon, B., Budahazi, G., McElroy, H.E., Gribskov, C.L., and Worland, S. (1994). Cell 77, 761–771.

Pearl, L.H., and Taylor, W.R. (1987). Nature 329, 351–354.

Perona, J.P., and Craik, C.S. (1995). Prot. Sci. 4, 337–360.

Qiu, X., Culp, J.S., DiLella, A.G., Hellmig, B., Hoog, S.S., Janson, C.A., Smith, W.W., and Abdel-Meguid, S. (1996). Nature 383, 275–279.

Tong, L., Wengler, G., and Rossmann, M.G. (1993). J. Mol. Biol. 230, 228–247.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., and Kent, S.B.H. (1989). Science 245, 619–621.