Biochemistry

 Volume 28, Number 26 December 26, 1989

Perspectives in Biochemistry

Proteolytic Processing of Polyproteins in the Replication of RNA Viruses†

Christopher U. T. Hellen,* Hans-Georg Kräusslich,‡ and Eckard Wimmer

Department of Microbiology, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794

Received June 28, 1989; Revised Manuscript Received September 12, 1989

Many animal and plant viruses depend on the action of virus-encoded proteinases at various stages in their replication. These enzymes are highly substrate-selective and cleavage-specific, in that they cleave large, virus-specified polypeptides called polyproteins at defined amino acid pairs. Logically, they are potential targets for inhibition of virus-specific proteolytic processing and consequently of viral replication. Inhibitors of viral proteinases may therefore emerge as major antiviral chemotherapeutic agents. Until recently, characterization of viral proteinases has been limited largely to genetic mapping and molecular genetic manipulation, but over the last 2 years several major advances have been made in the understanding of their structure and substrate specificity (Kräusslich et al., 1989b). Several viral proteinases have been purified to homogeneity, and the availability of polyprotein and peptide substrates has facilitated the study of their enzymatic and physical properties. New models relating the structure of viral proteinases to those of their cellular counterparts have been developed, and the three-dimensional structures of two retroviral proteinases have been determined.

A viral polyprotein is a precursor polypeptide that contains several distinct domains and is proteolytically processed to yield diverse structural and nonstructural proteins. Synthesis of such polyproteins is limited to positive-strand RNA viruses and retroviruses (whose genomic RNAs are of the same polarity as the viral mRNAs). Eukaryotic ribosomes normally initiate translation at a single AUG codon; consequently, mRNA species are translated to yield only a single polypeptide. Synthesis and subsequent proteolytic processing of a polyprotein can therefore be regarded as one of a variety of mechanisms by which these viruses express downstream cistrons. In addition, frame-shift or read-through by suppression of leaky termination codons can modify the length of a polyprotein. Other mechanisms include translation from subgenomic RNAs and segmentation of the genome. These strategies allow both for genetic economy and for differential expression of viral proteins. Proteolytic processing enables functionally different domains to be separated and cleavage products to be transported to different cellular compartments. It may also be important in regulating events in viral replication, such as uncoating, activation of replicative enzymes, and morphogenesis (Table II).

We have reviewed the subject recently (Kräusslich & Wimmer, 1988) and in this paper shall mainly discuss the current state of knowledge regarding polyprotein processing in the two best defined viral systems of eukaryotic cells: viz., picornaviruses and their plant virus relatives and retroviruses. Other viral systems will be included as appropriate and will be related to these two main systems.

Proteolysis in Picornavirus Protein Expression

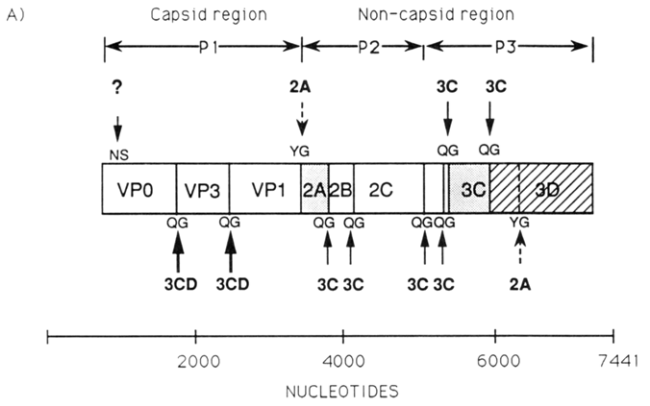
Picornaviridae are a family of small icosahedral viruses that cause a number of important disease syndromes. The family is currently divided into four genera: rhinovirus (the common cold virus), enterovirus (e.g., poliovirus), cardiovirus (e.g., encephalomyocarditis virus), and aphthovirus (foot-and-mouth disease virus). All have a positive-sense single-stranded monopartite genome which contains a single long open reading frame. This is translated to yield a large polyprotein which is subsequently cleaved to yield all structural and nonstructural proteins. Poliovirus was the first eukaryotic RNA virus for which the complete nucleotide seqence was determined, and its polyprotein was the first for which a detailed structure was obtained (Kitamura et al., 1981) [Figure 1; nomenclature according to Rueckert and Wimmer (1984)]. Processing of the polyprotein has subsequently been characterized in detail, and the mechanisms that have been elucidated are typical of many other viruses.

† Work in the authors' laboratories was supported in part by U.S. Public Health Service Grants AI 15122 and CA 28146 to E.W. from the National Institutes of Health and by a grant from Boehringer Ingelheim Pharmaceuticals, Inc.

* Corresponding author.

‡ Present address: Deutsches Krebsforschungszentrum, Institut für Virusforschung, Im Neuenheimer Feld 280, D-69 Heidelberg 1, West Germany.

0006-2960/89/0428-9881$01.50/0



	Site	*P6*	*P5*	*P4*	*P3*	*P2*	*P1*	*P1'*	*P2'*	*P3'*	*P4'*	*P5'*
B)	VPO/VP3	T	L	P	R	L	Q	G	L	P	V	M
	VP3/VP1	Q	K	A	L	A	Q	G	L	G	Q	M
	2A/2B	E	E	A	M	E	Q	G	I	T	N	Y
	2B/2C	P	Y	V	I	K	Q	G	D	S	W	L
	2C/3A	M	E	A	L	F	Q	G	P	L	Q	Y
	3A/3B	L	F	A	G	H	Q	G	A	Y	T	G
	3B/3C	R	T	A	K	V	Q	G	P	G	F	D
	3C/3D	Y	F	T	Q	S	Q	G	E	I	Q	W
C)	VP1/2A	K	D	L	T	T	Y	G	F	G	H	Q
	3C'/3D'	K	L	L	D	T	Y	G	I	N	L	P

FIGURE 1: Gene organization, processing scheme, and cleavage sites of the poliovirus polyprotein. (A) Proteolytic cleavages of the polyproteins occur between amino acid pairs indicated by standard single-letter code. Arrows above and below the polyprotein indicate sites that are cleaved in inter- and intramolecular reactions, respectively, by proteinases as indicated. The question mark indicates that the mechanism of cleavage at this site is not known. The positions of virus-encoded proteinases within the polyprotein are indicated by shaded boxes. The nomenclature of poliovirus proteins is according to Rueckert and Wimmer (1984). The amino acid residues at sites cleaved (B) by $3C^{pro}$ and $3CD^{pro}$ and (C) by $2A^{pro}$ are indicated by standard single-letter code and are described according to the nomenclature of Berger and Schechter (1970). The newly generated carboxy terminus, after cleavage of the peptide bond, is designated P1, preceded by the P2 residue, etc. The newly generated amino terminus is designated P1′, followed by the P2′ residue, etc.

Picornaviral capsids are composed of essentially equimolar amounts of four nonidentical polypeptides. They are synthesized as a precursor P1 whose N-terminus is myristoylated (Chow et al., 1987). In poliovirus (Toyoda et al., 1986), P1 is separated from the nascent polypeptide in a reaction catalyzed by the adjacent polypeptide $2A^{pro}$, a proteinase that hydrolyzes a Y–G bond at its own amino terminus and cleaves a second Y–G dipeptide, within the RNA polymerase 3D. The initial cleavage is probably an intramolecular event and must occur before the capsid precursor can be processed to yield capsomer proteins (Nicklin et al., 1987). The second cleavage is not essential for viral proliferation (Lee & Wimmer, 1988), so that the main function of $2A^{pro}$ is to sever the structural proteins of the viral nucleocapsid from the nonstructural proteins, allowing respectively assembly of particles and generation of a functional replication complex. A likely second function of $2A^{pro}$ is the rapid "shut off" of host cell protein synthesis which occurs on infection with poliovirus. This is associated with proteolytic cleavage of a component of the cap-binding complex, which is induced but not directly catalyzed by $2A^{pro}$ in vitro (Kräusslich et al., 1987) and in vivo (Sun & Baltimore, 1989). Amino acid sequence determination of the termini of poliovirus proteins established that all but one of the remaining cleavages of the polyprotein occurred at QG dipeptides (Nicklin et al., 1986, and references cited therein; Figure 1). Cleavage at QG sites in the P2 region of the polyprotein can probably be catalyzed by $3C^{pro}$ alone, but processing of the P1 capsid precursor requires additional sequences from the P3 region (Ypma-Wong & Semler, 1987). Protein 3CD (which contains the sequence of the polymerase 3D in addition to the 3C sequence) was able to cleave the P1 precursor efficiently at both sites in vitro (Jore et al., 1988; Ypma-Wong et al., 1988a; Figure 1). Similar results were obtained with proteins isolated from infected HeLa cells (Nicklin et al., 1988). Larger precursors (3BCD or P3) can also cleave P1 (Kuhn et al., 1988). It is therefore likely that the 3D sequence of $3CD^{pro}$ interacts with P1 in such a way that the QG sites in P1 can be recognized and cleaved. Cleavage by $3CD^{pro}$ of mutated P1 lacking a myristoylation signal is defective, which indicates that the interaction between P1 and $3CD^{pro}$ probably involves the myristic acid moiety (Kräusslich et al., 1989c). There are likely to be additional specific contacts between $3CD^{pro}$ and the structural domains of the capsid precursor, since even minor perturbations to the integrity of the P1 precursor are sufficient to impair cleavage (Ypma-Wong et al., 1988b, and references cited therein). These contacts may stabilize the precursor–proteinase complex, thus enabling $3CD^{pro}$ to cleave sites in the P1 precursor much more efficiently in vitro than $3C^{pro}$. $3C^{pro}$ is part of the nonstructural precursor and can catalyze all cleavages within it. However, the interaction between the P1 structural precursor and an appropriate proteinase may require a higher K_m, since they may have become spatially separated following the initial cleavage of structural from nonstructural precursors.

Processing of the poliovirus polyprotein serves three different functions, which can be correlated with three different proteolytic activities. First, structural and nonstructural precursors are separated from the nascent polyprotein by $2A^{pro}$, and second, release of different functional proteins is catalyzed by $3C^{pro}$ and $3CD^{pro}$. The last processing step in the formation of infectious particles is cleavage of VP0 to VP4 and VP2 at an N-S dipeptide (Figure 1), which occurs on encapsidation of viral RNA. It is not catalyzed by either $2A^{pro}$ or $3C^{pro}$. The three-dimensional structures of polio-, rhino-, and mengovirus each show the close proximity of a serine residue in VP2 to the carboxy end of VP4. It may act as a nucleophile, as in serine proteases; viral RNA could then act as a proton-abstracting base since there is no suitable His residue (Arnold et al., 1987). The ordered RNA in bean pod mottle virus (a plant virus relative of picornaviruses) is in a position and orientation that would be ideal to participate in this catalysis (Chen et al., 1989). However, the role of Ser10 in VP2 of poliovirus in the maturation cleavage has recently become questionable, since mutation of this serine to alanine or to cysteine yielded viable virus in which VP0 was cleaved (J. Harber and E. Wimmer, unpublished results).

All picornaviruses encode active and closely related 3C proteinases, but there is little similarity between the sequences of enzymatically active 2A proteinases of rhino- and enteroviruses and their counterparts in cardioviruses. In aphthoviruses, 2A consists of only 16 amino acid residues. A rapid primary cleavage reaction separates the structural precursor from the growing polyprotein of the latter two groups of viruses, but the mechanism by which it occurs has not been elucidated. Structural precursors can be further processed by $3C^{pro}$ in vitro to yield capsid proteins, but larger precursors

of $3C^{pro}$ are probably responsible for cleavage in vivo. For example, 3ABC is probably responsible for processing of the encephalomyocarditis virus capsid protein precursor (Jackson, 1986). This cleavage normally occurs in a defined stepwise manner, resulting in sequential release of the various capsid proteins (Shih & Shih, 1981). The mechanism by which this cascade is ordered is not known [although substitution of residues at one site has been shown to result in premature cleavage at another (Parks & Palmenberg, 1987)], but it is thought to be important in the formation of virion capsid structures.

Determinants of Picornavirus Substrate Recognition

Cleavage site recognition by $3C^{pro}$ of poliovirus is unusually stringent, for no corresponding picornaviral proteinase cleaves exclusively at one dipeptide (Figure 1). The residues occurring at cleavage sites within the polyproteins of other picornaviruses have been determined or predicted with reasonable certainty. It is clear that cleavage occurs within a small subset of dipeptides comprising Q-G, -S, -T, -V, -A, and -M and E-G and -S. Four of thirteen QG dipeptides and eight of ten YG dipeptides in the poliovirus polyprotein are not cleaved by $3C^{pro}$ and by $2A^{pro}$, respectively, so there must be additional determinants of cleavage site recognition. Moreover, there are clear differences in the efficiency of cleavage at different sites, resulting, for example, in the relative stability of 3AB and 3CD of poliovirus, and in the ordered stepwise cleavage of the encephalomyocarditis virus capsid precursor.

The specificity of $2A^{pro}$ and $3C^{pro}$ has been addressed by mutation of cleavage sites and by cleavage of peptide substrates. Lee and Wimmer (1988) noted that both bona fide $2A^{pro}$ cleavage sites are preceded by a Thr residue and have a Leu residue at position P4; this residue and other hydrophobic residues (I, V, and M) occupy the P4 position of VP1–2A cleavage sites in all other entero- and rhinoviruses. Substitution of a Thr by an Ala residue at position P2 of the 3C′/3D′ site abolished cleavage in vivo. Novel QA and naturally occurring QS and QG sites appear to be processed by encephalomyocarditis virus $3C^{pro}$ at normal rates in vitro [e.g., Parks et al. (1989)], but ten other mutated P1–P1′ sites were not cleaved, confirming the importance of residues at these positions. Interactions at positions other than P1 and P1′ are required since a QG dipeptide was not cleaved by purified poliovirus $3C^{pro}$ (Pallai et al., 1989). The P4 position may be important in cleavage site recognition: Nicklin et al. (1986) noted the prevalence of alanine or other aliphatic residues at this position, and Kuhn et al. (1988) found that some substitutions at this position impaired cleavage at the 3B/3C site. It is apparent that residues surrounding the QG dipeptide strongly influence cleavage efficiency, and this therefore appears to be a novel mechanism of regulation of gene expression. For example, the rate of cleavage of a peptide that corresponded to the 3C/3D site was increased by ca. 2 orders of magnitude by substitution of Thr by Ala at the P4 position (Pallai et al., 1989). Flanking Pro residues occur at most sites cleaved by $3C^{pro}$ within the encephalomyocarditis virus polyprotein, and it is interesting to note that polio peptides that are cleaved efficiently have Pro at the P2′ position, which may favor formation of a β-turn.

Determinants of Catalytic Activity

Inhibitor studies [references in Kräusslich and Wimmer (1988) and Sommergruber et al. (1989)] indicate that picornaviral 3C and 2A proteinases contain an active-site thiol group, suggesting that these enzymes constitute a class of cysteine proteinases. They are not related to the papain superfamily, since they are not inhibited by the characteristic inhibitor E-64 [e.g., Nicklin et al. (1988)] and show no sequence similarity with this family. Analysis of aligned sequences suggested that the viral proteinases are structurally related to trypsin-like serine proteinases and, more specifically, that His40, Asp85, and Cys147 of $3C^{pro}$ and His20, Asp38, and Cys109 of $2A^{pro}$ form the catalytic triad (Bazan & Fletterick, 1988). An evolutionary relationship has been proposed between a broad class of viral 3C-like Cys proteinases and the family of cellular trypsin-like Ser proteinases (Gorbalenya et al., 1989a,b; Bazan & Fletterick, 1989a,b); particularly notable is the replacement of the nucleophilic Ser by a spatially and functionally equivalent Cys residue in the catalytic triad of the viral enzyme. When a similar Ser-Cys substitution was introduced into the active sites of trypsin (Higaki et al., 1987) and subtilisin (Polgar & Bender, 1969; Neet et al., 1969), large reductions in activity were observed. However, additional changes may have a compensatory effect in viral proteinases, since substitution of the active-site Cys by Ser does not increase their activity as might have been predicted. Substitution of Cys147 by Ser was reported to inactivate $3C^{pro}$ of poliovirus in a bacterial expression system (Ivanoff et al., 1986); a similar Cys106-Ser substitution in polio $2A^{pro}$ reduced its activity by over 90% (M. Fäcke, H.-G. Kräusslich, and E. Wimmer, unpublished data). Viral proteinases function in controlled and limited proteolysis rather than in digestion, so that a reduction in turnover number coupled with a highly restricted substrate specificity may be a desirable property. Bazan and Fletterick (1988) suggested that the 3C and 2A subclasses of viral enzymes are homologous to the large (e.g., trypsin) and small (e.g., α-lytic protease) subclasses of trypsin-like proteases. Their model (derived by assignment of viral sequences which form the 12 components β-strands of the trypsin fold) enables prediction of structural features to be made. These are currently being investigated by X-ray crystallography of purified $3C^{pro}$ of poliovirus. Similarities between small bacterial and 2A-like proteinases may also involve the substrate-binding pocket. The primary specificity of α-lytic protease was relaxed and became much more permissive to large P1 residues as a result of substitutions that increased the size of the substrate binding pocket (Bone et al., 1989). Small residues occur naturally at equivalent positions in viral 2A proteinases, which may explain their ability to accommodate Tyr residues and to tolerate a number of substitutions at the P1 position (C. U. T. Hellen, C.-K. Lee, and E. Wimmer, unpublished results).

Proteolytic Processing of Plant Virus Polyproteins

There is a surprising degree of similarity between 3C and other nonstructural proteins of picornaviruses and proteins encoded by cowpea mosaic virus (Argos et al., 1974) and tobacco etch virus (Domier et al., 1987), which are members respectively of the comovirus and potyvirus groups of plant viruses. Moreover, the same order of nonstructural domains (putative helicase, VPg, proteinase, polymerase) is found in genomic RNAs of all three families of viruses. Although sequence similarity between como- and picornaviruses in the capsid regions is not apparent, capsid proteins of four picornaviruses and two comoviruses (cowpea mosaic virus and bean pod mottle virus) do have a common eight-stranded antiparallel β-barrel motif (Chen et al., 1989).

Comoviruses have a genome consisting of two positive-sense RNA molecules, both of which contain single large open reading frames; one encodes the capsid proteins while the other

Table I: Proteolytic Cleavage Sites in the Polyproteins Encoded (A) by Cowpea Mosaic Virus B-RNA, (B) by Cowpea Mosaic Virus M-RNA, (C) by Red Clover Mottle Virus M-RNA, and (D) by Bean Pod Mottle Virus M-RNA[a]

		sequence									
	site	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′
A	32K:58K	K	D	N	A	Q	S	S	P	V	I
	58K:VPg	S/G	A	E	P	Q	S	R	K	P	N
	VPg:24K	W	A	D	A	Q	M	S	L	D	Q
	24K:87K	I	A	Q	A	Q	G	A	E	E	Y
B	48/58:60	V	A	F	P	Q	M	E	Q	N	L
	VP37:VP23	G	A	I	A	Q	G	P	V	C	A
C	48/58:60	F	A	N	P	Q	T	D	T	D	L
	VP37:VP23	Q	A	E	A	Q	G	G	V	V	R
D	48/58:60	E	V	Q	A	Q	M	E	T	N	L
	VP37:VP23	G	T	I	P	Q	S	I	S	Q	Q

[a] Amino acid sequence data (A, B) from Wellink et al. (1986) and (D) from Chen et al. (1989) and G. Lomonossoff (personal communication); deduced amino acid sequence (C) from Shanks et al. (1986). The amino acid residues at cleavage sites are indicated by standard single-letter code and are described according to the nomenclature of Berger and Schechter (1970).

encodes nonstructural proteins. In contrast to picornaviruses, comoviruses synthesize their structual and nonstructural proteins as part of different polyproteins and therefore do not need a $2A^{pro}$-like activity. Potyviruses resemble picornaviruses in having monopartite positive-sense RNA genomes, which contain single large open reading frames (Figure 2). These are translated to yield polyproteins that are proteolytically processed to yield eight mature proteins (Dougherty & Carrington, 1988). The 24- and 49-kDa polypeptides of como- and potyviruses, respectively, are proteinases that resemble $3C^{pro}$ of picornaviruses. The cowpea mosaic virus proteinase can cleave all sites within both polyproteins in vitro. Remarkably, it also requires a cofactor for processing of the capsid precursor, but this is not the viral polymerase, as for poliovirus (Vos et al., 1988). The 49-kDa proteinase encoded by tobacco etch virus mediates its own autocatalytic release (probably in cis) and cleaves at three other positions in the polyprotein (Carrington & Dougherty, 1988; Carrington et al., 1988, and references cited therein). Potyviruses encode a second proteinase that is not related to the trypsin-like cysteine proteinases. It cleaves at a dipeptide (G-G) which is surrounded by residues that differ considerably from those at sites cleaved by the 49-kDa proteinase (Carrington et al., 1989; Figure 2). Cleavage by the 49-kDa proteinase of ten different potyviruses occurs at the dipeptide Q/(A, G, or S); all but one of these sites have H or F at the P2 position, and all but one have aliphatic residues at the P4 position. These potyvirus cleavage sites therefore resemble those of picornaviruses at P4, P1, and P1′ positions, but differ in that they extend over seven residues (Dougherty et al., 1989). The five sites recognized by the 49-kDa proteinase of tobacco etch virus are defined by the seven amino acid sequence: ExxYxQ/(G or S) (Figure 2). The role of these residues in defining a cleavage site has been investigated by substitution of residues at positions P7–P2′ (Dougherty et al., 1989, and references cited therein). Substitutions at absolutely conserved residues (P6, P3, P1) significantly reduced or even eliminated cleavage, while substitutions at other positions had little effect. Differences in nonconserved P4 and P2 positions of natural cleavage sites modulate their rate of cleavage and may therefore regulate the kinetics of formation of different gene products (Dougherty & Parks, 1989). Such a mechanism, which has been alluded to as a likely feature of processing of picornaviral polyproteins, would overcome the limitation that expression of a single polyprotein imposes on temporal regulation of expression. Cleavage sites within comovirus polyproteins resemble the 3C cleavage sites of picornaviruses at P4, P1, and P1′ positions (Table I). The trypsin/$3C^{pro}$ sequence/structure profiles developed by Bazan and Fletterick

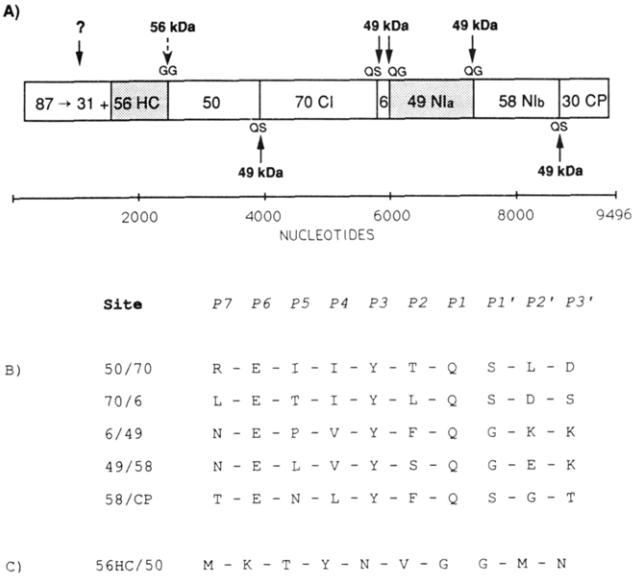


FIGURE 2: Gene organization, processing scheme, and cleavage sites of the tobacco etch virus polyprotein. (A) Proteolytic cleavages of the polyproteins occur between amino acid pairs indicated by standard single-letter code. Arrows above and below the polyprotein indicate sites that are cleaved in inter- and intramolecular reactions, respectively, by proteinases as indicated. The question mark indicates that the site and mechanism of cleavage are not known. The nomenclature of potyvirus proteins is as described by Domier et al. (1987), and the molecular weights are as described by Dougherty and Carrington (1988). The amino acid residues at sites cleaved (B) by the 49-kDa and (C) by the 56-kDa proteinases are described according to the nomenclature of Berger and Schechter (1970).

(1988) and by Gorbalenya and colleagues have been used to identify proteinase domains within polyproteins encoded by further families of plant (nepo-, sobemo-, and luteovirus) and animal (flavi-, pesti- and coronavirus) viruses (Table II; Bazan & Fletterick, 1989a,b; Gorbalenya et al., 1989a,b). All have genomes which encode long polyproteins and which have been shown to be processed to yield mature proteins. The predicted spatial disposition of active-site residues in these putative proteinases is similar to that of $3C^{pro}$, but surprisingly, proteinase domains in four of these virus groups have Ser-active centers.

RETROVIRIDAE

Retroviridae are a family of enveloped RNA viruses that have common morphological, biochemical, and physical properties (Teich, 1984) and differ from plus-strand RNA viruses in replicating via an obligatory DNA intermediate. The genomes of all replication-competent retroviruses consist of three major genetic elements that are arranged in the order

Table II: Viral Proteinases[a]

family	virus	enzyme	enzyme type	catalytic class	role in viral replication
picornavirus	polio	2A	trypsin-like (small subclass)	Cys-proteinase	separation of functionally different domains; mediates cleavage of cellular proteins
	polio	3C	trypsin-like	Cys-proteinase	release of nonstructural components from polyprotein
		3CD	trypsin-like	Cys-proteinase	release of capsid proteins from precursor
	FMDV	3C	trypsin/3C-like	Cys-proteinase	release of individual components from polyprotein; cleavage of cellular proteins
	FMDV	L	?	?	autocatalytic release from precursor; mediates cleavage of cellular proteins
	FMDV	?	?	?	separation of domains at 2A/2B junction
	polio	?	?	base-catalyzed?	maturation cleavage of VP0 to VP4 and VP2
comovirus	CPMV	24 kDa	trypsin/3C-like	Cys-proteinase	release of individual components from polyprotein
potyvirus	TEV	49 kDa	trypsin/3C-like	Cys-proteinase	release of individual proteins from polyprotein
	TEV	87 kDa	?	?	autocatalytic release from polyprotein
nepovirus	TBRV	23 kDa	trypsin/3C-like	Cys-proteinase	release of individual components from polyprotein
sobemovirus	SBMV	105 kDa	trypsin/3C-like	Ser-proteinase	release of individual components from 105-kDa P1 polyprotein
luteovirus	PLRV	70 kDa	trypsin/3C-like	Ser-proteinase	release of individual components from 115-kDa polyprotein?
flavivirus	YFV	ns3	trypsin/3C-like	Ser-proteinase	separation of individual components from polyprotein
pestivirus	HCV	p80	trypsin/3C-like	Ser-proteinase	separation of individual components from polyprotein
coronavirus	IBV	?	trypsin/3C-like	Cys-proteinase	release of individual components from polyprotein?
			trypsin-like	Cys-proteinase?	release of individual components from polyprotein?
togavirus	SV	capsid	?	Ser-proteinase	separation of functionally different domains
		ns2	?	?	separation of individual components from polyprotein destined for different subcellular compartments
retrovirus	HIV-1	PR	pepsin-like	Asp-proteinase	separation of individual components from polyprotein; early events in infection?
nodavirus	FHV	α-protein	?	?	maturation cleavage of capsid protein
adenovirus		19 kDa	?	Ser-proteinase	maturation cleavage of precursor proteins

[a] Abbreviations: polio, poliovirus; FMDV, foot-and-mouth disease virus; CPMV, cowpea mosaic virus; TEV, tobacco etch virus; TBRV, tobacco black ring virus; SBMV, southern bean mosaic virus; PLRV, potato leaf roll virus; YFV, yellow fever virus; HCV, hog cholera virus; IBV, infectious bronchitis virus; SV, Sindbis virus; HIV-1 human immunodeficiency virus, type 1; FHV, flockhouse virus. Proteinases encoded by specific viruses are not necessarily encoded by all members of a virus family. Questions marks indicate that the nature or identity of a specific viral proteinase has not been identified. Proteinases encoded by SBMV, PLRV, and IBV have been identified by sequence and structural-pattern analysis, but their sizes have not been determined. Processing of viral precursor proteins has been demonstrated in other RNA viruses, including birna-, reo-, calici-, and tymoviruses [see Kräusslich and Wimmer (1988) for references]. It has been suggested that DNA viruses, such as vaccinia virus and adeno- (see above) and hepadaviruses, encode proteinases that cleave virus-encoded proteins; recent experimental evidence does not support the assignment of proteinase activity to a sequence found within hepadaviruses, which was made on the basis of weak similarity between it and aspartic proteinases.

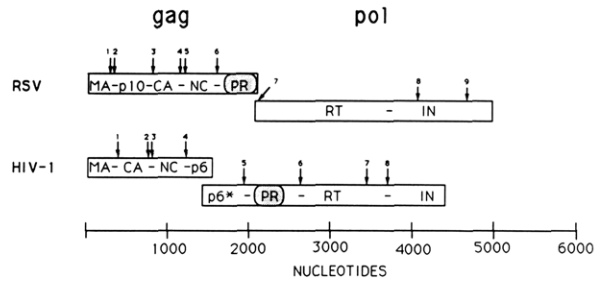
5′-*gag-pol-env*-3′. The *gag* (group-specific antigen) region encodes up to six structural proteins, which form the retroviral nucleocapsid and which are translated as a precursor from 35S mRNA. The *pol* region encodes the viral replication enzymes [reverse transcriptase (RT), integrase (IN), and in most instances, proteinase (PR), in the order PR-RT-IN] and is also translated from 35S mRNA as part of a *gag-pol* fusion polyprotein [Figure 3; nomenclature according to Leis et al. (1988)]. Synthesis of the *gag-pol* polyprotein is achieved either by suppression of an *amber* termination codon at the end of the *gag* gene (e.g., murine leukemia virus) or by ribosomal frame-shifting at one or two sites at the end of or within the *gag* gene (e.g., Rous sarcoma virus and HIV-1) [references in Kräusslich and Wimmer (1988); Figure 3). The infrequency of frame-shifting and suppression of termination codons leads to an overproduction of structural proteins compared to replicative enzymes.

The gene order of the *gag-pol* polyprotein is reminiscent of the poliovirus polyprotein; a proteinase maps to a location between structural and nonstructural proteins and severs these two regions. Proteolytic processing by the retrovirus-encoded proteinase is essential for a productive infectious cycle. Defective avian retroviruses which lack PR do not process the *gag* polyprotein (Hayman et al., 1979); cleavage was similarly abolished on mutation of the coding sequence of proteinases of avian, murine, and human retroviruses. PR-deficient mutants of murine leukemia virus (Crawford & Goff, 1985) and HIV-1 (Kohl et al., 1988) produced noninfectious virions containing unprocessed polyprotein, which indicates that processing of the polyprotein is not required for particle formation but is necessary to render these particles infectious. These observations are crucial if inhibitors of retroviral proteinases are to serve as antiviral drugs.

Structural and Biochemical Characterization of Retroviral Proteinases

The amino acid sequences of a number of purified retroviral PRs have been either wholly or partially determined by direct analysis. Comparison of these sequences with the deduced amino acid sequences of other retroviruses and related genetic elements revealed a conserved Asp-Thr(Ser)-Gly sequence, which is homologous with a sequence in the active site of the aspartic proteinase family (Toh et al., 1985). Aspartic proteinases have two homologous domains and may have evolved by gene duplication and fusion, whereas retroviral PRs are less than half as long and could thus correspond to only one such domain. Refined consensus templates for cellular and putative viral aspartic proteinases (established on the basis of patterns of residue conservation and known or predicted secondary structure) implied strong similarities between these domains. It was therefore proposed that retrovirus proteinases are dimers of identical PR polypeptide chains, with a fold similar to other aspartic proteinases; viral domains were predicted to differ principally in having shortened connecting loops between the strands of β-sheet that formed the core (Pearl & Taylor, 1987b). Similar parsimony was also noted when viral and cellular trypsin-like enzymes were compared (Bazan & Fletterick, 1988). Many of these proposals were confirmed following determination of the crystal structures of Rous sarcoma and HIV-1 proteinases.

The Rous sarcoma virus PR monomer is 124 amino acids long; it has an approximate intramolecular 2-fold symmetry and consists of two helices and of several β-strands connected by loops and turns (Figure 4). The strands in the core of the



RSV Sites	*P4*	*P3*	*P2*	*P1*	*P1'*	*P2'*	*P3'*
1	G	T	S	C	Y	H	C	G	T	A
2	P	P	Y	V	G	S	G	L	Y	P
3	P	V	V	A	M	P	V	V	I	K
4	I	A	A	A	M	S	S	A	I	Q
5	Q	P	L	I	M	A	V	V	N	R
6	P	P	A	V	S	L	A	M	T	M
7	R	A	T	V	L	T	V	A	L	H
8	T	F	Q	A	Y	P	L	R	E	A
9	S	P	L	F	A	G	I	S	D	W
HIV-1 Sites										
1	V	S	Q	N	Y	P	I	V	Q	N
2	K	A	R	V	L	A	E	A	M	S
3	T	A	T	I	M	M	Q	R	G	N
4	R	P	G	N	F	L	Q	S	R	P
5	V	S	F	N	F	P	Q	I	T	L
6	C	T	L	N	F	P	I	S	P	I
7	G	A	Q	T	F	Y	V	N	L	R
8	I	R	K	I	L	F	L	D	G	I

FIGURE 3: Coding regions from the genomes of Rous sarcoma virus (RSV) and human immunodeficiency virus, type 1 (HIV-1). Only the *gag* and *pol* regions have been shown, the latter being oriented relative to the former region, which has arbitrarily been placed in the first reading frame. Solid lines indicate start and end positions of open reading frames and stop codons within a functional reading frame. The nomenclature of retrovirus proteins is according to Leis et al. (1988). The amino-terminal product of the *pol* reading frame of the HIV-1 genome is designated p6*, and the positions of the proteinases within the polyproteins are indicated by stippled boxes. Cleavage sites within the RSV and HIV-1 polyproteins are numbered and indicated by arrows; amino acid residues at these positions are indicated by standard single-letter code and are described accoridng to the nomenclature of Berger and Schechter (1970).

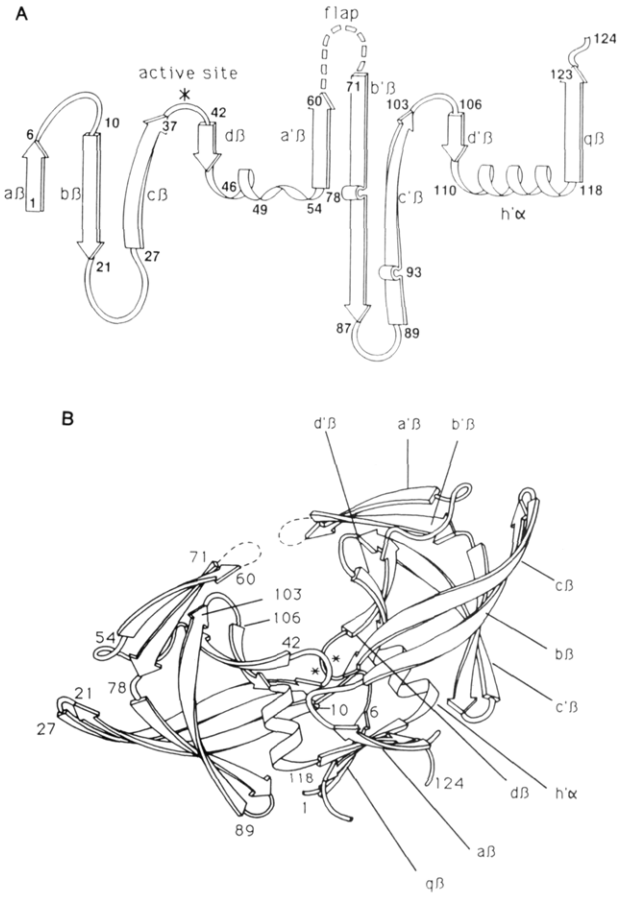


FIGURE 4: Schematic representation of (A) the secondary and (B) the tertiary structures of Rous sarcoma virus proteinase. Amino acid residues are numbered from the amino terminus; α-helices are depicted as helical ribbons and β-strands as arrows (labeled a–d, a′–d′, and q). The active site of the proteinase (Asp37–Ser38–Gly39) is located between cβ and dβ strands and is indicated by an asterisk. The disordered flap region (residues 61–70 in each monomer) is represented by a dashed line. Aspartic proteinases have a helix between d and a′ β-strands, which in Rous sarcoma virus proteinase is reduced to a single turn followed by a distorted loop (residues 50–55). Part B was drawn by J. Richardson, Duke University, and is used with permission.

monomer are organized into a sandwich of two four-stranded β-sheets; one is composed of antiparallel β-chains whereas the other comprises two superimposed ψ structures. The terminal strands of two monomers form a common four-stranded β-sheet (the dimer interface), which therefore differs from the six-stranded interface regions of pepsin-like proteinases. The active-site triplet of each subunit (residues 37–39) is located between cβ and dβ chains. Mutation of the Asp37 putative active-site residue confirmed its importance for catalytic activity (Kotler et al., 1988). The apposing residues of the active site are hydrogen-bonded in a "fireman's grip" configuration, which is characteristic of aspartic proteinases. A disordered "flap" region from each monomer (residues 61–70) projects over the active site, forming a large substrate-binding cleft; by contrast, pepsin-like proteinases have only one, N-terminal, flap. The subunits of the dimer are related by a nearly perfect 2-fold axis of symmetry which passes between the flaps, the active-site Asp residues, and the two C-terminal strands of the interface sheet. The resulting symmetrical active site contrasts with the asymmetric nature of retroviral substrates.

The HIV-1 PR is 99 amino acids long and is thus 25 residues shorter than the Rous sarcoma virus PR. The predicted similarity between the two PRs has recently been confirmed by analysis of the crystal structure (Wlodawer et al., 1989); they have the same fold, and there is an rms deviation of only 1.5 Å for 86 common C_α atoms; deletion of very few residues is necessary for proper alignment of the two PRs. Similar results were reported previously in another determination of the HIV-1 PR crystal structure, but the exact topology of the terminal strands could not be assigned accurately because of the lower resolution of that study (Navia et al., 1989). The structure determined by Wlodawer et al. (1989) indicates that Weber et al. (1989) accurately predicted protein subsites for each substrate residue (positions P4–P3′). Large regions of HIV-1 and Rous sarcoma virus PRs have structural analogues in microbial aspartic proteinases; over half the residues in an HIV-1 PR monomer could be fitted closely with the N- and C-terminal domains of microbial aspartic proteinases (Wlodawer et al., 1989).

Mutations have been made at each residue within HIV-1 PR, and their effects can now be related to the enzyme's structure. The importance of the active-site residue was confirmed by substitution of Asp25 [e.g., Kohl et al. (1988)]. Loeb et al. (1989b) found that three regions were highly sensitive to mutation. The first region comprises residues Ala22–Leu33, which surround the sequence noted by Toh et al. (1985). These residues form parts of the cβ and dβ strands and contain the active-site residue and the highly conserved Thr26 residue, which is believed to contribute toward dimer stabilization (Wlodawer et al., 1989). The second region

comprises residues Ile47–Gly52 and forms part of the flap, which probably interacts with the substrate (Weber et al., 1989). The third region comprises residues Thr74–Arg87, which form part of c′β and d′β strands. The flap and the superimposed ψ structures formed by the first and third regions are presumably the main structural and functional determinants of the enzyme.

Studies with proteinase inhibitors are consistent with the classification of retroviral proteinases as aspartic enzymes. Pepstatin A (a transition-state analogue inhibitor of aspartic proteinases) inhibited polyprotein processing (Katoh et al., 1987; Kräusslich et al., 1988; Roberts & Oroszlan, 1989) and also inhibited cleavage of synthetic peptide substrates by purified PR at inhibitory concentrations similar to those observed for renin (Kräusslich et al., 1989a; Meek et al., 1989; Richards et al., 1989). Characteristic inhibitors of aspartic proteinases supported the classification of HIV-1 PR as a member of this class of enzyme (Katoh et al., 1989; Meek et al., 1989). Biochemical experiments supported predictions from molecular modeling and structure determination that only dimers of retroviral PRs would be catalytically competent (Yoshinaka & Luftig, 1980; Nutt et al., 1988; Meek et al., 1989; Katoh et al., 1989). Results which suggested that the concentration of the PR-containing precursor is an important determinant of activity (Kräusslich et al., 1988) are consistent with the requirement for dimerization to generate the functional enzyme. A minimum cleavage concentration of PR (related to the dissociation constant of the dimer) may be required for activity. The pH optimum for cleavage of synthetic peptides by HIV-1 and RSV PR was 4.5–5.5 (Nutt et al., 1988; Billich et al., 1988; Kotler et al., 1989). Higher pH inhibited enzyme activity strongly (Richards et al., 1989), and the cytoplasmic neutral pH is therefore nonoptimal and may inhibit premature processing, an important point considering that cleavage of the retrovirus polyprotein normally only occurs in immature particles. The pH optimum of retroviral proteinases is considerably higher than that of cellular aspartic proteinases (e.g., pepsin, for which the optimum is ca. pH 3.5), with the sole exception of renin (whose optimum is pH 5.5–7.5). The different pH optimum of renin is due in part to substitution by Ala218 of the Thr or Ser residue common to most aspartic proteinases (Sielecki et al., 1989). It is therefore noteworthy that substitution of Ala40 by Ser in Rous sarcoma virus PR (which corresponds to Ala218 in renin) lowered the pH optimum by one pH unit (J. Leis, personal communication).

Determinants of Retroviral Cleavage Site Recognition

Retroviral proteinases are highly substrate-specific, since they correctly cleave homologous or closely related precursors (Yoshinaka et al., 1985), but not *gag* precursors, or peptides corresponding to *gag* or *pol* cleavage sites of distantly related retroviruses (Yoshinaka et al., 1986; Kräusslich et al., 1989a). Cleavage sites recognized by a number of different proteinases have been determined (Figure 3; Katoh et al., 1985; Henderson et al., 1988a,b; Hizi et al., 1989, and references cited therein), but surprisingly, no consistent pattern is apparent that would account for the high specificity for their natural substrates that distinguishes retroviral PRs from many other aspartic proteinases. However, consensus sequences and binding requirements at PR subsites have been suggested (Pearl & Taylor, 1987a; Henderson et al., 1988b; Weber et al., 1989). The cleavage site is defined by residues between P4 and P3′ positions; significant cleavage by HIV-1 PR was only observed for peptides spanning these residues [e.g., Darke et al. (1988) and Billich et al. (1988)] and did not increase as substrate length increased to include P5, P6, P4′, or P5′ residues (Moore et al., 1989). Cleavage of the HIV-1 MA/CA site was impaired by substitutions made between but not outside these positions (Partin et al., 1989). Tyr/Pro and Phe/Pro dipeptides appear frequently at P1 and P1′ positions, but the variety of other residues in these positions is remarkable: Leu/Ala, Leu/Phe, Met/Met, Met/Ser, and Phe/Leu dipeptides occur at HIV-1 cleavage sites (Figure 3). Most substitutions made at these positions of three HIV-1 *pol* sites impaired processing (Loeb et al., 1989a). The residues in the P2 and P4 positions are additional determinants of HIV-1 substrate specificity. The P2 position is commonly occupied by Asn residues, and even substitution of this residue by Gln at the MA/CA site abolished cleavage (K. Partin, C. Carter, and E. Wimmer, unpublished results). Structural analysis suggested that a small residue would be required to bind to the S4 site of HIV-1 PR. Ala, Ser, and Thr occur frequently at this position in natural cleavage sites, and a peptide that contained Phe at this position was not cleaved (Kräusslich et al., 1989a). Substitution of the Ser residue at the P4 position of the MA/CA cleavage site by Arg reduced the efficiency of cleavage, whereas substitution by Thr had no effect (K. Partin, C. Carter, and E. Wimmer, unpublished results). Pepstatin had a bulky isovaleryl group at the P4 position and has a K_i for HIV-1 PR that is 200 times worse than that of acetylpepstatin (Richards et al., 1989). The specificity of HIV-1 PR is broad enough for it to be able to accommodate peptide substrates in vitro that correspond to eight known *gag* and *pol* cleavage sites. Cleavage was efficient enough to suggest that these sites are also processed by HIV-1 PR in vivo (Darke et al., 1988). The variation in susceptibility to cleavage of peptides that correspond to different sites (Kräusslich et al., 1989a; Darke et al., 1988) suggests that residues which form cleavage sites may determine the order of cleavage in the processing cascade.

Proteolytic Processing Is a Determinant of Retroviral Morphogenesis

Retroviral polyproteins are not cleaved immediately after synthesis but are instead transported to the plasma membrane and are normally only processed once budding of immature particles occurs. The N-terminal glycine residues of many retroviral *gag* and *gag-pol* polyproteins are covalently linked to myristic acid (Schultz et al., 1988), and this modification directs the polyproteins to the membrane (Rhee & Hunter, 1987; Schultz & Rein, 1989). Membrane association of *gag* and *gag-pol* precursors probably concentrates and aligns them, promoting intermolecular interactions that may result in some self-assembly of *gag* molecules and dimerization of PR domains. Myristoylation is required for proteolytic processing of the *gag* precursors of murine leukemia and Mason–Pfizer monkey viruses (Rein et al., 1986; Rhee & Hunter, 1987). Göttlinger et al. (1989) found that myristoylation of HIV-1 *gag* and *gag-pol* polyproteins was not necessary for their cleavage in vivo when they were expressed at a high level (which may obviate the need for concentration by membrane association). However, myristoylation is necessary for HIV-1 assembly and consequently for production of infectious virus particles. Cleavage of the *gag* polyprotein is not required for budding of viral particles since immature particles released from retrovirus-transformed cells contain solely or largely uncleaved *gag* precursor (Yoshinaka & Luftig, 1977). Cleavage of polyproteins occurs largely after the release of virions from cells (Witte & Baltimore, 1978). Characteristic morphological changes associated with virion maturation (Gonda et al., 1985) are caused directly by cleavage of *gag*

precursor (Yoshinaka & Luftig, 1977), but the mechanism by which the processing cascade and budding of viral complexes from cells are initiated is not understood. Since HIV-1 PR is only active in dimeric form, it is possible that concentration of *gag-pol* precursor proteins at the plasma membrane is required to achieve a "minimal cleavage concentration". Premature processing, on the other hand, is undesirable because the liberated polyprotein domains would be devoid of their myristate "anchor" and would therefore diffuse back into the cytoplasm. Cleavage of the retrovirus polyprotein could occur as an intramolecular event in which two *gag-pol* polyproteins dimerize to cleave at their N- and C-termini or as an intermolecular event in which two such dimers cleave each other. It is difficult to envisage the major structural rearrangement of strands necessary for intramolecular cleavage of HIV-1 and RSV PRs, since the termini of both are ordered (Miller et al., 1989; Wlodawer et al., 1989). It is therefore likely that intermolecular cleavage, probably involving two polyprotein dimers, leads to the initial release of PR. Once released, PR would be confined within the budding particle; it could thus initiate a cascade of proteolytic cleavage events resulting in maturation of virus particles. Active enzyme can be isolated from mature, extracellular retrovirions. The packaged proteinase of equine infectious anemia virus is able to cleave the p11 NC protein to p6 and to p4 proteins when capsids are isolated at pH 7.6 (Roberts & Oroszlan, 1989). This step may be important in the early stages of the replication cycle, possibly including integration of viral DNA. It is not known whether it occurs in other retroviruses.

Conclusion

Numerous proteinases that are encoded by RNA viruses have been identified, and many have been well characterized (Table II). Most fall into one of two groups, related by sequence similarity and active-site topology to either trypsin-like serine or pepsin-like aspartic proteinases. $3C^{pro}$-like proteinases encoded by picorna-, como-, and potyvirus fall into the former category, whereas retroviral proteinases are of the latter type. Proteinases encoded by members of a number of other virus families have been identified, and although many have not been thoroughly characterized, it is likely that most will prove to be related to one or the other of these groups. However, there are viral proteinases, such as the nonstructural proteinase of Sindbis virus (Ding & Schlesinger, 1989), that are apparently not related to either of these groups.

Viral enzymes are much smaller than equivalent polypeptides of their cellular counterparts and also differ from most in having a lower turnover number and an unusually high degree of substrate specificity. This probably reflects the different functions that viral proteinases have evolved to fulfill but may fortuitously aid the design of specific inhibitors. Functions assigned to or suggested for virus-encoded proteinases include the separation of functionally different domains, the separation of components destined for different cellular components, maturation, and morphogenesis (Table II).

Cleavage sites have been identified and are defined by a hierarchy of parameters. The first to be identified was the importance of residues at the scissile bond, such as QG residues at P1–P1′ positions of poliovirus cleavage sites and hydrophobic residues in these positions for retroviruses. The importance of conserved residues in other positions indicates that viral cleavage sites are defined by an extended sequence of up to seven amino acids. Obvious additional determinants include the accessibility and flexibility of amino acid sequences surrounding cleavage sites. Attempts to identify all residues which determine a cleavage site by comparison of cleavage sites with a polyprotein are complicated by the heterogeneity at the same position within various sites. It is likely that this variation determines that rate and order of cleavage at the various sites, so that differential proteolytic release of specific precursors and products can be achieved despite the fact that synthesis of a polyprotein results in equimolar production of different domains. Different functions for cleavage intermediates and end products have been described (e.g., the different proteolytic specificities of $3C^{pro}$ and $3CD^{pro}$), and it is likely that other examples will be found in the future. Proteinases would thus appear to function in a regulatory capacity during the viral replication cycle.

The high degree of cleavage specificity of different viral proteinases raises the interesting possibility that they could be used as tools in molecular biological research, in an analogous manner to restriction enzymes. A number of examples have been described in which bona fide cleavage sites inserted into novel positions within unrelated proteins have been recognized and accurately cleaved by appropriate proteinases. For example, a $3C^{pro}$ site that can be cleaved by $3C^{pro}$ has been inserted into a surface loop of a viable poliovirus isolate (C. Mirzayan and E. Wimmer, unpublished results), and a site recognized by the tobacco etch potyvirus 49-kDa proteinase has similarly been introduced into a novel position within its polyprotein (Carrington & Dougherty, 1988).

Over the past decade viral proteinases have emerged as viable targets for chemotherapy, creating the exciting prospect that effective drugs will be developed to treat numerous viral diseases for which vaccines do not currently exist. The exquisite specificity of viral proteinases makes it unlikely that such drugs will interfere with normal cellular proteolytic processes, but the similarities between many viral proteinases make it likely that derivatives of an effective chemotherapeutic agent will be effective in inhibiting a whole family of viruses. Increasing appreciation of this potential has been parallelled by increasing awareness of the subtle interactions between viral substrates and proteinases and the consequent opportunities for sophisticated regulation of viral expression, replication, and morphogenesis that this affords.

Acknowledgments

We thank Alexander Wlodawer, Mariusz Jaskólski, and Kathryn Partin for critical reading of the manuscript and Alexander Wlodawer/Cold Spring Harbor Press for permission to use their figures. We are grateful to W. G. Dougherty and to the many other colleagues who provided unpublished information and preprints of their work.

References

Argos, P., Kamer, P., Nicklin, M. J. H., & Wimmer, E. (1984) *Nucleic Acids Res. 12*, 7251–7267.

Arnold, E., Luo, M., Vriend, G., Rossman, M. G., Palmenberg, A. C., Parks, G. D., Nicklin, M. J. H., & Wimmer, E. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 21–25.

Bazan, J. F., & Fletterick, R. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7872–7876.

Bazan, J. F., & Fletterick, R. J. (1989a) *FEBS Lett. 249*, 5–7.

Bazan, J. F., & Fletterick R. J. (1989b) *Virology 171*, 637–639.

Berger, A., & Schechter, I. (1970) *Philos. Trans. R. Soc. London, B 257*, 249–264.

Billich, S., Knoop, M. T., Hansen, J., Strop, P., Sedlacek, J., Mertz, R., & Moelling, K. (1988) *J. Biol. Chem. 263*, 17905–17908.

Bone, R., Silen, J. L., & Agard, D. A. (1989) *Nature 339*, 191–195.

Carrington, J. C., & Dougherty, W. G. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3391–3395.
Carrington, J. C., Cary, S. M., & Dougherty, W. G. (1988) *J. Virol. 62*, 2313–2320.
Carrington, J. C., Cary, S. M., Parks, T. D., & Dougherty, W. G. (1989) *EMBO J. 8*, 365–370.
Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kramer, G., Shanks, M., Lomonossoff, G. & Johnson, J. E. (1989) *Science 245*, 154–159.
Chow, M., Newman, J. F. E., Filman, D. J., Hogle, J. M., Rowlands, D. J., & Brown, F. (1987) *Nature 327*, 482–486.
Crawford, S., & Goff, S. P. (1985) *J. Virol. 53*, 899–907.
Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, T. M., Leu, C.-T., Lumma, P. K., Freidinger, R. M., Veber, D. F., & Sigal, I. S. (1988) *Biochem. Biophys. Res. Commun. 156*, 297–303.
Ding, M., & Schlesinger, M. J. (1989) *Virology 171*, 280–284.
Domier, L. L., Shaw, J. G., & Rhoads, R. E. (1987) *Virology 158*, 20–27.
Dougherty, W. G., & Carrington, J. C. (1988) *Annu. Rev. Phytopathol. 26*, 123–143.
Dougherty, W. G., & Parks, T. D. (1989) *Virology 172*, 145–155.
Dougherty, W. G., Cary, S. M., & Parks, T. D. (1989) *Virology 171*, 356–364.
Gonda, M. A., Wong-Staal, F., Gallo, R. C., Clements, J. E., Narayan, O., & Gilden, R. V. (1985) *Science 227*, 173–177.
Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M., & Koonin, E. V. (1989a) *FEBS Lett. 243*, 103–114.
Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., & Blinov, V. M. (1989b) *Nucleic Acids Res. 17*, 4847–4861.
Göttlinger, H. G., Sodroski, J. G., & Haseltine, W. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5781–5785.
Hayman, M. J., Royer-Pokora, B., & Graf, T. (1979) *Virology 92*, 31–45.
Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M., & Oroszlan, S. (1988a) in *Human Retroviruses, Cancer and AIDS: Approaches to Prevention and Therapy*, pp 135–147, Alan R. Liss, New York.
Henderson, L. E., Benveniste, R. E., Sowder, R. C., Copeland, T. D., Schultz, A. M., & Oroszlan, S. (1988b) *J. Virol. 62*, 2587–2595.
Higaki, J. N., Gibson, B. W., & Craik, C. S. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 615–621.
Hizi, A., Henderson, L. E., Copeland, T. D., Sowder, R. C., Krutzsch, H. C., & Oroszlan, S. (1989) *J. Virol. 63*, 2543–2549.
Ivanoff, L. A., Towatari, T., Ray, J., Korant, B. D., & Petteway, S. R., Jr. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5392–5396.
Jackson, R. J. (1986) *Virology 149*, 114–127.
Jore, J., de Geus, B., Jackson, R. J., Pouwels, P. H., & Enger-Valk, B. E. (1988) *J. Gen. Virol. 69*, 1627–1636.
Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T., & Oroszlan, S. (1985) *Virology 145*, 280–292.
Katoh, I., Yasunaga, T., Ikawa, Y., & Yoshinaka, Y. (1987) *Nature 329*, 654–656.
Katoh, I., Ikawa, Y., & Yoshinaka, Y. (1989) *J. Virol. 63*, 2226–2232.
Kitamura, N., Semler, B. L., Rothberg, P. G., Larsen, G. R., Adler, C. J., Dorner, A. J., Emini, E. A., Hanecak, R., Lee, J. J., van der Werf, S., Anderson, C. W., & Wimmer, E. (1981) *Nature 291*, 547–553.
Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M., & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4686–4690.
Kotler, M., Katz, R. A., Danho, W., Leis, J., & Skalka, A. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4185–4189.
Kräusslich, H.-G., & Wimmer, E. (1988) *Annu. Rev. Biochem. 57*, 701–754.
Kräusslich, H.-G., Nicklin, M. J. H., Toyoda, H., Etchison, D., & Wimmer, E. (1987) *J. Virol. 61*, 2711–2718.
Kräusslich, H.-G., Schneider, H., Zybarth, G., Carter, C. A., & Wimmer, E. (1988) *J. Virol. 62*, 4393–4397.
Kräusslich, H.-G., Ingraham, R. H., Skoog, M. T., Wimmer, E., Pallai, P. V., & Carter, C. A. (1989a) *Proc. Natl. Acad. Sci. U.S.A. 86*, 807–811.
Kräusslich, H.-G., Oroszlan, S., & Wimmer, E., Eds. (1989b) *Viral proteinases as targets for chemotherapy*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Kräusslich, H.-G., Hölscher, C., Reuer, Q., Harber, J., & Wimmer, E. (1989c) (submitted for publication).
Kuhn, R. J., Tada, H., Ypma-Wong, M.-F., Semler, B. L., & Wimmer, E. (1988) *J. Virol. 62*, 4207–4215.
Lee, C.-K., & Wimmer, E. (1988) *Virology 166*, 405–414.
Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M., Temin, H. M., & Vogt, V. (1988) *J. Virol. 62*, 1808–1809.
Loeb, D. D., Hutchison, C. A., Edgell, M. H., Farmerie, W. G., & Swanstrom, R. (1989a) *J. Virol. 63*, 111–121.
Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., & Hutchinson, C. A. (1989b) *Nature 340*, 397–400.
Meek, T. D., Dayton, B. D., Metcalf, B. W., Dreyer, G. B., Strickler, J. E., Gorniak, J. G., Rosenberg, M., Moore, M. L., Magaard, V. W., & Debouck, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1841–1845.
Miller, M., Jaskolski, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1989) *Nature 337*, 576–579.
Moore, M. L., Bryan, W. M., Fakhoury, S. A., Magaard, V. W., Huffman, W. F., Dayton, B. D., Meek, T. D., Hyland, L., Dreyer, G. B., Metcalf, B. W., Strickler, J. E., Gorniask, J. G., & Debouck, C. (1989) *Biochem. Biophys. Res. Commun. 159*, 420–425.
Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P. (1989) *Nature 337*, 615–620.
Neet, K. E., Nanci, A., & Koshland, D. E., Jr. (1968) *J. Biol. Chem. 243*, 6392–6401.
Nicklin, M. J. H., Toyoda, H., Murray, M. G., & Wimmer, E. (1986) *Bio/Technology 4*, 33–42.
Nicklin, M. J. H., Kräusslich, H.-G., Toyoda, H., Dunn, J. J., & Wimmer, E. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4002–4006.
Nicklin, M. J. H., Harris, K. S., Pallai, P. V., & Wimmer, E. (1988) *J. Virol. 62*, 4586–4593.
Nutt, R. F., Brady, S. F., Darke, P. L., Ciccarone, T. M., Colton, C. D., Nutt, E. M., Rodkey, J. A., Bennett, C. D., Waxman, L. H., Sigal, I. S., Anderson, P. S., & Veber, D. F. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7129–7133.
Pallai, P. V., Burckhardt, F., Skoog, M. T., Schreiner, K., Bax, P., Cohen, K., Hansen, G., Palladino, D. E. H., Harris, K. S., Nicklin, M. J. H., & Wimmer, E. (1989) *J. Biol. Chem. 264*, 9738–9741.
Parks, G. D., & Palmenberg, A. C. (1987) *J. Virol. 61*, 3680–3687.
Parks, G. D., Baker, J. C., & Palmenberg, A. C. (1989) *J. Virol. 63*, 1054–1058.
Partin, K., Kräusslich, H.-G., Bradley, J., Wimmer. E., & Carter, C. (1989) in *Modern Approaches to new Vaccines*

including Prevention of AIDS (Lerner, R., Ginsberg, H., Chanock, R. M., & Brown, F., Eds.) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Pearl, L. H., & Taylor, W. R. (1987a) *Nature 328*, 482.
Pearl, L., H., & Taylor, W. R. (1987b) *Nature 329*, 351–354.
Polgar, L., & Bender, M. L. (1969) *Biochemistry 8*, 136–141.
Rein, A., McClure, M. R., Rice, N. R., Luftig, R. B., & Schultz, A. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7246–7250.
Rhee, S. S., & Hunter, E. (1987) *J. Virol. 61*, 1045–1053.
Richards, A. D., Roberts, R., Dunn, B. M., Graves, M. C., & Kay, J. (1989) *FEBS Lett. 247*, 113–117.
Roberts, M. M., & Oroszlan, S. (1989) *Biochem. Biophys. Res. Commun. 160*, 486–494.
Rueckert, R. R., & Wimmer, E. (1984) *J. Virol. 50*, 957–959.
Schultz, A. M., & Rein, A. (1989) *J. Virol. 63*, 2370–2373.
Schultz, A. M., Henderson, L. E., & Oroszlan, S. (1988) *Annu. Rev. Cell Biol. 4*, 611–647.
Shanks, M., Stanley, J., & Lomonossoff, G. P. (1986) *Virology 155*, 697–706.
Shih, C. T., & Shih, D. S. (1981) *J. Virol. 40*, 942–945.
Sielecki, A. R., Hayakawa, K., Fujinaga, M., Murphy, M. E. P., Fraser, M., Muir, A. K., Carilli, C. T., Lewicki, J. A., Baxter, J. D., & James, M. N. G. (1989) *Science 243*, 1346–1351.
Sommergruber, W., Zorn, M., Blaas, D., Fessl, F., Volkmann, P., Maurer-Fogy, I., Pallai, P., Merluzzi, V., Matteo, M., Skern, T., & Kuechler, E. (1989) *Virology 169*, 68–77.
Sun, X.-H., & Baltimore, D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2143–2146.
Teich, N. (1984) in *RNA Tumor Viruses* (Weiss, R., Teich, N., Varmus, H., & Coffin, J., Eds.) pp 25–207, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Toh, H., Ono, M., Saigo, K., & Miyata, T. (1985) *Nature 315*, 691.
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.
Vos, P., Verver, J., Jaegle, M., Wellink, J., van Kammen, A., & Goldbach, R. (1988) *Nucleic Acids Res. 16*, 1967–1985.
Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M., & Wlodawer, A. (1989) *Science 243*, 928–931.
Wellink, J., Rezelman, G., Goldbach, R., & Beyreuther, K. (1986) *J. Virol. 59*, 50–58.
Witte, O. N., & Baltimore, D. (1978) *J. Virol. 26*, 750–761.
Wlodawer, A., Miller, M., Jaskôlski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616–621.
Yoshinaka, Y., & Luftig, R. B. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3446–3450.
Yoshinaka, Y., & Luftig, R. B. (1980) *J. Gen. Virol. 48*, 329–340.
Yoshinaka, Y., Katoh, I., Copeland, T. D., & Oroszlan, S. (1985) *J. Virol. 55*, 870–873.
Yoshinaka, Y., Katoh, I., Copeland, T. D., Smythers, G. W., & Oroszlan, S. (1986) *J. Virol. 57*, 826–832.
Ypma-Wong, M.-F., & Semler, B. L. (1987) *Nucleic Acids Res. 15*, 2069–2088.
Ypma-Wong, M.-F., Dewalt, P. G., Johnson, V. H., Lamb, J. G., & Semler, B. L. (1988a) *Virology 66*, 265–270.
Ypma-Wong, M.-F., Filman, D. J., Hogle, J. M., & Semler, B. L. (1988b) *J. Biol. Chem. 263*, 17846–17856.