



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Proteolytic Enzymes of the Viruses of the Family Picornaviridae

ERNST M. BERGMANN AND MICHAEL N. G. JAMES

*Department of Biochemistry and Medical Research Council of Canada Group in Protein Structure  
Function, University of Alberta, Edmonton, Alberta, T6G 2G7 Canada*

---

- I. Picornaviridae
- II. Viral Replication and Polyprotein Processing
- III. Picornaviral Proteinases
- IV. Conclusions and Implications for  
Antiviral Strategies
- References

## I. PICORNAVIRIDAE

The Picornaviruses constitute a large family of positive-sense, single-stranded RNA viruses (+RNA viruses) (Rueckert, 1996). There are more than 200 known viruses that belong to this family and are classified into six genera (Table I). These viruses share the major features of the viral replication cycle, including the central role of the specific proteolytic processing of a viral polyprotein (Palmenberg, 1990). Individual details of the viral replication and of the polyprotein processing distinguish the genera of the family Picornaviridae (Ryan and Flint, 1997).

Picornaviruses are small icosahedral viruses. There are examples of atomic resolution structures of individual viruses from four of the six genera (rhino-, entero-, cardio-, and aphtho-). The assembly of the precursor of the capsids is regulated by successive proteolytic cleavages of the structural proteins. The final assembly of the protomers into the procapsids requires the RNA genome and is not completely understood (Rueckert, 1996). A nonenzymatic, so-called maturation cleavage within the assembled procapsids then yields the infectious virus (Palmenberg, 1990).

TABLE I The Family Picornaviridae

Genus	Number of serotypes	Examples	Associated disease	Proteolytic enzymes
<i>Enterovirus</i>	93 hnPV 1-3; Cox A1-A23, A24, B1-B6; EV 1-9, 11-21, 24-28, 68-71	Polio Coxsackie Echo	Myelitis, carditis men- ingitis, encephalitis, herpangina, myalgia pleurodynia, pneumonia	2A, 3C
<i>Rhinovirus</i>	105	Rhino	Common cold	2A, 3C
<i>Aphthovirus</i>	7	FMDV	Foot-and-mouth disease of cloven-hoofed animals	L, 3C
<i>Cardiovirus</i>	2	EMCV	?	3C
<i>Hepatovirus</i>	1	HAV	Hepatitis A	3C
<i>Parechovirus</i>	2 EV22, EV23	Echo 22 Echo 23	Myocarditis	3C

Picornaviruses cause a wide variety of diseases in humans and animals (Couch, 1996; Hollinger and Ticehurst, 1996; Melnik, 1996). These range from relatively mild and widespread infections, such as the common cold and hepatitis A, to rare but often severe enteroviral diseases (Table I). There is some evidence that Picornavirus infections are also involved in the onset of severe autoimmune diseases such as myocarditis, diabetes, and multiple sclerosis (Carthy *et al.*, 1997; Steinmann and Conlon, 1997). Recently a mouse model for a demyelinating disease that resembles multiple sclerosis, provided a plausible, immunological mechanism for the triggering of such diseases by a viral infection (Miller *et al.*, 1997).

## II. VIRAL REPLICATION AND POLYPROTEIN PROCESSING

### A. THE PICORNAVIRAL LIFE CYCLE

The Picornaviruses release their single-stranded, mRNA-like genome into the cytosol of the host cell where it is translated into large polyproteins. The resulting polyproteins are then cleaved by specific viral proteinases into the structural and nonstructural proteins of the virus (Kräusslich and Wimmer, 1988). Figure 1 shows a simplified scheme of the life cycle of a typical Picornavirus.

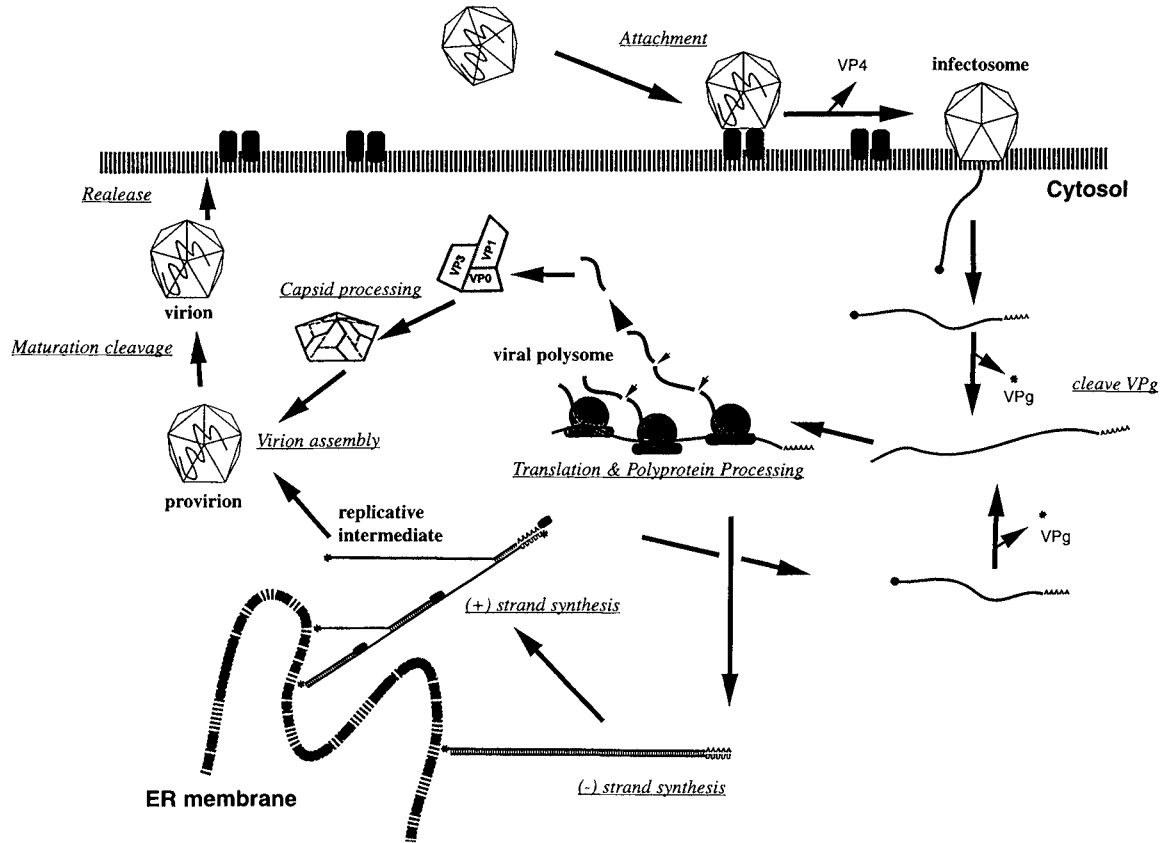


FIGURE 1 Simplified scheme of the replication cycle of a typical Picornavirus. At the center is the translation of the viral polyprotein and the specific, cotranslational polyprotein processing by the viral proteinases.

The virus binds to a specific receptor on the cell surface. The specific receptors are different for various Picornaviruses and in some cases are not yet known. Following attachment, the virus particles lose the VP4 protein and undergo a change that allows them to release their RNA genome into the cytosol of the host cell.

Picornaviral RNA genomes have a small, viral protein (VPg, the 3B gene product) covalently attached at the 5' terminus. After releasing the RNA into the cytosol of a host cell VPg is cleaved off by a cellular enzyme to yield a functional mRNA. The RNA is then translated into a large polyprotein which is cotranslationally proteolytically processed into the individual viral proteins (Palmenberg, 1990). Under normal conditions the full-length polyprotein is never found. Proteolytic processing of the polyprotein is accomplished by one or two viral proteinases which are themselves part of the polyprotein. The first proteolytic cleavage usually separates the structural proteins (P1 in Picornaviruses) from the remainder of the polyprotein.

The RNA genome is also replicated to yield negative-sense RNAs. The negative-sense RNAs then serve as templates for the production of viral genomes. Picornaviral RNA replication is accomplished by a multienzyme complex that includes the viral RNA-dependent RNA polymerase (3D), the putative viral RNA helicase (2C), and the picornaviral 3C proteinase (Porter, 1993; Wimmer *et al.*, 1993). Also part of this complex is the 3AB gene product. The 3A gene product presumably anchors the picornaviral RNA replication complex to the membrane of the smooth ER. Modification of the membrane structure of the host cell is a common feature of picornavirus infection (Bienz *et al.*, 1983, 1990; Teterina *et al.*, 1997a,b). The 3B gene product constitutes VPg—the viral genome-associated protein—which remains covalently bound to the 5'-end of the viral RNA genome (Wimmer, 1982). There is also evidence that some cellular proteins or their proteolytic cleavage products form part of the viral RNA replicase complex (Andino *et al.*, 1993; Xiang *et al.*, 1995; Gamarnik *et al.*, 1996; Parsley *et al.*, 1996).

Picornaviral 3C proteinases possess an RNA binding site and RNA binding activity that is distinct from its proteolytic activity (Hämmerle *et al.*, 1992; Andino *et al.*, 1993; Porter *et al.*, 1993; Walker *et al.*, 1995; Kusov and Gauss-Müller, 1997). It is common for the limited number of gene products of small RNA viruses to perform multiple, distinct functions. The exact function of the 3C proteinase within the RNA replicase complex is not clear. Apparently its RNA binding activity is required for the initiation of RNA replication. Certain proteolytic cleavages in some picornaviruses could be essential steps during RNA replication within the RNA replicase complex, e.g., the 3C-mediated cleavage of the RNA-associated VPg (3B) from the membrane anchor (3A) may be necessary to release the RNA from the membrane-bound replicase complex.

It is also believed that in some picornaviruses the cleavage of 3CD within the replicase complex after binding of the RNA is required to allow 3D to perform RNA replication (Harris *et al.*, 1992; Molla *et al.*, 1995).

Initially the structural proteins (P1) are cleaved from the viral polyprotein. Two more 3C-mediated cleavages (1AB|1C and 1C|1D or VP0|VP3 and VP3|VP1, respectively) are required within the capsid precursor. The resulting protomer then assembles into pentamers and the pentamers form the provirions by a poorly understood pathway that requires the VPg-linked RNA genome. A final maturation cleavage within the provirion (VP0 → VP2 + VP4) yields the infectious virus particles. This maturation cleavage is believed to be nonenzymatic and to require the presence of the packaged RNA genome (Palmenberg, 1990; Rueckert, 1996)

Many details such as the composition of the RNA replicase complex, the function of the individual components of the RNA replicase complex, the pathway of provirion assembly and so on are, even in the best studied viruses, not completely understood. There are also differences in many aspects of the viral life cycle between the individual genera of the Picornaviridae.

## B. POLYPROTEIN PROCESSING AND OTHER FUNCTIONS OF THE PICORNAVIRAL PROTEINASES

The genome of all picornaviruses carries at least one, more often two, genes encoding proteolytic enzymes (Ryan and Flint, 1997). The 3C gene product is the major processing proteinase in all picornaviruses. The primary function of the picornaviral proteinases is the cotranslational, specific cleavage of the viral polyprotein into the structural and nonstructural proteins. The individual proteolytic cleavages by the 3C proteinases within the picornaviral polyproteins are sequential; some sites are cleaved faster than others. The cleavage sites are identified by the sequence of the residues immediately preceding and following the scissile bond (approximately P<sub>4</sub> to P'<sub>2</sub> in the nomenclature of Schechter and Berger, 1967). The P<sub>1</sub> residue, immediately preceding the scissile peptide bond, is almost always a glutamine. The 3C proteinases of the individual picornaviruses also have sequence preferences for the residues at the P<sub>4</sub>, P<sub>2</sub>, P'<sub>1</sub>, and P'<sub>2</sub> sites of a cleavage site (Nicklin *et al.*, 1988; Long *et al.*, 1989; Pallai *et al.*, 1989; Weidner and Dunn, 1991; Malcolm, 1995). However, what distinguishes the good, preferred cleavages sites from the ones that are cleaved more slowly is not apparent from the peptide sequence. It is very likely that other factors, such as the accessibility and the local conformation, play a part in the determination of the sequence of cleavages.

The details of the polyprotein processing are one factor that distinguishes

the six different genera of the Picornaviridae (Ryan and Flint, 1997). Only a single proteinase, the 3C gene product, is present in the cardio-, hepato-, and parechoviruses. In the entero- and rhinoviruses the 2A gene product is a second proteolytic enzyme. An L proteinase at the amino-terminus of the polyprotein is a unique feature of the aphthoviruses.

The separation of the structural and nonstructural proteins is usually the primary cleavage event, but this is accomplished quite differently in the individual genera. In entero- and rhinoviruses 2A is a separate proteolytic activity. It performs the primary cleavage at its own amino-terminus which separates P1 from the nonstructural proteins. In the hepato- and parechoviruses the primary cleavage is a 3C-mediated cleavage at the amino-terminus of the 2B gene product (Jia *et al.*, 1993; Schultheiss *et al.*, 1994; Martin *et al.*, 1995; Schultheiss *et al.*, 1995a). It is not clear if the small 2A gene product has any function. In the aphtho- and cardioviruses the primary cleavage is at the carboxy-terminus of 2A. The 2A gene product is not a proteolytic enzyme in these viruses. The cleavage is presumably nonenzymatic and requires the carboxy-terminal residues of 2A (Palmenberg *et al.*, 1992; Donnelly *et al.*, 1997). The second proteinase present in the aphthoviruses, the L proteinase, only cleaves itself from the amino-terminus of the polyprotein (Strebel and Beck, 1986).

Larger precursors of the 3C proteinase, such as 3CD or 3ABC are also catalytically active proteinases (Ypma-Wong *et al.*, 1988; Harris *et al.*, 1992; Davis *et al.*, 1997). It has been shown in some systems that the presence of an additional domain can change the efficiency and specificity of the proteolytic activity. In poliovirus, 3CD is a proteinase with a distinct specificity (Ypma-Wong *et al.*, 1988). It cleaves at least some of the cleavage sites within the viral polyprotein more efficiently and is presumably required for the processing of the cleavage sites within the structural protein. In other picornaviruses other precursors may play a similar role. It has been suggested that 3ABC is a proteolytically active precursor of 3C in HAV (Harmon *et al.*, 1992; Schultheiss *et al.*, 1994).

Structural proteins of viruses in general are designed to form large assemblies, such as viral capsids. Therefore, they have to be synthesized as precursors, which are covalently modified before they can assemble. One very common form of modification, not only in small +RNA viruses, is proteolytic processing of the precursor (Kay and Dunn, 1990). This is one reason why proteolytic enzymes are among the most ubiquitous enzymatic activities expressed by viruses (Dougherty and Semler, 1993).

In the Picornaviruses the structural proteins are further proteolytically processed after they are separated from the nonstructural proteins. Two sequential, 3C-mediated cleavages are required within the capsid protein before the resulting 5S protomers can assemble into larger (14S) pentamers (Fig. 1). Final capsid assembly then requires the presence of RNA. In poliovirus the cleavages

of the capsid proteins within the protomer require the proteolytic activity of the precursor 3CD (Ypma-Wong *et al.*, 1988).

The picornaviral 3C proteinase cleaves itself out of the polyprotein. In experimental systems it was shown that this can be accomplished both in *cis*, when 3C is expressed as part of the polyprotein, or in *trans*, when 3C is expressed separately (Kräusslich and Wimmer, 1988; Harmon *et al.*, 1992). Kinetic evidence obtained with encephalomyocarditis virus (EMCV) also suggests that the cleavage in *cis* at both the amino and carboxy termini of 3C is intramolecular (Palmenberg and Rueckert, 1982). Another possible interpretation of these data is that the cleavages are performed by another 3C proteinase within a tight dimer or larger polymer. Further evidence for an intramolecular, autocatalytic cleavage of the 3C proteinase was provided by Hanecek *et al.* (1984).

The crystal structures of 3C proteinases have allowed one to deduce a structural model for an intramolecular cleavage of 3C at its own amino-terminus (Matthews *et al.*, 1994; Bergmann *et al.*, 1997). This model proposes that the amino-terminal helix, which is a unique feature of the 3C proteinase, folds out of the active site of 3C after 3C has cleaved its own amino-terminus. How and if the 3C proteinase could cleave its own carboxy-terminus in an intramolecular reaction is much less obvious.

An additional function of picornaviral proteinases is the inhibition or at least down-regulation of specific host cell functions which compete with the viral replication cycle (Ryan and Flint, 1997). The entero- and rhinoviral 2A proteinases cleave specifically one of the cellular proteins that forms part of the cap-recognition complex (eIF4G) (Lamphear *et al.*, 1993; Sommergruber *et al.*, 1994a; Haghighat *et al.*, 1996). This serves to down-regulate the translation of capped host cell mRNAs which competes with the translation of the picornaviral RNA genome. It is remarkable that the L proteinase of aphthoviruses, in spite of being a different proteinase, performs the same function. The entero-/rhinoviral 2A proteinase and the aphthoviral L proteinase-mediated cleavages of eIF4G occur in different places on the molecule (Kirchweger *et al.*, 1994). The picornaviruses which do not have a second proteolytic activity besides 3C do not cleave eIF4G or inhibit host cell translation by this mechanism. Hepatitis A virus even requires intact eIF4G for the translation of its own genome (Borman and Kean, 1997).

There are also reports of host cell proteins being substrates of the picornaviral 3C proteinases. Most of these cellular substrates of the 3C proteinases are involved in some aspect of cellular translation or replication (Ryan and Flint, 1997; Yalamanchili *et al.*, 1997).

Thus, there are three main functions of picornaviral proteinases: the specific processing of the viral polyprotein, covalent modification of the precursors of the viral capsid and down regulation of host cell processes by proteolytic



cleavage of host cell proteins (Gorbalenya and Snijder, 1996; Kay and Dunn, 1990; Kräusslich and Wimmer, 1988; Ryan and Flint, 1997). Cotranslational, specific processing of a viral polyprotein by a specific viral proteinase is an essential part of viral replication in +RNA viruses. This is true even for some families of +RNA viruses which have developed additional strategies to generate individual gene products from a single RNA genome, e.g., subgenomic RNAs or multiple ORFs. Proteolytic cleavage as a covalent modification of the precursors of viral structural proteins is even more common and occurs even in DNA viruses. Down-regulation of the host cell metabolism by specific cleavage of cellular proteins is a mechanism which is not found in all viruses. Given these important functions, it is not surprising that proteolytic enzymes are ubiquitous gene products in all +RNA and many other viruses.

### III. PICORNAVIRAL PROTEINASES

#### A. THE 3C PROTEINASE

##### 1. Structure

The major processing proteinase of the picornaviruses, the 3C proteinase, belongs to a new family of proteolytic enzymes: the chymotrypsin-like cysteine proteinases (Gorbalenya and Snijder, 1996). This had initially been predicted based on analysis of the sequence of the 3C gene product (Gorbalenya *et al.*, 1986, 1989; Bazan and Fletterick, 1988). This prediction was shown to be correct by the first crystal structure of a 3C proteinase (Allaire *et al.*, 1994). Refined crystal structures of 3C proteinases have now been published for the enzymes from hepatitis A virus (HAV), poliovirus (PV), and human rhinovirus (HRV) (Matthews *et al.*, 1994; Bergmann *et al.*, 1997; Mosimann *et al.*, 1997).

The three-dimensional structure of the 3C proteinases from HAV and PV are shown in Figs. 2 and 3, respectively. The two enzymes differ in size and belong to two subclasses of the 3C proteinases. The 3C gene product of HAV consists of 219 residues and the molecule from PV consists of 183 residues. In spite of the size difference, the core of the enzymes superimpose surprisingly well. The rms difference for the  $C_{\alpha}$ -atoms of 154 residues which superimpose closely is 1.85Å. This indicates that the core of the two domain structure of the 3C proteinase is fairly well conserved. Differences between the various 3C proteinases manifest in the length of the secondary structure elements and in the turns and loops that connect the  $\beta$ -strands and protrude from the core of the  $\beta$ -barrel domains.

In spite of being cysteine proteinases, the 3C proteinases belong structurally to the superfamily of chymotrypsin-like proteinases (Gorbalenya and Snijder,

1996). The structures of chymotrypsin-like proteinases are formed by two antiparallel  $\beta$ -barrels with the proteolytic active site at the domain interface. Both domains contribute to the catalytic residues in the active site. Both domains also participate in the binding of peptide substrates. The N-terminal domain is mostly involved in binding the substrate residues following the scissile peptide bond ( $P'_1$  to  $P'_3$ ) whereas the C-terminal domain forms the specific subsites for the substrate residues preceding the scissile bond ( $P_4$  to  $P_1$ ) (Perona and Craik, 1995).

The two domains of the chymotrypsin-like proteinases are usually described as six-stranded, antiparallel  $\beta$ -barrels, with the individual  $\beta$ -strands labeled aI-fI and aII-fII (Figs. 2 and 3, see color plates). An alternative description of the  $\beta$ -barrels is that of a sandwich of two orthogonal, four-stranded, antiparallel  $\beta$ -sheets (Chotia, 1984). The  $\beta$ -strands, which form the edge of the sheets, belong to both sheets and continue, sometimes uninterrupted, from one sheet to the other. As a result the two corners of the " $\beta$ -sandwich," which are formed by the edge-strands, are closed, while the other two corners are splayed (Chotia, 1984). In both  $\beta$ -barrel domains of the HAV 3C proteinase one of the edge strands is interrupted while the other continues from one  $\beta$ -sheet to the other (Fig. 2) (Bergmann *et al.*, 1997). In the N-terminal domain  $\beta$ -strand eI is interrupted by a single helical turn.  $\beta$ -strand bI forms a  $\beta$ -bulge at Val 28, allowing it to bend from one  $\beta$ -sheet to the other. The residue Val 28 is involved in the binding of peptide substrates by HAV 3C. In the C-terminal domain of HAV 3C the bII strand is interrupted by a short stretch of random coil structure, whereas the eII strand continues from one sheet to the other. There are seven defined  $\beta$ -strands in both of the domains of HAV 3C. In the two smaller  $\beta$ -barrels, which form the domains of the polio 3C, the edge strands continue uninterrupted from one sheet to the other (Fig. 3). There are six defined  $\beta$ -strands in each domain of the polio 3C proteinase (Mosimann *et al.*, 1997).

Two of the  $\beta$ -strands, bII and cII, of the C-terminal domain of the 3C proteinases are extended past the C-terminal  $\beta$ -barrel (Bergmann *et al.*, 1997). From the point where the two strands are no longer part of the  $\beta$ -barrel they form an antiparallel, two-stranded  $\beta$ -ribbon (light gray in Figs. 2 and 3). A defined  $\beta$ -bulge introduces a bend into this  $\beta$ -ribbon, which causes it to curl back toward the active site. The longer  $\beta$ -ribbon in the HAV 3C proteinase contributes to the residues involved in the catalytic mechanism and also to the binding of peptide substrates (see below). Because the  $\beta$ -ribbon is shorter in the poliovirus 3C, it only contributes to the  $P_4$  binding pocket and the proteolytic active site of polio 3C is much more accessible (Mosimann *et al.*, 1997). This  $\beta$ -ribbon is a unique feature of the 3C proteinase and replaces the "methionine loop" of the chymotrypsin-like serine proteinases. The corresponding topological feature is somewhat similar, but smaller, in some bacterial proteinases (e.g.,  $\alpha$ -lytic proteinase; Fujinaga *et al.*, 1985).

There are helices at the N- and C-termini of the 3C proteinases. The N-terminal helix packs against the C-terminal  $\beta$ -barrel and the C-terminal helix packs against the surface of the N-terminal domain. The two helices stabilize the structure like two latches (Bergmann *et al.*, 1997). The N-terminal  $\alpha$ -helix is a unique feature of the 3C proteinases among all chymotrypsin-like proteinases (Gorbalenya and Snijder, 1996). It has been speculated that it is important for the mechanism of a proposed intramolecular cleavage at the N-terminus of 3C (Matthews *et al.*, 1994; Bergmann *et al.*, 1997). In the proposed model for the N-terminal, intramolecular proteolytic cleavage, this helix is folded after 3C cleaves its own N-terminus. The favorable free energy of the folding of this stable helix may be required to fold the new N-terminus out of the active site in order to create the active proteinase (Bergmann *et al.*, 1997). The sequence of the residues which form the last turn of this helix is highly conserved throughout the picornaviral 3C genes ( $^{\text{K}}_R^{\text{R}}/_{\text{K}}\text{N}^{\text{L}}/_{\text{I}}$ ).

It is interesting that the structural and functional details of the proteolytic active site of the 3C proteinases are not the most conserved part of the 3C structure (Gorbalenya *et al.*, 1988). The 3C gene product constitutes one subunit of the picornaviral RNA replicase complex and has a distinct RNA binding site (Hämmerle *et al.*, 1992; Andino *et al.*, 1993; Leong *et al.*, 1993; Kusov *et al.*, 1997). The sequence of the residues which have been implicated in this second activity, KFRDI, is located in the domain connection of 3C, on the opposite site of the molecule from the proteolytic active site (Figs. 2b and 3b). It is completely conserved throughout the 3C gene sequences of all picornaviruses (Ryan and Flint, 1997). It was first shown for poliovirus that mutations within this sequence are deleterious for the viral replication and show two different phenotypes (Hämmerle *et al.*, 1992). These results can now be interpreted in light of the structures. The three charged residues within the consensus sequence (K82, R84, and D85 in poliovirus and K95, R97, and D98 in HAV) form part of the surface of the RNA binding site and are probably directly involved in RNA binding. The side-chains of the two highly conserved hydrophobic residues (F83 and I86 in poliovirus 3C and F96 and I99 in HAV) are packed into the interior of the molecule. They are part of the internal hydrophobic interactions that maintain the structure in this region and are important for this reason. The side-chain of the conserved phenylalanine interacts with a conserved glycine at the end of  $\beta$ -strand bI inside the N-terminal  $\beta$ -barrel. The sequence surrounding this glycine, LGV $^{\text{K}}_{\text{H}}\text{D}$ , is also highly conserved within the 3C genes. The residues in this sequence motif, from His 31 in poliovirus 3C and Lys 35 in HAV 3C on, form a reverse turn and connect  $\beta$ -strand bI and cI (Figs. 2b and 3b). They contribute to the surface of the RNA binding site of 3C. Presumably, also contributing to the molecular surface of the RNA binding site are the turns which connect  $\beta$ -strands dI and eI and dII and eII (Figs. 2b and 3b). The latter connection forms a single turn of a helix in HAV 3C (Fig. 2b).

One face of the N- and C-terminal helix each flanks the conserved residues within the domain connection and probably contributes to the RNA binding site. It appears likely, as was proposed by Ryan and Flint (1997), that the binding of RNA to this site would have an influence on the proteolytic processing of both the N- and C-termini of 3C. On the other hand, it is not known if binding of RNA to the RNA binding site of 3C affects the proteolytic activity. Only structural work on a complex of 3C and bound RNA could provide a definite answer to this question. Because the RNA binding site is on the opposite side of the molecule from the proteolytic active site, the structures suggest that it could be possible that the two activities are independent. However, the 3C structures also suggest a possible mechanism whereby binding of RNA in the RNA binding site of 3C could influence the proteolytic activity. The turns, which connect  $\beta$ -strands bI and cI and dII and eII, are probably involved in the specific binding of RNA. At the other end of each of these strands are residues which play important roles in the proteolytic activity. Slight conformational changes involving these  $\beta$ -strands could have a dramatic effect on the proteolytic activity.

## 2. Activity and Specificity

The picornaviral 3C proteinases are relatively slow enzymes when compared to some of the mammalian, extracellular serine proteinases. They have evolved to be very specific enzymes (Malcolm, 1995; Gorbalenya and Snijder, 1996; Ryan and Flint, 1997; Bergmann, 1998 and references therein).

The chymotrypsin-like proteinases belong to a large group of proteolytic enzymes in which the nucleophile is the oxygen or sulfur atom of the side-chain of a serine or cysteine residue, respectively. In these enzymes the general acid-base catalyst is a conserved histidine residue. It is generally accepted that the mechanism of these enzymes involves an acyl-enzyme intermediate formed between the nucleophile and the carbonyl of the P<sub>1</sub> residue of the substrate. Additional, so-called tetrahedral intermediates occur both during formation and hydrolysis of the acyl-enzyme intermediate. The tetrahedral intermediates carry a negative charge on the oxygen atom of the scissile peptide bond. In the catalytic reaction of the chymotrypsin-like serine proteinases the transition states leading to the tetrahedral intermediates are rate limiting and structurally resemble the tetrahedral intermediates.

Three chemical groups with distinct functions are typically found in the active sites of proteolytic enzymes (James, 1993; Ryan and Flint, 1997): a nucleophile, which attacks the carbonyl of the scissile peptide bond; a general acid-base catalyst, which assists in the attack and protonates the leaving group; and an electrophilic structure, which stabilizes the developing negative charge on the carbonyl. The latter structure is usually referred to as the oxyanion hole. In

the chymotrypsin-like serine proteinases it consists of a stretch of seven residues with a consensus sequence XGDSSG, where the serine is the nucleophile. The main-chain conformation of this structure orients the first and third peptide bonds so that they donate hydrogen bonds to the carbonyl of the scissile bond. These two hydrogen bonds of the oxyanion hole help to stabilize the developing negative charge on the carbonyl oxygen during the reaction (Whiting and Peticolas, 1994).

There are additional chemical groups in the active site of proteinases, the function of which is less clear. In the chymotrypsin-like serine proteinases the carboxylate of an aspartate residue interacts with the edge of the imidazole of the histidine general acid–base catalyst, which is opposite from the nucleophile ( $N^{\delta}$ ). Originally it was thought that this “third member of the catalytic triad” participates in a proton transfer, but it is now generally believed that its function is to maintain the orientation of the histidine general acid–base catalyst and possibly to stabilize its developing positive charge. There are chemical groups in similar positions to the carboxylate of the third member of the catalytic triad in the 3C proteinases, but the interactions with the histidine general acid–base catalyst are different (Fig. 4, see color plate).

It is generally accepted that the active sites of cysteine proteinases, such as the enzymes of the papain family, contain a thiolate–imidazolium ion pair that is stabilized over a wide pH range (Storer and Ménard, 1994). The active site of the 3C proteinases feature a thiol and an imidazole but in the structural context of a chymotrypsin-like proteinase. There is no direct experimental evidence for the charge and protonation state in the active site of the 3C proteinases. Even though the 3C proteinases belong to the superfamily of the chymotrypsin-like proteinases, they are cysteine proteinases. Whether the mechanism of 3C proteinases more closely resembles the mechanism of chymotrypsin-like serine proteinases or other cysteine proteinases or is unique is not clear.

Figure 4 shows the details of the active site residues of the 3C proteinases from HAV (Bergmann *et al.*, 1997) and PV (Mosimann *et al.*, 1997). The arrangement of the cysteine–histidine dyad and the oxyanion hole is similar to that observed in the chymotrypsin-like serine proteinases but due to the size of the sulfur nucleophile the active site is larger and the chemical groups are further apart.

The conserved glycine residue in the oxyanion hole of the wild-type 3C proteinases shows a conformation that is similar to the one seen for the corresponding residue in the chymotrypsin-like serine proteinases. This left-handed  $3_{10}$ -helical conformation requires a glycine in this position ( $\Phi = 95^\circ$ ,  $\Psi = -5^\circ$ ) (Bergmann *et al.*, 1997). In the chymotrypsin-like serine proteinases this conformation is maintained by interactions of the carbonyl of this peptide bond with other groups in the structure. The carbonyl of the corresponding residue

in the 3C proteinases (Pro 169 in HAV and Ala 144 in PV) does not make any interactions in the crystal structures. In the crystal structures of mutants of the nucleophilic cysteine of the HAV 3C proteinase this peptide bond is indeed flipped and the oxyanion hole has collapsed to a lower energy main-chain conformation (Allaire *et al.*, 1994). It appears that the presence of the nucleophilic sulfur atom itself is required to maintain the proper conformation of the oxyanion hole in the 3C proteinases. We believe that it is a negative charge on the nucleophilic sulfur that orients the peptide bonds of the oxyanion hole and take this as partial evidence for a mechanism involving a thiolate–imidazolium ion pair. Direct experimental evidence for the protonation state of the residues in the active site and the mechanism of the 3C proteinases is, however, lacking.

An aspartate or glutamate residue, which is in an equivalent position to the third member of the catalytic triad of the chymotrypsin-like serine proteinases, is present and conserved throughout the 3C proteinases (Gorbalenya *et al.*, 1988; Ryan and Flint, 1997). However, the interaction typical for the third member of the catalytic triad is not observed.

In HAV 3C the side-chain of Asp 84 points away from the imidazole of the general acid–base His 44. It is locked in interactions with other regions of the structure (Bergmann *et al.*, 1997). The position of the carboxylate of the third member of a catalytic triad is taken up by a water molecule in HAV 3C, which forms a hydrogen bond to the N<sup>δ</sup> of His 44 (Fig. 4a). This water molecule, the imidazole of His 44 and the nucleophilic S<sup>γ</sup> atom of Cys 172 of HAV 3C are in a common plane. Perpendicular to this plane and 3.0Å above it, is the side-chain of Tyr 143, which is located in the antiparallel  $\beta$ -ribbon of HAV 3C. Mutational studies have shown that Tyr 143 is important for the catalytic activity of HAV 3C. The side-chain of Tyr 143 does not form a hydrogen bond in the crystal structure of HAV 3C. It is perpendicular to the plane of the imidazole and it is 3.5Å away from the water molecule. We believe the side-chain of Tyr 143 is deprotonated and negatively charged in the structure of the HAV 3C proteinase. Presumably, an electrostatic interaction between His 44 and Tyr 143 helps to maintain the side-chain conformation of His 44 with the imidazole in the same plane as the nucleophilic S<sup>γ</sup> atom of Cys 172 and helps to stabilize a positive charge on the His 44 imidazole.

The conserved glutamate, which is present in the position that corresponds to the third member of a putative catalytic triad in polio virus 3C (Glu 71), does interact with the imidazole of the general acid–base catalyst His 40 (Mosimann *et al.*, 1997). The interaction is, however, unusual. Accepting a hydrogen bond from the N<sup>δ</sup> atom of His 44 is the *anti* lone electron pair of the carboxylate of Glu 71. This is similar to the structure of the 3C proteinase from rhinovirus (Matthews *et al.*, 1997).

Thus, the conserved features in the proteolytic active site of the picornaviral

3C proteinase are a cysteine–histidine dyad and an oxyanion hole that resembles that of the chymotrypsin-like serine proteinases remarkably well. Additional chemical groups have been shown to be important by mutagenesis experiments and are making interactions with mostly the histidine general acid–base catalyst. Their function is probably to maintain the orientation of the active site residues and to provide a specific electrostatic environment.

Chymotrypsin-like serine proteinases bind specific substrates in a canonical mode and with a specific conformation of the bound peptide substrate (Read and James, 1986; Bode and Huber, 1992). We now have evidence from structures of enzyme inhibitor complexes that the 3C proteinases bind peptide substrates in a similar conformation (Bergmann and James, manuscript in preparation). Furthermore, the specific recognition of the cleavage sites within the viral polyprotein by the 3C proteinases can be rationalized if one assumes a similar binding mode.

Chymotrypsin-like proteinases specifically bind 4–5 residues that precede the scissile bond and 2–3 residues that follow it in the sequence (i.e.,  $P_3$  to  $P'_3$ ). The residues from  $P_5$  to  $P_2$  of the substrate are usually in a  $\beta$ -strand conformation. The  $P_1$  residue adopts a main-chain conformation that corresponds to a tight  $3_{10}$  helix. This places the carbonyl of the scissile peptide bond into the oxyanion hole. The  $P'_1$  and  $P'_2$  residues are usually also in a  $\beta$ -conformation. This main chain conformation of a peptide substrate orients at least some of the peptide side-chains into specificity pockets, which are formed by the surface of the enzyme. While interactions between the enzyme and the main-chain of a bound substrate contribute significantly to the binding of a substrate, most of the specificity is provided by the interactions of the peptide side-chains in the specificity pockets of the enzyme (Fig. 5, see color plate).

The minimum size for a good substrate of a 3C proteinase is a hexapeptide with the specific  $P_4$  to  $P'_2$  residues. Sequence preferences for certain residues that distinguish the 3C cleavage sites in the picornaviral polyprotein can also be found for the  $P_4$  to  $P'_2$  residues (Pallai *et al.*, 1989; Weidner and Dunn, 1991; Bergmann, 1998 and references therein). Figure 5 shows a model of a hexapeptide substrate with the sequence of the primary cleavage site of the HAV polyprotein in the active site of HAV 3C. With the exception of the two hydrogen bonds that the carbonyl of the scissile peptide bond makes in the oxyanion hole, the main-chain interactions between the HAV 3C proteinases and the bound peptide are  $\beta$ -sheet interactions. In HAV 3C the substrate residues from  $P_5$  to  $P_2$  form an antiparallel  $\beta$ -sheet with  $\beta$ -strand eII. This form of substrate binding is common to all chymotrypsin-like proteinases. In HAV 3C there is also a parallel  $\beta$ -sheet interaction between the  $P_4$  to  $P_2$  residues of the substrate and enzyme residue from the extension of  $\beta$ -strand bII. This extension is not present in the smaller enteroviral 3C. Third, the  $P'_1$  and  $P'_2$  residues of the substrate of HAV 3C form an antiparallel  $\beta$ -interaction with the residues of the

enzyme, which form the  $\beta$ -bulge in strand bI of the enzyme. The conformation of  $\beta$ -strand bI is different in polio 3C and presumably forces a different main-chain conformation on the substrate. This could explain the unique preference of polio 3C for a P<sub>1</sub>' glycine residue.

All 3C proteinases share the preference for a glutamine residue in the P<sub>1</sub> position of a substrate. Sequence preferences for the other residues of a substrate from P<sub>4</sub> to P<sub>2</sub>' differ among the enzymes from the different genera. The major determinant of the primary specificity is a conserved histidine residue which is positioned inside the S<sub>1</sub> pocket of the enzymes (Fig. 4). This histidine residue (191 in HAV and 161 in hnPV) is conserved throughout the 3C genes of all Picornaviruses (Ryan and Flint, 1997). Models of substrates bound to 3C proteinases agree that this histidine residue donates a hydrogen bond to the carbonyl oxygen atom of the side-chain of a glutamine residue in the S<sub>1</sub> pocket. The environment of this histidine, which must contribute to the specific distinction between glutamine and glutamate, is, however, different in the crystal structures of HAV 3C and PV 3C. In HAV 3C His191 interacts, via buried water, with the side-chain of Glu 132. Bergmann *et al.* (1997) suggest that the deprotonation of this Glu 132, which is buried in the interior of the C-terminal domain, would be energetically expensive and unfavorable. Because the two residues interact, the protonation of His191 would also be unfavorable. In the enteroviral and rhinoviral enzymes a buried tyrosine residue performs a similar function (Mosimann *et al.*, 1997).

Comparison of the sequences of the natural cleavage sites of the HAV polyprotein reveals distinct sequence preferences for the residues in the P<sub>4</sub>, P<sub>2</sub>, and P<sub>1</sub> position of a 3C substrate (Bergmann, 1998). All natural cleavage sites in the HAV polyprotein have large hydrophobic residues (preferably Leu or Ile) in P<sub>4</sub>, serine or threonine in P<sub>2</sub>, and glutamine in P<sub>1</sub>. The 3C proteinase from poliovirus has a sequence preference for a small, hydrophobic residue in P<sub>4</sub>, glutamine in P<sub>1</sub>, and glycine in the P<sub>1</sub>' position of a peptide substrate.

Bergmann *et al.* (1997) suggest that His145 in HAV 3C can form a hydrogen bond to a serine or threonine residue in P<sub>2</sub> and is therefore responsible for the P<sub>2</sub> specificity (Fig. 5). The  $\beta$ -ribbon-contributing His145 in HAV 3C is shorter in the poliovirus enzyme so there is no equivalent residue in polio 3C. This correlates well with the fact that polio 3C does not show a sequence preference for a P<sub>2</sub> residue.

The hydrophobic S<sub>4</sub> pocket of the 3C proteinases is a cleft formed by  $\beta$ -strands eII and fII and the  $\beta$ -ribbon formed by the extension of  $\beta$ -strands bII and cII (Fig. 5). It is quite large in the HAV 3C proteinase. In polio 3C several of the hydrophobic residues that form this pocket are substituted by larger ones (e.g., Ala141 and Val200 in HAV 3C correspond to Leu125 and Phe170 in polio 3C). Therefore, the S<sub>4</sub> pocket in the polio 3C is smaller and polio 3C prefers smaller, hydrophobic side-chains in P<sub>4</sub>.



The larger 3C proteinase from HAV thus forms more extensive interactions with peptide substrates, both main-chain and side-chain. However, it is important to keep in mind that in the enteroviruses the 3CD precursor is a more active proteinase and is required for some of the cleavages of the polyprotein (Ypma-Wong *et al.*, 1988). Bergmann *et al.* (1997) suggested a model for the interactions of 3C with the 3D and 3AB domains in a larger precursor. In this model the 3D part of a 3CD precursor would be in a position to interact with the residues in the P<sub>2</sub> and P<sub>3</sub> position of a substrate and could influence the proteolytic activity (top left of the 3C molecule in Figs. 2a and 3a).

While the models of substrate binding to the 3C proteinases allow one to rationalize how the 3C proteinases recognize the specific cleavage sites within the polyprotein, it is not possible to explain the preference of some of the cleavage sites over others during the polyprotein processing. Presumably other factors besides the subsite specificity, such as the accessibility within the folded polyprotein, must play a part in the determination of the sequential polyprotein processing.

### 3. Inhibition

Inhibitors of the 3C proteinases usually combine a chemical functionality that covalently attaches to the nucleophilic thiol in the active site, with other groups which target some of the specific interactions between the proteinases and its substrates. Typical cysteine proteinase inhibitors such as iodoacetamide, N-ethylmaleimide, epoxides, and aldehydes are also effective against the 3C proteinases (Malcolm, 1995). More promising inhibitors are the fluoromethylketones and  $\gamma$ -aminovinylsulfones (Rasnik, 1996). Some of the best inhibitors available to date combine the latter functionalities with a peptidic specificity address which mimics the natural peptide specificity. A tetrapeptide fluoromethylketone inhibitor with the sequence Acetyl-Leu-Ala-Ala-Gln-FMK has been shown to be an effective inhibitor of the HAV 3C proteinase *in vitro* and *in vivo* (Morris *et al.*, 1997). It covalently attaches to the HAV 3C proteinase and is capable of reducing the production of progeny virus in infected cells.

Other functionalities that are now being investigated as inhibitors of the chymotrypsin-like cysteine proteinases include  $\alpha\beta$ -unsaturated carboxylesters,  $\beta$ - and  $\gamma$ -lactones, lactams, isatins (2,3-dioxindoles), and triterpene sulfates (Skiles and McNeil, 1990; Brill *et al.*, 1996). A cocrystal structure of the rhinovirus 3C proteinase with an isatin analog inhibitor in the active site shows that some of these compounds also covalently attach to the active site thiol and mimic the P<sub>1</sub> specificity determinant of a natural substrate (Webber *et al.*, 1996).

While the details of the specific enzyme substrate interactions gleaned

from the crystal structures of 3C proteinases provide valuable information for the design of effective inhibitors, there is little experimental evidence for the mechanism of the chymotrypsin-like cysteine proteinases. This kind of information would, however, be of great value in identifying potential chemical functionalities and inhibitors.

## B. THE ENTERO- AND RHINOVIRAL 2A PROTEINASE

The primary cleavage at the N-terminus of the 2A gene product that separates the structural and nonstructural proteins in the entero- and rhinoviruses is performed by the 2A gene product. Analysis of the sequence, mutational studies, and model-building studies have shown that the entero- and rhinoviral 2A proteinase is also a chymotrypsin-like cysteine proteinase but distinct from and only distantly related to 3C. The 2A proteinase is a smaller enzyme of 142 residues. It has been suggested, based on the results of mutational studies and structural models, that the active site of the 2A proteinases contains a catalytic triad of Cys106, His18, and Asp35, which more closely resembles that of the serine proteinases (Sommergruber *et al.*, 1989; Hellen *et al.*, 1991). The sequence alignments also seem to indicate a closer relationship of 2A to the small, bacterial serine proteinases (Sommergruber *et al.*, 1997).

The 2A proteinase has less stringent specificity requirements than the 3C proteinases (Skern *et al.*, 1991). Presumably this reflects the fact that its major function is to perform an intramolecular cleavage at its own N-terminus. Indeed, it has been found that amino acid changes in a substrate affect a *trans* activity but not the intramolecular *cis* activity (Hellen *et al.*, 1992).

The 2A proteinase is a zinc protein and the tightly bound zinc ion presumably plays a structural role (Sommergruber *et al.*, 1994b; Voss *et al.*, 1995). In the model of Sommergruber *et al.* (1997) the N-terminal  $\beta$ -barrel has fewer strands and presumably the zinc ion is therefore needed to stabilize the N-terminal domain of the small 2A proteinase.

A second function of the entero- and rhinoviral 2A proteinase is the specific cleavage of one of the proteins of the eukaryotic CAP-binding complex, eIF4G (Sommergruber *et al.*, 1994a; Haghghat *et al.*, 1996). This results in inhibition of the translation of capped, cellular mRNAs and preferential translation of the viral RNA.

The 2A gene product of entero- and rhinoviruses also has functions in addition to its proteolytic activity (Belsham and Sonnenberg, 1996). There is evidence that 2A forms a complex with other viral proteins and is involved in viral RNA translation and other aspects of viral replication (Molla *et al.*, 1993; Lu *et al.*, 1995; Cuconati *et al.*, 1998).

There is little experimental evidence for the catalytic mechanism of the 2A proteinase. The chemical functionalities which provide good inhibitors of the 3C proteinases, such as the fluoromethylketones, are also effective against other chymotrypsin-like cysteine proteinases. Because the specificity requirements of the 2A proteinases are less stringent, the design of specific inhibitors against this class of enzymes could be more difficult.

### C. THE L PROTEINASE OF THE APHTHOVIRUSES

The aphthoviruses have another distinct proteolytic activity besides the 3C proteinase. The gene coding for the L proteinase is located at the N-terminus of the polyprotein and precedes the structural proteins (Ryan and Flint, 1997). The L proteinase cleaves its own C-terminus (Strebel and Beck, 1986). *In vitro* this cleavage can occur in *cis* and *trans* (Medina *et al.*, 1993; Cao *et al.*, 1995). The aphthoviral L proteinase also cleaves the cellular eIF4G and thus causes inhibition of the translation of capped, cellular mRNAs. This function of the aphthoviral L proteinase is similar to the one performed by the entero- and rhinoviral 2A proteinases. However, the cleavage of eIF4G by the L proteinase occurs in a different position (Kirchweger *et al.*, 1994). In spite of these functions the L proteinase is not essential for the replication of the virus (Piccone *et al.*, 1995a).

The L proteinase is a cysteine proteinase and analysis of the sequence suggests that it belongs to the family of papain-like proteinases (Gorbalenya *et al.*, 1991). The enzyme is present in two forms which differ by size and originate from two different initiation codons in the viral genome. The Lb proteinase of foot-and-mouth disease virus (FMDV) consists of 173 residues and the Lab proteinase is 28 residues longer. Sequence analysis, site-directed mutagenesis, and modeling studies identified the nucleophile, the general acid–base catalyst, and the third member of the catalytic triad as Cys51, His148, and Asp164, respectively (Gorbalenya *et al.*, 1991; Piccone *et al.*, 1995b; Roberts and Belsham, 1995; Skern *et al.*, 1998). Sequence alignments also suggest that the side-chain of Asn46 contributes to the oxyanion hole (similar to Gln 19 of papain) (Ryan and Flint, 1997).

The C-terminus of the L proteinase has an extension, compared to the structure of papain, which has been predicted to adopt a helical conformation. Skern *et al.* (1998) suggest that this additional helix plays an important role for the mechanism of the intramolecular cleavage at the C-terminus of the L proteinase.

Crystallization of the Lb proteinase from FMDV has been reported but the crystal structure has not been published (Guarné *et al.*, 1996).

#### IV. CONCLUSIONS AND IMPLICATIONS FOR ANTIVIRAL STRATEGIES

The picornaviral 3C proteinases constitute an ideal target for the rational design of antiviral drugs. There is now a considerable amount of structural information for both enzymes and enzyme–inhibitor complexes. The details of the molecular interactions that are responsible for the specific substrate binding are reasonably well understood. Furthermore, the chymotrypsin-like cysteine proteinases constitute a unique class of enzymes with a distinct substrate specificity and are so far only found in +RNA viruses. Within these viruses the 3C proteinases perform a central and indispensable role during the viral life cycle and 3C proteinase inhibitors have the potential to limit the spread of viral infections (Morris *et al.*, 1997).

Neither the 2A nor the L proteinases are as attractive as targets for antiviral strategies. The activity of the L proteinase is apparently not as critical for viral replication. There is not as much structural information available for the entero-/rhinoviral 2A proteinase. The fact that the 2A proteinase activity is also less stringently specific could make the design of inhibitors more difficult.

While there are many Picornaviruses, and viruses from related families, that cause disease in humans, few of these are considered important targets for the design of antiviral drugs. Rhinoviruses cause at least half of all common colds in humans. But because there are other families of unrelated viruses that cause upper respiratory tract infections, which are essentially indistinguishable, effective drugs against rhinoviruses would only be useful in combination with simple analytical procedures to unambiguously identify rhinoviral infections. Such simple analytical procedures are not available at present (Couch, 1996).

There are safe and effective vaccines available against poliovirus and HAV. As a result of extensive worldwide vaccination, the incidence of poliomyelitis has been decreasing and currently there are realistic efforts underway to eradicate the disease completely. The introduction of an effective vaccine against HAV was very recent and it is too early to predict its effect. Wide-spread vaccination against HAV is currently not planned as hepatitis A is usually not a life-threatening disease. Because co-infections of chronic carriers of hepatitis B, C, and G with HAV appears to be dangerous, the observed increase in chronic infections with other forms of hepatitis may have an influence on future strategies to control hepatitis A.

It would be desirable to have antiviral drugs available against the more severe enteroviral infections. While most of the enteroviral infections are rare, they can have serious consequences. Because enteroviral infections occur infrequently, this is not considered an economically important target.

Several other families of +RNA viruses also carry 3C or 3C-like proteinases

(Wirblich *et al.*, 1995; Martín Alonso *et al.*, 1995; Tibbles *et al.*, 1996). Most notably, the Corona- and Caliciviridae cause upper respiratory tract infections and intestinal infections in humans. The viruses of these families are less well studied than the Picornaviruses but distantly related. The design of 3C proteinase inhibitors would in all likelihood also be useful toward the development of antiviral drugs against the 3C-like proteinases of the viruses of these families.

At present the mechanism by which some +RNA viruses, most notably the enteroviruses, can trigger severe autoimmune diseases are not well understood. It is also questionable whether inhibition of viral replication would prevent the disastrous consequences of the immune response at a later stage of an infection. Therefore, it is not clear whether antiviral drugs would be useful in the prevention of these diseases.

In conclusion, there is a wealth of experimental information available for the best-studied examples of the viruses of the Picornaviridae. This information provides an opportunity to design inhibitors against the viral 3C proteinase. Effective inhibitors of the picornaviral 3C proteinase have the potential to become effective antiviral drugs against human diseases such as the common cold, HAV, enteroviral infections, and diseases caused by related +RNA viruses.

## REFERENCES

- Allaire, M., Chernaia, M. M., Malcolm, B. A., and James, M. N. G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**, 72–76.
- Andino, R., Rickhof, G. E., Achacoso, P. L., and Baltimore, D. (1993). Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *EMBO J.* **12**, 3587–3598.
- Bazan, J. F., and Fletterick, R. J. (1988). Viral cysteine proteinases are homologous to the trypsin-like family of serine proteinases: Structural and functional implications. *Proc. Natl. Acad. Sci. USA* **85**, 7872–7876.
- Belsham, G. J., and Sonnenberg, N. (1996). RNA–protein interactions in regulation of picornavirus RNA translation. *Microbiol. Rev.* **60**, 499–511.
- Bergmann, E. M. (1998). Hepatitis A virus picornain 3C. In “Handbook of Proteolytic Enzymes” (A. D. Barrett, N. J. Rawlings, and F. Woesner, Eds.). Academic Press, London.
- Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A., and James, M. N. G. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: Specific proteinase activity and RNA recognition. *J. Virol.* **71**, 2436–2448.
- Bienz, K., Egger, D., Rasser, Y., and Bossart, W. (1983). Intracellular distribution of poliovirus proteins and the induction of virus-specific cytoplasmic structures. *Virology* **131**, 39–48.
- Bienz, K., Egger, D., Troxler, M., and Pasamontes, L. (1990). Structural organization of poliovirus RNA replication is mediated by viral proteins of the P2 genomic region. *J. Virol.* **64**, 1156–1163.
- Bode, W., and Huber, R. (1992). Natural protein proteinase inhibitors and their interactions with proteinases. *Eur. J. Biochem.* **204**, 433–451.
- Borman, A. M., and Kean, K. M. (1997). Intact eukaryotic initiation factor 4G is required for hepatitis A virus internal initiation of translation. *Virology* **237**, 129–136.
- Brill, B. M., Kati, W. M., Montgomery, D., Karwowski, J. P., Humphrey, P. E., Jackson, M., Clement,

- J. J., Kadam, S., Chen, R. H., and McAlpine, J. B. (1997). Novel triterpene sulfates from *fusarium compactum* using a rhinovirus 3C protease inhibitor screen. *J. Antibiot.* **49**, 541–546.
- Cao, X., Bergman, I. E., Füllkrug, R., and Beck, E. (1995). Functional analysis of the two alternative initiation sites of foot-and-mouth disease virus. *J. Virol.* **69**, 560–563.
- Carthy, C. M., Yang, D., Anderson, D. R., Wilson, J. E., and McManus, B. M. (1997). Myocarditis as systemic disease: New perspectives on pathogenesis. *Clin. Exp. Pharmacol. Physiol.* **24**, 997–1003.
- Chotia, C. (1984). Principles that determine the structure of proteins. *Annu. Rev. Biochem.* **53**, 537–572.
- Couch, R. B. (1996). Rhinoviruses. In “Fields Virology” (B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Channock, J. L. Melnick, T. P. Monath, B. Roizmann, and S. E. Straus, Eds.). Lippincott-Raven, Philadelphia.
- Cuconati, A., Xiang, W., Lahser, F., Pfister, T., and Wimmer, E. (1998). A protein linkage map of the P2 nonstructural proteins of poliovirus. *J. Virol.* **72**, 1297–1307.
- Davis, G. J., Wang, Q. M., Cox, G. A., Johnson, R. B., Wakulchik, M., Datson, C. A., and Villarreal, E. C. (1997). Expression and purification of recombinant rhinovirus 14 3CD proteinase and its comparison to the 3C proteinase. *Arch. Biochem. Biophys.* **346**, 125–130.
- Donnelly, M. L. L., Gani, D., Flint, M., Monaghan, S., and Ryan, M. D. (1997). The cleavage activity of aphtho and cardiovirus 2A proteins. *J. Gen. Virol.* **78**, 13–21.
- Dougherty, W. G., and Semler, B. L. (1993). Expression of virus-encoded proteinases: Functional and structural similarities with cellular enzymes. *Microbiol. Rev.* **57**, 781–822.
- Fujinaga, M., Delbaere, L. T. J., Brayer, G., and James, M. N. G. (1987). Refined crystal structure of  $\alpha$ -lytic protease at 1.7 Å resolution. *J. Mol. Biol.* **184**, 479–502.
- Gamarnik, A. V., and Andino, R. (1997). Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. *RNA* **3**, 882–892.
- Gorbalenya, A. E., and Snijder, E. J. (1996). Viral cysteine proteinases. *Perspect. Drug Disc. Design* **6**, 48–86.
- Gorbalenya, A. E., Blinov, V. M., and Donchenko, A. P. (1986). Poliovirus-encoded proteinase 3C: A possible evolutionary link between cellular serine and cysteine proteinase families. *FEBS Lett.* **194**, 253–257.
- Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M., and Koonin, E. V. (1989). Cysteine proteinases of positive strand RNA viruses and chymotrypsin-like serine proteinases: A distinct protein superfamily with a common structural fold. *FEBS Lett.* **243**, 103–114.
- Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1991). Putative papain-related thiol protease of positive strand RNA viruses: Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett.* **288**, 201–205.
- Guarné, A., Kirchweiger, R., Verdager, R., Liebig, H. D., Blaas, D., Skern, T., and Fita, I. (1996). Crystallization and preliminary X-ray diffraction studies of the Lb proteinase of foot-and-mouth disease virus. *Prot. Sci.* **5**, 1931–1933.
- Haghighat, A., Svitkin, Y., Novoa, I., Küchler, E., Skern, T., and Sonnenberg, N. (1996). The eIF4G-eIF4E complex is the target for direct cleavage by the rhinovirus 2A proteinase. *J. Virol.* **70**, 8444–8450.
- Hämmerle, T., Molla, A., and Wimmer, E. (1992). Mutational analysis of the proposed FG loop of poliovirus proteinase 3C identified amino acids that are necessary for 3CD cleavage and might be determinants of a function distinct from proteolytic activity. *J. Virol.* **66**, 6028–6034.
- Hanecak, R., Semler, B. L., Ariga, H., Anderson, C. W., and Wimmer, E. (1984). Expression of a cloned gene segment of poliovirus in *E. coli*: Evidence for autocatalytic production of the viral proteinase. *Cell* **37**, 1063–1073.
- Harmon, S. A., Updike, W., Xi-Ju, J., Summers, D. F., and Ehrenfeld, E. (1992). Polyprotein

- processing in *cis* and in *trans* by hepatitis A virus 3C protease cloned and expressed in *E. coli*. *J. Virol.* **66**, 5242–5247.
- Harris, K. S., Xiang, W., Alexander, L. S., Lane, W. S., Paul, A. V., and Wimmer, E. (1994). Interactions of poliovirus polypeptide 3CD<sup>Pro</sup> with the 5' and 3' termini of the poliovirus genome. *J. Biol. Chem.* **269**, 27004–27014.
- Hellen, C. U. T., Fache, M., Kräusslich, H. G., Lee, C., and Wimmer, E. (1991). Characterization of poliovirus 2A proteinase by mutational analysis: Residues required for autocatalytic activity are essential for induction of eukaryotic initiation factor 4F polypeptide p220. *J. Virol.* **65**, 4226–4231.
- Hellen, C. U. T., Lee, C., and Wimmer, E. (1992). Determinants of substrate recognition by poliovirus 2A proteinase. *J. Virol.* **66**, 3330–3338.
- Hollinger, F. B., and Ticehurst, J. R. (1996). Hepatitis A virus. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, B. Roizmann, and S. E. Straus, Eds.). Lippincott-Raven, Philadelphia.
- James, M. N. G. (1993). Convergence of active-centre geometries among the proteolytic enzymes. In "Proteolysis and Protein Turnover" (J. S. Bond and A. J. Barrett, Eds.). Portland Press, London.
- Jia, X.-Y., Summers, D. F., and Ehrenfeld, E. (1993). Primary cleavage of the HAV capsid protein precursor in the middle of the proposed 2A coding region. *Virology* **193**, 515–519.
- Kay, J., and Dunn, B. M. (1990). Viral proteinases: weakness in strength. *Biochim. Biophys. Acta* **1048**, 1–18.
- Kirchwegger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H. D., Sommergruber, W., Sobrino, F., Hohenadl, C., Blaas, D., Rhoads, R. E., and Skern, T. (1994). Foot-and-mouth disease virus leader proteinase: Purification of the Lb form and determination of its cleavage site on eIF4γ. *J. Virol.* **68**, 5677–5684.
- Kräusslich, H.-G., and Wimmer, E. (1988). Viral proteinases. *Ann. Rev. Biochem.* **57**, 701–754.
- Kusov, Y. Y., and Gauss-Müller, V. (1997). *In vitro* RNA binding of the hepatitis A virus proteinase 3C (HAV 3C<sup>Pro</sup>) to secondary structure elements within the 5' terminus of the HAV genome. *RNA* **3**, 291–302.
- Lamphear, B. J., Yan, R., Yang, F., Waters, D., Liebig, H.-D., Klump, H., Küchler, E., Skern, T., and Rhoads, R. E. (1993). Mapping the cleavage site in protein synthesis initiation factor eIF-4γ of the 2A proteases from human coxsackie virus and rhinovirus. *J. Biol. Chem.* **268**, 19200–19203.
- Leong, L. E. C., Walker, P. A., and Porter, A. G. (1993). Human rhinovirus 14 protease 3C (3C<sup>Pro</sup>) binds specifically to the 5'-noncoding region of the viral RNA. *J. Biol. Chem.* **268**, 25735–25739.
- Long, L. A., Orr, D. C., Cameron, J. M., Dunn, B. M., and Kay, J. (1989). A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase. *FEBS Lett.* **258**, 75–78.
- Lu, H. H., Li, X., Cuconati, A., and Wimmer, E. (1995). Analysis of picornavirus 2A (pro) proteins: Separation of proteinase from translation and replication functions. *J. Virol.* **69**, 7445–7452.
- Malcolm, B. A. (1995). The picornaviral 3C proteinases: Cysteine nucleophiles in serine proteinase folds. *Prot. Sci.* **4**, 1439–1445.
- Martin Alonso, J. M., Casais, R., Boga, J. A., and Parra, F. (1996). Processing of rabbit hemorrhagic disease virus polyprotein. *J. Virol.* **70**, 1261–1265.
- Martin, A., Escriou, N., Chao, S. F., Girard, M., Lemon, S. M., and Wychowski, C. (1995). Identification and site-directed mutagenesis of the primary (2A/2B) cleavage site of the hepatitis A virus polyprotein: Functional impact on the infectivity of HAV RNA transcripts. *Virology* **213**, 213–222.
- Mathews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L., and Worland, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site and means for cleaving precursor polyprotein. *Cell* **77**, 761–771.

- Medina, M., Domingo, E., Brangwun, J. K., and Belsham, G. J. (1993). The two species of the foot-and-mouth disease virus leader protein expressed individually, exhibit the same activities. *Virology* 194, 355–359.
- Melnick, J. L. (1996). Enteroviruses: Polioviruses coxsackie viruses, echoviruses, and newer enteroviruses. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Channock, J. L. Melnick, T. P. Monath, B. Roizmann, and S. E. Straus, Eds.). Lippincott-Raven, Philadelphia.
- Miller, S. D., Vanderlugt, C. L., Smith-Begolka, W., Pao, W., Yauch, R. L., Neville, K. L., Katz-Levy, Y., Carrizosa, A., and Kim, B. S. (1997). Persistent infection with Theiler's virus leads to CNS autoimmunity via epitope spreading. *Nat. Med.* 3, 1133–1136.
- Molla, A., Paul, A. V., Schmid, M., Jang, S. K., and Wimmer, E. (1993). Studies on dicistronic polioviruses implicate viral proteinase 2A<sup>Pro</sup> in RNA replication. *Virology* 196, 739–747.
- Morris, T. S., Frommann, S., Shechosky, S., Lowe, C., Lall, M. S., Gauss-Müller, V., Purcell, R. H., Emerson, S. U., Vederas, J. C., and Malcolm, B. A. (1997). *In vitro* and *ex vivo* inhibition of hepatitis A virus 3C proteinase by a peptidyl monofluoromethyl ketone. *Bioorg. Med. Chem.* 5, 797–807.
- Mosimann, S. C., Chernaia, M. M., Sia, S., Plotch, S., and James, M. N. G. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* 273, 1032–1047.
- Nicklin, M. J., Harris, K. S., Pallai, P. V., and Wimmer, E. (1988). Poliovirus proteinase 3C: Large-scale expression, purification and specific cleavage activity on natural and synthetic substrates *in vitro*. *J. Virol.* 62, 4586–4593.
- Pallai, P. V., Burkhardt, F., Shoog, M., Schreiner, K., Bax, P., Cohen, K. A., Hansen, G., Palladino, D. E., Harris, K. S., Nicklin, M. J., and Wimmer, E. (1989). Cleavage of synthetic peptides by purified poliovirus 3C proteinase. *J. Biol. Chem.* 264, 9738–9741.
- Palmenberg, A. C. (1990). Proteolytic processing of picornaviral polyprotein. *Annu. Rev. Microbiol.* 44, 602–623.
- Palmenberg, A. C., and Rueckert, R. R. (1982). Evidence for intramolecular self-cleavage of picornaviral replicase precursors. *J. Virol.* 41, 244–249.
- Palmenberg, A. C., Parks, G. D., Hall, D. J., Ingraham, R. H., Seng, T. W., and Pallai P. V. (1992). Proteolytic processing of the cardioviral P2 region: Primary 2A/2B cleavage in clone derived precursors. *Virology* 190, 754–762.
- Parsley, T. B., Towner, J. S., Blyn, L. B., Ehrenfeld, E., and Semler, B. L. (1997). Poly (rC) binding protein 2 forms a ternary complex with the 5'-terminal sequences of poliovirus RNA and the viral 3CD proteinase. *RNA* 3, 1124–1134.
- Perona, J. J., and Craik, C. S. (1995). Structural basis of substrate specificity in the serine proteinases. *Prot. Sci.* 4, 337–360.
- Piccone, M. E., Rieder, E., Mason, P. W., and Grubmann, M. J. (1995a). The foot-and-mouth disease leader proteinase gene is not required for viral replication. *J. Virol.* 69, 5376–5382.
- Piccone, M. F., Zellner, M., Kumosinski, T. F., Mason, P. W., and Grubman, M. J. (1995b). Identification of the active-site residues of the L proteinase of foot-and-mouth disease virus. *J. Virol.* 69, 4950–4956.
- Porter, A. G. (1993). Picornavirus nonstructural proteins: Emerging roles in virus replication and inhibition of host cell functions. *J. Virol.* 67, 6917–6921.
- Rasnick, D. (1996). Small synthetic inhibitors of cysteine proteinases. *Perspect. Drug Disc. Design* 6, 47–63.
- Read, R. J., and James, M. N. G. (1986). Introduction to the Protein Inhibitors: X-ray Crystallography. In "Proteinase Inhibitors" (A. J. Barrett and G. Salvesen, Eds.). Elsevier, Amsterdam.
- Roberts, P. J., and Belsham, G. J. (1995). Identification of critical amino acids within the foot-and-mouth disease virus leader protein, a cysteine protease. *Virology* 213, 140–146.
- Rueckert, R. R. (1996). Picornaviridae: The Viruses and their Replication. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Channock, J. L. Melnick, T. P. Monath, B. Roizmann, and S. E. Straus, Eds.). Lippincott-Raven, Philadelphia.



- Ryan, M. D., and Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *J. Gen. Virol.* **78**, 699–723.
- Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Schultheiss, T., Kusov, Y. Y., and Gauss-Müller, V. (1994). Proteinase 3C of hepatitis A virus (HAV) cleaves the HAV polyprotein P2-P3 at all sites including VP1/2A and 2A/2B. *Virology* **198**, 275–281.
- Schultheiss, T., Emerson, S. U., Purcell, R. H., and Gauss-Müller, V. (1995a). Polyprotein processing in echovirus 22—A first assessment. *Biochem. Biophys. Res. Commun.* **219**, 1120–1127.
- Schultheiss, T., Sommergruber, W., Kusov, Y. Y., and Gauss-Müller, V. (1995b). Cleavage specificity of purified recombinant hepatitis A virus 3C proteinase on natural substrates. *J. Virol.* **69**, 1727–1733.
- Skern, T., Fita, I., and Guarné, A. (1998). A structural model of picornavirus leader proteinase based on papain and bleomycin hydrolase. *J. Gen. Virol.* **79**, 301–307.
- Skiles, J. W., and McNeil, D. (1990). Spiro indolinone  $\beta$ -lactams, inhibitors of poliovirus and rhinovirus 3C-proteinases. *Tetrahedr. Lett.* **31**, 7277–7280.
- Sommergruber, W., Zorn, M., Blaas, D., Fessel, F., Volkmann, P., Mauser-Fogy, I., Pallai, P., Merluzzi, V., Matteo, M., Skern, T., and Kuchler, E. (1989). Polypeptide 2A of human rhinovirus type 2: Identification as a proteinase and characterization by mutational analysis. *Virology* **169**, 68–77.
- Sommergruber, W., Ahorn, H., Klump, H., Zoephel, A., Fessel, F., Blaas, D., Kuchler, E., Liebig, H.-D., and Skern, T. (1994a). 2A proteinases of coxsackie- and rhinovirus cleave peptides derived from eIF-4 $\gamma$  via a common recognition motif. *Virology* **198**, 741–745.
- Sommergruber, W., Casari, G., Fessel, F., Seipelt, J., and Skern, T. (1994b). The 2A proteinase of human rhinovirus is a zinc containing enzyme. *Virology* **204**, 815–818.
- Sommergruber, W., Seipelt, J., Fessel, F., Skern, T., Liebig, H.-D., and Casari, G. (1997). Mutational analyses support a model for the HRV2 2A proteinase. *Virology* **234**, 203–214.
- Steinmann, L., and Conlon, P. (1997). Viral damage and the breakdown of self-tolerance. *Nature Med.* **3**, 1085–1087.
- Storer, A. C., and Ménard, R. (1994). Catalytic mechanism in papain family of cysteine peptidases. *Meth. Enzymol.* **244**, 486–500.
- Strebel, K., and Beck, E. (1986). A second proteinase of foot-and-mouth disease virus. *J. Virol.* **58**, 893–899.
- Teterina, N. L., Bienz, K., Egger, D., Gorbalenya, A. E., and Ehrenfeld, E. (1997a). Induction of intracellular membrane rearrangements by HAV proteins 2C and 2BC. *Virology* **237**, 66–77.
- Teterina, N. L., Gorbalenya, A. E., Egger, D., Bienz, K., and Ehrenfeld, E. (1997b). Poliovirus 2C protein determinants of membrane binding and rearrangements in mammalian cells. *J. Virol.* **71**, 8962–8972.
- Tibbles, K. W., Brierley, I., Cavanagh, D., and Brown, T. D. K. (1996). Characterization *in vitro* of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the Coronavirus avian infectious bronchitis virus. *J. Virol.* **70**, 1923–1930.
- Voss, T., Meyer, R., and Sommergruber, W. (1995). Spectroscopic characterization of rhinoviral protease 2A: Zn is essential for structural integrity. *Prot. Sci.* **4**, 2526–2531.
- Walker, P. A., Leong, L. E. C., and Porter, A. G. (1995). Sequence and structural determinants of the interaction between the 5'-noncoding region of picornavirus RNA and rhinovirus protease 3C. *J. Biol. Chem.* **270**, 14510–14516.
- Webber, S. E., Tikhe, J., Worland, S. T., Fuhrmann, S. A., Hendrickson, T. F., Matthews, D. A., Love, R. A., Patick, A. K., Meador, J. W., Ferre, P. A., Brown, E. L., Delisle, D. M., Ford, C. E., and Binford, S. L. (1996). Design synthesis and evaluation of nonpeptide inhibitors of human rhinovirus 3C proteinase. *J. Med. Chem.* **39**, 5072–5882.

- Weidner, J. R., and Dunn, B. M. (1991). Development of synthetic peptide substrates for the poliovirus 3C proteinase. *Arch. Biochem. Biophys.* **286**, 402–408.
- Whiting, A. K., and Peticolas, W. L. (1994). Details of the acyl-enzyme intermediate and the oxyanion hole in serine protease catalysis. *Biochemistry* **33**, 552–561.
- Wimmer, E. (1982). Genome linked proteins of viruses. *Cell* **28**, 199–201.
- Wimmer, E., Hellen, C. U. T., and Cao, X. (1993). Genetics of poliovirus. *Ann. Rev. Genet.* **27**, 353–436.
- Wirblich, C., Sibilia, M., Boniotti, M. B., Rossi, C., Thiel, H.-J., and Meyers, G. (1995). 3C-like protease of rabbit hemorrhagic disease virus: identification of cleavage sites in the ORF1 polyprotein and analysis of cleavage specificity. *J. Virol.* **69**, 7159–7169.
- Xiang, W. S., Harris, K. S., Alexander, L., and Wimmer, E. (1995). Interaction between the 5'-terminal cloverleaf and 3AB/3CD<sup>pro</sup> of poliovirus is essential for RNA application. *J. Virol.* **69**, 3658–3667.
- Yalamanchili, D., Weidman, K., and Dasgupta, A. (1997). Cleavage of transcriptional activator Oct-1 by poliovirus encoded protease 3C pro. *Virology* **239**, 176–185.
- Ypma-Wong, M. F., Dewalt, P. G., Johnson, V. H., Lamb, J. G., and Semler, B. L. (1988). Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology*, **166**, 265–270.