

## Virus-encoded proteinases of the picornavirus super-group

Martin D. Ryan and Mike Flint

Department of Biochemistry, University of St Andrews, Irvine Building, North Street, St Andrews KY16 9AL, UK

### Introduction

Similarities in genome organization and replication strategy permit viruses to be organized into 'super-groups' – often including both animal and plant viruses. This article is the first in a planned series of reviews on virus-encoded proteinases of the major super-groups; in it we discuss the proteinases of the picornavirus supergroup (picorna-, como-, nepo-, poty- and bymoviruses). In forthcoming reviews the proteinases of the flavivirus group (flavi-, pesti-, hepatitis C and G viruses), alphavirus supergroup (alpha-, caliciviruses and a wide range of plant viruses), corona- and arteriviruses, retroviruses and DNA viruses will be discussed. Although several reviews have been published on virus-encoded proteinases (Wellink & van Kammen, 1988; Hellen *et al.*, 1989; Kay & Dunn, 1990; Palmenberg, 1990; Dougherty & Semler, 1993), we hope to provide the reader with an up-date to this rapidly advancing area of virology, and, give a slightly more molecular and evolutionary 'spin' to the content.

The understanding of virus-encoded proteinases was given fresh impetus by a series of publications detailing the alignment of such sequences with those of cellular proteinases. Work by A. E. Gorbalenya and colleagues and G. F. Bazan and R. J. Fletterick gave entirely novel and important new insights into the structural, mechanistic and evolutionary relationships amongst virus-encoded proteinases and with cellular proteinases. The first part of this review gives a brief outline of the catalytic mechanisms and the types of residues found in the active sites of each type of the cellular proteinases to assist the reader in interpretation of the sequence alignments – particularly conserved residues in regions thought to form the active sites of virus-encoded proteinases. Due to the large number of sequences available for certain virus proteinases it is not practicable to present alignments including all sequences. Full sequence alignments are available by e-mail (martin.ryan@st-and.ac.uk) or by anonymous FTP from ftp.st-and.ac.uk/pub/mdr1. Algorithms are available which may predict secondary structural features more accurately from multiple aligned sequences rather than from individual examples (Di Francesco *et al.*, 1996). Although structural data

are available for a very few virus-encoded proteinases, the design of many experiments may be assisted by the comparison of sequence alignments with the structures which are available for cellular proteinases, or related virus-encoded proteinases as they become available.

### Types of proteinases

Cellular proteinases have been organized into four types on the basis of their active site nucleophiles. It is clear, however, that, although only four types of nucleophiles are used, biology has arrived at many different solutions to the problem of forming structures capable of generating these nucleophilic species. Whilst the constraints of stereochemistry impose rigid structural criteria upon the functional groups of residues involved in catalysis, this is certainly not the case for the remainder of the structure. Indeed, the wheat serine carboxypeptidase II shows a folding pattern much more similar to the metalloproteinase carboxypeptidase A than that of a conventional serine proteinase (Liao & Remington, 1990).

Proteinases may be of the exo- or endo- variety: exo-proteinases remove residues from either the N or C termini of proteins (usually in a progressive manner), whereas the endoproteinases cleave at internal sites. The specificity of this reaction is imparted by the substrate binding pocket of the proteinase which interacts with substrate residues flanking the scissile bond. This interaction may lead to an extremely low specificity such as proteinase K ( $-X^{\downarrow}Y-$ ), where both residues are non-specific but X is preferred as an aliphatic, aromatic or other hydrophobic residue, or a much higher specificity.

The accepted system for reference to individual substrate binding sub-sites and residues of the substrate flanking the scissile bond is used here:  $S_1, S_2$  etc. refer to proteinase sub-sites which bind residues of the substrate progressively N-terminal of the scissile bond.  $S_1', S_2'$  etc. refer to proteinase sites binding substrate residues progressively C-terminal of the substrate scissile bond. Residues of the substrate are termed  $P_1, P_2$ , etc. progressively N-terminal of the scissile bond, and  $P_1', P_2'$ , etc. progressively C-terminal of the scissile bond (Schechter & Berger, 1967).

(i) *Serine proteinases.* Sequence alignment and structural studies have defined two major classes of serine proteinase; the 'large' sub-group connotes proteinases with similarities to

Author for correspondence: Martin Ryan.

Fax +44 1334 463400. e-mail martin.ryan@st-and.ac.uk

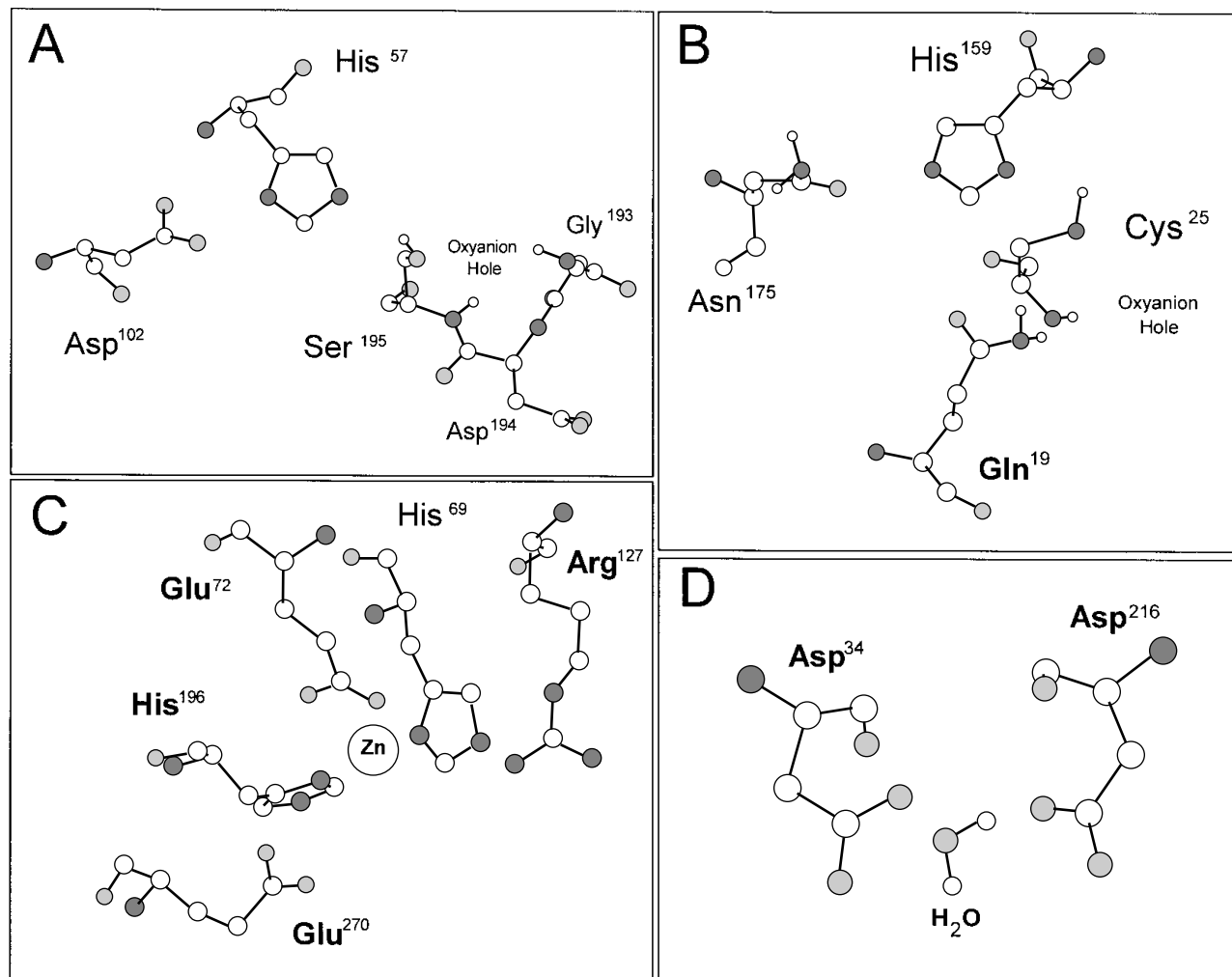


Fig. 1. Proteinase active sites. The catalytic triad of the serine proteinase trypsin is shown (Asp<sup>102</sup>, His<sup>57</sup> and Ser<sup>195</sup>) together with the hydrogen atoms of the nucleophilic hydroxyl group of Ser<sup>195</sup> and the main-chain amino groups of Ser<sup>195</sup> and Gly<sup>193</sup> which form the oxyanion hole (Brookhaven Acc. #3PTN; panel A). The analogous triad is shown for the cystine proteinase papain together with the hydrogen atoms of the nucleophilic sulphhydryl group of Cys<sup>25</sup> and the main-chain amino group of Cys<sup>25</sup> and the  $\epsilon$ -amino group of Gln<sup>19</sup> which together form the oxyanion hole (Brookhaven Acc. #9PAP; panel B). The zinc atom of the metalloproteinase carboxypeptidase A is shown along with ligands His<sup>69</sup>, Glu<sup>72</sup>, His<sup>196</sup>, Arg<sup>127</sup> (oxyanion stabilization) and Glu<sup>270</sup> which deprotonate the Zn-bound water (Brookhaven Acc. #5CPA; panel C). The catalytic aspartyl diad of the acid proteinase chymosin is shown along with the water molecule (Brookhaven Acc. #1CMS; panel D). Carbon atoms are shown represented by open circles whilst oxygen, nitrogen and sulphur atoms are shaded (lighter to darker, respectively) together with the pertinent hydrogen atoms (smaller circles).

trypsin whilst the 'small' sub-group connotes enzymes with similarities to subtilisin. The side-chain of the eponymous serine residue is unusually reactive, the hydroxyl group lying at the end of a 'catalytic triad' composed of an aspartate, a histidine and serine itself (Fig. 1, panel A). The secondary structural folds of the two classes differ somewhat – reflected by the order of the catalytic triad residues in the primary structure (trypsin – His<sup>57</sup>-Asp<sup>102</sup>-Ser<sup>195</sup>; subtilisin – Asp<sup>32</sup>-His<sup>64</sup>-Ser<sup>221</sup>). The structure of serine proteinases has recently been reviewed by Les & Fordham (1996). The catalytic mechanism, common to all classes, has been reviewed elsewhere (Steitz & Schulman, 1982; Warshel *et al.*, 1989) – the implications for virus-encoded proteinases are discussed below.

An important structural feature of the catalytic site is the 'oxyanion hole' – a site containing a series of pre-aligned dipoles which complement changes in the charge distributions during catalysis. The interaction between these dipoles and the peptide bond carbonyl oxygen is thought to promote the formation of the tetrahedral transition state and increase the rate of deacylation of the acyl-enzyme intermediate (Whiting & Peticolas, 1994). In trypsin these dipoles are derived from the main-chain amides of Ser<sup>195</sup> and Gly<sup>193</sup>. The substrate binding pocket is composed of sub-sites binding individual residues of the substrate – interactions which may extend some distance on either side of the scissile bond (four or more residues).

(ii) **Cysteine (thiol) proteinases.** In the thiol proteinases, such as papain, the sulphhydryl group of the active site cysteine shows high nucleophilicity – the sulphur atom forming a thiolate anion/imidazolium couple with His<sup>159</sup> at neutral pH. In papain the amide oxygen of the Asn<sup>175</sup> side-chain is hydrogen bonded to His<sup>159</sup> forming an Asn<sup>175</sup>-His<sup>159</sup>-Cys<sup>25</sup> triad (Fig. 1, panel B), analogous to the Asp-His-Ser triad described above. The Asn<sup>175</sup>-His<sup>159</sup> hydrogen bond is approximately collinear with the C $\gamma$ -C $\beta$  bond of His<sup>159</sup> allowing some rotation of the imidazole ring about this axis without disrupting the bond. Indeed, the function of Asn<sup>175</sup> may be modulation of the rotation of the imidazole ring during catalysis. The electrostatic contribution of the oxyanion pocket in cysteine proteinase catalysis appears to be somewhat less than that in serine proteinase catalysis (Menard *et al.*, 1995). In papain the oxyanion pocket dipoles are derived from the main-chain amide of Cys<sup>25</sup> and the side-chain amide of Gln<sup>19</sup>.

(iii) **Aspartyl (acid) proteinases.** In the acid proteinases, such as chymosin B or renin, a catalytic diad is observed composed of two aspartate residues. Catalysis proceeds by a general acid-general base mechanism in which a water molecule attacks the carbonyl of the scissile peptide bond – the aspartate side-chains assisting in the protonation of the scissile carbonyl oxygen and protonation of the amine leaving group to form the cleaved products (Fig. 1, panel D). In cellular enzymes the aspartates are from each half of a bi-lobed structure, with an approximate diad axis of symmetry passing through the active site centre. In this mechanism at no stage is the substrate covalently bound to the enzyme (as is the case for the proteinases described above).

(iv) **Metallo-proteinases.** It is quite possible that zinc is the catalytically active species in the natural form of all metallo-proteinases, although some such enzymes retain their activity when zinc is replaced by other metals such as cobalt. The active site of carboxypeptidase A contains a zinc atom bound by two histidine and one glutamate side-chains (His<sup>69</sup>, His<sup>196</sup>, Glu<sup>72</sup>; Fig. 1, panel C) and a fourth ligand, water. Catalysis proceeds by nucleophilic attack of the Zn<sup>2+</sup>-bound water on the scissile carbonyl bond. The active site contains a positively charged group (Arg<sup>127</sup>) stabilizing the oxyanion in addition to a general base (Glu<sup>270</sup>) which deprotonates the Zn<sup>2+</sup>-bound water. Like the acid proteinases at no stage during catalysis is the substrate covalently bound to the enzyme.

## Virus proteinases: general concepts

The development of translation systems *in vitro* permitted the first clear biochemical demonstrations that the specific proteolytic processing of RNA virus proteins, suggested some 10 years previously (Summers & Maizel, 1968), could be mediated by a virus-encoded rather than cellular proteinase (Pelham, 1978). Autoproteolysis is just one of many methods RNA viruses have developed to regulate protein biogenesis.

Polyprotein domains containing proteins with distinct, linked functions (capsid proteins, replicative proteins) which are separated co-translationally by autoproteolytic processing may be separated in other virus groups by the generation of sub-genomic mRNAs, or by being encoded on separate genome segments. The 'early'/'late' genetic (transcriptional) regulation of protein expression characteristic of many DNA viruses may be substituted by biochemical regulation: the activity or substrate specificity of virus-encoded proteinases may be altered during the infectious cycle to generate different sub-sets of proteins (functions) from a single precursor form. Indeed, different proteolytic processing pathways could occur simultaneously in different regions of infected cells with different concentrations of *trans*-acting factors – a spatial rather than temporal effect.

As our knowledge of RNA virus proteinases has increased a number of features have become apparent: (i) they may function as discrete proteins or, more commonly, as proteolytic domains of larger forms; (ii) these larger forms themselves may represent alternative processing products; (iii) proteinase activities may be different depending upon which type of larger form of the proteinase domain is present; (iv) their activity or substrate specificity may be modified by binding other virus proteins or RNA; (v) virus-encoded proteinases may cleave host-cell proteins modifying or inhibiting host-cell functions; (vi) their activation may be delayed until a special environment is encountered (i.e. during capsid morphogenesis); and (vii) they may cleave in an intramolecular manner (*in cis*), or in an intermolecular manner (*in trans*) – although the catalytic mechanism of such cleavages is the same.

Processing reactions occurring *in cis* are characteristically rapid and, since the cleavage site and the proteinase are parts of the same molecule, follow zero-order kinetics – the reaction being insensitive to dilution. The *cis* cleavage site may be proximal (either terminus of the proteinase) or distal – in the case of poliovirus 3C proteinase some hundreds of residues distal from the active site nucleophile in the primary structure. Processing reactions occurring *in trans* are characteristically slower, follow second-order kinetics, are sensitive to dilution, are more sensitive to inhibitors and generally show a greater sensitivity to sequence variations flanking the scissile amino acid pair.

In summary, virus-encoded proteinases are responsible for highly specific and regulated proteolysis of other virus protein precursor molecules and may control the biogenesis of different, alternative, functions from the same precursor. Additionally, they may have specific cellular protein targets which, when cleaved, result in a modification of cellular macromolecular processes to promote their own replication cycle. What makes them particularly interesting lies, perhaps, in the following aspects: (i) unlike cellular proteinases, they may show a high degree of regulation by interaction with other virus proteins and/or nucleic acid; (ii) structural and mechanistic aspects – in some cases (see 3C proteinase below)

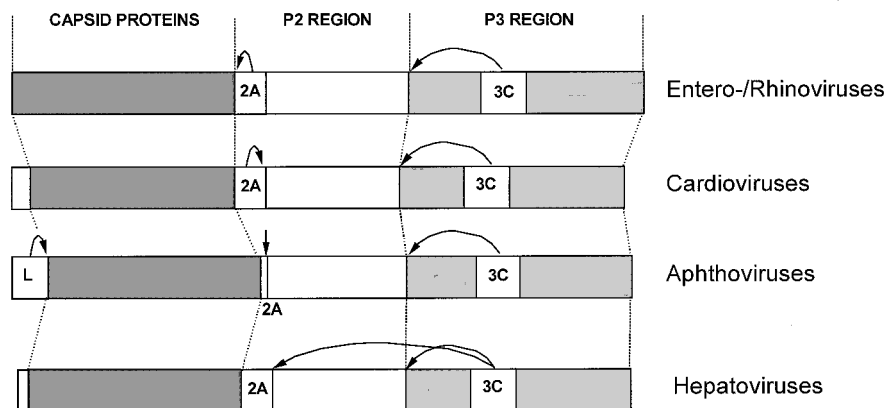


Fig. 2. Picornavirus polyprotein primary cleavages. The single, long open reading frames are shown (boxed areas) for the different picornavirus genera together with the sources of cleavage activity (curved arrows). Unlike the L proteinase of aphthoviruses, cardiovirus L protein is not a proteinase.

a combination of protein architecture and catalytic mechanism not observed in cellular enzymes, and in other cases a molecular mimicry of cellular enzymes.

### Picornavirus proteinases

The picornaviruses encode all of their proteins in a single, long open reading frame. Full-length translation products are not observed, however, due to extremely rapid co-translational, intramolecular (in *cis*), or 'primary' cleavages (Fig. 2). Subsequent 'secondary' processing of the primary cleavage products may occur either in *cis* or in an intermolecular fashion (in *trans*). Such processing of the replicative protein precursor forms (P2 and P3) may be regulated to follow alternative, mutually exclusive, pathways to generate different sub-sets of biochemical functions from the same type of precursor molecule. The genera within this family show differences in their genome organization (reviewed by Stanway, 1990), particularly in relation to their proteolytic activities. Picornavirus polyproteins contain three 'recognizable' types of proteinase – the L, 2A and 3C proteinases ( $L^{\text{pro}}$ ,  $2A^{\text{pro}}$  and  $3C^{\text{pro}}$ , respectively). Additional cleavages (1A/1B and aphtho-/cardiovirus 2A/2B) are mediated in an apparently enzyme-independent manner (see below). For other reviews on picornavirus polyprotein processing the reader is directed to Palmenberg (1990) and Lawson & Semler (1990) – the latter giving a concise account of the development of the field. Residue numbers referred to below are for those of poliovirus proteins or those residues from proteins of other virus groups which align with poliovirus. Picornavirus sequences referred to below may be obtained from the picornavirus WWW pages at the Institute for Animal Health, Pirbright Laboratory, UK (<http://pc0253.avri.bbsrc.ac.uk/index.htm>).

(i) *Aphthovirus L proteinase*. Burroughs *et al.* (1984) showed that aphtho- or foot-and-mouth disease virus (FMDV) polyproteins were processed by more than one type of proteinase.

The aphthoviruses, together with equine rhinoviruses (ERV) 1 and 2, possess a proteinase ( $L^{\text{pro}}$ ) at the N terminus of the polyprotein not found in other genera.  $L^{\text{pro}}$  cleaves co-translationally at its own C terminus (Strebel & Beck, 1986) and exists in at least two forms:  $Lab^{\text{pro}}$  and  $Lb^{\text{pro}}$  derived from initiation of translation at either of two in-frame AUG codons located 84 nucleotides apart (Clarke *et al.*, 1985; Sangar *et al.*, 1987). The  $Lb^{\text{pro}}$  form (and quite possibly the  $Lab^{\text{pro}}$  form) undergoes a post-translational modification by a carboxypeptidase B-like activity producing  $Lb'$  (and possibly  $Lab'$ ; Sangar *et al.*, 1988). Both the  $Lab^{\text{pro}}$  and  $Lb^{\text{pro}}$  forms are able to cleave at the L/P1 junction either in *cis* or in *trans* (Medina *et al.*, 1993; Cao *et al.*, 1995). Residue numbers quoted below refer to  $Lab$ .

Gorbalenya *et al.* (1991) suggested a relationship between  $L^{\text{pro}}$  and thiol proteinases. This hypothesis was confirmed by site-directed mutagenesis showing that Cys<sup>51</sup> and His<sup>148</sup> were the active site residues (Piccone *et al.*, 1995; Roberts & Belsham, 1995). In the former publication  $Lb^{\text{pro}}$  was modelled onto the papain three-dimensional structure and residues Glu<sup>76</sup> and Asp<sup>164</sup> were suggested to be involved in proteolysis by playing a role in substrate binding. Analysis of an alignment of FMDV  $L^{\text{pro}}$  with the equivalent region of the ERV-1 and -2 polyproteins (Fig. 3) shows complete conservation of the active site Cys<sup>51</sup> amongst the FMDV and ERV sequences but not His<sup>148</sup> – aligned for all sequences except ERV-2. In this case the only 'candidate' histidine residue appears to be misaligned by 9 residues implying a large deletion in the 14 residue strand that, in papain, joins the  $\alpha$ -helix (Gly<sup>148</sup> → Leu<sup>143</sup>) and  $\beta$ -sheet (Asp<sup>158</sup> → Ala<sup>163</sup>) features. The alignment predicts the oxyanion pocket to be composed of the side-chain amide of Asn<sup>46</sup>, which is totally conserved amongst all the virus sequences (analogous to Gln<sup>19</sup> of papain; Fig. 1) and the main-chain amide of Cys<sup>51</sup>. The third member of the catalytic triad of papain (Asn<sup>175</sup>) aligns with Asp<sup>164</sup> of the virus sequences and is totally conserved. Interestingly, this aspartate

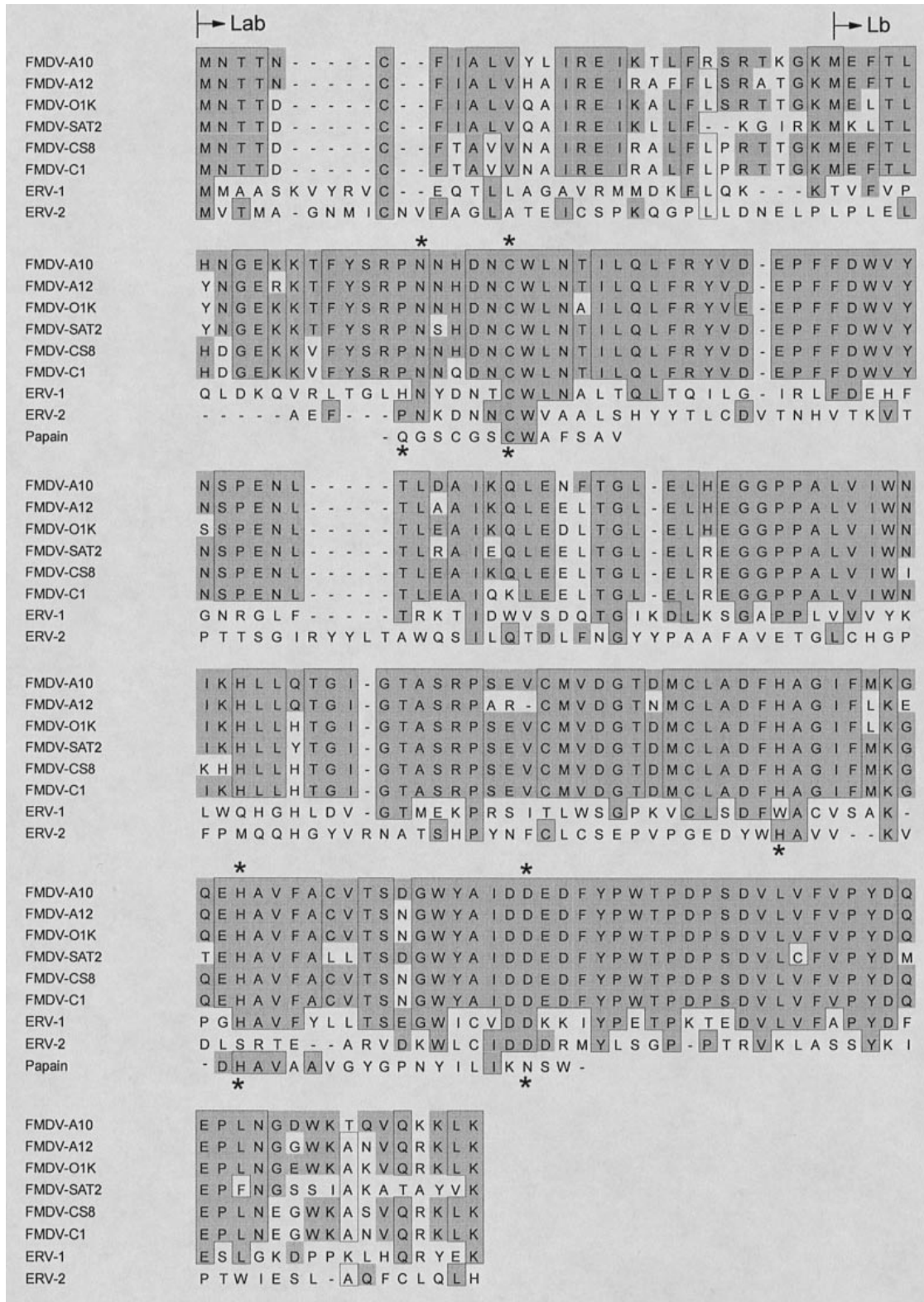


Fig. 3. L proteinase alignments. The L proteins of the foot-and-mouth disease viruses and equine rhinoviruses are aligned together with a partial alignment with the cellular thiol proteinase papain. Active site residues (and other residues referred to in the text) are indicated by asterisks. Identical residues are in boxed areas whereas similar residues are in shaded areas. Sequences were aligned (not papain) using CLUSTALW (Higgins *et al.*, 1991).

residue was mutated to (the papain-like) asparagine and found to be almost wild-type in L/P1 cleavage activity (Piccone *et al.*, 1995).

Not altogether surprisingly L<sup>pro</sup> has another function besides cleaving itself from the nascent polyprotein and this *trans* activity is the same as one identified previously for the 2A proteinase of the entero- and rhinoviruses – cleavage of the host-cell protein eIF-4G (discussed in greater detail in the following section). It is intriguing that this activity is conserved between the aphtho- and entero-/rhinoviruses but that (i) the proteinase with eIF-4G cleavage activity is transposed between the two types of polyprotein and (ii) aphthovirus L<sup>pro</sup> is recognizably of a thiol-type proteinase structure whilst the 2A<sup>pro</sup> is thought to resemble a proteinase of the sub-class of small serine proteinases – perhaps a striking example of convergent evolution. The cleavage site within eIF-4G differs between the two types of proteinase – L<sup>pro</sup> cleaving between residues Gly<sup>479</sup> and Arg<sup>480</sup>, whilst 2A<sup>pro</sup> cleaves between Arg<sup>486</sup> and Gly<sup>487</sup> (Kirchweger *et al.*, 1994).

(ii) **Entero-, rhinovirus 2A proteinase.** In the case of the entero- and rhinoviruses, a primary cleavage occurs between the P1 capsid protein precursor and the replicative domains of the polyprotein (P2, P3; Fig. 2) at a tyrosine–glycine scissile pair. This cleavage is mediated by a virus-encoded proteinase (2A<sup>pro</sup>), of some 17 kDa, cleaving at its own N terminus (Toyoda *et al.*, 1986; Sommergruber *et al.*, 1989). In addition, a processing intermediate form derived from the P3 precursor of both rhino- and enteroviruses may be processed by 2A<sup>pro</sup> at a tyrosine–glycine scissile pair within 3D to yield not the 3C proteinase (3C<sup>pro</sup>) and polymerase (3D<sup>pol</sup>) but 3C' and 3D' (McLean *et al.*, 1976; Rueckert *et al.*, 1979; Hanecak *et al.*, 1982). The significance of this cleavage is not clear and appears to be strain-specific – some polioviruses producing appreciable quantities of 3C' and 3D' whilst others appear not to utilize this pathway – although all entero- and rhinovirus polymerases sequenced to date conserve this tyrosine-glycine pair (interestingly, this pair is not conserved in the aphtho- or cardiovirus 3D<sup>pol</sup> sequences, which do not possess this type of 2A). Mutation of this pair did not, however, affect the growth of poliovirus – in tissue culture – and produced an apparently entirely parental phenotype (Lee & Wimmer, 1988).

The nature of this proteinase was not known but sequence similarities between the sub-class of small cellular serine proteinases and 2A<sup>pro</sup> were identified suggesting a 2A<sup>pro</sup> catalytic triad composed of His<sup>20</sup>, Asp<sup>38</sup> and, interestingly, an active site nucleophile of cysteine (109), rather than serine (Bazan & Fletterick, 1988). Indeed, inhibitor studies showed 2A<sup>pro</sup> to be inhibited by compounds known to be active against thiol proteinases: iodoacetamide and *N*-ethylmaleimide (Konig & Rosenwirth, 1988). The role of these residues in catalysis was confirmed by site-directed mutagenetic analyses (Sommergruber *et al.*, 1989; Hellen *et al.*, 1991; Yu & Lloyd, 1991, 1992). Recent characterizations of the enzyme have

shown 2A<sup>pro</sup> to be a zinc-containing enzyme (Sommergruber *et al.*, 1994a) and the metal ion to play an essential structural, rather than mechanistic, role (Voss *et al.*, 1995).

Analyses of the proteolysis of peptide substrates and site-directed mutagenesis experiments have shown the enzyme has preferred, rather than strict, requirements for specific amino acids near the cleavage site. Only branched amino acids (Val, Ile) were not readily accepted at the P1 site (Skern *et al.*, 1991), although substitution of the P1' residue (glycine in all entero- and rhinoviruses) results in inefficient P1/2A cleavage in *cis* – the effect being most marked by substitution with amino acids with bulky side-chains (Skern *et al.*, 1991; Hellen *et al.*, 1992). Cleavage in *trans*, however, at the P1/2A or 3C'/3D' sites showed substitution at the P2 position greatly reduced proteolysis (Lee & Wimmer, 1988; Hellen *et al.*, 1992).

A second function of 2A<sup>pro</sup> was identified as inhibition of host-cell protein synthesis (Bernstein *et al.*, 1985) which could be supplied in *trans* (or complemented; Bernstein *et al.*, 1986) and correlated with an earlier observation that this inhibition was mediated by cleavage of a 220 kDa polypeptide (latterly identified as eIF-4G) associated with the cap binding protein complex (Etchison *et al.*, 1982). Whilst reports in the literature indicated that 2A<sup>pro</sup> was not involved in eIF-4G cleavage or served as an activating factor for another (host-cell) proteolytic activity which cleaved eIF-4G (reviewed by Belsham & Sonenberg, 1996), other experiments with cloned, expressed, enzyme and eIF-4G substrate have shown that both 2A<sup>pro</sup> and L<sup>pro</sup> can cleave eIF-4G in a direct, rather than by an indirect, manner (Lamphear *et al.*, 1993; Sommergruber *et al.*, 1994b; Kirchweger *et al.*, 1994). The function of cleavage of eIF-4G by 2A<sup>pro</sup> is to inhibit the cap-dependent mode of translation used – resulting in host-cell 'shut-off' and selective translation of virus RNA [reviewed by Sonenberg (1990); Thatch (1992); Belsham & Sonenberg (1996)].

The 2A proteinase has been shown also to be a trans-activator of translation driven by a poliovirus internal ribosome entry sequence (IRES) at times when host-cell (cap-dependent) translation was not inhibited (Hambidge & Sarnow, 1992). Genetic analyses showed that mutations within the 5' non-coding region of virus RNA could be compensated by second-site mutations within the 2A region (Macadam *et al.*, 1994) at times when cap-dependent translation was abolished – interpreted as a direct involvement by 2A<sup>pro</sup> activation of cap-independent translation. The primary P1/2A cleavage mediated by 2A<sup>pro</sup> can be mimicked by the inclusion of a second IRES between P1 and 2A, or between 2A<sup>pro</sup> and 2B, to create a dicistronic poliovirus. Analyses of these viruses showed that although neither protein 2AB nor the entire P2 sequences need to be derived from the same cistron as the other replicative proteins, 2A<sup>pro</sup> appeared to be an essential component of virus RNA replication (Molla *et al.*, 1993). Analysis of viruses with point mutations introduced into 2A<sup>pro</sup> (Asp<sup>38</sup> → Glu; Tyr<sup>89</sup> → Leu) showed a loss of cleavage activity in *trans*, but not in *cis*, and no RNA replication (Yu & Lloyd, 1991; Yu *et al.*, 1995).

(iii) *Aphtho-, cardiovascular 2A protein*. The corresponding aphtho- and cardiovascular primary polyprotein cleavage occurs not between P1 and 2A, but at the C terminus of the 2A region between the capsid protein precursor ([P1-2A] – aphthoviruses; [L-P1-2A] – cardiovasculars) and 2B (Fig. 2). Precursors spanning the 2A/2B cleavage site are not detected during native polyprotein processing. The sequences involved in this 2A/2B cleavage do not conform to any known proteinase motifs and show no sequence similarities with 2A<sup>pro</sup> of the entero-/rhinoviruses, although comparable in size. Although 2A proteins are highly conserved amongst Theiler's murine encephalomyelitis (TME) viruses and amongst encephalomyocarditis (EMC) viruses, only the C-terminal region is highly conserved across the cardiovasculars. The C-terminal region of cardiovascular 2A is, however, highly similar to the much shorter 2A region of FMDV. The three C-terminal residues of aphtho- and cardiovascular 2A proteins are completely conserved (-NPG-) whilst the N-terminal residue (proline) of the 2B proteins of both groups is, again, completely conserved.

Cleavage at the 2A/2B site of FMDV or TMEV polyproteins was shown to require neither the L nor 3C proteinases (Clarke & Sangar, 1988; Roos *et al.*, 1989; Ryan *et al.*, 1989). Analyses of recombinant FMDV polyproteins and artificial reporter gene polyproteins showed that a 20 amino acid oligopeptide sequence corresponding to the FMDV 2A region (together with the N-terminal proline of protein 2B) was able to mediate a co-translational cleavage (Ryan *et al.*, 1991; Ryan & Drew, 1994). Other studies have shown that deletion of the N-terminal 66% of EMCV 2A did not abrogate cleavage at the 2A/2B site and that mutations within the conserved -NPGP- sequence at the extreme C terminus of EMCV 2A abolished cleavage activity (Palmenberg *et al.*, 1992). Cleavage at the 2A/2B site of TMEV was highly efficient when only 2A and 2B sequences were present (Batson & Rundell, 1991). The C-terminal 19 amino acids, together with the N-terminal proline of 2B from either FMDV, EMCV or TMEV, when inserted into an artificial polyprotein are able to mediate a co-translational cleavage with high efficiency (~95%; Donnelly *et al.*, 1997). This raises the question why do cardiovasculars possess a 2A region of some 150 amino acids whilst those of aphthoviruses are only 18 amino acids? Quite possibly the 2A protein of cardiovasculars plays another role in the infectious cycle.

The mechanism of the aphtho- and cardiovascular 2A protein-mediated cleavage is not clear. What is clear, however, is that this represents an entirely novel form of cleavage mechanism – data suggesting that this cleavage is a single turnover event and, therefore, not enzymic *sensu stricto* (unpublished observations). The alternative mechanisms of cleavage are: (i) 2A is a proteolytic element cleaving the Gly-Pro peptide bond; (ii) the 2A sequence represents a substrate for a host-cell proteinase specific for this particular Gly-Pro bond; or (iii) 2A interrupts the elongation cycle during protein synthesis such that a peptide bond is not actually formed between the glycine and proline residues – if this is the case the use of the term

'cleavage' in this particular section would assume an entirely different meaning. Site-directed mutagenesis/translation studies *in vitro* are in progress to determine how this 'type' of picornavirus 2A mediates cleavage.

(iv) *Hepatovirus, ECHO virus 22 and 23 2A protein*. The corresponding primary cleavage between the capsid protein precursor and the replicative proteins of hepatitis A virus is mediated not by hepatitis A virus 2A (for which no function is ascribed) but by the 3C proteinase (Jia *et al.*, 1993; Schultheiss *et al.*, 1994; see Fig. 2). Similarly, the function of the 2A protein of the highly divergent ECHO virus types 22 and 23 is not known. Expression of the ECHO-22 virus [P1-2ABC] protein *in vitro* did not reveal any proteolytic activity (Schultheiss *et al.*, 1995a).

(v) *Picornavirus 3C proteinase*. Virus-specific proteolytic activities were first described for EMCV (Lawrence & Thatch, 1975) and poliovirus (Korant *et al.*, 1979). This activity was shown to be associated with protein 3C in both EMCV (protein p22; Gorbalenya *et al.*, 1979; Palmenberg *et al.*, 1979) and poliovirus (protein P3-7c; Hanecak *et al.*, 1982). All picornaviruses possess a 3C proteinase (3C<sup>pro</sup>) which shows a high degree of sequence similarity across the genera. Like L<sup>pro</sup> and 2A<sup>pro</sup>, 3C<sup>pro</sup> mediates a single primary cleavage, although at a distal site – between 2C and 3A. One report suggests an alternative primary cleavage between poliovirus proteins 2A and 2B (Lawson & Semler, 1992) and we have unpublished data indicating a similar alternative primary cleavage between FMDV proteins 2B and 2C (M. D. Ryan & M. Flint, unpublished). In the case of the hepatitis A viruses, however, 3C<sup>pro</sup> mediates the cleavage between the capsid protein precursor and the replicative domains of the polyprotein (Schultheiss *et al.*, 1995b).

Unlike other picornavirus proteinases, 3C<sup>pro</sup> is responsible for a series of secondary cleavages resulting in the processing of the capsid and replicative protein precursors. In the case of the polioviruses 3C<sup>pro</sup>-mediated cleavages occur at Gln-Gly pairs – by no means all of such pairs present within the polyprotein are processed, however, their position within the polyprotein domains being a determining factor. The processing of the poliovirus capsid protein precursor is, in fact, mediated not by 3C<sup>pro</sup>, but by 3CD<sup>pro</sup> (Jore *et al.*, 1988; Ypma-Wong *et al.*, 1988). 3C<sup>pro</sup> from other genera is, however, able to process capsid protein precursors (Vakharia *et al.*, 1987; Clarke & Sanger, 1988; Parks *et al.*, 1989; Jia *et al.*, 1991; Harmon *et al.*, 1992) although the processing may be more efficient with 3CD<sup>pro</sup> (Ryan *et al.*, 1989). 3D is an RNA-dependent RNA polymerase but in this context (uncleaved 3CD) the presence of 3D serves to alter the substrate specificity of 3C<sup>pro</sup> in the processing of capsid protein precursors *in trans*. 3C also retains proteolytic activity in the precursor forms 3ABC and P3 (Jackson, 1986; Parks *et al.*, 1989). The auto-processing of 3CD<sup>pro</sup> to 3C and 3D can be greatly enhanced by the addition of purified 3AB protein, which also stimulates

processing of protein 2BC (Molla *et al.*, 1994). The extent to which 3C<sup>pro</sup> was able to process precursors from other species was studied by the construction of chimeric polyproteins in which the region of the poliovirus cDNA encoding 3C was replaced by the equivalent sequences from human rhinovirus 14 (HRV-14) or coxsackievirus B3 (CB3). The poliovirus P2 precursor could be processed by both the HRV-14 and CB3 3C<sup>pro</sup>s, but the P1 capsid protein precursor was not processed by either HRV-14 or CB3 3C<sup>pro</sup>s (Dewalt *et al.*, 1989).

In addition to processing virus proteins, picornavirus 3C<sup>pro</sup> has been demonstrated to cleave a number of host-cell proteins: histone H3 (Falk *et al.*, 1990; Tesar & Marquardt, 1990), transcription factor IIIC (Clark *et al.*, 1991), TATA-binding protein (Clark *et al.*, 1993; Das & Dasguta, 1993) and microtubule-associated protein 4 (Joachims *et al.*, 1995). Given the panoply of host-cell proteins present in an infected cell it is, perhaps, not surprising that some proteins are cleaved by 3C<sup>pro</sup>. The contribution to the efficiency of virus replication by such perturbations of host-cell macromolecular processes is, however, difficult to assess. These observations are, however, of great interest in the consideration of persistent or chronic picornavirus infections – cell-lines may indeed be generated expressing low levels of 3C<sup>pro</sup> (Lawson *et al.*, 1989; Martinez-Salas & Domingo, 1995).

Characterization of 3C<sup>pro</sup> using proteinase inhibitors, either with virus-infected cells, *in vitro* translation systems programmed with virus or transcript RNA or purified enzyme preparations, showed rather confusing inhibitor profiles. Inhibition of proteolytic activity was observed with both serine and thiol proteinase inhibitors (Summers *et al.*, 1972; Korant, 1972, 1973; Pelham, 1978; Gorbalenya & Svitkin, 1983; Korant *et al.*, 1985; Baum *et al.*, 1991). In addition, similarities between the large sub-class (trypsin-like) cellular serine proteinases and 3C proteinases were detected by sequence alignments (Gorbalenya *et al.*, 1986; Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1988, 1989*a, b*; Bazan & Fletterick, 1990). Such alignments predicted both a chymotrypsin (serine proteinase)-like fold and the identity of the residues which would form a catalytic triad analogous to that of serine proteinases (Fig. 4). The putative triad was predicted to be composed of (poliovirus numbering scheme): (i) His<sup>40</sup> – completely conserved amongst all sequences; (ii) an acidic residue – predictions implicated Asp<sup>85</sup> (Bazan & Fletterick, 1988), or either Asp<sup>71</sup> (aphtho-, cardioviruses) or Glu<sup>71</sup> (entero-, rhinoviruses; Gorbalenya *et al.*, 1989*a*); and, perhaps most interestingly (iii) the nucleophilic residue being a cysteine (I47) – rather than a serine residue (an identical type of substitution for the entero-/rhinovirus 2A proteinase as described above). Site-directed mutagenesis experiments confirmed the roles in catalysis of His<sup>40</sup> and Cys<sup>147</sup> and showed that such data supported the prediction of Gorbalenya: Asp<sup>71</sup> in the putative triad of aphthoviruses and Glu<sup>71</sup> in the triad of enteroviruses (Cheah *et al.*, 1990; Lawson & Semler, 1990; Hammerle *et al.*, 1991; Kean *et al.*, 1991, 1993; Grubman *et al.*, 1995).

An entirely unexpected property of this enzyme was discovered by the analysis of mutations suppressing the effect of a four base insertion within the 5' non-coding region of the RNA genome. Such suppresser mutations were shown to be active in *cis* and mapped within 3C<sup>pro</sup> (Andino *et al.*, 1990*a*). Latterly, it was demonstrated that these mutations were affecting the binding of 3CD<sup>pro</sup>, rather than 3C<sup>pro</sup>, to positive-strand RNA (Andino *et al.*, 1990*b*, 1993) and that these mutations mapped to multiple sites within the 3C<sup>pro</sup> primary structure (poliovirus numbering scheme): His<sup>31</sup>, Asp<sup>32</sup>, Lys<sup>82</sup>–Ile<sup>86</sup>, Thr<sup>154</sup>–Lys<sup>156</sup>, Ala<sup>171</sup> and Arg<sup>176</sup> (Andino *et al.*, 1990*a, b*, 1993; Leong *et al.*, 1993; Walker *et al.*, 1995). The RNA binding properties of 3CD<sup>pro</sup> are dependent upon the host-cell protein EF-1 $\alpha$  (p50) or p36 (an N-terminal fragment of p50). This host factor could be replaced by poliovirus protein 3AB and the [3AB:3CD<sup>pro</sup>] complex also was able to bind sequences at the 3' terminus of the poliovirus genome (Harris *et al.*, 1994).

(vi) *The atomic structure of picornavirus 3C proteinase – a catalytic dyad or triad?* The resolution of the atomic structures of the HRV-14 (Matthews *et al.*, 1994) and hepatitis A virus 3C proteinases (Allaire *et al.*, 1994) confirmed the predicted chymotrypsin-like fold of the enzyme but appeared (in some respects) to cloud, rather than resolve, the issue concerning the catalytic mechanism. The structures of the enzymes were similar and showed an N-terminal  $\alpha$ -helix not found in trypsin or  $\alpha$ -lytic proteinase (Fig. 4). In the case of the HRV-14 3C structure the side-chain of His<sup>40</sup> showed two alternative conformations. In one conformation the imidazole ring was oriented such that the N $\epsilon$ 2 proton hydrogen bonds to the thiol of Cys<sup>146</sup>, whereas in the other conformation the N $\epsilon$ 2 proton hydrogen bonds to the *syn*-oriented electron lone pair of the Glu<sup>71</sup> carboxylate – the authors suggesting that conformational mobility of the His<sup>40</sup> side-chain may play a role in catalysis. Glu<sup>71</sup> is totally conserved in alignments of entero- and rhinovirus 3C<sup>pro</sup>s (lineage 'A'; see Fig. 7, panel A, and below) with the exception of ECHO viruses 22 and 23. The Glu<sup>71</sup> side-chain is oriented differently from the (shorter) Asp side-chains in serine proteinases such that the carboxylate group forms a hydrogen bond using the *anti*-oriented electron lone pair. This type of hydrogen bond is unusual – all carboxyls in Asp-His-Ser triads are in the *syn*-orientation. This unusual conformation of the side-chain of Glu<sup>71</sup> is stabilized by an interaction by (i) the main-chain NH of His<sup>40</sup> and the O $\epsilon$ 1 of Glu<sup>71</sup> and (ii) the side-chain NH of Asn<sup>69</sup> and O $\epsilon$ 2 of Glu<sup>71</sup>. Asn<sup>69</sup> is not conserved amongst picornavirus 3C<sup>pro</sup>s alignments suggesting either: (i) substitution by serine or proline in the case of the enteroviruses, (ii) substitution by lysine in other rhinoviruses, (iii) absent (gap character) in aphthoviruses and (iv) substitution by phenylalanine in the hepatitis A viruses. In comparison with serine proteinases the HRV-14 3C<sup>pro</sup> oxyanion hole, formed by the -GxCG[G/S/A]- motif conserved amongst picornaviruses and the Cys<sup>146</sup> side-chain is displaced by some 1.2 Å from His<sup>40</sup>, although with a similar alignment.



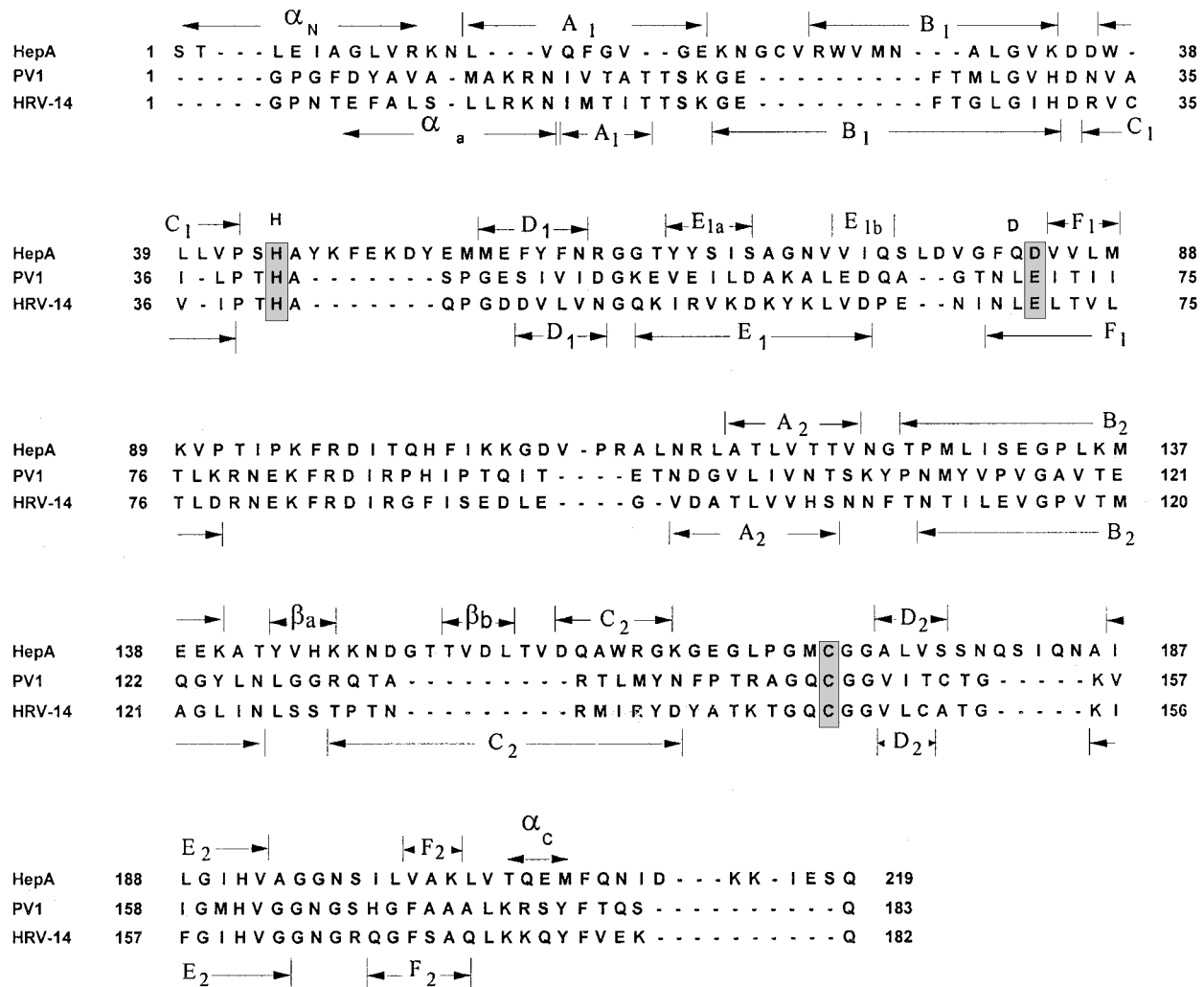


Fig. 4. 3C proteinase alignments. All available picornavirus 3C sequences were aligned using CLUSTALW. Only a representative sequence of the entero- (poliovirus type 1; PV1), rhino- (human rhinovirus-14; HRV-14) and hepatovirus (hepatitis A virus; HepA) groups is shown in alignment. The structural features determined for the hepatitis A virus and HRV-14 3C proteinases are shown above and below the alignment, respectively. Active site residues predicted initially by sequence alignment, latterly by determination of the atomic structures, are also shown (shaded boxes).

Molecular modelling analyses of the substrate binding pocket showed that the  $P_1$  residue (normally Gln or Glu) fits into a shallow pocket flanked by  $\beta E_2$ ,  $\beta F_2$  and the loop connecting  $\beta C_2$  to  $\beta D_2$  (see Fig. 4). The side-chain of the  $P_1$  residue is hydrogen bonded by His<sup>160</sup> (completely conserved) and Thr<sup>141</sup> (threonine in all entero-, rhino-, aphtho- and cardioviruses, serine in ECHO 22, 23 viruses and glycine in hepatitis viruses). Residues  $P_2$ – $P_5$  interact with sections of the chain between  $\beta B_2$  and  $\beta E_2$ , whilst residues  $P_1'$  and  $P_2'$  are positioned between  $\beta B_1$  and the loop composed of Thr<sup>143</sup>, Gly<sup>144</sup> and Gln<sup>145</sup>. The side-chain of the  $P_2$  residue is bound at the aperture of the cavity within the proteinase which can accommodate residues with bulky aromatic side-chains. The side-chain of the  $P_4$  residue is oriented towards a shallow pocket lined by the hydrophobic side-chains of Ile<sup>124</sup>, Leu<sup>126</sup>

and Phe<sup>169</sup> – favouring small, hydrophobic side-chains in the  $P_4$  position. These conclusions are in agreement with proteolysis studies using synthetic peptides (Long *et al.*, 1989). The  $P_1'$  and  $P_2'$  residues (normally Gly–Pro in the entero- and rhinoviruses) are thought to form a turn orienting the substrate toward a second side depression – without counterpart in the serine trypsin-like proteinases.

Mutations shown to affect RNA binding (see above) were located within the atomic structure on the opposite side to the catalytic site. The majority of these mutations were observed in a domain interconnecting the two lobes of the proteinase – located between the two (anti-parallel) helices of the N and C termini – and also clustered on the (exterior) loop connecting  $\beta D_2$  and  $\beta E_2$  – some 5 residues longer in hepatitis A virus than in HRV-14. Interestingly, this interconnecting region lies in

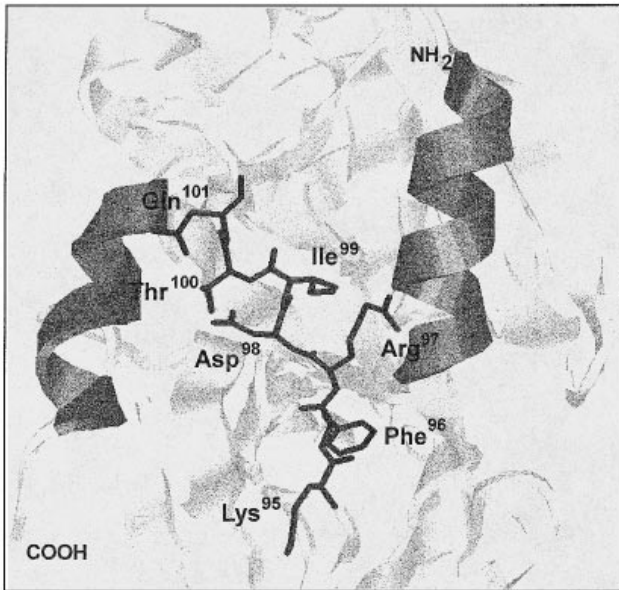


Fig. 5. Hepatitis A virus 3C proteinase RNA binding site. The domain interconnecting loop is shown with residues 95–101 shown as bold wireframe representations. The flanking N- and C-terminal helices are shown (darker shading).

close proximity to the N and C termini of the proteinase (Fig. 5). It is tempting to speculate that the RNA binding loop in such close proximity to the termini may well play a role in determining which of the alternative P3 processing pathways

is adopted – leading ultimately to the utilization of the polymerase 3D sequences either as a proteinase ( $3CD^{pro}$ ) or a polymerase ( $3D^{pol}$ ). The poorly understood mechanism whereby vRNA is packaged into the ‘correct’ capsids during, say, a mixed virus infection may be directly connected to these phenomena: (i) a *cis* interaction between  $3CD^{pro}$  and the RNA by which it is encoded, and (ii) the specificity of  $3CD^{pro}$  for the correct P1 capsid protein precursor substrate. Perhaps the [ $3CD^{pro}$ :vRNA] complex acts as a packaging ‘nucleation’ point – processing P1 and packaging vRNA simultaneously.

The structure of the hepatitis A virus 3C proteinase showed a high degree of similarity to that of HRV-14 with relative insertions/deletions (predicted from sequence alignments) corresponding to structural features such as connecting loops or relatively minor differences in the lengths of  $\beta$ -sheets. For the purposes of the following section the hepatitis A virus  $3C^{pro}$  numbering scheme will be used. The most interesting difference occurred in the nature of the catalytic site. In the case of the HRV-14 the three-dimensional structure was derived from protein expressed in *E. coli*, forming inclusion bodies and being re-folded from 6 M guanidine.HCl. In the case of the hepatitis A virus  $3C^{pro}$  the structure was determined from a double mutant  $Cys^{24} \rightarrow Ser$  and  $Cys^{172} \rightarrow Ala$ . The hepatitis A virus structure showed quite clearly, however, an alternative positioning of key residues within the active site to indicate a catalytic dyad (similar to the thiol proteinase papain), rather than a serine proteinase-like triad. The  $Asp^{84}$  of hepatitis A virus  $3C^{pro}$  in this lineage ‘B’ of picornavirus  $3C^{pro}$ s (see Fig.

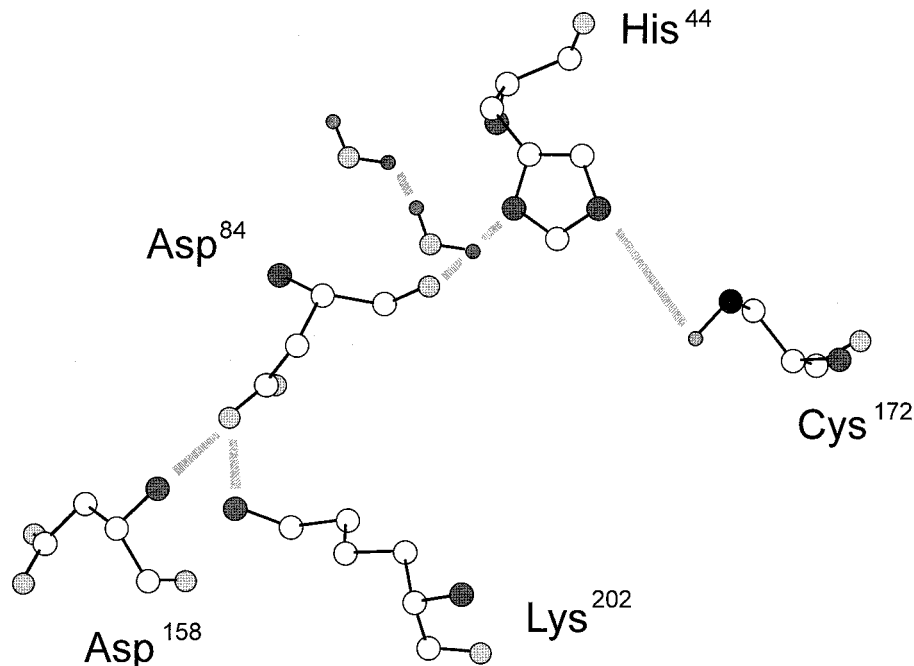


Fig. 6. Hepatitis A virus 3C proteinase active site. The positions of active site residues ( $His^{44}$ ,  $Cys^{172}$ ) and water molecules are shown. The interaction of  $Asp^{84}$  with  $Asp^{158}$  and  $Lys^{202}$  is also indicated (interactions between groups are indicated by dashed lines). Residue numbers quoted refer to the hepatitis A virus 3C sequence. Atom types are as represented as in Fig. 1.

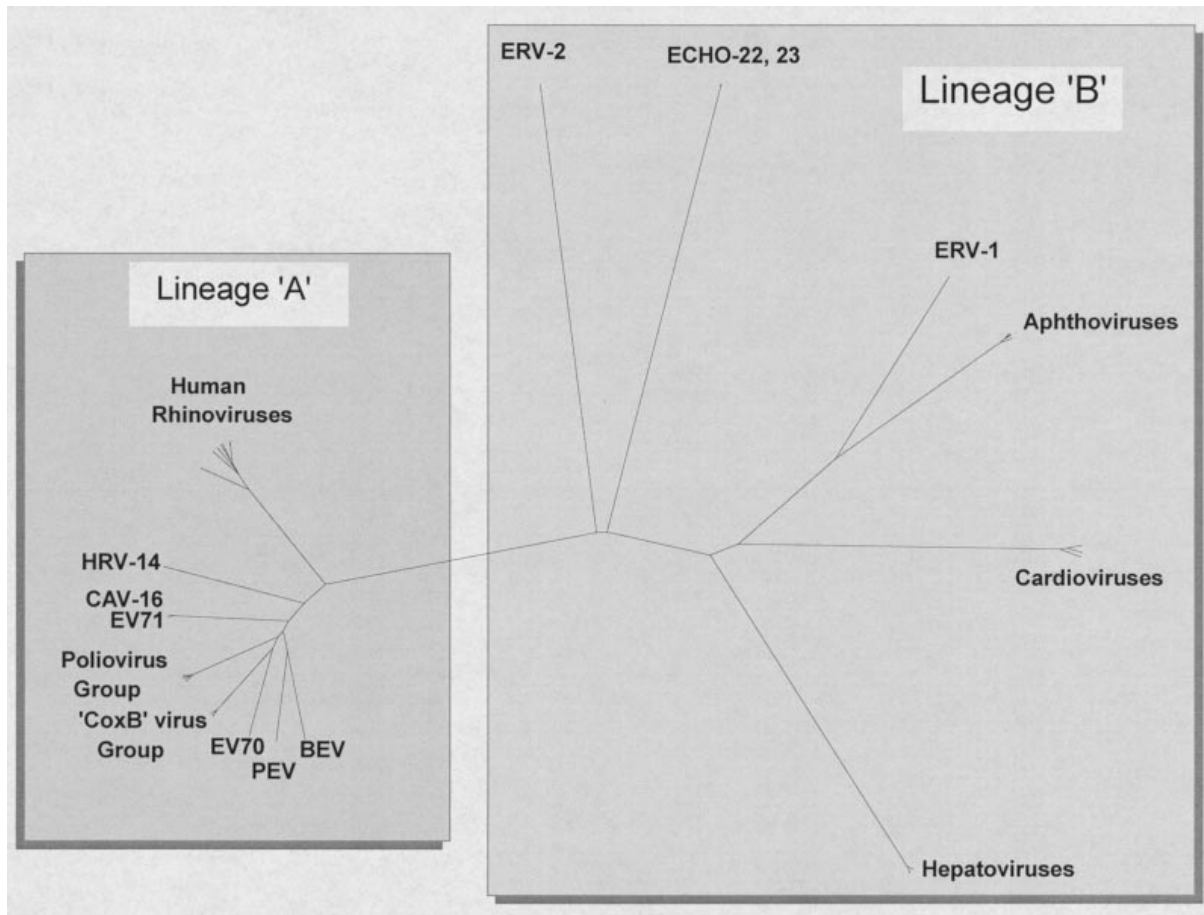


Fig. 7(A)

Fig. 7. For legend see page 710.

7, panel A) does not interact with His<sup>44</sup> as one would have expected but the carboxylate group of the side-chain points away from His<sup>44</sup> and is in strong interaction with the  $\epsilon$ -amino group of Lys<sup>202</sup> and the main-chain amide of Asp<sup>158</sup> (Fig. 6). Note: Lys<sup>202</sup> aligns with Ala<sup>171</sup> of HRV-14 – identified as being involved in RNA binding. The N $\delta$ I of His<sup>44</sup> is hydrogen bonded to a water molecule (occupying the position of the carboxylate group of the aspartate residues of a serine proteinase-like triad), both being in close proximity to the phenolic hydroxyl group of Tyr<sup>143</sup>. This water molecule in fact forms hydrogen bonds with three species: (i) another water molecule, (ii) the backbone carbonyl of Asp<sup>84</sup> and (iii) imidazole N $\delta$ I of His<sup>44</sup>.

The picture to emerge from the analysis of the hepatitis A virus 3C proteinase structure is one of a complex electrostatic environment in the active site. Asp<sup>84</sup> is completely conserved amongst lineage 'B' of proteinases (Fig. 7, panel A) and was assumed to play a catalytic role equivalent to Glu<sup>71</sup> of the entero- and rhinoviruses – as is observed in the HRV-14 atomic structure. The side-chain of Asp<sup>84</sup> does not, however, form a member of a catalytic triad but is oriented away from

the catalytic site – the position and function one might have expected it to have in order to be substituted by a water molecule. The observation that buried water molecules and their protein environments may be conserved amongst homologous serine proteinases suggests that they may be important elements of the whole structure and have been postulated to play a variety of roles – both structural and mechanistic (Sreenivasan & Axelsen, 1992). The 'sequestration' of the side-chain of Asp<sup>84</sup> by Lys<sup>202</sup> and Asp<sup>158</sup> would leave the N $\delta$ I of His<sup>44</sup> without a hydrogen bonding partner – fulfilled in the case of hepatitis A 3C proteinase by a water molecule.

Analyses of sequence alignments are shown as a rooted (Fig. 7, panel A) or unrooted phenogram (Fig. 7, panel B) with distances proportional to relatedness. Two lineages may be defined: lineage 'A' – those sequences with glutamate in alignment with position 71 of poliovirus; and lineage 'B' – those sequences with aspartate in this position. Lineage A is a closely related series of sequences comprising entero- and rhinoviruses. Lineage B is much more diverse encompassing the aphtho-, cardio- and hepatoviruses together with ECHO

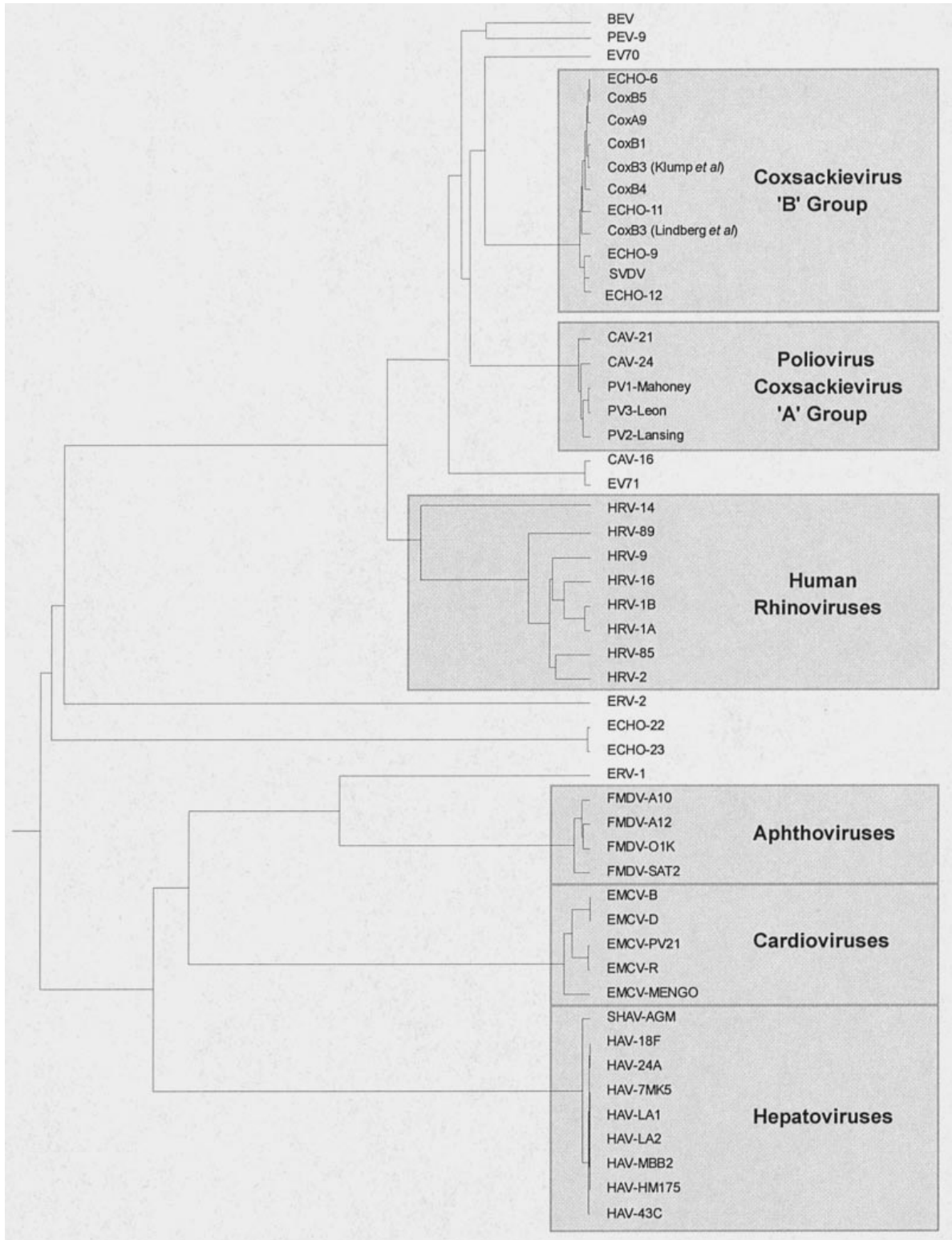


Fig. 7(B)

Fig. 7. Relationships between picornavirus 3C proteinases. Picornavirus 3C<sup>Pro</sup> sequences were aligned using CLUSTALW and the relatedness was determined using PROTDIST (Felsenstein, 1991). Relationships are shown as rooted (panel A) and unrooted (panel B) phenograms with branch lengths proportional to relatedness.

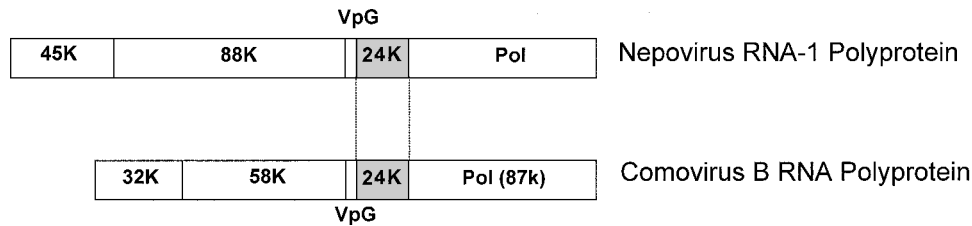


Fig. 8. Como- and nepovirus genome organization. Protein domains of the polyproteins encoded by comovirus B RNA and nepovirus RNA-1 are shown with 24K proteinases in shaded areas.

viruses 22, 23 (recently split off from the enteroviruses into an as yet unnamed new genus) and ERV-1 and -2. The strong interaction between Asp<sup>84</sup> and the side-chain of Lys<sup>202</sup> observed in the hepatitis A virus 3C proteinase structure cannot be conserved across other lineage B proteinases since lysine at position 202 is only observed in the hepatitis A virus 3C sequences in proteinase alignments.

### Como- and nepovirus 24K proteinases

The genomes of como- and nepoviruses comprise two positive-sense RNA molecules. The replicative proteins are encoded by comovirus B RNA and nepovirus RNA-1 (Fig. 8). Comovirus B RNA encodes a 200 kDa polyprotein which is co-translationally cleaved into 32 and 170 kDa (58K + VPg + 24K + 57K) proteins in an intramolecular reaction (*in cis*). The 170 kDa protein is subsequently cleaved into further processing intermediates and mature products. Nepovirus RNA-1 encodes a polyprotein of some 253 kDa showing the same organization as the comovirus B RNA polyprotein. The bipartite organization of these genomes may be regarded as functionally analogous to the primary proteolytic cleavage between the capsid and replicative protein domains of the picornavirus polyprotein. The enzyme responsible for the cleavages of the comovirus B and nepovirus RNA-1-encoded polyproteins has been identified as the 24K proteinase (Franssen *et al.*, 1984a; Verver *et al.*, 1987; Vos *et al.*, 1988; Margis *et al.*, 1991; Margis & Pinck, 1992). The comovirus 24K proteinase activity is modulated by interaction with other virus proteins. The primary cleavage products (32 and 170 kDa proteins) remain associated and the presence of the 32 kDa protein in the complex both inhibits the subsequent processing of the 170 kDa protein and enhances the proteolytic processing (*in trans*) of the polyprotein (capsid proteins) derived from the M RNA (Peters *et al.*, 1992). The 110 kDa product (uncleaved 24K proteinase and 87K polymerase – analogous to picornavirus 3CD) when expressed from a sub-genomic cDNA construct in translation mixtures was stable. If sequences upstream of the 24K proteinase were re-introduced into the cDNA construct the cleavage at the proteinase/polymerase junction was enhanced (Dessens & Lomonosoff, 1992), the effect being most marked when the C-terminal region of the 58K protein, the 58K/VPg cleavage site and entire VPg were

present – perhaps a similar effect to the poliovirus 3AB-stimulated cleavage of 3CD → 3C + 3D described above.

Alignment of como- and nepovirus 24K proteinase sequences together with picornavirus 3C<sup>pro</sup> (and other plant virus proteinases) allowed prediction of the serine proteinase chymotrypsin-like fold and the identity of functional residues (Franssen *et al.*, 1984b; Argos *et al.*, 1984; Greif *et al.*, 1988; Bazan & Fletterick, 1988, 1990; Gorbalenya *et al.*, 1989a). Site-directed mutagenic analyses of the comovirus proteinase resolved ambiguities as to the composition of the putative catalytic triad. Analysis of proteolytic activities of 24K proteinases with mutations at His<sup>40</sup>/His<sup>44</sup>, Glu<sup>76</sup>/Asp<sup>94</sup> and Cys<sup>166</sup> indicated a catalytic triad composed of His<sup>40</sup>, Glu<sup>76</sup> and Cys<sup>166</sup> (Dessens & Lomonosoff, 1991; see Fig. 9, panel A). The relatedness of the como- and nepovirus 24K sequences (together with the levivirus rice tungro spherical virus – see below) is shown in Fig. 9, panel B. The picornavirus 3C<sup>pro</sup> RNA binding domain (interconnecting loop linking βF<sub>1</sub> and βA<sub>2</sub>; see Figs 4 and 6) containing the highly conserved -KFRDI(R/T)- motif may be represented by the conserved -FxxD- motif in the como- and nepovirus 24K proteinases (Fig. 9, panel A) suggesting that these proteinases may also bind RNA.

### Parsnip yellow fleck and rice tungro spherical virus 3C-like proteinases

Attempts to align these 3C-like proteinase sequences with other virus proteinase sequences showed a close relationship between the rice tungro spherical virus proteinase (RTSV; Shen *et al.*, 1993) and those of the como- and nepoviruses (Fig. 9, panels A and B). The alignment clearly shows the putative catalytic triad and other features described above. The parsnip yellow fleck virus (PYFV; Turnbull-Ross *et al.*, 1992) 3C-like proteinase could also be aligned with the picornavirus 3C sequences (Fig. 10) and, again, showed the putative catalytic residues and substrate binding pocket region in good alignment. PYFV sequences aligning with the picornavirus RNA binding region (-KFRDI-) shows little similarity.

### Poty- and bymovirus proteinases

The potato virus Y family is composed of three genera: the monopartite poty- and rymoviruses and the bipartite bymo-

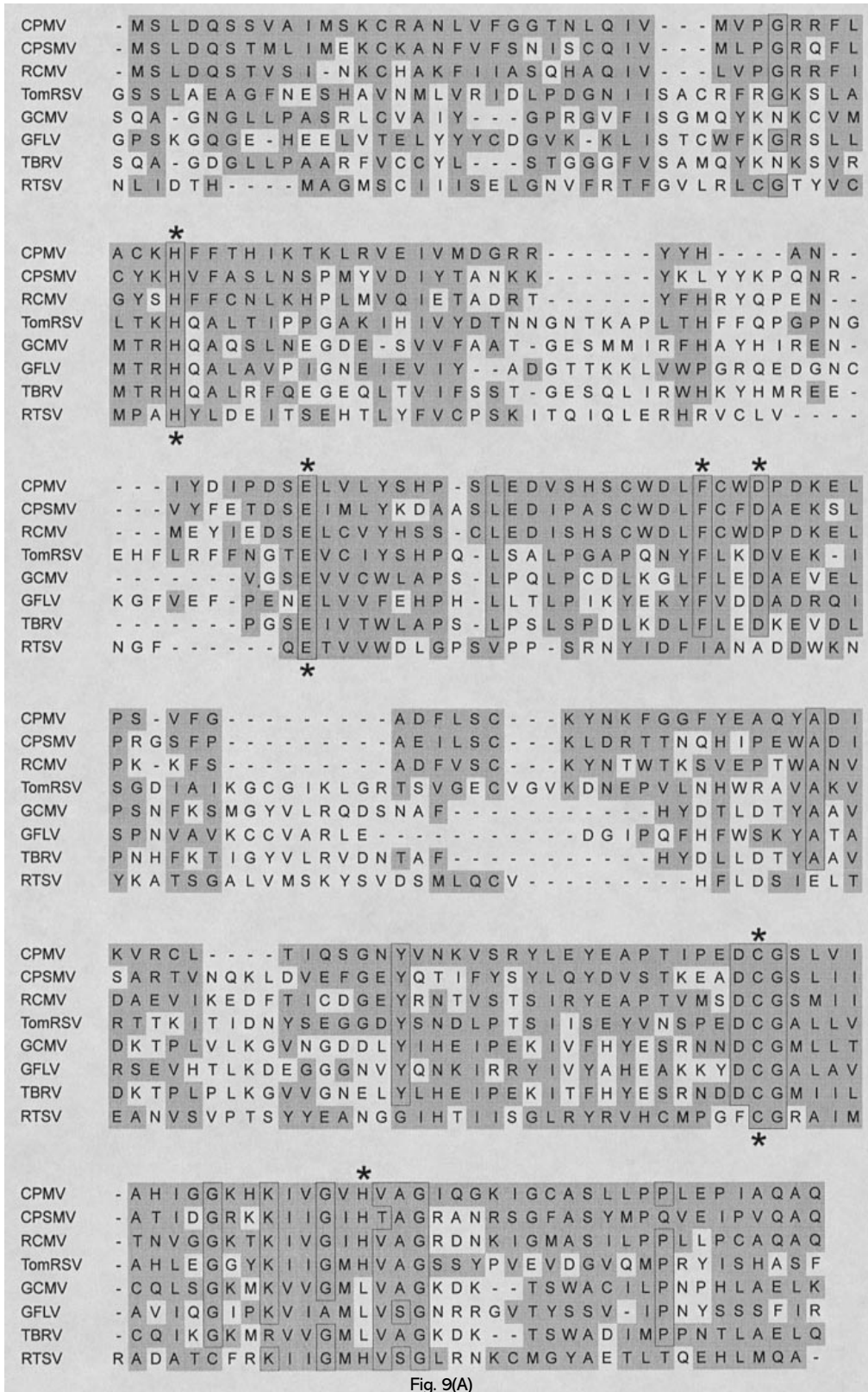


Fig. 9(A)

Fig. 9. For legend see facing page

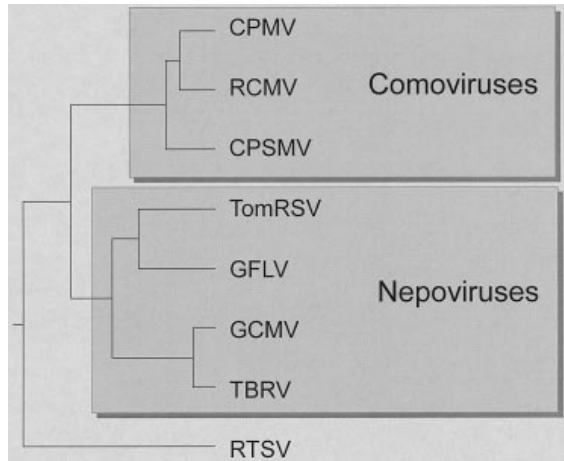


Fig. 9(B)

**Fig. 9.** Como- and nepovirus 24K proteinase alignments and relationships. The 24K proteinases of como- and nepoviruses were aligned using CLUSTALW. Comoviruses: cowpea mosaic virus (CPMV; Lomonosoff & Shanks, 1983), red clover mottle virus (RCMV; Shanks & Lomonosoff, 1990) and cowpea severe mosaic virus (CPSMV; Chen & Bruening, 1992). Nepoviruses: tomato ringspot virus (TomRSV; Rott *et al.*, 1995; Hans & Sanfaçon, 1995), grapevine fanleaf virus (GFLV; Ritzenthaler *et al.*, 1991), Hungarian grapevine chrome mosaic virus (GCMV; Le Gall *et al.*, 1989) and tomato black ring virus (TBRV; Greif *et al.*, 1988). Levivirus: rice tungro spherical virus (RTSV; Shen *et al.*, 1993). Active site residues (and other residues referred to in the text) are indicated by asterisks. Identical residues are in boxed areas whereas similar residues are in shaded areas (panel A). Phenogram plot with branch lengths proportional to relatedness (panel B).

viruses – with similar genome organizations (Fig. 11). Potyviruses encode all of their proteins in a single, long open reading frame whereas bymoviruses have two polyproteins – one encoded on each of the RNA segments (reviewed by Riechmann *et al.*, 1992). Early work using *in vitro* translation systems provided evidence that the protein profiles observed were derived from proteolytic processing of a polyprotein (Vance & Beachy, 1984; Yeh & Gonsalves, 1985) derived from a single, long open reading frame (Domier *et al.*, 1986; Allison *et al.*, 1986; Lain *et al.*, 1989; Maiss *et al.*, 1989; Robaglia *et al.*, 1989). The potyvirus polyprotein contains no less than three distinct proteolytic activities in the form of domains of larger proteins.

(i) **Nla proteinase.** The Nla (nuclear inclusion protein a) or 49K proteinase of tobacco etch virus (TEV) aggregates within the nucleus of infected cells to form plate-like crystalline inclusions. The *in vitro* expression of transcript RNA derived from a series of sub-genomic cDNAs showed the 49K protein to be a proteinase; moreover, this activity was located within the C-terminal half of the protein (Carrington & Dougherty, 1987*a*; Hellmann *et al.*, 1988). By similar analyses the Nla proteinase was shown to autoproteolytically process itself from the polyprotein (Carrington & Dougherty, 1987*b*). The cleavage sites within the TEV polyprotein all contained a -Glu-x-x-Tyr-x-Gln<sup>↓</sup>Ser/Gly sequence motif which forms a cleavage

‘cassette’ which may be included within, or between, proteins to form a restriction proteinase site – cleaved *in trans* by Nla proteinase (Carrington & Dougherty, 1988). The precise sequence of this heptapeptide determines the rate at which the site is cleaved (Dougherty & Parks, 1989; Dougherty *et al.*, 1989*a*). The 49K proteinase has also been used to construct self-processing artificial polyproteins in a manner similar to that of the FMDV 2A sequence (Marcos & Beachy, 1994; Ryan & Drew, 1994) and may be used as a ‘restriction proteinase’.

Alignments of potyvirus Nla sequences with other virus and cellular trypsin-like serine proteinases suggested a catalytic triad of histidine, aspartate and cysteine (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1989*a*; Fig. 12) with the histidine residue in the substrate binding pocket characteristic of a Glu-X substrate specificity. Site-directed mutagenic studies confirmed the importance of these residues for activity (Carrington *et al.*, 1988; Dougherty *et al.*, 1989*b*; Garcia *et al.*, 1990). A 27 kDa form of the proteinase was observed later in infection (Carrington & Dougherty, 1987*a, b*; Dougherty & Parks, 1991). The N-terminal VPg domain of the Nla protein is cleaved from the C-terminal 27K proteinase domain at a sub-optimal cleavage site. The proteolytic activity of the 27K form of the proteinase is similar to that of the 49K form (Dougherty & Parks, 1991). An unusual feature of the Nla proteinase is the relatively long C-terminal region following the catalytic cysteine residue. Nla proteinases are some 220 amino acids long whilst picornavirus 3C proteinases are some 183 amino acids – much of the extra length is accounted for in this potyvirus C-terminal extension. This C-terminal region does, however, undergo ‘trimming’ removing some 20–24 amino acids from the C terminus (Parks *et al.*, 1995; Kim *et al.*, 1995). The C-terminally truncated form of the enzyme was, however, some one-twentieth as active as the full-length form on test peptide substrates (Parks *et al.*, 1995). Previous studies had shown that the 113K precursor (uncleaved 6K–49K–58K, analogous to picornavirus 3ABDC) and its processing intermediates [(i) 107K (uncleaved 49K–58K, analogous to picornavirus 3BCD) (ii) 85K (uncleaved 27K–58K, analogous to picornavirus 3CD) and 54K (uncleaved 6K–49K, analogous to picornavirus 3ABC)] were all proteolytically active (Parks *et al.*, 1992).

(ii) **HC proteinase.** The HC proteinase has been the subject of a recent review (Maia *et al.*, 1996) and will, therefore, only be briefly discussed here. HC-Pro appears to perform only one cleavage within the potyvirus polyprotein – at its own C terminus (Carrington *et al.*, 1989*a, b*). The proteolytic domain of potyvirus HC-Pro may be aligned with the cysteine proteinase domain present within the polyprotein encoded by bymovirus RNA-2 (Davidson *et al.*, 1991; Fig. 13) and similarities to the autocatalytic p29 proteinase of the hypovirulence-associated dsRNA virus of chestnut blight fungus have been reported (Choi *et al.*, 1991). The order of the active

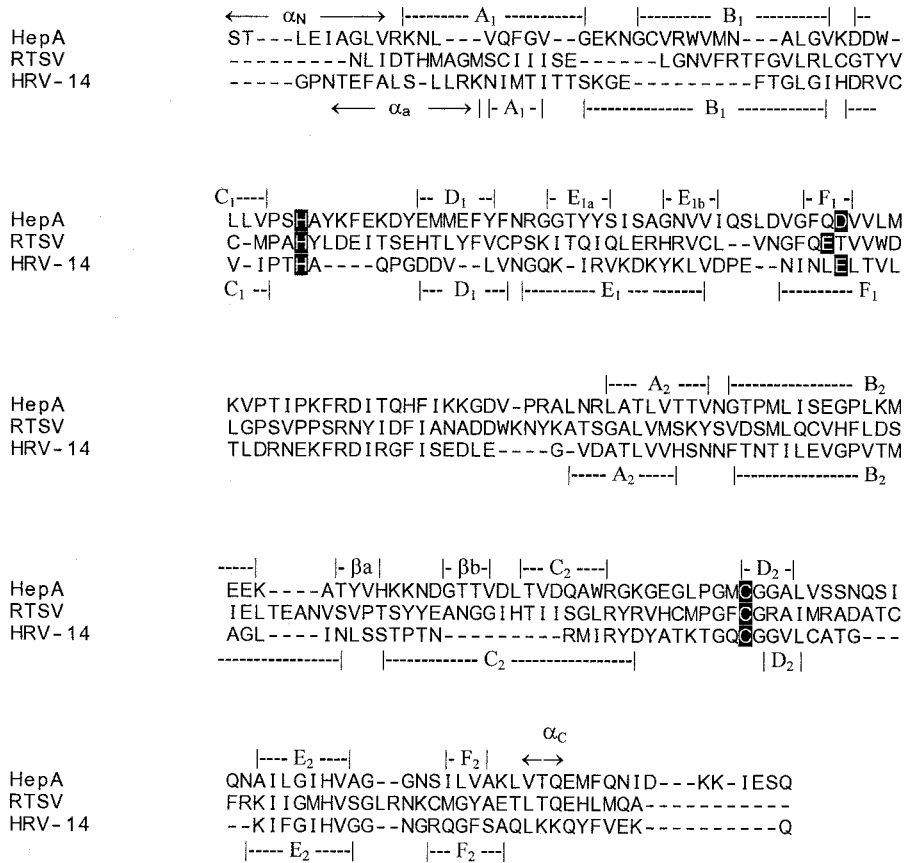


Fig. 10. Alignment of sequi- and picornavirus proteinases. The parsnip yellow fleck virus (PYFV; Turnbull-Ross *et al.*, 1992) 3C-like proteinase sequence was aligned with all available picornavirus 3C<sup>pro</sup> sequences. Only the alignment with the hepatitis A virus and HRV-14 sequences (with the structural features) is shown; active site residues are in white-on-black.

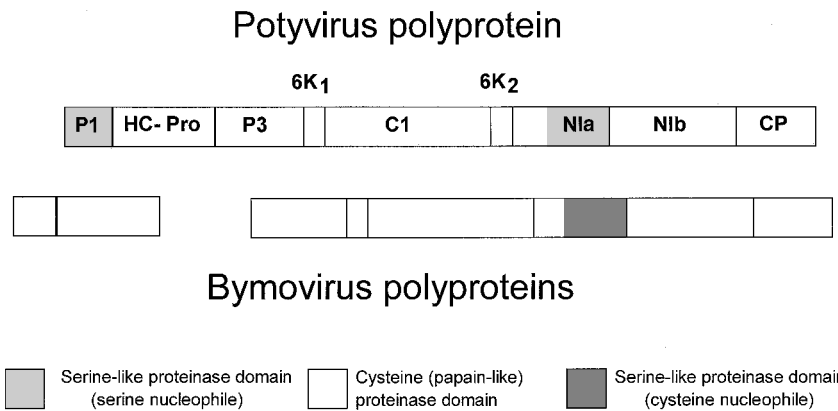


Fig. 11. Polyprotein organization of poty- and bymoviruses. The polyproteins encoded by the monopartite genome of the potyviruses and the bipartite genome of the bymoviruses. Boxed areas indicate mature processing products (virus proteinases or proteinase domains shown by shaded areas).

site nucleophiles Cys<sup>39</sup> and His<sup>113</sup> (numbering scheme used in Fig. 13) in the primary structure is taken as indicating a structural relationship with the thiol papain-like proteinases rather than having a serine proteinase chymotrypsin-like fold. It is interesting to note that the HC proteinase domains of the potyviruses, showing a high level of sequence similarity,

appear to have sequence similarities (-VRD-) with the RNA binding region of the picornavirus 3C proteinases (-V/FRD-; see Fig. 13). The equivalent region of the proteinase domain of the polyprotein encoded by bymovirus RNA-2 does not, however, appear to contain this motif. This could correlate with HC-Pro:RNA binding in *cis* during the replication of



TuMV	S--MFRGLRDYNP I SNN I CHL TNVSDGASNSLYGVG---FGPL I L TNR I L
TEV	S--LFGKPRDYNP I SST I CHL TNESDGHTTSLYGI G---FGPF I I TNK I L
BYMV	TLEASTG I L ---L SQVGV DVATRVGR I F I GTFNMNCYFYSDW I L VPG I L

TuMV	FERNNGELV I KSRHGEFV I KNTTQLHLLP I P---DR I L L L I R L P K D I P P F
TEV	FRRNNGTLL VQSLHGV FV KNTT L L Q Q H L I D---GR I M I I R M P K D F P P F
BYMV	-QDRSGNVT I Q-FPDQTVQT TTDALNANGVKRFYGL I V I A I R R P A I L R P-

TuMV	PQKLGF RQPEKGER I CMVGSNFQTKS I TSVVSETST I MPVENSQFVKHW I
TEV	PQKLKFR EPQREER I CLVT TNFQTKSMSSM VSDT S C T F P S S D G I F V K H W I
BYMV	--RTKL VKAFA I EEPV I AQMV F V D A Q G V R K F T Q S V R A R K E E N S G R W S H K I

TuMV	STKDGG I G S P M V S T K D G K I L G L H S L A N F Q N S I N Y F A A F P D D F A E K Y L H T I
TEV	QTKDGG I G S P L V S T R D G F I V G I H S A S N F T N T N N Y F T S V P K N F M E - L L T N Q
BYMV	STVLGM I G C Q F - W T L E R Q I D G I H V A T N Y T K R N E F Q P F T Q E V V D - F I N G P

\*

TuMV	EAHEWVKHWKYNTSAI SWGSLNIQASQPSGLFKVSKL I S--DL---DSTA
TEV	EAQQWVSGWRLNADSVL WGGHKVFM SKPEEPFQPVKEAT--QL---MNEL
BYMV	GTK I PYCPWV FDRPACGYASHTALFEKPTT L D I I HMQASDGLHNINNA I

TuMV	-----VYAQ
TEV	-----VYSQ
BYMV	EGFGSSLRGQ

Fig. 12(A)

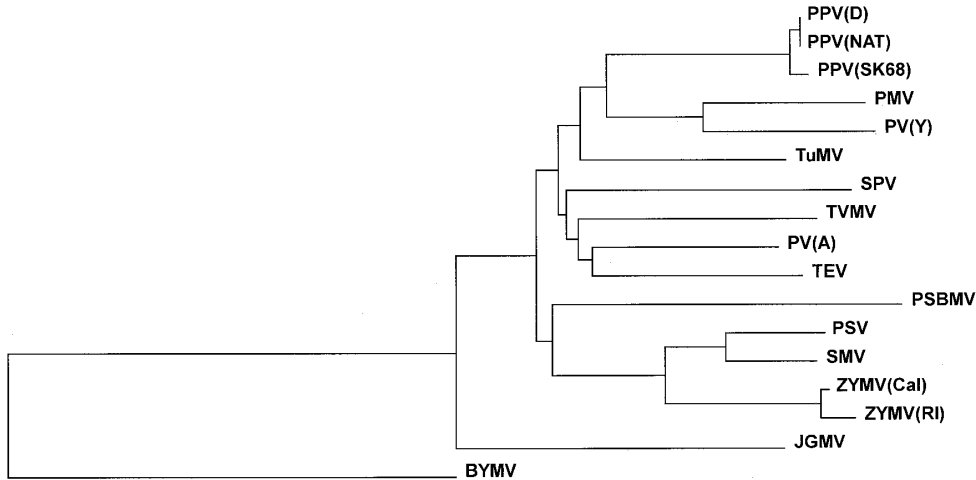


Fig. 12(B)

Fig. 12. Alignment of poty- and bymovirus Nla proteinases. The 27 kDa proteinase domains of the Nla proteins (VPg + 27K proteinase) were aligned using CLUSTALW. Active site residues (and other residues referred to in the text) are indicated by asterisks. The sites of C-terminal 'trimming' of TEV (-M/S-) and TuMV (-S/G-) are shown (arrows). The alignment shown represents that obtained by analysis of all the Nla sequences listed below, only three representative sequences being shown (panel A). The relatedness of all the aligned Nla sequences was analysed using TREECON (van der Peer & de Wachter, 1994; panel B). Sequences analysed: soybean mosaic virus (SMV; Ghabrial, *et al.*, 1990); barley yellow mosaic virus (BYMV; Kashiwazaki *et al.*, 1989); Johnson grass mosaic virus (JGMV; Gough & Shukla, 1993); pea seed-borne mosaic virus (PSBMV; Johansen *et al.*, 1991); peanut stripe virus (PSV; Gunasinghe *et al.*, 1994); pepper mottle virus (PMV; Vance *et al.*, 1992); plum pox virus strain D [PPV(D); Teycheney *et al.*, 1989]; aphid non-transmissible plum pox virus [PPV(NAT); Maiss *et al.*, 1989]; plum pox potyvirus isolate SK68 [PPV(SK68); Palkovics *et al.*, 1993]; Potato virus A [PV(A); Puurand *et al.*, 1994]; potato virus Y [PV(Y); Robaglia *et al.*, 1989]; shallot potyvirus (SPV; S. K. Zavriev and others, unpublished); tobacco etch virus (TEV; Allison *et al.*, 1986); tobacco vein mottling virus (TVMV; Domier *et al.*, 1986); turnip mosaic potyvirus (TuMV; Kim *et al.*, 1995); soybean mosaic virus (SMV; Ghabrial *et al.*, 1990); zucchini yellow mosaic virus strain California [ZYMV(Cal); R. Balint and others, unpublished, accession no. L31350]; zucchini yellow mosaic virus strain Reunion Island [ZYMV(RI); C. A. Baker and others, unpublished, accession no. L29569].



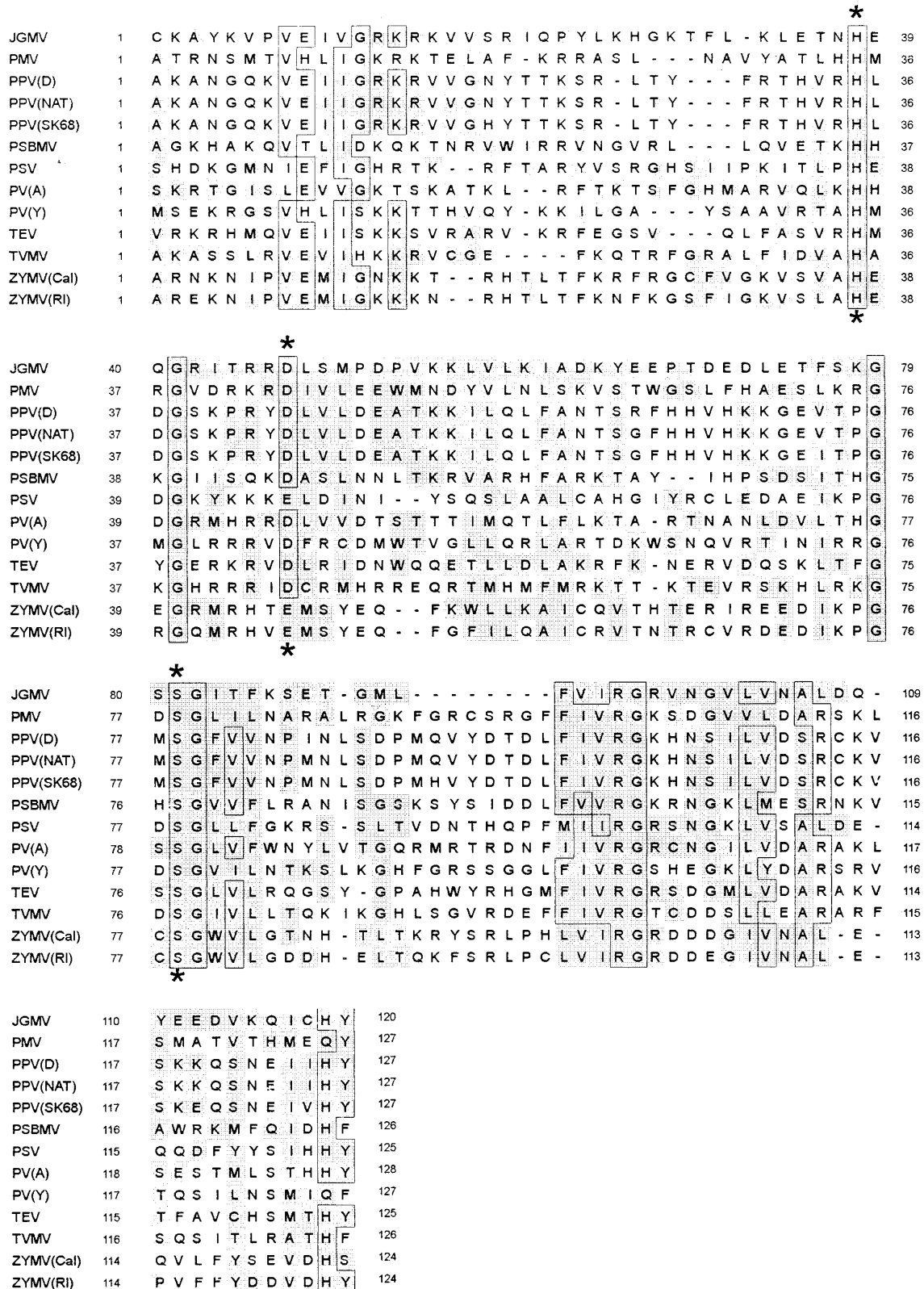


Fig. 14. Alignment of potyvirus P1 (35K) proteinases. The N termini of the sequences were chosen arbitrarily – the C termini correspond to authentic cleavage sites. The catalytic histidine and serine residues are shown together with the position of the conserved acidic residue (asterisks).

RNA since HC-Pro is encoded by RNA-1 (analogous to the picornavirus 3CD:RNA interactions), whereas the equivalent proteolytic domain in bymoviruses is encoded on RNA-2 which is replicated in *trans* and, therefore, is unable to interact with RNA in *cis* as a replicative function.

(iii) **P1 (P35) proteinase.** The existence of a third proteolytic activity in potyvirus polyprotein processing was demonstrated by analysis of the expression of recombinant TEV polyproteins in transgenic tobacco (Carrington *et al.*, 1990). The authors postulated that the cleavage between P35 and HC-Pro was mediated either by a cryptic proteinase or a host factor. Data from the same laboratory confirmed that the 35 kDa protein was a proteinase (Verchot *et al.*, 1991) and identified the residues essential for catalysis (Verchot *et al.*, 1992). The alignment of potyvirus P1 proteinases (Fig. 14) shows the catalytic His<sup>38</sup> and Ser<sup>81</sup> (numbering scheme used in Fig. 14) but the only position where an acidic residue is completely conserved is occupied by aspartate and glutamate (peanut stripe virus and zucchini yellow mosaic virus). The region C-terminal of Ser<sup>81</sup> with low conservation (absent in Johnson grass mosaic virus) would correspond to the prominent surface loop joining  $\beta D_2$  and  $\beta E_2$  of the hepatitis A virus 3C<sup>pro</sup> – a loop in close proximity to the N terminus of the proteinase.

## Concluding remarks

The adoption of the 'polyprotein strategy' reduces the genetic control of protein expression and increases the role of biochemistry in the biogenesis of virus proteins. Encoding proteins in a polyprotein eliminates any potential for the temporal control of the expression of the individual proteins via transcription. Within the picornavirus super-group of viruses it appears, therefore, that subtle control of proteolysis by the adoption of alternative processing pathways (which can lead to alternative sub-sets of biochemical functions from the same precursor molecule) provides an alternative to the other methods of control – primarily read-through of 'leaky' stop codons, ribosomal frame shifting and transcription of sub-genomic mRNAs. The co-evolution of a proteinase with its substrate(s) has large implications for the ability of such viruses to exchange genetic information via recombination since cognate enzyme/substrate pairs must be maintained for virus viability. In addition, the interaction with RNA and other virus proteins will form further barriers to successful recombination events.

The potyvirus P1 proteinase is currently the only proteinase within the picornavirus super-group that uses a serine as the active site nucleophile, all other proteinases using cysteine, either in the form of: (i) a papain-like thiol proteinase (i.e. aphthovirus L<sup>pro</sup>, potyvirus HC-Pro), or (ii) a serine proteinase chymotrypsin-like fold (i.e. picornavirus 3C<sup>pro</sup>, como- and nepovirus 24K and potyvirus NIa proteinases). The nature of the catalytic mechanism of the chymotrypsin-like

cysteine proteinases, even with the solving of two atomic structures, is not fully clear. Conservation of an acidic residue in viruses noted for their high mutation rates implies a strong selection for maintenance. That the acidic residue can be either aspartate or glutamate ( $pK_a$  differing by a pH unit) in the same enzyme from different, but closely related, viruses perhaps implies an important structural role rather than a direct involvement in catalysis – although analysis of sequence alignments further illuminates with what (presumably equally conserved) feature this conserved residue interacts.

The potential of these proteinases as targets for the rational design of anti-virus drugs is enormous and provides disease control strategies which transcend serotypes. The classical approach of designing transition-state analogues for the inhibition of proteolytic activity could now be supplemented by the design of drugs which could inhibit proteinase:RNA interactions or interactions with other virus proteins – essential for the control of proteolytic processing and, quite possibly, other stages in the replication cycle.

The authors wish to gratefully acknowledge Drs E. Bergmann, M. James and co-workers for access to data prior to publication concerning the molecular analysis of the hepatitis A virus 3C proteinase.

## References

- Allaire, M., Cheraia, M. M., Malcolm, B. A. & James, M. N. G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**, 72–76.
- Allison, R., Johnston, R. E. & Dougherty, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for synthesis of a single polyprotein. *Virology* **154**, 9–20.
- Andino, R., Rieckhof, G. E., Trono, D. & Baltimore, D. (1990a). Substitutions in the protease (3C<sup>pro</sup>) gene of poliovirus can suppress a mutation in the 5' noncoding region. *Journal of Virology* **64**, 607–612.
- Andino, R., Rieckhof, G. E. & Baltimore, D. (1990b). A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell* **63**, 369–380.
- Andino, R., Rieckhof, G. E., Achacoso, P. L. & Baltimore, D. (1993). Poliovirus RNA synthesis utilizes an RNP complex formed around the 5' end of viral RNA. *EMBO Journal* **12**, 3587–3598.
- Argos, P., Kamer, G., Nicklin, M. J. H. & Wimmer, E. (1984). Similarity between gene organisation and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. *Nucleic Acids Research* **12**, 7251–7267.
- Batson, S. & Rundell, K. (1991). Proteolysis at the 2A/2B Junction in Theiler's murine encephalomyelitis virus. *Virology* **181**, 764–767.
- Baum, E. Z., Bebernitz, G. A., Palant, O., Mueller, T. & Plotch, S. J. (1991). Purification, properties and mutagenesis of poliovirus 3C protease. *Virology* **185**, 140–150.
- Bazan, J. F. & Fletterick, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like serine proteases: structural and functional implications. *Proceedings of the National Academy of Sciences, USA* **85**, 7872–7876.
- Bazan, J. F. & Fletterick, R. J. (1990). Structural and catalytic models of trypsin-like viral proteases. *Seminars in Virology* **1**, 311–322.
- Belsham, G. J. & Sonenberg, N. (1996). RNA-protein interactions in

- regulation of picornavirus RNA translation. *Microbiological Reviews* **60**, 499–511.
- Bernstein, H. D., Sonenberg, N. & Baltimore, D. (1985).** Poliovirus mutant that does not selectively inhibit host cell protein synthesis. *Molecular and Cellular Biology* **5**, 2913–2923.
- Bernstein, H. D., Sarnow, P. & Baltimore, D. (1986).** Genetic complementation among poliovirus mutants derived from an infectious cDNA clone. *Journal of Virology* **60**, 1040–1049.
- Burroughs, J. N., Sangar, D. V., Clarke, B. E., Rowlands, D. J., Billiau, A. & Collen, D. (1984).** Multiple proteases in foot-and-mouth disease virus replication. *Journal of Virology* **50**, 878–883.
- Cao, X., Bergman, I. E., Fullkrieger, R. & Beck, E. (1995).** Functional analysis of the two alternative initiation sites of foot-and-mouth disease virus. *Journal of Virology* **69**, 560–563.
- Carrington, J. C. & Dougherty, W. G. (1987 a).** Small nuclear inclusion protein encoded by a plant potyvirus genome is a proteinase. *Journal of Virology* **61**, 2540–2548.
- Carrington, J. C. & Dougherty, W. G. (1987 b).** Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology* **160**, 355–362.
- Carrington, J. C. & Dougherty, W. G. (1988).** A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proceedings of the National Academy of Sciences, USA* **85**, 3391–3395.
- Carrington, J. C., Cary, S. M. & Dougherty, W. G. (1988).** Mutational analysis of tobacco etch virus polyprotein processing: cis and trans proteolytic activities of polyproteins containing the 49-kilodalton proteinase. *Journal of Virology* **62**, 2313–2320.
- Carrington, J. C., Cary, S. M., Parks, T. D. & Dougherty, W. G. (1989 a).** A second proteinase encoded by a plant potyvirus genome. *EMBO Journal* **8**, 365–370.
- Carrington, J. C., Freed, D. D. & Sanders, T. C. (1989 b).** Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and in vitro. *Journal of Virology* **63**, 4459–4463.
- Carrington, J. C., Freed, D. D. & Oh, C.-S. (1990).** Expression of potyviral polyproteins in transgenic plants reveals three proteolytic activities required for complete processing. *EMBO Journal* **9**, 1347–1353.
- Cheah, K.-C., Leong, L. E.-C. & Porter, A. G. (1990).** Site-directed mutagenesis suggests close functional relationship between human rhinovirus 3C cysteine protease and cellular trypsin-like serine proteases. *Journal of Biological Chemistry* **265**, 7180–7187.
- Chen, X. & Bruening, G. E. (1992).** Cloned DNA copies of cowpea severe mosaic virus genomic RNAs: infectious transcripts and complete nucleotide sequence of RNA 1. *Virology* **191**, 607–618.
- Choi, G. H., Pawlyk, D. M. & Nuss, D. L. (1991).** The autocatalytic protease p29 encoded by a hypovirulence-associated virus of the chestnut blight fungus resembles the potyvirus-encoded protease HC-Pro. *Virology* **183**, 747–752.
- Clark, M. E., Hammerle, T., Wimmer, E. & Dasgupta, A. (1991).** Poliovirus proteinase 3C converts an active form of transcription factor IIIC to an inactive form: a mechanism for inhibition of host cell polymerase III transcription by poliovirus. *EMBO Journal* **10**, 2941–2947.
- Clark, M. E., Lieberman, P. M., Berk, A. J. & Dasgupta, A. (1993).** Direct cleavage of human TATA-binding protein by poliovirus 3C in vivo and in vitro. *Molecular and Cellular Biology* **13**, 1232–1237.
- Clarke, B. E. & Sangar, D. V. (1988).** Processing and assembly of foot-and-mouth disease virus proteins using subgenomic RNA. *Journal of General Virology* **69**, 2313–2325.
- Clarke, B. E., Sangar, D. V., Burroughs, J. N., Newton, S. E., Carroll, A. R. & Rowlands, D. J. (1985).** Two initiation sites for foot-and-mouth disease virus polyprotein *in vivo*. *Journal of General Virology* **66**, 2615–2626.
- Das, S. & Dasgupta, A. (1993).** Identification of the cleavage site and determinants required for poliovirus 3Cpro-catalysed cleavage of human TATA-binding protein transcription factor TBP. *Journal of Virology* **67**, 3326–3331.
- Davidson, A. D., Proels, M., Schell, J. & Steinbiss, H.-H. (1991).** The nucleotide sequence of RNA 2 of barley yellow mosaic virus. *Journal of General Virology* **72**, 989–993.
- Dessens, J. T. & Lomonosoff, G. P. (1991).** Mutational analysis of the putative catalytic triad of the cowpea mosaic virus 24K protease. *Virology* **184**, 738–746.
- Dessens, J. T. & Lomonosoff, G. P. (1992).** Sequences upstream of the 24K protease enhances cleavage of the cowpea mosaic virus B RNA-encoded polyprotein at the junction between the 24K and 87K proteins. *Virology* **189**, 225–232.
- Dewalt, P. G., Lawson, M. A., Colonna, R. J. & Semler, B. L. (1989).** Chimeric picornavirus polyproteins demonstrate a common 3C proteinase substrate specificity. *Journal of Virology* **63**, 3444–3452.
- Di Francesco, V., Garnier, J. & Munson, P. J. (1996).** Improving protein secondary structure prediction with aligned homologous sequences. *Protein Science* **5**, 106–113.
- Domier, L. L., Franklin, K. M., Shahabuddin, M., Hellmann, G. M., Overmeyer, J. H., Hiremath, S. T., Siaw, M. F. E., Lomonosoff, G. P., Shaw, J. G. & Rhoads, R. E. (1986).** The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Research* **14**, 5417–5430.
- Donnelly, M. L. L., Gani, D., Flint, M., Monaghan, S. & Ryan, M. D. (1997).** The cleavage activities of aphtho- and cardiovirus 2A proteins. *Journal of General Virology* **78**, 13–21.
- Dougherty, W. G. & Parks, T. D. (1989).** Molecular genetic and biochemical evidence for the involvement of the heptapeptide cleavage sequence in determining the reaction profile at two tobacco etch virus cleavage sites in cell-free assays. *Virology* **172**, 145–155.
- Dougherty, W. G. & Parks, T. D. (1991).** Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology* **183**, 449–456.
- Dougherty, W. G. & Semler, B. L. (1993).** Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiological Reviews* **57**, 781–822.
- Dougherty, W. G., Cary, S. M. & Parks, T. D. (1989 a).** Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* **171**, 356–364.
- Dougherty, W. G., Parks, T. D., Cary, S. M., Bazan, J. F. & Fletterick, R. J. (1989 b).** Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology* **172**, 302–310.
- Etchison, D., Milburn, S. C., Ederly, I., Sonenberg, N. & Hershey, J. W. B. (1982).** Inhibition of HeLa cell protein synthesis following poliovirus infection correlates with the proteolysis of a 200,000-dalton polypeptide associated with eukaryotic initiation factor 3 and cap binding protein complex. *Journal of Biological Chemistry* **257**, 14806–14810.
- Falk, M. M., Grigera, P. R., Bergmann, I. E., Zibert, A., Multhaupt, G. & Beck, E. (1990).** Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host histone H3. *Journal of Virology* **64**, 748–756.
- Felsenstein, J. (1991).** PHYLIP manual version 3.4. University Herbarium, University of California, Berkeley, California.
- Franssen, H., Goldbach, R. & van Kammen, A. (1984 a).** Translation of

- bottom component RNA of cowpea mosaic virus in reticulocyte lysate: faithful proteolytic processing of the primary translation product. *Virus Research* **1**, 39–49.
- Franssen, H., Leunissen, J., Goldbach, R., Lomonosoff, G. P. & Zimmern, D. (1984b).** Homologous sequences in nonstructural proteins from cowpea mosaic virus and picornaviruses. *EMBO Journal* **3**, 855–861.
- Garcia, J. A., Lain, S., Cervera, M. T., Riechmann, J. L. & Martin, M. T. (1990).** Mutational analysis of plum pox potyvirus polyprotein processing by the NIa protease in *Escherichia coli*. *Journal of General Virology* **71**, 2773–2779.
- Ghabrial, S. A., Smith, H. A., Parks, T. D. & Dougherty, W. G. (1990).** Molecular genetic analyses of the soybean mosaic virus NIa proteinase. *Journal of General Virology* **71**, 1921–1927.
- Gorbalenya, A. E. & Svitkin, Y. V. (1983).** Encephalomyocarditis virus protease: purification and role of the SH groups in processing of the precursor of structural proteins. *Biochemistry (USSR)* **48**, 385–395.
- Gorbalenya, A. E., Svitkin, Y. V., Kazachkov, Y. A. & Agol, V. I. (1979).** Encephalomyocarditis virus-specific polypeptide p22 is involved in the processing of the viral precursor polypeptides. *FEBS Letters* **108**, 1–5.
- Gorbalenya, A. E., Blinov, V. M. & Donchenko, A. P. (1986).** Poliovirus-encoded proteinase 3C: a possible evolutionary link between cellular serine and cysteine proteinase families. *FEBS Letters* **194**, 253–257.
- Gorbalenya, A. E., Koonin, E. V., Blinov, V. M. & Donchenko, A. P. (1988).** Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Letters* **236**, 287–290.
- Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989a).** Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Letters* **243**, 103–114.
- Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989b).** N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. *Nucleic Acids Research* **17**, 3889–3897.
- Gorbalenya, A. E., Koonin, E. V. & 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 Letters* **288**, 201–205.
- Gough, K. H. & Shukla, D. D. (1993).** The nucleotide sequence of Johnson grass mosaic potyvirus genomic RNA. *Intervirology* **36**, 181–192.
- Greif, C., Hemmer, O & Fritsch, C. (1988).** Nucleotide sequence of tomato black ring virus RNA-1. *Journal of General Virology* **69**, 1517–1529.
- Grubman, M. J., Zellner, M., Bablanian, G., Mason, P. W. & Piccone, M. E. (1995).** Identification of the active-site residues of the 3C proteinase of foot-and-mouth disease virus. *Virology* **213**, 581–589.
- Gunasinghe, U. B., Flasiniski, S., Nelson, R. S. & Cassidy, B. G. (1994).** Nucleotide sequence and genome organization of peanut stripe potyvirus. *Journal of General Virology* **75**, 2519–2526.
- Hambidge, S. J. & Sarnow, P. (1992).** Translational enhancement of the poliovirus 5' noncoding region mediated by virus-encoded polypeptide 2A. *Proceedings of the National Academy of Sciences, USA* **89**, 10272–10276.
- Hammerle, T., Hellen, C. U. T. & Wimmer, E. (1991).** Site-directed mutagenesis of the putative catalytic triad of poliovirus 3C proteinase. *Journal of Biological Chemistry* **266**, 5412–5416.
- Hanecak, R., Semler, B. L., Anderson, C. W. & Wimmer, E. (1982).** Proteolytic processing of poliovirus polypeptides: antibodies to polypeptide P3–7c inhibit cleavage at glutamine glycine pairs. *Proceedings of the National Academy of Sciences, USA* **79**, 3973–3977.
- Hans, F. & Sanfaçon, H. (1995).** Tomato ringspot nepovirus protease: characterization and cleavage site specificity. *Journal of General Virology* **76**, 917–927.
- Harmon, S. A., Updike, W., Jia, X., Summers, D. F. & Ehrenfeld, E. (1992).** Polyprotein processing in cis and in trans by hepatitis A virus 3C protease cloned and expressed in *Escherichia coli*. *Journal of Virology* **66**, 5242–5247.
- Harris, K. S., Xiang, W., Alexander, L., Lane, W. S., Paul, A. V. & Wimmer, E. (1994).** Interaction of poliovirus polypeptide 3CD<sup>pro</sup> with the 5' and 3' termini of the poliovirus genome. *Journal of Biological Chemistry* **269**, 27004–27014.
- Hellen, C. U. T., Krausslich, H.-G. & Wimmer, E. (1989).** Proteolytic processing of polyproteins in the replication of RNA viruses. *Biochemistry* **28**, 9881–9890.
- Hellen, C. U. T., Facke, M., Krausslich, H., Lee, C. & Wimmer, E. (1991).** Characterisation of poliovirus 2A proteinase by mutational analysis: residues required for autocatalytic activity are essential for induction of eukaryotic initiation factor 4F polypeptide p220. *Journal of Virology* **65**, 4226–4231.
- Hellen, C. U. T., Lee, C. & Wimmer, E. (1992).** Determinants of substrate recognition by poliovirus 2A proteinase. *Journal of Virology* **66**, 3330–3338.
- Hellmann, G. M., Shaw, J. G. & Rhoads, R. E. (1988).** *In vitro* analysis of tobacco vein mottling virus NIa cistron: evidence for a virus-encoded protease. *Virology* **163**, 554–562.
- Higgins, D. E., Bleasby, A. J. & Fuchs, R. (1991).** CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences* **8**, 189–191.
- Jackson, R. J. (1986).** A detailed kinetic analysis of the *in vitro* synthesis and processing of encephalomyocarditis virus products. *Virology* **149**, 114–127.
- Jia, X.-Y., Ehrenfeld, E. & Summers, D. F. (1991).** Proteolytic activity of hepatitis A virus 3C protein. *Journal of Virology* **65**, 2595–2600.
- Jia, X.-Y., Summers, D. F. & Ehrenfeld, E. (1993).** Primary cleavage of the HAV capsid protein precursor in the middle of the proposed 2A coding region. *Virology* **193**, 515–519.
- Joachims, M., Harris, K. S. & Etchison, D. (1995).** Poliovirus protease 3C mediates cleavage of microtubule-associated protein 4. *Virology* **211**, 451–461.
- Johansen, E., Rasmussen, O. F., Heide, M. & Borkhardt, B. (1991).** The complete nucleotide sequence of pea seed-borne mosaic virus RNA. *Journal of General Virology* **72**, 2625–2632.
- Jore, J., De Geus, B., Jackson, R. J., Pouwels, P. H. & Enger-Valk, B. E. (1988).** Poliovirus protein 3CD is the active protease for processing of the precursor protein P1 *in vitro*. *Journal of General Virology* **69**, 1627–1636.
- Kashiwazaki, S., Hayano, Y., Minobe, Y., Omura, T., Hibino, H. & Tsuchizaki, T. (1989).** Nucleotide sequence of the capsid protein gene of barley yellow mosaic virus. *Journal of General Virology* **70**, 3015–3023.
- Kashiwazaki, S., Minobe, Y. & Hibino, H. (1991).** Nucleotide sequence of barley yellow mosaic virus RNA 2. *Journal of General Virology* **72**, 995–999.
- Kay, J. & Dunn, B. M. (1990).** Viral proteinases: weakness in strength. *Biochimica et Biophysica Acta* **1048**, 1–18.
- Kean, K. M., Tetrina, N. L., Marc, D. & Girad, M. (1991).** Analysis of putative active site residues of the poliovirus 3C protease. *Virology* **163**, 330–340.

- Kean, K. M., Howell, M. T., Grunert, S., Girard, M. & Jackson, R. J. (1993). Substitution mutations at the putative catalytic triad of the poliovirus 3C protease have differential effects on cleavage at different sites. *Virology* **194**, 360–364.
- Kim, D.-H., Park, Y. S., Kim, S. S., Lew, J., Nam, H. G. & Choi, K. Y. (1995). Expression, purification, and identification of a novel self-cleavage site of the N1a C-terminal 27-kDa protease of turnip mosaic potyvirus C5. *Virology* **213**, 517–525.
- Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H.-D., Sommergruber, W., Sobrino, F., Hohenadl, C., Blaas, D., Rhoads, R. E. & Skern, T. (1994). Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4 $\gamma$ . *Journal of Virology* **68**, 5677–5684.
- Konig, H. & Rosenwirth, B. (1988). Purification and partial characterization of poliovirus protease 2A by means of a functional assay. *Journal of Virology* **62**, 1243–1250.
- Korant, B. D. (1972). Cleavage of viral precursor proteins in vivo and in vitro. *Journal of Virology* **10**, 751–759.
- Korant, B. D. (1973). Cleavage of poliovirus-specific polypeptide aggregates. *Journal of Virology* **12**, 556–563.
- Korant, B. D., Chow, N., Lively, M. & Powers, J. (1979). Virus-specified protease in poliovirus-infected HeLa cells. *Proceedings of the National Academy of Sciences, USA* **76**, 2992–2995.
- Korant, B. D., Brzin, J. & Turk, V. (1985). Cystatin, a protein inhibitor of cysteine proteases alters viral protein cleavages in infected human cells. *Biochemical and Biophysical Research Communications* **127**, 1072–1076.
- Lain, S., Riechmann, J. L. & Garcia, J. A. (1989). The complete nucleotide sequence of plum pox potyvirus RNA. *Virus Research* **13**, 157–172.
- Lamphear, B. J., Yan, R., Yang, F., Waters, D., Liebig, H.-D., Klump, H., Keuchler, E., Skern, T. & Rhoads, R. E. (1993). Mapping of the cleavage site in protein synthesis initiation factor eIF-4 $\gamma$  of the 2A proteases from human coxsackievirus and rhinovirus. *Journal of Biological Chemistry* **268**, 19200–19203.
- Lawrence, C. & Thatch, R. E. (1975). Identification of a viral protein involved in post-translational maturation of the encephalomyocarditis virus capsid precursor. *Journal of Virology* **15**, 918–928.
- Lawson, M. A. & Semler, B. L. (1990). Picornavirus protein processing – enzymes, substrates, and genetic regulation. *Current Topics in Microbiology and Immunology* **161**, 49–87.
- Lawson, M. A. & Semler, B. L. (1991). Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. *Proceedings of the National Academy of Sciences, USA* **88**, 9919–9923.
- Lawson, T. G., Smith, L. L., Palmenberg, A. C. & Thach, R. E. (1989). Inducible expression of encephalomyocarditis virus 3C protease activity in stably transformed mouse cell lines. *Journal of Virology* **63**, 5013–5022.
- Lee, C.-K. & Wimmer, E. (1988). Proteolytic processing of poliovirus polyprotein: elimination of 2A<sup>pro</sup>-mediated, alternative cleavage of polypeptide 3CD by *in vitro* mutagenesis. *Virology* **166**, 405–414.
- Le Gall, O., Candresse, T., Brault, V. & Dunez, J. (1989). Nucleotide sequence of Hungarian grapevine chrome mosaic nepovirus RNA1. *Nucleic Acids Research* **17**, 7795–7807.
- Leong, L. E. C., Walker, P. A. & Porter, A. G. (1993). Human rhinovirus-14 protease 3C (3C<sup>pro</sup>) binds specifically to the 5′-noncoding region of the viral RNA. *Journal of Biological Chemistry* **268**, 25735–25739.
- Lesk, A. M. & Fordham, W. D. (1996). Conservation and variability in the structures of serine proteinases of the chymotrypsin family. *Journal of Molecular Biology* **258**, 501–537.
- Liao, D.-I. & Remington, S. J. (1990). Structure of wheat serine carboxypeptidase II at 3.5-Å resolution. *Journal of Biological Chemistry* **265**, 6528–6531.
- Lomonosoff, G. P. & Shanks, M. (1983). The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO Journal* **2**, 2253–2258.
- Long, L. A., Orr, D. C., Cameron, J. M., Dunn, B. M. & Kay, J. (1989). A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase. *FEBS Letters* **258**, 75–78.
- Macadam, A. J., Ferguson, G., Fleming, T., Stone, D. M., Almond, J. W. & Minor, P. D. (1994). Role for poliovirus protease 2A in cap independent translation. *EMBO Journal* **13**, 924–927.
- McLean, C., Matthews, T. J. & Rueckert, R. R. (1976). Evidence of ambiguous processing and selective degradation in the noncapsid proteins of rhinovirus 1a. *Journal of Virology* **19**, 903–914.
- Maia, I. G., Haenni, A.-L. & Bernardi, F. (1996). Potyviral HC-Pro: a multifunctional protein. *Journal of General Virology* **77**, 1335–1341.
- Maiss, E., Timpe, U., Briske, A., Jelkmann, W., Casper, R., Himmler, G., Mattanovich, D. & Katinger, H. W. D. (1989). The complete nucleotide sequence of plum pox virus RNA. *Journal of General Virology* **70**, 513–524.
- Marcos, J. F. & Beachy, R. N. (1994). In vitro characterization of a cassette to accumulate multiple proteins through synthesis of a self-processing polypeptide. *Plant Molecular Biology* **24**, 495–503.
- Margis, R. & Pinck, L. (1992). Effects of site-directed mutagenesis on the presumed catalytic triad and substrate-binding pocket of grapevine fanleaf nepovirus 24-kDa proteinase. *Virology* **190**, 884–888.
- Margis, R., Viry, M., Pinck, M. & Pinck, L. (1991). Cloning and *in vitro* characterisation of the grapevine fanleaf virus proteinase cistron. *Virology* **185**, 779–787.
- Martinez-Salas, E. & Domingo, E. (1995). Effect of expression of the aphthovirus protease 3C on viral infection and gene expression. *Virology* **212**, 111–120.
- Matthews, 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. & 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., Brangwyn, J. K. & 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.
- Menard, R., Plouffe, C., Lafamme, P., Vernet, T., Tessier, D. C., Thomas, D. Y. & Storer, A. C. (1995). Modification of the electrostatic environment is tolerated in the oxyanion hole of the cysteine proteinase papain. *Biochemistry* **34**, 464–471.
- Molla, A., Paul, A. V., Schmid, M., Jang, S. K. & Wimmer, E. (1993). Studies on dicistronic polioviruses implicate viral proteinase 2A<sup>pro</sup> in RNA replication. *Virology* **196**, 739–747.
- Molla, A., Harris, K. S., Paul, A. V., Shin, S. H., Mugavero, J. & Wimmer, E. (1994). Stimulation of poliovirus proteinase 3C<sup>pro</sup>-related proteolysis by the genome-linked protein VPg and its precursor 3AB. *Journal of Biological Chemistry* **269**, 27015–27020.
- Palkovics, L., Burgyan, J. & Balazs, E. (1993). Comparative sequence analysis of four complete primary structures of plum pox virus strains. *Virus Genes* **7**, 339–347.
- Palmenberg, A. C. (1990). Proteolytic processing of picornaviral polyprotein. *Annual Review of Microbiology* **44**, 603–623.
- Palmenberg, A. C., Pallansch, M. A. & Rueckert, R. R. (1979). Protease required for processing picornaviral coat protein residues resides in the viral replicase gene. *Journal of Virology* **32**, 770–778.

- Palmenberg, A. C., Parks, G. D., Hall, D. J., Ingraham, R. H., Seng, T. W. & Pallai, P. V. (1992). Proteolytic processing of the cardioviral P2 region: primary 2A/2B cleavage in clone-derived precursors. *Virology* **190**, 754–762.
- Parks, G. D., Baker, J. C. & Palmenberg, A. C. (1989). Proteolytic cleavage of encephalomyocarditis virus capsid region substrates by precursors to the 3C enzyme. *Journal of Virology* **63**, 1054–1058.
- Parks, T. D., Smith, H. A. & Dougherty, W. G. (1992). Cleavage profiles of tobacco etch virus (TEV)-derived substrates mediated by precursor and processed forms of the TEV N1a proteinase. *Journal of General Virology* **73**, 149–155.
- Parks, T. D., Howard, E. D., Wolpert, T. J., Arp, D. J. & Dougherty, W. G. (1995). Expression and purification of a recombinant tobacco etch virus N1a proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology* **120**, 194–201.
- Pelham, H. R. B. (1978). Translation of encephalomyocarditis virus RNA *in vitro* yields an active proteolytic processing enzyme. *European Journal of Chemistry* **85**, 457–462.
- Peters, S. A., Voorhorst, W. G. B., Wery, J., Wellink, J. & van Kammen, A. (1992). A regulatory role for the 32K protein in proteolytic processing of cowpea mosaic virus proteins. *Virology* **191**, 81–89.
- Piccone, M. E., Zellner, M., Kumosinski, T. F., Mason, P. W. & Grubman, M. J. (1995). Identification of the active-site residues of the L proteinase of foot-and-mouth disease virus. *Journal of Virology* **69**, 4950–4956.
- Puurand, Ü., Mäkinen, K., Paulin, L. & Saarma, M. (1994). The nucleotide sequence of potato virus A genomic RNA and its sequence similarities with other potyviruses. *Journal of General Virology* **75**, 457–461.
- Riechmann, J. L., Lain, S. & Garcia, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology* **73**, 1–16.
- Ritzenthaler, C., Viry, M., Pinck, M., Margis, R., Fuchs, M. & Pinck, L. (1991). Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. *Journal of General Virology* **72**, 2357–2365.
- Robaglia, C., Durand-Tardif, M., Tronchet, M., Boudazin, G., Astier-Manificier, S. & Casse-Delbart, F. (1989). Nucleotide sequence of potato virus Y (N strain) genomic RNA. *Journal of General Virology* **70**, 935–947.
- Roberts, P. J. & 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.
- Roos, R. P., Kong, W. & Semler, B. L. (1989). Polyprotein processing of Theilers murine encephalomyelitis virus. *Journal of Virology* **63**, 5344–5353.
- Rott, M. E., Gilchrist, A., Lee, L. & Rochon, D. (1995). Nucleotide sequence of tomato ringspot virus RNA 1. *Journal of General Virology* **76**, 465–473.
- Rueckert, R. R., Matthews, T. J., Kew, O. M., Pallansch, M. A., McLean, C. & Omilianowski, D. R. (1979). In *The Molecular Biology of Picornaviruses*, Chapter 6, pp. 113–125. Edited by R. Perez-Bercoff. New York: Plenum Press.
- Ryan, M. D. & Drew, J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO Journal* **13**, 928–933.
- Ryan, M. D., Belsham, G. J. & King, A. M. Q. (1989). Specificity of substrate-enzyme interactions in foot-and-mouth disease virus polyprotein processing. *Virology* **173**, 35–45.
- Ryan, M. D., King, A. M. Q. & Thomas, G. P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *Journal of General Virology* **72**, 2727–2732.
- Sangar, D. V., Newton, S. E., Rowlands, D. J. & Clarke, B. E. (1987). All FMDV serotypes initiate protein synthesis at two separate AUGs. *Nucleic Acids Research* **15**, 3305–3315.
- Sangar, D. V., Clark, R. P., Carroll, A. R., Rowlands, D. J. & Clarke, B. E. (1988). Modification of the leader protein (Lb) of foot-and-mouth disease virus. *Journal of General Virology* **69**, 2327–2333.
- Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. papain. *Biochemical and Biophysical Research Communications* **27**, 157–162.
- Schultheiss, T., Kusov, Y. Y. & Gauss-Muller, 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. & Gauss-Muller, V. (1995a). Polyprotein processing in echovirus 22 – a first assessment. *Biochemical and Biophysical Research Communications* **219**, 1120–1127.
- Schultheiss, T., Sommergruber, W., Kusov, Y. & Gauss-Muller, V. (1995b). Cleavage specificity of purified recombinant hepatitis A virus 3C proteinase on natural substrates. *Journal of Virology* **69**, 1727–1733.
- Shanks, M. & Lomonosoff, G. P. (1990). The primary structure of the 24K protease from red clover mottle virus: implications for the mode of action of comovirus proteases. *Journal of General Virology* **71**, 735–738.
- Shen, P., Kaniewska, M., Smith, C. & Beachy, R. (1993). Nucleotide sequence and genomic organization of rice tungro spherical virus. *Virology* **193**, 621–630.
- Skern, T., Sommergruber, W., Auer, H., Volkmann, P., Zorn, M., Liebig, H.-D., Fessl, F., Blaas, D. & Keuchler, E. (1991). Substrate requirements of a human rhinoviral 2A proteinase. *Virology* **181**, 46–54.
- Sommergruber, W., Zorn, M., Blaas, D., Fessl, F., Volkmann, P., Maurer-Fogy, I., Pallai, P., Merluzzi, V., Matteo, M., Skern, T. & Keuchler, E. (1989). Polypeptide 2A of human rhinovirus type 2: identification as a protease and characterisation by mutational analysis. *Virology* **169**, 68–77.
- Sommergruber, W., Casari, G., Fessl, F., Seipelt, J. & Skern, T. (1994a). The 2A proteinase of human rhinovirus is a zinc containing enzyme. *Virology* **204**, 815–818.
- Sommergruber, W., Ahorn, H., Klump, H., Seipelt, J., Zoepfel, A., Fessl, F., Krystek, E., Blaas, D., Keuchler, E., Liebig, H.-D. & Skern, T. (1994b). 2A proteinases of coxsackie- and rhinovirus cleave peptides derived from eIF-4 $\gamma$  via a common recognition motif. *Virology* **198**, 741–745.
- Sonenberg, N. (1990). Poliovirus translation. *Current Topics in Microbiology and Immunology* **161**, 23–42.
- Sreenivasan, U. & Axelsen, P. H. (1992). Buried water in homologous serine proteases. *Biochemistry* **31**, 2785–2791.
- Stanway, G. (1990). Structure, function and evolution of picornaviruses. *Journal of General Virology* **71**, 2483–2501.
- Steitz, T. A. & Schulman, R. G. (1982). Crystallographic and NMR studies of the serine proteinases. *Annual Review of Biophysics and Bioengineering* **11**, 419–444.
- Strebel, K. & Beck, E. (1986). A second protease of foot-and-mouth disease virus. *Journal of Virology* **58**, 893–899.
- Summers, D. F. & Maizel, J. V. (1968). Evidence for large precursor proteins in poliovirus synthesis. *Proceedings of the National Academy of Sciences, USA* **59**, 966–971.
- Summers, D. F., Shaw, E. N., Stewart, M. L. & Maizel, J. V. (1972).



- Inhibition of cleavage of large poliovirus-specific precursor proteins in infected HeLa cells by inhibitors of proteolytic enzymes. *Journal of Virology* **10**, 880–884.
- Tesar, M. & Marquardt, O. (1990).** Foot-and-mouth disease virus protease 3C inhibits cellular transcription and mediates cleavage of histone H3. *Virology* **174**, 364–374.
- Teycheney, P. Y., Tavert, G., Delbos, R., Ravelonandro, M. & Dunez, J. (1989).** The complete nucleotide sequence of plum pox virus RNA (strain D). *Nucleic Acids Research* **17**, 10115–10116.
- Thatch, R. E. (1992).** Cap-recap: the involvement of eIF-4F in regulating gene expression. *Cell* **68**, 177–180.
- Toyoda, H., Nicklin, M. J. H., Murray, M. G., Anderson, C. W., Dunn, J. J., Studier, F. W. & Wimmer, E. (1986).** A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell* **45**, 761–770.
- Turnbull-Ross, A. D., Reavy, B., Mayo, M. A. & Murant, A. F. (1992).** The nucleotide sequence of parsnip yellow fleck virus: a plant picornalike virus. *Journal of General Virology* **73**, 3203–3211.
- Vakharia, V. N., DeVaney, M. A., Moore, D. M., Dunn, J. J. & Grubman, M. J. (1987).** Proteolytic processing of foot-and-mouth disease virus polyproteins expressed in a cell-free system from clone derived transcripts. *Journal of Virology* **61**, 3199–3207.
- van der Peer, Y. & de Wachter, R. (1994).** TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer Applications in the Biosciences* **10**, 569–570.
- Vance, V. B. & Beachy, R. N. (1984).** Translation of soybean mosaic virus RNA *in vitro*: evidence of protein processing. *Virology* **132**, 271–281.
- Vance, V. B., Moore, D., Turpen, T. H., Bracker, A. & Hollowell, V. C. (1992).** The complete nucleotide sequence of pepper mottle virus genomic RNA: comparison of the encoded polyprotein with those of other sequenced potyviruses. *Virology* **191**, 19–30.
- Verchot, J., Koonin, E. & Carrington, J. C. (1991).** The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology* **185**, 527–535.
- Verchot, J., Herndon, K. L. & Carrington, J. C. (1992).** Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: identification of essential residues and requirements for autoproteolysis. *Virology* **190**, 298–306.
- Verver, J., Goldbach, R., Garcia, J. A. & Vos, P. (1987).** *In vitro* expression of a full-length DNA copy of cowpea mosaic virus B RNA: identification of the B RNA-encoded 24 kDa protein as a viral protease. *EMBO Journal* **6**, 549–554.
- Vos, P., Verver, J., Jaegle, M., Wellink, J., van Kammen, A. & Goldbach, R. (1988).** Two viral proteins involved in the proteolytic processing of cowpea mosaic virus polyproteins. *Nucleic Acids Research* **16**, 1967–1985.
- Voss, T., Meyer, R. & Sommergruber, W. (1995).** Spectroscopic characterization of rhinoviral protease 2A: Zn is essential for the structural integrity. *Protein Science* **4**, 2526–2531.
- Walker, P. A., Leong, L. E. C. & Porter, A. G. (1995).** Sequence and structural determinants of the interaction between the 5'-noncoding region of picornavirus RNA and rhinovirus protease 3C. *Journal of Biological Chemistry* **270**, 14510–14516.
- Warshel, A., Naray-Szabo, G., Sussman, F. & Hwang, J. (1989).** How do serine proteinases really work? *Biochemistry* **28**, 3629–3637.
- Wellink, J. & van Kammen, A. (1988).** Proteases involved in the processing of viral polyproteins. *Archives of Virology* **98**, 1–26.
- Whiting, A. K. & Peticolas, W. L. (1994).** Details of the acyl-enzyme intermediate and the oxyanion hole in serine protease catalysis. *Biochemistry* **33**, 552–561.
- Yeh, S.-P. & Gonsalves, D. (1985).** Translation of papaya ringspot virus RNA *in vitro*: detection of a possible polyprotein that is processed for capsid protein, cylindrical inclusion protein and amorphous-inclusion protein. *Virology* **143**, 260–271.
- Ypma-Wong, M. F., Dewalt, P. G., Johnson, V. H., Lamb, J. G. & Semler, B. L. (1988).** Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology* **166**, 265–270.
- Yu, S. F. & Lloyd, R. E. (1991).** Identification of essential amino acid residues in the functional activity of poliovirus 2A protease. *Virology* **182**, 615–625.
- Yu, S. F. & Lloyd, R. E. (1992).** Characterization of the roles of conserved cysteine and histidine residues in poliovirus 2A protease. *Virology* **186**, 725–735.
- Yu, S. F., Benton, P., Bovee, M., Sessions, J. & Lloyd, R. E. (1995).** Defective RNA replication by poliovirus mutants deficient in 2A protease cleavage activity. *Journal of Virology* **69**, 247–252.