Journal of General Virology (1999), **80**, 1879–1888. Printed in Great Britain

REVIEW ARTICLE

Virus-encoded proteinases of the *Togaviridae*

Edwin ten Dam,[1] Michael Flint[2] and Martin D. Ryan[1]

[1] School of Biomedical Sciences, University of St Andrews, Biomolecular Sciences Building, North Haugh, St Andrews KY16 9ST, UK
[2] School of Animal and Microbial Sciences, University of Reading, Whiteknights, PO Box 228, Reading RG6 6AJ, UK

Introduction

The *Togaviridae* are enveloped, positive-sense, single-stranded RNA viruses. The family is divided into two genera; the alphaviruses, typified by Sindbis virus (genome ~ 11·7 kb) and the rubiviruses, typified by rubella virus (genome ~ 9·7 kb). A major feature of the replication of toga viruses is the synthesis of a single sub-genomic RNA transcript, corresponding to the 3′ region of the genome (alphaviruses ~ 4·1 kb or '26S' RNA; rubiviruses ~ 3·3 kb; see Fig. 1). This mRNA transcript encodes a polyprotein comprising proteins involved in the encapsidation and envelopment of the virus. Togavirus non-structural (NS) proteins are encoded in the 5′ region of genome-length RNA, also in the form of a polyprotein (Fig. 1).

The ability of togaviruses to synthesize a sub-genomic RNA is in contrast with the viruses which were discussed previously in this series of reviews: the *Picornaviridae* (Ryan & Flint, 1997) and the *Flaviviridae* (Ryan *et al.*, 1998). In both the *Picorna-* and *Flaviviridae* all of their proteins are encoded in a single, long, open reading frame (ORF) with the virus proteins being generated by proteolytic 'processing' of the polyprotein. This processing is accomplished by virus-encoded proteinases: either wholly (*Picornaviridae*) or in combination with cellular proteinases (*Flaviviridae*). Whereas for picorna- and flaviviruses control over the biogenesis of virus proteins is solely post-transcriptional (by regulation of proteolytic processing), the togaviruses are able to regulate the biogenesis of encapsidation and replicative proteins in both a temporal and quantitative manner via transcription of the sub-genomic mRNA. Unlike the picorna- and flaviviruses, translation of togavirus replicative proteins does not obligate the translation of encapsidation functions, nor vice versa. The polyproteins derived from either genomic or sub-genomic RNAs are proteolytically processed: in the former by a proteinase domain contained within the NS polyprotein itself, in the latter by a combination of virus and host-cell proteolytic activities (alphavirus) or host-cell proteinases alone (rubivirus).

Reviews are available on other members of the alphavirus 'super-group': the caliciviruses (Clarke & Lambden, 1997) and the arteriviruses (Snijder & Meulenberg, 1998), together with the alpha- and rubiviruses themselves (Strauss & Strauss, 1994; Frey, 1994). The reader is directed to reviews on virus-encoded proteinases in general (Dougherty & Semler, 1993) and viral cysteine proteinases in particular (Gorbalenya & Snijder, 1996). A further resource is the comprehensive and systematic database of proteinases which classifies proteinases into clans and families ('MEROPS'; accessible on the worldwide web at http://www.bi.bbsrc.ac.uk/Merops/Merops.htm). This database also provides links to sequence databases and 3-D images. The proteinase nomenclature used here is from Barrett *et al.* (1998). Sequence alignments described in this paper are available from ftp://ftp.st-and.ac.uk/pub/mdr1 together with alignments from previous reviews.

Author for correspondence: Martin Ryan.
Fax +44 1334 462595. e-mail martin.ryan@st-and.ac.uk

Togavirus non-structural polyproteins

The alphavirus 5′-proximal long ORF encodes the replicative functions of the virus and is translated as an [nsP1–3] polyprotein. At low frequency (5–20% – temperature dependent) an amino acid may incorporated at the site of the 'leaky' opal (UGA) stop codon at the end of nsP3 (translational 'readthrough'), producing [nsP1–4] (de Groot *et al.*, 1990; Li & Rice, 1993; Strauss *et al.*, 1983). Although the identity of the amino acid incorporated at this site has not been confirmed by protein sequencing, it is likely to be arginine, cysteine or tryptophan by analogy with opal codon readthrough in retroviruses (Feng *et al.*, 1990). When readthrough has produced [nsP1–4], proteolytic processing may occur at a site (nsP3/nsP4) six amino acids downstream of the amino acid incorporated at the UGA codon (Fig. 2; reviewed by Strauss & Strauss, 1994). In certain alphaviruses this stop codon is absent altogether and a sense codon [CGA (arginine) or UGU (cysteine)] is found. The polyprotein is processed into the 'mature' products: nsP1–nsP4. The rubivirus 5′-proximal long ORF also encodes the replicative functions of the virus. In this case, however, three translation products are observed; a 200 kDa protein (p200), which may then be processed into two products – p150 (1301 aa; mol. mass 140 kDa) and p90 (904 aa; mol. mass 100 kDa) – at the Gly^{1301}/Gly^{1302} scissile pair (Chen *et al.*, 1996; Fig. 1). The order of the gene functions in the NS polyprotein is different between alpha- and rubiviruses. In alphaviruses functions have been ascribed as (N- to C-terminal): methyltransferase (nsP1), helicase/proteinase (nsP2), unknown 'X' motif (nsP3) and replicase (nsP4). In

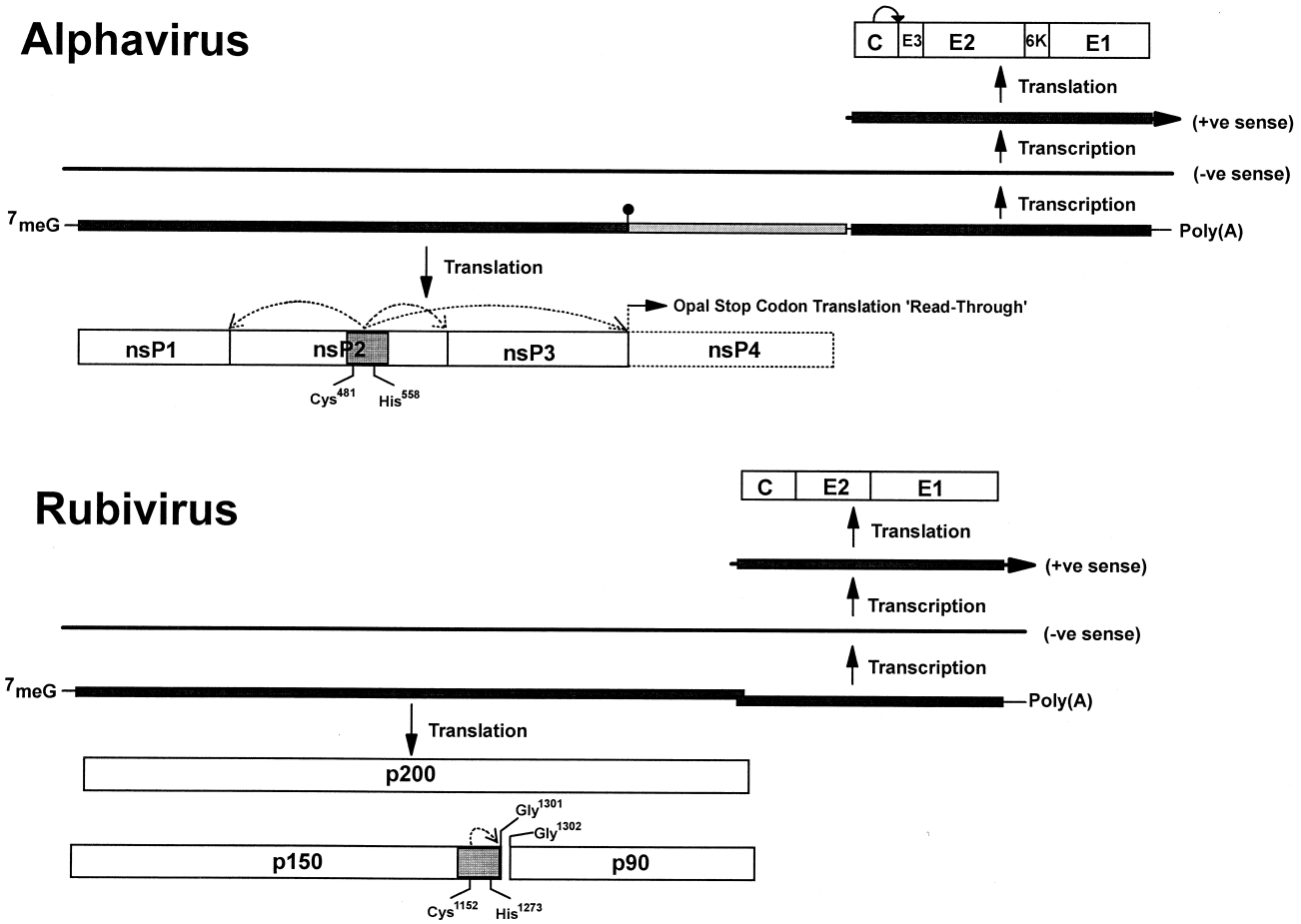

Fig. 1. Togavirus replication. Polyproteins are shown (boxed areas) together with the alphavirus nsP2 and rubivirus p150 proteinase domains (shaded areas). Sites of proteinase cleavage (curved arrows) and the proteinase active site cysteine/histidine residues are indicated.

rubiviruses a different order is observed: methyltransferase/'X'/proteinase/helicase/replicase. The evolutionary relationships between these (and other positive-strand RNA viruses) is discussed by Koonin & Dolja (1993) and Goldbach & Wellink (1988).

(*a*) Alphavirus nsP2 proteinase (clan CA, family C9)

The alphavirus nsP2 protein is thought to be a multifunctional protein. A helicase motif is present in the N-terminal portion of the protein whilst two lines of evidence from early studies showed that proteinase activity resided in the C-terminal portion of the protein. Firstly, temperature-sensitive (*ts*) proteinase mutations (Hahn *et al.*, 1989) and *ts* mutants deficient in processing of the polyprotein at the nsP1/nsP2, nsP2/nsP3 and nsP3/nsP4 sites (Hardy *et al.*, 1990) were mapped to this region. Secondly, the analysis of *in vitro* translation profiles derived from truncated or deleted forms of cDNA also showed the proteolytic activity to reside in the C-terminal portion of the molecule (residues 475–728 – Ding & Schlesinger, 1989; Hardy & Strauss, 1989; see Fig. 1).

Analysis of sequence alignments predicted the presence of a cysteine proteinase domain with the catalytic residues being Cys^{481} (the nucleophile) and His^{558} (Hardy & Strauss, 1989; Gorbalenya *et al.*, 1991; see Fig. 3A). Subsequent analysis of the proteolytic activities of site-directed nsP2 mutants confirmed that residues Cys^{481} and His^{558} were indeed critical for proteinase activity (Strauss *et al.*, 1992). These characteristics lead to the inclusion of nsP2 in clan CC (cysteine/histidine active site residues; Cys N-terminal to His). In many cysteine proteinases an asparagine residue is thought to play a role in catalysis (see Ryan & Flint, 1997, and references therein) and the possible importance of asparagine for the alphavirus nsP2 proteinase was examined. Mutation of all the conserved asparagine residues in the proteinase domain of nsP2 showed that none are absolutely required for proteolytic activity (Strauss *et al.*, 1992), demonstrating that this type of papain-like cysteine proteinase (PCP) is functional with a catalytic dyad of cysteine and histidine and that an asparagine residue is not apparently required.

The residue immediately C-terminal of His^{558} is tryptophan. This is in contrast to plant, animal and other virus papain-like

	nsP1 →			← nsP2				nsP2 →			← nsP3				nsP3 →				← nsP4			
	P4	*P3*	*P2*	*P1*	*P1′*	*P2′*	*P3′*	*P4′*	*P4*	*P3*	*P2*	*P1*	*P1′*	*P2′*	*P3′*	*P4′*	*P4*	*P3*	*P2*	*P1*	*P1′*	*P2′*	*P3′*	*P4′*
SIN	D	I	**G**	A	A	L	**V**	E	G	V	**G**	A	**A**	**P**	S	**Y**	*/R	L	T	G	V	**G**	G	**Y**	**I**	**F**	**S**
OCK	D	I	**G**	A	A	L	**V**	E	G	V	**G**	A	**A**	**P**	S	**Y**	*/R	L	T	G	V	**G**	G	**Y**	**I**	**F**	**S**
AURA	D	A	**G**	A	A	L	**V**	E	G	S	**G**	A	**A**	**P**	S	**Y**	*/R	L	T	G	V	**G**	G	**Y**	**I**	**F**	**S**
WHA	D	I	**G**	A	A	L	**V**	E	G	V	**G**	A	**A**	**P**	S	**Y**	*/R	L	T	G	V	**G**	G	**Y**	**I**	**F**	**S**
VEEV	E	A	**G**	A	G	S	**V**	E	E	A	**G**	C	**A**	**P**	S	**Y**	*/R	R	F	D	A	**G**	A	**Y**	**I**	**F**	**S**
EEEV	E	A	**G**	A	G	S	**V**	E	E	A	**G**	R	**A**	**P**	A	**Y**	*/R	R	Y	E	A	**G**	A	**Y**	**I**	**F**	**S**
WEEV	E	A	**G**	A	G	S	**V**	E	E	A	**G**	R	**A**	**P**	A	**Y**	*/R	R	Y	E	A	**G**	A	**Y**	**I**	**F**	**S**
ONN	R	A	**G**	A	G	I	**V**	E	R	A	**G**	C	**A**	**P**	S	**Y**	R	L	D	R	A	**G**	G	**Y**	**I**	**F**	**S**
IO	R	A	**G**	A	G	I	**V**	E	R	A	**G**	C	**A**	**P**	S	**Y**	R	L	D	R	A	**G**	G	**Y**	**I**	**F**	**S**
SFV	H	A	**G**	A	G	V	**V**	E	T	A	**G**	C	**A**	**P**	S	**Y**	R	L	G	R	A	**G**	A	**Y**	**I**	**F**	**S**
RRV	R	A	**G**	A	G	V	**V**	E	T	A	**G**	C	**A**	**P**	S	**Y**	*/R	L	G	R	A	**G**	A	**Y**	**I**	**F**	**S**
MID	R	A	**G**	A	G	V	**V**	N	T	A	**G**	C	**A**	**P**	S	**Y**	*/R	L	D	R	A	**G**	A	**Y**	**I**	**F**	**S**
BFV	R	A	**G**	E	G	V	**V**	E	P	A	**G**	S	**A**	**P**	A	**Y**	*/R	L	G	R	A	**G**	G	**Y**	**I**	**F**	**S**

Fig. 2. Cleavage sites of the alphavirus nonstructural proteins. Amino acid sequences at the sites of cleavage by nsP2 are shown. The universally conserved glycine at the P2 position is indicated in bold, amino acids conserved at each site are shaded and asterisks indicate opal stop codons at the nsP3/nsP4 junction. Virus sequences are as in Fig. 3, with chikungunya (CHICK – accession no. L37661), Ockelbo (OCK – accession no. M69205; Shirako *et al*., 1991), Igbo Ora (IO – accession no. AF079457), Middelburg (MID – Strauss *et al*., 1983) and Whataroa (WHA – see Strauss & Strauss, 1994).

proteinases where the side-chain of the residue immediately C-terminal of the active site histidine is small/aliphatic. It has been suggested that Trp^{559} in alphavirus nsP2 could function in place of asparagine by interaction with Trp^{482} (adjacent to the active site Cys^{481}) to maintain the conformation of the active site residues. Indeed, Trp^{559} has been shown by site-directed mutagenesis to be essential for catalytic activity (Strauss *et al*., 1992). By analogy, an interaction between these large, aromatic, residues could be present in the cysteine proteinase domains of other virus groups (rubi-, arteri-, tymo- and coronaviruses).

If one considers the extent of the alphavirus nsP2 proteinase domain to be comparable with other cysteine proteinase (domains), then the C-terminal region of nsP2 would contain an additional domain of some 210 amino acids. This region may possess yet another biochemical function, but analysis of the proteolytic properties of C-terminally deleted forms of nsP2 suggests that this region also has a role in proteolysis (Hardy & Strauss, 1989). It has been suggested that this C-terminal extension to the proteinase could act to mediate interaction with some, or all, of the substrate molecules (Dougherty & Semler, 1993), similar to the regulatory function of the 32 kDa protein of comoviruses (Peters *et al*., 1992; Vos *et al*., 1988) or protein NS2B of flaviviruses (Falgout *et al*., 1991). Suopanki *et al*. (1998) recently mapped the *ts*4 mutation in Semliki Forest virus to position Met^{781} in the C-terminal region of nsP2, and showed that a methionine-to-threonine mutation at this position inhibits P1234 processing and interferes with normal shut-off of minus-strand and 26S RNA synthesis.

Sequences flanking the NS polyprotein cleavage sites are shown in Fig. 2. Residues in the P3–P2–P1 positions are thought to be important for cleavage. The glycine residue at position P2 is completely conserved at all sites and was shown to be required for cleavage to occur (Shirako & Strauss, 1990). The alphavirus nsP2 proteinase domain has, therefore, a similar cleavage specificity to most other viral PCPs in cleaving between two short aliphatic residues. For residues immediately C-terminal of the cleavage site (P1′/P4′), similarities are observed between different virus polyproteins at any given cleavage site. This region is, however, dissimilar between the different cleavage sites for any particular polyprotein (see Fig. 2). Detailed kinetic data for the proteolysis at each of the different sites are unavailable. The presence or absence of the nsP1 and nsP3 domains in either target or cleaving polypeptide may, however, be highly influential in determining the rate of cleavage at a specific site (see below).

Much data are, however, available for the regulation of Sindbis virus nsP2 proteinase and its role in virus replication. A number of observations on the alphavirus polyprotein processing have been combined to account for aspects of virus replication (Hardy & Strauss, 1989; Shirako & Strauss, 1990; de Groot *et al*., 1990;Sawicki & Sawicki, 1993, 1994; Lemm & Rice, 1993*a*, *b*; Lemm *et al*., 1994; Wang *et al*., 1994; Dé *et al*., 1996).

(i) All nsP2-containing polypeptides are proteolytically active

(ii) The kinetics of cleavage of each of the nsP sites is different and may require particular nsP2-containing forms for processing *in trans*

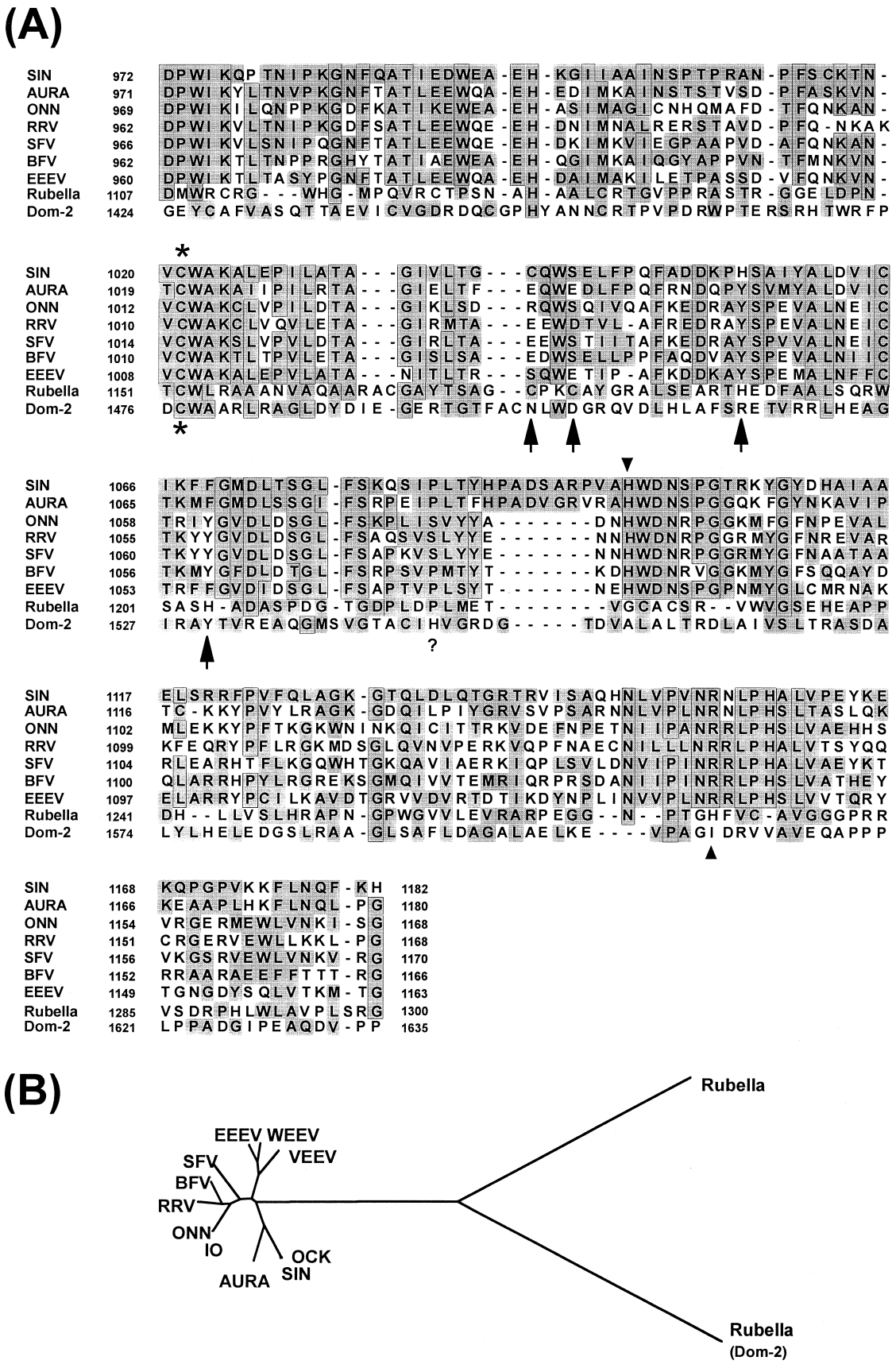

Fig. 3. For legend see opposite.

(iii) All nsP2-containing polypeptides were able to cleave the nsP1/2 site *in trans*

(iv) Proteinase forms containing nsP1 were unable to cleave the nsP3/4 site

(v) Only nsP3-containing proteinase forms were able to cleave the nsP3/4 site

(vi) Whilst nsP123^{pro} can cleave the nsP1/2 site, it cannot cleave the nsP2/3 site

(vii) nsP23^{pro} can cleave the nsP2/3 site very efficiently

(viii) Cleavage activity is not influenced by the presence or absence of nsP4

Using these proteolytic processing 'rules' an elegant model has been developed to account for the regulation of transcription of positive- and negative-sense RNA (de Groot *et al.*, 1990; Lemm *et al.*, 1994; Strauss & Strauss, 1994; Dé *et al.*, 1996). The presence of nsP1 and/or nsP3 sequences within the proteolytically active form serves to dramatically alter the substrate specificity of the proteinase form and, thereby, the different biochemical activities which are generated by the alternative processing pathways. Initiation complexes containing an intact nsP123 or nsP23 have an high activity in minus-strand synthesis, whereas upon cleavage of the nsP23 junction plus-strand synthesis is preferred. An *in vitro* replication system consisting of nsP4, an RNA template and various nsP123 mutants showed minus-strand production. Replication was reduced in constructs that increased processing at the nsP23 junction, whereas mutations that abolished processing at this site stimulated the formation of minus strands. Processing at the nsP12 site was allowed, as shown by replication in a reaction containing nsP1, P23 (protease-disabled) and nsP4 as separate peptides (Lemm *et al.*, 1998).

(*b*) Semliki Forest virus nsP4 protein

In Semliki Forest (SF), o'nyong-nyong and Igbo Ora viruses a sense (arginine) codon is found at the site corresponding to the opal stop codon delineating nsP3 and nsP4 in other alphaviruses. In these viruses, therefore, the entire ORF translated without suppression of a stop codon; the nsP4 protein is translated in equimolar ratios to nsP1–3. The SFV nsP3/nsP4 junction is Ala^{1817}/Tyr^{1818}, determined by N-terminal sequencing (Fig. 2; Kalkkinen *et al.*, 1981). Analyses of *in vitro* translation profiles derived from cDNA clones encoding SFV NS proteins were taken to show that cleavage of the SFV nsP3/4 site is not performed by the nsP2 proteinase. It was proposed that: (i) nsP4 is produced by a cleavage of nascent nsP1234, mediated by nsP4 and, (ii) proteolytic cleavages at the nsP1/2 and nsP2/3 sites were catalysed by the nsP2 proteinase (Takkinen *et al.*, 1991). The aspartate residue of a conserved – DTG – motif (characteristic of aspartyl proteinases), close to the N terminus of nsP4, was suggested to be the active-site residue of the putative proteinase activity associated with nsP4 mediating the nsP3/nsP4 cleavage (Takkinen *et al.*, 1990). Translation profiles from constructs encoding nsP4 with the N-terminal 102 residues deleted were, apparently, not subject to cleavage *in vitro*. Cleavage at the nsP3/nsP4 junction could not be mediated *in trans*; proteolysis could be inhibited by pepstatin A, but not by Zn^{2+} or by PMSF – consistent with the suggestion of an aspartyl proteinase mechanism (Takkinen *et al.*, 1991). It should be noted, however, that this scheme would necessitate the interaction of *two* nsP4 molecules to form the 'classical' acid proteinase active site and that such bimolecular interactions (tri-, including interaction with the substrate) are difficult to observe using *in vitro* translation systems due to the extremely low amounts of translation products synthesized. Interestingly, the nsP3/nsP4 junction maintains good homology with the viruses using readthrough (see Fig. 2) and should therefore be a substrate for the nsP2 protease.

(*c*) Rubivirus proteinase (clan CA, family C27)

Note: the numbering scheme used here refers to the corrected rubella virus sequence (Pugachev *et al.*, 1997). The analysis of sequence alignments predicted the NS polyprotein to contain a PCP domain with the catalytic residues being Cys^{1152} and His^{1273} (Gorbalenya *et al.*, 1991; Figs 1 and 3A). Subsequent site-directed mutagenesis experiments confirmed that these residues were critical for proteolytic activity (Chen *et al.*, 1996; Marr *et al.*, 1994; Yao *et al.*, 1998). The rubella virus proteinase domain mediates a single cleavage in the NS polyprotein at its own C terminus cleaving the p200 precursor into the p150 and p90 products (Fig. 1; Forng & Frey, 1995; Chen *et al.*, 1996). By deletion mutagenesis the proteinase domain has been shown to map between amino acids 1006 and 1508 of the non-structural polyprotein (Marr *et al.*, 1994).

Amino-terminal amino acid sequencing of the p90 cleavage product and site-directed mutagenetic studies showed the rubivirus NS polyprotein to be cleaved between residues Gly^{1301} and Gly^{1302} (Chen *et al.*, 1996; Yao *et al.*, 1998). This

Fig. 3. Sequence alignment of alphavirus nsP2, rubivirus p150 proteinase domain and rubivirus p90 domain 2. The active site nucleophile cysteine aligns amongst the alpha- and rubivirus sequences (*), although the active site histidine aligns amongst alphaviruses (▼), but not with active site histidine of the rubella virus p150 PCP sequence (▲). Residues comprising the proposed rubella virus p150 PCP zinc-binding site (Liu *et al.*, 1998) are also indicated (↑). The sequence of the N-terminal region of the rubella virus p90 protein (Gly^{1424}–Pro^{1636}) is also shown (Dom-2) to illustrate the poor alignment between the alpha- and rubivirus PCPs – a 'candidate' His of the Dom-2 sequence is also indicated (?) in panel (A). In panel (B) the similarity between sequences is shown, branch lengths being proportional to relatedness. Virus sequences used in the alignment are: Sindbis (SIN – Strauss *et al.*, 1984; AURA – Rumenapf *et al.*, 1995), o'nyong-nyong (ONN – Levinson *et al.*, 1990), Ross River (RRV – Faragher *et al.*, 1988), Semliki Forest (SFV – Garoff *et al.*, 1980*a*, *b*; Takkinen, 1986), Barmah Forest (BFV – accession no. U73745), Eastern equine encephalitis (EEEV – Volchkov *et al.*, 1991) and rubella (RUB – Dominguez *et al.*, 1990; Pugachev *et al.*, 1997). The numbering scheme refers to polyprotein positions.

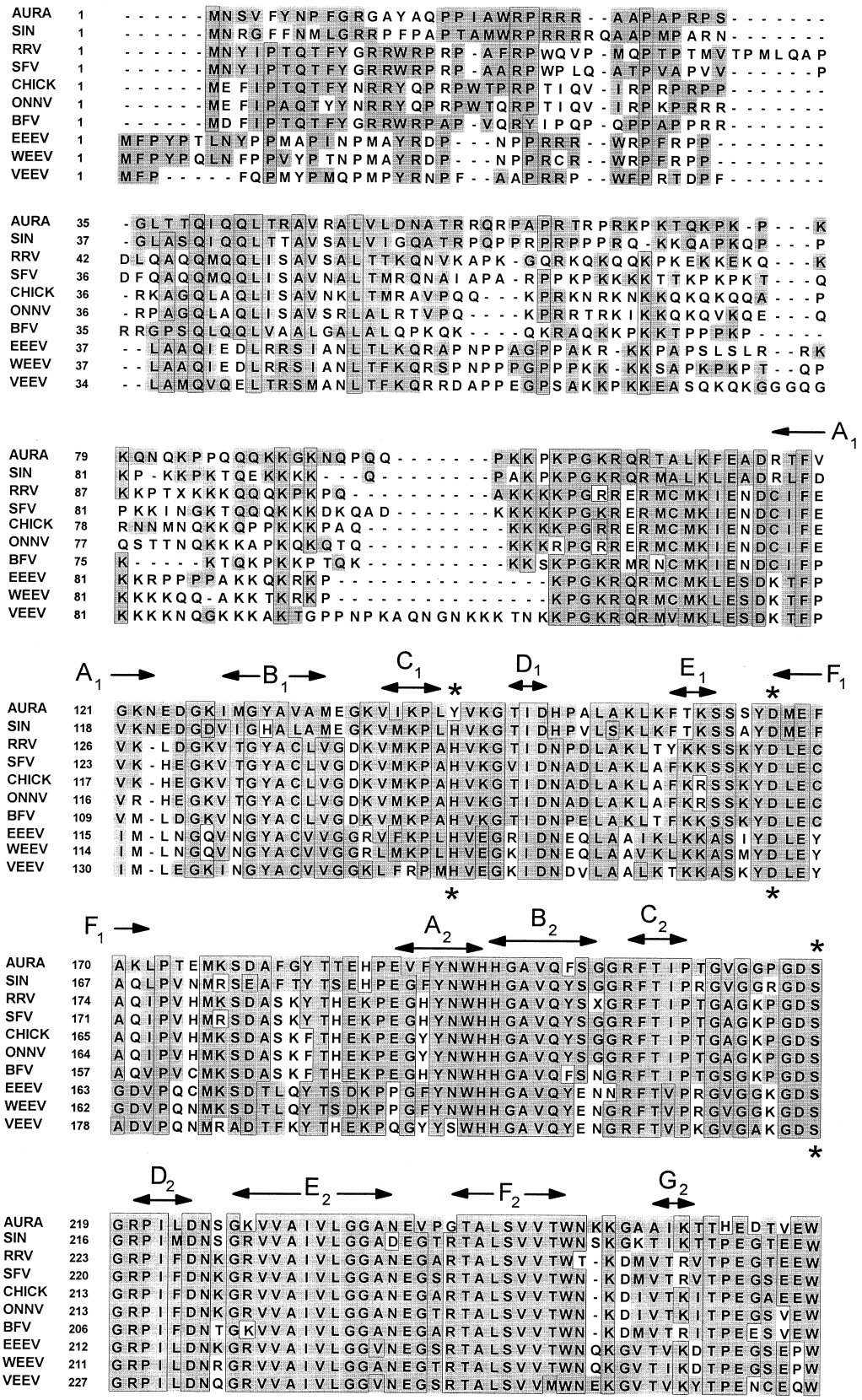

Fig. 4. For legend see facing page.

cleavage-site specificity is similar to other viral papain-like proteinases [mouse hepatitis virus PLP-1 – Gly/Val (Hughes *et al.*, 1995; Bonilla *et al.*, 1997); equine arteritis virus – Gly/Gly (Snijder *et al.*, 1994); potyvirus HC-Pro – Gly/Gly (Carrington *et al.*, 1989); hypovirulence associated virus p29 proteinase – Gly/Gly (Choi *et al.*, 1991; Shapira & Nuss, 1991). Characterization of the cleavage site by mutagenesis suggested that the P1 residue is the most important for cleavage, but that the P2 and P1′ residues also contribute to recognition of the cleavage site (Chen *et al.*, 1996).

The p150/p90 cleavage occurs *in cis*, although a recent analysis showed that the proteinase could also act *in trans*. Using an *in vitro* translation system co-translation studies showed cleavage *in trans* (Liu *et al.*, 1998) of a proteolytically active recombinant polyprotein (in which the mutated cleavage site was not capable of being cleaved) on another polyprotein (rendered proteolytically inactive by mutation) which was the substrate. The failure to detect cleavage *in trans* in earlier studies may have been caused by the proximity of the active site and the cleavage site; mutations created to destroy the cleavage site could also have affected structural integrity and inactivated the proteinase (Liu *et al.*, 1998).

The rubella virus proteinase domain was found not to function following translation in a rabbit reticulocyte lysate *in vitro* translation system, although all of the other viral PCPs do so. However, following the supplementation of the translation system with divalent cations such as Zn^{2+}, Cd^{2+} and Co^{2+}, the rubella virus proteinase functioned efficiently indicating that these cations are required either as direct cofactors in catalytic activity or as a structural component of the proteinase. The picornavirus 2A and hepatitis C virus NS3 proteinases have a zinc atom as a structural feature, the metal not playing a direct role in catalysis (Sommergruber *et al.*, 1994; Voss *et al.*, 1995; Kim *et al.*, 1996; Love *et al.*, 1996; Yan *et al.*, 1998). A putative zinc-binding motif has been identified in the sequence of the rubella virus proteinase domain (Liu *et al.*, 1998; Fig. 3A).

(*d*) Sequence alignments

The amino acid similarities/identities between disparate PCPs are very low although the presence of proteinase domains can be suspected by the presence of short motifs (including the Cys/His active site residues) and their spacing. Fig. 3 shows that the alphavirus nsP2 PCPs are conserved and are easily aligned. The rubivirus proteinase domain aligns with the alphavirus sequences very poorly. Indeed, a domain in the N-terminal region of the p90 protein (Gly^{1424}–Pro^{1636}) can be aligned with the alphavirus sequences very nearly as well as the 'authentic' proteinase domain (Fig. 3B)! Such PCP domains are found in a wide range of positive-strand RNA virus polyproteins and are discussed in the excellent review of Gorbalenya & Snijder (1996).

Structural polyprotein

The togavirus structural polyprotein comprises the nucleocapsid protein (C) and glycoproteins (Fig. 1). In such enveloped viruses the nucleocapsid protein must interact with both the RNA genome (packaging) and the cytoplasmic domains of the glycoproteins (particle morphogenesis) in addition to C–C interactions to form the capsid structure. In both alpha- and rubiviruses the nucleocapsid protein is N-terminal of the polyprotein. Early work showed that the translation products derived from alpha- and rubivirus sub-genomic mRNAs differed: the alphavirus structural polyprotein was very rapidly processed at the C/E3 site into two products whilst the rubivirus polyprotein was uncleaved. Supplementation of translation systems with microsomes produced full processing in both systems (reviewed by Strauss & Strauss, 1994; Frey, 1994). Processing of the rubivirus structural polyprotein is mediated, therefore, by host-cell signalases alone.

Alphavirus chymotrypsin-like proteolytic capsid protein (clan PA, family S3)

The first ~100 residues of the capsid protein are not highly conserved amongst alphaviruses. This region is rich in basic/proline residues (Fig. 4) and is thought to protrude into the interior of the capsid where it interacts with the genomic RNA via electrostatic interactions.

The activity which brings about the microsome-independent C/E3 polyprotein cleavage of the alphavirus structural polyprotein was identified as a proteolytic activity of the C protein itself. Site-directed mutagenetic experiments identified two components of the catalytic triad, His^{141} and Ser^{215} (Melancon & Garoff, 1987; Hahn & Strauss, 1990), the third member (Asp^{163}) latterly being identified by resolution of the atomic structure of the protein (Choi *et al.*, 1991; Tong *et al.*, 1993). This work confirmed earlier suggestions of a serine proteinase-like activity of the nucleocapsid protein (Boege *et al.*, 1981). Substitution of Ser^{215} by cysteine did not inhibit C/E3 processing, but did prevent formation of infectious Sindbis virus particles (Hahn & Strauss, 1990). The atomic structure of the capsid protein of another alphavirus, SFV, showed very similar features to the Sindbis structure (Choi *et al.*, 1997).

The capsid protein of alphaviruses cleaves only at its own C terminus in an intramolecular, co-translational, fashion (in *cis*) at a conserved Trp/Ser scissile pair. The C-terminal region in

Fig. 4. Sequence alignment of alphavirus capsid proteins. The atomic structure was determined from residue Arg^{114} (Choi *et al.*, 1991; Tong *et al.*, 1993). The active site Asp, His and Ser residues are indicated (*). Note: the Tyr residue in place of the active site His in the published AURA sequence is a sequencing error and, in fact, is His (J. H. Strauss, personal communication). The assignment of the β-sheets is that given in Tong *et al.* (1993). Virus sequences are as in Figs 1 and 2, with Venezuelan equine encephalitis (VEEV – Kinney *et al.*, 1989) and Western equine encephalitis (WEEV – Hahn *et al.*, 1988). The numbering scheme refers to polyprotein positions.

chymotrypsin-like serine proteinases is an α-helix some distance from the active site. The C-terminal region of the alphavirus capsid protein is, however, not helical and the structure terminates within the active site, the ultimate tryptophan residue remaining in the P1 position. This accounts for the absence of catalytic activity following the C/E3 cleavage *in cis*. Neither in the original, nor refined, atomic structures was it possible to determine whether, following proteolysis, a covalent bond was present between the carboxyl group of Trp^{264} and the side-chain of the nucleophile Ser^{215} – equivalent to an unresolved acyl-enzyme intermediate form. Amino acid sequencing of tryptic peptides, however, suggests that the C terminus of the protein is not covalently linked to the catalytic serine (Boege *et al*., 1980). The active site residues have a similar topology to other serine proteinases. The unusual folding of the C-terminal region into the active site exposes the catalytic Asp^{163} to the solvent. In other serine proteinases this active site residue is buried: shielding from the solvent may well be a requisite for the catalytic mechanism. During the co-translational cleavage another structural form may exist in which Asp^{163} is buried. The residue at the bottom of the substrate binding pocket, thought to mediate a large part of the substrate specificity, is valine whilst other residues surrounding the substrate binding site include a number of glycines. This creates a relatively hydrophobic pocket for the recognition of (and the space to accommodate) the bulky hydrophobic C-terminal tryptophan residue.

Concluding remarks

Proteolytic processing is an extremely common strategy amongst animal and plant positive-strand RNA viruses, only a few such plant viruses apparently generating all of their proteins from sub-genomic mRNAs. In many plant viruses (i.e. como-, nepo- and bymoviruses) genomes are multi-partite, each strand encoding a polyprotein. Some of the 'primary' polyprotein cleavages observed in the (single) polyproteins of monopartite viruses are, patently, not required in these viruses (see Ryan & Flint, 1997). In this review we have remarked upon two other strategies: synthesis of sub-genomic mRNA and readthrough of leaky stop codons. In general, non-enveloped positive-strand RNA viruses do not require the involvement of host proteinases in the processing of their polyproteins, whilst the enveloped viruses recruit host enzymes for the processing of precursors comprising encapsidation functions (i.e. proteolytic processing in the lumen of the endoplasmic reticulum), and, in some cases, for the processing of the non-structural protein precursor.

In this series of reviews we have discussed one of the major strategies adopted by positive-strand RNA viruses for the biogenesis of proteins during replication – polyprotein processing by virus-encoded proteinases. For the limited number of atomic structures available for the virus proteinases, remarkable structural similarities to their cellular counterparts are observed, although sequence similarities may be very low. In considering virus and cellular proteinases as a whole, however, the striking feature is the conservation of the catalytic mechanism and the active site stereochemistry within a (virus or cellular) proteinase *type*.

Perhaps the most noticeable difference between virus and cellular proteinases is a theme now developing in the literature: the subtle regulation of the activity and specificity of the virus enzymes by binding other virus proteins or virus RNA, and, the differences in activity which are observed between larger, uncleaved, forms containing proteolytic domains. These may be metastable, proteolytically active, and not necessarily precursor forms as such – i.e. $3CD^{pro}$ (picornaviruses) and the various nsP2-containing products described above.

Virus-encoded proteinases are an attractive target for antiviral drugs. Relatively simple assays may be developed which are suitable for automated high-throughput screening of combinatorial chemical libraries. The study of the regulation of these enzymes can, however, give major insights into the control of virus replicative processes and provide alternative strategies for future drug development.

References

Barrett, A. J., Rawlings, N. D. & Woessner, J. F. (editors) (1998). *Handbook of Proteolytic Enzymes*. London: Academic Press.

Boege, U., Wengler, G. & Wittmann-Liebold, B. (1980). Partial amino acid sequences of Sindbis and Semliki Forest virus-specific core proteins. *Virology* **103**, 178–190.

Boege, U., Wengler, G., Wengler, G. & Wittmann-Liebold, B. (1981). Primary structure of the core proteins of the alphaviruses Semliki Forest virus and Sindbis virus. *Virology* **113**, 293–303.

Bonilla, P. J., Hughes, S. A. & Weiss, S. R. (1997). Characterization of a second cleavage site and demonstration of activity in *trans* by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59. *Journal of Virology* **71**, 900–909.

Carrington, J. C., Freed, D. D. & Sanders, T. C. (1989). Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and in *vitro*. *Journal of Virology* **63**, 4459–4463.

Chen, J. P., Strauss, J. H., Strauss, E. G. & Frey, T. K. (1996). Characterization of the rubella virus nonstructural protease domain and its cleavage site. *Journal of Virology* **70**, 4707–4713.

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.

Choi, H.-K., Lu, G., Lee, S., Wengler, G. & Rossmann, M. G. (1997). Structure of Semliki Forest virus core protein. *Proteins: Structure, Function and Genetics* **27**, 345–359.

Clarke, I. N. & Lambden, P. R. (1997). The molecular biology of caliciviruses. *Journal of General Virology* **78**, 291–301.

de Groot, R. J., Hardy, W. R., Shirako, Y. & Strauss, J. H. (1990). Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in *vivo*. *EMBO Journal* **9**, 2631–2638.

Dé, I., Sawicki, S. G. & Sawicki, D. L. (1996). Sindbis virus RNA-negative mutants that fail to convert from minus-strand to plus-strand synthesis: role of the nsP2 protein. *Journal of Virology* **70**, 2706–2719.

Ding, M. X. & Schlesinger, M. J. (1989). Evidence that Sindbis virus NSP2 is an autoprotease which processes the virus nonstructural polyprotein. *Virology* **171**, 280–284.

Dominguez, G., Wang, C.-Y. & Frey, T. K. (1990). Sequence of the genome RNA of rubella virus; evidence for genetic rearrangement during togavirus evolution. *Virology* **177**, 225–238.

Dougherty, W. G. & Semler, B. L. (1993). Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiological Reviews* **57**, 781–822.

Falgout, B., Pethel, M., Zhang, Y. M. & Lai, C. J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *Journal of Virology* **65**, 2467–2475.

Faragher, S. G., Meek, A. D., Rice, C. M. & Dalgarno, L. (1988). Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. *Virology* **163**, 509–526.

Feng, Y.-X., Copeland, T. D., Oroszlan, S., Rein, A. & Levin, J. G. (1990). Identification of amino acids inserted during suppression of UAA and UGA termination codons at the *gag–pol* junction of Moloney murine leukemia virus. *Proceedings of the National Academy of Sciences, USA* **87**, 8860–8863.

Forng, R. Y. & Frey, T. K. (1995). Identification of the rubella virus nonstructural proteins. *Virology* **206**, 843–853.

Frey, T. K. (1994). Molecular biology of rubella virus. *Advances in Virus Research* **44**, 69–160.

Garoff, H., Frischauf, A. M., Simons, K., Lehrach, H. & Delius, H. (1980*a*). The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino-terminal region. *Proceedings of the National Academy of Sciences, USA* **77**, 6376–6380.

Garoff, H., Frischauf, A. M., Simons, K., Lehrach, H. & Delius, H. (1980*b*). Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. *Nature* **288**, 236–241.

Goldbach, R. & Wellink, J. (1988). Evolution of plus-strand RNA viruses. *Intervirology* **29**, 260–267.

Gorbalenya, A. E. & Snijder, E. J. (1996). Viral cysteine proteinases. *Perspectives in Drug Discovery and Design* **6**, 64–86.

Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. (1991). Putative papain-related thiol proteases 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.

Hahn, C. S. & Strauss, J. H. (1990). Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. *Journal of Virology* **64**, 3069–3073.

Hahn, C. S., Lustig, S., Strauss, E. G. & Strauss, J. H. (1988). Western equine encephalitis virus is a recombinant virus. *Proceedings of the National Academy of Sciences, USA* **85**, 5997–6001.

Hahn, Y. S., Strauss, E. G. & Strauss, J. H. (1989). Mapping of RNA-temperature sensitive mutants of Sindbis virus: assignment of complementation groups A, B, and G to nonstructural proteins. *Journal of Virology* **63**, 3142–3150.

Hardy, W. R. & Strauss, J. H. (1989). Processing the nonstructural polyproteins of Sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in *trans*. *Journal of Virology* **63**, 4653–4664.

Hardy, W. R., Hahn, Y. S., de Groot, R. J., Strauss, E. G. & Strauss, J. H. (1990). Synthesis and processing of the nonstructural polyproteins of several temperature-sensitive mutants of Sindbis virus. *Virology* **177**, 199–208.

Hughes, S. A., Bonilla, P. J. & Weiss, S. R. (1995). Identification of the murine coronavirus p28 cleavage site. *Journal of Virology* **69**, 809–813.

Kalkkinen, N., Laaksonen, M., Soderlund, H. & Jornvall, H. (1981). Radio-sequence analysis of in *vivo* multilabeled nonstructural protein ns86 of Semliki Forest virus. *Virology* **113**, 188–195.

Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. & Thomson, J. A. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **87**, 343–355.

Kinney, R. M., Johnson, B. J., Welch, J. B., Tsuchiya, K. R. & Trent, D. W. (1989). The full-length nucleotide sequences of the virulent Trinidad donkey strain of Venezuelan equine encephalitis virus and its attenuated vaccine derivative, strain TC-83. *Virology* **170**, 19–30.

Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical Reviews in Biochemistry and Molecular Biology* **28**, 375–430.

Lemm, J. A. & Rice, C. M. (1993*a*). Assembly of functional Sindbis virus RNA replication complexes: requirement for coexpression of P123 and P34. *Journal of Virology* **67**, 1905–1915.

Lemm, J. A. & Rice, C. M. (1993*b*). Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. *Journal of Virology* **67**, 1916–1926.

Lemm, J. A., Rumenapf, T., Strauss, E. G., Strauss, J. H. & Rice, C. M. (1994). Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *EMBO Journal* **13**, 2925–2934.

Lemm, J. A., Bergqvist, A., Read, C. M. & Rice, C. M. (1998). Template-dependent initiation of Sindbis virus RNA replication in *vitro*. *Journal of Virology* **72**, 6546–6553.

Levinson, R. S., Strauss, J. H. & Strauss, E. G. (1990). Complete sequence of the genomic RNA of o'nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees. *Virology* **175**, 110–123.

Li, G. & Rice, C. M. (1993). The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. *Journal of Virology* **67**, 5062–5067.

Liu, X., Ropp, S. L., Jackson, R. J. & Frey, T. K. (1998). The rubella virus nonstructural protease requires divalent cations for activity and functions in *trans*. *Journal of Virology* **72**, 4463–4466.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T. & Hostomska, Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **87**, 331–42.

Marr, L. D., Wang, C. Y. & Frey, T. K. (1994). Expression of the rubella virus nonstructural protein ORF and demonstration of proteolytic processing. *Virology* **198**, 586–592.

Melancon, P. & Garoff, H. (1987). Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. *Journal of Virology* **61**, 1301–1309.

Peters, S. A., Voorhorst, W. G., Wery, J., Wellink, J. & van Kammen, A. (1992). A regulatory role for the 32K protein in proteolytic processing of cowpea mosaic virus polyproteins. *Virology* **191**, 81–89.

Pugachev, K. V., Abernathy, E. S. & Frey, T. K. (1997). Improvement of the specific infectivity of the rubella virus (RUB) infectious clone: determinants of cytopathogenicity induced by RUB map to the nonstructural proteins. *Journal of Virology* **71**, 562–568.

Rumenapf, T., Strauss, E. G. & Strauss, J. H. (1995). Aura virus is a New World representative of Sindbis-like viruses. *Virology* **208**, 621–633.

Ryan, M. D. & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699–723.

Ryan, M. D., Monaghan, S. & Flint, M. (1998). Virus-encoded proteinases of the *Flaviviridae*. *Journal of General Virology* **79**, 947–959.

Sawicki, D. L. & Sawicki, S. G. (1993). A second nonstructural protein functions in the regulation of alphavirus negative-strand RNA synthesis. *Journal of Virology* **67**, 3605–3610.

Sawicki, D. L. & Sawicki, S. G. (1994). Alphavirus positive and negative strand RNA synthesis and the role of polyproteins in formation of viral replication complexes. *Archives of Virology Suppl.* **9**, 393–405.

Shapira, R. & Nuss, D. L. (1991). Gene expression by a hypovirulence-associated virus of the chestnut blight fungus involves two papain-like protease activities. Essential residues and cleavage site requirements for p48 autoproteolysis. *Journal of Biological Chemistry* **266**, 19419–19425.

Shirako, Y. & Strauss, J. H. (1990). Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123. *Virology* **177**, 54–64.

Shirako, Y., Niklasson, B., Dalrymple, J. M., Strauss, E. G. & Strauss, J. H. (1991). Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. *Virology* **182**, 753–764.

Snijder, E. J. & Meulenberg, J. J. M. (1998). The molecular biology of arteriviruses. *Journal of General Virology* **79**, 961–979.

Snijder, E. J., Wassenaar, A. L. & Spaan, W. J. (1994). Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *Journal of Virology* **68**, 5755–5764.

Sommergruber, W., Casari, G., Fessl, F., Seipelt, J. & Skern, T. (1994). The 2A proteinase of human rhinovirus is a zinc containing enzyme. *Virology* **204**, 815–818.

Strauss, J. H. & Strauss, E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiological Reviews* **58**, 491–562.

Strauss, E. G., Rice, C. M. & Strauss, J. H. (1983). Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. *Proceedings of the National Academy of Sciences, USA* **80**, 5271–5275.

Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133**, 92–110.

Strauss, E. G., De Groot, R. J., Levinson, R. & Strauss, J. H. (1992). Identification of the active site residues in the nsP2 proteinase of Sindbis virus. *Virology* **191**, 932–940.

Suopanki, J., Sawicki, D. L., Sawicki, S. G. & Kääriäinen, L. (1998). Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. *Journal of General Virology* **79**, 309–319.

Takkinen, K. (1986). Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus. *Nucleic Acids Research* **14**, 5667–5682.

Takkinen, K., Peranen, J., Keranen, S., Soderlund, H. & Kaariainen, L. (1990). The Semliki-Forest-virus-specific nonstructural protein nsP4 is an autoproteinase. *European Journal of Biochemistry* **189**, 33–38.

Takkinen, K., Peranen, J. & Kaariainen, L. (1991). Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein. *Journal of General Virology* **72**, 1627–1633.

Tong, L., Wengler, G. & Rossmann, M. G. (1993). Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *Journal of Molecular Biology* **230**, 228–247.

Volchkov, V. E., Volchkova, V. A. & Netesov, S. V. (1991). [Complete nucleotide sequence of the Eastern equine encephalomyelitis virus genome]. *Molekuliarnaia Genetika, Mikrobiologia, i Virusologa* **5**, 8–15.

Vos, P., Verver, J., Jaegle, M., Wellink, J., van Kammen, A. & Goldbach, R. (1988). Two viral proteins involved in the proteolytic processing of the 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.

Wang, Y.-F., Sawiki, S. G. & Sawiki, D. L. (1994). Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. *Journal of Virology* **68**, 6466–6475.

Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J. L., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L. C. & Chen, Z. (1998). Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2·2 Å resolution structure in a hexagonal crystal form. *Protein Science* **7**, 837–847.

Yao, J., Yang, D., Chong, P., Hwang, D., Liang, Y. & Gillam, S. (1998). Proteolytic processing of rubella virus nonstructural proteins. *Virology* **246**, 74–82.