Viruses and viral proteins

Nuria Verdaguer,[a]* Diego Ferrero[a] and Mathur R. N. Murthy[b]

[a]Institut de Biología Molecular de Barcelona, CSIC, Parc Científic de Barcelona, Baldiri i Reixac 15, 08028-Barcelona, Spain, and [b]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. *Correspondence e-mail: nvmcri@ibmb.csic.es

For more than 30 years X-ray crystallography has been by far the most powerful approach for determining the structures of viruses and viral proteins at atomic resolution. The information provided by these structures, which covers many important aspects of the viral life cycle such as cell-receptor recognition, viral entry, nucleic acid transfer and genome replication, has extensively enriched our vision of the virus world. Many of the structures available correspond to potential targets for antiviral drugs against important human pathogens. This article provides an overview of the current knowledge of different structural aspects of the above-mentioned processes.

1. Introduction

Viruses have evolved different strategies for their multiplication and propagation. The diversity and complexity of viral protein structures now available at near atomic resolution and stored in the Protein Data Bank (PDB; http://www.rcsb.org/pdb/) or in the specialized database Virus Particle Explorer (VIPER; http://viperdb.scripps.edu) (Carrillo-Tripp *et al.*, 2009) are the result of outstanding achievements of structural virologists. The PDB now treasures the three-dimensional structures of over 350 virus capsids, from more than 35 viral families and about 5000 protein structures coded by viral genomes. More than 90% of these structures have been determined by X-ray crystallography (PDB July 2014). These include not only structural proteins forming icosahedral capsids and other components of virus particles, such as proteins of cylindrical viruses or different components of tailed phages, but also many viral enzymes such as viral proteinases, and RNA and DNA polymerases.

Animal viruses may infect host cells by anchoring to an appropriate receptor molecule(s), which will trigger penetration of the entire virion or some of its components, always including the viral genome, into the cell. Unless the nucleic acid enters alone, viral genome replication and expression will require uncoating of the capsid and release of the internalized viral nucleic acid. Cell recognition, entry and uncoating frequently overlap and rely on quite different mechanisms depending on the virus species and host type.

In non-enveloped animal viruses, the mechanisms for genome delivery have not yet been characterized in detail, although a number of steps of the process have been described. Some well studied models include poliovirus (PV) and Human Rhinoviruses (HRV) both belonging to the *Picornaviridae* family that includes a large number of human
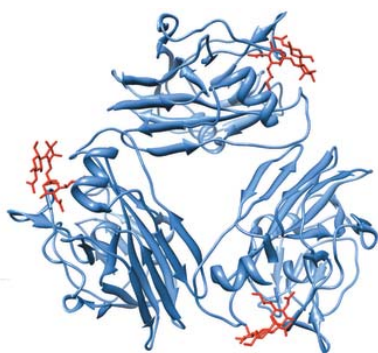
and animal pathogens. In these viruses, uncoating can be mediated by receptor binding, by low pH, or by the cooperative effect of both factors, and appears to be linked to the inherent stability and dynamics of the capsid. The X-ray and cryo-electron micrography structures of different HRV2 uncoating intermediates revealed the structural alterations that take place in the virus architecture, allowing the delivery of the genomic RNA: the acidic pH induces the conversion of the native HRV2 virions into an expanded, porous uncoating intermediate called an A-particle. This conversion is accompanied by major changes in the RNA organization and interactions at the inner capsid wall that would facilitate the subsequent RNA egress through large capsid disruptions produced at the particle twofold symmetry axes (Pickl-Herk *et al.*, 2013; Garriga *et al.*, 2012). Similar uncoating mechanisms have also been described for PV (Bostina *et al.*, 2011).

Enveloped animal viruses also use a two-step process to release their genetic material into the cell: first they bind to specific cell-surface receptors anchored to the target cell membrane and then they induce fusion of the viral and cell membranes. Binding to the cell-surface receptor in these viruses is mediated by a viral glycoprotein embedded in the viral lipid envelope which specifically interacts with some cellular molecule, which typically is a membrane-associated glycoprotein, carbohydrate or glycolipid. Detailed knowledge of virus–receptor interactions is essential to understand different aspects that determine viral tropism, spread and pathogenesis. During the last few years, a large number of high-resolution structures of viral proteins in complex with their specific receptors have been reported, shedding light on these aspects. Very recently, Wang and co-workers characterized the interactions between the newly identified Middle East respiratory syndrome coronavirus and its target cell by solving the structure of the receptor-binding domain of the viral envelope spike glycoprotein bound to its cellular receptor, the dipeptidyl peptidase 4 (Wang *et al.*, 2013). The structural information obtained might serve as a guide for the development of therapeutics against this novel coronavirus.

Fusion of viral and cell membranes may occur either at the cell surface or after internalization of the virus particle. The process is driven by specialized viral glycoproteins, called fusogens, which are in metastable conformations in the virus particle. Once activated, these proteins initiate a series of conformational changes, favouring the fusion of the two membranes. At the end of the fusion process, the viral fusogens adopt highly stable conformations. The free energy liberated during the transition from the metastable pre-fusion to the highly stable post-fusion conformation drives the fusion process. Remarkably, the virus–cell membrane-fusion process appears to follow essentially the same intermediate steps as in other membrane fusions that occur for instance in vesicular fusion or in cell–cell fusion. Recent data from Felix Rey and colleagues (Pérez-Vargas *et al.*, 2014) unveil a striking structural relationship between the *Caenorhabditis elegans* cellular fusion protein EFF-1, believed to be involved in the development of multicellular organisms, and the class II viral fusion proteins, indicating the importance of an intricate exchange of

genetic information between viruses and cells during the evolution of multicellular organisms.

Like animal viruses, bacteriophages also recognize receptors in the host-cell surface. However, because of the thickness and hardness of bacterial cell walls, the phage particles cannot enter the cells by endocytosis. Phages have evolved various mechanisms to transport their genomes across the bacterial cell wall. Tailed phages use a tail that penetrates the cell wall in a way similar to that of the needle of a hypodermic syringe. The nucleic acid genome is then injected into the cell, driven in part by the internal pressure built in the phage head during packaging of the nucleic acid. In contrast, tail-less phages rely on host organelles for genome transfer. Recent data from Michael Rossmann and colleagues (Sun *et al.*, 2014) demonstrate a novel mechanism of DNA delivery adopted by the tail-less coliphage φX174 that requires the tubular structure of DNA pilot protein H to be wide enough to allow the passage of two antiparallel strands of ssDNA: this tube acts as a tail for the translocation of the viral genome, but it protrudes from the virion only at the time of infection.

Plant viruses do not enter their hosts *via* receptor-mediated endocytosis because of the cell-wall barrier. They gain entry into cells by mechanical injury caused by the vectors that transmit them or by mechanical inoculation. From the primary site of infection, they move from cell to cell and systemically infect plants using specialized proteins called movement proteins encoded by their genome which alter the plasmodesmata, the communication channels between cells. However, some of the plant viruses have been shown to enter and replicate within yeast cells and also enter animal cells. For example, cowpea mosaic virus (CPMV) appears to preferentially enter cancer cells by interacting with a 54 kDa cell-surface protein, vimentin (Koudelka *et al.*, 2009). CPMV conjugated with the anti-cancer drug doxorubicin showed effective toxicity towards HeLa cells (Albajali *et al.*, 2013). The feasibility of using fluorescent mCherry-potato virus X as a probe for optical imaging in human cancer cells has been demonstrated (Shukla *et al.*, 2013). Several plant viral coat protein genes, when expressed in a suitable expression system, assemble into virus-like particles (VLPs). VLPs could be engineered biologically or by chemical modifications so as to possess specific surface characteristics. VLPs are increasingly being investigated for several applications because of their nanometre size range, symmetry, polyvalency, monodispersity, efficient and inexpensive production, biocompatibility and biodegradability (reviewed in Steinmetz, 2010). Studies performed on mice using CPMV and cowpea chrorotic mottle virus (CCMV) showed that both viruses entered different tissues. CPMV particles largely accumulated in the liver and spleen while CCMV was mostly found in the thyroid gland (Steinmetz, 2010).

RNA viruses typically encode their own RNA-dependent RNA polymerase (RdRP) to ensure genome replication within the infected cell. These enzymes are major targets for the development of antiviral compounds against the corresponding pathogens. A number of studies have revealed the structures of different RdRPs and provided a mechanistic

insight to the RNA-synthesis process in the case of single-stranded positive-sense RNA viruses and double-stranded RNA viruses. However, such information is still lacking for single-stranded negative-polarity RNA viruses. The most complete sets of structural data are from Picornaviruses and Caliciviruses. These structures provide high-resolution snapshots of the conformations adopted by the enzyme in the different steps of the catalytic cycle. Ng and colleagues provide structural and biochemical evidence in Norovirus for a range of conformational states of the polymerase associated with RNA-template–primer binding, NTP binding, catalysis, RNA translocation and pyrophosphate release (Ng *et al.*, 2008). Such structural information is essential for the design of new antiviral therapies.

2. Virus–receptor interactions

The initial event in the life cycle of a virus is its contact with the target cell. This encounter is mediated by interactions between components on the viral surface (capsid proteins or viral membrane anchored proteins) and several elements on the cellular membrane that can get attached to the virus. Most viruses establish contacts with cellular receptors which could be sialic acid bound molecules or proteins on the cell surface. Frequently, the first virus–cell contact is mediated by adhesion receptors that allow reversible attachment of the virus to particular cells or organs, mainly by means of electrostatic interactions. This initial interaction could lack specificity and is employed to enhance the encounter between virus and entry receptors, which irreversibly bind the virus, allowing them access to the intracellular milieu either by direct endocytosis/pinocytosis or by inducing fusion or penetration of the viral genome into the cell.

Characterization of the virus–receptor interactions has been a focus of sustained interest and many outstanding contributions of high-resolution X-ray structures of viral proteins in complex with their specific receptors have been reported during the last few years. Haemagglutinin (HA)–receptor interaction was extensively studied in different influenza virus subtypes. Influenza HA is synthesized in the infected cell as a polypeptide precursor (HA0) of about 550 amino acids that is proteolytically cleaved to generate the HA1 (approximately the N-terminal two thirds) and HA2 (the C-terminal third) chains that remain covalently linked by a disulfide bond. At the newly created HA2 N-terminus there is a stretch of hydrophobic amino acids, called the fusion peptide, which is inserted into the membrane during fusion (see the next section). The overall structure of the influenza HA resembles an elongated spike sticking out of the membrane. The distal head, formed exclusively by HA1 sequences, bears the receptor-binding site, formed by a shallow pocket exposed on an outward-forming surface. Different residues within this pocket mediate viral attachment to the α2,3- or α2,6-linked sialic acid moieties in avian or human cellular surfaces, respectively, as observed by X-ray crystallography in different complexes with receptor analogues. The sialic acid binding pocket of HA1 is formed by a loop–helix–loop structure composed of the 130-loop, 190-helix and 220-loop (Fig. 1).

A repertoire of crystal structures of HA–receptor complexes from different influenza outbreaks has recently been reported, providing a picture of the evolutionary path for the emergence of the H3N2 Influenza strain between 1968 and 2002 (Lin *et al.*, 2012). Furthermore, this research revealed a number of features that contribute to the changes in receptor affinity. In particular, high-resolution X-ray structures have shown how the point mutation Asp225 to Asn in the 220-loop diminishes the specific HA–receptor interactions, justifying how this fixed mutation decreases the human receptor binding in H3N2 viruses circulating since the last decade, thereby reducing the impact of this influenza subtype. Moreover, the structures of different HA–receptor complex analogues have revealed the key determinants for animal-to-human transmissibility of the highly pathogenic avian H5N1: a single substitution (Gln to Leu) promotes binding to the human receptor to the detriment of the avian receptor (Xiong *et al.*, 2013). Interestingly, this mutation leading to increased hydrophobicity was responsible for the high transmissibility of human pandemic H2 and H3 influenza subtypes.

Recent crystallographic studies also revealed details of virus–host interactions for Measles virus (MV), which is a single-stranded, negative-sense, enveloped RNA virus. Three seminal works provide a landscape of atomic interactions between MV haemagglutinin (MV-H) and the protein receptors. The structure of MV-H bound to human CD46 fragment (Santiago *et al.*, 2010) provided a rationale for the critical role of only two residues (Tyr481 and Gly546) in the alternative use of CD46 as a virus receptor, preserving the affinity for the primary SLAM receptor. The structure of the complex between MV-H and SLAM has also been determined (Hashiguchi *et al.*, 2011). In addition, the structure of MV-H bound to a third receptor protein, nectin-4 which has been identified recently, has also been reported (Zhang, Lu *et al.*, 2013), showing that nectin-4 binds to MV-H by means of its N-terminal domain, establishing extensive hydrophobic interactions in the same concave lateral groove of MV-H as the other receptors but with minor differences (Fig. 1). These high-resolution structures that illustrate the variability of interactions important for MV infection can also explain the extended virus tropism mediated by the alternative use of receptors or the receptor-specificity switching. Additionally, this structural information provides a framework for rational antiviral design.

Other crystallographic data have allowed characterization of the early steps of infection in important pathogens such as the single-stranded, positive-sense RNA Coronaviruses (CoV). The spike glycoprotein (S) of the CoV particle forms characteristic surface projections that are employed to interact with the target cell receptors. The determinants of CoV tropism are located in the distal globular domain of this viral membrane glycoprotein, which mediates attachment of the virus to the cell-surface molecules. In fact, the structure of the porcine respiratory coronavirus (PRCV) receptor-binding domain of the S protein in complex with its cell-receptor

protein, the aminopeptidase N (pAPN), shows that the conformation of the receptor-binding edge in the envelope S protein contains the determinants of their receptor-binding specificity (Fig. 1). This work also reports the interaction of a single glycan involved in protein interactions that must be conserved among a group of CoVs (Reguera *et al.*, 2012).

Very recently, Wang *et al.* (2013) determined the structure of the receptor-binding domain (RBD) present in the S protein of the newly identified Middle East respiratory syndrome coronavirus (MERS-CoV) in complex with its cellular receptor, the dipeptidyl peptidase 4 (DPP4). Interestingly, this receptor does not resemble any other CoV receptor in either sequence or structure. On the other hand, the MERS-CoV RBD appears to be different from the SARS-CoV reported earlier (Li *et al.*, 2005), despite the relative similarity in the core subdomain, contributing to the variance in receptor specificities (Fig. 1). The interphases of interactions consist of two regions distant from the DPP4 active site, where Li *et al.* (2005) identified relevant residues that support the viral–receptor interaction that will potentially guide therapeutic strategies against this virus.

3. Viral and eukaryotic fusion proteins

The viral fusion glycoproteins structurally characterized to date fall into three classes, although the membrane-fusion pathway seems to be very similar for all the enveloped viruses studied so far.

Class I fusogens are characterized by a seven-residue periodicity of nonpolar amino acids (called 'heptad repeats') that give rise to a central parallel trimeric α-helical coiled coil along the long axis of a rod-shaped molecule (Fig. 2; Igonet & Rey, 2012). The best characterized members of this class are the influenza virus HA (Wilson *et al.*, 1981; Bullough *et al.*, 1994) and the fusion protein (F) of paramyxoviruses (Yin *et al.*, 2005, 2006), but this class also includes fusion proteins
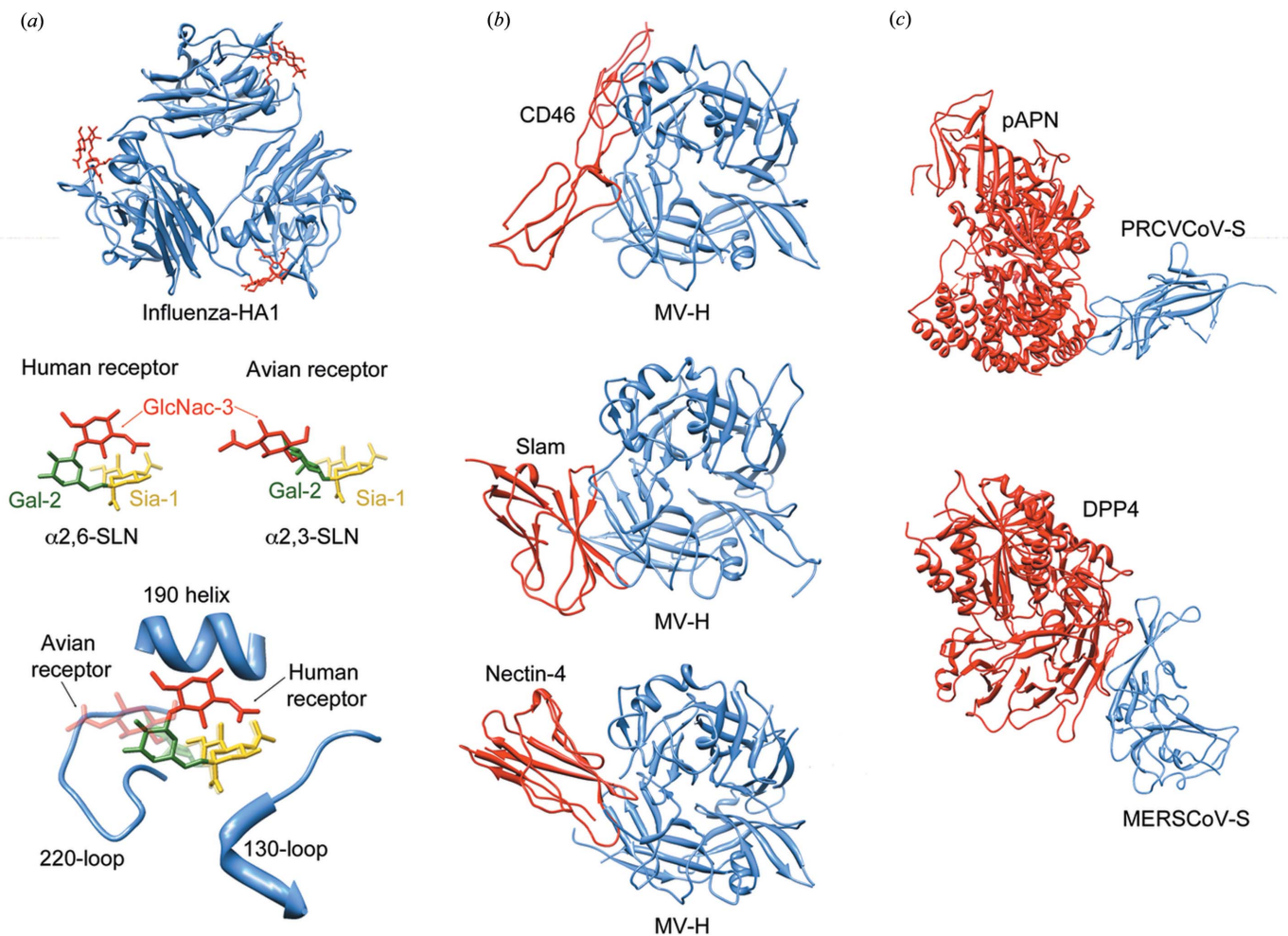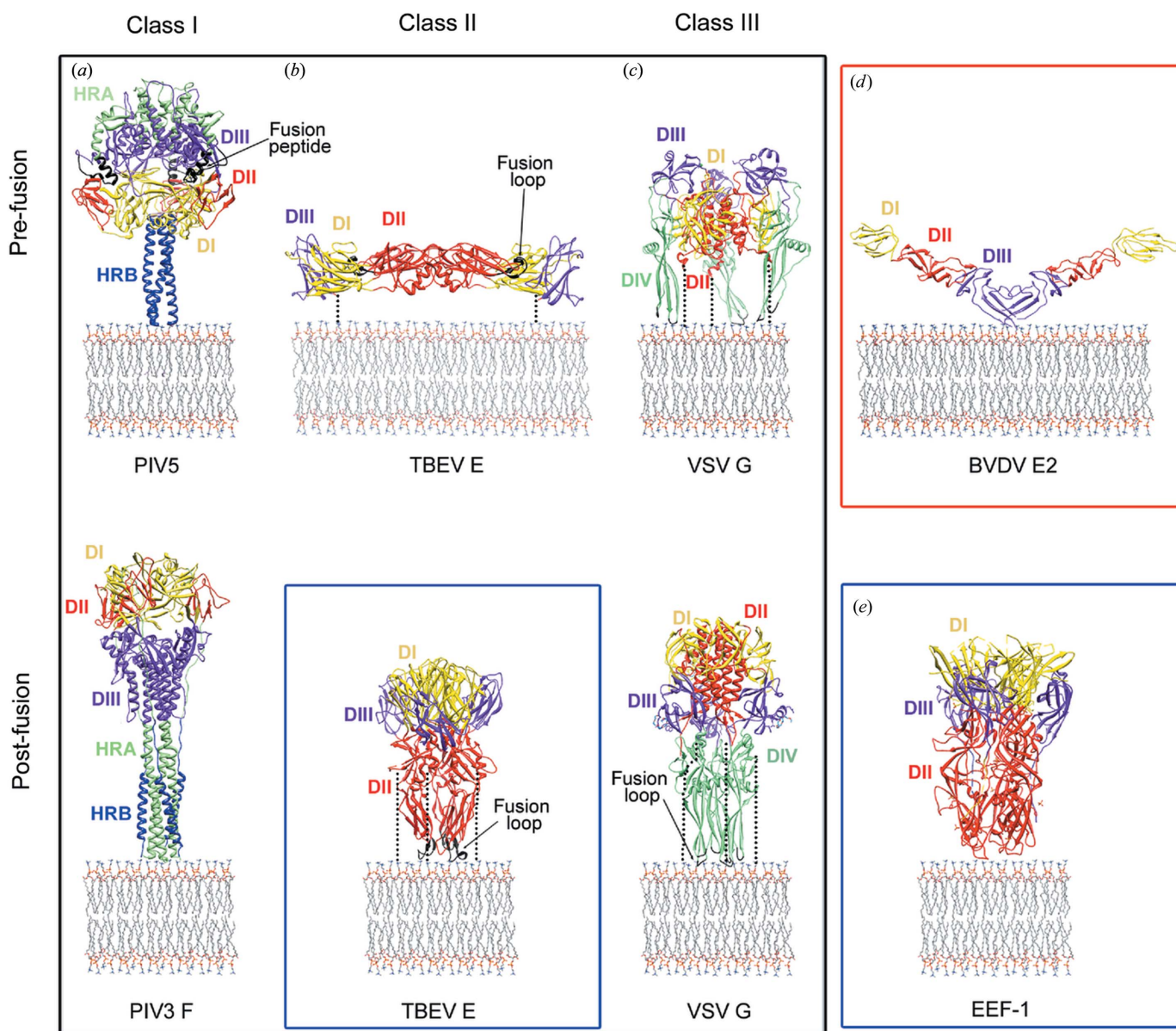

Figure 1
(*a*) The sialic acid binding pocket of influenza HA. A ribbon representation of an HA1 trimer is shown in the upper part. The structure of the bound human receptor analogue is shown as red sticks (PDB entry 4bh3, Xiong *et al.*, 2013). The components of the human (left) and avian (right) receptor analogues are shown in the middle part of the figure and a close-up of the receptor-binding pocket with the superimposition of the two ligands is shown in the lower part. (*b*) The MV-H protein bound to the receptors: CD46 (PDB entry 3inb, Santiago *et al.*, 2010; top), Slam (PDB entry 3alz, Hashiguchi *et al.*, 2011; middle) and nectin-4 (PDB entry 4gjt, Zhang *et al.*, 2013; bottom). Ribbon drawings, showing the MV-H molecules in blue and the different receptor molecules depicted in red. (*c*) Structure of the PRCV CoV S-APN complex (PDB entry 4f5c, Reguera *et al.*, 2012). (*d*) Structure of the complex MERS-CoV RBD-DPP4 complex (PDB entry 4l72, Wang *et al.*, 2013).

from retroviruses, coronaviruses and filoviruses (see Baquero *et al.*, 2013 for a review). Both the pre-fusion and post-fusion structures of class I fusion glycoproteins are trimeric. In most of the cases, the subunits constituting the trimer result from the proteolytic cleavage of a precursor into two fragments. The resulting C-terminal fragment, which is anchored in the viral membrane by a hydrophobic transmembrane (TM) domain, bears a hydrophobic fusion peptide at or near its amino-terminal end that is buried at a protein–protein interface in the pre-fusion state (Fig. 2). This peptide gets inserted into the target membrane during fusion. Thus, in the post-fusion

conformation, the shape of the trimeric molecule resembles an elongated hairpin-like structure with the fusion peptide and the TM domain located at the same end, as expected at the end of the fusion process.

Until recently, class II proteins had only been found in flaviviruses (protein E) and alphaviruses (protein E1), which share many key characteristics (reviewed in Modis, 2014). The first X-ray structure of a class II glycoprotein was that of the tick-borne encephalitis (TBE) flavivirus E protein ectodomain (Rey *et al.*, 1995), solubilized from virions by limited trypsin digestion. Similar structures are now available for the

Figure 2
The three classes of viral fusion glycoproteins. Ribbon diagrams of the pre-fusion (top) and post-fusion (bottom) conformations of the paramyxovirus (class I; PDB entries 2b9b, Yin *et al.*, 2006, and 1ztm, Yin *et al.*, 2005) (*a*); flavivirus (class II; PDB entries 1svb, Rey *et al.*, 1995, and 1urz, Bressanelli *et al.*, 2004) (*b*); and rhabdovirus (class III; PDB entries 2j6j, Roche *et al.*, 2007, and 2cmz, Cho *et al.*, unpublished work) (*c*) proteins. For each class of fusogen, the equivalent protein domains are highlighted with identical colours and explicitly labelled (DI, yellow; DII, red; DIII, purple; DIV and heptad repeat region A, HRA, light green; HRB, blue; fusion peptide or loops, black). (*d*) The novel fold of the envelope protein E2 from BVDV (PDB entry 4ild, Li *et al.*, 2013). (*e*) The structure of the eukaryotic fusogen EEF1 (PDB entry 4oje, Pérez-Vargas *et al.*, 2014).

ectodomain of dengue virus types 2 and 3. The polypeptide chain of the E protein is organized in three globular domains, essentially constituted by β-sheets (Fig. 2). The hydrophobic fusion loop in the virion is located at the tip of domain II, buried by interactions with domain III of the adjacent monomer in the EE dimer. These proteins are attached to the viral membrane *via* a C-terminal TM anchor, which is linked by a flexible 'stem' region to the ectodomain (Fig. 2). The envelope proteins from flaviviruses and alphaviruses assemble into icosahedral outer shells, but the mode of assembly differs in the two families, with alphaviruses forming canonical (*T* = 4) quasi-equivalent assemblies (Lescar *et al.*, 2001; Zhang *et al.*, 2011) and flaviviruses forming unusual non-equivalent icosahedral assemblies (Mukhopadhyay *et al.*, 2003; Zhang, Ge *et al.*, 2013).

Recent data show that the class II fold is more widely distributed than previously anticipated. Indeed, the crystal structures of the glycoprotein C (Gc) from Rift Valley fever virus (RVFV) reveal a class II fusion protein fold (Dessau & Modis, 2013). RVFV belongs to the phlebovirus genus in the *Bunyaviridae* family, unrelated to flaviviruses or alphaviruses (Modis, 2014). The structure of RVFV Gc is strikingly similar to flavivirus E structures. In particular, both viruses share the same head-to-tail configuration of the protein dimers, with the fusion loop buried at the dimer interface (Fig. 2). Also, the two fusion loops have the same tightly folded glycine-rich structure suggesting that phleboviruses may be evolutionarily related to alphaviruses and flaviviruses.

In another recent advance, the E1 protein of rubella virus, which belongs to the same *Togaviridae* family as alphaviruses, was found to have a class II fold, although with a more divergent structure that shows important differences in the fusion loops (DuBois *et al.*, 2013). In addition, rubella E1 does not form icosahedral assemblies.

Very recently, two independent groups have solved the structure of the larger envelope protein, E2, from the pestivirus bovine viral diarrhea virus (BVDV) (Fig. 2), another member of the *Flaviviridae* family (Fig. 2). The structure unveiled that E2 has marked differences with the rest of the class II fusion proteins, defining a new structural class of fusogens (El Omari *et al.*, 2013; Li *et al.*, 2013).

Class III proteins are trimeric before and after fusion and share structural characteristics with both class I and class II fusion glycoproteins (Fig. 2). A class III fold was identified in the structures of the ectodomains of the fusion glycoproteins G of vesicular stomatitis virus (VSV) and B (gB) of herpes simplex virus 1 (HSV-1). Their comparison revealed an unanticipated structural similarity between the two proteins, although no sequence similarity had previously been detected. The structures of the Epstein–Barr Virus (EBV) gB and baculovirus gp64 fusogens also belong to this class (reviewed in Baquero *et al.*, 2013). VSV G is the only class III fusion protein for which the X-ray structures of both the pre-fusion (Roche *et al.*, 2007) and post-fusion (Roche *et al.*, 2006) states have been determined. The VSV G protein possesses both receptor-binding and fusion-promoter activities. As in the case of influenza virus, binding of rabdovirus G to a poorly

characterized receptor at the cell surface induces endocytosis of the virus particle. Acidification of the endosome triggers G for membrane fusion. However, in contrast to all other fusion proteins, the low-pH inactivation of VSV G is reversible. The crystal structure of the G protein in the pre-fusion conformation revealed two fusion loops reminiscent of class II proteins that are oriented downwards towards the viral membrane. After low-pH exposure, the fusion domain moves upwards by flipping relative to the central core of the trimer to form an intermediate pre-hairpin structure. This is followed by the reversal of the molecule around the central rigid block formed by lengthening of the central helix in the grooves of the central core in an antiparallel manner (Fig. 2). This six-helix bundle has an obvious resemblance to those of the class I proteins. It is likely that the transition of VSV G from the pre-fusion to the post-fusion conformation involves disassembly of the trimer into the monomers and reassembly into trimers upon interaction of the fusion loops with the target membrane (Albertini *et al.*, 2012).

While vesicle fusion is required for a number of essential biological processes such as exocytosis and synaptic transmission, cell–cell fusion processes are crucial in development. In all cases, membrane fusion follows the same steps already described for virus–host interaction. In analogy with virus–cell fusion, vesicle and cell–cell fusion requires the formation of highly stable protein assemblies that provide the energy necessary to overcome the repulsive forces of membranes in close proximity. Also vesicle and cell–cell fusion, as in viral fusion, require higher-order multimerization of the fusion proteins. The main difference between virus–cell fusion and vesicle or cell–cell fusion is that in the former process the protein fusogen is present only in the viral membrane. In contrast, the proteins involved in vesicle fusion and cell–cell fusion are initially inserted in the two membranes predestined to fuse. Studies of intracellular fusion events have revealed two families of fusion proteins, the SNARE [soluble *N*-ethylmaleimide-sensitive factor (NSF) attachment protein receptors] (Südhof & Rothman, 2009) and the dynamin-like 'atlastin' GTPases (Bian *et al.*, 2011; Byrnes & Sondermann, 2011). In both cases, membrane merger results from trans-oligomerization of molecules anchored in the opposed membranes, followed by a conformational change that pulls the two membranes towards each other (reviewed in Moss *et al.*, 2011). In contrast to vesicle fusion, cell–cell fusion entails the same set of fusion proteins in the two membranes. For instance, the exceptional process of hypodermal cell fusion in *C. elegans* to form a large multi-nucleated syncytium of all skin cells is driven by the epithelial fusion failure 1 protein, EFF-1. Unlike SNAREs and viral fusogens, EFF-1 has been shown to be required in both cell membranes for fusion. Recently, the 2.6 Å resolution X-ray structure of the EFF-1 protein has been determined by Felix Rey and colleagues (Pérez-Vargas *et al.*, 2014). The EFF-1 trimer shows the same three-dimensional fold and quaternary conformation of post-fusion class II viral fusion proteins, although it lacks the nonpolar 'fusion loop', indicating that it does not insert into the target membrane and suggesting that membrane fusion driven by EFF-1 entails

trans-trimerization. The authors further show that the blocking of EFF-1 trimerization interferes with the fusion reaction. The study also provides evidence that the monomeric form of EFF-1 is metastable and that trimerization is irreversible, matching the properties of the pre- and post-fusion forms, respectively, of the viral counterparts, demonstrating an evolutionary link with viral fusion proteins (Pérez-Vargas *et al.*, 2014).

## 4. Genome delivery in bacteriophages

Bacteriophages show widely diverse structures and types of nucleic acid genomes; they have helical or icosahedral capsids and may not include a lipid envelope. Phages range in structural complexity from very small and simple non-enveloped icosahedral viruses (*e.g.* the ssDNA phage φX174) and long but simple helical viruses (filamentous phages), to large tailed viruses (*e.g.* the dsDNA phages T4 or φ29).

Bacteriophage tails are fascinating molecular machines specifically evolved to recognize host cells, penetrate the cell envelope barrier and deliver DNA into the cytoplasm (Fig. 3). The *Caudovirales* tails display very different sizes and morphologies with lengths ranging from ∼100 Å in some podophages to ∼8000 Å in some siphoviruses (reviewed in Fokine & Rossmann, 2013). Long tails of *Siphoviridae* and *Myoviridae* phages consist of the tail tip complex, which is responsible for host recognition and initiation of the infection process, the tail tube, which makes a conduit for genomic DNA, and terminator proteins, which terminate the tail assembly and create the binding interface for head attachment (Davidson *et al.*, 2012). The tail of *Myoviridae* phages also contains a contractile sheath surrounding the tail tube (Leiman *et al.*, 2010; Leiman & Shneider, 2012). The tail tip complex has different size and morphology in different phages. Phages which use protein receptors for cell binding usually have conical tail tips [*e.g.*, SPP1 or λ (Plisson *et al.*, 2007; Pell *et al.*, 2009)], whereas phages using polysaccharide receptors usually have elaborate baseplates at the distal end of the tail [*e.g.*, T4 or the lactococcal phages TP901-1 and p2 (Leiman *et al.*, 2010; Veesler *et al.*, 2012; Sciara *et al.*, 2010)]. Furthermore, phages usually have side-tail fibres or spikes, attached to the periphery of the tail tip complex, as well as a central tail spike (Fig. 3).

Bacteriophage baseplates vary in size and complexity. The most extensively studied *Myoviridae* baseplate is that of phage T4, which is composed of ∼140 polypeptide chains of at least 16 different proteins. The structures of T4 baseplate before and after tail contraction have been determined using cryo-electron microscopy (Kostyuchenko *et al.*, 2005; Leiman *et al.*, 2004) and the structures of nine baseplate proteins have been determined using X-ray crystallography (Leiman *et al.*, 2010). The T4 baseplate attaches six long tail fibres of ∼1450 Å length, which reversibly bind to the *Escherichia coli* lipopolysaccharide (LPS) and/or OmpC molecules, serving for primary host
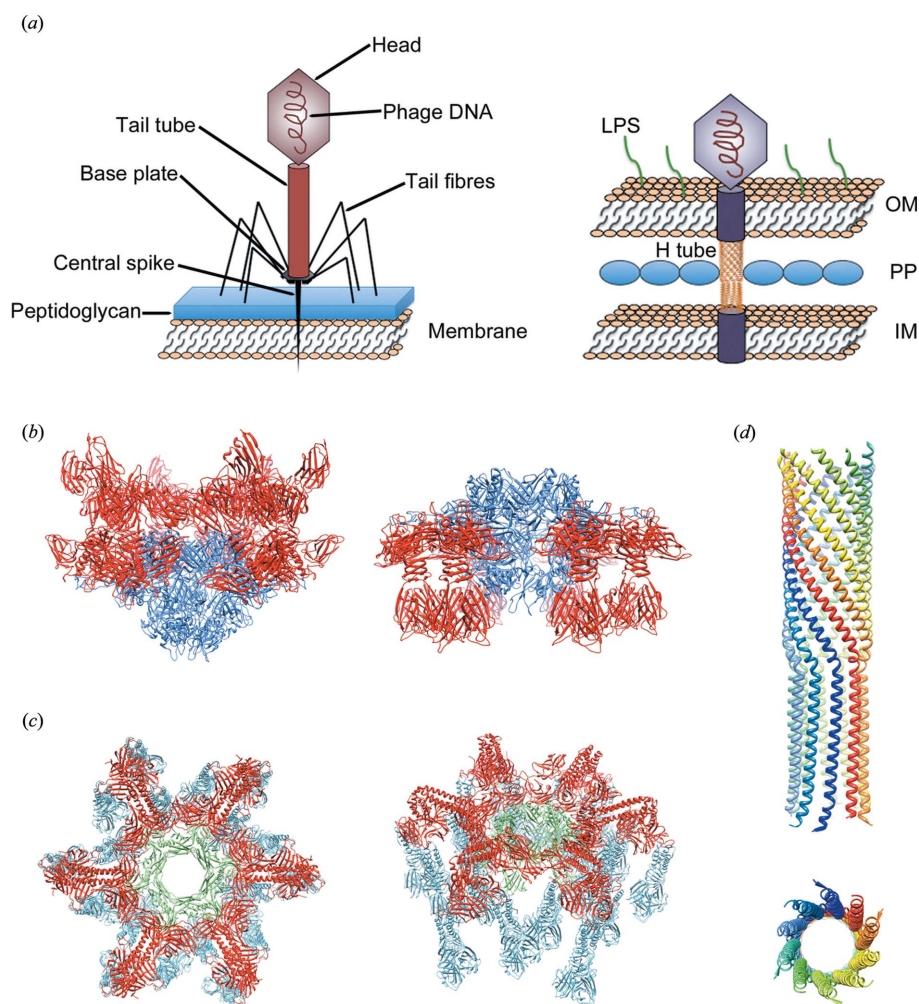
Figure 3
(*a*) Schematic representation of a tailed phage. LPS, lipopolysaccharide; OM, outer membrane; PP, periplasmic space; IM, inner membrane. (*b*) The baseplate of phage P2 before activation, with the receptor-binding domains of the receptor-binding protein (red) pointing upwards, away from the bacterium (left; PDB entry 2wzp, Sciara *et al.*, 2010). The movement of 200° by the receptor-binding protein that completely reverses the orientation. The right panel shows the baseplate of P2 after activation by calcium ions with the receptor-binding domains pointing downwards, towards the bacterium (PDB entry 2x53, Sciara *et al.*, 2010). (*c*) The baseplate of phage TP901-1 (PDB entries 4div and 4diw, Veesler *et al.*, 2012), top view looking down the phage tail tube axis (left) and lateral view (right). (*d*) The structure of the H protein coiled-coil tube (PDB entry 4jpp, Sun *et al.*, 2014). Ribbon representation of the structure, with the monomers individually coloured. (bottom) Top view of the tube, with the N-terminus pointing to the viewer.

recognition. It also has six short tail fibres attached to its periphery. Upon binding of the long tail fibres, a signal is transmitted to the baseplate causing the six short tail fibres to extend and bind irreversibly to the LPS, resulting in a conformational rearrangement where the baseplate switches from a dome-shaped form to a star-shaped conformation which, in turn, triggers the contraction of the tail sheath causing the specialized tip of the inner tail tube to puncture the outer membrane. The genome then passes through the tail tube into the cytoplasm (Leiman *et al.*, 2004; reviewed in Fokine & Rossmann, 2013).

Recent crystallographic and electron microscopy studies have revealed the structures of two *Siphoviridae* baseplates, those of phages p2 and TP901-1, infecting Gram-positive *Lactococcus lactis* (Sciara *et al.*, 2010; Veesler *et al.*, 2012). The baseplate of phage P2 is composed of three protein species. The central part of the baseplate is formed by a circular hexamer of ORF15 proteins with a central hole. A trimer of ORF16 is located at the bottom of the baseplate, forming a closed dome that does not allow DNA passage. Six ORF18 trimers are attached to the central ring, each trimer interacting with a carboxy-terminal extension of an ORF16 monomer. ORF18 is the receptor-binding protein (RBP) and electron microscopy reconstructions show that these proteins point upwards, towards the capsid in the free virion. In the presence of Ca^{2+}, a cation essential for infection, the RBP complex is rotated $200°$ to point downwards towards the host cell. This conformational change of the baseplate also leads to the separation of three ORF16 monomers, opening up a hole in the centre of the baseplate, presumably allowing passage of DNA into the host (Fig. 3).

The baseplate of phage TP901-1 is composed of multiple copies of four different proteins. The centre is a hexameric circular core formed by the Dit protein. From the core, six arms emanate, each arm being composed of a trimer of the BppU protein. The arms are α-helical until the elbow. The rest of the arm points downwards and forms an adaptor domain for receptor-binding proteins. Three trimeric receptor-binding proteins bind to each adaptor leading to a total of 54 sites for binding the host cell envelope saccharides (Fig. 3). The hole in the centre of the hexameric core is filled by a central tail fibre (Tal) as seen in the electron microscopy reconstruction (Veesler *et al.*, 2012). The receptor-binding proteins point downwards, *i.e.* towards the host bacterium, ready for adhesion. No Ca^{2+} ions are necessary for activation of TP901-1, suggesting that conformational changes are probably not needed for receptor binding. How receptor binding is related to DNA transfer in these phages is less clear; perhaps the strong binding with up to 54 receptor molecules pushes the central tail fibre against the cell wall and the force exerted by the cell wall against the tail is sensed by the other end of the tail fibre, which opens a hole at the end of the tail tube.

Very recently, Rossmann and colleagues have demonstrated that the small single-stranded (ss)DNA bacteriophage φX174 possesses a mechanism of DNA translocation similar to that described for the tailed phages. However, the φX174 'tail' protrudes from the virion only at the time of infection (Sun *et*

al., 2014). φX174 is a small icosahedral microvirus with a circular ss(DNA). X-ray and electron microscopy studies showed that the virus capsid has spikes on all pentameric vertices (McKenna *et al.*, 1992). The 260 Å-diameter capsid is constructed from 60 copies of the F protein. In addition, each of the 12 spikes consists of five G proteins, protruding 32 Å above the F-protein shell. The capsid also contained 60 copies of the DNA-binding protein J and ten to 12 copies of the DNA pilot protein H. Until now, the structure and location of the H proteins remained unknown as all structure determinations of this phage made use of the icosahedral symmetry. Rossmann and colleagues have now solved the crystal structure of the central domain of the H protein at 2.4 Å resolution. It consists of a 170 Å-long α-helical tube built from ten α-helices (Sun *et al.*, 2014). Each monomer is kinked at approximately the middle of the molecule, resulting in a tube with their amino termini arranged in a right-handed super-helical coiled coil and their carboxy termini arrayed in a left-handed super-helical coiled coil (Fig. 3). The N-terminal domain, which is slightly conical, has a minimum internal diameter of approximately 22 Å, whereas the internal diameter of the C-terminal domain is about 24 Å. With a minimum internal diameter of 22 Å, the H tube can easily accommodate two unpaired ssDNA strands with intercalated bases in a similar way to that in which the circular ssDNA genomes of filamentous bacteriophages are packaged into cylindrical fivefold symmetric capsids with comparable inner dimensions (Russel *et al.*, 2006). The H tube also appears long enough to span the periplasmic space, with both N- and C-terminal regions rich in Ala, Gly and Ser residues, which have high occurrence in transmembrane helices. Therefore, these regions would probably anchor the tube in the *E. coli* inner and outer membranes.

5. RNA-dependent RNA polymerases

RNA-dependent RNA polymerases (RdRPs) are the catalytic components of the RNA replication and transcription machineries and the central players in the life cycle of RNA viruses. RdRPs belong to the superfamily of template-directed nucleic acid polymerases, including DNA-dependent DNA polymerases (DdDP), DNA-dependent RNA polymerases and reverse transcriptases (RT). All theses enzymes share a cupped right-hand structure, including fingers, palm and thumb domains, and catalyse phosphodiester bond formation through a conserved two-metal-ion mechanism (Steitz, 1998). A structural feature unique to RdRPs is the 'closed-hand' conformation, as opposed to the 'open-hand' found in other polynucleotide polymerases. This closed-hand conformation is accomplished by interconnecting the fingers and thumb domains through the N-terminal portion of the protein and several loops protruding from fingers, named the fingertips, that completely encircle the active site of the enzyme (Ferrer-Orta *et al.*, 2009; Ng *et al.*, 2008). In the prototypic RdRPs, the closed 'right-hand' architecture encircles seven conserved structural motifs that are arranged in the order G, F, A, B, C, D and E containing a number of highly conserved amino acids

(Fig. 4). The only exception to this scheme is found in Birnavirus and Permutatetravirus polymerases, where motif C is encoded upstream of motif A (Garriga *et al.*, 2007; Ferrero *et al.*, 2012). Each of the seven motifs in the RNA polymerase domain adopts a specific fold that extends beyond the regions of sequence similarity into the so-called homomorphs for most RNA virus RNA polymerases (Lang *et al.*, 2013). Three well defined channels have been identified in the RdRP structures, serving as: the entry path for template (template channel) and for nucleoside triphosphates (NTP channel) and the exit path for the dsRNA product (central channel) (Fig. 4).

The recent explosion of structural and functional information on viral RdRPs has provided insights into both initiation of RNA synthesis and the elongation process. Correct initiation of RNA synthesis is essential for the integrity of the viral genome. There are two main mechanisms by which viral replication can be initiated: primer-independent or *de novo*, and primer-dependent initiation, reviewed in van Dijk *et al.* (2004). Briefly, in the *de novo* synthesis, one initiation nucleotide provides the 3′-hydroxyl for the addition of the next nucleotide whereas the primer-dependent initiation requires the use of either an oligonucleotide or a protein primer as provider of the hydroxyl nucleophile. The members of the *Picornaviridae* and *Caliciviridae* families use exclusively the protein-primed mechanism of initiation. In this process, a tyrosine residue provides the hydroxyl group for the formation of a phosphodiester bond with the first nucleotide. Enzymes using primer-dependent initiation display a more accessible active-site cavity, enabling them to accommodate the small VPg protein that acts as a primer protein in RNA synthesis. The foot-and-mouth disease virus (FMDV) VPg protein lines the RNA-binding cleft of the corresponding RdRP (Ferrer-Orta, Arias, Agudo *et al.*, 2006), positioning its Tyr3 OH group as a molecular mimic of the free 3′-OH group of a nucleic acid primer at the active site for nucleotidylylation, thereby initiating replication. In the presence of oligo-adenylate and UTP, the product of the reaction, VPg–UMP, can be observed in the crystal structure (Ferrer-Orta, Arias, Agudo *et al.*, 2006), at a position remarkably similar to the position of the template–primer RNA duplex (Ferrer-Orta *et al.*, 2004, 2007). After nucleoti-dylylation of VPg, some structural rearrangements of the RdRP follow, marking the transition from initiation to the elongation phase of RNA synthesis. By contrast, enzymes using the *de novo* initiation, Reovirus and Flavivirus RdRPs, contain additional structural elements that fill most of the active-site cavity, providing a support platform for the primer nucleotides, thereby enabling *de novo* initiation of RNA synthesis (Choi & Rossmann, 2009; Harrus *et al.*, 2010). Crucially, these protrusions are also able to undergo large conformational rearrangements to facilitate translocation of the RNA recently synthesized (Butcher *et al.*, 2001; Mosley *et al.*, 2012).

The replication elongation process can be roughly divided into three steps, including nucleotide selection, phosphodiester bond formation and translocation to the next nucleotide for the subsequent round of nucleotide addition. The structures of a large number of RdRP-RNA-rNTP replication–elongation complexes have been determined for different members of the *Caliciviridae* and *Picornaviridae* families. These studies have provided
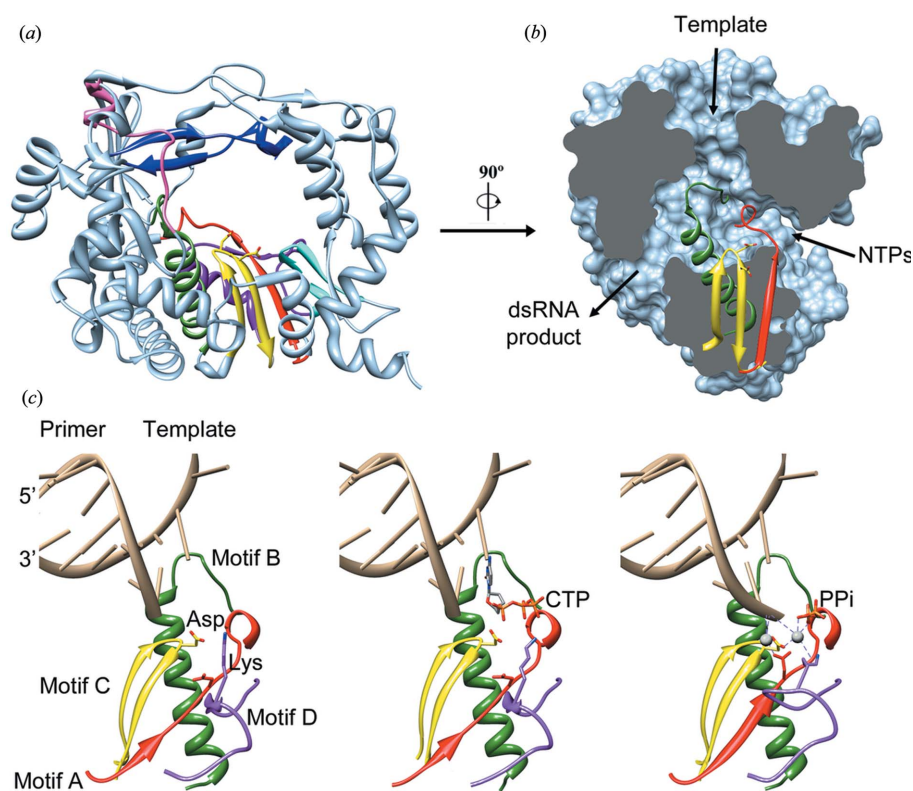
Figure 4
Overal structure of a viral RdRP. (*a*) Ribbon representation of a typical RdRP (FMDV 3D, PDB entry 1u09, Ferrer-Orta *et al.*, 2004). The seven conserved motifs are indicated in different colours: motif A, red; motif B, green; motif C, yellow; motif D, purple; motif E, cyan; motif F, blue; motif G, pink. The side chains of the catalytic Asp residues in the active site are also shown as sticks. (*b*) Lateral view of a surface representation of the enzyme (grey) that has been cut to expose the three channels that are the entry and exit sites of the different substrates and reaction products. The structural elements that support motifs A, B and C are also shown as ribbons. (*c*) Sequential structures illustrating the movement of the different residues within the palm domain from a binary RdRP-RNA open complex (left) to an RdRP-RNA-rNTP open ternary complex (middle), and a closed ternary complex (right). Image based on different poliovirus elongation complexes. The different structures correspond to the 3D-RNA (PDB entry 3ol6, Gong & Peersen, 2010), 3D-RNA-CTP open complex (PDB entry 3olb, Gong & Peersen, 2010) and 3D-RNA-CTP closed complex (PDB entry 3ol7, Gong & Peersen, 2010) structures of poliovirus elongation complexes, respectively.

important insight into the structural alterations associated with each catalytic step (Ferrer-Orta *et al.*, 2009; Lescar & Canard, 2009; Ng *et al.*, 2008; Zamyatkin *et al.*, 2008, 2009; Gong & Peersen, 2010; Gong *et al.*, 2013; Garriga *et al.*, 2013; Sholders & Peersen, 2014). These structures indicated that RdRPs use subtle rearrangements within the palm domain to fully configure the active site for catalysis upon correct rNTP binding (Fig. 4). Among these movements, a flexible loop at the N-terminus of motif B assists in the correct positioning of the template nucleotide in the active site, facilitating the binding of the incoming rNTP *via* base pairing to a fully prepositioned templating base and stacking with the primer (Ferrer-Orta *et al.*, 2007, 2009; Garriga *et al.*, 2013). Binding of the correct nucleotide then induces the realignment of β-strands in the palm subdomain that includes structural motifs A and C, resulting in the repositioning of the catalytic aspartate in motif A to allow interactions with both the metal ions required for RdRP function (Zamyatkin *et al.*, 2008, 2009; Gong & Peersen, 2010; Gong *et al.*, 2013). Recent data also suggest that a conserved lysine residue within motif D can coordinate the export of the inorganic phosphate (PPi) group from the active site, once catalysis has taken place (Yang *et al.*, 2012), thereby triggering the end of the reaction cycle and allowing translocation of RNA. Finally, structural and functional data in enteroviruses indicate that steric clashes between the motif-B loop and the template RNA would also promote translocation (Sholders & Peersen, 2014).

Furthermore, very recently Ng and colleagues provided evidence for conformational changes occurring in the calicivirus RdRPs during catalysis. These authors compared the crystal structures of the RNA-dependent RNA polymerase from the human Norovirus (NV) determined in more than ten different crystal forms in the presence and absence of divalent metal cations, nucleoside triphosphates, inhibitors and primer-template duplex RNAs. These structural comparisons revealed that, in addition to the active-site closure, the NV RdRP exhibits two other key changes: a rotation of the central helix in the thumb domain by 22°, resulting in the formation of a binding pocket for the primer RNA strand, and the displacement of the C-terminal tail region away from the central active-site groove, which also allows the rotation of the thumb helix (Fig. 4).

Oligomerization of RdRPs has been reported for a number of Picornavirus, Flavivirus and Calicivirus enzymes. Oligomerization may lead to cooperative template binding, lattice formation, and a stimulation of viral RNA polymerase activity *in vitro* (Spagnolo *et al.*, 2010; Lyle *et al.*, 2002; Högbom *et al.*, 2009), which in Flavivirus appears to be specific for the initiation of RNA synthesis (Luo *et al.*, 2000). Also inactive RdRPs can induce stimulation of activity and participate in array formation, which has led to the hypothesis that this mechanism has evolved to stabilize the enzyme during replication. The dimerization of the Hepatitis C virus (HCV) RdRp has been proposed to be mediated by the thumb domain (Chinnaswamy *et al.*, 2010), whereas in PV, residues in the N-terminal part of the polymerase domain and two

interface I aspartates appear to be crucially involved (Spagnolo *et al.*, 2010).

RdRPs synthesize RNA using an RNA template. This biochemical activity, not present in mammalian cells, offers the opportunity to identify very selective inhibitors of this viral enzyme. Antiviral drugs targeting the RdRPs may either directly inhibit polymerase activity or essential interactions with the RNA template, or RdRP–RdRP contacts promoting oligomerization, or interactions with other regulatory proteins. The detailed structural and mechanistic understanding of the conformational changes that occur during catalysis is essential not only for understanding viral replication at the molecular level but also for the design of novel inhibitors capable of trapping the enzyme in specific conformational states. The Flaviviruses, Hepatitis C virus, Dengue virus and West Nile virus, as well as calicivirus NV are clear illustrations of the extent of efforts directed towards developing drugs that inhibit viral replication (Powdrill *et al.*, 2010; Malet *et al.*, 2008; Gentile *et al.*, 2014; Caillet-Saguy *et al.*, 2014; Eltahla *et al.*, 2014).

6. Viral proteases

In most single-stranded RNA viruses, some double-stranded RNA viruses and retroviruses with polycistronic genomes, various functional protein domains are initially expressed as a single polyprotein. Cleavage of the polyprotein into individual functional units is essential for their survival. Hence, polyprotein processing is an integral step in the replication of these viruses. Extensive structural studies on proteases of animal viruses have been carried out as they are potential candidates for antiviral therapy. As of July 2014, the PDB has 307 entries of viral proteases of which 276 are X-ray crystal structures. When filtered at 90% sequence uniqueness, the number of proteases is still as large as 93 of which 60 are structures with bound ligands. Viral proteases possess many unique features which distinguish them from cellular and bacterial proteases. These features include stringent specificity, regulation of activity, novel constellation of residues at the active site, amino-acid sequence unrelated to other proteases of the same class, and adaptability to multiple roles.

Viral proteases are very stringent in their substrate specificities, unlike cellular proteases whose specificities are normally dependent on the P1 residue only (Kay & Dunn, 1990). In many viral proteases, the efficiency of cleavage at a particular site of the polyprotein is dependent on the sequence on either side of the cleavable bond and P1 to P4 and P1′, P2′, *etc.* are also important for specificity. The conformation and accessibility of the peptide bond also contribute to the susceptibility to cleavage. Each susceptible site in a polyprotein may have a different cleavage profile. This allows a way of controlling the cleavage events such that the proportions of individual domains are controlled and their activities regulated (Babé & Craik, 1997; Spall *et al.*, 1997). The stringent substrate specificity of viral proteases has led to their application in molecular biology as a reagent for removal of affinity tags from fusion proteins. The most widely used viral

protease for such applications is the tobacco etch virus NIa protease (Kapust & Waugh, 2000).

Many novel combinations of active-site residues have been identified among viral proteases. One such example is the serine-like cysteine protease family of viral proteases, which possesses a cysteine instead of a serine in the catalytic triad in a trypsin- or chymotrypsin-like fold (Bazan & Fletterick, 1988). The crystal structure of the HRV-14 3C^{pro} revealed the adjustments in the catalytic site required for accommodating a cysteine in the place of serine in the catalytic triad of serine-like cysteine proteases (Matthews *et al.*, 1994). The serine proteases of the cytomegalovirus family have a catalytic triad of histidine, histidine and serine (Qiu *et al.*, 1996). The members of the birnavirus family possess a protease with serine and lysine residues forming a catalytic dyad, similar to the Lon proteases in bacteria (Feldman *et al.*, 2006). The aspartic protease of HIV is also different compared to the cellular aspartic proteases in that it functions as an obligate dimer with the two catalytic aspartates contributed by two different subunits (Pearl & Taylor, 1987; Wlodawer *et al.*, 1989). A similar example of a composite active-site formation is the case of NS2^{pro} of HCV, which has a catalytic triad of histidine, glutamate from one monomer and cysteine from another (Lorenz *et al.*, 2006). The active-site geometry is similar to that in serine proteases. The existence of an altered catalytic triad might result in reduction of catalytic efficiency of viral proteases compared to cellular proteases. Such characteristics are suited to the viral life cycle since, instead of efficiency in proteolysis, viruses require greater stringency in specificity and regulation of function (Babé & Craik, 1997). In the cytomegalovirus protease with a novel catalytic triad, a slow proteolytic step may be optimal for viral assembly, as proteolysis should occur only after the capsid maturation, which may be a relatively slow process. Some of the viral proteases have a structural fold similar to serine proteases although there is no sequence similarity with cellular proteases. The Sesbania mosaic virus protease has a chymotrypsin-like fold and is shown to be similar to *Bacillus intermedius* glutamyl endopeptidase (Gayathri *et al.*, 2006). Athough the catalytic triad can be superposed well with serine proteases, the protease by itself is inactive (Satheshkumar *et al.*, 2005).

The catalytic function of viral proteases has been found to be regulated by other domains that may be attached covalently to the N- or C-terminus of the protease. These regulatory mechanisms also serve to control the time and place of proteolytic activity (Babé & Craik, 1997). It has been found that NS2A and NS4A act as cofactors for the regulation of activity of the NS3 proteases of dengue virus and hepatitis C virus, respectively. These cofactors also contain membrane-spanning domains, which might co-localize the proteases to the membrane before activation (Clum *et al.*, 1997). The oligomerization status of the viral proteases also appears to be variable compared to their cellular counterparts (Babé & Craik, 1997). The aspartyl protease of HIV is an obligate dimer. It has been demonstrated that the interaction of Sesbania mosaic virus (SeMV) protease with VPg in *cis*-

(Protease-VPg) leads to the activation of the protease (Satheshkumar *et al.*, 2005). Similarly VPg covalently linked to the N-terminus of pepper vein banding virus (PVBV) protease NIaPro (VPg-NIaPro) results in the enhancement of activity (Mathur & Savithri, 2012). In both these cases, the interaction of the disordered VPg domain alters the structure and function of the interacting proteases. It was shown that the poliovirus 3CD precursor, which consists of the protease domain 3C^{pro} and the polymerase domain 3D^{pol}, also exhibits protease activity similar to 3C^{pro}, but the polymerase is active only as 3D^{pol} after the elimination of the 3C^{pro} domain (Harris *et al.*, 1992). Dengue virus protease shows 3000- to 6000-fold activation in the presence of the NS2B cofactor (Yusof *et al.*, 2000).

Another feature of viral proteases is their extreme sequence diversity, not only amongst themselves but also with respect to cellular proteases. No significant sequence identity has been found between viral proteases and cellular proteases, though the structural fold might be maintained (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1988). Most of the viral protease X-ray crystal structures have therefore been determined by *ab initio* methods of phase determination using isomorphous replacement or anomalous dispersion. In most cases, only the catalytic residues are conserved. The Coronaviral proteases contain an additional domain consisting mainly of α-helices at the C-terminus which is essential for dimerization (Anand *et al.*, 2002, 2003; Yang *et al.*, 2003). The high diversity of viral proteases might be due to the accelerated replication of the viruses and absence of a proof-reading mechanism in the replication of viruses with an RNA genome (Babé & Craik, 1997).

Many viral proteases perform more than one role. Hence, there are certain features in these proteases that confer unique sequence and structural properties compared to normal cellular proteases. The proteases of the alphaviruses have a compact structure resulting from shortening of loops such that their assembly into capsids is facilitated (Choi *et al.*, 1996). RNA-binding motifs incorporated into the canonical protease domains are observed with many viral proteases of RNA viruses (Mosimann *et al.*, 1997). In addition to the main function of polyprotein processing, viral proteases also appear to be involved in shutting down of host-cell translation by acting upon integral components of the host translation machinery. For example, the viral protease 2A^{pro} encoded by the HRV genome cleaves the eukaryotic initiation factor eIF-4G of the host, preventing host protein synthesis (Gradi *et al.*, 1998; Lloyd *et al.*, 1988). Sindbis core protein is not only a protease undergoing autocleavage from the polyprotein followed by self-inactivation, it also assembles to form the viral capsid (Hahn & Strauss, 1990). Helper component protease (HCpro) has been shown to act as a suppressor of post-transcriptional gene silencing (reviewed in Marathe *et al.*, 2000). However, the protease function may be independent of suppressor activity. Antiviral drug discovery with proteases of the animal viruses as targets will be benefited by consideration of the special features of these proteases.

7. Conclusions and future perspectives

Since the first virus structures were determined at almost atomic resolution, about 35 years ago, virus X-ray crystallography has continued to provide an ever deeper and wider understanding of the virus world. However, many of the questions still remaining in virus structure and function will require an even bolder use of X-ray studies in combination with new experimental approaches, such as high-resolution electron microscopy and single-molecule techniques. Technological developments associated with synchrotron radiation with the advent of *in situ* diffraction and X-ray free-electron lasers, that will allow the possibility of single-virus-particle analysis, will undoubtedly influence our research on structural virology in the near future.

References

Albertini, A. A., Mérigoux, C., Libersou, S., Madiona, K., Bressanelli, S., Roche, S., Lepault, J., Melki, R., Vachette, P. & Gaudin, Y. (2012). *PLoS Pathog.* **8**, e1002556.

Aljabali, A. A., Shukla, S., Lomonossoff, G. P., Steinmetz, N. F. & Evans, D. J. (2013). *Mol. Pharm.* **10**, 3–10.

Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). *EMBO J.* **21**, 3213–3224.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). *Science*, **300**, 1763–1767.

Babé, L. M. & Craik, C. S. (1997). *Cell*, **91**, 427–430.

Baquero, E., Albertini, A. A., Vachette, P., Lepault, J., Bressanelli, S. & Gaudin, Y. (2013). *Curr. Opin. Virol.* **3**, 143–150.

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

Bian, X., Klemm, R. W., Liu, T. Y., Zhang, M., Sun, S., Sui, X., Liu, X., Rapoport, T. A. & Hu, J. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 3976–3981.

Bostina, M., Levy, H., Filman, D. J. & Hogle, J. M. (2011). *J. Virol.* **85**, 776–783.

Bressanelli, S., Stiasny, K., Allison, S. L., Stura, E. A., Duquerroy, S., Lescar, J., Heinz, F. X. & Rey, F. A. (2004). *EMBO J.* **23**, 728.

Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1994). *Nature (London)*, **371**, 37–43.

Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. (2001). *Nature (London)*, **410**, 235–240.

Byrnes, L. J. & Sondermann, H. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 2216–2221.

Caillet-Saguy, C., Lim, S. P., Shi, P. Y., Lescar, J. & Bressanelli, S. (2014). *Antiviral Res.* **105**, 8–16.

Carrillo-Tripp, M., Shepherd, C. M., Borelli, I. A., Venkataraman, S., Lander, G., Natarajan, P., Johnson, J. E., Brooks III, C. L. & Reddy, V. S. (2009). *Nucleic Acids Res.* **37**, D436–D442.

Chinnaswamy, S., Murali, A., Li, P., Fujisaki, K. & Kao, C. C. (2010). *J. Virol.* **84**, 5923–5935.

Choi, H. K., Lee, S., Zhang, Y. P., McKinney, B. R., Wengler, G., Rossmann, M. G. & Kuhn, R. J. (1996). *J. Mol. Biol.* **262**, 151–167.

Choi, K. H. & Rossmann, M. G. (2009). *Curr. Opin. Struct. Biol.* **19**, 746–751.

Clum, S., Ebner, K. E. & Padmanabhan, R. (1997). *J. Biol. Chem.* **272**, 30715–30723.

Davidson, A. R., Cardarelli, L., Pell, L. G., Radford, D. R. & Maxwell, K. L. (2012). *Adv. Exp. Med. Biol.* **726**, 115–142.

Dessau, M. & Modis, Y. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 1696–1701.

Dijk, A. A. van, Makeyev, E. V. & Bamford, D. H. (2004). *J. Gen. Virol.* **85**, 1077–1093.

DuBois, R. M., Vaney, M. C., Tortorici, M. A., Kurdi, R. A., Barba-Spaeth, G., Krey, T. & Rey, F. A. (2013). *Nature (London)*, **493**, 552–556.

El Omari, K., Iourin, O., Harlos, K., Grimes, J. M. & Stuart, D. I. (2013). *Cell Rep.* **3**, 30–35.

Eltahla, A. A., Lim, K. L., Eden, J. S., Kelly, A. G., Mackenzie, J. M. & White, P. A. (2014). *Antimicrob. Agents Chemother.* **58**, 3115–3123.

Feldman, A. R., Lee, J., Delmas, B. & Paetzel, M. (2006). *J. Mol. Biol.* **358**, 1378–1389.

Ferrero, D., Buxaderas, M., Rodriguez, J. F. & Verdaguer, N. (2012). *Acta Cryst.* F**68**, 1263–1266.

Ferrer-Orta, C., Agudo, R., Domingo, E. & Verdaguer, N. (2009). *Curr. Opin. Struct. Biol.* **19**, 752–758.

Ferrer-Orta, C., Arias, A., Agudo, R., Pérez-Luque, R., Escarmís, C., Domingo, E. & Verdaguer, N. (2006). *EMBO J.* **25**, 880–888.

Ferrer-Orta, C., Arias, A., Escarmís, C. & Verdaguer, N. (2006). *Curr. Opin. Struct. Biol.* **16**, 27–34.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmis, C., Domingo, E. & Verdaguer, N. (2004). *J. Biol. Chem.* **279**, 47212–47221.

Ferrer-Orta, C., Arias, A., Pérez-Luque, R., Escarmís, C., Domingo, E. & Verdaguer, N. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 9463–9468.

Fokine, A. & Rossmann, M. G. (2013). *Bacteriophage*, **4**, e28281.

Garriga, D., Ferrer-Orta, C., Querol-Audí, J., Oliva, B. & Verdaguer, N. (2013). *J. Mol. Biol.* **425**, 2279–2287.

Garriga, D., Navarro, A., Querol-Audí, J., Abaitua, F., Rodríguez, J. F. & Verdaguer, N. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 20540–20545.

Garriga, D., Pickl-Herk, A., Luque, D., Wruss, J., Castón, J. R., Blaas, D. & Verdaguer, N. (2012). *PLoS Pathog.* **8**, e1002473.

Gayathri, P., Satheshkumar, P. S., Prasad, K., Nair, S., Savithri, H. S. & Murthy, M. R. (2006). *Virology*, **346**, 440–451.

Gentile, I., Coppola, N., Buonomo, A. R., Zappulo, E. & Borgia, G. (2014). *Expert Opin. Investig. Drugs*, pp. 1–13.

Gong, P., Kortus, M. G., Nix, J. C., Davis, R. E. & Peersen, O. B. (2013). *PLoS One*, **8**, e60272.

Gong, P. & Peersen, O. B. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 22505–22510.

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M. & Donchenko, A. P. (1988). *FEBS Lett.* **236**, 287–290.

Gradi, A., Svitkin, Y. V., Imataka, H. & Sonenberg, N. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 11089–11094.

Hahn, C. S. & Strauss, J. H. (1990). *J. Virol.* **64**, 3069–3073.

Harris, K. S., Reddigari, S. R., Nicklin, M. J., Hämmerle, T. & Wimmer, E. (1992). *J. Virol.* **66**, 7481–7489.

Harrus, D., Ahmed-El-Sayed, N., Simister, P. C., Miller, S., Triconnet, M., Hagedorn, C. H., Mahias, K., Rey, F. A., Astier-Gin, T. & Bressanelli, S. (2010). *J. Biol. Chem.* **285**, 32906–32918.

Hashiguchi, T., Ose, T., Kubota, M., Maita, N., Kamishikiryo, J., Maenaka, K. & Yanagi, Y. (2011). *Nature Struct. Mol. Biol.* **18**, 135–141.

Högbom, M., Jäger, K., Robel, I., Unge, T. & Rohayem, J. (2009). *J. Gen. Virol.* **90**, 281–291.

Igonet, S. & Rey, F. A. (2012). *Cell*, **151**, 1634.

Kapust, R. B. & Waugh, D. S. (2000). *Protein Expr. Purif.* **19**, 312–318.

Kay, J. & Dunn, B. M. (1990). *Biochim. Biophys. Acta*, **1048**, 1–18.

Kostyuchenko, V. A., Chipman, P. R., Leiman, P. G., Arisaka, F., Mesyanzhinov, V. V. & Rossmann, M. G. (2005). *Nature Struct. Mol. Biol.* **12**, 810–813.

Koudelka, K. J., Destito, G., Plummer, E. M., Trauger, S. A., Siuzdak, G. & Manchester, M. (2009). *PLoS Pathog.* **5**, e1000417.

Lang, D. M., Zemla, A. T. & Zhou, C. L. (2013). *Nucleic Acids Res.* **41**, 1464–1482.

Leiman, P. G., Arisaka, F., van Raaij, M. J., Kostyuchenko, V. A., Aksyuk, A. A., Kanamaru, S. & Rossmann, M. G. (2010). *Virol. J.* **7**, 355.

Leiman, P. G., Chipman, P. R., Kostyuchenko, V. A., Mesyanzhinov, V. V. & Rossmann, M. G. (2004). *Cell*, **118**, 419–429.

Leiman, P. G. & Shneider, M. M. (2012). *Adv. Exp. Med. Biol.* **726**, 93–114.

Lescar, J. & Canard, B. (2009). *Curr. Opin. Struct. Biol.* **19**, 759–767.

Lescar, J., Roussel, A., Wien, M. W., Navaza, J., Fuller, S. D., Wengler, G., Wengler, G. & Rey, F. A. (2001). *Cell*, **105**, 137–148.

Li, F., Li, W., Farzan, M. & Harrison, S. C. (2005). *Science*, **309**, 1864–1868.

Li, Y., Wang, J., Kanai, R. & Modis, Y. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 6805–6810.

Lin, Y. P., Xiong, X., Wharton, S. A., Martin, S. R., Coombs, P. J., Vachieri, S. G., Christodoulou, E., Walker, P. A., Liu, J., Skehel, J. J., Gamblin, S. J., Hay, A. J., Daniels, R. S. & McCauley, J. W. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 21474–21479.

Lloyd, R. E., Grubman, M. J. & Ehrenfeld, E. (1988). *J. Virol.* **62**, 4216–4223.

Lorenz, I. C., Marcotrigiano, J., Dentzer, T. G. & Rice, C. M. (2006). *Nature (London)*, **442**, 831–835.

Luo, G., Hamatake, R. K., Mathis, D. M., Racela, J., Rigat, K. L., Lemm, J. & Colonno, R. J. (2000). *J. Virol.* **74**, 851–863.

Lyle, J. M., Bullitt, E., Bienz, K. & Kirkegaard, K. (2002). *Science*, **296**, 2218–2222.

Malet, H., Massé, N., Selisko, B., Romette, J. L., Alvarez, K., Guillemot, J. C., Tolou, H., Yap, T. L., Vasudevan, S., Lescar, J. & Canard, B. (2008). *Antiviral Res.* **80**, 23–35.

Marathe, R., Anandalakshmi, R., Smith, T. H., Pruss, G. J. & Vance, V. B. (2000). *Plant Mol. Biol.* **43**, 295–306.

Mathur, C. & Savithri, H. S. (2012). *Biochem. Biophys. Res. Commun.* **427**, 113–118.

Matthews, D. A. *et al.* (1994). *Cell*, **77**, 761–771.

McKenna, R., Xia, D., Willingmann, P., Ilag, L. L., Krishnaswamy, S., Rossmann, M. G., Olson, N. H., Baker, T. S. & Incardona, N. L. (1992). *Nature (London)*, **355**, 137–143.

Modis, Y. (2014). *Curr. Opin. Virol.* **5**, 34–41.

Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. (1997). *J. Mol. Biol.* **273**, 1032–1047.

Mosley, R. T., Edwards, T. E., Murakami, E., Lam, A. M., Grice, R. L., Du, J., Sofia, M. J., Furman, P. A. & Otto, M. J. (2012). *J. Virol.* **86**, 6503–6511.

Moss, T. J., Daga, A. & McNew, J. A. (2011). *Trends Cell Biol.* **21**, 416–423.

Mukhopadhyay, S., Kim, B. S., Chipman, P. R., Rossmann, M. G. & Kuhn, R. J. (2003). *Science*, **302**, 248.

Ng, K. K., Arnold, J. J. & Cameron, C. E. (2008). *Curr. Top. Microbiol. Immunol.* **320**, 137–156.

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

Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. & Davidson, A. R. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 4160–4165.

Pérez-Vargas, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Jamin, M., Raveh-Barak, H., Podbilewicz, B. & Rey, F. A. (2014). *Cell*, **157**, 407–419.

Pickl-Herk, A., Luque, D., Vives-Adrián, L., Querol-Audí, J., Garriga, D., Trus, B. L., Verdaguer, N., Blaas, D. & Castón, J. R. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 20063–20068.

Plisson, C., White, H. E., Auzat, I., Zafarani, A., São-José, C., Lhuillier, S., Tavares, P. & Orlova, E. V. (2007). *EMBO J.* **26**, 3720–3728.

Powdrill, M. H., Bernatchez, J. A. & Götte, M. (2010). *Viruses*, **10**, 2169–2195.

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

Reguera, J., Santiago, C., Mudgal, G., Ordoño, D., Enjuanes, L. & Casasnovas, J. M. (2012). *PLoS Pathog.* **8**, e1002859.

Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. (1995). *Nature (London)*, **375**, 291–298.

Roche, S., Bressanelli, S., Rey, F. A. & Gaudin, Y. (2006). *Science*, **313**, 187–191.

Roche, S., Rey, F. A., Gaudin, Y. & Bressanelli, S. (2007). *Science*, **315**, 843–848.

Russel, M. & Model, P. (2006). In *The Bacteriophages*, edited by R. Calendar, pp. 146–160. Oxford University Press.

Santiago, C., Celma, M. L., Stehle, T. & Casasnovas, J. M. (2010). *Nature Struct. Mol. Biol.* **17**, 124–129.

Satheshkumar, P. S., Gayathri, P., Prasad, K. & Savithri, H. S. (2005). *J. Biol. Chem.* **280**, 30291–30300.

Sciara, G., Bebeacua, C., Bron, P., Tremblay, D., Ortiz-Lombardia, M., Lichière, J., van Heel, M., Campanacci, V., Moineau, S. & Cambillau, C. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 6852–6857.

Sholders, A. J. & Peersen, O. B. (2014). *J. Mol. Biol.* **426**, 1407–1419.

Shukla, S., Ablack, A. L., Wen, A. M., Lee, K. L., Lewis, J. D. & Steinmetz, N. F. (2013). *Mol. Pharm.* **10**, 33–42.

Spagnolo, J. F., Rossignol, E., Bullitt, E. & Kirkegaard, K. (2010). *RNA*, **16**, 382–393.

Spall, V. E., Shanks, M. & Lomonossoff, G. (1997). *Semin. Virol.* **8**, 15–23.

Steinmetz, N. F. (2010). *Nanomedicine NBM*, pp. 634–641.

Steitz, T. A. (1998). *Nature (London)*, **391**, 231–232.

Südhof, T. C. & Rothman, J. E. (2009). *Science*, **323**, 474–477.

Sun, L., Young, L. N., Zhang, X., Boudko, S. P., Fokine, A., Zbornik, E., Roznowski, A. P., Molineux, I. J., Rossmann, M. G. & Fane, B. A. (2014). *Nature (London)*, **505**, 431–435.

Veesler, D., Spinelli, S., Mahony, J., Lichière, J., Blangy, S., Bricogne, G., Legrand, P., Ortiz-Lombardia, M., Campanacci, V., van Sinderen, D. & Cambillau, C. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 8954–8958.

Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K. C., Chen, Y. H., Zhang, L. & Wang, X. (2013). *Cell Res.* **23**, 986–993.

Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981). *Nature (London)*, **289**, 366–373.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989). *Science*, **245**, 616–621.

Xiong, X., Coombs, P. J., R Martin, S., Liu, J., Xiao, H., Mccauley, J. W., Locher, K., Walker, P. A., Collins, P. J., Kawaoka, Y., Skehel, J. J. & Gamblin, S. J. (2013). *Nature (London)*, **497**, 392–395.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13190–13195.

Yang, X., Smidansky, E. D., Maksimchuk, K. R., Lum, D., Welch, J.L., Arnold, J. J., Cameron, C. E. & Boehr, D. D. (2012). *Structure*, **20**, 1519–1527.

Yin, H. S., Paterson, R. G., Wen, X., Lamb, R. A. & Jardetzky, T. S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 9288–9293.

Yin, H. S., Wen, X., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. (2006). *Nature (London)*, **439**, 38–44.

Yusof, R., Clum, S., Wetzel, M., Murthy, H. M. & Padmanabhan, R. (2000). *J. Biol. Chem.* **275**, 9963–9969.

Zamyatkin, D. F., Parra, F., Alonso, J. M., Harki, D. A., Peterson, B. R., Grochulski, P. & Ng, K. K. (2008). *J. Biol. Chem.* **283**, 7705–7712.

Zamyatkin, D. F., Parra, F., Machín, A., Grochulski, P. & Ng, K. K. (2009). *J. Mol. Biol.* **390**, 10–16.

Zhang, X., Ge, P., Yu, X., Brannan, J. M., Bi, G., Zhang, Q., Schein, S. & Zhou, Z. H. (2013). *Nature Struct. Mol. Biol.* **20**, 105–110.

Zhang, R., Hryc, C. F., Cong, Y., Liu, X., Jakana, J., Gorchakov, R., Baker, M. L., Weaver, S. C. & Chiu, W. (2011). *EMBO J.* **30**, 3854–3863.

Zhang, X., Lu, G., Qi, J., Li, Y., He, Y., Xu, X., Shi, J., Zhang, C. W., Yan, J. & Gao, G. F. (2013). *Nature Struct. Mol. Biol.* **20**, 67–72.