Virus Research xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Atlas of coronavirus replicase structure

Benjamin W. Neuman[a,*], Peter Chamberlain[a], Fern Bowden[a], Jeremiah Joseph[b]

[a] School of Biological Sciences, University of Reading, Reading, UK
[b] Dart Neuroscience, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form 3 December 2013
Accepted 5 December 2013
Available online xxx

Keywords:
SARS coronavirus
Nonstructural protein
Protein structure
NMR
X-ray crystallography
Cryoelectron microscopy
Small angle X-ray scattering

ABSTRACT

The international response to SARS-CoV has produced an outstanding number of protein structures in a very short time. This review summarizes the findings of functional and structural studies including those derived from cryoelectron microscopy, small angle X-ray scattering, NMR spectroscopy, and X-ray crystallography, and incorporates bioinformatics predictions where no structural data is available. Structures that shed light on the function and biological roles of the proteins in viral replication and pathogenesis are highlighted. The high percentage of novel protein folds identified among SARS-CoV proteins is discussed.

1. Introduction

In the wake of the SARS crisis, a wave of structural proteomics swept the coronavirus research community. The focus of this effort was to understand the interplay between structure and function in what had been, until that time, a somewhat neglected branch of the positive-stranded viruses. The unusual aspect of the SARS proteomics at the time was its evenhandedness – rather than focusing exclusively on proteins with well-defined roles in pathogenesis, competing international teams attempted to solve structures and assign functions across the entire viral proteome.

This effort brought fresh attention to several little-known replicase cofactors, such as the European group's structure of the obscure but important RNA binding protein nsp9 (Egloff et al., 2004), the Chinese group's barrel-shaped 16-protein structure of nsp7+8 primase complex (Zhai et al., 2005) and the American group's long crawl through the giant multi-domain, multi-enzymatic protein nsp3 which found the first of three SARS-CoV macrodomain folds (Saikatendu et al., 2005).

Shortly after the outbreak, the sequence of the genome was completed and the 3-D structure of M^{pro}, the main protease essential for viral replication, was deposited in the Protein Data Bank (PDB). By 2007, 100 entries in the PDB were on 14 of the 28 SARS CoV proteins, and at present count there are 99 structures of coronavirus M^{pro} available in the PDB alone, providing an unprecedented database for investigators working on this and related viruses. This review summarizes the findings of functional and structural studies including those derived from cryoelectron microscopy, small angle X-ray scattering, NMR spectroscopy, and X-ray crystallography in an attempt to understand the function and biological roles of the proteins in viral replication and pathogenesis.

2. A note on functional organization

The new wealth of structural and functional information revealed that the coronavirus replicase, which is but one biologically successful example of the conserved nidovirus replicative machinery (Lauber et al., 2013), is not a patchwork amalgam of evolutionary jetsam, but an organized piece of biological machinery where proteins are generally organized into units with related functions (Fig. 1). The first two parts of the replicase, nsp1 and nsp2 are somewhat enigmatic, but appear to work by interfering with host defenses rather than by directly supporting virus replication. Subunits nsp3–6 contain all the viral factors that are necessary to form viral replicative organelles (Angelini et al., 2013), as well as two proteinases that are responsible for processing all of the viral replicase proteins (Ziebuhr et al., 2000). The small subunits nsp7–11 comprise the viral primer-making activities and provide other essential support for replication (Donaldson et al., 2007b; Imbert et al., 2006; Miknis et al., 2009). The final part of the replicase from nsp12–16 contains the remaining RNA-modifying enzymes needed for replication, RNA capping and proofreading.

* Corresponding author. Tel.: +44 0 118 378 8893.
E-mail address: b.w.neuman@reading.ac.uk (B.W. Neuman).

0168-1702/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.virusres.2013.12.004

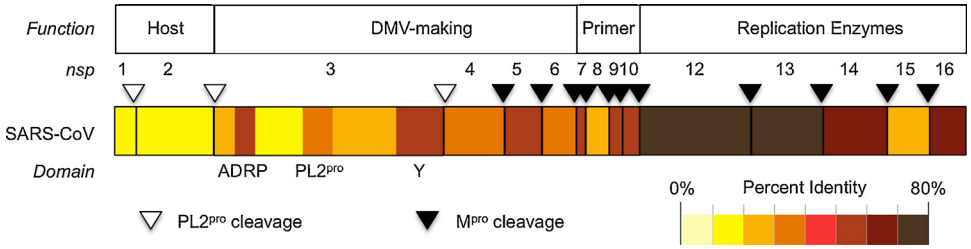

Fig. 1. Conservation of the SARS-CoV replicase. Replicase subunits, or domains for nsp3, were color-coded according to percent identity between homologous proteins of SARS-CoV and MERS-CoV. Alignments and identity calculations were performed using Clustal Omega (Sievers et al., 2011).

The organization of replicase has a sort of chronological logic to it. Nsp1–2 help to colonize the host, followed by Nsp3–6 which lay a foundation to organize and protect the replicative machinery. This is followed by the primer-making activities of nsp7–11 which also interact with downstream capping and RNA synthesis factors (Bouvet et al., 2010). Finally, in the proper framework, the RNA-synthesizing enzymes from the C-terminus of the replicase are able to function. While this may be an appealing way to think of the replicase, the reality is probably much more complex. The replicase proteins are all processed from large polyproteins, and therefore are produced at the same time. Because of this, the order in which different proteins are active during the viral replication cycle remains poorly understood.

The organization of the replicase also roughly follows a gradient of primary sequence conservation. Levels of sequence conservation among the different coronaviruses are highest at the 3′ end of the replicase gene, and the sequences are very divergent at the 5′ end, especially in nsp1–3, which are products of nsp3 PL^{pro} cleavage. The DMV-making proteins and the primase group of proteins show intermediate levels of conservation with the exception of the well-conserved nsp5M^{pro}. Fig. 1 illustrates amino acid conservation across the replicase using the comparison between SARS-CoV and MERS-CoV as an example.

3. The Atlas

3.1. Nsp1

Q2 See Box 1.

3.1.1. Structure

Nsp1 is the N-terminal cleavage product of the replicase polyprotein and is produced by the action of PL^{pro}. Nsp1 is not found in the gammacoronavirus or deltacoronavirus lineages, which encode a distant homolog of SARS-CoV nsp2 at the N-terminus of the replicase po lyprotein. This has led to a suggestion that nsp1 is useful as a group-specific marker (Snijder et al., 2003). SARS-CoV nsp1 is 179 residues long.

In the alphacoronaviruses, nsp1 (also known as p9) is a protein of about 110 residues, with 20–50% sequence identity among all alphacoronaviruses. The betacoronaviruses of subgroup A, such as murine hepatitis virus (MHV) and human coronavirus OC43, encode an nsp1 protein of about 245 residues, also known as p28 (Brockway and Denison, 2005). The nsp1 of SARS-CoV and its bat equivalents, which have been classified as the only members to date of the betacoronavirus subgroup B (Gorbalenya et al., 2006; Gorbalenya et al., 2004; Snijder et al., 2003), have 179 residues. Nsp1 sequences are divergent between subgroups of betacoronavirus, and no sequence similarity between SARS-CoV nsp1 and betacoronavirus subgroup A nsp1 proteins could be identified using standard searching tools such as PSI-BLAST.

Almeida et al. (2006, 2007) determined the NMR structure of the nsp1 segment from residue 13 to 128 and also showed that the polypeptide segments of residues 1–12 and 129–179 are flexibly disordered (PDB ID 2GDT; 2HSX) (Almeida et al., 2007). Residues 13–128 of nsp1 represents a novel α/β-fold formed by a mixed parallel/antiparallel 6-stranded β-barrel, an α-helix covering one opening of the barrel, and a 3_{10}-helix alongside the barrel (Fig. 2). NMR data indicate that full-length nsp1 has the same globular fold as the truncated nsp1, but with additional flexibly disordered regions that correspond to the N-terminal region (residues 1–12) and the long C-terminal tail (residues 129–179).

The C-terminal portion of SARS-CoV nsp1 is flexibly disordered. Interestingly, it has been determined that the C-terminal half of MHV nsp1 (Lys124–Leu241) is dispensable for viral replication in culture but is important for efficient proteolytic cleavage at the nsp1–2 peptide linkage by the papain-like protease and optimal viral replication (Brockway and Denison, 2005). Likewise, the long disordered terminus of SARS-CoV nsp1 are probably important for the efficient proteolytic processing of this protein from the nascent viral polyprotein chain.

The nsp1 of transmissible gastroenteritis virus (TGEV) was recently solved, and was found to contain a similar fold to SARS-CoV nsp1 (Jansson, 2013). This was surprising as there is no detectable homology between alphacoronavirus nsp1 proteins and betacoronavirus nsp1 proteins. However, the relationship of the structures suggests that coronavirus nsp1 proteins share a common evolutionary origin.

3.1.2. Function

In several coronaviruses, nsp1 suppresses host gene expression (Huang et al., 2011; Kamitani et al., 2006; Narayanan et al., 2008;

Box 1: Key nsp1 and nsp2 structures

Virus	Protein	Method	Accession	Reference
SARS-CoV	nsp1	NMR	2HSX	Almeida et al. (2007)
TGEV	nsp1	X-ray (1.49 Å)	3ZBD	Jansson (2013)
IBV	nsp2	X-Ray (2.5 Å)	3LD1	Yu et al. (2012)

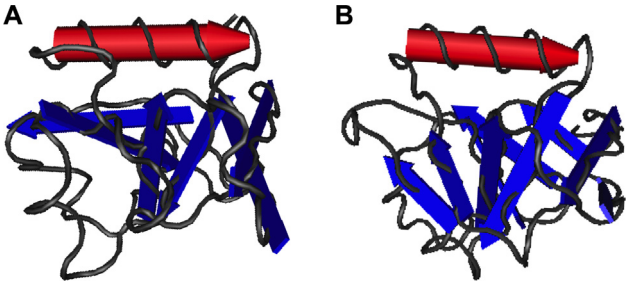

Fig. 2. Comparison of nsp1 structure in alpha- and betacoronavirus lineages. The SARS-CoV nsp1 structure comes from PDB entry 2HSX, and the TGEV nsp1 structure comes from PDB entry 3ZBD.