Protein-protein interactions leading to recruitment of the host DNA sliding clamp by the hyperthermophilic Sulfolobus islandicus rod-shaped virus 2

Andrew F. Gardner, Stephen D. Bell, Malcolm F. White, David Prangishvili, Mart Krupovic

JVI Accepts, published online ahead of print on 2 April 2014
J. Virol. doi:10.1128/JVI.00636-14

Protein-protein interactions leading to the recruitment of the host DNA sliding clamp by the hyperthermophilic archaeal virus SIRV2

Andrew F. Gardner[a], Stephen D. Bell[b], Malcolm F. White[c], David Prangishvili[d#], Mart Krupovic[d#]

[a] New England Biolabs, Ipswich, MA 01938, USA
[b] Molecular and Cellular Biochemistry Department, Biology Department, Indiana University, Bloomington, Indiana 47405, USA
[c] Biomedical Sciences Research Complex, University of St Andrews, Fife KY16 9ST, UK
[d] Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Département de Microbiologie, 75015 Paris, France

Running title: Recruitment of PCNA by SIRV2

Corresponding authors:
Mart Krupovic
E-mail: krupovic@pasteur.fr

David Prangishvili
E-mail: david.prangishvili@pasteur.fr

Word count for the abstract: 71
Word count for the text: 1,513

Abstract

Viruses infecting hyperthermophilic archaea typically do not encode DNA polymerases, raising questions regarding their genome replication. Here, using a yeast two-hybrid approach we have assessed interactions between proteins of *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) and the host-encoded proliferating cell nuclear antigen (PCNA), a key DNA replication protein in archaea. Five SIRV2 proteins were found to interact with PCNA, providing insights into the recruitment of host replisome for viral DNA replication.

Viruses infecting Archaea, the third domain of life, constitute one of the most enigmatic sections of the virosphere. Archaeal viruses, especially those infecting hyperthermophilic hosts thriving at temperatures above 70°C, are extremely diverse both morphologically and genetically (27, 28). The vast majority of them are unique to Archaea and do not resemble viruses infecting Bacteria or Eukaryotes. Furthermore, during the past few years it became apparent that the ways hyperthermophilic archaeal viruses interact with their hosts are also often unprecedented in the viral world (2, 3). Indeed, proteins encoded by these viruses typically do not share similarity with proteins in the sequence databases (28), often display new structural folds (18), and play unexpected roles in the viral life cycles (31). An interesting feature of hyperthermophilic archaeal viruses is the general absence of recognizable DNA polymerase genes in their genomes; among the 41 virus isolates for which genome sequences are available, only one encodes a DNA polymerase (26), raising a question as of how genome replication in these viruses is achieved. To answer this puzzling question, in the present study we have investigated how *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) recruits the cellular machinery for the replication of its genome.

SIRV2 is one of the most extensively studied archaeal viruses (29) and is the type species of the family *Rudiviridae* within the order *Ligamenvirales* (30). The rod-shaped SIRV2 virions are non-enveloped and are formed from linear double-stranded DNA genomes coated with one major and three minor structural proteins. SIRV2 infects acidophilic hyperthermophiles of the archaeal order *Sulfolobales*, namely *S. islandicus* (12) and *S. solfataricus* (23). The infection cycle starts with rapid virus adsorption to the pili-like structures on the host cell surface, which eventually leads to genome uncoating and internalization (33). Transcription of the viral genes commences within the first few minutes of infection (32) and is followed by the efficient replication of the viral DNA (2). SIRV2 virions, assembled in the cytoplasm, are released from the host ~8–10 hours post infection through large virus-encoded pyramidal structures (2). One of the least understood aspects of SIRV2 infection cycle is genome replication in the absence of virus-encoded DNA polymerase. The only proteins, which might be involved in the

replication of SIRV2 genome, are the Holliday junction resolvase P121 (9) and the endonuclease P119c (22) related to enzymes involved in the initiation of the rolling-circle replication of various plasmids and single-stranded DNA viruses (11, 15). However, the exact steps of SIRV2 genome replication in which the two enzymes participate as well as the involvement of cellular players in this process remain unclear.

Viruses that do not encode their own DNA polymerases rely of the replication machinery of the host. However, viral proteins often play an important role in directing the replisome to the viral replication origins. Indeed, a considerable number of euryarchaeal viruses encode proteins involved in the initiation of DNA replication, including replicative MCM helicases, Cdc6/Orc1 or PCNA (proliferating cell nuclear antigen) homologues (10, 14, 16, 17, 24, 34, 36). Since rudiviruses encode neither a DNA polymerse nor any of the above mentioned proteins involved in the initiation of DNA replication, we hypothesized that their genome replication should rely on the physical recruitment of the host replisome. To verify this possibility, we set out to investigate the interactions between SIRV2 proteins and the heterotrimeric *S. solfataricus* sliding clamp (SsoPCNA1–3) (7, 38). The latter was selected as a likely target because it is a key protein of DNA replication and repair in archaea and eukaryotes, allowing non-sequence-specific enzymes, such as replicative DNA polymerase PolB1, DNA ligase Lig1 and flap endonuclease FEN1, to associate with their DNA substrates (1, 20, 25). To explore which of the SIRV2 proteins might be involved in the recruitment of PCNA and, by extension, of the entire replisome, we employed yeast two-hybrid (Y2H) analysis. SsoPCNA subunits 1 (NP_341936), 2 (NP_342519), and 3 (NP_341944) were cloned into the 'bait' vector (pGBKT7) encoding GAL4 DNA-binding domain, while the viral 'prey' library was created by cloning the PCR-amplified SIRV2 genes into the NdeI/XmaI site of the pGADT7 vector encoding the GAL4 activation domain. All clones were verified by DNA sequencing. *Saccharomyces cerevisiae* AH109 (Clontech) was sequentially transformed with prey and bait plasmids and selected on nutritional media lacking either leucine or tryptophan, respectively. Controls with empty bait or prey vectors were performed to ensure that the fusion protein could not induce expression of selection gene in the absence of a protein partner. In the case of interaction between the bait and prey, the fusion protein is expected to activate the HIS3 expression and complement histidine auxotrophy, allowing growth on yeast minimal media lacking histidine.

Our Y2H screen has revealed five SIRV2-encoded proteins that interacted with the SsoPCNA: proteins P105a (NP_666544) and P84c (NP_666565) were found to interact with SsoPCNA1, whereas proteins P83a/b (NP_666535/NP_666588), P84c, P119a (NP_666536) and P121 (NP_666569) with SsoPCNA3 (Fig. 1). Notably, no interaction between viral proteins and SsoPCNA2 could be detected. Two of the PCNA-interacting SIRV2 proteins (P84c and P121) are conserved in all rudiviruses, two (P83a/b and

P105a) are encoded in a smaller subset of more closely related viruses, while P119a is restricted to SIRV2 (Table 1).

PCNA is known as a "molecular tool-belt" (38), which interacts with multiple proteins involved in DNA replication and repair (1, 20, 25). These interactions are often mediated via the PCNA-interacting protein (PIP) box, with the consensus sequence of Q*xxhxx*@@, where '*h*' represents hydrophobic amino acid residues, '@' corresponds to bulky aromatic residues, and '*x*' is any residue (37). Sequence analysis of the viral proteins, which were found to interact with SsoPCNA1 and SsoPCNA3, revealed the presence of potential PIP-boxes in all proteins except for P84c (Fig. 2). Interestingly, P119a was found to contain three PIP-boxes (Fig. 2D). Notably, whereas the presence of PIP boxes is suggestive of interaction with PCNA, the absence of the identifiable motif does not necessarily signify the reverse (i.e., lack of interaction), because (i) the exact sequence of PIP boxes is known to vary (37), and (ii) interactions might be mediated by motifs other than PIP (39). As has been previously observed for PCNA-interacting proteins from other viruses (13), the PIP-boxes of SIRV2 displayed variable correspondence to the consensus sequence, which might be important for orchestrating the sequential binding of different protein partners to PCNA.

To more critically scrutinize the predicted PIP boxes, we have investigated their exact location within the four PCNA-interacting viral proteins. For this purpose we have constructed structural homology models using Modeller v9.11 (19) for three of the SIRV2 proteins for which structural counterparts could be identified (Fig. 2A-C). This analysis suggested that PIP box of protein P121 is unlikely to be functional, since it is located within the core of the protein fold (Fig. 2C), and thus a different protein region must be responsible for the observed PCNA binding. Similarly, two of the three PIP boxes initially predicted in P119a are found within the secondary structure elements, which are likely to be integral for the tertiary protein structure. The PIP box of P83a/b is located in the loop region between the two hairpin-forming β-strands (Fig. 2A). By contrast, PIP box of P105a and PIP1 box of P119a are located at the extremities of the proteins, lending credence to their functionality.

Out of five PCNA-interacting SIRV2 proteins only one has been biochemically characterized; protein P121 is a Holliday junction resolvase (Hjr), which is implicated in the resolution of viral genome concatamers, a key step in SIRV2 genome replication (9). Interestingly, SsoPCNA has been previously shown to bind the Hjr of *S. solfataricus* and stimulate its enzymatic activity (8), suggesting that a similar effect might be exerted during the interaction with the viral Hjr. Another interesting protein, which has been identified as an SsoPCNA partner in our Y2H analysis, is P83a/b (genes *ORF83a* and *ORF83b* are located in inverted terminal repeats and encode identical proteins). RNA-seq analysis has shown that *ORF83a/b* transcripts are overwhelmingly dominant

starting with the first minutes of infection and remain abundant throughout the infection (32), suggesting an important role for P83a/b during the replication cycle of SIRV2. In the course of a structural genomics initiative the X-ray structure of P83a/b homolog from a closely related virus SIRV1 has been solved (21). The protein was found to form a hexameric ring and contain the helix-turn-helix DNA-binding domain, suggesting that it might be involved in DNA metabolism (18). Notably, previous Y2H screen for interactions among SIRV2 proteins revealed that P83a/b interacts with the viral Hjr, P121 (32), suggesting that P83a/b, P121 and SsoPCNA form a tripartite complex (Fig. 1B). Whereas *ORF83a/b* transcripts are abundant from the beginning of infection, those of *ORF121* appear later in infection, suggesting that the intricate interplay between viral proteins and the cellular PCNA might be regulated by both hierarchical strengths of PIP-box-mediated protein-protein interactions and transcriptional control of viral gene expression. Future studies will focus on revealing the molecular and structural details of the interactions between the host PCNA subunits and the five SIRV2 proteins identified here, which will clarify how SIRV2 hijacks the replisome of its host.

References

1. **Beattie, T. R., and S. D. Bell.** 2011. Molecular machines in archaeal DNA replication. Curr Opin Chem Biol **15:**614-619.
2. **Bize, A., E. A. Karlsson, K. Ekefjard, T. E. Quax, M. Pina, M. C. Prevost, P. Forterre, O. Tenaillon, R. Bernander, and D. Prangishvili.** 2009. A unique virus release mechanism in the Archaea. Proc Natl Acad Sci U S A **106:**11306-11311.
3. **Brumfield, S. K., A. C. Ortmann, V. Ruigrok, P. Suci, T. Douglas, and M. J. Young.** 2009. Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus *Sulfolobus* turreted icosahedral virus. J Virol **83:**5964-5970.
4. **Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay, and G. J. Barton.** 1998. JPred: a consensus secondary structure prediction server. Bioinformatics **14:**892-893.
5. **Dionne, I., and S. D. Bell.** 2005. Characterization of an archaeal family 4 uracil DNA glycosylase and its interaction with PCNA and chromatin proteins. Biochem J **387:**859-863.
6. **Dionne, I., N. J. Brown, R. Woodgate, and S. D. Bell.** 2008. On the mechanism of loading the PCNA sliding clamp by RFC. Mol Microbiol **68:**216-222.
7. **Dionne, I., R. K. Nookala, S. P. Jackson, A. J. Doherty, and S. D. Bell.** 2003. A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Mol Cell **11:**275-282.
8. **Dorazi, R., J. L. Parker, and M. F. White.** 2006. PCNA activates the Holliday junction endonuclease Hjc. Journal of molecular biology **364:**243-247.
9. **Gardner, A. F., C. Guan, and W. E. Jack.** 2011. Biochemical characterization of a structure-specific resolving enzyme from *Sulfolobus islandicus* rod-shaped virus 2. PLoS One **6:**e23668.
10. **Gorlas, A., E. V. Koonin, N. Bienvenu, D. Prieur, and C. Geslin.** 2012. TPV1, the first virus isolated from the hyperthermophilic genus Thermococcus. Environ Microbiol **14:**503-516.
11. **Ilyina, T. V., and E. V. Koonin.** 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic acids research **20:**3279-3285.

12. **Jaubert, C., C. Danioux, J. Oberto, D. Cortez, A. Bize, M. Krupovic, Q. She, P. Forterre, D. Prangishvili, and G. Sezonov.** 2013. Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. Open Biol **3:**130010.
13. **Kazlauskas, D., and C. Venclovas.** 2011. Computational analysis of DNA replicases in double-stranded DNA viruses: relationship with the genome size. Nucleic acids research **39:**8291-8305.
14. **Klein, R., U. Baranyi, N. Rossler, B. Greineder, H. Scholz, and A. Witte.** 2002. *Natrialba magadii* virus phiCh1: first complete nucleotide sequence and functional organization of a virus infecting a haloalkaliphilic archaeon. Mol Microbiol **45:**851-863.
15. **Krupovic, M.** 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Current opinion in virology **3:**578-586.
16. **Krupovic, M., P. Forterre, and D. H. Bamford.** 2010. Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. Journal of molecular biology **397:**144-160.
17. **Krupovic, M., S. Gribaldo, D. H. Bamford, and P. Forterre.** 2010. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. Mol Biol Evol **27:**2716-2732.
18. **Krupovic, M., M. F. White, P. Forterre, and D. Prangishvili.** 2012. Postcards from the edge: structural genomics of archaeal viruses. Adv Virus Res **82:**33-62.
19. **Marti-Renom, M. A., A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali.** 2000. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct **29:**291-325.
20. **Moldovan, G. L., B. Pfander, and S. Jentsch.** 2007. PCNA, the maestro of the replication fork. Cell **129:**665-679.
21. **Oke, M., L. G. Carter, K. A. Johnson, H. Liu, S. A. McMahon, X. Yan, M. Kerou, N. D. Weikart, N. Kadi, M. A. Sheikh, S. Schmelz, M. Dorward, M. Zawadzki, C. Cozens, H. Falconer, H. Powers, I. M. Overton, C. A. van Niekerk, X. Peng, P. Patel, R. A. Garrett, D. Prangishvili, C. H. Botting, P. J. Coote, D. T. Dryden, G. J. Barton, U. Schwarz-Linek, G. L. Challis, G. L. Taylor, M. F. White, and J. H. Naismith.** 2010. The Scottish Structural Proteomics Facility: targets, methods and outputs. J Struct Funct Genomics **11:**167-180.
22. **Oke, M., M. Kerou, H. Liu, X. Peng, R. A. Garrett, D. Prangishvili, J. H. Naismith, and M. F. White.** 2011. A dimeric Rep protein initiates replication of a linear archaeal virus genome: implications for the Rep mechanism and viral replication. J Virol **85:**925-931.
23. **Okutan, E., L. Deng, S. Mirlashari, K. Uldahl, M. Halim, C. Liu, R. A. Garrett, Q. She, and X. Peng.** 2013. Novel insights into gene regulation of the rudivirus SIRV2 infecting *Sulfolobus* cells. RNA Biol **10:**875-885.
24. **Pagaling, E., R. D. Haigh, W. D. Grant, D. A. Cowan, B. E. Jones, Y. Ma, A. Ventosa, and S. Heaphy.** 2007. Sequence analysis of an Archaeal virus isolated from a hypersaline lake in Inner Mongolia, China. BMC Genomics **8:**410.
25. **Pan, M., L. M. Kelman, and Z. Kelman.** 2011. The archaeal PCNA proteins. Biochem Soc Trans **39:**20-24.
26. **Peng, X., T. Basta, M. Haring, R. A. Garrett, and D. Prangishvili.** 2007. Genome of the *Acidianus* bottle-shaped virus and insights into the replication and packaging mechanisms. Virology **364:**237-243.
27. **Prangishvili, D.** 2013. The wonderful world of archaeal viruses. Ann Rev Microbiol **67:**565-585.
28. **Prangishvili, D., R. A. Garrett, and E. V. Koonin.** 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res **117:**52-67.
29. **Prangishvili, D., E. V. Koonin, and M. Krupovic.** 2013. Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in Archaea. Biochem Soc Trans **41:**443-450.

30. **Prangishvili, D., and M. Krupovic.** 2012. A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales". Arch Virol **157:**791-795.
31. **Quax, T. E., M. Krupovic, S. Lucas, P. Forterre, and D. Prangishvili.** 2010. The *Sulfolobus* rod-shaped virus 2 encodes a prominent structural component of the unique virion release system in Archaea. Virology **404:**1-4.
32. **Quax, T. E., M. Voet, O. Sismeiro, M. A. Dillies, B. Jagla, J. Y. Coppee, G. Sezonov, P. Forterre, J. van der Oost, R. Lavigne, and D. Prangishvili.** 2013. Massive activation of archaeal defense genes during viral infection. J Virol **87:**8419-8428.
33. **Quemin, E. R., S. Lucas, B. Daum, T. E. Quax, W. Kuhlbrandt, P. Forterre, S. V. Albers, D. Prangishvili, and M. Krupovic.** 2013. First insights into the entry process of hyperthermophilic archaeal viruses. J Virol **87:**13379-13385.
34. **Raymann, K., P. Forterre, C. Brochier-Armanet, and S. Gribaldo.** 2014. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. Genome Biol Evol **6:**192-212.
35. **Roberts, J. A., S. D. Bell, and M. F. White.** 2003. An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. Mol Microbiol **48:**361-371.
36. **Senčilo, A., D. Jacobs-Sera, D. A. Russell, C. C. Ko, C. A. Bowman, N. S. Atanasova, E. Osterlund, H. M. Oksanen, D. H. Bamford, G. F. Hatfull, E. Roine, and R. W. Hendrix.** 2013. Snapshot of haloarchaeal tailed virus genomes. RNA Biol **10:**803-816.
37. **Warbrick, E.** 2000. The puzzle of PCNA's many partners. Bioessays **22:**997-1006.
38. **Williams, G. J., K. Johnson, J. Rudolf, S. A. McMahon, L. Carter, M. Oke, H. Liu, G. L. Taylor, M. F. White, and J. H. Naismith.** 2006. Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus*. Acta Crystallogr Sect F Struct Biol Cryst Commun **62:**944-948.
39. **Xu, H., P. Zhang, L. Liu, and M. Y. Lee.** 2001. A novel PCNA-binding motif identified by the panning of a random peptide display library. Biochemistry **40:**4512-4520.

Figure legends

Figure 1. Results of the yeast two-hybrid interactions between SIRV2 proteins and SsoPCNA. (A) An example of yeast two-hybrid analysis demonstrating the interaction between the SIRV2 protein P83a/b and SsoPCNA subunit 3. VEC, empty vector. Interaction between the bait (PCNA3) and prey (P83a/b) activates the *HIS3* expression, complementing histidine auxotrophy and allowing growth on yeast minimal media lacking histidine (-HIS). (B) Schematic representation of the SsoPCNA interactome. SIRV2 proteins found to interact with SsoPCNA1 and 3 are represented by black spheres, while DNA replication and repair proteins of *S. solfataricus* previously reported to bind SsoPCNA are shown as grey spheres. Hjc, Holliday junction endonuclease (8); FEN1, flap structure-specific endonuclease 1 (7); Dpo4, DNA polymerase IV (6); RFC-S and -L, small and large subunits of the replication factor C, respectively (6); PolB1, replicative DNA polymerase (7); UDG1, uracil DNA glycosylase (5); Lig1, DNA ligase (7); XPF, nucleotide excision repair endonuclease (35).

Figure 2. Potential PCNA-interacting protein (PIP) boxes in rudiviral proteins. (A-C) Structural homology models of three PCNA-interacting SIRV2 proteins for which structural homologues could be identified. The locations of the regions corresponding to predicted PIP boxes are encircled. Qxxhxx@@ is the PIP consensus sequence, where 'h' represents hydrophobic amino acid residues, '@' corresponds to bulky aromatic residues, and 'x' is any residue (37). (A) P83a/b model was built using as a template the X-ray structure of SIRV1 protein P56a (PDB ID: 2X48). (B) P105a model was built using as a template the X-ray structure of lipothrixvirus SIFV protein ORF14 (PDB ID: 2H36). C. P121 model was built using as a template the X-ray structure of *Sulfolobus solfataricus* Holliday junction resolvase (PDB ID: 1HH1). D. The locations of the three PIP boxes in P119a. The secondary structure elements predicted using JPred (4) are indicated above the sequence (H, α-helixes, E, β-strands).

312 **Table 1.** Conservation of SIRV2 PCNA-interacting proteins in other viruses.

SIRV2	SIRV1	SRV	ARV1	SMRV1	Characteristics
P83a (NP_666535)	gp01	P57	–	–	Helix-turn-helix protein
P84c (NP_666565)	gp23	P75	gp22	gp22	C-terminal coiled-coil domain
P105a (NP_666544)	gp03	–	–	–	Novel fold; homologues in several lipothrixviruses and fusellovirus SSV6
P119a (NP_666536)	–	–	–	–	No homologues
P121 (NP_666569)	gp27	P116c	gp25	gp19	Holliday junction resolvase; homologues in numerous bacterial and archaeal Caudovirales

313 Genome accession numbers: SIRV2 (NC_004086); SIRV1 (NC_004087); SRV, *Stygiolobus* rod-shaped
314 virus (FM164764); ARV1, *Acidianus* rod-shaped virus 1 (NC_009965); SMRV1, *Sulfolobales* Mexican
315 rudivirus 1 (NC_019413).