Archaeal virus-host interactions

Tessa E.F. Quax

Thesis committee

Promotor
Prof. Dr John Van der Oost
Personal Chair at the Laboratory of Microbiology
Wageningen University, The Netherlands

Co-promotor
Dr David Prangishvili
Head of the unit of ´Biologie Moléculaire du Gène chez les Extrêmophiles´
Institut Pasteur, France

Other members
Prof. Dr Rob Lavigne, Leuven University, Belgium
Prof. Dr Vincent Marechal, UPMC, Université Paris 6, France
Prof. Dr Monique M. van Oers, Wageningen University
Prof. Dr Mike S.M. Jetten, Radboud University Nijmegen

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Archaeal virus-host interactions

Tessa E.F. Quax

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday December 6th 2013
at 11 am in the Aula.

Tessa E.F. Quax

Archaeal virus-host interactions
188 pages.

Ph.D. thesis, Wageningen University, Wageningen, NL (2013)
With references, with summaries in Dutch, French and English

ISBN 978-94-6173-783-0

Table of contents

Preface and outline **7**

Chapter 1 Introduction to the archaeal virus model: *Sulfolobus islandicus* rod-shaped virus 2 17

Chapter 2 Composition of the SIRV2 virion egress structure 37

Chapter 3 The simple and elegant design of an archaeal virion egress structure 47

Chapter 4 Supramolecular organisation and assembly of the heptagonal pyramid for virion egress in archaea 65

Chapter 5 Transcriptome analysis of the SIRV2 infection cycle 89

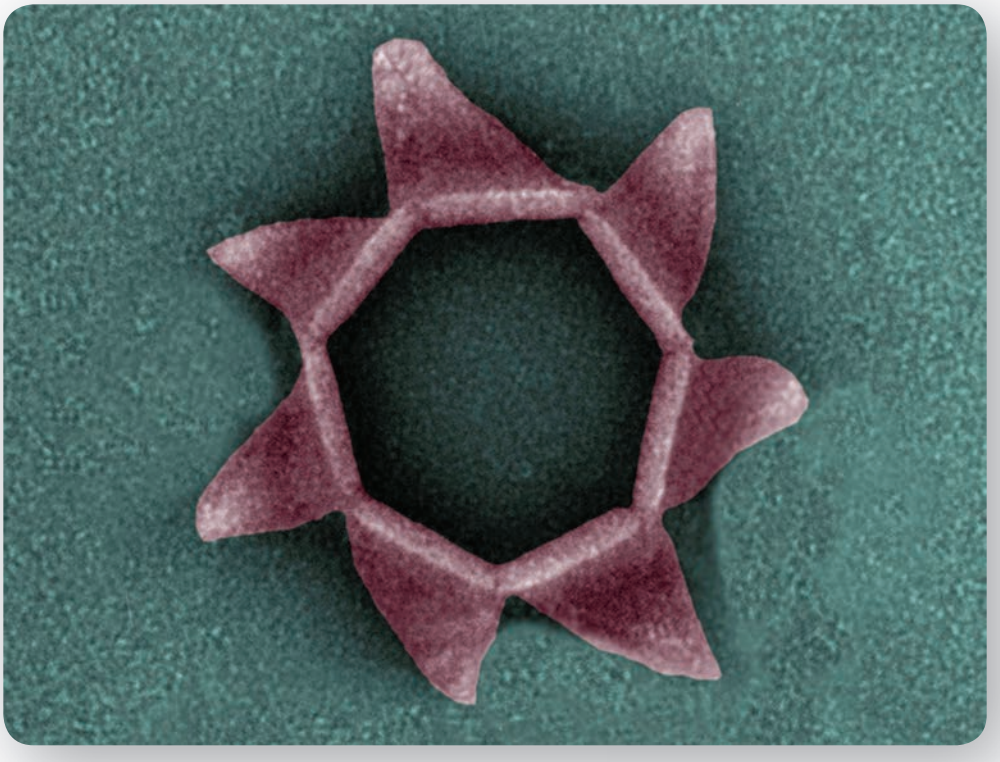
Chapter 6 Regulation of differential production of operon-encoded Cas proteins required for anti-viral defense 117

Chapter 7 Summary and general discussion 147

Appendices **161**

References 162
Co-author affiliations 174
Nederlandse samenvatting 176
Résumé Français 178
Acknowledgements 181
About the author 183
List of publications 184
Overview of completed training activities 185

Preface and outline



Preface

Invisible life

Up to the 17th century the existence of microorganisms, like bacteria and viruses, was unknown as they are invisible to the naked eye. The exploration of the wonder world of microorganisms could only begin after the development of a microscope with sufficient resolution. The pioneer Anthony van Leeuwenhoek constructed many self-made microscopes and was the first to visualize bacteria, which he observed in the material that he had scraped from his teeth (1). Another milestone was the discovery of non-cellular pathogens (later called viruses) by Martinus Beijerinck (2). Although viruses infecting bacteria were already described in the beginning of the 20th century independently by Frederick Twort and Felix d'Hérelle, it took until the 1950s and required the development of the electron microscope before the morphology of viruses was revealed (3-6).

Viruses are the most abundant biological entities on this planet and were isolated from practically all habitats. They can infect many (if not all) different cells, including microorganisms. Viruses infecting bacteria were instrumental for development of the discipline of molecular biology and served as tools to discover many basic aspects of cellular life (4). The subsequent progress of DNA sequencing techniques was one of the developments that led to overwhelming supporting evidence that life on earth is divided in three domains: Archaea, Bacteria and Eukarya (7). Archaea are the least characterised domain as archaea frequently thrive in extreme habitats and their cultivation may be a challenge (8, 9).

Life under extreme conditions

Extreme habitats are dominated by archaea (9). The first representatives of this domain were described in the 1970s and they appeared to form a bizarre collection of microbes, which were marked by their preference for particularly harsh environments (e.g. hyperthermic, hypersaline, acidic, alkaline). Because of these characteristics, archaea have attracted wide attention from astrobiologists in their search for extraterrestrial life (10-12). Later, archaea were also discovered in moderate environments where they were found to play key roles in biochemical cycles. Archaea are now recognized a major component of global ecosystems (13). It has been estimated that archaea constitute 20% of all microbial biomass in the sea (14).

Morphologically archaea resemble bacteria, since they lack distinctive intracellular organelles. Because of the lack of a nucleus, Archaea and Bacteria are named prokaryotes, in contrast to eukaryotes, which all contain a nucleus. This division refers to distinct levels of cellular organization rather than biological classification.

Archaea display numerous eukaryotic features, and the archaeal replication, transcription and translation machineries are for example basic versions of their eukaryotic counterparts (15, 16).

Because archaeal genomes code a mosaic of bacterial and eukaryotic functions, studies on archaea are important for understanding of the evolution of early life (17, 18).

Viruses of Archaea are unique

Similar to members of the other two domains of life, also archaea are infected by viruses. Although some early reports on virus-like particles from haloarchaea exist (19), archaeal viruses have only been intensively studied in the last two decades, concomitantly with the general recognition of Archaea as a separate domain (7). The hunt for archaeal viruses was pioneered by Wolfram Zillig and coworkers. Up to date it has resulted in about a hundred isolated viral species, which likely represent only the tip of the iceberg of archaeal viral diversity, since microscopical observation of archaeal habitats revealed a wealth of diverse virus-like particles (20, 21).

Ever since their discovery, archaeal viruses have intrigued scientists because of their unusual shapes, of which many are unique, including bottles, eggs and droplets (22). The majority of isolated archaeal viruses have dsDNA genomes (22). However, recently some ssDNA archaeal viruses were isolated (23-25). As for their structures, also genome sequences of archaeal viruses are typically very diverse and they share hardly any homology with sequences from extant databases (26). Due to the exceptional nature of archaeal viruses, they have already been assigned to ten different viral families, and this number is increasing rapidly as more archaeal viruses are being discovered and characterized (22). Especially viruses which are infecting extremophilic Crenarchaeota display an extraordinarily high morphological and genomic diversity in contrast to viruses from Euryarchaeota, which generally are thriving at moderate environments (27). Viruses of extremophilic archaea are stable under the same harsh conditions as the cells they infect, and therefore these viruses and their proteins are potentially very suitable for bionanotechnological applications (28). Furthermore, the study of archaeal viruses has proven instrumental for understanding viral diversity and evolution (29, 30).

Remarkable features of archaeal virus-host interactions

The described structures and genomes of archaeal viruses provide little insight in their infection cycle, due to the limited similarity of their gene sequences with those of viruses infecting Bacteria or Eukarya (26). For the majority of the archaeal viruses, aspects of interaction with the cells they infect remain obscure, including virion entry, exit, replication and interplay with anti-virus defense mechanisms. To tackle this issue, several archaeal viruses have emerged as suitable models, amongst which is *Sulfolobus islandicus* rod-shaped virus (SIRV)2 (31). This rod-shaped virus infects the hyperthermophilic *S. islandicus*, which thrives at high temperature (80 °C) and acidity (pH 3) (31). SIRV2 has been the focus of many recent studies, because of its remarkable proliferation cycle. Infection by this virus results in a degradation of the host genome. Moreover, formation of pyramidal shaped virion egress structures completes the infection cycle. In this

thesis, details of this exceptional virus-host interaction are described and in the last chapter comparisons are made with other known systems for virus release. In the first chapters focus is on characteristics of the unusual virion egress structure, including its unperceived geometry and features of its protein component. These chapters are followed by a whole transcriptome analysis aiming at identifying novel functions of SIRV2 genes important for interaction between virus and host. This analysis has revealed the wide spread activation of archaeal anti-viral defense systems, including the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) system.

Anti-viral defense system

In recent years it has been demonstrated that prokaryotes, in analogy to eukaryotes, have an extensive set of defense systems at their disposal to counteract viral attack. The CRISPR-Cas system is the most recently discovered example and shares some analogy with eukaryotic RNAi systems (32, 33). This system is present in about half the sequenced bacterial and almost all archaeal genomes. CRISPR-Cas defense relies on the incorporation of short sequences of foreign genetic elements between characteristic repeat sequences on prokaryotic genomes (34). These spacer repeat arrays are transcribed and the RNA processed, such that a small RNA molecule serves as a guide to recognize specifically the particular foreign element from which it was derived. Recognition of the foreign element occurs by a surveillance protein complex encoded by *cas* genes, which are associated with the CRISPR arrays. After its recognition, the foreign element is neutralized by binding and/or cleavage (34). Thus, the CRISPR-Cas system functions as both an adaptable and inheritable immune system.

Many different CRISPR-Cas systems exist, of which some target DNA and others RNA. Several types are exclusively present in archaeal genomes, while others are more universal. The various CRISPR-Cas systems all encode functional complexes of which the composition slightly differs. A common feature of these complexes is that they consists of several Cas proteins which typically display an uneven stoichiometry, i.e. the type I-E system consists of five subunits present in the ratio 1:2:6:1:1 (34, 35). Like most prokaryotic protein complexes, *cas* genes are encoded by operons. The founders of the operon concept, Monod, Jacob and Lwoff have described an "operon as a coordinated unit of genetic expression", which thus allows for co-transcription and co-translation (36). Therefore the mechanism by which differential expression of all subunits of the Cas complex is achieved remained a question which is addressed in Chapter 6 of this thesis. A regulatory mechanism is described which allows for differential production of Cas complexes and operon encoded proteins in general.

Thesis outline

This thesis is about the interaction between viruses and archaea and aims to reveal details of these interactions using SIRV2 as a model. The described studies address questions concerning the uniqueness of infection cycles of archaeal viruses and possible resemblance to other systems. Moreover, many of these findings have broad implications. Examples are the discovery of a universal membrane remodeling system, which can be exploited in a biotechnological setting, and the description of a prokaryotic regulatory mechanism for the production of operon encoded protein complexes.

Chapter 1

In the first chapter the archaeal virus SIRV2 is introduced. SIRV2 is a model for the study of archaeal virus-host interactions. An overview of characteristics of its infection cycle is presented. Flow cytometry experiments demonstrate that infection of *Sulfolobus islandicus* by SIRV2 results in degradation of the host genome, indicating that the rod-shaped virus is lytic. Analysis with electron microscopy of infected cells reveals that in the cytoplasm of each cell approximately 100 virions are assembled in several bundles, which fill almost the complete cell. More striking is the observation of several pyramidal shaped virion egress structures on the host cell surface. These Virus Associated Pyramids (VAPs) protrude with their tips into the extracellular medium and perforate the protective surface layer (S-layer) of the host cell. The cytoplasm is continuous in the interior of the VAPs. As the last stage of the infection cycle the VAPs open outwards, like the petals of a flower. In this fashion, apertures in the cell membrane are created, through which SIRV2 virions can escape from their host cell. After release of SIRV2 virions, these cells remain as empty ghosts.

All viruses face the challenging task of crossing the cell membrane at the end of the infection cycle. Different viruses have come up with various solutions to tackle this issue (e.g. by using the endolysin-holin system, or budding) (37, 38). However, the novel virus release mechanism described in this chapter is fundamentally different from all other reported strategies for virion egress and has therefore important implications for understanding the development and evolution of these systems.

Chapter 2

The discovery of the unusual archaeal virion egress structure raises questions concerning its composition. This is the focus of the second chapter of this thesis. The isolation of proteins from *S. islandicus* is described, which reveals specific accumulation of a 10 kDa protein in the membranes of SIRV2 infected cells. Using mass spectrometry, this protein is identified as the product of the viral gene *ORF98*. Accumulation of this protein, named SIRV2_P98 (PVAP), is so pronounced that

after infection it is the most abundant membrane protein as can be judged by visualization on SDS gels. Apart from accumulation of PVAP, the protein composition of membranes of infected cells seems to remain unmodified. Results of Edman degradation are presented, which show that the N-terminus of PVAP remains intact without any indication of (cleavage of) a signal sequence. PVAP has only a few homologues, mostly amongst the *Rudiviridae*, the viral family to which SIRV2 belongs. Strikingly, the only other PVAP homolog outside the *Rudiviridae* is present in the genome of *Sulfolobus* Turreted Icosahedral Virus (STIV)1 (39), apart from SIRV2 the only archaeal virus which has been demonstrated experimentally to employ the VAP based egress mechanism (40). Apart from PVAP, STIV1 and SIRV2 share only two other genes. Surprisingly, PVAP is not present in all rudiviral genomes and is also absent in the genome of STIV2, a close relative of STIV1. This indicates that during the evolution driven by the on-going arms race with host defense systems, morphological and egress systems in archaeal viruses likely are exchanged horizontally and evolve independently.

Chapter 3

The third chapter of this thesis deals with the elegant architecture of the VAPs. In addition, the composition of this structure is presented in detail. The findings described in Chapter 2 are the basis for the experimental setup in this chapter. Conveniently, VAPs can exist as stable discrete particles. These particles are isolated from crude membrane extracts using antibodies raised against PVAP. An electron microscopical survey of the particles, unveiled that isolated VAPs are present as intact polyhedra displaying a sevenfold rotational symmetry. Although not totally unheard of (e.g. 20S proteasome, scallop muscle myosin, alpha-hemolysin, chaperonin heat shock protein Hsp10 (41-44)), sevenfold symmetry is rare in nature. The geometry of isolated VAPs is always similar, although their sizes can differ considerably, suggesting that they develop by outwards growth of their facets.

Study of these isolated VAP particles allows the discovery of more aspects of its biology. Analysis of the composition of the isolated VAP, demonstrates that it consist exclusively of PVAP. Immunolabelling with antibodies raised against PVAP show the specific presence of this protein in VAPs of infected cells. PVAP is not only required but also sufficient for VAP formation, and it is likely that no viral or archaeal specific factors are required for this process. This is demonstrated by heterologous expression of this protein in *Escherichia coli* and *Sulfolobus acidocaldarius*. Astoundingly, in both systems, PVAP leads to the formation of pyramidal structures on the inner membrane and cell envelope respectively.

Thus, PVAP is capable of self-assembly into VAPs under the distinct growth conditions of hyperthermophilic acidophilic archaea and mesophilic bacteria. This is even more striking because of the fundamentally different lipid and protein composition of archaeal and bacterial membranes.

Chapter 4

The detailed supra-molecular structure and assembly process of the VAP are focus of Chapter 4. First of all, it is demonstrated that PVAP displays a high flexibility to form VAPs in virtually all biological membranes and can act as a universal membrane remodeling system. PVAP is expressed in several heterologous systems, including the eukaryote *Saccharomyces cerevisiae*. This demonstrates that VAPs can successfully assemble even in eukaryotes. In *S. cerevisiae* PVAP expression leads to VAP formation in different cellular membranes, including the Golgi, nuclear envelope and mitochondrial membranes. This suggests that PVAP does not follow the conventional pathways for membrane targeting. Correct insertion in membranes likely solely depends on the hydrophobicity of its N-terminal transmembrane segment.

Next, the ultrastructure of VAPs is described as observed with whole cell cryo-electron tomography. This analysis reveals the presence of two layers in VAPs, both in natural and in heterologous systems. The outer VAP layer appears continuous with the cell membrane, while the inner layer represents a protein sheet, which is facing the cytoplasm. In addition, the early development of VAPs is monitored, which shows nucleation of the structure on the cell membrane. When first formed, the VAPs are small (tens of nanometers). As the infection proceeds the VAPs grow by the expansion of their seven triangular facets, eventually reaching diameters of up to 250 nm.

Insight in particular functions of distinct PVAP domains is provided by analysis of several genetically modified versions of the protein. The three predicted C-terminal α-helices of PVAP monomers likely interact to form a dense protein sheet, which is visualized as the inner VAP layer. The N-terminal transmembrane segment of PVAP is important for interactions between PVAP monomers and the membrane. The existence of PVAP oligomers is demonstrated by analysis of purified PVAP on SDS-PAGE, which revealed the presence of several oligomers, including one with the estimated mass of a heptamer. Based on these results a model for VAP assembly is presented. The ability of PVAP monomers to interact with each other in various configurations seems a distinctive feature of this protein which might account for the unique assembly and geometry of the archaeal virion egress structure.

Chapter 5

In the fifth chapter of this thesis a transcriptome analysis of SIRV2 infection is presented. These results, in combination with a study of interactions between SIRV2 proteins, are used to advance current understanding of archaeal viral functions. This study provides a global map, defining host and viral gene expression during prokaryotic viral infection. As such, it provides in depth information on the expression profiles of all viral and host genes during a time series after SIRV2 infection. This approach is combined with yeast two-hybrid assays to analyze all possible interactions of viral proteins. The comprehension of the SIRV2 infection cycle is currently

hampered by the lack of predicted or confirmed functions for the majority of the encoded proteins. This is a common challenge for most archaeal viruses, due to low homology of their gene products with proteins from extant databases. To provide clues on the functions of SIRV2 genes in the infection cycle, the expression pattern of viral and host genes is monitored by RNA sequencing analysis. This high resolution analysis reveals some previously unrecorded events, such as the start of transcription of viral genes simultaneously from both ends of the SIRV2 dsDNA genome. The effect of SIRV2 infection on the host cell is severe and *S. islandicus* reacts to this threat by the massive activation of anti-viral defense systems, notably CRISPR-Cas and toxin anti-toxin systems. CRISPR-Cas systems are particularly abundant in archaea and typically the archaeal CRISPR arrays are very long, implying an important physiological role. This study reports on activation of archaeal CRISPR-Cas systems. In addition, it provides some first clues on a role of archaeal toxin anti-toxin systems in anti-viral defense. The observed different degree of activation of various defense systems highlights the specialized functions they perform. The presented information on individual gene expression and activation of all anti-viral defense systems of *S.islandicus* is expected to aid future studies, which are aiming to establish the function and interplay of the different systems *in vivo*.

Chapter 6

The peculiar stoichiometry of some anti-viral Cas defense complexes has provided the basis of the findings described in Chapter 6. In the previous chapter it is demonstrated that the functional complex of CRISPR-mediated defense is activated during SIRV2 infection. The subunits of the *Sulfolobus* Cas-related ribonucleoprotein complexes are composed of different subunits that are present in an uneven stoichiometry. This stoichiometry is particularly puzzling since *cas* genes are organized as operons. Evidently this is advantageous in case a protein complex with even stoichiometry is encoded. But expression of protein complexes with uneven stoichiometry requires an additional mode of control. Transcription analysis (Chapter 5) indicates at most minor differences in the transcription of subunit-encoding genes, not sufficient to explain the required differences at protein level. This suggests that regulation might take place by differential translation. These CRISPR associated complexes are not the only prokaryotic protein complexes with uneven stoichiometry. In fact, several operon encoded complexes with uneven stoichiometry perform roles in important cellular processes, like translation, energy production and secretion (45-47). Thus, regulation of the stoichiometry of such complexes seems very important for all prokaryotes. The recent advancement in development of high throughput techniques like RNA-sequencing and ribosome profiling, have provided a huge data set that can be used to gain further insight into the prokaryotic transcription and translation processes. Fifty years after the first description of the operon by Jacob and Monod, analysis of these datasets urge to refine the operon concept

and acknowledge the wide prevalence of differential translation to fine tune protein production from genes organized in operons.

A group of operon-encoded protein complexes with known stoichiometry is selected for this study and features influencing the efficiency of transcription and translation of their genes are assessed in a selection of >350 prokaryotic genomes. In agreement with the transcription analyzes of the *Sulfolobus* Cas complexes, this reveals that typically differential transcription cannot explain fluctuating protein levels. However, a clear correlation with factors influencing translation efficiency is observed, both at the level of initiation (RNA secondary structure, ribosome density profiles) and elongation (codon bias, codon co-occurrence). Since highly abundant subunits are not necessarily encoded by the first gene of the operon, *de novo* translation initiation of any cistron may occur irrespective of its position on a polycistronic mRNA. Experimental evidence for *de novo* translation initiation on intercistronic sites is provided using synthetic operons encoding reporter proteins. The findings described in this chapter suggest that tuning of the translation efficiency is a universal mode of control in bacteria and archaea, which allows for differential production of operon encoded genes.

Chapter 7

In the last chapter of this thesis a summary of the presented work is given. This is followed by a discussion of the state of the art of research on archaeal egress systems. The VAP-based egress system is compared with other (bacterial) systems and possible evolutionary advantages are discussed. Lines for future research and biotechnological applications are suggested. Due to the unique nature of the VAPs it is proposed to classify these non-capsid virus-encoded independent structures as 'virodomes'.

Next, the potential of SIRV2 as model to study archaeal viral biology is discussed followed by suggestions to study SIRV2 replication and the interplay with anti-viral defense mechanisms in more detail. Codon bias in combination with other factors is suggested as an *in silico* predictor of protein subunit stoichiometry. The chapter is concluded with a short discussion on the significance of archaeal anti-viral defense systems for the virus-host interplay in the natural environment.

Chapter 1

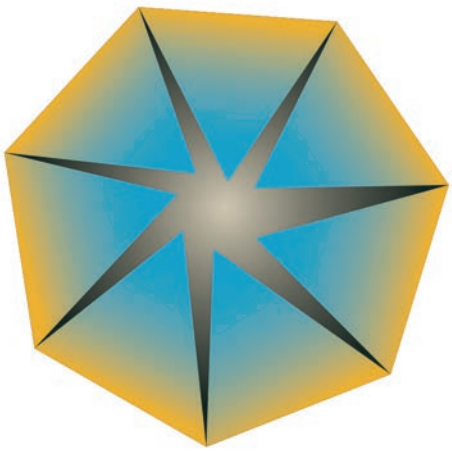
Introduction to the archaeal virus model:

Sulfolobus islandicus rod-shaped virus 2

Ariane Bize, Erik A. Karlsson, Karin Ekefjärd, Tessa E. F. Quax, Mery Pina,
Marie-Christine Prevost, Patrick Forterre, Olivier Tenaillon, Rolf Bernander, David Prangishvili

'A Unique *Virus* Release Mechanism in the Archaea'

Proceedings of the National Academy of Sciences USA. 2009 July 7. 106(27):11306-11



Abstract

Little is known about the infection cycles of viruses infecting cells from Archaea, the third domain of life. Here, we demonstrate that the virions of the archaeal *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) are released from the host cell through a novel biological mechanism, involving the formation of unique cellular structures. Large pyramidal virus-induced protrusions transect the cell envelope at several positions, rupturing the S-layer; they eventually open out, thus creating large apertures through which virions escape the cell. We also demonstrate that massive degradation of the host chromosomes occurs due to virus infection, and that virion assembly occurs in the cytoplasm. Furthermore, intracellular viral DNA is visualized by flow cytometry for the first time. The results show that SIRV2 is a lytic virus, and that the host cell dies as a consequence of elaborated mechanisms orchestrated by the virus. The generation of specific cellular structures for a distinct step of virus life cycle is known in eukaryal virus-host systems but it is unprecedented in cells from other domains.

Introduction

As for organisms belonging to the Bacteria and Eukarya, members of the domain Archaea are infected by specific viruses. The majority of archaeal viruses isolated so far contain dsDNA as the genetic material and infect hyperthermophilic hosts from the phylum Crenarchaeota (48). The diversity and uniqueness of these viruses at both the morphological and genetic levels are such, that they have been classified into seven new viral families (49). The knowledge on the biology of this exceptional group of viruses is still limited, partly due to the unique genetic content: very few genes have detectable functions or homologs in the databases (26).

In particular, little is known about relationships of crenarchaeal viruses with their hosts. Except for a few isolated cases (50-52), it is generally presumed that these viruses persist in the host cell in a carrier state, a non-lytic relationship in which virions are continuously secreted by the still-dividing cells (53). However, the classification of crenarchaeal viruses as chronic is based on indirect experimental evidence, such as a lack of optical density (OD) decrease and absence of cellular debris in infected cultures (*e. g.* (54, 55)). Detailed characterization of the infection cycle and the carrier state have not been specifically addressed in the scarce reports on crenarchaeal host-virus interactions (see *e. g.* (56)).

To study the nature of host-virus relationships in crenarchaea, we selected the non-enveloped, rod-shaped virus SIRV2 and its hyperthermophilic and acidophilic host, *Sulfolobus islandicus*. SIRV2, originally described as a carrier state, non-lysogenic virus (31), belongs to a common crenarchaeal virus family, the *Rudiviridae* (31, 55, 57-59), and contains a linear 35.5 kb dsDNA genome (60). The host belongs to a well-characterized crenarchaeal genus, *Sulfolobus(61, 62)*, from which also other viruses are known (49). We describe detailed *in vivo* effects of the virus on its host and, unexpectedly, demonstrate that SIRV2 is a cytocidal, lytic virus. Remarkably, an entirely novel virus release mechanism was encountered during the characterization, involving generation of unique pyramidal structures which, by opening out, cause local disruption of the cell envelope and allow virion escape. In addition, intracellular viral DNA was visualized by flow cytometry, and the technique was also used to demonstrate chromosome degradation in infected cells.

Results

Growth kinetics of SIRV2-infected cultures

OD and CFU values from uninfected and infected (multiplicity of infection (m.o.i.) about 7) cultures of *S. islandicus* were monitored over time. The effects of the virus were visible already 1.5 h post-infection (Fig. 1). Whereas uninfected cultures pursued normal growth with a generation time

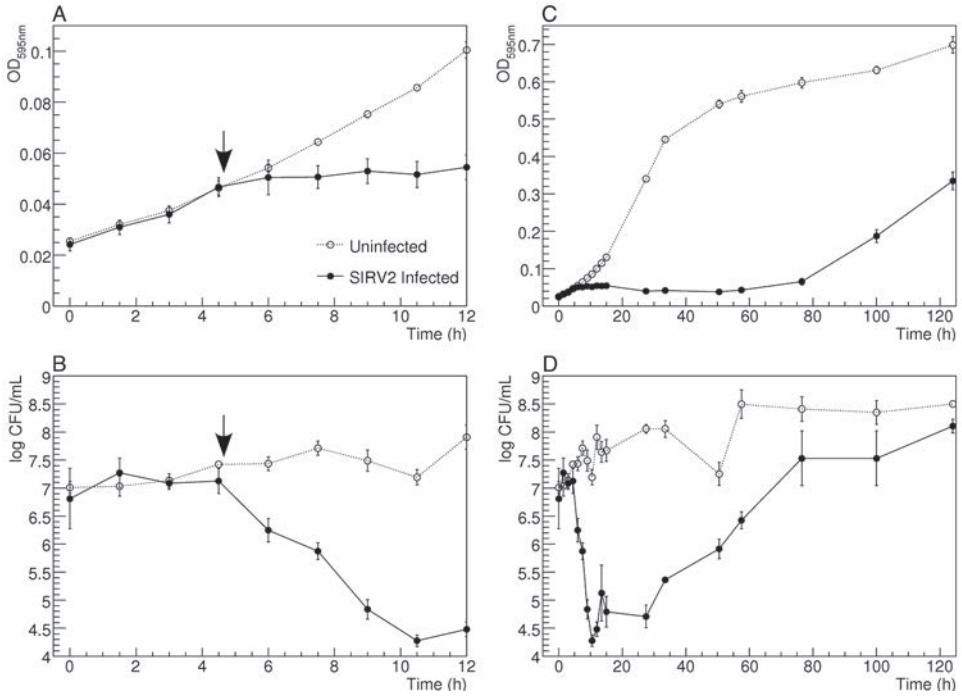


Figure 1. Impact of SIRV2 infection on the growth kinetics of *S. islandicus* cultures. Cultures infected at a moi of ≈7 (filled circles, continuous line) and uninfected cultures (empty circles, dotted lines), were launched in triplicates. Averages of the replicates ± 1 SD are shown. The vertical arrows in *A* and *B* correspond to virus addition (4.5 h). (**A**) $OD_{595\ nm}$, detail of the first hours. (**B**) Log transformation of the CFU titres, detail of the first hours. (**C**) $OD_{595\ nm}$ over the entire time course. (**D**) Log transformation of the CFU titers over the entire time course.

around 13 h, the OD in infected cultures remained constant for about 60 h (Fig. 1*A* and 1*C*), after which growth resumed (Fig. 1*C*). During this time period, the CFU values of uninfected controls remained constant or increased slightly. In contrast, the CFU values decreased dramatically in infected cultures, resulting in an about 1000-fold reduction at 6 h post-infection (Fig. 1*B*, 10.5 h). The CFU values also revealed growth of a minor cell population in infected cultures starting at early time points (Fig. 1*D*, from 15 h). This growth was initially not detectable in the OD measurements (Fig. 1*C*), due to the low concentration of this cell population at early time points. Thus, infection by SIRV2 has a pronounced effect on the host cultures, preventing growth of a majority of the cells.

To exclude the possibility that the results were linked to the high m.o.i. employed, or to the specific growth conditions, similar experiments were performed at low m.o.i. (about 10^{-3}), at different temperatures (70°C; 75°C; 78°C), pH (3.0; 3.5), medium richness (standard medium or 5-fold less rich medium), and with different host strains (*S. islandicus* strains KVEM10H3,

HVE10/4 and LAL14/1). No significant differences were observed (data not shown), indicating that the effects occurred independently of these parameters.

The cell population growing in the presence of SIRV2 consisted of cells completely resistant to SIRV2 infection, not producing any detectable infectious virions nor carrying the SIRV2 genome (data in Supporting Information (SI): Text and Fig. S1). This was consistent with the observation that the SIRV2 genome does not integrate into the host chromosome (31), and also excluded the possibility that SIRV2 established a carrier state relationship with its host. The high initial proportion of resistant cells suggested that specific mechanisms could be involved in their generation, in addition to random mutations, such as CRISPRs-related mechanisms (33).

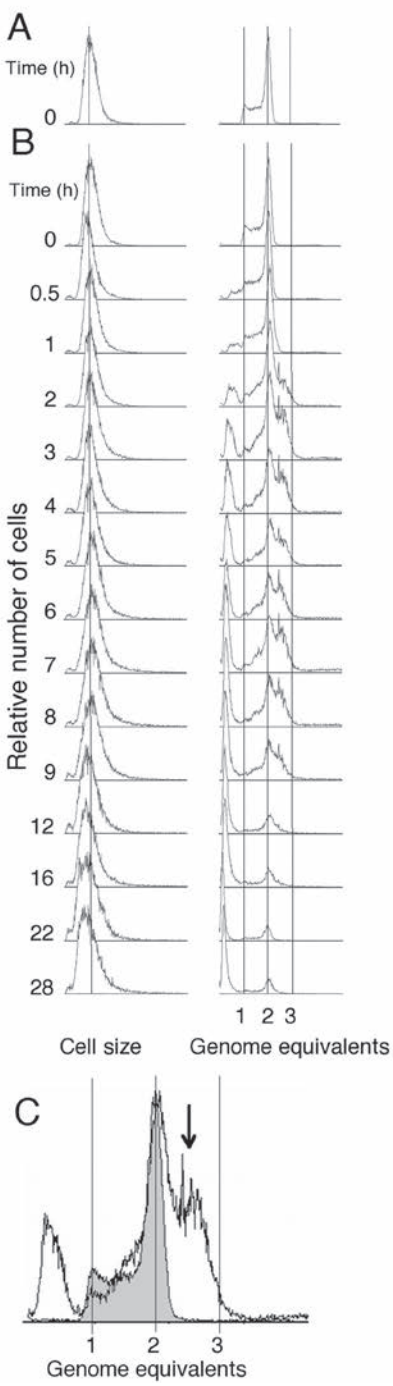
Flow cytometry analysis of infected cells over a time course

The cell size and intracellular DNA content in uninfected and SIRV2-infected cultures (m.o.i. about 10) over a time course were monitored by flow cytometry (Fig. 2, and Figs. S2 and S3).

The relative lengths of the *S. islandicus* cell cycle periods in the control cultures were found to be similar to those of other *Sulfolobus* species (61, 63), with the post-replicative phase occupying a large fraction of the generation time (68%, Fig. S4). Based on the results of flow cytometry, the genome size was roughly estimated to 2.6 Mb (Fig. S5). The average cell size (Fig. 2*A* left, Fig. S2) progressively decreased when the cultures approached stationary phase. In the infected cultures (Fig. 2B, left column), a cell size increase initially occurred in part of the cell population, evident as an extension of the distribution towards the right (6 - 8 h). Subsequently, the average cell size gradually decreased over time.

The DNA content distributions of the control cultures (Fig. 2*A* right, Fig. S2) were typical for exponentially growing *Sulfolobus* cells (63), with a majority of the cells containing 2 chromosomes. In the infected cultures (Fig. 2*B*, right column), cells with a very low DNA content (<<1 genome equivalent) started to appear at 0.5 h post-infection and then increased in proportion over time, while the proportion of cells containing 1 - 2 genome equivalents decreased. Thus, at 12 h, a large majority of the cell population contained no detectable intracellular DNA. The SIRV2 latent period is 8 - 10 h (below) and chromosome degradation, thus, occurred before virus release in a significant fraction of the cell population. Interestingly, the populations of chromosome-less cells and cells containing DNA were clearly separated and well defined (Fig. 2*B*, right column, Fig. S6). Thus, for a given infected cell, chromosome degradation must have occurred within a brief time interval.

In parallel to chromosome degradation, an increase in the total DNA content occurred in part of the cell population, evident as an extension of the 2-genome equivalents peak towards the right (Fig. 2*B*, right column, 2 h and onwards). This corresponded to newly synthesized viral DNA (below), and the increase was estimated to about 0.5 genome equivalents on average (Fig. 2*C*, arrow), at 3 h after infection.



The results demonstrate that infection by SIRV2 causes massive degradation of the host chromosome in virtually all infected cells during the first 12 h of infection, excluding the possibility that SIRV2 genomes are vertically transmitted between cell generations.

Links between the virus infection cycle, the kinetics of host chromosome degradation, and cell death

To discriminate between host chromosome and viral DNA, uninfected and infected cultures (m.o.i. about 15) were monitored by dot blot hybridizations, in addition to flow cytometry (Fig. S6). In an uninfected control culture, the percentage of chromosome-less cells did not exceed 5% (Fig. 3*A*), and tended to decrease over time. In infected cultures, chromosome-less cells began to accumulate in the first hours, and after 5 h, the percentage was about 40%, confirming that significant degradation occurred before virion release (at about 8 - 10 h, see below) and, at 11 h, more than 80% of the cells were chromosome-less. Subsequent degradation occurred at a lower rate and finally reached 97%, confirming that genome degradation occurred in most cells.

The intracellular amount of SIRV2 DNA (Fig. 3*B* and 3*D*) increased gradually and reached a

Figure 2. Flow cytometry time-course analysis of *S. islandicus* cells infected by SIRV2. (**A**) Representative cell size and DNA content distributions from an uninfected culture. (**B**) Cell size and DNA content distributions from a culture infected with SIRV2 (moi ≈10). The virus was added just after time point 0 h. (**C**) Visualization of intracellular SIRV2 DNA by flow cytometry at 3 h after infection. The DNA content distribution from an infected *S. islandicus* culture is shown against the distribution from an uninfected culture (translucent gray). The arrow indicates additional DNA in infected cells.

maximum around 8 h, followed by a large decrease up until 14 h. The initial increase presumably corresponded to viral DNA replication, and the decrease to virus release, indicating a latent period of about 8-10 h. Thus, a single round of infection occurred in the cultures at the high m.o.i. employed. To relate viral DNA production to the dynamics of chromosome degradation, the percentage of DNA-less cells appearing between successive time points was superimposed (Fig. 3*B*). A small peak of degradation, visible at 0.5 h post-infection, was most likely an artifact caused by the low signal-to-noise ratio for DNA-less cells in the very early time points. The major peak occurred at 11 h, in the middle of the virus release period. The use of a 16S rRNA gene probe combined with similar data analysis confirmed the chromosome degradation observed by flow cytometry (Fig. 3*C* and 3*E*).

To confirm the latent period of 8 - 10 h and to estimate the burst size, a one-step growth experiment was performed (Fig. S7). Virus release was shown to begin at around 8 - 10 h post-infection, confirming that no infectious virions were released prior to this time point, consistent with all other data. The burst size was estimated to 30±10 viruses per cell, in agreement with the estimates of intracellular SIRV2 DNA by flow cytometry. Finally, a membrane potential-sensitive probe (data in SI: text and Fig. S8) was used to confirm that cell death occurred in connection to virus release.

In conclusion, a single round of infection occurred when a high m.o.i. was used, and the SIRV2 latent period was about 8 - 10 h under the conditions employed. Massive host chromosome degradation occurred throughout the infection cycle, starting from the early stage, and cell death took place concomitantly with virus release. Thus, SIRV2 is a lytic virus that kills the host cell during the process of virus production and release.

Identification of novel cellular ultrastructures induced by SIRV2 infection

To obtain insights into the details of the virus-host interactions, infected cells were analyzed by scanning electron microscopy (SEM) and transmission electron microscopy (TEM). The cells were fixed at 10 h (just before virion release), 13 h (middle of release period) and 26 h (after release) post-infection. Uninfected cells in mid-exponential growth phase were used as control. For analysis with TEM, ultrathin sections of samples were prepared.

The irregular coccoid morphology of uninfected cells was typical for *Sulfolobales* species, with the cell envelope consisting of a lipid membrane and an S-layer (Figs. 4*A1*, 4*A2*). At 10 h post-infection, multiple pyramidal protrusions were observed on the cell surface by SEM (Fig. 4*B1*, arrows), which were absent in uninfected control cells. In thin sections these structures were observed by TEM as large angular protrusions associated with a local absence of S-layer at the cell envelope, (Figs. 4*B2*, 4*B3*, 4*B4*). Both with SEM and TEM, several such virus-associated pyramids (VAPs) were usually visible per cell (Figs. 4*B1*, 4*B2*). The pyramidal structure of the VAPs, suggested by SEM, was confirmed by TEM, showing a polygonal base in a plane parallel to the

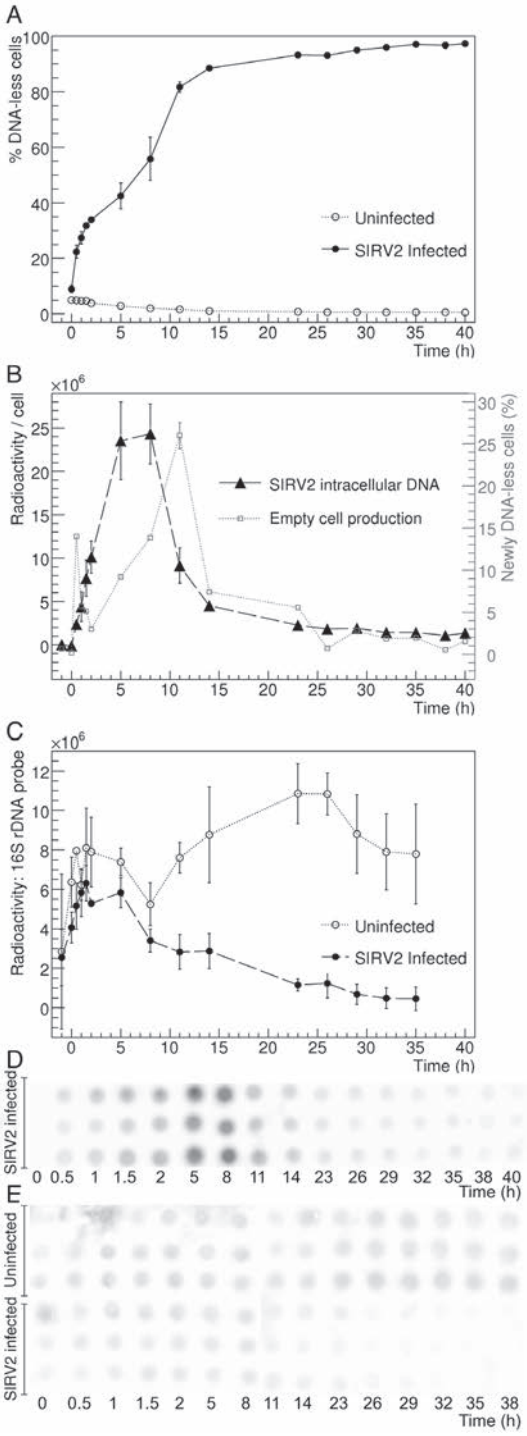


Figure 3. Links between the kinetics of host chromosome degradation and the SIRV2 infection cycle. Infected cultures and uninfected cultures were launched in triplicates. SIRV2 was added (moi ≈15) just after time point 0 h. Averages of 3 infection replicates ± 1 SD are shown in *A*–*C*. (**A**) Percentage of DNA-less cells in uninfected and infected cultures. The values were obtained by flow cytometry analysis, using data from 2-parameter distributions, gating them as illustrated in Fig. S6*B*. (**B**) Radioactivity/cell (filled triangles, discontinuous line, left axis), indicative of SIRV2 intracellular DNA in infected cultures, over a time course. Values in arbitrary units were obtained by quantifying the hybridization signal from each spot in the image shown in *D*. The percentage of DNA-less cells appearing between 2 successive time points (empty circles, dotted line, right axis) was also plotted, using the data from *A*. (**C**) Radioactivity/cell indicative of intracellular 16S rDNA amounts in uninfected cultures (empty circles, dotted line) or infected cultures (filled circles, discontinuous line). Values in arbitrary units were obtained by quantifying the hybridization signal from each spot in the image shown in *E*. (**D**) Autoradiogram of hybridization of spots of cells sampled from infected cultures with a SIRV2-specific probe. Each spot corresponds to the same approximate number of cells, based on OD measurements. The time course corresponds to horizontal lines, with the 3 replicates shown vertically for each time point. (**E**) Autoradiogram of hybridization of spots of cells sampled from uninfected and infected cultures with a 16S rRNA gene-specific probe. See *D* for additional explanations.

cell envelope (Fig. 4*B5*). In thin sections the VAPs often contained regions producing a denser staining (Fig. 4*B2,* arrows), localized at the tip of the pyramidal structure.

Dense aggregates of virions were visible by TEM within numerous cells from the infected culture (examples in Figs. 4*B4*, 4*B6*), showing that virion assembly occurred in the cytoplasm. Up to 3 densely packed aggregates, together containing up to about 150 virions, were detected in the cell sections, and occupied a high fraction of the intracellular volume. The higher number of virions compared to the estimates from the flow cytometry and one-step growth experiments (above), could be due to that viral DNA tightly packaged into virus particles is poorly stained, and that virions may still have been aggregated after release in the one-step experiments.

At 13 h post-infection, together with cells resembling the examples shown in Figs. 4*B1* and 4*B2*, cells lacking VAPs and displaying numerous perforations on cell surface were observed (not shown, similar to Figs. 4*C1*, 4*C2*), and 26 h post infection almost all cells were perforated and empty (Figs. 4*C1*, 4*C2*). The perforated cells appeared to have spherical morphotype, different from the native phenotype, suggesting alteration of intracellular organization. Thin section analysis of perforated cells displayed virion remains (Fig.4*C2*, inset) and disappearance of most of the cytoplamic content (Fig.4*C2*). The cell perforations were heterogeneous in size, and their majority visible in thin sections had a diameter in the range of 200 nm. TEM analysis of thin sections revealed that the perforations were delimited by C-shaped structures (Figs. 4*C2*, 4*C3*). Most likely, these represented the remains of VAPs. Apart from the perforations, the cell envelope appeared to be intact, with both the S-layer and the membrane visible (Figs. 4*C2*, 4*C3*). Notably, the characteristic structures at the boundary of the perforations of the lysed cells were sometimes observed detached from the cell envelope (Fig. 4*C4*, 4*C5*, 4*C6*). The resemblance of polygonal shapes in Figs. 4*B5*, 4*C5* and 4*C6*, as well as the similarity of the structures in Figs. 4*B3* and 4*C3*, supports the hypothesis that the structures in panel C represented remains of the VAPs shown in panel B. Thus the VAPs were apparently involved in perforation of the cell envelope. Since ongoing virus release could not be detected, this must have occurred within a brief time interval.

Discussion

We report the first detailed cellular study of the infection cycle of a crenarchaeal virus, and demonstrate that SIRV2 is a lytic virus. The virions are assembled in the cytoplasm of the host cell and, 8 - 10 hours post infection, start to be released through well-defined apertures in the cell envelope. Remarkably, formation of these openings is preceded and facilitated by the generation of virus-induced cellular structures of pyramidal shape, VAPs, located at the cell envelope and pointing outwards. The VAPs perforate the membrane and S-layer, and after disruption leave

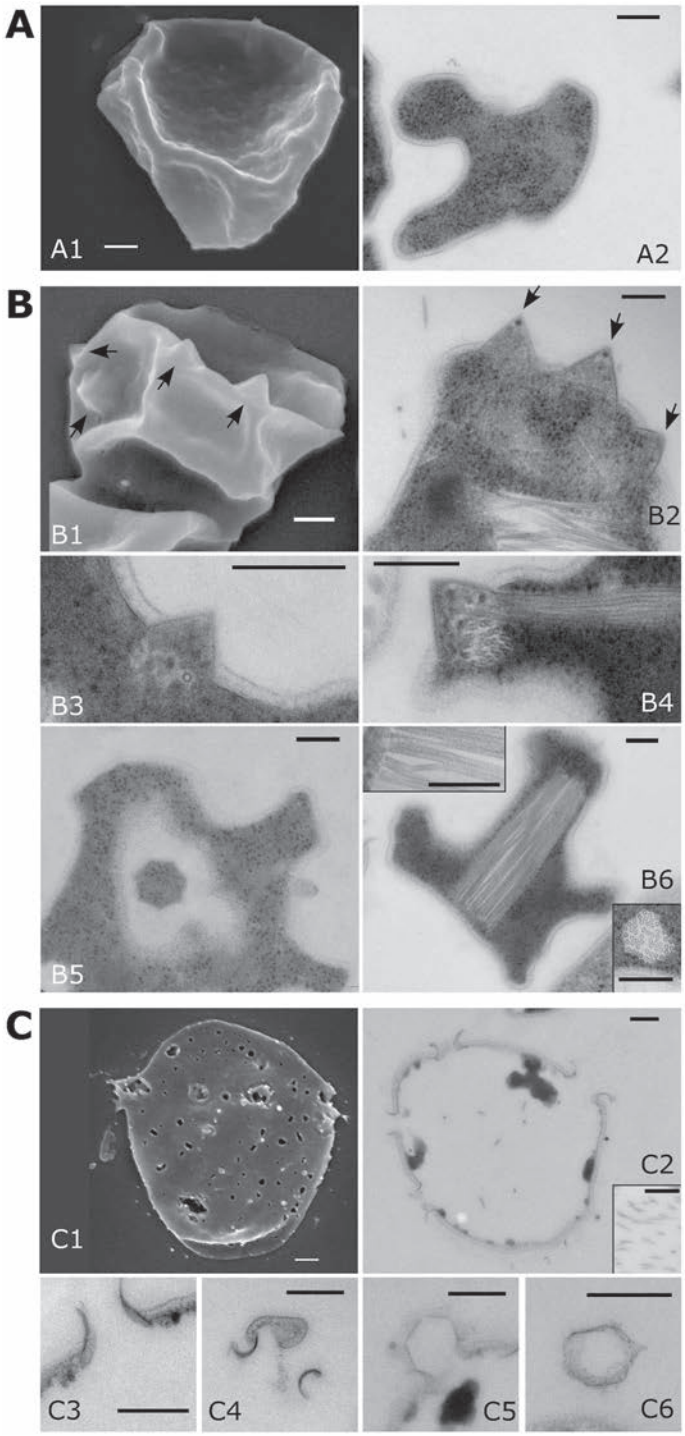


Figure 4. VAPs, detected by SEM and TEM on SIRV2 infected *Sulfolobus* cells. *A1*, *B1*, and *C1* micrographs were obtained by SEM, all other micrographs are TEM images from thin sections. (**A**) Uninfected cells. (**B**) Cells 10 h after infection. (*B2*, *B3*, *B4*, and *B6*) Thin sections in a plane perpendicular to the cell envelope. (*B5*) Thin section in a plane parallel to the cell envelope. (*B1* and *B2*) arrows indicate VAPs. (*B6 Insets*) Details of intracellular virion aggregates, sectioned according to a parallel (up) or perpendicular (down) plane. (**C**) Cells 26 h after infection. (*C2*, *C3*, and *C5*) Thin sections in planes perpendicular to the cell envelope. (*C5*) Disrupted VAP partly detached from cell envelope. (*C4* and C*6*) Thin sections of disrupted detached VAPs in different section planes. (*C2 Inset*) Virion remains inside a lysed cell. (Scale bars, 200 nm.)

behind apertures delimited by a ring structure of polygonal shape (Fig. 4C*5*, 4C*6*). Intracellular viral DNA was visualized by flow cytometry, and the same technique was used to show that host chromosome is completely degraded during the viral infection cycle. The combination of the data from one-step growth experiment, flow cytometry and TEM showed that chromosome degradation most likely occurred before the virion release, in the majority of the cell population. Together, all our results demonstrate that the host cells die as a consequence of specific and unique mechanisms orchestrated by the virus, rather than from general deleterious effects of the infection. The deduced viral life cycle is schematically illustrated in Fig. 5.

It is likely that a set of viral genes must control the formation of the VAPs and the generation of the apertures through which the virions are released. The genes might either directly code for the proteins involved, or modulate host-encoded mechanisms. The timing of VAP disruption and virus release must also be controlled by virus-encoded functions, such that cell lysis does not occur until the virions have been assembled, as for any lytic virus. Further, host chromosome degradation could also be an active mechanism, encoded by viral genes.

To our knowledge, the virus release mechanism identified here is unprecedented in virus biology. In lytic bacteriophages, the two main lysis strategies rely on the direct degradation of peptidogycan, *e. g.* with the holin-endolysin system (64), or on the inhibition of cell wall synthesis (65). Both strategies result in complete cell disruption, and do not involve a modification of the cell envelope in several localized regions, as reported here. To our knowledge, also for eukaryotes there are no reports on generation of distinct structures for cell perforation and viral release. Modification of intracellular membranes (endoplasmic reticulum, Golgi complex) does occur as a result of infection with certain eukaryotic RNA and DNA viruses, but this appears to be linked to viral replication rather than release (66). Recently, alteration of the *Sulfolobus* S-layer as a result of infection with the lytic icosahedral STIV virus was reported (67). It would be highly interesting if viruses that display little similarity in morphology and gene content would share a related mechanism for extrusion from the host cell.

The number and extent of elaborate modifications caused by SIRV2 on the host cell result in a radically transformed cell which can hardly be contemplated as the archaeon *Sulfolobus*. The whole infected cell rather appears to be converted into a complex viral factory, conceptually identical to those built by some eukaryotic viruses inside infected cells. In such cases, the structures of the factory are enclosed by a membrane to exclude cellular organelles. Ribosomes are however present and the factory is dedicated to viral genome replication and virion assembly (68, 69). The eukaryotic viral factories were suggested to constitute the genuine identity of viruses (70), which thus might be considered as a specific type of living organisms(70, 71). A weakness of this concept was the failure to observe viral factories in cells from other domains. SIRV2, as described above, constitutes a first example archaeal virus producing a transient viral factory, consisting of the whole transformed infected cell.

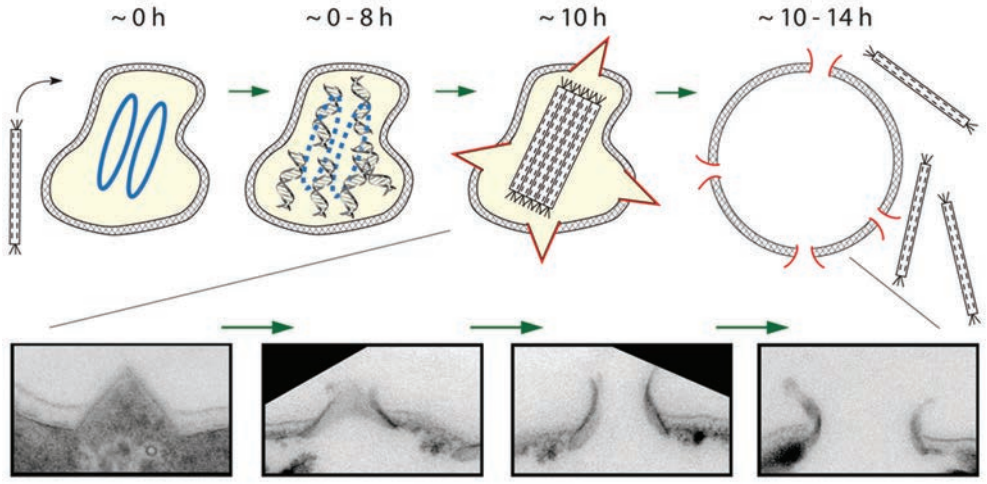


Figure 5. Schematic representation of the major stages of SIRV2 infection cycle in the *Sulfolobus* host cell. Times after infection are indicated in hours. At 0 h, 2 chromosomes of *Sulfolobus* are shown in blue. Later between 0 and 8 h, they degrade concomitantly with viral DNA synthesis (gray helices). At 10 h, the VAPs (shown in red) and the intracellular clusters of assembled virions are shown. Finally, at time points between 10 and 14 h, the VAPs open (remains of VAPs shown in red), the cell lyses, and the virions are extruded. The gradual opening out of VAPs (at time points between 10 and 14 h) is illustrated in more details with fragments from the TEM of thin sections.

Our results show that lytic cycles may be more common for crenarchaeal viruses than previously assumed (53), and that lytic properties may have been overlooked in other crenarchaeal viruses. The original notion that the carrier state host-virus relationship is dominant in crenarchaea was consistent with the suggestion that this life style would provide a durable intracellular refuge for the virus population in the harsh physico-chemical conditions at which cultured representatives of the Crenarchaeota thrive (T ~60-90°C, pH ~3.0 – 6.0). In contrast, our findings imply that virus particles can persist in such extreme ecosystems long enough to encounter a new host cell. The SIRV2 virions are well adapted to harsh environments, being extremely stable in various solvents and other inhospitable conditions (28, 59), and almost as stable at 80°C (unpublished data) as phages of mesophilic bacteria are at 37°C (72). Geothermal environments are extremely heterogeneous, due to a variety of gradients, dynamic movements and changes over time, and viruses may be trapped and preserved for long time periods in different environmental refuges in the absence of potential hosts. Finally, the fact that virus particles are apparently able to travel across the globe (73-75) also suggests that they are robust to variable environmental conditions, and display stability over very extended time periods in a variety of biotopes.

Experimental procedures

Virus, host strains and cultures

Virus stocks were prepared by PEG precipitation of the virions from the culture supernatants, followed by concentration and purification on Cesium chloride density gradients, as described in (76). The cells of *S. islandicus* LAL14/1 were grown in shaken 50 mL flasks at 78°C, pH 3.0, in rich medium as previously described (59). Colonies were obtained on Gelrite plates as described in (59). To infect cultures, the appropriate volume of virus solution was dialyzed against medium or water on 0.025 µm or 0.05 µm MF membrane filters (Millipore) and directly added to the liquid cultures during the early exponential phase (OD_{600nm} between 0.09 and 0.25). For the time-course experiments (growth kinetics, flow cytometry, dot blot hybridization, $DiBac_4(3)$ staining), all conditions were tested in triplicates. Six identical 50 mL cultures were launched by dilution of a same preculture. After overnight growth, SIRV2 was added to 3 of them at the appropriate m.o.i..

Titrations, OD and fluorescence measurements

To determine CFU values, culture samples were submitted to serial dilutions and 5 µL of each dilution were spotted on plates. After incubation, the colonies were counted in the last or last two positive spots. To determine the PFU values, the same method was used, except that 5 µL of each dilution were spotted on a fresh cell lawn. When required, the cells were removed by centrifugation before spotting. The cell lawns were prepared as described in (31), using a soft Gelrite overlay. After incubation, single plaques were counted in the last or last two positive spots. ODs were measured in 96-well round-bottomed culture microplates (TPP) in a Multiskan Ascent microplate photometer (Thermo LabSystems) at 595 nm, using 200 µL of culture.

Flow cytometry

Sampling and flow cytometry were performed as described in (63): the cells were fixed in 70% (v/v) ethanol and the intracellular DNA was stained with mithramycin A and Etd bromide. Samples were analyzed in a A40 Analyzer (Apogee, 25 mW solid-state laser, 405 nm wavelength). *S. islandicus* cell cycle was characterized preliminarily to the study of infected cultures (Fig. S2 and S3). For the study of infected cultures, a high m.o.i. was used (~10-15) to obtain as synchronous an infection as possible. At each time point, OD_{595nm} was measured and CFU titres were determined to control that the usual growth pattern was obtained. The distinct cell populations were identified based on the cell size distributions, DNA content distributions and 2D diagrams of cell size and DNA content. The data were gated, and several contours tested, to ensure the robustness of the analysis and of the identified cell populations. The proportion of empty cells over time was computed by gating the 2D-diagrams, similar to what is shown in Fig. S6. In Fig. 3A, the total

percentage of chromosome-less cells in the culture is shown. In Fig. 3B, for the curve related to chromosome-less cells, the difference between the values at time points T and T-1 is plotted, reflecting the production of empty cells between two successive time points.

Dot blot hybridization

Cells were washed once in cold medium, pelleted by low-speed centrifugation and stored at -20°C until further use. Cell pellets were resuspended in Tris-acetate pH 6.0 pre-cooled at 4°C. The suspension volume was adjusted for cell concentration to be roughly constant in all samples, on the basis of OD measurements. 4 µL of each sample were spotted on Hybond-N+ nylon membranes (Amersham Biosciences). The membranes were further prepared as for colony hybridization (77). The probes were generated by PCR. A ~240bp SIRV2 DNA fragment was generated using primer combination [5'-ACATGAAAAGTTAGAGAGATACAAACG(3872) 5'-TGGTTACCACTAGCTTCGCTAC(4086)] and a ~1300bp fragment of the 16S rDNA of *S. islandicus* LAL14/1 was generated by using primers 8aF and 1512uR (78). The probes were [^{32}P]-end-labelled with EasyTide [α-^{32}P]-dATP (PerkinElmer) using a random-primed DNA labeling Kit (Roche Applied Science), according to manufacturer's instructions. All hybridization steps were performed at 65°C in prewarmed solutions. After a minimum of 2h prehybridization followed by overnight hybridization, both performed in Church Buffer (7% SDS (w/v), 0.5M sodium phosphate pH 7.2, 1mM EDTA), membranes were washed 2 times 15 min in a solution of 2X SSC- 0.1% SDS, and 2 times 15 min in a solution of 0.5X SSC - 0.1% SDS. Membranes were exposed on a GP Phosphor Screen (Amersham Biosciences). The screen was scanned in a Molecular Dynamics Storm 860 (Amersham Biosciences). The images were analyzed with the ImageQuantTL software (Amersham Biosciences). After contrast and brightness adjustment, the radioactivity of each spot on the membranes was quantified, using the background removal option (local average). The images of Fig. 3D and 3E were processed with ImageQuantTL for contrast and brightness adjustment and with ImageJ software (http://rsbweb.nih.gov/ij/) for background removal, using the "sliding paraboloid" function.

Transmission Electron Microscopy

Cells were pelleted by low speed centrifugation. The cell pellet was fixed overnight at 4°C with 2.5% (w/v) glutaraldehyde in 20 mM Tris-acetate pH 6 buffer, post-fixed for 1h in 1% (w/v) OsO_4, and dehydrated in a graded series of ethanol solutions (25% to 100%). The cells were embedded in an epoxy resin and polymerized at 60°C for 48 h. Ultrathin sections (~60 nm) were cut on a Leica Ultratuc UCT microtome and deposited on carbon-coated copper grids. They were stained for 30 min with 2% (w/v) Uranyl acetate and for 5 minutes with 2.5% (w/v) Lead citrate. The grids were examined under a JEOL JEM-1010 transmission electron microscope (Tokyo, Japan)

operated at 80 kV. Images were recorded using an Eloise Keen View camera and the Analysis Pro software version 3.1 (Eloïse SARL, Roissy, France).

Scanning Electron Microscopy

Cells were pelleted by low-speed centrifugation and fixed overnight at 4 ˚C with 2.5% (w:v) glutaraldehyde in 0,1 M Tris buffer pH 6. Cells were adsorbed to polylysine-coated coverslips and post-fixed 1 hour in 1% osmium tetroxide solution. Samples were dehydrated through a graded series of ethanol dilutions (25% to 100%) and critical point dried using a Leica EM CPD030 device. The dried coverslips were sputtered with 15 nm gold palladium in a GATAN Ion Beam Coater prior to examination with a JOEL JSM-6700F field emission Scanning Electron Microscope (Tokyo, Japan) operated at 5 kV. Images were acquired from the upper SE detector (SEI).

Acknowledgements

We are indebted to Dr. Soizick Lucas-Staat for great help in virus preparation, to Léa Lepelletier for assistance in optical microscopy and to Professor Ryland F. Young for enlightening discussions. This work was supported by Ecole Nationale des Ponts et Chaussées, and Agence Nationale de la Recherche, France, by the Swedish Research Council and by a grant (Uppsala Microbiomics Center; www.microbiomics.se) from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas).

Supplemental text

Controls for the noncontamination of the host strain

Specific controls were designed to verify that the *S. islandicus* LAL14/1 strain stock was not contaminated by a distinct SIRV2 insensitive strain. It could have explained the growth of resistant cells in *S. islandicus* LAL14/1 cultures infected by SIRV2. The 16S rRNA gene of a SIRV2 resistant clone isolated from an infected culture was sequenced and verified to be 100% identical to the sequence obtained from an uninfected culture, over the 1,372 amplified bp. The Eco RI restriction patterns of genomes of 4 resistant clones isolated from infected cultures were verified to be identical to the pattern obtained using some uninfected *S. islandicus* LAL14/1 cells; as a reference, other closely related *Sulfolobus* strains were included and always produced distinguishable patterns. As a last control, the reference strain *S. islandicus* LAL14/1 was repurified: 4 clones were randomly selected after 5 successive streaking steps. The OD 595 nm of SIRV2 infected minicultures of these clones was monitored over time. It was verified that a

growth was detectable after a few days, indicating that repurifying the strain had not eliminated the presence/appearance of SIRV2 resistant cells. From these 3 controls, it was concluded that the resistant cells are variants of *S. islandicus* LAL14/1 strain.

Characterization of the cell population growing in the presence of SIRV2

To discriminate between the possibilities that cells growing in the presence of SIRV2 are either virus-resistant or contain virus in a carrier state, 23 clones were colony-purified from each of three SIRV2-infected *S. islandicus* LAL14/1 cultures and from uninfected control culture, resulting in total of 92 clones. To assess the production of infectious virions, supernatants from minicultures were spotted on a fresh cell lawn. No growth inhibition zones were detected, indicating that none of the 92 clones produced a detectable amount of virus, disfavoring the carrier state hypothesis. The sensitivity of the same 92 clones to SIRV2 infection was tested by growth in minicultures in the presence of SIRV2 added at a low moi (10^{-3}). In late exponential phase, the OD was measured and the virus concentration in the supernatants assessed by a spot test. As expected, the control clones from the uninfected culture were highly sensitive to SIRV2 infection: growth was strongly inhibited by the virus (Fig. S1A) and the spot tests were positive (Fig. S1B, clearings in the left column), each clone thus having produced a high amount of infectious virions. In contrast, the clones from the Inf1–3 cultures displayed no growth reduction (Fig. S1A) and the spot tests did not reveal an increase in virus concentration in the supernatants (Fig. S1B right columns). Thus, the cells able to grow in the presence of SIRV2 were totally resistant to reinfection and did not produce new infectious virions.

One-step growth curve

A 1-step growth experiment was performed to confirm the latent period and estimate the burst size. A low m.o.i. was used, such that the effect of infection by a single virus could be monitored. Virus release was shown to begin at 8–10 h after infection (Fig. S7), consistent with all other data. Although the precision of the PFU assay is limited, it clearly showed that virus production was significantly reduced after 13–15 h.

Kinetics of cell mortality after viral infection

Infected cells might remain metabolically active for some time despite a degraded genome. To determine when cell death exactly occurred, the membrane potential-sensitive dye DiBAC4(3) was used. The dye penetrates depolarized cells and binds to intracellular proteins and lipids, causing enhanced fluorescence. The ratio of fluorescence/OD in stained culture samples of constant volume was measured over time in uninfected and infected (m.o.i. 3) cultures (Fig. S8). The high ratios obtained at the initial time points in both controls and infected cultures were due to the low initial OD, and not to a high proportion of depolarized cells (Fig. S8). In infected

cultures only, a moderate peak was observed at 11 h (Fig. S8), corresponding to the middle of the virus release period as assessed by hybridization (Fig. S8). Surprisingly, the higher ratios in infected cultures were transitory, despite the persistence of chromosome-less cells as shown by flow cytometry. It is possible that intracellular macromolecules were degraded shortly after virion release, such that the fluorescence signal emitted by dead cells decreased strongly. The results suggest that the cells died concomitant with virus release, and that damage was not restricted to chromosome degradation.

Supplemental experimental procedures

Growth of isolated clones

For their growth in minicultures, 24-well flat-bottomed microplates (Corning-Costar) were used. The microplates were incubated at 78 °C and shaken at 150 rpm (Infors dry incubator) in a humid atmosphere.

Primers and PCRs

Six primer pairs specific for SIRV2 genome were used for the detection, by PCR, of SIRV2 genome potentially present in cells of isolated clones (each time, the 5 position in SIRV2 genome is indicated): [5′-ACATGAAAAGTTAGAGAGATACAAACG (3872); 5′-TGGTTACCACTAGCTTCGCTAC (4086)] was used to generate a probe for the dot blot hybridization experiments; [5′-GAGACGCAAAAGCAAGTCAAA (1707); 5′-TCCACTTCCCATATCTCCGA (2221)], [5′-AAAACAAATTGGAATAACGCC (10071); 5′-CTGCAATTTCATGTGGATCA (10616)], [5′-GAAATGCGGAAAGCGAGTTA (22972); 5′-TGGAGTTCTTGTGGAATCTGG (23482)], [5′-ATCTCCTGGCAAAATGGGA (27018); 5′-GCGGGCTTTATTGGAACTG (27479)], [5′-CTTCCATCTTGAACCCCTGA (32635); 5′-GAACTTGCGGAATTTTACGGA (33183)].

One-step growth curve

The virus was added to 1mL *S. islandicus* LAL14/1 culture in early exponential phase at a low m.o.i. (10^{-3}). After a 1.5h–2h incubation at 78 °C for adsorption of almost all viruses, the preparation was diluted in 200 mL prewarmed medium to prevent the occurrence of new adsorptions in the course of the experiment. The obtained diluted culture was further incubated at 78 °C. PFU titres of samples were regularly determined. The burst size was estimated by averaging the PFU titers from time points 0h–9h (before release) and from time points 13h–17h (after virus release), and then by dividing the latter value by the former one.

DiBAC4(3) staining

DiBAC4(3) was selected based on Beck and Huber (79) and purchased from Molecular Probes. It was verified by fluorescence microscopy (video inverted microscope Leica DM IRE2 HC FLUO, Metamorph software) that the growth medium was suitable for the probe, not causing any background nor suppressing the cell fluorescence; it was also verified that fluorescence from dead cells was significantly stronger than that of living cells, and that unstained cells were not autofluorescent. Afterwards, the changes in cell mortality were assessed by measuring the ratio fluorescence/OD 600nm of constant-volume culture samples. ODs were measured at 600 nm in the Infinite M200 microplate reader (Tecan). Fluorescence was measured in the same apparatus, the excitation wavelength was 490 nm, the emission wavelength 550 nm, and the manual gain was set to 120. Culture samples of 1.6 mL volume were allowed to cool down to room temperature for 10 min 1µL of DiBAC4 in ethanol (0.5 mg/mL) was added to each and the samples were incubated for 5 min in the dark. 200 µL of each sample were transferred into wells of a 96-well round-bottomed culture microplate (TPP) for OD and fluorescence measurements (see above). The ratio fluorescence/OD 600nm is strongly OD-dependent (Fig. S7); but for a given OD, the ratio is higher for higher rates of dead cells, so that the method is convenient.

Supplemental figures

Figures S2 to S6. See online figure at http://www.pnas.org/content/suppl/2009/06/19/0901238106.DCSupplemental/0901238106SI.pdf

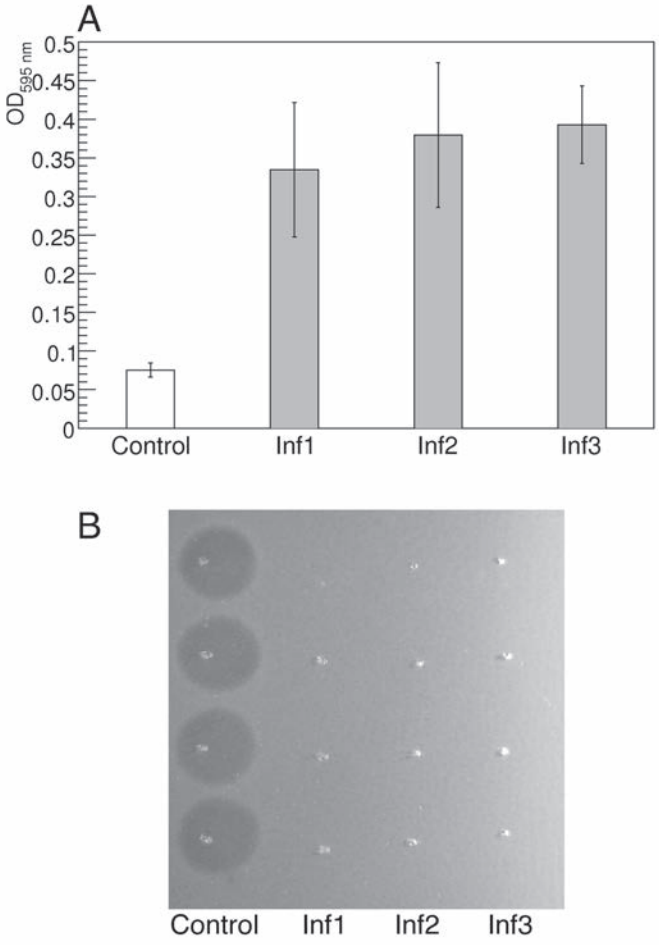


Figure S1. Sensitivity to SIRV2 infection of clones isolated from infected cultures. To obtain the clones, 23 colonies were isolated from an uninfected culture (control) and from each of 3 heavily infected cultures (Inf1, Inf2, and Inf3), resulting in a total of 92 clones. (**A**) OD 595 nm measurements after a 3-day growth of the clones in the presence of SIRV2 added at a m.o.i. of ~10^{-3}. Averages corresponding to the 23 clones 1 SD are shown. (**B**) Spot test on a lawn of *S. islandicus* LAL14/1 cells of 20-fold diluted supernatants obtained after growth of the clones in presence of SIRV2 (A). A clearing indicates the presence of infectious virions in the supernatant. The results for 4 of 23 clones are shown for each culture; identical results were obtained for all clones.

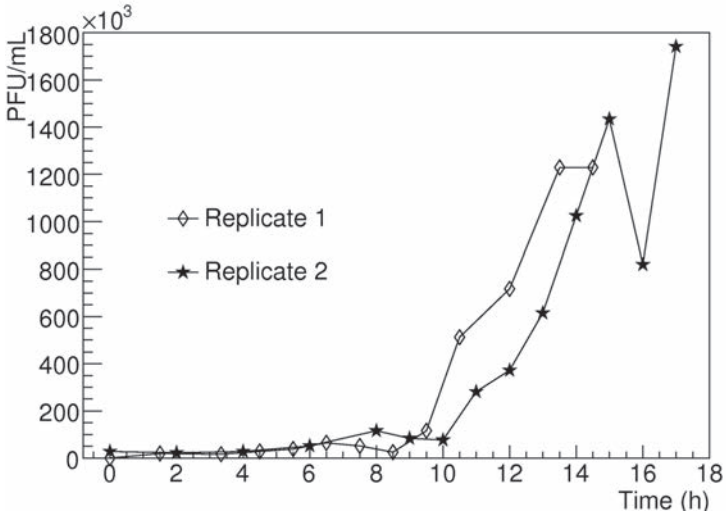


Figure S7. One-step growth curve of SIRV2 infection of *S. islandicus* LAL 14/1 cells. The virus was added at a m.o.i. of ~10^{-3}, and the experiment was repeated twice.

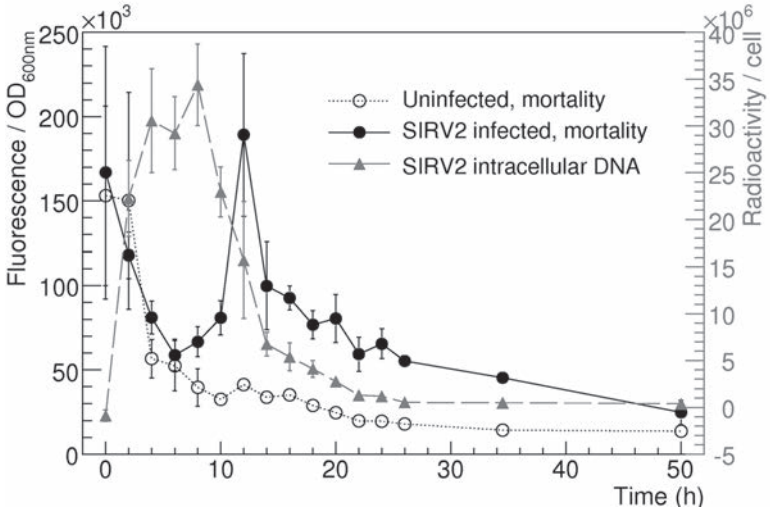


Figure S8. Impact of SIRV2 infection on cell mortality. Comparison of infected and noninfected cultures over time. The ratios of fluorescence/OD600 nm for infected (filled circles, continuous line) and uninfected (empty circles, dotted line) cultures reflect the changes in cell mortality. For the infected culture, the radioactivity/cell (filled triangles, discontinuous line, right axis) indicative of SIRV2 intracellular DNA amount was also plotted. Infected (moi ~5) and uninfected cultures were launched in triplicate, and averages of the replicates 1 SD are shown. In infected replicates, the virus was added at time point 0 h.

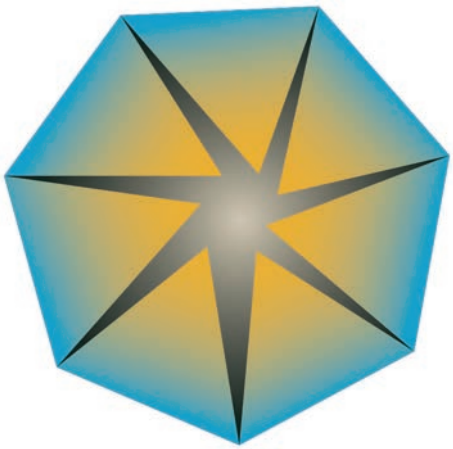
Chapter 2

Composition of the SIRV2 virion egress structure

Tessa E. F. Quax, Mart Krupovič, Soizick Lucas, Patrick Forterre, David Prangishvili

'The Sulfolobus rod-shaped virus 2 encodes a prominent structural component of the unique virion release system in Archaea'

Virology. 2010 Aug 15. 404:1-4



Abstract

Recently a unique mechanism of virion release was discovered in Archaea, different from lysis and egress systems of bacterial and eukaryotic viruses. It involves formation of pyramidal structures on the host cell surface that rupture the S-layer and by opening outwards, create apertures through which mature virions escape the cell. Here we present results of a protein analysis of *Sulfolobus islandicus* cells infected with the rudivirus SIRV2, which enable us to postulate SIRV2-encoded protein P98 as the major constituent of these exceptional cellular ultrastructures.

Introduction

Viruses of the Archaea, one of the three domains of life, are morphologically distinct from viruses of the Bacteria and Eukarya (48). In addition, they carry genomes where more than 90% of the genes are without predictable functions and detectable homologues in other viruses or cellular life forms (26). The lack of knowledge about gene functions can be partly attributed to very limited knowledge about life cycles of archaeal viruses and specific features of their interactions with the host cells. That these interactions have specific features has been recently demonstrated by revealing the novel mechanism of virion release employed by the two hyperthermophilic archaeal viruses, *Sulfolobus* rod-shaped virus 2, SIRV2 (83), and *Sulfolobus* turreted icosahedral virus, STIV (40).

Infection with the viruses SIRV2 and STIV leads to the formation of pyramid-shaped ultrastructures, virus-associated pyramids (VAPs), on the *Sulfolobus* host cell. They are localized on the cell envelope, and point outwards, perforating the S- layer ((40, 83), Fig. 1B). Such projections have not been documented for any bacterial or eukaryal virus-host system. Moreover, the pyramids appear to have seven-fold rotational symmetry, representing a peculiar case in the living world.

At later stages of infection, the VAPs open and the virions that have been preassembled in the cytoplasm are released through the apertures. The lysed cells persist in the form of empty spheres which, apart from the apertures, appear to be intact, with both the S-layer and the membrane being visible on electron micrographs ((40, 83), Fig. 1C). The apertures are delimited by polygonal shaped ring structures, which sometimes detach from the cell envelope (83). The observations indicated that the VAPs represent a distinct proteinatious structure. The large number of the VAPs per cell, a dozen or more (83), suggested a high abundance of their protein constituents within the membrane fraction of infected cells. We aimed at their identification by analyzing protein content of SIRV2-infected cells of *Sulfolobus islandicus,* as a function of cellular fraction and time.

Results and Discussion

Growth and infection of *S. islandicus* with SIRV2 were performed as described by Bize et al. (83), and the previous characterization of the viral life cycle guided the selection of time points for protein analysis and electron microscopy. Samples were collected at following time points post infection (p. i.): 0 h (start of infection), 3 h, 7 h (prior to virion release), 10 h (middle of virion release period) and 26 h (after virion release). At each time point the uninfected cell culture was used as a control. After mechanical disruption of cells by French press, three fractions were collected for

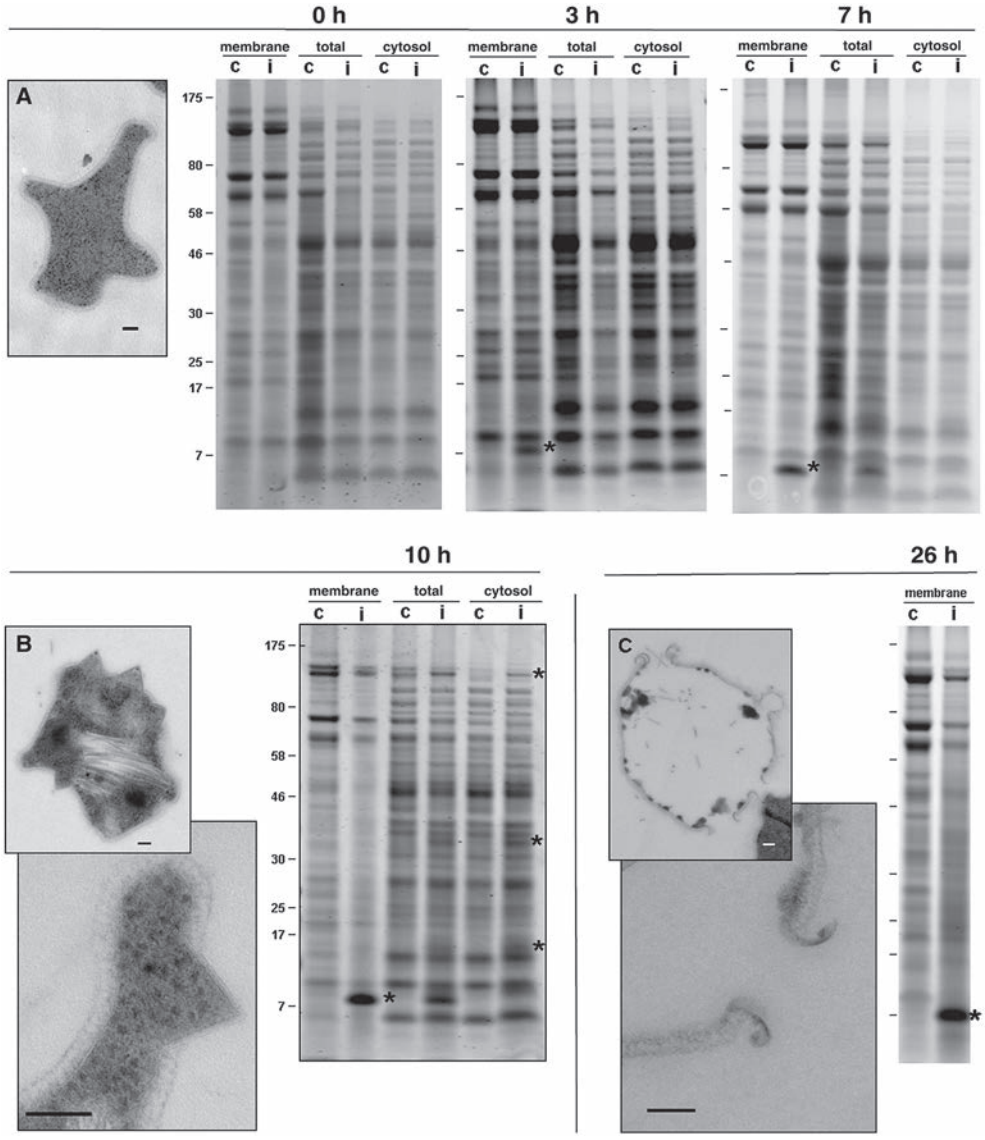


Figure 1. Coomassie-stained SDS-PAGE gels of different protein fractions of *S. islandicus* cells infected with the virus SIRV2. Proteins from the three fractions, ("membrane", "total", and "cytotosol") from uninfected control cells ("c") and infected cells ("i") were prepared as described in the text, and analyzed at time points post infection indicated on the top of each gel. Positions of proteins with known molecular masses (in kDa) are indicated with bars. Asterisks highlight protein bands in membrane and cytosol fractions which appear as a result of infection. Electron micrographs of thin sections of infected cells (**A**, **B**, and **C**) are displayed at corresponding time points. Bars, 100 nm.

protein analysis: the total cell lysate, the membrane fraction and the cytosol fraction. The latter two fractions were separated from each other by high-speed centrifugation at 100.000 x g, as described by Albers *et al*. (84). Proteins in each fraction were solubilized by incubation for 2 h at 37 ˚C in the presence of 1% n-Dodecyl β-D-maltoside. Following heat denaturation (90°C, 10 min, 0.5% SDS, 1.25% β-mercaptoethanol), proteins were separated by electrophoresis on 4-12% Bis-Tris polyacrylamide gradient gel (Invitrogen™), with 2-(N-morpholino)ethanesulfonicacid (MES) SDS running buffer (Invitrogen™) and visualized with Coomassie-based Instant Blue™ staining (Expedeon) according to the manufacturer's protocol. The SDS-PAGE gels are shown in Fig. 1, along with transmission electron micrographs of representative cells.

At time point 0, no difference was observed in protein patterns of infected and uninfected control cells (Fig. 1). However, at later time points, additional protein bands appeared in all fractions of infected cells. The most dramatic change was observed in the membrane fraction, where from 3 h p.i. onwards, an abundant protein with an apparent molecular mass of 10 kDa was detected in infected cells (Fig. 1). The relative intensity of this band, in comparison with the total protein content of the membrane fraction, increased from 0 at the time of infection to about 1.0 at 3 h p.i. and gradually reached about 2.0 at 26 h p.i., when all cells were perforated. This protein was identified as a product of SIRV2-ORF98 (NCBI RefSeq ID: NP_666583) by trypsin digestion and Matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF)-mass spectrometry (MS) as well as MS/MS analysis. MS was performed as described by Stingl et al. (85), and protein identifications were obtained using a combination of MS and MS/MS data from a 4800 Proteomics Analyzer (Applied Biosystem, USA) and the NCBI (20100119) protein database.

The protein P98 was not observed in the cytosol fraction. However, on an SDS-PAGE gel of this fraction, four protein bands appeared from 10 h p.i. onwards, in SIRV2 infected cells. The four proteins had estimated molecular masses of ~ 124, 35, 16, and 15 kDa (Fig. 1) and were analyzed by MALDI-TOF-MS and MS/MS. The 35 and 16 kDa proteins were identified as host proteins: The thermosome and the S-adenosyl-L-methionine-dependent methyltransfease of *S. islandicus* (NCBI RefSeq ID: YP_002829404 and YP_002829621, respectively). The 15 and 124 kDa proteins were SIRV2-encoded: The product of ORF131b, without putative function, but conserved in all rudiviruses, and the product of ORF1070, a minor structural protein of the virion (NCBI RefSeq ID: NP_666551 and NP_666572, respectively).

Edman degradation (86) was used to determine the N-terminal sequence of the protein P98 at 10 h p.i.. This confirmed the identity of the protein, and also revealed that P98 starts with AITLLE and therefore lacks the initiator methionine, as is often the case for archaeal proteins (87). We also found that P98 at 10 h p.i. is not glycosylated, as determined using the Pro-Q® Emerald 300 Glycoprotein Gel Stain Kit (Invitrogen) (Supplementary material). By contrast, virion proteins of the *Rudiviridae* family, to which SIRV2 belongs, are known to be extensively glycosylated (58).

In order to establish whether P98 is exposed on the surface of SIRV2-infected cells, its accessibility to externally added proteases was tested. Infected and uninfected cells were harvested at 10 h p.i. and washed twice with 100 mM Tris-HCl, pH 7.5, resuspended in the same buffer and incubated with α-chymotrypsin from bovine pancreas (Sigma-Aldrich® ; 50 ng per 1.5 x 10^{10} cells) for 3 h at 37 ˚C. After incubation, the protease was inactivated by heating for 10 min at 80 ˚C. The cells were washed twice with 20 mM bis-Tris propane, pH 6, and mechanically disrupted by French press. The cellular proteins were analyzed by SDS-PAGE (Fig. 2A). As a result of the protease treatment, P98 was no longer detectable among proteins of infected cells (Fig. 2A). In a control experiment, where SIRV2-infected cells were mechanically broken prior to protease treatment, all proteins were completely digested by α-chymotrypsin (Fig. 2B). These results demonstrate that intracellular content of the infected cells was not accessible to α-chymotrypsin in the course of the "protein shaving" experiment, and that P98 was degraded apparently due to its exposure on the cell surface.

Based on the data reported above, it appears highly likely that P98 of SIRV2 is the major constituent of the virus-associated pyramids. Firstly, it is the only protein to appear specifically in the membrane fraction of infected cells, where its accumulation throughout the viral cycle correlates well with the emergence of the VAPs, as can be judged from the TEM analysis (Fig. 1). Secondly, the protein is exposed on the cell surface and is not covered by the S-layer, in accordance with the data on rupturing of the S-layer by the VAPs (Fig. 1B).

The results which postulate SIRV2-P98 as a building block of the VAPs are in line with comparative genomics analysis. As mentioned above, VAPs highly similar in size and shape to those produced by SIRV2, were also observed in STIV infected cells (40). The comparative genome analysis of the two viruses, highly different in morphological properties, revealed three pairs of proteins encoded by both viruses. Only one pair, SIRV2-P98/ STIV-C92 is exclusive for the *Rudiviridae* (SIRV1/2 (31, 60), *Stygiolobus* rod-shaped virus, SRV (58)) and STIV (67). The results imply that STIV-C92, a homolog of SIRV2-P98, would be the protein constituent of the VAPs produced by STIV. Fig. 3 depicts a sequence alignment of SIRV2-P98 with homologous proteins from the rudiviruses (SIRV1, SRV) and the virus STIV. The higher similarity of the SIRV2 protein to its homologue in STIV than to that of the related rudivirus SRV (Fig. 3), suggests that the gene could have been transferred horizontally to STIV prior to the divergence of *Sulfolobus* and *Stygiolobus* rudiviruses. Sequence analysis of P98 revealed the presence of an N-terminal transmembrane domain (TMD; position 10–30; predicted with TMHMM (88) with probability cut-off of 0.7). P98 is predicted to be a Type II membrane protein, with the major part facing the extracellular side of the cytoplasmic membrane (probability of 0.92), by the Signal-Pred (89), a hidden Markov model-based tool trained specifically on archaeal protein sequences. These predictions are in accordance with the results obtained during the protease treatment of the surface-exposed proteins of SIRV2-infected cells (Fig. 2A).

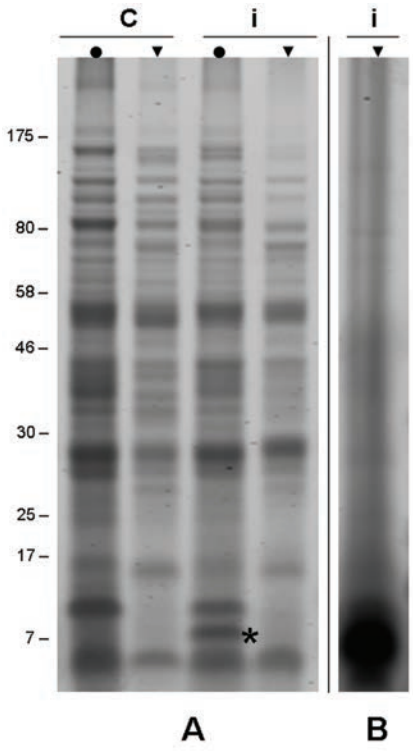


Figure 2. Coomassie-stained SDS-PAGE gels of proteins from protease treated cells of *S.islandicus*. (**A**) Proteins from uninfected cells ("c") and SIRV2-infected cells ("i") were separated by SDS-PAGE after incubation of intact cells with α-chymotrypsin (▼) or, as a control, after incubation with the assay mixture without the enzyme (●). (**B**) Proteins from SIRV2-infected cells were incubated with α-chymotrypsin after mechanical disruption of cells. Conditions of protease treatment in (A) and (B) were identical. Positions of proteins with known molecular masses (in kDa) are indicated with bars. An asterisk highlights the position of P98.

The virion release mechanisms exploited by SIRV2 and STIV apparently are not universal for hyperthermophilic archaeal viruses. Except for STIV and the *Rudiviridae* (SIRV1/2, SRV), no other archaeal virus carries a homologue of the SIRV2-ORF98. Moreover, even in the family *Rudiviridae*, one out of the four known species, the *Acidianus* rod-shaped virus 1, ARV (55), lacks a homologue of the gene. A homologous gene is also absent from the genome of STIV2, a close relative of STIV(90). It appears that in archaeal viruses morphogenetic and egress systems could be evolving independently. This situation resembles the one observed for bacterial viruses, where morphologically related viruses exploit different lysis systems, and vice versa, morphologically distinct viruses exploit similar lysis systems (37, 91).

The identification of the major player in the VAP-based archaeal virus release is crucial for initiating studies on the molecular mechanisms of this unique process. These studies promise to shed light on unknown aspects of membrane biology and basic questions regarding virus-host relationships.

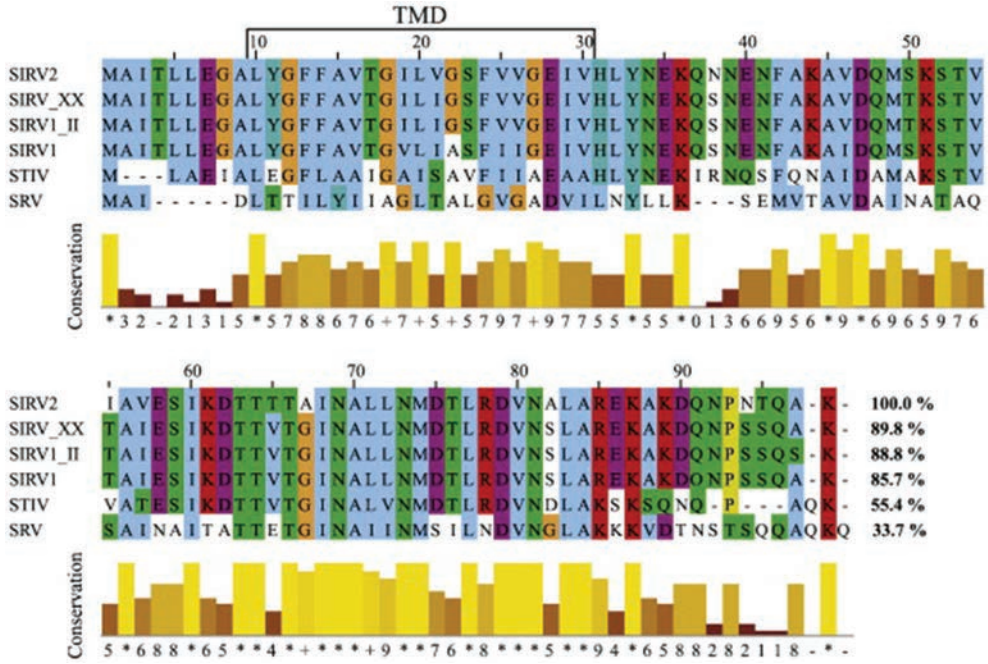


Figure 3. Sequence analysis of the P98-like proteins. The multiple sequence alignment was generated using PROMALS3D (92), manually edited and visualized using JalView (93). The alignment is coloured according to the standard ClustalX colouring scheme. The sequence conservation at each position is indicated at the bottom of the figure (the height of each bar is proportional to the conservation of physico-chemical properties for each column of the alignment). The transmembrane domain (TMD) is indicated above the alignment, while the pairwise identity values of the aligned sequences to the SIRV2-P98 protein are indicated on the right of the alignment. The unequal distribution of conservation along the sequence of P98-like proteins is indicated under the multiple alignment for the P98 homologue of SRV. Virus name abbreviations and protein accession numbers: SIRV2, *Sulfolobus islandicus* rod-shaped virus 2 (NP_666583); SIRV_XX, *Sulfolobus islandicus* rudivirus 1 variant XX (CAG38861); SIRV1_II, *Sulfolobus islandicus* rod-shaped virus 1 variant II (CAG28292); SIRV1, *Sulfolobus islandicus* rod-shaped virus 1 (NP_666630); STIV, *Sulfolobus* turreted icosahedral virus (YP_024995); SRV, *Stygiolobus* rod-shaped virus (CAQ58475).

Acknowledgements

We are thankful to Ariane Bize and Marie-Christine Prevost for providing electron micrographs, and Sonja Albers and Peter Redder for helpful discussions. This work was supported by the Agence Nationale de la Recherche (Programme Blanc) under the grant ANR-09-BLAN-0288.01. T. Q. was supported by an allocation from the Ministère de l'enseignement superieur et de la recherché of France and M. K. was supported by FEMS Research Grant.

Supplemental figure

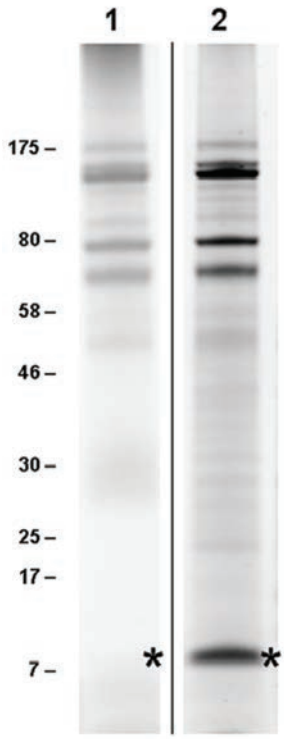


Figure S1. SDS-PAGE gel of membrane-associated proteins of *S. islandicus* cells infected with the virus SIRV2, 10 hours p.i. After electrophoresis, glycosylated proteins in the gel were selectively stained and visualized under UV light using the Pro-Q® Emerald 300 Glycoprotein Gel Stain Kit (Invitrogen) (**1**); later the same gel was stained by Coomassie, to visualize all proteins (**2**). Positions of proteins with known molecular masses (in kDa) are indicated with bars.' Asterisks highlight the position of P98.

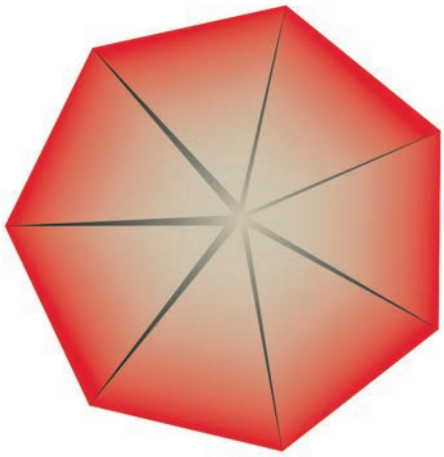
Chapter 3

The simple and elegant design of an archaeal virion egress structure

Tessa E F Quax, Soizick Lucas, Julia Reimann, Gerard Pehau-Arnaudet
Marie-Christine Prevost, Patrick Forterre, Sonja-Verena Albers, David Prangishvili.

'Simple and elegant design of a virion egress structure in Archaea'

Proceedings of the National Academy of Sciences USA. 2011 Feb 22. 108(8): 3354-3359



Abstract

Some viruses of Archaea employ an unusual egress mechanism that involves the formation of virus-associated pyramids (VAPs) on the host cell surface. At the end of the infection cycle, these structures open outwards and create apertures through which mature virions escape from the cell. In the present study, we describe in detail the structure and composition of VAPs formed by the *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) in cells of its hyperthermophilic archaeal host. It was shown that the VAP is a stable and autonomous assembly that can be isolated from membranes of infected cells and purified without affecting its structure. The purified VAPs were heterogeneous in size, which reflects the dynamics of VAP development in a population of infected cells. However, they had a uniform geometry and consisted of seven isosceles triangular faces, which form a baseless pyramid. Biochemical and immuno-electron microscopic analyzes revealed that the 10-kDa P98 protein encoded by the virus SIRV2 is the sole component of the VAP. By heterologous expression of the SIRV2-P98 gene, the VAPs were produced in *Sulfolobus acidocaldarius* and *Escherichia coli*. The results confirmed that P98 is the only constituent of the VAP and demonstrated that no other viral protein is involved in the assembly of pyramids. P98 was able to produce stable structures under conditions that ranged from moderate to extremely high temperatures (80 °C) and from neutral to extremely acidic pH conditions (pH 2), which is another remarkable property of this exceptional viral protein.

Introduction

The vast majority of known archaeal viruses carry double-stranded (ds) DNA genomes and differ morphologically from dsDNA viruses of the two other domains of life, Bacteria and Eukarya; an exception are archaeal head-and-tail viruses, related to the bacterial *Caudivirales* (48). Moreover, the genomes of the majority of archaeal viruses are also unique. The functions of more than 90% of putative genes cannot be identified, due to the lack of homologues in the extant databases (48) and limited knowledge of the biology of archaeal viruses. Archaeal viral cycles can have unusual features, which was recently demonstrated by the discovery of a unique virion release mechanism exploited by the *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) (83) and the *Sulfolobus* turreted icosahedral virus (STIV) (40).

Among archaeal viruses, SIRV2 and STIV are the best studied with respect to host cell interactions. It was demonstrated that both are lytic viruses and that SIRV2 causes massive degradation of the host chromosome. Virion assembly takes place in the cytoplasm and coincides with the appearance of numerous prominent virus-associated pyramids (VAPs) on the host cell surface, which point outwards and rupture the S-layer. Shortly after their formation, VAPs open to the outside and create large apertures through which the virions escape from the cell (83).

The discovery of this unique virion release system raised questions regarding the nature of the VAP and mechanism of its formation. Here, we report the isolation and purification of the VAPs formed by SIRV2 in *S. islandicus.* We provide a detailed characterisation of the morphology and constituents of the VAP and show that it is an individual stable structure formed by multiple copies of a single 10 kDa virus-encoded protein. Moreover, by heterologous expression of the corresponding viral gene, we were able to produce VAPs in *Sulfolobus acidocaldarius* and *Escherichia coli.*

Results

VAP isolation

The shape of the pyramidal egress structures, VAPs, which were reported to be found on the surface of SIRV2-infected *S. islandicus* cells (83), suggested that they might represent individual structures. To test this hypothesis, we analyzed the membrane fraction of infected cells that were disrupted using a French press. By transmission electron microscopy (TEM) we observed numerous individual particles which had pyramidal shapes and heptagonal bases (Fig. S1). The resemblance of their shape to that of VAPs observed *in vivo* (Fig. S2), supported the notion that VAPs represent autonomous structures. We isolated the particles using antibodies against protein 98 (P98) encoded by SIRV2, which we had previously postulated to be a constituent of

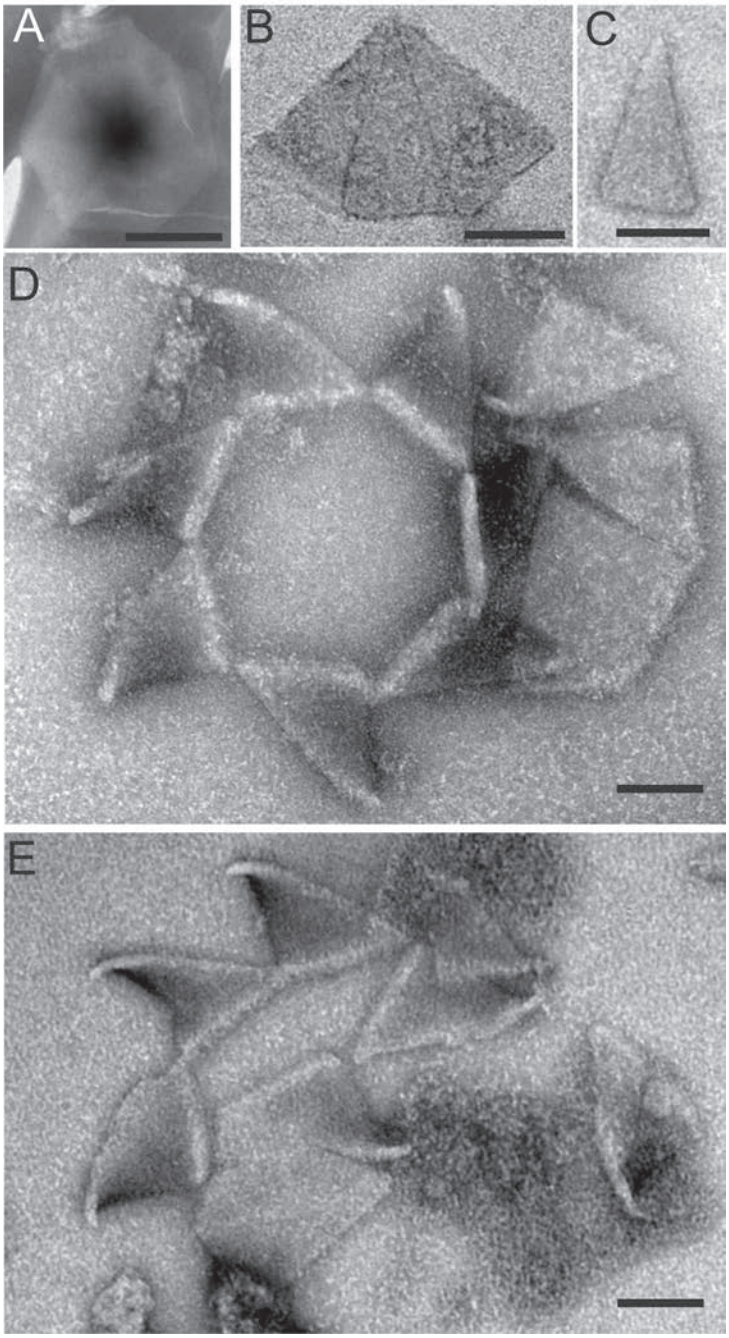


Figure 1. Negative contrast electron micrographs of isolated VAPs. (**A**) Top view and (**B**) side view of intact VAPs. (*C*) A single triangular face detached from a VAP. (**D**) Top view of a VAP in the open conformation. (**E**) Partially broken VAP in the open conformation (Scale bars, 100 nm.)

the VAP (94). Polyclonal antibodies were fused to magnetic beads and used to bind VAPs in crude membrane extracts of infected cells collected 10 hours post infection (h.p.i.), which corresponds to the start of virion release. In agreement with our expectations, it was indeed possible to precipitate and purify the VAPs in this manner. In the purified preparation, no structures other than VAPs were observed, except for a small proportion of S-layer fragments from the host cells. Moreover, the majority of the isolated structures appeared to be intact and similar in shape to the VAPs on the surface of the infected cells (83) (Fig. 1 *A, B*, and S2*A*). The intact purified VAPs could either be observed from the top, which revealed the seven-fold rotational symmetry of the structure (Fig. 1*A*), or from the side, in which case only three to four of the seven triangular faces of the VAPs were visible (Fig. 1*B*). Along with the intact VAPs, in the purified preparation single triangular subunits were also found to be detached from the structure (Fig. 1*C*). In addition, a small proportion of the VAPs were present in an "open" conformation (Fig. 1*D* and *E*). Characteristically, in the latter case, the seven faces of the VAPs were not strictly triangular but rather were slightly curved (Fig. 1*D* and *E*) and resembled the structure of the VAP *in vivo* after it opens (Fig. S2*B*). To exclude the possibility that this curvature was an artefact of sample preparation for negative staining, the VAPs were analyzed using cryo-EM. The faces of the open VAPs that were fixed in vitreous ice were also curved (Fig. 2*A*). The VAPs were best observed from the top, which revealed the heterogeneity of the diameter of the VAPs (Fig. 2*B*).

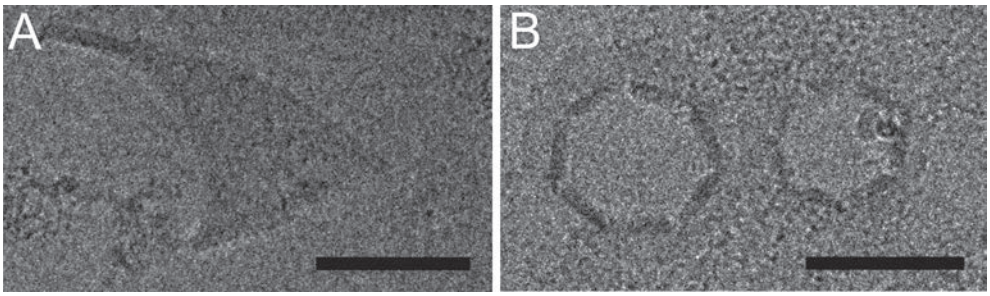


Figure 2. Cryo-electron micrographs of isolated VAPs. (**A**) Single curved face of a VAP. (**B**) Two VAPs in from the top (Scale bars, 100 nm.)

VAP constituents

The ability to isolate intact VAPs allowed us to study their constituents. The protein composition of the VAPs was analyzed by SDS-PAGE followed by SYPRO Ruby staining (Fig. 3*A*). The three visible bands in the purified VAP preparation were identified by Matrix-assisted laser desorption/ ionisation (MALDI)-time-of-light (TOF) mass spectrometry (MS) and MS/MS analyzes. The two larger proteins represented the light and heavy chains of the anti-P98 antibody used to purify the VAPs. The smallest protein (approximately 10 kDa, indicated by an asterisk) was identified as the product of SIRV2-ORF98 (NCBI RefSeq ID: NP_666583) (Fig. 3*A*).

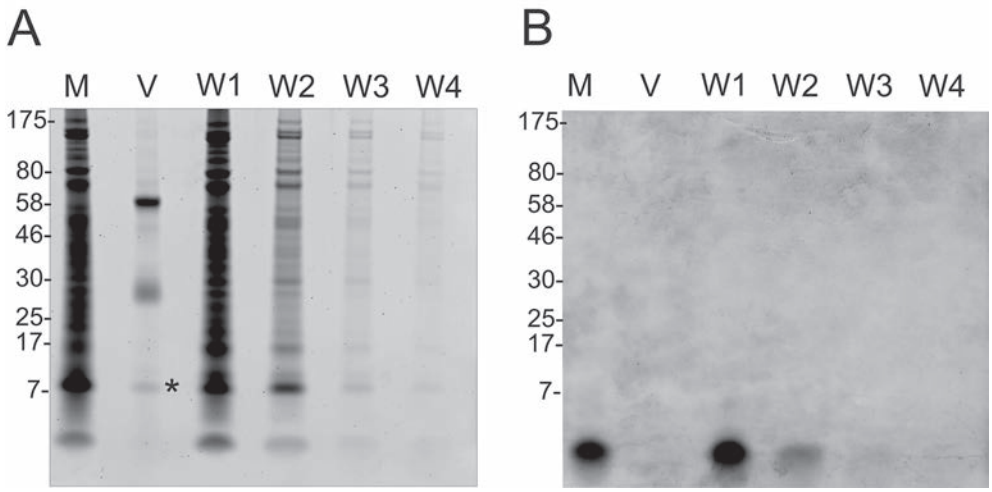


Figure 3. Isolation and constituents of VAPs. (*A*) SDS-PAGE of fractions obtained during immunoprecipitation and purification of VAPs. "M", crude membrane extract; "V", immunoprecipitated fraction; "W1-W4", four successive washes of the immunoprecipitate. Proteins were stained with SYPRO Ruby. (**B**) The same as in *A* but stained for phospholipids using Sudan Black. The P98 band in lane "V" is indicated with an asterisk (*). The positions of proteins with known molecular masses (in kDa) are indicated with bars.

To test the presence of phospholipids in the purified VAPs, a duplicate of the gel shown in Fig. 3*A* was stained using Sudan Black (Fig. 3*B*). Phospholipids could be observed in the membrane fraction and the first two washes of the precipitate but were not detectable in the purified VAP preparation.

Immunolabelling of VAPs

The localisation of SIRV2-P98 in the VAPs was verified by immuno-electron microscopy using polyclonal P98 antibodies and 10-nm protein A-gold particles (Fig. 4). The purified VAPs were labelled evenly across the structure, without concentration of the label in a particular area (Figs. 4*A*, *B* and *C*). As a control, similar labelling was performed on a crude membrane fraction of infected cells, which, in addition to VAPs, contained fragments of S-layers, membranes and virions. In this case, only VAPs were specifically labelled. Another control was performed by labelling purified VAPs with pre-immune serum and 10-nm protein A-gold particles. In this case, very low background labelling was observed (Fig. S3*A*).

The specific localisation of P98 in the VAPs was also confirmed in *S. islandicus* cells infected with SIRV2 h.p.i. Thawed ultrathin cryo-sections of chemically fixed cells were immunolabelled. VAPs in the ultra-thin sections had either an heptagonal appearance if the section was in the plane parallel to the cell surface and the base of the VAP or a triangular appearance if the sections were in other planes (83). As a result of labelling, the gold particles were arranged in

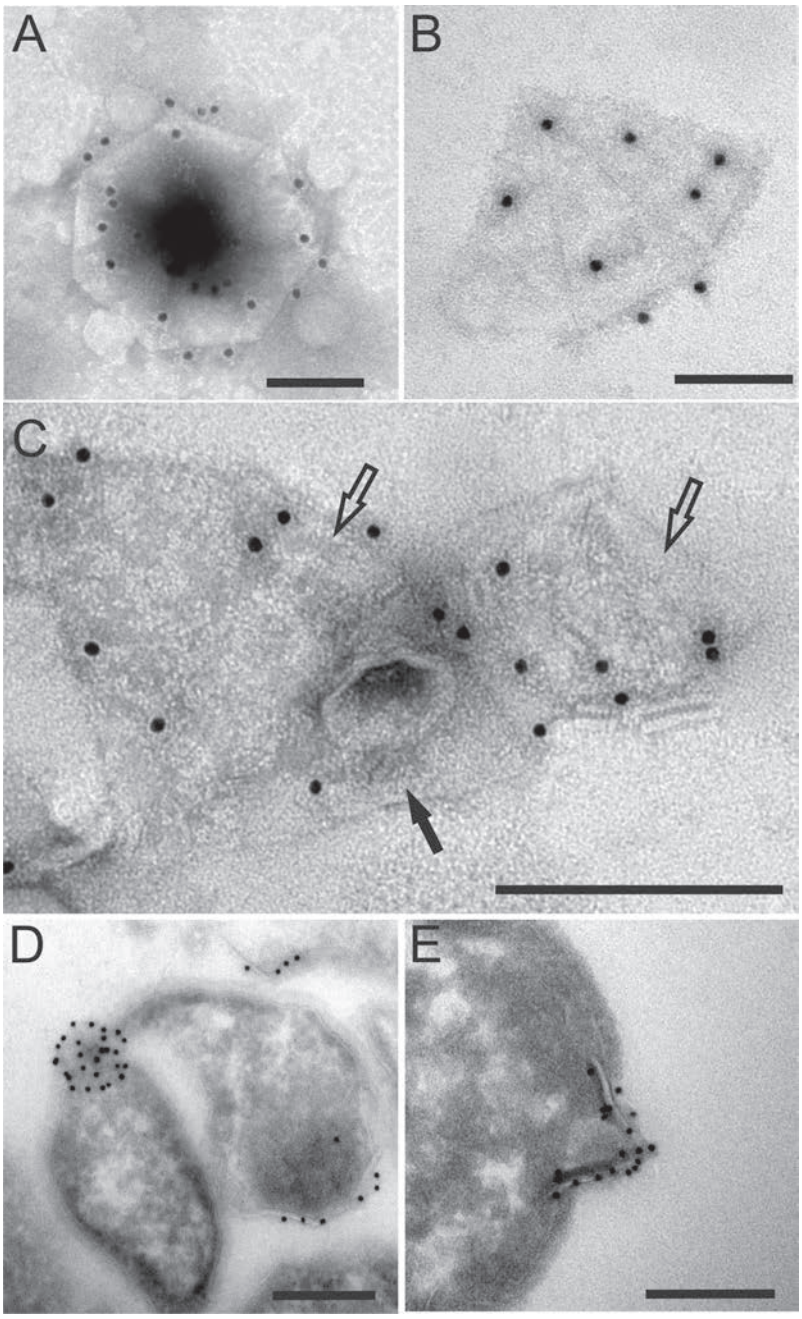


Figure 4. Negative contrast electron micrographs and P98 immunolabelling of VAPs. (**A**, **B**, and **C**) Purified VAPs seen from different angles. (**D** and **E**) Thawed cryo-sections through SIRV2-infected cells. The arrows in *C* indicate three individual VAPs, visualised from the side (empty arrows) and from below (filled arrows). (Scale bars in *A*-*C*, 100 nm; Scale bars in *D* and *E*, 200 nm.)

either a quasi-heptagonal shape (Fig. 4*D*) or along two sides of a VAP triangle (Fig. 4*E*). As a control, thawed ultra-thin cryo-sections of uninfected *S. islandicus* cells were labelled in a similar manner, and no labelling of any specific structure was observed.

VAP structure

Electron microscopy of isolated VAPs was used to produce a detailed description of their structure. Images of 150 intact or partially disrupted VAPs were measured. Independent of the sizes, the geometry of all VAPs was identical and represented a baseless pyramid with a heptagonal perimeter. Each of the seven faces of the pyramidal structure was an isosceles triangle, i.e., two sides were equal in length. The two angles at the base of the pyramid were 74°± 8°, and the one at the tip was 32° ± 3°. The structure shown in Fig. 5*A* represents a VAP in the closed form that was disrupted only along one seam between two of the seven faces. A two-dimensional schematic of the VAP, based on this image, is presented in Fig. 5*B*. Its folding, results in the three-dimensional structure shown in Fig. 5*C*. The structure perfectly matches all of the electron microscopic images of the VAPs taken from different angles.

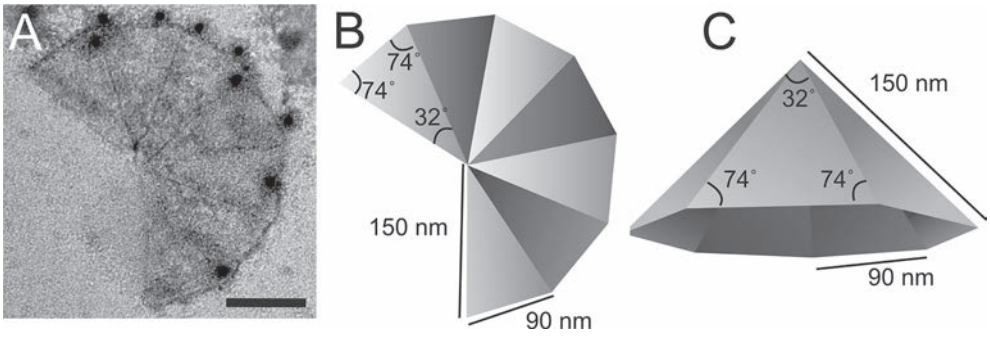


Figure 5. VAP structure. (**A**) Negative contrast electron micrograph and P98 immunolabelling of a partially disrupted VAPs. (**B**) Schematic representation of the image in *A*. (**C**) Schematic 3D reconstruction of the native VAP structure based on the images in *A* and *B*. The lengths and angles were measured from the image in *A*. (Scale bar, 100 nm.)

The data derived from isolated VAPs allowed us to conclude that the VAP is a baseless, hollow, pyramidal structure (Fig. 5*C*). This interpretation agrees with earlier reports of the structure of the VAP *in vivo*. Thin sections of infected cells suggest that cytoplasm is present in the interior of the VAP (Fig. S2*A*) (83).

The VAPs in the analyzed preparation were isolated from a quasi-synchronous population of host cells 10 h.p.i., and although they had identical geometries, they differed significantly in size. The lengths of the sides of the regular heptagonal perimeter of the VAP were measured using 150 images of isolated closed structures and ranged from about 15 nm to about 175 nm.

These structures were grouped into four size classes: (I) 0-50, (II) 50-100, (III) 100-150 and (IV) 150-200 nm (Fig. S4). The following percentages of VAPs in each class were found: class I, 22%; class II, 25%; class III, 47%; and class IV, 7% (Fig. S4). The most likely cause of the observed size distribution and the low percentage of class IV structures appears to be the proximity of the latter to the open state. This suggestion is based on the measurements of the side lengths of 20 open VAPs, of which 95% were in the range of 125-180 nm.

Overexpression of SIRV2-ORF98 in *Sulfolobus acidocaldarius*

Although P98 was found to be the sole constituent of the VAPs, it remained unclear whether it was sufficient for the formation of the pyramidal structure. To examine this possibility and to exclude the involvement of any other SIRV2-encoded factors in VAP assembly, SIRV2-ORF98 was expressed heterologously in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. This member of the genus *Sulfolobus* is resistant to SIRV2 and does not carry any extrachromosomal

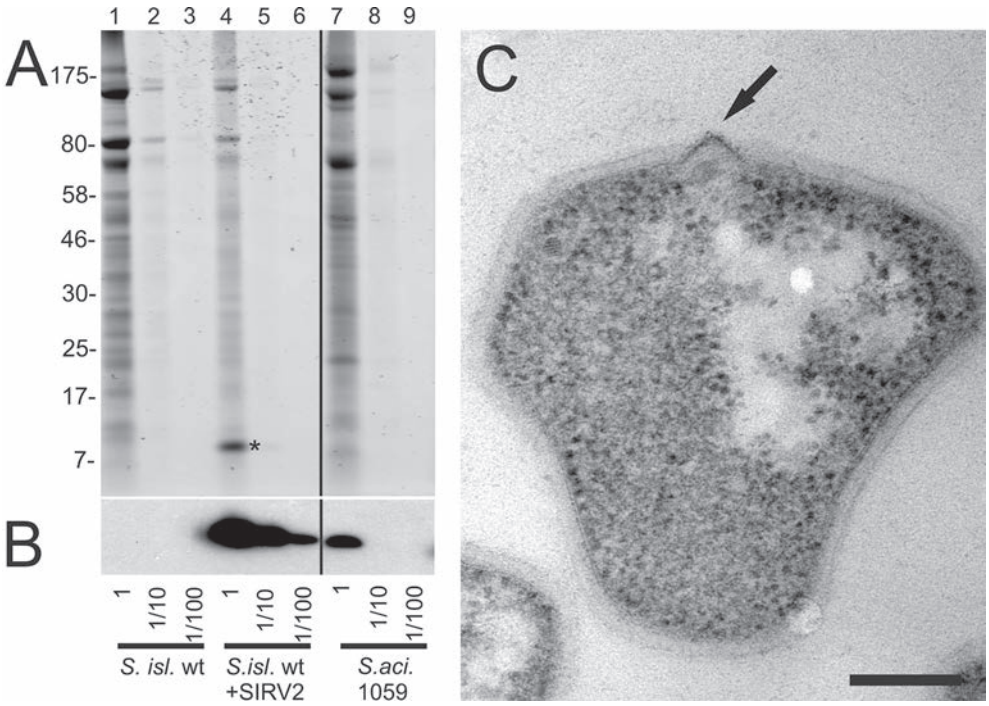


Figure 6. Heterologous expression of SIRV2-ORF98 in *S. acidocaldarius*. (**A**, **B**) SDS-PAGE of membrane fractions from the following samples: (1-3), uninfected *S. islandicus;* (4-6) SIRV2-infected *S. islandicus*, 10 h p.i; (7-9) *S. acidocaldarius* 1059 expressing SIRV2-ORF98. The samples were loaded without diluted and with dilutions of 1:10 and 1:100. The positions of proteins with known molecular masses (in kDa) are indicated with bars. The asterisk (*) shows the band containing P98. (**A**) Coomassie Blue-stained gel. (**B**) Western hybridisation of a duplicate of the gel in *A* with antibodies against SIRV2-P98. (**C**) Thin section through *S. acidocaldarius* 1059 expressing SIRV2-P98. The arrow indicates a VAP (Scale bar, 200 nm.)

element, or any genes homologous to those of SIRV2. For expression, SIRV2-ORF98 was cloned under the control of a maltose-inducible promoter in the pCMalLacS plasmid (95). Western blot revealed an approximately hundredth weaker signal than that produced by SIRV2 infection of *S. islandicus* (Figs. 6*A* and *B*). Remarkably, the analysis of ORF98-expressing *S. acidocaldarius* cells by electron microscopy showed the presence of VAPs on the cell membranes (Fig. 6*C*). The VAPs perforated the S-layer and were similar in size and geometry to the VAPs of SIRV2-infected *S. islandicus* cells. The low number of observed VAPs, i.e., about 1 in 100 cells, corresponded to the low level of ORF98 expression in *S. acidocaldarius*. It is noteworthy that only closed VAPs were observed in the heterologous system, whereas in the natural virus-host system VAPs in both closed and open states are present (2,3).

Overexpression of SIRV2-ORF98 in *Escherichia coli*

To exclude the possibility that some proteins common to members of the genus *Sulfolobus* are involved in VAP formation and to further test the self-assembly capacity of P98, the SIRV2-ORF98 gene was expressed in *E. coli*. The gene was cloned under a *lacS* promoter into the pSA4 plasmid (96). Expression was confirmed by Western hybridization, and the efficiency of P98 production was comparable to that of SIRV2-infected *S. islandicus* cells (Figs. 7*A* and *B*). The presence of the recombinant P98 was detected primarily in the membrane fraction. To our astonishment, P98 was not only produced in *E. coli* but was also capable of forming VAPs in these cells. Electron microscopy analysis of thin sections of P98-expressing cells of *E. coli* at 3, 7 and 21 hours after induction, revealed the presence of many pyramidal structures on the inner membrane, which protruded into the periplasmic space of the cell (Fig. 7*C* and *D*). The VAPs appeared to be similar in size and shape to those observed on the cell surface of SIRV2-infected *S. islandicus* and SIRV2-ORF98 expressing *S. acidocaldarius* (Fig. 7*C)*. Only closed VAPs were observed. Three hours after induction, the majority of cells contained VAPs; however, the cells stopped growing almost immediately after induction, and the proportion of VAP-containing cells did not change significantly in the following 21 hours. In addition, the number of observed VAPs per cell did not change over time. The number of VAPs was generally very high and correlated with the expression levels of the viral gene. In most sections, more than 30 VAPs were observed per cell (Fig. 7*D*). The formation of VAPs seemed to increase the surface area of the inner membrane, and the distance between the outer and inner membranes (Figs. 7*C* and *D*).

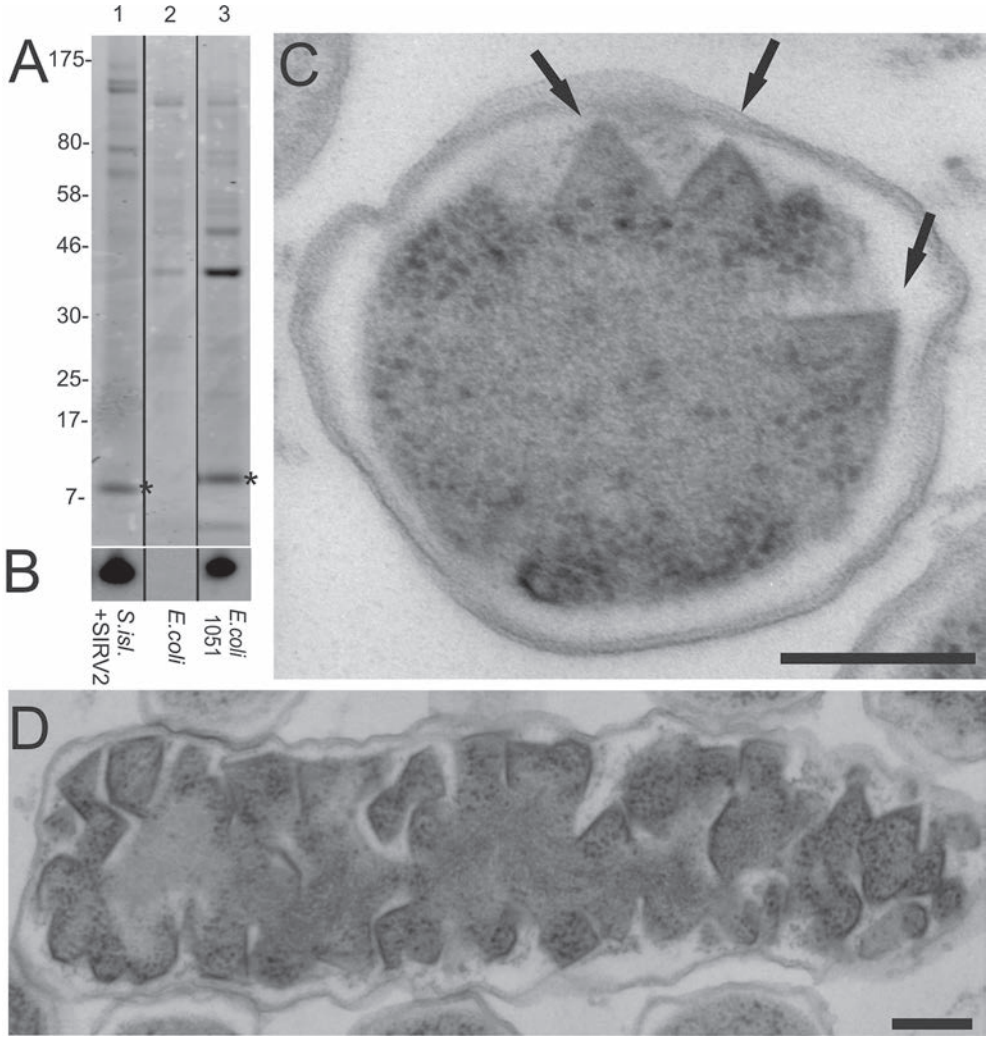


Figure 7. Heterologous expression of SIRV2-ORF98 in *E. coli*. (**A**, **B**) SDS-PAGE of membrane fractions from the following samples: (1) SIRV2-infected *S. islandicus*, 10 h.p.i.; (2) *E. coli;* (3) *E. coli* 1051 expressing SIRV2-ORF98. The positions of proteins with known molecular masses (in kDa) are indicated with bars. The asterisks (*) show the bands containing P98. (**A**) Coomassie Blue-stained gel. (**B**) Western hybridisation of a duplicate of the gel *A* with antibodies against SIRV2-P98. (**C**, **D**) Thin sections through *E. coli* 1051 expressing SIRV2-P98 (Scale bars, 200 nm.)

Discussion

The discovery of the exceptional VAP-based virion egress system exploited by some archaeal viruses (2, 3) has drawn attention to the underlying molecular machinery. Our finding that the VAPs represent separate structural units that can be isolated and purified from the membrane fraction of infected cells is crucial for understanding the nature and formation of these remarkable structures. Purification of VAPs was achieved by immunoprecipitation using antibodies against protein 98 of SIRV2, which was previously postulated to be a component of VAPs (94).

The VAPs were isolated and purified from a quasi-synchronised culture of SIRV2-infected *S. islandicus* at the start of virion release. The majority of purified VAPs appeared to be intact. Each VAP was composed of seven faces of isosceles triangles that together formed a baseless pyramid. In all cases, the angles of the faces were 74°± 8° and 32°± 3°. However, the diameters of the VAPs were heterogeneous. This heterogeneity most likely reflects the dynamics of VAP development in the analyzed cell culture. The observations suggest that VAP geometry does not change over the course of formation and that VAPs grow via gradual expansion of the triangular faces. It is noteworthy that structures of ordered 7-fold symmetry are extremely rare in nature (7, 8).

In addition to VAPs in the natural closed conformation, the purified preparation contained VAPs that were in the open state (Fig. 1). These VAPs might have originated from perforated cells. Indeed, the population of infected cells that was used for VAP isolation contained a small proportion of perforated cells with open VAPs. Alternatively, "opening" of VAPs could have been caused by mechanical shearing during the purification process. The faces of VAPs in the open state were curved, most likely due to the characteristics of the opening process. Another characteristic feature of the open VAPs was their large size, with diameters exceeding 250 nm. This suggests that VAPs might need to reach certain dimensions before they can be opened.

The analysis of the VAP constituents revealed that the structure is composed of multiple subunits of a SIRV2-encoded protein, SIRV2-P98. No lipid component could be detected by the methods used. The absence or extremely low proportion of lipids agrees with the fact that VAPs could be isolated as intact structures after extensive washing of the membrane fraction with detergents (Materials and Methods).

The disclosure of SIRV2-P98 as a sole component of the VAPs, allows suggesting that archaeal viruses which encode homologues of this protein may explore the same egress mechanism as SIRV2. This suggestion is in line with the VAP-based egress of the virus STIV1 (3), the genome of which contains ORF-C92 highly similar in sequence to SIRV2-ORF98 (4). However, besides STIV1 there are only two other viruses known to encode SIRV2-P98 homologues, both from the family *Rudiviridae* to which SIRV2 also belongs (4).

SIRV2-P98 was confirmed to be the only constituent of the VAP also by heterologous expression experiments. Expression of SIRV2-ORF98 in both *S. acidocaldarius* and *E. coli* led to the formation of VAPs on cell membranes (Figs. 5 and 6). These results also demonstrate that P98 is capable of self-assembling into pyramidal structures with seven-fold rotational symmetry. ORF98 was the only SIRV2 gene expressed in the mutant cells of *S. acidocaldarius* or *E. coli*. It is remarkable that the dramatic differences in the membrane composition and chemistry between bacterial and archaeal cells (97) did not affect the self-assembly of recombinant P98 into VAPs *in vivo*. The ability of P98 to form VAPs under the different conditions required for the growth of an hyperthermophilic, extremely acidophilic archaeon and a mesophilic bacterium is also remarkable.

The VAPs formed by heterologous expression of SIRV2-ORF98 in *E. coli* and *S. acidocaldarius* were similar in size and shape to those formed in *S. islandicus* infected with SIRV2. However, in the heterologous systems, VAPs were never observed in the opened state, unlike in the natural system, where VAPs eventually open and cause perforation of the infected cell. These observations suggest that at least one special factor is required for the process of VAP opening, which is absent in *S. acidocaldarius* and *E. coli.*. The nature and origin, viral or cellular, of this factor is currently unclear.

Our results demonstrate that the rudivirus SIRV2 encodes one autonomous structure in addition to the capsid, i.e., the VAP. The virus-encoded constituents of both structures can self-assemble. The major capsid protein self-assembles into filaments with the same diameter as the native linear virion (58), and VAP protein P98 self-assembles into pyramids. However, the functions of the two autonomous structures are different. Whereas the function of the former is DNA packaging and the formation of virus particles, the latter is specifically designed to release virus particles from the host cell. The molecular simplicity and elegance of the VAP design revealed in this study should aid in the future analysis of the elaborate molecular mechanisms of this unique virion release system in archaea.

Experimental procedures

Virus and host strains

The stock of the SIRV2 virus was prepared as described in ref. (83). The growth of *S. islandicus* LAL/14 and virus infection were also described in ref. (83).

VAP isolation

S. islandicus LAL 14/1 cells infected with SIRV2 and uninfected controls were harvested 10 h.p.i. The cells were disrupted using a French press, and the membranes were collected by

centrifugation at 100,000 g as described in ref. (94). VAPs were isolated from the membrane fraction using anti-P98 antibody. The customised P98 polyclonal peptide antibody was raised in rabbit against a peptide corresponding to the C-terminal region of P98 and affinity purified using this antigen (Eurogentec, Liege, Belgium). The specificity of the antibody was tested on a Western blot.

The antibodies were coupled to magnetic beads covered with protein A (Dynabeads Protein A, Invitrogen™ Dynal AS, Oslo, Norway) according to the manufacturer's protocol using non-denaturing elution. Alternatively, the membrane fraction was washed in 0.5% N-lauroylsarcosine (Sigma-Aldrich, St. Louis, USA) and centrifuged at 100,000 g, and the pellet was used for immunoprecipitation.

Protein and lipid analysis

The protein fractions were loaded onto a 4-12% polyacrylamide Bis-Tris gradient gel (Invitrogen™ Dynal AS, Oslo, Norway) using 2-(N-morpholino) ethanesulphonic acid SDS running buffer (Invitrogen™ Dynal AS, Oslo, Norway), and the proteins were visualised using either SYPRO Ruby (Invitrogen™ Dynal AS, Oslo, Norway) or Coomassie-based Instant Blue™ (Expedeon, Harston, UK) staining. The presence of lipids in the samples was detected by overnight staining in Sudan Black (Sigma-Aldrich, St. Louis, USA) according to the manufacturer's protocol. MS analysis was performed as described in ref. (94).

For Western hybridisation, proteins were transferred onto a PVDF membrane. P98 was visualised using 1:10000 dilution of the anti-P98 antibody and peroxidase-coupled goat anti-rabbit IgG (Sigma-Aldrich, St. Louis, USA) and the SuperSignal® West Pico Chemiluminescent Substrate (Thermo Scientific, Rockford, USA) on Amersham Hyperfilm™ ECL (GE Healthcare, Buckinghamshire, UK).

Negative contrast EM and immunolabelling of isolated VAPs

Purified VAPs were absorbed to grids with carbon-coated Formvar films and negatively stained for 1 min with 2% (wt/vol) uranyl acetate. For immunocytochemistry, after absorption on the grids, the samples were incubated in PBS with 1% BSA and labelled with the anti-P98 peptide antibody. The antibody-labelled protein was detected using 10-nm Protein A-colloidal gold (CMC, Utrecht, The Netherlands). All samples were observed on a Jeol 1200EX-II operated at 80 kV. Images were recorded and measured on a MegaView or KeenView camera (Olympus-SIS, Munster, Germany) using the ITEM software version 5.0 (Olympus-SIS, Munster, Germany).

Cryo-EM of isolated VAPs

Purified VAPs were spread on glow discharged Quantifoil R2/2 grids (Quantifoil Micro Tools GmbH, Jena, Germany) and cryofixed in liquid ethane. The specimens were transferred to a Gatan

626 DH cryoholder (Gatan, Pleasanton, USA) and examined on a Jeol2010F electron microscope (Jeol, Tokyo, Japan) operating at 200 kV. Images were recorded under low-dose conditions on a Gatan Ultrascan 4000 with Digital Micrograph (Gatan, Pleasanton, USA) version 1.83.842.

Immunolabelling of thawed cryosections

SIRV2-infected *S. islandicus* cells were fixed 10 h.p.i. with 4% formaldehyde in 0.1 M HEPES buffer, pH 6, for 5 h at 4 ˚C. The cells were pelleted by low speed centrifugation, embedded in 10% gelatine, cut into small blocks and infiltrated in 2.3 M sucrose at 4 ˚C. The blocks were mounted on aluminium pins and frozen in liquid nitrogen. Sections were cut with a nominal feed of 70 nm using a 35° angle diamond knife (Diatome, Biel, Switzerland) and a FC6 microtome (Leica Microsystems, Vienna, Austria). To pick up the ultrathin cryosections, a 1:1 mixture of 2.3 M sucrose and 2% methylcellulose was used.

Thawed sections were placed on formvar and carbon-coated copper grids, and sections were blocked in 1% BSA and labelled with the anti-P98 peptide antibody and 10-nm Protein A-colloidal gold (CMC, Utrecht, The Netherlands). Images were recorded on a KeenView camera (Olympus-SIS, Munster, Germany) using the ITEM software version 5.0 (Olympus-SIS, Munster, Germany).

Resin embedding of *Sulfolobus* cells

S. acidocaldarius MR31 pSVA1059 and SIRV2-infected *S. islandicus* were fixed with 2.5% (wt/vol) glutaraldehyde in 0.1 M HEPES buffer, pH 6 and 7, respectively. Post-fixation, dehydration, embedding in epoxy resin, sectioning and imaging with TEM were performed as described in ref. (83).

High pressure freezing and freeze substitution of *E.coli*

E.coli cells were taken up in capillary tubes (Leica, Vienna, Austria) as described in ref. 11. The filled tube was placed into the cavity of a brass planchette (Agar Scientific, Stanstad, UK), filled with 1-hexadecen, and immediately frozen with a HPM 010 (BalTec, now Abra Fluid AG, Widnau, Switzerland). Freeze-substitution was performed in anhydrous acetone (EMS, Hatfield, USA) containing 2% osmium tetroxide (Merck, Darmstadt, Germany) and 2% water. Substitution was performed at -90 °C for 24h, and at -60 °C and -30 °C for 8h each in an automated freeze substitution device (Leica AFS, Leica Microsystems, Vienna, Austria). Afterwards the temperature was raised to 0 °C and the samples washed with dry acetone and embedded stepwise in Epon and polymerized at 60°C for 48h.

Plasmid constructs and transformation of *S. acidocaldarius* and *E. coli*

For the overexpression of P98 in *S. acidocaldarius* M31, SIRV2-ORF98 (NCBI RefSeq ID: NP_666583) was amplified from SIRV2 genomic DNA and cloned into the *lacS* gene locus in the pSVA5 plasmid (98) using the NcoI and BamHI sites. SIRV2-ORF98 and the *araS* promoter were transferred from pSVA5 to the pCMalLacS plasmid (95) using the NcoI and EagI sites, which yielded pSVA1059.

S. acidocaldarius M31 was transformed with pSVA1059. The preparation of competent cells, methylation of the plasmid, and electroporation were carried out as described in ref. 13. The electroporator used was the Gene Pulser Xcell (Bio-Rad, Hercules, USA) with 1 mm cuvettes at 1500 V, 600 Ω and 25 µF. The cells were regenerated in Brock medium containing 0.1% enzymatically hydrolysed casein (tryptone, BD Biosciences, Franklin Lakes, USA) and 0.2% dextrin for 30 min at 75 ˚C. The cells were then streaked onto selective Gelrite® (Sigma-Aldrich, St. Louis, USA) plates lacking uracil. After 5 days, a preculture was grown under selective conditions from a single colony; 50 mL of medium containing inducer (0.2% maltose) was then inoculated with 1 mL preculture and grown until an OD_{600} of 1 was reached.

For the overexpression of SIRV2-ORF98 in *E. coli* Rosetta(DE3)pLys (Novagen® Merck KGaA, Darmstadt, Germany), the same gene was amplified from SIRV2 genomic DNA and cloned into the T7 promoter-driven expression vector pSA4 (96) using the NcoI and BamHI sites. The pSVA1051 vector contains an IPTG-inducible promoter that was used for the expression a C-terminal his-tagged protein.

Acknowledgements

We are grateful to Martin Sachse (Plate-Forme de Microscopie Ultrastructurale, Institute Pasteur) and Benjamin Meyer (Molecular Biology of Archaea, Max-Planck Institute for Terrestrial Microbiology) for experimental assistance. Also we are thankful to John van der Oost of (Wageningen University and Research centre) and Mart Krupovič (Biologie Moléculaire du Gène chez les Extrêmophiles, Intitute Pasteur) for helpful discussions. This work was supported by the Programme Blanc of the Agence Nationale de la Recherche (Paris, France) under grant ANR-09-BLAN-0288.01. T.Q. was supported by a fellowship from the Ministère de l'Enseignement supérieur et de la Recherche of France.

Supplemental figures

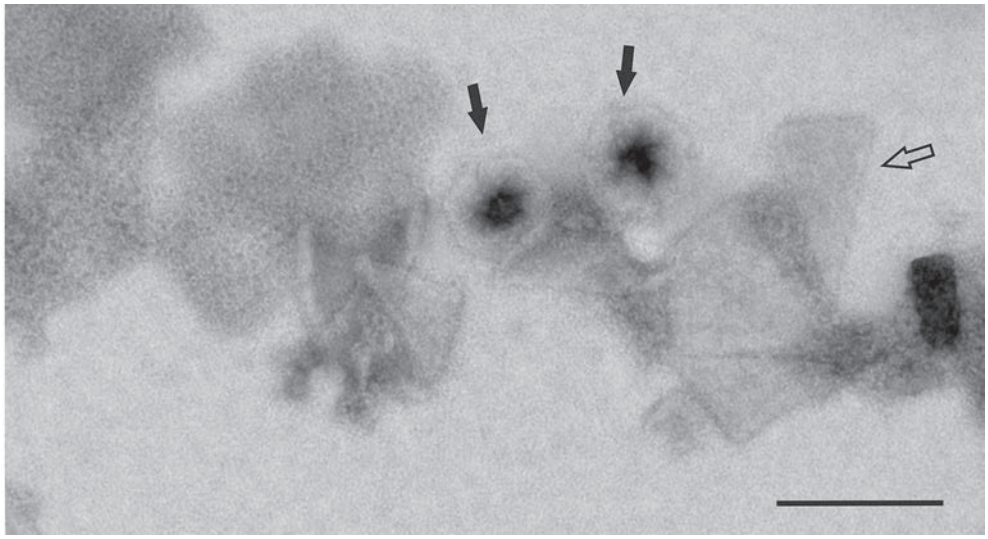


Figure S1. Negative contrast electron micrograph of VAPs observed in membrane extracts of SIRV2-infected cells of *S. islandicus*. VAPs as seen from top view (filled arrow) and side view (arrow) are indicated with filled and empty arrows, correspondingly. Parts of cellular S-layer are also visible. (Scale bar, 100 nm.)

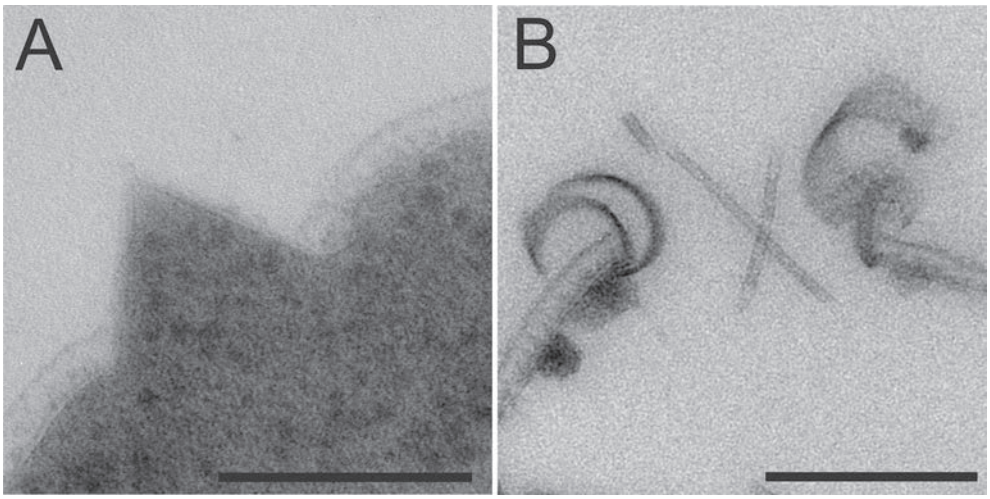


Figure S2. Thin sections through VAPs in (**A**) closed and (**B**) open conformation on cells surface of *S. islandicus* infected by SIRV2. (Scale bar, 100 nm.)

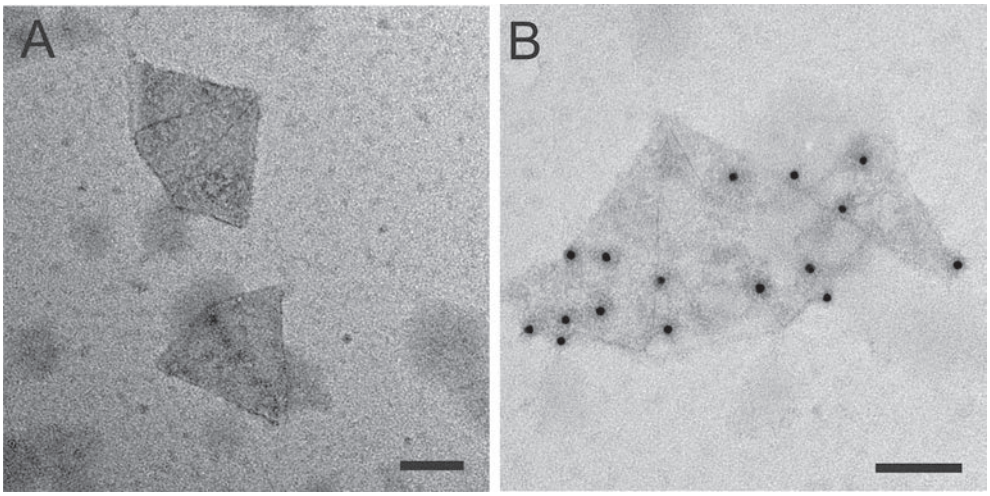


Figure S3. Negative contrast electron micrographs and immunolabelling of isolated VAPs with (**A**) pre-immune serum and (**B**) P98 antibodies. (Scale bars, 100 nm.)

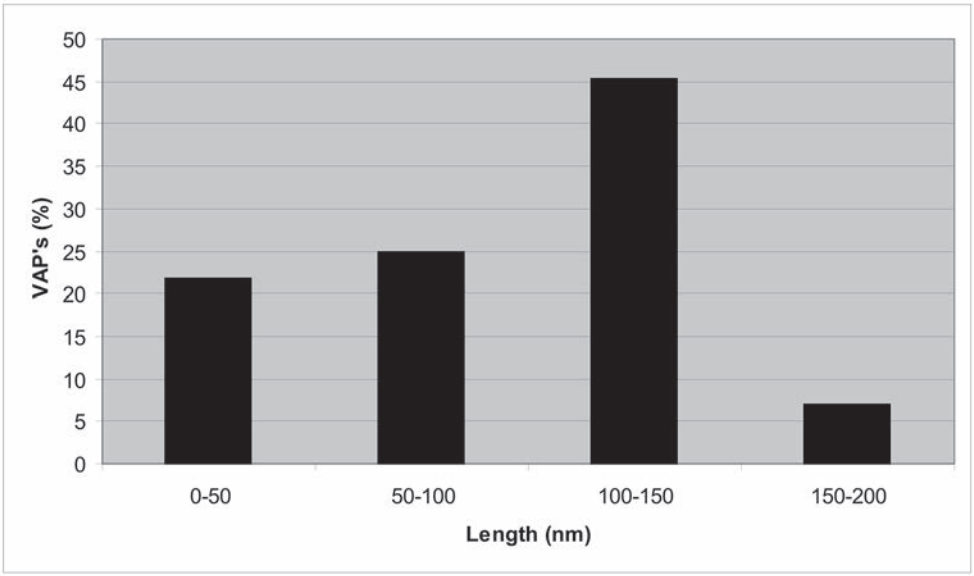


Figure S4. Size distribution of isolated VAPs in closed state. (X) Lengths of sides of the heptagonal perimeter of closed VAPs are divided in four size classes. (Y) The relative abundance of VAPs in each size class.

Chapter 4

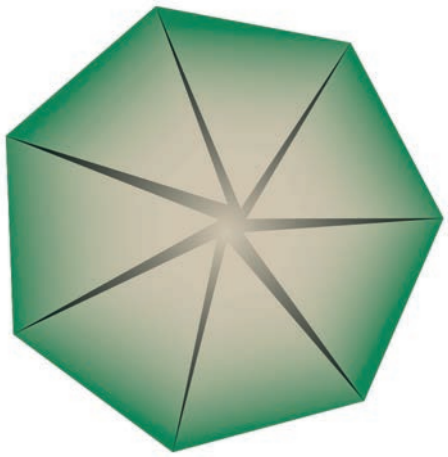
Supramolecular organisation and assembly of the heptagonal pyramid for virion egress in archaea

Bertram Daum *, Tessa E.F. Quax*, Martin Sachse, Deryck Mills, Julia Reimann, Özkan Yildiz, Sabine Häder, Cosmin Saveanu, Patrick Forterre, Sonja-Verena Albers, Werner Kühlbrandt, David Prangishvili

*contributed equally

'Supramolecular organisation and self-assembly of a universal pyramidal membrane remodelling system'

Submitted 2013



Abstract

Viruses have developed a wide range of strategies to escape from host cells in which they replicate. Some archaeal viruses employ for egress a pyramidal structure with sevenfold rotational symmetry. Virus-associated pyramids (VAPs) assemble in the host cell membrane from the virus-encoded protein PVAP, and open at the end of the infection cycle. We characterize this exceptional supramolecular assembly using a combination of genetic, biochemical and electron microscopic techniques. By whole-cell electron cryo-tomography, we monitor morphological changes in virus-infected host cells. Subtomogram averaging reveals the VAP structure. By heterologous expression of *PVAP* in cells from all three domains of life we demonstrate that the protein integrates indiscriminately into virtually any biological membrane, where it forms sevenfold pyramids. We identify the protein domains essential for VAP formation in PVAP truncation mutants by their ability to remodel the cell membrane. Self-assembly of PVAP into pyramids requires at least two different, in-plane and out-of-plane, protein interactions. Our findings allow us to propose a model of how PVAP arranges to form sevenfold pyramids and suggest how this small, robust protein may be used as a universal membrane remodeling system.

Introduction

Release of virus particles from infected cells is the last essential step of the viral replication cycle. In the course of this process, virions face the challenging task of crossing the cell envelope. Viruses have developed an arsenal of diverse strategies to overcome this problem. Most bacterial viruses induce lysis of the infected cell with help of the holin-endolysin system (99), while others disrupt the host cell envelope via inhibition of the murein biosynthesis pathway (100). The exceptional morphological and genomic properties of archaeal viruses (22, 27) suggested that their exit from host cells may also have unusual traits that are different from those of bacterial viruses. Indeed, it was shown that some lytic viruses of hyperthermophilic archaea, *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) and *Sulfolobus* turreted icosahedral virus 1 (STIV1), exploit a particularly exotic mechanism of virion egress (40, 101-104). During the infection cycle of these viruses, pyramidal protrusions with sevenfold rotational symmetry form in the host cell membrane. As the final step of the infection cycle the VAPs open outwards along the seams of their seven facets, creating large apertures through which the newly formed virions escape from the host cell (101, 102). VAPs consist of multiple copies of a ~10 kDa virus-encoded protein, which we term PVAP (SIRV2_P98) (102, 103, 105). Surprisingly, PVAP assembles into pyramidal structures even when expressed heterologously in archaeal and bacterial expression systems, demonstrating that no other viral proteins are required for VAP formation (102). The detailed mechanisms by which VAPs self-assemble in the membrane have remained obscure.

In the present study we used electron cryo-tomography to investigate morphological features of SIRV2 replication and the formation of VAPs at different time points after infection. By subtomogram averaging, we determined a first 3D map of the VAP. This map, in combination with secondary structure predictions of PVAP and expression of *wild type PVAP* or a variety of truncation mutants in archaeal, bacterial and eukaryotic cells allows us to propose a model of how PVAP arranges to form the sevenfold pyramids. These insights are fundamental for understanding how this intriguing mechanism can be exploited as a universal tool to engineer the formation and controlled opening of large pores in biological or artificial lipid bilayers.

Results

SIRV2 induces morphological changes of the host cell

We analyzed morphological changes in *S.islandicus* during SIRV2 infection and the time points of VAP formation and opening by whole cell electron cryo-tomography at 0.5, 3, 6 and 12 hours post infection (h.p.i.). This allowed us to monitor morphological changes at high resolution and to compare these with previous results obtained by thin sectioning of chemically fixated cells

during the final stages of SIRV2 infection (83). Up to 3 h.p.i. the morphology of *S. islandicus* cells resembled that of uninfected control cells and no virions were visible in the cytoplasm (Figs. 1*A*). This suggests that SIRV2 infects its host by DNA injection rather than by entering the cell as an intact virus particle. Electron-dense ~100 nm globules were the only conspicuous features observed in the cytoplasm of controls and infected cells (Fig.1*A*), similar to those reported for *S. solfataricus* cells (106). Each cell, infected or not, contained one or a few of these globules, which are therefore unrelated to virus replication. Their size and density above the cytoplasmic background indicates that they may be storage granules (107).

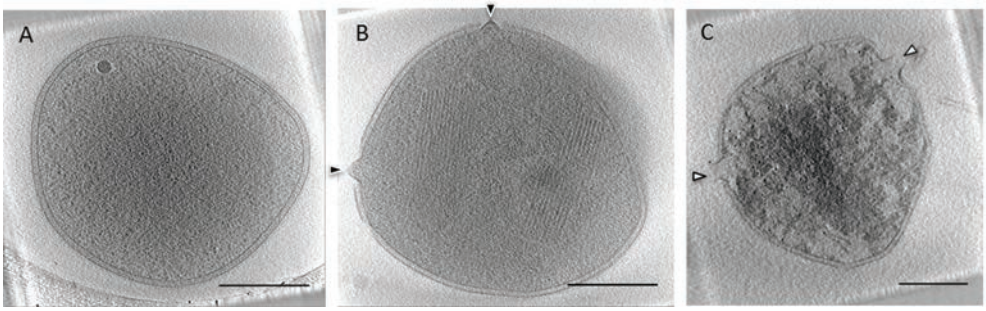


Figure 1. Morphological changes of *S. islandicus* during infection with SIRV2. Tomographic slices of typical archaeal cells at 0.5 (**A**), 3-6 (**B**) and 9 (**C**) hours post infection with SIRV2. Black arrowheads, closed VAPs; white arrowheads, open VAPs. Scale bars, 500 nm.

At 3-6 h.p.i., about half way through the infection cycle, newly assembled virions became visible in the cytoplasm as observed earlier (83). They were organized in up to three bundles per cell, each consisting of roughly 50 rod-shaped particles (Fig.1*B*). Starting from 3 h.p.i., VAPs of various sizes formed in the plasma membrane of the host cells, most of which had penetrated the S-layer (Fig. 1*B*, 2 *A-D*). The height of the VAPs (measured from the membrane to the tip of the pyramid) ranged from ~20nm to 150nm. At an early stage of formation, the VAPs in the plasma membrane had not yet punctured the S-layer of the host cell (Fig. *S1*), but they already had the distinct features of hollow heptagonal pyramids, corroborating the earlier assumption that VAPs grow by gradual expansion of their triangular facets (102). Very occasionally, VAPs contained a spherical storage granule (Fig. 2C). These granules likely correspond to the previously described intra-pyramidal bodies (IPB) in STIV-induced VAPs (106). Analysis of cells at several moments during the infection cycle, showed that at 6 h.p.i. a number of VAPs had opened. The fraction of open VAPs increased, until at 12 h.p.i. all VAPs had unfolded, similar to petals of a flower (Fig 1*C,* 2*E-H*). The VAPs appeared to open as the pyramidal structures broke along the seams of the triangular pyramid faces (Fig 2*F*). The VAP facets curved outwards with counter-clockwise handedness when viewed from the cell exterior (Fig. 2*F*), as had already been suggested by

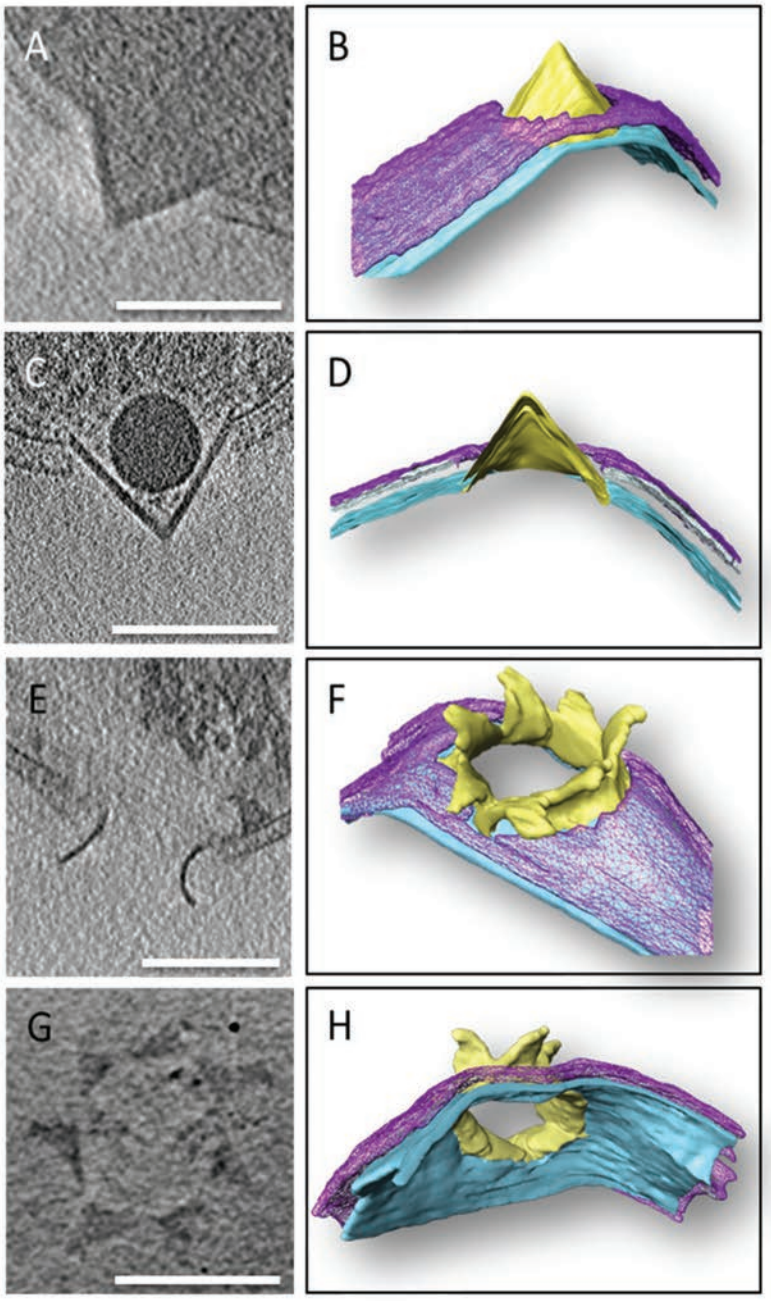


Figure 2. VAPs in closed and open conformation. (**A**, **C**, **E**, **G**) Tomographic slice and segmented, surface-rendered volumes (**B**, **D**, **F**, **H**) of VAPs in the membrane of SIRV2-infected *S. islandicus* cells. VAPs are either closed (**A-D**) or open (**E-H**). The S-layer is purple, the cell membrane blue and the VAP is yellow. Scale bars, 200nm

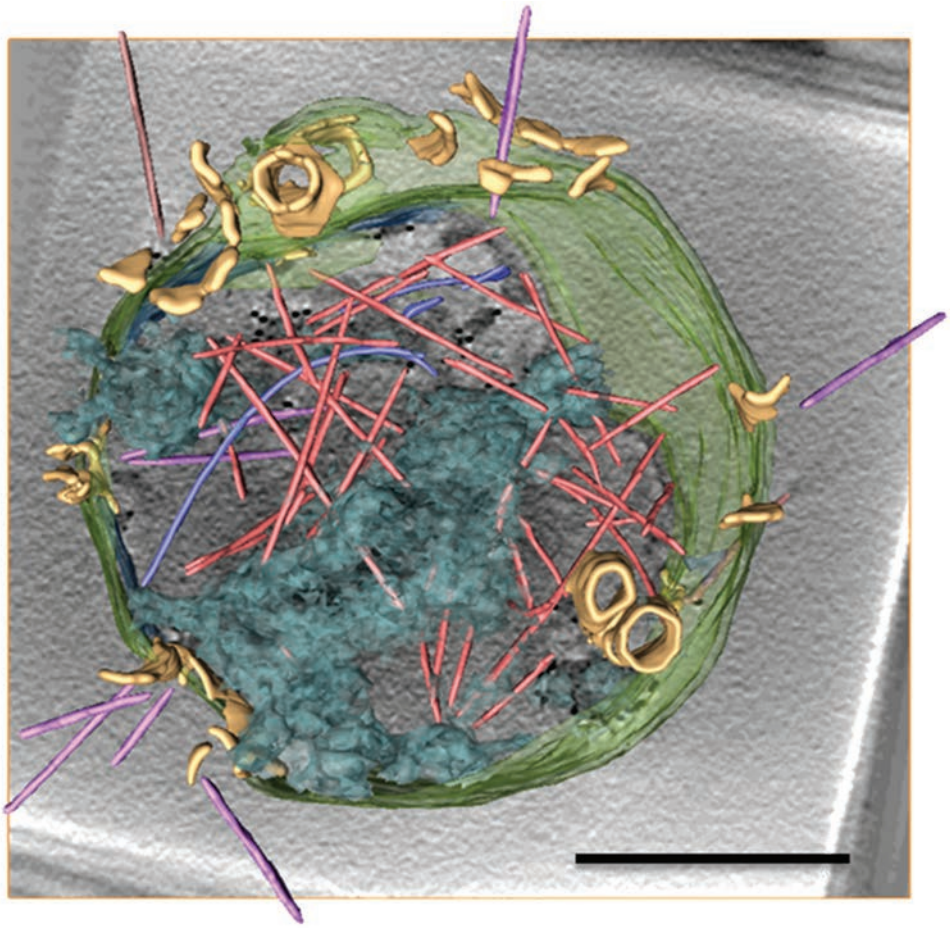


Figure 3. SIRV2 virion egress. Rendered tomographic volume of a SIRV2-infected *S. islandicus* cell, 12 h post infection. SIRV2 virions (orange, brown, purple) are released trough open VAPs (yellow) that create ~ 100 nm apertures in the plasma membrane and S-layer (green). Orange, virions inside the cell; purple, virions escaping from the cell; brown, virion outside the cell. Transparent blue, viral or host DNA.

electron micrographs of isolated VAPs (102). As the VAPs opened, the virion bundles disintegrated and egress of individual virions through the open VAPs was observed (Fig. 3).

VAP structure

The shape of isolated VAPs has previously been studied by negative staining and electron microscopy (102). To gain insight in the structure of VAPs *in situ,* closed or open VAPs were imaged by whole-cell electron cryo-tomography (Fig. 4*A*-*C*). This revealed that all VAPs, irrespective of their conformation or stage of assembly, consisted of two distinct layers. The ~4.5±1 nm outer layer was continuous with, and indistinguishable from, the cell membrane. The inner layer had

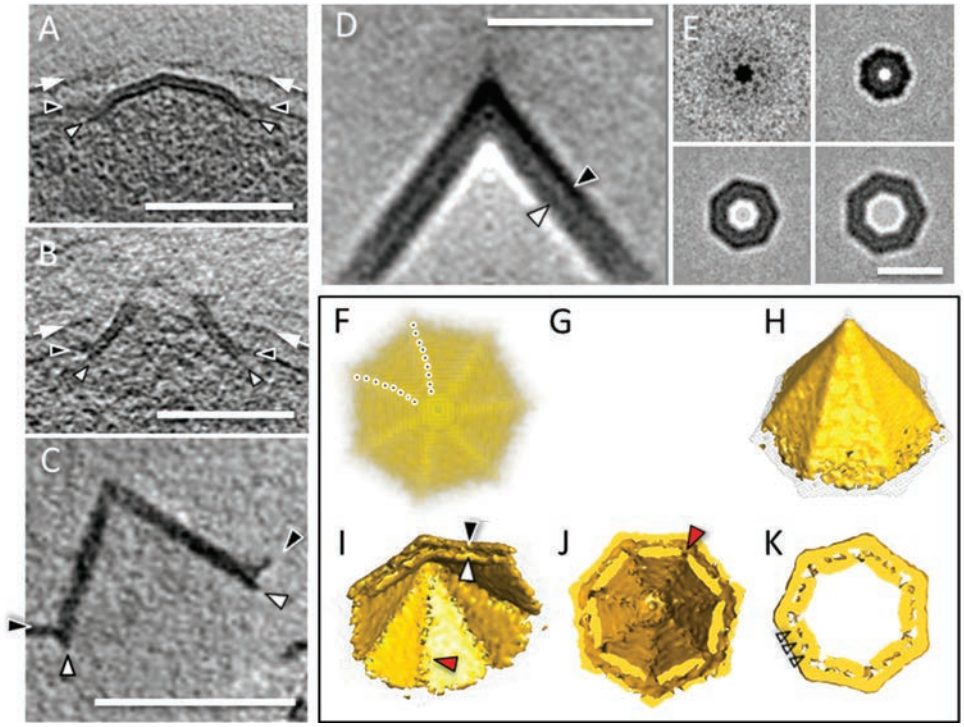


Figure 4. VAP structure. Tomographic slices of SIRV2-induced VAP of *S.islandicus* (**A**) and VAPs formed after PVAP expression in *E.coli* (**C**), indicating two layers, one of which is continuous with the cell membrane (black arrowheads), while the other (white arrowheads) forms a sheet at the cytoplasmic surface of the membrane. VAPs in *S. islandicus* protrude trough the S-layer (white arrows). (D-K) 3D map of VAP obtained by subtomogram averaging, with sevenfold symmetry applied. Tomographic slice perpendicular to the pyramid base (**D**) and successive tomographic slices parallel to the base (**E**) show the two layers in the walls of the pyramid. (**F**) Top-view of the 3D map in solid representation (128) shows that the edges of the of the seven pyramidal facets are slightly curved counter-clockwise (dotted lines). (**G-K**)Different orientations of the 3D map in surface representation. Transparent mesh and golden surface show different threshold levels. Black and white arrowheads indicate outer and inner layer, respectively. Red arrowhead indicates low density region at the edges of inner facets. Open arrowheads indicate connections between inner and outer layers of the VAP. Scale bars, 200 nm (**A-C**), 50 nm (D-K).

a thickness of ~4.0±1 nm (Fig. 4). The centre-to-centre distance between the two layers was 10.0±1 nm, leaving a 5.8±1 nm gap of lower density. At the base of each VAP, the inner layer extended up to 15 nm beyond the outer layer into the cytoplasm. This suggests that the inner layer consists of a protein sheet that is attached and runs parallel to the cytoplasmic membrane surface (Fig. 4).

In order verify that both layers of the VAP consist only of PVAP protomers, we analyzed archaeal (*S. acidocaldarius*) or bacterial (*E. coli*) *PVAP* expression mutants constructed in a previous study (102). Whole-cell electron cryo-tomography of transformed cells revealed the

distinctive two layers in all VAPs (Fig. 4C, S3), indicating that both are formed solely by the PVAP (Fig. S3).

Subtomogram averaging of VAPs

A 3D map of the VAP was obtained by averaging 57 tomographic subvolumes of closed VAPs in the membrane of *PVAP*-expressing *E. coli* cells. Sections through the averaged volume parallel to the membrane showed clear sevenfold symmetry (Fig. S4), which was then applied to the final average, which further improved the signal-to-noise ratio (Fig. 4*D-K* movie S1). Since VAPs had different sizes, only the upper parts of the pyramids, up to a height of ~62 nm and an outer diameter of ~96 nm, were used for averaging. At the tip, the inner opening angle of the pyramid was ~72°(Fig. 4*D*). As observed in tomographic slices, the structure consisted of two layers (Fig. 4*D*, *E*, *I-K*). The outer layer formed a continuous envelope, consisting of seven triangular facets (Fig. 4*F-K*). As seen from the outside, the facets had a perceptible counter-clockwise handedness (Fig. 4*F*) and each facet was slightly convex towards the inside (Fig. 4*F, G, J, K*). The angle at the tip of the triangular facets was 35°, in accordance with previous measurements on isolated VAP fragments (102). Overall, the entire structure had the appearance of a tent or teepee (Fig. 4 *G).*

The inner, cytosolic layer of the pyramid consisted of seven triangular sheets parallel to the membrane on the outside (Fig. 4*I-K*). At the junction of two triangular sheets, the structure was perforated (Fig. 4*I, J*), suggesting weak interactions along the inner edges. The perforated seams may serve as predetermined breaking points when the VAPs open. Cross-sections through the sevenfold averaged volume revealed narrow connections between the two layers of the pyramid, suggesting that they are physically linked (Fig. 4*K*).

Membrane remodelling by PVAP

In silico secondary structure predictions (108) and hydrophobicity analysis (88) of PVAP suggest that the protein consists of an N-terminal transmembrane helix (residues 5-34), followed by three hydrophilic α-helices of two to three turns each, separated by short linker regions (Fig. S2). The PSORT-server (109) indicated that PVAP does not contain a pre-sequence. This is consistent with a previous study of the N-terminal amino acid sequence of purified PVAP of SIRV2 with Edman degradation (86), which did not indicate a cleavable signal sequence (105). Thus, we assumed that the hydrophobicity of its predicted N-terminal transmembrane segment drives the spontaneous insertion of PVAP into the lipid bilayer. To test this hypothesis, *PVAP* was expressed in the eukaryote *Saccharomyces cerevisiae*. Cells were harvested 16 hours after induction of *PVAP* expression, high-pressure frozen and freeze-substituted. In addition, chemically fixed Tokuyasu sections were immunolabeled with antibodies raised against PVAP. All samples were analyzed by transmission electron microscopy. Surprisingly, VAPs were found in most, if not all, cellular

membranes. PVAP-specific antibodies labeled VAPs in the nuclear envelope, the endoplasmic reticulum, Golgi apparatus, intracellular vesicles and mitochondria (Fig. 5).

Role of PVAP domains in VAP assembly

To identify which parts of the PVAP are required for VAP assembly, truncated mutants lacking the last 10, 20, 30, 40 or 70 C-terminal residues (*ΔC10, ΔC20, ΔC30, ΔC40* or *ΔC70*) were constructed (Fig. 6). EM analysis of *E.coli* cells transfected with these constructs revealed VAPs only in case of the *ΔC10* mutant (Fig. 6*)*.

By contrast, VAPs did not form after truncation of 20 to 70 C-terminal residues (PVAP*ΔC20, ΔC30, ΔC40 or ΔC70*, Fig. 6), corresponding to one to three C-terminal α-helical segments. Instead, expression of these constructs resulted in protein aggregates in most cases. In addition, constructs lacking 20-40 C-terminal residues caused the inner membrane of *E.coli* to form large invaginations, suggesting that these variants still interact with the membrane (Fig 6*B*). The effect was most pronounced for PVAPΔC20. In contrast, PVAPΔC70 did not produce any membrane invaginations (Fig 6*B*).

Expression of a *PVAP* construct lacking the predicted N-terminal transmembrane helix (*PVAPΔN30*) likewise did not result in VAP formation and no membrane interaction was evident (Fig. 6*B*), indicating that the N-terminal transmembrane domain is indeed required for membrane insertion of PVAP protomers.

We asked if the PVAP transmembrane domain is essential for VAP formation or could be replaced by any other transmembrane domain. To characterize the role of the PVAP transmembrane domain in VAP formation we constructed a chimera by fusing the *E. coli* flagellar regulator *Flk*, a gene encoding a single trans-membrane helix inner membrane protein (110, 111) to *PVAPΔN30*, replacing the N-terminal transmembrane helix (residues 1-30) of PVAP (110, 111). After expression, this fusion construct (PVAPtmFlk) was indeed inserted into the membrane, as judged by Western blot analysis of cell fractions with SIRV2-PVAP antibody (Fig S5). However, there was no evidence of VAPs in these cells (Fig 6*B*).

Taken together these findings indicate that the N-terminal domain is essential for membrane insertion of PVAP and for the interaction between PVAP protomers, which results in the assembly of a protein sheet on the inner membrane surface. The C-terminal domain of PVAP (except the last 10 residues, which are predicted to be disordered) is required for VAP formation. Without this domain, the protein aggregates instead of forming VAPs.

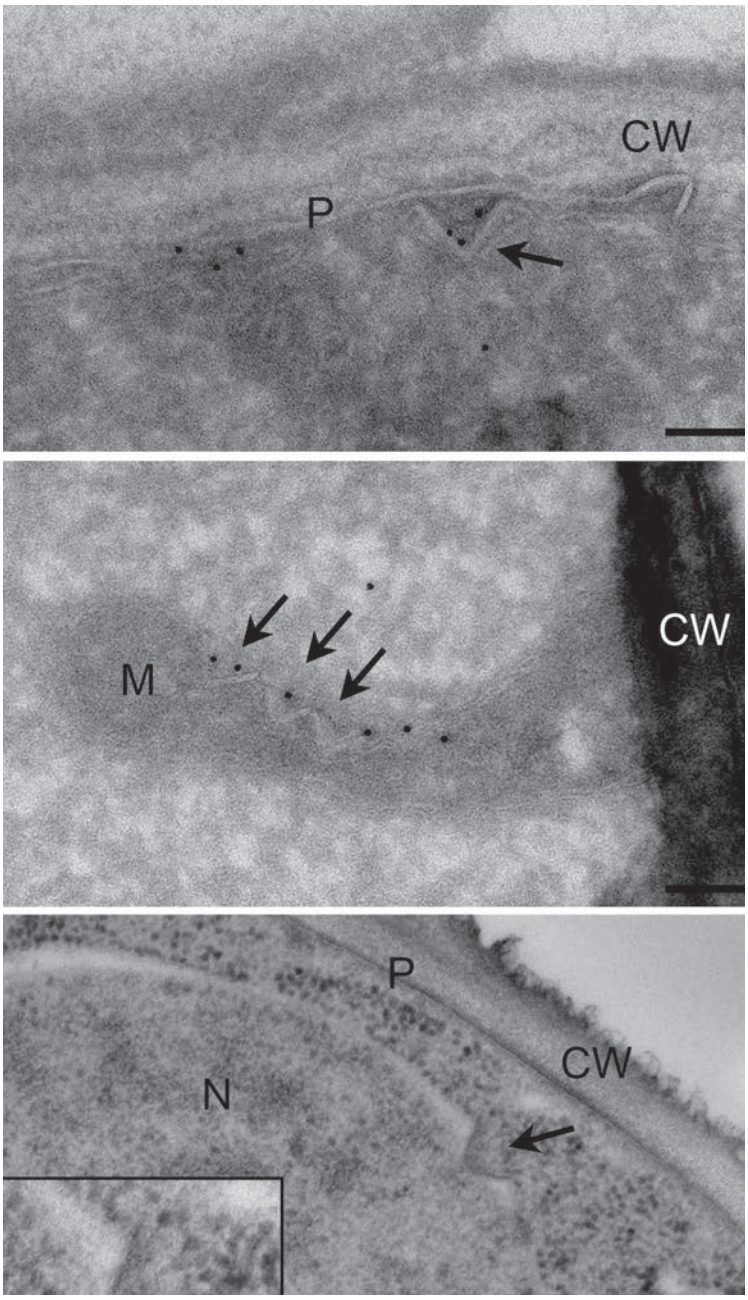


Figure 5. VAP formation in *S. cerevisiae*. *PVAP* expression in *S. cerevisiae* causes VAP formation in various cellular membranes. (**A**, **B**) Immuno-labelling of unfrozen cryo-sections with anti-PVAP antibodies. (**A**) VAP in the endoplasmic reticulum. (**B**) VAPs in mitochondrial membranes (**C**) Freeze-substituted cell with VAP in the nuclear envelope. The inset shows an enlarged VAP. P, plasma membrane; CW, cell wall; M, mitochondrion; N, nucleus. Arrows indicate VAPs. Scale bars, 200 nm.

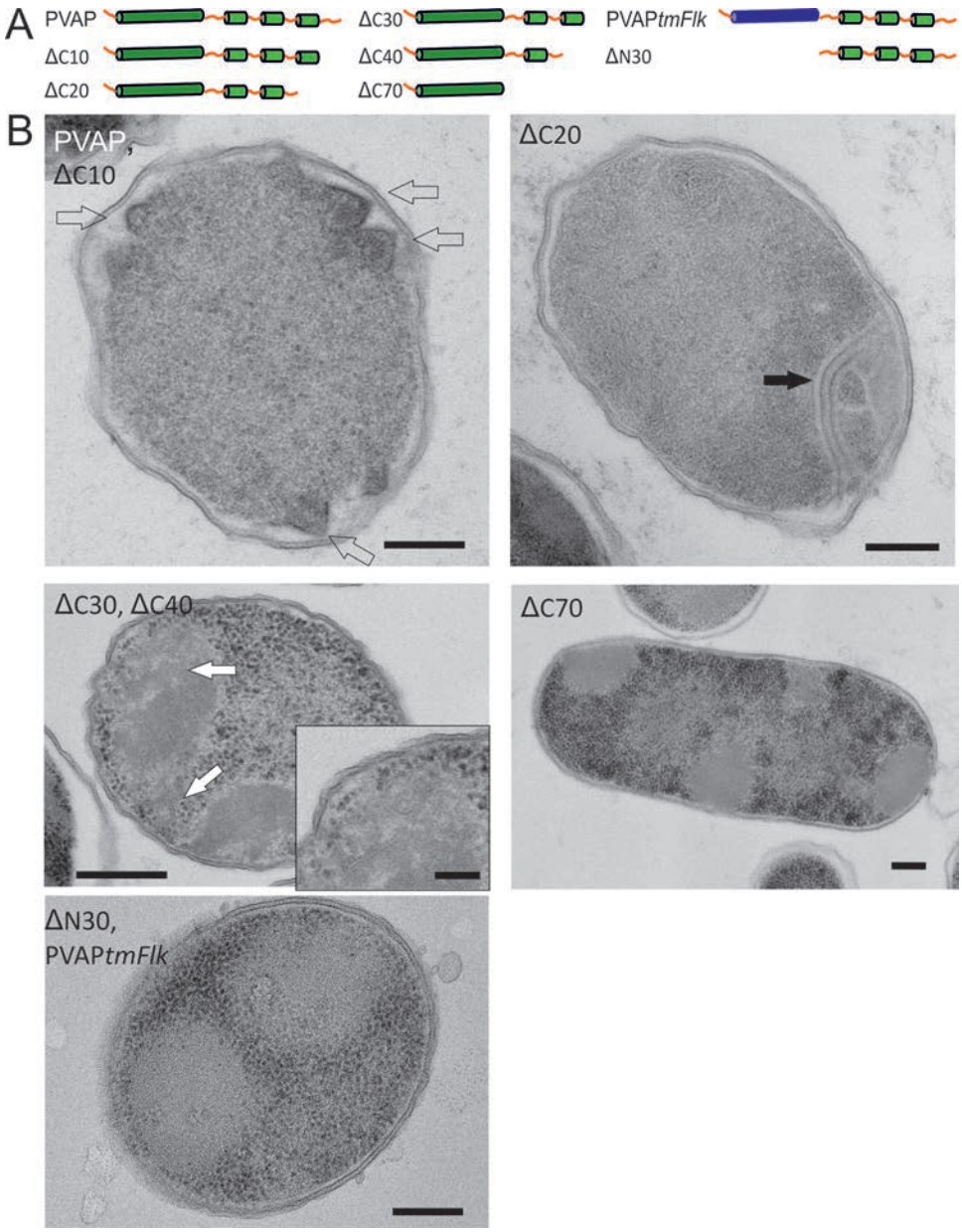


Figure 6. Expression of *PVAP* variants in *E.coli*. Several PVAP truncation mutants were constructed. (**A**) Schematic representation of PVAP constructs expressed in *E.coli*. (PVAP) wild type PVAP. (PVAP*ΔC10*) PVAP lacking 10, (PVAP*ΔC20*) 20, (PVAP*ΔC30*) 30, (PVAP*ΔC40*) 40, or (PVAP*ΔC70*) 70 C-terminal residues. (PVAP*tmFlk*) PVAP construct in which the transmembrane segment is replaced by that of the *E.coli* membrane protein Flk. (PVAP*ΔN30*) PVAP lacking the N-terminal transmembrane segment. (**B**) Electron micrographs of thin sections trough *E.coli* cells expressing PVAP constructs as in A. VAPs (open arrows) and invagination of the membrane (black arrow) are shown; Bars, 200 nm; bar in inset, 100 nm.

PVAP oligomers

To characterize the oligomerization of PVAP *in vitro*, we fused a His-tag to the C-terminus and expressed the protein heterologously in *E. coli*. Isolated membranes were solubilized with the detergent N-laurylsarcosine. PVAP was purified by nickel affinity chromatography and size exclusion chromatography. The single peak in the gel filtration profile corresponds to a molecular mass of ~70 kDa (Fig S6). SDS-PAGE analysis of peak fractions show discrete PVAP bands at ~10, 20, 30 and 70 kDa (Fig S6), indicating that in detergent solution PVAP forms different oligomers, the largest of which is most likely a heptamer. A PVAP heptamer is also suggested by gel filtration chromatography.

Discussion

The VAP, an archaeoviral egress structure that takes the shape of a large sevenfold pyramid in the host membrane, is without precedent or parallel in biology. It consists of multiple copies of PVAP, a 10 kDa membrane protein, which forms VAPs in the membrane, evidently without the need for any other cellular component.

How can a small, simple protein like PVAP assemble unaided into such an elaborate structure? To address this intriguing question, we have investigated the VAP structure itself by electron cryo-tomography, and have studied the membrane insertion and biochemical properties of PVAP. We have demonstrated that PVAP forms VAPs in archeal, bacterial and eukaryotic membranes, into which it inserts indiscriminately, and that, with the exception of the last 10 C-terminal residues, the entire length of the protein is required for VAP assembly. Finally, we have shown that PVAP forms oligomers, most likely heptamers, in detergent solution.

PVAP is a universal membrane remodeling system

Sequence analysis of PVAP suggested that the protein does not contain a signal sequence and thus most likely integrates spontaneously into the archaeal membrane. A similar mechanism of membrane insertion has been found for tail-anchored (TA) proteins (112) and for bacterial pore-forming toxins (bPTFs) (113). TA proteins are indigenous proteins, which contain a single C-terminal transmembrane segment. They are inserted into their target membrane in a Sec-independent, but organelle-specific manner, occasionally aided by cytoplasmic chaperones (112). Similar to PVAP, bPTFs are expressed as monomers and insert into the target membrane, where they assemble into pore-forming oligomers, either to kill other bacteria or, in case of pathogens, to lyse the host membrane and thus to aid bacterial proliferation (113).

Overexpression of *PVAP* in the archaeon *S. acidocaldarius*, the bacterium *E. coli* and the eukaryote *S. cerevisiae* resulted in the formation of VAPs in the plasma membranes of all hosts. Even more remarkably, VAPs were observed in virtually all cellular membranes of the eukaryote *S. cerevisiae*, including the nuclear envelope, the ER, the mitochondrial outer membrane and the plasma membrane. This demonstrates that, in contrast to other known type of protein spontaneously integrating into membranes, PVAP is able to insert into practically any biological lipid bilayer, solely by virtue of its N-terminal transmembrane segment. Once inserted into the bilayer, it forms sevenfold pyramids, irrespective of fundamental differences in lipid or protein composition of the target membrane. These characteristics render PVAP a unique, universal membrane remodeling tool.

Supramolecular organisation of VAPs

Whole-cell electron cryo-tomography and subtomogram averaging revealed that the VAPs consist of the same two layers of roughly equal thickness in all endogenous and heterologous expression hosts. The outer layer was continuous with the plasma membrane, whereas the inner layer formed a discontinuous sheet at the cytoplasmic membrane surface. As PVAP must insert with its N-terminal TM segment into the plasma membrane, the outer layer most likely consists of multiple copies of this part of PVAP plus interspersed membrane lipid. In turn, the cytoplasmic protein sheet must consist of the tightly associated C-terminal domains of the protein. The low-density region between the two layers would then account for the linker region between the cytoplasmic domains and the N-terminal trans-membrane segments (Fig. 7A).

Recently it was suggested that the opening of STIV-induced VAPs depends on polymerization of ESCRT-III homologs, resulting in the 'stripping' of VAPs from their cytoplasmic membrane (114). Our observation that the outer layer of the double-layered VAP structure is continuous with the plasma membrane (Fig 4) does not support this model. Moreover, global analysis of host gene expression during the SIRV2 infection cycle documented that ESCRT-III-like proteins were down-regulated in infected cells (115). Thus, it is possible, although not very likely, that mechanisms of VAP opening differ in SIRV2-infected and STIV-infected cells.

Our tomograms of closed and open VAPs offer intriguing insights into the mechanism by which the pyramids open to facilitate viral egress. The seven edges of the closed pyramid are slightly curved in a right-handed fashion (Fig. 4 F). In the open state, the edges of the seven individual facets are curved in the same way (Fig. 4 F). In the open VAP, each facet curls outwards (Fig. 4 F). This suggests that in the closed state, the VAP structure is under mechanical tension. This tension is likely to provide the energy required for VAP opening, which occurs along the seams of the pyramid. At these seams, the inner layer of the VAP is perforated (Fig. 4 I-K), most likely to provide predetermined break points for VAP opening.

VAP assembly and opening

We propose a model for VAP assembly (Fig. 7). Upon synthesis in the cytosol, PVAP integrates spontaneously into the cell membrane, depending solely on the hydrophobicity of its N-terminal transmembrane helix segment. In the membrane, intimate interactions between PVAP protomers result in the formation of protein sheets, which consist of the observed two layers. We propose that this interaction involves both the N-terminal transmembrane helices and the C-terminal hydrophilic PVAP domains, as indicated by the expression of truncation constructs.

In order to assemble into a pyramid instead of a flat sheet, at least two different kinds of interactions between PVAPs are necessary, one in-plane interaction within the triangular facets, and one out-of-plane interaction, at the edges of the pyramid. The interaction at the edges is

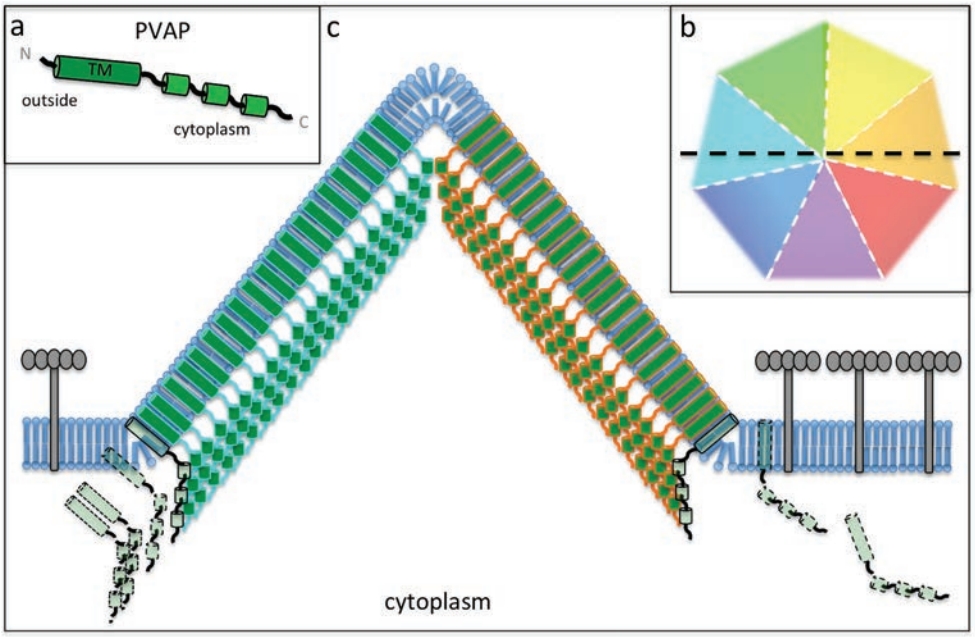


Figure 7. Model of VAP assembly. (**A**) Predicted secondary structure of PVAP, with N-terminal transmembrane helix and three short C-terminal α-helices. (**B**) Schematic top view of the VAP, with its seven facets in different colours. Strong interactions between individual PVAP protomers stabilise the integrity of each facet, whereas weak interactions at the seams (white dashed lines) between two neighbouring facets form predetermined break points. The black dashed indicates the cross section through the VAP in in C. (**C**) PVAP protomers (green) insert spontaneously into the plasma membrane (blue), with the short C-terminal helices exposed to the cytoplasm. Close contacts between PVAP molecules in each facet (aqua and orange outline) exclude S-layer proteins (grey), and combine into a protein sheet below the plasma membrane. Addition of PVAP units at the base pushes the pyramid outwards.

evidently weaker than the in-plane interaction, so that the pyramids open at these pre-formed seams.

At present it is unknown if the pyramids are built in a one-by-one self-assembly process from individual PVAP protomers or if, instead, protomers first assemble into heptamers upon membrane insertion, and that the heptamers then combine into pyramids in a second stage of assembly. The prior formation of heptamers in the membrane is suggested by the gel filtration experiments, which show one homogenous peak of ~70 kDa in detergent solution. Given that detergent mimics the hydrophobic membrane environment, it is not unlikely that the same interactions that give rise to the heptamer in a detergent micelle would also promote the formation of heptamers in the membrane, which may thus be the building blocks of the pyramids.

VAP opening presumably involves a host or virus-specific factor, since the pyramids only open in virus-infected *Sulfolobus* cells but remain closed in PVAP-expressing bacteria and yeast. With this factor, the PVAP system can be developed into a universal system for the controlled opening of large, ~100 nm apertures in any lipid bilayer. Such a system would evidently have great potential for basic research, biotechnological applications, and therapy.

Experimental Procedures

Virus and host strains

The SIRV2 virus stock was prepared and the *S. islandicus* LAL 14/1 strain grown as described previously (101). For analysis of morphological changes upon viral infection, cells were synchronized by dilution of precultures in fresh medium. The cultures were grown for ~12 hours until an OD of 0.1-0.2 was reached and SIRV2 was added directly to the cultures as described (101).

Plasmid constructs and transformation of *S. acidocaldarius*

For the expression of *PVAP* in *S. acidocaldarius* M31, SIRV2_ORF98 (NCBI RefSeq ID: NP_666583) was amplified from SIRV2 genomic DNA and cloned into the lacS gene locus in the pMZ1 plasmid (116) using the NcoI and BamHI sites. SIRV2_ORF98 (PVAP) and the araS promoter were transferred from pMZ1 to the pSVA1450 plasmid (117) using the NcoI and EagI sites, which yielded pTQ26. pTQ26 was transformed to *S. acidocaldarius* M31. The preparation of competent cells, methylation of the plasmid, and electroporation were carried out as described (118) using a Bio-Rad Gene Pulser Xcell electroplater with 1-mm cuvettes at 1,500 V, 600 Ω, and 25 µF. Selection of *PVAP* expressing colonies and induction of expression were performed as before (102).

Plasmid constructs and transformation of *E.coli*

For the overexpression of *SIRV2_ORF98* (*PVAP*) C-terminal truncation mutants in *E. coli* Rosetta(DE3)pLys (Novagen Merck), SIRV2_ORF98 was amplified from SIRV2 genomic DNA with different reverse primers resulting in PCR products of 297 (full length), 267 (-10 AA), 237 (-20 AA), 207 (-30 AA), 177 (-40 AA), 87 (-70 AA) bp. PCR amplification of a 216 bp product starting 81 bp downstream of the ATG resulted in a PVAP mutant lacking the N-terminal transmembrane domain. The same sequence was used for a fusion with the 75 bp transmembrane segment of the *E.coli Flk* gene, preceded by the 27 N-terminal bp of PVAP. This oligonucleotide sequence was synthesized by GeneArt® (Invitrogen). All *PVAP* gene mutants were cloned into the T7 promoter–driven expression vector pSA4 using the NcoI and BamHI sites(119). The pSA4 vector contains an isopropyl β-d-1-thiogalactopyranoside–inducible promoter that was used for the expression of a C-terminally His-tagged protein. Analysis of *PVAP* expression cultures by high pressure freezing and freeze substitution was performed as described (102). The location of the Flk-PVAP fusion was analyzed by isolation of membranes as described in ref (105), followed by SDS PAGE, western blotting and immunolabelling with antibodies raised against PVAP as described (102).

Plasmid constructs and transformation of *S. cerevisiae*

For the overexpression of *SIRV2_PVAP* (*PVAP*) in *Saccharomyces cerivisiae, SIRV2_PVAP* was amplified from SIRV2 genomic DNA and cloned in the expression vector pCM190 (120) using PstI and XbaI sites. *S. cerivisiae* was transformed with the plasmid according to Gari et al (120). After selection on plates without uracil a single colony was picked and grown at 30 °C overnight in a pre-culture in selective uracil-free medium, and with 10 microgram/ml doxycyclin. After one day, cells were diluted 1/1000 in medium without doxycyclin.

Immuno-electron microscopy

Yeast cells were fixed with 4% paraformaldehyde in 0.1M Hepes, pH 5.4 for 2 h at RT. The cells were then washed with 50 mM NH_4Cl in PBS to quench free aldehyde groups and pelleted in 12% gelatin in PBS. The gelatine pellet was solidifed on ice and cut into small blocks, which were infiltrated over night at 4°C with 2.3M sucrose for cryoprotection, mounted on aluminium pins and frozen in liquid nitrogen. Thin sections were cut with a UC6/FC6 (Leica microsystems, Vienna, Austria) and picked up in a 1:1 mixture of 2.3M sucrose and 2% methylcellulose (121). Labelling for PVAP was done as described previously (102).

High-pressure freezing and freeze-substitution

E.coli cells were taken up in cellulose capillary tubes (Leica Microsystems GmbH, Vienna, Austria) as described in (122). *S. cerevisiae* cultures were concentrated by filtration. Tubes or

cell concentrates were placed into brass planchettes filled with 1-hexadecen (Agar Scientific, Stansted, United Kingdom). Samples were frozen in a HPM 010 freeze-fracture device (Baltec, Switzerland).

Freeze-substitution was performed in anhydrous acetone containing 2% osmium tetroxide (Merck, Germany). Small cracks were introduced under liquid nitrogen in solid 1-hexadecen by pre-cooled fine point forceps (No 5, Dumont, Switzerland) to allow perfusion of the substitution mix. Freeze-substitution was carried out at -90°C for 24h, and at -60°C and -30°C for 8h each in a freeze substitution device (Leica Microsystems GmbH, Vienna, Austria). Afterwards the temperature was raised to 0°C and the samples were washed with dry acetone and embedded stepwise in EPON. After heat polymerization thin sections were cut with an Ultracut UCT microtome (Leica Microsystems GmbH, Vienna, Austria). Sections were collected on 200 mesh Formvar-coated copper grids and post-stained with 4% uranylacetate and Reynold's lead citrate. Images were recorded with a JEOL 1010 electron microscope at 80 kV equipped with an Olympus Keen View camera (Olympus Soft imaging systems, Münster, Germany).

Whole cell cryo-tomography

Cells were harvested shortly after infection, concentrated by low-speed centrifugation (3000 rpm for 10 min) and plunge-frozen directly in the growth medium. For this, cell pellets were resuspended in an equal volume of fresh medium. Before freezing, suspensions were mixed with an equal volume of 10 nm colloidal protein-A gold suspension (Aurion, Wageningen, The Netherlands). 3 µl of this mixture were added to a 300 mesh R2/2 glow-discharged Quantifoil grid, blotted and rapidly injected into liquid ethane.

Tomograms were recorded with a Polara G2 Tecnai field emission transmission electron microscope (FEI, Hillsboro, USA) operated at 300 kV, equipped with a Gatan Tridiem energy filter and 2x2 k CCD camera (Gatan Inc., Pleasanton, USA). Zero-loss filtered images were collected using the Digital Micrograph software (Gatan Inc., Pleasanton, USA). Tomographic tilt series were recorded with the FEI tomography software (FEI Company, Hillsboro, USA). Tilt series were generally collected in a range of -60° to +60° in steps of 1.5° or 2°, at 6-9 µm defocus and magnifications of 41.000x, 34.000x or 27.500x, corresponding to a pixel size of 0.5766nm, 0.709nm or 1.073nm, respectively. Tomograms were reconstructed using the IMOD software package (123) and de-noised by non-linear anisotropic diffusion (NAD) (124).

Subtomogram averaging

For subtomogram averaging of VAPs, 57 pyramid volumes were cut out, aligned and averaged using the PEET software (125, 126) as described before (127). The average volumes were averaged further by applying 7-fold rotational symmetry. 3D maps obtained by subtomogram averaging were displayed and analyzed in 3Dmod (IMOD,(123)) or UCSF chimera (128).

PVAP purification

The codon-optimized gene coding for PVAP was synthetized (Genscript) and inserted in to the expression plasmid pET26b (Novagene) using the restriction sites NdeI and XhoI. *E. coli* BL21DE3/Rosetta/pLysS cells were transformed with the resulting plasmid and transformants were selected using kanamycin (Kan, 50µg/ml) and chloramphenicol (Cam, 34µg/ml). An overnight pre-culture of a single colony was transferred to 12L "terrific broth" (TB) medium containing Kan and Cam and incubated in the shaker at 160 rpm until the optical density at 600nm (OD_{600}) reached 0.8 - 1.0. Protein expression was induced with 1mM IPTG (isopropyl-β-D-1-thio-galacatopyranoside) at 37°C. After 2 hours, cells were pelleted, resuspended in lysis buffer (50 mM Tris pH 7.0, 300 mM NaCl, 0.5 mM PMSF) and disrupted with a Microfluidizer (M-110L, Microfluidics Corp., Newton, MA). Unbroken cells were removed by 30 min. centrifugation at 14,000g. The membrane fraction was pelleted by centrifugation of the supernatant at 100.000g for 90 min. at 4°C ,resuspended in 50 mM Tris pH 7.0, 300 mM NaCl and diluted to a protein concentration of 5 mg/ml. Membranes were solubilized by adding of N-laurylsarcosine to a final concentration of 1.5%. Non-solubilized protein was removed by centrifugation for 60 minutes at 100.000g. The supernatant containing the His-tagged protein was loaded onto a Ni-NTA column and unspecifically bound proteins were removed in several washing steps. The protein was eluted in buffer containing 500 mM imidazole and concentrated with a 30 kDa cutoff prior to loading onto a gel filtration column (Superdex75). Purified protein was eluted using 50 mM Tris pH 7.0, 300 mM NaCl and 0.05 % N-laurylsarcosine as running buffer.

Acknowledgements

We thank John van der Oost and Alain Jacquier for helpful discussions. T.E.F.Q. was supported by a fellowship from the Ministère de l'Enseignement supérieur et de la Recherche of France. WK and BD acknowledge financial support from the Max Planck Society.

Supplemental figures

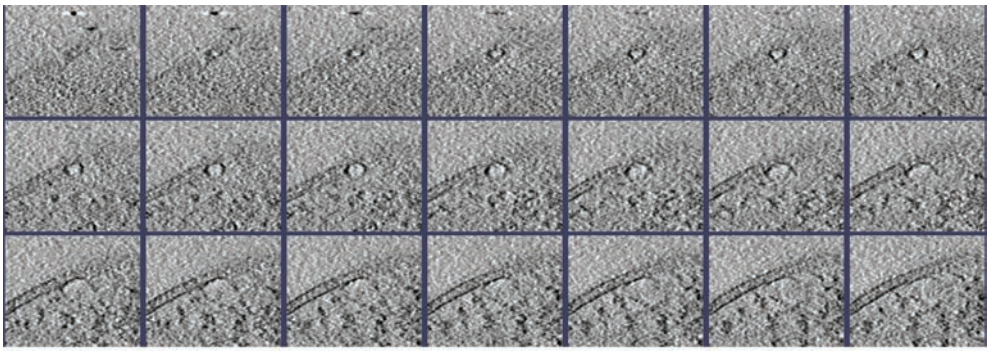


Figure S1. Early stage of VAP assembly. Consecutive tomographic slices through nascent VAP at 3 h.p.i. in *S. islandicus* host cell. Black arrows indicate VAP. PM, plasma membrane. Scale bar: 100 nm

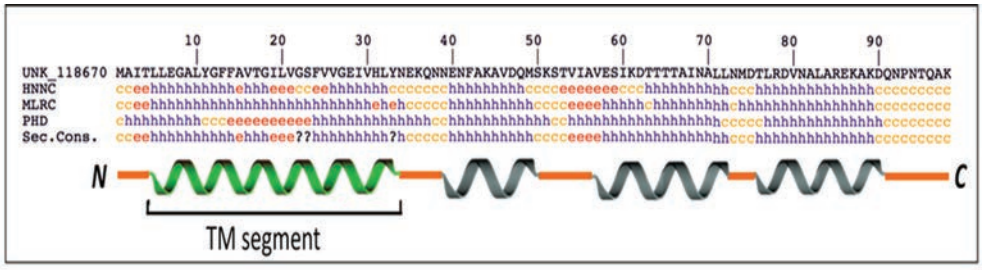


Figure S2. Secondary structure prediction of PVAP. Sequence of wild type PVAP (black lettering) with predicted secondary structure as indicated. Blue, predicted α-helices (**a**); yellow, coils (**c**); red, strands (**e**). TM segment, predicted trans-membrane segment.

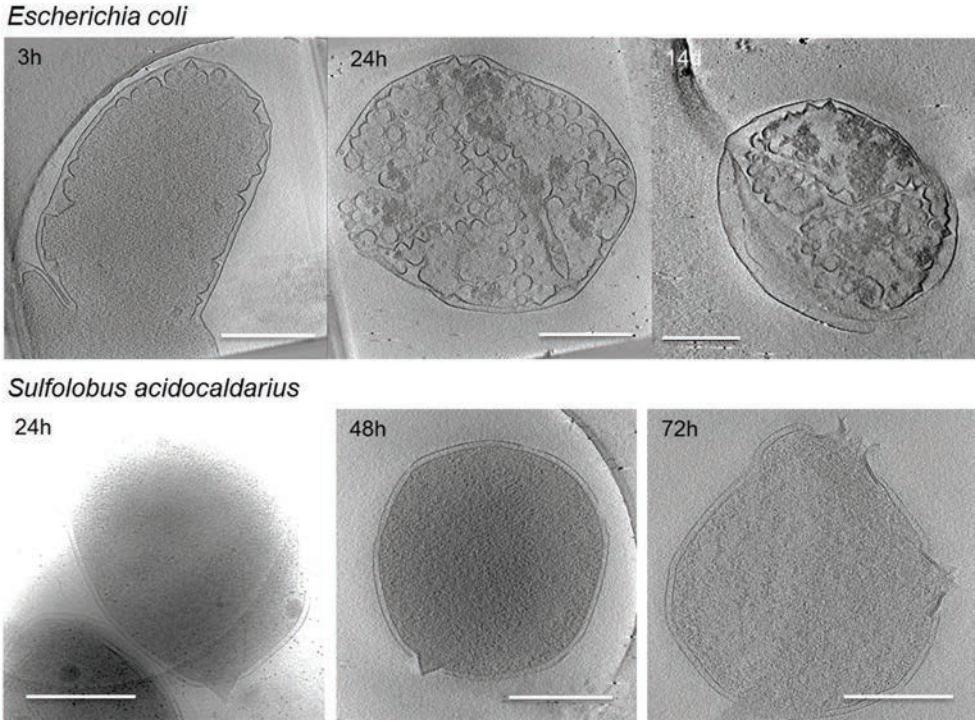


Figure S3. VAPs in *E. coli* and *S. acidocaldarius*. Tomographic slices through *E. coli* (upper panel) and *S. acidocaldarius* (lower panel) cells expressing PVAP. VAPs in *E.coli* remain closed for longer than 2 weeks. VAPs in *S. acidocaldarius* open ~72 hours after PVAP induction; bars, 500 nm.

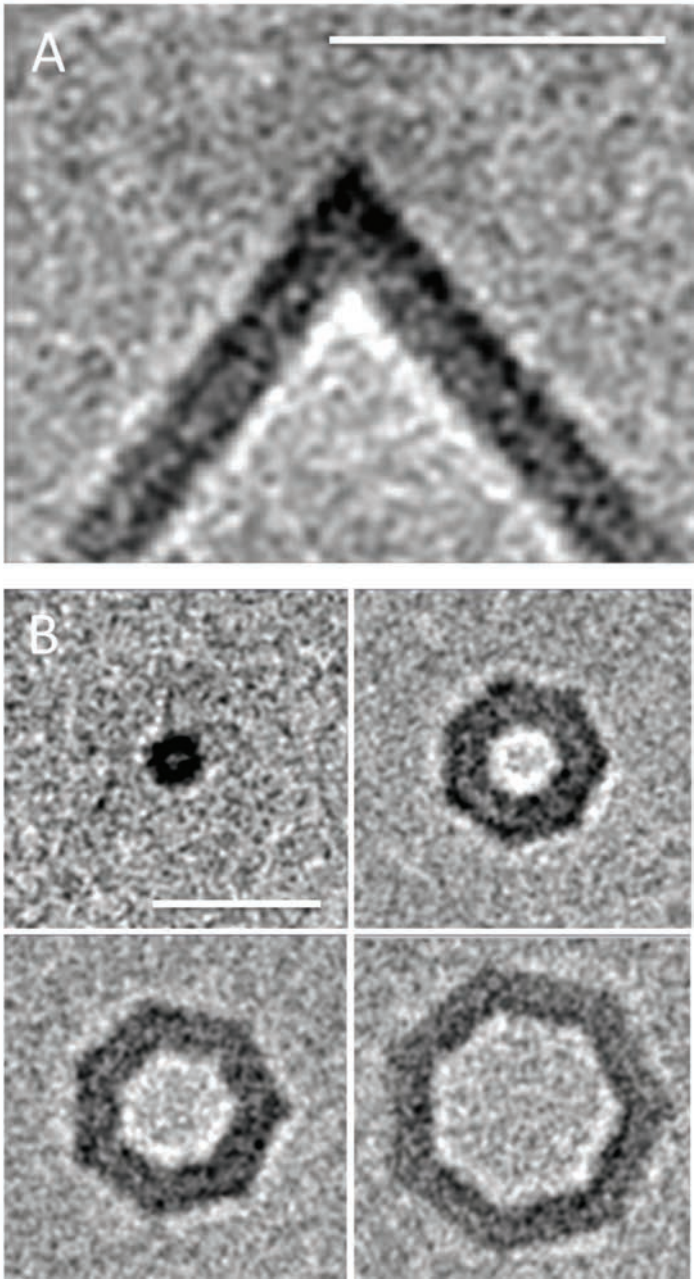


Figure S4. Unsymmetrised subtomogram average of VAPs. Average of 57 VAPs obtained from *E. coli* cells overexpressing PVAP. Slices run perpendicular (**A**) or parallel (**B**) to the base of the pyramid. The unsymmetrized average shows clear 7-fold symmetry around an axis perpendicular to the base. Scale bars: 50 nm.

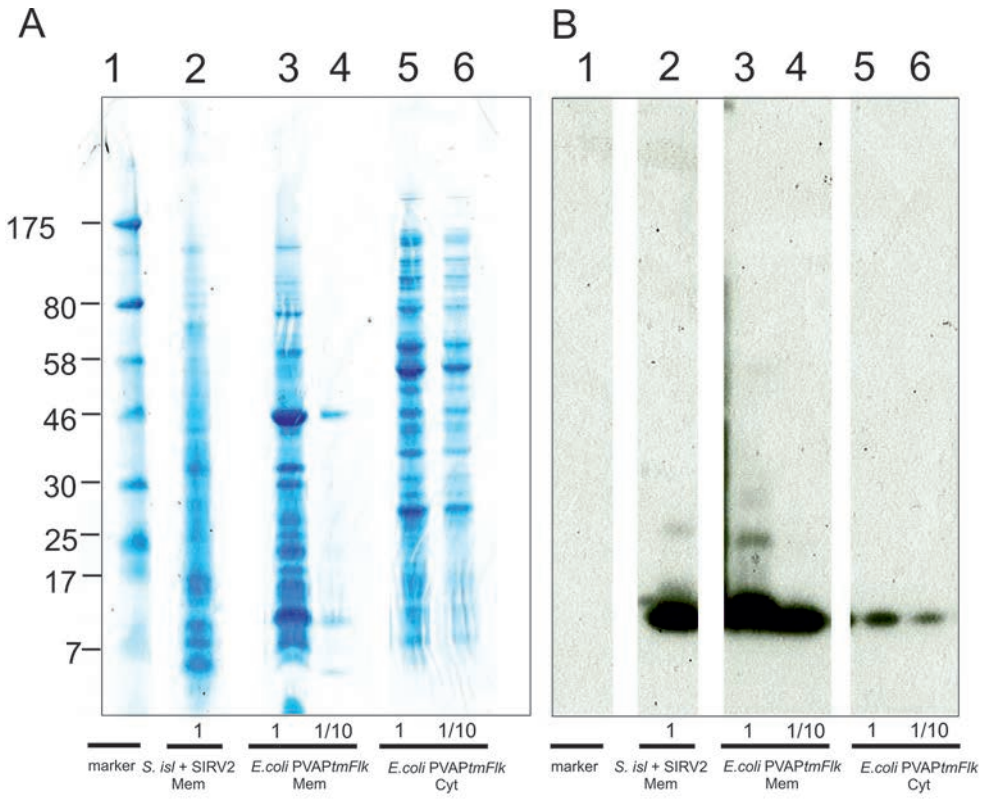


Figure S5. Flk-PVAP fusion protein localizes to *E.coli* membranes. SDS/PAGE of: 1, marker; 2, SIRV2-infected *S. islandicus*, 10 hpi; 3-4, membrane fraction of *E. coli* expressing Flk-PVAP fusion (PVAP*tmFlk*); 5-6 cytosolic fraction of *E.coli* expressing PVAP*tmFlk*. The *E. coli* samples were loaded undiluted or diluted 1:10. The positions of proteins with known molecular mass (in kDa) are indicated by bars based on the marker loaded in the first lane. (**A**) Coomassie Blue–stained gel. (**B**) Western blot of a duplicate gel with antibodies against SIRV2-PVAP. Mem, membrane fraction; Cyt, cytosolic fraction

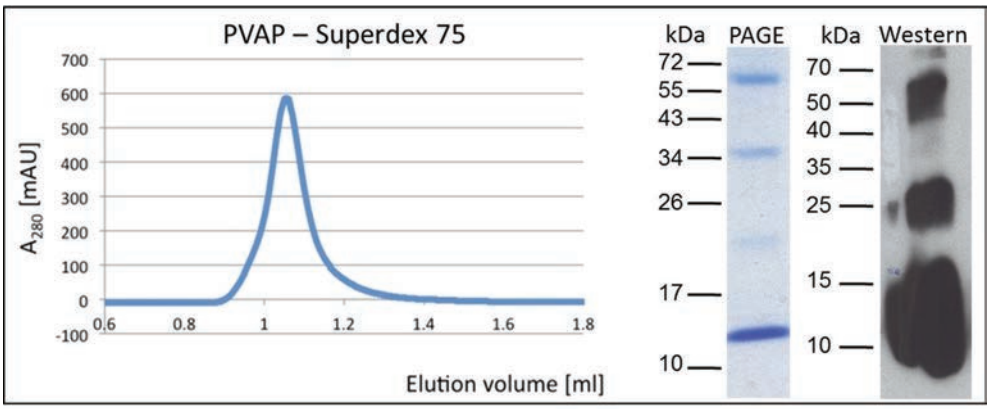


Figure S6. Analytical size exclusion chromatography of PVAP. Elution profile of PVAP on the Superdex75 column shows a symmetrical peak eluting at 70 kDa. SDS-PAGE of the peak fraction and Western-Blot analysis with antibodies against PVAP indicate 4 protein bands at the level expected for the PVAP monomer, dimer, trimer and heptamer.

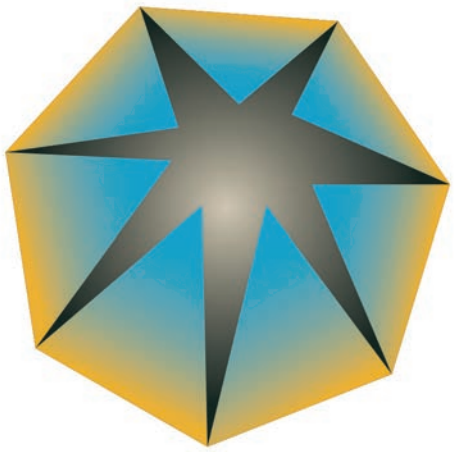
Chapter 5

Transcriptome analysis of the SIRV2 infection cycle

Tessa .E.F. Quax, Marleen Voet, Odile Sismeiro, Marie-Agnes Dillies, Bernd Jagla, Jean-Yves Coppee, Guennadi Sezonov, Patrick Forterre, John van der Oost, Rob Lavigne, David Prangishvili

'Massive activation of archaeal defense genes during viral infection'

Journal of Virology. 2013 Aug. 87(15): 8419-28



Abstract

Archaeal viruses display unusually high genetic and morphologic diversity. Studies of these viruses proved to be instrumental for the expansion of knowledge on viral diversity and evolution. The *Sulfolobus islandicus* Rod Shaped Virus 2 (SIRV2) is a model to study virus-host interactions in archaea. It is a lytic virus that exploits a unique egress mechanism based on formation of remarkable pyramidal structures on the host cell envelope. Using whole transcriptome sequencing we present here a global map defining host and viral gene expression during the infection cycle of SIRV2 in its hyperthermophilic host *S. islandicus* LAL14/1. This information was used, in combination with a yeast two hybrid analysis of SIRV2 protein interactions, to advance current understanding of viral gene functions. As a consequence of SIRV2 infection, transcription of more than one third of *S. islandicus* genes was differentially regulated. While expression decreased of genes involved in cell division, those playing a role in anti-viral defense were activated on large scale. Expression of genes belonging to Toxin-Antitoxin and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas systems was specifically pronounced. The observed different degree of activation of various CRISPR-Cas systems highlights the specialized functions they perform. The information on individual gene expression and activation of anti-viral defense systems is expected to aid future studies aiming at detailed understanding of functions and interplay of these systems *in vivo*.

Introduction

The knowledge on virus-host interactions in the third domain of life, the Archaea, remains limited despite increasing numbers of described archaeal viruses. The available information mainly concerns unusual morphological and genomic properties of these viruses and structures of proteins they encode (129, 130). Several viruses infecting hyperthermophilic Crenarchaeota have emerged as suitable models to study molecular details of the life cycle of archaeal viruses, e.g. *Sulfolobus* turreted icosahedral virus (STIV) (39), Sulfolobus spindle-shaped virus (SSV) (131) and *Sulfolobus islandicus* rod-shaped virus (SIRV) (31). They cover examples of lytic (STIV and SIRV2) and temperate (SSV) viruses and fundamentally differ from each other in virion morphology and genome organization. For these viruses and their hosts, attempts have been made to analyze patterns of gene expression in the course of the infection cycle (51, 132-134). The temperate spindle-shaped virus SSV1 has a circular dsDNA genome, which can integrate into the host genome, establishing a lysogeny which can be reverted by UV-irradiation (135, 136). A chronological regulation of transcription of viral genes was observed during a microarray study following induction of SSV1 replication caused by UV irradiation. In this case, hardly any differences of host gene expression were detected (132). A completely different situation was encountered when studying transcription of the lytic viruses STIV and SIRV2 after infection of *Sulfolobus* host cells, by microarray and Northern hybridization analysis, respectively (51, 134). In both cases, little temporal control of viral gene expression was detected. During STIV infection, a high proportion of host genes was either up or down regulated (133). Genes involved in transcription, translation and anti-viral defense mechanisms were over represented amongst the differentially expressed genes (133).

The virus SIRV2 is an appealing model for the study of archaeal virus-host interactions, since the infection has a pronounced and clear effect on the host cell (83, 102, 105). SIRV2 is a member of the family *Rudiviridae* and infects the hyperthermophilic archaeaon *S.islandicus* LAL14/1, which thrives at 80 °C and pH 3. The virus genome is dsDNA of ~35 kb and encompasses 54 open reading frames (ORFs) (60, 137). SIRV2 is a lytic virus and degradation of the host DNA occurs after infection, i.e. 5 hours post infection (h.p.i.) in ~40% of cells the chromosome is degraded (83). At late stages of the infection cycle, multiple pyramidal shaped structures of up to 200 nm in diameter are observed on the surface of each infected cell. These Virus Associated Pyramids (VAPs) comprise sevenfold rotational symmetry and consist of multiple copies of the viral encoded protein SIRV2_P98 (NP_666583.1) (102, 105). At this stage of the infection cycle, mature virions are present in the cell in 2-3 bundles of up to 50-100 virions (83). As the final step of the infection cycle, the VAPs open up outwards creating large apertures through which the mature virions escape the host cell (83). Thus, cell morphology and metabolism are dramatically affected by SIRV2 infection.

We studied the interplay between SIRV2 and its host by monitoring changes in expression of the viral and host genes during the infection cycle using deep transcriptome sequencing (RNAseq). In addition, we performed a yeast two hybrid screen, the results of which could be used in combination with gene expression profiles to predict roles of viral proteins in currently unknown processes. This approach uncovered a mild temporal regulation of viral gene expression, but dramatic changes of gene expression of the host. More than one third of all host genes were differentially transcribed, with a clear bias towards genes involved in cell division and defense against foreign genetic elements.

Results

SIRV2 infection of *S.islandicus* and RNA isolation

Cells of *S. islandicus* LAL14/1 were infected by the virus SIRV2 in conditions causing infection of 95% of the cell population (5) and resulting in severe retardation of cell growth (data not shown). Electron microscopical analysis of a time series of cells after infection revealed typical morphology of SIRV2 infected cells, with bundles of mature virions in the cytoplasm and the virion egress structures, VAPs, appearing on the cell surface (Fig. 1).

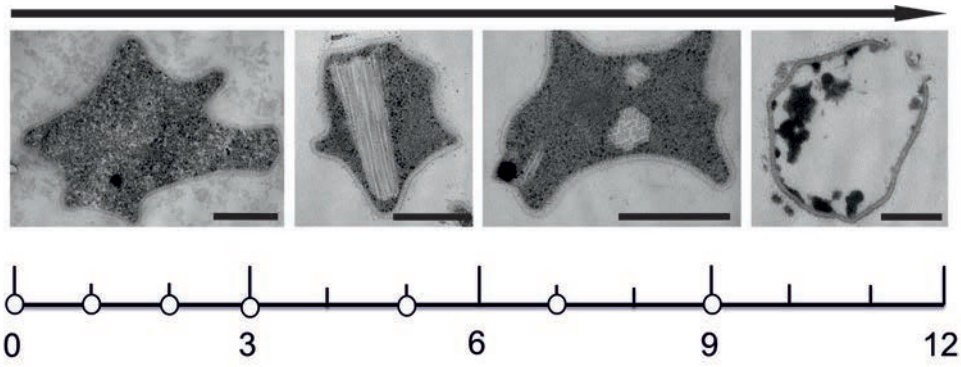


Figure 1. Time scale of SIRV2 infection cycle in the *Sulfolobus* host cell, indicating time points of RNA isolation (white dots) in hours post infection. Above the time scale are shown electron micrographs of thin sections of representative SIRV2 infected *S.islandicus* cells at each stage of infection. During the first hours after infection (t<3 h.p.i) cells appear similar to uninfected control cells. Later bundles of virions assemble in the cytoplasm (t=3-6 h.p.i), followed by VAP formation (t= 6-9 h.p.i.) and eventually opening of the VAPs and release of the virions (t= 10-12 h.p.i.). Bars, 500 nm.

Time points for RNA analysis were selected based on the available information on the SIRV2 infection cycle (83). Total RNA was isolated 0, 1, 2, 3, 5, 7 and 9 h.p.i. from two biological duplicates. For control experiments, RNA was isolated at the same time points from two independent uninfected control cultures. Whole-transcript sequencing was applied using the Illumina RNAseq technology. The resulting reads were mapped to coding sequences of the *S.islandicus* LAL14 and SIRV2 genomes. Reads that mapped to ribosomal RNAs were discarded (Methods). The fraction of reads mapping to rRNA sequences was approximately the same for all samples. This yielded a total of ~4 million whole-transcript reads per sample mapped to viral and host genomes.

Identification of differentially regulated genes

For each gene, the mapped reads of biological duplicates were averaged and the accumulation levels of specific RNAs were compared between all samples of each time point/condition using both DEseq and EdgeR analysis programs (138, 139). Very similar sets of significantly differentially expressed genes were obtained with both programs. Differences were only observed for poorly expressed genes or for genes that displayed little change in expression. Genes marked as significantly differentially expressed by both programs were retained for further analysis.
Uninfected control samples of each time point had highly similar gene expression profiles. In contrast, pronounced variations of gene expression were observed between the control and infected samples. More than one third of all *S. islandicus* genes were significantly up or down regulated in infected cells, when compared with control cells (Fig. S3). The gene expression profiles of biological duplicates displayed extensive similarity, indicating that the differential expression is not a direct consequence of the SIRV2 induced chromosome degradation. In addition, a prominent variation was observed between gene expression of infected cells sampled at various time points after infection. Comparing infected samples harvested at different times, between 10 and 50% of all genes was differentially expressed (Fig. S3).

In figure 2, a heat map is presented which shows gene expression profiles of *S.islandicus* genes of all analyzed samples. The control samples cluster together, which demonstrates the extended similarity in gene expression between uninfected samples. The infected samples also cluster, but at a very different position, stressing the variation in gene expression profile with respect to the control cultures (Fig. 2). One exception is the first harvested infected sample (t=0 h.p.i.) of which RNA was isolated just after addition of the virus; as expected, this sample closely resembles the uninfected samples. Divergence is observed between expression patterns of infected samples collected at different time points, which is indicative of extensive temporal regulation of expression in the course of the infection cycle (Fig. 2).

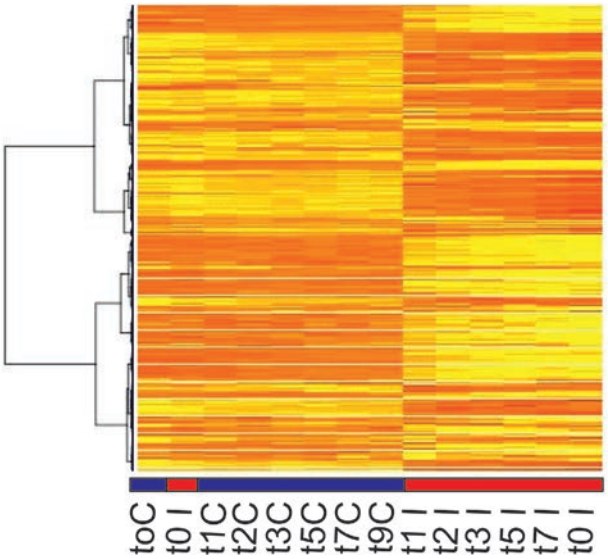


Figure 2. Heat map of *S.islandicus* gene expression patterns of uninfected control (blue line) and SIRV2 infected (red line) cells. Up and down regulated genes are showed in yellow and red respectively. Genes are clustered according to their expression profile in all samples. t, the time (hours) after infection RNA was isolated. C, uninfected control cells. I, SIRV2 infected cells.

Viral genes

Substantial numbers of reads mapping to the SIRV2 genome were identified in infected samples, but never in control samples. The number of viral reads increased steadily in the course of infection, reaching a plateau around 5 h.p.i. at which time point ~20% of all transcripts (excl. rRNA hits) mapped to the viral genome (Fig. 3). Although the total number of reads mapping to the viral genome was still relatively low in the first sample taken after infection (t=0 h.p.i.), a clear trend could be observed in the genomic location. Viral reads at t=0 h.p.i. mapped typically to ORFs at the distal ends of the linear dsDNA genome, indicating that transcription starts simultaneously from both ends of the SIRV2 genome (Fig. 4 and Table S2). Expression was exclusively observed from the predicted open reading frames (ORFs) and hardly any reads mapped to the non-coding terminal repeat regions of the genome. ORF83a (NP_666535.1) and ORF83b (NP_666588.1) were the highest expressed genes at this time point and are located on either end of the linear dsDNA genome. They have identical nucleotide sequences and could only be distinguished by differing sequences in their untranslated regions, showing that ORF83a and b are expressed approximately in 2:1 ratio. All other expressed genes at t=0 h.p.i. were located close to the genome termini: ORF119a (NP_666536.1), ORF103a (666537.1) and ORF119b (NP_666587.1), of which the former codes a protein with unknown function and the latter two code a proteins belonging to the DUF1374 superfamily, of which several members are present in the genomes of SIRV2 and other

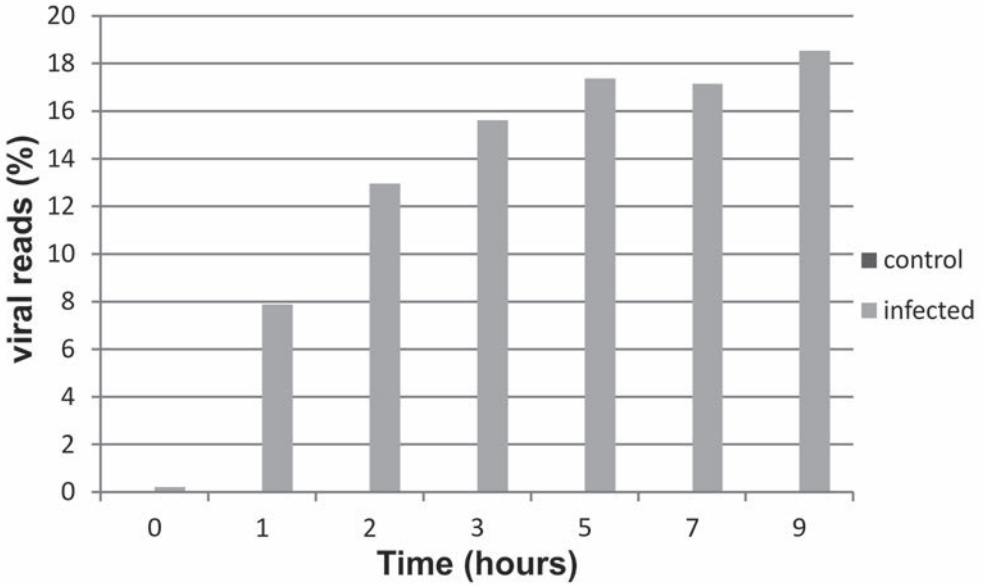


Figure 3. Percentage of viral transcripts increases in time. A plot is shown in which reads mapping to the viral genome are shown as the percentage of the total number of detected reads. Depicted are control (dark grey) and SIRV2 infected (light grey) samples of different time points after infection.

archaeal viruses. In samples from all time points (except t=0 h.p.i.), significant expression of practically all viral genes was detected. Very poor expressed genes (<250 RPKM) were ORF76 (NP_666559.1) and ORF119c (NP_666550.1), of which the former is coding for a protein of unknown function and the latter for a Rep protein proposed to initiate SIRV2 replication (140). A negligible number of reads mapped to regions outside annotated ORFs, besides those from intergenic regions on polycistronic messengers. However, there were a few exceptions: the region between ORF119c (NP_666550.1) and ORF131b (NP_666551.1), and the region between ORF156 (NP_666563.1) and ORF64 (NP_666564.1). A moderate number of reads mapped in the 5' to 3' direction to both regions. Both transcripts likely contain protein coding genes, which were not predicted in the original annotation due to their short length (<150 bp) (137). Expression of the majority of viral genes increased in the course of infection. A few reached the highest level of expression at one or two h.p.i. and decreased later during infection (Fig. 3, Table S2). In all infected samples (except t=0 h.p.i.) the gene expression appeared to be randomly distributed over the SIRV2 genome. At one h.p.i., the most abundantly transcribed genes were: ORF83a, ORF83b , ORF56b (NP_666549.1), ORF108 (NP_666585.1) and ORF103a (NP_666537.1) (Fig 4 and Table S2). ORF56b encodes the transcriptional regulator SvtR, which acts as a repressor of a number of viral genes, most importantly ORF98 and ORF1070 (NP_666572.1) that code for

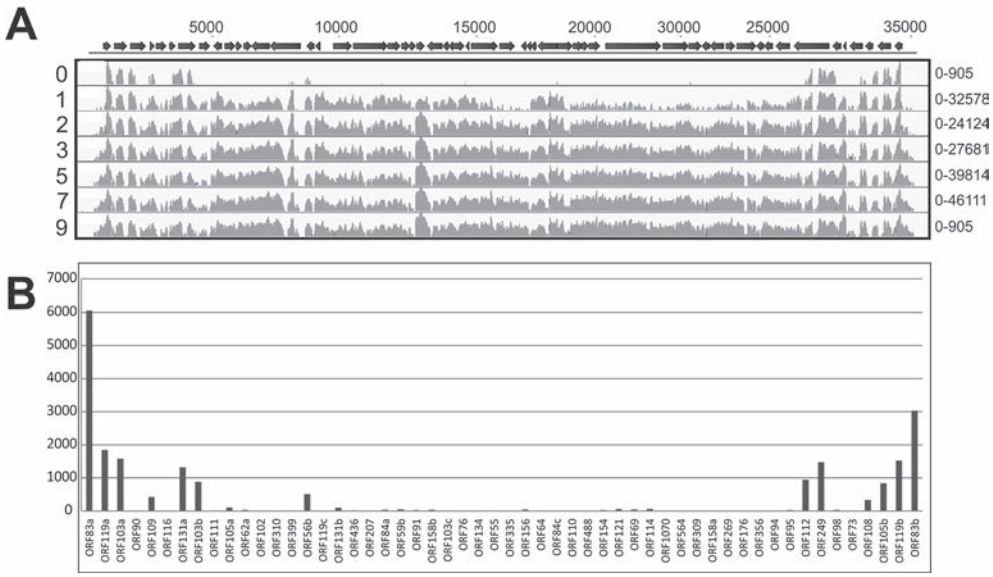


Figure 4 Transcription profiles of SIRV2 genes during viral infection. (**A**) RNA was isolated from *S. islandicus* cells at several hours post infection, indicated by the numbers on the left associated with each transcription map. On the top a schematic representation of the SIRV2 genome is shown, of which the numbers above indicate the base pair position. Arrows represent ORFs. The Reads per kilobase of transcript (RPKM) mapping to the SIRV2 genome are depicted in light grey on a log scale. Minimum and maximal values of the y-axis of each transcription map are shown on the right. (**B**) Detail of SIRV2 gene expression directly after infection (t=0). RPKM are depicted on the y-axis for all 54 annotated genes shown on the x-axis in the order which they reside on the SIRV2genome. For detailed information on expression of individual viral genes see Table S2.

the VAP and tail fibre proteins respectively (141). ORF103a and ORF108 code for proteins with unknown functions.

At late stages of infection, expression of genes coding for structural proteins increased, such that at the end of infection cycle ~35% of viral reads mapped to the gene encoding the major coat protein, ORF134 (NP_666560.1) and ~13 % to the VAP gene, ORF98, encoding the component of the pyramidal egress structures. Other abundantly expressed genes at this stage were ORF83a and b and ORF110 (NP_666566.1), a gene of unknown function.

Yeast two hybrid screen of viral proteins

An array-based yeast two hybrid screen was performed, to provide additional information about possible functions of viral gene products. All 54 SIRV2 genes and six truncated genes lacking trans-membrane domains were cloned to both bait and prey vectors and transferred to yeast for an ORFeome array-based screen. Yeast two hybrid screens were performed for each gene both in prey and in bait vector to limit the number of false positive interactions and quantified by a standard α-galactosidase assay. Several protein interactions between SIRV2 proteins were

identified (Table 1). The majority of them reflected intramolecular associations of SIRV2 proteins. In addition, some interactions between different SIRV2 proteins were identified. The proteins of unknown function from ORF95 (NP_666580.1) and ORF73 (NP_666584.1) were found to interact. Binding was observed also between two different predicted glycosyltransferases, encoded by ORF356 (NP_666578.1) and ORF335 (NP_666562.1). Interestingly, interactions were detected between the highly expressed predicted DNA binding proteins from ORF83a\b and the Holliday junction resolvase encoded by ORF121 (NP_666569.1).

Table 1. Yeast two-hybrid interaction analysis of SIRV2.

Confirmed heterotypic interactions				
ORF83a	DNA binding protein	↔	ORF121	Holliday junction resolvase
ORF335	GT1 glycosyltransferase	↔	ORF356	glycosyltransferase
ORF121	Holliday junction resolvase	↔	ORF83b	DNA binding protein
ORF95	Unknown protein	↔	ORF73	Unknown protein
Confirmed homotypic interactions				
ORF103a	Unknown protein			
ORF90	With transmembrane domain		(ORF90 without TMD shows no interaction)	
ORF131a	Unknown protein			
ORF84a	Unknown protein			
ORF91	Unknown protein			
ORF154	GCN5 acetyltransferase			
ORF69	Unknown protein			
ORF108	Unknown protein			

Host gene response to viral infection

After infection with SIRV2, extensive changes in *S. islandicus* gene expression were observed. About 30-50% of all host genes were differentially expressed in infected samples compared to uninfected control cultures, and these differences were very pronounced (Table S3). The numbers of up and down regulated genes were approximately the same. However, the degree of increase in expression was much higher amongst the up regulated genes compared to the degree of decrease in expression of down regulated genes (Table S3). In datasets from all time points, a relatively high fraction (~50-80%) of identical genes were marked as significantly differentially expressed (Fig 5). Venn diagrams, representing the extent of similarity between samples, show slightly more uniformity in the down regulated compared to the up regulated genes from distinct time points (Fig 5).

An analysis was performed with GOseq on differentially regulated genes (142), to detect over- represented functional categories of genes. One functional category was significantly over represented amongst genes of which expression increased after infection (from one to five

h.p.i) . Genes involved in 'defense mechanisms' belong to this category (Table S3). Among the down regulated genes (from one to three h.p.i.) a single functional group was significantly over represented. Its members are genes involved in 'cell cycle control, cell division and chromosome partitioning' (Table S3). Both these functional categories are discussed in more detail below.

Proteins involved in cell division

Genes belonging to the functional category 'cell cycle control, cell division and chromosome partitioning' were down regulated after infection. This category has a fairly small number of members, with only 14 genes present on the *S.islandicus* LAL14/1 genome. Half of these genes are annotated as ATPases involved in chromosome partitioning. However, four out of these seven genes encoding ATPases were up regulated after infection and evidently did not contribute to the over representation of this category amongst the down regulated genes (Table S3). The majority of the other genes from this category share sequence similarity with eukaryotic genes that code members of the ESCRT-III sorting complex. In *S.islandicus* these genes are part of the *cdv* operon that comprises *cdvA*, *cdvB* and *cdvC* and codes for the *Sulfolobus* cell division machinery, responsible for the constriction of dividing cells (143, 144). Interestingly, the complete *cdv* operon was down regulated around tenfold as a result of infection (Table S3). Moreover, the three other *cdvB* paralogs present in different regions on the *S.islandicus* LAL14/1 genome, were down regulated three to tenfold as a consequence of infection (Table S3).

CRISPR-Cas

Following SIRV2 infection, expression augmented of genes belonging to the functional category 'defense mechanisms' (Table S3). Many members of this category, encode CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas systems; operons I-A SiL_0385-0392, I-A SiL_0393-0397, I-D SiL_0606-0609, I-D SiL_0610-0613, III-Bα SiL_0786-0793 and III-Bβ SiL_0600-0605 (Table S3) (145). The CRISPR-associated (*cas*) genes play essential roles in the CRISPR mediated prokaryotic adaptive immune system that can protect cells against invasion of mobile genetic elements, i.e. plasmids and viruses (32-34). The *S. islandicus* LAL14/1 genome contains five complete and one incomplete CRISPR-*cas* arrays. The complete CRISPR-*cas* arrays consist of a CRISPR array and adjacently located *cas* genes. Unique spacer sequences matching specifically to foreign genetic elements, are located in-between the CRISPR repeat sequences (34). Six *cas* operons are encoded: two I-A subtypes, two III-B subtypes (III-Bα SiL_0786-0793 is incomplete, lacking the CRISPR array) and two I-D subtypes consisting of two adjacent gene clusters coded in opposite direction (145, 146). In uninfected control cells reads mapping to all individual *cas* genes were detected, although the expression levels of *cas* operons differed considerably. One I-A (SiL_0385-0392) and the III-Bα SiL_0786-0793 operons are highly

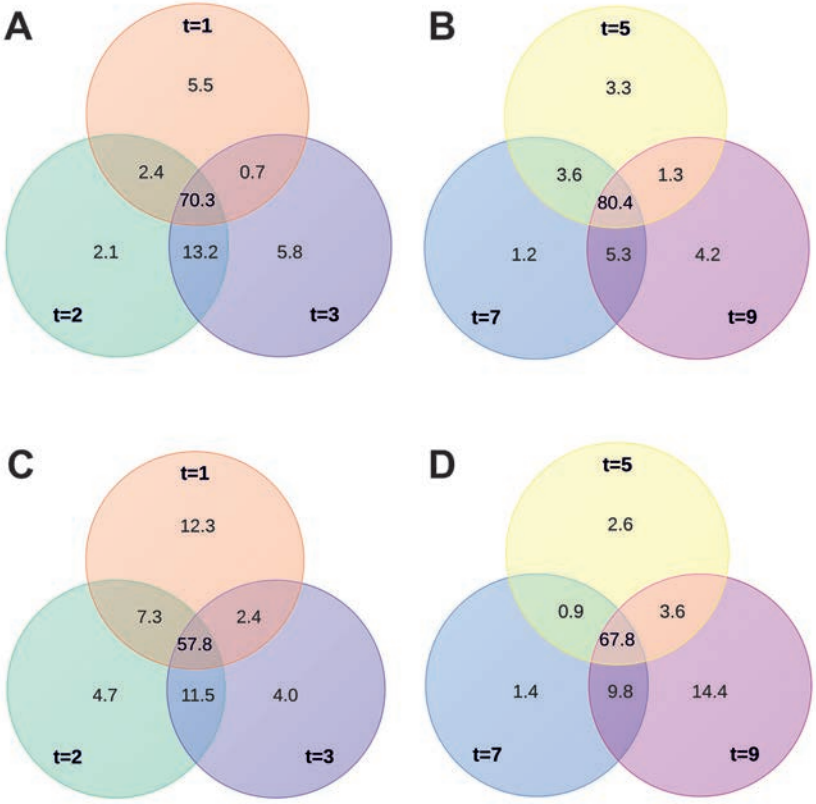


Figure 5 Correspondence between sets of differentially regulated genes harvested at various time points, comparing SIRV2 infected and uninfected control cells. Venn diagrams depict the percentage of identical and unique genes amongst the differentially regulated groups taken at 1 to 9 h.p.i. (t=1 to t=9). (**A**, **B**) down and (**C**, **D**) up regulated genes.

Figure 6 CRISPR-*cas* expression increases after SIRV2 infection. Schematic representation of the six *cas* operons and five associated CRISPR arrays present on the *S.islandicus* LAL14/1 genome. Operons of type I-A (red), I-D (green) and III-B (blue) are present. Numbers within arrows indicate the gene names of all genes of which each operon consists. Number above the arrows represent the number of reads mapping to genes of the represented operon in control (left) and SIRV2 infected (right) cells. Colours of these numbers indicate expression increase (green) or decrease (red) upon infection.

expressed, the other type III-B operon (III-Bβ SiL_0600-0605) and one I-D operon (I-D SiL_0606-0609) are moderately expressed (Fig 5, Table S5). Under these conditions, expression of the other I-A operon (I-A SiL_0393-0397) and part of the I-D operon (I-D SiL_0610-0613) could hardly be detected (Fig 5, Table S7). The CRISPR arrays located adjacent to each *cas* operon were expressed to roughly similar degree as the *cas* genes that they are clustered with (Fig 6, Table S4). All CRISPR arrays contain several spacers matching to different viruses, mostly *Rudiviridae*. Only the CRISPR array 2, adjacent to operon I-A SiL_0393-0397, contains in addition to viral matching spacers, many which target plasmids (145). This CRISPR array is hardly expressed (Fig 6, Table S4).

Interestingly at one h.p.i., expression of the majority of *cas* operons strongly increased. The I-A SiL_0385-0392 and I-D SiL_0606-0609 operons were up regulated around tenfold, the I-A SiL_0393-0397 and I-D SiL_0610-0613 operons approximately fivefold, and the III-Bβ SiL_0600-0605 operon twofold. Expression of the I-D SiL_0606-0609 operon increased during infection, while expression of the other *cas* operons reached the highest level already one h.p.i. and remained stable in the course of the infection cycle (Table S4). Expression of *cas* genes as a result of SIRV2 infection was so pronounced that 3.7% of all mapped reads originated from *cas* operons. The incomplete III-B α SiL_0786-0793 operon, without CRISPR array adjacent to it, represents the only *cas* operon which is slightly down regulated after infection (~twofold) (Fig 6, Table S4). Thus, expression of all CRISPR arrays increased after SIRV2 infection and typically expression kept increasing during the viral infection cycle, in contrast to the adjacent *cas* genes which usually reached highest expression levels quickly after the onset of infection (Table S4).

Toxin-antitoxin

Together with the CRISPR-cas systems, toxin-antitoxin (TA) systems belong to the functional category 'defense mechanisms'. Many TA genes were abundantly up regulated in SIRV2 infected cells (Table S3). Prokaryotic TA systems are widespread mobile two-gene elements, which are subdivided in three families based on the nature and mode of action of the antitoxin (147). The family II TA systems consist of an antitoxin protein that counteracts the negative effect of the toxin protein, which is usually a nuclease that is more stable than the antitoxin (148). Family II TA operons are widespread in archaea and especially amongst members of the order *Sulfolobales* (147). The genome of *S.islandicus* LAL14/1 carries 16 TA operons of the family II type VapBC (virulence-associated protein) (145). In addition 6 TA operons are present of another recently described family: HEPN-NT (Higher Eukaryotes and Prokaryotes Nucleotide-binding - nucleotidyltransferases) (145, 147). Of this HEPN-NT family, fifteen separate antitoxin coding genes are found on the *S.islandicus* LAL14/1 genome and they do not appear to be associated with toxins. In uninfected *S.islandicus* LAL14/1 cells the majority of TA loci are moderately transcribed, although expression levels between loci can differ considerably (Table S5). In SIRV2 infected samples of different time points (except t=0 h.p.i.) expression increased of 11 out of 16

VapBC and 3 out of 6 HEP-NT loci. In most cases both the genes coding for the antitoxin and the toxin, were up regulated and expressed to similar extent. Gene expression of only two antitoxins decreased after SIRV2 infection.

Since gene expression of TA gene clusters previously was reported to increase after heat shock together with stress response genes (149), we checked for changes in expression of known stress response genes (i.e. HSP, USP, proteasome). After SIRV2 infection, expression of these genes decreased (~threefold), or the genes appeared non responsive and expression remained constant (Table S3).

Insertion Sequence elements

Just one functional category was marked as over represented amongst up regulated genes after viral infection. It is possible that specific sub groups were not scored as over represented in the GOseq analysis, because they are part of a large functional category to which many genes belong. Therefore, over representation of sub groups was checked manually. Amongst the genes of which expression increased after infection, a high proportion of insertion sequence (IS) elements was detected (12-15% of all up regulated genes). These belong to the large functional category 'replication, recombination and repair' (Table S3).

The *S.islandicus* LAL14/1 genome contains a high number of IS elements, similar to the genomes of *S.islandicus* strains HVE10/4 and REY15A, and of *Sulfolobus solfataricus*. The latter genome contains approximately 200 IS elements, of which active transposition has been observed (150). In the *S.islandicus* LAL14/1 genome 53 predicted IS elements are present of which only seven encode intact transposases. However, since IS elements with mutated transposases can be mobilized by transposases of the same family acting *in trans*, a total of 31 IS elements in the *S. islandicus* LAL14/1 genome could potentially be mobile (145). In uninfected cells the majority of IS elements were transcribed. As a result of infection, expression of many IS elements from different families increased, such that they formed the largest group of up regulated genes (Table S3). Most actively expressed were members of the IS1048 group (IS630 family) (Table S3).

Induction of transcription of IS elements after virus infection is peculiar and has not been reported previously. However, pronounced increase of expression of IS elements of *S. solfataricus* has been observed as a result of a heat shock and UV irradiation (151) Therefore, high expression of IS elements was suggested to be associated with the general stress response of *S. solfataricus* (149).

Discussion

Using in-depth transcriptome analysis, we have monitored the dramatic changes in gene expression occurring during infection of the archaeon *S.islandicus* with the virus SIRV2. The amount of viral transcripts increases in time, until it constitutes approximately 20 % of all mRNA in the infected cells. Transcription of these genes has significant impact on host cell morphology and metabolism. The host responds to this threat by a vast change in gene expression (~50% of genes). Most of the host genes that are strongly activated upon infection are implied to function in defense against viruses, thus unveiling a regulatory mechanism that aims at counteracting the viral attack.

Viral gene expression

SIRV2 gene transcription starts from the two distal ends of the SIRV2 linear dsDNA genome. For some bacterial viruses a slightly similar situation is reported and early transcription was observed from genes located on the termini of linear dsDNA genomes of the viruses phi29 and PRD1, from the families *Podoviridae* and *Tectiviridae* respectively (152-154). In the case of SIRV2 this pattern of gene expression might be a consequence of the mode of viral infection that remains currently obscure. Since the two ends of the SIRV2 virion are identical, the binding of the virion and entry of the viral genome could potentially occur from both sides. In this case, it would be advantageous when ORF83a and b, obviously important at early stages of infection, are located at either end of the genome and are readily available for transcription. Alternatively, the location of genes with identical complementary sequences on both ends of the genome might be advantageous for SIRV2 genome replication, which includes formation of head-to-head and tail-to-tail replicative intermediates (60). The exact nucleotide identity between SIRV2_ORF83a and b and between their homologues in SIRV1 suggests the presence of selection pressure to maintain this trait.

The genes highly transcribed later during SIRV2 infection are distributed evenly across the viral genome. Little temporal regulation is observed and expression of the majority of genes starts immediately after infection and subsequently steadily increases. This is in contrast to the viral gene expression of the temperate archaeal virus SSV, of which the expression pattern during the first hours after UV induction is dominated by only one early transcript (132). The qualitative results of previously performed Northern blot hybridization and primer extension analysis of genome transcription of the rudiviruses SIRV1 and SIRV2 correspond with our finding that almost all viral genes are transcribed after infection (134).

Implications for comprehension of viral gene functions

In general, the expression of SIRV2 genes corresponds very well with predicted or confirmed gene functions. Transcripts of structural genes (encoding proteins of the viral capsid or egress structure), like ORF98, ORF134 and ORF1070 are most abundant late in the infection cycle, while expression of the transcriptional regulator SvtR, which represses ORF98 and ORF1070, peaks soon after infection (141). Indeed expression of ORF98 and ORF1070 is inversely proportional to that of SvtR .

Yeast two-hybrid analysis uncovered several homotypic interactions of SIRV2 proteins, all of them without predicted functions. Two different glycosyl transferases were found to interact and correspondingly, glycosyl transferases are known to be capable of protein complex formation (155, 156).

ORF119c encodes a Rep protein that was implied to initiate viral genome replication and aid in resolution of replicative intermediates (140). The viral encoded Holliday junction-resolving enzyme was suggested to be required for recovery from situations in which linear SIRV2 genome replicative intermediates are generated by Rep cleavage failure (140). Surprisingly, expression of ORF119c peaks at the end of the infection cycle, is very poor and is approximately 3000-fold lower than that of the Holliday junction resolvase ORF121 (NP_666569.1). This suggests that the Rep protein is not required for the analyzed mode of viral replication and might be implemented in currently unknown aspects of the viral cycle.

In addition to confirming already proposed functions of genes, the here presented expression data also suggest an important role for some uncharacterized viral genes. Most of early expressed genes located on the genome termini have unknown functions. SIRV2_ORF103a and SIRV2_ ORF119b have a homologue in *Acidianus* Filamentous Virus 1 (AFV1), AFV1_ORF99, of which the crystal structure was determined (157). AFV1_ORF99 displays a completely novel fold from which no function prediction could be derived (157). Interestingly, ORF83a and ORF83b are the first transcribed genes after infection. These genes have identical nucleotide sequences and are located in opposite direction on each outer end of the genome. Although the expression decreases after one h.p.i., the number of transcripts mapping to ORF83a and ORF83b remains very high. Since ORF83 encodes a protein with predicted DNA binding capacity and is the first gene to be transcribed, it is tempting to speculate that it might have an essential function in viral genome replication. In the closely related SIRV1, the ORF83 homologues SIRV1_ORF56a and b are also located on the extremities of the linear genome (60). The crystal structure has been determined for the ORF83 homolog in SIRV1 (ORF56a), which is lacking 27 amino acids on the N-terminus compared to ORF83 (158). The protein has a hexameric configuration and it contains an unusually conserved C-terminal cysteine, which might be involved in subunit-subunit cross linking. The C-terminal half of the protein displays a classical HTH domain, which is often found amongst DNA binding proteins and specifically transcriptional regulators (158). This is

reconcilable with a function in genome replication, although the DNA binding activity of ORF83 has not been experimentally characterized. For initiation of SIRV2 replication nicking activity would be required as well. The potential involvement of ORF83 in SIRV2 genome replication is enforced by the detected interaction between ORF83 and ORF121 by yeast two-hybrid screen. ORF121 is coding for the Holliday junction resolvase that is implied in the resolution of replicative intermediates (159). Thus, there seems to be a link between ORF83 and the late steps of genome replication.

Host response to viral infection

The consequence of SIRV2 infection could be detected by the adaptation of the *S. islandicus* gene expression profile. Specifically, there was intensive decrease of expression from genes involved in cell division, notably ESCRT-III homologs and components of the *cdv* operon. *CdvB* and *cdvC* have homologues in the ESCRT-III complex involved in budding of luminal vesicles in eukaryotes (160). The decrease in expression of these genes is likely to result from the DNA degradation caused by SIRV2 infection, since the cdv proteins are shown to be under control of checkpoint systems, which inhibit cell division in response to DNA damage (143). Interestingly, it was reported that after infection with the archaeal virus STIV, *cdv* genes were significantly up regulated in the host *Sulfolobus solfataricus* (51), implying an important function in the STIV infection cycle.

Activation of CRISPR-Cas systems

Recent discoveries of novel prokaryotic anti-viral defense mechanisms have raised awareness that 'simple' prokaryotes, in analogy to eukaryotes, have a range of sophisticated immune systems at their disposal. The CRISPR-Cas system is of this the most recently described example (34, 161). CRISPR-Cas systems are very abundant amongst archaea and often multiple different systems are present on the same archaeal genome. The repeat spacer arrays are relatively long and can make up 1% of archaeal genomes (132, 162), which indicates the importance of these systems for the fitness of archaeal cells. Hyperthermophilic archaeal CRISPR-Cas systems have been studied extensively using *Pyrococcus furiosus* and *S. solfataricus* as models (163, 164). Both organisms encode various Cas complexes, a subset of which has been analyzed *in vitro* (163-165). Transcription of CRISPR arrays was detected in *S. solfataricus* and *P. furiosus* (166, 167). However, behaviour of archaeal CRISPR- Cas systems during viral infection has hardly been studied until now. Nevertheless, during a recent proteomics analysis a number of Cas proteins were detected after STIV infection, suggesting that there might be activation of CRISPR-Cas as a consequence viral attack (133). The presence of so many different CRISPR-Cas systems in archaeal cells has raised questions about the diverse roles they might have during viral infection.

We observed that SIRV2 infection resulted in massive activation of the CRISPR-Cas systems present in *S. islandicus*. The *S.islandicus* LAL14/1 genome contains six cas operons, coding for complexes of three different types (I-A, I-D and III-B). Expression of five out of six *cas* operons increased as a result of viral infection. Interestingly, expression of almost all *cas* operons augmented three to tenfold, just like the associated CRISPR arrays, making them by far the most pronounced up regulated genes after infection. All systems are activated directly after viral infection. Two operons, belonging to the I-A and I-D subtype respectively, were considerably transcribed in uninfected cells and after viral infection their activation was most pronounced of all Cas-operons. After viral infection, expression decreased of the type III-B Cmrα complex, without adjacent CRISPR array (168). The type III-B Cmrβ complex was moderately expressed in non-infected cells and increased slightly after infection. Thus, the type III-B operons appear to react to a lesser extent to SIRV2 infection. These results indicate that type I-A and I-D operons, which code DNA targeting Cas complexes, play a more important role during SIRV2 infection than the type III-B encoded complexes.

CRISPR arrays of *S.islandicus* LAL14/1 do not possess any spacers perfectly matching to the SIRV2 genome, which is probably the reason why cells do not recover from a SIRV2 infection despite the widespread activation of CRISPR-Cas systems.

The described findings demonstrate that the different CRISPR-Cas systems present in archaea probably have specialized roles which could result in a 'tailor-made' defense reaction for each different type of foreign genetic element. Some Cas operons (I-A SiL_0385-0392 and III-Bα SiL_0786-0793) are actively transcribed in uninfected cells indicating that the encoded Cas proteins might be continuously present in cells to act as 'watchers at the gate' which can target foreign genetic elements directly when they access the cell. In contrast, other Cas-systems are hardly expressed without activation signal arising from viral infection, which is the case for the I-A SiL_0393-0397 operon and adjacent CRISPR array, containing a high proportion of plasmid targeting spacers.

Expression of Toxin Antitoxin systems

As a result of SIRV2 infection toxin-antitoxin (TA) clusters of the host genome are activated. Such clusters are ubiquitously present on prokaryotic genomes (147, 169) and they are proposed to function in programmed cell death and in stress response (149, 170). In addition, it was demonstrated that bacterial TA systems play a role in abortive infection, during which cells commit altruistic suicide after viral infection to protect the clonal population (171, 172). TA suicide/dormancy systems were proposed to be linked with diverse immunity systems in prokaryotes to provide robustness to the antivirus response (173). Indeed, in *S. islandicus* LAL14/1, as in many prokaryotes, a high proportion of TA loci are found in the close proximity of CRISPR-*cas* operons, suggesting their involvement in defense against viruses (173). However, the involvement of

archaeal TA systems in anti-viral defense has not yet been demonstrated, although numerous TA-encoding genes are present in archaeal genomes. Interestingly, the most actively transcribed *S. solfataricus* gene as a consequence of STIV infection has an unknown function (51), but homology searches suggest it might be an antitoxin part of a TA system. SIRV2 infection results in a severe increase of expression of many TA loci, of which the majority is VapBC like (Family II). The observation that SIRV2 infection did not cause increased expression of any of the regular stress response genes suggests that high TA expression in this case is not linked with basal stress response and might be a specific reaction to viral infection.

Here we have presented a global map defining host and viral gene expression during the infection cycle of the virus SIRV2 in its host, the hyperthermophilic archaeon *S. islandicus* LAL14/1. This pioneering study provides in depth information for all viral and host genes during a time series after prokaryotic viral infection. This work has demonstrated once more the power of the recently developed RNAseq approach. These findings, in combination with performed yeast two hybrid analysis of the viral ORFeome, have corroborated predictions for functions of some SIRV2 genes, while suggesting novel functions for others. Moreover, it enables the detailed study of a host response to a viral infection, showing massive activation of host anti-viral defense genes, most importantly the CRISPR-Cas and TA systems, in the presented case. This information on individual gene expression and activation of all the anti-viral defense systems is expected to aid future studies aiming to establish the function and interplay of the different systems *in vivo*.

Experimental procedures

Growth and infection of S. islandicus

Sulfolobus islandicus strain LAL14/1 was grown and infected by the virus SIRV2 as described previously (83).

Transmission electron microscopy

SIRV2 infected *S.islandicus* cells were prepared for electron microscopy at distinct time points after addition of the virus (i.e. 0, 5, 9 and 12 h.p.i.). Cells were fixed with 2.5% (wt/vol) glutaraldehyde in 0.1 M Hepes buffer (pH 6.5). Postfixation, dehydration, embedment in epoxy resin, sectioning, and transmission electron microscopy imaging were performed as described previously (83).

Yeast two hybrid

The yeast two-hybrid analysis was essentially performed as described by Rajagopala and Uetz (174). Briefly, open reading frames of SIRV2 were amplified (Table S1) and cloned to pENTR

vectors for Gateway™ transfer to the bait (pGBT9g) and prey (pGAD424g) vectors. The bait vectors were transformed to *Saccharomyces cerevisiae* AH109 Mata and the prey vectors to *S. cerevisiae* Y187 Matα. An auto activation assay was performed, which showed that addition of 3 mM 3-Amino-1,2,4-triazole (3-AT) was most optimal for the reduction of background growth of the strains. Under these conditions none of the bait constructs resulted in autonomous activation of the reporter constructs. We used an eight-clone pooled, array-based mating screening. The reciprocal screen (exchanging bait & prey) was performed and verifications of positive interactions were done by sequencing and recloning of the initial constructs to *S. cerevisiae* strain AH109 and independent Y2H verification.

RNA isolation and library preparation

Four identical cultures of *S. islandicus* LAL14/1, were inoculated in 500 ml medium with 5 ml of preculture. After ~12 hours of growth, when the optical density (OD) was in the range of 0.1-0.2, two of the cultures were infected with SIRV2, while the other two served as uninfected controls. Total RNA was isolated from these cultures at different time points after addition of the virus (0, 1, 2, 3, 5, 7, 9 h.p.i.), with the *mir*Vana™ Isolation Kit (Ambion) using the manufacturers protocol for total RNA isolation. The RNA quality was checked with the 2100 Bioanalyzer (Agilent). 5 µg of total RNA of each sample was directly used for RNAseq library preparation. Directional libraries were prepared using the TruSeq SmallRNA sample prep kit, set A and B (Illumina), according to manufactures instructions. The total RNA was chemically fragmented with Ambion reagent (AM8740), followed by purification on RNeasy columns (Qiagen, #74204). After treatment with phosphatase and polynucleotide kinase, RNA was purified on RNeasy columns (Qiagen, #74204). The fragmented RNA was then ligated with 3'- and 5'- TruSeq adapters, as described in the manufacturers protocol. Synthesis of cDNA was performed by reverse transcription. The cDNA products were specifically amplified by 11 cycles of PCR and products were purified on Agencourt AMPure XP beads (Beckman Coulter Genomics, # A63881). The resulting libraries were checked on a Bioanalyzer DNA1000 chip (Agilent). Libraries were sequenced to generate single-end 50 bases reads using the Illumina Hiseq 2000 in a multiplexed 51 +7 bases single read using a TruSeq SR cluster kit v3 cBot HS (Illumina, # GD-401-3002) and a TruSeq SBS kit v3 HS 50 cycles (Illumina, # FC-401-3002).

Read mapping

Reads were cleaned from adapter sequences and from sequences of low quality using an in-house program. Only sequences with a minimum length of 30 nucleotides were considered for further analysis. Bowtie ((175), version 0.12.7,) was used to align the reads to the reference genomes: *S.islandicus* LAL14/1 and SIRV2 (60, 145). Gene annotations for *SIRV2 and S. islandicus LAL14/1* were downloaded from GenBank (accession numbers: NC_004086.1 and CP003928.1

respectively). Reads mapping to rRNA were discarded. For each gene, reproducible sites that overlapped its protein-coding region, as well as those residing in the intergenic region upstream of its beginning, were associated with the gene. Viral genes of which the maximal expression level did not exceed 250 Reads Per Kilobase of transcript per Million mapped reads (RPKM) were considered lowly expressed genes.

Statistical and GO analysis

DEseq (138) and EdgeR (139) were used to determine significantly up and down regulated genes comparing all samples of each time point for each condition (infected or control), and between all samples of each condition for each time point. Genes marked as differentially regulated by both methods ($P<0.05$) were kept for further analysis. Venn diagrams were constructed using BioVenn (176). For each list of up or down regulated genes between samples a Gene Ontology (GO) analysis was performed using GOseq (142) on the 10% of genes with the lowest P-values. Genes were sorted to functional groups according to the COG data available on the NCBI site: http://www.ncbi.nlm.nih.gov/COG/grace/fiew.cgi. ArCOGs were linked with *S.islandicus* LAL14/1 genes based on homology of these with other *S.islandicus* strains available in the arCOG database (177).

RNAseq data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-1660.

Acknowledgements

We are thankful to Pieter-Jan Ceyssens and Kira Makarova for discussion. This work was supported by the l'Agence National de la Recherche, France, in the framework of Programme Blanc and by the Netherlands Organization for Scientific Research (NWO) via an ALW TOP grant (854.10.003). T.E.F.Q. was supported by a fellowship from the Ministère de l'Enseignement supérieur et de la Recherche of France. M.V. & R.L. are supported by a grant of the FWO Vlaanderen (G.0323.09).

Supplemental Tables

Table S1. Primers for amplification of SIRV2 ORFs for yeast two-hybrid screen

SIRV2ORF83aF	CACCATGAAAGTCGAGAGATACAAATG
SIRV2ORF119aF	CACCATGCACATATGTAAAAGTGGTGAG
SIRV2ORF103aF	CACCATGAAAAAGATGAAATTCGAAACGT
SIRV2ORF90F	CACCATGGATGAAGATCTTTTAGTTGAAG
SIRV2ORF109F	CACCATGAATGTAGAAAATCAAGTAG
SIRV2ORF116F	CACCATGAGAAATATGGGCTATCAATATA
SIRV2ORF131aF	CACCATGAGTCAAAAAGTCGAATTTCC
SIRV2ORF103bF	CACCATGAGTTTTTCGTATATATCAATTAGTG
SIRV2ORF111F	CACCATGAGGAGAAAATTAGGAAAAAGAC
SIRV2ORF105aF	CACCATGGAATTTGAAGATTTAGATGTAG
SIRV2ORF62aF	CACCATGAAGATTCTAGTAGATAATG
SIRV2ORF102F	CACCATGATTTTCAATTCGCCAATTGTTC
SIRV2ORF310F	CACCATGAAATTAGTATTTGAAATAACATC
SIRV2ORF399F	CACCATGGAATTCGAAAGAAAAAGTTCTG
SIRV2ORF56bF	CACCATGCAAACTCAAGAACAATCTAAAC
SIRV2ORF119cF	CACCATGGATTTGAAAAAAGTTTTAAATTTTC
SIRV2ORF119cF-TRUNC	CACCATGGATACTGAAAAAATTTTTAAAAAATTTCGAC
SIRV2ORF131bF	CACCATGGCCTCATTAAAACAAATAATAG
SIRV2ORF436F	CACCATGAGTGAAAACACACAAC
SIRV2ORF207F	CACCATGGTAAATATGAATTATGAAG
SIRV2ORF84aF	CACCATGAGGAATATGAGTCAAATAGAAG
SIRV2ORF59bF	CACCATGATGAAGATTATTACATTTAAAATTCC
SIRV2ORF91F	CACCATGACTGATTATAAAAACGGAATAAAG
SIRV2ORF158bF	CACCATGATTTTATCAGATAGAG
SIRV2ORF103cF	CACCATGGAAATTGATCTAAAGAATGAATGTAG
SIRV2ORF76F	CACCATGCATATCTTTGTGAATAACTTCTTG
SIRV2ORF134F	CACCATGGCAAAAGGTCACACATCAAG
SIRV2ORF55F	CACCATGGCACTATTAGGATATGAATGTC
SIRV2ORF335F	CACCATGAAAACTGCAATTTTAACTATG
SIRV2ORF156F	CACCATGGTAGCTAAAGGATTTTATATGTG
SIRV2ORF64F	CACCATGATATCTTATTATTATGATGAAAAAG
SIRV2ORF84cF	CACCATGAATTATCTGAGGAGGAAAGTG
SIRV2ORF84cF-TRUNC	CACCATGAAAATTAAGCAAGCATTAAAAG
SIRV2ORF110F	CACCATGAAAATTGAAGATCCTTTTC
SIRV2ORF488F	CACCATGACACTGTATGACATTTATATAC
SIRV2ORF154F	CACCATGAATCCAAAATATGAAATTGAAG
SIRV2ORF121F	CACCATGAACATTAGACAATCCG
SIRV2ORF69F	CACCATGTCAGAGCAAAAGAGAAG
SIRV2ORF114F	CACCATGAATAAAGTCTATTTGGCAAATG
SIRV2ORF1070F	CACCATGTTTATAATTTTGAAAAGAAAC
SIRV2ORF564F	CACCATGATATATTATATCATGCCTTTAC
SIRV2ORF309F	CACCATGTCTTCAACTTGTAATCCAATTAC
SIRV2ORF158aF	CACCATGAATTATGATGATTATTTTTG
SIRV2ORF269F	CACCATGAGTGTAACTTATACTTCAATTTC
SIRV2ORF269F-TRUNC	CACCATGAATACTTTAAATAAGTTAACAC
SIRV2ORF176F	CACCATGAAAATTTTCACTTTCGTAGG
SIRV2ORF356F	CACCATGCAAAAAACTATTTTCTATGTC
SIRV2ORF94F	CACCATGGTGAAAAAAATGGAGTTTGAAG
SIRV2ORF95F	CACCATGAATTTGAAAAAAGTTAAACGAATTATAG
SIRV2ORF112F	CACCATGAGTTTTTATATAATGTATATTAAGGTG
SIRV2ORF249F	CACCATGGAGGTAAAACAGATAAAGAAG
SIRV2ORF98F	CACCATGGCTATAACATTATTAGAAGGAG
SIRV2ORF98F-TRUNC	CACCATGGGAGAAATTGTCCACTTATATAATG
SIRV2ORF73F	CACCATGTTTTTTGAAGATTCTAATATACAAG
SIRV2ORF108F	CACCATGGGGAAAAAAATGGATGATATAC
SIRV2ORF105bF	CACCATGATAGATGAAAAGGCACAAG
SIRV2ORF119bF	CACCATGCACATATGTAAAAGTGAGG
SIRV2ORF83bF	CACCATGAAAGTCGAGAGATACAAATG
SIRV2ORF83aR	CTAACAACTCTCCAAATATCTCCTA
SIRV2ORF119aR	TTAATGATACTTAATCATTTCAAC
SIRV2ORF103aR	TTACTTTTTGACTTTTCTCCAGAAATG
SIRV2ORF90R	TCATCTCAAATTTTTCACTTTATACAG
SIRV2ORF90R-TRUNC	TCAAGAATTATTATCGACAAATTTTTTATC
SIRV2ORF109R	TTATTGACATTTAAATATAAGTCTTG
SIRV2ORF116R	TTAGGAAATCTTATTAATAGATTTTTC
SIRV2ORF131aR	TTAATATCTAGAAATCTCTGG
SIRV2ORF103bR	TTAGTTTTTTATTAATTTGATCTTATG
SIRV2ORF111R	TTATTTATTATTATTGTTTATCAAATTCTG

SIRV2ORF105aR	TTAAATATTGTTTTCTTCATCCTC
SIRV2ORF62aR	TTAAAGAACTTTAATCCTCTTTATTCC
SIRV2ORF102R	TTAGAAATATAAATTTGTTATCC
SIRV2ORF310R	TTATATTTCTAATTCTATCTC
SIRV2ORF399R	CTATTTTGGCATGACATTTATTTTAC
SIRV2ORF56bR	TTAACCGCCTCGTTTTTGCAAATATTC
SIRV2ORF119cR	TTATGACATTTTTATATCTTTTTTTG
SIRV2ORF131bR	TTAAAACTCCTCCTCAACTG
SIRV2ORF436R	TTACCATCCTCCTAAATTGCTAAATC
SIRV2ORF207R	TTAAAAAAGTGATATAATGCATTTTTG
SIRV2ORF84aR	TCATTTTAAATCACCATTCCCGAAAAC
SIRV2ORF59bR	TCAGAACTTTATTTTCTCAACTTTTG
SIRV2ORF91R	TCAGACATTTAAATCACCTAAC
SIRV2ORF158bR	TTAATCTTCTTTTGCTAAAGTTAC
SIRV2ORF103cR	TCAGTTCTTATTTTCAGTTTCTGAC
SIRV2ORF76R	TTAGTTAAGCAGATAAGAC
SIRV2ORF134R	TTAACTTACATATCCAGTTGGGCTTC
SIRV2ORF55R	TTATTCTTCTTCAGCCTTTATTTC
SIRV2ORF335R	TCATTCTAAGAATCTAGTATATAAATTTC
SIRV2ORF156R	CTAAAGTAAATAAATCGTTACTTTATTTAG
SIRV2ORF64R	TTACCATAATCCCCATTCCTTACTATC
SIRV2ORF84cR	TCATTTTTTATCTAACCTCC
SIRV2ORF110R	TCACTCATCTTCACTTTCCTCCTCAG
SIRV2ORF488R	CTATCTATTTGCAATATTTTGAAAATTC
SIRV2ORF154R	TCACTTATTAAGATATTTTACATAAACTTC
SIRV2ORF121R	TTAGCTGTTAATTCCGTATTTAAATTTG
SIRV2ORF69R	TCATTTTTCCTCACCTTCAACTG
SIRV2ORF114R	TTAGTCTACAATTATCTCAAAAAATTG
SIRV2ORF1070R	TTAGTAAGCACTTCTTAATTGCC
SIRV2ORF564R	TCACGCGACATAAATTGTAGTTAATTG
SIRV2ORF309R	TTATTGACAATTATTATTTATTATTTGTG
SIRV2ORF158aR	CTATAACTTGCAAAATACAAATTCTC
SIRV2ORF269R	TCATGACTGAACTACCTCAAATAAC
SIRV2ORF269R-TRUNC	TCAAGCATTTATTCCAGTTTTTTGTAATTGTG
SIRV2ORF176R	TTATTTCTCCATTGGAGTTTTTAC
SIRV2ORF356R	TTAGTTAGTATATTTTTCTACTAC
SIRV2ORF94R	TTAACTAGATCTCCAAAAATCGATTTC
SIRV2ORF95R	TTATCTCTTTTCACGAAATAAATC
SIRV2ORF112R	TTAACTTTCTAAAACAGATTTAACTTG
SIRV2ORF249R	TTAATCATTTCCTGACCGTCTTGG
SIRV2ORF98R	TTATTTAGCTTGCGTATTTGGATTTTG
SIRV2ORF73R	TCACAATTTTTTAGTACGCAATATTTG
SIRV2ORF108R	TTATCTTGGATATCTAATTAATTG
SIRV2ORF105bR	TTATGAATTTGAAGTAATTGATTTTTCTAG
SIRV2ORF119bR	TTAAATATTGTCTTCTTCATCTTC
SIRV2ORF83bR	CTAACAACTCTCCAAATATCTC

Table S2. SIRV2 gene expression during the infection of *S.islandicus* LAL14/1.
Depicted is the absolute number of reads per kilobase of transcript (RPKM) mapping to each SIRV2 ORF. t=0-t=9, samples harvested 0 to 9 h.p.i.

gp number	GenBank accesion number	Gene	t0	t1	t2	t3	t5	t7	t9
1	NP_666535.1	ORF83a	6062	234957	114130	102725	124763	122335	145288
2	NP_666536.1	ORF119a	1844	18097	18415	11469	10876	10111	15075
3	NP_666537.1	ORF103a	1579	55032	31152	22256	19755	17921	21598
4	NP_666538.1	ORF90	4	507	1158	1205	1289	1441	2145
5	NP_666539.1	ORF109	426	7488	6345	4976	5411	5708	7832
6	NP_666540.1	ORF116	7	1094	1521	1460	1339	1125	1256
7	NP_666541.1	ORF131a	1318	20975	17605	11782	10816	9846	13568
8	NP_666542.1	ORF103b	881	11264	6907	4864	4417	4423	5649
9	NP_666543.1	ORF111	15	1415	1126	644	457	460	708
10	NP_666544.1	ORF105a	105	28146	30401	24217	16604	14208	14841

gp number	GenBank accesion number	Gene	t0	t1	t2	t3	t5	t7	t9
11	NP_666545.1	ORF62a	32	10257	14635	12437	10466	7601	8116
12	NP_666546.1	ORF102	8	1851	4317	3877	3743	3191	3673
13	NP_666547.1	ORF310	3	466	3531	5612	8494	9407	11545
14	NP_666548.1	ORF399	10	2309	20762	28828	39014	36691	37579
15	NP_666549.1	ORF56b	506	119860	176877	133456	94962	65015	74368
16	NP_666550.1	ORF119c	0	31	79	79	110	154	249
17	NP_666551.1	ORF131b	98	27559	34150	28102	18297	12604	14973
18	NP_666552.1	ORF436	19	2846	3650	2924	2843	2616	3124
19	NP_666553.1	ORF207	6	1901	2803	2604	3205	3503	4165
20	NP_666554.1	ORF84a	39	7712	5906	3390	1892	1347	1976
21	NP_666555.1	ORF59b	50	14450	10386	5819	3911	3039	4725
22	NP_666556.1	ORF91	29	6406	5422	2982	2246	1987	3207
23	NP_666557.1	ORF158b	38	7832	8260	6496	5186	4419	5761
24	NP_666558.1	ORF103c	16	2220	2019	1444	1143	1093	1468
25	NP_666559.1	ORF76	0	110	108	152	104	154	188
26	NP_666560.1	ORF134	11	1632	204021	352185	500990	494474	642358
27	NP_666561.1	ORF55	15	6298	9896	8268	8152	9753	14292
28	NP_666562.1	ORF335	8	2326	6049	6090	6034	5322	5277
29	NP_666563.1	ORF156	47	9939	15146	12438	12519	12339	16213
30	NP_666564.1	ORF64	3	215	2754	3867	5249	5859	7190
31	NP_666565.1	ORF84c	4	429	5384	7465	10196	10388	13808
32	NP_666566.1	ORF110	9	1719	30817	40884	57344	63616	82868
33	NP_666567.1	ORF488	1	43	3249	4059	5243	5220	6576
34	NP_666568.1	ORF154	28	6298	7481	5776	3848	2620	3288
35	NP_666569.1	ORF121	60	16213	35186	30426	24630	20333	23433
36	NP_666570.1	ORF69	48	13629	32895	32119	30890	26281	26026
37	NP_666571.1	ORF114	64	12965	29720	26565	30049	27729	30912
38	NP_666572.1	ORF1070	2	480	10110	14126	21404	24211	30762
39	NP_666573.1	ORF564	1	190	3791	5621	8573	9382	12388
40	NP_666574.1	ORF309	1	39	780	1259	1807	1924	2521
41	NP_666575.1	ORF158a	0	269	6945	9208	13972	15128	19898
42	NP_666576.1	ORF269	2	446	12175	14788	22518	26154	31785
43	NP_666577.1	ORF176	6	857	4217	4319	4684	4201	4379
44	NP_666578.1	ORF356	1	545	4185	4970	6543	6419	6980
45	NP_666579.1	ORF94	9	2116	2588	2237	1693	1291	1446
46	NP_666580.1	ORF95	28	6684	7016	5738	3925	2604	2470
47	NP_666581.1	ORF112	944	21650	15916	11745	10063	8770	13575
48	NP_666582.1	ORF249	1477	36410	35947	24487	22987	22802	30834
49	NP_666583.1	ORF98	35	5141	64921	94658	156902	189222	239929
50	NP_666584.1	ORF73	0	291	432	401	448	392	534
51	NP_666585.1	ORF108	332	61453	49118	30234	20815	14979	19408
52	NP_666586.1	ORF105b	840	5937	4530	2846	2818	2835	4261
53	NP_666587.1	ORF119b	1525	19554	17692	10939	10603	11636	17028
54	NP_666588.1	ORF83b	3031	117479	57065	51363	62382	61168	72644

Table S3. See table online:
http://jvi.asm.org.ezproxy.library.wur.nl/content/87/15/8419/suppl/DCSupplemental

Table S4. Expression of CRISPR associated genes in *S.islandicus* LAL14/1 cells challenged with SIRV2 infection.
Of each operon are shown the type to which the encoded CRISPR system belongs, as well as the individual gene and protein names of all genes in this operon. CRISPR numbers represent repeat-spacer arrays, uninfected control samples. I, SIRV2 infected samples. t0-t9, samples harvested 0 to 9 h.p.i.

operon	gene	protein	t0C	t0I	t1C	t1I	t2C	t2I	t3C	t3I	t5C	t5I	t7C	t7I	t9C	t9I
I-A	SiL_0385	Cas6	132	139	147	1174	137	1472	143	1559	108	1764	110	1824	151	2322
	SiL_0386	Casx	633	813	985	5799	961	7399	902	7353	928	7878	888	9011	873	10969
	SiL_0387	HD	394	503	656	3530	657	4566	687	4936	576	4896	532	4976	553	5341
	SiL_0388	Cas3	459	637	761	5086	801	6188	812	6464	770	6798	745	6959	745	7364
	SiL_0389	Cas5	787	1014	1262	10618	1384	13321	1448	13718	1395	14072	1288	13157	1248	13520
	SiL_0390	Csa2	1825	2304	2926	26188	3168	28588	3221	28211	3133	28439	2896	25435	2787	21752
	SiL_0391	Csa5	1640	2365	2561	33417	2847	32729	2740	32186	2662	32562	2653	29023	2583	24348
	SiL_0392	Csa3	102	115	115	333	121	386	115	379	131	342	113	400	105	395
I-A	SiL-0993	Csa3	6	5	8	5	8	11	6	15	2	23	7	20	6	37
	SiL-0994	Cas4	14	19	17	75	12	91	16	103	14	77	15	84	14	81
	SiL-0995	Cas2	11	25	4	71	29	58	8	57	13	63	12	71	15	83
	SiL-0996	Cas1	17	20	13	62	14	70	21	84	15	84	23	85	13	85
	SiL-0997	Csa1	5	6	2	7	6	11	6	12	1	10	3	12	2	18
III-B	SiL_0786	Cmr3	763	738	760	406	722	414	716	386	660	379	735	363	730	321
	SiL_0787	Cmr2	669	642	738	323	705	383	703	397	742	411	723	397	699	364
	SiL_0788	Cmr6	891	766	965	398	1028	507	1046	540	1083	617	997	609	1021	560
	SiL_0789	Cmr1	765	744	884	385	940	485	966	486	989	554	948	567	881	528
	SiL_0790	Cmr5	2415	2906	2987	1446	3564	1867	3121	1760	3508	2023	3518	2225	3196	2132
	SiL_0791	Cmr4	902	941	1035	538	1194	655	1211	656	1253	775	1226	745	1107	706
III-B	SiL_0600	Cmr3	250	304	307	848	368	1089	389	1136	385	1255	370	1062	372	1063
	SiL_0601	Cmr2	192	214	251	441	275	631	281	702	303	758	276	669	294	711
	SiL_0602	Cmr6	75	100	105	217	126	325	128	298	140	375	144	331	140	409
	SiL_0603	Cmr5	557	856	898	1889	1153	2728	979	2709	1126	2770	1212	3127	1305	4383
	SiL_0604	Cmr1	131	153	186	363	223	533	247	592	244	619	230	599	238	582
	SiL_0605	Cmr4	132	171	215	399	242	665	270	688	282	714	249	654	279	713

operon	gene	protein	t0C	t0I	t1C	t1I	t2C	t2I	t3C	t3I	t5C	t5I	t7C	t7I	t9C	t9I
I-D	SiL_0606	Cas3	143	154	219	1025	212	2242	216	3568	206	4878	196	5176	230	4873
	SiL_0607	Csc1	136	161	220	1226	245	2643	234	3792	242	4912	246	4849	235	4665
	SiL_0608	Csc2	1023	1219	1856	10439	2089	21157	2006	27497	2137	35689	2073	35991	1948	32103
	SiL_0609	cscA	236	297	317	3378	393	7077	378	9214	418	11656	422	11419	348	11071
I-D	SiL_0610	Csa3	36	30	50	191	44	200	40	171	49	135	53	101	28	61
	SiL_0611	Cas6	15	18	23	68	33	89	36	75	27	61	22	46	20	38
	SiL_0612	Cas1	3	6	6	49	2	64	3	61	7	61	5	65	4	71
	SiL_0613	cscB	20	19	22	156	26	161	24	188	32	173	24	178	19	192
I-A		CRISPR-1	16972	15337	15251	22509	13349	27778	13972	30632	14648	37093	15698	41785	15249	57281
I-A		CRISPR-2	4	3	5	38	2	51	6	50	1	86	7	65	6	92
III-B		CRISPR-3	1366	1167	1312	1771	1187	2323	1212	2551	1345	3249	1496	3589	1388	5249
III-B		CRISPR-4	1878	1674	1708	2246	1589	2789	1800	3278	1940	4648	2229	5725	2111	8443
I-D		CRISPR-5	1	3	2	20	1	17	1	21	2	21	0	37	4	18

Table S5 Predicted toxin and antitoxin genes in the *S.islandicus* LAL14/1 genome.
Of each gene the average amount of reads (given in RPKM) in uninfected control (C) and SIRV2 infected (I) cells, and their fold change in expression (FC) is depicted. Significantly up (green) or down (red) regulated genes are colour coded in the FC column. TA pairs are show in light/dark grey.

Toxin-antitoxin genes					
Family II: VapBC like					
gene	**arCOG**	**control**	**infected**	**FC**	
SiL_2040	arCOG00713	211	943	4,5	toxin
SiL_2041	arCOG00815	2423	5096	2,1	antitoxin
SiL_2042	arCOG07235	298	334	1,1	antitoxin
SiL_2043	arCOG00727	34	45	1,3	toxin
SiL_2080	arCOG00820	122	148	1,2	antitoxin
SiL_2081	arCOG06028	300	412	1,4	toxin
SiL_2134	arCOG00815	2615	8314	3,2	antitoxin
SiL_2135	arCOG02730	1310	4746	3,6	toxin
SiL_2176	arCOG00729	153	104	0,7	toxin
SiL_2177	arCOG02218	636	361	0,6	antitoxin
SiL_2253	arCOG02731	173	543	3,1	toxin
SiL_2254	arCOG00824	204	626	3,1	antitoxin
SiL_2575	arCOG02222	26	87	3,3	toxin
SiL_2576	arCOG00818	3	12	3,5	antitoxin
SiL_0190	arCOG03845	10	40	4,2	antitoxin
SiL_0382	arCOG00818	85	229	2,7	antitoxin
SiL_0383	arCOG02221	42	159	3,8	toxin
SiL_0359	arCOG02219	352	269	0,8	toxin
SiL_0360	arCOG08550	294	188	0,6	antitoxin
SiL_0413	arCOG02681	288	676	2,3	antitoxin
SiL_0414	arCOG02219	45	135	3,0	toxin
SiL_0431	arCOG00726	480	2403	5,0	toxin
SiL_0432	arCOG02217	448	2325	5,2	antitoxin
SiL_0619	arCOG00729	16	60	3,9	toxin
SiL_0620	arCOG02218	129	555	4,3	antitoxin
SiL_0631	arCOG00727	52	92	1,8	toxin
SiL_0632	arCOG02217	109	235	2,1	antitoxin
SiL_0634	arCOG00815	140	216	1,5	antitoxin
SiL_0635	arCOG02730	51	73	1,4	toxin
SiL_0734	arCOG07589	363	1285	3,5	toxin
SiL_0735	arCOG08216	1333	3607	2,7	antitoxin
SiL_0914	arCOG05414	568	3151	5,5	antitoxin
SiL_0915	arCOG03706	491	3091	6,3	toxin

Family II: HEPN-NT					
gene	**arCOG**	**control**	**infected**	**FC**	
SiL_0103	arCOG01191	292	603	2,1	antitoxin
SiL_0487	arCOG01203	198	251	1,3	toxin
SiL_0488	arCOG01192	63	134	2,1	antitoxin
SiL_0497	arCOG03707	40	88	2,2	antitoxin
SiL_0628	arCOG02106	27	97	3,6	toxin
SiL_0629	arCOG02109	25	94	3,7	antitoxin
SiL_0651	arCOG03721	276	520	1,9	antitoxin
SiL_0772	arCOG03721	601	2013	3,3	antitoxin
SiL_0781	arCOG03712	26	107	4,1	antitoxin
SiL_0782	arCOG01197	165	246	1,5	toxin
SiL_0783	arCOG01191	90	139	1,6	antitoxin
SiL_0784	arCOG03712	31	87	2,8	antitoxin
SiL_0795	arCOG03712	523	590	1,1	antitoxin
SiL_0798	arCOG03721	330	343	1,0	antitoxin
SiL_0885	arCOG03721	93	83	0,9	antitoxin
SiL_0925	arCOG01197	709	731	1,0	toxin
SiL_0926	arCOG01191	121	101	0,8	antitoxin
SiL_0927	arCOG01192	8	42	5,1	antitoxin
SiL_0944	arCOG01191	255	293	1,1	antitoxin
SiL_0945	arCOG07292	146	229	1,6	toxin
SiL_0947	arCOG01204	182	249	1,4	toxin
SiL_0948	arCOG01191	80	110	1,4	antitoxin
SiL_1888	arCOG04066	361	331	0,9	toxin
SiL_1975	arCOG03721	1325	316	0,2	antitoxin
SiL_2312	arCOG03712	217	373	1,7	antitoxin
SiL_2474	arCOG03721	448	1019	2,3	antitoxin
SiL_2475	arCOG03721	867	1701	2,0	antitoxin

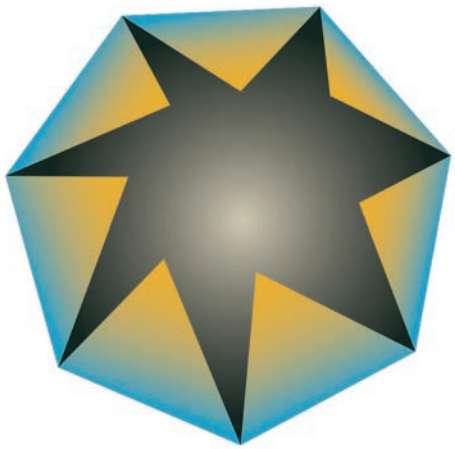
Chapter 6

Regulation of differential production of operon-encoded Cas proteins required for anti-viral defense

Tessa E.F. Quax, Yuri I. Wolf, Jasper J. Koehorst, Omri Wurtzel, Richard van der Oost, Wenqi Ran, Fabian Blombach, Kira S. Makarova, Stan J.J. Brouns, Anthony C. Forster, E. Gerhart H. Wagner, Rotem Sorek, Eugene V. Koonin, John van der Oost.

'Differential translation tunes uneven production of operon-encoded proteins'

Cell Reports. 2013 Sep 12. 4 (5): 938-944



Abstract

Clustering of functionally related genes in operons allows for co-regulated gene expression in prokaryotes. This is advantageous when equal amounts of gene products are required. Production of protein complexes with an uneven stoichiometry, however, requires tuning mechanisms to generate subunits in appropriate relative quantities. Using comparative genomic analysis, we show that differential translation is a key determinant of modulated expression of genes clustered in operons and that codon bias generally is the best *in silico* indicator of unequal protein production. Variable ribosome density profiles of polycistronic transcripts strongly correlate with differential translation patterns. In addition, we provide experimental evidence that *de novo* initiation of translation can occur at intercistronic sites, allowing for differential translation of any gene irrespective of its position on a polycistronic messenger. Thus, modulation of translation efficiency appears to be a universal mode of control in bacteria and archaea that allows for differential production of operon-encoded proteins.

Introduction

CRISPR-Cas

The whole transcriptome analysis presented in the previous chapter of this thesis, has revealed wide spread activation of anti-viral defense systems as a consequence of SIRV2 infection, notably CRISPR-Cas. Genomic CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) arrays consist of palindromic repeats which are regularly interspaced by short sequences which are matching to foreign genetic elements, like plasmids and viruses (34). These CRISPR arrays are transcribed and the precursor CRISPR RNA (pre-crRNA) is processed, resulting in small crRNA molecules that can be used as guides to specifically target complementary foreign genetic elements. Associated with these chromosomal CRISPR arrays, operons of *cas* genes are located that encode Cas proteins (34). Some Cas proteins are involved in the incorporation of new spacers in the CRISPR array and while others are required for the targeting of foreign genetic elements using the small processed spacers as guide. The role of CRISPR-Cas systems in adaptive immunity was discovered in 2005, and in the last years considerable efforts have been made by various groups to biochemically characterize the Cas proteins (32, 33, 178). Many different CRISPR-Cas systems exist, which all code for slightly different Cas proteins (146). In a single genome, multiple different CRISPR-Cas systems may be present, as is the case for *S.islandicus* that contains six CRISPR-*cas* arrays of three different types. The hallmark of type I and type III CRISPR systems is that the majority of their Cas proteins is forming a functional complex that targets invading foreign genetic elements by binding and/or cleavage of the alien DNA or RNA. These complexes typically contain an uneven stoichiometry; i.e. the type I-E Cascade complex has five protein subunits present in the ratio 1:2:6:1:1 (35). The three different *S. islandicus* Cas complexes have currently undetermined stoichiometry, Types I-A, I-D and III-B. All Cas complexes are encoded on operons. Whole transcriptome sequencing of *S. islandicus* cells shows that *cas* genes are co-transcribed as a polycistronic mRNA, consistent with the classical operon concept. Thus, to match requirements of the complex stoichiometry, the tuning of protein concentrations likely occurs on a different level.

The uneven stoichiometry of the Cas complexes is not unusual. In fact, many protein complexes with uneven stoichiometry play key roles in essential cellular processes (45, 179-185). The majority of them are, like the Cas complexes, encoded on operons, raising questions about the regulation of their stoichiometry. To address this question, in this chapter we perform comparative analyzes of prokaryotic genomes to study the possible correlation of subunit stoichiometry with features influencing gene expression.

Operons

The operon concept was developed over 50 years ago by Jacob and Monod (186). Their pioneering analyzes revealed a fundamental and characteristic feature of prokaryotic genome organization, i.e. clustering of functionally related genes (187). The operon organization allows for co-regulated gene expression (188-190). This is evidently advantageous when equimolar amounts of gene products are required, for instance to generate multi-subunit complexes with an even stoichiometry. However, a substantial number of operon-encoded multi-subunit complexes have an uneven stoichiometry and many of these play key roles in cellular processes like protein translation, secretion, energy conservation and anti-viral defense (45, 180, 185, 191). Although it is anticipated that a specific tuning mechanism is required to generate subunits of these complexes in appropriate relative quantities, the elucidation of its molecular basis is a long standing issue.

Control of subunit stoichiometry can theoretically be established at three levels: transcription, translation, and/or protein turnover. Only a few proteolysis substrates have been recognized to date and comparison of the protein degradation rates proteins awaits the generation of comprehensive proteomic pulse-chase databases (192). Although a minor contribution of different protein half-life values cannot be ruled out, it is considered most likely that prokaryotes avoid substantial energy loss by controlling different rates of subunit biosynthesis in order to obtain the appropriate relative quantities. Hence, uneven subunit stoichiometry of operon-encoded protein complexes is likely to be controlled by fine tuning of differential transcription and/or translation rates. In the classical operon model multiple genes/ cistrons are transcribed on a single polycistronic messenger, resulting in the same levels of mRNA segments that are part of the operon mRNA (186). Indeed, many full length polycistronic mRNAs have been identified experimentally, and in these cases differential transcription or stability cannot account for uneven protein output. The recent development of whole transcriptome sequencing has allowed for a more detailed analysis of operon transcription, which has revealed widespread internal transcription initiation and/or termination sites (193, 194).

Another possibility for differential production of operon-encoded proteins involves regulation of translation efficiency. Translation efficiency dependins on a range of features hidden in the non-coding and coding fragments of transcripts' nucleotide sequences and tuning may occur at the level of translation initiation and translation elongation (195-205).

Besides analysing relative production of the polypeptide end products, until recently no high throughput methods were available for the direct monitoring of translation efficiency. The recent development of a ribosome density profiling method provides insight into transcriptome-wide translation efficiency and offers the exciting possibility to study operon-encoded protein translation in greater detail (202, 206). In this Report, we present results of a comparative genomic

analysis of operon-encoded proteins showing that differential translation is a key determinant of modulated expression of individual genes that are part of operons.

Results

Selection of datasets

Expression of a given gene is influenced by many determinants, each of which plays a role in producing appropriate levels of the encoded proteins (199, 207-209). To understand how differential production of operon-encoded proteins is achieved, one should assess the relative contribution of each of these key controlling features to the expression of the individual genes within an operon. Apart from the rapidly growing database of archaeal and bacterial genomes (http://www.ncbi.nlm.nih.gov/genome/) the recent availability of several prokaryotic transcriptome data sets (194, 210-214) as well as bacterial ribosome density profiles (202, 206), provide the data that is required for addressing the issue of tuning uneven subunit stoichiometry.

A set of well-conserved operon-encoded protein complexes from prokaryotes was selected to allow for the identification of factors that correlate best with, and thus may be causative for, differential protein production. Ten operons were chosen on the basis of their established uneven protein stoichiometry, and because they are conserved in many bacterial or archaeal genomes. In addition, two operons were included as controls because they encode complexes in which all subunits are present in equal amounts (Tables S1 and S2). Comparative analyzes were performed on a set of 1055 bacterial and archaeal genomes, of which we selected a subset of 383 to avoid biases caused by the close relationships among some of the available genomes, e.g. at the species or sub-species level (Tables S1 and S2). All selected operons encode protein complexes that play key roles in important cellular processes (e.g. translation, secretion, energy production and defense) (45, 179-185, 191, 215).

Differential transcription

We first assessed whether differential transcription or, more specifically, different levels of mRNA segments encoding the cistrons in the selected set of operons, reflects differential production of subunits of complexes with uneven stoichiometry. To this end we used the high-throughput whole transcriptome sequencing data and tiling-array expression data for representative microbes including three bacteria and an archaeon (194, 210-214). Differences in the mRNA levels corresponding to genes within an operon could arise from internal transcription initiation/ termination and/or processing and differential decay of polycistronic mRNA (216). In addition, experimentally determined transcription start-site (TSS) maps for each organism were analyzed to identify potential alternative transcriptional units within the selected clusters. In the majority of

the analyzed operons (90%), deep transcriptome sequencing data showed that genes encoding different proteins were transcribed to similar levels (Table S1, Fig. S1). Thus, modulation at the transcription level, i.e. generating non-stoichiometric mRNA segment levels, may contribute to some extent, but does not appear to be a dominant factor in tuning the differential production of proteins.

Differential translation- *in silico* analysis

As the minor effect observed at transcription level cannot account for drastically different protein output from operons, we set out to analyze the contribution of differential translation. Previous analysis of (mono-cistronic) transcripts in prokaryotes has shown that several factors could contribute to the overall efficiency of the translation process. We analyzed the correlation of each of the factors with protein subunit stoichiometry, using the aforementioned dataset of well characterized complexes.

Translation initiation

Conventional translation initiation in bacteria involves binding of the 30S ribosomal subunit to the ribosome binding site (RBS) of an mRNA. This is generally dependent on the Shine-Dalgarno (SD) sequence which base-pairs with the anti-SD sequence in 16S rRNA to guide selection of the correct start codon. The rate of translation initiation depends on, (i) the strength of interaction of SD/anti-SD base-pairing (217), and on (ii) the accessibility of the ribosome binding site (involving primarily SD and/or the start codon), which is negatively affected by stable secondary structure (205). In the absence of a canonical SD-motif, the codon following the initiation codon may affect translation initiation efficiency (200).

By calculating the RNA hybridization energy between the SD sequences and the anti-SD sequences (G) of the genes within the selected operons, the majority of the RBS's failed to reveal statistically significant differences between genes in operons (Table S2). Similarly, analysis of adenine enrichment of the second codon showed no association with stoichiometry of the complex subunits (Table S2). These observations suggest that the affinity of the SD/anti-SD interaction and the nature of the second codon do not play major roles in differentiating translation efficiency.

Next, the propensity of forming secondary RNA structure (E) in RBS regions of genes (-20 to +20 bp relative to the start codon) of selected operons was calculated. To compute the operon-specific values, accounting simultaneously for genome-specific and operon-specific biases, the mean for the respective operon was subtracted from individual gene values (ΔE value) and the differential was adduced over the complete set of operons of the given type. This analysis revealed a moderate but significant correlation between subunit stoichiometry and mRNA folding (ΔE: $r_{Spearman}$=0.57, P=0.0092; Figs. 1A and B, S2, S3 and Table S2). In addition, we

calculated the propensity of forming secondary RNA structure in the region -4 to +37 bp relative to the start codon which showed a similar trend as the calculations for the other region (ΔE: $r_{Spearman}$=0.47, *P*=0.0188; Table S2). These results indicate that the structural accessibility of RBS might contribute to the differential gene expression in several of the selected operons.

Translation elongation

The efficiency (accuracy and/or rate) of translation elongation depends on the coding sequence of a gene (196, 201). Codon bias reflects differences between isoacceptor tRNAs with respect to abundance, amino acid charging, and kinetics (197, 218, 219). In addition, translation is enhanced by co-occurrence of isoaccepting codons (195, 204), i.e. codons that are recognized by the same tRNA (by Watson-Crick base pairing or by wobbling, which allows non Watson-Crick base pairing between two nucleotides of interacting RNA molecules). A comparative analysis of codon usage and co-occurrence was performed in genes of the selected operons (Table S2). For each analyzed genome, the frequency of optimal codons (220), the classical Codon Adaptation Index (CAI) (203), the percentage of optimal codons (F) and codon co-occurrence (Co) (195) were calculated. A significant correlation was observed between protein subunit stoichiometry and family-averaged scores for codon co-occurrence (ΔCo: $r_{Spearman}$=0.63, *P*=0.0099) (Figs. 1A and D, S2, S3 and Table S2). The results obtained for CAI and F were in excellent agreement and indicated that, in all selected operons, the genes encoding the most abundant proteins (>2 copies per complex) contain the highest percentage of optimal codons and display the highest CAI values (Figs. 1A and C, S2, S3 and Table S2). A threshold value of 0.02 for the family average value ΔF was determined to minimize the error rate for the prediction of high-stoichiometry subunits, resulting in prediction accuracy of 96% for the analyzed set of operons (Fig. S4). Overall, for all selected operons, there was a strong correlation between codon usage and protein stoichiometry (for ΔF: $r_{Spearman}$=0.75, *P*=0.0002; for ΔCAI: $r_{Spearman}$=0.71, *P*=0.0002; Experimental procedures).

Ribosome profiling

To determine *in vivo* translation rates, we analyzed ribosome density profiles of *Escherichia coli* and *Bacillus subtilis* that were recently reported by Weissman and colleagues (202). The ribosome profiling strategy allows for quantitative monitoring of protein production *in vivo* because ribosome density values closely correlate with translation efficiencies if the cross-linked ribosomes are evenly distributed over the coding sequence (206). On the other hand, local peaks in density characterize sequences with lower translation efficiency due to ribosome stalling (202). For the selected operons, we found a strong correlation of ribosome density with protein subunit stoichiometry (Figs. 1A and E, S5 and Table S2). This finding is in perfect agreement with our conclusion that uneven production of subunits of operon-encoded protein complexes is tuned by differential translation.

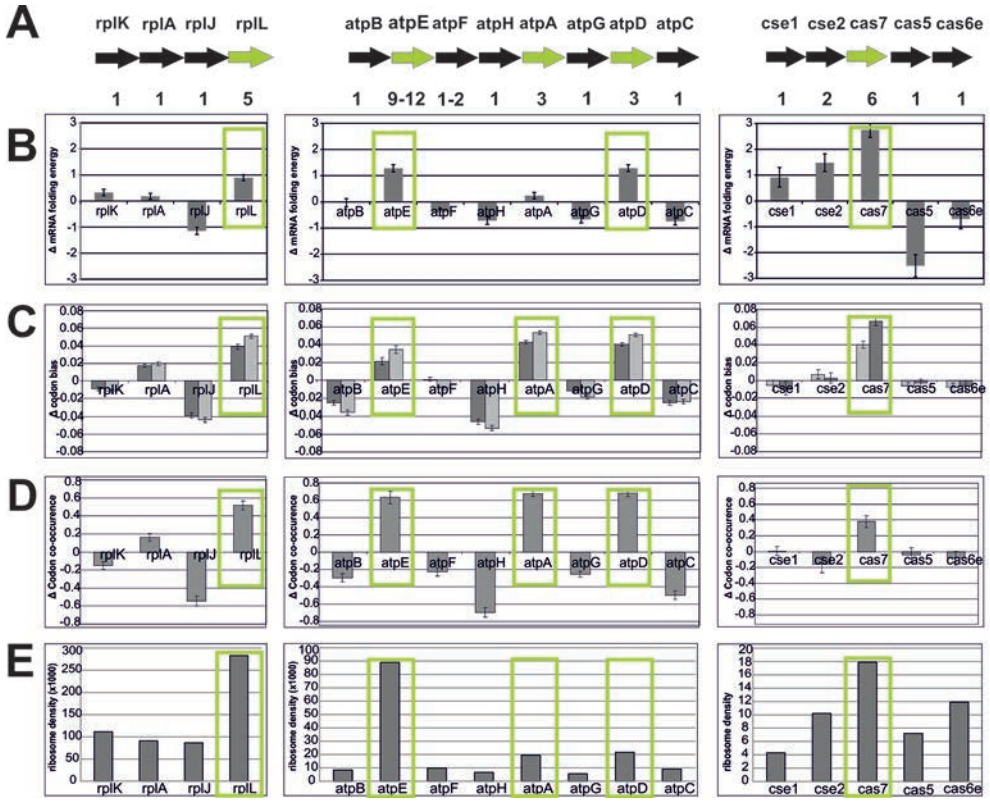


Figure 1. Cistron properties correlated with stoichiometry of operon-encoded protein complexes. The analysis of the ribosomal protein operon L7/L12 (high expression), the F-type ATPase (moderate expression) and the Type I-E Cascade complex (low expression) are depicted. (**A)** selected operons (block arrows), and the stoichiometry in the corresponding protein complexes; (**B)** predicted mRNA folding energy (ΔE) of the RBS region of each cistron (-20 to +20 bp relative to the start codon). **(C)** codon bias; ΔF (optimal codon usage) is shown by dark grey, and ΔCAI (Codon Adaptation Index) is shown by light grey; **(D)** codon co-occurrence (ΔCo); **(E)** ribosome density profiles per gene. Error bars indicate standard deviations. The green arrows represent genes in each operon that encode the most abundant subunit(s), and green rectangles denote the corresponding positive deviations in codon bias (ΔF>0.02), codon co-occurrence (highest value), low RNA folding potential (highest energy value) and/or ribosome density (highest value). See also Figs. S1-S5 and Table S2.

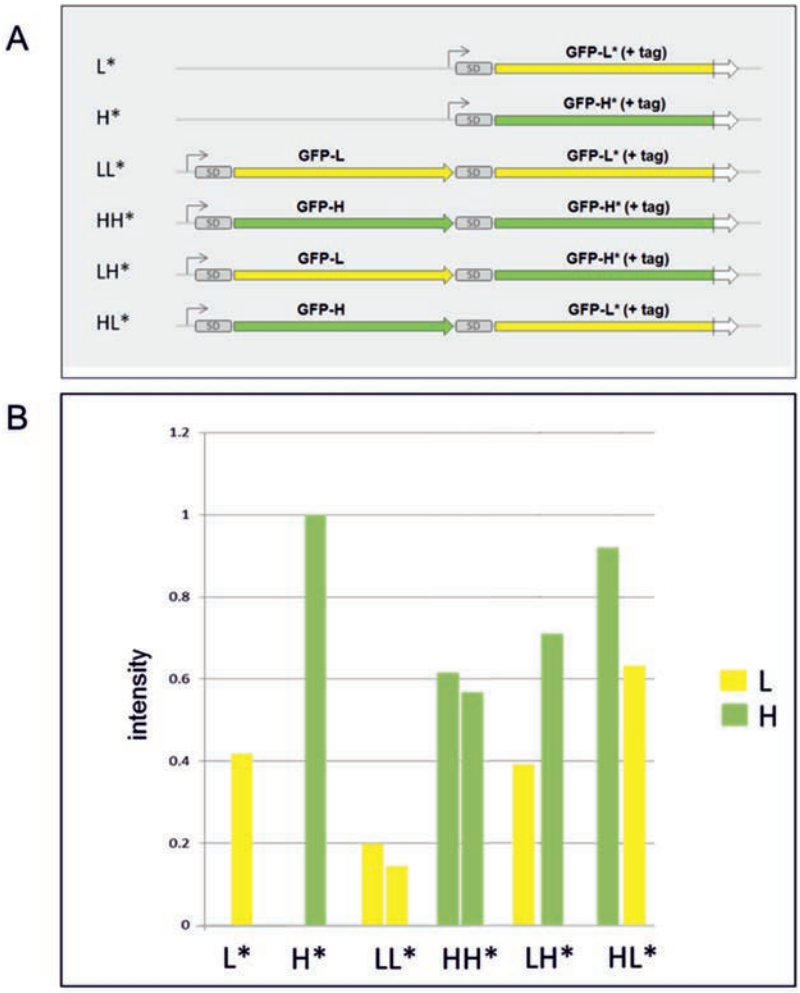


Figure 2. Translation efficiency influences protein expression from individual genes within operons. (**A)** expression constructs under control of the T7 promoter (grey arrow) encoding different combinations of two identical GFP polypeptides, with synonymous mutations, resulting in low (yellow block arrow, *gfp*L) or high (green block arrow, *gfp*H) translation efficiency; the single genes, as well as the downstream gene of the operons are translational fusions to a Strep-tag (white block arrow; in construct name indicated by an asterisk, *); SD, SD-sequence (Fig. S6). **(B)** Quantification of Western blot with GFP antibody on whole cell lysates of the variant GFP-expressing constructs after expression in *E.coli*; equal total amounts of cellular proteins were loaded in order to allow comparison between different samples (Fig. S6).

De novo translation initiation

The fact that genes coding for abundant subunits do not necessarily occupy the first position in the respective operons (Figs. 1, S2, S3 and S5) implies that internal translation initiation is required to allow for elevated translation rates of a downstream cistron. To investigate the possibility of internal initiation of translation, we used operons consisting of two synthetic Green Fluorescent Protein (GFP)-encoding genes with synonymous mutations (*gfp*L and *gfp*H, with Low and High expression, respectively, Figs. 2A and S6). After induction of expression in *E.coli*, detection by Western blot analysis showed that moderate levels of GFP variants were produced (Figs. 2B and S6). When *gfp*L and *gfp*H were combined in one operon, the expression level of *gfp*H was substantially higher than that of *gfp*L, irrespective of its position in the operon (Figs. 2B and S6), confirming that the position of a cistron in the operon does not appreciably affect protein expression levels. It is concluded that the differential protein production of GFP-L and GFP-H resulted from differences in translation efficiency, in accord with internal *de novo* translation initiation.

Discussion

In line with the definition of an operon, our analyzes indicate that differential expression of individual genes in the selected operons shows only limited dependence on differential transcription. Rather, we provide evidence that expression of genes in operons is predominantly controlled at the translation level. While protein degradation may not be the most economical method for a cell to balance protein ratios, we cannot rule out its importance.

Elevated overall translation of a cistron, potentially at any position on a poly-cistronic mRNA, requires enhanced translation initiation rates. Indeed, in several instances we found significant correlations between subunit stoichiometry and the mRNA folding energy in the RBS region of the analyzed genes (Figs. 1, S2, S3 and Table S2). However, a relatively high number (~20%) of false positives was obtained, i.e. genes with the highest folding energy that did not code for the most abundant subunit. This might be a consequence of difficulties to predict the correct mRNA structure, since the setting of the RNA fold analysis (i.e. the selected sequence window to be analyzed) has a strong influence on the outcome of the structure prediction (221). Instead, subunit abundance showed the strongest correlation with the fraction of optimal codons in a gene (Figs. 1, S2, S3 and Table S2). Thus, codon optimization is the most reliable *in silico* indicator of subunit stoichiometry of operon-encoded complexes.

Complementary to the *in silico* analyzes, a meta-analysis of experimentally determined ribosome density profiles confirms the widespread occurrence of differential translation that explains operon-encoded protein expression. The observation of increased ribosome densities

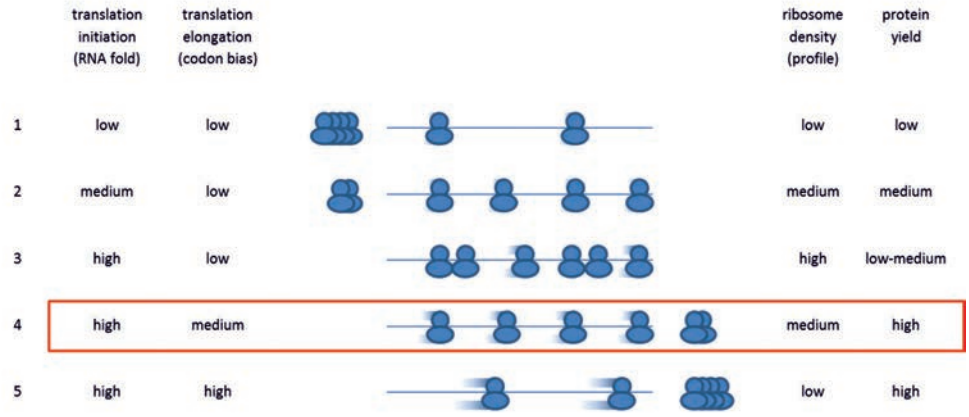


Figure 3. Models describing tuning of the translation process. Five scenarios are depicted with different rates of translation initiation and/or translation elongation. The corresponding ribosome density profiles and the expected protein yield are based on the assumption that differential transcription is insignificant (as has been demonstrated experimentally for the majority of the operons analyzed in this study). Scenario-1 relatively little protein is produced (e.g. single subunit per operon-encoded complex), scenario-2 relatively little expression is required (no significantly different codon bias). Scenario-3 a hypothetical case when high translation initiation results in ribosome jamming as the elongation rate is not optimized. Scenario-4 relatively much protein is produced (e.g. multiple subunits per operon-encoded complex): high translation initiation (low RNA fold), elevated but not maximal elongation rate (codon adaptation), elevated but not maximal ribosome densities (experimental profiles), and high protein yield (experimental protein complex stoichiometry values). Scenario-5 might also lead to high protein yields, but the detected elevated ribosome densities suggest that elongation rates are not maximal, favouring scenario 4 (box) as approaching the *in vivo* situation of translation-controlled overexpression.

implies that translation initiation is significantly increased for genes encoding highly abundant proteins in the selected operons. The correlation between codon adaptation and subunit stoichiometry reflects enhanced translation elongation rates in the cistrons coding for abundant subunits (201) (Fig. 3). The increased rate of translation elongation could contribute to the avoidance of ribosome crowding (198). These findings suggest that internal translation initiation and termination are essential for differential translation of cistrons at any position within polycistronic transcripts (Fig. 4 and movies S1 and S2). The occurrence of *de novo* translation initiation at inter-cistronic RBS's has previously been deduced from the analysis of some bacterial operons, and translational coupling was only demonstrated for a subset of the genes in an operon (222, 223). In the present study we provide *in silico* and *in vivo* data that indicate the frequent occurrence of *de novo* inter-cistronic initiation of translation (Fig. 1 and Table S2). Moreover, we demonstrate that uncoupled translation of polycistronic messengers allows for differential translation (Figs. 2 and S6). This initial experimental support for the model can serve

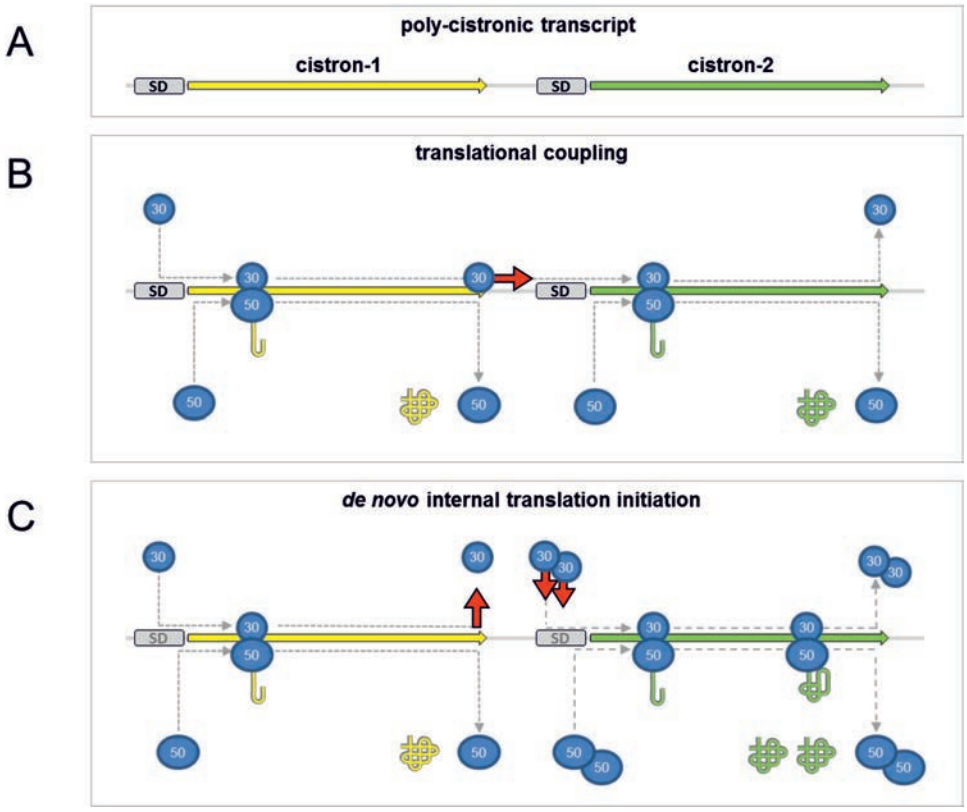


Figure 4. Model for translation of polycistronic messengers in prokaryotes. **(A)** Transcript consisting of two cistrons, each preceded by a SD-sequence. **(B)** Translational coupling where the 30S ribosomal subunit remains associated (red arrow) after termination, and 50S joins for re-initiation, resulting in stoichiometric output from both cistrons (Movie S1). **(C)** *de novo* internal recruitment (red arrows) of both 50S and 30S subunits allows for differential translation initiation rates between cistrons; depending on the translation elongation rate (codon bias) of each cistron, this may result in different ribosome density profiles (Movie S2); yellow and green cistrons have low and high translation efficiency, respectively (Fig. 2).

as a prelude to wider, subsequent experimental testing using both natural operon sequences and synthetic biology approaches.

Apart from the selection of 10 widespread operons analyzed in this study, it should be emphasized that many more (if not all) prokaryotic protein complexes with uneven stoichiometry are likely to rely on differential translation for tuning of their protein levels. Similarly, operons that encode enzymes of metabolic pathways might employ differential translation in case the enzymes are required in uneven quantities. Overall, we conclude that modulation of translation efficiency is a universal mode of control in bacteria and archaea that allows for differential protein expression of operon-encoded multi-subunit complexes with uneven stoichiometry. This fundamental principle can be applied to predict the ratios between protein subunits of

uncharacterized complexes (e.g. CRISPR-Cas complexes) as well as for design of synthetic operons.

Experimental procedures

Transcriptome analysis

The expression of genes and their division to transcriptional units were extracted from studies providing high-throughput cDNA-sequencing data (for *E. coli, S. solfataricus,* and *P. aeruginosa)*, transcription start-sites (TSS) maps (for *E. coli, S. solfataricus, P. aeruginosa,* and *L. monocytogenes*), and tiling array data (*L. monocytogenes*) (194, 210-214). High-throughput cDNA sequencing reads were mapped to the corresponding genome (Genbank: NC_000913, NC_002754, NC_008463 for *E. coli, S. solfataricus, and P. aeruginosa* PA14, respectively) with Novoalign V2.07 with default parameters, as previously described (224). Gene coordinates were extracted from the reference genomes, and gene expression values were calculated as the number of mapped sequencing reads that overlapped the gene normalized by library size and gene length, previously described as reads per kilobase of gene model per million sequencing reads (RPKM) (225). Genes were assigned to different transcriptional units according to the available TSS-maps and cDNA sequencing data (194, 213, 214). When a TSS was found upstream to a gene or when previous transcriptome annotation indicated that a gene begins an operon it was allocated to a new transcriptional unit (TU). The genes downstream to it, without TSSs, were assigned to the same TU (212).

SD/anti-SD

Genes constituting 12 different operons encoding multi-subunit complexes of known stoichiometry were obtained from the NCBI database. Of 1055 available prokaryotic genomes a selection was made to avoid a bias due to the close resemblance of the available genomes, so that 383 genomes remained for further analysis. The RNA hybridization energy (kcal/mol) between the potential Shine-Dalgarno sequences located at position (-17) to (-1) relative to the start codon of the gene, and the ten 3'-terminal nucleotides of the 16S rRNA was predicted using the UNAfold program (226) . In addition, we estimated the scores for the occurrence of adenine at the 1st, 2nd and 3rd codon position (0.40, 0.23 and 0.15, respectively) (G) from previously reported data (200), and scored the 2nd codon of each gene in the operons. Family-specific score shifts were calculated in the same manner as for codon bias values (ΔG).

mRNA secondary structure

The same genes of the selected operons were extracted from the NCBI database from the set of 383 prokaryotic genomes as described for the SD/anti-SD analysis. The correlation between protein subunit stoichiometry and the accessibility of Shine-Dalgarno sequence (227) and of the start codon of each cistron was computed using the UNAFold program (226) for positions (-20) to (+20) relative to the CDS start codon and (-4) to (+37) as suggested previously (205). The lowest energy (kcal/mol) of predicted structure(s) was used to characterize the folding propensity of the RBS region of each cistron in the mRNA. Family mean folding energy shift values were calculated in the same manner as the codon bias values.

Codon bias

The potential correlation of codon bias with differential translation of polycistronic messengers was computed relative to the codon composition of annotated ribosomal protein genes in each organism. Genes constituting 12 different operons encoding multi-subunit complexes of known stoichiometry were obtained from the NCBI database from the set of 383 prokaryotic genomes as described for the SD/anti-SD analysis. For each analyzed genome, ribosomal protein genes were used to identify 'optimal' codons for each of the 18 amino acids with multiple synonymous codons. The relative frequency of occurrence of these codons within individual genes was calculated (220). To compute the family-specific codon bias, accounting simultaneously for genome-specific and operon-specific biases, the mean for the respective operon was subtracted from this frequency (ΔF value) and adducted over the complete set of operons of the given type. In similar fashion the Codon Adaptation Index (CAI) (203) values were computed for each coding sequence in the selected operons and family specific values were calculated by subtracting the mean of the operon from values of individual genes.

Codon co-occurrence

On the same set of genes from selected operons mentioned above, analysis of codon co-occurrence was performed as described before (195). Briefly, for each series of codons, coding for the same amino acid, the frequency of consecutive pairs (isoaccepting codons recognized by identical tRNA's) was computed across the genome (ignoring codons that code for different amino acids). For each such pair the score characterizing the deviation of the observed frequency from its random independent expectation was calculated; then scores for all genes were calculated using these score tables. Family mean codon co-occurence shift values were calculated in the same manner as the codon bias values.

Statistical analysis

Spearman rank correlation coefficients were calculated for stoichiometry and either the ΔF, ΔCAI, ΔCo or ΔE values; p-value was estimated by permuting the data. The optimal threshold values were determined by minimizing the error rate (the sum of false positives and false negatives) in the dataset of operons with uneven stoichiometry.

Plasmid constructs and protein detection

An oligonucleotide sequence was synthesised by GeneArt (Germany), which contained a bicistronic operon under control of the T7 promoter. The two cistrons were C-terminally his tagged *gfp* genes differing in synonymous codon usage, with an additional C-terminal Strep-tag attached to the second cistron. The SD-motifs were identical for both cistrons. The sequences of *gfp*L, translated with low efficiency, and *gfp*H, translated with high efficiency, are identical to GFP_156 and GFP_182 of a previous study, respectively (205). The % optimal codons (0.35 vs 0.55) and codon co-occurrence (2.23 vs 4.71) values differ considerably between both genes. For the complete sequence of the synthetic oligonucleotide, see Fig. S6. This synthetic operon was cloned in the vector pET-52b(+) (Novagen) using SphI and AvrII/NheI sites. Restriction sites creating compatible overhangs 5'and 3'of each *gfp* gene, enabled cloning of six constructs containing one or two *gfp* genes in different combinations. We used pET-52(b) containing a commercial *gfp* gene optimized for high expression and the empty pET-52(b) plasmid as positive and negative controls, respectively. Each construct was transformed into *E.coli* BL21(DE3) and induced with IPTG. After two hours, cell pellets were harvested, resuspended in 50 mM TrisHCl, 150 mM NaCl pH7.4, sonicated and centrifuged 30 min at 10000 rpm. Protein concentrations in the cell free lysate were determined using Roti®-Nanoquant and equal amounts of protein were loaded on 15% SDS TrisHCl gels and run in Tris-Glycine buffer. ColorPlus Prestained Protein Marker Broad Range #7709 (New England Biolabs) was used as a marker. Total proteins were detected with Coomassie Brilliant Blue R250 staining. A duplicate of the gel was used for Western hybridization and proteins were transferred onto a PVDF membrane. GFP was detected using a 1:1000 dilution of mouse anti-GFP IgG (Roche) and peroxidase-coupled sheep anti-mouse IgG (GE Healthcare) and the SuperSignal West Pico Chemi-luminescent Substrate (Thermo Scientific). Visualisation was done on a *ChemiDoc* XRS+ imaging system (Bio-Rad) and quantification was performed with Image Lab software.

Acknowledgements

This project was funded by the Netherlands Organization for Scientific Research (NWO) via an ALW TOP grant (854.10.003) to J.v.d.O., and an ALW Veni grant (863.08.014) to S.J.J.B. R.S was supported by an ERC-StG grant and O.W by an Azrieli PhD fellowship. T.E.F.Q. was supported by a fellowship from the Ministère de l'Enseignement supérieur et de la Recherche de France. W.R., K.S.M., Y.I.W and E.V.K. were supported by the U.S. Department of Health and Human Services intramural program (NIH, National Library of Medicine). A.C.F. and E.G.H.W. were supported by grants from the Swedish Research Council, including the Uppsala RNA Research Centre. We thank Jonathan Weissman and Gene-Wei Li (UCSF) for providing ribosome profiling data and stimulating discussions.

Supplemental information

Supplemental Figures

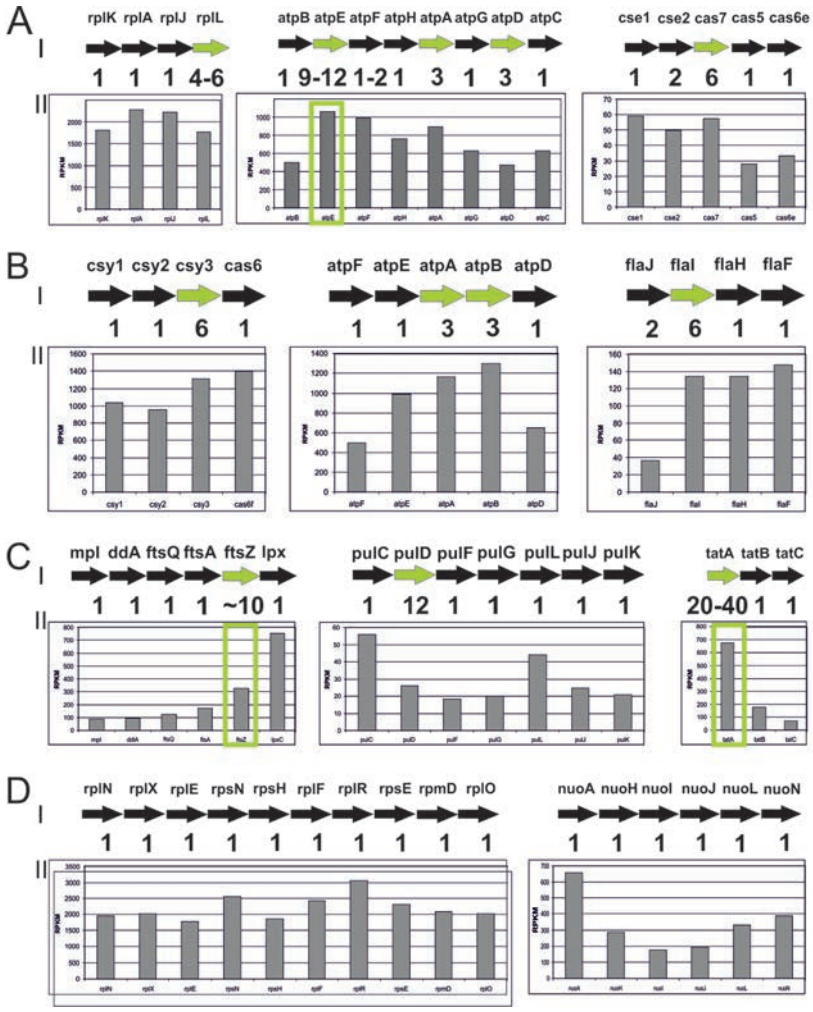


Figure S1. Transcription of operons. Analyses of mRNA levels of the following operons are depicted; **(A)** ribosomal protein operon L7/L12 (180), the F-type ATPase (45) and the Type I-E Cascade complex (185). **(B)** the Type I-F Cascade complex (*P. aerugionosa*) (228), the V-type ATPase (179), the archaeal flagellum (*S.solfataricus*). **(C)** the FtsZ operon (182), the Type II secretion system (191), TAT secretion system (184). **(D)** the ribosomal protein operon L5 (180) and the NADH dehydrogenase (181). **I**, the selected operons (block arrows), and the stoichiometry in the corresponding protein complexes are shown. **II**, RNAseq data shown in Reads Per Kb per Million (RPKM) from three representative prokaryotes. All RNAseq data shown in this figure is originating from *E.coli* unless specified otherwise. The light grey arrows are genes in each operon that encode the most abundant subunit(s), and light grey rectangles highlight corresponding positive transcription values (RKPM).

Figure S2. Cistron properties correlated with stoichiometry of operon-encoded protein complexes. Translational factors that contribute to fine tuning of uneven stoichiometry potentially occur. As representative examples are depicted the analyses of the **(A)** Type I-A, I-B, I-C Cascade complex (229), Type I-F Cascade complex (228), the archaeal flagellum (183) **(B)** the V-type ATPase (179), the FtsZ operon (182). I, selected operons (block arrows), and the stoichiometry in the corresponding protein complexes. II, mRNA folding energy in the region around the RBS (-20 to +20 bp relative to the start codon) (ΔE) as measured by UNAFold program. III codon bias (ΔF optimal codons (dark grey), and ΔCAI (light grey)), IV codon co-occurrence (Δ Co) Error bars indicate standard deviations. The light grey arrows represent genes in each operon that encode the most abundant subunit(s), and light grey rectangles highlight corresponding positive deviations in codon bias (ΔF > 0.02), codon co-occurrence (highest value) and/or low RNA folding potential (highest energy value).

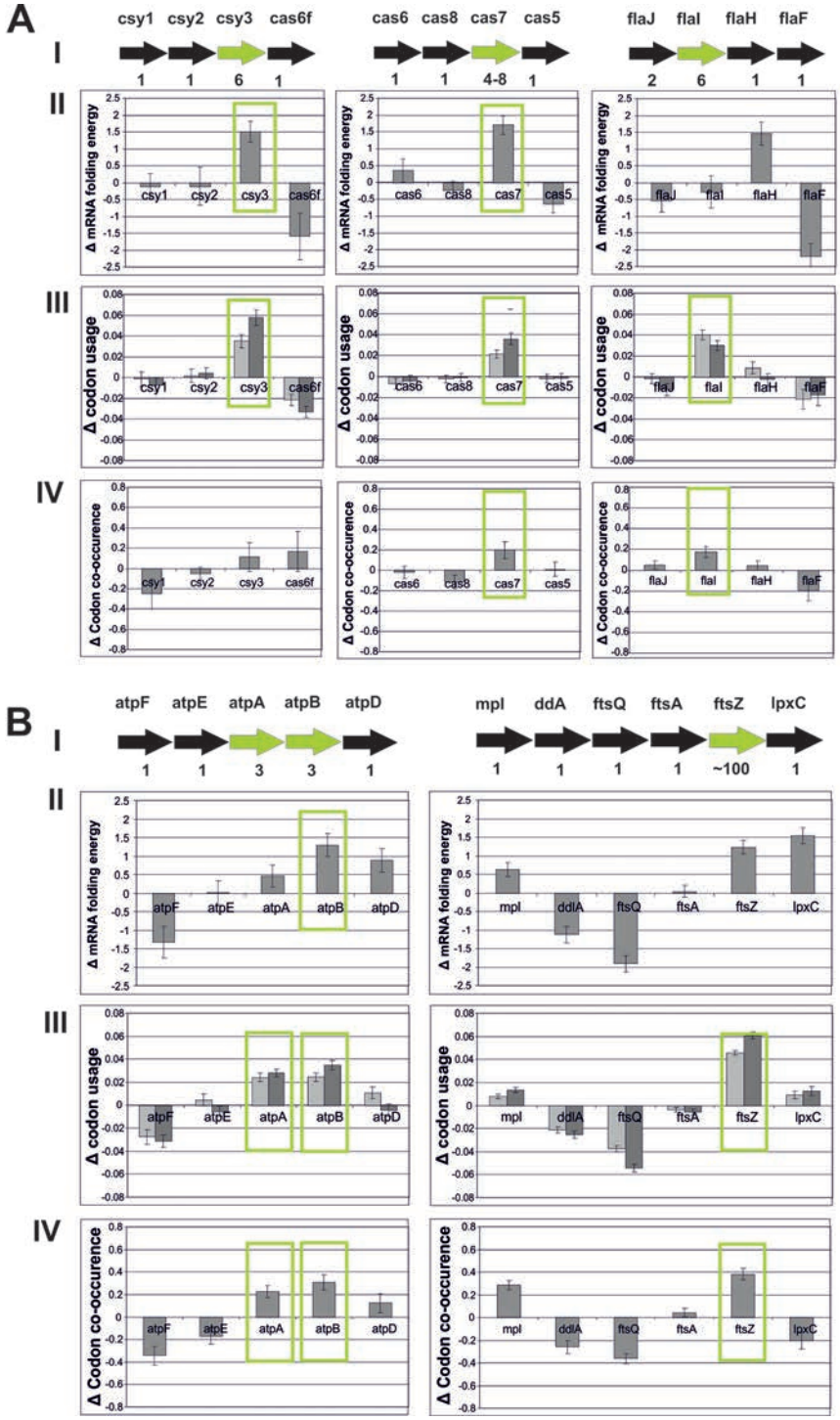
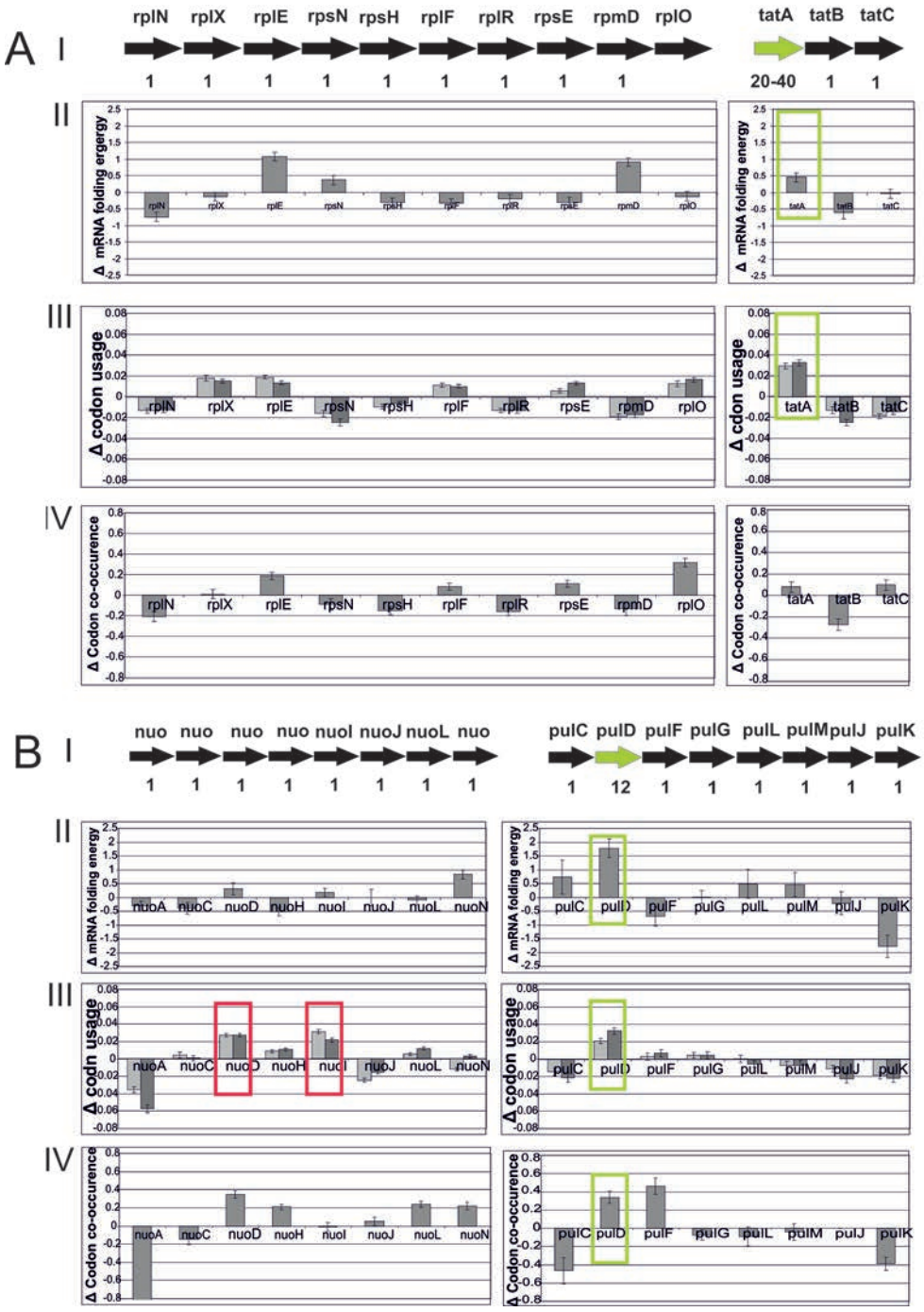


Figure S3. Cistron properties correlated with stoichiometry of operon-encoded protein complexes. Translational factors that contribute to fine tuning of uneven stoichiometry potentially occur. As representative examples are depicted the analyses of the **(A)** the ribosomal protein operon L5 (180) and the TAT secretion system (184) **(B)** the NADH dehydrogenase (181) and Type II secretion complex (191). I, selected operons (block arrows), and the stoichiometry in the corresponding protein complexes. II, mRNA folding energy in the region around the RBS (-20 to +20 bp relative to the start codon) (ΔE) as measured by UNAFold program. III codon bias (ΔF optimal codons (dark grey), and ΔCAI (light grey)), IV codon co-occurrence (Δ Co) Error bars indicate standard deviations. The light grey arrows represent genes in each operon that encode the most abundant subunit(s), and light grey rectangles highlight corresponding positive deviations in codon bias (ΔF > 0.02), codon co-occurrence (highest value) and/or low RNA folding potential (highest energy value); false positives (ΔF > 0.02, stoichiometry < 2; see Figure S4) are indicated with dark grey rectangles.



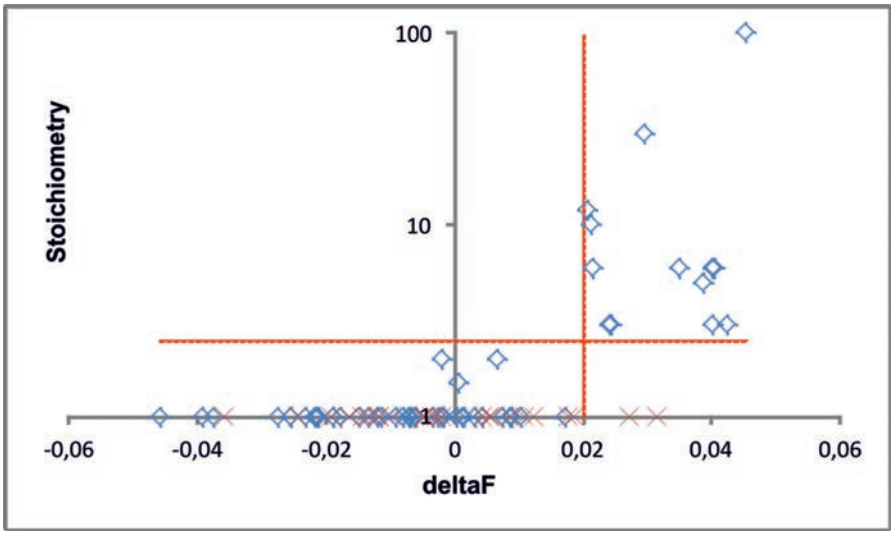


Figure S4. Codon bias and subunit stoichiometry in bacterial and archaeal operons. The horizontal axis shows the family mean codon usage shift (ΔF), i.e. the fraction of optimal codons; Fig. 1C); the vertical axis shows the (approximate) number of subunits in the complex. Diamonds show genes belonging to the complexes with uneven stoichiometry; crosses show genes belonging to the complexes with even stoichiometry. The dotted lines show the ΔF and stoichiometry thresholds providing the best separation between 13 true positives (upper right quadrants), 41 true negatives (lower left quadrants, respectively) and 2 false positives (lower right quadrant; discussed in text). Details on the selected complexes are provided in Table S2.

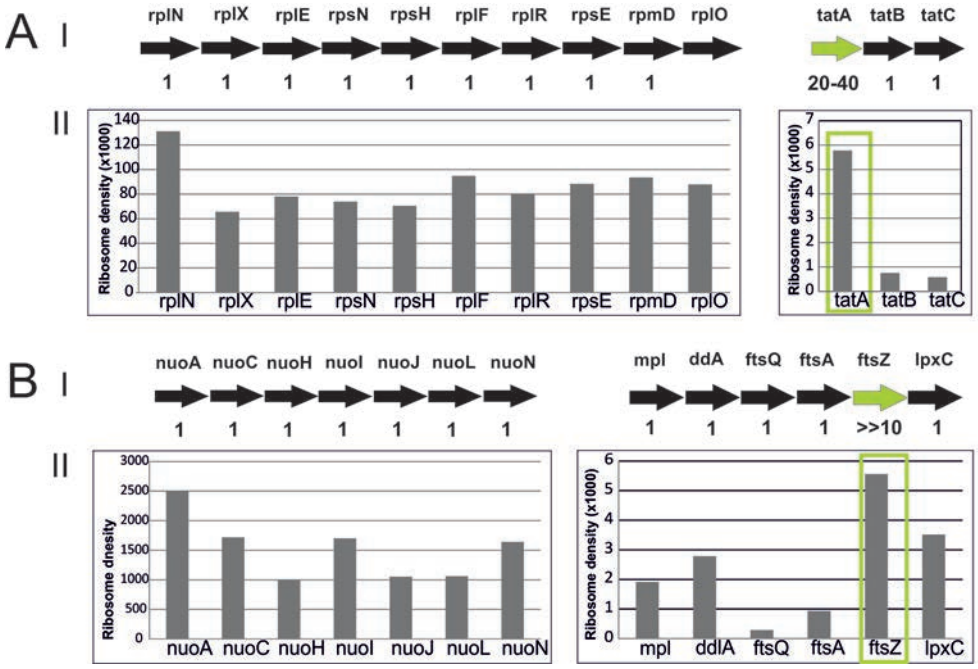


Figure S5. Ribosome density profiles of polycistronic transcripts coding for protein complexes with uneven stoichiometry. The analysis of selected *E.coli* operons. Depicted are (**A**) the ribosomal protein operon L5 (180) and the TAT secretion system (184). (**B**) The NADH dehydrogenase (181) the FtsZ operon (182). **I,** selected operons (block arrows), and the stoichiometry in the corresponding protein complexes; **II,** Density of ribosomes per cistron; The light grey arrows represent genes in each operon that encode the most abundant subunit(s), and light grey rectangles denote the corresponding positive deviations of ribosome density.

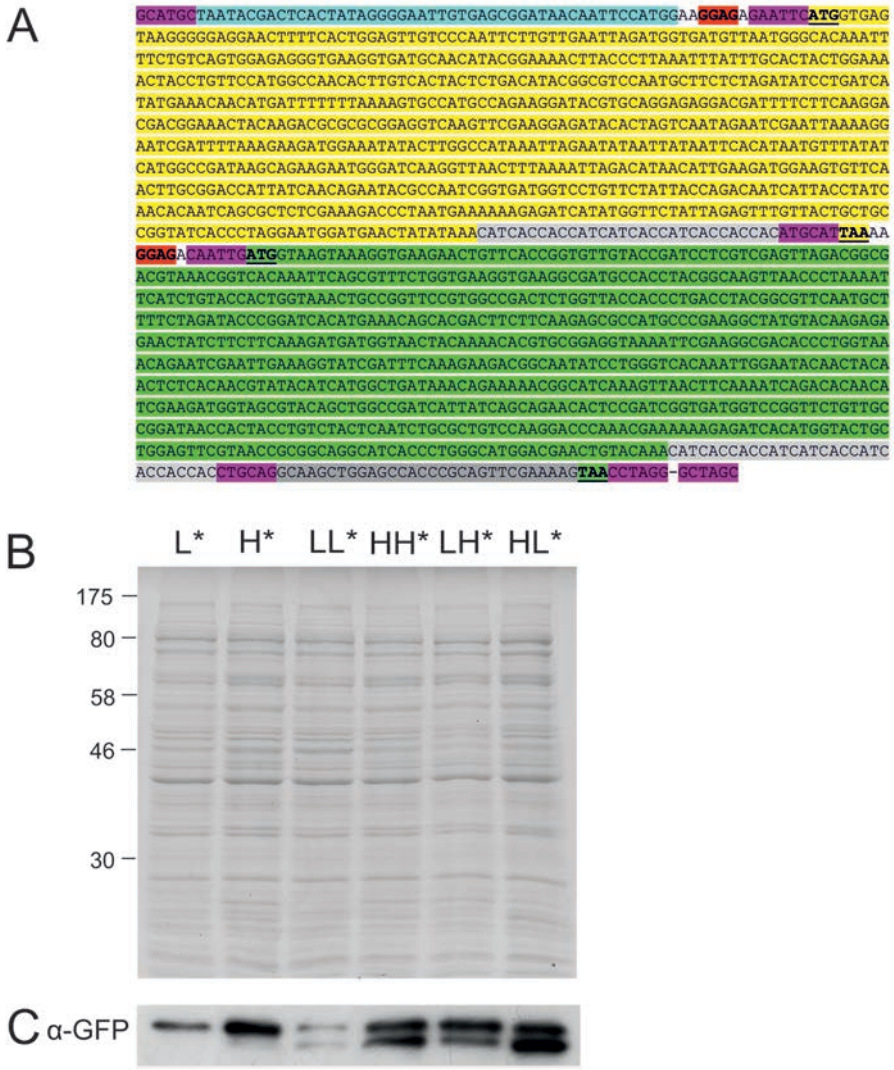


Figure S6. Expression of synthetic operons. (A) Sequence of synthetic operon as depicted in Fig. 2a, with two GFP genes with synonymous mutations which was cloned in pET52b (LH*). Relevant features are indicated: *gfp*L (low efficiency of translation; yellow), *gfp*H (high efficiency of translation; green), cloning sites (purple; GCATGC, SphI; Compatible GAATTC/ CAATTG, EcoRI/ MfeI; Compatible ATGCAT/ CTGCAG, NsiI/ PstI; CCTAGG, AvrII; GCTAGC, NheI), T7-promoter (blue) SD- motif (bold, red), start/stop codons (bold, underlined), His-10-tag (light gray), strep-tag (dark gray). Theoretical molecular weights are: 28.6 kDa (GFP-L) and 29,8 kDa (GFP-H+strep-tag). **(B)** Coomassie stained SDS-PAGE gels of whole cell lysate of *E. coli* containing pET52b-based expression constructs (Fig. 2A) encoding different combinations of two identical GFP polypeptides, with synonymous mutations, resulting in low (*gfp*L) or high (*gfp*H) efficiency of translation. The SDS-PAGE gel is a replicate of the one used for Western blot (C). Based on a marker (#7709, NEB®) molecular masses (kDa) are indicated. **(C)** Western blot with GFP antibody on whole cell lysates of the variant GFP-expressing constructs after expression in *E.coli*; the upper band corresponds to the slightly larger Strep-tagged GFP-variant (*).

Supplemental Movies

Supplemental Movie S1- follow the link: http://vimeo.com/surfrender/protein-synth-even

Even stoichiometry by translational coupling. This movie shows transcription of an operon consisting of two genes (yellow and orange), and translation of the polycistronic messenger. Translational coupling results in even stoichiometry of the produced proteins.

Supplemental Movie S2- follow the link: http://vimeo.com/surfrender/protein-synth-uneven

Uneven stoichiometry by *de novo* internal translational initiation. Like movie 1, this movie shows transcription of an operon consisting of two genes (yellow and green, like the one described in Fig. 2), and translation of the polycistronic messenger. Because of intercistronic initiation involving both newly recruited 30S and the 50S ribosomal subunits, and because of optimal translation efficiency in the second cistron, this results in higher protein output from to the second reading frame of the operon mRNA: uneven stoichiometry.

Supplemental Tables

Table S1. RNA expression of selected prokaryotic operons with known stoichiometry.
Subunit stoichiometry values are based on experimental analyzes. Highlighted in light grey are abundant subunits (stoichiometry column), as well as experimental transcription values. Abbreviations: TU, transcription units; RPKM, reads per kilobase of gene model per million sequencing reads. NA, not available (operon (grey) or gene (black) not present in genome).

	Gene information		Transcription data												
			E. coli			*P. aerugionosa*			*L. monocytogenes*		*S. Solfataricus*			Additional data		
Complex	Profile ID	Stoichi-ometry	Locus	RPKM	# TU	Locus	RPKM	# TU	Locus	# TU	Locus	RPKM	# TU	Locus	RPKM	# TU
Archaeal flagellum	flaJ	2	NA	NA	NA	NA	NA	NA	NA	NA	SSO2315	36,8	1	NA	NA	NA
	flaI	6	NA	NA	NA	NA	NA	NA	NA	NA	SSO2316	133,7	1	NA	NA	NA
	flaH	1	NA	NA	NA	NA	NA	NA	NA	NA	SSO2318	134,3	1	NA	NA	NA
	flaF	1	NA	NA	NA	NA	NA	NA	NA	NA	SSO2319	147,2	1	NA	NA	NA
V-type ATPase	atpF	1	NA	NA	NA	NA	NA	NA	NA	NA	SSO0560	494,7	1	NA	NA	NA
	atpE	1	NA	NA	NA	NA	NA	NA	NA	NA	SSO0561	992,8	2	NA	NA	NA
	atpA	3	NA	NA	NA	NA	NA	NA	NA	NA	SSO0563	1169	3	NA	NA	NA
	atpB	3	NA	NA	NA	NA	NA	NA	NA	NA	SSO0564	1305	3	NA	NA	NA
	atpD	1	NA	NA	NA	NA	NA	NA	NA	NA	SSO0566	649,7	3	NA	NA	NA
Type I-A, I-B, I-C Cascade	cas6	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	cas8	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	cas7	~4-8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	cas5	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Type I-F Cascade	csy1	1	NA	NA	NA	PA14_33330	1036	2	NA	NA	NA	NA	NA	NA	NA	NA
	csy2	1	NA	NA	NA	PA14_33320	955,3	2	NA	NA	NA	NA	NA	NA	NA	NA
	csy3	6	NA	NA	NA	PA14_33310	1314,9	3	NA	NA	NA	NA	NA	NA	NA	NA
	cas6f	1	NA	NA	NA	PA14_33300	1398,8	3	NA	NA	NA	NA	NA	NA	NA	NA
FtsZ system	mpl	1	b0091	88,7	1	PA14_57330	369,1	1	NA	NA	NA	NA	NA	NA	NA	NA
	ddlA	1	b0092	96,9	1	PA14_57320	292,5	1	NA	NA	NA	NA	NA	NA	NA	NA
	ftsQ	1	b0093	124,3	1	PA14_57300	216,8	1	NA	NA	NA	NA	NA	NA	NA	NA
	ftsA	1	b0094	172,4	1	PA14_57290	524,7	1	NA	NA	NA	NA	NA	NA	NA	NA
	ftsZ	>>10	b0095	325,7	2	PA14_57275	1562,5	2	NA	NA	NA	NA	NA	NA	NA	NA
	lpxC	1	b0096	752,4	2	PA14_57260	1540,3	2	NA	NA	NA	NA	NA	NA	NA	NA

Complex	Gene information		Transcription data												
			E. coli			*P. aerugionosa*			*L. monocytogenes*		*S. Solfataricus*			Additional data		
	Profile ID	Stoichi-ometry	Locus	RPKM	# TU	Locus	RPKM	# TU	Locus	# TU	Locus	RPKM	# TU	Locus	RPKM	# TU
Type II secretion system	pulC	1	b3324	55,9	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	pulD	12	b3325	26	1	NA	NA	NA	NA	NA	NA	NA	NA	PA14_55450	3,8	1
	pulF	1	b3327	18,4	1	PA14_24010	140,3	1	NA	NA	NA	NA	NA	PA14_55430	2,7	1
	pulG	1	b3330	20,1	1	PA14_24050	135,3	2	NA	NA	NA	NA	NA	PA14_55490	6,8	1
	pulL	1	b3333	44	1	PA14_24080	191,5	3	NA	NA	NA	NA	NA	NA	NA	1
	pulM	1	NA	NA	NA	PA14_24100	164,9	3	NA	NA	NA	NA	NA	PA14_55460	3,5	1
	pulJ	1	b3331	24,6	1	PA14_24060	189,8	2	NA	NA	NA	NA	NA	NA	NA	1
	pulK	1	b3332	20,7	1	PA14_24070	114,3	2	NA	NA	NA	NA	NA	PA14_55480	1,3	1
TAT secretion system	tatA	20-40	b3836	674,9	1	PA14_66960	333,4	1	lmo0362	1	SSO0482	2234	1	SSO11972	1619	1
	tatB	1	b3838	175,4	1	NA	NA	NA	NA	NA	SSO0483	663,6	1	NA	NA	NA
	tatC	1	b3839	71,7	1	PA14_66980	72,3	1	lmo0361	1	SSO0484	65,7	1	SSO3108	311,7	2

Table S2. Translation characteristics of selected operons with known stoichiometry.
Subunit stoichiometry values are based on experimental analyzes. Highlighted in light grey are abundant subunits (stoichiometry column), as well as factors potentially correlating with elevated production: ΔG, ΔF, ΔCAI, ΔCo, Δ E and ribosome profiling values above thresholds (see text). Highlighted in dark grey are apparent false positives (codon bias, RNA folding) as discussed in text. Abbreviations: CAI, codon adaptation index.

	Gene information				Codon bias						Stability mRNA folding -20..20		Stability mRNA folding -4..37		Shine-Dalgarno		2nd Codon		Ribosome profiling	
Complex	Profile ID	Annotation	Stoichiometry	Nr of analyzed genomes	Δ % optimal codons	Z-score Δ % optimal codons	Δ CAI	Z-score Δ CAI	Δ % codon co-occurrence	Z-score Δ % codon co-occurrence	Δ mRNA folding energy	z-score ΔmRNA folding energy	Δ mRNA folding energy	z-score ΔmRNA folding energy	delta	z-score	delta	z-score	*E.coli*	*B. subtilis*
Archaeal flagellum	flaJ	Archaeal flagella assembly protein J	2	46	0,00	-0,39	-0,01	-2,11	0,05232	1,32	-0,63	-1,86	-0,55	-1,64	0,732	2,15	-0,0781	-2,68	NA	NA
	flaI	Type IV secretory pathway, VirB11 component	6	45	0,04	8,75	0,03	6,29	0,17744	3,29	-0,34	-0,88	-0,28	-0,59	-0,127	-0,45	-0,0175	-0,55	NA	NA
	flaH	Predicted ATPase involved in biogenesis of archaeal flagella	1	45	0,01	1,50	0,00	-0,45	0,04213	0,88	0,62	1,68	1,46	4,33	-0,195	-0,7	-0,0281	-1,15	NA	NA
	flaF	Putative archaeal flagellar protein F	1	44	-0,02	-2,22	-0,02	-1,79	-0,19255	-1,83	-1,32	-3,66	-2,20	-5,86	-0,607	-1,36	-0,1381	-5,98	NA	NA
V-type ATPase	atpF	Archaeal/vacuolar-type H+-ATPase subunit F	1	57	-0,03	-4,28	-0,03	-5,83	-0,34241	-4,06	-0,254	-1,08	-1,32	-4,00	-0,797	-3,12	0,1604	5,52	NA	NA
	atpE	Archaeal/vacuolar-type H+-ATPase subunit E	1	60	0,00	0,88	-0,01	-1,31	-0,16996	-2,36	-0,044	-0,14	0,01	0,02	1,02	3,22	-0,0096	-0,55	NA	NA
	atpA	Archaeal/vacuolar-type H+-ATPase subunit A	3	60	0,02	6,13	0,03	8,32	0,22369	4,11	1,561	6,22	0,45	1,46	0,236	0,91	0,0072	0,34	NA	NA
	atpB	Archaeal/vacuolar-type H+-ATPase subunit B	3	59	0,02	6,39	0,03	8,66	0,30682	4,53	-0,169	-0,51	1,29	4,00	-0,603	-2,77	0,1017	3,08	NA	NA
	atpD	Archaeal/vacuolar-type H+-ATPase subunit D	1	50	0,01	2,06	0,00	-0,82	0,12314	1,47	-0,303	-0,88	0,88	2,58	0,662	2,47	-0,1016	-4,16	NA	NA
Type I-A, I-B, I-C Cascade	cas6	CRISPR/Cas system-associated RAMP superfamily protein Cas6	1	51	-0,01	-1,38	0,00	-0,85	-0,02086	-0,35	0,356	0,78	0,35	1,03	-0,296	-0,78	0,0499	1,95	NA	NA
	cas8	CRISPR/Cas system-associated RAMP superfamily protein Cas7	1	82	0,00	-0,73	0,00	-0,22	-0,11232	-1,64	-1,377	-3,65	-0,22	-0,87	-2,098	-6,48	-0,0069	-0,32	NA	NA
	cas7	CRISPR/Cas system-associated RAMP superfamily protein Cas7	~4-8	68	0,02	5,47	0,04	6,65	0,19762	2,36	1,757	6,78	1,70	6,21	-0,214	-0,72	0,0283	1,23	NA	NA
	cas5	CRISPR/Cas system-associated RAMP superfamily protein Cas5	1	75	0,00	-0,45	0,00	-0,33	0,00927	0,13	-0,417	-1,11	-0,63	-2,34	-0,321	-1,1	0,023	0,8	NA	NA

	Gene information				Codon bias						Stability mRNA folding -20..20		Stability mRNA folding -4..37		Shine-Dalgarno		2nd Codon		Ribosome profiling	
Complex	Profile ID	Annotation	Stoichiometry	Nr of analyzed genomes	Δ % optimal codons	Z-score Δ % optimal codons	Δ CAI	Z-score Δ CAI	Δ % codon co-occurrence	Z-score Δ % codon co-occurrence	Δ mRNA folding energy	z-score ΔmRNA folding energy	Δ mRNA folding energy	z-score ΔmRNA folding energy	delta	z-score	delta	z-score	*E.coli*	*B. subtilis*
Type I-F Cascade	csy1	CRISPR/Cas system-associated protein Csy1	1	16	0,00	-0,23	-0,01	-1,13	-0,25182	-1,7	-0,25	-0,62	-0,12	-0,30	0,175	0,27	-0,017	-0,32	NA	NA
	csy2	CRISPR/Cas system-associated RAMP superfamily protein Csy2	1	15	0,00	0,26	0,00	0,74	-0,04878	-0,8	-0,46	-0,86	-0,11	-0,20	-1,98	-3,06	-0,0113	-0,22	NA	NA
	csy3	CRISPR/Cas system-associated RAMP superfamily protein Csy3	6	16	0,03	5,59	0,06	7,91	0,11174	0,78	2,00	4,58	1,51	4,90	0,481	1,04	-0,122	-2,28	NA	NA
	cas6f	CRISPR/Cas system-associated RAMP superfamily protein Cas6f	1	17	-0,02	-3,90	-0,03	-5,96	0,16868	0,85	-0,60	-1,06	-1,59	-2,30	-0,673	-1,3	-0,0444	-1,09	NA	NA
FtsZ system	mpl	UDP-N-acetylmuramate-alanine ligase	1	176	0,01	3,68	0,01	5,47	0,28375	6,8	0,234	1,09	0,64	3,33	0,074	0,41	0,1486	8,74	1918	NA
	ddlA	D-alanine-D-alanine ligase and related ATP-grasp enzymes	1	146	-0,02	-7,96	-0,03	-7,06	-0,25895	-4,51	-0,304	-1,3	-1,12	-4,92	-0,272	-1,65	0,0454	2,58	2787	NA
	ftsQ	Cell division septal protein	1	191	-0,04	-14,41	-0,05	-16,41	-0,36216	-7,75	-2,434	-11,07	-1,91	-8,78	0,131	0,74	-0,0583	-3,51	288	NA
	ftsA	Actin-like ATPase involved in cell division	1	188	0,00	-1,65	-0,01	-2,27	0,04191	1,09	0,625	3,21	0,04	0,25	0,037	0,21	-0,0151	-1,05	941	NA
	ftsZ	Cell division GTPase	>>10	189	0,05	19,86	0,06	19,86	0,38305	7,95	1,251	7,62	1,23	6,61	-0,742	-4,43	-0,1097	-8,09	5561	NA
	lpxC	UDP-3-O-acyl-N-acetylglucosamine deacetylase	1	114	0,01	2,64	0,01	3,11	-0,20383	-2,81	1,001	3,96	1,54	7,27	1,185	5,09	0,0169	0,94	3517	NA

	Gene information				Codon bias						Stability mRNA folding -20..20		Stability mRNA folding -4..37		Shine-Dalgarno		2nd Codon		Ribosome profiling	
Complex	Profile ID	Annotation	Stoichiometry	Nr of analyzed genomes	Δ % optimal codons	Z-score Δ % optimal codons	Δ CAI	Z-score Δ CAI	Δ % codon co-occurrence	Z-score Δ % codon co-occurrence	Δ mRNA folding energy	z-score ΔmRNA folding energy	Δ mRNA folding energy	z-score ΔmRNA folding energy	delta	z-score	delta	z-score	*E.coli*	*B. subtilis*
Type II secretion system	pulC	Type II secretory pathway; component PulC	1	37	-0,01	-3,27	-0,02	-3,87	-0,46266	-3,27	0,558	1,39	0,74	1,20	0,573	1,48	-0,0986	-2,62	NA	NA
	pulD	Type II secretory pathway; component PulD	12	80	0,02	6,69	0,03	8,88	0,3402	5,08	1,695	5,34	1,78	5,13	0,803	3,92	0,0613	2,39	NA	NA
	pulF	Type II secretory pathway; component PulF	1	88	0,00	0,78	0,01	1,54	0,46083	5,22	-0,344	-1,13	-0,69	-2,04	0,313	1,24	-0,2129	-10,09	NA	NA
	pulG	Type II secretory pathway; pseudopilin PulG	1	215	0,00	1,32	0,00	1,04	-0,0738	-1,29	0,304	1,26	0,01	0,03	-0,099	-0,69	0,0318	1,94	NA	NA
	pulL	Type II secretory pathway; component PulL	1	32	0,00	0,12	-0,01	-1,36	-0,08916	-0,87	0,527	0,94	0,48	0,92	-0,173	-0,39	0,0085	0,2	NA	NA
	pulM	Type II secretory pathway; component PulM	1	47	-0,01	-1,56	-0,01	-1,35	-0,04109	-0,45	0,297	0,73	0,47	1,09	-1,006	-3,31	0,1374	4,21	NA	NA
	pulJ	Type II secretory pathway; component PulJ	1	50	-0,01	-2,86	-0,02	-4,50	-0,00012	0	-1,365	-2,95	-0,22	-0,52	0,117	0,37	0,0468	1,37	NA	NA
	pulK	Type II secretory pathway; component PulK	1	77	-0,02	-4,84	-0,02	-4,76	-0,3889	-5,35	-1,997	-5,32	-1,78	-4,43	-0,583	-2,18	0,0205	0,82	NA	NA
TAT secretion system	tatA	Sec-independent protein secretion pathway component TatA	20-40	209	0,03	10,42	0,03	10,60	0,07566	1,5	0,57	3,85	0,45	3,33	-0,791	-5,67	-0,0689	-7,33	5780	5350/65
	tatB	Sec-independent protein secretion pathway component TatB	1	137	-0,01	-4,06	-0,02	-8,02	-0,27648	-5,46	-0,60	-3,51	-0,61	-3,23	0,195	1,28	-0,0703	-6,59	755	NA
	tatC	Sec-independent protein secretion pathway component TatC	1	231	-0,02	-7,47	-0,01	-5,45	0,09552	2,02	-0,16	-1,05	-0,04	-0,28	0,6	4,83	0,104	10,31	580	278/3

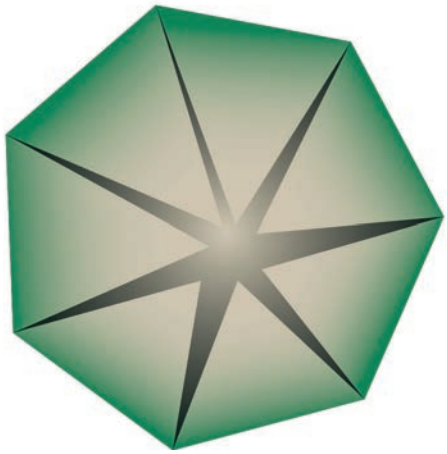
Chapter 7

Summary and general discussion

Partly adapted from:
David Prangishvili and Tessa E F Quax

'Exceptional virion release mechanism: one more surprise from Archaeal viruses'

Current Opinion in Microbiology. 2011 Jun. 14(3):315-20.



Summary

The work presented in this thesis provides novel insights in several aspects of the molecular biology of archaea, bacteria and their viruses.

Three fundamentally different groups of viruses are associated with the three domains of life. Archaeal viruses are characterized by a particularly high morphological and genetic diversity. Some archaeal viruses, such as *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2), have quite remarkable infection cycles. As described in **Chapter 1**, infection with SIRV2 results in the formation of large virus associated pyramids (VAPs) on the host cell surface. The structures open in the final step of the infection cycle, creating large apertures to release the rod-shaped viruses that have matured in the cytoplasm. This virus release mechanism is unique and does not resemble egress mechanisms of bacterial and eukaryotic viruses. Analysis of the protein composition of SIRV2 infected cells, as outlined in **Chapter 2**, revealed the strong accumulation of the virus encoded protein PVAP in membranes after infection, suggesting involvement in VAP formation. The VAPs can be isolated as discrete particles, as demonstrated in **Chapter 3**. Electron microscopic survey of these particles showed that they are baseless pyramids with a heptagonal perimeter. This geometry is exceptional and especially the sevenfold symmetry is very rare in nature (20S proteasome, myosin). The structures can have various sizes, probably reflecting different developmental stages. This suggests that they grow by the gradual expansion of the triangular facets. Analysis of the protein composition of the structures revealed the exclusive presence of PVAP and anti-bodies raised against this protein labeled specifically the VAPs on thin sections of infected cells as observed in electron microscopy. PVAP is sufficient for VAP formation, which was demonstrated by expression of the protein and successful assembly of pyramidal structures, in the archaeon *S. acidocaldarius* and the bacteria *Escherichia coli*. Further analysis of PVAP truncation mutants as outlined in **Chapter 4**, showed that besides the 10 C-terminal amino acids, all domains of the protein are essential for VAP formation. PVAP can form oligomers of several sizes, including those of a heptamer, which probably act as nucleation points for VAP formation on the cell membrane. Analysis of the truncation mutants indicated that both the C and N terminal domain are important for interaction between monomers. Detailed observation with whole cell cryo-electron tomography of VAPs formed in the natural and heterologous system, revealed the presence of two layers in the structure. The outer one is continuous with the cell membrane. The inner layer facing the cytoplasm, presumably represents a protein sheet formed by tight interactions between the C-terminal domain of PVAP connected with a short linker region to the membrane. The sheets are slightly bended, giving the complete structure the appearance of a teepee. At the junction of two triangular sheets, the structure is perforated, creating predetermined breaking points. Furthermore, in this chapter data is presented which underlines the unique nature of this protein, since it is able to form VAPs successfully in

archaeal, bacterial and eukaryotic membranes, which all fundamentally differ in protein and lipid composition. In case of expression in *Saccharomyces cerevisiae*, VAPs are formed on all membranes, including those of mitochondria, suggesting that the protein inserts spontaneously in membranes. Thus, PVAP serves as a universal membrane remodeling system, which might be exploited for biotechnological purposes, such as the development into a universal system for the controlled opening of ~100 nm apertures in any lipid bilayer.

Production of VAPs is one of the dramatic consequences that SIRV2 infection has on the host cell. Whole transcriptome sequencing allowed determination of a global map of virus and host gene expression during the infection cycle, which is presented in **Chapter 5**. Directly after infection, transcription of viral genes starts simultaneously from both genome termini. All possible protein interactions between all SIRV2 proteins were assayed with yeast two-hybrid and these results were used to advance current knowledge on SIRV2 genes functions, of which the majority is still unidentified. The host cells respond to viral infection by adapting expression of more than 30% of its genes. Genes involved in cell division are down regulated, while those playing a role in anti-viral defense are activated. Specifically, for the first time massive activation of toxin anti-toxin and CRISPR-Cas systems is observed in an archaeal system. The different degree of expression and activation of the various systems highlights the specialized functions they perform.

The CRISPR-associated multi-subunit ribonucleoprotein complexes that are crucial for the CRISPR mediated anti-viral defense, generally have an uneven stoichiometry, i.e. the 4-6 different protein subunits are present in different quantities. Just as most functionally related bacterial and archaeal genes, the *cas* genes are clustered in operons, which allow for co-expression (as has indeed been observed in the transcriptome analysis described in Chapter 5). This is advantageous when equal amounts of gene products are required, such as is the case for protein complexes with even stoichiometry. However, a substantial number of important protein complexes contain uneven stoichiometry. Employing comparative genomics, in **Chapter 6**, it is shown that differential translation is a key determinant of modulated expression of genes clustered in operons and that codon bias generally is the best *in silico* indicator of unequal protein production. In addition, analysis of protein production from genes with synonymous mutations from synthetic operons, provides evidence that initiation of translation can occur at intercistronic sites. The widespread occurrence of modulation of translation efficiency, suggests that this is a universal mode of control in bacteria and archaea that allows for differential production of operon-encoded proteins.

General Discussion

Archaeal virion egress

VAP assembly

In the first three chapters of this thesis, the molecular mechanism for egress of the archaeal virus SIRV2 is described, which relies on the formation of Virus Associated Pyramids (VAPs) (Fig. 1). The protein constituent of the VAPs, SIRV2-PVAP, is predicted to form a protein with four α-helices, of which one resides in the membrane as outlined in Chapter 4. The other three C-terminal α-helices are thought to undergo intimate interactions resulting in a protein sheet just below the cell membrane, which is visible as the inner VAP layer in cryo tomograms. Tango software predicts that the PVAP hydrophobic trans membrane segment can form β-sheets at the expense of α-helices (230). Thus, there is a possibility that PVAP acts as a prion like protein with amyloid character, converting from α-helix to β-sheet conformation upon multimerization. Since α-helical conformations form more rapidly, but β-sheets are typically more stable, this strategy might allow for fast and efficient formation of the unusual VAP structure.

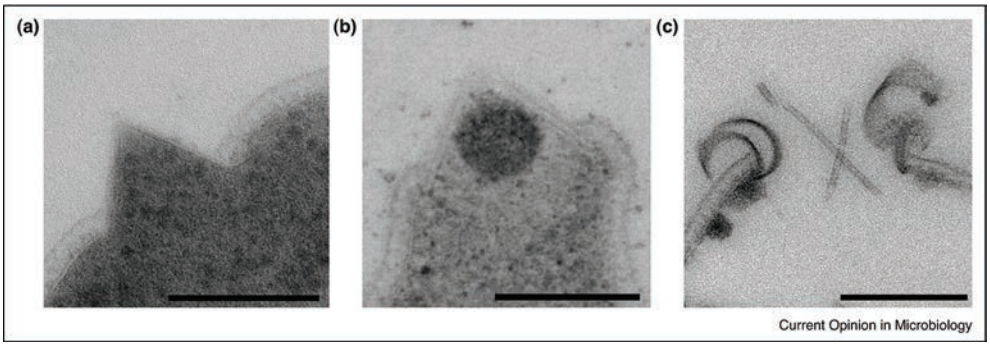


Figure 1. VAPs are part of a unique virus release mechanism in archaea. Thin sections through VAPs in (**A**) and (**B**) closed and (**C**) open conformation on cell surface of *S. islandicus* infected with SIRV2. bars, 100 nm.

A crystal structure of PVAP would be very valuable for future studies on the assembly mechanism of VAPs. Since the PVAP trans membrane segment is essential for VAP formation, it would be desirable to obtain a structure of the intact protein. This is a challenge due to the hydrophobicity of the membrane segment; therefore 2-D crystallography might be an effective alternative strategy.

Flexibility of VAP based virion egress system

The tight interaction of PVAP monomers results in the exclusion of Surface (S)-layer proteins from the site of VAP assembly, since they are anchored in the cell membrane. This mechanistic approach to pinch the cell wall might provide a strategy which facilitates extension of the virus host range, because this mechanism is independent of enzyme specificity in contrast to the lysines coded by bacterial viruses. However, the VAP egress mechanism does not allow virions to cross the peptidoglycan cell wall of bacteria, as this requires enzymatic degradation. Indeed heterologous expression of PVAP in *E.coli* never resulted in punctured cell walls by VAPs.

Biotechnological applications

Results presented in Chapter 3 and 4 of this thesis, demonstrate that SIRV2-PVAP has the unique property to self-assemble into pyramidal structures of sevenfold symmetry in archaea, bacterial and eukaryotic cells. Therefore PVAP acts as a universal membrane remodelling system and is expected to form VAPs in all biological membranes.

These remarkable protein properties could be exploited in a biotechnological setting. First of all, VAPs could be used in targeted drug delivery, when the isolated particles are conjugated with specific compounds destined for transport. This method of drug delivery is already being explored for the SIRV2 virions and it was shown that specific compounds could be conjugated to the virion particles (28). Alternatively, liposomes exposing VAPs on their surface could be used for the same purpose. Liposomes are at present being used for targeted delivery of drugs (231). Specific opening of liposomes at the site of drug delivery, would provide several advantages over the currently used methods of membrane fusion. Alternatively, PVAP might be exploited as a universal cell lysis system or as an *in vitro* gate lock system which opens lipid layers under specific conditions (international patent: PCT/EP2012/050902).

VAP opening

Understanding the mechanism of VAP opening is important from a fundamental perspective, but also for the realization of the above mentioned biotechnological applications of the PVAP protein. Although the geometry and size of the VAPs in the heterologous systems were similar to those to SIRV2 induced VAPs, opening was never observed in these systems, except in *S. acidocaldarius* after incubation of PVAP expressing cells for more than 72 hours (Quax et al, unpublished). Thus, at least in Sulfolobales, PVAP seems sufficient for VAP opening in certain situations, which might be linked to the growth conditions of the cells. It has been suggested that the ESCRT-III like archaeal cell division machinery is involved in the opening of STIV induced VAPs (232). A model is proposed in which ESCRT-III like proteins polymerize and strip the STIV induced VAP from the inner membrane layer (232). Since the SIRV2 induced VAPs were clearly shown to contain a protein layer below the outer membrane embedded layer (Chapter 4) and ESCRT-III like proteins

were specifically down regulated after SIRV2 infection (Chapter 5), it may be that, at least in SIRV2 infected cells, opening of VAPs relies on a different mechanism. To draw a firm conclusion about similarities and differences, the involvement of ESCRT in VAP opening should be tested in the SIRV2 system as well.

Egress mechanisms of archaeal viruses

The sole building block of the VAPs has few homologues in archaeal viruses (105). Homologues are present in genomes of nviruses that are member of the family Rudiviridae, to which SIRV2 also belongs, namely SIRV1 (31) and *Stygiolobus* rod-shaped virus (SRV) (58), as well in the icosahedral virus STIV1 unrelated to the rudiviruses (39). The SIRV2_PVAP homologue of STIV1 was also experimentally shown to be the constituent of the VAPs induced in *S. solfataricus* cells infected by STIV (103). PVAP encoded by STIV1 and SIRV2 share only ~50% sequence identity on the protein level, but it was demonstrated that they are functional homologues by the expression of chimeras of the two proteins which results in VAP formation (233). However, the two PVAP proteins from STIV1 and SIRV2 are not interchangeable in the context of the STIV replication cycle, since expression of the chimeras did not result in virion production (233).

It is likely that the VAP-based virion release system is characteristic to all viruses carrying the gene for SIRV2-PVAP homologue. Remarkably, STIV2, a close relative of STIV1 does not encode a homologue of SIRV2-PVAP (234), nor does *Acidianus* rod-shaped virus 1, ARV1 (55), a member of the *Rudiviridae*. Apparently, the morphogenetic and egress systems are evolving independently in archaeal viruses, and the VAP-based virion release mechanism is not universal for them. To gain insight into the diversity and evolution of virion egress mechanisms in general it is essential to obtain information on the mechanism of virion egress of other archaeal viruses. Interesting candidates for such a study would be ARV1 and STIV2, since these two viruses are the only members of their viral family that are not encoding PVAP and should thus rely on a different egress mechanism. Moreover, it can be expected that enveloped viruses, such as *Acidianus* Filamentous Virus 1, acquire their lipid envelope during a budding based egress mechanism.

Comparison with virion release mechanisms of bacterioviruses

The vast majority of bacterial dsDNA viruses under laboratory conditions escape the infected cell by disrupting the latter with the aid of the holin-endolysin system, named after the two proteins essential in this process (64). The endolysin is a muralytic enzyme capable of degrading the peptidoglycan cell wall surrounding the bacterial cell, and the holin is a small virus-encoded protein that forms large non-specific lesions in the cytoplasmic membrane of the infected cell (64). For many viruses of gram-negative bacteria the system also encompasses the proteins Rz and Rz1, which are thought to form a complex connecting the cytoplasmic and the outer membranes (37, 64).

In contrast, bacterial ssDNA and ssRNA viruses encode only a single protein sufficient for cell lysis. This type of protein was shown to inhibit cell wall synthesis, eventually leading to cell disruption (65).

Not all bacterioviruses are lytic. Some, like members of the family *Inoviridae*, exit without killing or lysing their bacterial host. They establish persistent infection in which the host continues to grow and divide while producing and releasing virions (235). The ssDNA virus genome is extruded through the cytoplasmic membrane, where it is coated with capsid proteins. Following the assembly, virions are transported through the outer membrane of gram-negative bacteria through gated channels, which can be either virus (for M13-like viruses) or host (for CTXϕ-like viruses) encoded (236). In the case of the bacteriovirus f1, these channels are formed by multimers of the viral protein pIV, homologous to the protein EpsD, an outer membrane component of type II secretion system (237). The protein is able to self-assemble and the complex can be isolated (238, 239).

The VAP-based virion release mechanism of archaeal viruses significantly differs in all molecular characteristics from the three above-described egress systems of bacterioviruses.

A new class of virus structures

The geometry of the VAP is unrecorded in the living world. Moreover, structures with seven fold symmetry are rare in nature and the two known examples comprise the archaeal 20S proteasome (42) and the scallop muscle myosin filaments (240).

Not only is the geometry of VAPs special, exceptional is also the fact that this stable independent structure is encoded by the virus. In addition to the capsid, the VAP is an autonomous structure encoded by the virus SIRV2 (Fig. 2). The constituents of both structures are able to self-assemble in heterologous systems. The major coat protein of the rudiviruses self-assembles into long filaments with the same diameter as the capsid (58) and the VAP protein PVAP into pyramids with seven fold symmetry. The functions of the two viral structures are completely different though (Fig 3). The capsid serves for the packaging and protection of the viral genome, while the VAP is employed by the virus to escape the cell (Fig. 3). The discovery of the remarkable VAP has raised awareness of the existence of a class of non-capsid virus-encoded autonomous structures employed for virion release. It is proposed to name them "virodomes". VAPs can represent only one type of archaeal "virodomes" and many more could be discovered in the course of future studies on archaeal virus-host interactions. In bacterial systems, the two-dimensional holin rafts can possibly be considered as "virodomes".

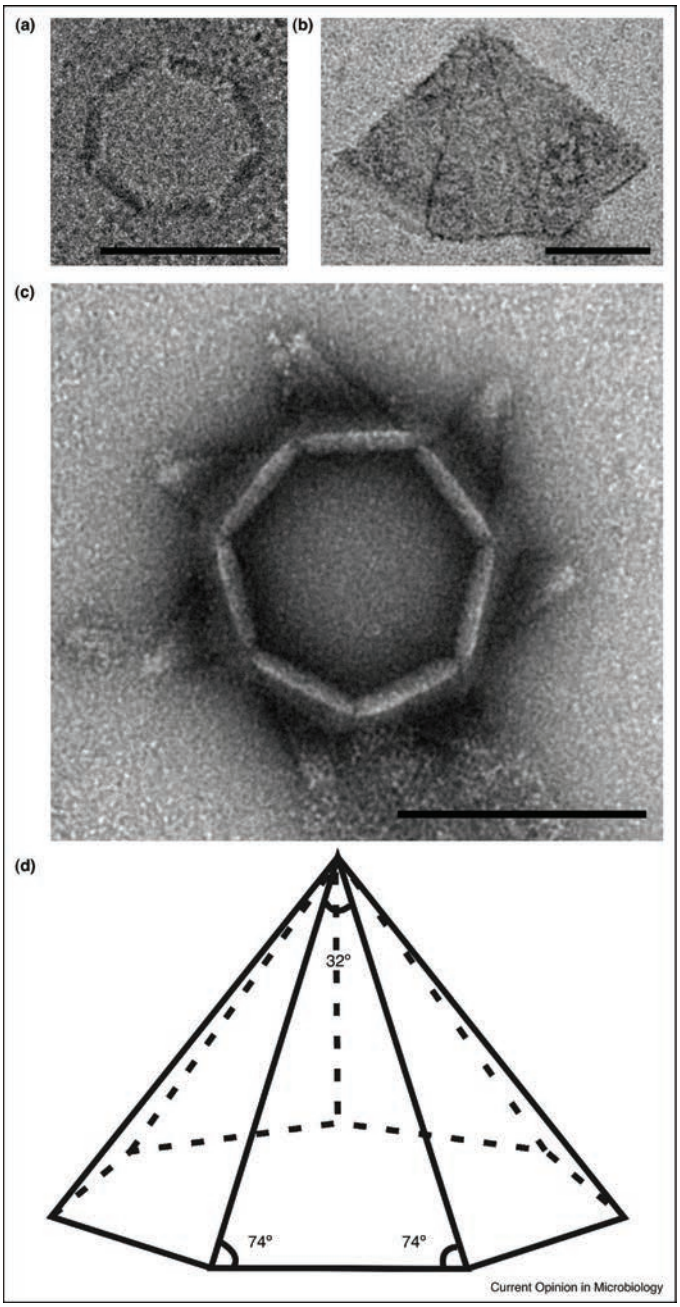


Figure 2. VAPs are independent particles. Negative contrast electron micrographs of isolated VAPs. **(a)** Top view and **(b)** side view of intact VAPs. **(c)** Top view of a VAP in the open conformation. **(d)** Schematic representation of the native VAP structure based on measurements of ~150 images of isolated VAPs. bars, 100 nm.

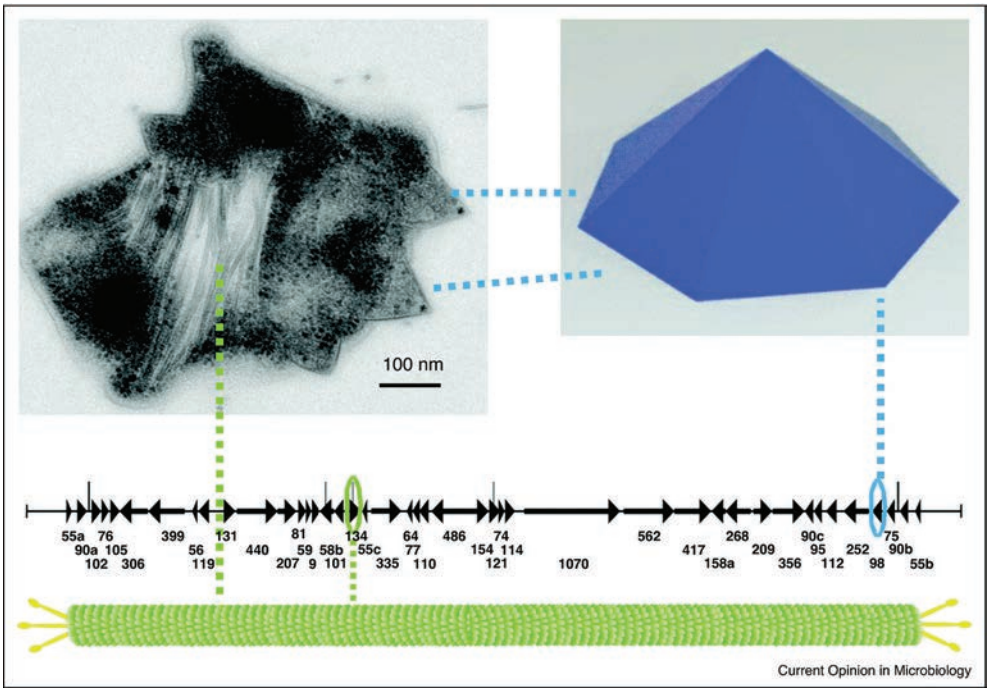


Figure 3. VAPs are non-capsid virus-encoded autonomous structures. Schematic representation of the SIRV2 genome. Highlighted are the two genes encoding virus structures. ORF134 encodes the major capsid protein of the virion protein, and ORF98 encodes the VAP protein.

Archaeal anti-viral defense

Model archaeal virus

Studies on the VAP dependent virion egress mechanism have demonstrated the power of SIRV2 as a model to explore archaeal virus-host interactions. Since the discovery of the first archaeal virus in the 1980', the number of described archaeal viral species is steadily increasing and new morphotypes keep being discovered (129). This increasing number of isolated archaeal viruses requires standardization of studies on them and the development of suitable models to characterize and compare different aspects of their infection cycle in analogy with the T-series of bacterial viruses. *Sulfolobus* Spindle-shaped Virus (SSV) (131), STIV1 (39) and SIRV2 (31) have already emerged as models to study archaeal viral biology, because of the high virus titers that can be obtained and the clear effect on the host cell of infection or induction with these. Using these model systems several aspects of archaeal viral biology have been unraveled, which demonstrates both the unique nature of archaeal viruses as their relation with viruses infecting members of the other two domains (40, 83, 102, 240, 241).

SIRV2 replication

The whole transcriptome analysis and yeast two hybrid assays described in Chapter 5 aid in the prediction of novel SIRV2 gene functions. Examples are the two identical genes located on the SIRV2 genome termini, which are encoding the protein SIRV2_P83. This protein is implied in SIRV2 genome replication or transcription, based on the expression profile and its binding to the viral encoded Holliday Junction Resolvase (Chapter 5). The biochemical characterization of this protein will be crucial for a better understanding of its possible role and the replication mechanism of SIRV2. The predicted DNA binding activity needs to be tested, just as possible nicking activity. Nicking activity would be important for initiation of replication of the linear covalently closed SIRV2 genome (242). SIRV2 is not coding its own DNA polymerase and other proteins are involved in replication such as cdc6 (cell division cycle) , MCM (mini chromosome maintenance) proteins and PCNA (proliferating cell nuclear antigen) (15, 137, 243). As a consequence, genome replication of SIRV2 relies on host factors. Although none of these host genes appears to be significantly up regulated after SIRV2 infection, in would be important for the understanding of the P83 function to assess possible interaction with these host replication proteins. A first clue to the possible link between P83 and host replication factors is the reported binding between P83 and PCNA (Gardner, Prangishvili et al, personal communication).

Interplay of anti-viral defense systems in archaea

Infection with SIRV2 results in wide spread activation of anti-viral defense systems (Chapter 5). Specific activation of toxin-antitoxin (TA) modules after viral infection, suggests a role in anti-viral defense. The role of TA based anti-viral defense has not yet been studied in archaea, but the detection of up regulation of an anti-toxin gene during STIV1 infection suggests that it is important to investigate the role of TA modules in archaeal virus defense in more detail (51). In addition it would be interesting to assess the *in vivo* role of the Sulfolobus TA modules during viral challenge. It was suggested that TA systems act in concert with other defense systems, like CRISPR-Cas to provide robustness to the anti-viral response and function as a 'last rescue' when all other systems fail (173). One hour after SIRV2 infection, CRISPR-Cas and TA systems were both highly activated. Therefore a transcriptome study of infected cells at short time intervals closer to the onset of viral infection might provide insight into the temporal sequence in which the two systems are activated. The different transcription and activation of several CRISPR-Cas systems during SIRV2 infection might allow for a tailor-made defense reaction depending on the properties of the foreign genetic element, which invades a cell. This assumption could be analyzed by whole transcriptome sequencing during challenge with different invading viruses and plasmids to gain insight into possible specialization of different defense systems. The richness of anti-viral defense systems encoded on the *S.islandicus* genome, renders the SIRV2-*islandicus* system a very suited model to study the differentiated roles and interplay of these systems *in vivo.*

Deciphering the interplay between them will be essential to provide an answer to the question why the majority of archaeal genomes carry so many different defense systems.

Codon usage as an in silico predictor of protein subunit stoichiometry

CRISPR anti-viral defense systems rely on Cas complexes for successful targeting of foreign nucleic acids. Typically, a Cas complex consist of several proteins present in uneven ratio's encoded on operons, as is the case for the majority of bacterial and archaeal functionally related proteins. Results presented in Chapter 6 of this thesis demonstrate that differential production of Cas proteins, is achieved by differential translation, in order to maintain the stoichiometry of the complex (Fig. 4). Besides the Cas complexes many more protein complexes that play roles in important cellular processes contain uneven stoichiometry. The analysis presented in Chapter 6 indicates that differential translation is a universal mode of control for the production of operon-encoded complexes in prokaryotes (Fig. 5). Codon bias is a suitable *in silico* indicator to predict subunit stoichiometry of protein complexes. Analysis of codon usage in combination with other influential features, like prediction of mRNA structure, might prove a good strategy to obtain information on subunit stoichiometry, in case structural data is lacking. However, due to the complexity of the various factors influencing translation, currently, predictions will still need verification by experimental approaches, either holistic by ribosome profiling (Chapter 6) or specifically by biochemical/structural analysis of the protein complex.

A future challenge in synthetic biology, is the exact determination of all factors influencing translation efficiency in such a way that this can be employed to improve (heterologous) protein expression from a given gene. Currently, 'optimization of codon usage' is employed to enhance gene expression (219, 244). In practice, however, results are often not satisfying, probably because the artificial adaptation of codon usage interferes with other factors which influence protein expression, such as the mRNA folding energy, the presence of Shine Dalgarno-like sequences in the gene and other still unidentified elements (202, 205, 219). To understand how the 'perfect' codon landscape can be constructed to improve gene expression, engineering approaches might be employed to construct and analyze a library of non-coding flanking regions and coding regions with synonymous mutations.

Since viruses are largely dependent of host factors for transcription and translation, generally they share similar codon usage with the cells they infect. It will be interesting to assess if viruses also employ a tuning mechanism relying on differential translation to achieve varying production levels of proteins required in different ratios. Suitable cases to test this hypothesis are the capsid proteins of viruses of which the virion structure has been determined. The recent development of the ribosome density profiling technique offers the exciting possibility to define a global map of all translating ribosomes and might be employed to assess the influence of differential translation in determining production of viral proteins (202, 206).

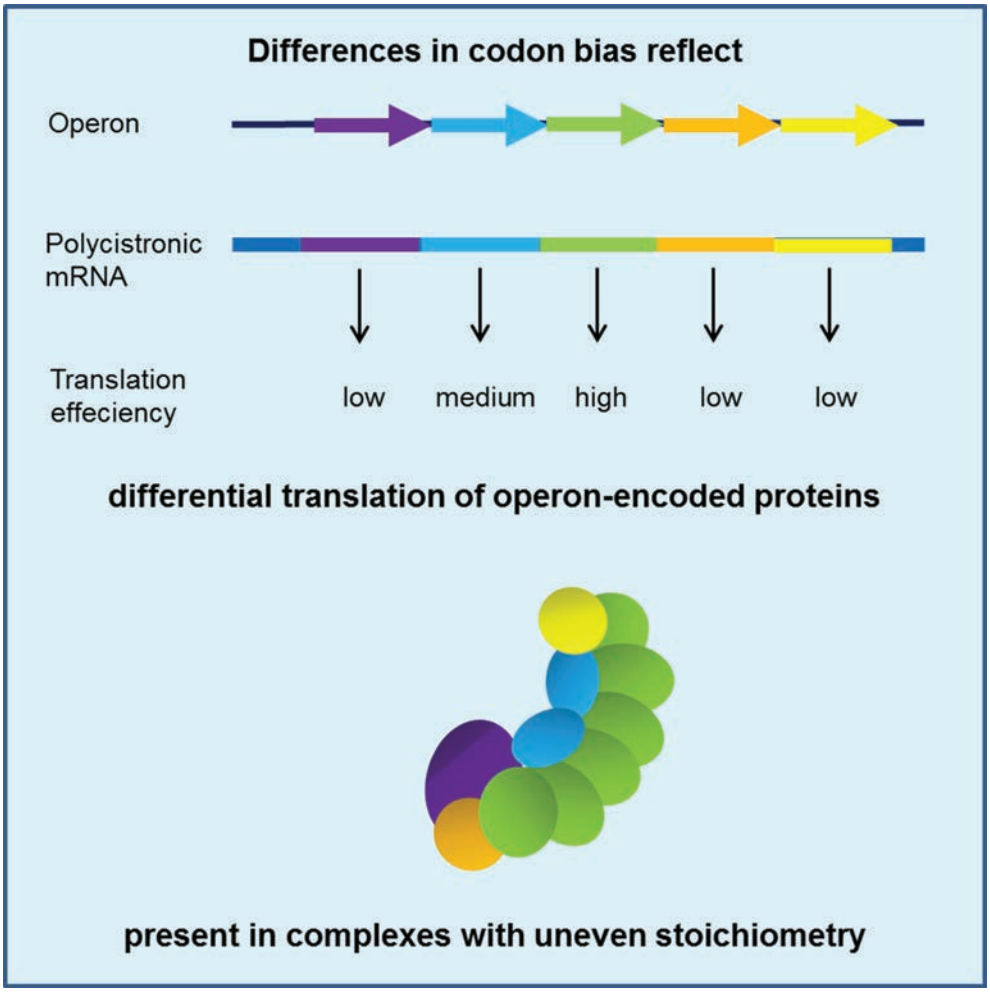


Figure 4. Graphical overview of the process of differential translation of the operon-encoded Cas proteins, which results in a protein complex with uneven stoichiometry that is important for CRISPR-mediated antiviral defense.

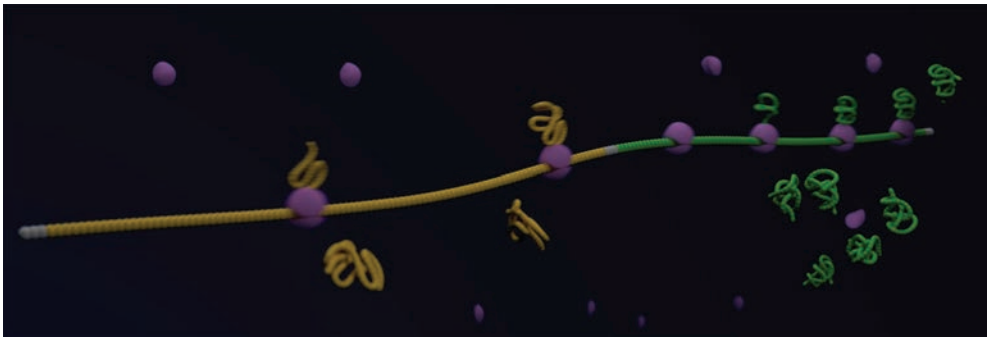


Figure 5. Differential translation of operons is a universal mode of control for the production of operon encoded proteins. Schematic representation of a mRNA containing two cistrons with synonymous mutations resulting in translation with low (yellow) and high (green) efficiency. Ribosomes are depicted in purple.

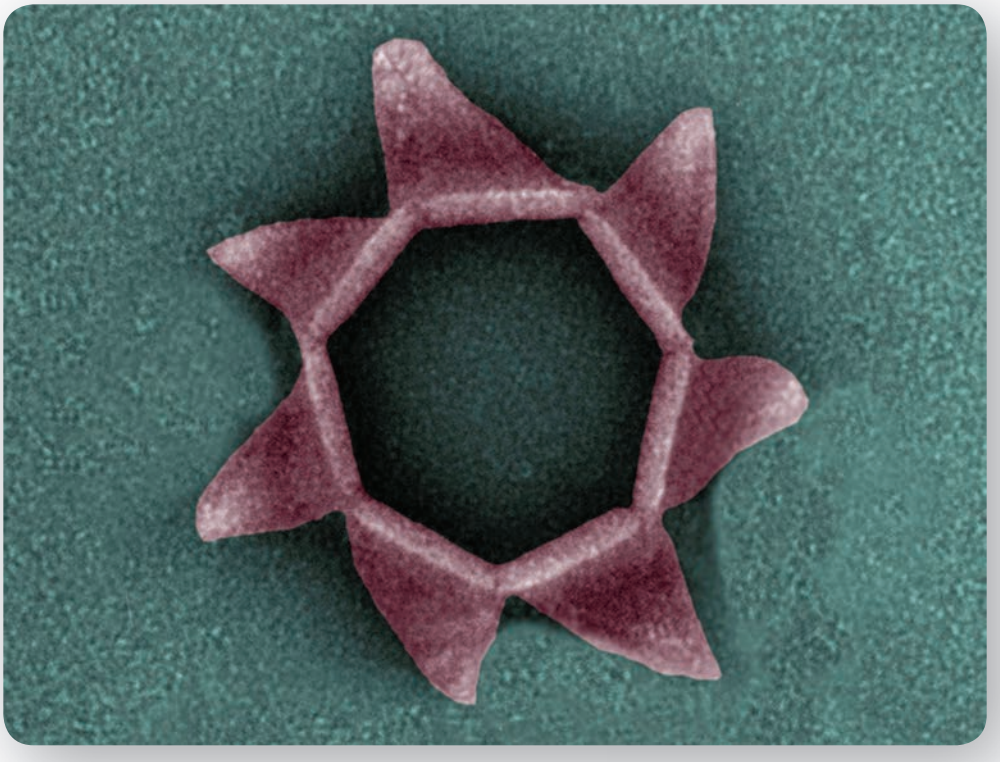
Arms race between archaea and viruses in the natural environment

Activation of *cas* genes as described in Chapter 5 of this thesis, was reported also during infection with the archaeal virus STIV (245). It is noteworthy is that activation of *cas* genes and a severe host response were only observed when a low susceptible host strain was used (10% of cells susceptible to infection) (245). In contrast, when a highly susceptible host strain was challenged with STIV, Cas proteins were not activated (133). Since this susceptible host strain has not been fully sequenced, it cannot be excluded that mutations in CRISPR-*cas* operons have occurred. The *Sulfolobus* host, which is typically used for studies on the STIV infection cycle, has been selected based on high susceptibility to the viral infection, as is the case for other archaeal model viruses like SIRV2 (51, 83). Important in this respect is that the *S. solfataricus* strain, which was selected based on susceptibility to SIRV2 infection, carries a large deletion spanning several CRISPR-*cas* operons (246). Infection of these strains with virus, generally results in an almost completely infected cell population (51, 83). This might not reflect the natural situation, since it would result in rapid extinction of the particular host in an environment where there is constant threat of viral infection. The equilibrium in which host and virus populations exists in the natural environment might reflect the 'stable carrier state' that has been proposed to be the mode in which the majority of archaeal viruses is maintained in host cells (48, 53). Viruses and cells constantly co-evolve. They have to be highly inventive and yet efficient, to maintain themselves and remain one step ahead of the other. The remarkable virion egress structure and the CRISPR-Cas anti-viral defense system both represent inventive and elegant products of the ongoing arms race between archaea and their viruses.

Concluding remarks

The discovery and characterization of the VAP-based virion release mechanism has shown that archaeal viruses, along with exceptional morphological and genomic properties, possess unusual life cycles. The VAP dependent virion egress mechanism is unique for archaeal viruses. On the other hand, viral infection is counteracted using anti-viral defense mechanisms, which are also ubiquitously present in bacteria. The role of differential translation on regulation of subunit stoichiometry of operon encoded Cas complexes was found to be an universal strategy to produce operon-encoded proteins of bacteria and archaea in appropriate quantities. Thus, the results described in this thesis support the notion of the domain-specific nature of the archaeal virosphere and yet also revealed common features shared with the other domains of life.

Appendices



References

1. James J (2004) [Van Leeuwenhoek's discovery of bacteria: a look too far ahead]. *Ned Tijdschr Geneeskd* 148(52):2590-2594.
2. Bos L (1999) Beijerinck's work on tobacco mosaic virus: historical context and legacy. *Philos Trans R Soc Lond B Biol Sci* 354(1383):675-685.
3. F dH (1917) Sur un microbe invisible antagoniste des bacilles dysentériques. *Comptes rendus Acad Sci Paris* 165:373-375.
4. Summers WC (2006) *The Bacteriophages* (Oxford University Press, New York) 2 Ed.
5. Twort FW (1915) An investigation on the nature of the ultra-microscopic viruses. *Lancet* II:1241-1243.
6. Ackermann HW (2007) 5500 Phages examined in the electron microscope. *Archives of virology* 152(2):227-243.
7. Woese CR, Kandler O, & Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* 87(12):4576-4579.
8. Huber R, Huber H, & Stetter KO (2000) Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS microbiology reviews* 24(5):615-623.
9. Valentine DL (2007) Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nature reviews. Microbiology* 5(4):316-323.
10. Orange F, *et al.* (2009) Experimental silicification of the extremophilic Archaea Pyrococcus abyssi and Methanocaldococcus jannaschii: applications in the search for evidence of life in early Earth and extraterrestrial rocks. *Geobiology* 7(4):403-418.
11. Stetter KO (1996) Hyperthermophiles in the history of life. *Ciba Found Symp* 202:1-10; discussion 11-18.
12. Moissl-Eichinger C (2011) Archaea in artificial environments: their presence in global spacecraft clean rooms and impact on planetary protection. *ISME J* 5(2):209-219.
13. Leininger S, *et al.* (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442(7104):806-809.
14. DeLong EF & Pace NR (2001) Environmental diversity of bacteria and archaea. *Syst Biol* 50(4):470-478.
15. Barry ER & Bell SD (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev* 70(4):876-887.
16. Werner F (2007) Structure and function of archaeal RNA polymerases. *Mol Microbiol* 65(6):1395-1404.
17. Brochier-Armanet C, Forterre P, & Gribaldo S (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. *Current opinion in microbiology* 14(3):274-281.
18. Martijn J & Ettema TJ (2013) From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem Soc Trans* 41(1):451-457.
19. Torsvik T & Dundas ID (1974) Bacteriophage of Halobacterium salinarium. *Nature* 248(450):680-681.
20. Sime-Ngando T, *et al.* (2011) Diversity of virus-host systems in hypersaline Lake Retba, Senegal. *Environ Microbiol* 13(8):1956-1972.
21. Rachel R, *et al.* (2002) Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch Virol* 147(12):2419-2429.

22. Pina M, Bize A, Forterre P, & Prangishvili D (2011) The archeoviruses. *FEMS Microbiol Rev* 35(6):1035-1054.
23. Mochizuki T*, et al.* (2012) Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proceedings of the National Academy of Sciences of the United States of America* 109(33):13386-13391.
24. Pietila MK, Roine E, Paulin L, Kalkkinen N, & Bamford DH (2009) An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. *Mol Microbiol* 72(2):307-319.
25. Bolduc B*, et al.* (2012) Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *J Virol* 86(10):5562-5573.
26. Prangishvili D, Garrett RA, & Koonin EV (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res* 117(1):52-67.
27. Prangishvili D, Forterre P, & Garrett RA (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* 4(11):837-848.
28. Steinmetz NF*, et al.* (2008) Site-specific and spatially controlled addressability of a new viral nanobuilding block: Sulfolobus islandicus rod-shaped virus 2. *Adv. Funct. Mater.* 18:2478-3486
29. Forterre P (2012) Darwin's goldmine is still open: variation and selection run the world. *Front Cell Infect Microbiol* 2:106.
30. Forterre P & Prangishvili D (2009) The origin of viruses. *Res Microbiol* 160(7):466-472.
31. Prangishvili D*, et al.* (1999) A novel virus family, the Rudiviridae: Structure, virus-host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. *Genetics* 152(4):1387-1396.
32. Brouns SJ*, et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960-964.
33. Barrangou R*, et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709-1712.
34. Westra ER*, et al.* (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet* 46:311-339.
35. Jore MM*, et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18(5):529-536.
36. Jacob F & Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3:318-356.
37. Krupovic M & Bamford DH (2008) Holin of bacteriophage lambda: structural insights into a membrane lesion. *Mol Microbiol* 69(4):781-783.
38. Martin-Serrano J & Neil SJ (2011) Host factors involved in retroviral budding and release. *Nat Rev Microbiol* 9(7):519-531.
39. Rice G*, et al.* (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proceedings of the National Academy of Sciences of the United States of America* 101(20):7716-7720.
40. Brumfield SK*, et al.* (2009) Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus sulfolobus turreted icosahedral virus. *J Virol* 83(12):5964-5970.
41. Al-Khayat HA, Morris EP, & Squire JM (2009) The 7-stranded structure of relaxed scallop muscle myosin filaments: support for a common head configuration in myosin-regulated muscles. *J Struct Biol* 166(2):183-194.
42. Forster A, Whitby FG, & Hill CP (2003) The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. *EMBO J* 22(17):4356-4364.
43. Mande SC, Mehra V, Bloom BR, & Hol WG (1996) Structure of the heat shock protein chaperonin-10 of Mycobacterium leprae. *Science.* 271(5246):203-207.

44. Song L, *et al.* (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274(5294):1859-1866.

45. Abrahams JP, Leslie AG, Lutter R, & Walker JE (1994) Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370(6491):621-628.

46. Dunkle JA, *et al.* (2011) Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* 332(6032):981-984.

47. Johnson TL, Abendroth J, Hol WG, & Sandkvist M (2006) Type II secretion: from structure to function. *FEMS Microbiol Lett* 255(2):175-186.

48. Prangishvili D, Forterre P, & Garrett RA (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* 4(11):837-848.

49. Prangishvili D BT, Garrett RA (2008) Crenarchaeal viruses: Morphotypes and genomes. *Encyclopedia of Virology.*, ed Mahy BWJ vRM (Elsevier, Oxford), pp pp. 587–595.

50. Janekovic D WS, Holz I, Zillig W, Gierl A, Neumann H (1983) TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaebacterium Thermoproteus tenax. *Molecular and General Genetics* 192(1): 39-45.

51. Ortmann AC, *et al.* (2008) Transcriptome analysis of infection of the archaeon Sulfolobus solfataricus with Sulfolobus turreted icosahedral virus. *J Virol* 82(10):4874-4883.

52. Prangishvili D, *et al.* (2006) Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J Mol Biol* 359(5):1203-1216.

53. Prangishvili D & Garrett RA (2005) Viruses of hyperthermophilic Crenarchaea. *Trends Microbiol* 13(11):535-542.

54. Schleper C, Kubo K, & Zillig W (1992) The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. *Proceedings of the National Academy of Sciences of the United States of America* 89(16):7645-7649.

55. Vestergaard G, *et al.* (2005) A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus Acidianus. *Virology* 336(1):83-92.

56. Contursi P, *et al.* (2006) Characterization of the Sulfolobus host-SSV2 virus interaction. *Extremophiles* 10(6):615-627.

57. Rice G, *et al.* (2001) Viruses from extreme thermal environments. *Proceedings of the National Academy of Sciences of the United States of America* 98(23):13341-13345.

58. Vestergaard G, *et al.* (2008) Stygiolobus rod-shaped virus and the interplay of crenarchaeal rudiviruses with the CRISPR antiviral system. *J Bacteriol* 190(20):6837-6845.

59. Wolfram Zillig, *et al.* (1993) Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. *System Appl Microbiol.* 16(4):609-628.

60. Peng X, *et al.* (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. *Virology* 291(2):226-234.

61. Bernander R (2007) The cell cycle of Sulfolobus. *Mol Microbiol* 66(3):557-562.

62. Duggin IG & Bell SD (2006) The chromosome replication machinery of the archaeon Sulfolobus solfataricus. *J Biol Chem* 281(22):15029-15032.

63. Bernander R & Poplawski A (1997) Cell cycle characteristics of thermophilic archaea. *J Bacteriol* 179(16):4963-4969.

64. Wang IN, Smith DL, & Young R (2000) Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol* 54:799-825.

65. Bernhardt TG, Wang IN, Struck DK, & Young R (2002) Breaking free: "protein antibiotics" and phage lysis. *Res Microbiol* 153(8):493-501.

66. Miller S & Krijnse-Locker J (2008) Modification of intracellular membrane structures for virus replication. *Nat Rev Microbiol* 6(5):363-374.

67. Fulton J*, et al.* (2009) Genetics, biochemistry and structure of the archaeal virus STIV. *Biochem Soc Trans* 37(Pt 1):114-117.

68. Fontana J, Lopez-Montero N, Elliott RM, Fernandez JJ, & Risco C (2008) The unique architecture of Bunyamwera virus factories around the Golgi complex. *Cellular microbiology* 10(10):2012-2028.

69. Novoa RR*, et al.* (2005) Virus factories: associations of cell organelles for viral replication and morphogenesis. *Biology of the cell / under the auspices of the European Cell Biology Organization* 97(2):147-172.

70. Claverie JM (2006) Viruses take center stage in cellular evolution. *Genome Biol* 7(6):110.

71. Raoult D & Forterre P (2008) Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol* 6(4):315-319.

72. De Paepe M & Taddei F (2006) Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biol* 4(7):e193.

73. Breitbart M, Miyake JH, & Rohwer F (2004) Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett* 236(2):249-256.

74. Sano E, Carlson S, Wegley L, & Rohwer F (2004) Movement of viruses between biomes. *Appl Environ Microbiol* 70(10):5842-5846.

75. Snyder JC*, et al.* (2007) Virus movement maintains local virus population diversity. *Proceedings of the National Academy of Sciences of the United States of America* 104(48):19102-19107.

76. Bettstetter M, Peng X, Garrett RA, & Prangishvili D (2003) AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. *Virology* 315(1):68-79.

77. Sambrook J FE, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab Press, Plainview, NY) 2nd Ed.

78. Eder W, Ludwig W, & Huber R (1999) Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of kebrit deep, red Sea. *Archives of microbiology* 172(4):213-218.

79. Beck P HR (1997) Detection of cell viability in cultures of hyperthermophiles. *FEMS Microbiol Lett* 48:11-14.

80. Wold S, Skarstad K, Steen HB, Stokke T, & Boye E (1994) The initiation mass for DNA replication in Escherichia coli K-12 is dependent on growth rate. *EMBO J* 13(9):2097-2102.

81. Chen L*, et al.* (2005) The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. *J Bacteriol* 187(14):4992-4999.

82. She Q*, et al.* (2001) The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proceedings of the National Academy of Sciences of the United States of America* 98(14):7835-7840.

83. Bize A*, et al.* (2009) A unique virus release mechanism in the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* 106(27):11306-11311.

84. Albers SV*, et al.* (1999) Glucose transport in the extremely thermoacidophilic Sulfolobus solfataricus involves a high-affinity membrane-integrated binding protein. *J Bacteriol* 181(14):4285-4291.

85. Stingl K*, et al.* (2008) In vivo interactome of Helicobacter pylori urease revealed by tandem affinity purification. *Molecular & cellular proteomics : MCP* 7(12):2429-2441.

86. Edman P (1949) A method for the determination of amino acid sequence in peptides. *Arch Biochem* 22(3):475.

87. Falb M*, et al.* (2006) Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey. *J Mol Biol* 362(5):915-924.

88. Krogh A, Larsson B, von Heijne G, & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 305(3):567-580.

89. Bagos PG, Tsirigos KD, Plessas SK, Liakopoulos TD, & Hamodrakas SJ (2009) Prediction of signal peptides in archaea. *Protein engineering, design & selection : PEDS* 22(1):27-35.

90. Happonen LJ*, et al.* (2010) Familial relationships in hyperthermo- and acidophilic archaeal viruses. *J Virol* 84(9):4747-4754.

91. Krupovic M & Bamford DH (2007) Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. *BMC Genomics* 8:236.

92. Pei J, Tang M, & Grishin NV (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res* 36:W30-34.

93. Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.

94. Prangishvili D & Quax TE (2011) Exceptional virion release mechanism: one more surprise from archaeal viruses. *Current opinion in microbiology* 14(3):315-320.

95. Berkner S, Wlodkowski A, Albers SV, & Lipps G (Inducible and constitutive promoters for genetic systems in Sulfolobus acidocaldarius. *Extremophiles* 14(3):249-259.

96. Albers SV, Szabo Z, & Driessen AJ (2003) Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J Bacteriol* 185(13):3918-3925.

97. van de Vossenberg JL, Driessen AJ, & Konings WN (1998) The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* 2(3):163-170.

98. Albers SV*, et al.* (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon Sulfolobus solfataricus. *Appl Environ Microbiol* 72(1):102-111.

99. Krupovic M & Bamford DH (2008) Holin of bacteriophage lambda: structural insights into a membrane lesion. *Mol Microbiol* 69(4):781-783.

100. Bernhardt TG, Wang IN, Struck DK, & Young R (2002) Breaking free: "protein antibiotics" and phage lysis. *Res Microbiol* 153(8):493-501.

101. Bize A*, et al.* (2009) A unique virus release mechanism in the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* 106(27):11306-11311.

102. Quax TE*, et al.* (2011) Simple and elegant design of a virion egress structure in Archaea. *Proceedings of the National Academy of Sciences of the United States of America* 108(8):3354-3359.

103. Snyder JC, Brumfield SK, Peng N, She Q, & Young MJ (2011) Sulfolobus turreted icosahedral virus c92 protein responsible for the formation of pyramid-like cellular lysis structures. *J Virol* 85(13):6287-6292.

104. Prangishvili D & Quax TE (2011) Exceptional virion release mechanism: one more surprise from archaeal viruses. *Curr Opin Microbiol* 14(3):315-320.

105. Quax TE, Krupovic M, Lucas S, Forterre P, & Prangishvili D (2010) The Sulfolobus rod-shaped virus 2 encodes a prominent structural component of the unique virion release system in Archaea. *Virology* 404(1):1-4.

106. Fu CY*, et al.* (2010) In vivo assembly of an archaeal virus studied with whole-cell electron cryotomography. *Structure (London, England : 1993)* 18(12):1579-1586.

107. Iancu CV*, et al.* (2010) Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J Mol Biol* 396(1):105-117.

108. Sapay N, Guermeur Y, & Deleage G (2006) Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *Bmc Bioinformatics* 7.

109. Nakai K & Kanehisa M (1991) Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* 11(2):95-110.

110. Borgese N & Righi M (2010) Remote origins of tail-anchored proteins. *Traffic* 11(7):877-885.

111. Karlinsey JE, Pease AJ, Winkler ME, Bailey JL, & Hughes KT (1997) The flk gene of Salmonella typhimurium couples flagellar P- and L-ring assembly to flagellar morphogenesis. *J Bacteriol* 179(7):2389-2400.

112. Borgese N & Fasana E (2011) Targeting pathways of C-tail-anchored proteins. *Biochim Biophys Acta* 1808(3):937-946.

113. Bischofberger M, Iacovache I, & van der Goot FG (2012) Pathogenic pore-forming proteins: function and host response. *Cell host & microbe* 12(3):266-275.

114. Snyder JC, Samson RY, Brumfield SK, Bell SD, & Young MJ (2013) Functional interplay between a virus and the ESCRT machinery in Archaea. *Proc Natl Acad Sci U S A*.

115. Quax TE, *et al.* (2013) Massive Activation of Archaeal Defense Genes during Viral Infection. *Journal of virology* 87(15):8419-8428.

116. Zolghadr B, Weber S, Szabo Z, Driessen AJ, & Albers SV (2007) Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in Sulfolobus solfataricus. *Mol Microbiol* 64(3):795-806.

117. Wagner et al, submitted).

118. Berkner S & Lipps G (2007) Characterization of the transcriptional activity of the cryptic plasmid pRN1 from Sulfolobus islandicus REN1H1 and regulation of its replication operon. *J Bacteriol* 189(5):1711-1721.

119. Albers SV, Szabo Z, & Driessen AJ (2003) Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J Bacteriol* 185(13):3918-3925.

120. Gari E, Piedrafita L, Aldea M, & Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. *Yeast* 13(9):837-848.

121. Liou W, Geuze HJ, & Slot JW (1996) Improving structural integrity of cryosections for immunogold labeling. *Histochem Cell Biol* 106(1):41-58.

122. Hohenberg H, Mannweiler K, & Muller M (1994) High-pressure freezing of cell suspensions in cellulose capillary tubes. *J Microsc* 175(Pt 1):34-43.

123. Kremer JR, Mastronarde DN, & McIntosh JR (1996) Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* 116(1):71-76.

124. Frangakis AS & Hegerl R (2001) Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *J Struct Biol* 135(3):239-250.

125. Nicastro D (2009) Cryo-electron microscope tomography to study axonemal organization. *Methods in cell biology* 91:1-39.

126. Nicastro D, *et al.* (2006) The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* 313(5789):944-948.

127. Davies KM, Anselmi C, Wittig I, Faraldo-Gomez JD, & Kuhlbrandt W (2012) Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. *Proc Natl Acad Sci U S A* 109(34):13602-13607.

128. Pettersen EF, *et al.* (2004) UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.

129. Pina M, Bize A, Forterre P, & Prangishvili D (2011) The archeoviruses. *FEMS microbiology reviews* 35(6):1035-1054.

130. Krupovic M, White MF, Forterre P, & Prangishvili D (2012) Postcards from the edge: structural genomics of archaeal viruses. *Adv Virus Res* 82:33-62.

131. Palm P, *et al.* (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium Sulfolobus shibatae. *Virology* 185(1):242-250.

132. Frols S, Gordon PM, Panlilio MA, Schleper C, & Sensen CW (2007) Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. *Virology* 365(1):48-59.

133. Maaty WS*, et al.* (2012) Proteomic analysis of Sulfolobus solfataricus during Sulfolobus Turreted Icosahedral Virus infection. *J Proteome Res* 11(2):1420-1432.

134. Kessler A, Brinkman AB, van der Oost J, & Prangishvili D (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon sulfolobus. *Journal of bacteriology* 186(22):7745-7753.

135. Muskhelishvili G, Palm P, & Zillig W (1993) SSV1-encoded site-specific recombination system in Sulfolobus shibatae. *Mol Gen Genet* 237(3):334-342.

136. Schleper C, Kubo K, & Zillig W (1992) The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. *Proceedings of the National Academy of Sciences of the United States of America* 89(16):7645-7649.

137. Prangishvili D, Koonin EV, & Krupovic M (2013) Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in Archaea. *Biochem Soc Trans* 41(1):443-450.

138. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.

139. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.

140. Oke M*, et al.* (2011) A dimeric Rep protein initiates replication of a linear archaeal virus genome: implications for the Rep mechanism and viral replication. *J Virol* 85(2):925-931.

141. Guilliere F*, et al.* (2009) Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. *J Biol Chem* 284(33):22222-22237.

142. Young MD, Wakefield MJ, Smyth GK, & Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11(2):R14.

143. Lindas AC, Karlsson EA, Lindgren MT, Ettema TJ, & Bernander R (2008) A unique cell division machinery in the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* 105(48):18942-18946.

144. Samson RY*, et al.* (2011) Molecular and structural basis of ESCRT-III recruitment to membranes during archaeal cell division. *Mol Cell* 41(2):186-196.

145. Jaubert C*, et al.* (2013) Genomics and genetics of Sulfolobus islandicus LAL14/1, a model hyperthermophilic archaeon. *Open Biol* 3(4):130010.

146. Makarova KS*, et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9(6):467-477.

147. Makarova KS, Wolf YI, & Koonin EV (2009) Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct* 4:19.

148. Yamaguchi Y, Park JH, & Inouye M (2011) Toxin-antitoxin systems in bacteria and archaea. *Annual review of genetics* 45:61-79.

149. Tachdjian S & Kelly RM (2006) Dynamic metabolic adjustments and genome plasticity are implicated in the heat shock response of the extremely thermoacidophilic archaeon Sulfolobus solfataricus. *Journal of bacteriology* 188(12):4553-4559.

150. Martusewitsch E, Sensen CW, & Schleper C (2000) High spontaneous mutation rate in the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by transposable elements. *Journal of bacteriology* 182(9):2574-2581.

151. Frols S*, et al.* (2007) Response of the hyperthermophilic archaeon Sulfolobus solfataricus to UV damage. *Journal of bacteriology* 189(23):8708-8718.

152. Davison BL, Murray CL, & Rabinowitz JC (1980) Specificity of promoter site utilization in vitro by bacterial RNA polymerases on Bacillus phage phi 29 DNA. Transcription mapping with exonuclease III. *J Biol Chem* 255(18):8819-8830.

153. Loskutoff DJ & Pene JJ (1973) Gene expression during the development of Bacillus subtilis bacteriophage phi29. II. Resolution of viral-specific ribonucleic acid molecules. *Journal of virology* 11(1):87-97.

154. Grahn AM, Bamford JK, O'Neill MC, & Bamford DH (1994) Functional organization of the bacteriophage PRD1 genome. *Journal of bacteriology* 176(10):3062-3068.

155. Bu S*, et al.* (2008) Interaction between two putative glycosyltransferases is required for glycosylation of a serine-rich streptococcal adhesin. *Journal of bacteriology* 190(4):1256-1266.

156. Noffz C, Keppler-Ross S, & Dean N (2009) Hetero-oligomeric interactions between early glycosyltransferases of the dolichol cycle. *Glycobiology* 19(5):472-478.

157. Goulet A*, et al.* (2009) The thermo- and acido-stable ORF-99 from the archaeal virus AFV1. *Protein Sci* 18(6):1316-1320.

158. Oke M*, et al.* (2010) The Scottish Structural Proteomics Facility: targets, methods and outputs. *Journal of structural and functional genomics* 11(2):167-180.

159. Gardner AF, Guan C, & Jack WE (2011) Biochemical characterization of a structure-specific resolving enzyme from Sulfolobus islandicus rod-shaped virus 2. *PloS one* 6(8):e23668.

160. Makarova KS, Yutin N, Bell SD, & Koonin EV (2010) Evolution of diverse cell division and vesicle formation systems in Archaea. *Nature reviews. Microbiology* 8(10):731-741.

161. Horvath P & Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327(5962):167-170.

162. Lillestol RK, Redder P, Garrett RA, & Brugger K (2006) A putative viral defence mechanism in archaeal cells. *Archaea* 2(1):59-72.

163. Hale CR*, et al.* (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139(5):945-956.

164. Zhang J*, et al.* (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45(3):303-313.

165. Hale CR*, et al.* (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* 45(3):292-302.

166. Tang TH*, et al.* (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus. *Mol Microbiol* 55(2):469-481.

167. Hale C, Kleppe K, Terns RM, & Terns MP (2008) Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA* 14(12):2572-2579.

168. Deng L, Garrett RA, Shah SA, Peng X, & She Q (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Mol Microbiol* 87(5):1088-1099.

169. Pandey DP & Gerdes K (2005) Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res* 33(3):966-976.

170. Gerdes K, Christensen SK, & Lobner-Olesen A (2005) Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* 3(5):371-382.

171. Pecota DC & Wood TK (1996) Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. *Journal of bacteriology* 178(7):2044-2050.

172. Fineran PC*, et al.* (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proceedings of the National Academy of Sciences of the United States of America* 106(3):894-899.

173. Makarova KS, Anantharaman V, Aravind L, & Koonin EV (2012) Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biology direct* 7:40.

174. Rajagopala SV & Uetz P (2011) *Network Biology: Methods and applications, Methods in Molecular Biology* (Springer, New York, USA) pp 1-30.

175. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.

176. Hulsen T, de Vlieg J, & Alkema W (2008) BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 9:488.

177. Wolf YI, Makarova KS, Yutin N, & Koonin EV (2012) Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biology direct* 7:46.

178. Wiedenheft B*, et al.* (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477(7365):486-489.

179. Beyenbach KW & Wieczorek H (2006) The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. *J Exp Biol* 209: 577-589.

180. Dunkle JA*, et al.* (2011) Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* 332(6032):981-984.

181. Efremov RG, Baradaran R, & Sazanov LA (2010) The architecture of respiratory complex I. *Nature* 465(7297):441-445.

182. Errington J (2003) Dynamic proteins and a cytoskeleton in bacteria. *Nat Cell Biol* 5(3):175-178.

183. Ghosh A & Albers SV (2011) Assembly and function of the archaeal flagellum. *Biochem Soc Trans* 39(1):64-69.

184. Jakob M, Kaiser S, Gutensohn M, Hanner P, & Klosgen RB (2009) Tat subunit stoichiometry in Arabidopsis thaliana challenges the proposed function of TatA as the translocation pore. *Biochim Biophys Acta* 1793(2):388-394.

185. Jore MM*, et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology* 18(5):529-536.

186. Jacob F & Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology* 3:318-356.

187. Koonin EV (2009) Evolution of genome architecture. *Int J Biochem Cell Biol* 41(2):298-306.

188. Grundy FJ & Henkin TM (2006) From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol* 41(6):329-338.

189. French SL, Santangelo TJ, Beyer AL, & Reeve JN (2007) Transcription and translation are coupled in Archaea. *Mol Biol Evol* 24(4):893-895.

190. Brenner S, Jacob F, & Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190:576-581.

191. Johnson TL, Abendroth J, Hol WG, & Sandkvist M (2006) Type II secretion: from structure to function. *FEMS Microbiol Lett* 255(2):175-186.

192. Gur E, Biran D, & Ron EZ (2011) Regulated proteolysis in Gram-negative bacteria--how and when? *Nat Rev Microbiol* 9(12):839-848.

193. Koide T*, et al.* (2009) Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Syst Biol* 5:285.

194. Wurtzel O*, et al.* (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res* 20(1):133-141.

195. Cannarozzi G*, et al.* (2010) A role for codon order in translation dynamics. *Cell* 141(2):355-367.

196. Drummond DA & Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134(2):341-352.

197. Fredrick K & Ibba M (2010) How the sequence of a gene can tune its translation. *Cell* 141(2):227-229.

198. Tuller T, *et al.* (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141(2):344-354.

199. Timmermans J & Van Melderen L (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cell Mol Life Sci* 67(17):2897-2908.

200. Stenstrom CM, Jin H, Major LL, Tate WP, & Isaksson LA (2001) Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in Escherichia coli. *Gene* 263(1-2):273-284.

201. Gingold H & Pilpel Y (2011) Determinants of translation efficiency and accuracy. *Mol Syst Biol* 7:481.

202. Li GW, Oh E, & Weissman JS (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484(7395):538-541.

203. Sharp PM & Li WH (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15(3):1281-1295.

204. Shao ZQ, Zhang YM, Feng XY, Wang B, & Chen JQ (2012) Synonymous codon ordering: a subtle but prevalent strategy of bacteria to improve translational efficiency. *PLoS One* 7(3):e33547.

205. Kudla G, Murray AW, Tollervey D, & Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324(5924):255-258.

206. Ingolia NT, Ghaemmaghami S, Newman JR, & Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324(5924):218-223.

207. Goldberger RF, Deeley RG, & Mullinix KP (1976) Regulation of gene expression in prokaryotic organisms. *Adv Genet* 18:1-67.

208. Kaberdin VR & Blasi U (2006) Translation initiation and the fate of bacterial mRNAs. *FEMS microbiology reviews* 30(6):967-979.

209. Marzi S, *et al.* (2007) Structured mRNAs regulate translation initiation by binding to the platform of the ribosome. *Cell* 130(6):1019-1031.

210. Cho BK, *et al.* (2009) The transcription unit architecture of the Escherichia coli genome. *Nat Biotechnol* 27(11):1043-1049.

211. Mendoza-Vargas A, *et al.* (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. *PLoS One* 4(10):e7526.

212. Toledo-Arana A, *et al.* (2009) The Listeria transcriptional landscape from saprophytism to virulence. *Nature* 459(7249):950-956.

213. Wurtzel O, *et al.* (2012) Comparative transcriptomics of pathogenic and non-pathogenic Listeria species. *Mol Syst Biol* 8:583.

214. Wurtzel O, *et al.* (2012) The single-nucleotide resolution transcriptome of Pseudomonas aeruginosa grown in body temperature. *PLoS Pathog* 8(9):e1002945.

215. Wiedenheft B, *et al.* (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* 108(25):10092-10097.

216. Li Y & Altman S (2004) Polarity effects in the lactose operon of Escherichia coli. *Journal of molecular biology* 339(1):31-39.

217. Vellanoweth RL & Rabinowitz JC (1992) The influence of ribosome-binding-site elements on translational efficiency in Bacillus subtilis and Escherichia coli in vivo. *Mol Microbiol* 6(9):1105-1114.

218. Forster AC (2012) Synthetic biology challenges long-held hypotheses in translation, codon bias and transcription. *Biotechnology journal*.

219. Welch M, *et al.* (2009) Design parameters to control synthetic gene expression in Escherichia coli. *PLoS One* 4(9):e7002.

220. Ikemura T (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *Journal of molecular biology* 146(1):1-21.

221. Lange SJ, *et al.* (2012) Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res* 40(12):5215-5226.

222. Oppenheim DS & Yanofsky C (1980) Translational coupling during expression of the tryptophan operon of Escherichia coli. *Genetics* 95(4):785-795.

223. McCarthy JE (1990) Post-transcriptional control in the polycistronic operon environment: studies of the atp operon of Escherichia coli. *Mol Microbiol* 4(8):1233-1240.

224. Avrani S, Wurtzel O, Sharon I, Sorek R, & Lindell D (2011) Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* 474(7353):604-608.

225. Mortazavi A, Williams BA, McCue K, Schaeffer L, & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621-628.

226. Markham NR & Zuker M (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* 453:3-31.

227. Jacob WF, Santer M, & Dahlberg AE (1987) A single base change in the Shine-Dalgarno region of 16S rRNA of Escherichia coli affects translation of many proteins. *Proceedings of the National Academy of Sciences of the United States of America* 84(14):4757-4761.

228. Wiedenheft B, *et al.* (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* 108(25):10092-10097.

229. Lintner NG, *et al.* (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J Biol Chem* 286(24):21643-21656.

230. Rousseau F, Schymkowitz J, & Serrano L (2006) Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* 16(1):118-126.

231. Muller RH & Keck CM (2004) Challenges and solutions for the delivery of biotech drugs--a review of drug nanocrystal technology and lipid nanoparticles. *J Biotechnol* 113(1-3):151-170.

232. Snyder JC, Samson RY, Brumfield SK, Bell SD, & Young MJ (2013) Functional interplay between a virus and the ESCRT machinery in Archaea. *Proc Natl Acad Sci U S A*.

233. Snyder JC, *et al.* (2013) Insights into a viral lytic pathway from an archaeal virus-host system. *Journal of virology* 87(4):2186-2192.

234. Happonen LJ, *et al.* (Familial relationships in hyperthermo- and acidophilic archaeal viruses. *J Virol* 84(9):4747-4754.

235. Russel M (1995) Moving through the membrane with filamentous phages. *Trends Microbiol* 3(6):223-228.

236. Marciano DK, Russel M, & Simon SM (2001) Assembling filamentous phage occlude pIV channels. *Proceedings of the National Academy of Sciences of the United States of America* 98(16):9359-9364.

237. Davis BM, *et al.* (2000) Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXphi. *Science* 288(5464):333-335.

238. Linderoth NA, Simon MN, & Russel M (1997) The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. *Science* 278(5343):1635-1638.

239. Opalka N, *et al.* (2003) Structure of the filamentous phage pIV multimer by cryo-electron microscopy. *J Mol Biol* 325(3):461-470.

240. Al-Khayat HA, Morris EP, & Squire JM (2009) The 7-stranded structure of relaxed scallop muscle myosin filaments: support for a common head configuration in myosin-regulated muscles. *J Struct Biol* 166(2):183-194.

241. Redder P, *et al.* (2009) Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. *Environ Microbiol* 11(11):2849-2862.

242. Blum H, Zillig W, Mallok S, Domdey H, & Prangishvili D (2001) The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. *Virology* 281(1):6-9.

243. Klimczak LJ, Grummt F, & Burger KJ (1985) Purification and characterization of DNA polymerase from the archaebacterium Sulfolobus acidocaldarius. *Nucleic Acids Res* 13(14):5269-5282.

244. Gustafsson C*, et al.* (2012) Engineering genes for predictable protein expression. *Protein expression and purification* 83(1):37-46.

245. Maaty WS*, et al.* (2012) Global analysis of viral infection in an archaeal model system. *Front Microbiol* 3:411.

246. Okutan E*, et al.* (2013) Novel insights into gene regulation of the rudivirus SIRV2 infecting Sulfolobus cells. *RNA Biol* 10(5).

Co-author affiliations

Institut Pasteur, Paris, 75015 France:
Ariane Bize, Patrick Forterre, Soizick Lucas, Mart Krupovic, Guennadi Sezonov, Mery Pina, David Prangishvili, Biologie Moléculaire du Gène chez les Extrêmophiles

Gerard Pehau-Arnaudet, Marie-Christine Prevost, Martin Sachse, Institut Pasteur, Plate-Forme de Microscopie Ultrastructurale

Marie-Agnes Dillies, Jean-Yves Coppee, Bernd Jagla, Odile Sismeiro , Plate-forme Transcriptome et Epigénome

Cosmin Saveanu, Institut Pasteur, Génétique des Interactions Macromoléculaires

Université Paris VII, 75018 Paris, France
Olivier Tenaillon, INSERM U722, Faculté de médecine Xavier Bichat

Wageningen University, 7603 HB Wageningen, The Netherlands
Fabian Blombach, Stan J. J. Brouns, Jasper J. Koehorst, John van der Oost, Richard van der Oost, Laboratory of Microbiology, Wageningen University

Max-Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany
Sonja-Verena Albers, Julia Reimann, Molecular Biology of Archaea

Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany
Bertram Daum, Sabine Häder, Werner Kühlbrandt, Deryck Mills, Özkan Yildiz, Department of Structural Biology

KU Leuven, 3001 Leuven, Belgium
Rob Lavigne, Marleen Voet, Laboratory of Gene Technology, Department of Biosystems

Uppsala University, SE-752 36 Uppsala, Sweden
Erik A. Karlsson, Karin Ekefjärd, Rolf Bernander, Department of Molecular Evolution, Evolutionary Biology Center

Anthony C. Forster, E. Gerhart H. Wagner, Department of Cell and Molecular Biology

National Center for Biotechnology Information, Bethesda, MD 20894, USA
Eugene V. Koonin, Kira S. Makarova, Wenqi Ran,Yuri I. Wolf, National Library of Medicine, National Institutes of Health

Weizmann Institute of Science, Rehovot 76100, Israel,
Rotem Sorek, Omri Wurtzel, Department of Molecular Genetics

Nederlandse samenvatting

Virussen zijn de meest voorkomende biologische deeltjes op deze planeet en er wordt geschat dat er minstens 10 keer zoveel virussen zijn dan micro-organismen. Als je alle virussen op elkaar zou stapelen, creëert dit een toren, die zó hoog is, dat deze 200 lichtjaren in het heelal zou reiken. Virussen worden op allerlei plaatsen gevonden, zelfs op plekken met een zeer extreem klimaat, zoals in hete geisers, waar kokend water uit het binnenste van de aarde aan de oppervlakte komt. De virussen, die voorkomen op deze plaatsen infecteren 'extremofiele' gastheer cellen die net als de virussen aangepast zijn aan het extreme klimaat. Al het leven op aarde wordt ingedeeld in drie groepen: archaea, bacteriën en eukaryoten. De meeste organismen, die leven in extreme milieus, zijn archaea. De virussen welke archaea infecteren vormen een aparte groep die gekenmerkt wordt door een hoge verscheidenheid. Over het algemeen zijn zowel de vormen van deze virussen, als hun genetisch materiaal zeer divers. Sommige virussen zien er zelfs uit als een kleine druppel of een fles.

Dit proefschrift gaat over het virus SIRV2, dat het archaeon *Sulfolobus islandicus* infecteert, een organisme dat geïsoleerd is uit een hete bron op IJsland. SIRV2 heeft een speciale infectie cyclus die beschreven staat in **hoofdstuk 1**. Na infectie vermeerdert het virus zich in de cel. Om uit de cel te ontsnappen heeft het virus een bijzondere strategie ontwikkeld die gebaseerd is op de vorming van piramide structuren (VAPs) op het celmembraan van het archaeon. Deze structuren openen aan het einde van de infectiecyclus, waardoor er grote gaten in het celmembraan ontstaan, waardoor het virus de cel kan verlaten.

In **hoofdstuk 2** wordt beschreven dat de piramide structuren bestaan uit het virus eiwit PVAP. Deze piramides kunnen geïsoleerd worden als individuele deeltjes los van de cel. Resultaten uit **hoofdstuk 3** laten zien dat deze deeltjes de vorm hebben van een holle piramide met een heptagonale omtrek en een zevenvoudige symmetrie, uniek in de natuur. De piramide vormige VAPs ontstaan op het celmembraan en ze worden groter door de aangroei aan de onderkant van de zeven driehoekige vlakken. Het eiwit PVAP kan uit zichzelf piramides vormen, zoals werd aangetoond door dit eiwit tot expressie te brengen in een ander archaeon en in een bacterie.

In **hoofdstuk 4** wordt een model gepresenteerd voor het bouwproces van de VAP. Microscopie experimenten en een mutatieanalyze van het eiwit PVAP laat zien dat het N-terminale domein in het membraan integreert en de buitenste laag van de VAP vormt. Het C-terminale domein van het eiwit is essentieel om interacties aan te gaan tussen de verschillende monomeren van het eiwit, wat nodig is om de piramide te vormen. Dit gedeelte van het eiwit is te vinden in de binnenste laag van de piramides. Expressie van PVAP in de eukaryoot bakkersgist (*Saccharomyces cerevisiae*) leidt tot formatie van VAPs op alle cellulaire membranen, inclusief het kernmembraan

en de mitochondriën. Dit toont aan dat het eiwit spontaan in allerlei membranen kan integreren en functioneert als een universeel systeem om de vorm van membranen te veranderen. Dit systeem zou daarom gebruikt kunnen worden voor biotechnologische toepassingen, zoals de specifieke afgifte van medicijnen in het menselijk lichaam.

De virus-gastheer interactie wordt verder onder de loep genomen in **hoofdstuk 5**. Alle RNA in geïnfecteerde cellen wordt gesequenced om informatie te verkrijgen over expressiepatronen van alle genen van de gastheer en het virus. Daarnaast is er gescreend voor alle mogelijke interacties tussen virus en gastheereiwitten. Expressie van de virale genen begint gelijktijdig vanaf beide uiteinden van het lineaire virus genoom. Als reactie op de virusinfectie worden genen van de geïnfecteerde cel geactiveerd, welke een rol spelen bij afweer tegen virussen. Een belangrijk afweer mechanisme van bacteriën en archaea is het CRISPR-Cas systeem. Dit afweercomplex bestaat uit verschillende Cas eiwitten, die aanwezig zijn in een ongelijke aantallen. De genen, die coderen voor de Cas eiwitten zijn georganiseerd in operons. Genen met gerelateerde functies worden in de genomen van bacteriën en archaea vaak naast elkaar gegroepeerd in operons aangetroffen. Deze organisatie faciliteert de co-expressie van genen. Deze co-expressie is voordelig wanneer alle eiwitten gelijktijdig en in gelijke hoeveelheden geproduceerd moeten worden. Wanneer echter, zoals in het geval van het Cas complex, eiwitten in verschillende verhoudingen nodig zijn, moet daartoe een extra vorm van regulatie gebruikt worden. De resultaten van de analyze gepresenteerd in **hoofdstuk 6** laten zien dat een verschil in translatie efficiëntie van eiwitten gecodeerd op operons, verantwoordelijk is voor de productie van deze eiwitten in ongelijke verhouding. In het algemeen kan een verschil in de eiwitproductie het beste voorspeld worden door het codongebruik in de genen binnen één operon te analyzeren. Naast het Cas complex zijn er nog veel meer belangrijke eiwitcomplexen met ongelijke stoichiometrie, die gecodeerd worden op operons. Het beschreven mechanisme blijkt geconserveerd in de meeste archaea en bacteriën en het zorgt voor de verschillen in productie van het merendeel van de eiwitcomplexen, die zijn gecodeerd op operons.

Résumé Français

Les résultats présentés dans cette thèse donnent un nouveau regard sur plusieurs aspects de biologie moléculaire des archées, des bactéries et des leurs virus. Il est admis que chacun des trois domaines du vivant est associé à un groupe spécifique des virus. Les virus spécifiques aux archées ont la particularité d'être morphologiquement et génétiquement extrêmement divers. Certains de ces virus, comme le virus SIRV2 infectant *Sulfolobus islandicus*, ont des cycles d'infection remarquablement complexes. Comme décrit dans le **chapitre 1**, l'infection par SIRV2 aboutie à la formation sur la surface de la cellule hôte de grandes structures pyramidales associées à ce virus (VAP). Ces structures s'ouvrent à la fin du cycle, produisant de larges pores qui permettent la libération des particules virales, en forme de bâtonnets, assemblés au cours du cycle dans le cytoplasme. Le mécanisme de libération du virus SIRV2 est unique et ne ressemble à aucun des mécanismes connus pour les virus des bactéries et eucaryotes. Comme présenté dans le **Chapitre 2**, l'analyse de la composition protéique des cellules infectées par SIRV2, a révélé une forte accumulation de la protéine membranaire PVAP codée par le virus, ce qui suggère son implication directe dans la formation des VAP. Les VAP peuvent être isolées sous forme de structures compactes, comme ceci a été démontré dans le **chapitre 3**. Étude au microscope électronique de ces structures a démontré qu'elles correspondent à des pyramides heptagonales creuses. Cette géométrie est extrêmement inhabituelle et la symétrie septuple est assez rare dans le monde du vivant (ex. 20S du protéasome, la myosine). Ces structures peuvent avoir des tailles différentes, ce qui correspond probablement aux différents stades de leur développement. Cette observation permet de supposer qu'elles grandissent par la croissance progressive des facettes triangulaires qui les forment. L'analyse de la composition protéique des VAP démontre qu'elles sont composées d'une seule protéine, PVAP. Les anticorps dirigés contre cette protéine marquent spécifiquement les VAP comme observé au microscope électronique sur les coupes fines de cellules infectées. La protéine PVAP est suffisante pour la formation des VAP. Les VAP en structures pyramidales ont été observées après l'expression hétérologue du gène viral correspondant chez l'archée *S. acidocaldarius* et chez la bactérie, *Escherichia coli*. Une analyse complémentaire des mutants portant des versions raccourcies du gène VAP, telles que décrites dans le **chapitre 4**, a démontré que, exceptés les 10 derniers acides aminés C-terminaux, tous les domaines de la protéine sont essentiels pour la formation de VAP. La PVAP peut former des oligomères de plusieurs tailles, y compris ceux d'un heptamère, qui agissent vraisemblablement comme points de nucléation pour la formation des VAP sur la membrane cellulaire. L'analyse des mutants de délétions a indiqué que les deux domaines, C – et N-terminaux, sont essentiels pour l'interaction entre les monomères de PVAP. L'observation détaillée à l'aide de la technique de tomographie cryo-électronique a

permis de conclure que les VAP formées dans les systèmes naturels ou hétérologues ont deux couches dans leur structure. La couche extérieure correspond à la membrane cellulaire. La couche intérieure tournée vers le cytoplasme représente vraisemblablement un feuillet protéique formé par les interactions entre les domaines C-terminal des PVAPs connectées à la membrane à l'aide d'une courte région linker. Les feuillets sont légèrement pliés, ce qui donne à la structure finale l'apparence d'un tipi. A la jonction de deux plaques triangulaires la structure est perforée, ce qui crée des points destinés à la rupture. En outre, les données discutées dans ce chapitre soulignent le caractère unique de cette protéine, car elle est capable de former des VAP tout à fait normales dans les membranes des bactéries, des archées et des eucaryotes malgré le fait que ces membranes sont fondamentalement différentes par leurs compositions lipidiques et protéiques. Exprimées chez *Saccharomyces cerevisiae*, les VAPs sont formées sur toutes les membranes, y compris celles des mitochondries, ce qui suggère que cette protéine s'insert dans les membranes de façon spontanée. Ainsi, les PVAP peuvent représenter un système universel de remodelage des membranes qui pourrait être exploité à des fins biotechnologiques, par exemple pour le développement d'un système universel permettant la formation contrôlée de grandes ouvertures de ~ 100 nm dans une bicouche lipidique. Production des VAP est l'une des conséquences les plus marquantes de l'infection par SIRV2 des cellules hôtes. L'approche de transcriptome a permis d'établir une carte globale de l'expression des gènes viraux et ceux de l'hôte pendant le cycle d'infection. Ces résultats sont présentés dans le **chapitre 5**. Immédiatement après l'infection, la transcription des gènes viraux commence simultanément à partir de deux extrémités de son génome linéaire. Toutes les interactions possibles entre l'ensemble des protéines codées par SIRV2 ont été analysées par l'approche du double-hybride chez la levure et les résultats obtenus ont permis de mieux comprendre les fonctions codées par le génome de SIRV2 pour lequel le rôle de plus de la moitié des gènes reste inconnu. Les cellules hôtes répondent à l'infection virale, en modifiant l'expression de plus de 30% de leurs gènes. Les gènes impliqués dans la division cellulaire sont régulés à la baisse, tandis que ceux qui jouent un rôle dans la défense antivirale sont activés. Plus précisément, pour la première fois l'activation massive des gènes codant pour le système anti-toxine mais également le système CRISPR-Cas a été observée chez les archées. Les différents degrés d'expression de ces différents systèmes mettent en évidence l'importance des fonctions qu'ils codent. Les complexes ribonucléoprotéiques multiproteiques associés au système de défense antivirale CRISPR ont généralement une stoechiométrie non proportionnelle, où les 4-6 types de sous-unités protéiques qui le composent sont présents dans des quantités non proportionnelles. Comme la plupart des gènes procaryotiques impliqués dans la même fonction, les gènes CAS sont regroupés dans les opérons permettant leur co-expression coordonnée (comme cela a en effet été observé dans l'analyse du transcriptome décrit dans le **chapitre 5**). Ceci est avantageux lorsque des quantités égales de produits de gènes sont nécessaires, comme c'est le cas pour les complexes protéiques avec la stœchiométrie proportionnelle. Cependant,

les complexes CRISPR ainsi que de nombreux autres complexes protéiques cellulaires importants (ribosome, ATPase, le système de sécrétion de la protéine TAT) sont des complexes avec une stœchiométrie non proportionnelle pour sous unités qui les composent. Comme discuté dans le **chapitre 6**, à l'aide de l'approche de la génomique comparative, il a été démontré que la traduction différentielle est un mécanisme clef permettant l'expression modulée des gènes regroupés dans opérons. Egalement, le biais des codons est généralement le meilleur indicateur *in silico* et ce paramètre permet d'expliquer la production inégale des protéines à partir de la même molécule de l'ARNm. En outre, l'analyse de l'expression des gènes portant à la suite de la mutagenèse des codons synonymiques a apporté la preuve que l'initiation de la traduction peut se produire à partir des sites inter-cistroniques. L'universalité du phénomène de modulation de l'efficacité de la traduction suggère qu'il s'agit d'une voie universelle de contrôle chez les bactéries et chez les archées et ce mécanisme permet une production différentielle des protéines codées par les gènes organisés en opérons.

Acknowledgements

The work in this thesis has been supported by a lot of people. It is impossible to list them all here individually. Nevertheless I would like to thank some persons in particular.

First of all I would like to thank my supervisors John van der Oost and David Prangishvili.

David, thank you for introducing me to the wonderful world of archaea and their viruses, a subject which I am working on with tremendous pleasure. I am very grateful for all the opportunities you have offered me to design my own lines of research and to start several collaborations, which have been instrumental for the progress of my projects. It has been very inspiring to work with you at the Institut Pasteur, which is so rich in (viral) history.

John, heel erg bedankt voor het warme welkom in je groep in Wageningen. Je 'beroemde' enthousiasme werkt aanstekelijk en is inspirerend voor iedereen om je heen. Dit enthousiasme is het fundament van de positieve werksfeer die er binnen BacGen heerst. Het was heel spannend om samen een nieuw onderzoeksterrein te ontdekken en ik denk dat ik voor ons beiden spreek als ik zeg dat we hier veel van geleerd hebben.

David, John, I am very happy with the trust and freedom that you have both given to me, which allowed the special realization of this joint PhD project. Even though the administrative load of the project has been a challenge, the extra scientific opportunities, which it has provided made it definitively worth the effort.

In this respect I also would like to thank the members of the Frans-Nederlandse Academie for their expertise and advice on the construction of the joint PhD project.

Much of the work described in this thesis could not have been realized without the contribution of many collaborators.

Werner Kühlbrandt, Bertram Daum, thank you very much for hosting me at the Max Planck Institute of Biophysics in Frankfurt. The cryo-EM equipment and expertise at your institute is impressive and our fruitful collaboration has helped the project some major leaps forward.

Sonja-Verena Albers, Julia Reimann, it was great to be welcomed by you at the Max Planck Institute of Terrestrial Microbiology in Marburg. I have enjoyed the nice atmosphere and the efficient working environment, during which I have constructed many mutants. Sonja, thank you very much for all your helpful advice.

Rob Lavigne, Marleen Voet, thanks for welcoming me at the University of Leuven and acquainting me with the possibilities of interactomics to study virus-host interactions.

Eugene Koonin, Yuri Wolf, Kira Makarova, thank you for the commitment to our project and insightful advice. Yuri, special thanks to you for the effort you have put in the many analyzes you have run to answer the high number of questions of John and me.

Also I would like to thank all members of the Genomics and the Electron Microscopy Platform of the Institut Pasteur. Martin Sachse, I am very grateful for your help during my PhD research.

It has been great to work in both research groups at the Institut Pasteur and at Wageningen University. Past and present members of BMGE, it has been very pleasant working together. Ana, without you I would have been hopelessly lost in 'French bureaucracy'. Thanks for you continuing help. Patrick, I appreciate very much your 'evolutionary goggles' in discussions, which provoke me to think outside the box. Mery, Tomo, you always make me laugh and I hope that despite the distance we will be able to continue seeing each other.

Past and present members of BacGen, thank you for the good atmosphere in the group. Stan en Matthijs, heel erg bedankt het delen van jullie expertise en het geven van nuttig advies.. Elleke, zonder jouw mentale steun als fietsmaatje was ik waarschijnlijk menigmaal natgeregend achtergebleven langs de kant van de weg tussen Ede en Wageningen. Bas, Nico en Raymond, het was heel erg lachen om met jullie een kamer te delen. Ik hoop dat ik ondanks de vele dagen waarop mijn bureau onbezet was, en de regen die ik meebracht op de andere dagen, toch een leuke kamergenoot ben geweest. Al het ondersteunend personeel van Microbiologie: heel erg bedankt, zonder jullie zou het een chaos worden.

Thanks also to my students: Elena Rensen, Jasper Koehorst, Jérémie Chaligné. It was nice to work together.

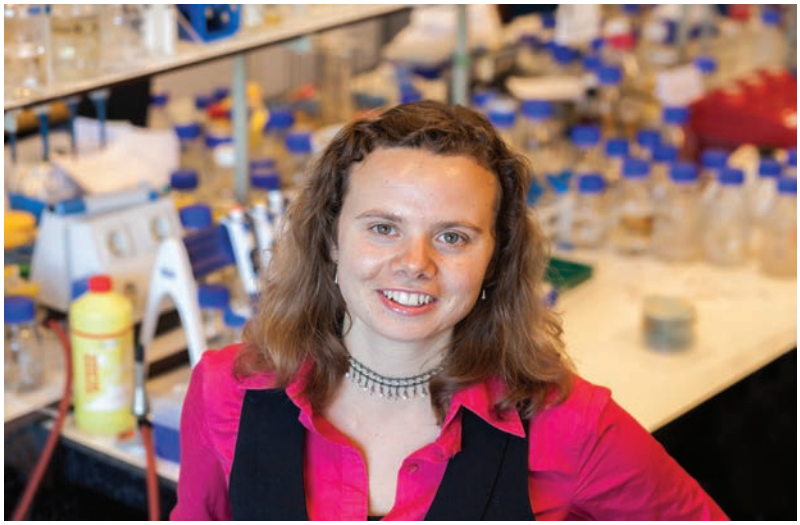
Then, I also want to thank many friends in Paris, especially those who have offered me a place to sleep when I was homeless again: Inga, Elise, Sukhi, Julia, thanks for the support, advice and the clean bed you have offered. Roel, het was leuk om samen een appartement in Vanves te delen en super handig om samen door het joint PhD traject te gaan, waarbij we "best practices" van elkaar konden afkijken.

Also many thanks to friends in the Netherlands and at other places in the world. It was great to share experiences and I appreciate it very much that we stayed in contact despite the (temporal) distance. Xenia, thanks for your support. Mathijs, heel fijn om samen het wel en wee van het leven van een AIO te bespreken. Linda, heel erg bedankt voor het luisterend oor, wanneer ik weer eens aan het doordraven ben over een wetenschappelijke futiliteit.

Pap, mam, Robin, Silvan, Ivo en alle andere familie, heel erg bedankt voor jullie steun en bruikbaar advies. Remco, bedankt voor je begrip en flexibiliteit bij het omgaan met onze ongewone woon/werk situatie. Ik ben heel erg blij met jou.

About the author

Tessa Evita Frederike Quax was born in 1986 on September 22 in Leiden, The Netherlands. After completing the gymnasium at the Dr Aletta Jacobs College at Hoogezand-Sappemeer *cum laude*, she started studying Biology at Utrecht University. In 2007 she obtained her Bachelor degree *cum laude* and started the Master program Biomolecular Sciences at Utrecht University. Next to this master, she was selected for participation in the excellent tracée Xtrack of the Graduate School for Life Sciences of Utrecht University. During the first year of her master program she performed a nine-month internship in the Microbiology research group, where she studied mushroom formation under supervision of Prof. Dr Han Wösten. In this period she became highly interested in archaea. In the second year of her master she opted for an internship of six months at the Institut Pasteur in the group of Dr David Prangishvili and Prof. Dr Patrick Forterre to study viruses infecting archaea. In 2009 she completed her Master *cum laude* and received the Xtrack award for her research proposal to study the CRISPR-Cas anti-viral defense system. She applied successfully for a personal PhD grant of the French government to continue her work on archaeal viruses at the Institut Pasteur, which allowed her to study a novel virus release mechanism in detail. Because of her interest in anti-viral defense, she initiated collaboration with the group of Prof. Dr John van der Oost of the Microbiology department of Wageningen University. This collaboration was accommodated by the construction of a 'joint PhD project', during which Tessa Quax performed research both at the Institut Pasteur and at Wageningen University. Results of this joint PhD project are presented in this thesis.



List of publications

Daum B.*, **Quax T.E.F**.*, Sachse M., Mills D., Reimann J., Yildiz Ö., Häder S., Saveanu C., Forterre P., Albers S.-V., Kühlbrandt W. and Prangishvili D. (2013) Supramolecular organisation and self-assembly of a universal pyramidal membrane remodelling system, *submitted*.

Pina M., Basta T., **Quax T.E.F**., de Cian A., Baconnais S., Cortez D., Lambert S., Le Cam E., Bell S.D., Forterre F. and Prangishvili D. (2013) Eukaryotic traits of the genome replication mechanism of the hyperthermophilic archaeal virus AFV1, *in revision*

Quemin E., Lucas S., Daum B., **Quax T.E.F**., Kühlbrandt W., Forterre P., Albers S.-V., Prangishvili D. and Krupovic M. (2013). First insights into the entry process of hyperthermophilic archaeal viruses, *Journal of Virology, in press*

Quax T.E.F., Wolf Y.I., Koehorst J.J., Wurtzel O., van der Oost R., Ran W., Blombach F., Makarova K.S., Brouns S.J.J., Forster A.C., Wagner E.G.H., Sorek R., Koonin E.V. and van der Oost J. (2013) Differential translation tunes uneven production of operon-encoded proteins, *Cell Reports*, 4:1-7.

Quax T.E.F., Voet M., Sismeiro O., Dillies M.-A., Jagla B., Coppee J.-Y., Sezonov G., Forterre. P, van der Oost J., Lavigne R. and Prangishvili D. (2013) Massive activation of archaeal defense genes during viral infection. *Journal of Virology*, 87(15): 8419-28

Snyder J.C., Brumfield S.K., Kerchner K.M., **Quax T.E.F**, Prangishvili D. and Young M.J. (2013) Insights into a viral lytic pathway from an archaeal virus-host system. *Journal of Virology*, 87(4):2186-92

Prangishvili D. and **Quax T.E.F.** (2011). Exceptional virion release mechanism: one more surprise from Archaeal viruses. *Current Opinion in Microbiology*, 4(3):315-20.

Quax T.E.F., Lucas S., Reimann J., Pehau-Arnaudet G., Prevost M.-C., Forterre P., Albers S.-V. and Prangishvili, D (2011). Simple and elegant design of a virion egress structure in Archaea. *Proceedings of the National Academy of Sciences USA,* 108(8): 3354-3359

Quax T.E.F., Krupovič M., Lucas S., Forterre P. and Prangishvili, D. (2010). The Sulfolobus rod-shaped virus 2 encodes a prominent structural component of the unique virion release system in Archaea. *Virology* 404:1-4

Bize A., Karlsson E.A., Ekefjärd K., **Quax T.E.F.**, Pina M., Prevost M.-C., Forterre P., Tenaillon O., Bernander R. and Prangishvili D. (2009).A Unique Virus Release Mechanism in the Archaea. *Proceedings of the National Academy of Sciences USA* 106(27):11306-11

*Equal contribution.

Overview of completed training activities

Discipline specific activities

Meetings

- ALW meeting Molecular Genetics. 2013. Lunteren, The Netherlands.*
- Thermophiles conference. 2013. Regensburg, Germany.*
- Gordon conference, Archaea: Ecology, Metabolism & Molecular Biology. 2013. Lucca, Italy.**
- Microbiology days of the Institut Pasteur. 2013. Vichy, France.*
- EMBO/EMBL symposium, The complex life of mRNA. 2012. Heidelberg, Germany.
- Molecular Biology of Archaea III meeting. 2012. Marburg, Germany.**
- Annual Meeting NWO study group Protein Research, Nucleic Acids and Lipids & Membranes. 2012. Veldhoven, The Netherlands.**
- EMBO conference, Viruses of Microbes II. 2012. Brussel, Belgium.*
- Microbiology days of the Institut Pasteur. 2012. Paris, France.**
- ALW meeting Molecular Genetics. 2011. Lunteren, The Netherlands.*
- Thermophiles conference. 2011. Montana, USA.*
- French Archaea days. 2011. Paris, France.*
- CRISPR meeting, Mechanisms and Applications. 2010. Wageningen, The Netherlands.
- Extremophiles conference. 2010. Ponta Delagada, Portugal.**
- Microbiology days of the Institut Pasteur. 2010. Paris, France.**

* Oral presentation

** Poster presentation

Courses

- Guest Scientist – Archaeal protein expression. 2012. Max Planck Institute for Terrestrial Microbiology, Marburg, Germany.
- Guest Scientist – Whole Cell Cryo-Electron Tomography. 2012. Max Planck Institute of Biophysics, Frankfurt, Germany.
- Radiation course. 2012. Wageningen University
- EMBO Practical Course – Electron Microscopy & Stereology in Cell Biology. 2010. Oslo, Norway.
- Guest Scientist – Cryo-electron Microscopy. 2010. The Scripps Institute, La Jolla, USA

– Guest Scientist – Hands on course Genetic Tools *Sulfolobus*. 2010. Max Planck Institute for Terrestrial Microbiology, Marburg, Germany.

General courses

– Voice and presentation skills training, Wageningen University.
– Mobilizing the scientific network, Wageningen University.
– Career orientation, Wageningen University.

Optionals

– PhD study trip 2011 to Japan and China.
– Preparing PhD Research Proposal.
– Bacterial Genetics Group Meetings (weekly).
– Laboratory of Microbiology PhD Meetings (biweekly).

The research in this thesis was financially supported by a fellowship from the Ministère de l'Enseignement supérieur et de la Recherche de France, an ALW TOP grant of the Netherlands Organization for Scientific Research (NWO) and a Van Gogh grant of the French-Dutch Academy

Cover: Zeno van der Broek
Printing: Gildeprint