Extremophiles (2006) 10:17–28
DOI 10.1007/s00792-005-0492-x

REVIEW

Georg Lipps

Plasmids and viruses of the thermoacidophilic crenarchaeote *Sulfolobus*

Received: 17 March 2005 / Accepted: 15 July 2005 / Published online: 6 January 2006

Abstract The crenarchaeote *Sulfolobus* spp. is a host for a remarkably large spectrum of viruses and plasmids. The genetic elements characterized so far indicate a large degree of novelty in terms of morphology (viruses) and in terms of genome content (plasmids and viruses). The viruses and conjugative plasmids encode a great number of conserved proteins of unknown function due to the lack of sequence similarity to functionally characterized proteins. These apparently essential proteins remain to be studied and should help to understand the physiology and genetics of the respective genetic elements as well as the host. *Sulfolobus* is one of the best-studied archaeons and could develop into an important model organism of the crenarchaea and the archaea.

Key words Plasmid · *Sulfolobus* · pRN1 · pNOB8 · SIRV2 · SIFV

Introduction

Three types of genetic elements are discussed in this review: Viruses, cryptic plasmids and conjugative plasmids. Viruses from *Sulfolobus* have been reviewed (Prangishvili et al. 2001; Snyder et al. 2003; Prangishvili and Garrett 2004), latest developments will be reported here. The research on cryptic plasmids has focussed on the plasmid pRN1 from *Sulfolobus islandicus*, which could serve as a model for several other cryptic plasmids of the pRN plasmid family. The proteins encoded by the plasmid pRN1 have been studied biochemically allowing to set up a model how pRN1 is replicated. *Sulfolobus* is also a host for conjugative plasmids; however, these plasmids have been less well studied. In the last section the latest developments in the field of shuttle vector construction are reported. Although *Sulfolobus* is one of the best-studied archaeons, genetic studies with *Sulfolobus* are still at its infancy.

Sulfolobus spp. belong to the *Crenarchaeota*, one of the two major branches of the archaea. The genus *Sulfolobus* was first described by Brock et al. (1972). He isolated several *Sulfolobus* strains from acidic thermal habitats in Yellowstone National Park, El Salvador, Italy and Dominica. *Sulfolobus* strains are thermoacidophiles with an optimal growth temperature above 70°C and an optimal pH of 3. *Sulfolobus* species are aerobes, which can grow chemolitho(auto)trophically by oxidizing elementary sulphur or hydrogen sulphide. They can also grow heterotrophically. Although being extremophiles, *Sulfolobus* strains can be relatively easily grown in the laboratory without the need of specialized equipment. Their doubling time in liquid complex media is a few hours. They can also be cultivated on plates prepared with Gelrite gum, a polysaccharide that does not melt at 80°C in the presence of divalent cations.

The physiology of *Sulfolobus* has been studied in more detail than that of many other members of the archaea (Schafer 1996; Gao et al. 1998; Lower et al. 2000; Hjort and Bernander 2001; Reilly and Grogan 2001; She et al. 2001a). Studies on transcription (Qureshi et al. 1997; Bell et al. 1998; Tang et al. 2005), translation (Cammarano et al. 1985; Condo et al. 1999), replication (Pisani et al. 2000; Dionne et al. 2003; Lao-Sirieix and Bell 2004; Lundgren et al. 2004; Pucci et al. 2004; Robinson et al. 2004; Savino et al. 2004) and repair (Constantinesco et al. 2002; Roberts et al. 2003; Constantinesco et al. 2004; Roberts and White 2005) in *Sulfolobus* have deepened our understanding of information processing in the archaea. For biochemical studies the thermophilic character of *Sulfolobus* is yet another advantage. Proteins from thermophiles are usually resistant to heat denaturation and to proteolysis simplifying the purification of recombinant proteins

Communicated by D. A. Cowan

G. Lipps
Institute of Biochemistry, University of Bayreuth,
Universitätstrasse 30, 95440 Bayreuth, Germany
E-mail: georg.lipps@uni-bayreuth.de
Tel.: +49-921-552419
Fax: +49-921-552432

produced in mesophilic hosts. Their stability and conformational rigidity is often advantageous for crystallographic studies. The high number of genomes already sequenced also reflects the great interest in *Sulfolobus* (Table 1). The genomes of *S. solfataricus* P2 (3.0 Mb), *Sulfolobus tokodaii* str. 7 (2.7 Mb) and *Sulfolobus acidocaldarius* (2.3 Mb) have been completely sequenced (Kawarabayasi et al. 2001; She et al. 2001b; Chen et al. 2005), and the sequencing of *S. islandicus* strain HVE10/4 is under way.

The ease of cultivation, its thermophilic character, the availability of several genome sequences and the already accumulated data on *Sulfolobus* strains indicate that *Sulfolobus* could develop into a model organism of the *Crenarchaeota* and possibly the archaea on the whole. However, the genetic methods for studying *Sulfolobus* strains are not that advanced as for the halophilic and methanogenic *Euryarchaeota*, e.g. *Haloferax* and *Methanosarcina*. The research on the genetic elements of *Sulfolobus* could therefore provide the scientific grounds to advance the genetic studies with *Sulfolobus*.

Sulfolobus is a host for several novel viruses

The viruses of *Sulfolobus* belong to four families: *Fuselloviridae* (SSV1, SSV2 and SSV RH and SSV K1), *Rudiviridae* (SIRV1 and SIRV2), *Lipothrixviridae* (SIFV) and *Guttaviridae* (SNDV). In addition a recently discovered icosahedral virus from the Yellowstone National Park could not be assigned to a virus family (Rice et al. 2004). All *Sulfolobus* viruses described so far are non-lytic. Therefore, the viruses do not need to resist to an acidic and hot environment for longer periods.

Sulfolobus shibatae virus 1: the best-studied Fuselloviridae

The first member of this new family was the virus SSV1 (*Sulfolobus shibatae* virus 1) which was isolated from a *Sulfolobus shibatae* strain in Beppu, Japan (Palm et al. 1991). This virus is the best-studied virus of the genus *Sulfolobus* and can serve as a model for the three related viruses SSV2 (14.8 kb), a fusellovirus isolate from Kamchatka (SSV K1, 17.4 kb) and an isolate from Yellowstone National Park (SSV RH, 16.5 kb) (Wiedenheft et al. 2004). The *Fuselloviridae* are lemon shaped particles of 60 × 100 nm size (Zillig et al. 1996). Recently, a fusellovirus was also found in a *Pyrococcus abyssi* strain (PAV1) (Geslin et al. 2003). Moreover a fusellovirus-like particle was observed for the halophile archaeon *Haloarcula hispanica* (Bath and Dyall-Smith 1998) and for the euryarchaeote *Methanococcus voltae* A3 (Wood et al. 1989).

The SSV1 genome of 15.5 kb is maintained as circular extrachromosomal element in the cell as well as site-specifically integrated into the host arginyl–tRNA gene. SSV1 virus production can be induced by exposure to UV light and mitomycin C. The induction is mediated by the transcription at the promoter T_{ind}; however, details on the activation mechanism have not been elucidated. The 11 transcripts of the virus genome have been mapped and led the basic for an early definition of the consensus sequences for archaeal promoters and terminators (Reiter et al. 1987, 1988). Pure viral SSV1 DNA was also used to develop an electrotransformation procedure for *Sulfolobus* (Schleper et al. 1992).

The SSV1 genome has 34 open reading frames (ORFs), which are tightly packed in the genome (Fig. 1). Eighteen genes code for proteins conserved within the *Fuselloviridae* (Wiedenheft et al. 2004). However, only four genes of SSV1 could be assigned a function up to now. Three structural proteins could be identified biochemically through protein sequencing of purified virus particles (Palm et al. 1991). Remarkably only for a single gene a functional assignment could be made based on sequence homology to known proteins. This gene encodes a type I tyrosine recombinase, which catalyses the site-specific integration of the viral genome into an arginyl–tRNA gene of the host. Following integration into the anticodon loop of the arginyl–tRNA, the arginyl–tRNA gene is restored by a complementing sequence of

Table 1 Genomes of some viruses and plasmids of *Sulfolobus*

Element	Size	Predicted ORFs	Conserved hypothetical proteins	Functionally assigned genes	Characterized proteins
SSV1 (*Fuselloviridae*)	15.5 kb	34	18	6	Integrase, adaptor protein D-63, winged-helix protein F-93
SIRV2 (*Rudiviridae*)	35.5 kb	54	44	5	dUTPase, Hjc resolvase
SIFV (*Lipothirxviridae*)	41 kb	74	12 (SIRV1/2), 7 (AFV1)	7	
SNDV (*Guttaviridae*)	ca 20 kb				
SITV	17.6 kb	36	3	1	
pRN1 (cryptic)	5,350 b	6	3	3	CopG, DNA-binding protein, replication protein
pNOB8 (conjugative)	41.2 kb	46	13	11	

The genomes of the genetic elements are densely packed with ORFs. For most of the genes a functional assignment was not possible and only a few gene products have been characterized biochemically. A comparative analysis of the genomes allows to identify highly conserved ORFs. This table clearly shows the large discrepancy between the exact knowledge of the genome sequences on the one hand and the lack of characterized gene products on the other hand

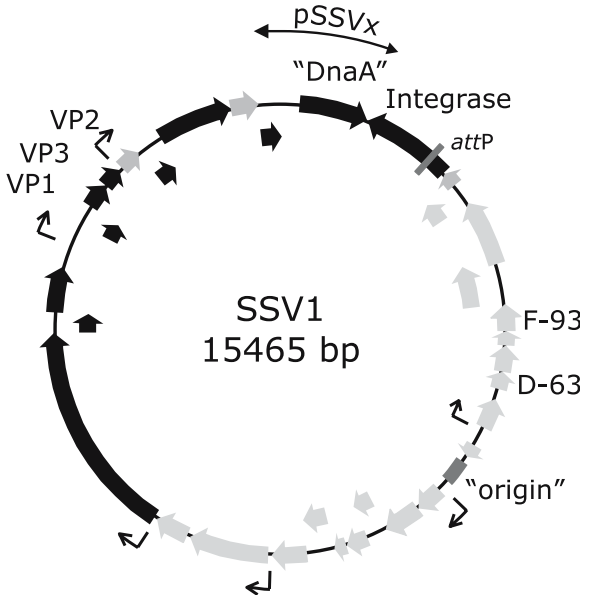

Fig. 1 Genome of the Fuselloviridae SSV1. *Black arrows* depict ORFs, which are conserved in all four sequenced *Fuselloviridae* from *Sulfolobus*. VP1-3 are structural proteins of the virus. The protein VP2 is specific for SSV1. The adaptor protein D-63 and the winged-helix protein F-93 are labelled. The integrase is the only protein, for which a function could be derived, based on sequence alignments. The gene contains the viral attachment site *attP*. The putative homologue of DnaA as well as the putative origin of replication are indicated. Transcription start sites are depicted as *thin arrows*. The two conserved genes, which are also found on the plasmid–virus hybrid pSSVx, are *highlighted*. Small ORFs are not drawn to scale

the viral attachment site but the integrase gene is interrupted. It is not known whether the gene products of the interrupted gene can catalyse the excision of the viral genome. All *Fuselloviridae* characterized so far integrate site-specifically into tRNA genes possibly because the tRNA genes are not susceptible to genomic drift. The integrase is the only protein of SSV1 that has been studied biochemically in more detail. The recombinant protein has sequence-specific endonuclease activity as well as ligase activity (Muskhelishvili et al. 1993; Serre et al. 2002). Both activities are compatible with the integration process.

Recently the crystal structures of the 63 amino acid protein D-63 and the 93 amino acid protein F-93 from SSV1 have been determined. The former protein is dimeric and has homologues in SSV2 and SSV RH. It forms a four-helix bundle and is believed to participate in protein–protein interaction (Kraft et al. 2004a). Neither its functional role nor its protein interaction partners could be derived from the protein structure. The protein F-93 is conserved in SSV K1 and SSV2 and is also a dimeric protein. The protein adopts a winged-helix DNA-binding fold suggesting that it could function as a transcription factor. Its DNA target site is currently unknown (Kraft et al. 2004b).

SSV1 also exists in the cell in an episomal state. The origin of replication has not been defined but indirect evidence suggests that the replication origin is located around the putative ORF b-49 (Palm et al. 1991; Cannio et al. 1998). Furthermore, a DnaA homologue has been predicted (Koonin 1992). One further aspect is remarkable considering the genome organization: genes encoding for cysteine-free genes are clustered in one section of the genome and led to the speculation that SSV1 arouse from an early viral genome fusion event.

The plasmid–virus hybrid pSSVx is another example for a fusion event (Arnold et al. 1999). The genome of pSSVx contains three highly conserved ORFs of the pRN plasmid family (see next section) and two conserved ORFs from the *Fuselloviridiae* (see Fig. 1). It is believed that pSSVx replicates in the host cell as plasmid. Upon superinfection with the viruses SSV1 or SSV2 infectious virus particles of two different sizes are released. The smaller virus particle probably contains the smaller genome of pSSVx whereas the larger particle contains either the viral genome of SSV1 or SSV2.

Rudiviridae

Rudiviridae is a viral group specific for the genus *Sulfolobus*. Only two members have been described and were named SIRV1 and SIRV2 for *S. islandicus* rod shaped virus. These virus particles are stiff rods with a diameter of 23 nm and a length of about 800–900 nm. The virus particles have no envelope. At the ends of the rods tail fibres of 28 nm length are found. The linear double-stranded DNA of the virus is compacted in a superhelical conformation by a 16 kDa structural DNA-binding protein (Prangishvili et al. 1999).

The genomes of both *Rudiviridae* are very similar; however, 13 ORFs are unique to SIRV1 and three ORFs of SIRV2 are not found in SIRV1. Whereas the virus SIRV2 is stably propagated, the virus SIRV1 forms variants. Depending on the host strain particular variants dominate the variant population. These variants differ in gene size, content and order. At least two mechanisms appear to be responsible for the generation of variants: small 12 bp long introns as well as recombination events (Peng et al. 2004).

The genome of SIRV2 has a size of 35.5 kb and contains 54 ORFs. Most of the ORFs are not similar to characterized proteins. However, about a third of the genes have homologues in the genome of another *Sulfolobus* virus (SIFV, see next section). In addition several ORFs have sequence similarity to eukaryal viruses, including poxviruses and Chlorella virus (Peng et al. 2001).

Five genes could be assigned a function: a structural DNA-binding 16 kDa protein, which was identified by amino acid sequencing (Prangishvili et al. 1999), a dUTPase, which was expressed in *Escherichia coli* and characterized biochemically (Prangishvili et al. 1998b), a Holiday junction resolvase, whose resolvase activity on cruciform DNA was analysed in vitro (Birkenbihl et al. 2001) as well as two putative glycosyl transferases. The

Rudiviridae have an extremely low GC content of only 25%. The physiological role of the viral-encoded dUTPase could be to lower the intracellular concentration of dUTP to prevent incorporation into the viral genomes during replication. The ends of the linear DNA are covalently closed (Blum et al. 2001). Extensive repeat structures, e.g. 1.6 kb for SIRV2, are found at both termini, and head-to-head as well as tail-to-tail replication intermediates have been identified. Possibly the replication of the viruses is initiated by nicking of one strand at the covalently closed termini. The viruses are maintained at a copy number of about 20 in the cell and do not integrate into the host genome (Prangishvili et al. 1999).

The transcripts of SIRV1 and SIRV2 have been mapped (Kessler et al. 2004). In both viruses transcription starts immediately after infection (30 min) at multiple starts sites. Nearly all viral genes are simultaneously expressed and there seems to be little temporal regulation on the transcription. The viral promoters contain a characteristic GTC motif just downstream of the putative TATA boxes. It has been suggested that this motif could be a *cis*-regulatory element (Kessler et al. 2004).

Lipothrixviridae

Viruses from the *Lipothrixviridae* family are propagated from several crenarchaeal species: *Sulfolobus* (SIFV), *Acidianus* (AFV and DAFS) and *Thermoproteus* (TTV1, TTV2 and TTV3) (Arnold et al. 2000b; Bettstetter et al. 2003; Snyder et al. 2003). These viruses have a lipid envelope. The *Sulfolobus islandicus filamentous virus* (SIFV) forms long flexible virus particles with an elliptical cross-section with a diameter of 16–24 nm. The filaments are 2 μm long and possess six tail fibres each at their ends. The virus core contains two major basic proteins, which seem to be involved in packing the linear double-stranded DNA of the virus. The N-termini of both structural proteins were sequenced and their genes, which form a transcription unit, could be identified. The virus core is surrounded by a 4 nm lipid envelope, which is partly derived from host phospholipids. SIFV is non-lytic and does not integrate into the host genome. The viral genome has been sequenced except both termini, which could not be subcloned since they are modified in an unknown manner. The 41 kb genome is tightly packed with 74 ORFs. Two genes encode putative helicases of superfamily 2, further three viral proteins have glycosyl transferase motifs, and could be involved in modifying host lipids with activated sugar compounds (Arnold et al. 2000b).

Guttaviridae

The virus SNDV (*Sulfolobus neozealandicus droplet-shaped virus*) is so far the only representative of the provisional virus family *Guttaviridae*. The virus infects *Sulfolobus* strains from New Zealand but not from Iceland. The infection is non-lytic and unstable as strains are easily cured. The virus particles are droplet-shaped with a length of 110–185 nm and a width of 70–95 nm. A beard of thin tail fibres is observed at their pointed ends. The viral genome, which is maintained in an episomal state, is double-stranded circular DNA and has a size of 20 kb. Viral DNA isolated from virus particles is methylated at the *dam*-sites most probably by a virally encoded methylase (Arnold et al. 2000a).

The *Sulfolobus* turreted icosahedral virus

The STIV (*Sulfolobus* turreted icosahedral) virus was isolated from an acidic hot spring in Yellowstone National Park and could be propagated in *Sulfolobus solfataricus* strains P1 and P2. The virus particles have a diameter of 74 nm. In addition, 12 turret-like projections extend 13 nm above the viral surface. The STIV has a double-stranded DNA genome with a size of 17.6 kb. Thirty-six ORFs were assigned but only three ORFs show similarity to other proteins in the databases. The gene of the major capsid protein, which has a mass of 37 kDa, could be identified. This protein is believed to adopt a similar fold as the major capsid proteins of adenoviruses (Rice et al. 2004).

Plasmids

Two types of plasmids have been reported for the genus *Sulfolobus*: rather small cryptic plasmids with genome sizes of 5–14 kb and a family of conjugative plasmids with genomes larger than 25 kb. The first *Sulfolobus* plasmids were discovered by Zillig and co-workers by a systematic screening for genetic elements in geothermal waters in Iceland (Zillig et al. 1994).

Cryptic plasmids

The first cryptic plasmid completely sequenced was the plasmid pRN1 which was isolated from the *S. islandicus* strain REN1H1 (Keeling et al. 1996). This strain also harbours the plasmid pRN2 and it could be shown that both plasmids replicate independently (Purschke and Schaefer 2001). In the following years further plasmids from Iceland were sequenced: pRN2 and pHEN7 as well as the virus–plasmid hybrid pSSVx. Strikingly all these plasmids belong to a single plasmid family which was termed pRN (Peng et al. 2000). Another member of the family is the plasmid pDL10, which has been isolated from the crenarchaeote *Acidianus ambivalens*. The plasmids of the plasmid family share three highly conserved genes, whose proteins have been characterized biochemically (see below). Plasmids of the pRN-type were also found integrated in the genome of *S. solfataricus* and *S. tokodaii* (She et al. 2001a, 2002, 2004). The integration of these plasmids were probably mediated by

a site-specific integrase since the genes for the integrase and the attachments sites could be detected. In the case of the integrated plasmid pXQ1, the integration event into the *S. solfataricus* genome leads to the disruption of the integrase gene as is the case for the virus SSV1 (see above). Excision of the plasmid is therefore unlikely and no free plasmid nor an empty integration site could be detected by PCR (She et al. 2001a). The genomic attachment sites for pXQ1 (*S. solfataricus*), pST1 and pST3 (*S. tokodaii*) are inside tRNA genes. The latter element has its plasmidal attachment side outside the integrase gene, therefore the integrase gene remained intact after integration and could in principle catalyse the excision of the plasmid from the genome (She et al. 2002). The same type of non-partioning integrases is also encoded on conjugative plasmids (see below).

With the sequencing of plasmids from other geographic regions the diversity of cryptic *Sulfolobus* plasmids increased. The plasmid pIT3 (AY591755.1) isolated from a *S. solfataricus* strain from Italy and the plasmids pTAU4, pORA1 and pTIK4 obtained from *Sulfolobus* strains from New Zealand (Greve et al. 2004b) are different from the pRN family plasmids. These plasmids have only some features with the pRN-type plasmids in common (see below).

The fundamental property of all pRN plasmids is that they share three conserved ORFs (Fig. 2). These genes are termed in this review according to the gene names of plasmid pRN1, i.e. *orf56*, *orf80* and *orf904*. Up to now no pRN plasmid has been discovered outside the crenarchaeal phylum. Moreover, plasmids of the pRN family are not related to any other known eubacterial or archaeal plasmid.

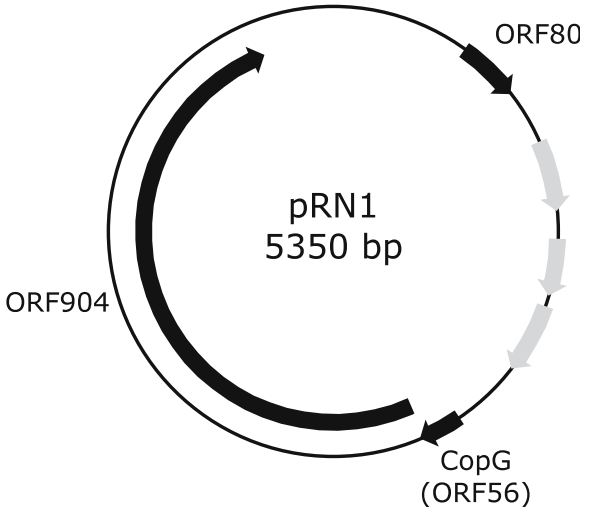

Fig. 2 Genome of the cryptic plasmid pRN1. The three highly conserved ORFs of the plasmid family pRN are depicted as *black arrows*. The copy-control gene *copG* appears to be cotranscribed with the gene of the replication protein ORF904. The DNA-binding proteins CopG and ORF80 bind within their own promoter region. The exact physiological function of ORF80 is, however, unknown. The replication origin has not been determined

The proteins encoded by the *orf56* homologues contain the DNA-binding domain of the ribbon–helix–helix fold (pfam01402). The best-studied protein of this domain is the CopG protein from pLS1, which participates in the copy-control of the plasmid. CopG is a sequence-specific DNA-binding protein which binds to its own promoter and represses its own expression as well as the expression of the following replication protein RepA (del Solar et al. 1990). In addition to transcriptional repression by CopG, a RNA counter-transcript is thought to participate in copy-control as well (Rasooly and Rasooly 1997). The structure of this protein repressor, which is the shortest reported to date, has a homodimeric ribbon–helix–turn–helix fold. The helix–turn–helix structure is involved in dimerization and not DNA binding. The DNA-binding interface of these proteins is an anti-parallel β-sheet that is formed from both subunits of the dimeric proteins.

The ORF56 protein was purified from recombinant *E. coli* cells and characterized. The highly basic protein is a 56 amino acid protein with a molecular mass of 6.5 kDa (Lipps et al. 2001b). It could be shown that the ORF56 protein is a sequence-specific double-stranded DNA-binding protein, which binds with high affinity and moderate specificity to an inverted repeat upstream of its own start codon. The protein is dimeric in solution and assembles into a tetramer on the cognate DNA. Primer extension analysis of RNA isolated from *S. islandicus* REN1H1 cells containing the plasmid pRN1 revealed that there is a transcriptional start site 33 bases upstream of the start codon (Lipps, unpublished). The inverted repeat of the ORF56 binding site is located within these 33 bases, suggesting that ORF56 acts as a repressor for transcription initiation upstream of *orf56*. The gene downstream of *orf56*, *orf904*, could form a transcription unit with *orf56*. As suggested by bioinformatic analysis of the *S. solfataricus* genome, translation initiation within an operon requires a ribosome-binding site (Tolstrup et al. 2000). In fact a Shine–Dalgarno sequence could be identified upstream of a putative start codon of *orf904* further supporting that *orf56* and *orf904* are transcribed in a bicistronic message. The physiological function of ORF56 could be to down-regulate the expression of the genes *orf56* and *orf904*, the latter coding for the replication protein of the plasmid (see below). By this mechanism, the initiation of the plasmid replication could be limited leading to copy number control.

Preliminary structure analysis of the recombinant protein by multidimensional NMR suggests that the ORF56 protein adopts a ribbon–helix–turn–helix structure (Lipps, unpublished). This succession of secondary structure elements is also found in the CopG protein of the plasmid pLS1. Since the ORF56 homologues seem to be structurally similar to the CopG protein and also have DNA-binding activity, which could be involved in plasmidal copy-control, I suggest to rename the respective genes of the crenarchaeal plasmids as *copG*. The cryptic plasmids from Italy (pIT3) and New Zealand

(pTAU4, pORA1 and pTIK4) might also encode CopG proteins and in several instances putative *copG* genes were identified upstream of a large ORF which could be involved in plasmidal replication initiation. But the sequence similarity to the among the CopG proteins is generally very low and the corresponding proteins are often not detected using a standard BLASTP search ($E = 10$).

In contrast, the *orf80* genes of the plasmids are the most conserved genes of the plasmid family pRN and homologues are also found on the cryptic plasmid pTIK4 and pTAU4 but not on the plasmid pORA1 and pIT3. In addition *orf80* homologues are also highly conserved within the conjugative plasmids of *Sulfolobus* (see below). By using the search tool for conserved domains the *orf80* homologue of the plasmid pDL10 produced a weakly significant hit to the winged-helix arsenical resistance repressor domain (conserved domain cd00090). It is currently not known whether the ORF80 proteins belong to the winged-helix DNA-binding proteins. The *orf80* gene of the plasmid pRN1 codes for a basic 9.5 kDa protein. Based on the amino acid sequence the protein can be divided in two parts. The N-terminal 40 amino acids are predicted to have a helical secondary structure and contain several conserved leucines in the correct distances for forming a leucine zipper. The C-terminal 30 amino acids contain a high number of conserved basic amino acids. This part of the protein could therefore fold into a basic DNA-binding domain.

The ORF80 protein from pRN1 has been expressed and purified from *E. coli* (Lipps et al. 2001a). The dimeric protein binds double-stranded DNA in a sequence-specific manner. By DNase I footprinting two binding sites on the plasmid pRN1 were detected. The binding sites are very similar and have the core consensus sequence TTAAnnnnnnnTTAA. Using small synthetic DNA fragments containing the consensus sequence-specific binding to these fragments could be shown by electrophoretic mobility shift assay (EMSA) as well as by fluorescence titrations. With a dissociation constant of 140 nM under physiological salt concentration the binding of the ORF80 protein to DNA is of rather low affinity (Lipps et al. 2001a).

The two DNA-binding sites detected by DNase I footprinting are separated by 65 bp and are found upstream of the *orf80* gene. Most interestingly upstream of the *orf80* homologues ORF80 binding sites could be identified on all pRN plasmids as well as on some conjugative plasmids, and in all cases the binding sites were separated by about 65 bp (see Fig. 5). The conservation of not only the binding site sequence but also the distance of the binding sites points to an important structural or functional role of both binding sites. EMSA experiments indicate that a rather large albeit specific DNA–protein complex is formed on DNA molecules containing both binding sites. In line with this observation a highly cooperative binding was observed with these DNA substrates. It is therefore suggested that ORF80 assembles into a rather large protein–DNA complex on a single site of the plasmid pRN1 (Lipps et al. 2001a). The structural and functional significance of this putative complex is currently unknown. The physiological function of ORF80 remains obscure; nevertheless the *orf80* homologues are partly annotated as *plr*A meaning *pl*asmid regulatory gene A.

The third conserved ORF of the plasmid family pRN, the gene *orf904*, is by far the largest ORF of the plasmid family. The corresponding genes occupy about a third to a half of the plasmids and encode proteins of about 900–1,000 amino acids. Within the C-terminal part of the proteins a helicase domain of superfamily 3 can be identified. The best-studied protein containing this domain is the bacteriophage P4 α protein, which is a multifunctional replication protein harbouring a helicase domain in its C-terminal part (Ziegelin et al. 1993, 1995).

The gene *orf904* from the plasmid pRN1 encodes a 106 kDa protein. The N-terminal 40 amino acids are not found in the corresponding proteins of the other plasmids. Possibly translation starts at the alternative start codons amino acids valine 14 (GTG), valine 22 (GTG) or methionine 24 (ATG). The ORF904 protein was expressed in *E. coli* and purified as N-terminally his-tagged protein. The recombinant protein proved to be a multifunctional protein with helicase, primase and DNA polymerase activity (Lipps et al. 2003). The helicase activity of ORF904 has 3′–5′ polarity (Lipps, unpublished). The ATPase activity of ORF904 is stimulated in the presence of double-stranded DNA. Single-stranded DNA and RNA are only weak activators for the ATPase activity.

Most surprisingly, ORF904 is also able to carry out DNA polymerization in a primer extension assay. The kinetics of DNA polymerization suggests that ORF904 is not a processive DNA polymerase and that its specific activity is rather low. The enzyme does neither have a 3′–5′ (proofreading) nor a 5′–3′ exonuclease activity. In marked contrast to all DNA polymerases ORF904 is able to polymerize DNA de novo, i.e. without a primer. The primase activity of ORF904 prefers deoxynucleotides as precursors about 20-fold over ribonucleotides suggesting that in vivo ORF904 uses deoxynucleotides for priming. The primers synthesized by ORF904 have a length of 8 bases and can be extended by the polymerase activity of ORF904 upon prolonged incubation (Lipps et al. 2003).

From the amino acid sequence no similarity to known DNA polymerases and primases could be detected. The domain structure was investigated by deletion studies, which revealed that the N-terminal 255 amino acids are sufficient to carry out DNA polymerization. To further pin down the active site of polymerization, point mutants of the conserved acidic residues were constructed. Mutant proteins, in which the amino acids aspartate 111, glutamate 113 or aspartate 171 were changed to alanine, are inactive in DNA polymerization and primase activity. These findings strongly suggest that these amino acids are critical for enzymatic catalysis

and that primase and DNA polymerase activity use the same active site (Lipps, unpublished). The corresponding domain of the protein was therefore termed prim/pol domain. The domain could be expressed separately in *E. coli* and the crystal structure could be determined with high resolution (Lipps et al. 2004). The structure consists of a DNA-binding domain and a catalytic subdomain containing the active site residues identified by mutational analysis. The residue aspartate 111 binds a manganese ion suggesting a metal-dependent reaction mechanism as observed for DNA polymerases (Steitz et al. 1994). Surprisingly the prim/pol domain has structural similarity to the catalytic subunit of the archaeal primase from *Pyrococcus furiosus* (Augustin et al. 2001). Both enzymes share a conserved core, which forms the active site. The sequence conservation between both enzymes is very low and cannot be detected by similarity searches. Only four residues of the active site are conserved. In addition six conserved amino acids are found at homologous surface positions outside the active site. The structural similarity could be a result of convergent evolution. Alternatively both enzymes could have evolved from an ancestral enzyme, which could have been a simple DNA replicase. Notably both enzymes, the replication protein and the pyrococcal primase, are able to polymerize DNA de novo, a realistic property at the advent of DNA replication.

The prim/pol domain is also found on the cryptic plasmids pORA1 and pIT3. On both plasmids this domain is the N-terminal part of a large protein of more than 800 amino acids. The remainder of the protein is not clearly related to the helicase domain of ORF904. However, the AAA+ ATPase region comprising the Walker motifs, which are also found in helicases of superfamily 3, could be detected on the putative replication proteins (see Fig. 3). It is noteworthy that the cryptic plasmid pTAU4 has a distinct type of replication protein. This plasmids codes for a mini-chromosome-maintenance (MCM) protein and does not have an N-terminal prim/pol domain. The MCM-proteins are believed to be the replicative helicases for archaea and eukaryotes. They are related to the superfamily 3 helicases (Koonin 1993) suggesting that all cryptic plasmid sequenced so far require helicase activity for replication. It is currently unknown whether the helicase activity is needed for unwinding the replication origin or for replication fork movement.

The observed biochemical activities of the gene products encoded by pRN1 now allow us to draft a model how pRN1 is replicated. The replication protein ORF904 can be assumed to be central to the plasmidal replication. It is suggested that ORF904 alone or in concert with another protein, possibly ORF80, recognizes the origin of replication. Subsequently, the helicase

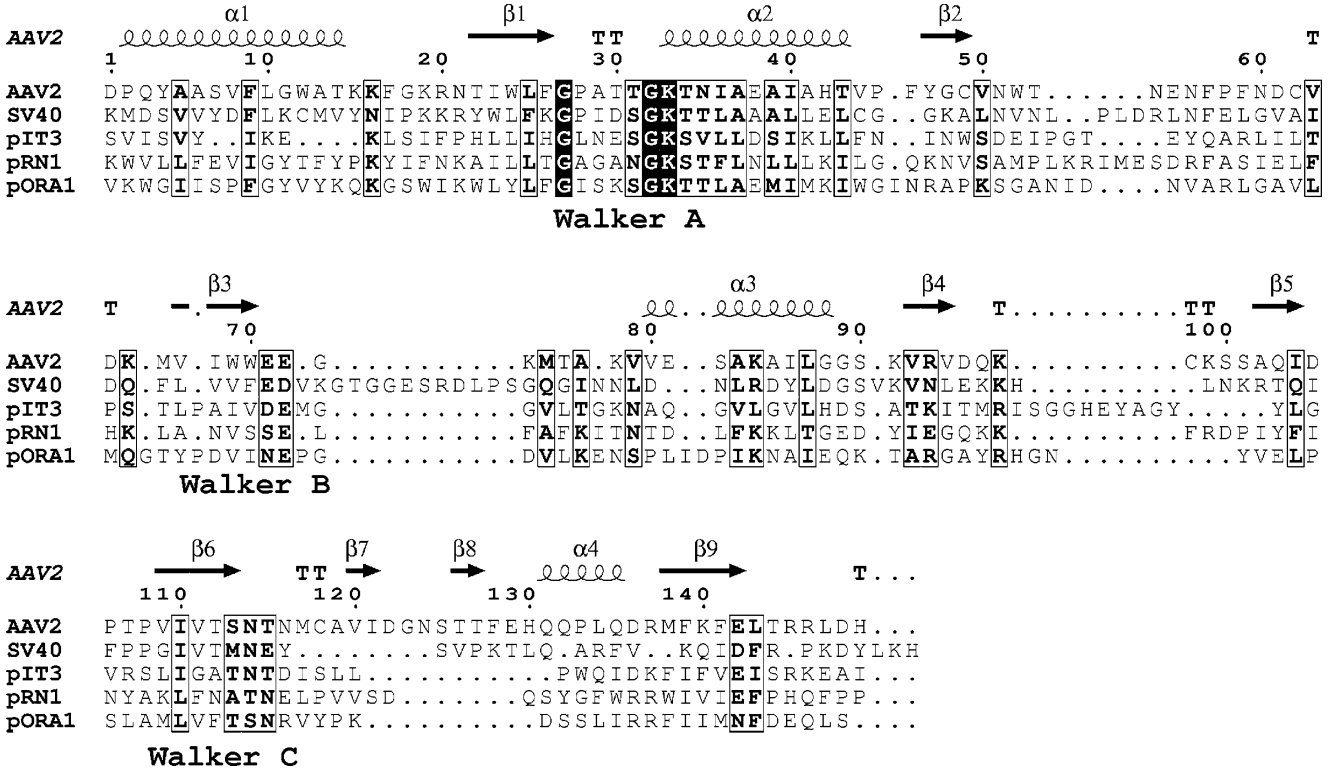

Fig. 3 Alignment of the putative helicase domains of the replication proteins from the cryptic plasmids. The helicase domains of the replication proteins are only very weakly conserved. Aligned are the central part of the proteins based on an structural alignment of two superfamily 3 helicase structures, the helicase of the adeno-associated virus 2 (*AAV2*) and the large T antigen of SV40 (simian virus). The Walker A motif can be clearly detected. In contrast the Walker B and C motifs are only weakly conserved. The secondary structure elements (α α-helix, β β-strand, *T* turn) of the AAV2 structure are indicated above the alignment

domain of ORF904 will melt the replication origin using chemical energy from the hydrolysis of ATP and the prim/pol domain of ORF904 will synthesize a primer. Theoretically ORF904 alone could replicate the complete plasmid. However, low processivity and the lack of proofreading activity suggest that the host replication machinery carries out DNA replication after the primer has been synthesized. The CopG protein (ORF56) could limit the intracellular concentration of the replication protein and thereby down-regulate plasmidal replication initiation. The function of the ORF80 protein is still a mystery, possibly ORF80 has a structural role, e.g. it could mark the replication origin.

Conjugative plasmids

Sulfolobus is also a host for conjugative plasmids. The first conjugative plasmid discovered was the plasmid pNOB8 which was isolated from a Japanese strain of *Sulfolobus* (Schleper et al. 1995). The plasmid is transferred from a donor cell to another cell by extensive cellular contacts. Conjugation is quite efficient and the plasmid is able to spread through the whole culture. The plasmid is replicated in the host cell to a high copy number of about 20–40, leading to growth retardation of the culture. Upon prolonged growth the plasmid undergoes extensive genetic variation and is also cured in many *Sulfolobus* strains. The genome of the plasmid pNOB8 and its major deletion variant pNOB8-33 has been determined (She et al. 1998). The deletion variant probably results from a recombination event at an 85 bp long direct repeat. The deletion variant is less stable at low copy numbers which is explained by the lack of a putative ParA and ParB protein (see Fig. 4) which are lost after recombinational deletion (She et al. 1998). The genetic dynamics of the conjugative plasmids have also been studied using plasmids of the pING family from *S. islandicus* strains (Stedman et al. 2000). The variants pING4 and pING6 are derived from pING1 by integration of genomic IS elements. Two conjugation defective deletion variants pING2 and –3 were also discovered. For pING2 recombination at two inverted repeat TAAACTGGGGAGTTTA within the plasmid pING4 could have caused the deletion. Identical recombination sites were recently identified in other conjugative plasmids (Greve et al. 2004a) and the similar motif TTTACTGGGGAGTAAA is found seven times on the conjugative plasmid pTC. Both putative recombination motifs are a perfect inverted repeat separated by four guanosines.

Conjugative plasmids seem to be widespread in the genus *Sulfolobus*. Screening of 300 isolates from Iceland resulted in the discovery of 11 conjugative plasmids (Prangishvili et al. 1998a). The conjugative plasmids discovered in Iceland are about 25–35 kb in size. The plasmids have similar characteristics as the plasmid pNOB8 including high copy number, growth retardation and high genetic variation. The genome sequence of five

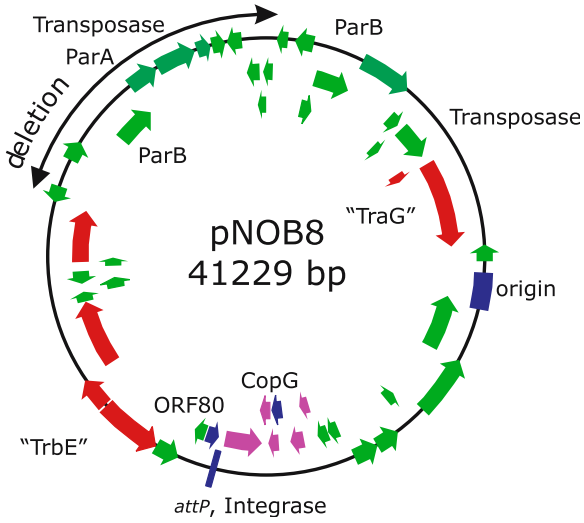

Fig. 4 Genome of the conjugative plasmid pNOB8. The roughly 50 ORFs are colour-coded according to conservation and putative function. Shown in *red* are the six highly conserved genes of the conjugative plasmids with a suggested function in the conjugation process, in *violet* are five conserved genes which are believed to participate in replication initiation or maintenance. The ORFs from the *copG* and the *orf80/plrA* gene are shown in *blue*. Homologues of these genes are also found on cryptic plasmids and viruses. ORFs were labelled when their amino acid sequence produces a significant match to a characterized protein. On the plasmid pNOB8, two transposase genes and three genes which appear to be involved in plasmid partitioning (ParA, ParB) are found. These genes are not present on the other sequenced conjugative plasmids. Parts of two ORFs are similar to the bacterial proteins TrbE and TraG which are involved in type IV DNA transport and are accordingly labelled 'TrBE' and 'TraG'. The conjugative plasmid also encodes a tyrosine integrase. The attachment site *attP* is found upstream. The putative replication origin is shown in *blue* and the major recombinational deletion leading to the variant pNOB8-33 is indicated by an *arrow*. Small ORFs are not drawn to scale

conjugative plasmids from Iceland (pING1, pKEF9, pHVE14, pARN3 and pARN4) have been determined and analysed (Greve et al. 2004a). Moreover, the sequence of a conjugative plasmid from China, named pTC, has been published (NC_005969.1).

Comparisons of the genome sequences of the conjugative plasmids sequenced allow to delineate the essential genes of this plasmid family (Greve et al. 2004a). The highly conserved genes are clustered in two conserved sections (Fig. 4). One section contains six genes and it is proposed that this section encodes the core of the archaeal conjugation apparatus. While none of the proteins have been studied biochemically yet, the amino acid sequence of two proteins resembles partly the bacterial proteins TraG and TrbE, two proteins which participate in the transport of single-stranded DNA across bacterial membranes. In both archaeal proteins Walker A and B motifs could be detected suggesting that both proteins can hydrolyse nucleotides and provide chemical energy for the transport process. A further protein in this section contains a large number of putative transmembrane helices. This protein could be implicated in forming a pore across the archaeal membrane. All theses genes are

not only conserved in the isolates from Iceland but are also found in the Japanese isolate pNOB8 and in the Chinese conjugative plasmid pTC underpinning the ubiquitous conservation of the whole section and suggesting that this section forms a minimal functional unit for DNA transport across membranes. These genes could make up the essential core of the archaeal conjugation apparatus, which appears to be substantially different from bacterial conjugation.

A second section contains seven conserved genes. For this section of the genome the assignment of specific function to the conserved genes remain difficult. Of the seven genes only four genes, namely the *copG*, *orf80/plrA*, the integrase gene and the first gene of the putative operon could also be detected on the recently sequenced Chinese isolate pTC. The genes of this section might participate in replication initiation. One further protein resembles a putative plasmidal replication initiation protein and another protein might contain a leucine zipper (Greve et al. 2004a). A protein with sequence similarity to the CopG DNA-binding protein is found in this section. Interestingly, the CopG proteins are also conserved in the cryptic plasmids of *Sulfolobus* (see above). Therefore, the CopG proteins could participate in regulating the replication initiation of the cryptic as well as the conjugative plasmids. Interestingly, CopG homologues are also found in the viral genomes from AFV1, SITV, SIRV1 and SIRV2. Furthermore homologues of the ORF80 protein from the cryptic plasmid pRN1 can be detected in all conjugative plasmids. It seems that this protein carries out an important function but its precise physiological role remains unknown (see above). Putative ORF80 binding site could be detected on some of the conjugative plasmids (see Fig. 5). The hallmark of the ORF80 binding sites of the pRN plasmid family is their palindromic character and the fact that two binding sites are separated by about 65 bases. These characteristics are also valid for the proposed binding sites of the conjugative plasmids.

The only gene for which a specific function could be derived in this section is a non-partitioning integrase. In the genomes of *S. tokodaii* and *S. acidocaldarius* copies of conjugative plasmids were detected suggesting that the integrase could have integrated the plasmids in the genome of the hosts (Greve et al. 2004a). The presence of a conjugative plasmid in the genome of *S. acidocaldarius* could supply the host with the conjugation machinery and might explain the conjugational marker transfer observed for this organism (Aagaard et al. 1995; Grogan 1996; Reilly and Grogan 2001). Moreover, the integration and excision of pNOB8 in two glutamyl–tRNA genes of *S. solfataricus* has been recently described (She et al. 2004) suggesting that the integrase gene of pNOB8 is functional.

Although there is no experimental evidence for the replication origin of the conjugative plasmids, Greve et al. detected a conserved region which is highly enriched in repeats and is AT-rich (Greve et al. 2004a). Moreover, nucleotide skew analysis supports a replication origin in this region. The putative replication origin which is downstream of the section for the putative genes for archaeal conjugation (Fig. 4) is conserved in the pING deletion variants pING2 and pING3 the latter comprising only 6 kbp.

Viruses and plasmids used as genetic tools

Although quite a large number of genetic elements have been characterized for *Sulfolobus*, they have not been used extensively to conduct genetic studies. Several groups have developed shuttle vectors that can be propagated in *E. coli* and *Sulfolobus* strains.

One shuttle vector was constructed by cloning the *Pyrococcus* rolling-circle plasmid pGT5 (Erauso et al. 1996) and the alcohol dehydrogenase gene from *S. solfataricus* into pBR322. This vector is stably maintained at a low copy number in *E. coli* (under ampicillin selection) and in *S. acidocaldarius* and *P. furiosus* in the presence of butanol or benzyl alcohol (Aravalli and Garrett 1997).

The shuttle vector pEXSs is based on a heat-stable mutant of the hygromycin phosphotransferase, which was generated by error-prone PCR for this purpose (Cannio et al. 1998). The antibiotic hygromycin appears to be suitable as a positive selection marker, since this antibiotic, unlike most antibiotics, is not rapidly degraded under the acidic conditions and the elevated temperatures. To ensure replication in *Sulfolobus* the plasmid contains the putative origin of replication of the virus SSV1. The plasmid is maintained at about one to two copies per cell (Cannio et al. 1998). Recently this shuttle vector was used to express heterologously an

```
pRN1      (419) aggcaaacagtaTTAAtaaagcgTTAAtcctacctcc.30.caaaagatatTTAAcagtctgTTAAtcctactttacATG
pTC      (3325) atgagaagagcaTTACtaattgaGTAAacctactccc.30.cgaagggacaTTACtaaataaGTAAacctaat.25.GTG
pARN4   (24116) atggaaataaggTTAAagaaattTTAAccctgaaccc.30.tttaagatagTTAAagattttTTAAtcttactttacgca
pKEF9   (26369) atggaaataataTTACcattttaGTAAacctactgga.30.ggtaatactaTTACgacattaGTAAacctaaa.27.ATG
pARN3   (24382) atggaaataataTTACcattttaGTAAacctactgga 30.ggtaatactaTTACgacattaGTAAacctatt.27.ATG
```

Fig. 5 Putative binding sites of the ORF80 homologues. Aligned are the putative binding sites (*underlined*) of ORF80 homologues from the conjugative plasmids pTC, pARN4, pKEF9 and pARN3. In the *first line* the experimentally determined binding sites for the plasmid pRN1 are shown. The binding sites are upstream of the *orf80/plrA* genes (start codon in capitals) except for the plasmid pARN4. Here the start codon is situated between both binding sites. A single binding site consists of an eight base palindrome (*bold letters*) separated by seven bases. The distance between both binding sites is 65 bp. The functional significance for the invariant spacing of the sites is unknown

alcohol dehydrogenase gene in *Sulfolobus* (Contursi et al. 2003) and to complement a genomic β-galactosidase deletion (Bartolucci et al. 2003).

Shuttle vectors based on the complete SSV1 genome have also been presented. With an elegant strategy, singly cut SSV1 DNA was produced and ligated into a pBluescript vector. Among 21 possible integration sites, only a single integration site appeared to be tolerated and produced viable viruses after transformation in *S. solfataricus* (Stedman et al. 1999). The copy number of the shuttle vector termed pKMSD48 was high with about 20–40 copies per cell. Since the virus spreads through the culture by infection there is no need for a selective marker.

Plasmids based on the virus SSV1 have also been constructed by Schleper and co-workers. The vector pMJ03 is the most versatile vector of their constructs. This plasmid contains the β-galactosidase gene *lacS* under the control of a heat shock promoter and the genes *pyrEF* enabling selection using uracil-free growth media with suitable *Sulfolobus* strains (Jonuscheit et al. 2003). In contrast to the plasmid pKMSD48, pMJ03 is solely maintained as an integrated copy in the *Sulfolobus* chromosome. The vector is very stable under selective conditions, i.e. growth without uracil. The vector can be used for reporter gene studies and for the expression of proteins in *Sulfolobus* since the heat shock promoter is strong and inducible (Jonuscheit et al. 2003).

A shuttle vector based on a chromosomal origin of replication (*ori*C) was recently described (Contursi et al. 2004). In this work a 1.1 kb intergenic region between the *S. solfataricus cdc6* gene and *mcm* gene was identified as a replication origin. Although this replication origin is different from the replication origins determined by others (Contursi et al. 2004; Lundgren et al. 2004; Robinson et al. 2004), autonomous replication was observed for the plasmid. The pUC-based plasmid contained the intergenic region as replication origin and the hygromycin phosphotransferase gene as selection marker for *Sulfolobus*. The vector was stably maintained in *S. solfataricus* and had a copy number around one. No recombination or rearrangements have been observed (Contursi et al. 2004).

All shuttle vectors constructed so far have not been extensively used by other laboratories. Clearly, further improvements are required and will help that the vectors are used more avidly by the scientific community. Together with the recent development of a homologous recombination method for *Sulfolobus* (Worthington et al. 2003; Schelert et al. 2004) and the availability of microarrays (Lundgren et al. 2004) it is expected that genetic studies with *Sulfolobus* will provide new insights into the biology and physiology of *Sulfolobus*.

Concluding remarks

In the past years quite a number of viruses and plasmids of *Sulfolobus* have been discovered. The viruses of *Sulfolobus* belong to five novel virus families. The genomes of viruses and plasmids encode a large number of conserved proteins whose functions still remain to be unveiled. The discovery of a novel replication protein on the cryptic plasmid pRN1 and the delineation of an archaeal conjugation apparatus should stimulate the research on the orphan genes in this area and shall also be highly rewarding.

Acknowledgments I thank G. Krauss, D. Prangishvili and S. Berkner for critically reading the manuscript and R. Garrett for the sequences of the plasmids pTIK4 and pTAU4 prior to release. I also thank the anonymous reviewers for their helpful suggestions. My research is supported by the Deutsche Forschungsgemeinschaft.

References

Aagaard C, Dalgaard JZ, Garrett RA (1995) Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron-cells of *Sulfolobus acidocaldarius*. Proc Natl Acad Sci USA 92:12285–12289

Aravalli RN, Garrett RA (1997) Shuttle vectors for hyperthermophilic archaea. Extremophiles 1:183–191

Arnold HP, She Q, Phan H, Stedman K, Prangishvili D, Holz I, Kristjansson JK, Garrett R, Zillig W (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon Sulfolobus is a hybrid between a plasmid and a virus. Mol Microbiol 34:217–226

Arnold HP, Ziese U, Zillig W (2000a) SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon Sulfolobus. Virology 272:409–416

Arnold HP, Zillig W, Ziese U, Holz I, Crosby M, Utterback T, Weidmann JF, Kristjanson JK, Klenk HP, Nelson KE, Fraser CM (2000b) A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon Sulfolobus. Virology 267:252–266

Augustin MA, Huber R, Kaiser JT (2001) Crystal structure of a DNA-dependent RNA polymerase (DNA primase). Nat Struct Biol 8:57–61

Bartolucci S, Rossi M, Cannio R (2003) Characterization and functional complementation of a nonlethal deletion in the chromosome of a beta-glycosidase mutant of *Sulfolobus solfataricus*. J Bacteriol 185:3948–3957

Bath C, Dyall-Smith ML (1998) His1, an archaeal virus of the Fuselloviridae family that infects *Haloarcula hispanica*. J Virol 72:9392–9395

Bell SD, Jaxel C, Nadal M, Kosa PF, Jackson SP (1998) Temperature, template topology, and factor requirements of archaeal transcription. Proc Natl Acad Sci USA 95:15218–15222

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

Birkenbihl RP, Neef K, Prangishvili D, Kemper B (2001) Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV2. J Mol Biol 309:1067–1076

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 281:6–9

Brock TD, Brock KM, Belly RT, Weiss RL (1972) Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Microbiol 84:54–68

Cammarano P, Teichner A, Londei P, Acca M, Nicolaus B, Sanz JL, Amils R (1985) Insensitivity of archaebacterial ribosomes to protein synthesis inhibitors. Evolutionary implications. EMBO J 4:811–816

Cannio R, Contursi P, Rossi M, Bartolucci S (1998) An autonomously replicating transforming vector for *Sulfolobus solfataricus*. J Bacteriol 180:3237–3240

Chen L, Brugger K, Skovgaard M, Redder P, She Q, Torarinsson E, Greve B, Awayez M, Zibat A, Klenk HP, Garrett RA (2005) The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. J Bacteriol 187:4992–4999

Condo I, Ciammaruconi A, Benelli D, Ruggero D, Londei P (1999) Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. Mol Microbiol 34:377–384

Constantinesco F, Forterre P, Elie C (2002) NurA, a novel 5′–3′ nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea. EMBO Rep 3:537–542

Constantinesco F, Forterre P, Koonin EV, Aravind L, Elie C (2004) A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea. Nucleic Acids Res 32:1439–1447

Contursi P, Cannio R, Prato S, Fiorentino G, Rossi M, Bartolucci S (2003) Development of a genetic system for hyperthermophilic Archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. FEMS Microbiol Lett 218:115–120

Contursi P, Pisani FM, Grigoriev A, Cannio R, Bartolucci S, Rossi M (2004) Identification and autonomous replication capability of a chromosomal replication origin from the archaeon *Sulfolobus solfataricus*. Extremophiles 8:385–391

del Solar GH, Perez MJ, Espinosa M (1990) Plasmid pLS1-encoded RepA protein regulates transcription from repAB promoter by binding to a DNA sequence containing a 13-base pair symmetric element. J Biol Chem 265:12569–12575

Dionne I, Nookala RK, Jackson SP, Doherty AJ, Bell SD (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Mol Cell 11:275–282

Erauso G, Marsin S, Benbouzid RN, Baucher MF, Barbeyron T, Zivanovic Y, Prieur D, Forterre P (1996) Sequence of plasmid pGT5 from the archaeon *Pyrococcus abyssi*: evidence for rolling-circle replication in a hyperthermophile. J Bacteriol 178:3232–3237

Gao YG, Su SY, Robinson H, Padmanabhan S, Lim L, McCrary BS, Edmondson SP, Shriver JW, Wang AH (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. Nat Struct Biol 5:782–786

Geslin C, Le Romancer M, Erauso G, Gaillard M, Perrot G, Prieur D (2003) PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "*Pyrococcus abyssi*". J Bacteriol 185:3888–3894

Greve B, Jensen S, Brugger K, Zillig W, Garrett RA (2004a) Genomic comparison of archaeal conjugative plasmids from *Sulfolobus* . Archaea 1:231–239

Greve B, Jensen S, Phan H, Brugger K, She Q, Zillig W, Garrett RA (2004b) Novel RepA-MCM proteins encoded in plasmids pTAU4, pORA1 and pTIK4 from Sulfolobus neozealandicus. Archaea (published online 29.11.04)

Grogan DW (1996) Exchange of genetic markers at extremely high temperatures in the archaeon *Sulfolobus acidocaldarius*. J Bacteriol 178:3207–3211

Hjort K, Bernander R (2001) Cell cycle regulation in the hyperthermophilic crenarchaeon *Sulfolobus acidocaldarius*. Mol Microbiol 40:225–234

Jonuscheit M, Martusewitsch E, Stedman KM, Schleper C (2003) A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. Mol Microbiol 48:1241–1252

Kawarabayasi Y, Hino Y, Horikawa H, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Otsuka R, Nakazawa H, Takamiya M, Kato Y, Yoshizawa T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Masuda S, Yanagii M, Nishimura M, Yamagishi A, Oshima T, Kikuchi H (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. DNA Res 8:123–140

Keeling PJ, Klenk HP, Singh RK, Feeley O, Schleper C, Zillig W, Doolittle WF, Sensen CW (1996) Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1. Plasmid 35:141–144

Kessler A, Brinkman AB, Van der OJ, Prangishvili D (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon sulfolobus. J Bacteriol 186:7745–7753

Koonin EV (1992) Archaebacterial virus SSV1 encodes a putative DnaA-like protein. Nucleic Acids Res 20:1143

Koonin EV (1993) A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic Acids Res 21:2541–2547

Kraft P, Kummel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM (2004a) Structure of D-63 from sulfolobus spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. J Virol 78:7438–7442

Kraft P, Oeckinghaus A, Kummel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM (2004b) Crystal structure of F-93 from Sulfolobus spindle-shaped virus 1, a winged-helix DNA binding protein. J Virol 78:11544–11550

Lao-Sirieix SH, Bell SD (2004) The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3′-terminal nucleotidyl transferase activities. J Mol Biol 344:1251–1263

Lipps G, Ibanez P, Stroessenreuther T, Hekimian K, Krauss G (2001a) The protein ORF80 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein. Nucleic Acids Res 29:4973–4982

Lipps G, Stegert M, Krauss G (2001b) Thermostable and site-specific DNA binding of the gene product ORF56 from the *Sulfolobus islandicus* plasmid pRN1, a putative archael plasmid copy control protein. Nucleic Acids Res 29:904–913

Lipps G, Rother S, Hart C, Krauss G (2003) A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. EMBO J 22:2516–2525

Lipps G, Weinzierl AO, von Scheven G, Buchen C, Cramer P (2004) Structure of a bifunctional DNA primase-polymerase. Nat Struct Mol Biol 11:157–162

Lower BH, Bischoff KM, Kennelly PJ (2000) The archaeon *Sulfolobus solfataricus* contains a membrane-associated protein kinase activity that preferentially phosphorylates threonine residues in vitro. J Bacteriol 182:3452–3459

Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R (2004) Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. Proc Natl Acad Sci USA 101:7046–7051

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

Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology 185:242–250

Peng X, Holz I, Zillig W, Garrett RA, She Q (2000) Evolution of the Family of pRN Plasmids and their Integrase-mediated Insertion into the Chromosome of the Crenarchaeon *Sulfolobus solfataricus*. J Mol Biol 303:449–454

Peng X, Blum H, She Q, Mallok S, Brugger K, Garrett RA, Zillig W, Prangishvili D (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology 291:226–234

Peng X, Kessler A, Phan H, Garrett RA, Prangishvili D (2004) Multiple variants of the archaeal DNA rudivirus SIRV1 in a single host and a novel mechanism of genomic variation. Mol Microbiol 54:366–375

Pisani FM, De Felice M, Carpentieri F, Rossi M (2000) Biochemical characterization of a clamp-loader complex homologous to eukaryotic replication factor C from the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Mol Biol 301:61–73

Prangishvili D, Garrett RA (2004) Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. Biochem Soc Trans 32:204–208

Prangishvili D, Albers SV, Holz I, Arnold HP, Stedman K, Klein T, Singh H, Hiort J, Schweier A, Kristjansson JK, Zillig W (1998a) Conjugation in archaea: frequent occurrence of conjugative plasmids in Sulfolobus. Plasmid 40:190–202

Prangishvili D, Klenk HP, Jakobs G, Schmiechen A, Hanselmann C, Holz I, Zillig W (1998b) Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. J Biol Chem 273:6024–6029

Prangishvili D, Arnold HP, Gotz D, Ziese U, Holz I, Kristjansson JK, Zillig W (1999) A novel virus family, the Rudiviridae: structure, virus–host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. Genetics 152:1387–1396

Prangishvili D, Stedman K, Zillig W (2001) Viruses of the extremely thermophilic archaeon Sulfolobus. Trends Microbiol 9:39–43

Pucci B, De Felice M, Rossi M, Onesti S, Pisani FM (2004) Amino acids of the *Sulfolobus solfataricus* mini-chromosome maintenance-like DNA helicase involved in DNA binding/remodeling. J Biol Chem 279:49222–49228

Purschke WG, Schaefer G (2001) Independent replication of the plasmids pRN1 and pRN2 in the archaeon *Sulfolobus islandicus*. FEMS Microbiol Lett 200:97–102

Qureshi SA, Bell SD, Jackson SP (1997) Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. EMBO J 16:2927–2936

Rasooly A, Rasooly RS (1997) How rolling circle plasmids control their copy number. Trends Microbiol 5:440–446

Reilly MS, Grogan DW (2001) Characterization of intragenic recombination in a hyperthermophilic archaeon via conjugational DNA exchange. J Bacteriol 183:2943–2946

Reiter WD, Palm P, Yeats S, Zillig W (1987) Gene expression in archaebacteria physical mapping of constitutive and uv-inducible transcripts from the sulfolobus virus-like particle ssv1. Mol Gen Genet 209:270–275

Reiter WD, Palm P, Zillig W (1988) Transcription termination in the archaebacterium Sulfolobus: signal structures and linkage to transcription initiation. Nucleic Acids Res 16:2445–2459

Rice G, Tang L, Stedman K, Roberto F, Spuhler J, Gillitzer E, Johnson JE, Douglas T, Young M (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc Natl Acad Sci USA 101:7716–7720

Roberts JA, White MF (2005) An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. J Biol Chem 280:5924–5928

Roberts JA, Bell SD, White MF (2003) An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. Mol Microbiol 48:361–371

Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell 116:25–38

Savino C, Federici L, Johnson KA, Vallone B, Nastopoulos V, Rossi M, Pisani FM, Tsernoglou D (2004) Insights into DNA replication: the crystal structure of DNA polymerase B1 from the archaeon Sulfolobus solfataricus. Structure 12:2001–2008

Schafer G (1996) Bioenergetics of the archaebacterium Sulfolobus. Biochim Biophys Acta 1277:163–200

Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P (2004) Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. J Bacteriol 186:427–437

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. Proc Natl Acad Sci USA 89:7645–7649

Schleper C, Holz I, Janekovic D, Murphy J, Zillig W (1995) A multicopy plasmid of the extremely thermophilic archaeon Sulfolobus effects its transfer to recipients by mating. J Bacteriol 177:4417–4426

Serre MC, Letzelter C, Garel JR, Duguet M (2002) Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. J Biol Chem 277:16758–16767

She Q, Phan H, Garrett RA, Albers SV, Stedman KM, Zillig W (1998) Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon. Extremophiles 2:417–425

She Q, Peng X, Zillig W, Garrett RA (2001a) Gene capture in archaeal chromosomes. Nature 409:478

She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der OJ (2001b) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc Natl Acad Sci USA 98:7835–7840

She Q, Brugger K, Chen L (2002) Archaeal integrative genetic elements and their impact on genome evolution. Res Microbiol 153:325–332

She Q, Shen B, Chen L (2004) Archaeal integrases and mechanisms of gene capture. Biochem Soc Trans 32:222–226

Snyder JC, Stedman K, Rice G, Wiedenheft B, Spuhler J, Young MJ (2003) Viruses of hyperthermophilic Archaea. Res Microbiol 154:474–482

Stedman KM, Schleper C, Rumpf E, Zillig W (1999) Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. Genetics 152:1397–1405

Stedman KM, She Q, Phan H, Holz I, Singh H, Prangishvili D, Garrett R, Zillig W (2000) pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. J Bacteriol 182:7014–7020

Steitz TA, Smerdon SJ, Jager J, Joyce CM (1994) A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. Science 266:2022–2025

Tang TH, Polacek N, Zywicki M, Huber H, Brugger K, Garrett R, Bachellerie JP, Huttenhofer A (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol Microbiol 55:469–481

Tolstrup N, Sensen CW, Garrett RA, Clausen IG (2000) Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. Extremophiles 4:175–179

Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M (2004) Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. J Virol 78:1954–1961

Wood AG, Whitman WB, Konisky J (1989) Isolation and characterization of an archaebacterial viruslike particle from *Methanococcus voltae* A3. J Bacteriol 171:93–98

Worthington P, Hoang V, Perez-Pomares F, Blum P (2003) Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Bacteriol 185:482–488

Ziegelin G, Scherzinger E, Lurz R, Lanka E (1993) Phage P4 alpha protein is multifunctional with origin recognition, helicase and primase activities. EMBO J 12:3703–3708

Ziegelin G, Linderoth NA, Calendar R, Lanka E (1995) Domain structure of phage P4 alpha protein deduced by mutational analysis. J Bacteriol 177:4333–4341

Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Hain J, Lanzendoerfer M, Kristjansson JK (1994) Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. Syst Appl Microbiol 16:609–628

Zillig W, Prangishvilli D, Schleper C, Elferink M, Holz I, Albers S, Janekovic D, Gotz D (1996) Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic archaea. FEMS Microbiol Rev 18:225–236