Extremophiles (1998) 2:391–393

ORIGINAL PAPER

Patrick J. Keeling · Hans-Peter Klenk · Rama K. Singh
Margret E. Schenk · Christoph W. Sensen
Wolfram Zillig · W. Ford Doolittle

Sulfolobus islandicus plasmids pRN1 and pRN2 share distant but common evolutionary ancestry

Received: November 19, 1997 / Accepted: March 10, 1998

Abstract The complete sequence of the plasmid pRN2 from the thermoacidophile *Sulfolobus islandicus* has been determined. The plasmid was found to be circular and 6959 bp in length. *S. islandicus* harbors another endogenous plasmid, pRN1, and comparison of pRN1 and pRN2 revealed that these two plasmids are essentially homologous, although very distantly related. pRN1 and pRN2 share several stretches of highly conserved noncoding DNA and three common open reading frames. Two of these reading frames are likely related to replication, one encoding a large protein with a helicase domain similar to viral helicases, and the other a copy number control protein, CopG.

Key words Archaebacteria · Archaea · *Sulfolobus* · Plasmid · Thermophile

Introduction

Sulfolobus islandicus strain REN1H1 is a sulfur-dependent thermophile isolated from the Reykjanes sulfataric field in Iceland (Zillig et al. 1994). At the time it was discovered, *S. islandicus* was observed to harbor two endogenous plasmids of approximately 5.5 and 6.9 kb. These plasmids were shown to have a high copy number in *S. islandicus* (roughly 20 and 35 per cell, respectively), and capable of transformation and replication in both *S. islandicus* and *S. solfataricus* but not *S. acidocaldarius* (Schlepper 1993; Zillig et al. 1994).

The small size, high copy number, and fairly broad host range make pRN1 and pRN2 attractive candidates as potential *Sulfolobus* vectors. Plasmids from a number of Archaea have been identified and characterized in varying detail (Schalkwyk 1993; Schleper et al. 1995; Erauso et al. 1996; Elferink et al. 1996; Zillig et al. 1996), and some have been successfully used to develop shuttle vectors (see Table 2 in Schalkwyk 1993). However, much of this work has concentrated on plasmids from the archaeal group known as Euryarchaeota. The *Sulfolobales* belong to the Crenarchaeota, which are only distantly related to Euryarchaeota. Although many mesophilic Crenarchaeota are now being discovered (see, for instance, DeLong et al. 1994), the majority of characterized Crenarchaeota are thermophiles, and many are also acidophiles. Often these present practical difficulties for laboratory manipulations because they require extremely high temperature and pressure or have low tolerance for oxygen. *Sulfolobales* are adapted to thermoacidophily, but are nevertheless among the most easily manipulated Crenarchaeota as the result of their comparatively moderate growth requirements.

The complete sequence of pRN1 was determined previously (Keeling et al. 1996). It was found to be predominantly composed of open reading frames, including one that spanned more than half its length. This large reading frame encodes a putative protein of 904 amino acids with a domain similar to helicases used for initiating replication by a diversity of eukaryotic viruses and some bacteriophages (Gorbalenya et al. 1990; Ilyina et al. 1992). The only other fragment of pRN1 with detectable homology to any known sequence was a small open reading frame immediately upstream of this large helicase gene. The product of this small gene was shown be related to CopG, a copy number control protein used by numerous eubacterial plasmids (del Solar et

Communicated by G. Antranikian

P.J. Keeling (✉)[1] · H.-P. Klenk[2] · M.E. Schenk · W.F. Doolittle
Department of Biochemistry, Dalhousie University, Halifax, N.S., Canada, B3H 4H7

R.K. Singh · C.W. Sensen
National Research Council of Canada – Institute for Marine Biosciences, 1411 Oxford Street, Halifax, N.S., Canada, B3H 3Z1

W. Zillig
Max-Plank-Institut für Biochemie, 82152 Martinsried, Germany

Present address:
[1]Department of Biology, Indiana University, Bloomington IN 47405, USA
Tel. +1-812-855-2549; Fax +1-812-855-6705
e-mail: pkeeling@sunflower.bio.indiana.edu
[2]Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Georg-August-University, Goettingen, Grisebachstr. 8, 37077 Goettingen, Germany

al. 1993). In these eubacterial plasmids, the *cop*G gene is located upstream of a gene encoding a replication initiator protein, and the two genes are expressed from a common promoter. The CopG protein binds to this promoter and represses the expression of both proteins, thus controlling the replication of the plasmid. The position of the helicase and *cop*G-like genes in pRN1 was reminiscent of these eubacterial plasmids, and led to the suggestion that pRN1 might operate in a similar fashion (Keeling et al. 1996).

Here we have determined the sequence of the second, larger, endogenous plasmid of *S. islandicus*, pRN2. This plasmid appears to be homologous to pRN1, but nevertheless quite divergent. By comparing the two plasmids, highly conserved regions that are likely important to maintenance have been identified, and conversely, potentially disruptable sites to develop pRN1 or pRN2 as cloning vectors have also been identified in nonconserved regions.

Materials and methods

pRN2 was originally cloned at a unique *Kpn*I restriction site (Zillig et al. 1994). Restriction subclones were constructed from this clone using the enzymes *Eco*RI, *Sph*I, *Pst*I, *Cla*I, *Hin*cII, and *Hin*dIII. These clones as well as the original *Kpn*I clone were sequenced using dye-terminator chemistry on an ABI 373A automated sequencer (Foster City CA), providing most of the sequence on one strand. The remaining gaps, and the complementary strand sequence, as well as confirmation of sequences flanking restriction subclone junctions, were obtained by primer walking. To confirm that the *Kpn*I site initially used to clone pRN2 is indeed a unique site, the 475-bp region around it was amplified from native plasmid by PCR using the primers GAAGGTGGCTAATTATAGAG and AACGCCCCATAGTTAATCGT. The single product of this reaction was found to be of the expected size and sequence.

Results and discussion

Sulfolobus islandicus strain REN1H1 carries two endogenous cryptic plasmids, pRN1 and pRN2. The sequence of pRN1 was determined previously (Keeling et al. 1996), and we have now also determined the sequence of the larger plasmid, pRN2. This sequence was found to be 6959 bp in length, and its map is circular. The predicted size and the physical map derived from the enzymes used for subcloning (not shown) were both in close agreement with those predicted from the sequence.

The dominant feature of pRN2 is its similarity in organization and sequence to pRN1 (Fig. 1). Three open reading frames (ORF) are clearly homologous between pRN1 and pRN2, and two of these are likely a coregulated unit with functional analogs in eubacterial plasmids. The downstream ORF of this pair is very large, 904 and 980 codons in pRN1

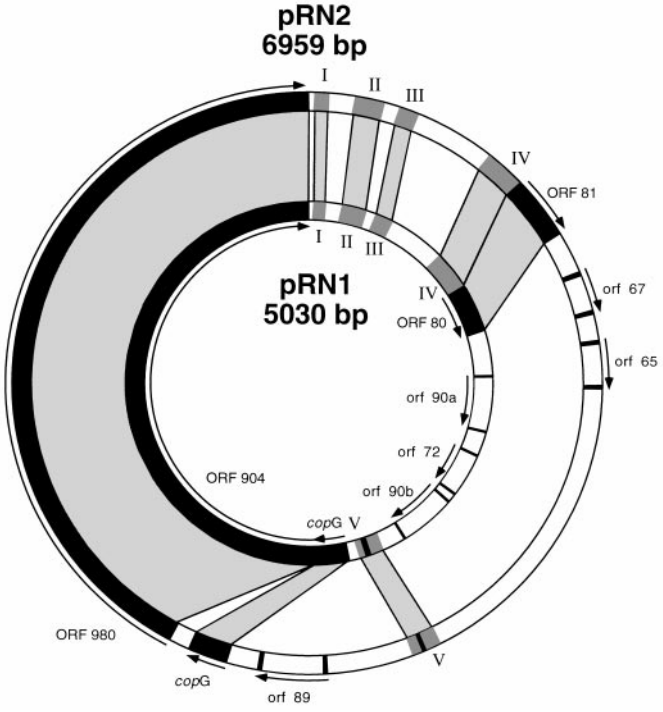

Fig. 1. Maps of pRN1 (*inner circle*) and pRN2 (*outer circle*). Detectably homologous regions are indicated by a *shaded area* between the two maps. Open reading frames shared between the two are shown in *black*: these are ORF80/81, *cop*G, and the helicase-like gene, ORF904/980. Other open reading frames that are not shared and have not been identified are labeled "*orf*" and are shown as *open boxes*. The direction of transcription of all these reading frames is clockwise, as indicated by *arrows*. Conserved noncoding regions are *shaded* and labeled with roman numerals *I–V*. The poly-cytosine track is represented by a *black hatch* in *block V*

and pRN2, respectively. This ORF was recognized in pRN1 as having a carboxy-terminus ATP/GTP binding site characteristic of helicase domains common to many eukaryotic and eubacterial viruses (Gorbalenya et al. 1990; Keeling et al. 1996). Such helicases are often part of episome primase-helicase complexes, and often the primase activity maps to another domain of the same large protein (Ilyina et al. 1992). The amino-terminus of the pRN1 and pRN2 helicase-like ORFs are very highly conserved with respect to one another, but they do not resemble any known primase family, or indeed any other known protein.

Immediately upstream of this helicase-like ORF in both pRN1 and pRN2 is a second shared ORF, similar to a copy number control element *cop*G. CopG is a helix-turn-helix protein used by numerous eubacterial plasmids to suppress the initiation of plasmid DNA replication by controlling the expression level of a replication initiation protein (reviewed in del Solar et al. 1993). Typically, *cop*G is encoded immediately upstream of the gene for the initiator, and controls replication by binding to their common promoter, suppressing transcription of both genes (del Solar and Espinosa 1992; Kwak and Weisblum 1994). In both pRN1 and pRN2 the *cop*G-like gene is located immediately upstream of the

large helicase-like gene, suggesting that the pRN plasmids also use this strategy and that the helicase is a replication initiation protein.

A third gene shared by pRN1 and pRN2 has not previously been recognized, and points out the advantage of having the two sequences. These genes, ORF80 in pRN1 and ORF81 in pRN2, are 78% identical at the amino acid level, but not demonstrably homologous to any other known sequence, and are of unknown function. Similarly, there are a handful of stretches of noncoding DNA with higher than expected identity between pRN1 and pRN2 that would otherwise not be recognized as significant (blocks I–IV in Fig. 1). These conserved blocks range between 50 to 200 bp in length, with a minimum of 67% and as high as 89% nucleotide identity between the corresponding fragments of pRN1 and pRN2. Block V in particular draws attention to itself, partly because it is isolated from other conserved features of the plasmids, and partly because of the unusual characteristics of its sequence. This conserved region is approximately 130 bp in length, and includes an unusual poly-cytosine track flanked by a pyrimidine-rich sequence in one direction and a purine-rich sequence in the other. In addition, the sequence around this region contains several iterations of the octomer CTAACTCT, which appears six times near the poly-cytosine track in pRN1 and four times in pRN2. This repeat and the poly-cytosine track were noted as possibly important in pRN1 (Keeling et al. 1996), and their appearance in pRN2 supports this.

From Fig. 1 it can be seen that the two *Sulfolobus* plasmids are roughly colinear from the start of *cop*G clockwise to the end of ORF80/ORF81. The three shared genes and all the conserved stretches of noncoding DNA except block V fall between these points, and the spacing and order of these features is also conserved. Conversely, with the exception of block V, the remaining sequence is not detectably conserved between pRN1 and pRN2, and the majority of the size variation can also be found here. The size variation and lack of obviously conserved features in this area make it an attractive target for disruption.

Acknowledgments This project was funded by the Canadian Genome Analysis and Technology program, and a grant to W.F.D. from the Medical Research Council (MRC). P.J.K. was the recipient of a MRC studentship. W.F.D. is a fellow of the Canadian Institute for Advanced Research. P.J.K. thanks G.I. McFadden for providing facilities for the preparation of this manuscript. This publication is registered as NRCC #39782 by the National Research Council of Canada. The sequence of pRN2 has been deposited in GenBank under the accession number U93082.

References

del Solar G, Espinosa M (1992) The copy number of plasmid pLS1 is regulated by two trans-acting plasmid products: the antisense RNA II and the repressor protein, RepA. Mol Microbiol 6:83–94

del Solar G, Moscoso M, Espinosa M (1993) Rolling circle-replicating plasmids from gram-positive and gram-negative bacteria: a wall falls. Mol Microbiol 8:789–796

DeLong EF, Wu KY, Prezelin BB, Jovine RV (1994) High abundance of Archaea in Antarctic marine picoplankton. Nature (Lond) 371:695–697

Erauso G, Marsin S, Benbouzid-Rollet N, 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

Elferink MG, Schleper C, Zillig W (1996) Transformation of the extremely thermoacidophilic archaeon *Sulfolobus solfataricus* via a self-spreading vector. FEMS Microbiol Lett 137:31–35

Gorbalenya AE, Koonin EV, Wolf YI (1990) A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett 262:145–148

Ilyina TV, Gorbalenya AE, Koonin EV (1992) Organization and evolution of bacterial and bacteriophage primase-helicase systems. J Mol Evol 34:351–357

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

Kwak JH, Weisblum B (1994) Regulation of plasmid pE194 replication: control of *cop-rep*F operon transcription by Cop and of *rep*F translation by countertranscript RNA. J Bacteriol 176:5044–5051

Schalkwyk L (1993) Halobacterial genes and genomes. In: Kates M, Kushner DJ, Matheson AT (eds) The biochemistry of Archaea. Elsevier, Amsterdam, pp 467–496

Schleper C (1993) Genetische Elemente und Transformation des extrem thermophilen Archaeons *Sulfolobus solfataricus*. Dissertation der Fakultät für Biologie der Ludwig-Maximilians-Universität, München

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

Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Hain J, Lanzendörfer 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