PLASMID **35,** 141–144 (1996)
ARTICLE NO. 0016

Complete Nucleotide Sequence of the *Sulfolobus islandicus* Multicopy Plasmid pRN1

PATRICK J. KEELING,*[,1] HANS-PETER KLENK,*[,2] RAMA K. SINGH,†
OISIN FEELEY,† CHRISTA SCHLEPER,‡[,3] WOLFRAM ZILLIG,‡
W. FORD DOOLITTLE,* AND CHRISTOPH W. SENSEN†

**Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7; †National Research Council of Canada—Institute for Marine Biosciences, 1411 Oxford Street, Halifax, Nova Scotia, Canada B3H 3Z1; and ‡Max-Plank-Institut für Biochemie, 82152 Martinsried, Germany*

Received October 9, 1995; revised December 10, 1995

The complete sequence of the 5350-bp plasmid pRN1 from the crenarchaeote *Sulfolobus islandicus* has been determined. This plasmid is the first to be sequenced from this group of thermoacidophilic archaebacteria (Archaea) and its high copy number and wide host range make it a good candidate for a cloning vector. pRN1 contains several open reading frames, including one that spans over half the plasmid and has significant similarity to the helicase domain of viral primase proteins. Directly upstream of this putative primase is a homologue of Cop, a family of small proteins from promiscuous eubacterial plasmids which control copy number by repressing the expression of the replication initiation protein. In eubacterial plasmids *cop* is found upstream of the replication initiator protein. The location of a *cop* homologue upstream of a primase-like gene in pRN1 suggests that it controls DNA replication in a manner similar to these eubacterial plasmids, but does so using a mixture of componants from plasmids and viruses.

Plasmid pRN1 is one of four multicopy plasmids found by Zillig *et al.* (1994) in crenarchaeotes, the thermoacidophilic branch of the archaebacteria. pRN1 and pRN2 were isolated as a mixture from *Sulfolobus islandicus* strain REN1H1 during a systematic screening of acidic springs, mud holes, and water of solfataric fields in Iceland for genetic elements that might be used to develop cloning vectors for Sulfolobales. At only 5350 bp, pRN1 is the smallest episome presently known in the Sulfolobales. The host range of pRN1 is also known to include *S. solfataricus,* the species that will likely become the model of crenarchaeote genetics. Mutants have been described in *S. solfataricus* (Dollard, 1994), plating techniques are well worked out (Grogan, 1989), an infective phage is known (Schleper *et al.,* 1992), electroporation protocols have been developed and used with pRN1 (Schleper, 1993; Zillig *et al.,* 1994), and a genome sequencing project is underway. As part of this development, we have determined the complete nucleotide sequence of pRN1 in order to identify potential coding regions and control sequences.

A plasmid with pRN1 cloned into an *Escherichia coli* vector (pRN1-pUC18 of Zillig *et al.,* 1994) was digested with *Sac*I, *Eco*RI, *Pst*I, *Cla*I, and *Sph*I and the resulting fragments were subcloned in pBluescript (SK^+) (Stratagene). Subclones were sequenced using the universal M13 forward and reverse primers on an ABI 373A sequencing machine using dye-terminator biochemistry. Gap-filling and second-strand sequencing were carried out by primer walking on the original full-length pRN1-pUC18 clone, which also confirmed the sequence of junction points between the various subclones.

The pRN1-pUC18 construct used to make these subclones was originally isolated from *S. islandicus* by digesting cccDNA with *Sac*I, which appears to have a single recognition site in pRN1. This linearized the plasmid and the resulting 5.5-kb fragment was cloned. To

[1] To whom correspondence should be addressed.
[2] The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850.
[3] Present address: Biology Department, University of California at Santa Barbara, Santa Barbara, CA 93106.

0147-619X/96 $18.00

eliminate the possibility that there are actually multiple closely spaced *Sac*I sites in pRN1, we reisolated the native plasmid by cesium chloride gradient centrifugation (Zillig *et al.*, 1994), recloned it at a different position, and sequenced the region around the *Sac*I site, revealing that the sequence of the native plasmid does indeed concur with that of pRN1-pUC18.

The complete sequence of pRN1 is 5350 bp and is represented by the map in Fig. 1. On one strand there are six putative open reading frames, including two with putative TTG start codons, which are known in *Solfolobus* to be functional and rather abundant (Klenk *et al.*, 1991; H.-P. K., unpublished observations). By far the largest open reading frame, *orf904*, encompasses 2712 bp, over half of the plasmid. Sequence comparison of this sequence with prosite (Bairoch and Bucher, 1994) and sequence databases using the BLASTP and BLASTX programs (Altschul *et al.*, 1990) reveals that the region between amino acids 550 and 700 contains an ATP/GTP binding motif (PS 00017) and shares significant similarity to the carboxy terminal domain of a class of viral proteins. ORF904 shares the greatest similarity to the highly conserved ORF2 from a family of *Streptococcus thermophilis* bacteriophages (Brüssow *et al.*, 1994) and the primase proteins from phage P4 and phi-R73 (Sun *et al.*, 1991). This region of the P4 primase corresponds to a helicase domain which is conserved across a diverse group of primases found mostly in eukaryotic papoviruses, paroviruses, herpesviruses, and some RNA viruses (see Gorbalenya *et al.*, 1990, and references within). However, the amino terminal primase domain of the P4 protein is not related to this family of viral primases, but to a family of bacterial proteins represented by DnaG (Ilyina *et al.*, 1992). The amino terminal domain of ORF904 is not noticeably similar to either, suggesting further mixing of domains in primase–helicase systems.

The only other ORF for which a significant match could be found was *orf56*, which lies immediately upstream of *orf904*, overlapping by 20 bp. The predicted product of this small ORF is very similar to a family of regulatory proteins found in a class of broad-host-range eubacterial plasmids which replicate by a rolling-circle mechanism. These plasmids are typified by pLS1 from *Streptococcus* species, pE194 from *Bacillus subtilis*, and pWVO1 from *Lactococcus lactis* (del Solar *et al.*, 1993; Kwak and Weisblum, 1994; Leenhouts *et al.*, 1991). In pLS1 this protein is known as CopG, a repressor that acts by binding a 13-bp symmetrical operator, thereby blocking transcription of the *copG-repB* cistron. RepB is the replication initiation protein, and by suppressing its expression, CopG controls the plasmid copy number (del Solar *et al.*, 1990; del Solar and Espinosa, 1992). In pRN1, *orf56* and *orf904* are in the same respective orientation as *copG* and *repB*, and also appear to share a promoter. If the identification of a helicase-like domain in the 3′ end of ORF904 is accurate, then this large ORF may be a multifunctional replication initiator which is regulated by the binding of ORF56 to their common promoter, thereby controlling plasmid copy number. In support of this is an imperfectly symmetrical 13-bp sequence which encompasses the start site of ORF56 at position 2144.

The untranslated regions of pRN1 are generally rich in repeats and stem loops, as is the case in the control regions of many eubacterial plasmids (Dempsey *et al.*, 1995; del Solar and Espinosa, 1992). In one case, between 1920 and 2150, there is a large stable structure with a terminal loop containing an unusual track of 17 consecutive cytosine residues (see Fig. 1). There are, however, no significant matches to any of the replication origin consensus sequences known for the family of eubacterial plasmids with similarity to pRN1 (such as pLS1 or pE194; reviewed in Gruss and Ehrlich, 1989). This is perhaps not surprising in light of the similarity of the putative initiation protein (ORF904) to *viral* proteins which are not found on any other plasmid to date.

The similarities noted between pRN1 and other plasmids and viruses tentatively suggest that many of the strategies and machinery used in plasmid and viral replication may transcend domain boundaries. This is particularly interesting given the differences between gene expression in eubacteria, eukaryotes, and archae-

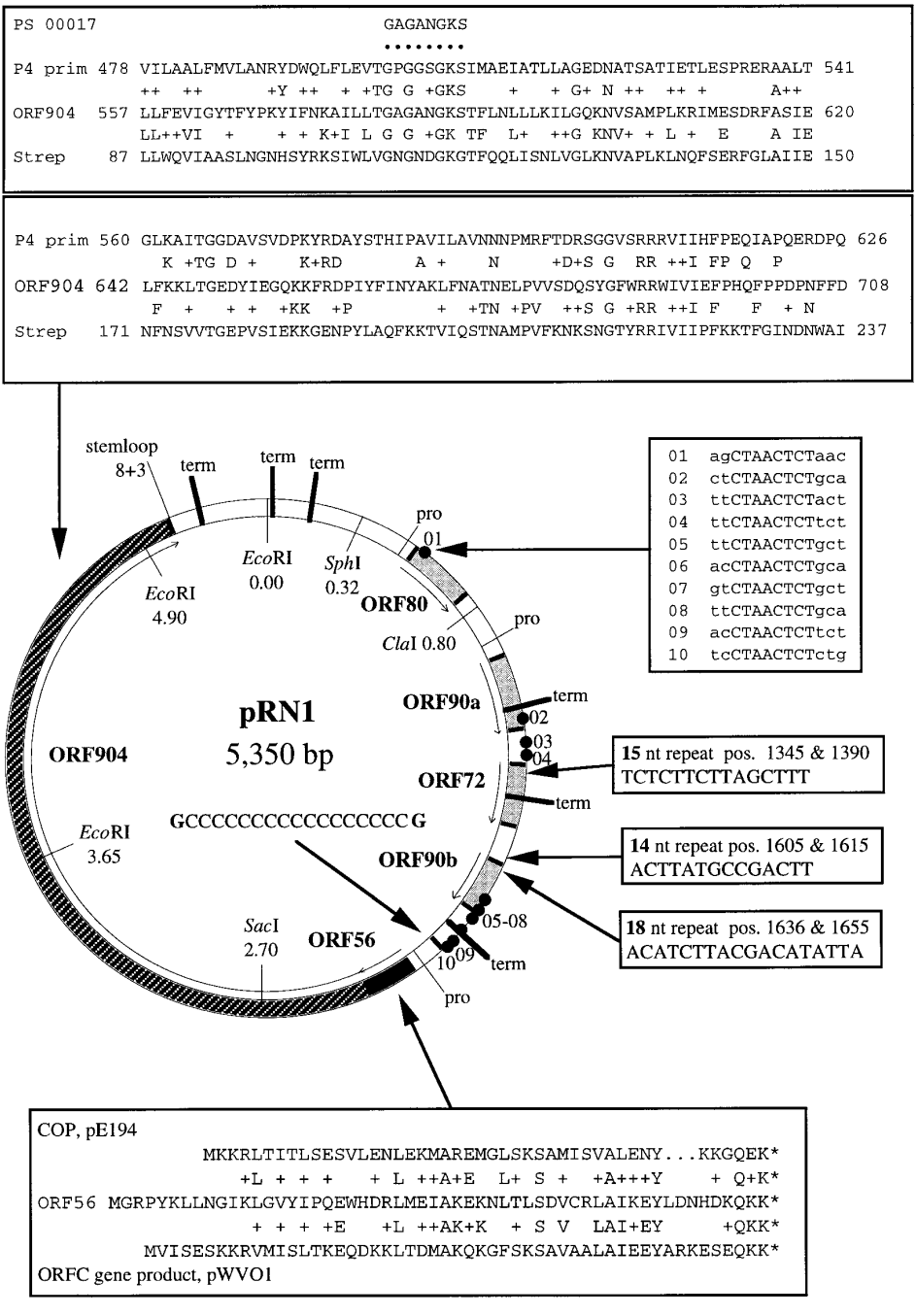

FIG. 1. Map of pRN1 including several direct repeats and the cytosine track, as well as putative transcription promoters (pro), terminators (term), and unidentified ORFs (light shading). *orf90a* and *orf72* begin with TTG rather than ATG, but may nevertheless represent true coding sequences as TTG start sites are used by numerous *Sulfolobus* genes. Two identifyable genes, *orf904* (dark shading) and *orf56* (black) are also shown. These ORFs share a common promoter-like sequence, overlap by 20 bp, and are followed by a terminator suggesting that they are cotranscribed. Boxed areas show regions of similarity between these two genes and other proteins, and draw attention to the position of particular features of the nucleotide sequence such as repeats and a track of cytidine residues.

bacteria (Keeling and Doolittle, 1995) and the adaptation to hyperthermophily, which are each serious obstacles for a mobile element colonizing a new host. The integration of proteins found in either plasmids or viruses in pRN1 also deserves notice: this system seems to underscore the *ad hoc* nature of mobile elements, which may use whatever tools are available to replicate.

ACKNOWLEDGMENTS

This work was done in the context of the *Sulfolobus* Genome Sequencing Project and was supported by the Canadian Genome Analysis and Technology Program (CGAT). P.J.K. is supported by a studentship from the Medical Research Council (MRC). W.F.D. is a fellow and C.W.S. and W.Z. are associates of the Canadian Institute for Advanced Research (CIAR). The sequence of pRN1 has been deposited in GenBank under Accession No. U36383. We thank P. Forterre for helpful comments on the manuscript.

REFERENCES

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W., AND LIPMAN, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

BAIROCH, A., AND BUCHER, P. (1994). PROSITE: Recent developments. *Nucleic Acids Res.* **22,** 3583–3589.

BRÜSSOW, H., PROBST, A., FRÉMONT, M., AND SIDOTI, J. (1994). Distinct *Streptococcus thermophilis* bacteriophages share an extremely conserved DNA fragment. *Virology* **200,** 854–857.

DEL SOLAR, G. H., PEREZ-MARTIN, J., AND 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.

DEL SOLAR, G., AND 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. AND ESPINOSA, M. (1993). Rolling circle-replicating plasmids from gram-positive and gram-negative bacteria: A wall falls. *Mol. Microbiol.* **8,** 789–796.

DEMPSEY, L. A., ZHAO, A. C., AND KAHN, S. A. (1995). Localization of start sites of lagging-strand replication of rolling-circle plasmids from Gram-positive bacteria. *Mol. Microbiol.* **15,** 679–688.

DOLLARD, C. D. (1994). Development of a genetic system for the extremely thermophilic archaeon *Sulfolobus.* Master's thesis, Dalhousie University, Halifax.

GORBALENYA, A. E., KOONIN, E. V., AND WOLF, Y. I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett.* **262,** 145–148.

GROGAN, D. W. (1989). Phenotypic characterization of the archaebacterial genus *Sulfolobus:* Comparison of five wild-type strains. *J. Bacteriol.* **171,** 6710–6719.

GRUSS, A. AND EHRLICH, S. D. (1989). The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. *Microbiol. Rev.* **53,** 231–241.

ILYINA, T. V., GORBALENYA, A. E. AND KOONIN, E. V. (1992). Organization and evolution of bacterial and bacteriophage primase-helicase systems. *J. Mol. Evol.* **34,** 351–357.

KEELING, P. J., AND DOOLITTLE, W. F. (1995). Archaea: Narrowing the gap between prokaryotes and eukaryotes. *Proc. Natl. Acad. Sci. USA* **92,** 5761–5764.

KLENK, H.-P., SCHWASS, V., AND ZILLIG, W. (1991). Nucleotide sequence of the genes encoding the L30, S12, and S7 equivalent ribosomal proteins from the archaebacterium *Thermococcus celer. Nucleic Acids Res.* **19,** 6047.

KWAK, J. H., AND WEISBLUM, B. (1994). Regulation of plasmid pE194 replication: control of *cop-repF* operon transcription by Cop and of *repF* translation by countertranscript RNA. *J. Bacteriol.* **176,** 5044–5051.

LANGER, D., HAIN, J., THURIAUX, P., AND ZILLIG, W. (1995). Transcription in *Archaea:* Similarity to *Eukarya. Proc. Natl. Acad. Sci. USA* **92,** 5277–5281.

LEENHOUTS, K. J., TOLNER, B., BRON, S., KOK, J., VENEMA, G., AND SEEGERS, F. M. L. (1991). Nucleotide sequence and characterization of the broad-host-range Lacctococcal plasmid pWVO1. *Plasmid* **26,** 55–66.

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., KUBO, K., AND ZILLIG, W. (1992). The particle SSV1 form 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.

SUN, J., INOUYE, M., AND INOUYE, S. (1991). Association of a retroelement with a P4-like cryptic prophage (retronphage phiR73) integrated into the selenocystyl tRNA gene of *Escherichia coli. J. Bacteriol.* **173,** 4171–4181.

ZILLIG, W., KLETZIN, A., SCHLEPER, C., HOLZ, I., JANEKOVIC, D., HAIN, J., LANZENDOERFER, M., AND KRISTJANSSON, J. K. (1994). Screening for *Sulfolobales,* their plasmids, and their viruses in Icelandic solfataras. *System. Appl. Microbiol.* **16,** 609–628.

Communicated by D. K. Chattoraj