*Published online 17 September 2009* *Nucleic Acids Research, 2009, Vol. 37, No. 20* ***6831–6848***
*doi:10.1093/nar/gkp742*

# DNA translocation activity of the multifunctional replication protein ORF904 from the archaeal plasmid pRN1

**Martin Sanchez[1], Markus Drechsler[2], Holger Stark[3] and Georg Lipps[1,4,*]**

[1]Department of Biochemistry, [2]Department of Macromolecular Chemistry II, University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, [3]Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Goettingen and [4]University of Applied Research of Northwestern Switzerland, Gründenstrasse 40, 4132 Muttenz, Switzerland

Received April 30, 2009; Revised and Accepted August 24, 2009

## ABSTRACT

**The replication protein ORF904 from the plasmid pRN1 is a multifunctional enzyme with ATPase-, primase- and DNA polymerase activity. Sequence analysis suggests the presence of at least two conserved domains: an N-terminal prim/pol domain with primase and DNA polymerase activities and a C-terminal superfamily 3 helicase domain with a strong double-stranded DNA dependant ATPase activity. The exact molecular function of the helicase domain in the process of plasmid replication remains unclear. Potentially this motor protein is involved in duplex remodelling and/or origin opening at the plasmid replication origin. In support of this we found that the monomeric replication protein ORF904 forms a hexameric ring in the presence of DNA. It is able to translocate along single-stranded DNA in 3′–5′ direction as well as on double-stranded DNA. Critical residues important for ATPase activity and DNA translocation activity were identified and are in agreement with a homology model of the helicase domain. In addition we propose that a winged helix DNA-binding domain at the C-terminus of the helicase domain could assist the binding of the replication protein specifically to the replication origin.**

## INTRODUCTION

Plasmids and other mobile genetic elements such as bacteriophages and viruses have proved to be excellent model systems to investigate the fundamentals of information processing in bacteria and eukarya. The temperate phage λ has been intensively studied and basic concepts of gene regulation were developed with this model. Likewise eukaryotic and bacterial replication has been studied with the help of bacteriophages (e.g. T7), plasmids (ColE1) and viruses (SV40) (1). Most of the studied plasmids and viruses are from eukaryotes and bacteria. In contrast our knowledge on archaeal plasmids and viruses is still rather limited.

Archaeal plasmids are known for the three major branches of the *Euryarchaeota* [e.g. pC2A from *Methanosarcina acetivorans* (2), pNRC100 from *Halobacterium* sp. NRC-1 (3) and pGT5 from *Pyrococcus abyssi* GE5 (4)]. Among the *Crenarchaeota* plasmids are known for the genera *Sulfolobus*, *Acidianus* and *Thermofilum*. Most archaeal plasmids are only poorly studied (5). Nevertheless the sequencing of the plasmidal genomes allowed predictions to be made concerning the regulation and replication of some of these plasmids. For the genus *Sulfolobus* about a dozen plasmids have been sequenced. They can be provisionally divided into two groups. One group comprises conjugative plasmids with genome sizes between 25 and 42 kb. These plasmids appear to encode a cluster of conserved genes which could be responsible for conjugative DNA transport (6). The other group consists of smaller plasmids (genome sizes between 5 and 14 kb). Plasmids obtained from isolates from Iceland founded the novel plasmid family pRN (see below). The plasmids obtained from isolates in Italy and New Zealand (7) are similar, however, they do not share all conserved open reading frames of the pRN family. It is therefore less clear if they also belong to the pRN plasmid family.

The best-studied archaeal plasmids from a biochemical point of view are the plasmids pGT5 from *Pyrococcus abyssi* and pRN1 from *Sulfolobus islandicus*. The plasmid pGT5 (3444 bp) encodes for a protein with sequence similarity to the bacterial initiator proteins of rolling circle replication. In fact it could be shown that this gene product has nicking and closing activities at the predicted double-stranded replication origin (8).

*To whom correspondence should be addressed. Tel: +41 61 4674301; Fax: +41 61 4674701; Email: georg.lipps@fhnw.ch

Our group has studied the plasmid pRN1, a 5350 bp long circular plasmid of the crenarchaeote *Sulfolobus islandicus* (9). The plasmid shares three highly conserved open reading frames with several other related crenarchaeal plasmids, which have been grouped into the novel plasmid family pRN (10). Sequence analysis suggests that these plasmids are unrelated to hitherto known bacterial plasmids. We expressed the respective genes of the highly conserved open reading frames in *E. coli* and characterised the gene products biochemically. Two of the proteins (ORF56 and ORF80) are sequence-specific DNA-binding proteins. The protein ORF56 is a dimeric, basic 6.5 kDa protein which binds upstream of its own gene. This protein could act as a repressor and participate at the copy number control of the plasmid (11,12). The protein ORF80 binds to two neighbouring binding sites upstream of its own gene. It has been suggested that it forms a large protein–nucleic acid complex. However, the function of the protein or the complex is not known (13). Although this gene is highly conserved and homologues are also found in conjugative plasmids of *Sulfolobus* as well as in crenarchaeal viruses, interruption of the gene unexpectedly does not appear to prevent plasmidal replication (14).

The third conserved gene encodes the 106 kDa replication protein ORF904. The gene occupies roughly half of the coding capacity of the plasmid pRN1. We could show that the enzyme is multifunctional with ATPase, primase and DNA polymerase activity (15). The replication protein consists of at least two conserved domains. The novel N-terminal prim/pol domain has primase and DNA polymerase activities (15,16). We could recently show that the primase activity is highly sequence specific requiring a GTG motif in the template DNA. Rather unusually, the primer formed by the replication protein consists mainly of deoxynucleotides with the exception of the first base which is a ribonucleotide (17). Possibly the primase activity of the replication protein synthesises the first primer for plasmid genome replication initiation. The steps preceding priming are much less clear. The double-stranded plasmid DNA must be unwound at the replication origin in order to deliver single-stranded DNA to the primase active site. *In vivo* this process can be additionally impeded by the positive supercoiling of the plasmid DNA as observed for thermophilic plasmids (18).

In general, the process of origin opening is not well-understood even for chromosomal replication. For bacteria as well as for archaea and eukarya the formation of large nucleoprotein complexes is made responsible for the destabilisation of duplex DNA which ultimately leads to origin opening and assembly of the replisome (19–21). Involved in this process are $AAA^+$-ATPases such as the eukaryotic origin recognition complex subunits Orc and the bacterial DnaA protein. Interestingly, another type of $AAA^+$-ATPase protein is responsible for DNA replication initiation of viruses and other genetic elements: the superfamily 3 helicases were originally discovered in the genome of several viruses (22,23). These proteins may unwind the origin of replication and may also act as replicative helicases. The large T-antigen from the SV40 virus and the homologous proteins from adenoassociated virus and the papilloma virus are the best-studied superfamily 3 helicases. Crystal structures of the respective proteins show that these helicases function as hexameric rings and that the DNA is transported through the central pore (24,25). The directional transport of DNA is assisted by conformational changes of β-hairpins located in the pore and are driven by ATP hydrolysis. The molecular mechanism of duplex opening at the replication origin by superfamily 3 helicases however remains unclear.

In order to get insight into the functional organization of the replication protein and to understand the process of replication initiation of the plasmid pRN1, we constructed and characterised several deletion and point mutants. Based on these studies we show that the replication protein contains an additional winged helix DNA-binding domain and we identified critical residues for DNA translocation activity. We propose that the replication protein ORF904 is an autonomous replication initiation protein which could perform several steps for initiation including origin binding, duplex opening and primer synthesis at the replication origin.

## EXPERIMENTAL

### Purification of ORF904, point and deletion mutants

The complete *orf904* gene was amplified by PCR and cloned into pET28c (Novagen) in-frame with the N-terminal hexahistidine tag of the vector. Transformed *E. coli* BL21 (DE3) CodonPlus cells (Stratagene) were grown in Luria Broth at room temperature, and induced with 1 mM IPTG. Washed cells were lysed in 50 mM sodium phosphate pH 7.0 by sonication. After centrifugation, the crude extract was adjusted to 300 mM NaCl, 2 M urea and pH 7.0 and loaded onto a Talon column (Clontech) equilibrated with 50 mM sodium phosphate pH 7.0, 300 mM NaCl, 2 M urea and 6 mM imidazole. ORF904 was step-eluted with 150 mM imidazole in the starting buffer, diluted 1:1 in $ddH_2O$ and then directly loaded onto a Fractogel EMD-sulphate column (Merck) developed with a linear 1 M NaCl gradient in 25 mM sodium phosphate pH 7.0. ORF904 elutes in a sharp peak at 700 mM NaCl. Pooled fractions were dialysed against 50 mM sodium phosphate pH 7.0, 150 mM NaCl, 1 mM DTT and 40% glycerol.

The point mutants of ORF904 were constructed with two overlapping mutagenic primers, multiple cycle extension and *Dpn*I digestion of parental DNA. PCR generated stretches of the gene *orf904* were confirmed by DNA sequencing. Deletion mutants of ORF904 were constructed by cloning a PCR generated fragment carrying the sequence of interest into the wild-type vector with the respective enzymes. Purification of the helicase domain deletion mutants was done as for the wild-type ORF904.

Deletion mutants of the putative origin binding domain were cloned without the N-terminal hexahistidine tag. Purification was achieved by first incubating the crude extract at 70°C for 10 min and removal of the denatured proteins by centrifugation. The supernatant was loaded

onto a Fractogel EMD-sulphate column. Further purification was done as for ORF904.

The preparations of the larger proteins (Supplementary Figure S1) contained minor amounts of fragments. These proteins are unlikely to stem from the expression host *E. coli* since their molecular weight differed in the preparation of the various deletion mutants. The impurities are also unlikely to disturb our ATPase and helicase assays as we did not observe any ATPase activity with the highly sensitive radioactive ATPase assay neither for the arginine finger mutants (this work) nor for the Walker A mutant (15).

Protein concentrations were determined by using the theoretical extinction coefficient of the respective protein at 280 nm.

### Gel filtration

Ten micrograms of protein K752G844 were diluted in 100 µl of elution buffer (25 mM potassium phosphate pH 7.0, 150 mM KCl) and loaded on a Bio-Sil TSK 250 G3000SW HPLC gel filtration column at 0.5 ml/min. The protein eluted as a sharp peak after 48.3 min. The native molecular weight was determined to be 10.8 kDa with the aid of standard proteins. Since K752G844 has a molecular weight of 11 kDa, the protein is a monomer in solution.

### ATPase assay

Reactions were performed in a 10 µl reaction volume containing reaction buffer (25 mM Tris–HCl pH 7.5, 1 mM DTT, 10 mM $MgCl_2$), 0.2 mM ATP, 0.2 nM [$\gamma$-$^{32}$P]-ATP (5000 Ci/mmol) and 0.5 µM of the indicated protein. Reactions were incubated at 80°C for 20 min before 2 µl aliquots of each sample were analysed on thin-layer PEI cellulose plates (Machery Nagel) developed with 0.8 M lithium chloride in 1 M formic acid. Quantification was performed by instantimaging (Packard Imager). Replicate measurements showed that the standard deviation is <10% for this assay.

### Translocation assay

Biotinylated oligodeoxynucleotides were labelled at the 5′-end using [$\gamma$-$^{32}$P]-ATP and T4 polynucleotide kinase. Ten nanomolar single- or double-stranded substrate were incubated with 180 nM streptavidin in reaction buffer in a final volume of 10 µl at 37°C for 30 min, followed by addition of 6.5 µM free biotin, 0.5 µM protein and 2 mM ATP. After incubation for 30 min at 50°C, 3 µl of loading buffer (0.1% xylene cyanol, 0.1% bromphenol blue, 0.1 mM EDTA, 1% SDS, 75% glycerol) was added to each sample, followed by electrophoretic analysis on a native 8% polyacrylamide gel in 0.5× TBE. The gels were analysed using instantimaging.

### Helicase assay

Oligodeoxynucleotides were labelled as for the translocation assay. To generate double-strand substrates, the labelled oligodeoxynucleotides were hybridised with the complementary or partially complementary oligodeoxynucleotide by incubating an equimolar mixture of the oligodeoxynucleotides for 5 min at 99°C, followed by cooling down to room temperature. Helicase assay reaction mixtures (10 µl) contained reaction buffer, 2 mM ATP, 10 nM of $^{32}$P-labelled substrate, 0.5 µM protein and 100 nM of unlabelled oligodeoxynucleotide preventing reannealing of the displaced, labelled strand. The competitor strand was present from the beginning of the reaction.

The reactions were incubated for 30 min at 60°C and stopped by addition of 3 µl of loading buffer, then loaded on a 8% polyacrylamide gel in 0.5× TBE and run with constant power at 10 W. After the electrophoresis, the gel was analysed by instantimaging.

### Triple-helix assay

Oligodeoxynucleotides were labelled as for the translocation assay. To generate the triple-helix substrate the labelled oligodeoxynucleotide TFO was hybridised with *Ssp*I linearized pMJ5 (26). An equimolar mixture was incubated in buffer MM (25 mM MES, pH 5.5, 10 mM $MgCl_2$) for 15 min at 57°C and left to cool to room temperature overnight. Triple-helix assay reaction mixtures (10 µl) contained reaction buffer (25 mM Tris–HCl pH 7.5, 1 mM DTT, 10 mM $MgCl_2$), 2 mM ATP, 5 nM of $^{32}$P-labelled substrate and 0.05 µM protein. The reactions were incubated for 1–120 min at 45°C, stopped by addition of 3 µl of triplex loading buffer [15% (w/v) glucose, 3% (w/v) SDS, 250 mM MOPS pH 5.5, 1 mg/ml Proteinase K, 0.4 mg/ml bromphenol blue] and analysed on an 8% polyacrylamide gel (40 mM Tris–acetate, 5.0 mM sodium acetate, 1.0 mM $MgCl_2$, pH 5.5) at 10 V/cm for 2 h at 4°C. After the electrophoresis the gel was analysed by instantimaging.

### Fluorescence anisotropy titration

Fluorescence anisotropy measurements were performed on a Perkin Elmer LS-50B fluorescence spectrometer equipped with a thermo jacket. As substrate we used two oligodeoxynucleotides (5′-TCTTCTGTGCACT CTTC-3′, 17 bases; 5′-TTGCCATTGGAAT CGGTCC AATGTA-3′, 25 bases) carrying a fluorescein via a hexyl linker at the 3′-end and the corresponding double-stranded substrate.

Three to ten micromolars of protein was added to a solution with 20 nM fluorescein-labelled DNA. The concentration of the protein was then serially decreased by 30% in each titration step down to a final concentration of 5–90 nM by replacing aliquots of the protein/DNA solution in the cuvette with buffer containing 20 nM DNA. Titration was carried out in 20 mM Tris–HCl pH 7.5, 0.01% Tween-20, 100 mM KCl (standard buffer) in a 1 ml quartz cuvette thermostated at 25°C. The sample was excited at 495 nm using a 15 nm slit and a vertical polarising filter. The vertical and horizontal emission was monitored at 526 nm with a slit of 20 nm and a cut-off filter of 515 nm. The grating factor was calculated with fluorescein as the depolarising sample. At each titration point the anisotropy was measured at least three times

with 10 s integration time. Data was analysed with the Hill model of binding:

$$r = (r_{AB} - r_A)\frac{B^n}{K^n + B^n} + r_A \quad \mathbf{1}$$

with $r$, $r_{AB}$, $r_A$ the measured anisotropy, the anisotropy of the protein–DNA complex and of the free DNA, respectively. $B$ is the protein concentration, $K$ is the half-maximal concentration of binding and $n$ is the Hill coefficient. In the case of non-cooperative binding ($n = 1$), $K$ is the dissociation constant of the protein–DNA complex. All titrations were done at least twice.

### Circular dichroism spectroscopy

Far UV circular dichroism (CD) spectra of the deletion mutants were recorded at temperatures ranging from 20 to 105°C on a Jasco J-600 spectropolarimeter, using a 0.1 cm quartz cell containing 3–18 μM of the respective protein dissolved in 1× PBS (138 mM NaCl, 2.7 mM KCl, 4.7 mM $Na_2HPO_4$, 1.4 mM $KH_2PO_4$) to reveal if the protein is natively folded and to estimate the α-helical secondary structure fraction of the protein.

The final spectra were corrected for the spectrum of the buffer and converted into mean residue ellipticity in units of deg cm$^2$ dmol$^{-1}$. Other experimental settings were 100 nm/min scan speed, 1.0 nm bandwidth, 0.2 nm resolution, 20 mdeg sensitivity and 1 s response.

### Glycerol gradient sedimentation experiment

To analyse the quaternary structure of the protein, glycerol gradient sedimentation experiments were performed. For this purpose buffer A [17% (v/v) glycerol; 2 mM $MgCl_2$; 1× PBS buffer] and buffer B [34% (v/v) glycerol; 2 mM $MgCl_2$; 1× PBS; 0.15% (v/v) glutaraldehyde] were mixed with an FPLC LKB gradient pump 2249 to prepare 11 ml of a gradient ranging from 17 to 34% glycerol. The protein sample solution (150 μl) containing 6–11 μM protein in 1× PBS and 10 mM $MgCl_2$ was incubated for 30 min at 60°C and then carefully pipetted onto the glycerol gradient. In some samples, 1–2 μM of single- or double-stranded DNA (49 bases) and 2 mM ATP or AMP-PnP were added. In some cases, we also added 0.15% glutaraldehyde in order to crosslink the protein complexes. The centrifugation was carried out in a Centrikon T-1055 ultracentrifuge with a TST 41.14 rotor for 16 h at 10°C and 40 000 r.p.m. After the run, fractions of 500 μl were collected from the bottom. The fractions were analysed with a sensitive microplate based Bradford assay. The comparison with a reference sample (50 μg BSA or 50 μg β-amylase and 50 μg thyroglobulin) allowed to estimate the molecular weight of the proteins in the distinct fractions.

### Cryogenic transmission electron microscopy (Cryo-TEM)

For cryo-TEM studies, a 2 μl drop of the sample (0.5 μM protein, 0.5 μM 49 bp DNA, 1 mM AMP-PnP in 10 mM $MgCl_2$, 25 mM potassium phosphate, 150 mM KCl, 1 mM β-mercaptoethanol, pH 7) was placed on an hydrophilised lacey carbon filmed copper TEM grid (200 mesh, Science Services, Munich, Germany). Most of the liquid was removed with blotting paper, leaving a thin film stretched over the carbon lace holes. The specimens were vitrified by rapid immersion into liquid ethane cooled to ~90 K by liquid nitrogen in a temperature controlled freezing unit (Zeiss Cryobox, Zeiss NTS GmbH, Oberkochen, Germany). The temperature was monitored and kept constant in the chamber during all of the sample preparation steps. After freezing the specimens, the remaining ethane was removed using blotting paper. The specimen was inserted into a cryo-transfer holder (CT3500, Gatan, Muenchen, Germany) and transferred to a Zeiss EM922 EFTEM instrument. Examinations were carried out at temperatures ~90 K. The transmission electron microscope was operated at an acceleration voltage of 200 kV. Zero-loss filtered images ($\Delta E = 0$ eV) were taken under reduced dose conditions (100–1000 e/nm$^2$). All images were registered digitally by a bottom mounted CCD camera system (Ultrascan 1000, Gatan, Munich, Germany) combined and processed with a digital imaging processing system (Gatan Digital Micrograph 3.10 for GMS 1.5).

### Transmission electron microscopy

The complex was prepared for electron microscopy using the GraFix procedure that combines a gradient centrifugation with a mild chemical fixation (glutaraldehyde, ~0.1%) (27). Following fractionation, the sample was adsorbed to a thin layer of carbon film, stained with uranyl formate (2% w/v) (28) and air-dried. Images were taken on a F415 slow-scan CCD camera (TVIPS GmbH, Gauting, Germany) at 90 000× magnification in a Philips CM200 FEG operated at 160 kV acceleration voltage. Two-thousand three-hundred fifty-one individual macromolecular images were selected from the micrographs, translationally aligned and subjected to multivariate statistical analysis (29) and classification (30). The first two eigenimages clearly reveal 6-fold symmetry (31) (see Figure 7C). Characteristic 6-fold symmetric views are obtained after multi-reference alignment (32), MSA based classification and subsequent averaging (Figure 7E). Image processing was performed with the IMAGIC-V software package (IMAGE Science GmbH, Berlin) (33).

## RESULTS

### Domain structure of the replication protein

Analysis of the amino acid sequence of the replication protein immediately reveals two conserved domains (Figure 1A): an N–terminal prim/pol domain (amino acids 47–247, pfam09250) and an ATPase domain (amino acids 408–827, COG3378). The latter domain encompasses the conserved domain D5_N (amino acids 417–509, pfam08706). Closer inspection reveals that the central part of the ATPase domain (amino acids ~500–800) is related to several superfamily 3 (SF3) helicases, i.e. the bacteriophage P4 α protein and the poxvirus D5 proteins. The structures of several SF3 helicases have been determined and we used these structures to construct

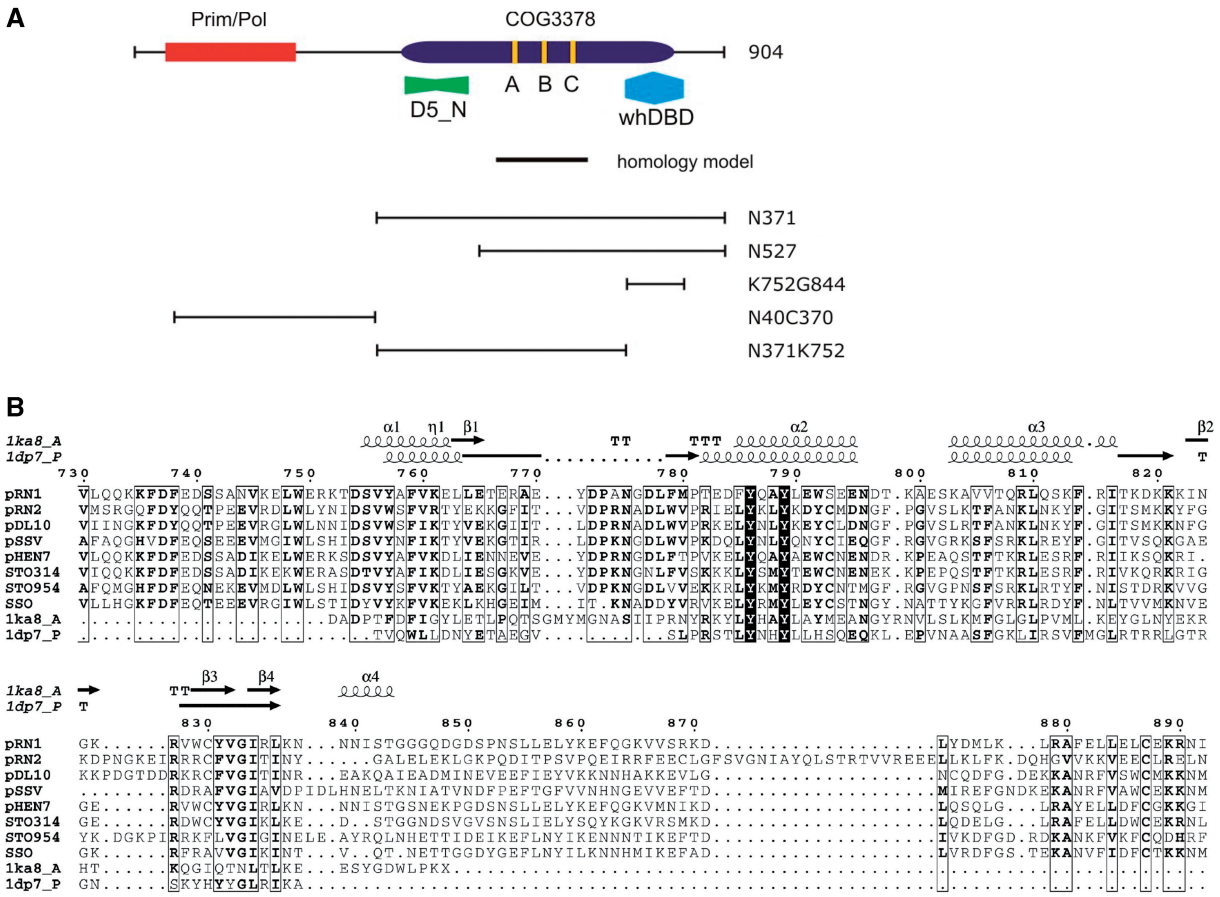


**Figure 1.** Domain organization of the replication protein ORF904. (**A**) The 904 amino acid replication protein from pRN1 has an N-terminal prim/pol domain (pfam09250) and a C-terminal helicase domain of superfamily 3 (COG3378) which includes the D5_N conserved domain (pfam08706). Our analysis suggests an additional winged helix DNA-binding domain (pfam03288 and pfam02257) in the C-terminal part of the protein. The Walker A–C motifs within the helicase domain are indicated as well as the region of the protein which was modelled. The positions of the deletion mutants N40C370, N371, N527, N371K752 and K752G844 are also shown. (**B**) Alignment of the replication proteins from the plasmids pRN1, pRN2, pDL10, pSSVx and pHEN7 with integrated plasmids in the genomes of *Sulfolobus tokodaii* and *Sulfolobus solfataricus* and with two winged helix DNA-binding proteins whose structure has been determined. The two winged helix DNA-binding domains are from the eukaryotic transcriptional regulator Rfx [1dp7 (35)] and from the bacteriophage P4 α protein [1ka8 (47)]. The secondary structures elements of these two proteins are given above the alignment. The Rfx winged helix DNA-binding domain makes contacts with the DNA over the helix α3 (recognition helix) and over the wing (β-hairpin formed by β2 and β3). Similar DNA-binding contacts are observed in the winged helix DNA-binding domain of the archaeal Cdc6/Orc1 proteins (21,44). The alignment of the replication proteins underscores that the putative DNA-binding domain is conserved and that the conservation declines after the end of the DNA-binding domain.

a homology model of the core of the helicase domain of the replication protein ORF904 (amino acids 554–695), see below.

In addition, our analysis suggests the presence of a winged-helix DNA-binding domain at the C-terminal end of the protein. This part of the protein (amino acids 752–844) is still highly conserved within the pRN plasmid family and remarkably the sequence conservation drops significantly just after the end of the proposed winged helix domain (Figure 1B). Noteworthy, the bacteriophage P4 α protein also contains a winged helix DNA-binding domain C-terminal to the SF3 helicase domain. The structure of this domain as well as the structure of the eukaryotic transcriptional activator regulatory factor X (Rfx) were aligned to the replication protein with the sensitive HHpred algorithm, a profile–profile hidden Markov model based alignment procedure (34). Two tyrosine residues in the middle of helix α2 are completely conserved in the replication proteins and in both winged helix DNA-binding domains. The specific function of these residues is not known. In any case they do not appear to contribute directly to DNA binding. The recognition helix α3 and the following wing make most of the DNA contacts as observed in the Rfx-DNA co-crystal (35).

To investigate whether this part of the replication protein folds into a separate domain with DNA-binding activity, we constructed four deletion mutants of this domain. These mutants began at amino acid D737 or K752 and ended either at G844 or at the native C-terminus. The protein D737G844 was mostly insoluble but the other deletion constructs could be obtained in reasonable amounts and were characterised in more detail. Preliminary experiments showed that the protein from amino acid K752 to G844 bound DNA better than

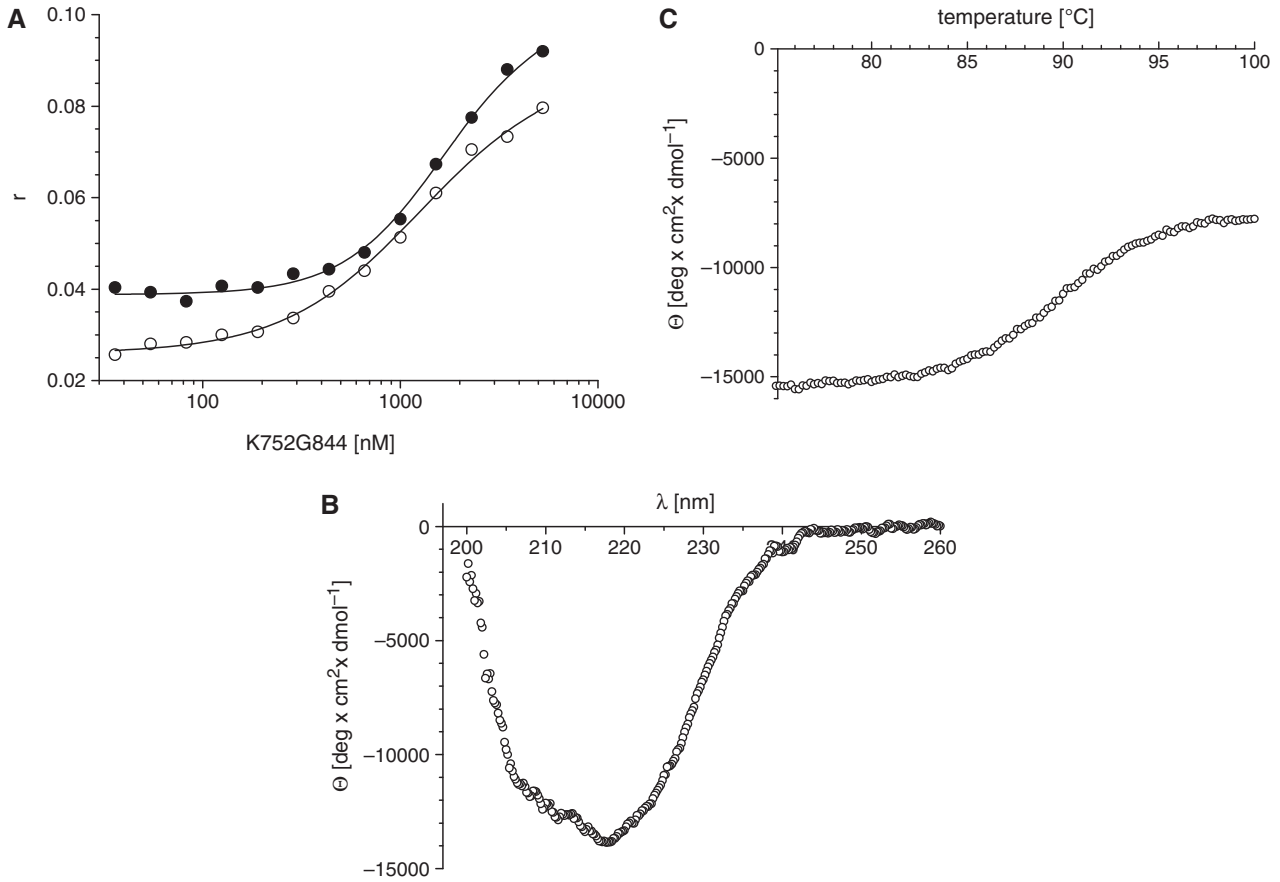


**Figure 2.** The deletion mutant K752G844 is a stable folded protein with DNA-binding activity. (**A**) The DNA-binding strength was determined with short single- and double-stranded DNA by fluorescence anisotropy titrations. The experimental values (filled circle: 17 bp substrate and open circle 17 bp substrate) were analysed with the Hill binding model. (**B**) CD spectroscopy of the deletion mutant K752G844 at 20°C. (**C**) Thermal unfolding of K752G844 monitored by circular dichroism spectroscopy at 218 nm.

the other deletion mutants. We therefore characterised the DNA-binding activity of this mutant which represents the minimal winged-helix domain. To determine whether the winged-helix domain displays DNA-binding activity we used fluorescence anisotropy measurements. This assay allows to precisely determine the dissociation constants of the protein–DNA equilibria. As substrates, we employed a 17 base long oligodeoxynucleotide labelled with a fluorescein at its 3′-end and the same oligodeoxynucleotide hybridised to its complementary strand. The minimal winged-helix domain bound to both substrates with low affinity (half-maximal binding at 1.6 μM single-stranded DNA and 2.1 μM double-stranded DNA) under intermediate salt concentration, i.e. 100 mM potassium chloride. With both substrates, we observed positive cooperativity ($n_{Hill}$ = 1.5 for single-stranded DNA and $n_{Hill}$ = 2.1 for double-stranded DNA, Figure 2A) suggesting that more than one protein molecule is involved in the DNA–protein equilibrium. In an attempt to investigate whether the winged-helix domain exhibits sequence specific DNA-binding activity, we incubated the protein with various fragments of pRN1. Unfortunately, we were unable to detect a preferential binding to any of the tested fragments by neither the winged helix domain protein nor by the full-length protein. In the context of plasmid replication, a sequence-specific DNA binding of the initiator protein of replication is expected. By independent genetic evidence we could identify that the region downstream of the *orf904* gene is the only non-coding region which cannot be deleted without losing the ability to replicate (Berkner and Lipps, unpublished). We therefore specifically analysed with DNAse I and with permanganate footprinting whether the full-length or the winged-helix domain deletion mutant bind sequence specifically 3′ of the *orf904* gene. We were however unable to detect a clear footprint in this region.

On a sizing column, the deletion mutant elutes as a monomer (data not shown). Possibly the protein dimerises on the DNA. Noteworthy two monomers of the transcriptional activator Rfx bind to a single binding-site on double-stranded DNA (35).

Circular dicroism (CD) spectroscopy allows the investigation of protein folding since α-helices and β-strands give rise to typical spectral features. The CD spectra of K752G844 show negative ellipticity at 220 nm indicative for α-helices (Figure 2B). From the molar ellipticity, the helical fraction can be estimated to be ∼20%. Further support that the deletion mutant K752G844 folds into a compact domain comes from a thermal

**Table 1.** DNA-binding activities of ORF904 full-length protein and deletion mutants

| | *K* (nM)/*n* ss | *K* (nM)/*n* ds | *K* (nM)/ss, ATP | *K* (nM)/*n* ds, ATP |
|---|---|---|---|---|
| ORF904 | 130–330/1.0–1.1 | 620–820/1.0–1.1 | 260–450/1.5 | 700–800/1.7–1.8 |
| N40C370 | 720–780/1.3 | 4270–4300/0.8–0.9 | 5000–6500 | >14 000 |
| N371 | 800–920/1.1–2.3 | 590–660/1.2–2.0 | 270–390/1.2–1.8 | 290–300/1.5–1.6 |
| N371K752 | 2400–2700 | 1700–1900 | 430–490/1.7–2.2 | 820–900/1.3–1.4 |
| K752G844 | 1500–1600/1.5–2.4 | 1700–2100/1.0–2.0 | | |

The dissociation constants *K* were determined by fluorescence anisotropy measurements with a single- or double-stranded 25-nt long DNA substrate. In most cases we observed positive cooperativity; in these cases, the data were analysed with the Hill equation and the concentration of half-maximal binding and the hill coefficient (*n*) are reported. Proteins with the helicase domain or the prim/pol domain were also tested in the presence of ATP. Shown is the range of at least two replicate measurements.

denaturation experiment. The CD signal is stable up to about 83°C (Figure 2C).

In aggregate our experiments underscore that the replication protein ORF904 contains an evolutionary conserved putative winged helix DNA-binding domain at its C-terminal end. The isolated domain folds and binds double- as well as single-stranded DNA. Unfortunately, we have not been able to identify the cognate binding site maybe because the isolated domain could lack cooperative interactions which could increase the affinity to DNA in the context of the whole protein (see also 'Discussion' section).

### The replication protein ORF904 binds DNA through at least three domains

To study the relative contributions of the three domains of ORF904 (Figure 1) to the DNA-binding activity, we constructed further deletion mutants. The prim/pol domain starts at amino acid 47 and ends at 247, however, our functional characterization of the primase activity in several deletion mutants allowed us to define the minimal domain for primase. This analysis shows that the amino acids 40–370 are required for full primase activity (17). In addition to this mutant comprising the minimal primase domain (termed N40C370) we constructed a deletion mutant starting at amino acid 371 (called N371) and a further deletion mutant starting at amino acid 527, before the core of the helicase domain. The latter deletion mutant was called N527. Another deletion mutant started at amino acid 371 and ended at amino acid 752, i.e. just before the winged helix domain. All deletion mutants were recombinantly expressed and purified essentially as the full-length protein (Supplementary Figure S1).

Next, we determined the binding affinity of the full-length protein and of the deletion mutants with the fluorescence anisotropy assay (Table 1). As substrate, we employed a 25 base long oligodeoxynucleotide labelled with a fluorescein at its 3′-end and the same oligodeoxynucleotide hybridised to its complementary strand. In the absence of ATP the full-length protein binds single- as well as double-stranded DNA alike or better than all deletion mutants. However, the deletion mutants differ in their affinity to single- versus double-stranded DNA. The prim/pol domain (N40C370) prefers single-stranded DNA and the helicase domain (N371 and N371K752, respectively) rather prefers double-stranded DNA. These findings are congruent with the enzymatic activities of both domains: the primase domain has a single-stranded substrate and the ATPase activity of ORF904 is better stimulated by double-stranded DNA (see below). Upon addition of ATP the binding affinities to the DNA change considerably. The prim/pol domain binds about a one order of magnitude weaker to single- and double-stranded DNA. This result is unexpected since ATP is required for the formation of the primer. In contrast, the addition of ATP strengthens the DNA binding of the isolated helicase domain (N371 or N371K752) but not in the context of the full-length protein. Whereas in the absence of ATP, the DNA-binding activities of the isolated domains add up synergistically to the DNA-binding strength of the full-length protein, in the presence of ATP the full-length protein binds less strong to DNA than the deletion mutant N370. It thus appears that the prim/pol domain is able to occlude the DNA-binding site of the helicase domain in the presence of ATP. These observations suggest that the binding of ATP to the prim/pol and/or the helicase domain has a strong effect on the conformation of the full-length protein. We tried to detect these conformational changes by limited proteolysis. However, using trypsin and chymotrypsin, we did not find significant changes of the proteolytic susceptibility of the full-length protein when comparing the protein in the presence and absence of ATP and/or DNA.

### ORF904 does not show an ATPase activity characteristic for processive helicases

Next, we studied the ATPase activity of the replication protein ORF904. We have previously observed that the ATPase activity of ORF904 is stimulated by double-stranded DNA but only weakly by single-stranded DNA and not by RNA and that an intact Walker A motif is required for ATPase activity (15). Since SF3 helicases form hexameric rings (36), we were interested to investigate whether ORF904 prefers long circular or short linear DNA as substrate. The first substrate could support a processive enzyme in a topological restrained configuration; in contrast, the second substrates would only allow short helicase movements and would require frequent rebinding and reassembly. Both substrates stimulate the ATPase activity of ORF904 equally well under a range of DNA substrate concentrations (Figure 3). Even under limiting DNA concentrations under which the rebinding

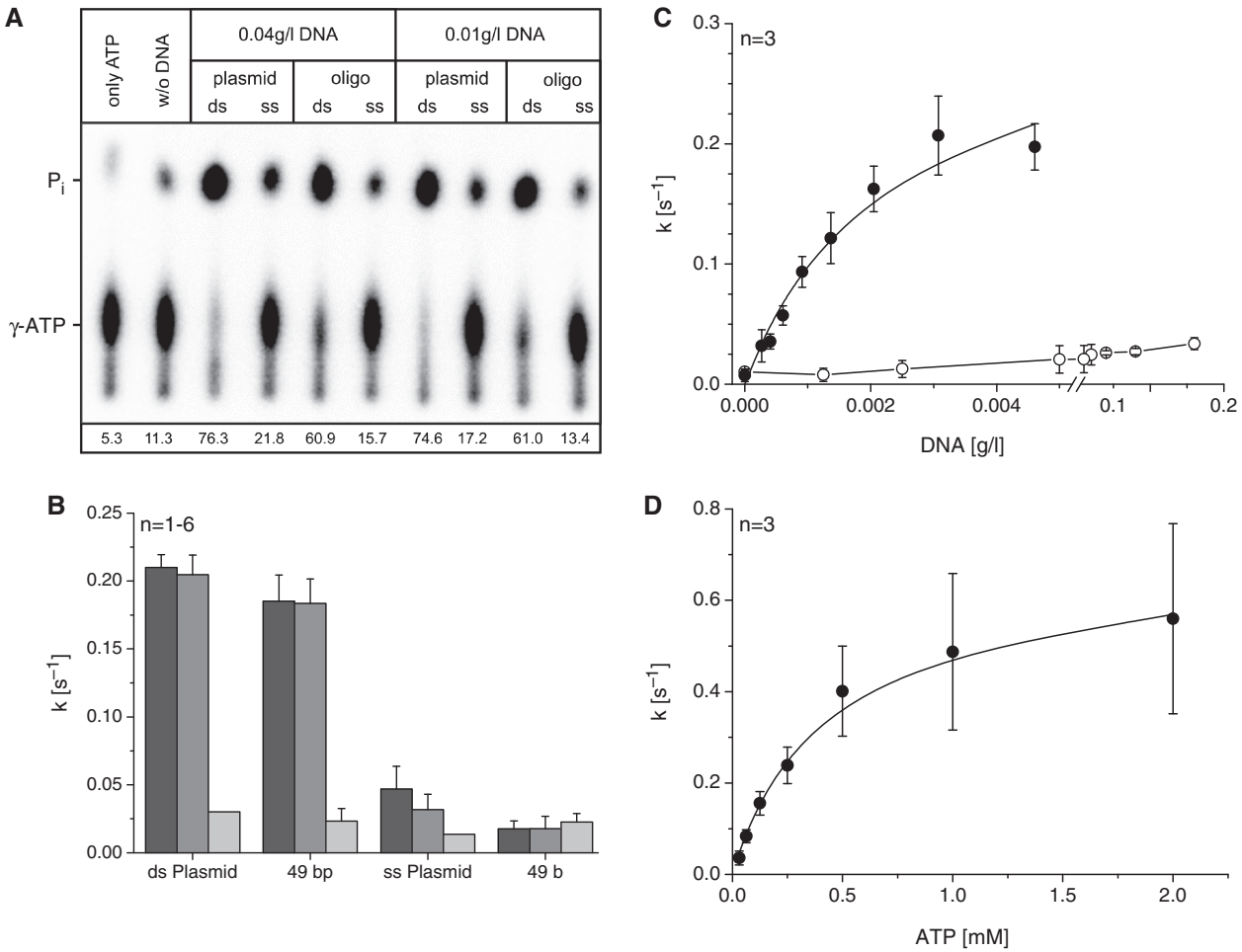


**Figure 3.** ATPase activity of the replication protein. (**A**) The ATPase activity of the replication protein was determined with [$\gamma$-$^{32}$P]-ATP followed by thin layer chromatography. Short linear and long circular double-stranded DNA strongly stimulates the ATPase activity. The assay was performed for 20 min at 80°C in the presence of 0.2 mM ATP. The amount of hydrolysis quantified with an Instantimager is given below the autoradiography. (**B**) Graphical representation of the substrate dependant ATPase activity 40 mg/l DNA (black) 10 mg/l DNA (dark grey) and 1 mg/l DNA (light grey). (**C**) Michaelis Menten kinetics of the ATPase activity. The experiment was performed at a constant ATP concentration of 0.16 mM. The apparent $K_m$ is 2.5 mg/l 49 bp DNA (filled circles). The same $K_m$ value was determined under saturating ATP concentrations of 2 mM ATP. With single-stranded DNA (open circle) the ATP hydrolysis is very weak. (**D**) Michaelis–Menten kinetics of the ATPase activity. The experiments were performed at a constant DNA concentration of 40 mg/l 49 bp DNA. The apparent $K_m$ is 0.42 mM ATP. Shown are the arithmetic means of three replicate measurements and as error bar the standard deviation.

to the DNA substrate is probably the slowest step the plasmid DNA is not preferred. Under these experimental conditions, a helicase with high processivity would display a higher ATPase activity with a plasmid than with short linear substrates. We quantified the ATPase activity in the presence of defined single- and double-stranded short substrates under varying ATP and DNA concentrations. We found that ssDNA stimulates the ATPase activity about 10× lower than dsDNA and that the half-maximal concentrations are at ~2.5 mg/l dsDNA (~70 nM) and 0.42 mM ATP (Figure 3). We previously measured a considerable higher $K_m$ value for ATP of 3 mM and a turnover number of 3.8 $s^{-1}$ (15). The reasons for the discrepancy with the present more detailed kinetic analysis are not known. The deletion mutant N527 was unable to hydrolyse ATP but the deletion mutant N371 displayed ATPase activity comparable to the full-length protein (data not shown).

## ORF904 has DNA translocation activity

In our previous work, we were not able to detect helicase activity of ORF904 when we used helicase substrates based on oligodeoxynucleotides (with 3′ and 5′ overhangs) hybridised to single-stranded plasmids (15). In contrast, when we now use a short-forked helicase substrate with a 25 base-long overhang and with a 25 bp-long duplex region, we consistently observed a weak unwinding activity of ORF904. However, the unwinding activity is quite low and we also observe destabilization of the forked substrate in the presence of protein but the absence of ATP (see below and Table 2).

Our failure to detect helicase activity on these substrates could also be related to the multidomain character of the protein. The N-terminal prim/pol domain is able to bind to single-stranded DNA (see above). It is therefore possible that under the experimental conditions (excess

**Table 2.** Unwinding rates for the full-length replication protein ORF904 and the deletion mutant N371

| Substrate | Protein | ATP (mM) | % Unwinding/min | % Net unwinding/min |
|---|---|---|---|---|
| Holliday junction | N371 | 0 | 0,0 | |
| | | 2 | 1.7–2.0 | 1.9 |
| | ORF904 | 0 | 0.9–1.1 | |
| | | 2 | 1.7–1.8 | 0.8 |
| Fork | N371 | 0 | 0.1 | |
| | | 2 | 3.1–3.4 | 3.2 |
| | ORF904 | 0 | 1.2–1.9 | |
| | | 2 | 5.6 | 4.1 |
| Triple-helix | N371 | 0 | 0.1–0.3 | |
| | | 2 | 1.1–1.3 | 1.0 |
| | ORF904 | 0 | 0.2–0.3 | |
| | | 2 | 1.2–1.3 | 1.0 |

The unwinding rates (shown are the ranges) were calculated from replicate kinetic experiments ($n \geq 2$, see Figure 4 for a representative gel). The unwinding reactions were analysed in the linear range of the assay (Holliday junction: 5–20 min, fork: 2–10 min, triple-helix: 0–15 min). The full-length protein has a high background of DNA destabilisation in the absence of ATP with the fork and the Holliday junction substrate.

of protein over DNA) the replication protein destabilises the duplex DNA by preferentially binding to single-stranded DNA.

We therefore used the deletion mutant N371 devoid of the prim/pol domain but still active in the ATPase assay (data not shown) to investigate the unwinding activity further. We first tested whether the deletion mutant N371 is able to unwind a forked substrate. The time course of the reaction in presence and absence of ATP clearly demonstrates that N371 possesses helicase activity (Figure 4) and that the substrate is stable at 60°C during the time course of the experiment. With this protein there is, in contrast to the full-length protein, only a marginal destabilization of the duplex in the absence of ATP (see also Table 2). A forked helicase substrate is the preferred substrate of ring-shaped helicases, such as the SF3 helicases or the replicative bacterial helicases of family 4. These helicases could bind to one of the single-stranded regions of the substrate and then actively transport the single-stranded DNA through the central pore. The other strand of the DNA is sterically excluded therefore leading to unwinding of the duplex stem of the forked substrate. Consistent with this unwinding mechanism, we do not observe unwinding activity with blunt-ended short duplex DNA and with DNAs presenting either a 3′ or a 5′ overhang neither with the full-length protein nor with the deletion mutant N371 (data not shown).

The analysis of the ATPase activity suggested that double-stranded DNA is the preferred substrate of the replication protein. It therefore seems unlikely that the replication protein translocates preferentially on single-stranded DNA since this type of DNA is only a poor stimulator of the ATPase activity. To investigate whether the replication protein is also able to translocate on double-stranded DNA we used a Holliday junction as a substrate for the helicase assay. This substrate does not offer a single-stranded region for initial binding. When we use the Holliday junction as substrate we observe unwinding (Figure 4). Possibly, the replication protein binds to one of the arms and when translocating against the junction two forks are released which are then subsequently unwound. An excess of unlabelled oligodeoxynucleotide ensures that the released labelled strand does not reanneal back to the Holliday junction. Since we do not observe any intermediates during the assay, it seems that the unwinding of the forks proceeds faster than the initial unwinding of the junction.

These assays showed that ORF904 possess weak unwinding activity and can translocate along single- and double-stranded DNA. In both cases, the substrates offered branch points which might be recognized by the protein. Next, we employed the triple-helix substrate without any branchings. In this experiment, ORF904 is incubated with a double-stranded DNA to which a labelled oligodeoxynucleotide is annealed. When this substrate is incubated with protein and ATP a partial but significant unwinding is observed with the full-length protein and the helicase domain protein N371 (Figure 5). It therefore appears that this protein can translocate along the double-stranded plasmid and displace the annealed third strand. While the unwinding activity observed with this artificial substrate is probably not relevant for plasmid replication, the results of these experiments reinforce that ORF904 has motor activity on double-stranded DNA. In the context of plasmid replication duplex unwinding or duplex distortion by some sort of mechanical force is needed to melt the replication origin.

The polarity of the translocation movement cannot be determined neither with the fork substrate nor with the Holliday junction substrate and with the triple helix assay. We therefore used the biotin-streptavidin displacement assay (37) to determine the polarity of the replication enzyme. As substrates we used oligodeoxynucleotides which carry a biotin moiety either at the 3′- or 5′-end. These substrates are first incubated with streptavidin and then with excess biotin. When incubated together with a ring-shaped helicase and ATP, the translocation of the helicase on the single-stranded DNA is able to displace the streptavidin tetramer. The full-length protein and the deletion mutant N371 are able to displace the streptavidin at the 5′-end but not at the 3′-end (Figure 6). Thus the protein translocates in 3′–5′ direction on single-stranded DNA. The polarity is the same as for other helicases of superfamily 3. The streptavidin displacement and the triple-helix assay were also used to determine which nucleotides are able to fuel the translocation activity: ATP, CTP and all dNTPs are able to serve as energy source. CTP and UTP are used less efficiently (data not shown).

Since double-stranded DNA stimulates the ATPase activity better than single-stranded DNA, we were interested to investigate whether ORF904 also translocates on double-stranded DNA in this assay. For this purpose a biotinylated oligodeoxynucleotide was hybridised with a complementary oligodeoxynucleotide. However although the Holliday junction substrate and the triple-helix assay clearly demonstrate that the replication protein is able to translocate along double-stranded DNA, the replication

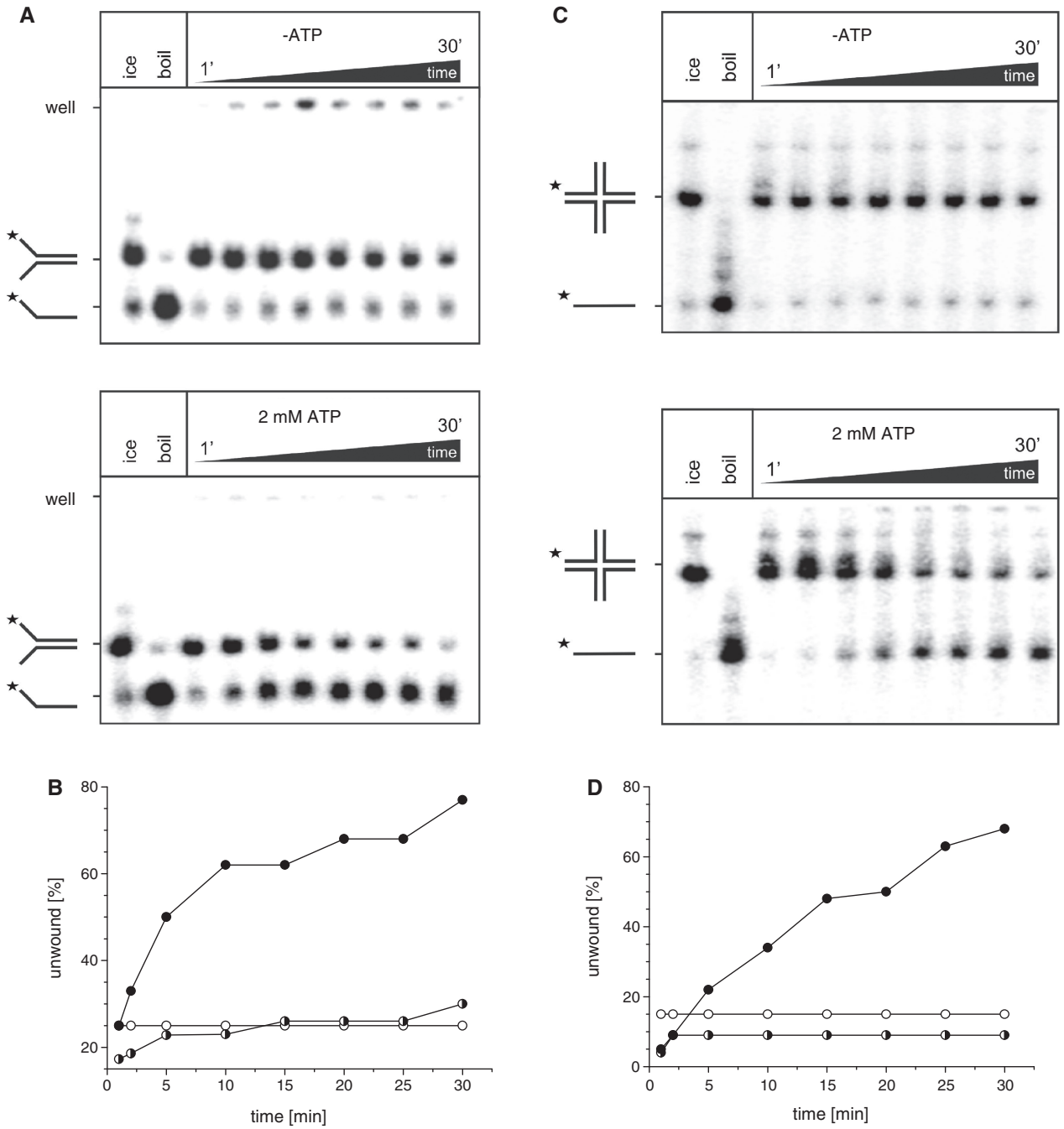


**Figure 4.** The deletion mutant N371 unwinds a forked substrate and a Holliday junction substrate. (**A**) Helicase assay with 0.5 μM N371 and the forked substrate. The kinetics of unwinding are followed over 30 min and analysed by native electrophoresis. (**B**) Graphical representation of the data in panel A. Solid circle: with protein and with 2 mM ATP, half-filled circle: with protein but without ATP, open circle: control reaction without ATP and without protein. (**C** and **D**) As panel A and B, but with Holliday junction substrate.

protein was unable to displace the streptavidin on double-stranded DNA. Possibly when encountering the 'road-block' streptavidin the replication protein looses track of the sugar phosphate backbone of the 3′–5′ strand and then slides backwards or disassembles. Under these conditions a productive streptavidin displacement will not be observed.

On the whole the helicase and translocation assays show that the replication protein can track along single- and double-stranded DNA but also that the protein does not appear to be a processive helicase as would be expected for a replicative helicase involved in plasmidal genome replication.

## Quaternary structure of ORF904

We have used gel-filtration chromatography and sedimentation analysis to investigate the oligomerization behaviour of ORF904. Most of the protein is lost on the

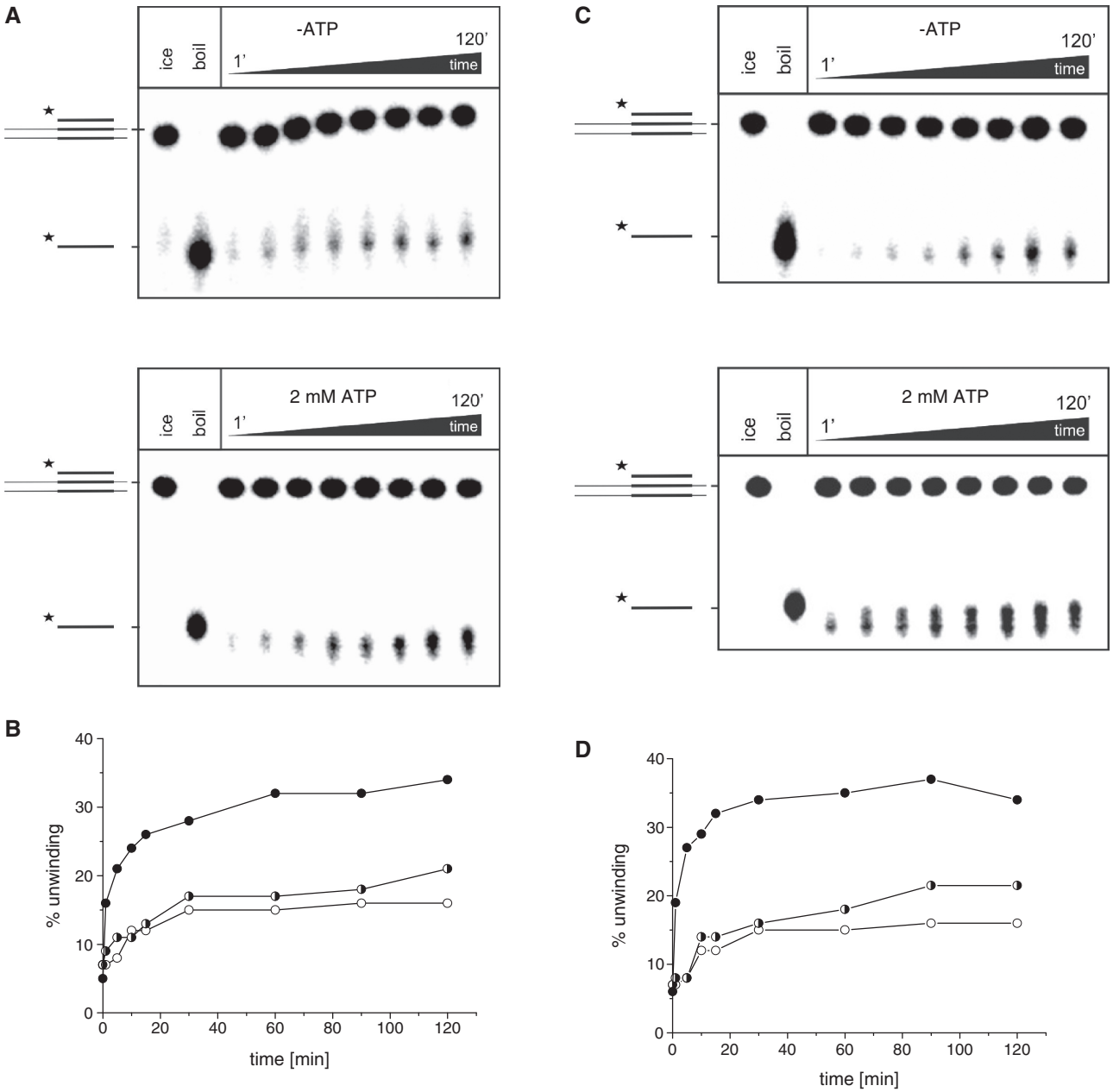


**Figure 5.** The full-length protein and the deletion mutant N371 unwind a triple helix substrate. (**A**) ORF904 (0.05 μM) was incubated with the triple helix substrate at 45°C.The kinetics of unwinding are followed over 120 min and analysed by native electrophoresis. (**B**) Graphical representation of the data in panel A. Solid circle: with protein and with 2 mM ATP, half-filled circle: with protein but without ATP, open circle: control reaction without ATP and without protein. (**C** and **D**) As panel A and B, but with the deletion mutant N371.

sizing column but the remaining protein eluted as a monomer. We suspect that ORF904 has a tendency to oligomerise and then aggregate on the column. The sedimentation analysis with the glycerol gradient clearly shows that ORF904 is a monomer in solution (data not shown). Even when the ultracentrifugation was carried out in the presence of ATP or AMPPnP and single- or double-stranded DNA, a multimeric protein species could never be clearly identified (Supplementary Figure S2). However, when we incubated the replication protein together with short double-stranded DNA and AMPPnP and analysed the reaction by cryo electron microscopy particles of a size of 12 nm were present (Figure 7). The formation of these particles was also observed with the deletion mutant starting at amino acid 371; in contrast no particles were present when either AMPPnP or DNA was missing or when the shorter deletion mutant N527 was tested. The latter sample may also serve as a negative control indicating that an unrelated *E. coli* protein is not responsible for the formation of these distinct particles. Single-particle analysis of the uranyl-acetate stained complexes revealed a ring-shaped structure with 6-fold symmetry. The particles have a diameter of ~15 nm. Possibly these particles represent hexamers of

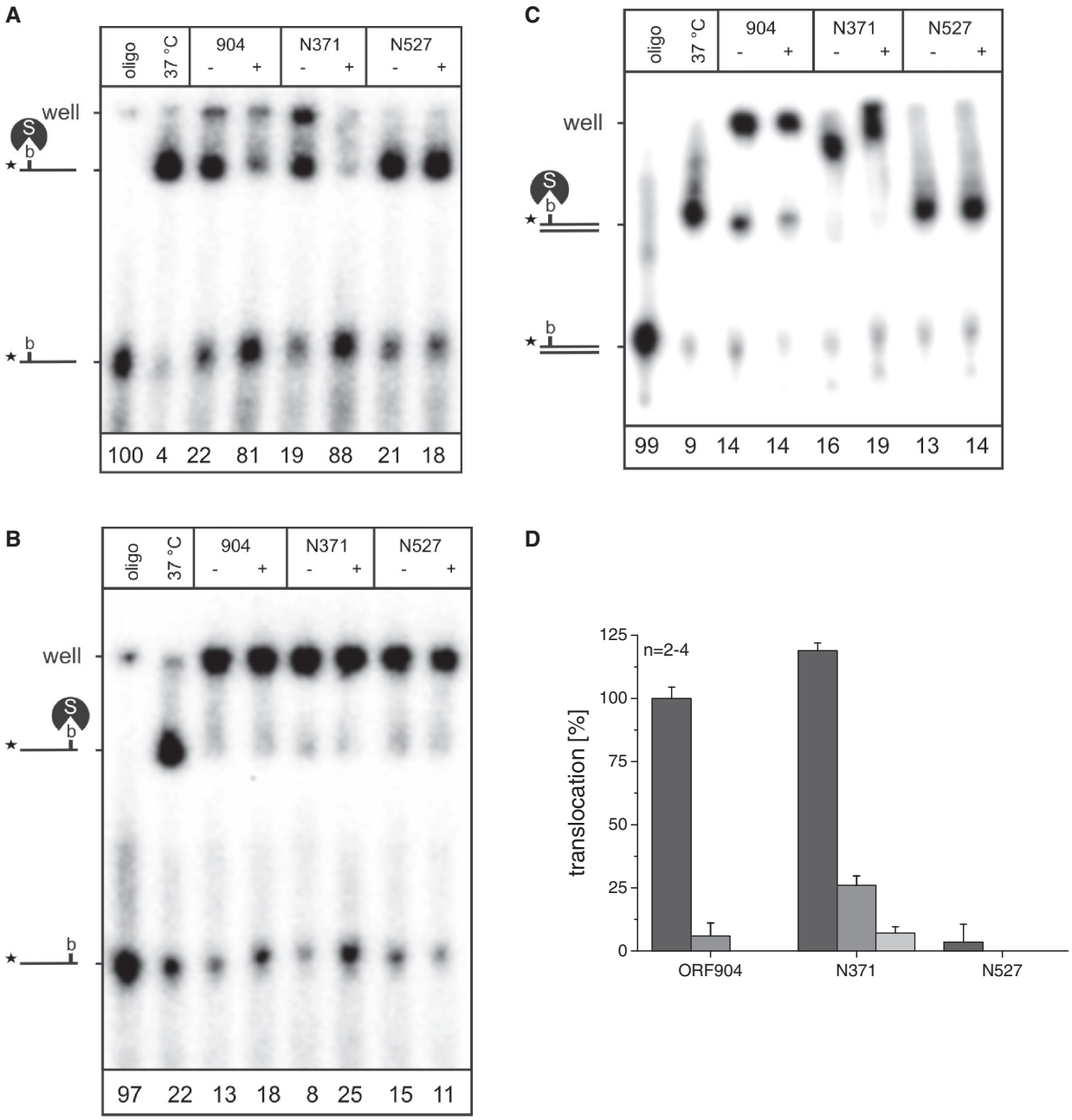


**Figure 6.** The replication protein has 3′–5′ translocation polarity. (**A**) Biotinylated oligodeoxynucleotide (lane oligo) is incubated with streptavidin and excess biotin (lane 37°C). In additional reactions the replication enzyme is added at 0.5 μM without (lane −) or with 2 mM ATP (lane +). The percentage of radioactivity corresponding to the biotinylated oligodeoxynucleotide was quantified and is given below the gel. Consistent ATP dependant streptavidin displacement was only seen with the single-stranded substrate with biotin at the 5′-end (**A**). In contrast, a displacement is neither observed for the oligodeoxynucleotide with the biotin moiety at the 3′-end (**B**) nor for the double-stranded substrate (**C**). Graphical representation of the translocation activity relative to the wild-type protein with the single-stranded DNA with biotin at the 3′-end (**D**): 3′-end, single-stranded (black); 3′-end, double-stranded (dark grey); 5′-end, single-stranded (light grey). Shown is the arithmetic mean of two to four replicate measurements and the standard deviation (error bars).

the replication protein assembled around DNA which are arrested by addition of the non-hydrolysable ATP analogue AMP-PnP. Since these complexes are neither observed by gel-filtration nor by sedimentation, the interaction between the subunits is probably quite weak. The quaternary structure observed with two independent electron microscopy techniques is consistent with the structures of other SF3 helicases and explains the translocation activity as observed in the helicase assays and the streptavidin displacement assays.

### A structural model of the helicase domain of ORF904

The amino acid sequence of ORF904 shows high sequence similarity to the helicase domain of the bacteriophage P4 α protein and allows identification of an ATPase domain (amino acids 408–827, COG3378) within the C-terminal

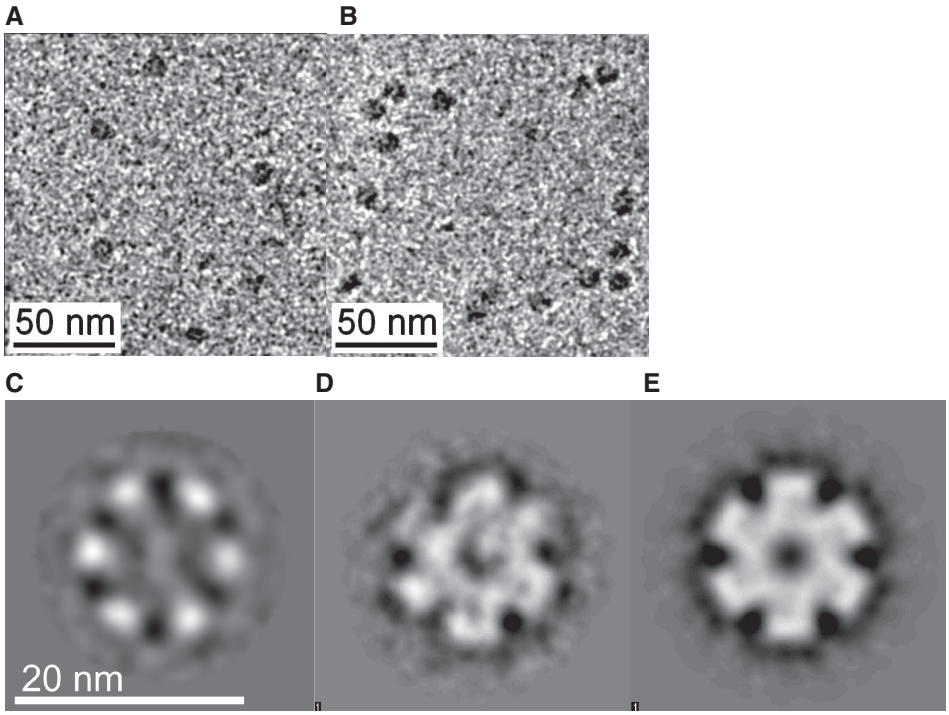


**Figure 7.** Electron microscopy of ORF904. (**A** and **B**) Cryo-TEM images of ORF904. ORF904 (A) and the deletion mutant N371 (B) were incubated with short duplex DNA and AMP-PnP. Complexes of 15 nm diameter can only be seen when ORF904 or N371 were incubated together with double-stranded DNA. No complexes were seen when the deletion mutant N527 was used or AMP-PnP or DNA were not included during preparation. (**C–E**) Single-particle analysis of negatively stained complexes. The eigenimage (C) suggests a 6-fold symmetry of the complex. The first classsum is shown in (D) and with symmetry imposed in (E).

part of ORF904. This conserved domain is distantly related to viral helicases of the superfamily 3 for which structural information are available. We therefore used the structures of the viral helicases from papilloma virus type 18 (PDB 1TUE), adenoassociated virus 2 (PDB 1S9H) and Simian Virus 40 large T antigen (PDB: 1SVL) as templates for homology modelling. Since ORF904 did not display significant sequence similarity to any of these helicases outside the core $AAA^+$-ATPase domain and the three helicases differ structurally outside the core domain we restrained our modelling efforts to the core of the helicase domain (amino acids 554 to 695 of ORF904). The model had an rms value of ~1.8 Å to the helicases of the papilloma virus and the adenoassociated virus and an rms value of 0.9 Å to the SV40 large T antigen. The SV40 large T antigen structure allowed us to construct a hexameric model with bound nucleotides. We used this homology model (Figure 8) to identify structural features which are important for the helicase activity of the superfamily 3 helicases, namely, a β-hairpin which reaches into the central pore of the hexameric ring and contacts the DNA to be transported and an arginine residue that assists the hydrolysis of the ATP bound at the neighbouring subunit as well as a putative lysine finger which also contacts the nucleotide pocket *in trans*.

### Characterisation of helicase domain point mutants

Within the β-hairpin we mutated the positively charged residues K656, K657 and R659 and constructed the respective alanine point mutants of the wild-type protein. The ATPase and translocation activities of these point mutants were analysed. We found that all point mutants and even the double-mutant K656A/K657A are still able to hydrolyse ATP. Since the β-hairpin is distant from the nucleotide-binding site these findings are not unexpected.

When we tested the ATPase activity of these point mutants, we unexpectedly found that these mutants have a relaxed specificity for duplex DNA (Figure 9). Whereas the wild-type protein is only stimulated well by double-stranded DNA, the β-hairpin mutants also display stimulated ATPase activity in the presence of single-stranded DNA. The magnitude of the observed effect is surprising. ATP hydrolysis requires an arginine finger from the neighbouring subunit (see below). Therefore multimerisation of the replication protein in the presence of DNA, e.g. formation of a hexameric ring as can be seen by electron microscopy, is a prerequisite to allow ATP hydrolysis to occur. The wild-type protein appears to form an ATPase competent ring around single-stranded DNA rather inefficiently; however, changing a single amino acid in the β-hairpin significantly improves the formation of a multimeric assembly which hydrolyses ATP in the presence of single-stranded DNA. It appears that the cumulative effect of six changed amino acids in the central pore of the ring is the cause of this rather large effect.

One of the point mutants, K657A, is unable to translocate on single-stranded DNA in the streptavidin displacement assay (Figure 9). We therefore suggest that

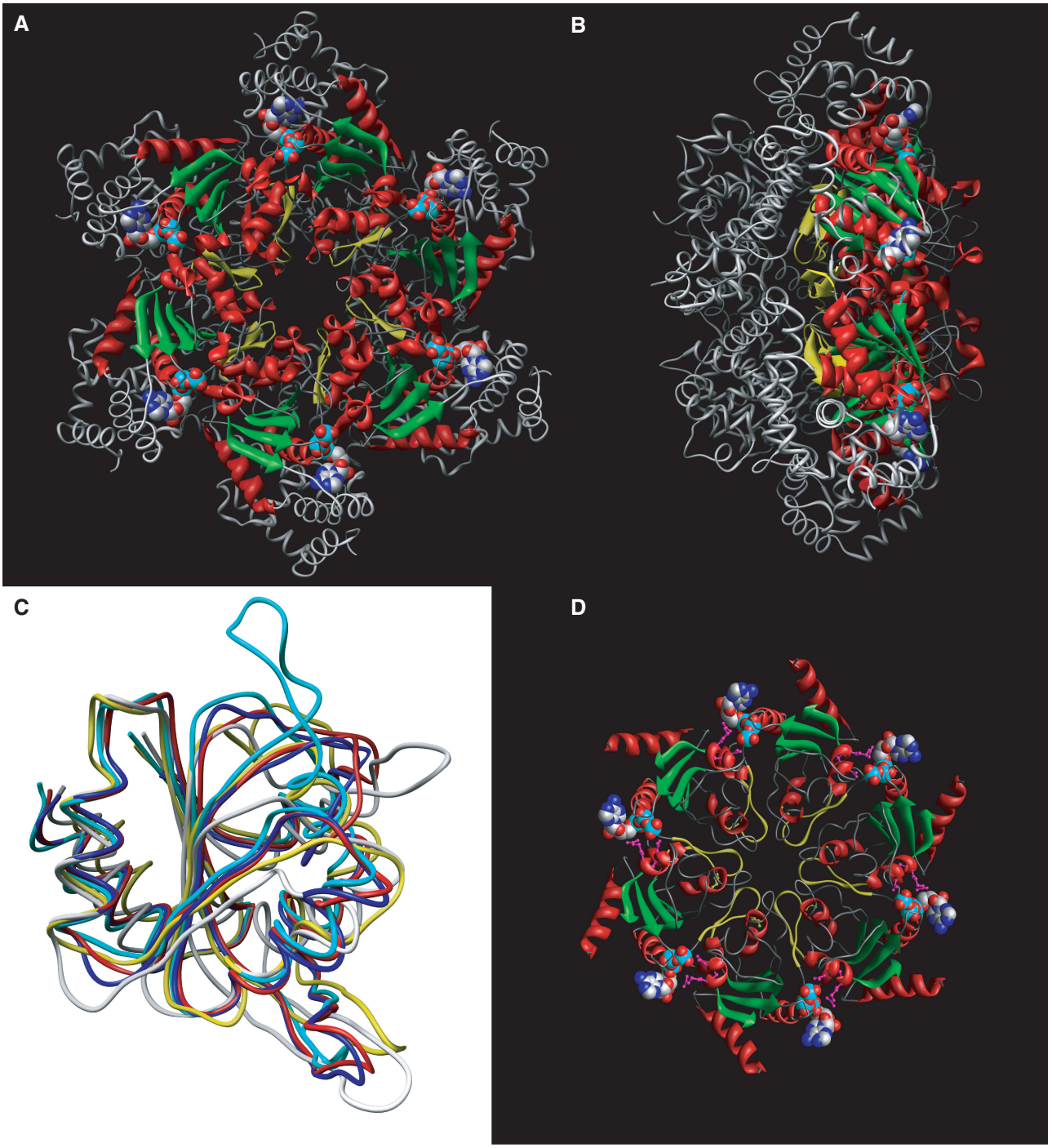


**Figure 8.** Model structure of ORF904. (**A** and **B**) Top and side view of the SV40 large T-antigen hexamer (PDB 1SVM). The core of the helicase domain with the $AAA^+$-ATPase fold is highlighted in green (β-strands) and red (α-helices). The ATP molecules are shown as spheres. The β-hairpin (amino acids 507-520) is shown in yellow. (**C**) Superposition of the $AAA^+$-ATPase domain of the three template structures. Papilloma virus type 18 (PDB 1TUE, amino acids 460–595, yellow), adenoassociated virus 2 (PDB 1S9H, amino acids 309–450, light grey) and Simian Virus 40 large T antigen (PDB: 1SVM, amino acids 400–546, cyan). Two models of the helicase core of ORF904 (amino acids 554–695) are shown in red and blue, respectively. (**D**) Hexameric model of the helicase core of ORF904 based on the quaternary structure of the SV40 large T antigen [coloured as in (A)]. The putative β-hairpin (yellow) starts at amino acid 653 and ends at amino acid 666.

this amino acid is critical for DNA transport through the pore. This residue could make a favourable electrostatic interaction with the negatively charged backbone. Disruption of this interaction might impede the DNA transport.

The proteins of the better-studied viral helicases have a single arginine residue ('arginine finger') which is implicated in ATP hydrolysis and intersubunit coordination required for DNA transport. At the homologous position in the crenarchaeal replication proteins and

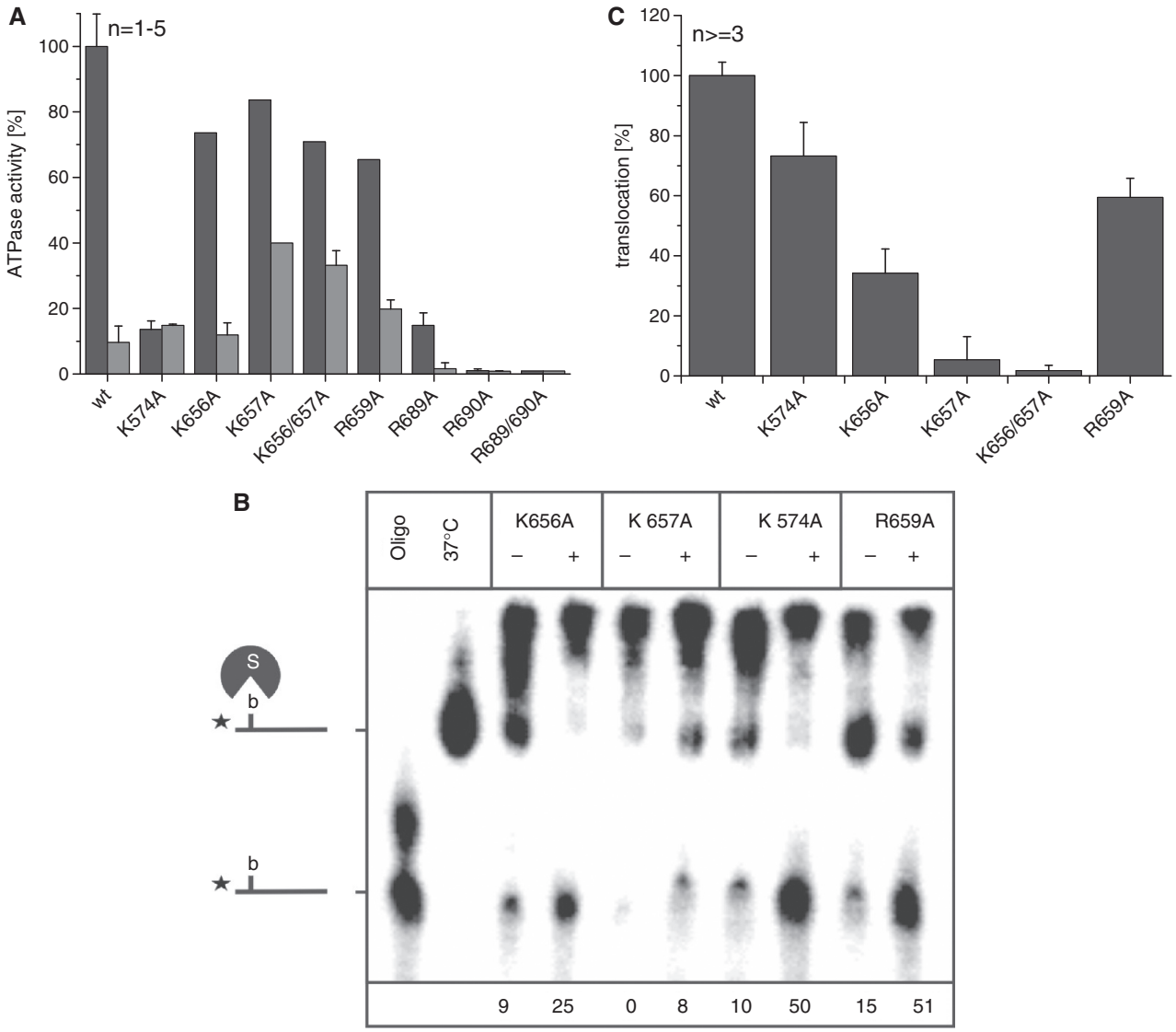


**Figure 9.** ATPase activity and translocation activity of point mutants. (**A**) ATPase activity of the point mutants in the presence of short double-stranded (black) and single-stranded (grey) DNA. The activity is expressed relative to the ATPase activity of the wild-type protein with a 49 bp DNA substrate. The β-hairpin mutants are better stimulated by a 49 base long oligodeoxynucleotide than the wild-type protein. The low activity of the mutant R690A is in agreement with R690 acting as an arginine finger (see text for details). (**B**) Translocation activity of the point mutants. The percentage of radioactivity corresponding to the biotinylated oligodeoxynucleotide was quantified and is given below the gel. (**C**) Graphical representation of the translocation activity relative to the wild-type protein. The point mutant K657A and the double mutant K656/657A are inactive in DNA translocation. Possibly the residue K657 is able to bind via an ionic interaction with the negatively charged backbone of the DNA.

in the bacteriophage P4 α protein, we found two neighbouring arginine residues. These two residues R689 and R690 as well as a putative 'lysine finger' K574 were mutated to alanine and the point mutants were studied. The point mutant R690A is almost unable to hydrolyse ATP suggesting that this residue is the *bona fide* arginine finger which directly contacts the ATP and stabilises the transition state during ATP hydrolysis (Figure 9). In contrast there is some residual ATPase activity with the point mutant R689A. It therefore appears that this residue could disturb the local structure and impede ATP hydrolysis only indirectly. Both point mutants as well as the double-mutant R689A/R690A are unable to translocate on single-stranded DNA as seen in the streptavidin displacement assay (data not shown).

The 'lysine finger' K574 appears to be not critical for the function of the replication enzyme, since ATP is still hydrolysed and the protein K574A is still active in the streptavidin displacement assay.

## DISCUSSION

The replication protein ORF904 is a multifunctional protein with two well-characterised functional domains: an N-terminal prim/pol domain with a highly sequence specific primase activity and a C-terminal helicase

domain of superfamily 3 with a 3′–5′ DNA translocation activity. The latter domain is associated with the functional uncharacterized D5_N subdomain and the putative winged-helix DNA-binding domain (see Figure 1). Although this protein and other distantly related proteins have been studied in detail, the exact molecular mechanism how the replication protein ORF904 participates in replicating the plasmid is still not well understood.

The replication origin of the plasmid pRN1 and those of the other crenarchaeal plasmids are not known. In the past it has been speculated that pRN1 is replicated through a rolling circle mechanism and bioinformatic evidence for a double- and a single-stranded origin have been presented (38). However, currently there is no experimental and bioinformatic support for an endonuclease which is the essential enzyme for initiation of rolling circle replication at the double-stranded origin (39). In addition, the ability of the replication protein ORF904 to synthesise a primer (40) argues against a rolling circle replication mechanism.

Typical plasmidal replication origins contain an AT-rich unwinding region and iterons which are targeted by a replication protein (41). The identification of an AT-rich region is problematic with a plasmid having a GC content of 37% and we were not able to discern a putative replication origin of pRN1 based on the above criteria. However, recent genetic experiments involving deletions of various parts of the pRN1 plasmid suggest that the region 3′ to the replication protein gene is essential (Berkner and Lipps, unpublished results). Possibly this non-coding part of the plasmid could function as a replication origin. But even with a closer look at this region, we were not able to identify iterons in this part of the plasmid.

Another unsettled issue concerning the replication of pRN1 is the question which protein recognises the replication origin. The plasmid pRN1 encodes two sequence-specific DNA-binding proteins: the protein ORF56 binds upstream of its own gene and appears to be involved in regulating the expression of the co-transcript *orf56*/*orf904*. This protein might therefore be involved in regulating the copy number of the plasmid. The second DNA-binding protein is the protein ORF80. This protein is highly conserved and binds to two neighbouring binding sites upstream of its own gene. We have proposed that this protein might recognize the replication origin (42). However, the interruption of the gene is tolerated (14) and we could also delete the binding sites and the complete gene without destroying the replication of the resulting shuttle plasmid in *Sulfolobus* (Berkner and Lipps, unpublished results). It therefore appears that either ORF56 has a dual function (acting as repressor and as replication initiation protein) or that the replication protein ORF904 recognises the replication origin on its own, possibly through its putative winged helix DNA-binding domain. In our view the latter possibility is more congruent with the experimental data: winged helix DNA-binding proteins are very common among plasmidal, viral and chromosomal replication initiator proteins (43). Secondly, the bacteriophage P4 α protein has a similar functional organisation as the protein ORF904. Both proteins have a N-terminal domain with primase activity—the α protein has a bacterial primase domain (DnaG) and ORF904 has a prim/pol domain—followed by the superfamily 3 helicase domain and a winged helix DNA-binding domain. The organisation of these three functions in a single polypeptide in two evolutionarily distant replicons is remarkable and suggests that this organisation is advantageous for a concerted molecular mechanism of replication initiation. Possibly this type of protein together with some host proteins such as a single-stranded DNA-binding protein could make up a simplified version of an 'orisome' capable of preparing the DNA for replication fork assembly.

Recently, the structure of two archaeal replication initiator proteins Cdc6/Orc1 in complex with origin DNA has been determined by two groups (21,44). These proteins have an $AAA^+$-ATPase domain followed by a winged-helix DNA-binding domain. Surprisingly, these proteins not only bind the DNA through the winged-helix DNA-binding domain but they also contact the DNA with a helical insertion in the canonical $AAA^+$ domain ['steric wedge' (45)]. This insert is specifically found in the chromosomal replication initiator proteins form bacteria, archaea and eukarya and is situated between the Walker A and B motifs of the $AAA^+$ domain. In contrast, the superfamily 3 helicases do not have this insert with two α-helices but they possess a different insert, a β-hairpin located between the Walker B and C motif [also called presensor-1 β-hairpin (PS1BH) (46)]. This β-hairpin is found in the mini chromosome maintenance proteins and the superfamily 3 helicases. In these proteins, the β-hairpin is believed to assist the directional transport of DNA through the central pore of the hexameric ring-shaped proteins. Remarkably the helical insert of the chromosomal initiator proteins, i.e. DnaA and Cdc6/Orc1, and the β-hairpin of the SF3/MCM proteins are on the same side of the protein (Supplementary Figure S3). We therefore suggest that the monomeric replication protein ORF904 could initially contact the DNA through the prim/pol domain, the winged-helix DNA-binding domain and the β-hairpin of the $AAA^+$ domain. Conformational changes of the replication protein, possibly assisted by ATP binding and hydrolysis could lead to duplex destabilisation of the origin DNA. In later steps, multimerisation of the replication protein around single DNA strands could prepare the DNA for priming and subsequent fork assembly.

This model could reconcile the high ATPase activity of the replication protein in presence of double-stranded DNA with its low translocation activity on these substrates. In this model, ATP is not only hydrolysed during DNA transport through the central pore but also when multimeric protein assemblies bind to double-stranded DNA and use the chemical energy for duplex distortion. However under *in vitro* conditions, a productive origin unwinding is not observed due to the lack of appropriate conditions of the assay or the lack of host proteins which are needed for subsequent steps of assembly.

The results presented here stress the central role of the replication protein ORF904 for replication initiation and

give a framework for further biochemical and genetic studies on replication initiation in this model system.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Deutsche Forschungsgemeinschaft grant Li193/6 (to G.L.); Deutsche Forschungsgemeinschaft (SFB 481) (to M.D.). Funding for open access charge: Deutsche Forschungsgemeinschaft.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Kornberg,A. and Baker,T.A. (1992) *DNA Replication 2*. W.H. Freeman and Company, New York.
2. Sowers,K.R. and Gunsalus,R.P. (1988) Plasmid DNA from the acetotrophic methanogen Methanosarcina acetivorans. *J. Bacteriol.*, **170**, 4979–4982.
3. Ng,W.L. and Dassarma,S. (1993) Minimal replication origin of the 200-kilobase Halobacterium plasmid pNRC100. *J. Bacteriol.*, **175**, 4584–4596.
4. Erauso,G., Marsin,S., Benbouzid,R.N., Baucher,M.F., Barbeyron,T., Zivanovic,Y., Prieur,D. and 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.
5. Garrett,R.A., Redder,P., Greve,B., Brugger,K., Chen,L. and She,Q. (2005) Archaeal plasmids. In Funnel,B.E. and Phillips,G.J. (eds), *Plasmid Biology*. ASM Press, Washington, DC, pp. 377–392.
6. Greve,B., Jensen,S., Brugger,K., Zillig,W. and Garrett,R.A. (2004) Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. *Archaea*, **1**, 231–239.
7. Greve,B., Jensen,S., Phan,H., Brugger,K., She,Q., Zillig,W. and Garrett,R.A. (2005) Novel RepA-MCM proteins encoded in plasmids pTAU4, pORA1 and pTIK4 from Sulfolobus neozealandicus. *Archaea*, **1**, 319–325.
8. Marsin,S. and Forterre,P. (1998) A rolling circle replication initiator protein with a nucleotidyl-transferase activity encoded by the plasmid pGT5 from the hyperthermophilic archaeon Pyrococcus abyssi. *Mol. Microbiol.*, **27**, 1183–1192.
9. Keeling,P.J., Klenk,H.P., Singh,R.K., Feeley,O., Schleper,C., Zillig,W., Doolittle,W.F. and Sensen,C.W. (1996) Complete nucleotide sequence of the Sulfolobus islandicus multicopy plasmid pRN1. *Plasmid*, **35**, 141–144.
10. Peng,X., Holz,I., Zillig,W., Garrett,R.A. and 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.
11. Lipps,G., Stegert,M. and Krauss,G. (2001) 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.
12. Berkner,S. and Lipps,G. (2007) Characterization of the Transcriptional Activity of the Cryptic Plasmid pRN1 from Sulfolobus islandicus REN1H1 and Regulation of Its Replication Operon. *J. Bacteriol.*, **189**, 1711–1721.
13. Lipps,G., Ibanez,P., Stroessenreuther,T., Hekimian,K. and Krauss,G. (2001) 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.
14. Berkner,S., Grogan,D., Albers,S.V. and Lipps,G. (2007) Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for Sulfolobus acidocaldarius and Sulfolobus solfataricus, model organisms of the (cren-) archaea. *Nucleic Acids Res.*, **35**, e88.
15. Lipps,G., Rother,S., Hart,C. and Krauss,G. (2003) A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. *EMBO J.*, **22**, 2516–2525.
16. Lipps,G., Weinzierl,A.O., von Scheven,G., Buchen,C. and Cramer,P. (2004) Structure of a bifunctional DNA primase-polymerase. *Nat. Struct. Mol. Biol.*, **11**, 157–162.
17. Beck,K. and Lipps,G. (2007) Properties of an unusual DNA primase from an archaeal plasmid. *Nucleic Acids Res.*, **35**, 5635–5645.
18. Forterre,P., Bergerat,A. and Lopez,G.P. (1996) The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol. Rev.*, **18**, 237–248.
19. Mott,M.L. and Berger,J.M. (2007) DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.*, **5**, 343–354.
20. Clarey,M.G., Erzberger,J.P., Grob,P., Leschziner,A.E., Berger,J.M., Nogales,E. and Botchan,M. (2006) Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nat. Struct. Mol. Biol.*, **13**, 684–690.
21. Dueber,E.L., Corn,J.E., Bell,S.D. and Berger,J.M. (2007) Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science*, **317**, 1210–1213.
22. 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.
23. Ilyina,T.V. and Koonin,E.V. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res.*, **20**, 3279–3285.
24. Gai,D., Zhao,R., Li,D., Finkielstein,C.V. and Chen,X.S. (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell*, **119**, 47–60.
25. Enemark,E.J. and Joshua-Tor,L. (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, **442**, 270–275.
26. Firman,K. and Szczelkun,M.D. (2000) Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *EMBO J.*, **19**, 2094–2102.
27. Kastner,B., Fischer,N., Golas,M.M., Sander,B., Dube,P., Boehringer,D., Hartmuth,K., Deckert,J., Hauer,F., Wolf,E. *et al.* (2008) GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods*, **5**, 53–55.
28. Golas,M.M., Sander,B., Will,C.L., Luhrmann,R. and Stark,H. (2003) Molecular architecture of the multiprotein splicing factor SF3b. *Science*, **300**, 980–984.
29. van Heel,M. and Frank,J. (1981) Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, **6**, 187–194.
30. van Heel,M. and Stoffler-Meilicke,M. (1985) Characteristic views of E. coli and B. stearothermophilus 30S ribosomal subunits in the electron microscope. *EMBO J.*, **4**, 2389–2395.
31. Dube,P., Tavares,P., Lurz,R. and van Heel,M. (1993) The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.*, **12**, 1303–1309.
32. Sander,B., Golas,M.M. and Stark,H. (2003) Corrim-based alignment for improved speed in single-particle image processing. *J. Struct. Biol.*, **143**, 219–228.
33. van Heel,M., Harauz,G., Orlova,E.V., Schmidt,R. and Schatz,M. (1996) A new generation of the IMAGIC image processing system. *J. Struct. Biol.*, **116**, 17–24.
34. Soding,J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics.*, **21**, 951–960.
35. Gajiwala,K.S., Chen,H., Cornille,F., Roques,B.P., Reith,W., Mach,B. and Burley,S.K. (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature*, **403**, 916–921.
36. Hickman,A.B. and Dyda,F. (2005) Binding and unwinding: SF3 viral helicases. *Curr. Opin. Struct. Biol.*, **15**, 77–85.
37. Morris,P.D., Byrd,A.K., Tackett,A.J., Cameron,C.E., Tanega,P., Ott,R., Fanning,E. and Raney,K.D. (2002) Hepatitis C virus NS3 and simian virus 40 T antigen helicases displace streptavidin from 5′-biotinylated oligonucleotides but not from 3′-biotinylated oligonucleotides: evidence for directional bias in translocation on single-stranded DNA. *Biochemistry*, **41**, 2372–2378.

38. Kletzin,A., Lieke,A., Urich,T., Charlebois,R.L. and Sensen,C.W. (1999) Molecular analysis of pDL10 from Acidianus ambivalens reveals a family of related plasmids from extremely thermophilic and acidophilic archaea. *Genetics*, **152**, 1307–1314.
39. Khan,S.A. (1997) Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.*, **61**, 442–455.
40. Lipps,G. (2004) The replication protein of the Sulfolobus islandicus plasmid pRN1. *Biochem. Soc. Trans.*, **32**, 240–244.
41. del Solar,G., Giraldo,R., Ruiz-Echevarria,M.J., Espinosa,M. and Diaz-Orejas,R. (1998) Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.*, **62**, 434–464.
42. Lipps,G. (2006) Plasmids and viruses of the thermoacidophilic crenarchaeote Sulfolobus. *Extremophiles*, **10**, 17–28.
43. Giraldo,R. (2003) Common domains in the initiators of DNA replication in bacteria, archaea and eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol. Rev.*, **26**, 533–554.
44. Gaudier,M., Schuwirth,B.S., Westcott,S.L. and Wigley,D.B. (2007) Structural basis of DNA replication origin recognition by an ORC protein. *Science*, **317**, 1213–1216.
45. Erzberger,J.P., Mott,M.L. and Berger,J.M. (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.*, **13**, 676–683.
46. Iyer,L.M., Leipe,D.D., Koonin,E.V. and Aravind,L. (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.*, **146**, 11–31.
47. Yeo,H.J., Ziegelin,G., Korolev,S., Calendar,R., Lanka,E. and Waksman,G. (2002) Phage P4 origin-binding domain structure reveals a mechanism for regulation of DNA-binding activity by homo- and heterodimerization of winged helix proteins. *Mol. Microbiol.*, **43**, 855–867.