GINS, a central nexus in the archaeal DNA replication fork

Nina Marinsek^{1*}, Elizabeth R. Barry^{1*}, Kira S. Makarova², Isabelle Dionne¹, Eugene V. Koonin²
& Stephen D. Bell^{1†}

¹MRC Cancer Cell Unit, Hutchison MRC Research Centre, Cambridge, UK, and ²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

In eukaryotes, the GINS complex is essential for DNA replication and has been implicated as having a role at the replication fork. This complex consists of four paralogous GINS subunits, Psf1, Psf2, Psf3 and Sld5. Here, we identify an archaeal GINS homologue as a direct interaction partner of the MCM helicase. The core archaeal GINS complex contains two subunits that are poorly conserved homologues of the eukaryotic GINS subunits, in complex with a protein containing a domain homologous to the DNA-binding domain of bacterial RecJ. Interaction studies show that archaeal GINS interacts directly with the heterodimeric core primase. Our data suggest that GINS is important in coordinating the architecture of the replication fork and provide a mechanism to couple progression of the MCM helicase on the leading strand with priming events on the lagging strand.

Keywords: replication; MCM; GINS; archaea; DNA

EMBO reports (2006) 7, 539–545. doi:10.1038/sj.embor.7400649

INTRODUCTION

The archaeal DNA replication apparatus is fundamentally related to that of eukaryotes (Kelman & Kelman, 2003). In general, the archaeal machinery is a simplified version of that found in eukaryotes; for example, the presumptive replicative helicase, the MCM complex, is a heterohexameric in eukaryotes but a homo-multimer in archaea.

The eukaryotic GINS complex is composed of four subunits, Psf1, Psf2, Psf3 and Sld5, that are represented in all sequenced eukaryotic genomes (Kanemaki *et al*, 2003; Kubota *et al*, 2003; Takayama *et al*, 2003). Sequence comparisons show that these four subunits are distantly related to one another and, accordingly,

seem to be ancient paralogues (Makarova *et al*, 2005). By contrast, most archaea encode a single GINS homologue.

GINs is essential for both initiation and elongation phases in eukaryotic DNA replication; however, its molecular basis of action is unclear at present. GINS binding to origins of replication requires MCM and is important for recruitment of the DNA polymerase (Kanemaki *et al*, 2003; Kubota *et al*, 2003; Takayama *et al*, 2003). There is increasing evidence that, although initially recruited to replication origins, GINS moves with the replication fork (Calzada *et al*, 2005). In the current work, we identify an archaeal GINS complex and show that it directly physically interacts with the archaeal MCM. Additional interaction studies implicate GINS as a central nexus in the replication fork, with the potential to act as a bridge between MCM and primase.

RESULTS AND DISCUSSION

In the course of a two-hybrid screen to identify potential binding partners of the MCM protein of the hyperthermophilic archaeon *Sulfolobus solfataricus*, we identified open reading frame SSO0772 as a candidate interaction partner. As the initial construct identified did not contain the full-length SSO0772 open reading frame, we prepared plasmid constructs that would express full-length SSO0772 protein fused to GAL4 DNA-binding domain (DBD) or activation domain (AD) in yeast. Yeast cells were transformed with these constructs together with MCM-DBD- and MCM-AD-expressing plasmids. As can be seen in Fig 1A, the SSO0772 protein interacts with MCM in both DBD and AD fusion contexts. Next, we tested the ability of SSO0772 to interact with amino- and carboxy-terminally truncated MCM. The results shown in Fig 1B indicate that SSO0772 interacts with the N-terminal 266 residues of MCM.

As shown previously, SSO0772 is one of the archaeal homologues of the eukaryotic GINS subunits (Makarova *et al*, 2005). Iterative database searches using the PSI-BLAST program with the predicted protein sequence of SSO0772 showed that it is most similar to Psf2 and Psf3 but also shows significant, though weaker, similarity to the other two GINS components, Sld5 and Psf1. Previous analyses have shown that the four eukaryotic GINS subunits are related to one another and, accordingly, must have evolved from a common ancestor by gene duplication events

¹MRC Cancer Cell Unit, Hutchison MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK

²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, Room 5N503, 8600 Rockville Pike, Bethesda, Maryland 20894, USA

*These authors contributed equally to this work

†Corresponding author. Tel: +44 1223 763311; Fax: +44 1223 763311; E-mail: sb419@hutchison-mrc.cam.ac.uk

Fig 1 | An archaeal GINS homologue binds to MCM. (A) Yeast two-hybrid analysis showing interaction between SSO0772/Gins23 and MCM. Yeast cells were transformed with plasmids containing DNA-binding domain (DBD) or activation domain (AD) fused to MCM (M) or the SSO0772/Gins23 open reading frame (G). 'V' indicates empty vector. (B) Two-hybrid analyses with constructs containing the amino-terminal 266 amino acids of MCM (N) or residues 267–686 of MCM (C) fused to DBD and the SSO0772/Gins23 fused to AD. (C) Diagram of the genomic locus encoding MCM and SSO0772 (Gins23). Arrows indicate the direction of transcription. The *gins23* and *mcm* open reading frames overlap by 11 bp. SSO0772 encodes a putative RNA-binding protein. Black triangles indicate the positions of the primers used in the reverse transcription-PCR (RT-PCR) shown in (D). (D) Ethidium bromide-stained agarose gel of the products of RT-PCR using primer pairs A + B or A + C with or without prior treatment of RNA with reverse transcriptase. Size markers are shown (M) and the size (in kilobases) of relevant markers is indicated on the left.

(Makarova *et al*, 2005). As SSO0772 is most similar to Psf2 and Psf3, we propose the name *gins23* for this gene and will use this nomenclature in the following discussion.

In the *S. solfataricus* genome (She *et al*, 2001), the *gins23* gene is located immediately upstream of the gene for the MCM protein (Fig 1C). Indeed, the two open reading frames overlap by 11 bp. This organization is suggestive of co-transcription of the *gins23* and *mcm* genes. We tested this hypothesis by performing reverse transcription-PCR (RT-PCR) using primers within the *gins23* and *mcm* open reading frames. The data confirm that a transcript containing Gins23 and MCM coding sequences can be detected in RNA isolated from *S. solfataricus* cells, indicative of coexpression of the genes (Fig 1D).

Gins23 was expressed in *Escherichia coli* as a recombinant C-terminally hexahistidine-tagged protein. The protein was highly insoluble and could only be purified in the presence of 8 M urea. The insolubility did not seem to be a function of the presence of the His₆ tag on the protein, as native-length, non-tagged, protein expressed in *E. coli* was similarly insoluble (data not shown). The purified His-tagged protein was used to generate rabbit polyclonal antisera. The antisera recognized a single protein of 21 kDa in extracts prepared from *S. solfataricus* cells (Fig 2A). Western blotting of whole-cell extracts prepared from logarithmically growing and stationary-phase *S. solfataricus* (Fig 2B) showed that, like MCM, the protein was still present in non-replicating cells, in contrast to the initiator protein Cdc6-1 (Robinson *et al*, 2004).

To test whether the MCM-Gins23 interaction detected in the yeast two-hybrid assay was representative of a physiologically relevant interaction in *S. solfataricus* cells, we carried out immunoprecipitation assays. The results indicate that Gins23 and MCM can be co-immunoprecipitated. Further, the insensitivity of the co-immunoprecipitation to the presence of ethidium bromide or DNaseI (data not shown) indicates that DNA is unlikely to be acting as a bridge between MCM and Gins23 (Fig 2C). We tested the ability of MCM and Gins23 to co-precipitate from extracts prepared from cells in early and mid-logarithmic growth and also from stationary phase. We find a very modest increase in co-immunoprecipitation of MCM with Gins23 from the extract prepared from mid-logarithmic cells, perhaps indicative of a higher proportion of cycling cells in this population (Fig 2D).

Whereas recombinant Gins23 was chronically insoluble in *E. coli* cells, native Gins23 seemed to be highly soluble in extracts from *S. solfataricus* cells prepared in salt concentrations ranging from 50 mM to 1 M NaCl (data not shown). We speculated therefore that the endogenous Gins23 might be associated with other proteins in *S. solfataricus*, thereby facilitating its solubility. To test this, we undertook purification of the endogenous Gins23. First, we fractionated crude cell extract by anion exchange chromatography and then used immunoaffinity chromatography to isolate endogenous Gins23-containing complex. As can be seen in Fig 3A, three proteins eluted specifically from the anti-Gins23 antisera column. These were identified by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry as Gins23 and the products of the SSO0295 and SSO1049 open reading frames. The immediate genomic environment of SSO0295 does not contain any genes of known or predicted function in DNA replication. Intriguingly, however, as shown in Fig 3B, SSO1049 is encoded immediately downstream of the gene

Fig 2 | (A) Anti-Gins23 antisera recognize a single band in *Sulfolobus* whole-cell extracts. Western blots were performed on 10 µg of whole-cell extracts with pre-immune (Pre) or anti-Gins23 (Imm) antisera. (B) Western blotting of extracts prepared from cells collected at the indicated optical density, corresponding to early, mid- and late logarithmic growth. (C) Immunoprecipitation of an endogenous GINS–MCM complex using anti-Gins23 antisera. Immunoprecipitation assays were performed, as indicated, in the presence or absence of 24 µg/ml ethidium bromide with either pre-immune (PI) or anti-Gins23 (I) antisera. Lanes marked 'In' contain input material. Unbound and bead-bound fractions were recovered and analysed by western blotting using anti-Gins23 sera (upper panel) or anti-MCM sera (lower panel). (D) Immunoprecipitation of the endogenous GINS–MCM complex using anti-Gins23 antisera from extracts prepared from cells in early (OD_{600 nm} = 0.052) and mid-logarithmic (OD_{600 nm} = 0.72) growth and stationary phase (OD_{600 nm} = 1.02). Labelling is as in (C).

for the catalytic subunit of the heterodimeric DNA primase, in an operon-like organization that also contains the gene for the sliding clamp subunit PCNA2 (Dionne *et al*, 2003) in addition to ribosome- and translation-associated factors.

PSI-BLAST searches indicated that the protein encoded by SSO0295 was related to the DBD of the bacterial RecJ family of single-stranded DNA exonucleases (Yamagata *et al*, 2002); however, SSO0295 lacked a counterpart to the C-terminal, catalytic domain of RecJ. Hereinafter, we refer to this protein as RecJdbh, after RecJ DNA-binding domain homologue (data not shown). Extensive PSI-BLAST searches with the sequence of

SSO1049 showed that it was a highly diverged member of the GINS family of proteins (Fig 3C), more closely related to Psf1 and Sld5 than to Psf2 and Psf3. Accordingly, we shall refer to this gene as *gins15*.

To test that the immunoaffinity approach used to purify the complex described above had not led to artefactual co-purification of the three proteins, we undertook a partial purification of the putative Gins23-containing complex using conventional chromatography (Fig 3D). As can be seen in Fig 3E, the two GINS family members and RecJdbh co-purified in eight steps, supporting the notion that they exist as a stable complex in cells.

We next performed further interaction studies to investigate the architecture of the Gins23–Gins15–RecJdbh complex and to determine whether further interactions could be detected with other DNA replication-associated proteins. A two-hybrid matrix indicated that Gins15 could interact with itself, with Gins23 and with RecJdbh. As described above, Gins23 interacted with MCM and also with both large and small subunits of the archaeal primase (Fig 4A). No interactions could be detected between any of Gins23, Gins15 or RecJdbh and Cdc6-1, Cdc6-2, Cdc6-3, SSB, UDG1, DNA polB1, Fen1, the two RFC subunits or any of the three *Sulfolobus* PCNA subunits (data not shown).

To test further the interactions with primase and MCM, we coexpressed Gins23 and Gins15 in *E. coli*. This permitted purification of a soluble complex containing both proteins in an equimolar ratio (Fig 4B). Analytical gel filtration gave an apparent molecular mass of 67 kDa. As a dimer of Gins23 and Gins15 would have a mass of 37 kDa, it seems that the core archaeal GINS complex is tetrameric, containing a dimer of dimers. The tetrameric structure of the archaeal GINS complex is in good agreement with the proposed quaternary structure of the eukaryal GINS complex (Kubota *et al*, 2003; Takayama *et al*, 2003).

The purified archaeal GINS complex was next used in 'pull-down' assays with His₆-tagged primase (tagged on the PriL subunit; Lao-Sirieix & Bell, 2004) and His₆-tagged MCM. As can be seen in the upper panels of Fig 4C,D, GINS was retained on columns containing MCM or primase, but not by the matrix alone, or by matrix containing PCNA2 (not shown). Finally, we tested whether addition of the *S. solfataricus* GINS complex to MCM helicase assays or primase assays could have any influence on the activity of primase or MCM. However, no significant effects could be detected (data not shown).

Our results describe the first characterization of an archaeal GINS complex. As shown in Fig 4E, our data indicate that the complex is a tetramer containing two copies of Gins15 and Gins23. As Gins15 can self-interact and also bind to Gins23, we propose that a Gins15 homodimer lies at the heart of the complex, with each Gins15 protomer binding to one Gins23 subunit. These data are in agreement with the proposed architecture of the eukaryal GINS complex (Kanemaki *et al*, 2003; Kubota *et al*, 2003; Takayama *et al*, 2003).

The interaction of Gins23 with primase and MCM suggests that GINS could act as a molecular bridge between MCM and the primase. We have tested for, but failed to detect, direct interactions between the archaeal primase and MCM (Fig 4A, and C,D, lower panels). This contrasts with the situation in bacteria, in which the primase DnaG forms a functional complex with the helicase DnaB (Soultanas, 2005). However, there is an important difference between the bacterial helicase and the MCM

C

Xenlae_37999519_Psf2	66 DVEKLEAIRDER	3 TFTPMPSPY-YMELTKLLNHA	7 DEIRLVLDVMDTRIAKRL	18 TLMBINTIGTFESLN	165	14 VTIPNPSLD	10 PNFSLPVSPLML	INLKQRC	56
Sacscce_6322389_YJL072C	93 TTKELDRKQYK	2 PDRFSELPMNVLARILFNKA	8 HELRQKICLRIRQIKVLK	15 SLLEINELRPITEIMD	190	24 IKIPPRITTR	27 NMVNLSTEVVLWIL	LLKQGS	83
Xenlae_2769489_Psf3	82 --PKYIREGWRTV	7 VDLHKMGPH-YYPGSQLNLPD	5 -EIAKTIQTQVGRFRIMD	17 DELERSLFRAGQRLNA	179	30 LPQCIESGFP	12 SIPEGRKMLPLWL	KGLVDNK	74
Saccer_37362690_YOL146W	74 LPDMFSTKVMNA	7 LDLHSINSH-FPSLAIKNIMLF	4 --LANVSEILLQRAQLNH	32 EEMKEIYKSHESYKD	186	15 PPCKFQVDIP	11 PITKNKLSLPLWL	RLIAIVG	58
Pyrhor_14591002_PH177	46 FIEBAKLTYLEM	3 EPQELPDNF-YDKVKLIEEVL	14 KRKNHVKLMKTRLEKIFY	15 TREERELVKKSLQQA	149	2 EVEVSVRFRV	3 EWKAGTIVLVKDDFRLI	YAG	37
Pyraby_14521638_PAB0956	51 VIGEDIRILNEER	3 PLSAIPENL-YEKAIFYIYLL	21 KNLKKYVLMKDIRFNKILE	15 APQERRIVLEISIRNE	161	6 KALIPVRVLK	3 NWKEGMILLEWKK	KELWEG	41
Xenlae_14590503_PHO0610	49 VIGEDIRILNEER	3 PLSAIPENL-YEKAIFYIYLL	21 KNLKKYVLMKDIRFNKILE	15 APQERRIVLEISIRNE	161	6 KALIPVRVLK	3 NWKEGMILLEWKK	KELWEG	41
Pyrfur_18976855_PFO483	49 VIGEDIRILNEER	3 PLSAIPENL-YEKAIFYIYLL	21 KNLKKYVLMKDIRFNKILE	15 APQERRIVLEISIRNE	161	6 KALIPVRVLK	3 NWKEGMILLEWKK	KELWEG	41
Thekod_57641554_TK1619	49 VIGEDIRILNEER	3 PLSAIPENL-YEKAIFYIYLL	21 KNLKKYVLMKDIRFNKILE	15 APQERRIVLEISIRNE	161	6 KALIPVRVLK	3 NWKEGMILLEWKK	KELWEG	41
Pyrar_18312313_PAB0965	58 SPKDVATLKYAQ	3 YPVKLPEGF-YPRVRIITILLG	8 RAVICEVRELLIERVKIAM	18 TPBEKALLTSMNATTS	158	13 VFRDVSPLS	3 SVQTNVAEVPFLKL	KLADMD	48
Aerper_14600516_APE0187	70 DIDYVNLVHYEV	6 SLVGVPSPD-YLRVKELVEDLD	15 ENVERAIMQISERRIVMMR	14 TVENLVYSMAEEVVK	176	26 KDHPRLPPTG	3 NARRGEVELPRWQ	RMLREKG	61
Sulsol_15897675_SSO0772*	83 SIEDLGRILFOER	7 SLVPLGKDF-TSRVOLYLETLR	11 RKSIGILNEITIKIRKLQ	15 TEEELLYKTKIKQLKE	187	40 EDWNGSSITT	3 ELVKGENCPLQWL	IILEGK	75
Sultoc_15920685_ST0469	68 SVBELGRILFOER	7 SLVPLGKDF-TSRVOLYLETLR	11 RKSIGILNEITIKIRKLQ	15 TEEELLYKTKIKQLKE	187	25 DMGPFDGDF	3 SLTKGEDETKLWL	DELCKKS	60
Sulaci_70606687_Saci_0901	67 TIDELGRILFOER	7 SLVPLGKDF-TSRVOLYLETLR	11 RKSIGILNEITIKIRKLQ	15 TEEELLYKTKIKQLKE	187	24 EDWNGSSITT	3 ELVKGENCPLQWL	IILEGK	75
Xenlae_29365477_Sld5	18 LPAELINKLEEA	19 CVMEQLNEM-BONLIRAKPG	11 ERIRVLSYIRSRMLKLEK	20 SPPEFAFAKEYMTMTET	28	DSVPLRVKE	19 DNEVGRKHLIRVTRIAP	VASG	216
Saccer_45270874_YDR489W	53 VSPQDFSDLMKS	19 RLINRISMQ-SQLIENISMGL	27 ERIRVLSYIRSRMLKLEK	22 SKDEIKYHDTSLINLK	32	NKVPFTHANG	33 ELTIGIHYMYREVIRDL	RLDD	289
Xenlae_29365479_Psf1	0 MPCEKATLIREL	17 QLEEMKAL-YEQNADNNEAK	18 LRNEPCTVAVLYDLIRLIR	17 STEMDQMNQYKSLAT	21	SLAIEVRCLR	13 LLKKNCHPLPRWK	BOLIRQC	189
Saccer_6320216_YDR013W	0 MYGDLGNKLVLEA	25 NILKEVSNL-RKNTYLKEQQ	21 -NRNCKLAVQRLRTGILDS	24 SHQQRVLYKXCIDLITD	19	DVIRIDVRVLK	12 NILIKGPFVFRQSD	ERLIQGG	203
Arcful_11498930_AFI332	1 NLDELILLTEKNR	1 QLSKIDDF-YEKLEERIAELE	18 RTLKLQKIFELRTGILDS	18 CSDBRVFKKLIDILKE	13	SDRVLVRKIK	12 KLREPVVITLPQLN	DALIKGG	166
Metthe_15679311_MTH1311	5 PFKGKRIQKKEK	3 GLARVGDF-YERVSYIEDLL	15 RDTQKLAFCERREHKITD	29 TPEEELVYSLVETIRE	53	EDIEVTLIFE	15 PFRPQIVITLPSLN	NVFINAK	223
Metjan_15668423_MJ0248	24 MYSELKNYFFBIE	3 KLLKLDDF-YDRIEYIKNIK	5 ERVKYVYELKRLIRYKALY	8 LPEELNIHALENIVVE	24	NDIDVVKVKE	12 DLNKNVLSLDRKI	HILEKHR	179
Metkan_20093854_MK0416	4 SABR-LLEDLKEF	9 RINPQIMER-RTPRGWLREW	36 ERSTDREDDVLYRRALLRY	38 FAEAQLLYPTPDVED	32	LEVYVVRV	-G	15 PSKEGLLVGVETA	RILSERG
Metmaz_21227912_MM1810	1 DIBGTSQTYLKEK	2 TLKTIPADF-YLEAEKYIRELE	21 ERALSDELTIFMKRIGKAIT	19 LPAEKLVLVDVLQGLEA	62	EYVVRVRLK	13 TVNAE VVVLPQLN	TGLIKRN	221
Metace_20089534_MA0647	5 DIBETSYLYLKEK	2 TLKTIPADF-YLEAEKYIRELE	21 ERALSDELTIFMKRIGKAIT	19 LPAEKLVLVDVLQGLEA	62	EYVVRVRLK	13 TVNAE VVVLPQLN	TGLIKRN	221
Metmar_45359273_MMP1710	3 MYKRLDAPFNEF	3 GLSDLPENP-YKDAENYLETIE	5 KRKYVLSNLSFRIYVNC	8 TPDEKRVISFTEETCYG	43	SDIDIVRVIT	12 VLKKNVLSLDRKI	KILEKHN	177
Metbar_73669082_Mbar_A1570	1 DIBETSYLYLKEK	2 TLKTIPADF-YLEAEKYIRELE	21 ERALSDELTIFMKRIGKAIT	19 LPAEKLVLVDVLQGLEA	62	DEYVVRVRLK	13 TVNAE VVVLPQLN	TGLIKRN	220
Pyrfur_18893033_PFO982	1 DIBVLRRLLEKEL	3 ELTKIEEFP-YDDLESPKALE	22 SLVQSTAREILKIRLHKIVD	10 VGDGKKIPAVLTAFING	24	GILEAYLLRV	14 PFKAGLVVLPRST	IRVLQRD	177
Pyrhor_3257074_PH0666	5 DIBVLRRLLEKEL	3 ELTKIEEFP-YDDLESPKALE	22 SLVQSTAREILKIRLHKIVD	10 VGDGKKIPAVLTAFING	24	GILEAYLLRV	14 PFKAGLVVLPRST	IRVLQRD	177
Pyraby_5458796_PAB1466	1 DIBVLRRLLEKEL	3 ELTKIEEFP-YDDLESPKALE	22 SLVQSTAREILKIRLHKIVD	10 VGDGKKIPAVLTAFING	24	GILEAYLLRV	14 PFKAGLVVLPRST	IRVLQRD	177
Thekod_57158795_TK0536	1 DIBVLRRLLEKEL	3 ELTKIEEFP-YDDLESPKALE	22 SLVQSTAREILKIRLHKIVD	10 VGDGKKIPAVLTAFING	24	GILEAYLLRV	14 PFKAGLVVLPRST	IRVLQRD	177
Theaci_16082074_Tai1042	8 LIESEPTAIYABQ	3 KIEIDPDI-YRUVANSDJO	22 OTLKQPSPPQRFKIVA	14 TDEKEPEKIHNTTE	64	GRVLVMTIE	12 SHENLVLVPLQDL	SILAKK	226
Thevol_13541384_TVN0553	11 LIESEPTAIYABQ	3 KIEIDPDI-YRUVANSDJO	22 OTLKQPSPPQRFKIVA	14 TDEKEPEKIHNTTE	64	GRVLVMTIE	12 SHENLVLVPLQDL	SILAKK	226
Pictor_48477338_PFO0266	8 IIRNKLAQILSER	3 RINKDSNF-YSEIRISIMQLO	20 DELKLFHSEYELPFSKIAK	13 TPEKTFYENLKREPMNI	41	NEVILTRITE	12 YLHKGVVYLRKGL	DILIKRN	200
Haloba_15790471_Vng1471C	1 NLIDDLAAQTQER	3 GLQDLQESF-YADVAFIESLE	6 ETABOTVAYIERRVGVKIV	16 TAEERDLVGLDVARIED	106	TARVVRVITD	13 DLACGVVITLPEAN	DPLIEKD	268
Halmar_55229930_rnAC0293	3 QLQQLRETF-YEDAGQFQCLR	18 GEIRSLACKLVFKRIKITE	16 LPEKLLVNLVYELVSK	15 IMGGRVRLK	13	LVYAGVLSLV	ESNKTIRLKP		159
Natpha_76802630_NP3992A	1 NLDELQSVQARER	3 DLQQLRPSF-YREVGAFVOELT	6 ETABOTVAYIERRVGVKIV	16 TAEERDLVGLDVARIED	106	TARVVRVITD	13 DLACGVVITLPEAN	DPLIEKD	268
Nanequ_41615206_NEQ0419	2 SIREKRYLEQES	3 KLIELDNF-LYDLVSTIKTIK	18 GEIRSLACKLVFKRIKITE	16 LPEKLLVNLVYELVSK	15	IMGGRVRLK	13 LVYAGVLSLV	ESNKTIRLKP	159
Pyrar_18313778_PAB3035	0 MIKELRWLLABE	3 VLSDPPFNS-YAETABQVKEL	14 EHISSTLTIMELRKKIIN	11 TAEERLIRPTVKIRHA	7	QHVYVVFILT	14 PFSRGLAKLPLRD	KDLREEG	152
Aerper_14600709_APE0437	0 ----MRLLSER	2 AGGPYKAVS-GSHEVLGSERL	13 RDVC-SAYLERLLRSGIA	14 GVDRDLVLEALGSVFEY	22	IMGGRVRLK	13 LVYAGVLSLV	ESNKTIRLKP	159
Sulsol_15897916_SSO1049*	0 ----MDLKVKEK	3 EITEIKLEB-IKIVITLKKSK	12 KPLSELASSLPFLRLSKVLE	5 GDFDFIDFIFKILKQFY	2	LLGGRYIYN	16 RVNAGVLSLV	PMRELP	138
Sultoc_15921178_ST0942	0 ----MDLKVKEK	3 EITEIKLEB-IKIVITLKKSK	12 KPLSELASSLPFLRLSKVLE	5 GDFDFIDFIFKILKQFY	2	LLGGRYIYN	16 RVNAGVLSLV	PMRELP	138
Sulaci_70607043_Saci_1278	0 ----MDLKVKEK	3 EITEIKLEB-IKIVITLKKSK	12 KPLSELASSLPFLRLSKVLE	5 GDFDFIDFIFKILKQFY	2	LLGGRYIYN	16 RVNAGVLSLV	PMRELP	138
consensus/90% ALL	0 ----MDLKVKEK	3 EITEIKLEB-IKIVITLKKSK	12 KPLSELASSLPFLRLSKVLE	5 GDFDFIDFIFKILKQFY	2	LLGGRYIYN	16 RVNAGVLSLV	PMRELP	138
jpred_cons	0 ----MDLKVKEK	3 EITEIKLEB-IKIVITLKKSK	12 KPLSELASSLPFLRLSKVLE	5 GDFDFIDFIFKILKQFY	2	LLGGRYIYN	16 RVNAGVLSLV	PMRELP	138

D

Soluble extract

↓

Precipitate with PEI to 0.1%

↓

Resuspend in 10 mM Tris 1 M NH₄SO₄

↓

Butyl Sepharose

↓

Mono Q HR10/10 (pH 8.0)

↓

Superdex 200 (HiLoad 26/60)

↓

Hydroxyapatite CHT-5-1

↓

HiTrap heparin

↓

Mono Q (pH 6.0)

◀ **Fig 3** | Isolation of a Gins23-containing complex from *Sulfolobus solfataricus* extracts. (A) Immunoaffinity purification of GINS. The sizes of markers (M) are shown on the left of the panel. Input material is shown (In). The eluted material from columns coupled to pre-immune (PI) or anti-Gins23 (I) sera is indicated. The three specific proteins that elute from the anti-Gins23 column are indicated on the right of the panel with genome annotation number and proposed name. (B) Diagram of the operon-like structure that contains the SSO1049/gins15 open reading frame. The arrow indicates the direction of transcription. (C) Alignment of conserved blocks of SSO1049/gins15 with archaeal and eukaryal GINS family members. Sequences are denoted by abbreviated species name, GI number and gene name or identifier. Species are abbreviations as follows: Aerper, *Aeropyrum pernix*; Arcful, *Archaeoglobus fulgidus*; Halmar, *Haloarcula marismortui*; Haloba, *Halobacterium* sp; Metace, *Methanosarcina acetivorans*; Metbar, *Methanosarcina barkeri*; Metjan, *Methanocaldococcus jannaschii*; Metkan, *Methanopyrus kandleri*; Metmar, *Methanococcus maripaludis*; Metmaz, *Methanosarcina mazei*; Metthe, *Methanothermobacter thermautotrophicus*; Nanequ, *Nanoarchaeum equitans*; Natpha, *Natronomonas pharaonis*; Pictor, *Picrophilus torridus*; Pyrab, *Pyrococcus abyssi*; Pyraer, *Pyrobaculum aerophilum*; Pyrfur, *Pyrococcus furiosus*; Pyrhor, *Pyrococcus horikoshii*; Sacer, *Saccharomyces cerevisiae*; Sulaci, *Sulfolobus acidocaldarius*; Sulsol, *Sulfolobus solfataricus*; Sultok, *Sulfolobus tokodaii*; Theaci, *Thermoplasma acidophilum*; Thekod, *Thermococcus kodakarensis*; Thevol, *Thermoplasma volcanium*; Xenlae, *Xenopus laevis*. Eukaryotic species names are colour-coded blue, Euryarchaeal names are coded magenta and Crenarchaeal names are coded orange. The two proteins characterized in this work, Gins23 (SSO0772) and Gins15 (SSO1049), are denoted by asterisks. The names of experimentally characterized proteins are shown in red. The positions of the first and the last residue of the aligned region in the corresponding protein are indicated for each sequence. The numbers within the alignment represent poorly conserved inserts that are not shown. The vertical dashed line separates the permuted region in the Psf2/Psf3 orthologues that contain a circular permutation (Makarova et al, 2005). Positions with identical amino acids are boldfaced. The colouring is based on the consensus (calculated for all sequences in the alignment) shown underneath the alignment; 'h' indicates hydrophobic residues (ACFILMVWYHRK), 't' indicates turn-forming residues (ASTDNVGPERRK), 'p' indicates charged residues (STEDKRNQH), 's' indicates small residues (AGCVDS) and '+' indicates positively charged residues (KR). The secondary structure was predicted using the JPRED program (Cuff et al, 1998). 'H' indicates α -helix and 'E' indicates extended conformation (β -strand). (D) Scheme employed to partially purify the endogenous GINS complex. (E) Proteins eluting from the final Mono Q column; fraction numbers are given below the figure. The contaminating species GDH was identified as NAD-specific glutamate dehydrogenase (SSO1907).

complex. Bacterial helicases have 5' to 3' polarity and so, in the classical model of the replication fork, track along the lagging strand. The physical association of bacterial primase and helicase therefore serves to couple initiation of Okazaki fragment synthesis on the lagging strand with replication fork progression. In contrast, MCM has 3' to 5' polarity and thus would track along the leading strand. By interacting with the trailing N-terminal domains of the MCM helicase and serving as a bridge to the primase, GINS might facilitate coupling of MCM progression on the leading strand with the action of primase on the lagging strand template (Fig 4E). Given the conservation of GINS and MCM and the requirement for GINS in both initiation and elongation phases of replication, we propose that this architecture is conserved in eukarya (Kanemaki et al, 2003; Kubota et al, 2003; Takayama et al, 2003). Indeed, recent work from Karim Labib and colleagues has identified an MCM- and GINS-containing 'replisome progression complex' in yeast (K. Labib, personal communication). Finally, the interaction of a homologue of the DBD of RecJ with the archaeal GINS complex is intriguing. In *E. coli*, RecJ is important in the processing of stalled replication forks (Courcelle et al, 2003). It is tempting to speculate that, in *Sulfolobus*, RecJdbh, by virtue of its interaction with the GINS complex, might have a role in detection and signalling of stalled replication fork structures. Thus, archaeal GINS might have dual roles in facilitating monitoring and maintenance of replication fork progression.

METHODS

Yeast two-hybrid screen. This was performed using a library of *S. solfataricus* genomic DNA as described previously (Dionne & Bell, 2005). The bait construct contained the full-length open reading frame of *S. solfataricus* MCM in the pGBKT7 vector (Clontech, Mountain View, CA, USA).

Immunoprecipitation. Frozen cell pellets of exponentially growing *S. solfataricus* were resuspended in 20 ml of 1 × TBS

(150 mM NaCl, 10 mM Tris-HCl (pH 8.0)) per gram of pellet. Triton X-100 was added to 0.1% to lyse the cells. After a 5 min incubation on ice, lysates were clarified (35,000g, 20 min, 4 °C) twice and passed through a 0.45 μ m syringe filter. Filtered lysates (100 μ l) were incubated with 2 μ l of the specified antisera and 24 μ g/ml ethidium bromide, if applicable, at 4 °C. After 90–120 min, 50 μ l of protein A beads equilibrated in TBST (1 × TBS, 0.1% (v/v) Tween 20) and 20 μ g/ml ethidium bromide, if applicable, were added, and the incubation was repeated. After 90–120 min, the supernatants were removed and saved for analysis. Beads and associated material were washed on ice four times in 1 ml TBST over a period of 30 min, resuspended in 50 μ l 1 × LB (1.05% (w/v) SDS, 62.5 mM Tris-HCl (pH 6.8), 10% (v/v) glycerol, 0.35 M β -mercaptoethanol, 0.0125% (w/v) bromophenol blue), boiled (5 min), resolved by SDS-polyacrylamide gel electrophoresis and immunoblotted.

Immunopurification of native GINS complex. A 5 g portion of *S. solfataricus* P2 biomass was resuspended in 25 ml of 10 mM Tris (pH 8.0), following which Triton X-100 was added to 0.1%, incubated for 5 min at 4 °C and then sonicated. The extract was clarified by centrifugation (15 min, 4 °C, 35,000g). The extract was dialysed overnight against 10 mM Tris (pH 8.0), re-clarified by centrifugation and applied to an 18 ml Q-Sepharose column, then washed and eluted with a linear gradient to 10 mM Tris (pH 8.0) and 1 M NaCl. Then, 10 ml fractions were collected and Gins23-positive fractions identified by western blotting. The three peak fractions were pooled, dialysed against 10 mM Tris (pH 8.0), 150 mM NaCl and 0.1% Tween 20, and 15 ml was applied to either specific or control immunoaffinity columns. Immunoaffinity columns were prepared using 1 ml HiTrap rProtein A columns to which 2 ml of either pre-immune or anti-Gins23 antisera was bound and crosslinked using dimethylpimelimidate. Following application to the columns, the columns were washed with 10 mM Tris (pH 8.0) and 700 mM NaCl and eluted with glycine

(pH 2.5) + 0.5% SDS. Then, 1 ml fractions were collected and concentrated by trichloroacetic acid precipitation before resolution by SDS-polyacrylamide gel electrophoresis. The GINS complex eluted after 4 column volumes of glycine (pH 2.5) + 0.5% SDS had been applied. Protein species were identified by MALDI-TOF mass spectrometry in the MRC Laboratory of Molecular Biology.

Fig 4 | Architecture of GINS and a model for the replication fork. (A) Summary of two-hybrid analyses. ‘++’ indicates growth on media lacking histidine and adenine and ‘+’ indicates growth on media lacking histidine. (B) Gel filtration chromatography of purified recombinant Gins23–Gins15 complex on a Superdex 200 column. The 280 nm absorbance profiles of size standards BSA, ovalbumin (Ov) and chymotrypsinogen (Ch) are shown in the upper part of the figure, and the profile of Gins23–Gins15 is shown below. A Coomassie-stained gel of the peak fractions is shown below. (C) Ni-NTA agarose-mediated pull-down assays using 100 ng of purified Gins23–Gins15 complex (upper panel) or heterodimeric primase (lower panel) with either beads alone or beads + His₆-tagged MCM. A measure of 10 and 1 ng of Gins23–Gins15 is shown in the input lanes. Protein was detected by western blotting with anti-Gins23 antisera. (D) Ni-NTA agarose-mediated pull-down assays using 100 ng of purified Gins23–Gins15 complex (upper panel) or MCM (lower panel) with either beads alone or beads + His₆-tagged primase (tagged on PriL). Input lanes and detection are as above. (E) A model for the replication fork complex based on the interaction data we have obtained. The stoichiometry of RecJdbh in the complex is unknown at present.

Chromatographic purification of native GINS complex. During purification, the Gins23 complex was followed by western blotting of fractions. A measure of 100 ml of extract was prepared from 20 g of biomass as described above. Following clarification, polyethyleneimine was added slowly to a final concentration of 0.1% and the solution was incubated with mixing for 30 min at 4 °C. Precipitated material was collected by centrifugation as before and resuspended in 50 ml of 10 mM Tris (pH 8.0) and 1 M NH₄SO₄ and dialysed overnight against 5 l of the same buffer. The resolubilized material was clarified as above and applied to a 20 ml Butyl Sepharose column at 2 ml/min, washed and eluted with 200 ml gradient to 10 mM Tris (pH 8.0). Gins23-containing fractions were dialysed overnight into 10 mM Tris (pH 8.0) and 50 mM NaCl and applied to a Mono Q HR10/10 column. The column was developed with a gradient to 10 mM Tris (pH 8.0) and 300 mM NaCl, positive fractions pooled and concentrated to 8 ml in Vivaspin 5k concentrator before application to Superdex 200 column in 10 mM Tris (pH 8.0) and 150 mM NaCl at 2 ml/min. Positive fractions were pooled, dialysed against 10 mM sodium phosphate (pH 6.6), applied to hydroxyapatite column and eluted with a gradient to 500 mM sodium phosphate (pH 6.6). Positive fractions were pooled and dialysed against 100 volumes of 10 mM Tris (pH 6.0) and 50 mM NaCl, and passed through a 5 ml HiTrap heparin column (the Gins23 complex eluted in the flowthrough). This material was then applied to a 1 ml Mono Q HR5/5 and eluted with a linear gradient to 10 mM Tris (pH 6.0) and 250 mM NaCl.

Reverse transcription-PCR. An *S. solfataricus* culture was grown to an OD₆₀₀ of 0.27. RNA was extracted using the RNeasy Mini Kit (Qiagen, Crawley, UK) and treated with 7.5 U RNase-free DNaseI (Roche, Basel, Switzerland) for 1 h at 37 °C in a 50 µl reaction containing 0.25 × NEBuffer 3 (New England Biolabs, Ipswich, MA, USA), and cleaned up with the RNeasy Mini Kit (Qiagen). A 10 µg portion of RNA was used with the OneStep RT-PCR Kit (Qiagen) under the following PCR conditions: 45 s denaturation, 45 s annealing (60 °C), 29 cycles. Sequences of the *gins23*- and *mcm*-specific primers are available on request.

Purification of recombinant proteins. The Gins23 and Gins15 open reading frames were cloned into the pET-Duet vector (Novagen, EMD Biosciences, San Diego, CA, USA). This was transformed into BL21 AI (Stratagene, La Jolla, CA, USA) and protein expression induced according to the manufacturer's instructions. Cell lysate was prepared in 10 mM Tris (pH 8.0) and 300 mM NaCl and clarified by centrifugation. The soluble material was heated to 80 °C for 30 min and re-centrifuged. The supernatant was dialysed against 10 mM Tris (pH 8.0) and 1 M NH₄SO₄ and purified over Butyl Sepharose, Mono Q and Superdex 200 columns as described above, before concentration in a Vivaspin concentrator. His-tagged and untagged MCM and primase were purified as described previously (Lao-Sirieix & Bell, 2004; McGeoch *et al*, 2005).

Supplementary information is available at *EMBO reports* online (<http://www.emboreports.org>).

ACKNOWLEDGEMENTS

We thank K. Labib for communicating results before publication and for invaluable discussions. We are indebted to S.Y. Peak-Chew of the MRC laboratory of Molecular Biology for mass spectrometry. Work in S.D.B.'s laboratory is funded by the Medical Research Council and the Nuffield Foundation.

REFERENCES

- Calzada A, Hodgson B, Kanemaki M, Bueno A, Labib K (2005) Molecular anatomy and regulation of a stable replisome eukaryotic DNA at a paused replication fork. *Genes Dev* **19**: 1905–1919
- Courcelle J, Donaldson JR, Chow KH, Courcelle CT (2003) DNA damage-induced replication fork regression and processing in *Escherichia coli*. *Science* **299**: 1064–1067
- Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**: 892–893
- Dionne I, Bell SD (2005) Characterisation of an archaeal family 4 uracil DNA glycosylase and its interaction with PCNA and chromatin proteins. *Biochem J* **387**: 859–863
- Dionne I, Nookala RK, Jackson SP, Doherty AJ, Bell SD (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Cell* **11**: 275–282
- Kanemaki M, Sanchez-Diaz A, Gambus A, Labib K (2003) Functional proteomic identification of DNA replication proteins by induced proteolysis *in vivo*. *Nature* **423**: 720–724
- Kelman LM, Kelman Z (2003) Archaea: an archetype for replication initiation studies? *Mol Microbiol* **48**: 605–616
- Kubota Y, Takase Y, Komori Y, Hashimoto Y, Arata T, Kamimura Y, Araki H, Takisawa H (2003) A novel ring-like complex of *Xenopus* proteins essential for the initiation of DNA replication. *Genes Dev* **17**: 1141–1152
- Lao-Sirieix SH, Bell SD (2004) The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. *J Mol Biol* **344**: 1251–1263
- Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin EV (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acid Res* **33**: 4626–4638
- McGeoch AT, Trakselis MA, Laskey RA, Bell SD (2005) Organization of the archaeal MCM complex on DNA and implications for helicase mechanism. *Nat Struct Mol Biol* **12**: 756–762
- Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116**: 25–38
- She Q *et al* (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA* **98**: 7835–7840
- Soultanas P (2005) The bacterial helicase–primase interaction: a common structural/functional module. *Structure* **13**: 839–844
- Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev* **17**: 1153–1165
- Yamagata A, Kakuta Y, Masui R, Fukuyama K (2002) The crystal structure of exonuclease RecJ bound to Mn²⁺ ion suggests how its characteristic motifs are involved in exonuclease activity. *Proc Natl Acad Sci USA* **99**: 5908–5912