High-fidelity chromosomal DNA replication is fundamental to all forms of cellular life and requires the complex interplay of a wide variety of essential and non-essential protein factors in a spatially and temporally co-ordinated manner. In eukaryotes, the GINS complex (from the Japanese go-ichi-ni-san meaning 5-1-2-3, after the four related subunits of the complex Sld5, Psf1, Psf2 and Psf3) was recently identified as a novel factor essential for both the initiation and elongation stages of the replication process. Biochemical analysis has placed GINS at the heart of the eukaryotic replication apparatus as a component of the CMG [Cdc45–MCM (minichromosome maintenance) helicase–GINS] complex that most likely serves as the replicative helicase, unwinding duplex DNA ahead of the moving replication fork.

GINS homologues are found in the archaea and have been shown to interact directly with the MCM helicase and with primase, suggesting a central role for the complex in archaeal chromosome replication also. The present review summarizes current knowledge of the structure, function and evolution of the GINS complex in eukaryotes and archaea, discusses possible functions of the GINS complex and highlights recent results that point to possible regulation of GINS function in response to DNA damage.

Key words: CMG complex, chromosome replication, DNA helicase, DNA replication, GINS.

INTRODUCTION

Chromosomal DNA replication in all cells requires the complex interplay of a large number of essential and non-essential protein factors in a spatially and temporally co-ordinated manner [1]. In eukaryotic cells, replication is initiated at multiple sites (replication origins) on each chromosome. Assembly of the replication machinery begins during G1, but multiple regulatory mechanisms ensure that origin DNA unwinding and replication initiation does not occur until S-phase and that replication occurs once and only once in every cell cycle. As the origin DNA is unwound, the replication machinery (replosome) is assembled and the replication forks move bidirectionally from the origin into the non-origin DNA [1,2].

Much of what is known of the events of replication initiation has come through work on experimental organisms such as the budding yeast Saccharomyces cerevisiae, the fission yeast Schizosaccharomyces pombe and the African clawed frog Xenopus laevis. Numerous factors essential for replication initiation have been identified and their relationships to one another deciphered in these model systems [1].

In the present review I summarize our current understanding of the function, structure and evolution of the most recently identified essential replication factor in eukaryotic cells, the tetrameric GINS complex [3–5]. GINS (the name is an acronym for go-ichi-ni-san, the Japanese for 5-1-2-3, after the four subunits of the complex Sld5, Psf1, Psf2 and Psf3) is essential for the initiation and elongation stages of chromosome replication [3–5] and serves as a key component of a protein complex that unwinds the duplex DNA ahead of the moving replication fork and around which many additional components of the replisome appear to assemble [6,7]. The function of the GINS complex has been analysed by a variety of methods in a wide range of species and the three-dimensional structure of the human GINS complex has been solved by X-ray crystallography [8–10]. Archaeal GINS homologues have also been identified [11–13], providing insights into GINS complex evolution and core function, whereas evidence from yeast and mammalian cells point to a possible secondary role for GINS in chromosome segregation [14] and highlights the importance of GINS in development and disease avoidance [15–19]. Finally, recent results suggest that the GINS function might be regulated in response to DNA damage [20].

STEPWISE ASSEMBLY OF THE REPLICOSOME

The early events of chromosome replication are best understood in the budding yeast, although it is highly likely that key features of the initiation mechanism will be conserved across evolution [1,2]. In budding yeast, replication origins are bound throughout the cell cycle by a conserved six-subunit protein complex known as ORC (origin recognition complex). During the G1-phase of the cell cycle, ORC is bound by the Cdc6 protein which then recruits two additional factors, Cdt1 and the MCM (minichromosome maintenance) helicase, to form the pre-RC (pre-replicative complex) (shown in Figure 1). Assembly of the pre-RC is tightly regulated, with protein phosphorylation in particular playing an important role. Pre-RC assembly, also known as replication licensing, occurs during G1, when levels of the inhibitory CDK (cyclin-dependent kinase) are low, thereby

---

Abbreviations used: ATM, ataxia-telangiectasia mutated; ATR, ATM and Rad3 related; BiFC, bimolecular fluorescence complementation; CDK, cyclin-dependent kinase; ChIP, chromatin immunoprecipitation; CPC, chromosomal passenger complex; E, embryonic day; EM, electron microscopy; GFP, green fluorescent protein; GINS, go-ichi-ni-san; HBD, hormone-binding domain; HDF, human dermal fibroblast; Hsp90, heat-shock protein of 90 kDa; MCM, minichromosome maintenance; ORC, origin recognition complex; PCNA, proliferating-cell nuclear antigen; pre-LC, pre-loading complex; pre-RC, pre-replicative complex; RecJdhh, RecJ DNA-binding domain homologue; RPC, replisome progression complex; siRNA, small interfering RNA; S-CDK, S-phase CDK; ssDNA, single-stranded DNA; YFP, yellow fluorescent protein.

1 email stuart.macneill@st-andrews.ac.uk
ensuring that replication occurs once, and only once, per cell cycle [1].

Subsequent to pre-RC assembly, replication initiation is promoted by the action of S-CDK (S-phase CDK) and a second protein kinase, DDK (Cdc7-Dbf4) [1,2]. Identification of S-CDK and DDK substrates is key to understanding how replication initiation is regulated. S-CDK substrates are abundant in yeast but two essential proteins, Sld2 and Sld3, have been shown to comprise the minimal set required for replication initiation [21,22]. Phosphorylation of Sld3 by S-CDK on Thr600 and Ser622 is essential for cell viability, substituting these two residues with non-phosphorylatable alanine residues blocks entry into S-phase. The effect of the phosphorylation of Sld3 is to generate a binding site from another essential replication factor, Dpb11 (equivalent to TOPBP1 in human cells). The Dpb11 protein possesses four BRCT domains that act in pairs to bind to phosphopeptidases [23] and it is the N-terminal pair that is responsible for binding to phosphorylated Sld3 (Figure 1) [21,22]. The importance of Thr600 and Ser622 phosphorylation in mediating Dpb11 binding is shown by the fact that fusion of non-phosphorylatable Sld3 to Dpb11 produces a protein that is able to rescue loss-of-function mutations in both genes and partially bypass the requirement for S-CDK activity to initiate S-phase [22]. Note that Sld3 also associates with another essential replication factor, Cdc45, which is likely to be recruited to the pre-RC via its interactions with MCM (Figure 1). Interestingly, a mutant form of Cdc45 (the Cdc45-JET protein) can also partially bypass the requirement for S-CDK, probably by promoting tighter Sld3–Dpb11 interactions [21].

The second essential S-CDK substrate in yeast is Sld2 (equivalent to RecQL4 in human cells). This protein is phosphorylated on multiple sites by S-CDK, but Thr49 phosphorylation is crucial for Sld2 function [24]. The effect of Thr49 phosphorylation is also to create a binding site for Dpb11, this time via the C-terminal Dpb11 BRCT domains (Figure 1). By combining a mutant form of the Sld2 protein in which Thr49 is substituted with an aspartate residue to mimic phosphorylation with the Sld3–Dpb11 fusion protein described above, it is possible to completely bypass the requirement for S-CDK in replication initiation [22]. Similar results were seen when the phosphomimicking Sld2 mutant was combined with cells expressing the Cdc45-JET protein and overproducing Dpb11 [21]. Dpb11, therefore, bridges phosphorylated Sld2 and Sld3, but to what end? Although yet to be described in detail, Sld2 is reportedly [25] part of a complex (termed the pre-loading complex or pre-LC) which includes Dpb11, the leading strand polymerase Polε and the subject of the present review, GINS, suggesting perhaps that the function of Dpb11 in bridging Cdc45–Sld3 with Sld2 is to recruit the pre-LC components to the origin to activate the MCM helicase (Figure 1). The role of GINS in MCM helicase function will be described in detail below.

Once the replisome is activated, the replication forks move off bidirectionally from the origin DNA into the surrounding sequences (Figure 1). ORC remains bound at the origin, apparently intact, at least in yeast cells, whereas MCM, Cdc45, GINS and Polε move with the fork [3,5,26–28]. The remainder of the factors described above dissociate from the replisome and in some cases are degraded or excluded from the nucleus. Meanwhile, numerous additional protein factors – some essential, some not – are recruited around the Cdc45–MCM–GINS core [29] to form a larger macromolecular assembly known as the RPC (replisome progression complex) [6,7,30]. This is discussed further below.

**IDENTIFICATION OF THE GINS COMPLEX**

**Eukaryotic GINS**

Identification of the GINS complex was reported by two groups working with budding yeast, but using very different experimental approaches. As part of their ongoing genetic analysis of the role of
Dpb11 in replication initiation, Araki and co-workers [31] isolated mutations that were synthetically lethal in combination with the temperature-sensitive *dbp11-1* allele. These fell into a number of distinct complementation groups, including *SLD2* and *SLD3* described above, and also *SLD5* (*SLD* is an acronym for synthetic lethal with *dbp11-1*). The *SLD5* gene was cloned and shown to encode a protein of 294 amino acids of unknown function. To identify factors that might interact with *SLD5*, Takayama et al. [5] first isolated temperature-sensitive alleles of *SLD5*, then screened for multicopy suppressors of one of these, *sld5-12*, and in this way isolated the *PSF1* gene encoding a 208-amino-acid protein. Next, a temperature-sensitive *psf1* allele, *psf1-1*, was generated and another new gene, *PSF3*, encoding a 194-amino-acid protein, isolated as a multicopy suppressor. Although not isolated in their screens, *SLD5* was also shown to rescue *psf1-1*, albeit weakly.

These genetic interactions suggested strongly that the products of the budding yeast *SLD5*, *PSF1* and *PSF3* genes might interact with each other. To test this, an epitope-tagged form of the Psf1 protein was affinity-purified under native conditions from exponentially growing yeast cells and was shown to co-purify with Sld5 and Psf3, as well as with a novel 213-amino-acid protein later designated Psf2 [5]. Overproduction of *PSF2*, like *PSF1*, was also shown to be able to suppress to *sld5-12*. Further biochemical analysis showed that all four proteins were stoichiometric components of a 100 kDa complex present at constant levels through the cell cycle; the complex was named GINS.

While Takayama et al. [5] were treading a traditional yeast molecular genetic path, Labib and co-workers [3] took an altogether different approach. Previous work from a consortium of laboratories [32] had identified a significant number of budding yeast genes with essential functions but where comparative protein sequence analysis gave no clues to their cellular roles. In order to identify members of this group that were specifically required for cell-cycle progression, temperature-sensitive mutant forms of many of these essential proteins were constructed by fusing a heat-inducible degron cassette [33] to their N-termini [3]. At higher temperatures, the unfolded degron is recognized by the Ubr1 protein, targeted for ubiquitylation and subsequently degraded by the proteasome. Three of the four subunits of the GINS complex were identified by this approach: Sld5, Psf1 and Psf2 (note that these were designated Cdc105, Cdc101 and Cdc102 in the original paper describing this work) [3]. Purification of Psf2 from yeast cell extracts demonstrated the physical association of this protein with Sld5 and Psf1 and allowed the subsequent identification of Psf3 [3].

In parallel with identification of budding yeast GINS, Takisawa and colleagues [4] reported the isolation of the *Xenopus* GINS proteins and showed that these too formed a heterotetrameric GINS complex with a molecular mass of ~100 kDa.

**FUNCTION OF GINS**

**Budding yeast**

Each of the four GINS proteins is essential for cell viability in budding yeast. Inactivation of individual proteins by temperature-shifting cells carrying temperature-sensitive mutant alleles of *sld5* or *psf1* [5] or heat-inducible degron alleles of *sld5* or *psf2* [3] results in cell-cycle arrest, with the arrested cells having a single nucleus and displaying the large-budded (dumbbell) shape that is typical of yeast replication mutants. Chromosomal DNA replication is blocked under these conditions, particularly in cells expressing the Sld5 and Psf2 degron fusion proteins [3].

To investigate GINS function in greater detail, ChIP (chromatin immunoprecipitation) [34] was used to ask whether the GINS proteins associated with replicating DNA in cells proceeding slowly through S-phase at low temperature or at low concentrations of the ribonucleotide reductase inhibitor HU (hydroxyurea). In both cases, GINS proteins were found to associate with origin sequences in early S-phase, before moving with the replication fork from origin to non-origin DNA as replication continued [5,28]. In addition to this, GINS was shown by density substation experiments to be required both for replication fork establishment at origins and, importantly, also for replication fork progression, demonstrating that the GINS complex is not merely a passenger at the fork but an active participant [3]. Consistent with GINS being required for replication fork establishment at origins, neutral/neutral two-dimensional gel analysis showed firing of the budding yeast replication origin *ARS1* to be reduced in *sld5* and *psf1* mutants [5].

The origin association of GINS is dependent on both Sld3 and Dpb11 function: when *sld5-3* or *dbp11-26* cells are released into S-phase from G1 at their restrictive temperatures, no binding of GINS to origins can be detected by ChIP [5], suggesting that this event is downstream of Sld3 and Dpb11 function (see Figure 1). However, GINS function is required for Dpb11 to associate with origins, indicating that GINS and Dpb11 origin binding are mutually dependent events. As noted above, there are indications that GINS and Dpb11 are present in a distinct complex, termed the pre-LC; along with Sld2 and Polc [25]; it is tempting to speculate that the observed interdependency of GINS and Dpb11 reflects the consequences of disruption of this complex.

GINS is also required for Cdc45 chromatin binding [5], but not for initial recruitment of Cdc45 to origins during late G1. However, GINS is required to allow Cdc45 to move away from origins as replication progresses, as well as for the stable association of Cdc45 with origin sequences in cells in which replication is stalled by inactivation of GINS using a heat-inducible degron-tagged Psf2 protein [28].

**Fission yeast**

The four subunits of GINS in the fission yeast *S. pombe* are similar in sequence to their budding yeast counterparts and are presumed to form a similar heterotetrameric complex. As in budding yeast, ChIP has been used to show that GINS associates with origin DNA in early S-phase [35].

Several conditional lethal mutant forms of the fission yeast GINS proteins have been used for analysis of *S. pombe* GINS function, including traditional temperature-sensitive alleles of *psf2* [14,36] and *psf3* [35], and hormone-dependent alleles of *psf1* and *psf2* [37,38]. The first of these to be isolated was the temperature-sensitive *psf2-209* allele, from a screen for mutations that prevented re-replication of chromosomal DNA caused by an inappropriate reduction of CDK activity [36]. When shifted to the restrictive temperature, *psf2-209* cells display a somewhat heterogeneous phenotype: the cells are elongated, show some nuclear abnormalities (discussed further below) and have DNA contents in the 1C–2C range, suggesting that either entry into S-phase or S-phase progression is delayed. Consistent with the latter possibility, *psf2-209* cells require the presence of a functional DNA-damage checkpoint for viability. The cells probably also display defects in pre-meiotic S-phase [36].

In contrast with *psf2-209*, *psf3-1* was isolated in a targeted screen for *psf3* mutations. Cells carrying this temperature-sensitive allele undergo cell-cycle arrest when shifted to their restrictive temperature, accumulating at early time points with
1C DNA content that is consistent with an early S-phase block [35]. At later time points, cells with 2C DNA are seen, indicating that some replication is possible in this mutant, as well as cells with <1C DNA that are thought to be the result of cell division occurring in the absence of complete DNA replication. The sld3-1 mutation can be suppressed by overproduction of Sld5, Psf1 and Psf2, and also by Sld2 [35], and is synthetically lethal at the normal permissive temperature with mutant alleles of cut5 and cdc20, encoding the fission yeast Dpb11 homologue and the catalytic subunit of Polɛ respectively, as well as with sld3 and cdc45. The GINS complex appears also to be partially destabilized in psf3-1 cells; the mutant Psf3-1 protein fails to co-purify with FLAG-tagged Psf2, although pulldown of Sld5 and Psf1 is unaffected, indicating that Psf3 is not required for formation of an Sld5–Psf1–Psf2 sub-complex.

As in budding yeast, binding of Cdc45 to origins (assayed by ChIP) is abolished by GINS inactivation in psf3-1 cells. The same is true of Cut5 and Drc1 (the fission yeast orthologues of Dpb11 and Sld2 respectively) and Polɛ, but not MCM and Sld3, both of which associate normally with origin DNA. Conversely, both Cut5 and Sld3 are required for GINS binding to origins, but Cdc45 is not [35].

Studies in several organisms have shown that fusing a β-oestradiol HBD (hormone-binding domain) to a target protein can result in the activity of the protein becoming regulated by the hormone [39]. In certain cases, this appears to be caused by Hsp90 (heat-shock protein of 90 kDa) binding to the HBD in the absence of hormone and blocking protein–protein interactions by steric hindrance. The inactivation seen is reversible, as subsequent addition of β-oestradiol causes Hsp90 to dissociate from the HBD. In fission yeast, fusing the β-oestradiol HBD to Psf1 and Psf2 produces cells that require the constant presence of β-oestradiol in the growth medium for cell viability [37,38]. Removal of β-oestradiol causes rapid cell-cycle arrest in early S-phase and in the case of Psf2, nuclear delocalization of the protein (Psf1 was not analysed in this way). The Psf1–HBD mutant has been used to show that GINS function is required for Cdc45 chromatin loading at S-phase initiation (as above), as well as for maintenance of Cdc45 binding during the elongation stages of replication [38]. GINS inactivation by β-oestradiol removal also results in a failure to load Polɛ on to chromatin, as seen previously [35]. Polɛ is loaded normally in these conditions, but is not released from chromatin as S-phase cannot be completed in the absence of GINS function.

Finally, interactions between GINS and MCM have been investigated using BiFC (bimolecular fluorescence complementation) in fission yeast [40]. This technique relies on the ability of the N- and C-terminal fragments of the YFP (yellow fluorescent protein) to form a fluorescent complex when brought together by the association of two interacting partners [41]. Using BiFC, specific interactions between GINS and MCM mediated by Psf1 and Mcm4 have been observed on chromatin during S-phase, consistent with the role of the proteins at the replication fork [40]. However, GINS–MCM interactions were also seen in other phases of the cell cycle, perhaps due to the fact that once they form a complex, the two fragments of YFP are extremely reluctant to dissociate [41].

**Xenopus**

The frog *Xenopus laevis* has proved a very powerful model for biochemical studies of eukaryotic chromosome replication. Central to this has been the ability of extracts prepared from *Xenopus* eggs to efficiently replicate exogenously added DNA templates such as sperm chromatin or purified plasmid DNA and to do so in a manner that preserves cell-cycle control of replication licensing [42].

*Xenopus* GINS was the first higher eukaryotic GINS complex to be functionally characterized [4]. The four subunits of frog GINS are approx. 20–25% identical (40–50% similar) at the amino acid level to the budding yeast proteins. Antibodies raised against Sld5 co-immunoprecipitate Psf1, Psf2 and Psf3, indicating that the four proteins co-exist in a complex. Further analysis showed the complex to have a molecular mass of ~100 kDa, consistent with the expected heterotetrameric structure of GINS, and to be globular with a ring- or C-shape (discussed below) [4].

To test whether GINS was required for chromosome replication, the ability of Sld5-depleted *Xenopus* egg extracts to replicate exogenously added sperm chromatin was assayed by monitoring the incorporation of labelled nucleotide analogues into chromosomal DNA in Sld5-depleted and mock-depleted S-phase egg extracts [4]. Depending on the exact method used, either little or no incorporation of the labelled nucleotides was seen in extracts depleted for Sld5, showing that Sld5, and most probably the other subunits of GINS, are required for efficient replication initiation in *Xenopus* egg extracts. Replication activity could be restored by adding-back recombinant GINS. Further analysis using the egg extract system showed that GINS became associated with sperm chromatin sometime after pre-RC formation in an S-CDK-dependent manner and dissociated from chromatin as replication progressed [4]. The timing of GINS association with chromatin was indistinguishable from that of Cdc45 and, indeed, the two binding events are interdependent: GINS does not become chromatin-associated in Cdc45-depleted extracts and, as in budding yeast, Cdc45 does not become chromatin-associated in Sld5-depleted extracts. In both cases, lack of chromatin binding leads to a failure to load the replicative polymerases Polɛ and Polɛ [4]. Cut5 was also required for GINS and Cdc45 chromatin loading (as in fission yeast, Cut5 is the orthologue of Dpb11; see Figure 1).

**Human**

Genes encoding four subunits of human GINS were identified by comparative sequence analysis and their products shown to form a ~100 kDa heterotetrameric complex with 1:1:1:1 stoichiometry when expressed in recombinant form [43,44]. In addition, monoclonal antibodies raised against Psf2 have been used to pulldown the GINS complex from human cell extracts [45]. As in yeast, the complex is present throughout the cell cycle, but varies in level in different cell types with significantly higher concentrations seen in immortalized and transformed cells, probably due, at least in part, to simultaneous transcriptional up-regulation of the four genes [45]. Even in untransformed cell lines, such as Wi38 human fibroblasts, there are ~100 000 GINS complexes per cell. As cells leave the cell cycle due to contact inhibition of growth, these levels drop markedly, so that the four GINS subunits are undetectable in cell extracts prepared from G0 non-cycling cells. Protein turnover probably has an important influence here, as the half-lives of the GINS protein are in the 6–8 h range [45]. While the level of GINS is constant throughout the cell cycle, a significant fraction of the complex becomes chromatin-associated during S-phase, as is the case in yeast and *Xenopus*. This chromatin-associated material co-purifies with MCM and Cdc45 from extracts prepared from nuclease-treated S-phase chromatin [45], consistent with the existence of CMG complexes in human cells. siRNA (small
interfering RNA) is a powerful tool for dissecting gene function in mammalian cells [46,47]. Transfection of siRNAs directed against the GINS subunits individually into HeLa cells reduces GINS levels to <10% of that seen in control transfected cells and inhibiting the expression of any one subunit of the complex results in degradation of the others [45]. Cell proliferation and DNA replication (measured by uptake of labelled precursors and flow cytometry) is significantly impaired in HeLa and HDF (human dermal fibroblast) cells transfected with GINS-targeted siRNAs, with defects seen at both the initiation and elongation stages of replication and the apparent accumulation of DNA damage [45,48]. Pre-RC assembly is unaffected in HDF cells, as judged by the loading of MCM on to chromatin, whereas Cdc45 recruitment to chromatin is delayed and the maximal level of chromatin-bound Cdc45 reduced [48]. Further studies suggest that GINS depletion in HeLa human cells leads to mitotic defects also [14], although this is not seen in untransformed HDF cells [48]; this will be discussed further shortly.

Recombinant human GINS has been used in both biochemical and structural studies [8–10,43,44]. In EMSAs (electrophoretic mobility-shift assays), GINS has been shown to be able to bind to DNA with a preference for single-stranded molecules or molecules containing stretches of ssDNA (single-stranded DNA), including bubble structures [43]. To date, this remains the only evidence for direct GINS–DNA interactions. Using surface plasmon resonance and co-immunoprecipitation, human GINS has been shown to interact directly with the dimeric primase component of Polα-primase, the enzyme responsible for synthesis of the RNA–DNA primers that initiates synthesis of the leading strand and each Okazaki fragment on the lagging strand [44]. A similar interaction has been observed with the archaeal GINS and primase proteins (see below) [12]. In addition to binding to primase, human GINS also appears capable of stimulating its activity: a 10-fold stimulation is seen when GINS is present in a 200-fold molar excess over primase in assays using singly primed single-stranded M13 DNA [44].

GINS is a component of the CMG complex

The experiments described above clearly place GINS at the replication fork, performing an essential role for the initiation and elongation stages of replication, but what does the GINS complex actually do? To date, no stand-alone enzymatic activity has been ascribed to GINS. Instead, GINS appears to perform its function as part of the three-component CMG complex. CMG, an acronym for Cdc45–MCM–GINS, was first identified by Moyer et al. [29] in the course of purifying high molecular mass complexes from Drosophila embryo extracts that contained the essential Cdc45 protein [29]. In these experiments, Cdc45 co-purified with ten other proteins in a complex with an apparent molecular mass of 700 kDa – the six subunits of the MCM helicase (Mcm2–Mcm7) and the four subunits of Drosophila GINS (Sld5, Psf1, Psf2 and Psf3). Only ~5% of the total amount of Cdc45 and GINS in the extract was in the CMG complex and even less of the MCM helicase (~1%), suggesting that assembly of the CMG might be regulated by, for example, post-translational regulation of one or other of the components. Nevertheless, as isolated from Drosophila embryo extracts, the CMG complex is stable and, more importantly, biochemically active as a DNA helicase, which raised the possibility that the CMG complex was the elusive replicative helicase in eukaryotic cells [29].

Further evidence for CMG being the replicative helicase comes from studies of plasmid replication in Xenopus egg extracts. In these assays, plasmid DNA is incubated in an extract (termed HSS for high-speed supernatant) derived from egg cytoplasm to promote pre-RC formation, after which a concentrated NPE (nucleoplasmic extract) is added to bring about replication initiation [49]. Initiation occurs independently of DNA sequence in this system, restricting detailed analysis of protein–DNA interactions at replication origins. To circumvent this restriction, Pacek et al. [50] synthesized plasmid DNA that contained a biotin group at a single site, to which a streptavidin moiety was bound. The presence of the biotin–streptavidin pairing is sufficient to bring about replication fork pausing at the labelled site, allowing ChIP to be used to monitor the presence or absence of individual target proteins in the paused replisome. In addition, it had previously been shown that adding the DNA polymerase inhibitor aphidicolin results in the uncoupling of DNA synthesis from DNA unwinding in these reactions [49]; the uncoupled DNA helicase continues to unwind the duplex while nascent DNA synthesis ceases. Under these conditions, the helicase alone would be expected to be present at the pause site on the plasmid. Using ChIP, it was shown that the MCM helicase was indeed present at the pause site, as were Cdc45 and the GINS complex. All three replicative DNA polymerases (α, δ and ε) were also present but not in the presence of aphidicolin, when synthesis and unwinding are uncoupled. It is highly likely that the MCM, Cdc45 and GINS proteins seen at the pause site are in the form of the CMG.

GINS is a component of the replisome progression complex

In addition to making up the CMG complex, it is now clear that Cdc45, MCM and GINS are part of a larger complex present in S-phase at the replication fork. Tandem affinity purification of GINS from budding yeast cell extracts brings down, in addition to Cdc45 and MCM, a specific set of proteins that mediate or regulate replication fork progression [6,30]. This large set of proteins has been designated the RPC. The list of co-purifying proteins identified in these experiments is long, but all six subunits of MCM are present, as is Cdc45. Also present are Top1 and Csm3 (which allow forks to pause when encountering protein barriers to replication), Mcq1 (which mediates checkpoint signalling in response to fork stalling), Ctf4 (which interacts with Polα-primase and is required for the establishment of sister chromatid cohesion), Top1 (topoisomerase I, which is thought to remove positive supercoils generated by DNA unwinding ahead of the fork), Mcm10 (which also interacts with Polα-primase), FACT (the histone chaperones Spt16 and Pob3) and, under certain purification conditions, Polα-primase itself [7]. Interestingly, evidence from budding yeast suggests that each RPC contains only a single MCM helicase at its core and therefore only a single Cdc45 and GINS in the extract was in the CMG complex and even less of the MCM helicase (~1%), suggesting that assembly of the CMG might be regulated by, for example, post-translational regulation of one or other of the components. Nevertheless, as isolated from Drosophila embryo extracts, the CMG complex is stable and, more importantly, biochemically active as a DNA helicase, which raised the possibility that the CMG complex was the elusive replicative helicase in eukaryotic cells [29].

Further evidence for CMG being the replicative helicase comes from studies of plasmid replication in Xenopus egg extracts. In these assays, plasmid DNA is incubated in an extract (termed HSS for high-speed supernatant) derived from egg cytoplasm to promote pre-RC formation, after which a concentrated NPE (nucleoplasmic extract) is added to bring about replication initiation [49]. Initiation occurs independently of DNA sequence in this system, restricting detailed analysis of protein–DNA interactions at replication origins. To circumvent this restriction, Pacek et al. [50] synthesized plasmid DNA that contained a biotin group at a single site, to which a streptavidin moiety was bound. The presence of the biotin–streptavidin pairing is sufficient to bring about replication fork pausing at the labelled site, allowing ChIP to be used to monitor the presence or absence of individual target proteins in the paused replisome. In addition, it had previously been shown that adding the DNA polymerase inhibitor aphidicolin results in the uncoupling of DNA synthesis from DNA unwinding in these reactions [49]; the uncoupled DNA helicase continues to unwind the duplex while nascent DNA synthesis ceases. Under these conditions, the helicase alone would be expected to be present at the pause site on the plasmid. Using ChIP, it was shown that the MCM helicase was indeed present at the pause site, as were Cdc45 and the GINS complex. All three replicative DNA polymerases (α, δ and ε) were also present but not in the presence of aphidicolin, when synthesis and unwinding are uncoupled. It is highly likely that the MCM, Cdc45 and GINS proteins seen at the pause site are in the form of the CMG.

GINS STRUCTURES

Primary structures

As noted above, the four subunits of eukaryotic GINS are distantly related to one another at the protein sequence level and most probably descended from a common ancestor, that is, they are paralogous. This was first shown by Koonin and colleagues [11] who identified GINS homologues in archaea also (discussed in detail below). Interestingly, the four eukaryotic GINS proteins can
be recognized as being of two types that differ in the arrangement of conserved protein sequences (Figure 2A). Two regions of protein sequence conservation can be identified in Sld5 and Psf1 that correspond, as we shall see shortly, to distinct domains in the three-dimensional structures of the proteins. In Psf2 and Psf3, both of these domains (termed the A- and B-domains) are present, but their order in the primary sequence of the proteins is reversed relative to that seen in Sld5 and Psf1: the larger N-terminal A-domain in Sld5 and Psf1 is C-terminally located in Psf2 and Psf3, whereas the smaller C-terminal B-domain of Sld5 and Psf1 is N-terminally located in Psf2 and Psf3 (Figure 2A). Presumably the four genes encoding the eukaryotic GINS subunits evolved from an ancestral gene by two rounds of gene duplication with the first round being followed by a rearrangement (circular permutation) of one of the two duplicated genes that reversed the domain order [11]. In archaea, proteins with A-B domain order predominate (see Figure 2B), perhaps suggesting that the ancestral GINS protein was of this type, although other interpretations are possible, particularly if there have been archaeological lineage-specific losses of genes encoding B-A-type proteins.

Crystal structures

In independent studies, three groups have solved crystal structures of human GINS with resolutions varying from 2.3 Å to 3.2 Å (1 Å = 0.1 nm) [8–10]. The structures (PDB codes 2E9X, 2EHO and 2Q9Q) are highly similar to one another, but none shows the entire structure of the complex. The overall shape of the heterotetrameric complex has been variously described as resembling a trapezium [9], an elongated spindle with a visible central hole [8] or of being of elliptical shape [10]. Figure 3 shows the overall structure of the complex (an interactive version of this can be seen at http://www.BiochemJ.org/bj/425/0489/bj4250489add.htm), while for clarity, Figure 4 shows views of individual subunits and domains.

The overall dimensions of the complex are ~110 Å × 60 Å × 60 Å, the central hole is perhaps 5–10 Å in diameter [8–10]. The four subunits are arranged in two layers: Sld5 and Psf1 form the top layer and Psf2 and Psf3 the bottom layer. This arrangement of subunits is consistent with observations made with recombinant Xenopus and human GINS proteins [4,43], where Sld5, Psf1 and Psf2 have been shown to be able to form a stable trimer in the absence of Sld3, and Sld5 and Psf2 a stable dimer in the absence of Psf1 and Psf3, although it should be noted that there is no evidence that such sub-complexes exist in vivo.

All four subunits are related to one another in structure, as they are in sequence, and the arrangement of subunits creates a vertical pseudo 2-fold axis in the centre of the complex (indicated by the broken line in Figure 3). Consistent with the bioinformatic analysis discussed above, each subunit is composed of two distinct domains joined by a linker region of variable length: the predominantly α-helical A-domain and the largely β-stranded B-domain. As indicated above, the A-domain is N-terminal in Sld5 and Psf1 but C-terminal in Psf2 and Psf3. The A-domains comprise five α-helices in Sld5, Psf1 and Psf3 and four α-helices in Psf2 (a β-strand replaces one of the five helices seen in the other four subunits) and form an arch shape (this is particularly noticeable in Sld5 and Psf1; see Figure 4A) [8–10]. Heterodimerization of the A-domains of the top layer subunits Sld5 and Psf1, and of the bottom layer subunits Psf2 and Psf3, creates the vertical interface in the tetrameric complex and creates a 2-fold pseudo-symmetrical axis (indicated in Figure 3).

Key to the structure of each individual A-domain is an intrasubunit bidentate hydrogen bond formed, in the otherwise markedly
hydrophobic environment, between a highly conserved arginine residue in the α3 helix and a highly conserved glutamate residue in the α5 helix. Replacement of the arginine residue in fission yeast Psf2 with a lysine residue is the cause of the conditional-lethal (temperature-sensitive) phenotype of the psf2-209 allele [36], highlighting the importance of the hydrogen bond for the Psf2 structure and function.

The smaller B-domain comprises an extended N-terminal part which includes a single α-helix followed by a globular core which comprises up to five β-strands, arranged in two antiparallel β-sheets, and a single α-helix (shown for Sld5 and Psf2 in Figure 4B) [8–10]. Although the A-domain fold is reported to be distantly related to part of the structure of the cytoskeletal protein spectrin [9], no domains similar to the B-domain have been found.

In the tetramer, the B-domains of Psf2 and Psf3 create, through their interactions with the A-domain α1 and α3 helices of Sld5 and Psf1, the horizontal interface of the top and bottom layers (Figure 3) [8–10]. The effects of disrupting this interface can be seen with the budding yeast psf1-1 allele [5]: this mutation maps to an arginine residue in the Psf1 α3 helix that would normally be expected to protrude into a hydrophobic pocket in the B-domain of Psf3. Replacement of the arginine residue with glycine renders the mutant protein temperature-sensitive. In contrast with the N-terminal B-domains of Psf2 and Psf3, the C-terminal B-domain of Sld5 is located between the A-domains of Sld5 and Psf2 on the front surface of the complex (see Figure 3, right-hand panel). Deletion of the Sld5 B-domain weakens the horizontal interface between the top and bottom layers of the complex without affecting the vertical interfaces between Sld5 and Psf1 or Psf2 and Psf3 [9].

Absent from all three crystal structures is the C-terminal B-domain of Psf1. In two of three cases [9,10], the B-domain was removed from the complex prior to crystallization, whereas in the third [8], the B-domain was present in the crystallized material but was invisible in the electron density maps. Interestingly, biochemical experiments indicate that this domain does not interact stably with a tetrameric complex comprising C-terminally truncated Psf1 and full-length Sld5, Psf2 and Psf3, prompting the suggestion that it is tethered to the core complex by its linker region only [9]. Although not required for core complex formation, the Psf1 B-domain does have an essential function, however, and may be involved in protein–protein interactions with other components of the replication machinery. Addition of recombinant human GINS to egg extracts depleted of endogenous GINS restores full replication activity, but only when Psf1 is present in full-length; human GINS complexes containing C-terminally truncated Psf1 are incapable of replicating sperm chromatin, even when recombinant Psf1 B-domain is added in trans. The truncated complexes fail to stably associate with chromatin and Cdc45 chromatin binding is also greatly reduced, consistent with previous observations of interdependent chromatin binding of GINS and Cdc45 in Xenopus extracts [4]. Pols also fails to bind chromatin, whereas ORC and MCM loading (pre-RC formation) is unaffected (see Figure 1).

Note that, in addition to the Psf1 B-domain, the last few residues of Psf2, including putative phosphorylation sites for the ATM (ataxia-telangiectasia mutated) and ATR (ATM and Rad3 related) kinases, are also missing from all three structures (discussed further below).

**EM (electron microscopy) structures**

Prior to the publication of the crystal structures of the human GINS complex, two EM studies of GINS structure were reported. In the original report on the identification and characterization of Xenopus GINS, transmission EM images of rotary-shadowed GINS complexes (reproduced in Figure 5) were presented that appear to show a ring- or C-shaped structure with an estimated diameter of 100 Å and a central channel or hole with an estimated diameter of 40 Å [4]. A second EM study, this time of the recombinant human GINS complex, showed a C-shaped complex of broadly similar dimensions [43]. Neither of these EM structures is entirely consistent with the crystal structures; in particular, the size of the central channel is significantly smaller in the crystal structures and may not be large enough to accommodate single- or double-stranded DNA, as was previously suggested [4,43]. The reasons for these differences are unclear, but may reflect the limitations of the low-angle rotary shadowing technique used to prepare the GINS complexes for EM analysis [9]. Should it exist,
Figure 4 The human GINS subunits possess a common fold

(A) Crystal structures of the four individual human GINS subunits. The N-terminal A-domains of Sld5 and Psf1 form an arch shape with the C-terminal B-domain positioned above (seen for Sld5 only, as the Psf1 B-domain is not present in the solved structure). The arch shape of Psf2 and Psf3 is less apparent, but is formed by the A-domain helices and resembles that of Sld5 and Psf1. In Psf2 and Psf3, however, the B-domain is located at one end of the arch, at the N-termini of the proteins. (B) The isolated globular cores of the B-domains of Sld5 (left-hand side) and Psf2 (right-hand side). The structures shown are derived from PDB 2E9X [9] and drawn using MacPyMOL (DeLano Scientific).

the function of the central channel is unknown; a possible role for a short peptide from the N-terminus of Psf3 in gating access to the channel has been proposed [8], but what might be being gated is unclear.

ARCHAEAL GINS

The archaean comprise the third domain of life on Earth and possess chromosomal replication machinery that resembles a highly simplified form of that found in eukaryotic cells [51]. This apparent simplicity, coupled with the ease with which archaean replication factors, particularly those derived from hyperthermophilic archaean organisms, can be expressed and purified in recombinant form, has led to the adoption of the archaean as an important model system for eukaryotic chromosome replication. At least three major archaean phyla have been identified: the Crenarchaeota (which include only hyperthermophilic organisms), the Euryarchaeota (includes methanogens, halophiles, thermoacidophiles and some hyperthermophilic) and the Thaumarchaeota [52]. The Korarchaeota may represent a fourth phylum [53]. Among the replication proteins relevant to the present review that are conserved between the archaean and eukaryotes are ORC and Cdc6 (in archaean, the functions of these two factors appear to reside in one protein which has homology with both ORC and Cdc6), the MCM helicase (homohexameric in archaean, rather than heterohexameric), primase and GINS [51].

Identification and distribution of archaean GINS homologues

Archaean GINS homologues were first identified by bioinformatic methods [11] and are found encoded by all archaean species
whose genomes have been sequenced to date (S. A. MacNeill, unpublished work). As in eukaryotic cells, two types of GINS proteins are recognizable on the basis of the optimization of the conserved A- and B-domains (shown in Figure 2A). All archaeal organisms encode a protein with an A-B domain order resembling eukaryotic Sld5 and Psf1. These have been named Gins15 [12] or Gins51 [13]; in the present review, the latter nomenclature is used, consistent with the 5-1-2-3 origin of the term GINS (go-i-ichi-ni-san). In many cases, the gene encoding the Gins51 protein is found adjacent in the genome to the gene encoding PriS, the small subunit of primase, the enzyme responsible for synthesis of the RNA primer at the 5′ end of each Okazaki fragment on the lagging strand as well as at the 5′ end of the leading strand at the replication origin, or the gene encoding the polymerase processivity factor and sliding clamp PCNA (proliferating-cell nuclear antigen), or both. Figure 2(B) shows the genome context of genes encoding Gins51 proteins in species representative of the individual orders of the four putative archaeal phyla. In seven of 14 species shown, the gene encoding the primase subunit PriS is located immediately upstream of that encoding Gins51, whereas in three species, the gene encoding PCNA is found immediately upstream. In the three crenarchaeal organisms, all three genes are linked, whereas in the sole sequenced korarchaeal species, Korarchaeum cryptofilum, the gene encoding the large subunit of primase is also present (Figure 2B). In the remaining four species, the genes are not linked. Further analysis of gene context has established an intriguing link between replication and translation, discussion of which lies outside the scope of the present review [54,55].

Unlike the Gins51 proteins, Gins23 proteins, with their characteristic B-A domain order similar to Psf2 and Psf3, have been found encoded by some, but not all, archaeal species. Most of the euryarchaeia, including the methanogenic archaea (represented by the species indicated by Mth, Mma, Mac and Mhu in Figure 2B), the halorarchea (Hma) and the nanoarchaeia (Neq), appear to lack Gins23. This may simply indicate that the sequences of the Gins23 proteins in these organisms have diverged to the point where they can no longer be detected on the basis of amino-acid-sequence similarity: the known archaeal Gins23 proteins are already highly divergent. However, if Gins23 proteins are truly absent, intriguing questions are raised regarding the structure and stability of the GINS complex. As we shall see shortly, where both Gins51 and Gins23 proteins are present, the archaeal GINS complex is a tetramer made up of homodimers of each protein [12,13]. Given the important role for the N-terminal B-domains of Psf2 and Psf3 in forming the horizontal interface of the eukaryotic GINS complex (shown in Figure 3), it may be the case that in the absence of Gins23, the Gins51 protein is physically incapable of assembling into a stable two-layered, tetrameric structure like that found in eukaryotes and presumably also in archaea that encode both Gins51 and Gins23. Thus the GINS complex in these organisms may be dimeric. Since Gins51 proteins can be capable of homodimer formation when expressed in recombinant form (S. A. MacNeill, unpublished work), this question will only be answered by purification of GINS complexes from native protein extracts prepared from archaeal organisms that apparently lack Gins51. Interestingly, two of the species that possess genes encoding both Gins51 and Gins23 (Cenarchaeum symbiosum and K. cryptofilum, indicated as Csy and Kcr in Figure 2B) have been proposed to be members of deep-branching, ancient archaean orders [52,53]. If this is true, it implies that the gene encoding the Gins23 protein was present in the last common ancestor of the eukaryotic species and then lost at a later stage of evolution, after the archaean and eukaryotic lineages had diverged.

Biochemical activity

To date, two GINS complexes from evolutionarily distinct hyperthermophilic species, the crenarchaeote Sulfolobus solfataricus and the euryarchaeote Pyrococcus furiosus, have been characterized biochemically [12,13]. Both species encode both Gins51 and Gins23 proteins (Figure 2B). In S. solfataricus, Gins23 has been shown to interact with the non-catalytic N-terminal domain of the MCM helicase in yeast two-hybrid assays (indeed, this was how the gene was isolated, in a two-hybrid screen with MCM as the bait) and by co-immunoprecipitation from S. solfataricus cell extracts [12]. The genes encoding these two proteins overlap on the chromosome by 11 bp (Figure 2B). Gins23 also interacts in yeast two-hybrid assays and in vitro pull-down experiments with both subunits of primase. Purification of Gins23 from native protein extracts identified two further interacting factors, one of which was shown to be a GINS family protein and named Gins51. The gene encoding Gins51 is immediately adjacent in the chromosome to that encoding the small catalytic subunit of primase (Figure 2B).

When co-expressed in recombinant form, S. solfataricus Gins23 and Gins51 form a tetrameric GINS complex comprising a dimer of dimers. This complex presumably adopts a structure very similar to that seen with the human GINS complex (Figure 3), but with the difference that the top and bottom layers are composed of Gins51 and Gins23 homodimers respectively, rather than Sld5–Psf1 and Psf2–Psf3 heterodimers. All other aspects of the structure of the complex, such as the role of the Psf2 and Psf3 B-domains in creating, through their interactions the A-domains of Sld5 and Psf1 respectively, the horizontal interface between the top and bottom layers, might be expected to be preserved in the S. solfataricus GINS structure.

The other protein that co-purified with GINS displayed sequence similarity to the DNA-binding domain of the bacterial RecJ family of ssDNA exonucleases [12]. The RecJdBh protein (RecJ DNA-binding domain homologue) remained associated with the GINS protein through eight chromatographic steps, suggesting a very stable interaction, but its function is unknown [12]. Nevertheless, it is tempting to speculate that the potential ability of the RecJdBh to bind to ssDNA may be significant in the context of the likely role and location of the GINS complex at the replication fork.

The other archaeal GINS to be analysed biochemically is that of P. furiosus [13]. When expressed in recombinant form, the P. furiosus Gins51 and Gins23 proteins also form a tetrameric GINS complex in a 2:2 molar ratio. As seen with the S. solfataricus proteins, the P. furiosus Gins23 protein interacts in a two-hybrid assay with the MCM helicase, and the two proteins can be co-immunoprecipitated from cell extracts. The corresponding genes are located next to one another on the chromosome also (Figure 2B). Two-hybrid analysis also shows the Gins51 protein to interact with the P. furiosus Orc1/Cdc6 homologue, and ChIP studies using anti-Gins23 sera show the GINS complex to be associated with the P. furiosus replication origin in exponentially growing, but not stationary phase, cells [13].

Interestingly, recombinant P. furiosus GINS has been shown in in vitro assays to be able to stimulate MCM ATPase and helicase activities when present at up to a 4-fold molar excess over MCM [13], echoing the situation in eukaryotic cells with the CMG complex [29]. Unlike eukaryotic cells, however, there are no indications yet that the archaea have the potential to encode a homologue of Cdc45, the third component of the CMG complex. It is possible that this protein is present in archaean cells but that it has diverged in sequence to the point where it cannot be detected by even the most thorough database
searching. Alternatively, Cdc45 may only have evolved after the eukaryotic and archaean lineages diverged in evolution. In this scenario, archaean organisms might make do without a Cdc45-like function or that function might be provided by an altogether different protein, one such as RecJdbh perhaps. Taken together, the results from *S. solfataricus* and *P. furiosus* [12,13] place the archaean GINS complex at the heart of the replication apparatus, making contact with Orc1/Cdc6 at origins and with primase and MCM in the moving replisome, and in doing so underline the credentials of the archaean as simplified models for understanding the complexities of chromosome replication in eukaryotes.

**LINKS TO MITOSIS**

In eukaryotic cells, evidence exists pointing to a possible role for GINS in chromosome segregation that appears distinct from its role in S-phase [14]. The human Survivin protein forms a complex (called the CPC or chromosomal passenger complex) with the mitotic protein kinase Aurora B, the inner centromere protein INCENP and a protein called borealin, that apparently acts to co-ordinate multiple events during mitosis and cytokinesis [56]. The CPC localizes on the spindle in mitotic metaphase before relocating to the spindle midzone in anaphase. Survivin is phosphorylated by Aurora B and may have a role in regulating the activity of the kinase, which has itself been implicated in a number of processes including monitoring spindle tension, ensuring correct chromosome bi-orientation and spindle disassembly [56].

In fission yeast, the *psf2*+ gene was identified as a multicopy suppressor of a temperature-sensitive mutation, *bir1-46*, in the Survivin homologue Bir1 [14]. Overexpression of the Psf2 protein also rescues additional *bir1-46* mutant phenotypes including TBZ sensitivity (TBZ is a microtubule poison), suggesting a possible role for the protein in mitosis or cytokinesis. In support of such a role, a significant proportion of psf2Δ cells display chromosome missegregation phenotypes in mitosis, such as unequal nuclear division or lagging chromosomes. The same is true of cells in which *psf2* mRNAs are depleted by promoter shut-off, in cells carrying the temperature-sensitive *psf2-209* allele described above, and in *sls5Δ* cells. In addition, the localization of Bir1 in anaphase and telophase is disrupted in *sls5Δ* and *psf2Δ* cells. Human Psf2–EGFP [EGFP is enhanced GFP (green fluorescent protein)] fusion proteins localize in part to the spindle in metaphase and anaphase and to the midzone in cytokinesis, and, in transformed human HeLa cells, siRNA depletion of Psf2 also causes defects in the maintenance of nuclear morphology and chromosome missegregation. Fission yeast *psf2–GFP* may also localize to the mitotic spindle [14].

Taken together, these results suggest that GINS may have an important role in mitosis and cytokinesis in eukaryotic cells. The key question, however, is to what extent is such a role distinct and separable from the S-phase function of GINS? Are the observed mitotic defects simply a consequence of trying to separate incompletely replicated chromosomes? Attempts have been made in fission yeast to addres this issue, by waiting until S-phase is complete before depleting or inactivating GINS in G2 and then following the effects on the subsequent mitosis, but without clear-cut results [14]. In addition, a more recent study, this time using untransformed HDFs rather than HeLa cells, failed to observe aberrant mitotic phenotypes upon simultaneous siRNA depletion of Psf1 and Psf2 [48]. The explanation for this difference in behaviour is unclear, but begs further investigation.

**GINS FUNCTION IN VERTEBRATE DEVELOPMENT**

Several studies have appeared that characterize the requirement for GINS during vertebrate development [15–19]. In *Xenopus*, expression patterns of *SLD5*, *PSF1*, *PSF2* and *PSF3* have been analysed in *in situ* hybridization during embryonic development [18]. These studies showed the expression patterns of the four genes to be generally coincident, suggesting that they share some common regulation. Interestingly, high levels of expression do not always correspond with high rates of cell proliferation, again possibly hinting at roles for GINS outside of S-phase. Injection of anti-sense oligonucleotides (morpholinos) into *Xenopus* embryos to knockdown *PSF2* function results in impairment of a number of developmental processes, among which eye development has been studied in detail [19].

In adult mice, *PSF1* expression is readily detected in activity-proliferating tissues, such as bone marrow and thymus, and also in the reproductive tissues, the testis and ovary. Immunohistochemical studies of testis show the Psf1 protein to be located in immature cell populations, including the blastocysts and haematopoietic progenitor cells [16]. Targeted disruption of one *Psf1* allele in mice results in the expected reduction of *Psf1* transcript levels, but has no gross phenotypic consequences. However, crossing heterozygous *Psf1+/-* mice does not produce viable *Psf1+/-* offspring, with *Psf1+/-* embryos failing to develop post-implantation beyond embryonic stage E5.5 (where E is embryonic day), when blastocysts are present. It is likely that viability through the early stages of embryo development, up to E5.5, is assured by the presence of maternal *Psf1* transcripts. As the embryos reach stage E5.5, *Psf1* expression from the maternal mRNA is insufficient for continued cell growth [16]. In culture, *Psf1+/-* blastocysts display defects in cell proliferation and impaired DNA replication, as judged by failure to incorporate the nucleotide analogue BrdU (bromodeoxyuridine) into DNA. Similar phenotypes are seen in homozygous *Cdc45+/-* embryos [57]. Although the heterozygous *Psf1+/-* embryos are viable and do not display gross phenotypic defects, closer investigation of the processes of haemopoiesis has revealed that *Psf1* haploinsufficiency leads to impaired stem-cell proliferation in response to haemopoietic stress and consequent reduced survival [17]. Underlining the importance of GINS in normal development, it has also been reported that *PSF2* expression is significantly elevated in certain types of human hepatic carcinomas [15]. Whether this contributes directly to tumorigenesis is unclear at present, but as mutations in *MCM4* and down-regulation of *MCM2* have already been shown to cause cancer in mice [58–60], it would not be surprising if deregulation or mutation of GINS were to have similar consequences.

**REGULATION OF GINS ACTIVITY**

The activities of various components of the replication machinery are tightly regulated, to ensure once-per-cell-cycle replication, for example, or to allow the replication machinery to respond to the ever-present threat of DNA damage, and recent results indicate that this is also likely to be the case for GINS [20]. Human Psf2 was identified in an extensive proteomic-based survey of proteins phosphorylated by the ATM and ATR protein kinases [20], two key mediators of the DNA-damage response with overlapping, but non-redundant, functions [61,62]. In the case of Psf2, two sites of phosphorylation were identified: Thr180 and Ser182. ATM and ATR recognize serine and threonine residues in SCDs (SQ/TQ cluster domains) [63]; Thr180 and Ser182 are found in the sequence SQ/TQ,
close to the C-terminus of Psf2 (and absent from all three crystal structures of the human GINS complex) [8–10].

At present it is not known what effect these phosphorylation events have on GINS function, nor is it known how they contribute to the cellular DNA-damage response. Neither phosphorylation site is particularly well-conserved across evolution. Budding yeast Psf2, for example, does not contain any SQ/TQ motifs, although fission yeast Psf2 does have an SQ close to its C-terminus. It would be interesting, and relatively straightforward technically, to determine whether mutating this site affected the ability of the cells to deal with the DNA damage in S-phase.

Another aspect of GINS function in higher eukaryotes that has been little investigated is the potential role for differential expression or alternative splicing in generating different isoforms of the GINS proteins. In a recent study, differences in PSF1 transcript structure in testis were described that could potentially give rise to the expression of N-terminally truncated Psf1 proteins [64]. Direct biochemical evidence for the existence of such proteins is lacking, however, so it remains to be seen whether variants of this type have any part to play in GINS function.

CONCLUSIONS AND PERSPECTIVES

Since its discovery in 2003 [3–5], GINS has moved rapidly to centre stage in our attempts to fully understanding the workings of the chromosome replication machinery in eukaryotic and archaean cells. GINS is undoubtedly a key component of the likely replicative helicase, the CMG complex [29,50], but what exactly does GINS contribute to the activity of the CMG? What is its biochemical function? The structure of the GINS complex in isolation [8–10] offers no clues to this. The requirement for GINS for replication elongation suggests an intimate association with processive DNA unwinding, but more detailed speculation on GINS function is hampered by our lack of understanding of the mechanism of the MCM helicase itself. As discussed elsewhere [65], various models have been proposed for how MCM might act at the fork, largely depending on whether the protein is present as a hexamer or a double hexamer, and whether the ring-shaped MCM complex encircles single- or double-stranded DNA. The elegant demonstration that the RPC contains only a single MCM complex [6] probably resolves the first of these issues, but the single-stranded versus double-stranded DNA debate continues. One intriguing model posits that the MCM hexamer is not actually a helicase at all, i.e. that it is not intrinsically able to unwind duplex DNA, but simply a double-stranded DNA translocase [65]. In this model, duplex DNA is pumped in an ATP-dependent manner through the central channel of the MCM hexamer without being unwound. However, exit from the channel for duplex DNA is blocked by the presence of an unknown protein and it is some part of the structure of this additional factor (termed the ploughshare after the main cutting blade of a plough) that sterically separates the two strands. Could GINS harbour the ploughshare? The ability of recombinant human GINS to bind preferentially ssDNA [43] would certainly be consistent with GINS being located in proximity to unwound DNA emerging from the CMG, but aside from this, evidence for a precise role for GINS is lacking. A complete crystal structure of the CMG, ideally complexed with a synthetic DNA substrate, would be hugely informative in this regard, but the technical challenges involved in generating sufficient material in a form appropriate for crystallization are substantial, and there is no guarantee that crystallization itself would be successful. Until these challenges are met, further genetic and biochemical analysis of GINS function in a variety of experimental systems is required and will no doubt provide important insights into the function of this essential protein complex.

FUNDING

Work in the author’s laboratory is funded by the Scottish Universities Life Sciences Alliance (SULSA).

REFERENCES

40 Blow, J. J. (2001) Control of chromosomal DNA replication in the early Xenopus embryo. EMBO J. 20, 3293–3297
SUPPLEMENTARY ONLINE DATA

Structure and function of the GINS complex, a key component of the eukaryotic replisome

Stuart A. MacNEILL

Biomedical Sciences Research Complex, School of Biology, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, U.K.

Table S1 Entrez GeneIDs for genes in Figure 2(B) in the main paper

<table>
<thead>
<tr>
<th>Species</th>
<th>Gins23</th>
<th>MCM</th>
<th>PCNA</th>
<th>PriL</th>
<th>PriS</th>
<th>Gins51</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sso</td>
<td>1455037</td>
<td>1455038</td>
<td>1454088</td>
<td>1454089</td>
<td>1454090</td>
<td></td>
</tr>
<tr>
<td>Ape</td>
<td>1445714</td>
<td>1445715</td>
<td>1444627</td>
<td>1444626</td>
<td>1444625</td>
<td></td>
</tr>
<tr>
<td>Pae</td>
<td>1465397</td>
<td>1463798</td>
<td>1463797</td>
<td>1463796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mth</td>
<td></td>
<td>1471029</td>
<td></td>
<td></td>
<td>1471029</td>
<td></td>
</tr>
<tr>
<td>Tac</td>
<td></td>
<td></td>
<td>1456561</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mac</td>
<td></td>
<td></td>
<td>1472540</td>
<td>1472539</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mhu</td>
<td></td>
<td></td>
<td>3922510</td>
<td>3922509</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hma</td>
<td></td>
<td></td>
<td>3128913</td>
<td>3128974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atu</td>
<td></td>
<td></td>
<td></td>
<td>1484558</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mna</td>
<td></td>
<td></td>
<td>2761382</td>
<td>2761381</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plu</td>
<td>1468324</td>
<td>1468323</td>
<td>1468848</td>
<td>1468847</td>
<td>2654353</td>
<td></td>
</tr>
<tr>
<td>Neq</td>
<td>6371904</td>
<td>6371903</td>
<td></td>
<td>6372034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Csy</td>
<td>6094628</td>
<td>6094629</td>
<td>6094806</td>
<td>6094807</td>
<td>6094808</td>
<td></td>
</tr>
<tr>
<td>Kcr</td>
<td>6094628</td>
<td>6094629</td>
<td>6094806</td>
<td>6094807</td>
<td>6094809</td>
<td></td>
</tr>
</tbody>
</table>

Received 5 October 2009/6 November 2009; accepted 20 November 2009
Published on the Internet 15 January 2010, doi:10.1042/BJ20091531

1 email stuart.macneill@st-andrews.ac.uk