Structural diversity and dynamics of genomic replication origins in *Schizosaccharomyces pombe*

Cristina Cotobal, Mónica Segurado
and Francisco Antequera*

Instituto de Microbiología Bioquímica, CSIC/Universidad de Salamanca,
Edificio Departamental, Salamanca, Spain

DNA replication origins (ORI) in *Schizosaccharomyces pombe* colocalize with adenine and thymine (A+T)-rich regions, and earlier analyses have established a size from 0.5 to over 3 kb for a DNA fragment to drive replication in plasmid assays. We have asked what are the requirements for ORI function in the chromosomal context. By designing artificial ORIs, we have found that A+T-rich fragments as short as 100 bp without homology to *S. pombe* DNA are able to initiate replication in the genome. On the other hand, functional dissection of endogenous ORIs has revealed that some of them span a few kilobases and include several modules that may be as short as 25–30 contiguous A+Ts capable of initiating replication from ectopic chromosome positions. The search for elements with these characteristics across the genome has uncovered an earlier unnoticed class of low-efficiency ORIs that fire late during S phase. These results indicate that ORI specification and dynamics varies widely in *S. pombe*, ranging from very short elements to large regions reminiscent of replication initiation zones in mammals.

The EMBO Journal (2010) 29, 934–942. doi:10.1038/emboj.2009.411; Published online 21 January 2010

Subject Categories: cell cycle; genome stability & dynamics

Keywords: cell cycle; DNA replication; genome organization; replication dynamics; replication origin

Introduction

Eukaryotic replication origins (ORIs) are chromosomal regions in which the origin recognition complex (ORC) and other initiator proteins assemble to initiate DNA replication (DePamphilis, 2005). Although endowed with the same function, there are significant structural differences and sequence degeneracy between individual ORIs even within the same genome, making it difficult to predict their localization on the basis of their sequence. Among eukaryotes, ORIs have been studied in great detail in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* because of their compact genomes and the feasibility of genetic approaches. In *S. cerevisiae*, a consensus sequence 11–17 bp long (Theis and Newlon, 1997) is required for origin activity, although only a

small proportion of all of these elements colocalizes with active ORIs in the genome (Breier *et al*, 2004). Even in this yeast, ORIs are degenerated in the sense that auxiliary elements, that are also required for ORI activity, vary between different ORIs (Theis and Newlon, 2001; Dershowitz *et al*, 2007).

In *S. pombe*, it has long been known that autonomously replicating sequence (ARS) elements have a high content in adenine and thymine (A+T) (Maundrell *et al*, 1988) and deletion analyses have failed to identify any essential consensus element (Clyne and Kelly, 1995; Dubey *et al*, 1996). A+T richness has also been found to be a property of genomic ORIs (Dubey *et al*, 1994; Okuno *et al*, 1997; Gómez and Antequera, 1999) and this feature led to the identification of A+T-rich islands, which colocalize with active genomic ORIs in 90% of the cases tested. A+T-rich islands were initially defined as regions 0.5–1 kb long with an A+T content ranging between 72 and 75% (Segurado *et al*, 2003). Recent studies have shown that the majority of these regions bind ORC *in vivo* (Hayashi *et al*, 2007) and colocalize with distinct peaks of early replication during the S phase (Feng *et al*, 2006; Heichinger *et al*, 2006; Mickle *et al*, 2007). See Mickle *et al* (2007) for a detailed comparison between predictions of *S. pombe* genomic ORIs and for the different names given to them using different approaches.

In contrast with the diversity of ORI sequences, the six subunits of ORC are conserved among all eukaryotes studied (Bell and Dutta, 2002). An important exception arose with the identification of the gene encoding the Orc4 subunit of *S. pombe* (Chuang and Kelly, 1999). This protein contains a carboxy domain of about 450 amino-acids homologous to the Orc4 protein in other eukaryotes in addition to a—so far exclusive—amino-domain of a similar size that includes nine AT-hook motifs capable of binding A+T-rich stretches 4–8 bp long *in vitro* regardless of their primary sequence (Reeves and Beckerbauer, 2001). This domain is not required for the assembly of ORC *in vitro*, but mutant complexes lacking it are unable to bind to ORIs (Lee *et al*, 2001) and this is probably the reason why the expression of the carboxy domain alone cannot rescue the lethal phenotype caused by the absence of the wild-type (wt) protein (Chuang *et al*, 2002).

This two-component system made up of A+T-rich ORIs and Orc4p provides a reasonable explanation of how ORC is targeted to specific sites in the genome to initiate replication. Nevertheless, this general model still leaves unanswered the question of how ORIs are specified in the *S. pombe* genome; that is which are the minimal requirements in terms of size and base composition for a genomic region to recruit ORC and initiate replication. This question is relevant because only a small proportion of intergenic regions displays detectable ORI activity in the genome (Gómez and Antequera, 1999; Segurado *et al*, 2003) and bind ORC *in vivo* (Hayashi *et al*, 2007) despite having an average composition as high as 70% A+T.

In this study, we have found that ORIs in the *S. pombe* genome range from short A+T-rich elements to initiation

*Corresponding author. Instituto de Microbiología Bioquímica, CSIC/Universidad de Salamanca, Edificio Departamental, Campus Miguel de Unamuno, Salamanca 37007, Spain. Tel.: +34 923 121778; Fax: +34 923 224876; E-mail: cpg@usal.es

Received: 20 July 2009; accepted: 21 December 2009; published online: 21 January 2010

zones made up of several degenerate elements each of which has the potential to drive replication individually. These results dramatically reduce earlier estimates of ORI size in *S. pombe* based on plasmid assays and have led to the finding of a new class of low-efficiency and late-firing genomic ORIs.

Results

Artificial A + T-rich DNA fragments specify active replication origins in *S. pombe*

Earlier work from several laboratories has proposed that ARS elements in *S. pombe* are large and lack consensus sequences (Maundrell *et al*, 1988; Clyne and Kelly, 1995; Dubey *et al*, 1996; Okuno *et al*, 1997; Dai *et al*, 2005) and deletion experiments have identified important elements for replication initiation in plasmids (Clyne and Kelly, 1995; Dubey *et al*, 1996; Okuno *et al*, 1999) and in the chromosome (Takahashi *et al*, 2003). These studies have invariably used *S. pombe* DNA either as a donor for ARS assays or have manipulated endogenous genomic sequences to identify relevant functional elements. We have taken a more radical approach to address the question of what are the minimum requirements to specify active ORIs by asking whether custom-made synthetic DNA fragments covering a range of sizes and A + T content could mimic the activity of endogenous ORIs.

As a first step, we synthesized three sets of DNA fragments 100 bp long with an A + T content of 75, 83 and 90%, respectively. Every 100 bp monomer within each class had a different sequence and none of them derived from *S. pombe* or showed significant similarity to any sequence in its genome (see Materials and methods for construction and cloning details and Supplementary Figures 1 and 2 for sequences). The three sets of fragments were self-ligated independently and cloned into the 2.9 kb pBSK+ plasmid lacking an *S. pombe* replication origin, but harbouring the *ura4* gene as a selectable marker. Ligation mixtures were used directly to transform an *S. pombe ura-d18* strain auxotroph for uracil, and cells that incorporated plasmids capable of autonomous replication were selected on minimal medium plates. Sequence analysis of cloned inserts from a random selection of the transformant colonies showed that the minimum size of fragments displaying ARS activity was approximately 100, 200 and 500 bp in the 90, 83 and 75% A + T sets, respectively (Figure 1A).

These results indicated that A + T-rich DNA sequences alien to the *S. pombe* genome could drive efficient plasmid replication. It remained possible, however, that cells might have selected fragments containing some cryptic sequence pattern meaningful for them, but unidentifiable to us. To test this possibility, we cloned 15 individual fragments of between 102 and 472 bp and 83% A + T into the 2.9 kb pBSK+ plasmid and recovered them in *E. coli*. ARS analysis in *S. pombe* showed that all fragments longer than 200 bp were active, ruling out the possibility that specific sequence elements were required (Figure 1B). In addition, and consistent with the results in Figure 1A, none of the six fragments approximately 100 bp long was active, probably because of their small size. To test whether these fragments could be active in the context of a larger plasmid, we repeated the ARS assay after cloning them into the 5.3 kb pJK148 plasmid.

Figure 1 Plasmid replication by artificial ARS elements. (A) Size of ARS fragments of 90% (red), 83% (yellow) and 75% (green) A + T content capable of replicating the pBSK+ plasmid harbouring the *ura4* gene as selectable marker. Sequences of the 27 fragments are shown in Supplementary Figure 1. (B) ARS activity of 15 fragments from 102 to 472 bp long with an 83% A + T content. Efficiencies are relative to the positive control *ars1* (1.2 kb) (black bar). Plasmid pBSK+ with the *ura4* gene without any ARS element was used as a negative control (C-). The results are average of two independent experiments. The sequences of all the fragments are shown in Supplementary Figure 2. (C) Inactive monomers (m) (a, b and c) from panel (B) display increasing ARS activity when self-ligated as dimers (d) or trimers (t). *Ars1* and C-, positive and negative controls as in (B).

None of the six fragments generated any colony except the 111 bp long that generated a reduced number of very small and heterogeneous colonies (data not shown). To test whether the lack of ARS activity of these fragments was indeed because of their small size, we self-ligated three different inactive 100 bp monomers (a, b and c in Figure 1B) into dimers and trimers and cloned them in the pBSK+ plasmid. Figure 1C shows that the three sets of oligomers made up of individually inactive monomers reached an ARS activity comparable with that of the positive control *ars1* depending on their size.

Could these ARS elements of arbitrary sequence also initiate replication in the genome? To answer this question, we selected 11 fragments of between 100 and 700 bp and 75, 83 or 90% A + T identified in Figure 1A and targeted them by homologous recombination to position 108845 in chromosome III, between the *SPCC330.19c* and *SPCC330.03c* genes (Figure 2A). The nearest predicted ORIs were 32 kb upstream and 10 kb downstream from this site (Segurado *et al*, 2003; Hayashi *et al*, 2007) and we have earlier confirmed by two-dimensional electrophoresis that this region does not contain a replication origin (Gómez and Antequera, 1999; Figure 2B, wt). Integration of the *ura4* gene by itself did not change the passive replication pattern of this locus (Figure 2B, *ura4*). To test whether this region was permissive for replication initiation, we showed that *ars1* was equally efficient as an ORI when integrated into this region (Figure 2B, *ars1*) as in its endogenous location (Gómez and Antequera, 1999). Analysis of the integrated exogenous fragments showed that 9 out of 11 were capable of initiating replication. Only one out of the three fragments of 75% A + T showed weak ORI activity

Figure 2 Artificial ORIs initiate replication in the *S. pombe* genome. (A) Diagram of the insertion site (black arrow) of artificial ARS elements in chromosome III. White arrowheads indicate nearest predicted ORIs. rDNA represents ribosomal RNA locus at the left end of the chromosome. The expanded region below shows genes surrounding the insertion site. Bracket indicates the genomic region centred on the insertion site tested by two-dimensional electrophoresis. (B) Replication of the 3.8 kb Xba I restriction fragment in wild-type cells (wt), and after integration of the *ura4* marker or *ars1*. The presence of a bubble replication arc in *ars1*, which indicates the presence of an active ORI, is indicated by a black arrow. Bubble arcs in all the other 2D gels in this work are also indicated by black arrows. (C–E) Replication pattern on insertion of the indicated fragments from Figure 1A with a 75, 83 and 90% A+T content, respectively.

(Figure 2C, 75.5) despite their relatively large size, suggesting that fragments of this base composition probably represent the lowest boundary required for genomic ORIs to be specified. This notion was supported by the fact that fragments with 75% A+T yielded a significantly lower number of transformants than those of 83 and 90% in plasmid ARS assays (data not shown). In contrast, all eight 83 and 90% A+T fragments tested were capable of initiating replication efficiently (Figure 2D and E). The presence of the *ura4* marker linked to the integrated fragments was not required for ORI activity, as shown by the unchanged activity of the 83.1 fragment on removal of *ura4* from this region in the genome (Figure 2D, 83.1–*ura4*).

Taken together, these results indicated that A+T-rich fragments of arbitrary sequence could drive replication initiation as efficiently as endogenous ORIs, and confirmed the lack of the requirement of a specific sequence pattern. A second conclusion was that fragments as short as 100–200 bp could drive robust ORI activity, indicating that genomic ORIs could be specified by regions much shorter than those found in plasmid ARS assays (Maundrell *et al*, 1988; Dubey *et al*, 1996; Okuno *et al*, 1999; Dai *et al*, 2005).

Replication initiation zones in the *S. pombe* genome are made up of individually active redundant elements

Many *S. pombe* chromosomal ORIs colocalize with regions of 0.5–1 kb with a high A+T content that we have called A+T-rich islands (Segurado *et al*, 2003). The results depicted in Figures 1 and 2 showed, however, that efficient replication could be driven from shorter regions and this raised the question of whether A+T-rich islands were an absolute requirement for replication initiation. To address this question, we chose the intergenic region encompassing A+T island 1090 (Figure 3A). This region is located between the

vma5 and *SPAPB2B4.04c* genes in chromosome I and colocalizes with a strong ORI (Figure 3C, wt). The NcoI restriction fragment of 5.6 kb in wt cells tested by two-dimensional electrophoresis encompassed the A+T island and the entire intergenic region, making it impossible to distinguish whether independent elements could contribute to ORI activity. As a first step in the functional dissection of this region, we assayed the ARS activity of five non-overlapping fragments across the regions 1, 5, 6, 7 and 8 in Figure 3A. We found that fragment 5 (321 bp and 79% A+T), which spanned the A+T island, was active, consistent with the results in Figure 1 (Figure 3B). In addition, fragment 1 (1.1 kb) displayed an even higher activity despite the fact that its global A+T content was 68%, which is very similar to the 70% intergenic average. Cleavage of this fragment into two halves generated fragment 2, which was inactive, and fragment 4, which retained over 50% of the activity. Close sequence examination revealed the presence of an uninterrupted stretch of 29 A+Ts (29W) at the 5' end of this fragment. Deletion of this element generated fragment 3 and resulted in the complete loss of ARS activity (Figure 3B). The 29W stretch also lacked ARS activity by itself when cloned into the plasmid. Fragments 6, 7 and 8 were also inactive despite their relatively large size.

To evaluate the relative contribution of fragments 5 and 29W to ORI activity, we deleted them independently or in combination in their chromosomal context. Two-dimensional electrophoresis revealed that in both cases replication initiation was diminished, but not abolished, indicating the presence of other elements capable of sustaining residual activity (Figure 3C). Deletion of half of the intergenic region encompassing fragments 5 and 1 (which included 29W) ($\Delta 5 + \Delta 1$) resulted in a further reduction in activity, which probably derived from cooperation between fragments 6, 7

Figure 3 Functional dissection of the intergenic region encompassing A + T-rich island 1090. (A) A + T content in 300 bp windows (step 50 bp) across the region between the divergently transcribed genes *SPAPB2B4.04c* (red) and *vma5* (blue) genes. A + T island 1090 is indicated by an arrow. Vertical lines represent restriction sites for *NcoI* flanking the 5.6 kb region tested by two-dimensional electrophoresis. The genomic intergenic average is 70% A + T (red line). Fragments tested for ARS activity in panel (B) are indicated by numbered colour bars and the 29 bp stretch of A + Ts by a yellow triangle (29W). (B) ARS activity of fragments in panel (A) relative to *ars1* (black bar). (C) Two-dimensional electrophoresis of the region indicated in panel (A) in exponential wild-type cells (wt) and in cells deleted for region 5 ($\Delta 5$), 29W ($\Delta 29W$), both simultaneously ($\Delta 5 + \Delta 29W$), and regions 5 and 1 combined ($\Delta 5 + \Delta 1$).

Figure 4 ORI specification by individual modules at an ectopic location. (A) Two-dimensional gel analysis of fragments 5 and 29W integrated at the insertion site (IS, black arrow) in chromosome III shown in panel (B). (B) A 1 kb region encompassing the insertion site (ISR 1 kb), the 29W element alone (29W) and the 29W element integrated into the ISR 1 kb fragment (ISR + 29W) were cloned into the pBSK + *ura4* vector and tested for ARS activity. The histogram shows that neither of the two individual fragments showed any ARS activity, but the composite fragment yielded the same number of colonies as the positive control *ars1* (black bar). (C) Colony growth in plates of minimal medium after 4 days at 32°C. Cells transformed with the plasmid carrying the ISR + 29W fragment grew at a much lower rate than the *ars1* control plasmid. Neither fragment generated any transformants when assayed independently (top panels).

and 8, each of which lacked ARS activity when assayed individually. Indeed, when these three fragments were ligated into a plasmid to generate a unique fragment in the genomic configuration (6–8), their ARS activity was comparable with that of *ars1*, although the colonies grew at a much smaller rate comparable with that shown in Figure 4C (see later).

To test whether fragments 5 and 29W contained enough information to drive replication by themselves in the genome,

we integrated them independently at the ectopic location in chromosome III (Figure 4B). Fragment 5 drove replication from this site (Figure 4A, 5/Chr III), consistent with the observation that fragments of a comparable size and A + T content were also active in this locus (Figure 2). Contrary to our expectations, however, we found that the 29W element also generated an active, although less efficient, ORI on integration in the same site (Figure 4A, 29W/Chr III). To confirm the presence of the weak bubble arc relative to

the passive replication arc, we repeated this 2D gel in triplicate from cultures grown from independent colonies with the same result in the three cases (data not shown). This unanticipated result raised the question of how such a small fragment could specify a genomic ORI despite the fact of being inactive when assayed in a plasmid. We speculated that 29W might recruit ORC through the AT-hooks of the Orc4p subunit and that this binding could be stabilized by the surrounding context of the intergenic region that had a higher A + T content than when cloned in the plasmid. To test this possibility, we measured the ARS activity of the 29W element cloned in the middle of a 1 kb fragment encompassing the insertion site region (ISR, Figure 4B). This region was unable to initiate replication in the genome (Figure 2B, wt) and did not display ARS activity in a plasmid either (Figure 4C, ISR). However, insertion of 29W into this inactive fragment (ISR + 29W) generated a similar number of colonies in the ARS assay as the positive *ars1* control, although the colonies grew at a much slower rate (Figure 4C), which is consistent with the different efficiency of *ars1* and 29W to initiate replication on integration in chromosome III (compare *ars1* in Figure 2B with 29W/Chr III in Figure 4A). The inability of 29W to sustain ARS activity by itself could explain why short elements on this kind had not been identified earlier in ARS assays. These results supported the hypothesis that 29W could recruit ORC to DNA and the complex could be further stabilized by binding to flanking regions and implied that the minimum size of a DNA sequence to specify an active ORI could be even shorter than the 100–200 bp found in Figures 1 and 2.

We tested this possibility directly by chromatin immunoprecipitation (ChIP) with an antibody against the Orc4 subunit of ORC. The results of four immunoprecipitation assays from independent cultures of the *S. pombe* strain carrying the ectopic 29W insertion and another four of the wt strain as a control showed a 23% increase in the amount of immunoprecipitated Orc4 in the strain harbouring the ectopic 29W element relative to the wt control (Supplementary Figure 3). Although small, this increase relative to background is consistent with the ORC binding profile of other poly-W ORIs found in an earlier ChIP-on-chip genome-wide analysis using antibodies against Orc4, Orc1 and Mcm6 (Hayashi *et al*, 2007).

Poly-W stretches colocalize with late-firing genomic ORIs

There are 460 non-overlapping stretches of 29 contiguous A + Ts in the *S. pombe* genome, approximately half of which map to intergenic regions devoid of A + T-rich islands and do not colocalize with the sites of ORC binding (Hayashi *et al*, 2007). Considering that perhaps even shorter stretches could drive replication, an immediate implication of these results is that the number of active ORIs in the genome could be much higher than current estimates (Segurado *et al*, 2003; Heinricher *et al*, 2006; Hayashi *et al*, 2007; Mickle *et al*, 2007). To test this possibility, we selected 12 genomic regions that contained uninterrupted stretches of homogenous Ts (or As) or a combination of both bases (W) between 21 and 41 bp long and analysed them by two-dimensional electrophoresis. All regions tested were at least 12 kb distant from the nearest predicted ORI (Segurado *et al*, 2003; Hayashi *et al*, 2007). The results depicted in Figure 5 show that 8 out of 12 regions displayed bubble arcs of an intensity comparable with that generated by the 29W element when inserted in chromosome III (Figure 4A, 29W/Chr III). In sharp contrast with ORIs colocalizing with A + T-rich islands such as 1090 (Figure 3C, wt and Segurado *et al*, 2003), a common feature of poly-W ORIs was their low efficiency. This was indicated by the low ratio of the intensity between bubble and fork arc signals that, in some cases, bordered the limits of detection (Figure 5).

A general correlation between ORI efficiency and time of activation during S phase has been described in *S. pombe* (Heinricher *et al*, 2006). To address whether poly-W ORI firing was biased towards any particular period during S phase, we synchronized *S. pombe cdc10-129* temperature-sensitive mutant cells by arresting them in G1 at the restrictive temperature. On release at the permissive temperature, we monitored the firing of the strong origin *AT1090* and five of the poly-W ORIs identified in Figure 5 at the indicated times during the ensuing S phase (see Supplementary Figure 4 for the FACS analysis). As shown in Figure 6, poly-W ORIs 31W, 39T and 41W fired at 60 min, and 33T and 28W at 90 min into S phase, whereas bubble intermediates were already detectable in the 30 min sample in origin *AT1090*. As discussed below, the low efficiency and late firing of poly-W ORIs is probably because of a low affinity for ORC, as

Figure 5 Short poly-W tracts colocalize with active ORIs. Two-dimensional gel electrophoresis of 12 genomic regions including runs of contiguous thymines (T) or combined thymines and adenines (W) in asynchronous, exponentially growing wild-type cells. Figures indicate the number of nucleotides in each track. The genomic localization of poly-W stretches and of the restriction fragments analysed are indicated in Supplementary Figure 6.

Figure 6 Poly-W ORIs fire late during S phase. *S. pombe cdc10-129* temperature-sensitive mutant cells were arrested in G1 at 36°C for 4 h. Samples were taken at the indicated times during S phase and the activation time of ORIs 1090, 31W, 39T, 41W, 33T and 28W was monitored by two-dimensional electrophoresis. Samples from different time points contained the same amount of DNA and autoradiographs were exposed for the same length of time. Differences in the intensity of passive replication arcs are due to the variation in the amount of replication intermediates at different times during S phase, consistent with the FACS analysis of the progress across S phase shown in Supplementary Figure 4. The same membrane was hybridized successively with probes against 1090, 31W and 28W ORIs. Independent membranes were used for the remaining hybridizations. A sample from exponential asynchronous *cdc10-129* cells (Exp) growing at 25°C was also included in the analysis.

indicated by the fact that only three out of the eight positives in Figure 5 (21T, 33T and 34T) were identified as sites of ORC binding by ChIP analysis (Hayashi *et al*, 2007).

Discussion

ORI specification in *S. pombe*

An essential and as yet unanswered question in the eukaryotic DNA replication field is the identification of the elements that specify active ORIs in the genome. In *S. pombe*, plasmid replication assays have established that ARS fragments range between 0.5 and over 3 kb (Maundrell *et al*, 1988; Clyne and Kelly, 1995; Dubey *et al*, 1996; Okuno *et al*, 1997; Dai *et al*, 2005), and it is assumed that similar requirements would apply to genomic ORIs. In fact, size is often cited as one of the major differences between ORIs in *S. pombe* and *S. cerevisiae*, in which ARS elements can be shorter than 200 bp (Van Houten and Newlon, 1990; Marahrens and Stillman, 1992). Contrary to this assumption, our results show that poly-W tracks as short as 30 bp long are sufficient to specify replication initiation in the context of

intergenic regions (Figure 4) and to reliably predict the localization of active ORIs in the *S. pombe* genome (Figure 5). An immediate consequence of these results is that current estimates of ORIs in the genome should be expanded, and at the same time raises the question of why this class of ORIs has not been detected in earlier studies.

Segurado *et al* (2003) found that 90% of A + T-rich islands colocalized with relatively strong ORIs (e.g. see A + T island 1090 in Figure 3). However, because that study used a lower size limit of 500 bp, based on earlier ARS data in the literature, many of the ORIs specified by short poly-W tracks were probably excluded from that prediction. Hayashi *et al* (2007) identified 460 sites of ORC binding by ChIP followed by microarray hybridization, many of which colocalized with A + T-rich islands. However, five out of the eight weak ORIs in Figure 5 (31W, 34W, 28W, 39T and 41W) were not detected in that study, indicating that a large proportion of this class of ORIs either bind ORC very weakly or do so only in a small proportion of the cells, such that their identification falls below the threshold of reliable detection by ChIP. A third study based on microarray hybridization using DNA

replicated in the presence of hydroxyurea predicted the existence of 401 high-efficiency and 503 low-efficiency ORIs in *S. pombe* (Heichinger *et al*, 2006). It is difficult to ascertain how many of these low-efficiency ORIs colocalize with poly-W tracks because the average density of the microarray used (about 1 probe/kb) often left gaps spanning several intergenic regions between two adjacent probes. In addition, hydroxyurea prevents the firing of late ORIs in *S. cerevisiae* and in *S. pombe* (Santocanale and Diffley, 1998; Kim and Huberman, 2001; Mickle *et al*, 2007), and it is likely that these conditions might have prevented the detection of a significant proportion of the late-firing poly-W ORIs. In any case, our results lend support to the study of Heichinger *et al* (2006), and we would expect some overlap between both independent predictions.

The presence or absence of poly-W ORIs contributes to explaining why some intergenic regions contain active, although low-efficiency, ORIs, whereas some do not do so at a detectable level in plasmids (Dai *et al*, 2005) and in the chromosome (Gómez and Antequera, 1999). This suggestion is supported by the fact that only 4 out of 15 intergenic regions devoid of ORI activity earlier analysed (Gómez and Antequera, 1999; Segurado *et al*, 2002) contained a poly-W tract between 20 and 25 bp.

Narrow and extended replication initiation zones in *S. pombe* correlate with the time of firing during S phase

Our results point to a broad diversity in the way ORIs can be specified in *S. pombe*. Although poly-W tracts are probably close to the minimum sequence information capable of binding ORC through the Orc4 subunit (Supplementary Figure 3; Maher and Nathans, 1996), other ORIs are specified by the combined contribution of several elements across several kilobases, such as origin *AT1090* (Figure 3). The arrangement in several independent modules is also present in *ars 2004*, in which individual deletions of three A+T-rich non-overlapping regions reduced, but did not prevent, ORI activity in the genome (Takahashi *et al*, 2003). This organization is consistent with the fact that ORC binds at various sites across *ars 2004* and *ars 3001* (Kong and DePamphilis, 2002; Takahashi *et al*, 2003) and that *in vitro* footprinting analysis has shown that ORC protects a wide region in the *ars1* origin (Lee *et al*, 2001). It is likely that many ORIs associated with A+T-rich islands (such as origin *AT1090*, *ars 2004*, *ars 3001* and *ars1*) have an extended organization and this could be the reason why the average size of intergenic regions in which this class of ORIs lie is twice as long (2 kb) as those devoid of them (data not shown).

Although at a different scale, the functional organization of these regions is similar to that of the 55 kb initiation zone downstream from the hamster *DHFR* gene (Dijkwel *et al*, 2002). Both regions are made up of individual elements whose deletion does not abolish ORI activity (Figure 3; Kalejta *et al*, 1998) and which have the potential to initiate replication from ectopic positions in the genome (Figure 4; Altman and Fanning, 2001). The picture that emerges from these observations is that ORI organization in *S. pombe* is comparable with that found in mammalian genomes, in which short *Lamin B2* (Abdurashidova *et al*, 2000) and β -globin human ORIs (Kitsberg *et al*, 1993) coexist with larger initiation zones such as the 55 kb region downstream from

the *DHFR* gene in the hamster (Dijkwel *et al*, 2002) and the mouse β -globin region (Aladjem *et al*, 2002).

Our results also show that the different organization of ORIs in *S. pombe* correlates with their activity and dynamics of activation during S phase. A+T island ORIs fire early during S phase (Figure 6, *AT 1090*; Segurado *et al*, 2002; Heichinger *et al*, 2006) and their average efficiency is around 30% as estimated by two-dimensional electrophoresis (Segurado *et al*, 2002), single DNA molecule analysis (Patel *et al*, 2005) and the increase in the DNA content (Heichinger *et al*, 2006). Very few ORIs in *S. pombe* fire in >50% of the cells in any given S phase and this represents another significant difference relative to *S. cerevisiae*, in which efficiencies much higher than 50% are not unusual (Newlon *et al*, 1993; Yamashita *et al*, 1997). A possible explanation for this discrepancy is that the large amount of low-efficiency ORIs compete with higher affinity ORIs to bind Orc4p in *S. pombe* having as a consequence a reduction in the global average ORI efficiency.

The great diversity in the organization and efficiency of ORIs in *S. pombe* challenge what could be defined as a 'binary' model with a clear cut distinction of a locus as either being an ORI or not. Our results are more consistent with a model in which the propensity of different loci to initiate replication varies as a continuum across a wide range depending on their specific arrangement of A+T-rich elements and, probably, on other parameters such as nucleotide pools, chromatin organization and the transcriptional activity of the region. This scenario is consistent with the results from single DNA molecule analysis of 12 ORIs in *S. pombe* whose frequency of firing ranged from 15 to 53% (Patel *et al*, 2005). An even broader range of origin efficiency was found in the microarray-based genome analysis of Heichinger *et al* (2006). We would expect that poly-W ORIs, because of their lower affinity for ORC, would be more vulnerable to the influence of these parameters and would show a lower efficiency and a wider range of variability in their activity.

It has recently been proposed that the temporal pattern of activation is determined by differences in the time of recruitment of ORC to ORIs during the earlier M/G1 phase, such that ORIs that bind ORC earlier are activated more efficiently and earlier on than those that bind ORC later during the licensing period. These initial differences are paralleled by the subsequent assembly of the MCM complex and pre-initiation factors such as Cdc45p, Hsk1p and Dfp1p (Wu and Nurse, 2009). It is reasonable to assume that the binding affinity of ORC to DNA through the AT-hooks of Orc4p will depend on the specific architecture of each ORI. This assumption is supported by the observation that removal of one of the A+T-rich modules (region III) of *ars 2004* (Okuno *et al*, 1999; Takahashi *et al*, 2003) reduces and delays the binding of Orp1p, Mcm4p and Cdc45p and results in decreased initiation efficiency and late firing during S phase (Wu and Nurse, 2009). The only poly-W ORI in Figure 6 earlier identified by ChIP was 33T, which colocalizes with ORI 2142 and was classified as a late origin by Hayashi *et al* (2007). The remaining four origins map at late replicating domains except the 31W ORI that is located in chromosome III, most of which has been reported to be early replicating. These observations are consistent with the finding that poly-W ORIs replicate late, and suggest that, without excluding the possible influence of other parameters, the temporal programme of

ORI firing in *S. pombe* would ultimately depend on the number and A+T richness of the modules that make up different ORIs.

Materials and methods

Design, construction and cloning of synthetic ORIs

Five different fragments of 100 bp and an A+T content of 75% and another five fragments of 90% A+T were PCR amplified from the mitochondrial DNA of *Schizosaccharomyces octosporus* with specific primers 20 nucleotides long harbouring an additional 5' extension of 10 nucleotides including an Xba I restriction site. The five amplified monomers of each class were digested with Xba I, mixed together and ligated to generate two independent populations of oligomers with an A+T content of 75 and 90%. In the case of fragments with an A+T content of 83%, monomers 108 nucleotides long were generated by chemical synthesis allowing for a high degree of degeneracy in their internal sequence. A constant sequence 20 nucleotides long was included at their ends harbouring Xba I and EcoRV sites for PCR amplification and cloning into a plasmid vector. PCR-amplified fragments from each of the three sets were ligated independently and cloned into the plasmid vectors indicated below. Small deviation from exact multiples of monomer sizes in Figures 1 and 2 are due to differences in the amplified PCR fragments and to cloning constraints. Sequences of the 75, 83 and 90% A+T monomers are shown in Supplementary Figure 5. Sequences of artificial ORIs in Figure 1A and B are indicated in Supplementary Figures 1 and 2, respectively. Sequences of all the remaining PCR primers used, of genomic deletions and insertions, and particular cloning details for all the constructs described in this work are available on request.

ARS assays

All fragments to be tested for ARS activity were cloned in the 2.9 kb pBSK+ or the 5.3 kb pJK148 plasmids harbouring a 1.7 kb Hind III fragment including the *S. pombe ura4* gene under its own promoter. The ARS-positive control carried, in addition, an EcoRI 1.2 kb fragment including the *ars1 S. pombe* replication origin. *S. pombe ura4-d18* cells auxotrophs for uracil were transformed with 200 ng of plasmid after the lithium acetate protocol described by Moreno and Norbury (1997). Transformed cells were selected on plates of minimal medium at 32°C for 3–4 days. ARS activity was expressed as the percentage of transformant colonies relative to the number generated by the positive control carrying the *ars1* replication origin.

Targeting of ARS elements to chromosome III by homologous recombination

Sequences tested in Figure 2 were ligated to the 1.7 kb Hind III fragment including the *ura4* gene and the resulting constructs were cloned into the Eco RV site in the centre of a fragment between positions 108349 and 109320 in chromosome III such that there would be 0.5 kb on each side to facilitate their integration at this position in the genome by homologous recombination. A total of 200 to 500 ng of the resulting constructs were used to transform *S. pombe ura-d18* cells as described above. Correct integration into the targeted site was confirmed by PCR, Southern hybridization and DNA sequencing of the relevant genomic region in all transgenic strains.

References

- Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, Falaschi A (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science* **287**: 2023–2026
- Aladjem MI, Rodewald LW, Lin CM, Bowman S, Cimbara DM, Brody LL, Epner EM, Groudine M, Wahl GM (2002) Replication initiation patterns in the beta-globin loci of totipotent and differentiated murine cells: evidence for multiple initiation regions. *Mol Cell Biol* **22**: 442–452
- Altman AL, Fanning E (2001) The Chinese hamster dihydrofolate reductase replication origin beta is active at multiple ectopic

Genomic deletions across intergenic region 1090

Selected fragments across the AT 1090 intergenic region were replaced with the *ura4* gene by transforming *S. pombe ura-d18* cells with the Hind III 1.7 kb fragment including *ura4* described above flanked at both ends by regions about 200–400 bp homologous to the flanks of the regions to be deleted and *ura+* colonies were selected on plates of minimal medium. In a second step, the *ura4* gene was removed by transforming them with the same recombination cassettes without the Hind III 1.7 kb fragment between them. Transformed *ura-* colonies were selected on plates of minimal medium containing 225 mg/l of uracil and 1 g/l of 5-fluoroorotic acid at 32°C during 3–4 days.

Two-dimensional gel electrophoresis and cdc10-129 cell synchronization

DNA isolation and two-dimensional electrophoresis conditions were identical to those described by Segurado *et al* (2002). For the synchronization analysis (Figure 6), a 2.5 l culture of *S. pombe cdc10-129* temperature-sensitive cells was arrested in G1 at 36°C 4 h and then released at 25°C. Five samples of 500 ml (5×10^9 cells) were taken at the indicated times (Figure 6) and processed for 2D electrophoresis. See Supplementary Figure 4 for FACS analysis of the synchronous S phase.

Chromatin immunoprecipitation

ChIP analysis was performed as described by Pidoux *et al* (2004) with some modifications. Exponential *S. pombe* wt cells and cells carrying the ectopic 29W element in chromosome III were fixed with 1% formaldehyde for 30 min at 32°C. Cells were disrupted in Fast Prep (1 pulse at speed 4.5 for 40 s and 2 pulses at speed 5 for 40 s each). Cells were chilled on ice for 2–3 min between each disruption step. Cell extracts were sonicated on wet ice to shear chromatin to size of 500–1000 nucleotides using a Diagenode Bioruptor Sonicator (3 cycles of 10 min each with alternating pulses of 5 s on/off). Samples were incubated overnight at 4°C with 5 µg of a polyclonal antibody against the carboxy terminus of the Orc4 protein. Samples were purified with the GFX PCR DNA and Gel Band kit (GE Healthcare). ChIP and whole cell extract control samples were resuspended in 50 µl of sterile water before being used as template for Q-PCR analysis.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We are grateful to Elisa de Castro, Laura Marín and Mar Sánchez for help with 2D gel analysis, chromatin immunoprecipitation and ARS experiments, and also to Luis Quintales for genome bioinformatic analysis and to María Gómez for helpful comments on the manuscript. CC was supported by a postgraduate fellowship from the Spanish Ministerio de Ciencia e Innovación (MICINN) and this work was funded by grants BFU2008-01919BMC and Consolider-Ingenio CSD2007-00015 (from MICINN).

Conflict of interest

The authors declare that they have no conflict of interest.

- Chuang RY, Kelly TJ (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci USA* **96**: 2656–2661
- Clyne RK, Kelly TJ (1995) Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *EMBO J* **14**: 6348–6357
- Dai J, Chuang RY, Kelly TJ (2005) DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc Natl Acad Sci USA* **102**: 337–342
- DePamphilis ML (2005) Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle* **4**: 70–79
- Dershowitz A, Snyder M, Sbia M, Skurnick JH, Ong LY, Newlon CS (2007) Linear derivatives of *Saccharomyces cerevisiae* chromosome III can be maintained in the absence of ARS elements. *Mol Cell Biol* **27**: 4652–4663
- Dijkwel PA, Wang S, Hamlin JL (2002) Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies. *Mol Cell Biol* **22**: 3053–3065
- Dubey DD, Kim SM, Todorov IT, Huberman JA (1996) Large, complex modular structure of a fission yeast DNA replication origin. *Curr Biol* **6**: 467–473
- Dubey DD, Zhu J, Carlson DL, Sharma K, Huberman JA (1994) Three ARS elements contribute to the ura4 replication origin region in the fission yeast, *Schizosaccharomyces pombe*. *EMBO J* **13**: 3638–3647
- Feng W, Collingwood D, Boeck ME, Fox LA, Alvino GM, Fangman WL, Raghuraman MK, Brewer BJ (2006) Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* **8**: 148–155
- Gómez M, Antequera F (1999) Organization of DNA replication origins in the fission yeast genome. *EMBO J* **18**: 5683–5690
- Hayashi M, Katou Y, Itoh T, Tazumi M, Yamada Y, Takahashi T, Nakagawa T, Shirahige K, Masukata H (2007) Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *EMBO J* **26**: 1327–1339
- Heichinger C, Penkett CJ, Bahler J, Nurse P (2006) Genome-wide characterization of fission yeast DNA replication origins. *EMBO J* **25**: 5171–5179
- Kalejta RF, Li X, Mesner LD, Dijkwel PA, Lin HB, Hamlin JL (1998) Distal sequences, but not ori-beta/OBR-1, are essential for initiation of DNA replication in the Chinese hamster DHFR origin. *Mol Cell* **2**: 797–806
- Kim SM, Huberman JA (2001) Regulation of replication timing in fission yeast. *EMBO J* **20**: 6115–6126
- Kitsberg D, Selig S, Keshet I, Cedar H (1993) Replication structure of the human beta-globin gene domain. *Nature* **366**: 588–590
- Kong D, DePamphilis ML (2002) Site-specific ORC binding, pre-replication complex assembly and DNA synthesis at *Schizosaccharomyces pombe* replication origins. *EMBO J* **21**: 5567–5576
- Lee JK, Moon KY, Jiang Y, Hurwitz J (2001) The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc Natl Acad Sci USA* **98**: 13589–13594
- Maher JF, Nathans D (1996) Multivalent DNA-binding properties of the HMG-1 proteins. *Proc Natl Acad Sci USA* **93**: 6716–6720
- Marahrens Y, Stillman B (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**: 817–823
- Maundrell K, Hutchison A, Shall S (1988) Sequence analysis of ARS elements in fission yeast. *EMBO J* **7**: 2203–2209
- Mickle KL, Ramanathan S, Rosebrock A, Oliva A, Chaudari A, Yompakdee C, Scott D, Leatherwood J, Huberman JA (2007) Checkpoint independence of most DNA replication origins in fission yeast. *BMC Mol Biol* **8**: 112
- Moreno S, Norbury C (1997) Cloning cell cycle regulatory genes by transcomplementation in yeast. *Methods Enzymol* **283**: 44–59
- Newlon CS, Collins I, Dershowitz A, Deshpande AM, Greenfeder SA, Ong LY, Theis JF (1993) Analysis of replication origin function on chromosome III of *Saccharomyces cerevisiae*. *Cold Spring Harb Symp Quant Biol* **58**: 415–423
- Okuno Y, Okazaki T, Masukata H (1997) Identification of a predominant replication origin in fission yeast. *Nucleic Acids Res* **25**: 530–537
- Okuno Y, Satoh H, Sekiguchi M, Masukata H (1999) Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Mol Cell Biol* **19**: 6699–6709
- Patel PK, Arcangioli B, Baker SP, Bensimon A, Rhind N (2005) DNA replication origins fire stochastically in fission yeast. *Mol Biol Cell* **17**: 308–316
- Pidoux A, Mellone B, Allshire R (2004) Analysis of chromatin in fission yeast. *Methods* **33**: 252–259
- Reeves R, Beckerbauer L (2001) HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim Biophys Acta* **1519**: 13–29
- Santocanale C, Diffley JF (1998) A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. *Nature* **395**: 615–618
- Segurado M, Gómez M, Antequera F (2002) Increased recombination intermediates and homologous integration hot spots at DNA replication origins. *Mol Cell* **10**: 907–916
- Segurado M, de Luis A, Antequera F (2003) Genome-wide distribution of DNA replication origins at A+T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep* **4**: 1048–1053
- Takahashi T, Ohara E, Nishitani H, Masukata H (2003) Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast. *EMBO J* **22**: 964–974
- Theis JF, Newlon CS (1997) The ARS309 chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. *Proc Natl Acad Sci USA* **94**: 10786–10791
- Theis JF, Newlon CS (2001) Two compound replication origins in *Saccharomyces cerevisiae* contain redundant origin recognition complex binding sites. *Mol Cell Biol* **21**: 2790–2801
- Van Houten JV, Newlon CS (1990) Mutational analysis of the consensus sequence of a replication origin from yeast chromosome III. *Mol Cell Biol* **10**: 3917–3925
- Wu PY, Nurse P (2009) Establishing the program of origin firing during S phase in fission yeast. *Cell* **136**: 852–864
- Yamashita M, Hori Y, Shinomiya T, Obuse C, Tsurimoto T, Yoshikawa H, Shirahige K (1997) The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* **2**: 655–665