1

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
Campus Miguel de Unamuno
37007-Salamanca. Spain

Character count: **39569** (including spaces, excluding references and suppl. material)

Subject category: Genome stability and dynamics

Running title: Specification of replication origins in *S. pombe*

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

ABSTRACT

DNA replication origins (ORI) in *Schizosaccharomyces pombe* colocalize with A+T-rich regions and previous 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 adenines and thymines capable of initiating replication from ectopic chromosome positions. The search for elements with these characteristics across the genome has uncovered a previously 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.

Keywords: Cell cycle / DNA replication / Genome organization / Replication dynamics / Replication origin

INTRODUCTION

Eukaryotic replication origins (ORIs) are chromosomal regions where 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 base pairs 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 ARS (Autonomus Replication Sequence) 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; Gómez and Antequera, 1999; Okuno et al., 1997) 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 kb to 1 kb long with an A+T content ranging between 72% and 75% (Segurado et al., 2003). Recent studies have shown that the great 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 aminoacids 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 absence of the wild-type 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 on the genome to initiate replication. Nevertheless, this general model still leaves unanswered the question of how ORIs are specified in the *S. pombe* genome; i.e., 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 display 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 zones made up of several degenerate elements each of which has the potential to drive replication individually. These results dramatically reduce previous 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*

Previous 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 (Okuno et al., 1999; Clyne and Kelly, 1995; Dubey et al., 1996) 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 minimun requirements to specify active ORIs by asking whether custom-made synthetic DNA fragments covering a range of sizes and A+T content could mimmic 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 a *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 bp, 200 bp 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 bp 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). Also, and consistent with results in Figure 1A, none

of the 6 fragments approximately 100 bp long was active, probably due to their small size. To test

whether these fragments were active in the context of a larger plasmid, we repeated the ARS assay

after cloning them into the 5.3 kb pJK148 plasmid. None of the 6 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 due to

their small size, we self-ligated 3 different inactive 100 bp monomers (a, b and c in Figure 1B)

into dimers and trimers. Figure 1C shows that the three sets of oligomers made up of individually

inactive monomers reached an ARS activity comparable to 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 bp 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 3, 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 previously confirmed by two dimensional electrophoresis

that this region does not contain a replication origin (Gómez and Antequera, 1999 and 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 3 fragments of 75% A+T showed weak ORI activity (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). By contrast, all eight 83% and 90% A+T fragments tested were capable of initiating replication efficiently (Figure 2D, 2E). 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 upon 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 and 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 to 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.

1   To address this question, we chose the intergenic region encompassing A+T island 1090 (Figure

2   3A). This region is located between the *vma5* and *SPAPB2B4.04c* genes in chromosome I and

3   colocalizes with a strong ORI (Figure 3C, wt). The NcoI restriction fragment of 5.6 kb in wild

4   type cells tested by two dimensional electrophoresis encompassed the A+T island and the entire

5   intergenic region, making it impossible to distinguish whether independent elements could

6   contribute to ORI activity. As a first step in the functional dissection of this region, we assayed the

7   ARS activity of 5 non-overlapping fragments across the region (1, 5, 6, 7 and 8 in Figure 3A). We

8   found that fragment 5 (321 bp and 79% A+T), which spanned the A+T island, was active,

9   consistent with results in Figure 1 (Figure 3B). In addition, fragment 1 (1.1 kb) displayed an even

10   higher activity despite the fact that its global A+T content was 68%, which is very similar to the

11   70% intergenic average. Cleavage of this fragment into two halves generated fragment 2, which

12   was inactive, and fragment 4, which retained over 50% of the activity. Close sequence

13   examination revealed the presence of an uninterrupted stretch of 29 adenines and thymines (29W)

14   at the 5' end of this fragment. Deletion of this element generated fragment 3 and resulted in the

15   complete loss of ARS activity (Figure 3B). The 29W stretch also lacked ARS activity by itself

16   when cloned into the plasmid. Fragments 6, 7 and 8 were also inactive despite their relatively

17   large size.

18   To evaluate the relative contribution of fragments 5 and 29W to ORI activity, we deleted them

19   independently or in combination in their chromosomal context. Two dimensional electrophoresis

20   revealed that in both cases replication initiation was diminished but not abolished, indicating the

21   presence of other elements capable of sustaining residual activity (Figure 3C). Deletion of half of

22   the intergenic region encompassing fragments 5 and 1 (which included 29W) ($\Delta 5 + \Delta 1$) resulted in

23   a further reduction in activity which probably derived from cooperation between fragments 6, 7

24   and 8, each of which lacked ARS activity when assayed individually. Indeed, when these three

25   fragments were ligated into a plasmid to generate a unique fragment in the genomic configuration

1   (6-8) their ARS activity was comparable to that of *ars1*, although the colonies grew at a much

2   smaller rate comparable to that shown in Figure 4B (see later).

3   To test whether fragments 5 and 29W contained enough information to drive replication by

4   themselves in the genome, we integrated them independently at the ectopic location in

5   chromosome 3 (Figure 4B). Fragment 5 drove replication from this site (Figure 4A, 5/Chr III),

6   consistent with the observation that fragments of a comparable size and A+T content were also

7   active in this locus (Figure 2). Contrary to our expectations, however, we found that the 29W

8   element also generated an active, although less efficient, ORI upon integration in the same site

9   (Figure 4A, 29W/Chr III). To confirm the presence of the weak bubble arc relative to the passive

10   replication arc, we repeated this 2D gel in triplicate from cultures grown from independent

11   colonies with the same result in the three cases (data not shown). This unanticipated result raised

12   the question of how such a small fragment could specify a genomic ORI despite the fact of being

13   inactive when assayed in a plasmid. We speculated that 29W might recruit ORC through the AT-

14   hooks of the Orc4p subunit and that this binding could be stabilized by the surrounding context of

15   the intergenic region that had a higher A+T content than when cloned in the plasmid. To test this

16   possibility we measured the ARS activity of the 29W element cloned in the middle of a 1 kb

17   fragment encompassing the insertion site region (ISR, Figure 4B). This region was unable to

18   initiate replication in the genome (Figure 2B, wt) and did not display ARS activity in a plasmid

19   either (Figure 4C, ISR). However, insertion of 29W into this inactive fragment (ISR+29W)

20   generated a similar number of colonies in the ARS assay as the positive *ars1* control, although the

21   colonies grew at a much slower rate (Figure 4C), which is consistent with the different efficiency

22   of *ars1* and 29W to initiate replication upon integration in chromosome III (compare *ars1* in

23   Figure 2B with 29W/Chr III in Figure 4A). The inability of 29W to sustain ARS activity by itself

24   could explain why short elements on this kind had not been identified previously in ARS assays.

25   These results supported the hypothesis that 29W could recruit ORC to DNA and the complex

1   could be further stabilized by binding to flanking regions and implied that the minimum size of a

2   DNA sequence to specify an active ORI could be even shorter than the 100-200 bp found in

3   Figures 1 and 2.

4   We tested this possibility directly by chromatin immunoprecipitation (ChIP) with an antibody

5   against the Orc4 subunit of ORC. Results of four immunoprecipitation assays from independent

6   cultures of the *S. pombe* strain carrying the ectopic 29W insertion and another 4 of the wild type

7   strain as a control showed a 23% increase in the amount of immunoprecipitated Orc4 in the strain

8   harboring the ectopic 29W element relative to the wild type control (Supplementary Figure 3).

9   Although small, this increase relative to background is consistent with the ORC binding profile of

10   many poly W ORIs found in a previous a ChIP-on-chip genome-wide analysis using antibodies

11   against Orc4, Orc1 and Mcm6 (Hayashi et al., 2007).

12

13   **Poly W stretches colocalize with late-firing genomic ORIs**

14   There are 460 non-overlapping stretches of 29 contiguous adenines and thymines in the *S. pombe*

15   genome, approximately half of which map to intergenic regions devoid of A+T-rich islands and do

16   not colocalize with sites of ORC binding (Hayashi et al., 2007). Considering that perhaps even

17   shorter stretches could drive replication, an immediate implication of these results is that the

18   number of active ORIs in the genome could be much higher than current estimates (Segurado et

19   al., 2003; Heichinger et al., 2006; Hayashi et al., 2007; Mickle et al., 2007). To test this

20   possibility, we selected 12 genomic regions that contained uninterrupted stretches of homogenous

21   Ts (or As) or a combination of both bases (W) between 21 and 41 bp long and analyzed them by

22   two dimensional electrophoresis. All regions tested were at least 12 kb distant from the nearest

23   predicted ORI (Segurado et al., 2003; Hayashi et al. 2007). Results depicted in Figure 5 show that

24   8 out of 12 regions displayed bubble arcs of an intensity comparable to that generated by the 29W

25   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* (Heichinger 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. Upon release at the permissive temperature, we monitored the firing of the strong origin *AT1090* and 5 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 minutes, and 33T and 28W at 90 minutes into S phase, while bubble intermediates were already detectable in the 30 minutes sample in origin *AT1090*. As discussed below, the low efficiency and late firing of poly W ORIs is probably due to a low affinity for ORC, as indicated by the fact that only 3 out of the 8 positives in Figure 5 (21T, 33T and 34T) were identified as sites of ORC binding by chromatin immunoprecipitation analysis (Hayashi et al., 2007).

1 **DISCUSSION**

2 **ORI specification in *S. pombe***

3 An essential and as yet unanswered question in the eukaryotic DNA replication field is the

4 identification of the elements that specify active ORIs in the genome. In *S. pombe*, plasmid

5 replication assays have established that ARS fragments range between 0.5 kb and over 3 kb

6 (Maundrell, 1988; Clyne and Kelly, 1995; Dubey et al. 1996; Okuno et al. 1997; Dai et al., 2005),

7 and it is assumed that similar requirements would apply to genomic ORIs. In fact, size is often

8 cited as one of the major differences between ORIs in *S. pombe* and *S. cerevisiae*, where ARS

9 elements can be shorter than 200 bp (Marahrens and Stillman, 1992; Van Houten and Newlon,

10 1990). Contrary to this assumption, our results show that poly W tracks as short as 30 bp long are

11 sufficient to specify replication initiation in the context of intergenic regions (Figure 4) and to

12 reliably predict the localization of active ORIs in the *S. pombe* genome (Figure 5). An immediate

13 consequence of these results is that current estimates of ORIs in the genome should be expanded,

14 and at the same time this raises the question of why this class of ORIs has not been detected in

15 previous studies.

16 Segurado et al. (2003) found that 90% of A+T-rich islands colocalized with relatively strong ORIs

17 (see, for example, A+T island 1090 in Figure 3). However, because that study used a lower size

18 limit of 500 bp, based on previous ARS data in the literature, many of the ORIs specified by short

19 poly W tracks were probably excluded from that prediction. Hayashi et al. (2007) identified 460

20 sites of ORC binding by chromatin immunoprecipitation followed by microarray hybridization,

21 many of which colocalized with A+T-rich islands. However, 5 out of the 8 weak ORIs in Figure 5

22 (31W, 34W, 28W, 39T and 41W) were not detected in that study, indicating that a large

23 proportion of this class of ORIs either bind ORC very weakly or do so only in a small proportion

24 of the cells, such that their identification falls below the threshold of reliable detection by

25 chromatin immunoprecipitation. 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 contribute to explaining why some intergenic regions contain active, although low-efficiency, ORIs while 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 previously analyzed (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*. While poly W tracts are probably close to the minimum sequence information capable of binding ORC through the Orc4 subunit (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*, where 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

1   at various sites across *ars 2004* and *ars 3001* (Takahashi et al., 2003; Kong and DePamphilis,

2   2002) and that *in vitro* footprinting analysis has shown that ORC protects a wide region in the

3   *ars1* origin (Lee et al., 2001). It is likely that many ORIs associated with A+T-rich islands (such

4   as origin *AT1090*, *ars 2004*, *ars 3001* and *ars1*) have an extended organization and this could be

5   the reason why the average size of intergenic regions where this class of ORIs lie is twice as long

6   (2 kb) as those devoid of them (data not shown).

7   Although at a different scale, the functional organization of these regions is similar to that of the

8   55 kb initiation zone downstream from the hamster *DHFR* gene (Dijkwel et al., 2002). Both

9   regions are made up of individual elements whose deletion does not abolish ORI activity (Figure 3

10   and Kalejta et al., 1998) and which have the potential to initiate replication from ectopic positions

11   in the genome (Figure 4 and Altman and Fanning, 2001). The picture that emerges from these

12   observations is that ORI organization in *S. pombe* is comparable to that found in mammalian

13   genomes, where short *Lamin B2* (Abdurashidova et al., 2000) and *β-globin* human ORIs (Kitsberg

14   et al., 1993) coexist with larger initiation zones such as the 55 kb region downstream from the

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

16   2002).

17   Our results also show that the different organization of ORIs in *S. pombe* correlates with their

18   activity and dynamics of activation during S phase. A+T island ORIs fire early during S phase

19   (Figure 6 AT 1090, and Segurado et al., 2002; Heichinger et al., 2006) and their average efficiency

20   is around 30% as estimated by two dimensional electrophoresis (Segurado et al., 2002), single

21   DNA molecule analysis (Patel et al., 2005) and the increase in the DNA content (Heichinger et al.,

22   2006). Very few ORIs in *S. pombe* fire in more than 50% of the cells in any given S phase and this

23   represents another significant difference relative to *S. cerevisiae*, where efficiencies much higher

24   than 50% are not unusual (Newlon et al., 1993; Yamashita et al., 1997). A possible explanation for

25   this discrepancy is that the large amount of low efficiency ORIs compete with higher affinity ORIs

1 to bind Orc4p in *S. pombe* having as a consequence a reduction in the global average ORI

2 efficiency.

3 The great diversity in the organization and efficiency of ORIs in *S. pombe* challenge what could be

4 defined as a "binary" model with a clear cut distinction of a locus as either being an ORI or not.

5 Our results are more consistent with a model where the propensity of different loci to initiate

6 replication varies as a continuum across a wide range depending on their specific arrangement of

7 A+T-rich elements and, probably, on other parameters such as nucleotide pools, chromatin

8 organization and the transcriptional activity of the region. This scenario is consistent with results

9 from single DNA molecule analysis of 12 ORIs in *S. pombe* whose frequency of firing ranged

10 from 15% to 53% (Patel et al. 2005). An even broader range of origin efficiency was found in the

11 microarray-based genome analysis of Heinchinger et al. (2006). We would expect that poly W

12 ORIs, because of their lower affinity for ORC, would be more vulnerable to the influence of these

13 parameters and would show a lower probability and a wider range of variability in their activity.

14 It has recently been proposed that the temporal pattern of activation is determined by differences

15 in the time of recruitment of ORC to ORIs during the previous M/G1 phase, such that ORIs that

16 bind ORC earlier are activated more efficiently and earlier on than those that bind ORC later

17 during the licensing period. These initial differences are paralleled by the subsequent assembly of

18 the MCM complex and pre-initiation factors such as Cdc45p, Hsk1p and Dfp1p (Wu and Nurse,

19 2009). It is reasonable to assume that the binding affinity of ORC to DNA through the AT-hooks

20 of Orc4p will depend on the specific architecture of each ORI. This assumption is supported by

21 the observation that removal of one of the A+T-rich modules (region III) of *ars 2004* (Okuno et

22 al., 1999; Takahashi et al., 2003) reduces and delays the binding of Orp1p, Mcm4p and Cdc45p

23 and results in decreased initiation efficiency and late firing during S phase (Wu and Nurse, 2009).

24 These observations are consistent with the finding that poly W ORIs replicate late, and suggest

25 that, without excluding the possible influence of other parameters, the temporal programme of

1   ORI firing in *S. pombe* would ultimately depend on the number and A+T richness of the modules

2   that make up different ORIs.

3

4

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 5 fragments of 90% A+T were PCR amplified from the mitochondrial DNA of *Schizosaccharomyces octosporus* with specific primers 20 nucleotides long harboring 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 harboring 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 constrains. Sequences of the 75%, 83% and 90% A+T monomers are shown in Supplementary Figure 5. Sequences of artificial ORIs in Figures 1A and 1B 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 upon 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 harboring 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

1 were transformed with 200 ng of plasmid following the lithium acetate protocol described by

2 Moreno and Norbury (1997). Transformed cells were selected on plates of minimal medium at

3 32 ºC for 3-4 days. ARS activity was expressed as the percentage of transformant colonies

4 relative to the number generated by the positive control carrying the *ars1* replication origin.

5

6 **Targeting of ARS elements to chromosome III by homologous recombination**

7 Sequences tested in Figure 2 were ligated to the 1.7 kb Hind III fragment including the *ura4*

8 gene and the resulting constructs were cloned into the Eco RV site in the centre of a fragment

9 between positions 108349 and 109320 in chromosome III such that there would be 0.5 kb on

10 each side to facilitate their integration at this position in the genome by homologous

11 recombination. Two to five hundred ng of the resulting constructs were used to transform *S.*

12 *pombe ura-d18* cells as described above. Correct integration into the targeted site was confirmed

13 by PCR, Southern hybridization and DNA sequencing of the relevant genomic region in all

14 transgenic strains.

15

16 **Genomic deletions across intergenic region 1090**

17 Selected fragments across the *AT 1090* intergenic region were replaced with the *ura4* gene by

18 transforming *S. pombe ura-d18* cells with the Hind III 1.7 kb fragment including *ura4* described

19 above flanked at both ends by regions about 200-400 bp homologous to the flanks of the regions

20 to be deleted and *ura+* colonies were selected on plates of minimal medium. In a second step,

21 the *ura4* gene was removed by transforming them with the same recombination cassettes without

22 the Hind III 1.7 kb fragment between them. Transformed *ura-* colonies were selected on plates of

23 minimal medium containing 1 g/l of 5-Fluoroorotic acid (5-FOA) at 32 ºC during 3-4 days.

24

25

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 liter culture of *S. pombe cdc10-129* temperature sensitive cells was arrested in G1 at 36 ºC 4 hours and then released at 25 ºC. Five samples of 500 ml (5 x 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)

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% formaldehide for 30 minutes at 32º C. Cells were disrupted in Fast Prep (1 pulse at speed 4.5 for 40 seconds and 2 pulses at speed 5 for 40 seconds each). Cells were chilled on ice for 2-3 minutes 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 minutes each with alternating pulses of 5 seconds on/off). Samples were incubated overnight at 4º C with 5 micrograms 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 we resuspended in 50 microliters of sterile water before being used as template for Q-PCR analysis.

3

4

1 **FIGURES LEGENDS**

2 **Figure 1. Plasmid replication by artificial ARS elements**

3 **A)** Size of ARS fragments of 90% (red), 83% (yellow) and 75% (green) A+T content capable of

4 replicating the pBSK+ plasmid harbouring the *ura4* gene as selectable marker. Sequences of the 27

5 fragments are shown in Supplementary Figure 1.

6 **B)** ARS activity of 15 fragments from 102 bp to 472 bp long with an 83% A+T content.

7 Efficiencies are relative to the positive control *ars1* (1.2 kb) (black bar). Plasmid pBSK+ with the

8 *ura4* gene without any ARS element was used as a negative control (C-). Results are average of two

9 independent experiments. The sequences of all the fragments are shown in Supplementary Figure 2.

10 **C)** Inactive monomers (m) a, b and c from panel B display increasing ARS activity when self-

11 ligated as dimers (d) or trimers (t). *Ars1* and C-, positive and negative controls as in B.

12

13 **Figure 2. Artificial ORIs initiate replication in the *S. pombe* genome**

14 **A)** Diagram of the insertion site (black arrow) of artificial ARS elements in chromosome III. White

15 arrowheads indicate nearest predicted ORIs. rDNA represents ribosomal RNA locus at the left end

16 of the chromosome. The expanded region below shows genes surrounding the insertion site.

17 Bracket indicates the genomic region centered on the insertion site tested by two dimensional

18 electrophoresis.

19 **B)** Replication of the 3.8 kb Xba I restriction fragment in wild-type cells (wt), and after integration

20 of the *ura4* marker or *ars1*. The presence of a bubble replication arc in *ars1*, which indicates the

21 presence of an active ORI, is indicated by a black arrow. Bubble arcs in all the other 2D gels in this

22 work are also indicated by black arrows.

23 **C, D, E)** Replication pattern upon insertion of the indicated fragments from Figure 1A with a 75%,

24 83% and 90% A+T content, respectively.

25

1

2

3   **Figure 3. Functional dissection of the intergenic region encompassing A+T-rich island 1090**

4   **A)** A+T content in 300 bp windows (step 50 bp) across the region between the divergently

5   transcribed genes *SPAPB2B4.04c* (red) and *vma5* (blue) genes. A+T island 1090 is indicated by an

6   arrow. Vertical lines represent restriction sites for NcoI flanking the 5.6 kb region tested by two

7   dimensional electrophoresis. The genomic intergenic average is 70% A+T (red line). Fragments

8   tested for ARS activity in panel B are indicated by numbered colour bars and the 29 bp stretch of

9   adenines and thymines by a yellow triangle (29W).

10 **B)** ARS activity of fragments in panel A relative to *ars1* (black bar)

11 **C)** Two dimensional electrophoresis of the region indicated in panel A in exponential wild-type

12 cells (wt) and in cells deleted for region 5 (Δ5), 29W (Δ29W), both simultaneously (Δ5+Δ29W),

13 and regions 5 and 1 combined (Δ5+Δ1).

14

15 **Figure 4. ORI specification by individual modules at an ectopic location**

16 **A)** Two-dimensional gel analysis of fragments 5 and 29W integrated at the insertion site (IS, black

17 arrow) in chromosome III shown in panel B.

18 **B)** A 1 kb region encompassing the insertion site (ISR 1 kb), the 29W element alone (29W), and the

19 29W element integrated into the ISR 1 kb fragment (ISR+29W) were cloned into the pBSK+/*ura4*

20 vector and tested for ARS activity. The histogram shows that neither of the two individual

21 fragments showed any ARS activity but the composite fragment yielded the same number of

22 colonies as the positive control *ars1* (black bar).

23 **C)** Colony growth in plates of minimal medium after 4 days at 32 ºC. Cells transformed with the

24 plasmid carrying the ISR+29W fragment grew at a much lower rate than the *ars1* control plasmid.

25 Neither fragment generated any transformants when assayed independently (top panels).

1

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 analyzed 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 ºC for 4 hours. 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. 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 ºC was also included in the analysis. FACS analysis of the progress across S phase is shown in Supplementary Figure 4.

REFERENCES

Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M. and Falaschi, A. (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science* **287**: 2023-2026.

Aladjem, M.I., Rodewald, L.W., Lin, C.M., Bowman, S., Cimbora, D.M., Brody, L.L., Epner, E.M., Groudine, M. and Wahl, G.M. (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, A.L. and Fanning, E. (2001) The Chinese hamster dihydrofolate reductase replication origin beta is active at multiple ectopic chromosomal locations and requires specific DNA sequence elements for activity. *Mol Cell Biol* **21**: 1098-1110.

Bell, S.P. and Dutta, A. (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* **71**: 333-374.

Breier, A.M., Chatterji, S. and Cozzarelli, N.R. (2004) Prediction of *Saccharomyces cerevisiae* replication origins. *Genome Biol* **5**: R22.

Chuang, R.Y., Chretien, L., Dai, J. and Kelly, T.J. (2002) Purification and characterization of the *Schizosaccharomyces pombe* origin recognition complex: interaction with origin DNA and Cdc18 protein. *J Biol Chem* **277**: 16920-16927.

Chuang, R.Y. and Kelly, T.J. (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci U S A* **96**: 2656-2661.

Clyne, R.K. and Kelly, T.J. (1995) Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *EMBO J* **14**: 6348-6357.

Dai, J., Chuang, R.Y. and Kelly, T.J. (2005) DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc Natl Acad Sci U S A* **102**: 337-342

1   DePamphilis, M.L. (2005) Cell cycle dependent regulation of the origin recognition complex. *Cell*

2   *Cycle* **4**: 70-79

3   Dershowitz, A., Snyder, M., Sbia, M., Skurnick, J.H., Ong, L.Y. and Newlon, C.S. (2007) Linear

4   derivatives of *Saccharomyces cerevisiae* chromosome III can be maintained in the absence of

5   ARS elements. *Mol Cell Biol* **27**: 4652-4663.

6   Dijkwel, P.A., Wang, S. and Hamlin, J.L. (2002) Initiation sites are distributed at frequent intervals

7   in the Chinese hamster dihydrofolate reductase origin of replication but are used with very

8   different efficiencies. *Mol Cell Biol* **22**: 3053-3065.

9   Dubey, D.D., Kim, S.M., Todorov, I.T. and Huberman, J.A. (1996) Large, complex modular

10   structure of a fission yeast DNA replication origin. *Curr Biol* **6**: 467-473.

11 Dubey, D.D., Zhu, J., Carlson, D.L., Sharma, K. and Huberman, J.A. (1994) Three ARS elements

12   contribute to the ura4 replication origin region in the fission yeast, Schizosaccharomyces

13   pombe. *EMBO J* **13**: 3638-3647.

14 Feng, W., Collingwood, D., Boeck, M.E., Fox, L.A., Alvino, G.M., Fangman, W.L., Raghuraman,

15   M.K. and Brewer, B.J. (2006) Genomic mapping of single-stranded DNA in hydroxyurea-

16   challenged yeasts identifies origins of replication. *Nat Cell Biol* **8**: 148-155.

17 Gómez, M. and Antequera, F. (1999) Organization of DNA replication origins in the fission yeast

18   genome. *EMBO J* **18**: 5683-5690.

19 Hayashi, M., Katou, Y., Itoh, T., Tazumi, M., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige,

20   K. and Masukata, H. (2007) Genome-wide localization of pre-RC sites and identification of

21   replication origins in fission yeast. *EMBO J* **26**: 1327-1339.

22 Heichinger, C., Penkett, C.J., Bahler, J. and Nurse, P. (2006) Genome-wide characterization of

23   fission yeast DNA replication origins. *EMBO J* **25**: 5171-5179.

1 Kalejta, R.F., Li, X., Mesner, L.D., Dijkwel, P.A., Lin, H.B. and Hamlin, J.L. (1998) Distal

2   sequences, but not ori-beta/OBR-1, are essential for initiation of DNA replication in the

3   Chinese hamster DHFR origin. *Mol Cell* **2**: 797-806.

4 Kim, S.M. and Huberman, J.A. (2001) Regulation of replication timing in fission yeast. *EMBO J*

5   **20**: 6115-6126.

6 Kitsberg, D., Selig, S., Keshet, I. and Cedar, H. (1993) Replication structure of the human beta-

7   globin gene domain. *Nature* **366**: 588-590.

8 Kong, D. and DePamphilis, M.L. (2002) Site-specific ORC binding, pre-replication complex

9   assembly and DNA synthesis at *Schizosaccharomyces pombe* replication origins. *EMBO J* **21**:

10   5567-5576.

11 Lee, J.K., Moon, K.Y., Jiang, Y. and Hurwitz, J. (2001) The *Schizosaccharomyces pombe* origin

12   recognition complex interacts with multiple AT-rich regions of the replication origin DNA by

13   means of the AT-hook domains of the spOrc4 protein. *Proc Natl Acad Sci U S A* **98**: 13589-

14   13594.

15 Maher, J.F. and Nathans, D. (1996) Multivalent DNA-binding properties of the HMG-1 proteins.

16   *Proc Natl Acad Sci U S A* **93**: 6716-6720.

17 Marahrens, Y. and Stillman, B. (1992) A yeast chromosomal origin of DNA replication defined by

18   multiple functional elements. *Science* **255**: 817-823.

19 Maundrell, K., Hutchison, A. and Shall, S. (1988) Sequence analysis of ARS elements in fission

20   yeast. *EMBO J* **7**: 2203-2209.

21 Mickle, K.L., Ramanathan, S., Rosebrock, A., Oliva, A., Chaudari, A., Yompakdee, C., Scott, D.,

22   Leatherwood, J. and Huberman, J.A. (2007) Checkpoint independence of most DNA

23   replication origins in fission yeast. *BMC Mol Biol* **8**: 112.

24 Moreno, S. and Norbury, C. (1997) Cloning cell cycle regulatory genes by transcomplementation in

25   yeast. *Methods Enzymol* **283**: 44-59.

1 Newlon, C.S., Collins, I., Dershowitz, A., Deshpande, A.M., Greenfeder, S.A., Ong, L.Y. and

2   Theis, J.F. (1993) Analysis of replication origin function on chromosome III of *Saccharomyces*

3   *cerevisiae*. *Cold Spring Harb Symp Quant Biol* **58**: 415-423.

4 Okuno, Y., Okazaki, T. and Masukata, H. (1997) Identification of a predominant replication origin

5   in fission yeast. *Nucleic Acids Res* **25**: 530-537.

6 Okuno, Y., Satoh, H., Sekiguchi, M. and Masukata, H. (1999) Clustered adenine/thymine stretches

7   are essential for function of a fission yeast replication origin. *Mol Cell Biol* **19**: 6699-6709.

8 Patel, P.K., Arcangioli, B., Baker, S.P., Bensimon, A. and Rhind, N. (2005) DNA Replication

9   Origins Fire Stochastically in Fission Yeast. *Mol Biol Cell* **17**: 308-316.

10 Pidoux, A., Mellone, B. and Allshire, R. (2004) Analysis of chromatin in fission yeast. *Methods* **33**:

11   252-259

12 Reeves, R. and Beckerbauer, L. (2001) HMGI/Y proteins: flexible regulators of transcription and

13   chromatin structure. *Biochim Biophys Acta* **1519**: 13-29.

14 Santocanale, C. and Diffley, J.F. (1998) A Mec1- and Rad53-dependent checkpoint controls late-

15   firing origins of DNA replication. *Nature* **395**: 615-618.

16 Segurado, M., Gómez, M. and Antequera, F. (2002) Increased recombination intermediates and

17   homologous integration hot spots at DNA replication origins. *Mol Cell* **10**: 907-916

18 Segurado, M., de Luis, A. and Antequera, F. (2003) Genome-wide distribution of DNA replication

19   origins at A+T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep* **4**: 1048-1053.

20 Takahashi, T., Ohara, E., Nishitani, H. and Masukata, H. (2003) Multiple ORC-binding sites are

21   required for efficient MCM loading and origin firing in fission yeast. *EMBO J* **22**: 964-974.

22 Theis, J.F. and Newlon, C.S. (1997) The ARS309 chromosomal replicator of *Saccharomyces*

23   *cerevisiae* depends on an exceptional ARS consensus sequence. *Proc Natl Acad Sci U S A* **94**:

24   10786-10791.

1 Theis, J.F. and Newlon, C.S. (2001) Two compound replication origins in *Saccharomyces*

2   *cerevisiae* contain redundant origin recognition complex binding sites. *Mol Cell Biol* **21**: 2790-

3   2801.

4 Van Houten, J.V. and Newlon, C.S. (1990) Mutational analysis of the consensus sequence of a

5   replication origin from yeast chromosome III. *Mol Cell Biol* **10**: 3917-3925.

6 Wu, P.Y. and Nurse, P. (2009) Establishing the program of origin firing during S phase in fission

7   Yeast. *Cell* **136**: 852-864.

8 Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H. and Shirahige, K.

9   (1997) The efficiency and timing of initiation of replication of multiple replicons of

10   *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* **2**: 655-665.

11
12

13

14

15