A multiprotein complex necessary for both transcription and DNA replication at the β-globin locus

Subhradip Karmakar1, Milind C Mahajan1,2,*, Vincent Schulz3, Gokul Boyapati1 and Sherman M Weissman1,*

1Department of Genetics, The Anlyan Center, Yale University School of Medicine, New Haven, CT, USA, 2Department of Genetics, Yale Center for Genome Analysis, Yale University, Orange, CT, USA and 3Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

DNA replication, repair, transcription and chromatin structure are intricately associated nuclear processes, but the molecular links between these events are often obscure. In this study, we have surveyed the protein complexes that bind at β-globin locus control region, and purified and characterized the function of one such multiprotein complex from human erythroleukemic K562 cells. We further validated the existence of this complex in human CD34+ cell-derived normal erythroid cells. This complex contains ILF2/ILF3 transcription factors, p300 acetyltransferase and proteins associated with DNA replication, transcription and repair. RNAi knockdown of ILF2, a DNA-binding component of this complex, abrogates the recruitment of the complex to its cognate DNA sequence and inhibits transcription, histone acetylation and usage of the origin of DNA replication at the β-globin locus. These results imply a direct link between mammalian DNA replication, transcription and histone acetylation mediated by a single multiprotein complex.

The EMBO Journal (2010) 29, 3260–3271. doi:10.1038/emboj.2010.204; Published online 31 August 2010

Subject Categories: chromatin & transcription; genome stability & dynamics

Keywords: multiprotein complex; purification; replication; transcription

Introduction

Interaction between nuclear processes such as modifications of chromatin structure, DNA replication, repair and transcription have been described in several eukaryotic organisms (Gottipati and Helleday, 2009; Rampakakis et al., 2009). Maintenance of genome integrity during transcription and DNA replication is accomplished by repair mechanisms such as transcription-coupled repair (Hanawalt and Spivak, 2008), replication-associated homologous recombination and translesion synthesis (Bridges, 2005; Lehmann, 2005; Hanawalt and Spivak, 2008).

The formation of pre-replication complexes at initiation of replication (IR) sites is initiated by the origin recognizing complex (ORC)-mediated recruitment of cdc6, cdt1 and the MCM2–7 complex during G1 phase of the cell cycle and the origins are licensed to initiate a single round of DNA synthesis per cell cycle. During G1/S transition, CDC7 kinase and cyclins E/A recruit additional components, including CDC45, GINS and replicative DNA polymerase to form the pre-initiation complexes (Rampakakis et al., 2009).

Sequence-specific loading of replication proteins and firing of replication origins occurs in bacteria and yeast as well as animal viruses (Kohzaki and Murakami, 2005). In contrast, ORC from metazoan cells shows very little sequence specificity other than some preference for AT-rich sequences, so that mechanisms must exist to account for the non-random distribution of origins of replication. In several animal virus infections, transcription factors have been shown to recruit the host-cell ORC to specific sites (Guo et al., 1996; Ito et al., 1996; Murakami et al., 2007). In some metazoans, transcription is known to influence DNA replication (Danis et al., 2004; Xie and Orr-Weaver, 2008). In mammalian cells, actively transcribing genes are often replicated early in S phase (Dimitrova, 2006; Falkenberg et al., 2007). Many mammalian IR sites are AT-rich (Gilbert, 2001) or contain a region of AT-rich sequence, but the mechanisms leading to initiation of DNA replication from specific loci are not fully understood.

The β-globin locus contains five globin genes: an embryonic gene (ε), two foetal globin genes (γG and γA) and two adult-globin genes (δ and β), and a pseudogene (εβ), with the order of the genes corresponding to their time of expression during development. The expression of these genes is strongly dependent on a locus control region (LCR) 50 kb upstream of the β-globin gene. The LCR, in turn, contains four erythroid-specific DNase hypersensitive sites (HS1–4) that include evolutionarily conserved sequences (Figure 1A). The fifth DNase HS5 occurs in erythroid as well as several other non-erythroid haematopoietic cell systems. The β-globin locus harbours a strong IR site for DNA replication between the δ- and β-globin genes that has been used as a model system for studying mammalian DNA replication (Figure 1A) (Aladjem, 2004). Two or more independent modules exist within this β-globin IR (Wang et al., 2004). In addition, initiation of DNA replication has been described at the 3′ enhancer of the β-globin locus and at γ-globin genes (Kamath and Leffak, 2001; Aladjem et al., 2002; Buzzin et al., 2005). In the chicken globin locus, the HS4 sequence of the LCR is also reported to initiate DNA replication (Prioireau et al., 2003).

The globin cluster is embedded in a complex of olfactory genes that are non-functional in non-erythroid lineages.
Selective activation of the transcription of the globin cluster in erythroid cells involves participation of trans-factors that are thought to function not only at the globin promoters, but also at the LCR. The LCR was shown to be a binding region for several large protein complexes (Mahajan et al., 2007) and orchestrates molecular events to form an active chromatin hub (Zhou et al., 2006; Kooren et al., 2007) and transcriptional factory (Mitchell and Fraser, 2008).

The presence of intergenic deletions adjacent to globin genes of β-thalassemia patients provides genetic evidence for the presence of the globin IR. The Lepore deletion that removes sequences between the β- and δ-globin genes that contain an IR region results in passive replication of the locus from an outside origin (Kitsberg et al., 1993). In Hispanic thalassemia in which the LCR and upstream sequences are deleted, the β-globin locus replicates late in S phase with replication originating outside of the globin locus (Forrester et al., 1990; Mechali, 2001). Interestingly, in Lepore patients, the direction of replication is from 5′ to 3′ relative to globin gene transcription, whereas it is in the opposite orientation in...
Hispanic thalassemia (Kitsberg et al., 1993; Mechali, 2001). The different directions of DNA replication in these two types of deletions suggest that in addition to the IR sequences, upstream sequences can act as origins of DNA replication for the β-globin locus. However, in some experimental systems, deletion of the LCR neither changed the IR site nor the timing of replication (Cimборa et al., 2000) and the choice of IRs may be changed indirectly, for example by delaying replication firing at a potential IR site, causing replication initiated from an external region to extend across the IR before firing occurs.

In our survey of sequence-specific protein complexes in erythroid cells, we noted that antibodies against an MCM protein disrupted a sequence-specific complex that bound to an oligonucleotide from the HS4 region of the β-globin LCR (data not shown). Deletion analyses have shown that HS4 functions in maintaining normal levels of globin mRNA production (Simon et al., 2001; Fedosyuk and Peterson, 2007) and that deletion of this site may affect the timing of replication at the β-globin locus (Simon et al., 2001), but the specifics of how HS4 functions and the function of the hypersensitive sites have not been worked out. We hypothesized that proteins recruited to the LCR might functionally co-operate with IRs in the initiation of transcription and replication. Therefore, we performed a systematic analysis for novel protein complexes from erythroid cells that bind to the globin locus HS4 region. For this purpose, we performed tiling electrophoretic mobility shift assay (EMSA) scans to isolate erythroid K562 cell line-specific DNA–protein complexes. We used a series of overlapping oligonucleotides (HS4–1 through HS4–10) tiling the core 350 bp HS4 region of the β-globin LCR and carried out EMSA assays with nuclear extracts from HeLa and K562 cells (Figure 1B). This tiling EMSA scan led to the identification of a K562-specific multiprotein complex binding to an AT-rich HS4–9 oligonucleotide. We have purified a multiprotein complex that is sequence specifically associated with this HS4–9 sequence of the HS4 and identified its protein components by mass spectrometry. This analysis uncovered a complex of at least 16 components. Follow-up experiments suggest that this DNA-associated replication and transcription complex (DARTT) affects replication, chromatin modification and transcription.

**Results**

Identification of a high-molecular weight protein complex associated with an AT-rich sequence in the HS4 region of the β-globin LCR

In order to identify the multiprotein complexes associated with the β-globin LCR HS1–4, we performed a tiling EMSA scan using a series of 40 overlapping 35-mer (average) double-stranded synthetic oligonucleotides, tiling four LCR DNase HS1–4 sites with 7 bp overlaps. These were used in EMSA with nuclear extracts prepared from HeLa and the human erythroleukemic cell line K562 to identify K562-specific DNA–protein interactions at the core HS1–4 sequences. In this study, we have focused on the identification and characterization of the K562-specific DNA–protein interactions formed on the core HS4 sequence (Figure 1B). In this tiling EMSA screen of the 350 bp HS4 sequence, we have used 10 oligonucleotides HS4–1 through HS4-10 that span the entire core HS4 region of the LCR. This EMSA screen of HS4 revealed three prominent K562-specific EMSA bands with HS4–5, HS4–7 and HS4–9 oligonucleotides. The native molecular weight of the protein(s) associated with these EMSA bands were estimated by Superose-6 column-based gel exclusion chromatography. When we passed the K562 nuclear extracts through the Superose-6 size exclusion column, the protein complex associated with HS4–9 oligonucleotide was eluted with an apparent molecular weight of >2 MDa (Figure 1C). To further test whether the large protein complex associated with this HS4–9 sequence is indeed sequence specific, we carried out competitive EMSA assays. In these assays, we tested for the competitive displacement of the radioactive 32P-labelled HS4–9-associated EMSA bands with molar excesses of non-radioactive (cold) HS4–9 oligonucleotide as well as several control oligonucleotides of the same base composition (Figure 1D). We found efficient competition with the non-radioactive HS4–9 oligonucleotide alone and not with other control oligonucleotides, indicating sequence specificity of the protein complex associated with HS4–9 oligonucleotide sequence fragment (Figure 1D).

Purification of the protein complex associated with the HS4–9 EMSA

Having established that the HS4–9 sequence-associated DNA-binding activity is sequence specific and involves a large protein complex (Figure 1), we undertook the biochemical purification of this protein complex from K562 nuclear extracts. Our purification strategy involved the enrichment of DNA-binding proteins from the K562 nuclear extracts using heparin-Sepharose chromatography followed by molecular sieving on a Superose-6 sizing column for isolation of large protein complexes. These in turn were fractionated on a DEAE-Sephacel column. The HS4–9-binding proteins eluted from the DEAE-Sephacel ion-exchange column were purified using HS4–9 oligo-affinity chromatography (Figure 2A). Extensive purification was achieved after the final HS4–9 DNA oligonucleotide-affinity column step. Active EMSA fractions eluted from the DNA-affinity column were analysed by SDS–PAGE. Intensities of various bands as seen after staining the SDS gel suggested that the proteins were present in near stoichiometric ratios (Figure 2B).

MS/MS analysis of the trypsin-digested SDS–PAGE fragments revealed that the complex contained DARTT components (Figure 2B and C). Western blot analysis of the chromatographic fractions showed co-elution of the DARTT components during the purification steps (Supplementary Figure 1A–D). Among the several protein identified, we validated the presence of ILF2, ILF3, MCM5, p300 and RAD50 in the DARTT complex by co-immunoprecipitation and immunodepletion experiments (Figure 4). Tip49a and Tip49b are involved in transcription and DNA damage response, but were not studied in further detail, as we could not obtain suitable antibodies for co-immunoprecipitation and ChIP experiments. However, the presence of Tip49 as well as other components was confirmed by an alternative procedure using immunopurification of DARTT by p300 immunoaffinity chromatography (Supplementary Figure 6). In addition, we established stable K562 cell lines expressing C-terminal 6XHis-3XFLAG-tagged ILF2. We used the nuclear extracts from these cell lines for the purification of the HS4–9-associated protein complex. We obtained the same set of proteins in this complex, when we used a modified purification procedure in which we used successively heparin agarose
chromatography, superose FPLC, immunoaffinity purification with an anti-FLAG antibody, TALON column purification and finally the specific HS4-9 oligonucleotide-affinity column (Supplementary Table 1). Sucrose gradient fractionation of nuclear extracts also showed that the components of DARRT migrated as a single high-molecular weight peak (Supplementary Figure 9).

To establish that the DARRT complex is responsible for the EMSA bands with the HS4–9 oligonucleotide, we used gel supershift assays with antibodies against DARRT components (Figure 2D). Antibodies against p300 and RAD50 neutralized the EMSA bands, whereas antibodies against MCM5 and ILF2 showed supershifts. These data show in vitro association of DARRT with the HS4–9 oligonucleotide. ChIP-qPCR analysis using antibodies against ILF2, MCM5, RAD50 and p300 confirmed the in vivo association of the DARRT complex with the HS4 region (Figure 3; Supplementary Figure 2B). In addition to HS4, our ChIP qPCR results showed small but significant enrichment of ILF2, RAD50 and MCM5 at the γ-globin promoter and at the β-globin replication origin β-Rep-1 in K562 cells (Figure 3A). The p300 ChIP qPCR indicated strong binding to the HS4 region, and weaker but significant binding to other LCR and promoter sites, suggesting that it bound to multiple sites at the β-globin (Figure 3A).

To test whether DARRT is also associated with normal erythroid cells, we carried out ChIP analysis using antibodies against ILF2, ILF3, MCM5 and RAD50 in human CD34+ derived primary erythroid cells (Figure 3B). All antibodies showed a significant enrichment at the HS4 site. In addition, in the normal erythroid cells, ILF2, ILF3 and MCM5 showed substantial enrichment at the β-Rep-1 region of the globin IR. This suggests redundancy in the binding sites for DARRT in the normal β-globin complex and this redundancy may account for some of the variation in effects of local deletions on globin replication (Forrester et al., 1990; Kitsberg et al., 1993; Cimbora et al., 2000; Mechali, 2001).

Co-immunoprecipitation (co-IP) and immunodepletion experiments were carried out to test whether DARRT exists...
**ILF2 recruits DARRT to its target DNA sequence**

NF45/ILF2 associates with NF90/ILF3 as a heterodimer in the nucleus and regulates IL-2 gene transcription by binding to the antigen receptor response element/NF-AT DNA target sequence (Corthesy and Kao, 1994; Kao et al., 1994). To test whether these proteins are responsible for the recruitment of DARRT to its target DNA sequence on HS4, we established stable K562 lines expressing tetracycline-inducible ILF2 and MCM5 shRNA. Addition of doxycycline to these cells resulted in significant knockdown of ILF2 and MCM5 proteins (Figure 4D). Nuclear extracts from these ILF2 knockdown cells had significantly reduced DARRT DNA-binding activity to the HS4–9 oligo as revealed by EMSA (Figure 4E). siRNA-mediated double knockdown of ILF2 and ILF3 resulted in complete loss of DARRT EMSA-binding activity (Figure 4E), indicating that these DARRT components are necessary for binding to the HS4–9 DNA sequence. Furthermore, recombinant ILF2 was found to bind in vitro to HS4, supporting an important function for ILF2 in DARRT recruitment to HS4 (Supplementary Figure 2A). To investigate the effect of ILF2 knockdown on the recruitment of other DARRT components to the HS4 region, we performed ChIP qPCR for MCM5 and p300 in K562 cells expressing ILF2 shRNA. The results showed a significant reduction in the ChIP signal at HS4 and other sites on the locus (Figure 4F; Supplementary Figure 2B). Collectively, these results suggest that ILF2 has an important function in recruiting the DARRT components to the HS4 sites.

**DARRT components are essential for high levels of globin mRNA expression**

DARRT contains transcription factors such as ILF2 and ILF3 (Corthesy and Kao, 1994) and other components such as Tip49a, Tip49b and MCM proteins that have been implicated in the regulation of transcription (Snyder et al., 2005, 2009; Jha and Dutta, 2009) and chromatin remodeling (Dziak et al., 2003). Therefore, we investigated whether DARRT regulates β-globin gene transcription by analysing the effects of ILF2, RAD50, MCM5 or ORC2 depletion on β-globin mRNA expression in K562 cells. K562 cells actively transcribe embryonic (ε)- and foetal (γ)-globin genes, whereas the adult globin (β) remains essentially inactive. MCM5, ORC2 and ILF2 knockdown exhibited significant inhibition of ε- and γ-globin expression (Figure 5A). RT–PCR experiments in K562 cells (Supplementary Figure 7) after siRNA-mediated ILF2 or MCM5 knockdown showed no significant changes in the levels of GATA1, GATA2, NFE2, EKLF (KLFL1), TAL1, p18-MAF, PU.1, BACH1, LMO2, EPOR, YY1, LMO4 and RUNX1 mRNAs, suggesting that knockdown of ILF2 and MCM5 messages do not significantly alter the erythroid properties of K562 cells. Hence, the effect of ILF2 and
MCM5 knockdown may directly affect the ε- and γ-globin transcription in these cells (Supplementary Figure 7).

Knockdown of MCM5 and ILF2 proteins also resulted in a severe inhibition of K562 cell growth (Supplementary Figure 5B). Hence, we tested whether the decrease in ε- and γ-globin expression was a secondary effect of reduced cell proliferation. This was addressed when we slowed the growth of K562 cells by serum starvation (that is culturing in 0.5% FBS) to levels comparable with those seen with ILF2 or MCM5 depletion (Supplementary Figure 5A), we did not observe any significant change in the mRNA levels of the ε- and γ-globins (Supplementary Figure 5C). To further rule out the possible cell cycle arrest effect on the down regulation of ε- and γ-globin genes in MCM5- and ORC2-knocked down cells, we treated the K562 cells with three different cell cycle inhibitors, namely roscovitine, nocodazole and hydroxyl urea (Supplementary Figure 11). Treatment of roscovitine and nocodazole did not have any effect on the globin gene transcription. Treatment of hydroxyurea previously reported to increase γ-globin transcription in K562 and primary erythroid cells (Tang et al, 2005) significantly increased the γ-globin transcription in our cells (Supplementary Figure 11). These data suggest that reduced ε- and γ-globin transcription may not be an effect of cell cycle arrest, but are specific effects of MCM5 and ORC2 knockdown and that ILF2 and MCM5 have an essential and specific function for globin transcription.

In addition, knockdown of ILF2 in MEL cells produced significant reduction in mRNA of adult β-major and β-minor
Among promoters, acetylations at K5 and K12 were present in marked reduction of acetylation at the promoters, and among enhancers, acetylations at K8, K14 and K56 were present in a reduction of acetylation at the enhancers. The presence of p300 histone acetyltransferase in the DARRT complex suggests that the complex may have a function in regulation of chromatin structure through histone acetylation. We used ChIP qPCR to study the effects of ILF2 knockdown on the acetylation of the four histone H4 residues (H4K5, H4K8, H4K12 and H4K16) for several DARRT-binding regions. Interestingly, the pattern of acetylation in control cells was different for each of these residues (Figure 5E-H). Among promoters, acetylations at K5 and K12 were present in γ- and ε-globin genes, respectively. K8 acetylation was present at both the γ- and ε-globin genes. Among the HS sites of the LCR, K5, K12 and K16, acetylations were present mainly at the HS3 region, whereas none of the HS sites had acetylation at K8. Among the replication origin sites, we detected significant amounts of K8 acetylation at β-Rep-1 and 1.8 kb downstream of the HS4 sequence. ILF2 knockdown resulted in marked reduction of acetylation at the promoters, and
ILF2 shRNA expressing K562 cells. Non-silencing (NS) shRNA-expressing cells served as negative control, whereas MCM5-knocked down cells as positive control. The results show that knockdown of ILF2 abolishes the signals for DNA origins of replication in the lamin-B2 locus.

To further investigate a potential function for DARRT in DNA replication, we performed a replicating nascent-strand abundance assay in wild type and ILF2- and MCM5-knocked down K562 cells (Figure 6). This assay depends on isolation of single-stranded nascent DNA fragments that are protected from λ-exonuclease digestion because they retain an RNA primer at their 5’ end. We tested known IR sites in the β-globin locus (Kamath and Leffak, 2001; Aladjem et al., 2002; Wang et al., 2004; Buzina et al., 2005), and several novel lesser IR sites that we have identified at HS1 and between the HS4 and HS3 regions by analysing the abundance of replicating short single-stranded RNA–DNA hybrids (Figure 6). In this assay, we found similar enrichment patterns of nascent replicating DNA molecules with both 300–600 and 600–1200 bp DNA fragments. As expected, we did not observe the enrichment of the nascent replicating strands when we size selected 2–4 kb DNA fragments (Supplementary Figure 10B), and there was no enrichment when the single-stranded DNA (ssDNA) was treated with RNase before λ-exonuclease digestion (Supplementary Figure 10A). MCM5 knockdown resulted in marked decrease of replication initiation at the IR sites in the β-globin locus as well as at the lamin-B locus, which contains a well-studied origin of replication (Figure 6E) (Lucas et al., 2007), whereas ILF2 depletion specifically abrogated replication initiation at the IR of the β-globin locus, but not at the lamin-B locus (Figure 6E). As ILF2 knockdown partially silenced the β-globin cluster, we wished to test whether the method

Figure 6 Nascent-strand abundance assay for mapping origins of DNA replication. Graphs show the qPCR quantization of nascent strands originating from replicating foci at the β-globin locus. Short nascent single-stranded DNA was isolated as described in Materials and methods. For each qPCR, 3.5 ng of the DNA were used. The primers used for the qPCR are listed in Supplementary data. Fold enrichment shown in the y axis is the PCR signal intensity over the signal from an equal amount of non-replicating DNA obtained from the flow through of a BND-cellulose column. The shRNA construct harbouring K562 cells were grown in the presence of doxycycline for 9 days, at the end of which they were used for the nascent-strand abundance assay. Solid lines in the graph represent ILF2 (A), MCM5 (B), RAD50 (C) and non-silencing (D) shRNA-expressing K562 cells. Non-silencing (NS) shRNA-expressing cells served as negative control, whereas MCM5-knocked down cells as positive control. Graphs with dotted lines represent normal (control) K562 cells. (E) Effect of ILF2 and MCM5 knockdown cells on the origin of DNA replication at lamin-B2 gene. The values in each point in the graphs represent the average of two independent biological experiments performed in duplicates. The results show that knockdown of ILF2 abolishes the signals for DNA origins of replication in the β-globin locus, but not in the lamin-B2 locus.

©2010 European Molecular Biology Organization
we used for detecting replication origins was still effective when the β-globin cluster was embedded in transcriptionally inactive silenced chromatin. To test this, we examined the β-globin IRs in three cell types in which globin is not expressed and the chromatin is in a relatively closed conformation. The method we used readily detected the β-globin IR in the cell lines that were not transcribing the globin mRNA (Supplementary Figure 8A). These data indicated that ILF2, which recruits DARRT to the HS4 region of the LCR, is essential for firing only from a sub-set of DNA replication origins. RAD50 knockdown, however, did not have any significant effect on origin firing (Figure 6C), suggesting that some components of DARRT are dispensable for its function in replication. The MCM proteins and ORC2 are well-known components of the initiation complex for DNA replication, and perhaps ILF2 enables origin firing at the β-globin locus at least partly by recruiting these components of DARRT to this locus.

**Discussion**

We have identified a protein complex DARRT that appears to have an important function in DNA replication, histone modification and transcriptional regulation at the β-globin locus, which suggests that this complex may mediate cross-talk between these processes. Several subunits of DARRT have known functions in DNA replication, transcription and repair (Snyder et al., 2005, 2009; Zhao et al., 2005; Merrill and Gromeier, 2006; Shi et al., 2007; Pei et al., 2008; Jha and Dutta, 2009; Sakamoto et al., 2009). Out of the several DARRT proteins, ILF2 or the ILF2/3 heterodimer are candidate DNA-binding components that function to recruit DARRT to its target DNA sequence in HS4. Interestingly, the ILF2/ILF3 heterodimer has previously been reported to be associated with DNA-PK that is part of the double-stranded break repair complex (Ting et al., 1998). DARRT also contains several well-known DNA repair proteins including DNA-PK, RAD50, MRE11 and APEX proteins. The significance of DNA repair proteins in DARRT remains to be determined as RAD50 and the chromatin is in a relatively closed conformation. The shRNA knockdown studies confirm this, indicating that, apart from transcription, a sub-set of DNA replication at the globin IR as well as globin gene transcription. The shRNA knockdown studies confirm this, indicating that, apart from transcription, a sub-set of DNA replication origins are targeted by DARRT. Although we cannot absolutely exclude the alternative possibility that ILF2 knockdown acts indirectly through effects on some undiscovered factor elsewhere in the genome that is necessary for use of the globin origin of replication and transcription of the globin genes, it seems simpler to propose that the effects are a result of actions of the DARRT complex containing ILF2 directly at the globin locus where we know it binds.

Although a majority of the replication initiation sites contain AT-rich sequences (Gilbert, 2001), the mechanism for the choice of initiation sites for DNA replication is not clearly known. Frequent usage of a fairly large numbers of cryptic sites, competition between sites and flexibility of origin usage implies that multiple factors determine the final outcome of origin selection. Recruitment of DARRT-associated DNA replication initiation complexes to specific DNA sequences provides an example of transcription factor-mediated recruitment of a part of the pre-initiation complex. A few examples have been reported in animal viral systems, wherein sequence-specific transcription factors recruit the viral-encoded replication initiator/helicase to its replication
origin DNA sequence (Guo et al., 1996; Ito et al., 1996; Murakami et al., 2007) or stabilize the pre-initiation complex (Mul and Van der Vliet, 1992; van Leeuwen et al., 1997). In the case of EBV, EBNA1 loads cellular ORC into viral oriP sites (Scherers et al., 2001; Ritzi et al., 2003). In the case of human c-myc replication origin sequences, deletion of transcription factor-binding sites abolished the initiation of DNA replication (Ghosh et al., 2004). The MYC transcription factor was recently reported to be associated with active DNA replication origins independent of its transcriptional activity (Dominguez-Sola et al., 2007).

In summary, our data shows that an ILF2 containing multiprotein complex has a multiplex function in regulation of transcription, initiation of DNA replication and histone modification at the β-globin locus. This presumably is a consequence of sequence-specific recruitment of the DARRT complex to sites within the globin locus including the globin LCR HS4. This complex also contains several proteins implicated in DNA repair. Overall, the results imply an unexpected linkage between the transcription, locus-specific initiation of DNA replication and potentially DNA repair.

Materials and methods

Cell culture and media

KS62 and HeLa cells were grown in RPMI 1640 (with l-glutamine supplemented with 10% FBS and antibiotics and anticytotoxins (Invitrogen). The expansion and erythroid differentiation of CD34+ cells was as previously described (Majaian et al., 2009).

EMSA and oligonucleotide probes

The oligonucleotides used in the EMSA assays are listed in Supplementary data. EMSA and gel supershift procedures were performed according to the published protocol (Majaian and Weissman, 2002). The tilting EMSA screen for identification of KS62-specific DNA–protein interactions was carried out by designing a set of short synthetic double-stranded oligonucleotides from the phylogenetically conserved core HS4 region of the LCR. The average size of the oligonucleotides was 35 bp with 7 bp overlap between the adjacent sequences.

Biochemical purification of the DARRT complex

All purification procedures were carried out in the cold room and all buffers were pre-chilled on ice and supplemented with a cocktail of protease inhibitors (Roche). The composition of buffers and other solutions used for the purification procedures are described in Supplementary data. Cells (15 billion) were grown in spinner flasks and log phase cells were harvested by spinning down at 2000 r.p.m. and resuspended in ice cold hypotonic buffer (Buffer A) followed by nuclei lysis assay. The lysate was prepared by swelling the cells for 30 min in ice cold hypotonic buffer (Buffer A) followed by disruption with a Dounce homogenizer to lyse the cells. Nuclei were collected by centrifugation at 3000 r.p.m. and the nuclear envelope was disrupted in Buffer B. This suspension was then centrifuged at 15 000 r.p.m. for 30 min to collect the nuclear lysate. Salts were dialysed out over night against Buffer C or by passing the lysate through PD10 desalting columns followed by 10 min centrifugation (15 000 r.p.m. in a Sorvall DuPont RCSC centrifuge). Nuclear lysate was then used for the purification of the HS4–9-binding protein complex. During the purification, the HS4–9-binding proteins were tracked by EMSA assay in all elution fractions. The column run on SDS–PAGE was stained with Colloidal Coomassie Blue (Sigma-Aldrich) stain. Near equimolar staining of the protein bands was observed. The SDS–PAGE bands were excised from the gel and digested with trypsin. The tryptic peptides obtained from each protein digestion were sequenced by MS/MS for protein identification.

Mapping the origins of DNA replication

A nascent-strand abundance assay was carried out in asynchronously growing cells to determine the origins of DNA replication. The protocol was adapted from previously published procedures (Buzina et al., 2005; Gerbi, 2005). The compositions of the buffers are described in Supplementary data. Briefly, genomic DNA was prepared by suspending 108 cells in 100 ml of TEN buffer and further purified by using a chloroform extraction and a phenol:chloroform:isoamylic alcohol extraction step. The column elution fractions that displayed positive EMSA activity were selected for the next column fractionation. About 100 mg of the crude KS62 nuclei extract prepared as described above was loaded on a 30 ml bed volume 25 cm long heparin-Sepharose column pre-equilibrated with Buffer C. The HS4–9-associated DNA sequence was bound to the heparin-Sepharose column as the flow through did not have any EMSA activity. The heparin-Sepharose column-bound proteins were batch eluted with 150 ml each of 0.2, 0.42 and 0.6 M NaCl in Buffer C. Seventy five 2 ml fractions were collected during each salt elution and assayed for the HS4–9-associated EMSA activity. We found HS4–9-binding proteins only in the 0.42 M NaCl fractions, which were then concentrated and passed over a FPLC-based Superose-6 sizing column pre-equilibrated with Buffer C. Fractions that showed active HS4–9-binding eluted in the 2–5 SMDa size range. These were further purified over an ion-exchange DEAE-Sephaloc column that was also pre-equilibrated with Buffer C that contained 0.1 M NaCl. The high-molecular weight HS4–9 DNA-binding proteins were bound to the DEAE-Sephaloc column as the flow through tested negative for the EMSA activity. The column was washed extensively with 200 ml of 0.1 M NaCl in Buffer C. The proteins bound to the DEAE-Sephaloc column were batch eluted with the 0.2 M NaCl salt fractions and tested for HS4–9 DNA sequence-binding activity by EMSA. We found HS4–9 DNA-binding activity only in the protein fractions eluted with 0.2 M NaCl. The EMSA-positive fractions were pooled and dialysed against Buffer C. The preparation was then passed through a scrambled HS4–9 oligonucleotide affinity column as judged by the lack of EMSA activity in the flow through. The column was then washed serially with 30 ml of Buffer C, 15 ml of poly-dIdC in Buffer C at 0.1 mg/ml concentration and finally with 15 ml of Buffer C. The bound proteins were eluted with 0.2, 0.3 and 0.4 M NaCl in Buffer C and tested for the HS4–9-binding protein by EMSA. Fifteen 0.5 ml fractions were collected at each batch elution. We detected EMSA activity only in the 0.2 NaCl elution fractions. These positive EMSA fractions were pooled, concentrated and displayed on an SDS–PAGE gel. The proteins run of SDS–PAGE were stained with Colloidal Coomassie Blue (Sigma-Aldrich) stain. Near equimolar staining of the protein bands was observed. The SDS–PAGE bands were excised from the gel and digested with trypsin. The tryptic peptides obtained from each protein digestion were sequenced by MS/MS for protein identification.
References


Bridges BA (2005) Error-prone DNA repair and translesion synthesis: focus on the replication fork. DNA Repair (Annu) 4: 618–619, 634


Dimitrova DS (2006) Nuclear transcription is essential for specifica-
tion of mammalian replication origins. Genes Cells 11: 829–844


Gottipati P, Helleday T (2009) Transcription-associated recombina-
tion in eukaryotes: link between transcription, replication and recombination. Matagena 24: 203–210


Conflict of interest

The authors declare that they have no conflict of interest.


Mul YM, Van der Vliet PC (1992) Nuclear factor I enhances adenovirus DNA replication by increasing the stability of a preinitiation complex. EMBO J 11: 751–760


