Proteomic analysis of DNA synthesis on a structured DNA template in human cellular extracts: Interplay between NHEJ and replication-associated proteins.

Regine JANEL-BINTZ^{1*}, Lauriane KUHN^{2*}, Philippe FRIT³, Johana CHICHER², Jérôme WAGNER¹, Lajos HARACSKA⁴, Philippe HAMMANN², Agnes M CORDONNIER¹

1. Biotechnologie et Signalisation Cellulaire, Université de Strasbourg, UMR7242, CNRS, Illkirch 67412, France

2. Institut de Biologie Moléculaire et Cellulaire du CNRS, Plateforme Protéomique Strasbourg - Esplanade, FR1589, 67084 Strasbourg, France

3. Institut de Pharmacologie et Biologie Structurale, IPBS, Université de Toulouse, CNRS, UPS, Toulouse, France. Equipe Labellisée Ligue Contre le Cancer 2018, Toulouse, France.

4. Institute of Genetics, Biological Research Center, HU-6726 Szeged, Hungary

*These authors contributed equally to this work.

Corresponding author: agnes.cordonnier@unistra.fr

Received: 29/05/2019; Revised: 19/12/2019; Accepted: 29/01/2020

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/pmic.201900184](https://doi.org/10.1002/pmic.201900184).

This article is protected by copyright. All rights reserved.

Running Title: Interplay between NHEJ and replication-associated proteins.

Abbreviations: SIR, short inverted repeat; NHEJ, Non-Homologous End-Joining; dsO, double-stranded oligonucleotide; ssDNA, single-stranded DNA; DSB, double-strand break; pUC118.ss, single-stranded pUC118; V(D)J recombination, variable (V), diversity (D) and joining (J) gene segment recombination

Keywords: DNA synthesis; DNA-PK complex; SIR, short inverted repeat

Abstract

It has been established that short inverted repeats trigger base substitution mutagenesis in human cells. However, how the replication machinery deals with structured DNA is unknown. We have previously reported that in human cell-free extracts, DNA primer extension using a structured single-stranded template is transiently blocked at DNA hairpins. Here, we report the proteomic analysis of proteins bound to the DNA template and provide evidence that the DNA-PK complex (DNA-PKcs and the Ku heterodimer) recognizes, and is activated by, structured single-stranded DNA. Hijacking the DNA-PK complex by double-stranded oligonucleotides results in a large removal of the pausing sites and an elevated DNA extension efficiency. Conversely, DNA-PKcs inhibition results in its stabilization on the template,

along with other proteins acting downstream in the Non-Homologous End-Joining pathway, especially the XRCC4-DNA Ligase 4 complex and the cofactor PAXX. Retention of NHEJ factors to the DNA in the absence of DNA-PKcs activity correlates with additional halts of primer extension, suggesting that these proteins hinder the progression of the DNA synthesis at these sites.

Overall these results raise the possibility that, upon binding to hairpins formed onto ssDNA during fork progression, the DNA-PK complex interferes with replication fork dynamics *in vivo*.

Significance statement

All over the course of DNA synthesis, single-stranded DNA intermediates are produced, leading to the formation of secondary structures. It has been shown that these structures represent replication hurdles that induce replication fork stalling or collapse *in vivo*. Using an *in vitro* primer extension assay in human cell-free extracts, we have shown that short inverted repeats in a single-stranded template DNA trigger a transient pausing of the DNA polymerase mimicking the *in vivo* situation. In order to identify the proteins recruited onto the template DNA, we have set up a protein capture assay coupled to mass spectrometry analysis. We provide evidence that the DNA-PK complex (DNA-PKcs and the Ku heterodimer) recognizes, and is activated by, structured single-stranded DNA. DNA-PKcs inhibition results in its stabilization on the DNA template, along with other proteins acting downstream in the Non-Homologous End-Joining pathway, especially the XRCC4-DNA Ligase 4 complex and the cofactor PAXX. These proteins, bound to DNA hairpin, counteract primer extension catalyzed by cellular extracts. This result opens up the possibility that the DNA-PK complex, owing to its ability to bind onto DNA secondary structures, interferes with replication fork dynamic *in vivo*, even in absence of DNA double-strand breaks.

Introduction

Non-B DNA structures have been implicated in genome instability associated with several pathological disorders ^[1]. All over the course of DNA synthesis, single-stranded DNA (ssDNA) intermediates, that have a propensity to form secondary structures, are produced. Among them, inverted repeats may fold into DNA hairpins with the palindromic sequence and the intervening spacer sequence forming the stem and the loop, respectively. It has been shown that these structures represent replication hurdles that lead to replication fork stalling or collapse *in vivo* ^[2]. Of particular interest, short inverted repeats (SIRs with a size <100 bases) are very abundant in the human genome with variable densities ranging from 2000 to 3000 SIRs per Mb throughout the referenced human genome ^[3]. They were shown to impede DNA replication forks in mammalian cells and to be hotspots for formation of double-strand breaks (DSBs) and deletion mutations ^[4], as previously reported for long inverted repeats ^[5]. Surprisingly, the spacer sequences, corresponding to the loop of the hairpins show a higher level of base substitution mutagenesis than their surrounding sequences ^[3,6]. Yet, the mechanism responsible for this intrinsic high mutability is currently unknown.

Using an *in vitro* primer extension assay in order to analyze DNA

Accepted Article

synthesis catalyzed by cellular extracts we have previously shown that short inverted repeats in a single-stranded template DNA induce a transient pausing of the DNA polymerase ^[7,8], mimicking the *in vivo* situation ^[4]. In order to identify the proteins recruited onto the template DNA, we have set up a protein capture assay coupled to mass spectrometry analysis. Our data suggest that components of the NHEJ pathway, bound to structured single-stranded DNA, counteract primer extension catalyzed by cellular extracts *in vitro*. This result opens up the possibility that the DNA-PK complex, owing to its ability to bind onto DNA secondary structures, interferes with processes occurring on single-stranded DNA, in the absence of double-strand DNA breaks.

Experimental Procedures

A full description of Materials and Methods can be found in SI Materials and Methods.

Results

1. Characterization of proteins bound to the structured single-stranded template DNA during primer extension

We examined the replication of a circular single-stranded template DNA (pUC118.ss) in a primer extension assay catalyzed by human cell-free extracts. The primer is 5' labeled with ³²P to monitor product formation on a denaturing

gel. As previously shown ^[7,8], the analysis of the elongation products after 10 minutes incubation reveals the presence of several bands indicating stalling of the DNA synthesis at secondary structures. The first hairpin encountered by the replication machinery represents a major block of DNA elongation as about 60 % of the elongation products are arrested at this site at the 10 minutes time point (Fig. 1A). We have previously shown that the annealing of oligonucleotides complementary to this hairpin sequence releases the blockage by impairing this hairpin formation ^[8]. A depletion/reconstitution experiment shows that DNA polymerase δ is the DNA polymerase responsible for primer extension in this assay.

As a first step toward the identification of the proteins recognizing the structured ssDNA, as well as those involved in the primer extension reaction in cellular extracts, a biotinylated primer hybridized to the pUC118.ss template was immobilized on streptavidin beads. After washing the beads in order to remove non-specific proteins, the proteins bound to the template DNA were analyzed by liquid chromatography coupled to mass spectrometry without any labeling procedure. Two widely used label-free quantification methodologies were conducted to relatively quantify peptides and proteins across the samples: Spectral Counting (SC, Suppl. Tables 1-6a) and eXtracted Ion Chromatogram (XIC, Suppl. Tables 1-6b). Specific and dedicated post-

processing packages and statistical tools were applied to both types of MS datasets and we focused our attention on proteins as statistically enriched by both SC and XIC methodologies.

We compared the proteins captured after 10 minutes incubation of cell extracts in standard conditions, with and without the template/primer, so as to distinguish proteins specifically bound (adjusted p-value < 0.05) to the primed DNA from those that were non-specifically enriched. An additional control experiment performed in the absence of dNTPs allows distinction between DNA binding and replication associated proteins (Suppl. Tables S1a and b).

The Venn diagram (Fig. 1B) depicts proteins found enriched on the primed template DNA versus the control (-DNA) in the presence of dNTPs in at least two out of three independent experiments either by Spectral Count (SC) or eXtracted Ion Chromatogram (XIC) methods. Gene Ontology enrichment analysis assigns 50-60% of the 26 proteins found enriched on the template DNA by both SC (Suppl. Tables S2a-S4a) and XIC methods (Suppl. Tables S2b-S4b), to the biological processes of DNA repair and DNA replication, as expected. Table 1 displays the proteins found enriched by both methods in one of these three experiments (Suppl. Tables S4a and b). The single-stranded DNA binding complex RPA (RPA70, RPA32 and RPA14) is specially identified in the

Accepted Article

samples containing DNA, as well as other DNA repair proteins that have been recruited to the template due to their ability to bind to single-stranded DNA or to RPA. In addition, some *bona fide* replisome components, PCNA and PCNA binding proteins are recovered (highlighted in bold in the Venn diagram Fig. 1B). Accordingly, we have clearly established that PCNA^[16] is required for the primer extension reaction in this system. We noticed that DNA polymerase δ subunits are not found to be enriched. This result probably reflects the fact that only a small fraction of DNA molecules undergoes replication in these experimental conditions (see Fig. 2C).

Surprisingly, proteins involved in Non-Homologous End-joining (NHEJ), DNA-PKcs along with the Ku complex, have been reproducibly and significantly detected by both the SC and XIC methods, in the presence as well as in the absence of dNTPs (Fig. 1C). Similarly, protein PAXX belonging also to the NHEJ repair pathway, is found significantly enriched in the absence of dNTPs but in the presence of dNTPs the numbers of peptides are too low to be considered as significant by Spectral Counting (Fig. 1C). Nevertheless, a time course analysis by Western blot unambiguously confirms the binding of DNA-PK as well as PAXX to the DNA template independently of the primer extension (Fig. 1D).

During the repair of DNA double-strand breaks, the Ku heterodimer binds to the free DNA ends protecting them from nuclease attack and recruits DNA-PKcs and downstream repair factors, allowing DNA ends processing and ligation ^[17]. The presence of the DNA-PK complex onto the DNA template in our experimental conditions reinforces a previous report showing that structured single-stranded DNA activates purified DNA-PK *in vitro* ^[18]. As previously proposed by others ^[19], Ku may enter the DNA template by inducing a distortion of the hairpin loop on the structured ssDNA. The stem of the DNA hairpin will provide a portion of dsDNA that further strengthens the association of the Ku dimer to the DNA hairpin.

In order to assess that the proteins belonging to the core NHEJ process (DNA-PKcs, the Ku dimer and PAXX) are able to bind to structured single-stranded DNA, we performed a capture assay using a 5' biotinylated linear single-stranded DNA fragment encompassing the first DNA hairpin encountered by the replication machinery (Fig. 1E). As a control, the sequence of the stem of the hairpin was mutated in order to impair the DNA folding. The impact of these mutations is validated by the pattern of primer extension catalyzed by cellular extracts. Indeed pausing of the DNA extension at this site is no longer observed upon incubation of cellular extracts with the mutated template (Fig. S1). As shown in figure 1E, DNA-PKcs, Ku80, Ku70 and PAXX are

recovered efficiently in a pull-down assay with the hairpin DNA (lane 2), while these proteins are barely detected when the mutated DNA is employed as bait (lane 1). This experiment shows that the binding of these NHEJ factors to the hairpin DNA is specific.

In addition to these DNA repair and DNA replication-associated proteins, a number of proteins belonging to the family of Dynamins, involved in the cytoskeleton dynamics, are specifically recruited onto the substrate. A similar result has already been noticed in a proteomic study of DNA-ends specific proteins ^[20], but the functional relevance of this unexpected association awaits further investigations. Finally, some heterogeneous nuclear ribonucleoproteins were recovered that reflects their ability to bind to ssDNA with high affinity ^[21].

2. Double-stranded oligonucleotides (dsO) prevent the NHEJ proteins from assembling on the template DNA

In order to assess the DNA-PK interference with DNA synthesis in cell-free extracts, we used double-stranded oligonucleotides (dsO) that have been shown to trap the DNA-PK complex ^[22]. The capture of DNA-PKcs, Ku70 and Ku80 by the competitor dsO is supported by the mass spectrometry analysis, as the amount of these DNA-bound proteins dramatically drops down upon addition of the competitor dsO (Table 1, Suppl. Tables Sa and b).

Accepted Article

It has to be noted that the addition of dsO to the cellular extracts has no significant impact on the amount of the other captured proteins, except that replication-associated proteins are more abundant (Table 1: replication proteins-1). Moreover, additional proteins belonging to this pathway are at present found significantly enriched on the template DNA by both methods (Table 1: replication proteins-2), such as the POLD2 subunit of the DNA polymerase δ as well as PCNA binding proteins, PAF15, the mismatch repair protein MSH2, the DNA ligase 1 (DNLI1), the RNase H2 complex and DTL (CDT2), a component of the E3 ubiquitin ligase CRL4^{CDT2} regulating the degradation of some PCNA interacting proteins.

The addition of dsO causes a large decrease in the intensity of the pausing bands at the DNA hairpins and an increase of the amount of the replication product (Fig. 2A left panel) in a dose-dependent manner. The elevated overall elongation efficiency in the presence of dsO is confirmed by the measurement of labeled α -dCTP incorporation during primer extension (Fig. 2C). Remarkably, complementation of the extract with purified Ku dimer restores the pausing bands in a dose-dependent manner (Fig. 2A right panel) indicating that these proteins are indeed responsible for the halt of the replication complex.

In conclusion, we found that addition of dsO in cellular extracts diverts the DNA-PK complex from the ssDNA template and that purified Ku dimer reverts the dsO effect. This result strongly suggests that the binding of these proteins to the ssDNA template physically counteracts the progression of the DNA polymerase.

3. Impact of DNA-PK inhibition on primer extension of structured single-stranded DNA in cellular extracts

During NHEJ, the assembly of DNA-PKcs at DSB ends via the Ku proteins triggers the activation of DNA-PK, thereby phosphorylating a number of substrates including DNA-PKcs itself. Autophosphorylation of several residues around T2609 within the ABCD cluster and around S2056 within the PQR cluster of DNA-PKcs regulates end-processing and end-ligation steps^[23–25]. The observation that DNA-PK binds to the DNA template prompted us to investigate whether DNA-PK is activated by structured single-stranded DNA in cellular extracts. Figure 3A shows that the unprimed ssDNA triggers phosphorylation of DNA-PKcs on both residues S2056 and T2609, whereas these phosphorylations are no longer observed when the extracts were pre-incubated with the specific inhibitor of DNA-PKcs NU7441. These data

demonstrate DNA-PK activation and auto-phosphorylation upon addition of the structured single-stranded DNA, accordingly to a previous report ^[18].

In order to explore the influence of DNA-PK activity on primer extension, the reaction was performed in the presence or absence of the DNA-PK inhibitor at different concentrations. An ATM inhibitor KU-55933 was used in parallel as a control. Figure 3B (upper panel) shows that in the presence of the DNA-PKcs inhibitor, additional arrested elongation products are readily detectable, particularly at the base of the first hairpin (depicted by an open triangle in Fig. 3B and 3C). A quantitative analysis shows that this band represents 12.5% and 4.1% of the whole elongation products at the 5 min and 20 min time points respectively, whereas it is quite undetectable in the absence of the inhibitor (Fig. 3D). A time course experiment discloses the transitory nature of the pausing sites, both in the absence and presence of the DNA-PK inhibitor. Indeed, longer incubation times allow replication to progress beyond the pausing sites, as shown by the increase of the intensity of bands corresponding to longer products (Fig. 3C, upper panel).

Substantial inhibition of the catalytic activity of DNA-PKcs by the inhibitor is demonstrated by a net decrease of RPA32 phosphorylation at Serine 4 and Serine 8 (Fig. 3B and Fig. 3C, lower panels), described as DNA-PKcs

phosphorylation target sites *in vivo* after replication stress^[26]. Our data show that concentrations beyond 10 μ M of the DNA-PK inhibitor are appropriate to efficiently inhibit DNA-PK activity in our experimental conditions, as well as to reveal the impact of this inhibition on the primer elongation, while lower doses are less efficient. On the other hand, inhibition of ATM, another serine/threonine kinase, also member of the PIKK family, has no impact on the DNA synthesis, and only a small effect on RPA32 phosphorylation at the highest inhibitor concentration (Fig. 3B).

With the aim of investigating if the absence of RPA phosphorylation might be responsible for the emergence of the additional stalling sites, we examined the profile of primer extension catalyzed by extracts depleted in RPA32 versus mock-depleted extracts (Fig. S2A). Optimal RPA-immunodepletion is performed in two successive rounds of incubation of an MRC5 cell extract with an antibody directed against RPA32 and G-Sepharose beads. The three subunits of the RPA heterotrimer, namely RPA70, RPA32 and RPA14, are simultaneously depleted by this procedure as shown by immunoblotting of the depleted extracts (Fig. S2B). While the global efficiency of elongation is slightly reduced in the RPA-depleted extract compared to that in the mock-depleted extract, the general profiles of primer elongation are similar (Fig. S2A). Thus, RPA is not implicated in the formation of the additional

stalling sites induced by the inhibition of DNA-PKcs activity.

Overall, these results establish that DNA-PK is activated by single-stranded circular DNA in cellular extracts where it phosphorylates RPA32 at the S4/S8 sites. The analysis of the elongation products shows a similar global efficiency of DNA synthesis in the presence or absence of the DNA-PKcs inhibitor, indicating that inhibition of DNA-PKcs had no major impact neither on the RFC-dependent loading of PCNA nor on the binding of PCNA-associated factors. However, the appearance of an additional pausing site at the base of the first DNA hairpin encountered by the replication machinery, strongly argues in favor of a mobilization of proteins at this site. When the Ku dimer is hijacked from the substrate upon incubation of the extract with dsO, the DNA-PKcs inhibitor has no longer any effect on the pattern of DNA synthesis (Fig. S2C), again suggesting that the association of DNA-PK at this site is necessary for the emergence of this additional pausing site.

4. NHEJ factors (XRCC4, DNA ligase 4 and PAXX) accumulate on the template DNA *in vitro* when the DNA-PK activity is compromised

To monitor how the DNA-bound proteins change in response to the DNA-PK inhibitor, a protein capture assay was performed on the extending DNA, in the absence and presence of the DNA-PKcs inhibitor at the 10 minutes

time point. The mass spectrometry analysis reveals that the DNA-PKcs inhibitor induces a significant specific retention to the ssDNA of the ligation complex, consisting of PAXX, DNA ligase 4 and XRCC4 (Fig. 4A and Suppl. Tables S6a and b). Accordingly to the mass spectrometry, analysis of captured proteins by immunoblot supports the accumulation of PAXX, DNA ligase 4-XRCC4 proteins on the template DNA in the presence of the DNA-PK inhibitor (Fig.4B). It has been shown that multiple protein-protein interactions collectively help to the formation of a supramolecular complex during NHEJ, in which the DNA-PK and ligation complexes co-exist at DNA ends ^[27,28]. In addition, impairment of the kinase activity upon pharmacological inhibition of DNA-PKcs increases its residence time at DSBs ^[29]. This provides an explanation for the retention of DNA-PKcs together with the ligation complex onto the circular ssDNA in the presence of the inhibitor as illustrated in the Western blot (Fig. 4B).

Discussion

The binding ability of the Ku dimer (Ku70/Ku80) to a wide variety of DNA end configurations underlies the central role of DNA-PK in DSB repair and V(D)J recombination, the mechanism responsible for generating antigen receptor diversity.

The results presented here argue to another less explored DNA binding

capacity of Ku to hairpins formed at small inverted repeats. Indeed, our data support the idea that the arrested primer extension products observed both in the presence and absence of the DNA-PK inhibitor mark the physical interaction of DNA-PK with a DNA hairpin. The small effect of the DNA-PK inhibitor on DNA elongation merely reflects the increase of the residence time of DNA-PK on DNA hairpins and the presence of additional NHEJ factors when the phosphorylation of DNA-PKcs is inhibited. Depending on the configuration of the DNA hairpin, the Ku dimer may persist even in the presence of active DNA-PK. This is the case of the band depicted by an asterisk, which represents the major elongation block. The displacement of the DNA-PK complex from the ssDNA by the dsO is concomitant with the large removal of the pausing sites, both in the presence and absence of the DNA-PK inhibitor. The activation of DNA-PK by structured ssDNA raises the possibility that DNA-PK interferes with some DNA transactions occurring on ssDNA *in vivo*, independently of DSB.

Binding ability of Ku at hairpins

The crystal structure of the Ku heterodimer shows that its double ring shape exactly fits around the dsDNA. The affinity of Ku for DNA ends is independent of the DNA sequence and precise structure, but requires a minimum length of 14 nucleotides^[30]. Interestingly, it has been demonstrated

in vitro^[19] that Ku can additionally bind to a hairpin substrate lacking free ends with high affinity (with a K_d value in the nM range), albeit a little weaker than to double-stranded oligonucleotides.

A few examples have demonstrated the biological importance of the binding ability of Ku to hairpins *in vivo*. In lymphoid cells, coding ends with a perfect hairpin configuration are generated during V(D)J recombination. DNA-PKcs/ARTEMIS complex opens the hairpins, thereby allowing end processing and assembly of coding ends by non-homologous DNA end-joining. While it has been shown that the entire process could be reconstituted *in vitro* in the absence of Ku^[31], it is likely that Ku binds to the hairpin *in vivo* to enhance the affinity of DNA-PKcs/ARTEMIS for the target DNA substrate and to activate DNA-PKcs^[32]. Finally, the functional relevance of the ability of Ku to bind to stem loop structures *in vivo* has been established for RNA hairpins. Indeed, the binding of Ku to the p53 5'UTR represses p53 mRNA translation^[33]. In addition, Ku associates to a specific stem-loop structure of the telomerase RNA in both yeast^[34] and human^[35] and this interaction is crucial for recruiting the active reverse transcriptase complex^[36].

We show here that, in cellular extracts, the NHEJ complex including DNA-PKcs, Ku, XRCC4/Ligase4 and PAXX is stably interacting with structured

Accepted Article

single-stranded DNA when DNA-PKcs is inhibited. PAXX (Paralog of XRCC4 and XLF), a recently identified NHEJ factor, shares structural similarity with XRCC4 and XLF [37–39]. Together PAXX and XLF promote efficient DNA end-ligation through distinct mechanisms, with XLF enhancing Ligase 4 recruitment while PAXX stabilizes Ku at DNA ends [40]. The first hairpin encountered by the replication machinery is composed of a 15 nucleotides long stem including two mismatches and a 5 nucleotides loop, which represents a proper configuration for Ku binding. Interestingly, PAXX, but not XLF, was specially identified in our experiments, suggesting that these proteins may have distinct requirements for stably interacting with Ku-bound DNA. Accordingly, the association of PAXX with a Ku-bound 15-bp dsDNA *in vitro* requires a 5' or 3' short stretch of bare DNA that can be either single or double-stranded [41]. The DNA surrounding both sides of the SIR could thus help the association of PAXX with Ku bound to the hairpin DNA. Delineating the structural determinants of Ku binding to DNA and RNA hairpins would potentially allow to uncover new functions of Ku in the modulation of replication and gene expression.

Potential consequences of the binding of NHEJ factors to SIRs during replication

Here, we show that, in cellular extracts, primer extension using a

Accepted Article

circular ssDNA template is hindered by the physical association of NHEJ factors to SIRs. This result highlights a possible interference of these factors in the replication fork dynamics *in vivo*. Further experiments are necessary to know if the transient association of NHEJ factors and/or DNA-PK activation, could trigger some remodeling of the replisome, thus modulating its properties. This could explain the targeted mutagenesis observed *in vivo* [3]. On the other hand, prolonged fork stalling at some of these sites may promote the action of the nuclease ARTEMIS that would create a one-ended DSB by cleaving the hairpin loop. Consistently, it has recently been shown that ARTEMIS and XPF-ERCC1 function independently to produce DSBs early after treatment of the cells with hydroxyurea, an inhibitor of dNTP biogenesis that induces ssDNA formation by replicative helicase. This DSB induction promotes replication fork restart^[42,43]. Our data, showing that DNA-PKcs is targeted to SIRs in cellular extracts, provide a possible explanation for the activation of the DNA-PKcs that is essential for ARTEMIS endonuclease activity, when cells have to overcome a replication stress.

In conclusion, our investigation confirms and emphasizes the propensity of the DNA-PK complex to bind to structured ssDNA. These data open new avenues to elucidate the crosstalk between NHEJ and replication proteins and to understand the role of DNA-PK in replication and S-phase checkpoint

signaling.

Fundings:

Institut National du Cancer [PLBIO 2016/INCa_10493] and European Union and European Regional Development Funds GINOP-2.3.2-15-2016-00020.

The mass spectrometry instrumentation was funded by the University of Strasbourg, IdEx “Equipement mi-lourd” 2015.

Data availability:

Raw data are available via ProteomeXchange with identifier PXD011156.

References

- [1] A. Aguilera, B. Gómez-González, *Nat. Rev. Genet.* **2008**, *9*, 204.
- [2] I. Voineagu, V. Narayanan, K. S. Lobachev, S. M. Mirkin, *Proc. Natl. Acad. Sci.* **2008**, *105*, 9936.
- [3] X. Zou, S. Morganella, D. Glodzik, H. Davies, Y. Li, M. R. Stratton, S. Nik-Zainal, *Nucleic Acids Res.* **2017**, *45*, 11213.
- [4] S. Lu, G. Wang, A. Bacolla, J. Zhao, S. Spitser, K. M. Vasquez, *Cell Rep.* **2015**, *10*, 1674.
- [5] H. Kurahashi, H. Inagaki, T. Ohye, H. Kogo, T. Kato, B. S. Emanuel, *DNA Repair* **2006**, *5*, 1136.
- [6] I. Georgakopoulos-Soares, S. Morganella, N. Jain, M. Hemberg, S. Nik-Zainal, *Genome Res.* **2018**, DOI 10.1101/gr.231688.117.
- [7] V. Schmutz, J. Wagner, R. Janel-Bintz, R. P. Fuchs, A. M. Cordonnier, *DNA*

Repair Amst **2007**, *6*, 1726.

- [8] V. Schmutz, R. Janel-Bintz, J. Wagner, D. Biard, N. Shiomi, R. P. Fuchs, A. M. Cordonnier, *Nucleic Acids Res.* **2010**, *38*, 6456.
- [9] L. I. Huschtscha, R. Holliday, *J Cell Sci* **1983**, *63*, 77.
- [10] A. M. Cordonnier, A. R. Lehmann, R. P. Fuchs, *Mol Cell Biol* **1999**, *19*, 2206.
- [11] O. Bombarde, C. Bobby, D. Gomez, P. Frit, M.-J. Giraud-Panis, E. Gilson, B. Salles, P. Calsou, *EMBO J.* **2010**, *29*, 1573.
- [12] J. Chicher, A. Simonetti, L. Kuhn, L. Schaeffer, P. Hammann, G. Eriani, F. Martin, *PROTEOMICS* **2015**, *15*, 2417.
- [13] C. Carapito, A. Burel, P. Guterl, A. Walter, F. Varrier, F. Bertile, A. Van Dorsselaer, *PROTEOMICS* **2014**, *14*, 1014.
- [14] J. Gregori, L. Villarreal, A. Sánchez, J. Baselga, J. Villanueva, *J. Proteomics* **2013**, *95*, 55.
- [15] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, C. von Mering, *Nucleic Acids Res.* **2015**, *43*, D447.
- [16] R. Janel-Bintz, J. Wagner, L. Haracska, M. C. M. Mah-Becherel, M. Bichara, R. P. Fuchs, A. M. Cordonnier, *PLoS ONE* **2012**, *7*, e36004.
- [17] A. J. Davis, B. P. C. Chen, D. J. Chen, *DNA Repair* **2014**, *17*, 21.
- [18] S. Soubeyrand, H. Torrance, W. Giffin, W. Gong, C. Schild-Poulter, R. J. G. Haché, *Proc. Natl. Acad. Sci.* **2001**, *98*, 9605.
- [19] D. Arosio, S. Cui, C. Ortega, M. Chovanec, S. Di Marco, G. Baldini, A. Falaschi, A. Vindigni, *J. Biol. Chem.* **2002**, *277*, 9741.
- [20] V. Berthelot, G. Mouta-Cardoso, N. Hégarat, F. Guillonneau, J.-C. François, C. Giovannangeli, D. Praseuth, F. Rusconi, *Nucleic Acids Res.* **2016**, *44*, 4721.

- [21] S. Pinol-Roma, Y. D. Choi, M. J. Matunis, G. Dreyfuss, *Genes Dev.* **1988**, *2*, 215.
- [22] Y. Yuan, S. Britton, C. Delteil, J. Coates, S. P. Jackson, N. Barboule, P. Frit, P. Calsou, *Nucleic Acids Res.* **2015**, gkv894.
- [23] D. W. Chan, B. P.-C. Chen, S. Prithivirajasingh, A. Kurimasa, M. D. Story, J. Qin, D. J. Chen, *Genes Dev.* **2002**, *16*, 2333.
- [24] Q. Ding, Y. V. R. Reddy, W. Wang, T. Woods, P. Douglas, D. A. Ramsden, S. P. Lees-Miller, K. Meek, *Mol. Cell. Biol.* **2003**, *23*, 5836.
- [25] X. Cui, Y. Yu, S. Gupta, Y.-M. Cho, S. P. Lees-Miller, K. Meek, *Mol. Cell. Biol.* **2005**, *25*, 10842.
- [26] S. Liu, S. O. Opiyo, K. Manthey, J. G. Glanzer, A. K. Ashley, C. Amerin, K. Troksa, M. Shrivastav, J. A. Nickoloff, G. G. Oakley, *Nucleic Acids Res.* **2012**, *40*, 10780.
- [27] K. Yano, D. J. Chen, *Cell Cycle* **2008**, *7*, 1321.
- [28] J. Cottarel, P. Frit, O. Bombarde, B. Salles, A. Négrel, S. Bernard, P. A. Jeggo, M. R. Lieber, M. Modesti, P. Calsou, *J. Cell Biol.* **2013**, *200*, 173.
- [29] N. Uematsu, E. Weterings, K. Yano, K. Morotomi-Yano, B. Jakob, G. Taucher-Scholz, P.-O. Mari, D. C. van Gent, B. P. C. Chen, D. J. Chen, *J. Cell Biol.* **2007**, *177*, 219.
- [30] M. Falzon, J. W. Fewell, E. L. Kuff, *J. Biol. Chem.* **1993**, *268*, 10546.
- [31] Y. Ma, U. Pannicke, K. Schwarz, M. R. Lieber, *Cell* **2002**, *108*, 781.
- [32] M. S. Schlissel, *Cell* **2002**, *109*, 1.
- [33] A. Lamaa, M. Le Bras, N. Skuli, S. Britton, P. Frit, P. Calsou, H. Prats, A. Cammas, S. Millevoi, *EMBO Rep.* **2016**, *17*, 508.
- [34] T. S. Fisher, A. K. P. Taggart, V. A. Zakian, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1198.
- [35] N. S. Y. Ting, *Nucleic Acids Res.* **2005**, *33*, 2090.

- [36] H. Chen, J. Xue, D. Churikov, E. P. Hass, S. Shi, L. D. Lemon, P. Luciano, A. A. Bertuch, D. C. Zappulla, V. Géli, J. Wu, M. Lei, *Cell* **2018**, *172*, 331.
- [37] M. Xing, M. Yang, W. Huo, F. Feng, L. Wei, W. Jiang, S. Ning, Z. Yan, W. Li, Q. Wang, M. Hou, C. Dong, R. Guo, G. Gao, J. Ji, S. Zha, L. Lan, H. Liang, D. Xu, *Nat. Commun.* **2015**, *6*, 6233.
- [38] T. Ochi, A. N. Blackford, J. Coates, S. Jhujh, S. Mehmood, N. Tamura, J. Travers, Q. Wu, V. M. Draviam, C. V. Robinson, T. L. Blundell, S. P. Jackson, *Science* **2015**, *347*, 185.
- [39] A. Craxton, J. Somers, D. Munnur, R. Jukes-Jones, K. Cain, M. Malewicz, *Cell Death Differ.* **2015**, *22*, 890.
- [40] X. Liu, Z. Shao, W. Jiang, B. J. Lee, S. Zha, *Nat. Commun.* **2017**, *8*, 13816.
- [41] S. K. Tadi, C. Tellier-Lebègue, C. Nemoz, P. Drevet, S. Audebert, S. Roy, K. Meek, J.-B. Charbonnier, M. Modesti, *Cell Rep.* **2016**, *17*, 541.
- [42] J. Unno, M. Takagi, J. Piao, M. Sugimoto, F. Honda, D. Maeda, M. Masutani, T. Kiyono, F. Watanabe, T. Morio, H. Teraoka, S. Mizutani, *Cancer Sci.* **2013**, *104*, 703.
- [43] R. Bétous, T. Goullet de Rugy, A. L. Pelegri, S. Queille, J.-P. de Villartay, J.-S. Hoffmann, *PLoS Genet.* **2018**, *14*, DOI 10.1371/journal.pgen.1007541.

Figure 1: NHEJ proteins interact with structured single-stranded DNA.

(A) DNA synthesis reactions were carried out on primed pUC118.ss 10 min in MRC5 cell-free extracts (1), mock-depleted (2) and POLD2/POLD3 depleted extracts (3) complemented with 8 ng, 4 ng or 2 ng of purified human Pol δ . (*): major pausing site of DNA synthesis at the first hairpin encountered by the replication machinery. Lower panel: Western blot analysis of cell-free extracts.

(B) The Venn diagram depicts proteins found enriched on the primed template DNA versus the control (-DNA) in the presence of dNTPs in at least two out of three independent experiments either by Spectral Count (SC) or eXtracted Ion Chromatogram (XIC) methods. Replication-associated proteins found enriched only in the presence of dNTPs are highlighted in bold. The histogram illustrates Gene Ontology enrichment of the proteins found enriched by both methods. GOBP: Gene Ontology biological processes. GOMF: Gene Ontology molecular functions (GOMF).

(C) Log₂(fold change) of NHEJ proteins enriched on the DNA versus the control (-DNA) either in the absence or presence of dNTPs. *: adjusted p-value < 0.05.

(D) Western blot analysis of proteins captured on the DNA during a time course experiment.

(E) Pull down of proteins using 5'-biotinylated single-stranded DNA fragments (153 bases) bound to magnetic streptavidin beads. Captured proteins were analyzed by immunoblotting with antibodies as indicated.

Figure 2: Impact of double-stranded oligonucleotides (dsO) on DNA synthesis in MRC5 cell-free extracts.

(A) DNA synthesis reactions in MRC5 cell-free extracts (20 μg) were carried out 15 min on primed pUC118.ss in the absence or presence of dsO and purified Ku dimer, as indicated. (**): elongation product from the 5' end of the primer up

to the *PvuII* restriction site (306 nucleotides). Quantification of the band corresponding to the major pausing site (*) over the whole elongation products is shown below the gels.

(B) Western blot analysis of the amount of endogenous Ku within MRC5 cell-free extract. Note that the His tag on purified Ku70 reduces its electrophoretic mobility compared to the endogenous protein.

(C) Quantification of replication efficiency in the presence or absence of dsO. Error bars represent S.D. of two experiments.

Figure 3: DNA-PK activation by structured single-stranded DNA in MRC5 cell-free extracts and impact of DNA-PKcs inhibition on primer extension.

(A) MRC5 cell extracts were incubated with or without pUC118.ss, in the absence or presence of DNA-PKcs inhibitor (25 μM). The indicated proteins were analyzed by Western blotting.

(B) and (C) DNA synthesis reactions in MRC5 cell extracts were carried out on primed pUC118.ss in the absence or presence of DNA-PK (DNA-PKi) and ATM (ATMi) inhibitors, as indicated. (*): major pausing site of DNA synthesis.

(>): additional stalling site caused by the DNA-PKcs inhibitor.

Duplicate samples were analyzed by electrophoresis through a 7 M urea / 8% polyacrylamide gel (upper panels) or subjected to immunoblotting with anti-

RPA32 and anti-phospho S4/S8 RPA antibodies, as indicated (lower panels).

(D) Quantification of the additional pausing site (>) over the whole elongation products at 5 and 20 min in the absence and presence of DNA-PK inhibitor.

Error bars represent S.D. of 3 experiments using 3 independent cell extracts.

Figure 4: The DNA-PKcs inhibitor induces the mobilization of NHEJ proteins on the primer template DNA.

(A) Log2(fold change) of NHEJ proteins enriched on DNA versus the control (- DNA) in the absence or presence of DNA-PKcs inhibitor (25 μ M).

: adjusted p-value < 0.05. () : adjusted p-value for XRCC4 = 0.055

(B) Western blot analysis of proteins captured at 10 min on the DNA in the absence or presence of different doses of DNA-PK inhibitor.

Table 1: Proteins enriched on the extending DNA template in the absence or in the presence of dsO

Gene name (Protein name)	Log2(Fold-Change) + DNA / - DNA		Log2(Fold-Change) + DNA +dsO / - DNA	
	SP Table S4a	(XIC) Table S4b	SP Table S5a	(XIC) Table S5b
Replication-associated proteins 1				
PCNA	3.00*	(4.14*)	3.62*	(5.58*)
RFC1	0.93*	(1.10*)	0.90*	(1.27*)
RFC2	0.88*	(1.19*)	0.83	(1.01)
RFA1 (RPA70)	4.53*	(5.30*)	4.32*	(5.07*)
RFA2 (RPA32)	5.16*	(5.39*)	4.94*	(5.05*)
RFA3 (RPA14)	4.61*	(5.82*)	4.22*	(5.25*)
A8K1E1 (MSH3)	3.54*	(1.55*)	4.42*	(2.66*)
CAF1A	6.37*	(1.60*)	7.10*	(2.10*)
CAF1B	2.26*	(2.27*)	3.23*	(1.92*)
RNH2B	4.90*	(4.23*)	6.82*	(6.64*)
DNMT1	1.38*	(1.13*)	2.40*	(1.72*)
Replication-associated proteins 2				
DPOD2 (POLD2 of Pol δ)	1.26	(1.01)	3.17*	(2.12*)
PAF15	4.28	(5.38*)	5.36*	(6.26*)
DNL11 (LIG1)	3.82	(2.54*)	7.49*	(3.83*)
MSH2	0.71	(0.63)	1.34*	(1.27*)
B4E0E6 (DTL/CDT2)	1.86	(4.50*)	4.92*	(4.83*)
RNH2A	3.82	(2.52*)	7.01*	(4.16*)
RNH2C	3.53	(2.83*)	5.62*	(4.33*)

M0QXA7 (WIZ)	2.50* (0.57)	3.44* (1.29*)
DNA Repair proteins		
PRKDC (DNA-PKcs)	0.89* (1.48*)	-0.45 (-0.44)
XRCC6 (Ku70)	0.80* (1.65*)	-0.25 (-0.24)
XRCC5 (Ku80)	0.90* (1.57*)	-0.13 (-0.38)
PAXX	1.86 (3.17*)	n.i n.i
RAD50	1.54* (1.66*)	0.97* (0.87)
MRE11	1.16* (1.35*)	0.64 (0.81)
NBN (NBS1)	1.98* (1.56*)	1.28 (0.93)
RMI1	4.46* (2.94*)	4.09 (2.69*)
RECQ1 (RECQL)	1.17* (1.66*)	0.87* (1.09)
TOP3A	6.06* (2.44*)	6.29* (2.37*)
WRN	5.03* (1.42*)	5.04* (2.68*)
TONSL	5.94* (1.81*)	6.82* (2.17*)
MMS22	4.63* (2.24*)	4.92* (2.85*)
SMAL1 (SMARCA1)	6.62* (2.98*)	6.14* (2.78*)
Dynamins		
DYN2	1.47* (1.89*)	-0.11 (-0.01)
DYN3	1.25* (1.89*)	-0.42 (-0.01)
Heterogenous Ribonucleoproteins		
HNRPD	2.71* (1.96*)	3.20* (2.88*)
A0A087WZV1 (HNRPA/B)	1.62* (1.68*)	2.16* (3.35*)
B2R5W2 (HNRPC1/C2)	1.52* (2.16*)	1.47* (2.23*)
ROA2 (HNRPA2/B1)	0.67 (0.86)	1.12* (1.82*)
HNRPU	0.66* (0.82)	1.41* (2.03*)
A8K4T9 (HRNPC-like)	1.64 (1.96*)	2.05* (1.76*)
HNRDL	2.09 (0.82)	2.70* (1.63*)
Other proteins		
TTF2	1.37* (1.35*)	1.29* (1.11)

RBM3	1.60*	(1.62*)	1.81*	(1.89*)
HM20A	n.i	n.i	4.79*	(2.51*)
PURA	2.58*	(3.21*)	2.26*	(2.86*)
PURB	2.77*	(3.89*)	2.72*	(3.79*)

Bold numbers highlight the proteins found enriched in both SC and XIC. The replication-associated proteins (1) are significantly enriched on the DNA versus the control, whereas the replication-associated proteins (2) are specifically enriched in the presence of dsO. NHEJ proteins are highlighted in grey. *: adjusted p-value < 0.05; n.i.: non identified