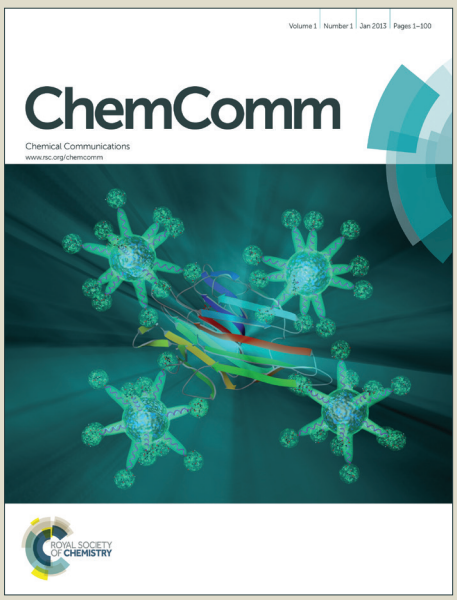
# ChemComm

Accepted Manuscript





ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

# Identification of novel interactors of human telomeric G-quadruplex DNA†

Bruno Pagano,*[a] Luigi Margarucci,[b] Pasquale Zizza,[c] Jussara Amato,[a] Nunzia Iaccarino,[a] Chiara Cassiano,[b] Erica Salvati,[c] Ettore Novellino,[a] Annamaria Biroccio,[c] Agostino Casapullo[b] and Antonio Randazzo[a]

**A chemoproteomic-driven approach was used to investigate the interaction network between human telomeric G-quadruplex DNA and nuclear proteins. We identified novel G-quadruplex binding partners, able to recognize these DNA structures at chromosome ends, suggesting a possible, and so far unknown, role of these proteins in telomere functions.**

Telomeres are DNA-protein complexes at the ends of linear eukaryotic chromosomes. Human telomeric DNA, which consists of tandem repeats of the short TTAGGG sequence, is double-stranded for most of its length, except for a single-stranded protrusion at the 3'-end.[1] Mammalian telomeric DNA is associated to shelterin, a protein complex that protects chromosome ends from being recognized and repaired as double strand breaks and from triggering DNA damage responses.[2] The 3'-end G-rich overhang is also implicated in cancer progression, being the substrate for telomerase, a DNA polymerase that elongates telomeric DNA in cancer cells, leading to cellular immortality.

The telomeric DNA can form a large loop structure, known as telomeric loop,[3] and unusual secondary structures called G-quadruplexes (G4s).[4] These motifs comprise a structure of π-stacked tetrads formed by the coplanar arrangement of four Hoogsteen-paired guanines.[5] The formation of such guanine-based motifs is particularly favoured under physiological conditions, with respect to pH and the presence of metal cations (i.e. $K^+$ and $Na^+$). Telomeric G4s have been shown to have regulatory roles for telomere extension and maintenance.[4] The formation of such structures renders the G-rich single-stranded overhang inaccessible to telomerase, thus inhibiting telomere extension. Interest in the more general significance of G4s has expanded during the past decade to include G4 structures in oncogene promoter sequences,[4,6] 5'-UTR regions and introns,[7] as well as in a number of fragile/breakpoint regions.[8] The broad concept of G4s has been recently validated by their direct visualization in human cells,[9,10] and by the evidence that these structures can be stabilized in cells by small molecules, emerging as a novel approach to cancer therapeutics and other diseases.[6,11]

The analysis of the G4-proteins interaction network can be considered a crucial point to clarify the elusive biological mechanisms in which such relevant DNA structures could be implicated and the consequent potential involvement in the field of drug discovery. Besides telomerase, several proteins have been shown to associate with the unfolded single-stranded telomeric DNA and exert a number of biological functions.[2,12] Some proteins are also able to unfold telomeric G4, playing critical roles in promoting or blocking interactions between telomere and telomerase.[13] The discovery of these proteins raises interesting questions regarding the dynamic nature and function of such structures within the genome, especially at telomeres.

MS-based chemical proteomics is one of the most powerful approaches among the new emerging analytical techniques in the field of target discovery, combining affinity-based chromatography with mass spectrometry and bioinformatics. In this method, the molecule of interest, previously biotinylated or linked to a solid support through a spacer arm, is used as a probe to fish out its interactors from a complex proteome such as a cell lysate or a tissue extract. Once selected and eluted, these cellular targets are identified by high-resolution tandem-MS and bioinformatics analyses.[14] A following validation of the specific interaction profile of the probe molecule provided by the chemical proteomics analysis is usually needed and done through *in vitro* and/or in cell orthogonal biophysical assays.[15]

In this report, we describe the interactome identification, by a chemoproteomic-driven approach, of a truncation of the G4-forming human telomeric DNA (Fig. 1), and the preliminary *in vitro* validation of the main selected targets. Moreover, a complementary biological analysis revealed the *in vivo* localization of these interactions, thus suggesting a so far unknown physiological role of G4 structures in human cells. The 26-mer truncation of human telomeric DNA sequence d(TTAGGGTTAGGGTTAGGGTTAGGGTT) ($tel_{26}$) was chosen to study the interaction between G4-forming telomeric DNA and cellular proteins. Since our chemical proteomic approach took advantage of the biotin-based affinity chromatography, we employed a chemically modified $tel_{26}$ sequence carrying a biotin tag at 5'-end. As generally done in chemoproteomic procedures,

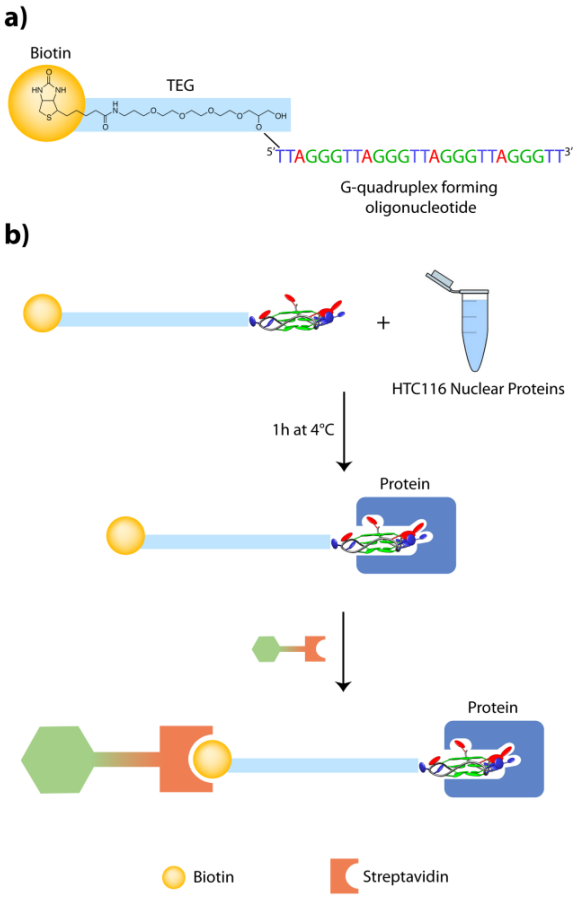



**Fig. 1** a) Structure of biotin-tel$_{26}$ used in this study. b) Schematic overview of chemical proteomics workflow for the identification of G-quadruplex DNA interactors.

a spacer arm was introduced between the G4 and the biotin moiety to avoid steric hindrance that could hamper protein interactions (Fig. 1). The biotinylated scrambled oligonucleotide d(GAAGTGTGTGTGTGTGTGTGTGAA), having a sequence sharing the same length and base composition with tel$_{26}$, but unable to fold into a G4 structure, was used as a control, to exclude proteins with a general DNA-binding ability, as well as proteins binding to G-rich sequences, but not specific for the G4 motif.

The first step was to analyze the G4 structure formed by the biotinylated human telomeric sequence (hereafter referred to as biotin-tel$_{26}$) in comparison with tel$_{26}$, together with the check of the unfolded structure of the biotinylated scrambled oligonucleotide (hereafter referred to as biotin-scr$_{26}$).

As far as telomeric DNA is concerned, the presence of $K^+$, certainly the most relevant cation in the aqueous environment of living cells, induces its folding into a variety of G4 topologies *in vitro*, thus exhibiting a broad structural diversity and polymorphism compared to duplex DNA.[16] Although the exact arrangement of the human telomeric G4 in physiologically relevant conditions has not yet been unambiguously determined, there is evidence that cell-mimicking conditions significantly stabilize the parallel G4 form over the others, regardless of sequence length, thus suggesting the parallel fold as the

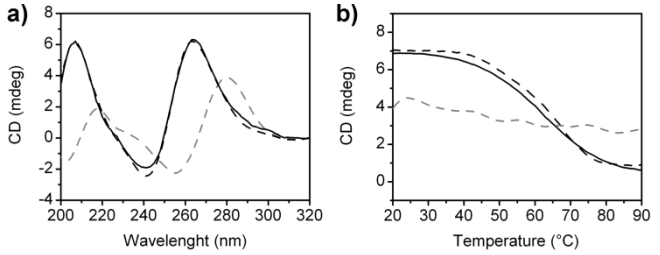



**Fig. 2** a) CD spectra recorded at 20 °C and b) CD melting profiles of tel$_{26}$ (black solid line), biotin-tel$_{26}$ (black dashed line) and biotin-scr$_{26}$ (gray dashed line).

predominant structure in the overcrowded solvent conditions of a cell.[17] Therefore, since it has been shown that a high DNA concentration promotes the parallel G4 folding of human telomeric sequence,[18] we prepared the oligonucleotide samples in high concentration conditions (Experimental Section, ESI†).

To verify that tel$_{26}$ and biotin-tel$_{26}$ actually adopt the expected folding, circular dichroism (CD) experiments were performed (Fig. 2). Indeed, CD provides a useful tool for detecting the presence and the overall topology of a G4 structure.[19] CD spectra of tel$_{26}$ and biotin-tel$_{26}$ showed a positive band at 264 nm and a negative band at around 240 nm indicating, in both cases, the presence of a parallel-stranded G4 structure. On the other hand, the CD spectrum of biotin-scr$_{26}$ lacks these characteristics, revealing that, as expected, it cannot form a G4 structure. The stability of tel$_{26}$ and biotin-tel$_{26}$ G4s was investigated by CD melting experiments. All the thermal denaturations were monitored at the wavelength of maximum CD intensity. In both cases, the melting experiments showed a sigmoidal transition curve indicative of the formation of a G4 structure (Fig. 2). The melting temperatures were found to be 63.5 ± 0.5 °C for tel$_{26}$ and 65.0 ± 0.5 °C for biotin-tel$_{26}$, clearly indicating that the chemical modification does not destabilize the G4. No distinct transition was observed for the biotin-scr$_{26}$ control oligonucleotide, indicating that it does not fold into a G4 structure but remains in a random coil. All CD data recorded for biotin-tel$_{26}$ were in line with those for tel$_{26}$, thus indicating that both oligonucleotides form the same parallel G4 structure under our experimental conditions.

Once characterized the oligonucleotides, the next step was to isolate and identify the proteins binding to the G4-forming human telomeric motif. In order to obtain data with relevant and focused biological significance, the experiments were carried out on nuclear extracts from HCT116 cells (see Experimental Section, ESI†). A sample of the nuclear extracts recovered from cells was incubated with biotin-tel$_{26}$ for 1 h at 4 °C under shaking, allowing the interactions between the G4-forming oligonucleotide and its potential partner(s) to take place. The recovery of the DNA probe was done by the addition of streptavidin-modified beads, taking advantage of the strong and specific affinity between streptavidin and the biotin moiety. After 1 h of incubation, the solid phase was then collected by centrifugation. The beads were extensively washed, with phosphate saline buffer, to remove the weakly bound proteins, while the tightly bound interactors were eluted by boiling the beads with the SDS-PAGE loading buffer. The eluted interactors were finally separated by 1D gel electrophoresis (Fig. S1, ESI†). The gel runs, conveniently stained with Coomassie,

were cut and submitted to in gel digestion with trypsin. The same procedure was applied to a sample of biotin-scr$_{26}$ and of the biotinylated linker alone as control experiments, and all experiments were performed in duplicate.

The peptide mixtures, recovered after trypsin digestion, were analyzed by nano-HPLC-MS/MS and then the resulting MS data, in form of peak lists, were submitted to database search for protein identification. Since we were interested in finding proteins that only bound to the G4-forming DNA, but not to an unfolded DNA sequence (unspecific binders), a protein list containing the potential G4 interactors was obtained after superimposition and subsequent exclusion of the proteins shared by biotin-tel$_{26}$ and controls (biotin-scr$_{26}$ and biotinylated linker). The proteins were then classified on the basis of the number of peptides and Mascot score (Table 1). Among the best hits, since we were looking for proteins never identified previously as G4 binding partners, high mobility group B1 protein (HMGB1), far upstream element-binding protein 2 (KHSRP or FUBP2) and lamin B1 (LMNB1) were selected as candidates for specific binding to the G4. Conversely, XRCC6 (also known as Ku protein), RFA1 (or RPA1) and ROAA (or hnRNP A/B) were discarded since they were already known to interact with telomeric DNA or G-quadruplex-forming sequences.[20]

The following steps were then addressed to confirm a direct interaction between the selected targets and the tel$_{26}$ sequence. First, an immune-blotting analysis was performed to validate the chemical proteomics results, and a relevant enrichment of the selected proteins was evident in the G4 lane when compared to the control experiment (Fig. S2, ESI†). Next, the affinity between the candidate proteins and the non-biotinylated G4-forming tel$_{26}$ oligonucleotide was monitored and measured by SPR analysis, giving calculated dissociation constants of 2.8 × $10^{-6}$, 1.6 × $10^{-6}$, and 1.4 × $10^{-6}$ M for HMGB1, KHSRP and LMNB1 respectively, as evidence of a relevant direct interaction between the counterparts (Fig. S3, ESI†). Notably, the interactions take place without a consequent unwinding of the G4 structure, as shown by CD analysis (Fig. 3 and Fig. S4, ESI†). To the best of our knowledge, this is the first report on the identification of proteins able to recognize the parallel arrangement of telomeric G4.

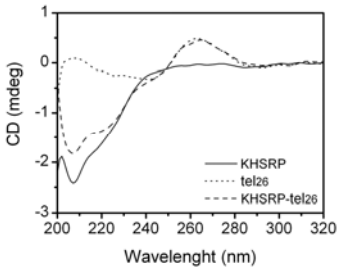



**Fig. 3** Representative CD spectrum of a G4-protein complex recorded at 1:1 molar ratio (50 nM). CD spectra of the protein and tel$_{26}$ alone are also shown.

The association of HMGB1, KHSRP and LMNB1 proteins with the telomeric tract was also evaluated *in vivo* by fluorescence microscopy. Interestingly, we found that HMGB1 and KHSRP proteins were distributed in the nucleus in a punctate pattern and several of these spots colocalized with the telomere repeat binding factor 1 (TRF1) (Fig. 4), a well-established marker for interphase telomeres,[21] and with an anti-G4 antibody[10] recognizing, for more than the 80%, the telomeres (Fig. S5, ESI†). At the same time, analysis of LMNB1 revealed that, in spite of a more diffuse nuclear distribution with a marked

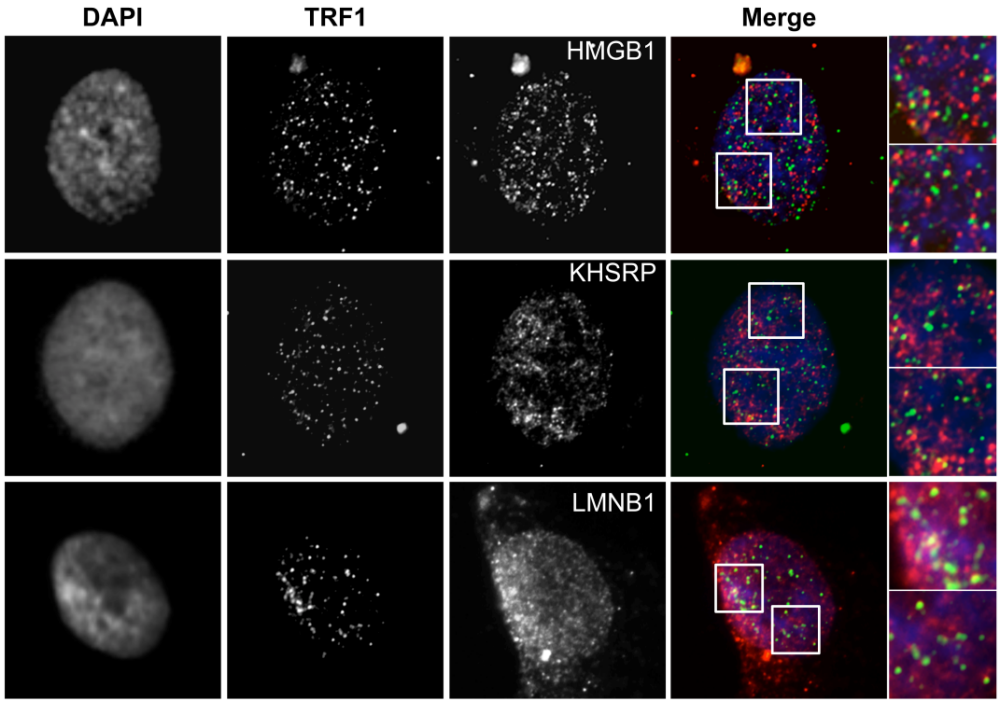



**Fig. 4** HMGB1, KHSRP and LMNB1 colocalize with TRF1 in vivo. Representative IF images acquired by using a Leica Deconvolution microscope (magnification 100x) are shown. Discrete foci deriving from the colocalization of HMGB1 (red, upper panels), KHSRP (red, middle panels) and LMNB1 (red, lower panels) with TRF1 (green) are clearly visible into the nucleus (blue). Enlarged views of colocalization foci are reported on the right of each merged picture.

**Table 1** List of potential binding partners of human telomeric G4[a]

| SwissProt Code | Mass (Da) | Experiment 1 SCR (pep n°) | Experiment 2 SCR (pep n°) |
|---|---|---|---|
| XRCC6_HUMAN | 70084 | 444 (35) | 758 (53) |
| HMGB1_HUMAN | 25049 | 88 (9) | 513 (25) |
| FUBP2_HUMAN | 73355 | 126 (9) | 150 (9) |
| RFA1_HUMAN | 68723 | 67 (7) | 137 (10) |
| LMNB1_HUMAN | 66653 | 64 (6) | 76 (8) |
| ROAA_HUMAN | 36316 | 52 (3) | 115 (10) |
| EHD4_HUMAN | 61365 | 70 (6) | 81 (7) |
| FUBP1_HUMAN | 67690 | 49 (5) | 163 (8) |
| MCM7_HUMAN | 81884 | 53 (5) | 26 (7) |
| SEPT9_HUMAN | 65646 | 41 (3) | 61 (9) |
| PABP2_HUMAN | 32843 | 48 (2) | 50 (7) |
| CPSF6_HUMAN | 59344 | 35 (3) | 97 (5) |
| RPN1_HUMAN | 68641 | 20 (2) | 33 (6) |
| LRC47_HUMAN | 64004 | 19 (2) | 30 (6) |
| ESRP1_HUMAN | 76449 | 42 (3) | 100 (4) |
| RBP56_HUMAN | 62021 | 70 (3) | 39 (4) |
| SARNP_HUMAN | 23713 | 50 (2) | 54 (2) |
| EDF1_HUMAN | 16359 | 49 (1) | 64 (3) |
| MIC1_HUMAN | 75668 | 39 (1) | 41 (1) |

[a] see ESI† for the complete list of proteins identified.

perinuclear staining, also this protein showed detectable nuclear spots colocalizing with telomeres (Fig. 4 and Fig. S5, ESI†). Finally, immunofluorescence analysis performed on cells exposed to RHPS4 showed that telomeric localization of the proteins was not affected by G4 stabilization (Fig. S6, ESI†).

Interestingly, although these proteins have not been associated with interactions to such DNA structure so far, some of them have been supposed to be involved in telomere functions. For example, it was recently shown that knockdown of HMGB1 in mouse embryonic fibroblasts results in a decreased telomerase activity, chromosomal abnormalities and telomere dysfunction.[22] Moreover, the LMNB1 protein, a major structural component of the nucleus, appears to be involved in the regulation of many nuclear functions and it has been shown to have roles in cellular proliferation and senescence, raising important questions on how nuclear lamin might interfere with human telomeres.[23]

In summary, our data identify HMGB1, KHSRP and LMNB1 as three novel G4 interactors able to recognize such DNA motifs at chromosome ends, without unfolding them, suggesting a possible, and so far undescribed, function of these proteins. Preliminary SPR data showed a good specificity of HMGB1 for the telomeric G4 relative to other biologically relevant G4s (Fig. S7, ESI†), though further investigations are required to address this point. The present study ultimately leads toward a more holistic view of the molecular interplay at telomeric level and lays the basis for further studies aiming at elucidate the biological relevance of such interactions.

This work was supported by AIRC (grant nos. 11567, 11947, 14150). We thank P.M. Landscorp for providing anti-G4 antibody.

## Notes and references

[a] *Department of Pharmacy, University of Naples "Federico II", Via D. Montesano 49, 80131 Napoli, Italy. E-mail: bruno.pagano@unina.it.*
[b] *Department of Pharmacy, University of Salerno, Via Giovanni Paolo II 132, 84084 Fisciano (SA), Italy.*
[c] *Experimental Chemotherapy Laboratory, Regina Elena National Cancer Institute, 00158 Rome, Italy.*

† Electronic Supplementary Information (ESI) available: Detailed experimental methods, figures and the complete list of proteins identified. See DOI: 10.1039/c000000x/

1 W. E. Wright, V. M. Tesmer, K. E. Huffman, S. D. Levene and J. W. Shay, *Genes Dev.*, 1997, **11**, 2801; V. L. Makarov, Y. Hirose and J. P. Langmore, *Cell*, 1997, **88**, 657.
2 T. de Lange, *Genes Dev.*, 2005, **19**, 2100.
3 J. D. Griffith, L. Comeau, S. Rosenfield, R. M. Stansel, A. Bianchi, H. Moss and T. de Lange, *Cell*, 1999, **97**, 503.
4 M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770.
5 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402.
6 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discov.*, 2011, **10**, 261.
7 J. L. Huppert, A. Bugaut, S. Kumari and S. Balasubramanian, *Nucleic Acids Res.*, 2008, **36**, 6260; J. Eddy and N. Maizels, *Nucleic Acids Res.*, 2008, **36**, 1321.
8 Y. L. Wu and R. M. Brosh, *FEBS J.*, 2010, **277**, 3470; S. De and F. Michor, *Nat. Struct. Mol. Biol.*, 2011, **18**, 950.
9 G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nature Chem.*, 2013, **5**, 182.
10 A. Henderson, Y. L. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh, D. Sen and P. M. Lansdorp, *Nucleic Acids Res.*, 2014, **42**, 860.
11 S. A. Ohnmacht and S. Neidle, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 2602; E. Salvati, P. Zizza, A. Rizzo, S. Iachettini, C. Cingolani, C. D'Angelo, M. Porru, A. Randazzo, B. Pagano, E. Novellino, M. E. Pisanu, A. Stoppacciaro, F. Spinella, A. Bagnato, E. Gilson, C. Leonetti and A. Biroccio, *Nucleic Acids Res.*, 2014, **42**, 2945.
12 M. J. McEachern, A. Krauskopf and E. H. Blackburn, *Annu. Rev. Genet.*, 2000, **34**, 331.
13 A. J. Zaug, E. R. Podell and T. R. Cech, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 10864; F. Wang, M. L. Tang, Z. X. Zeng, R. Y. Wu, Y. Xue, Y. H. Hao, D. W. Pang, Y. Zhao and Z. Tan, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 20413.
14 L. Margarucci, M. C. Monti, A. Tosco, R. Riccio and A. Casapullo, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 3960; C. Cassiano, L. Margarucci, R. Esposito, R. Riccio, A. Tosco, A. Casapullo and M. C. Monti, *Chem. Commun.*, 2014, **50**, 6043.
15 U. Rix and G. Superti-Furga, *Nat. Chem. Biol.*, 2009, **5**, 616; U. Rix, M. Gridling and G. Superti-Furga, *Methods Mol. Biol.*, 2012, **803**, 25.
16 J. X. Dai, M. Carver and D. Z. Yang, *Biochimie*, 2008, **90**, 1172.
17 G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876; H. Q. Yu, X. B. Gu, S. Nakano, D. Miyoshi and N. Sugimoto, *J. Am. Chem. Soc.*, 2012, **134**, 20060.
18 D. Renciuk, I. Kejnovska, P. Skolakova, K. Bednarova, J. Motlova and M. Vorlickova, *Nucleic Acids Res.*, 2009, **37**, 6625.
19 A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo and M. Webba da Silva, *Angew. Chem. Int. Ed.*, 2011, **50**, 10645; A. Randazzo, G. P. Spada and M. Webba da Silva, *Top. Curr. Chem.*, 2013, **330**, 67.
20 T. R. Salas, I. Petruseva, O. Lavrik, A. Bourdoncle, J. L. Mergny, A. Favre and C. Saintomé, *Nucleic Acids Res.*, 2006, **34**, 4857; S. Cogoi, M. Paramasivam, B. Spolaore and L. E. Xodo, *Nucleic Acids Res.*, 2008, **36**, 3765; Y. He and R. Smith, *Cell. Mol. Life Sci.*, 2009, **66**, 1239; C. Sissi, B. Gatto and M. Palumbo, *Biochimie*, 2011, **93**, 1219; A. Prakash, A. Natarajan, L. A. Marky, M. M. Ouellette and G. E. Borgstahl, *J. Nucleic Acids*, 2011, **2011**, 896947; T. Zhang, H. Zhang, Y. Wang and L. B. McGown, *Anal. Bioanal. Chem.*, 2012, **404**, 1867.
21 C. Thiriet and J. J. Hayes, *Mol. Cell*, 2005, **18**, 617.
22 E. Polanska, Z. Dobsakova, M. Dvorackova, J. Fajkus and M. Stros, *Chromosoma*, 2012, **121**, 419.
23 T. Shimi, V. Butin-Israeli, S. A. Adam, R. B. Hamanaka, A. E. Goldman, C. A. Lucas, D. K. Shumaker, S. T. Kosak, N. S. Chandel and R. D. Goldman, *Genes Dev.*, 2011, **25**, 2579; O. Dreesen, A. Chojnowski, P. F. Ong, T. Y. Zhao, J. E. Common, D. Lunny, E. B. Lane, S. J. Lee, L. A. Vardy, C. L. Stewart and A. Colman, *J. Cell Biol.*, 2013, **200**, 605.

