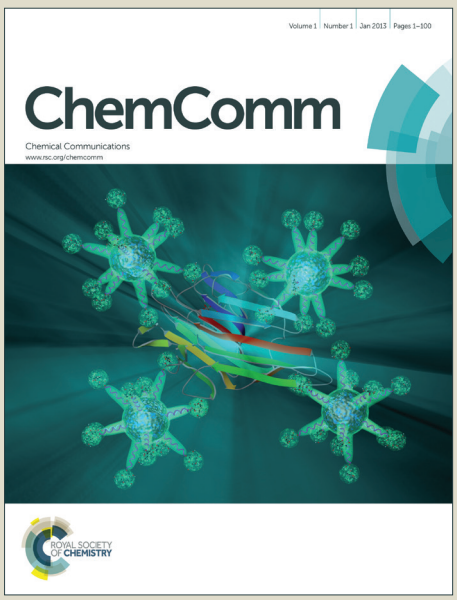
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

ChemComm

RSCPublishing

COMMUNICATION

Yeast transcription co-activator Sub1 and its human homolog PC4 preferentially bind to G-quadruplex DNA

Cite this: DOI: 10.1039/xoxxoooooox

Received ooth January 2015,
Accepted ooth January 2015

DOI: 10.1039/xoxxoooooox

www.rsc.org/

Jun Gao[§a], Boris L. Zybailov[§a], Alicia K. Byrd[a], Wezley C. Griffin[a], Shubeena Chib[a], Samuel G. Mackintosh[a], Alan J. Tackett[a], and Kevin D. Raney[*a]

Using a G-quadruplex bait, we identified the transcription co-activator Sub1 as a G-quadruplex binding protein by quantitative LC-MS/MS and demonstrated *in vivo* G-quadruplex binding by ChIP. *In vitro*, Sub1, and its human homolog PC4, bind preferentially to G-quadruplexes. This provides a possible mechanism by which G-quadruplexes can influence gene transcription.

G-quadruplexes (G4) are highly stable four-stranded nucleic acid structures, consisting of stacked planar guanine tetrads held together through Hoogsteen hydrogen bonds [1,2]. G4 structures are known to be stabilized by large monovalent cations, preferably potassium [3]. G4-DNA structures were recently detected in cells [4,5]. Computational analysis showed that G4-DNA sequences are predominantly located within telomeric regions and promoter regions (within 1kb upstream of the transcription start site, TSS) [6,7], indicating strongly that G4 structures have roles in the regulation of gene expression. Recently, several proteins have been found to bind tightly to G4-DNA including the Pif1 helicase [8], human helicases XPB and XPD [9], and RNA helicase RHAU [10]. Nucleolin [11,12] and SP1 [13] bind to G4-DNA and may regulate transcription from specific genes. Genome-wide regulation of transcription by G4DNA is supported by transcription profiling in the presence of G4DNA binding ligands[14,15,16]. Herein, we report that the yeast multifunctional, chromatin associated protein Sub1, and its human homolog PC4 [17] preferentially bind to G4-DNA.

We performed affinity purification from sonicated *Saccharomyces cerevisiae* whole cell lysate by using a G4-DNA sequence conjugated to biotin and bound to streptavidin Dynabeads (Figure 1a, 1b). For comparison, we used an oligonucleotide composed of 15 thymidine residues (ssDNA). The isolated proteins were separated by SDS-PAGE followed by mass spectrometric analysis. Spectral-count comparative proteomics [18] was applied to establish the level of significance and enrichment of each protein between two sets of affinity purified samples. Figure 1c shows a representative MS/MS spectrum of a peptide originating from the sample isolated by the G4-DNA pull-down. With peptide sequence coverage of 52% (Table S1 in ESI†), we identified the yeast transcription co-activator Sub1 as one of the major G4-DNA binding proteins. Selectivity of Sub1 for G4-DNA was supported by 24 ± 2 spectral counts in the G4-DNA affinity purified samples compared to zero spectral counts in each of the ssDNA affinity purified samples (Figure 1d). In addition, we identified G4p1 (also named as Arc1p) as a G4-DNA binding protein (Table S2 in ESI†). G4p1/Arc1p has been reported previously to have affinity for G4 nucleic acids[19], demonstrating the reliability of our approach.

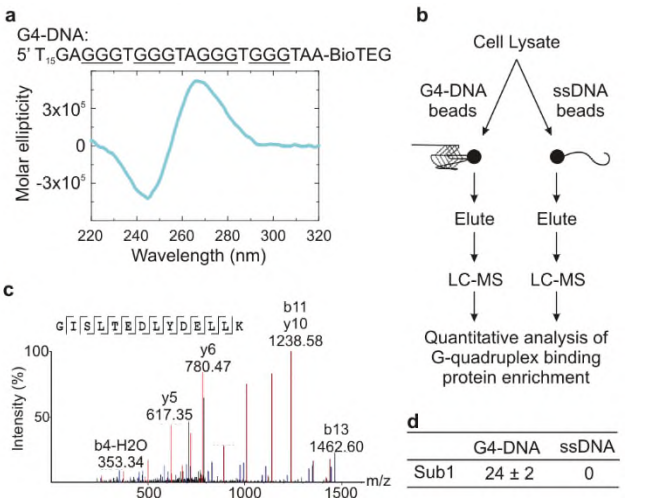


Figure 1. Identification of Sub1 as a G4-DNA binding protein. a) DNA sequence of the biotin labelled G4-DNA bait and circular dichroism spectrum of this sequence. The minimum near 240 nm and the maximum near 260 nm suggest the formation of a parallel G4-DNA. b) Diagram of affinity purification and experimental strategies. c) Representative peptide MS/MS spectrum (m/z=804.92, z=2, RT=33.06, ppm=0.4) of Sub1 from LC-MS/MS analysis. d) Spectral counts representing Sub1 in LC-MS/MS analysis, values are average ± errors calculated from duplicate experiments. Calculated P-value = 0.0001.

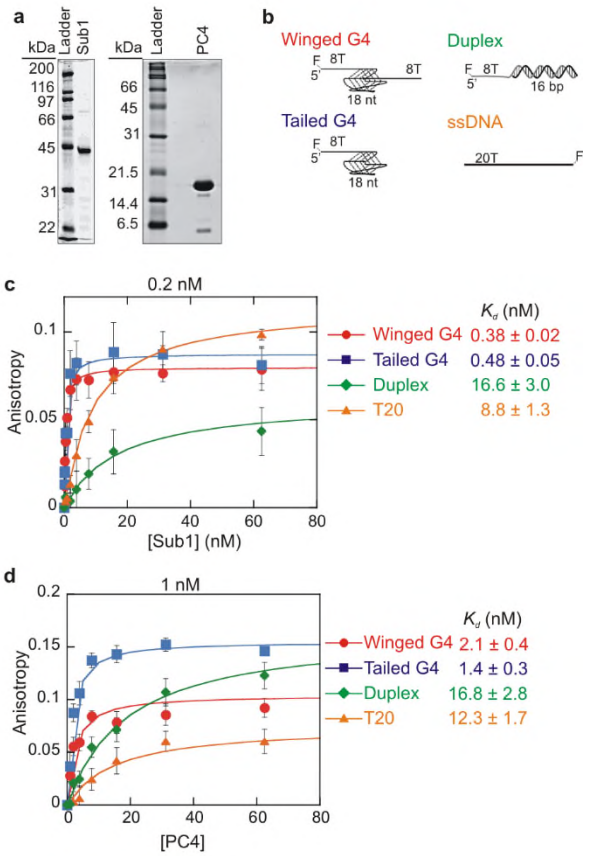


Figure 2. Sub1 and PC4 bind to G4-DNA substrates more tightly than to duplex DNA and ssDNA *in vitro*. a) SDS-PAGE of purified recombinant Sub1 and PC4 proteins. b) Diagram of DNA substrates. Sequences are listed in the experimental methods. c) Recombinant Sub1 binding to DNA substrates. d) Recombinant PC4 binding to DNA substrates. The K_d values (mean ± standard derivation) were determined by averaging the anisotropy data from three experiments and fitting data to the quadratic equation.

Next, we over-expressed and purified recombinant Sub1 protein from *E. coli*, and studied its binding affinity to different conformations of DNA (Figure 2a, 2b). Sub1 binds to G4-DNA very tightly with a dissociation constant (K_d) value of 0.38 ± 0.02 nM for the winged G4-DNA, and 0.48 ± 0.05 nM for the tailed G4-DNA (Figure 2c). Binding to G4-DNA is around 20-fold tighter than binding to the ssDNA (K_d value of 8.8 ± 1.3 nM), and 40 fold tighter than its binding affinity for the tailed duplex DNA (K_d value of 16.6 ± 3.0 nM). These results demonstrate that Sub1 preferentially binds to G4-DNA *in vitro*.

Sub1 is a suppressor of TFIIB mutations [20], and has strong homology to the human multifunctional transcription positive co-activator 4 (PC4) [21,22]. We overexpressed and purified recombinant PC4 protein from *E. coli*, and studied its binding to different forms of DNA (Figure 2a, 2d). We found that PC4 binds to G4-DNA with low-nanomolar affinity (K_d value of 2.1 ± 0.4 nM for the winged G4-DNA, and K_d value of 1.4 ± 0.3 nM for the tailed G4-DNA), which is about 5-9 fold tighter than its binding to the ssDNA (K_d value of 12.3 ± 1.7 nM), and 8-12 fold tighter than its binding to the tailed duplex DNA (K_d value of 16.8 ± 2.8 nM). These results demonstrate that PC4, like Sub1, binds to G4-DNA preferentially over ssDNA and duplex DNA *in vitro*.

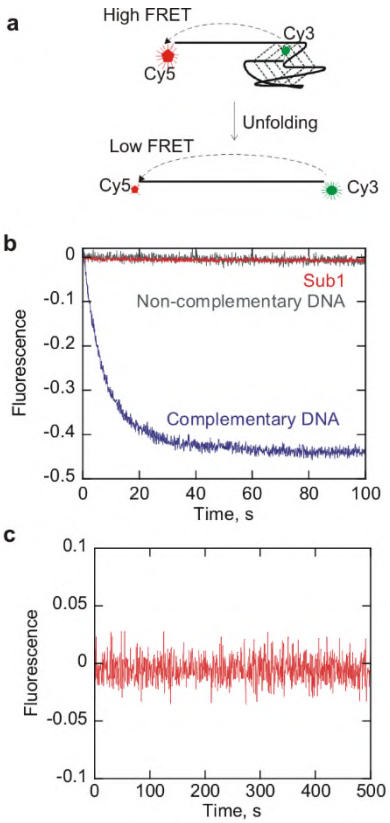


Figure 3. Sub1 binding does not cause G4 unfolding. a) 3'Cy3-5'Cy5 tailed G4-DNA is in a high FRET state when the G4 is folded; upon unfolding, FRET decreases. b) Binding of 400 nM Sub1 to 25 nM G4-DNA does not result in a FRET decrease. The complementary oligonucleotide unfolds the G4-DNA, while the non-complementary oligonucleotide does not. c) FRET remains stable when the incubation time of Sub1 with the G4-DNA is extended.

We then tested whether the binding of Sub1 unwinds a G4-DNA substrate, by monitoring the FRET in a 3'Cy3-5'Cy5 tailed G4 DNA substrate (Figure 3a). As reported previously, the G4-DNA structure can be disrupted by a complementary oligonucleotide [23], observed by the decrease in FRET upon the addition of the complementary oligonucleotide (Figure 3b). No change in FRET was observed with a non-complementary oligonucleotide or Sub1 (Figure 3b), indicating that the G4-DNA remains stable upon Sub1 binding, even with an extended incubation time (Figure 3c). This demonstrates that binding of Sub1 to G4-DNA does not destabilize the quadruplex.

Sub1 interacts with RNA polymerase II [24,25] and III [26] and plays an essential regulatory role in transcription initiation, elongation, and mRNA 3'-end formation [24,27,28]. We therefore tested whether Sub1 can bind to a G4 region on chromosomal DNA *in vivo*. Examination of the genome-wide localization data for Sub1 [26] suggests that the protein does not localize exclusively to G4-forming sequences, consistent with its tight binding to ssDNA and dsDNA. However, we identified several Sub1 binding loci that contain predicted G4-forming sequences [15]. One such region is the YDR544C locus near the telomeric region in the yeast Chromosome IV (Figure 4a, 4b). A yeast cell line expressing protein A-tagged Sub1 (SUB1-TAP) was used to perform quantitative PCR of chromatin immuno-precipitated (ChIP) DNA for comparison with untagged BY4741 cells (control). SUB1-TAP yielded a two-fold enrichment within 200 bp of the G4-region at the YDR544C locus (Figure 4c), demonstrating

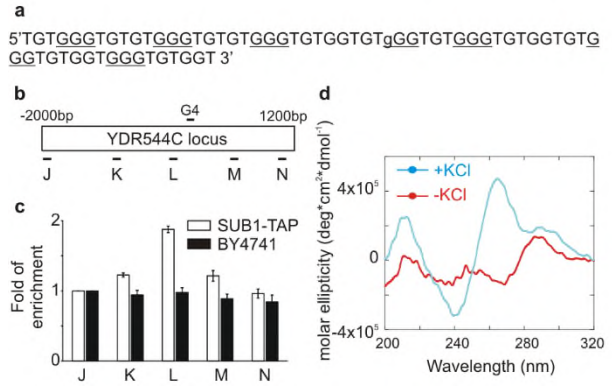


Figure 4. Sub1 preferentially binds to G4-DNA *in vivo*. a) DNA sequence of a G4 region in the yeast YDR544C locus. The lower case "g" is the position "zero" for measuring relative distance of the qPCR regions in this locus. b) Diagram of YDR544C locus and qPCR analysis regions. The relative positions of DNA labelled as J, K, L, M, and N are located upstream 2kb, upstream 1kb, upstream 200bp, downstream 500bp, and downstream 1.1kb respectively. c) qPCR analysis of ChIP-DNA samples from TAP-tagged (SUB1-TAP) and BY4741 (control) cells. Data (mean ± standard derivation) were obtained from triplicate experiments. d) The oligonucleotide from panel a folds into G-quadruplex structures *in vitro* in the presence of KCl. The peaks at both 260 nm and 290 nm suggest that either a mixture of parallel and antiparallel G4 forms or the structure is a hybrid G4 conformation.

that Sub1 preferentially binds to this region *in vivo*. A synthetic oligonucleotide identical to the sequence of this region (Figure 4a) was found to form a G4 DNA structure *in vitro* in the presence of KCl (Figure 4d). These results suggest that Sub1 can bind to endogenous G-quadruplexes *in vivo*.

In conclusion, we have shown that yeast Sub1 and its human homolog PC4 preferentially bind to G4 DNA *in vitro*. Binding of Sub1 to G4 DNA does not destabilize the G4 DNA structure. By using a targeted genome localization experiment, we also demonstrated that Sub1 can bind to a G4 DNA sequence *in vivo*. Both Sub1 and PC4 are global modulators of RNA polymerase II and III transcription with important roles in transcription initiation, elongation and termination [17]. In addition, each protein is highly abundant and multifunctional with proposed roles in DNA repair [29], replication [30], and chromatin condensation [31]. By defining G-quadruplexes as key genomic targets of these factors, our data provides a protein-based mechanism by which G4-DNA structures can broadly influence regulation of gene transcription and other DNA metabolic processes.

This work was supported by grants from the National Institutes of Health (R01 GM098922), and the UAMS Research Council. Core facilities were supported in part by National Institutes of Health grants (R01 GM106024, P30 GM103450, P20 GM103429, and UL1TR000039). We thank Dr. Sebastiaan Werten for the pET11a-PC4 plasmid, Dr. Galina Grazko and Dr. Christine Conesa for helpful discussions regarding the genome-wide binding data for Sub1, Dr. Amit Ketkar and Dr. Stephanie Byrum for technical consultation.

Notes and references

[a] Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, 4301 West Markham Street, Slot 516, Little Rock, AR 72205, U.S.A.

§ Both authors contributed equally to this work.

* Correspondence and requests for materials should be addressed to Kevin D. Raney, Email: RaneyKevinD@uams.edu, Telephone: +1 (501) 686-5244, Fax: +1 (501) 686-8169.

† Electronic Supplementary Information (ESI) available: Detailed experimental methods, and tables. See DOI: 10.1039/c000000x/

1 G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.
2 M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770–780.
3 S. Haider, G. N. Parkinson and S. Neidle, *J. Mol. Biol.*, 2002, **320**, 189–200.
4 G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182–186.
5 A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh Jr, D. Sen and P. M. Lansdorp, *Nucleic Acids Res.*, 2014, **42**, 860–869.
6 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2007, **35**, 406–413.
7 N. Maizels and L. T. Gray, *PLoS Genet.*, 2013, **9**, e1003468.
8 K. Paeschke, M. L. Bochman, P. D. Garcia, P. Cejka, K. L. Friedman, S. C. Kowalczykowski and V. A. Zakian, *Nature*, 2013, **497**, 458–462.
9 L. T. Gray, A. C. Vallur, J. Eddy and N. Maizels, *Nat. Chem. Biol.*, 2014, **10**, 313–318.
10 J. P. Vaughn, S. D. Creacy, E. D. Routh, C. Joyner-Butt, G. S. Jenkins, S. Pauli, Y. Nagamine and S. A. Akman, *J. Biol. Chem.*, 2005, **280**, 38117–38120.
11 L. A. Hanakahi, H. Sun and N. Maizels, *J. Biol. Chem.*, 1999, **274**, 15908–15912.
12 L. A. Dempsey, H. Sun, L. A. Hanakahi and N. Maizels, *J. Biol. Chem.*, 1999, **274**, 1066–1071.
13 E.-A. Raiber, R. Kranaster, E. Lam, M. Nikan and S. Balasubramanian, *Nucleic Acids Res.*, 2012, **40**, 1499–1508.
14 P. Kumar, V. K. Yadav, A. Baral, P. Kumar, D. Saha and S. Chowdhury, *Nucleic Acids Res.*, 2011, **39**, 8005–8016.
15 S. G. Hershman, Q. Chen, J. Y. Lee, M. L. Kozak, P. Yue, L.-S. Wang and F. B. Johnson, *Nucleic Acids Res.*, 2008, **36**, 144–156.
16 H. Fernando, S. Sewitz, J. Darot, S. Tavaré, J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2009, **37**, 6716–6722.
17 C. Conesa and J. Acker, *RNA Biol.*, 2010, **7**, 287–290.
18 B. Zybailov, G. Friso, J. Kim, A. Rudella, V. R. Rodríguez, Y. Asakura, Q. Sun and K. J. van Wijk, *Mol. Cell. Proteomics*, 2009, **8**, 1789–1810.
19 J. D. Frantz and W. Gilbert, *J. Biol. Chem.*, 1995, **270**, 20692–20697.
20 R. Knaus, R. Pollock and L. Guarente, *EMBO J.*, 1996, **15**, 1933–1940.
21 N. L. Henry, D. A. Bushnell and R. D. Kornberg, *J. Biol. Chem.*, 1996, **271**, 21842–21847.
22 S. Werten and D. Moras, *Nat. Struct. Mol. Biol.*, 2006, **13**, 181–182.
23 M. K. Raghuraman and T. R. Cech, *Nucleic Acids Res.*, 1990, **18**, 4543–4552.
24 T. W. Sikorski, S. B. Ficarro, J. Holik, T. Kim, O. J. Rando, J. A. Marto and S. Buratowski, *Mol. Cell*, 2011, **44**, 397–409.
25 A. García, E. Rosonina, J. L. Manley and O. Calvo, *Mol. Cell. Biol.*, 2010, **30**, 5180–5193.
26 A. Tavenet, A. Suleau, G. Dubreuil, R. Ferrari, C. Ducrot, M. Michaut, J.-C. Aude, G. Dieci, O. Lefebvre, C. Conesa and J. Acker, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 14265–14270.
27 A. García, A. Collin and O. Calvo, *Mol. Biol. Cell*, 2012, **23**, 4297–4312.
28 E. Rosonina, I. M. Willis and J. L. Manley, *Mol. Cell. Biol.*, 2009, **29**, 2308–2321.
29 K. Batta, M. Yokokawa, K. Takeyasu and T. K. Kundu, *J. Mol. Biol.*, 2009, **385**, 788–799.
30 Z. Q. Pan, H. Ge, A. A. Amin and J. Hurwitz, *J. Biol. Chem.*, 1996, **271**, 22111–22116.
31 C. Das, K. Hizume, K. Batta, B. R. P. Kumar, S. S. Gadad, S. Ganguly, S. Lorain, A. Verreault, P. P. Sadhale, K. Takeyasu and T. K. Kundu, *Mol. Cell. Biol.*, 2006, **26**, 8303–8315.