

Electronic Supplementary Information (ESI) for the manuscript

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

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}

^a Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, 4301 West Markham Street, Slot 516, Little Rock, AR 72205, USA.

[§] 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.

Experimental Methods

Yeast Strains --- the *S. cerevisiae* strains YPH499 (*MATa*, *ura3-52*, *lys2-801*, *ade2-101*, *trp1-D63*, *his3-D200* *leu2-D1*) was purchased from ATCC (ATCC204679), RSC0045 (*SUB1-TAP* in *BY4741* background) was purchased from Open Biosystems (YSC1177), and *BY4741* wild type strain was purchased from ATCC.

DNA-conjugated Dynabeads Preparation--- 3'-biotinylated oligonucleotide (G4: TTT TTT TTT TT TGA GGG TGG GTA GGG TGG GTAA-BioTEG), and 3'-biotinylated single-stranded DNA T₁₅-BioTEG were purchased from Integrated DNA Technologies. Gel-purified G4 oligonucleotide was folded into the G-quadruplex DNA structure at 60°C overnight in the presence

of 10 mM Tris pH 7.5 and 100 mM KCl. The successful folding into G-quadruplex DNA structure was confirmed by circular dichroism using a JASCO 715 Spectropolarimeter. The folded G4, and the control T₁₅ oligonucleotide were conjugated to Streptavidin-Coupled M-280 Dynabeads[®] (Life Technologies) per the manufacturer's instructions. 1 mg of the beads was washed with 10 mM Tris pH 7.5, 1 mM EDTA, 300 mM KCl and resuspended to the final concentration of 5 mg/ml. Next, on equal volume of 4 μM biotinylated DNA, 10 mM Tris pH 7.5 with 100 mM KCl was added, followed by incubation at room temperature for 20 min with gentle agitation. The beads with the bound DNA were washed 3 times with 5 mM Tris pH 7.5, 0.5 mM EDTA, 150 mM KCl.

DNA-Dynabeads Affinity Purification of Proteins for Mass Spectrometry Analysis ---

Yeast YPH499 wild-type cells were grown in 2 liters YPG media at 30°C to OD₆₀₀ near 1.5. Cells were harvested by centrifugation at 2600x g for 20 minutes and washed once with ice-cold water. Cells were re-suspended in 1/10 (volume/cell weight) of ice-cold 20 mM Hepes pH 7.5 then snap-frozen in liquid nitrogen. Frozen cells were subjected to cryogenic grinding using a Retsch MM301 Mill at the frequency of 30 per second, 3 minutes per round, for 6 rounds. Ground cell powders were re-suspended in IP buffer [20 mM Tris, pH 7.5, 300 mM KCl, 0.1% Tween-20, with a protease inhibitor cocktail (Sigma) added] at 5 ml per gram of cells, then subjected to sonication. Sonicated samples were centrifuged at 2600x g at 4 °C for 10 minutes. DNA-conjugated Dynabeads[®] were washed once with the 1 ml of the IP buffer. The yeast cell lysates (5 ml per 1 mg of beads) were incubated with the DNA-conjugated Dynabeads in 15-ml tubes at 4 °C for 4 hours with gentle inversion. Beads were collected by magnetic rack, washed using the IP buffer 4 times, and eluted by Laemmli Sample Buffer (BioRad) with boiling for 5 minutes. Samples were resolved by 4-15% Mini-PROTEAN TGX precast Gels (BioRad), then subjected to LC-MS/MS analysis.

LC-MS/MS Analysis (including the bioinformatics data analysis) --- Protein gel bands were excised and subjected to in-gel trypsin digestion as follows. Gel slices were destained in 50% methanol, 100 mM ammonium bicarbonate, followed by reduction in 10 mM Tris [2-carboxyethyl] phosphine and alkylation in 50 mM iodoacetamide. Gel slices were then dehydrated in acetonitrile, followed by addition of 100 ng porcine trypsin in 100 mM ammonium bicarbonate and incubation at 37 °C for 12-16 hours. Peptide products were acidified in 0.1% formic acid. Tryptic peptides were separated by reverse phase Jupiter Proteo resin (Phenomenex) on a 100 x 0.075 mm column using a nanoAcquity UPLC system (Waters). Peptides were eluted using a 30 min gradient from 98:2 to 40:60 buffer A:B ratio. [Buffer A = 0.1% formic acid, 0.05% acetonitrile; buffer B = 0.1% formic acid, 75% acetonitrile.] Eluted peptides were ionized by electrospray (1.9 kV) followed by MS/MS analysis using collision induced dissociation on an LTQ Orbitrap Velos mass spectrometer (Thermo). MS data were acquired using the FTMS analyzer in profile mode at a resolution of 60,000 over a range of 375 to 1500 m/z. MS/MS data were acquired for the top 15 peaks from each MS scan using the ion trap analyzer in centroid mode and normal mass range with normalized collision energy of 35.0. Proteins were identified by database search using PEAKS v7 (Bioinformatics Solutions Inc.). Using the InChorus feature within the PEAKS v7 software suit, results from three different search engines were integrated – PEAKS, Mascot, and OMSSA. The filter thresholds were set such so that the overall peptide-to-spectrum match false positive rate was below 1%.

Identification of G4-interacting Proteins Using Spectral Counting --- For the purpose of determining confident G4 DNA interactors, we used a spectral counting approach^{1,2} and G-test of independence³ comparing spectral counts in the G4 DNA-bound sample, to the proteins in the T₁₅-bound sample. Two biological replicates were analyzed. Proteins that were significant at 5%

significance level (after Benjamini-Hochberg correction for multiple testing) in each of the two replicates were considered as strong G4-binders.

pSUMO-SUB1 Plasmid Construction --- Oligonucleotides (5'-ATT GCA CGT CTC AAG GTA TGT CAT ATT ACA ACA GGT ATA GGA ACA AAA G-3' and 5'-ACA ACG CGT CTC CTC GAG TTA CTA TTA TTC TTC TTC ACT TAT GTC GTC TTC AGC CTT G-3') were used for PCR using purified S288C yeast genomic DNA as the template. The PCR product was purified by Wizard SV PCR clean-up kit (Promega), and subjected to BsmB-I (New England Biolabs) digestion at 55 °C overnight. The digested PCR fragment was gel purified by Wizard SV Gel Extraction Kit (Promega), and ligated to BsaI and XhoI linearized pSUMO-T7-Kan vector (Lifesensors Inc.) at 16 °C overnight. Ligation product was transformed into DH5 α Competent *E. coli* (MCLAB), and plated on the LB agar plates containing 50 μ g/ml kanamycin for selection. Plasmids were purified by using QIAprep Spin Miniprep kit (QIAGEN), and confirmed by DNA sequencing at UAMS DNA Sequencing core facility.

Recombinant Sub1 Protein Purification --- Purified pSUMO-SUB1 plasmid was transformed into RosettaTM 2 (DE3) Competent *E. coli* (EMD Millipore) and plated on LB agar plates containing 50 μ g/ml kanamycin and 25 μ g/ml chloramphenicol for selection at 37 °C overnight. Clones were inoculated into 100 ml NZCYM media containing 50 μ g/ml kanamycin and 25 μ g/ml chloramphenicol, and grown at 37 °C to an OD₆₀₀ around 1. Cultures then were transferred into 6 liters of auto-induction media⁴ containing 100 μ g/ml kanamycin and 100 μ g/ml chloramphenicol and grown at 37 °C to OD₆₀₀ = 0.8, then auto-induced at 25 °C overnight. Cells were harvested by centrifugation at 3000x *g* for 30 minutes at 4 °C. All the subsequent steps in this protein purification were at 4°C. Cell pellets were re-suspended and homogenized in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 0.1 g/ml lysozyme, 1

mM PMSF, 2.5 mM β -mercaptoethanol, pH 8.0) at the volume of 10 ml per gram of cells, passed through a Microfluidizer (Microfluidics Inc.), followed by sonication. Cell lysates were pelleted by ultra-centrifugation at 148,000x g for 90 minutes. Supernant was loaded onto a 20 ml Ni-NTA Agarose (MCLAB) column, washed with 400 ml wash buffer (50 mM NaH_2PO_4 , 300 mM NaCl, 20 mM imidazole, 10% glycerol, 2.5 mM β -mercaptoethanol, pH 8.0), followed by elution and fractionation using imidazole buffer (50 mM NaH_2PO_4 , 300 mM NaCl, 250 mM imidazole, 10% glycerol, 2.5 mM β -mercaptoethanol, pH 8.0). Samples from each elution fraction were analyzed by 10% SDS-PAGE. Fractions containing SUMO-Sub1 were pooled, and dialyzed to 50 mM NaH_2PO_4 , 300 mM NaCl, 10 mM imidazole, 10% glycerol, 2.5 mM β -mercaptoethanol, pH 8.0. During dialysis, Ulp1 protease was added to cleave the SUMO tag. The completion of Ulp1 digestion was confirmed by 10% SDS-PAGE. The Ulp1 treated samples were loaded into 20 ml Ni-NTA agarose column (MCLAB), and the flow-through was collected by fractionator. Samples from each fraction were analyzed by 10% SDS-PAGE. Fractions containing recombinant Sub1 were pooled and diluted into HQ buffer (20 mM NaH_2PO_4 , 75 mM NaCl, 10% glycerol, 2.5 mM β -mercaptoethanol, pH 8.0). Samples were loaded onto a 20 ml Macro-Prep[®] High-Q (Bio-Rad) column, washed with 240 ml of HQ buffer, and eluted with a linear salt gradient from 75mM NaCl to 1M NaCl in HQ buffer. Fractions containing purified recombinant Sub1 were pooled and concentrated by centrifugation using Amicon Ultracel Centrifugal Filters (Millipore), followed by dialysis into storage buffer (20 mM Hepes pH 7.5, 150 mM KCl, 1 mM EDTA, 5 mM DTT, 20% glycerol).

Recombinant PC4 Protein Purification --- The pET11a-PC4 plasmid was a generous gift provided from Dr. Sebastiaan Werten⁵. Recombinant PC4 protein was induced and purified as previously described⁵.

Fluorescence Anisotropy Binding Assay --- All oligonucleotides were gel purified ⁶. G4 DNA was prepared as described above for the DNA conjugated Dynabead preparation. The duplex DNA substrate was formed by mixing equal quantities of the individual strands, heating the sample to 95 °C for 5 minutes, and slow cooling to room temperature. Following duplex formation, the sample was purified by native PAGE. Oligonucleotide sequences were T₂₀: 5'-T₂₀-FAM-3', duplex: 5'-FAM-T₈CGCTGATGTCGCCTGG-3' annealed to: 5'-CCAGGCGACATCAGCG-3', tailed G4: 5'-FAM-T₉GGGTGGGTAGGGTGGGT-3', and winged G4 5'-FAM-T₉GGGTGGGTAGGGTGGGT₉-3'. DNA, at the concentration indicated in the figure legends, was incubated in reaction buffer (10 mM Tris, pH 7.5, 200 mM KCl, 1 mM EDTA, 1 mM dithiothreitol, 0.01 mg/ml bovine serum albumin) with increasing concentrations of protein for 30 min. Anisotropy was measured using a Perkin Elmer 1420 Multilabel Counter Victor³V, and the data was fit to the quadratic equation to determine the K_d value using KaleidaGraph software.

G4-DNA unwinding --- All concentrations are final. 25 nM 3'Cy3-5'Cy5 tailed G4 (5'-Cy5-TTT TTT TTT TTT TTT GAG GGT GGG TAG GGT GGG TAA-Cy3-3') was mixed with 400 nM Sub1 in reaction buffer (in 25 mM Hepes pH 7.5, 50 mM KCl, 2 mM β -mercaptoethanol, 10 mM MgCl₂) in an SX.18MV stopped flow reaction analyzer (Applied Photophysics). The assay mixture was excited at 550 nm and FRET was measured after a 665 nm cut-on filter (Newport Corporation, #51330). The control reactions were performed by mixing 25 nM 3'Cy3-5'Cy5 tailed G4 in reaction buffer with 10 μ M oligonucleotide complementary to the G4 forming region (5'-TTA CCC ACC CTA CCC ACC CTC A-3') or 10 μ M non-complementary oligonucleotide (5'-TTA CAC ACA CTA CAC ACA CTC A-3').

Chromatin Immuno-precipitation (ChIP) --- SUB1-TAP (Sub1 containing a C-terminal tandem affinity purification tag) or BY4741 (control) yeast cells were grown in four liters of SD

media at 30 °C to mid-log phase. Cells were cross-linked with 1.25% formaldehyde at 30 °C for 5 minutes and quenched by glycine at a final concentration of 0.125 M for 5 minutes at 30 °C. Cells were harvested by centrifugation at 2600x g for 30 minutes at 4 °C and washed once with ice-cold water. Cells were re-suspended in 1/10 (volume/cell weight) of ice-cold 20 mM Hepes pH 7.5 and 1.2% PVP, snap-frozen in liquid nitrogen. Frozen cells were cryogenically ground by Retsch MM301 Mill at the frequency of 30 per second, 3 minutes per round, for 6 rounds. Ground cell powders were re-suspended in IP buffer (20 mM Hepes pH 7.5, 300 mM KCl, 2 mM MgCl₂, 0.1% Tween-20, 1x Roche protease inhibitors cocktail) at 5 ml per gram of cells and sonicated by Bioruptor at 30 seconds on/off for two rounds of 15 minutes in 4 °C. Sonicated samples were centrifuged at 2600x g at 4 °C for 15 minutes. Supernants (5 ml per sample) were incubated with IgG-conjugated M-270 Epoxy Dynabeads[®] at 4 °C for 5 hours with rotation. Beads were collected by magnetic rack and washed with IP buffer 6 times before elution with IP elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) at 65 °C overnight. Eluted ChIP samples were subjected to proteinase K digestion at 37 °C for 2 hours. Then the ChIP DNA samples were purified and extracted by Wizard SV Gel Extraction Purification Kit (Promega).

Quantitative PCR (qPCR) Analysis -- qPCR was performed using SsoAdvanced[™] Universal SYBR Green Supermix (BioRad) and CFX96 Real-time System (BioRad) following manufacturer's instruction. The name and sequence of oligonucleotides used for qPCR were: YDR544C Up2KbFor 5' - TGT TGA CTG GTA TTG CCC C -3'; YDR544C Up2KbRev 5'- GCG TTT GTC TCT AGT TTG CG -3'; YDR544C Up1KbFor 5'- TGG GAA TGG AAA TAG GAT GCC -3'; YDR544C Up1KbRev 5'- CAC CTC ACT GTC GTA ACA CTC -3'; YDR544C Up200bpFor 5'- GTG GTA GTG GAG TTG GAT ATG G -3'; YDR544C Up200bpRev 5'-ATT ACC CTA CCT CCC CAC TC-3'; YDR544C Down500bpFor 5'-AGA AGT TGT AGG CTA

AGC GC-3'; YDR544C Down500bpRev 5'-AGG ATT CTG TTC GTT GCT CAG-3'; YDR544C Down1.1KbFor 5'-AGT GCG GTC ATT CTA CGA AG-3'; YDR544C Down1.1KbRev 5'-TCC TAT CCA TTG CAC CAG TTC-3'; Tel15 For 5'-TGT ATT GCA TGC TGG ATG GT-3'; Tel15 Rev 5'-AAT GTA TTA CCC CGC CGA AT-3'.

References:

- 1 B. Zybaylov, A. L. Mosley, M. E. Sardi, M. K. Coleman, L. Florens and M. P. Washburn, *J. Proteome Res.*, 2006, **5**, 2339–2347.
- 2 B. Zybaylov, M. K. Coleman, L. Florens and M. P. Washburn, *Anal. Chem.*, 2005, **77**, 6218–6224.
- 3 B. Zybaylov, 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.
- 4 F. W. Studier, *Protein Expr. Purif.*, 2005, **41**, 207–234.
- 5 S. Werten, F. W. Langen, R. van Schaik, H. T. Timmers, M. Meisterernst and P. C. van der Vliet, *J. Mol. Biol.*, 1998, **276**, 367–377.
- 6 P. D. Morris, A. J. Tackett, K. Babb, B. Nanduri, C. Chick, J. Scott and K. D. Raney, *J. Biol. Chem.*, 2001, **276**, 19691–19698.

Supplementary Table S1. List of Sub1 peptides identified by LC-MS/MS.

Rank	Validated	Peptide	Unique	Scan ID	m/z	RT	Mass	ppm	#Spec	Score	S.E.	S.P.	S.M.	S.O.	Start	End	PTM
1	TRUE	K.GISLTEDLYDELLK.H	YES	F20:4425	804.922	33.06	1607.8296	0.3	2	99.95	PMO	90.5	108.8	1.54E-10	80	93	
2	TRUE	K.IAEPEPEPVTQLQAK.K	YES	F19:3089	809.937	25.36	1617.8613	-1	3	99.95	PMO	80.9	59.77	6.45E-06	238	252	
3	TRUE	R.EANATLIIPGQAGR.K	YES	F19:3144	705.888	25.69	1409.7625	-0.7	3	99.95	PMO	79.3	60.02	1.95E-05	190	203	
4	TRUE	K.KEDIVSNINESK.D	YES	F19:1996	688.356	18.69	1374.6990	-0.8	1	99.95	PMO	62.7	60.06	1.84E-04	253	264	
5	TRUE	K.EKIAEPEPEPVTQLQAK.	YES	F20:2894	938.507	24.41	1874.9990	-0.3	2	99.95	PMO	73.1	59.18	4.58E-05	236	252	
6	TRUE	R.EYYLDSSTGEM(+15.99)	YES	F20:1964	860.89	18.74	1719.7661	-0.7	4	99.94	PM	79.6	63.63	-	64	78	o
7	TRUE	R.NINLIDIR.E	YES	F20:3969	485.788	30.88	969.5606	0.3	1	99.94	PMO	40.5	47.04	8.89E-04	56	63	
8	TRUE	K.KKPAPPTLLPHEENIQN	YES	F20:3583	571.312	28.6	2281.2178	0	5	99.94	PmO	37.1	31.53	1.29E-05	170	189	
9	TRUE	R.LNIDEALR.R	YES	F21:3492	472.264	27.02	942.5134	-0.2	4	99.93	PMo	39.8	44.77	4.41E-02	96	103	
10	TRUE	R.EYYLDSSTGEMKPGK.K	YES	F20:2523	568.932	22.19	1703.7712	0.8	1	99.03	Po	25.7	-	2.73E-02	64	78	

The 1% false discovery rate threshold corresponds to the PEAKS score of 23, Mascot score of 45, and OMSSA score of 1e-4
 The search engines that successfully identify a give peptide are shown in the column named "S.E." (Search Engines), as follows:
 P - Peaks, M- Mascot, O - OMSSA. Lower case letter indicate identification below the threshold score.

Annotation:

Rank - peptides are ranked by descending InChorus confidence score

Validated- the MS/MS spectrum for the top scoring spectrum-to-peptide match for each peptide was manually inspected

Peptide - peptide sequence

Unique - only peptide uniquely matched to the protein were chosen

Scan ID - scan identifier from the raw mass spectrum file

m/z - experimentally observed mass over z ratio

RT - observed chromatographic retention time, in minutes

Mass - calculated mass of the molecular ion

ppm - deviation of the observed mass from theoretical, in parts per million (matches with absolute errors above 3 ppm were excluded from analysis)

#Spec - number of total MS/MS spectra per peptides

Score - InChorus score

S.E. (Search Engine) - Spectra were search with three different search engines - Peaks, Mascot, and OMSSA, and the results were integrated using Peaks v7 software suit with InChorus feature.

S.P. (Score. Peaks) - Peaks peptide identification score

S.M. (Score.Mascot) - Mascot peptide identification score

S.O. (Score.OMSSA) - OMSSA identification score

Start, End - position within the protein amino acid sequence

PTM - o - methionine oxidation

Supplementary Table S2. List of G4p1 peptides identified by LC-MS/MS.

Rank	Validated	Peptide	Unique	Scan ID	m/z	RT	Mass	ppm	#Spec	Score	S.E.	S.P.	S.M.	S.O.	Start	End	PTM
1	TRUE	K.EQSAQAAQWESVLK.S	YES	F20:3889	787.894	30.44	1573.7737	0.2	1	99.95	PMO	75.2	62.45	9.62E-07	23	36	
2	TRUE	K.VASIANAQVR	YES	F20:1631	514.796	16.67	1027.5774	-0.1	1	99.93	PMo	49.2	45.41	3.27E-01	367	376	
3	TRUE	K.HPDADSLYVSTIDVGDE	YES	F20:3880	758.017	30.4	2271.0293	-0.4	1	99.93	PmO	58.7	36.38	3.02E-04	222	242	
4	TRUE	K.FESLIISK.Y	YES	F20:3366	468.774	27.27	935.5327	-0.4	2	99.9	Pmo	41.2	37.06	4.91E-02	8	15	
5	TRUE	R.TVC(+57.02)SGLVK.H	YES	F20:1486	432.236	15.8	862.4583	-0.2	1	97.18	P	32.8	-	-	243	250	c
6	TRUE	K.APEKPKPSAIDFR.V	YES	F20:2605	485.936	22.68	1454.7881	-0.7	1	91.11	P	27	-	-	200	212	

The 1% false discovery rate threshold corresponds to the PEAKS score of 23, Mascot score of 45, and OMSSA score of 1e-4
The search engines that successfully identify a give peptide are shown in the column named "S.E." (Search Engines), as follows:
P - Peaks, M- Mascot, O - OMSSA. Lower case letter indicate identification below the threshold score.

Annotation:

Rank - peptides are ranked by descending InChorus confidence score

Validated- the MS/MS spectrum for the top scoring spectrum-to-peptide match for each peptide was manually inspected

Peptide - peptide sequence

Unique - only peptide uniquely matched to the protein were chosen

Scan ID - scan identifier from the raw mass spectrum file

m/z - experimentally observed mass over z ratio

RT - observed chromatographic retention time, in minutes

Mass - calculated mass of the molecular ion

ppm - deviation of the observed mass from theoretical, in parts per million (matches with absolute errors above 3 ppm were excluded from analysis)

#Spec - number of total MS/MS spectra per peptides

Score - InChorus score

S.E. (Search Engine) - Spectra were search with three different search engines - Peaks, Mascot, and OMSSA, and the results were integrated using Peaks v7 software suit with InChorus feature.

S.P. (Score. Peaks) - Peaks peptide identification score

S.M. (Score.Mascot) - Mascot peptide identification score

S.O. (Score.OMSSA) - OMSSA identification score

Start, End - position within the protein amino acid sequence

PTM - c - carbamidomethylated cysteine