PNAS
www.pnas.org

Supplementary Information for

Transcriptional elongation factor Paf1 core complex adopts a spirally-wrapped solenoidal topology

Pujuan Deng, Yuqiao Zhou, Junyi Jiang, Haojie Li, Wei Tian, Yinghua Cao, Yan Qin, Jaehoon Kim, Robert G. Roeder, Dinshaw J. Patel and Zhanxin Wang

Email: wangz@bnu.edu.cn; pateld@mskcc.org

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1812256115

SI Materials and Methods

Protein expression and purification. Open reading frames of all subunits (Ctr9, Paf1, Leo1, Rtf1 and Cdc73) in both *S. cerevisiae* and *M. thermophila* were chemically synthesized with optimized codon for efficient bacterial expression.

Subunits of scPaf1 complex were subcloned into three compatible vectors for coexpression in *E.coli*. Full-length scCtr9 was cloned into the pETDuet-1 vector without any tag. His-SUMO-tagged scRtf1 and scCdc73 without any tag were inserted into two multiple cloning sites of the RSFDuet-1 vector, respectively. SUMO-tagged scPaf1 and SUMO-tagged scLeo1 were cloned into two multiple cloning sites of the pCDFDuet-1 vector, respectively. These three plasmids were co-expressed in *E.coli* strain BL21(DE3) at 37 °C till the OD_{600} reached around 1.0. Then after cooling the cells at 20 °C for around an hour, 0.2 mM IPTG and 0.1 mM $ZnCl_2$ were added to induce expression overnight. Cells were harvested by centrifugation at 5,000 rpm for 10 minutes. Cell pellets were resuspended with the buffer containing 20 mM Tris-pH 8.0, 200 mM NaCl and 20 mM imidazole and sonicated for 5 minutes. The supernatant was fractionated by centrifugation of the cell lysate at 18,000 rpm for an hour. Histidine-SUMO-tagged target protein was isolated through a nickel-charged HiTrap Chelating FF column from GE healthcare. Both the SUMO and the histidine-SUMO tag was then cleaved by incubating with histidine-tagged ULP1 protease. The samples were then loaded directly onto a heparin column to remove bound DNA. Target protein was separated by increasing the salt concentration of the low salt buffer (20 mM Tris-pH 7.0, 200 mM NaCl, 2 mM DTT) from 200 mM to 1 M NaCl through a linear gradient. The eluted target protein was then purified on a HiLoad 200 16/600 gel-filtration column. Finally, the purified sample was loaded onto a Mono Q 5/50 column to remove excess subunits. After these purification steps, the target protein is of high purity and homogeneity. The purified protein was concentrated to around 16 mg/ml and stored in a -80 °C freezer.

The mtPaf1 complex was also reconstituted by the coexpression method. MtCtr9 (1-1094) was cloned into a hexahistidine-MBP-TEV-tagged RSFDuet-1 vector. The mtRtf1 and mtCdc73 without any tag were cloned into two multiple

cloning sites of the pETDuet-1 vector, respectively. For better expression, mtPaf1 and mtLeo1 were linked by a TEV protease site and cloned into a SUMO-tagged pCDFDuet-1 vector. To reconstitute the 5-subunit mtPaf1 complex, above three plasmids were coexpressed in *E.coli*. The purification procedures are the same as for the scPaf1 complex.

For coexpression of the ternary complex of mtCtr9-mtPaf1-mtCdc73, mtCtr9 (31-967) was inserted into a hexahistidine-GST-TEV-tagged pGEX-6p-1 vector. SUMO-tagged mtPaf1 (1-120) and SUMO-tagged mtCdc73 (155-227) were cloned into two multiple cloning sites of a pCDFDuet-1 vector, respectively. The selenomethionine-labeled mtCtr9 (31-967)-mtPaf1 (1-120)-mtCdc73 (155-227) complex is expressed in a methionine auxotrophic B834(DE3) strain. One liter cells with the OD_{600} of 1.2 grown from regular LB media were harvested by centrifugation at 4,000 rpm for 10 minutes, washed twice with ddH_2O and used to inoculate 4 L methionine-depleted Medium Base supplemented with 50 mg/L L-selenomethionine and Nutrient Mix (SelenoMet, Molecular Dimensions). After shaking for an additional 30 min at 37 °C, the cells were induced with addition of 0.2 mM IPTG and shaken overnight at 20 °C.

The binary complex of mtCtr9 (31-967)-mtPaf1 (1-120) was obtained by coexpressing SUMO-tagged mtPaf1 (1-120) in pCDFDuet-1 plasmid and hexahistidine-GST-TEV-tagged mtCtr9 (31-967) pGEX-6p-1 plasmid in *E.coli*.

All mutations/truncations were induced by PCR method. The mtCdc73 (155-227) and its mutants were all cloned into a hexahistidine-MBP-TEV-tagged RSFDuet-1 vector. The target protein was first purified with a nickel-charged HiTrap Chelating FF column. Then, the eluted target protein was purified on a HiLoad 200 16/600 gel-filtration column.

Crystallization and structure determination. Crystallization was carried out using the hanging-drop, vapor-diffusion method by mixing equal volume of protein and well solution. Crystals of the ternary complex containing mtCtr9(31-967), mtPaf1(1-120) and mtCdc73(155-227), (designated as the core mtPaf1 complex), were grown at 20 °C by mixing 0.5 µl protein at the concentration of 16

mg/ml with 0.5 μl crystallization buffer containing 0.2 M potassium phosphate dibasic and 20% PEG 3,350.

Crystals of the selenomethionine-labeled core mtPaf1 complex were obtained at 20 °C by mixing 0.5 μl protein with 0.5 μl well solution, which contains 0.2 M potassium citrate tribasic monohydrate and 20% PEG 3,350.

Crystals of the selenomethionine-labeled mtPaf1 core complex bearing a single L200M mutation in Cdc73, or bearing both the I40M mutation in Paf1 and the L186M mutation in Cdc73 were all grown at 20 °C in the buffer containing 0.2 M potassium citrate tribasic monohydrate and 20% PEG 3,350.

For above four kinds of crystals, well solution with added 12% 2,3-butanediol was used as the cryo-protectant.

Data sets for the crystals of the native proteins were collected at the Shanghai Synchrotron Radiation Facility (SSRF) beamline BL19U1 in China. Data sets for crystals of the selenomethionine-labeled wild-type samples were collected at SSRF beamline BL17U1 at the wavelength of 0.97917 Å. Data sets for both kinds of selenomethionine-labeled mutant samples were collected at SSRF beamline BL18U1 at the wavelength of 0.97930. All the data sets were processed by the program HKL2000 (1). Structure determination was carried out by PHENIX (2) using the data sets of the selenomethionine-labeled crystals of the wild-type samples through the SAD method. Twenty one out of twenty two selenium atoms were identified, which were used to solve the initial phase with a partial model. The unidentified selenomethionine residue is located at the N-terminus of Paf1, which is disordered. The partial model was manually rebuilt by Coot (3), and further refined by PHENIX (2) and Refmac5 (4) in the CCP4 suite (5). After an almost complete model was built, the selenomethionine-labeled model was used as the starting model for the 2.9 Å native data set for further refinement. To identify I40M introduced in Paf1, L186M and L200M introduced in Cdc73, anomalous difference Fourier maps were calculated by the CCP4 suite (5). Anomalous signals of the mutation sites were identified, which further verify the accurateness of the model.

Electrophoretic Mobility-Shift Assay (EMSA). For a better resolution, 100 picomoles of single stranded DNA or 20 picomoles of all the other forms of DNAs tested were mixed with increasing amount of recombinant Paf1 complex samples in the buffer containing 20 mM Tris-pH 8.0, 200 mM NaCl and 2 mM DTT. After incubation at 4 °C for 20 minutes, the mixture was loaded onto a 1.2% agarose gel in the TAE buffer for electrophoresis and detected by ethidium bromide staining. The sequence of the 40-mer single stranded DNA is: GGGCGGCGGG-$(T)_{20}$-GGCGGGGCGG. The sequence of one strand of the 20-mer double stranded DNA is: GCAGGTCCATTCGGGAATTA. The sequence of the forward strand of the fork DNA is: GGGCGGCGGG-$(T)_{20}$. The reverse strand of the fork DNA is: CCCGCCGCCC-$(T)_{20}$. The sequence of the forward strand of the DNA bubble is: GGGCGGCGGG-$(T)_{20}$-GGCGGGGCGG. The sequence of the reverse strand of the DNA bubble is: CCGCCCCGCC-$(T)_{20}$-CCCGCCGCCC. All EMSA experiments were repeated at least three times.

N/C-terminus analysis and Edman sequencing. The crystals of the scPaf1 complex were picked and washed there times with the crystal mother liquor. All the picked crystals were dissolved in the buffer containing 20 mM Tris-pH 8.0, 200 mM NaCl and 2 mM DTT, and incubated at 95 °C for 5 minutes. Samples for N/C-terminus analysis were separated by SDS-PAGE. Single bands from the polyacrylamide gel were picked and used for the analysis.

Samples for Edman sequencing were separated by SDS-PAGE. Then the unstained samples were transferred onto a polyvinylidene difluoride (PVDF) membrane. The blotted membrane was stained by coomassie blue to visualize the transferred bands. After the PVDF membrane was fully dried, separated bands were excised and sequenced by the Protein and Chemistry Platform at Tsinghua University.

***In vitro* reconstitution of the complexes.** Three-, four- or five-subunit Paf1 complexes can also be obtained through *in vitro* reconstitution method. To reconstitute the ternary complex of mtCtr9-mtPaf1-mtCdc73 used for the

mutational analysis, binary complex of mtCtr9-mtPaf1 and mtCdc73 or its mutants were prepared separately. Binary complex of mtCtr9(31-967)-mtPaf1(1-120) was obtained through a coexpression method as described above. The mtCdc73(155-227) fragment and its mutants were expressed as His-MBP-tagged protein for better staining. Purified mtCtr9-mtPaf1 and mtCdc73 or its mutants were mixed at 1:1.5 ratio on ice for half an hour, then the mixture was loaded onto a Superdex 200 column. Complex fractions were pooled and concentrated for further analysis.

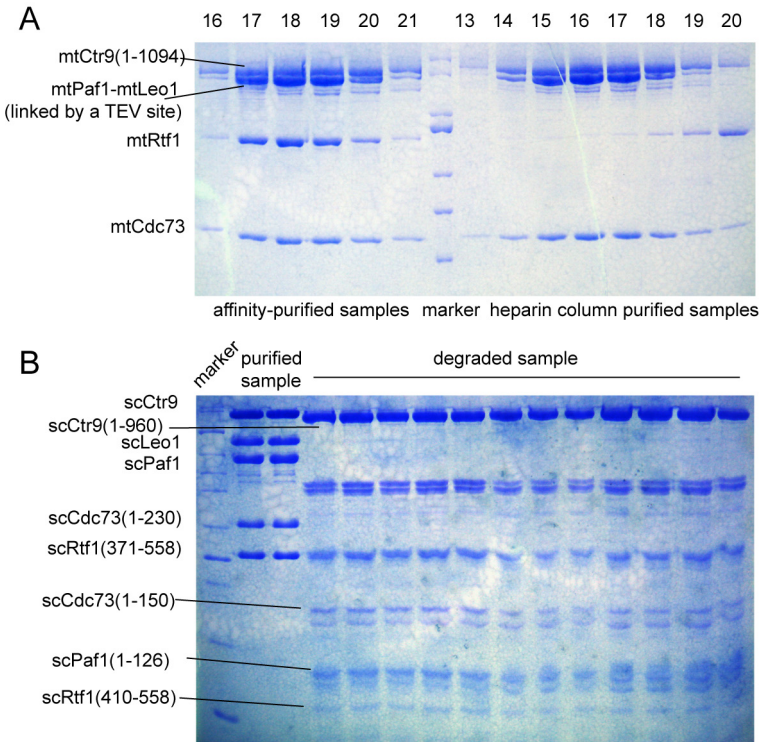

Fig. S1. SDS-PAGE analysis of different combinations of Paf1 complexes. (A) Left, gel-filtration results of the five-subunit mtPaf1 complex obtained by a one-step affinity purification procedure. (B) SDS-PAGE analysis of the crystallized samples of the five-component scPaf1 complex. Four degraded bands identified by the N/C-terminus analysis and Edman sequencing are pointed out.

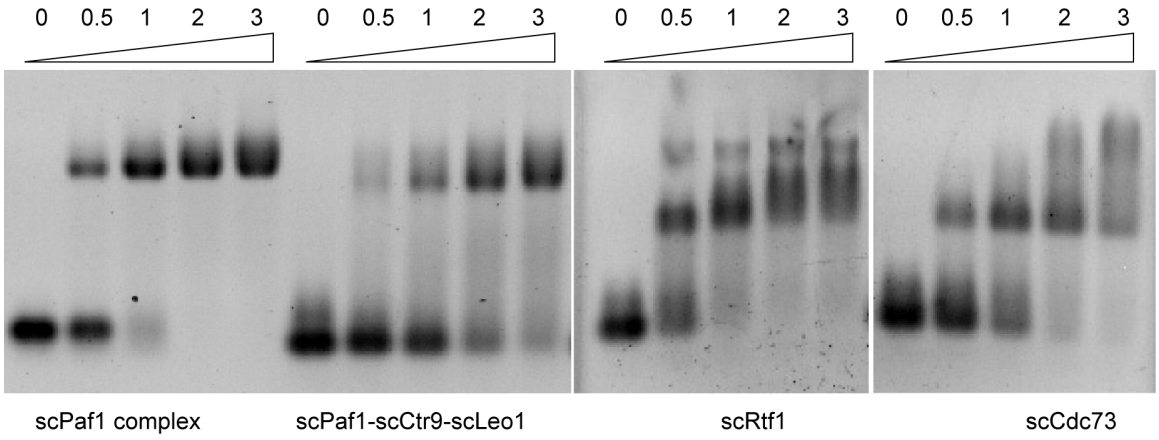

Fig. S2. EMSA results of the subunits or sub-complex of scPaf1 complex with the bubble-containing DNA. The same bubble-containing DNA was used to do the EMSA analysis with the full-length scPaf1 complex, the scPaf1-scCtr9-scLeo1 ternary complex, scRtf1 and scCdc73, respectively. Protein to DNA ratio was indicated above.

Table S1. Data collection and refinement statistics

Crystal	Native mtPaf1C mtCtr9(31-967)/ mtPaf1(1-120)/ mtCdc73(155-227)	Selenomethionine-labeled mtPaf1C	Selenomethionine-labeled mtPaf1C + I40M (mtPaf1) + L186M (mtCdc73)	Selenomethionine-labeled mtPaf1C + L200M (mtCdc73)
Beam line	SSRF-BL19U1	SSRF-BL17U1	SSRF-BL18U1	SSRF-BL18U1
Wavelength	0.97890	0.97917	0.97930	0.97930
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell				
a, b, c (Å)	99.9, 237.8, 62.7	99.1, 238.4, 62.6	99.6, 240.9, 62.7	99.5, 239.9, 62.6
α, β, γ (^{o})	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.0-2.90 (2.95-2.90)[a]	50.0-3.30 (3.36-3.30)	50.0-3.40 (3.46-3.40)	50.0-3.20 (3.36-3.30)
R_{sym}	0.075 (1.231)	0.190 (1.547)	0.149 (1.161)	0.142 (1.15)
I/σ (I)	24.7 (1.4)	16.6 (1.6)	15.8 (1.8)	18.5 (1.7)
Completeness (%)	100.0 (100.0)	99.3 (99.9)	99.9 (99.8)	100.0 (99.9)
Redundancy	10.4 (10.3)	10.9 (11.5)	12.1 (11.1)	12.9 (12.7)
Number of unique reflections	34480	22840	21731	23669
R_{work}/R_{free} (%)	24.2/30.5			
Non-H atoms	8372			
Ramachandran statistics				
In preferred regions (%)	95.8			
In allowed regions (%)	4.2			
Average B factors (Å^2)				
mtCtr9	94.8			
mtPaf1	81.5			
mtCdc73	103.5			
Bond lengths (Å)	0.01			
Bond angles (^{o})	1.37			

[a] Highest resolution shell (in Å) shown in parentheses.

References

1. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307-326.
2. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
3. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
4. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-255.
5. Winn MD*, et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235-242.