Supplementary Information

Precise Sequencing of Single Protected-DNA Fragment Molecules for Profiling Distribution and Assembly of Proteins on DNA

*Zheng Yuan[a,b], Dapeng Zhang[a], Fangzhi Yu[a,b], Yangde Ma[a,b], Yan Liu[a,b], Xiangjun Li[b], Hailin Wang[a,b,c,d]**

a. State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, 100085, China

b. University of Chinese Academy of Sciences, Beijing, 100049, China

c. Institute of Environment and Health, Jianghan University, Wuhan, Hubei, 430056, China

d. Institute of Environment and Health, Hangzhou, Institute for Advanced Study, UCAS, Hangzhou, 310000, China

*To whom correspondence should be addressed. E-mail: hlwang@rcees.ac.cn

Contents

Material and methods

Chemicals and reagents.

Escherichia. coli single-stranded DNA-binding protein (SSB) and S1 nuclease were from Promega (Madison, WI, USA). The supernuclease and deoxyribonuclease I (DNase I) were supplied by Sino Biological Inc. (Beijing, China) and Worthington Biochemical Corp. (NY, USA) respectively, and the powders were dissolved in 1× TH buffer (20 mM Tris-HCl, pH 7.5) prior to the use. Quick CIP (calf intestine phosphatase), proteinase K (molecular biology grade) and 10 × Cutsmart® buffer (20 mM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 100 μg/ml BSA, pH 7.9, 25 °C) were purchased from New England Biolabs Inc. (Ipswich, MA, USA). DH5α *E. coli* competent strain was bought from Weidi Biothchnology Corp. (Shanghai, China). CV17-Zero Background pTopo-Blunt Simple Cloning Kit was supplied by Aidlab Biotechnologies Corp. (Beijing, China). All oligodeoxynucleotides were synthesized and HPLC purified by Sangon Biological Engineering Technology and Services (Shanghai, China) and the related sequences are listed in Table S1.

Ampicillin sodium salt was supplied by Ameresco (Solon, OH, USA). DNA extraction reagent (phenol/chloroform/isoamyl alcohol = 25:24:1, pH>7.8), Sodium acetate (3M sterile solution, pH 5.2) and 40% acrylamide (mass ratio of acrylamide to bis-acrylamide: 29:1) were bought from Solarbio Science and Technology Co. Ltd. (Beijing, China). Nucleic acid co-precipitating reagent glycogen (20 μg/mL) and Qubit™ 1x dsDNA HS Assay Kit were obtained from Thermo Fisher Scientific (Rockford, IL, USA). Other chemical reagents were purchased from National Pharmaceutical Group Chemical Reagent Company (Beijing, China) unless otherwise stated. All solutions were prepared using ultrapure water from a Purelab Ultra Elga Labwater system (VWS Ltd., UK) with an electrical resistivity of 18.2 MΩ·cm and purified through a 0.45 μm filter.

Equipment.

A Bio-Rad T100 System (Bio-Rad Laboratories, Foster City, CA, USA) was used to maintain the temperature at 37 °C for DNA-protein interactions and supernuclease cutting reactions, and at 25 °C for vector ligation. A K30 dry bath incubator supplied by Allsheng instrument Co. Ltd. (Hangzhou, China) was used to control the temperature of other reactions. A BG-verMIDI standard vertical gel tank (Baygene Biotech Company Limited, Beijing, China) was used to prepare polyacrylamide gel and perform electrophoresis. Odyssey CLX infrared imaging system (LI-COR Biotechnology, Lincoln, NE, USA) was used for infrared fluorescence scanning with a 700 nm excitation.

High-resolution cutting of ssDNA and dsDNA by nucleases.

To seek optimized nuclease cutting assay, 400 fmol supernuclease or DNase I was added to the solution of 20 nM ssDNA (46TCy5-ss90mer) or dsDNA (46TCy5-ds90mer) prepared with a volume of 20 μL in 1× TH buffer, plus 10 mM Mg^{2+} or 10 mM Ca^{2+}. Following incubation at 37 °C for designated time, 50 mmol/L EDTA was added to terminate the cutting reaction. Then the final samples were mixed with 10% (final concentration) glycerol and resolved on a 16% native polyacrylamide gel at 200 V for 100 min.

Enzymatic cutting of unprotected DNA and recovery of protected DNA fragments.

DNA substrates (20 nM) were mixed with SSB (tetramer, 10 - 50 nM) in 1× TH buffer, plus 10 mM Mg^{2+}. After 10 min incubation at 37 °C, 20 U supernuclease was added to remove unprotected DNA segments for the indicated time. The total reaction volume was 80 μL. Then, phenol/chloroform extraction solution (200μL) was added as quickly as possible and vigorously shocked manually for 60 seconds. After the addition of 120 μL ultrapure water, the solution was centrifuged (13,800 × g) for 5.0 min, and the upper aqueous phase was taken out and mixed with one-tenth volume 3.0 M NaAc (pH 5.2), 40 μg glycogen and a triple volume of cold ethanol. Then the mixed solution was frozen at -80 °C for 1-2 hours. Subsequently DNA pellets were isolated by centrifugation at 13,800

× g for 15 min and desalted by adding 600 μL 75% alcohol (alcohol/water: 3:1, v/v) before further twice centrifugation at 13,800 × g for 5.0 min at 4 °C. After air-drying at room temperature for 10 min, the collected DNA pellets were dissolved in 1× TH buffer, plus 10 mM NaCl.

Preparation of blunt-ended dsDNA template.

Full length of complementary ssDNA (excessive) was mixed with the recovered ssDNA fragments and the mixed solution was heated at 95 °C for 5.0 min and then slowly cooled down to 25 °C for 1.0 hour for triggering hybridization. After gentle centrifugation at 4 °C, the hybridized solution was mixed with 20 U S1 nuclease, 1.0 U Quick CIP in 1× Cutsmart® buffer and incubated at 37 °C for 50 min. The generated dsDNA template was enriched by phenol/chloroform extraction followed by ethanol precipitation. The recovered DNA pellets were resuspended in 30 μL cool ultrapure water. Each experiment was repeated three times. To evaluate the reproducibility, the yield of generated dsDNA templates was measured using Qubit™ 1 × dsDNA HS Assay Kit. Qubit™ 4 Fluorometer was used for this measurement.

T-clone sequencing.

The recovered dsDNA template (2.0 μL) was mixed with 30 ng (24.4 fmol) pTOPO-Blunt Simple vector and 1.0 μL 10× Enhancer from Aidlab at 25 °C for 10 min. The ligated plasmids were transformed into 100 μL DH5α *E. coli* competent cells by heat shock. The transformed cells (50 μL) were plated in Luria-Bertani (LB) medium on solid agar plates containing 100 μg/mL ampicillin and cultured in a 37 °C incubator overnight. For SPDFMS assay of a DNA-protein interaction reaction, 70-287 grown monoclones were randomly picked up using a quartering method and then each was cultured with 500 μL LB medium containing 100 μg/mL ampicillin in a 37 °C incubator. Then plasmid DNA from each cultured clone were isolated using well-established alkaline lysis method[1]. Briefly, the cultured E. coli carrying pTOPO-plasmid is harvested, and then lysed with a SDS/NaOH solution followed by a rapid acidification using concentrated

potassium acetate, which causes the precipitation of protein and chromosomal DNA. The plasmids in the supernatant were precipitated using 70% ethanol (v/v) and recovered. The isolated plasmids were subjected to Sanger sequencing using the designed primer 3.X by ABI 3730xl 96-capillary DNA analyzer (Applied Biosystems, Inc., USA). The sequences of DNA fragments were inferred by aligning with the tested DNA substrate and un-ligated pTOPO-Blunt Simple vector by SnapGeneTM 1.1.3v. All sequencing data were handled with Microsoft Excel 2016 for follow-up displaying.

HeatMap for the binding rate statistic of each nucleotide.

The additive binding probability (P_n) of nth nucleotide for a tested DNA substrate can be calculated as:

$$P_n = L_n/T \quad (1)$$

Where T is the total number of protected-DNA fragments randomly chosen for modified T-clone Sanger sequencing. The protected fragments were obtained from high-resolution removal of un-occluded segments for the complexes of proteins and tested DNA substrate. L_n is the present number of the nth nucleotide in all obtained sequences of protected-DNA fragments. The binding probability of all nucleotides was listed in Microsoft Excel 2016 according to the sequence order of the tested DNA substrate. By the use of Microsoft Excel 2016, the HeatMaps were made by the three color scales of the conditional formatting function.

Table S1 The sequences of the synthesized oligodeoxynucleotides used in this work. The marked red T indicates a Cy5 labelling, and the marked blue nucleotide and orange nucleotide indicate a phosphorylation and TMR labelling, respectively.

probe name	Sequence (5' to 3')
marker10	TTGTCCCCAG
marker20	TTGTCCCCAGCATTTAAAAC
marker40	TTGTCCCCAGCATTTAAAACCTCTGCCGTAAGCGATGTCC
marker60	TTGTCCCCAGCATTTAAAACCTCTGCCGTAAGCGATGTCCTGG CCCCTCCTCAGCACCTT
46TCy5-ss90mer	TTTCCTACCTTAAGATCCTTCCAGTCTCCGCCGGCCAGTGTTAT TTTTAGAGCTCATACCATTCGCCAATTTCTTCGCACGTTAGTCT TT
com-ss90mer	AAAGACTAACGTGCGAAGAAATTGGCGAATGGTATGAGCTCT AAAAATAACACTGGCCGGCGGAGACTGGAAGGATCTTAAGGT AGGAAA
5'Cy5-ss80(+)	TTGTCCCCAGCATTTAAAACCTCTGCCGTAAGCGATGTCCTGG CCCCTCCTCAGCACCTTATCCTTTCCGAACCGAATTT'
ss80(+)	TTGTCCCCAGCATTTAAAACCTCTGCCGTAAGCGATGTCCTGG CCCCTCCTCAGCACCTTATCCTTTCCGAACCGAATTT
5'-ss50(-)	GGAGGGGCCAGGACATCGCTTACGGCAGAGGTTTTAAATGCT GGGGACAA
Mid-ss50(-)	AGGATAAGGTGCTGAGGAGGGGCCAGGACATCGCTTACGGCA GAGGTTTT
3'-ss50(-)	AAATTCGGTTCGGAAAGGATAAGGTGCTGAGGAGGGGCCAGG ACATCGCT
5'P-ss80(+)	TTGTCCCCAGCATTTAAAACCTCTGCCGTAAGCGATGTCCTGG CCCCTCCTCAGCACCTTATCCTTTCCGAACCGAATTT
3'P-ss80(+)	TTGTCCCCAGCATTTAAAACCTCTGCCGTAAGCGATGTCCTGG CCCCTCCTCAGCACCTTATCCTTTCCGAACCGAATTT
5'3'P-ss80(+)	TTGTCCCCAGCATTTAAAACCTCTGCCGTAAGCGATGTCCTGG CCCCTCCTCAGCACCTTATCCTTTCCGAACCGAATTT
ss80(-)	AAATTCGGTTCGGAAAGGATAAGGTGCTGAGGAGGGGCCAGG ACATCGCTTACGGCAGAGGTTTTAAATGCTGGGGACAA

mirror ss80(+)	TTTAAGCCAAGCCTTTCCTATTCCACGACTCCTCCCCGGTCCTG TAGCGAATGCCGTCTCCAAAATTTACGACCCCTGTT
mirror ss80(-)	AACAGGGGTCGTAAATTTTGGAGACGGCATTCGCTACAGGAC CGGGGAGGAGTCGTGGAATAGGAAAGGCTTGGCTTAAA
ss80(++)	TGAACCCATCAAATTTTCTCTCGAACTGTCCCCCGTATCTTCCT TCGCACCGTCGCTAGGGGCATGCATTCACTACCATC
ss80(--)	GATGGTAGTGAATGCATGCCCCTAGCGACGGTGCGAAGGAAG ATACGGGGGACAGTTCGAGAGAAAATTTGATGGGTTCA
SPDFMS30(-)	TATCCAGTAGGCGGGTTACTTACACGATAA
TMR-30(-)	TATCCAGTAGGCGGGTTACTTACACGATAA
SPDFMS110(+)	TTATCGTGTAAGTAACCCGCCTACTGGATATTGTCCCCAGCAT TTAAAACCTCTGCCGTAAGCGATGTCCTGGCCCCTCCTCAGCA CCTTATCCTTTCCGAACCGAATTT
SPDFMS110(-)	AAATTCGGTTCGGAAAGGATAAGGTGCTGAGGAGGGGCCAGG ACATCGCTTACGGCAGAGGTTTTAAATGCTGGGGACAATATCC AGTAGGCGGGTTACTTACACGATAA
5’ds30-ss80	SPDFMS110(+)/PSMPD-30(-)
TMR-5’ds30-ss80	SPDFMS110(+)/TMR-30(-)
Primer 3.X	TACCGCTGTTGAGATCCAGTTC

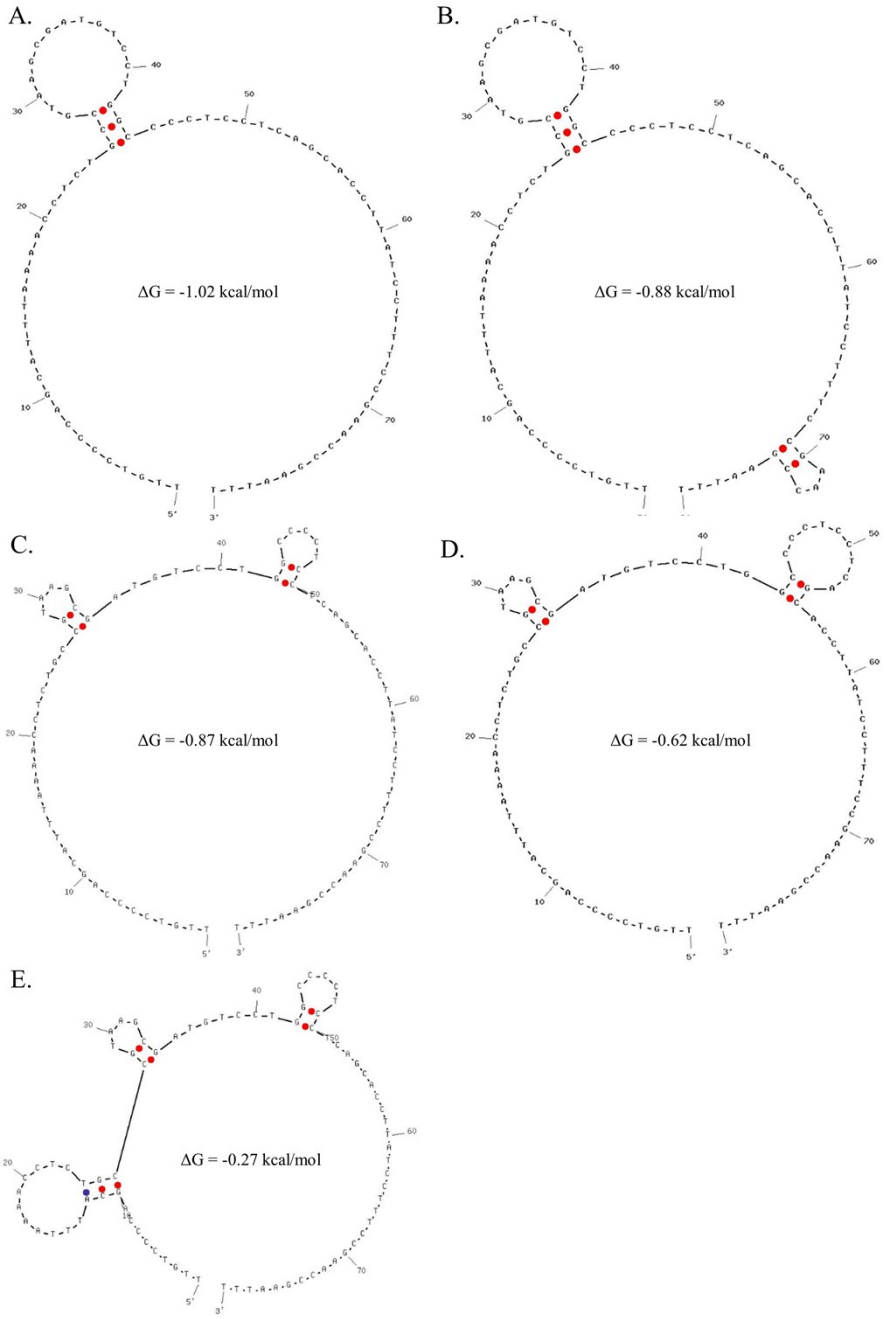

Fig. S1 Secondary conformation prediction for ss80(+) and single-stranded segment of the DNA substrate 5'ds30-ss80. The secondary conformation and the free energy were evaluated by using the Quickfold function of UNAFold[2]. The red and blue dots indicate the formed intramolecular hydrogen-bond. The short connected base-pairs and the estimated values of low Gibbs free energy indicate that no stable secondary conformation exists in the single-stranded segment of 5'ds30-ss80.

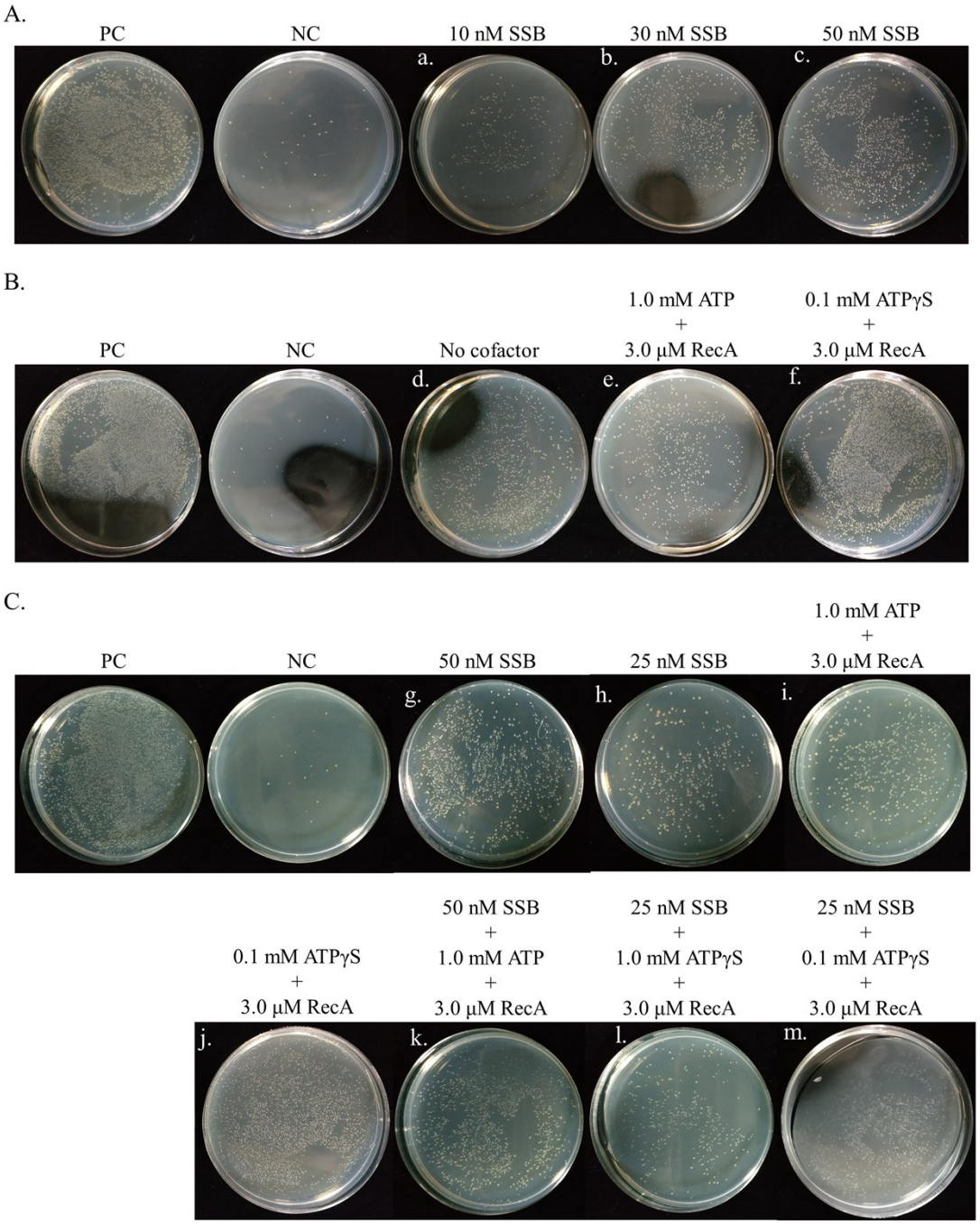

Fig. S2 The mono-clonal growth status of SPDFMS assay. (A), related to Fig. 5; (B), related to Fig. 6; and (C), related to Fig. 7. PC: positive control, obtained directly from the modified T-clone growing for the DNA substrate 5'ds30-ss80 without supernuclease digestion; NC: negative control, obtained from the modified T-clone growing for the products of the DNA substrate 5'ds30-ss80 by supernuclease digestion.

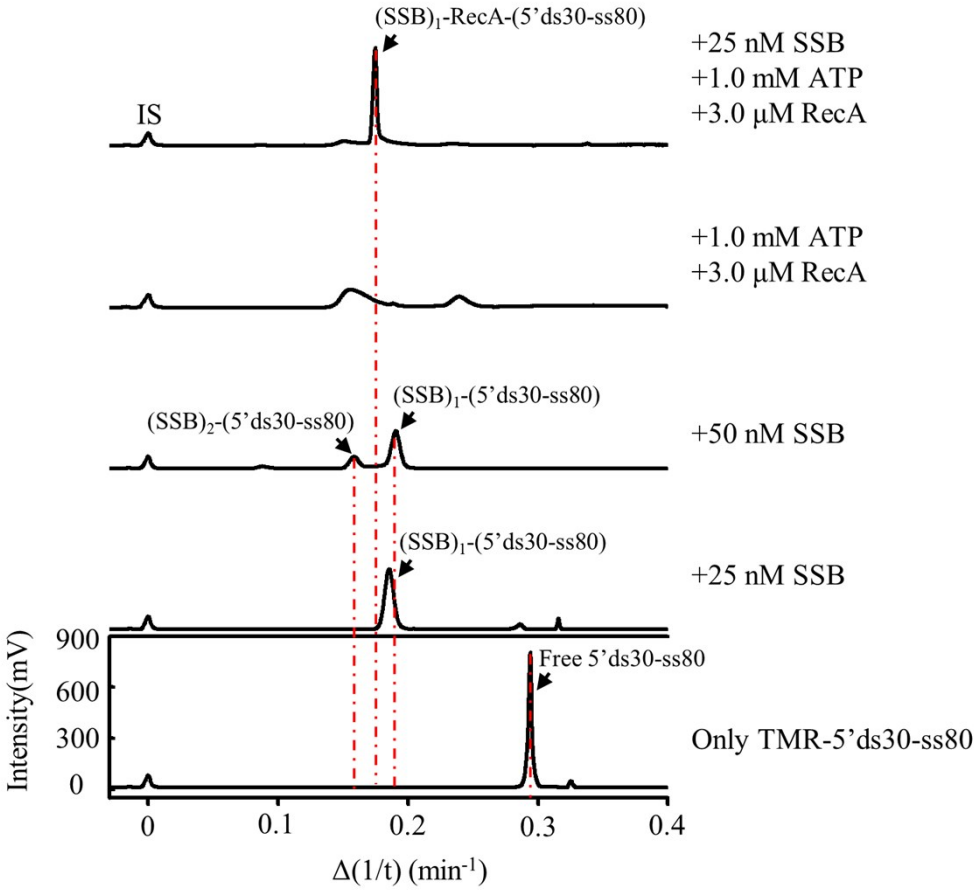

Fig. S3 CE-LIF assay of protein-DNA complexes in the presence of SSB, RecA and ATP, showing the formation of a SSB-RecA-ssDNA tertiary complex. The concentration of SSB (tetramer) was 25 nM. The reaction proceeded for 5.0 min. TMR-5'ds30-ss80 was used as the DNA substrate. The TMR-5'ds30-ss80 was prepared by hybridizing TMR-30(-) with SPDFMS110(+) and was purified as previously described [3]. The CE-LIF analysis was undertaken according to a previous work [4]. $\Delta(1/t) = 1/t_0 - 1/t$, and t_0 and t represents the migration time of IS and the complexes, respectively. As shown in this figure, the complex $(SSB)_1$-$(RecA)_n$-5'ds30-ss80 is identified due to the fact that it has an electrophoretic mobility between that of $(SSB)_1$-5'ds30-ss80 and $(SSB)_2$-5'ds30-ss80.

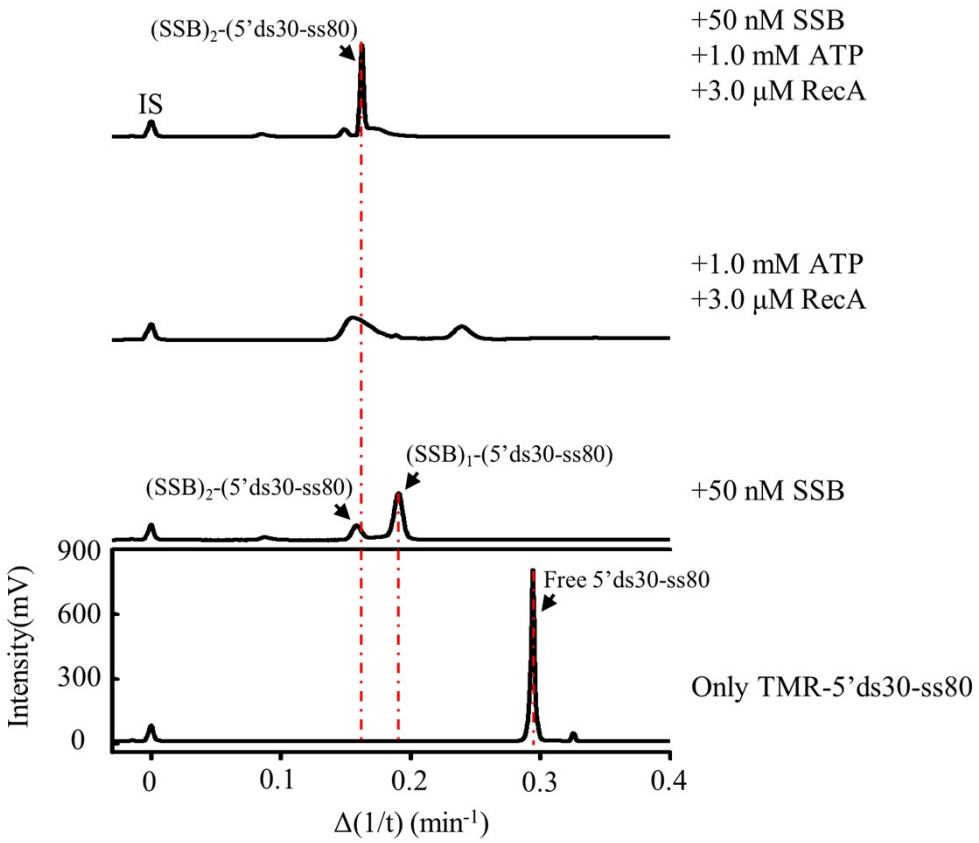

Fig. S4 CE-LIF assay of protein-DNA complexes in the presence of SSB, RecA and ATP, showing the formation of $(SSB)_{35}$-related 2:1 SSB-DNA complex promoted by RecA assembly. The concentration of SSB (tetramer) was 50 nM. The reaction proceeded for 5.0 min. TMR-5'ds30-ss80 was used as the DNA substrate.

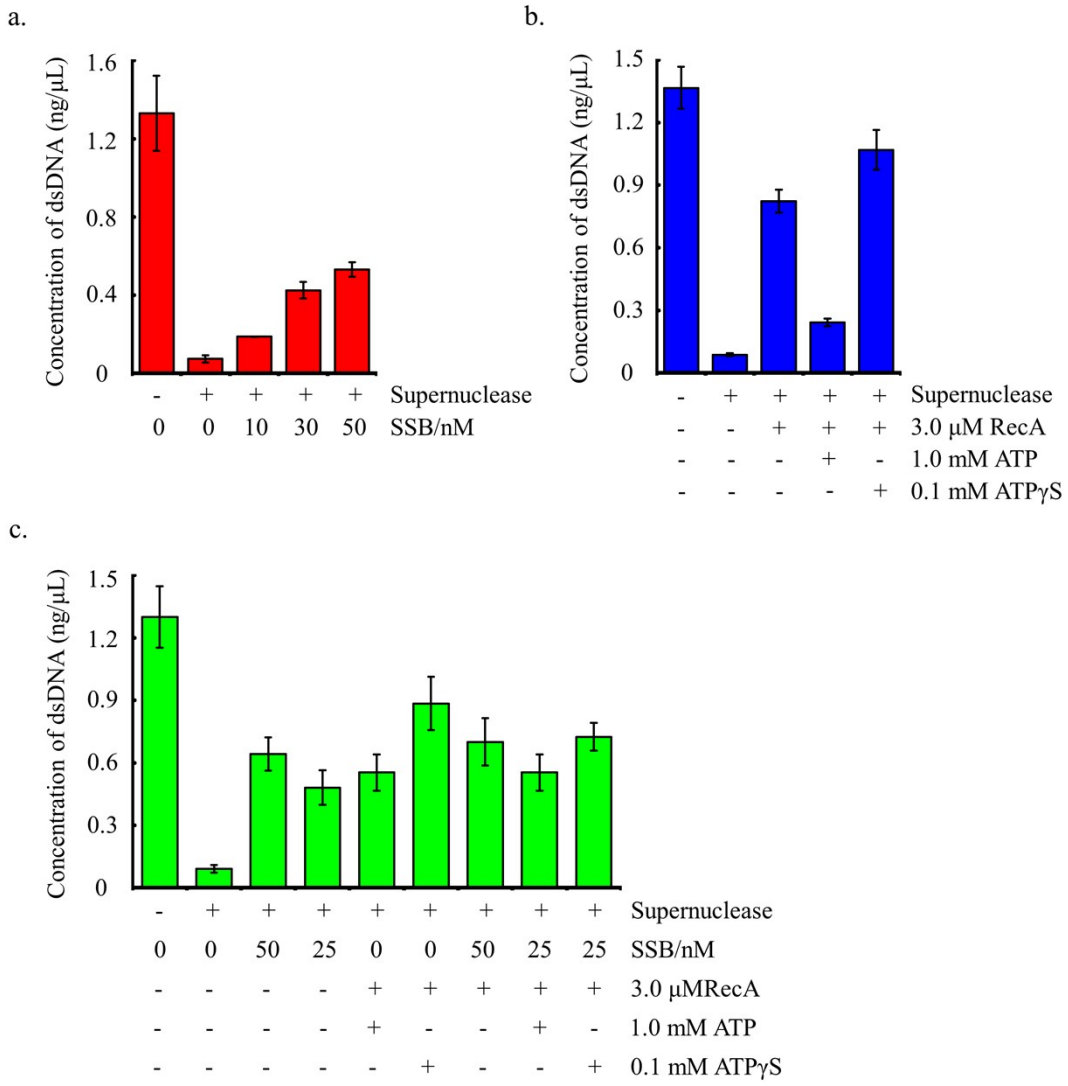

Fig. S5 Yield of dsDNA templates generated for T-clone Sanger sequencing in all SPDFMS assays listed in the main text. (a), Related to Fig. 6. (b), Related to Fig. 7. (c), Related to Fig. 8. The error bar indicates the standard deviation. Each experiment was independently repeated three times (n = 3).

Reference

1. S. Ehrt and D. Schnappinger, in *E. coli Plasmid Vectors: Methods and Applications*, eds. N. Casali and A. Preston, Humana Press, Totowa, NJ, 2003, DOI: 10.1385/1-59259-409-3:75, pp. 75-78.
2. N. R. Markham and M. Zuker, *Nucleic acids research*, 2005, **33**, W577-W581.
3. A. Petrov, T. Wu, E. V. Puglisi and J. D. Puglisi, in *Method Enzymol.*, ed. J. Lorsch, Academic Press, 2013, vol. 530, pp. 315-330.
4. B. Zhao, D. Zhang, C. Li, Z. Yuan, F. Yu, S. Zhong, G. Jiang, Y. G. Yang, X. C. Le, M. Weinfeld, P. Zhu and H. Wang, *Cell discovery*, 2017, **3**, 16053.