Single-molecule binding characterization of primosomal protein PriA involved in replication restart

Tzu-Yu Lee[1], Yi-Ching Li[2], Min-Guan Lin[2], Chwan-Deng Hsiao[2,*], and Hung-Wen Li[1,*]

[1]Department of Chemistry, National Taiwan University, Taiwan
[2]Institute of Molecular Biology, Academia Sinica, Taiwan
[*]To whom correspondence should be sent: hsiao@gate.sinica.edu.tw (C.-D. Hsiao) or hwli@ntu.edu.tw (H.-W. Li)

Supplemental information

Kinetic analysis

Intensity calculation and colocalization analysis were performed using Imscroll functions as previously described.[1] The start of a colocalization event is marked when a spot is detected within 1.5-pixel distance to the DNA molecule, while the end of this colocalization event corresponds to when no spot is found within 2.25 pixels. We utilized ebFRET to find the change points in intensity. The intensity time traces were linearly scaled to the interval (0.1, 0.9) to use ebFRET. The scaled traces were analyzed by the script version of ebFRET (https://github.com/ebfret/ebfret-script), with settings of the number of states from 1 to 4 and 2 restarts. The best number of states were selected from the results, which is the lowest possible number to separate the average of each state by at least 2.5 standard deviations. Finally, the Viterbi paths were scaled back to the original intensity.

Dwell time extraction was performed using custom python scripts (https://gitlab.com/leejuyuu/python_for_imscroll). The states of molecules were determined majorly from Viterbi paths output from ebFRET. Non-zero intensity states were validated by $\geq 90\%$ of the interval corresponds to colocalization signal. Traces that cannot pass this criterion were discarded. The remaining traces were further verified by visual inspection. The intensity change points were calculated as the midpoint of Viterbi path change. Then dwell times were then determined from differences of these time points.

Survival curves were determined with python scripts calling "survfit" from the survival pacakge of R. Survival model parameters were determined using maximum likelihood estimation by non-linear optimization. The survival function for association dwell time is double-exponential $S(t) = Ae^{-k_1 t} + (1 - A)e^{-k_2 t}$, where A, k_1, and k_2 are optimization parameters. As the slower component likely corresponds non-specific binding rate, we determined the apparent association rate

constant as $\max\{k_1, k_2\}$. For the dissociation dwell time, the survival function is single-exponential $S(t) = e^{-kt}$, where k is the optimization parameter.

The electrostatic surface features of *Gst*PriA and *Ec*PriA

To assess the surface potential of *Gst*PriA and *Ec*PriA molecules, we first had to establish their protein structures. Unfortunately, the actual tertiary structures of *Gst*PriA and *Ec*PriA remain unknown, so we attempted to generate homology models in the SWISS-MODEL Server[2] and using *Kp*PriA (PDB:4NL4) as the template. The DNA binding domain (DBD) of PriA contains two functional elements: a 3' DNA-binding domain (3' BD), and an unusual circularly-permuted winged helix (WH) domain.[3] The 3' BD possesses a pocket that can recognize and accommodate the 3' terminal nucleotide of one oligonucleotide[4], and the WH domain can bind partial duplex DNA.[5] Both the 3' BD and WH domains functionally cooperate to show a higher DNA-binding affinity than either individual domain alone.[5] The DBD domain of *Kp*PriA exhibits a highly basic charge on its surface and it was speculated that this was important for binding the DNA replication fork.[3] As shown in Fig. S1, the electrostatic surface of *Gst*PriA and *Ec*PriA include prominent basic patches presented on the 3'BD and WH. To identify the sequence variation between species, we aligned the PriA sequence from Gram-negative and -positive bacteria. A significant portion of the basic patch on the 3' BD domain is evolutionarily well conserved among PriA proteins (Fig. S1C).

Identifier	Sequence	Modification	Manufacturer
f307	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GGG TAC CGA GCT CGA ATT CAC TGG CCG TCG /3Bio/	3' biotin	IDT
f308	CGA CGG CCA GTG AAT TCG AGC TCG GTA CCC GCC AGC CAC AGT CGT GGC CAT TGC CAT ATG	None	MDBio
f311	/5Alex488N/CAT ATG GCA ATG GCC ACG ACT GTG GCT GGC	5' Alexa Fluor 488	IDT
f312	CGA CGG CCA GTG AAT TCG AGC TCG GTA CCC GCC AGC CAC AGT CGT GGC CAT TGC CAT ATT /3AlexF488N/	3' Alexa Fluor 488	IDT
f313	CGA CGG CCA GTG AAT TCG AGC TCG GTA CCC TTT TTG CCA GCC ACA GTC GTG GCC ATT GCC ATA TG	None	MDBio
f314	AGG GAT CCT CTA GAG TCG ACC TGC AGG CAT	None	MDBio
f315	ATG CCT GCA GGT CGA CTC TAG AGG ATC CCT GGG TAC CGA GCT CGA ATT CAC TGG CCG TCG/3Bio/	3' biotin	IDT

Table S1: DNA oligonucleotide sequences

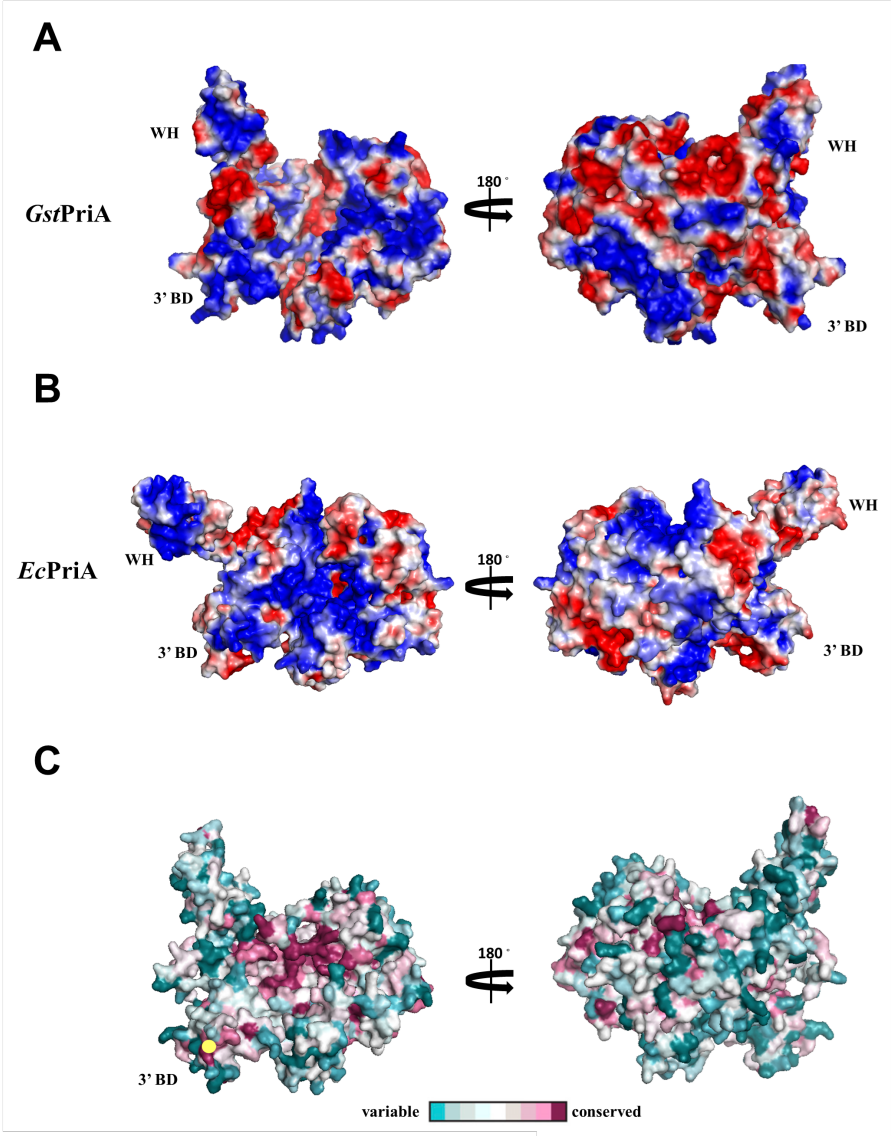

Figure S1: Molecular surfaces showing the electrostatic potentials for *Gst*PriA and *Ec*PriA. (A and B) *Gst*PriA and *Ec*PriA are shown with positive charges indicated in blue and negative charges in red. The 3' BD and WH domains are labeled, respectively. (C) Evolutionary conservation of *Gst*PriA (a conservation scale, from variable to invariable among 5 PriA protein sequences, is shown below the structure). Multiple amino acid sequence alignment of PriA from different bacterial species: Gram-negative bacteria: *E. coli* PriA (*Ec*PriA; GenBank: BAA00491.1), *K. pneumoniae* PriA (*Kp*PriA; GenBank: CDO16358.1), Gram-positive bacteria: *G. stearothermophilus* PriA (*Gst*PriA; GenBank: KZE96245.1), *B. subtilis* (*Bsu*PriA; GenBank: ASK23644.1), *S. pneumoniae* PriA (*Sp*PriA; GenBank: AAL00384.1). The sequence alignment was constructed in the ClustalW and the ConSurf web server.

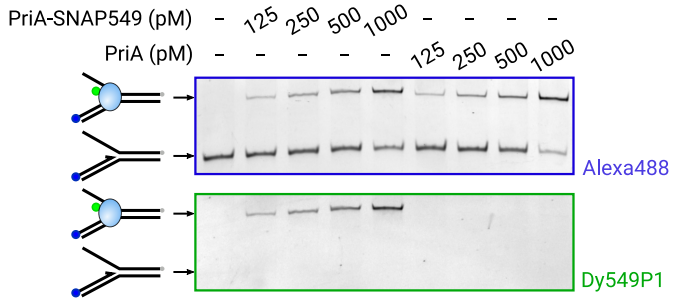

Figure S2: EMSA result of PriA and PriA-SNAP549 on DNA substrate b. The upper panel is the fluorescence signal from Alexa Fluor 488 label on the DNA. The lower panel is the signal from the Dy549P1 label on PriA-SNAP549.

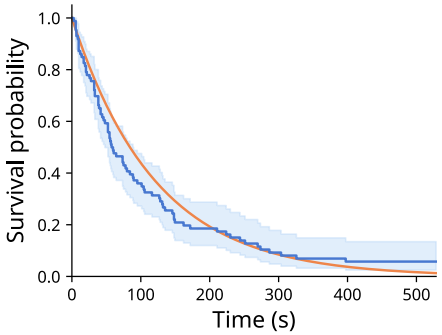

Figure S3: The photobleaching time distribution of Dy549P1. Measured from surface-anchored Dy549P1-labeled DNA. The fitted photobleaching rate is $0.0082\,\mathrm{s}^{-1}$.

References

[1] L. J. Friedman and J. Gelles, *Methods*, 2015, **86**, 27–36.

[2] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede, *Nucleic Acids Res.*, 2014, **42**, W252–W258.

[3] B. Bhattacharyya, N. P. George, T. M. Thurmes, R. Zhou, N. Jani, S. R. Wessel, S. J. Sandler, T. Ha and J. L. Keck, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, **111**, 1373–1378.

[4] K. Sasaki, T. Ose, N. Okamoto, K. Maenaka, T. Tanaka, H. Masai, M. Saito, T. Shirai and D. Kohda, *EMBO J.*, 2007, **26**, 2584–2593.

[5] T. Tanaka, T. Mizukoshi, K. Sasaki, D. Kohda and H. Masai, *J. Biol. Chem.*, 2007, **282**, 19917–19927.