Electronic Supplementary Material (ESI) for Soft Matter.
This journal is © The Royal Society of Chemistry 2022

Electronic Supplementary Information for

Aggregation or Phase Separation can be Induced in Highly Charged Protein by Small Charged Biomolecules

Minchae Kang[a], Minsoo Kim[b], Min Wook Kim[c], Yewon Shin[a], Jejoong Yoo[b,*] and Sang Hak Lee[a,*]

a Department of Chemistry, Pusan National University, Busan 46241, Korea
b Department of Physics, Sungkyunkwan University, Suwon 16419, Korea
c School of Computer Science and Engineering, Pusan National University, Busan 46241, Korea

Materials and Methods

Cell culture

The transformation of GFP plasmid DNA into BL21 causes the BL21 to express the protein. When the OD_{600} of BL21 was > 1, the cell pellet was obtained through centrifugation. After removing the supernatant, we added 0.01% sepiolite and 200 ng of plasmid DNA to the cell pellet and agitated the mixture for 90 seconds. The mixture was spread on an LB agar plate with ampicillin. We incubated the agar plate at 37 ℃ and selected a colony to culture.

The transformed BL21 cells were incubated for longer than 18 hours at 37 ℃. At OD_{600}=0.1, 500uM Isopropyl β-D-1-thiogalactopyranoside (IPTG, Aldrich) was added to an LB medium and incubated for an additional five hours. One hour after adding the IPTG to the cell medium, we added 50uM Carbonyl Cyanide *m*-Chlorophenylhydrazone (CCCP, Aldrich) in order to remove (or reduce) ATP in the cells. To test for the effect of polyamine on protein aggregation, we added 100 mM of polyamine (spermine or spermidine, Alfa Aesar) to the LB medium three hours after adding the CCCP.

Plasmids

The following GFP plasmids were purchased from Addgene: pET-6xHis-pos(28) GFP (addgene #89250), pET-6xHis-pos(16)GFP (addgene #89249) and pET-(neg27)GFP -ggs9-Cre-6xHis (addgene #89253)).

Detecting polyamine

We added Polyamine-RED (FDV-0020, Bio-REV Pte. Ltd.) to the medium 10 minutes before imaging.

Microscopy

We set up an imaging chamber for cell imaging as previously reported.[1] On a cover slip, we dropped 2uL of the cell-containing medium, then covered it with 10% agarose gel (1*1*0.1 cm). The sample was sealed using epoxy glue to prevent it from drying. The imaging chamber with sample was placed under a microscope (TE-2000, Nikon) and images were taken with a 100 x oil immersion objective lens and an iXon EMCCD camera (Andor). To then specifically image the GFP, we used a 488 nm laser (06-MLD, Cobolt) to excite the protein and used an optical filter (HQ592/100M, Chroma) and to specifically image polyamine-red, 561 nm laser (06-DPL, Cobolt) was used and the emission was selectively observed using HQ608/60M (Chroma). We used ImageJ to view and analyze the images.

Cell Counts

To remove any subjective determination in counting viable cells, we used a machine-learning program to discriminate viable from dead cells. We adopted the mid-2020 version of the Poly-YOLO deep learning model for object detection to arrive at the number of active cells. The Poly-YOLO algorithm uses a polygonal grid rather than a square grid, thereby more accurately detecting the objects. We customized a python script from https://gitlab.com/irafm-ai/poly-yolo and trained the model with a test set of 100 cell images. The hyperparameters of the model were Intersection over Union (IoU) with a score of 0.1 and a threshold score of 0.01.

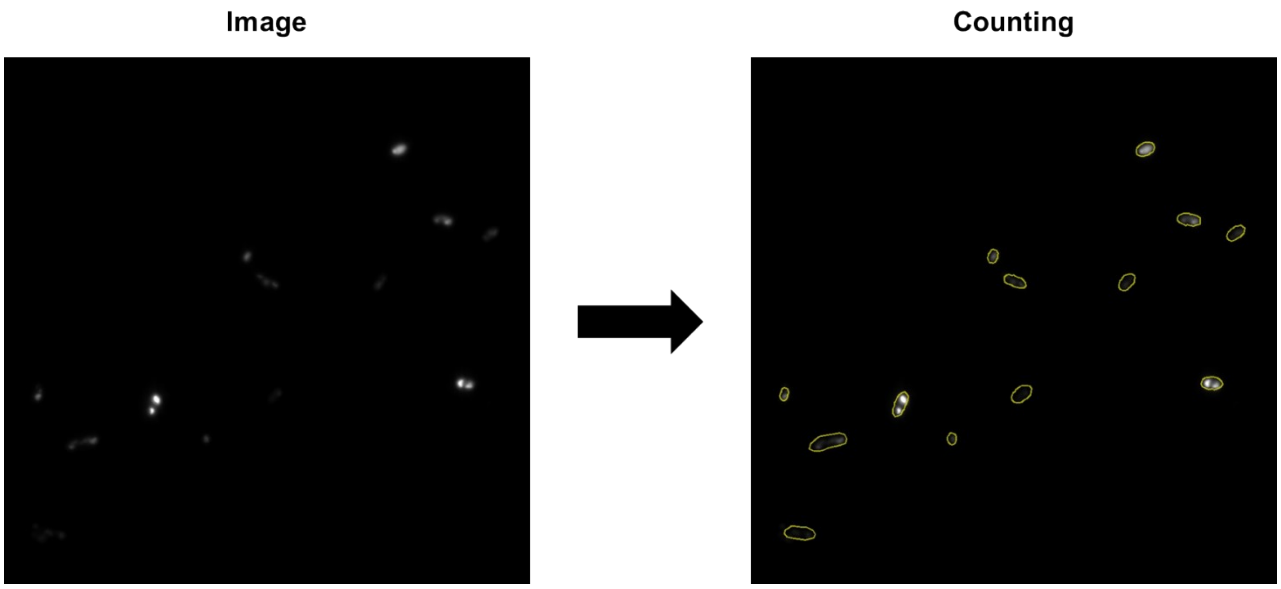

Figure S1. The raw images of GFP-expressed E-coli cells and the analyzed image using Poly-YOLO

Machine learning algorithm

YOLO is a single neural network that is utilized to detect objects within a larger space. YOLO divides a given image into regions (512-512) and predicts bounding boxes and probabilities for each region. If the center of an object falls into a bounding box, that box is responsible for detecting that object. These boxes are weighted according to predicted probabilities. If the probability is close to 1, this means that there is a very high probability that there is an object. YOLO also calculates conditional class probability to determine the class of objects. These probabilities are conditioned on the region containing an object. YOLO can predict one set of class probabilities for each region, regardless of the number of boxes. Thus, YOLO can detect the cells in a given image without any subjective bias. The code is available in the supporting information.

MD simulation protocol

For all MD simulations, we used GROMACS 2020.2 package[2] and Amber ff99sb-ildn-phi force field[3, 4] with the TIP3P water model, combined with the CUFIX corrections that improve the charge-charge interactions of ions,[5] proteins,[6, 7] and nucleic acids.[8, 9] The parameters for spermine and ATP were taken from the previous works by Dai et al.[10] and Meagher et al.[11], respectively. All simulations were performed under a constant isothermal-isobaric (NPT) ensemble [12, 13] at 300K temperature and 1 bar pressure. Van der Waals forces were computed using 10 Å to 12 Å switch scheme. Long-range Coulomb forces were computed using the particle-mesh Ewald (PME) summation method[14] of 12 Å real-space cutoff and 1.2 Å grid spacing. Covalent bonds of hydrogen in non-water and water molecules were constrained using LINCS[15] and SETTLE[16] algorithms. We used 2 fs time step during MD simulation and saved atomic coordinates every 20 ps for analysis.

MD method for the free energy calculations

We manually substituted residues of GFP(2b3p) to construct positively (+28) and negatively (-27) charged GFP. We prepare five systems: two GFP (+28) with 10 spermines and 100 Cl ions; two GFP (-27) with 15 spermines and 12 Cl ions; two GFP (-27) with 20 spermines and 32 Cl ions; two GFP (+28) with 15 ATPs; two GFP (-27) with 10 ATPs and 88 Na ions. All systems are in a periodic triclinic box ($a = b = c \approx 16\,nm,\ \alpha = \beta = 60°, \gamma = 90°$). We energy-minimized each system for 2500 steps and equilibrated it for 80 ns with the harmonic position restraints using a force constant of 1,000 kJ/mol. For the measurements of the free energy, we used the umbrella sampling method, varying the distance between the center of masses from 3 nm to 6 nm with a 0.3-nm spacing. For each window, starting from the pre-equilibrated system, we performed the simulation for 130 ns. Because the distance between proteins was adjusted to the target distance within 10 ns at most, we discarded the first 10 ns of each window from the analysis. For example, see Figure S5a for the distribution of distance in each window in the case of GFP (-27) in the presence of spermine. Using the samplings, we computed the free energy profiles in two different ways: the weighted histogram analysis method (WHAM) included in the Gromacs package and the integration of window-averaged forces with respect to the distance. Because two methods result in almost

identical results (see Figure S5b for example), we used the outcomes of the force integration in the main text. To estimate the error bars, we computed four free energy curves by dividing the 120-ns trajectories into four non-overlapping trajectories and computed the standard errors of four free energy curves.

Results

Cell counts

GFP Charge	16		28		-27	
	Cell	Aggregate	Cell	Aggregate	Cell	Aggregate
IPTG	303	278	1094	577	576	90
CCCP	601	188	623	254	965	638
Spd 100mM	447	128	670	351	601	562
Spm 100mM	364	134	405	195	336	242

Table 1 Cell counts and the number of those cells in which one or more protein aggregate is present.

Internal cell aggregate counts

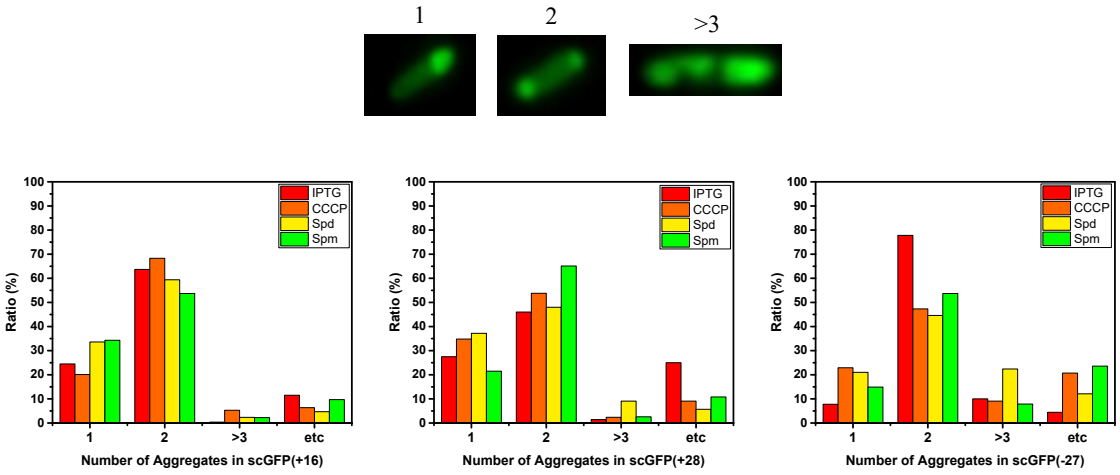

Figure S2 Counting protein condensate spots in a cell. Fluorescence images on top show the example of each protein condensate spot. The histogram results from scGFP (+16, +28, -27) show that the majority are one or two spots.

FRAP (Fluorescence Recovery After Photobleaching)

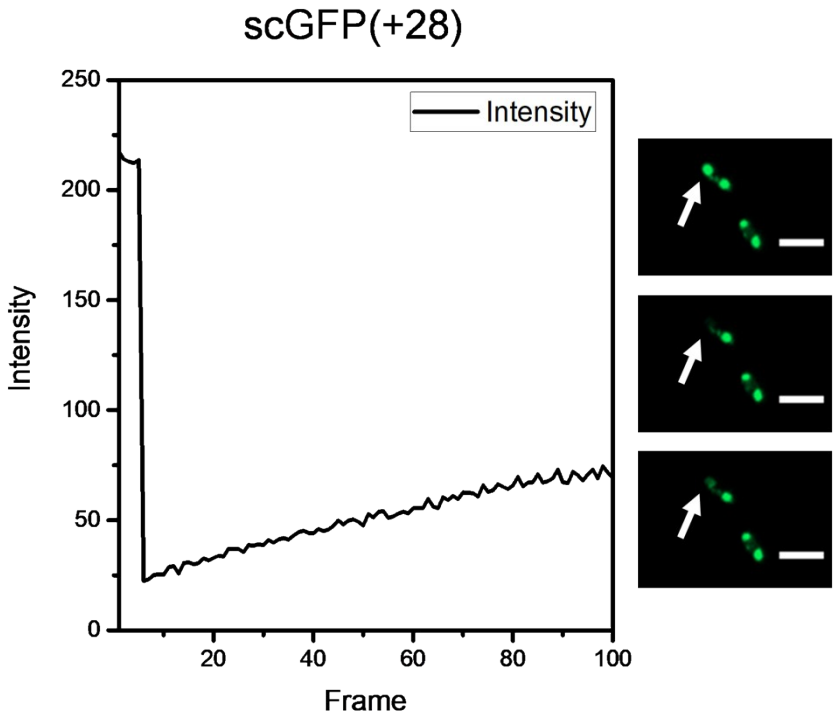

Figure S3 Representative Fluroescence recovery after photobleaching result. The trace shows that fluorescence intensity very slowly and partially recovered, but it did not show full recovery.

High Performance Liquid Chromatography (HPLC) Methods

We cultured the cells in 5 mL media and added 2M NaOH 1.5mL (Sigma), raising the pH over 13. To make chemical benzoylation, we added 5ul of benzoyl chloride (Sigma, 259950) and stirred the mixture for 20 minutes to complete the reaction. We added 2mL of saturated NaCl (Sigma, S9888) solution to the mixture. We dissolved the mixture in ether and filtered the mixture in a column filter two times and then took the sample through the evaporation stage with nitrogen. The sample was again dissolved in 30% acetonitrile/water. The solution was mounted in an HPLC to separate benzoyl-polyamine. Reference spermidine were dissolved in distilled water, derivatized with NaOH and benzoyl chloride, and separated using the same procedure.[17]

HPLC results

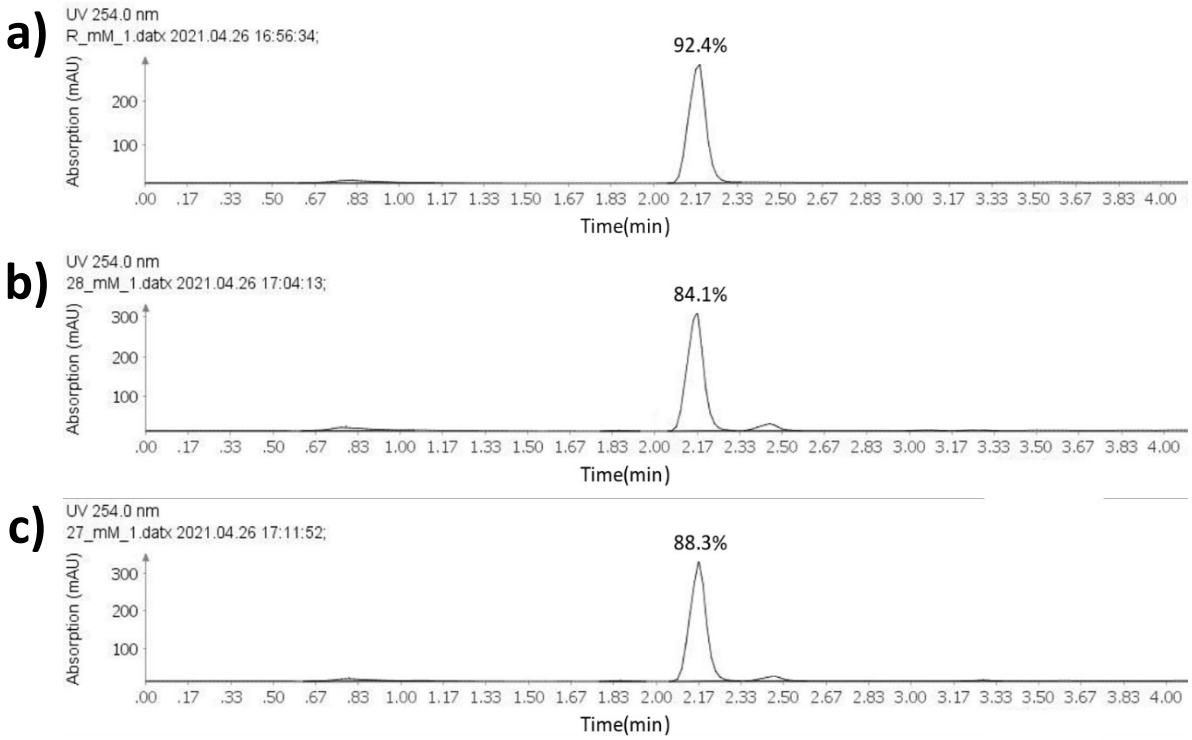

Figure S4 HPLC results after treating the cells with 100mM spermidine. We separated benzoylated spermidine on a reverse-phase column using a solution of 30:70 acetonitrile/water as mobile phase and detected by UV absorption at 254 nm. Results are shown: a) reference, b) positive 28 charge GFP, and c) negative 27 charge GFP. From 92.4% of 100mM spermidine in the reference, the percentage decreased to 84.1% and 88.3% after cell culture.

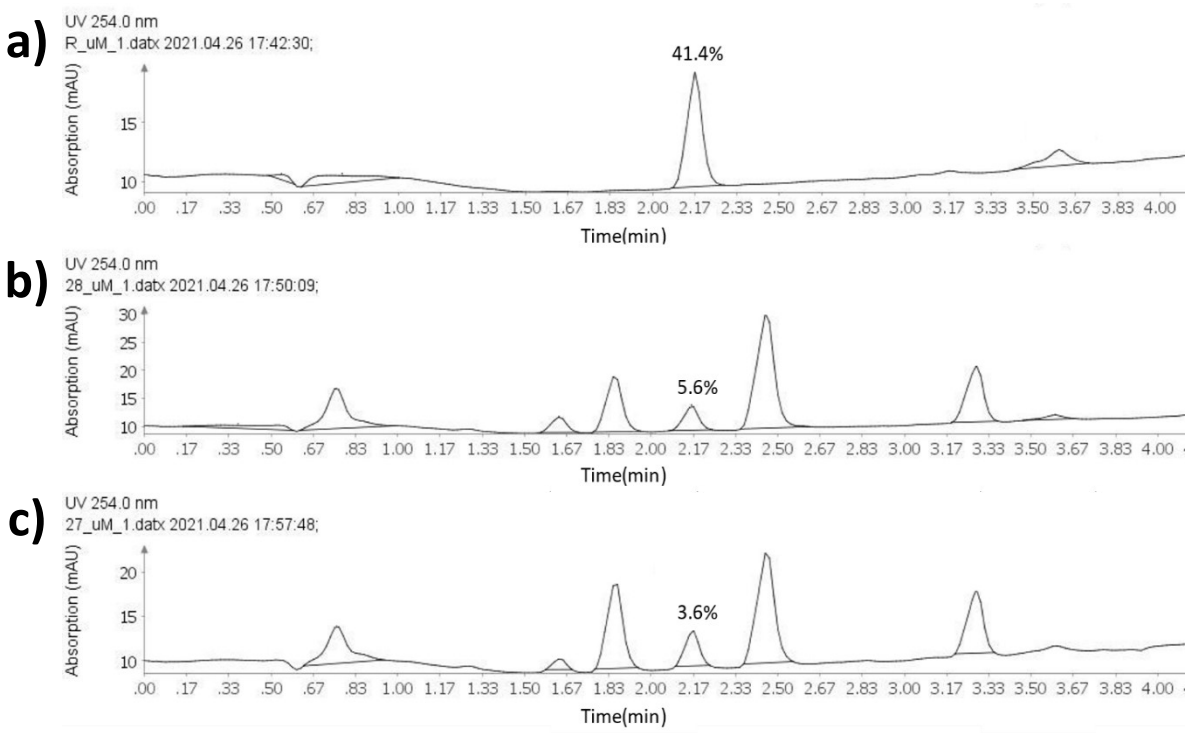

Figure S5 Result of HPLC after treating a micromolar scale of spermidine to cell, showing a) reference, b) positive 28 charge GFP and c) negative 27 charge GFP. The percentage of 100uM spermidine decreased from 41.4% in the sample to 5.6% and 3.6% after cell culture.

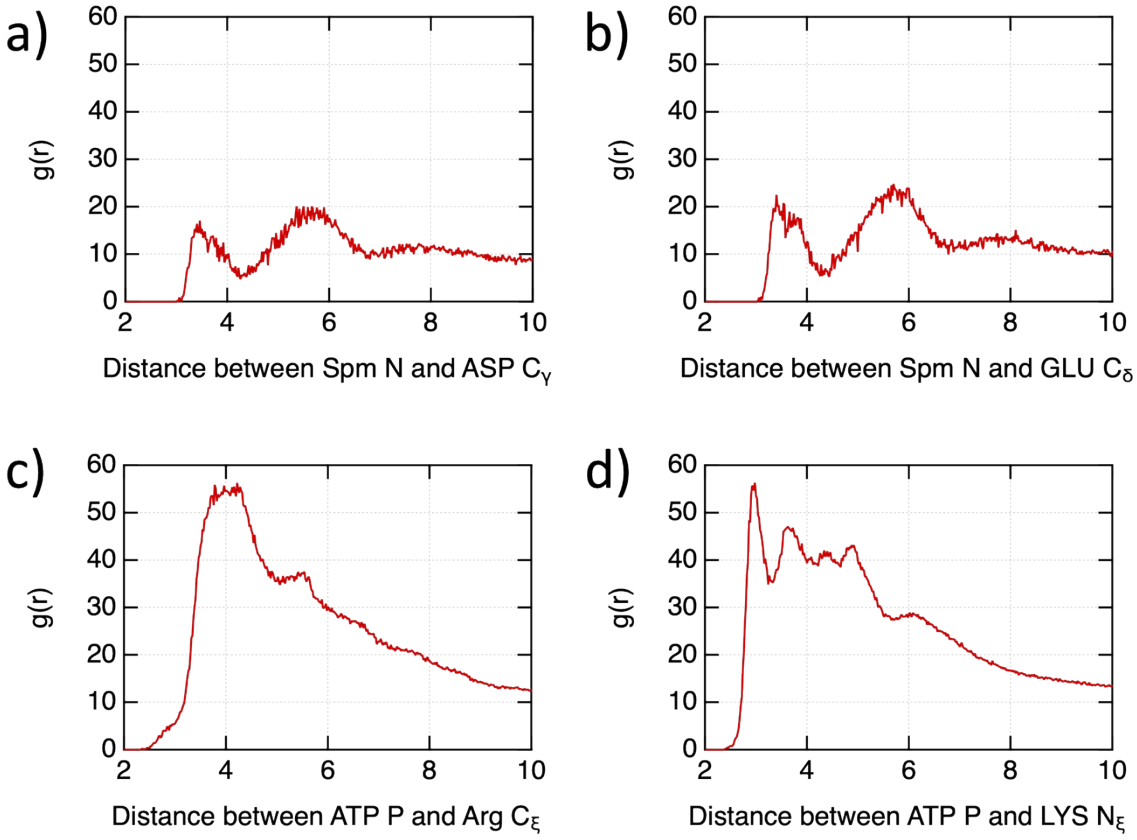

Figure S6 Interaction of spermine and ATP molecules with charged residues of GFP. a,b) Radial distribution functions of carboxylate carbon atoms of Asp (a) and Glu (b) with respect to the nitrogen atoms of spermine. c,d) Radial distribution functions of guanidinium carbon atoms of Arg (c) and amine nitrogen atoms of Asp (d) with respect to the phosphorus atoms of ATP.

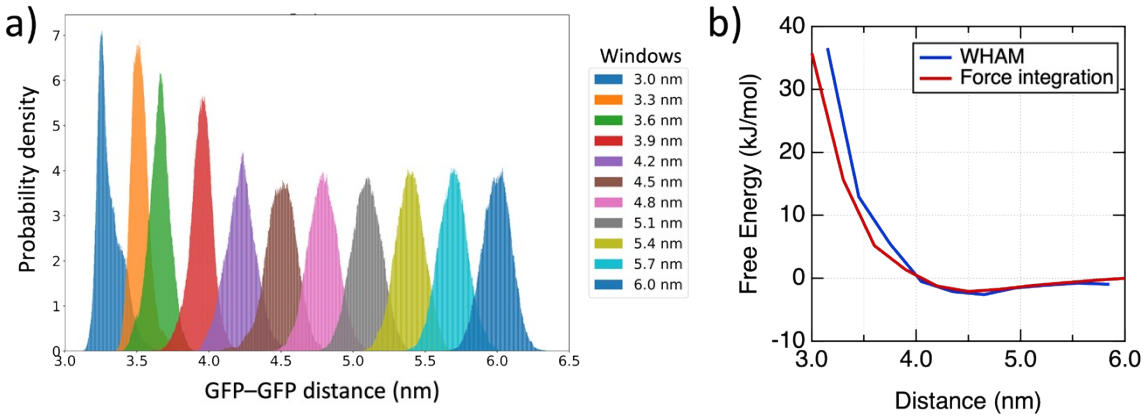

Figure S7 Computation of the free energy using the Umbrella sampling method. a) The probability density functions of distance in each window in the simulations of a pair of GFP (-27) proteins in the presence of spermine. b) The comparison of two free energy profiles using the WHAM method (blue curve) and the force integration method (red curve).

Reference

1. S. O. Skinner, L. A. Sepulveda, H. Xu and I. Golding, *Nat. Protoc.*, 2015, **10**, 1457-1457.
2. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.
3. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J Am Chem Soc*, 1995, **118**, 2309-2309.
4. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950-1958.
5. J. J. Yoo and A. Aksimentiev, *J Phys Chem Lett*, 2012, **3**, 45-50.
6. J. Yoo and A. Aksimentiev, *J Phys Chem Lett*, 2016, **7**, 3812-3818.
7. J. Yoo and A. Aksimentiev, *J Chem Theory Comput*, 2016, **12**, 430-443.
8. J. Yoo and A. Aksimentiev, *Phys Chem Chem Phys*, 2018, **20**, 8432-8449.
9. J. Yoo and A. Aksimentiev, *Nucleic Acids Res*, 2016, **44**, 2036-2046.
10. L. Dai, Y. G. Mu, L. Nordenskiold and J. R. C. van der Maarel, *Phys Rev Lett*, 2008, **100**.
11. K. L. Meagher, L. T. Redman and H. A. Carlson, *J Comput Chem*, 2003, **24**, 1016-1025.
12. M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182-7190.
13. S. Nosé and M. L. Klein, *Mol. Phys.*, 2006, **50**, 1055-1076.
14. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
15. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463-1472.
16. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952-962.
17. Y. H. Mei, *J. Liq. Chromatogr.*, 1994, **17**, 2413-2418.