Supporting Information

Reduced native state stability in crowded cellular environment due to protein-protein interactions

Ryuhei Harada[a], Naoya Tochio[b], Takanori Kigawa[b], Yuji Sugita[a,c,d], and Michael Feig[c,e,*]

[a]RIKEN Advanced Institute for Computational Science, 7-1-26 Minatojima-minami machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan, [b]RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan, [c]RIKEN Quantitative Biology Center, International Medical Device Alliance (IMDA) 6F, 1-6-5 Minatojima-minami machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan, [d]RIKEN Advanced Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, [e]Department of Biochemistry & Molecular Biology and Department of Chemistry, Michigan State University, East Lansing, MI, 48824, United States.

Experimental methods

Protein expression and purification

The DNAs encoding the headpiece subdomain (the amino acid sequence from L792 to F826, referred to from L2 to F36 in this paper; villin) of the chicken villin protein (SwissProt accession number P02640) and the segment B1 (from T228 to E282, referred to from T2 to E56 in this paper; protein G) of streptococcal protein G (P06654) were cloned into the expression vector pCR2.1 (Invitrogen, USA), as a fusion with an N-terminal histidine tag and a SUMO (small ubiquitin-related modifier) protein using the two-step PCR(1). After cleavage of the tag, these samples contained one extra Met residue at each N-terminus. The ^{15}N-labeled (villin) and non-labeled (protein G) fusion proteins were synthesized by the cell-free protein synthesis system(2-4), and were firstly purified using a HisTrap HP Column (GE Healthcare, USA). Tag sequence was cleaved by SUMO protease and then separated using HisTrap HP column (GE Healthcare). The isolated protein G and villin were further purified using HiTrap Q or HiTrap SP columns (GE Healthcare), respectively, dialyzed with water, and lyophilized. For NMR spectroscopy, we have prepared the variety of protein samples; 1 and 32 mM ^{15}N labeled villin, and 32 mM ^{15}N labeled villin with 16 mM non-labeled protein G in 50 mM HEPES buffer (pH 7.0) containing 10% D_2O. The pH of all the NMR samples were adjusted so as to the final values were between pH7.0 and pH7.1 before lyophilization.

NMR spectroscopy

All NMR spectra were recorded at 298 K on a 600 MHz Bruker Avance spectrometer equipped with CryoProbe. The backbone chemical shift assignments of villin were accomplished by using 3D HNCA and HNCO spectra(5). All spectra were processed using NMRPipe(6), and the programs Kujira(7) and NMRView(8) were employed for optimal visualization and spectral analyses.

Figure S1 The three dimensional structures and amino-acid sequences of protein G and villin headpiece subdomain (villin). The structures of (a) protein G and (b) villin are shown with secondary structures. Purple and yellow regions indicate α-helices and β-sheets. The amino acid sequences of (c) protein G and (d) villin are shown with acidic (red), basic (blue), polar (green) and hydrophobic (black) residue types.

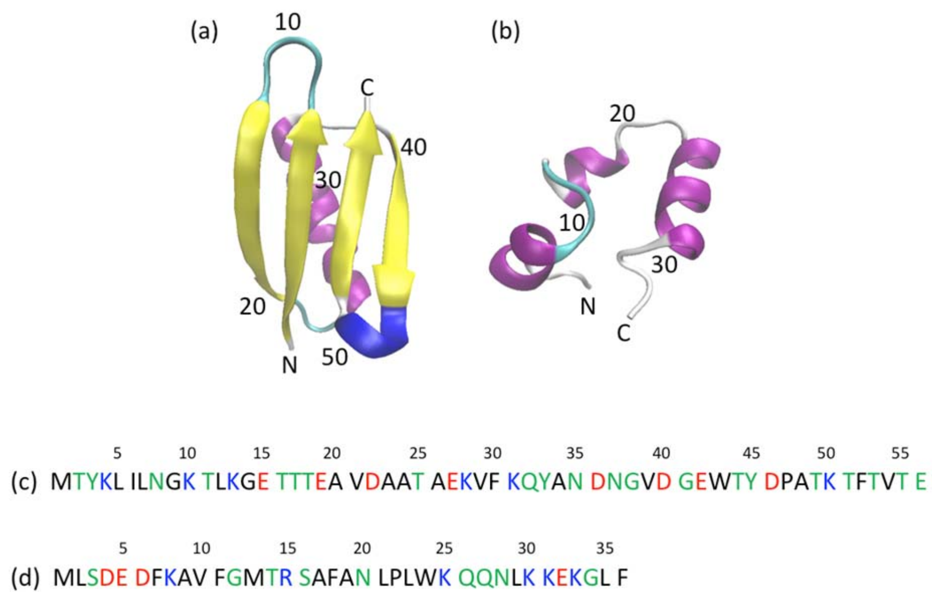
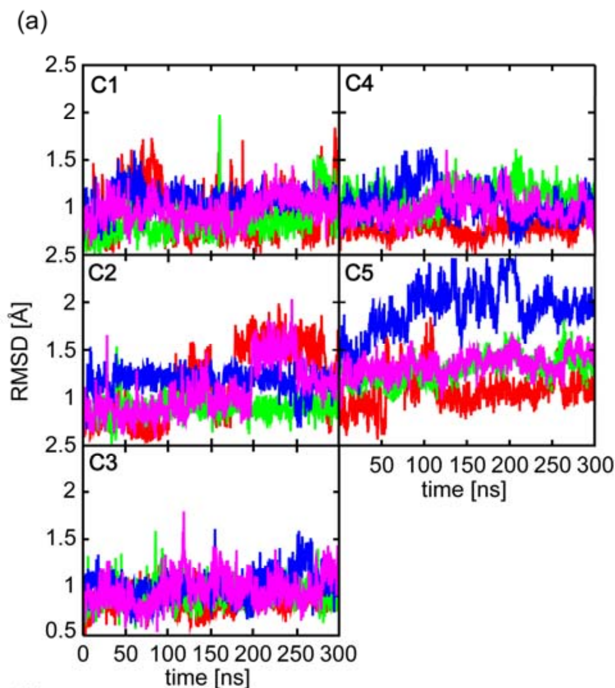


Figure S2 (a) RMSD from native structures of protein-G and (b) villin during 300 ns at different protein concentrations (C1-C5). Each MD simulations contained 4 protein-G and 8 villin and thus 4 (red, blue, green and magenta) and 8 (red, blue, green, magenta, yellow, orange, light blue and black) lines are shown in (a) and (b).



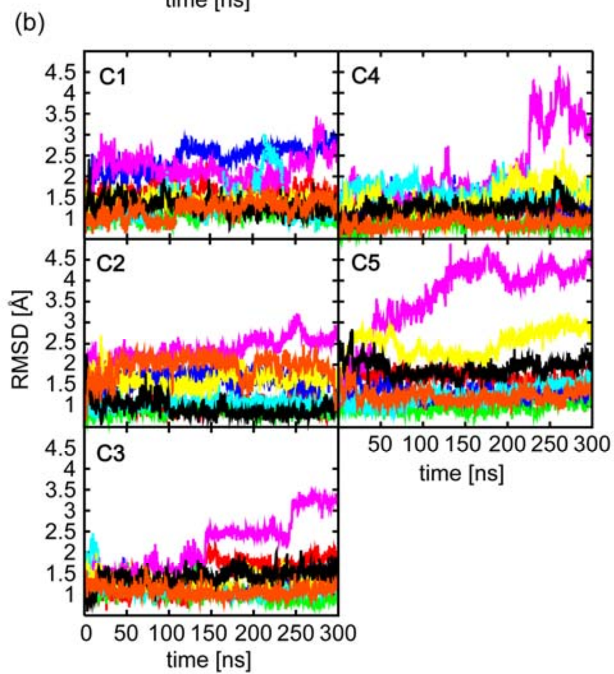


Figure S3 Potential of mean force (PMF) of protein G as a function of RMSD and Rg at non-crowded (NC) and different crowded conditions (C1-C5). The free energy values are scaled by $k_B T$. The native region is defined with values of 0 Å < RMSD < 1 Å and 10 Å < Rg < 11 Å.

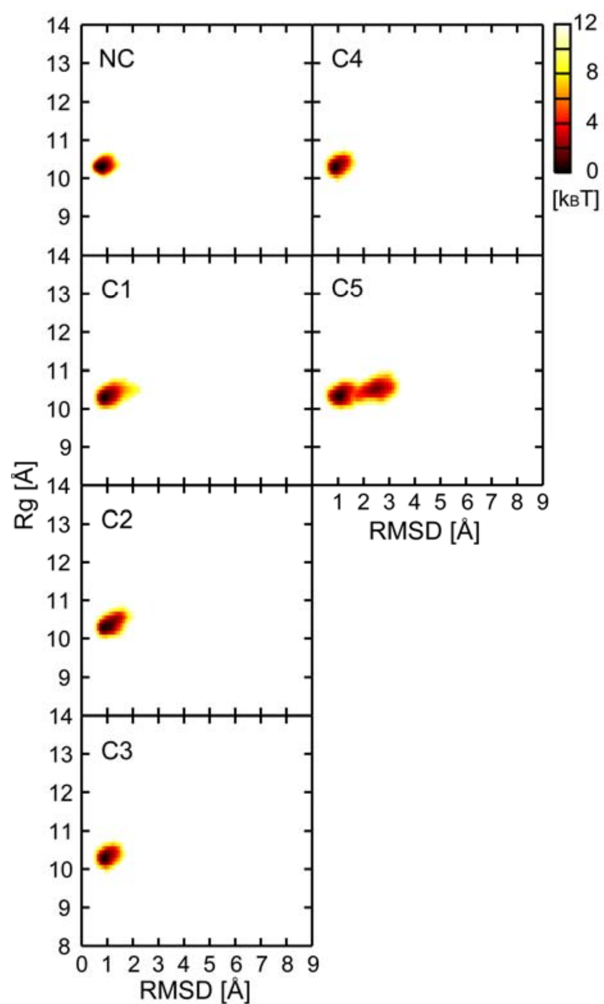


Figure S4 α-helical content for each residue from the crowding simulations: C1 (red), C2 (green), C3 (blue), C4 (purple) and C5 (light blue), respectively. Secondary structures were identified by the STRIDE algorithm(9).

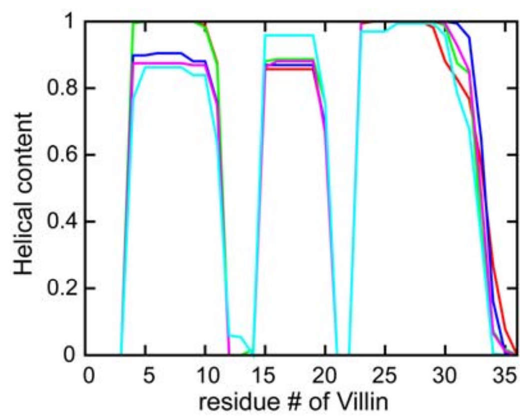


Figure S5 The destabilized structure of protein G (red) under the most crowded environment (C5) superimposed onto the native structure (green).

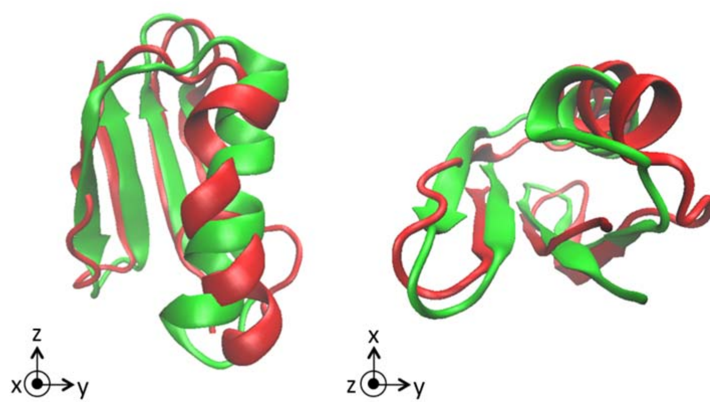


Figure S6 PMF at the most crowded condition (C5) projected onto the free-energy as a function of RMSD and Rg compared with those from the high temperature MD simulation at 500 K (HT) during 300 ns and the urea denaturation MD simulation during 300 ns (Urea). All values of free energy are scaled by k_BT. The structures from high temperature MD simulations (green) with minimum RMSDs after superimposition to LM1-LM4 (red) are also shown.

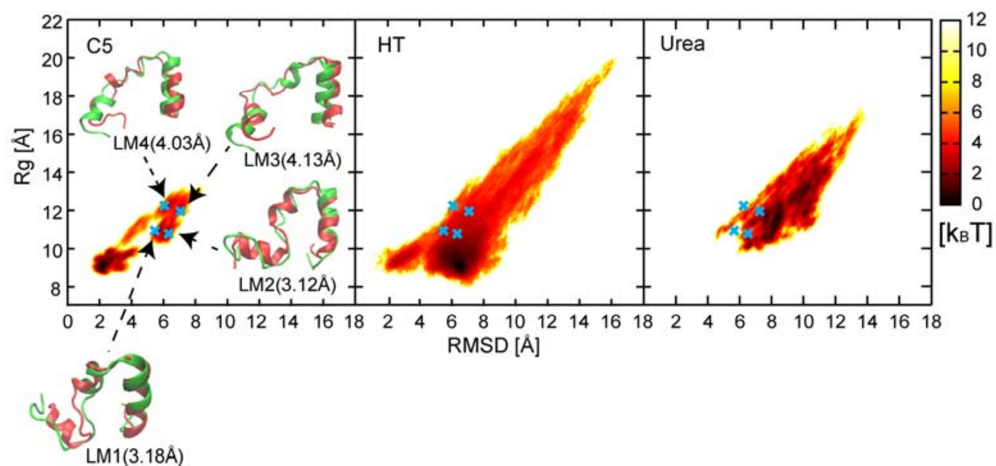
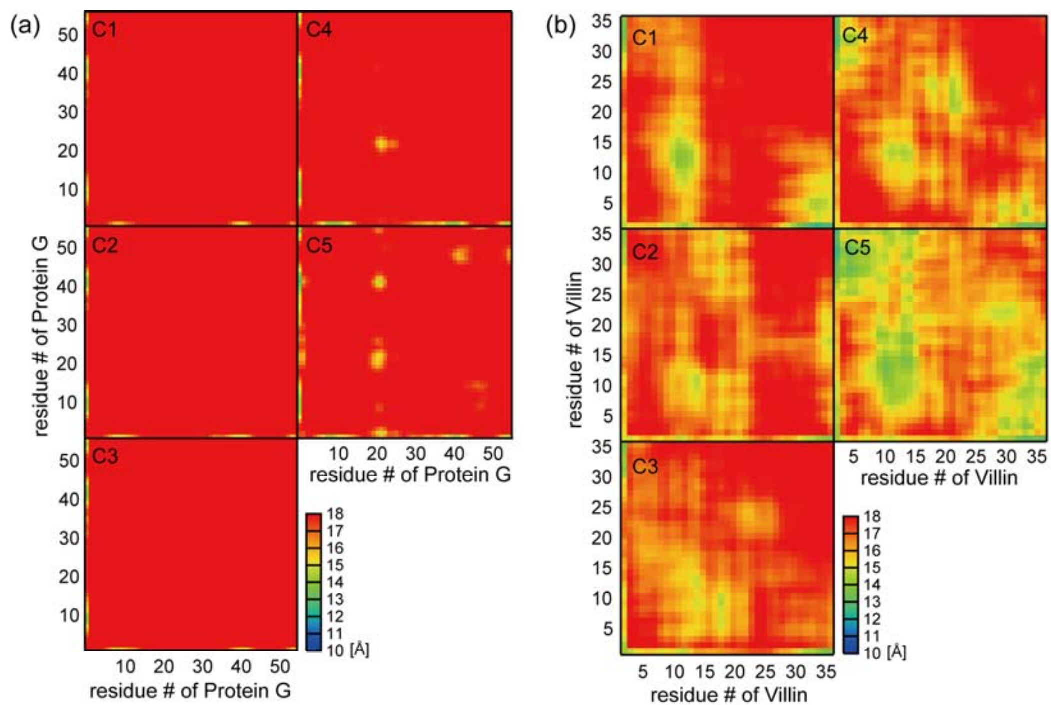


Figure S7 (a) The correlation maps of residues between protein G and protein G, (b) between villin and villin for different protein concentrations (C1-C5). Each color indicates a minimum distance between C_α atoms of residues averaged over the 300 ns simulations.



References:

1. Yabuki T, *et al.* (2007) A robust two-step PCR method of template DNA production for high-throughput cell-free protein synthesis. *J Struct Funct Genomics* 8:173-191.
2. Yokoyama J, Matsuda T, Koshiba S, Tochio N, Kigawa T (2011) A practical method for cell-free protein synthesis to avoid stable isotope scrambling and dilution. *Analytical Biochemistry* 411:223-229.
3. Matsuda T, *et al.* (2007) Improving cell-free protein synthesis for stable-isotope labeling. *J Biomol NMR* 37:225-229.
4. Kigawa T, *et al.* (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* 442:15-19.
5. Ikura M, Kay LE, Bax A (1990) A Novel-Approach for Sequential Assignment of H-1, C-13, and N-15 Spectra of Larger Proteins - Heteronuclear Triple-Resonance 3-Dimensional Nmr-Spectroscopy - Application to Calmodulin. *Biochemistry* 29:4659-4667.
6. Delaglio F, *et al.* (1995) Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J Biomol NMR* 6:277-293.
7. Kobayashi N, *et al.* (2007) KUJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. *J Biomol NMR* 39:31-52.
8. Johnson BA, Blevins RA (1994) Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data. *J Biomol NMR* 4:603-614.
9. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* 32:W500-W502.