Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information

pH-Dependent Cooperativity and Existence of a Dry Molten Globule in the Folding of a Miniprotein BBL

Zhi Yue, and Jana Shen*
Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201-1075, USA
Corresponding author: jana.shen@rx.umaryland.edu

Data analysis

All analyses were conducted by CHARMM[1] (version c37b1) or GROMACS[2] (version 4.5.5). The trajectory snapshots were aligned to the backbone atoms of the NMR model (first entry in the PDB 1W4H[3]).

RMSD of the backbone atoms (N, C_α, C), excluding the unstructured tails (residues 126–132, 170) were calculated following least-square fitting to the corresponding backbone of the NMR model[3]. RMSD was calculated using the CHARMM command COOR rms or the GROMACS tool g_rms.

The hydrophobic core used to calculate R_g is comprised of Ile135, Leu138, Leu139, Leu144, Ala146, Ala148, Ile149, Leu158, Vla163, His166, and Leu167[4–6]. Calculation used the CHARMM command CORREL or the GROMACS tool g_gyrate.

α-helices span residues 133-141 (NT-helix) and 160-168 (CT-helix), respectively. The interhelical angle was calculated using the CHARMM command COOR helix.

Secondary structure was identified using the GROMACS tool do_dssp based on the DSSP algorithm[7]. The helicity of a residue is defined as the frequency of occurrence of the α-helical conformation.

Clustering was performed using the GROMACS tool g_cluster based on the GROMOS algorithm[8]. Backbone atoms (N, C_α, C) excluding the unstructured tails (residues 126–132, 170) were used for the least-squares fit and RMSD calculation in cluster analysis. Large cutoff such as 3.5 or 5 Å was used to split wide-separated states. To reveal finer differences, cutoff of 1.5 or 1 Å was used.

Two non-neighboring residues i and j ($j > i+1$) are defined as in contact if the center-of-geometry (COG) distance between side-chain heavy atoms (C_α for Gly residues) is closer than 6.5 Å[9]. A contact was considered native if it occurred with at least 50% probability in the control simulation. With native contacts defined, q can be calculated using the CHARMM command COOR dmat for CHARMM trajectories and an in-house Tcl script for VMD[10] for GROMACS trajectories.

Supplementary Tables and Figures

Table S1: Comparison of published NMR models of BBL[a]

PDB	Sequence	Mutation	NMR conditions			RMSD (Å)	R_{g} (Å)	Reference
			T (K)	pH	Salt (mM)			
1W4H	126–170	WT	298	7.0	200	0.0	6.0	Fersht[3]
2CYU	131–169	WT	278	5.3	0	1.8	7.0	Munõz[11]
1BBL	131–167	WT	298	5.3	-	1.6	6.8	Gronenborn[4]
1BAL	121–170	WT	298	5.3	-	1.7	6.8	Gronenborn[4]
2BTH	126–170	H166W	298	6.5	200	0.9	6.2	Fersht[3]
2BTG	126–170	H166W	298	6.5	200	0.8	6.3	Fersht[3]
2WAV	126–170	H142W	298	7.0	200	0.9	6.2	Fersht[12]
2WXC	126–170	H142W	298	7.0	200	1.1	6.1	Fersht[12]

[a]RMSD refers to the root-mean-squared deviation of backbone atoms of residues 133–167 with respect to the NMR model (first entry in PDB 1W4H).

Table S2: Summary of calculated pK_{a} values for BBL[a]

Residue	Experiment [b]	p$K_{\mathrm{a}}^{\mathrm{N}}$	p$K_{\mathrm{a}}^{\mathrm{DMG}}$	p$K_{\mathrm{a}}^{\mathrm{D}}$
D129	3.88±0.02	3.2	3.4	3.4
E141	4.46±0.04	3.5	3.5	3.5
H142	6.47±0.04	6.8	7.3	6.9
D145	3.65±0.04	3.2	3.5	3.3
E161	3.72±0.05	3.4	3.1	3.5
D162	3.18±0.04	1.0	2.8	3.4
E164	4.50±0.03	3.4	3.7	3.2
H166	5.39±0.02	6.0	6.4	6.7

[a]N, DMG, and D denote native, dry molten globule, and denatured state, respectively. Titration was also performed for all Lys and Arg residues: Lys150, Lys160, Lys165, Lys169, Arg136, Arg137, Arg157, Arg160. Their pK_{a}'s are not listed here as they were protonated at all pH conditions in the duration of simulation. [b]Experimental pK_{a}'s were determined by Fersht and coworkers using NMR titration[13,14].

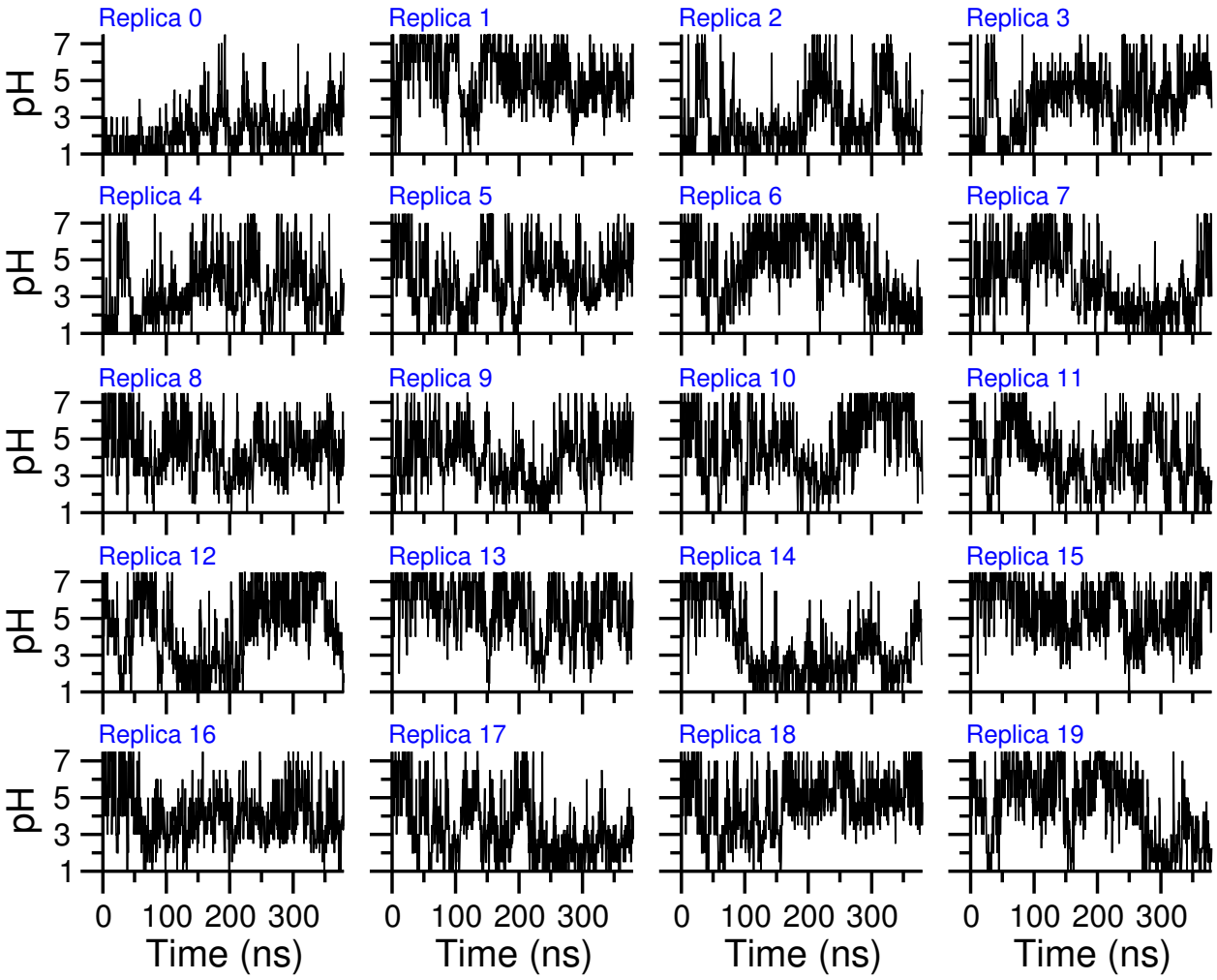

Figure S1: **Multiple replicas walked through the entire pH ladder.** pH condition for each replica as a function of simulation time. Sampled every 500 ps.

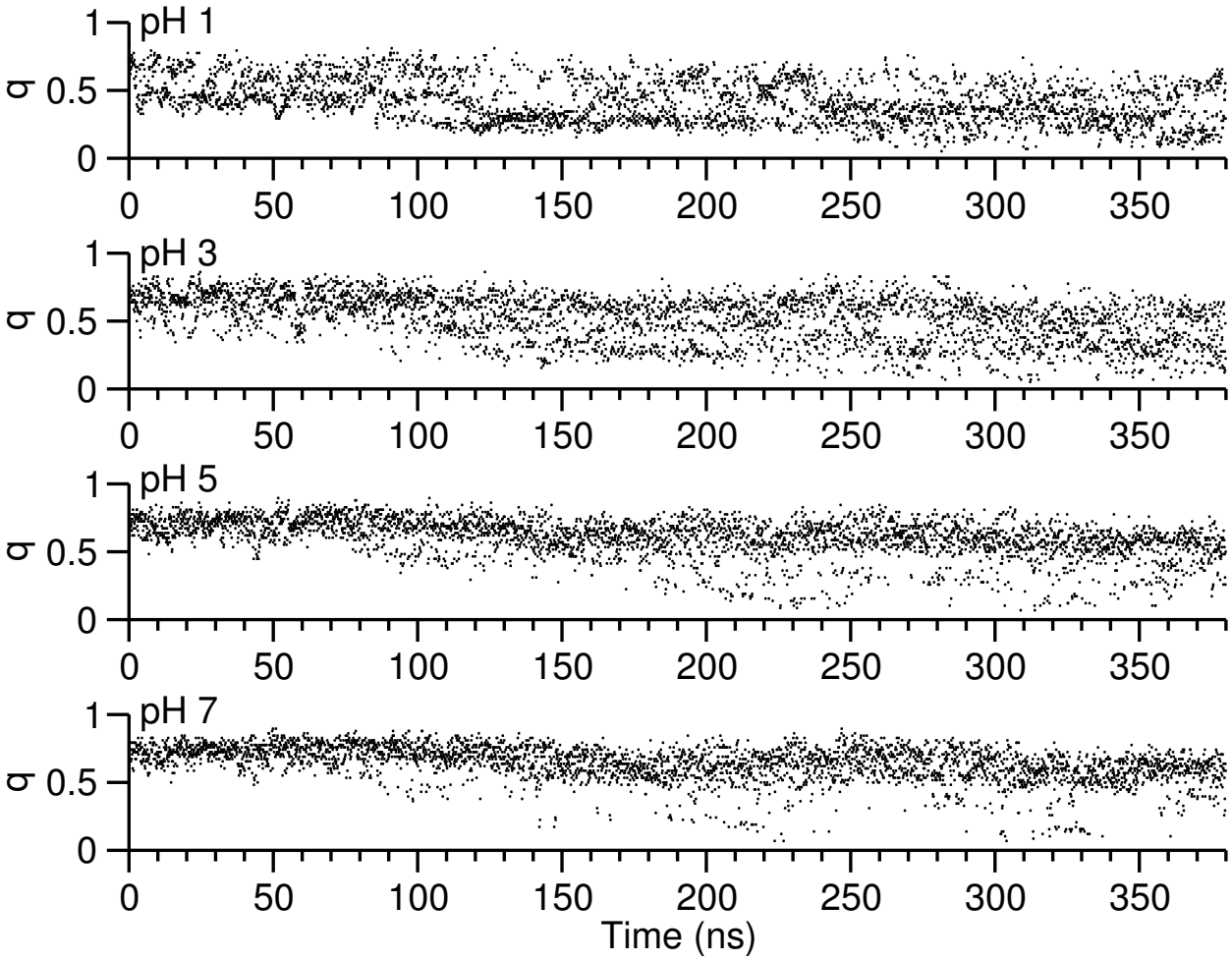

Figure S2: **Fraction of native contacts is converged.** The fraction of native contacts at different pH conditions as a function of simulation time. Definition is given in the main text.

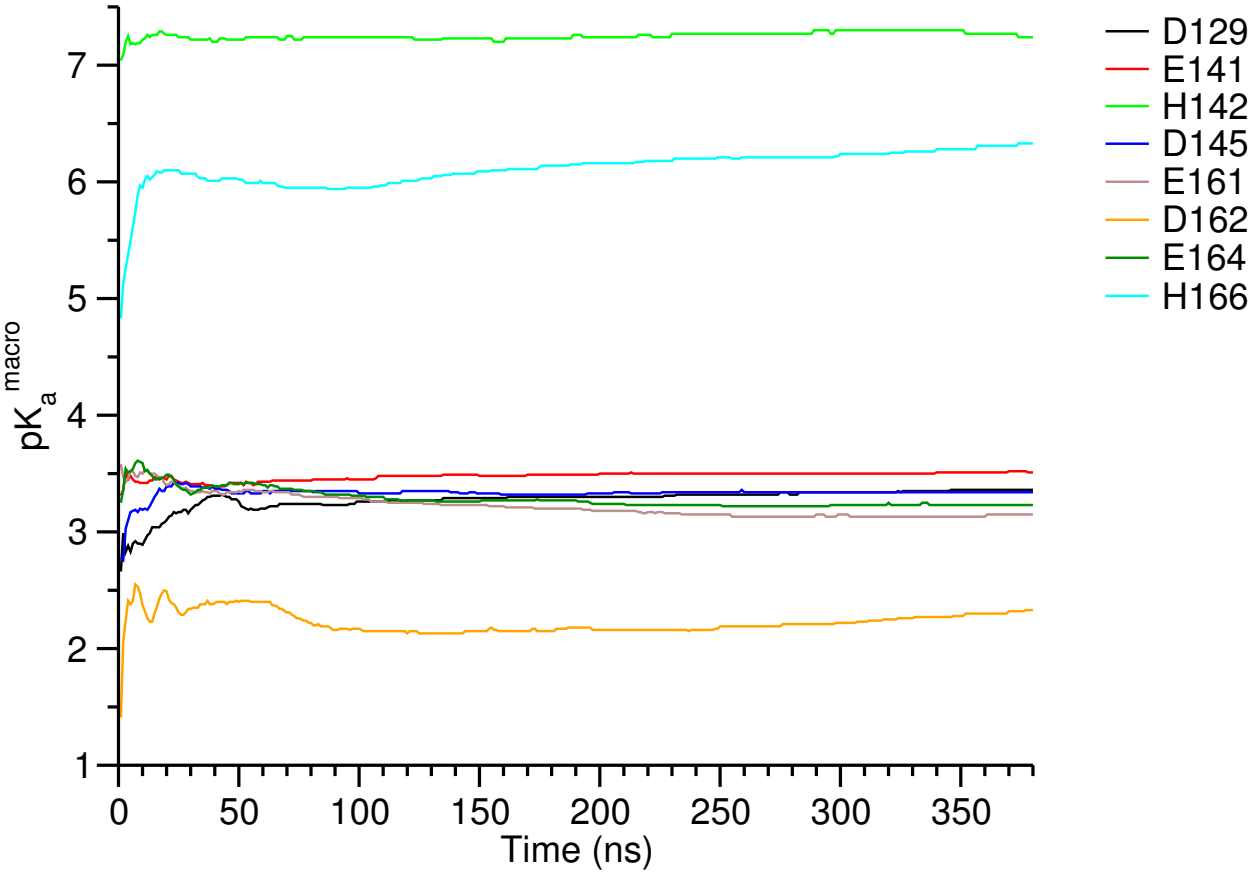

Figure S3: **Macroscopic pK_a's are converged.** Cumulatively calculated pK_a's for all Asp/Glu/His residues as a function of simulation time. Lys and Arg sidechains were always protonated at all pH conditions and are thus not shown in the figure.

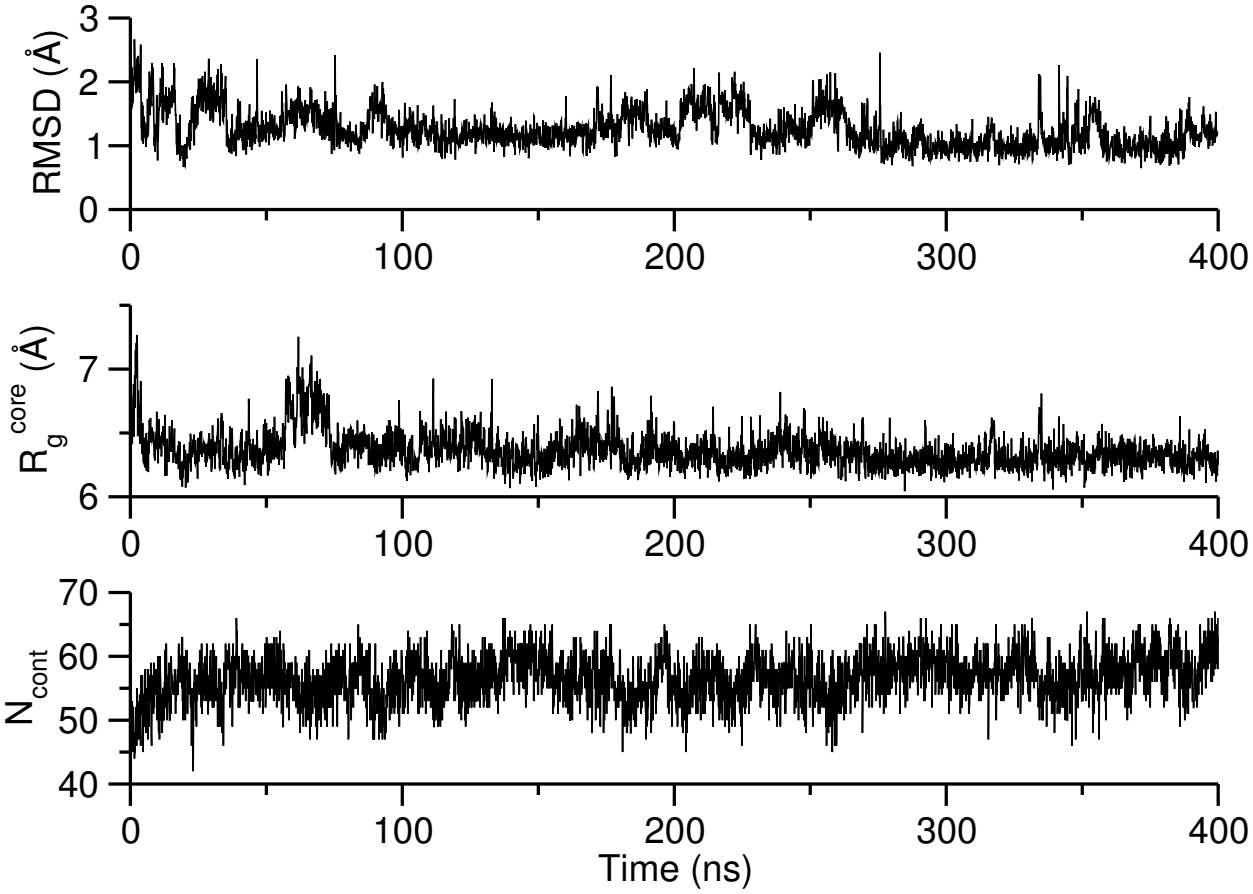

Figure S4: **Stability and convergence of the control simulation.** Time series of the backbone RMSD (top), radius of gyration of the hydrophobic core, and the total number of side-chain contacts (bottom). Sampled every 100 ps.

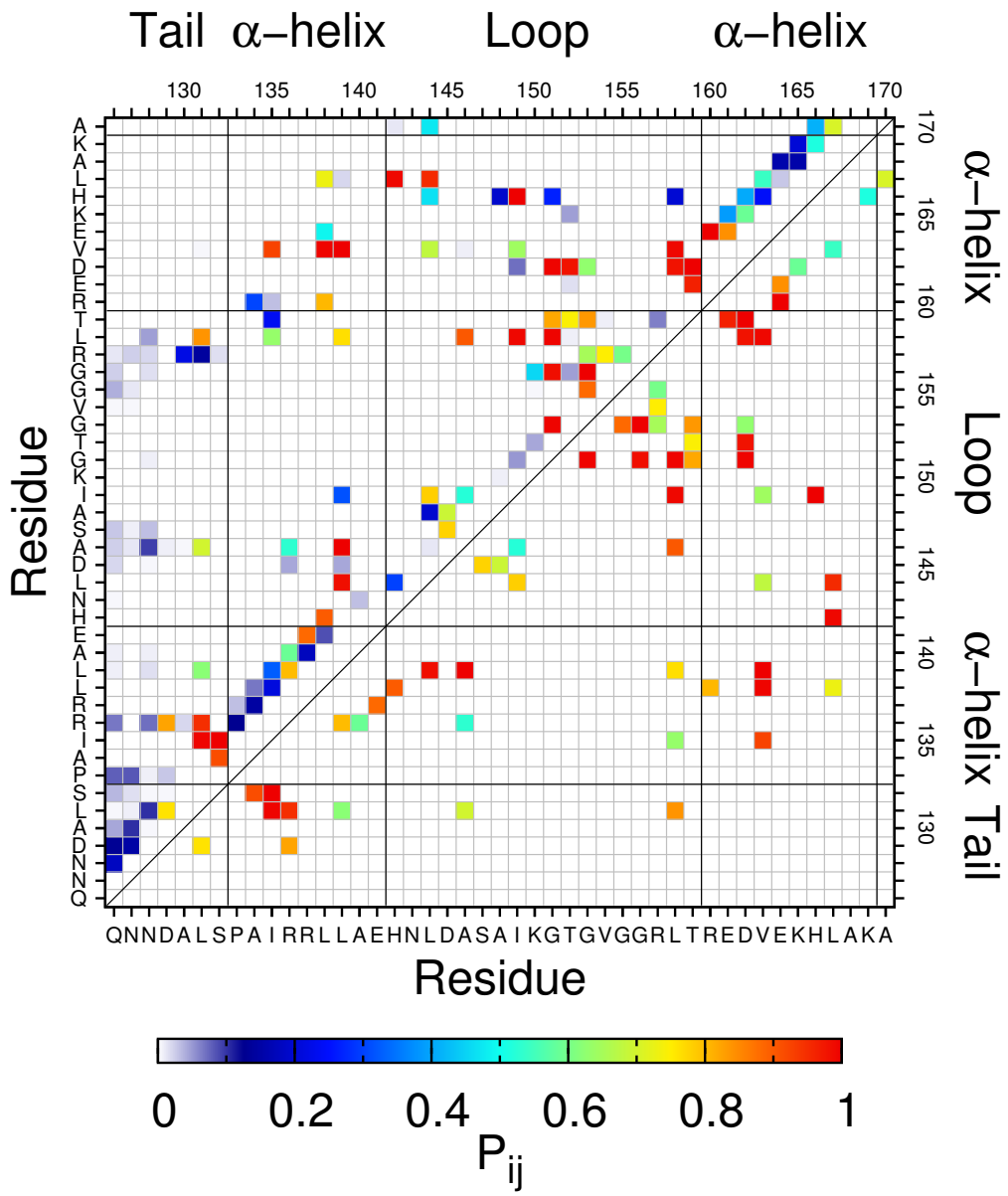

Figure S5: **Native contacts in BBL.** Upper triangle: Map of contacts. The contact occupancy is color coded from blue to red. Two residues are considered in contact if the center of geometry distance between the side chain heavy atoms is within 6.5 Å. Lower triangle: Map of native contacts. A contact is considered native if the occupancy is above 0.5.

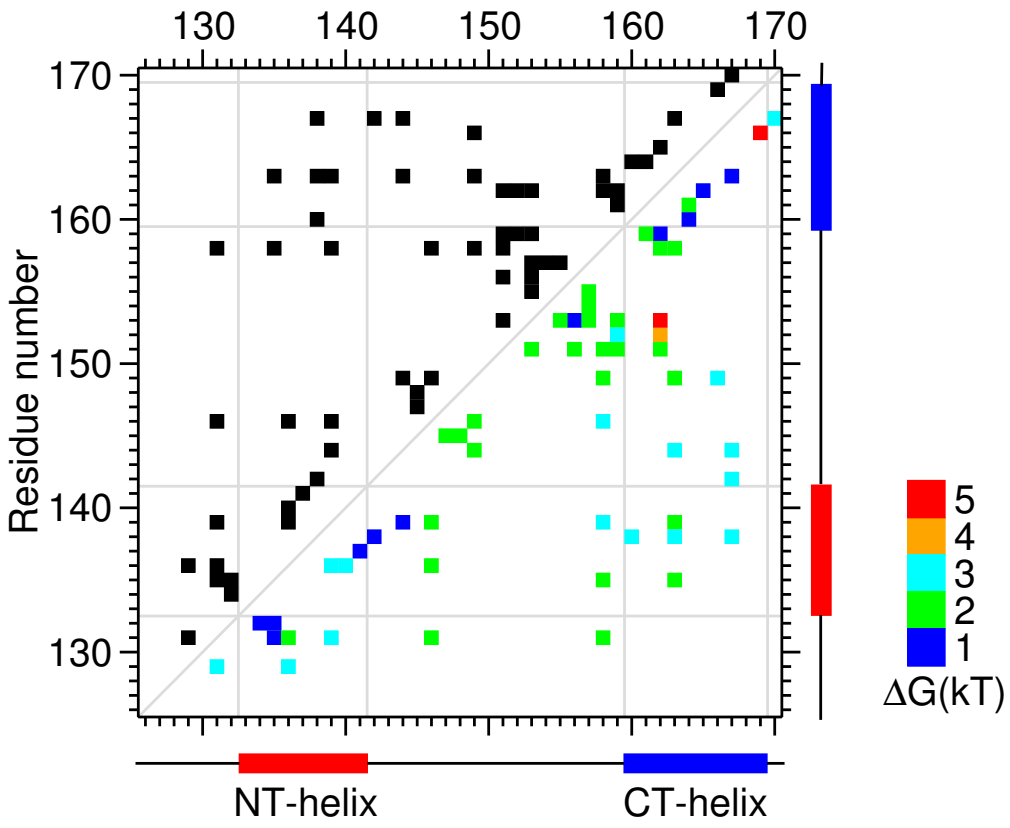

Figure S6: **Native contacts in the denatured state of BBL.** Upper triangle: Map of native contacts. Lower triangle: Destabilization of native contacts in the denatured state. A destabilization free energy, defined as the contact occupancy relative to that in the native state, is calculated for each contact.

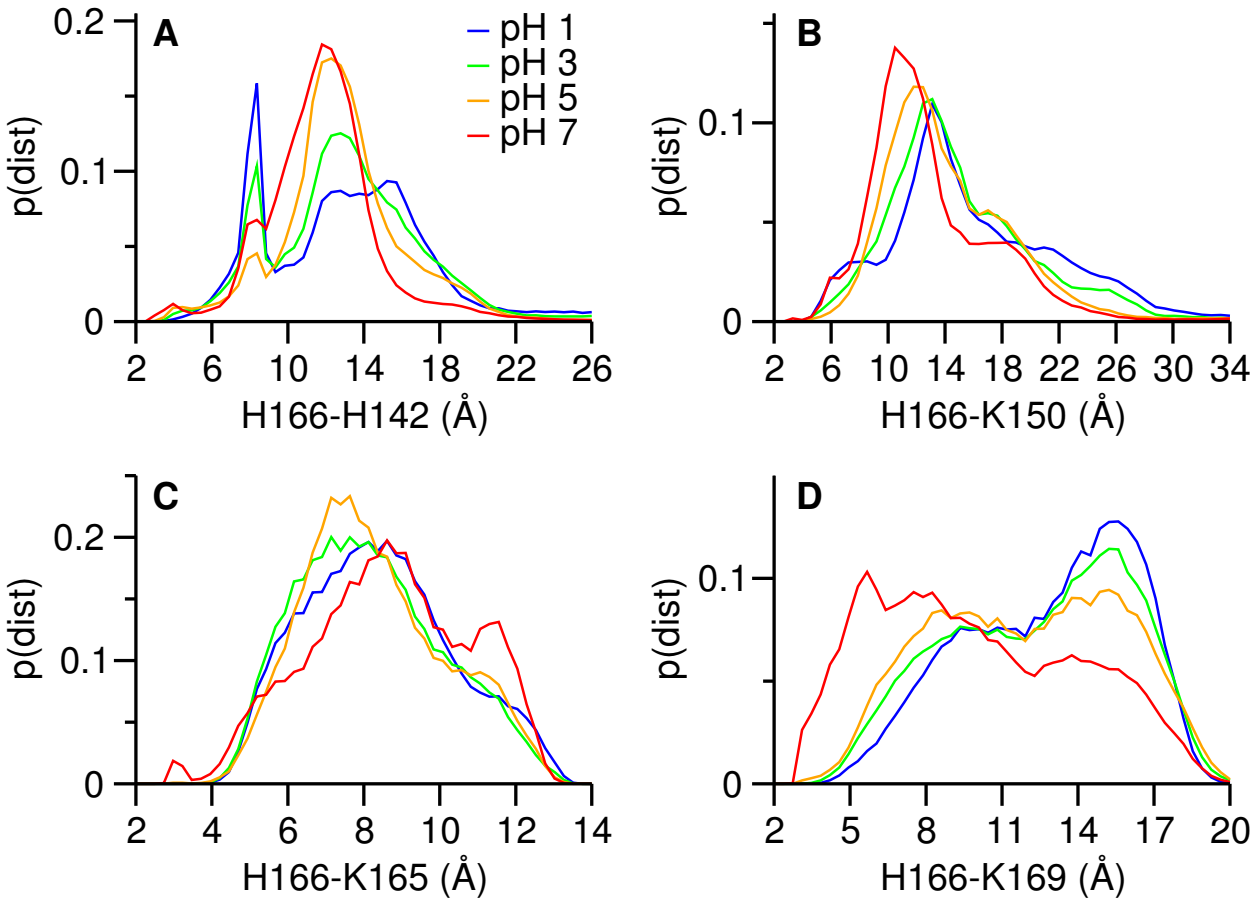

Figure S7: **His166 experiences repulsion from Lys169 once it becomes protonated.** Minimum distance between the sidechain nitrogen atoms of His166 and Lys169.

References

[1] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartles, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. K. T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.

[2] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.

[3] N. Ferguson, T. D. Sharpe, P. J. Schartau, S. Sato, M. D. Allen, C. M. Johnson, T. J. Rutherford and A. R. Fersht, *J. Mol. Biol.*, 2005, **353**, 427–446.

[4] M. A. Robien, G. M. Clore, J. G. Omichinski, R. N. Perham, E. Appella, K. Sakaguchi and A. M. Gronenborn, *Biochemistry*, 1992, **31**, 3463–3471.

[5] G. Zuo, J. Wang and W. Wang, *Proteins*, 2006, **63**, 165–173.

[6] H. Liu and S. Huo, *J. Phys. Chem. B*, 2012, **116**, 646–652.

[7] W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.

[8] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew. Chem. Int. Ed.*, 1999, **38**, 236–240.

[9] J.-E. Shea and C. L. Brooks, III, *Annu. Rev. Phys. Chem.*, 2001, **52**, 499–535.

[10] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

[11] M. Sadqi, D. Fushman and V. Muñoz, *Nature*, 2006, **442**, 317–321.

[12] H. Neuweiler, T. D. Sharpe, T. J. Rutherford, C. M. Johnson, M. D. Allen, N. Ferguson and A. R. Fersht, *J. Mol. Biol.*, 2009, **390**, 1060–1073.

[13] E. Arbely, T. J. Rutherford, T. D. Sharpe, N. Ferguson and A. R. Fersht, *J. Mol. Biol.*, 2009, **387**, 986–992.

[14] E. Arbely, T. J. Rutherford, H. Neuweiler, T. D. Sharpe, N. Ferguson and A. R. Fersht, *J. Mol. Biol.*, 2010, **403**, 313–327.