Supplementary Information

MCCE2: Improving Protein pK_a Calculations with Extensive Side Chain Rotamer Sampling

*Yifan Song, Junjun Mao and M. R. Gunner**

Physics Department J-419

City College of New York

138th St. and Convent Ave.

New York, NY 10031

Supplementary Material S1. Rotamer making. The number of conformers is given for hen egg white lysozyme (4LZT). Default values for each step are given. Underlined values can be modified in running parameter input file.

Step	Number of conformers	Default rules
	Lysozyme (129 amino acids)	
1. Prepare protein.	128[a]	Delete exposed water and ions; add NTR and CTR; add missing atoms to any group as needed.
2. Build multiconformer protein		
a. Initial minimization	620	5 initial minimization cycles based on ideal conformers closest to the native ones
b. Isosteric conformers	908	Asn, Gln, His
c. Heavy atom rotamer	25707	Make heavy atom rotamers at 60° intervals
Self-energy pruning	3659	Delete rotamers with ≥10Kcal total LJ and torsion energy to atoms in backbone and in the same conformer.
d. Packing and pruning	1201	Repack 5000 times. Keep conformers within 2.5 Kcal/mol of the minimum energy conformers.
e. Protons and ionization states	2012	Heavy atom rotamers duplicated for each ionization state; polar protons added into each torsion minima.
f. Heavy atom relaxation	4218	Energy minimization for neighboring pairs of all-atom conformers. 10 rounds with 100 1fs minimization steps.
g. Hydroxyl optimization	6063	Hydroxyl hydrogens optimized to have favorable interactions with neighboring residue conformers.
h. Rotamer pruning	2301	Cluster conformers differing by less than 2Å atomic distance; and 1.5 kcal/mol LJ and 1.5 Kcal/mol electrostatic pairwise interactions, retaining only one.

[a]GLY has no side chain conformers, and NTR and CTR are counted as separate residues.

Supplementary Material S2. List of proteins analyzed, the PDB files used and source of the experimental pK_as

Protein	Exp pK_a ref	PDB	Method	# of Models	Ligands kept	Deleted ligands
Ribonuclease α-sarcin precursor	[1, 2]	1DE3	NMR	20		
Apo E2	[3]	**1LE2**	X-ray	1		
Apo E3	[4]	**1NFN**	X-ray	1		
Apo E4	[3]	**1GS9**	X-ray	1		
B1 Domain of protein G	[5]	1GB1	NMR	60		
		1PGA	X-ray	1		
		1PGB	X-ray	1		
B2 Domain of protein G	[5]	2IGH	NMR	24		
		1IGD	X-ray	1		
		2IGD	X-ray	1		
Barnase	[6-10]	**1A2P**	X-ray	3		Surface Zn
		1B2X	X-ray	3		Surface Zn
		1BNR	NMR	20		
		1FW7	NMR	20		
Pancreatic trypsin inhibitor precursor (BPTI)	[11, 12]	1BPI	X-ray	1		Surface PO4
		1JV8	NMR	23		
		4PTI	X-ray	1		
Calbindin D9k	[13]	1KQV	NMR	30		LA(III) not affecting benchmarked residues
		1KSM	NMR	1		LA(III)
		1IG5	X-ray	1		Mg
		4ICB	X-ray	1		Ca
Cardiotoxin A5	[14]	**1KXI**	X-ray	2		
CD2d1	[15]	**1HNG**	X-ray	2		
Chymotrypsin	[16]	**2TGA**	X-ray			Ca
Chymotrypsin Inhibitor 2	[17]	**2CI2**	X-ray	1		
		2SNI	X-ray	1		Ca, only chain I is used
		3CI2	NMR	20		
Cyclophilin	[18]	**2CPL**	X-ray			
		1OCA	NMR	20		
Epidermal growth factor	[19]	1EGF	NMR	16		
		1EPG	NMR	1		
		1EPH	NMR	10		
		1EPI	NMR	1		
		1EPJ	NMR	5		
FKBP	[18]	1FKS	NMR			
fungal beta cryptogein	[20]	1BEG	NMR	18		
		1BEO	X-ray	1		
		1BXM	X-ray	1		ERG

hen egg white lysozyme	[21-23]	1E8L	NMR	50		
		1LSE	X-ray	1		
		4LZT	X-ray	1		NO3
human DNA polymerase lambda lyase domain	[24]	1NZP	NMR	8		
		1XSN	X-ray	1		
MutT	[25]	1PPX	NMR	20		8OG, MO2
		1MUT	NMR	15		
Myoglobin - horse	[26, 27]	**1DWR**	X-ray	1	HEM	CMO,SO4
		1WLA	X-ray	1	HEM	SO4
		1YMB	X-ray		HEM	SO4
Myoglobin - sperm whale	[26, 28]	1MYF	NMR	12	HEM	
		1A6K	X-ray	1	HEM	SO4
		1A6M	X-ray	1	HEM	Oxy,SO4
phage T4 lysozyme	[29]	**2LZM**	X-ray	1		
phage T4 lysozyme M102K mutant	[30]	1L54	X-ray	1		
Phosphonoacetaldehyde hydrolase	[31]	1FEZ	X-ray	4		Mg, WO4
		1RQL	X-ray	2		Mg, VSO
RNase A	[32, 33]	2AAS	NMR	32		
		7RSA	X-ray	1		TBU
		9RAT	X-ray	1		
		3RN3	X-ray	1		SO4
RNase H1	[34, 35]	1RCH	NMR	8		
		1GOA	X-ray	1		
		2RN2	X-ray	1		
RNase SA	[36]	1C54	NMR	20		
		1LNI	X-ray	2		GOL,SO4
		1RGG	X-ray	2		SO4
RNase T1	[37-40]	1IYY	NMR	24		
		1I0V	X-ray	1		ADP,MG,POB
		1BVI	X-ray	4		2GP,Ca
		1RGA	X-ray	1		3GP, Ca, G
Sea anemone neurotoxin III	[41]	1ANS	NMR	28		
snake erabutoxin b	[42]	1ERA	NMR	1		
		1FRA	NMR	14		
		3EBX	X-ray	1		SO4
Staphylococcal Nuclease	[43]	**1STY**	X-ray			
Streptomyces Subtilisin Inhibitor	[44]	**2SIC**	X-ray	1		Ca
		3SIC	X-ray	1		Ca
		5SIC	X-ray	1		Ca
		3SSI	X-ray	1		

Sso7d	[45]	1JIC	NMR	1	
		1SSO	NMR	1	
		1BBX	NMR	2	
		1C8C	*X-ray*	*1*	*Not used in this study due to the methylated residue 1 and disordered CTR*
Thioredoxin (reduced)	[46]	**1ERT**	X-ray	1	
		3TRX	NMR	1	
		4TRX	NMR	33	
Thioredoxin (oxidized)	[46]	**1ERU**	X-ray	1	
		1AUC	X-ray	1	
Turkey ovomucoid inhibitor	[47, 48]	1OMT	NMR	50	
		1OMU	NMR	50	
		1PPF	X-ray	1	BMA,FUC,GAL,GLC,MAN,NAG -- complex structure
		1TUR	NMR	12	
		1TUS	NMR	12	
Tyrosin phosphatase	[49, 50]	1BVH	NMR	15	
		1DG9	X-ray	1	EPE
		1Z12	X-ray	1	VO4
		1Z13	X-ray	1	MOO
		1PNT	X-ray	1	PO4
Xylanase BA	[51]	**1H4G**	X-ray	2	FXP,SO4
		1H4H	X-ray	4	XYP,XYS, residue 1 changed from Glu to Gln
		1QH6	X-ray	2	FXP
		1QH7	X-ray	2	XYP
Xylanase BC	[52]	1BVV	X-ray	1	DFX,XYP
		1C5H	X-ray	1	
		1XNB	X-ray	1	SO4

Bold: PDB files used for the dataset with 1 X-ray structure/protein.

References:

1. Perez-Canadillas, J.M., et al., Biochemistry, 1998. **37**(45): p. 15865-76.
2. Garcia-Mayoral, M.F., et al., Biochemistry, 2003. **42**(45): p. 13122-33.
3. Lund-Katz, S., et al., J Lipid Res, 2001. **42**(6): p. 894-901.
4. Lund-Katz, S., et al., J Biol Chem, 2000. **275**(44): p. 34459-64.
5. Khare, D., et al., Biochemistry, 1997. **36**(12): p. 3580-9.
6. Oliveberg, M., V.L. Arcus, and A.R. Fersht, Biochemistry, 1995. **34**(29): p. 9424-33.
7. Loewenthal, R., J. Sancho, and A.R. Fersht, J Mol Biol, 1992. **224**(3): p. 759-70.
8. Loewenthal, R., et al., J. Mol. Biol., 1993. **232**: p. 574-583.
9. Mossakowska, D.E., K. Nyberg, and A.R. Fersht, Biochemistry, 1989. **28**(9): p. 3843-50.
10. Sali, D., M. Bycroft, and A.R. Fersht, Nature, 1988. **335**(6192): p. 740-3.
11. March, K.L., et al., Biochemistry, 1982. **21**(21): p. 5241-5251.

12. Richarz, R. and K. Wuthrich, Biochemistry, 1978. **17**: p. 2263-2269.
13. Kesvatera, T., et al., J Mol Biol, 1996. **259**(4): p. 828-39.
14. Chiang, C.M., et al., Biochemistry, 1996. **35**: p. 9177-9186.
15. Chen, H.A., et al., Biochemistry, 2000. **39**(23): p. 6814-24.
16. Markley, J.L. and I.B. Ibanez, Biochemistry, 1978. **17**(22): p. 4627-40.
17. Tan, Y., et al., J. Mol. Biol., 1995. **254**: p. 980-92.
18. Yu, L. and S.W. Fesik, Biochim Biophys Acta, 1994. **1209**(1): p. 24-32.
19. Kohda, D., T. Sawada, and F. Inagaki, Biochemistry, 1991. **30**: p. 4896-4900.
20. Gooley, P.R., et al., J Biomol NMR, 1998. **12**(4): p. 523-34.
21. Bartik, K., C. Redfield, and C.M. Dobson, Biophys J, 1994. **66**(4): p. 1180-4.
22. Kuramitsu, S. and K. Hamaguchi, J Biochem (Tokyo), 1980. **87**(4): p. 1215-9.
23. Takahashi, T., H. Nakamura, and A. Wada, Biopolymers, 1992. **32**: p. 897-909.
24. Gao, G., et al., Biochemistry, 2006. **45**(6): p. 1785-94.
25. Harris, T.K., et al., Biochemistry, 2000. **39**(7): p. 1655-74.
26. Kao, Y.H., et al., Biophys J, 2000. **79**(3): p. 1637-54.
27. Papa, S., et al., Ann N Y Acad Sci, 1992. **671**: p. 345-358.
28. Bashford, D., et al., Biochemistry, 1993: p. 8045-8056.
29. Anderson, D.E., W.J. Becktel, and F.W. Dahlquist, Biochemistry, 1990. **29**(9): p. 2403-8.
30. Dao-pin, S., et al., Biochemistry, 1991. **30**(49): p. 11521-9.
31. Zhang, G., et al., Biochemistry, 2002. **41**(45): p. 13370-7.
32. Rico, M., et al. *Proceedings of the 2nd International Meeting.* 1991. Sant Feliu de Guixols, Girona, Spain.
33. Baker, W.R. and A. Kintanar, Arch Biochem Biophys, 1996. **327**(1): p. 189-99.
34. Oda, Y., M. Yoshida, and S. Kanaya, J Biol Chem, 1993. **268**(1): p. 88-92.
35. Oda, Y., et al., Biochemistry, 1994. **33**(17): p. 5275-84.
36. Laurents, D.V., et al., J Mol Biol, 2003. **325**(5): p. 1077-92.
37. Koumanov, A., et al., Eur Biophys J, 2001. **30**(3): p. 198-206.
38. Spitzner, N., et al., Eur Biophys J, 2001. **30**(3): p. 186-97.
39. Inagaki, F., et al., J Biochem (Tokyo), 1981. **89**(4): p. 1185-95.
40. Giletto, A. and C.N. Pace, Biochemistry, 1999. **38**(40): p. 13379-84.
41. Norton, R.S., et al., Biochem J, 1993. **293 (Pt 2)**: p. 545-51.
42. Inagaki, F., et al., Eur J Biochem, 1978. **89**(2): p. 433-42.
43. Lee, K.K., C.A. Fitch, and E.B. Garcia-Moreno, Protein Sci, 2002. **11**(5): p. 1004-16.
44. Fujii, S., K. Akasaka, and H. Hatano, J Biochem (Tokyo), 1980. **88**(3): p. 789-96.
45. Consonni, R., et al., Biochemistry, 2003. **42**(6): p. 1421-9.
46. Forman-Kay, J.D., G.M. Clore, and A.M. Gronenborn, Biochemistry, 1992. **31**(13): p. 3442-52.
47. Schaller, W. and A.D. Robertson, Biochemistry, 1995. **34**(14): p. 4714-4723.
48. Swint-Kruse, L. and A.D. Robertson, Biochemistry, 1995. **34**(14): p. 4724-4732.
49. Zhou, M.M., J.P. Davis, and R.L. Van Etten, Biochemistry, 1993. **32**(33): p. 8479-86.
50. Tishmack, P.A., et al., Biochemistry, 1997. **36**(39): p. 11984-94.
51. Betz, M., et al., Biochemistry, 2004. **43**(19): p. 5820-31.
52. Joshi, M.D., A. Hedberg, and L.P. McIntosh, Protein Sci, 1997. **6**(12): p. 2667-70.

Supplementary Material S3. Calculated pK_as - FULL MCCE single X-ray structure; 33 proteins

Protein	Res	Calc pK1	Calc pK2	exp pK	Comments	DSSP
Apo E2	LYS 69	12.6		10.1		H
Apo E2	LYS 72	10.6		10		H
Apo E2	LYS 75	11.3		10		H
Apo E2	LYS 95	10.1		10.2		H
Apo E2	LYS 143	10.1		9.4		H
Apo E2	LYS 146	9.2		9.9		H
Apo E2	LYS 157	11.5		10.9		H
Apo E3	LYS 69	11.7		10.4		H
Apo E3	LYS 72	10.7		10		H
Apo E3	LYS 75	11.8		10.1		H
Apo E3	LYS 95	10.2		10.1		H
Apo E3	LYS 143	9.7		9.5		H
Apo E3	LYS 146	9.7		9.2		H
Apo E3	LYS 157	12.0		11.1		H
Apo E4	LYS 69	11.2		10.1		H
Apo E4	LYS 72	10.4		10		H
Apo E4	LYS 75	10.5		10.1		H
Apo E4	LYS 95	10.7		10.1		H
Apo E4	LYS 143	9.8		9.9		H
Apo E4	LYS 146	9.8		9.4		H
Apo E4	LYS 157	11.9		10.9		H
B1 Domain of protein G	TYR 3	15.6		11	Lower Limit	E
B1 Domain of protein G	LYS 4	11.8		11	Lower Limit	E
B1 Domain of protein G	LYS 10	11.9		11	expt val approximate	S
B1 Domain of protein G	LYS 13	11.8		11	Lower Limit	E
B1 Domain of protein G	GLU 15	5.0		4.4		E
B1 Domain of protein G	GLU 19	4.1		3.7		E
B1 Domain of protein G	ASP 22	2.3		2.9		S
B1 Domain of protein G	GLU 27	4.9		4.5		H
B1 Domain of protein G	LYS 28	10.9		10.9		H
B1 Domain of protein G	TYR 33	11.1		11	expt val approximate	H
B1 Domain of protein G	ASP 36	4.9		3.8		H
B1 Domain of protein G	ASP 40	4.4		4		L
B1 Domain of protein G	GLU 42	4.0		4.4		E
B1 Domain of protein G	TYR 45	15.8		12	Lower Limit	E
B1 Domain of protein G	ASP 46	3.8		3.6		E
B1 Domain of protein G	ASP 47	3.3		3.4		T
B1 Domain of protein G	GLU 56	3.7		4		L
B2 Domain of protein G	TYR 8	14.8		11	Lower Limit	E
B2 Domain of protein G	LYS 9	11.5		11	expt val approximate	E
B2 Domain of protein G	LYS 15	12.4		11	Lower Limit	S
B2 Domain of protein G	LYS 18	11.6		11	Lower Limit	E
B2 Domain of protein G	GLU 20	4.8		4.3		E
B2 Domain of protein G	LYS 24	11.1		10.7		E
B2 Domain of protein G	ASP 27	3.0		2.9		S

B2 Domain of protein G	GLU	29	4.3		4.2	H
B2 Domain of protein G	GLU	32	4.9		4.6	H
B2 Domain of protein G	LYS	33	10.8		11 expt val approximate	H
B2 Domain of protein G	TYR	38	11.9		11 expt val approximate	H
B2 Domain of protein G	ASP	41	5.1		3.9	H
B2 Domain of protein G	ASP	45	4.4		4	L
B2 Domain of protein G	TYR	50	16.3		12 Lower Limit	E
B2 Domain of protein G	ASP	51	3.6		3.6	E
B2 Domain of protein G	ASP	52	3.2		3.4	T
B2 Domain of protein G	GLU	61	4.7		4.2	L
Bacterial phosphono-acetaldehyde hydrolase	*LYS*	*53*	*7.8*		*9.3 pH dependence of inactivation*	*H*
Bacterial proteinase inhibitor Ssi	*HIS*	*43*	*1.7*		*3.15 A - see below*	*L*
Bacterial proteinase inhibitor Ssi	HIS	106	4.9		6	T
Barnase	ASP	8	3.6		3.3	H
Barnase	ASP	12	4.0		3.8	H
Barnase	HIS	18	7.6		7.73	S
Barnase	ASP	22	3.9		3.3	T
Barnase	GLU	29	5.3		3.75	H
Barnase	ASP	44	4.2		3.35	H
Barnase	ASP	54	3.8		2.2	E
Barnase	GLU	60	3.9		3	T
Barnase	GLU	73	2.1	4.4	2.2	E
Barnase	ASP	75	2.4		3.1 change boundary	E
Barnase	ASP	86	5.8		4.2	L
Barnase	ASP	93	1.2		2 Upper Limit	T
Barnase	ASP	101	2.0	1.1	2	S
Barnase	HIS	102	6.2		6.3	T
Barnase	CTR		3.6		3.3	ND
BPTI	NTR		6.9		7.9	ND
BPTI	ASP	3	3.5		3.55	G
BPTI	GLU	7	3.9		3.85	S
BPTI	TYR	10	10.2		9.46	L
BPTI	LYS	15	10.7		10.4	L
BPTI	TYR	21	10.5		9.94	E
BPTI	TYR	23	12.2		11	E
BPTI	LYS	26	10.7		10.4	T
BPTI	TYR	35	10.4		10.6	E
BPTI	LYS	41	11.3		10.8	L
BPTI	LYS	46	10.5		10.4	S
BPTI	GLU	49	4.0		3.91	H
BPTI	ASP	50	2.7		3.2	H
BPTI	CTR		3.7		3.1	ND
Calbindin D9k	LYS	1	11.3		10.6	L
Calbindin D9k	GLU	4	4.1		3.8	H
Calbindin D9k	GLU	5	2.8		3.4	H
Calbindin D9k	LYS	7	11.6		11.4	H

Calbindin D9k	GLU	11	4.9	4.7		H
Calbindin D9k	LYS	12	11.2	11.1		H
Calbindin D9k	LYS	16	11.9	10.1		T
Calbindin D9k	GLU	17	3.5	3.62		S
Calbindin D9k	LYS	25	11.7	11.7		H
Calbindin D9k	GLU	26	3.7	4.1		H
Calbindin D9k	LYS	29	11.0	11.4		H
Calbindin D9k	LYS	41	10.7	10.9		L
Calbindin D9k	ASP	47	3.9	3		H
Calbindin D9k	GLU	48	4.4	4.6		H
Calbindin D9k	LYS	55	12.2	12.1		T
Calbindin D9k	GLU	64	3.7	3.8		H
Calbindin D9k	LYS	71	9.8	10.7		H
Calbindin D9k	LYS	72	10.6	11.3		H
Calbindin D9k	CTR		3.9	3.2		ND
Cardiotoxin A5	HIS	4	6.2	5.6		E
Cardiotoxin A5	GLU	17	3.8	4		T
Cardiotoxin A5	ASP	42	3.2	3.2		S
Cardiotoxin A5	ASP	59	1.8	2.3	Upper Limit	T
CD2d1	ASP	2	2.3	3.5		L
CD2d1	ASP	25	3.5	3.53		T
CD2d1	ASP	26	4.0	3.58		T
CD2d1	ASP	28	4.0	3.57		E
CD2d1	GLU	29	5.1	4.51		E
CD2d1	GLU	33	4.0	4.2		E
CD2d1	GLU	41	6.7	6.53		E
CD2d1	GLU	56	4.4	3.95		E
CD2d1	ASP	62	6.2	4.18		L
CD2d1	ASP	71	3.8	3.2		G
CD2d1	ASP	72	5.2	4.14		G
CD2d1	ASP	94	3.1	3.83		E
CD2d1	GLU	99	4.3	4.1		L
CD2d1	CTR		4.0	3.11		ND
Chymotrypsin Inhibitor 2	GLU	23	4.3	3.2		B
Chymotrypsin Inhibitor 2	GLU	26	2.4	3.3		G
Chymotrypsin Inhibitor 2	GLU	33	3.2	4.2		H
Chymotrypsin Inhibitor 2	GLU	34	3.8	3.8		H
Chymotrypsin Inhibitor 2	GLU	45	3.2	3.85		T
Chymotrypsin Inhibitor 2	GLU	60	3.1	3.5		L
Chymotrypsin Inhibitor 2	ASP	64	4.8	3.8		T
Chymotrypsin Inhibitor 2	ASP	71	2.1	2.8		L
Chymotrypsin Inhibitor 2	ASP	74	4.4	4.95		S
chymotrypsinogen	His	40	3.7	4.6		E
chymotrypsinogen	His	57	8.8	7.3		G
Cyclophilin	HIS	54	1.4	4.2	Upper Limit	L
Cyclophilin	HIS	70	6.3	5.8		S
Cyclophilin	HIS	92	-2.8	4.2	Upper Limit	L
Cyclophilin	HIS	126	5.2	6.3		S
Fungal beta cryptogein	NTR		9.0	7.43		ND

Fungal beta cryptogein	TYR	12	10.4		11.5	expt val approximate	H
Fungal beta cryptogein	ASP	21	2.1		2.49		S
Fungal beta cryptogein	ASP	30	5.0		2.51		H
Fungal beta cryptogein	TYR	33	18.1		12	Lower Limit	L
Fungal beta cryptogein	TYR	47	14.0		12	Lower Limit	H
Fungal beta cryptogein	LYS	61	10.4		10.1	Ambiguous assignment	H
Fungal beta cryptogein	ASP	72	4.0		2.61		B
Fungal beta cryptogein	TYR	85	10.0		10.4		H
Fungal beta cryptogein	TYR	87	17.5		12	Lower Limit	H
Fungal beta cryptogein	LYS	94	10.7		9.4	Ambiguous assignment	H
Fungal beta cryptogein	CTR		4.2		3.51		ND
Hen egg white lysozyme	LYS	1	10.9		10.8		L
Hen egg white lysozyme	GLU	7	3.0		2.85		H
Hen egg white lysozyme	LYS	13	11.1		10.5		H
Hen egg white lysozyme	HIS	15	6.0		5.36		T
Hen egg white lysozyme	ASP	18	2.3		2.66		T
Hen egg white lysozyme	TYR	20	11.6		10.3		B
Hen egg white lysozyme	TYR	23	10.7		9.8		B
Hen egg white lysozyme	LYS	33	9.4		10.4		H
Hen egg white lysozyme	GLU	35	6.1		6.2		H
Hen egg white lysozyme	ASP	48	3.5		1.6		T
Hen egg white lysozyme	ASP	52	4.2		3.68		E
Hen egg white lysozyme	ASP	66	1.0		0.9		L
Hen egg white lysozyme	ASP	87	2.1		2.1		S
Hen egg white lysozyme	ASP	87	2.1		2.07		S
Hen egg white lysozyme	LYS	96	11.3		10.8		H
Hen egg white lysozyme	LYS	97	11.0		10.3		H
Hen egg white lysozyme	ASP	101	5.2		4.08		T
Hen egg white lysozyme	LYS	116	10.2		10.2		T
Hen egg white lysozyme	ASP	119	3.9		3.2		L
Hen egg white lysozyme	CTR		3.3		2.75		ND
Human DNA polymerase lambda lyase domain	LYS	312	9.6	8.1	9.5		H
Human thioredoxin (ox)	GLU	6	3.8		4.9		L
Human thioredoxin (ox)	GLU	13	4.2		4.4		H
Human thioredoxin (ox)	ASP	16	3.9		4.2		H
Human thioredoxin (ox)	ASP	20	3.2		3.8		T
Human thioredoxin (ox)	ASP	26	10.7	13.0	8.1	change boundary	E
Human thioredoxin (ox)	GLU	47	4.5		4.3		H
Human thioredoxin (ox)	GLU	56	2.4		3.2		E
Human thioredoxin (ox)	***Asp 58/60/61***		***2.7, 3.9, 5.2 (coupled expt values)***				L
Human thioredoxin (ox)	*ASP 58*		*2.5*	*7.3*			*E*
Human thioredoxin (ox)	*ASP 60*		*3.8*				*T*
Human thioredoxin (ox)	*ASP 61*		*5.0*	*3.4*			*T*
Human thioredoxin (ox)	ASP	64	3.8		3.2		H
Human thioredoxin (ox)	GLU	68	4.3		5.1		H
Human thioredoxin (ox)	GLU	70	4.2		4.8		T
Human thioredoxin (ox)	GLU	88	2.8		3.6		E
Human thioredoxin (ox)	GLU	95	3.7		4.1		H

Human thioredoxin (ox)	GLU	98	3.4	3.9		H
Human thioredoxin (ox)	GLU	103	4.1	4.5		H
Human thioredoxin (red)	GLU	6	4.3	4.8		L
Human thioredoxin (red)	GLU	13	4.0	4.4		H
Human thioredoxin (red)	ASP	16	3.7	4		H
Human thioredoxin (red)	ASP	20	3.4	3.8		T
Human thioredoxin (red)	ASP	26	10.7	9.9		E
Human thioredoxin (red)	HIS	43	7.4	5.5		H
Human thioredoxin (red)	GLU	47	3.6	4.1		H
Human thioredoxin (red)	GLU	56	2.9	3.1		E
Human thioredoxin (red)	***ASP***	***58/60/61***	***2.8, 4.2, 5.3 (coupled expt values)***			L
Human thioredoxin (red)	*ASP*	*58*	*2.7*	*7.5*		*E*
Human thioredoxin (red)	*ASP*	*60*	*5.2*			*T*
Human thioredoxin (red)	*ASP*	*61*	*4.8*			*T*
Human thioredoxin (red)	ASP	64	3.7	3.2		H
Human thioredoxin (red)	GLU	68	3.9	4.9		H
Human thioredoxin (red)	GLU	70	3.7	4.6		T
Human thioredoxin (red)	GLU	88	2.9	3.7		E
Human thioredoxin (red)	GLU	95	4.2	4.1		H
Human thioredoxin (red)	GLU	98	4.2	3.9		H
Human thioredoxin (red)	GLU	103	4.4	4.4		H
Myoglobin - horse	HIS	36	6.1	7.8		L
Myoglobin - horse	HIS	48	5.9	5.62		T
Myoglobin - horse	HIS	81	7.0	6.94		L
Myoglobin - horse	HIS	113	6.0	5.87		H
Myoglobin - horse	HIS	116	6.6	6.79		H
Myoglobin - horse	HIS	119	5.6	6.56		S
Myoglobin - sperm whale	HIS	12	7.4	6.5		H
Myoglobin - sperm whale	HIS	24	-0.5	5	Upper Limit	H
Myoglobin - sperm whale	HIS	36	7.0	8		L
Myoglobin - sperm whale	HIS	48	6.5	5.6		T
Myoglobin - sperm whale	HIS	64	1.8	5	Upper Limit	H
Myoglobin - sperm whale	HIS	81	6.9	6.9		L
Myoglobin - sperm whale	HIS	82	3.8	5	Upper Limit	L
Myoglobin - sperm whale	HIS	97	5.8	5.6		S
Myoglobin - sperm whale	HIS	113	6.5	5.4		H
Myoglobin - sperm whale	HIS	116	6.8	6.7		H
Myoglobin - sperm whale	HIS	119	5.6	6.2		L
Phage T4 lysozyme	HIS	31	7.9	9.1		E
RNase A	GLU	2	3.0	2.6		L
RNase A	GLU	9	4.6	4		H
RNase A	HIS	12	4.1	6		H
RNase A	ASP	14	1.8	1.8		L
RNase A	ASP	38	3.2	3.5		S
RNase A	HIS	48	7.8	6.1		L
RNase A	GLU	49	4.4	4.7		S
RNase A	ASP	53	4.2	3.7		H
RNase A	ASP	83	5.3	3.3		E
RNase A	GLU	86	4.3	4		E

RNase A	HIS	105	6.8	6.5		L
RNase A	GLU	111	4.2	3.5		E
RNase A	HIS	119	5.5	6.5		E
RNase A	ASP	121	2.6	3		E
RNase A	CTR		2.8	2.3		?
RNase H1	GLU	6	4.0	4.5		E
RNase H1	ASP	10	8.2	6.1		E
RNase H1	GLU	32	4.2	3.6		E
RNase H1	GLU	48	1.8	4.4		H
RNase H1	GLU	57	3.7	3.2		H
RNase H1	GLU	61	3.3	3.9		S
RNase H1	HIS	62	6.7	7		L
RNase H1	GLU	64	4.5	4.4		E
RNase H1	ASP	70	3.6	2.6		L
RNase H1	HIS	83	5.9	5.5		H
RNase H1	ASP	94	4.2	3.2		T
RNase H1	ASP	102	0.9	2	Upper Limit	H
RNase H1	ASP	108	3.8	3.2		H
RNase H1	HIS	114	-0.1	5	Upper Limit	S
RNase H1	GLU	119	5.2	4.1		E
RNase H1	HIS	124	7.9	7.1		S
RNase H1	HIS	127	7.3	7.9		L
RNase H1	GLU	129	3.3	3.6		H
RNase H1	GLU	131	4.6	4.3		H
RNase H1	ASP	134	2.7	4.3		H
RNase H1	GLU	135	4.7	4.3		H
RNase H1	GLU	147	4.5	4.2		L
RNase H1	ASP	148	-1.6	2	Upper Limit	L
RNase H1	GLU	154	4.5	4.4		L
RNase H1	CTR		3.4	3.4		ND
RNase SA	NTR		8.6	9.14		ND
RNase SA	ASP	1	3.2	3.44		L
RNase SA	GLU	14	5.3	5.05		H
RNase SA	ASP	17	4.7	3.72		H
RNase SA	ASP	25	5.1	4.87		T
RNase SA	TYR	30	10.5	11.3		S
RNase SA	ASP	33	0.6	2.39		T
RNase SA	GLU	41	3.5	4.14		T
RNase SA	TYR	49	10.1	10.6		T
RNase SA	TYR	51	15.8	11.5	Lower Limit	S
RNase SA	TYR	52	19.4	11.5	Lower Limit	L
RNase SA	HIS	53	8.9	8.27		E
RNase SA	GLU	54	4.2	3.42		E
RNase SA	TYR	55	13.8	11.5	Lower Limit	E
RNase SA	GLU	74	5.0	3.47		S
RNase SA	GLU	78	4.9	3.13		L
RNase SA	ASP	79	7.4	7.37		E
RNase SA	TYR	80	18.1	11.5	Lower Limit	E
RNase SA	TYR	81	13.2	11.5	Lower Limit	E

RNase SA	ASP	84	2.7		3.01		S
RNase SA	HIS	85	5.9		6.35		T
RNase SA	TYR	86	14.0		11.5	Lower Limit	T
RNase SA	ASP	93	3.8		3.09		E
RNase SA	CTR		2.1		2.43		ND
RNase T1	ASP	15	4.2		3.52		H
RNase T1	HIS	27	8.4		7		H
RNase T1	GLU	28	6.1		5.61		H
RNase T1	ASP	29	5.5		4.26		H
RNase T1	GLU	31	4.6		5.36		L
RNase T1	HIS	40	8.5		7.9		E
RNase T1	GLU	46	4.0		3.62		T
RNase T1	ASP	49	5.0		4.22		L
RNase T1	GLU	58	2.5		3.96		E
RNase T1	ASP	66	4.6		3.9		S
RNase T1	ASP	76	4.1		0.5	B - see below	E
RNase T1	GLU	82	2.8		3.27		T
RNase T1	HIS	92	6.1		7.3		S
RNase T1	GLU	102	5.3		5.3		E
Snake erabutoxin b	HIS	6	0.1		2.8	Upper Limit	L
Snake erabutoxin b	TYR	25	13.5		12	Lower Limit	E
Snake erabutoxin b	HIS	26	5.6		5.8		E
Staph Nuclease	HIS	8	7.6		6.52		L
Staph Nuclease	HIS	46	6.4		5.86		L
Staph Nuclease	HIS	121	6.6		5.3		T
Staph Nuclease	HIS	124	4.6		5.73		H
Turkey ovomucoid inhibitor	NTR		6.0		8		ND
Turkey ovomucoid inhibitor	ASP	7	2.1		2.6	Upper Limit	L
Turkey ovomucoid inhibitor	GLU	10	2.5		4.2		T
Turkey ovomucoid inhibitor	TYR	11	9.2	11.5	10.2		L
Turkey ovomucoid inhibitor	LYS	13	11.2	14.1	9.9		L
Turkey ovomucoid inhibitor	GLU	19	3.7		3.2		L
Turkey ovomucoid inhibitor	TYR	20	10.3		11.1		L
Turkey ovomucoid inhibitor	ASP	27	3.3		2.3	Upper Limit	T
Turkey ovomucoid inhibitor	LYS	29	11.4		11.1		L
Turkey ovomucoid inhibitor	TYR	31	14.9		12.5	Lower Limit	E
Turkey ovomucoid inhibitor	LYS	34	10.6		10.1		H
Turkey ovomucoid inhibitor	GLU	43	3.9		4.8		H
Turkey ovomucoid inhibitor	HIS	52	7.3		7.5		E
Turkey ovomucoid inhibitor	LYS	55	11.1		11.1		L
Tyrosin phosphotase	HIS	66	6.8		8.3		T
Tyrosin phosphotase	HIS	72	7.3		9.2		L
Xylanase BA	ASP	5	4.6		3.84		S
Xylanase BA	HIS	11	7.4	0.7	6.5	change boundary	E
Xylanase BA	ASP	12	5.2		3.94		T
Xylanase BA	ASP	15	3.2		3.35		E
Xylanase BA	GLU	17	3.5		4.31		E
Xylanase BA	ASP	21	2.8		3.46		L
Xylanase BA	HIS	32	5.4		6.7		S

Xylanase BA	HIS	60	4.0	6.2	4		H
Xylanase BA	ASP	90	4.0		3.88		T
Xylanase BA	GLU	94	4.6	8.1	3.94		E
Xylanase BA	ASP	99	2.2		2.7	Upper Limit	E
Xylanase BA	ASP	118	2.5		2.7	Upper Limit	T
Xylanase BA	ASP	123	0.4		2.7	Upper Limit	E
Xylanase BA	GLU	126	4.7		4.51		E
Xylanase BA	HIS	162	-2.5		2.7	Upper Limit	H
Xylanase BA	GLU	167	2.3		3.58		H
Xylanase BA	GLU	178	4.5	7.1	4.1	change boundary	E
Xylanase BA	GLU	184	5.3		6.5		E
Xylanase BA	CTR		4.2		6.52		ND
Xylanase BC	ASP	11	2.2		2.5		L
Xylanase BC	GLU	78	4.3	7.2	4.6		E
Xylanase BC	ASP	101	1.1		2	Upper Limit	T
Xylanase BC	ASP	106	4.6		2.7		E
Xylanase BC	ASP	119	3.3		3.2		T
Xylanase BC	ASP	121	4.7		3.6		S
Xylanase BC	HIS	149	-1.3		2.3	Upper Limit	H
Xylanase BC	HIS	156	6.7		6.5		T
Xylanase BC	GLU	172	6.7	4.1	6.7		E

References to experimental values in Table S2
pK2 provided when bimodal equation provides significantly better fit to calculated titration
Comments
Upper or lower limit indicates the residue pK could not be seen in experimental titration
Change boundary:Rerun with MCCE selected exposed conformer substituted for experimental input positions used to calculate the corrected boundary
Other notes refer to ambiguity of assignments.

A Exp pK = 3.2. Protonation of His occurs with denaturation of the protein. Therefore, pKa for the wildtype protein should be <3.2. High desolvation energy is consistent with the low solvent accessibility found in the NMR measurements. (Fujii et al., 1980, J Biochem (Tokyo))
B Small ppm shift is found in Spitzner et al. at ph 4 and 8, ambiguous data in Koumanov et al. Pfeiffer et al. suggest it is protonated at pH 5.5

DSSP assignment: H helix; E b-sheet;L loops; ND termini not defined; S,T etc other

Supplementary Material S4. Comparson of pK_as calculated by MCCE, MD/GB/TI and PROPKA

			pK_a - $pK_{a,sol}$			Error		
PDB	**Res**		**Exp**	**MCCE**	**MD**	**PROPKA**	**MCCE**	**MD**	**PROPKA**
3rn3	Asp	14	-2.2	-2.7	-0.3	-2.6	-0.5	1.9	-0.4
4lzt	Asp	87	-1.9	-2.0	-0.8	-1.5	-0.1	1.1	0.4
1ppf[a]	*Asp*	*27*	*-1.8*	*-0.8*	*0.6*	*-1.6*	1.0	**2.4**	0.2
1xnb	Asp	11	-1.5	-2.0	-0.1	-2.0	-0.5	1.4	-0.5
1beo	Asp	21	-1.5	-2.0	-1.1	-2.6	-0.5	0.4	-1.1
4lzt	Asp	18	-1.3	-1.7	-0.8	-1.2	-0.4	0.5	0.1
1xnb	Asp	106	-1.3	0.5	-1.2	-1.0	1.8	0.1	0.3
1pga	Asp	22	-1.1	-0.6	-0.4	-1.8	0.5	0.7	-0.7
3rn3	Asp	121	-0.9	-1.8	1.0	-0.3	-0.9	1.9	0.6
1a2p	Asp	75	-0.9	-1.6	0.3	-5.3	-0.7	1.2	**-4.4**
2rn2	Asp	94	-0.8	0.1	0.3	-1.3	0.9	1.1	-0.5
1pga	Asp	47	-0.6	0.2	-0.1	-1.3	0.8	0.5	-0.7
3rn3	Asp	53	-0.3	0.2	1.0	-0.6	0.5	1.3	-0.3
4lzt	Asp	52	-0.3	0.2	2.0	-1.0	0.5	**2.3**	-0.7
1pga	Asp	36	-0.2	1.0	0.6	-0.1	1.2	0.8	0.1
2trx	Asp	20	-0.2	0.2	0.0	-1.5	0.4	0.2	-1.3
1de3	Asp	59	0.1	-0.8	-0.1	-0.8	-0.9	-0.2	-0.9
1de3	Asp	57	0.3	0.0	-0.1	0.0	-0.3	-0.4	-0.3
2rn2	Asp	10	2.1	4.6	2.7	3.0	**2.5**	0.6	0.9
2trx	Asp	26	4.1	2.9	3.7	1.2	-1.2	-0.4	**-2.9**
3rn3	Glu	2	-1.8	-3.3	-2.0	-1.7	-1.5	-0.2	0.1
4lzt	Glu	7	-1.5	-1.5	0.0	-0.7	0.0	1.5	0.8
1ppf	Glu	19	-1.2	-0.8	-0.3	-0.5	0.4	0.9	0.7
2rn2	Glu	57	-1.2	-0.5	0.2	-1.8	0.7	1.4	-0.6
1a2p	Glu	60	-1.2	-0.5	0.7	-0.6	0.7	1.9	0.6
3rn3	Glu	111	-0.9	-0.1	0.5	0.2	0.8	1.4	1.1
2rn2	Glu	129	-0.8	-1.3	-0.5	-0.9	-0.5	0.3	-0.1
2rn2	Glu	61	-0.5	-1.2	0.5	-0.8	-0.7	1.0	-0.3
3rn3	Glu	9	-0.4	0.1	1.3	0.2	0.5	1.7	0.6
2bca	Glu	26	-0.3	-0.1	0.4	0.4	0.2	0.7	0.7
1ppf	Glu	10	-0.3	-2.0	0.7	0.0	**-1.7**	1.0	0.3
2rn2	Glu	119	-0.3	0.6	1.6	-0.9	0.9	1.9	-0.6
1pga	Glu	27	0.1	0.9	1.3	-0.1	0.8	1.2	-0.2
1ppf	Glu	43	0.4	-0.5	0.3	0.1	-0.9	-0.1	-0.3
1de3	Glu	96	0.7	1.4	2.4	0.6	0.7	1.7	-0.1
1ans	Glu	20	1.0	0.2	0.7	0.1	-0.8	-0.3	-0.9
1rga	Glu	28	1.5	0.6	1.2	-0.3	-0.9	-0.3	-1.8
4lzt	Glu	35	1.8	1.1	1.8	0.6	-0.7	0.0	-1.2
1hng	Glu	41	2.3	1.9	2.3	0.3	-0.4	0.0	-2.0
1xnb	Glu	172	2.3	0.7	2.7	2.9	-1.6	0.4	0.6
2snm[b]	*Lys*	*66*	*-4.1*	*-11.5*	*-3.1*	*-2.6*	**-7.4**	1.0	1.5
1l54[b]	*Lys*	*102*	*-3.9*	*-9.7*	*-0.9*	*-0.3*	**-5.8**	**3.0**	**3.6**
1mut[a]	*Lys*	*39*	*-2.1*	*1.3*	*0.3*	*0.0*	**3.4**	**2.4**	**2.1**
1nfn	Lys	146	-1.3	-0.8	-0.8	0.0	0.5	0.5	1.3
1fez[a]	*Lys*	*53*	*-1.2*	*-3.4*	*1.8*	*-2.4*	**-2.2**	**3.0**	-1.2

1gs9	Lys	146	-1.1	-0.7	-0.8	0.0	0.4	0.3	1.1
1le2	Lys	143	-1.1	-0.5	-0.2	-0.5	0.6	0.9	0.6
1nfn	Lys	143	-1.0	-0.8	-0.3	0.0	0.2	0.7	1.0
1nzp	Lys	312	-1.0	0.3	0.7	-0.2	1.3	1.7	0.8
1gs9	Lys	143	-0.6	-0.7	-0.3	0.0	-0.1	0.3	0.6
1le2	Lys	146	-0.6	-1.5	-0.5	-0.1	-0.9	0.1	0.5
1ppf	Lys	34	-0.4	0.2	-0.2	-0.3	0.6	0.2	0.1
4lzt	Lys	33	-0.1	-0.8	-0.4	-0.2	-0.7	-0.3	-0.1
2bca	Lys	41	0.3	0.3	0.2	-0.1	0.0	-0.1	-0.4
4lzt	Lys	96	0.3	1.3	0.1	-0.3	1.0	-0.2	-0.6
1pga	Lys	28	0.4	0.6	0.2	-0.6	0.2	-0.2	-1.0
2bca	Lys	16	0.4	2.4	0.7	-0.6	**2.0**	0.3	-1.0
1ppf	Lys	55	0.6	0.7	0.2	0.0	0.1	-0.4	-0.6
2bca	Lys	7	0.7	1.0	0.9	0.0	0.3	0.2	-0.7
2bca	Lys	55	1.3	1.5	1.3	0.0	0.2	0.0	-1.3
3ebx[a]	*His*	*6*	*-3.5*	*-5.3*	*-4.6*	*0.0*	-1.8	-1.1	**3.5**
3ssi[b]	*His*	*43*	*-3.1*	*-6.3*	*-3.1*	*-1.4*	**-3.2**	0.0	1.7
1stn	His	121	-1.0	0.8	1.4	-1.0	1.8	**2.4**	0.0
4lzt	His	15	-0.9	-0.4	1.9	1.0	0.5	**2.8**	1.9
1ert	His	43	-0.8	0.1	0.2	0.0	0.9	1.0	0.8
1de3	His	137	-0.5	-1.0	1.3	-4.3	-0.5	1.8	**-3.8**
3rn3	His	48	-0.2	1.4	4.9	-3.1	1.6	**5.1**	**-2.9**
3rn3	His	119	0.2	-0.8	-0.5	0.2	-1.0	-0.7	0.0
3rn3	His	12	-0.3	-2.2	0.5	-4.5	-1.9	0.8	**-4.2**
1de3	His	104	0.2	0.6	2.3	0.1	0.4	**2.1**	-0.1
1de3	His	36	0.5	0.3	0.7	0.2	-0.2	0.2	-0.3
2rn2	His	62	0.7	0.3	0.7	0.7	-0.4	0.0	0.0
2rn2	His	124	0.8	1.0	-1.1	0.1	0.2	-1.9	-0.7
1de3	His	50	1.4	1.1	3.1	-3.8	-0.3	1.7	**-5.2**
1rga	His	92	1.5	0.1	1.2	-0.4	-1.4	-0.3	-1.9
1rga	His	40	1.6	2.1	1.6	2.5	0.5	0.0	0.9
2rn2	His	127	1.6	1.0	1.6	0.8	-0.6	0.0	-0.8
1dg9	His	66	2.0	0.5	1.3	1.3	-1.5	-0.7	-0.7
2lzm	His	31	2.8	1.6	0.9	1.3	-1.2	-1.9	-1.5
1dg9	His	72	2.9	1.0	3.2	1.6	-1.9	0.3	-1.3
RMSD[1]				0.94	1.24	1.40			
RMSD[2]				1.46	1.27	1.45			

Calculated pK_a shift using MD/GB/TI (Simonson et al. J Am Chem Soc, 2004) and PROPKA (Li et al. Proteins, 2005) methods as reported in Stanton and Houk's study (Stanton and Houk, J Chem Theory Comput, 2008).
Errors greater then 2pH units in bold.
a: The experimental pK_a is out of the range of the titration or is measured by protein activity.
b: the measured pK_a is coupled to protein denaturation.
RMSD[1]: The RMSD between calculations and experiments for 73 residues, excluding (a) and (b).
RMSD[2]: The RMSD between calculations and experiments for 76 residues, excluding (a).