Supporting Material for

"Rapid calculation of protein pK_a values using Rosetta"

Krishna Praneeth Kilambi and Jeffrey J. Gray

METHODS

Experimental pK_a dataset

Main dataset

A diverse set of experimental pK_a values were collected for Asp, Glu, His, Tyr and Lys from previous studies (1–4) (see Table S4). Residues missing absolute pK_a values but instead reported to be within a specific range of pH values were omitted. The set comprised 264 amino acid residues, including 76 aspartates, 73 glutamates, 76 histidines, 17 tyrosines and 22 lysines. The protein structures containing these residues include 33 x-ray crystal structures, 5 NMR structures and one structure determined using neutron diffraction. 94% of the selected x-ray crystal structures were determined at a resolution of 2.2 Å or better. Water and ligands were removed from the structures.

Staphylococcal nuclease dataset

A second smaller test set comprised pK_a values of all wild-type and mutant residues from either the standard or the highly stable Δ+PHS variant of *staphylococcal* nuclease for which crystal structures were available. The dataset included Asp residues at positions 19, 21, 40 and 95 and Glu residues at positions 10, 43, 52, 57, 67, 73, 75, 101, 122, 129 and 135 from the wild-type Δ+PHS SNase (3BDC). The dataset also included Lys mutants at positions 25 (3ERQ), 34 (3ITP), 38 (2RKS), 62 (3DMU), 66 (2SNM), 72 (2RBM), 92 (1TT2), 103 (3E5S), 104 (3C1F) and 125 (3C1E); Glu mutants at positions 25 (3EVQ), 38 (3D6C), 66 (1U9R), 72 (3ERO), 91 (3D4D), 92 (1TQO) and 104 (3H6M); and Asp mutants at positions 66 (2OXP) and 92 (2OEO).

Standard Rosetta score function

The standard Rosetta score function includes a Lennard–Jones potential to model the Van der Waals potential (with parameters modified from CHARMM19), the Lazaridis–Karplus Gaussian exclusion implicit solvation model (5), an orientation-dependent hydrogen-bonding term (6), a term based on the probability of proximity of two amino acids (7), a statistical potential for the free energies of the amino-acid side-chain rotamers (8) and reference energies for each amino acid that are summed to approximate the free energy of the denatured state (9).

pK_a derivation

During titration in Rosetta-pH, when $\mathrm{pH} = \mathrm{p}K_a$,

$$E_{\text{total}}^{\text{prot}} = E_{\text{total}}^{\text{deprot}}$$

$$E_{\mathrm{pH}}^{\mathrm{prot}} + \sum_{\substack{\text{energies} \\ \text{except pH}}} E_i^{\mathrm{prot}} = E_{\mathrm{pH}}^{\mathrm{deprot}} + \sum_{\substack{\text{energies} \\ \text{except pH}}} E_i^{\mathrm{deprot}}$$

$$E_{\mathrm{pH}}^{\mathrm{prot}} - E_{\mathrm{pH}}^{\mathrm{deprot}} = - \sum_{\substack{\text{energies} \\ \text{except pH}}} (E_i^{\mathrm{prot}} - E_i^{\mathrm{deprot}})$$

$$E_{\mathrm{pH}}^{\mathrm{prot}} - E_{\mathrm{pH}}^{\mathrm{deprot}} = - \sum_{\substack{\text{energies} \\ \text{except pH}}} (E_i^{\mathrm{prot}} - E_i^{\mathrm{deprot}})$$

$$2.3 k_B T (\mathrm{pH} - \mathrm{Ip}K_{\mathrm{a}}) = - \sum_{\substack{\text{energies} \\ \text{except pH}}} (E_i^{\mathrm{prot}} - E_i^{\mathrm{deprot}})$$

Substituting $\mathrm{pH} = \mathrm{p}K_{\mathrm{a}}$, we get

$$\mathrm{p}K_{\mathrm{a}} = \mathrm{Ip}K_{\mathrm{a}} - \frac{1}{2.3 k_B T} \sum_{\substack{\text{energies} \\ \text{except pH}}} (E_i^{\mathrm{prot}} - E_i^{\mathrm{deprot}})$$

where $E_{\mathrm{total}} = E_{\mathrm{vdw}} + E_{\mathrm{sol}} + E_{\mathrm{elec}} + E_{\mathrm{hbond}} + E_{\mathrm{dun}} + E_{\mathrm{pH}} + E_{\mathrm{ref}}$ and the summations are taken over all score types excluding the protonation potential (E_{pH}).

E_i^{prot} and E_i^{deprot} are scores of the protonated and deprotonated variants, and $\mathrm{Ip}K_a$ is the unperturbed intrinsic $\mathrm{p}K_{\mathrm{a}}$ value of the model compound in solution. The 'deprotonated' residues include the -1 charge states of Asp, Glu and Tyr, and the neutral His and Lys residues. All other alternate residue charge states are considered to be 'protonated'. While the total scores $E_{\mathrm{total}}^{\mathrm{prot}}$ and $E_{\mathrm{total}}^{\mathrm{deprot}}$ are intended to represent free energies, they only include enthalpy and the water entropy (via the solvation score), and neglect protein entropy.

Rosetta command line

The calculations in the paper were carried out using Rosetta's developer revision number 46351. All algorithms will be distributed in the next Rosetta release (summer/fall 2012). The Rosetta command-line arguments and scores used for the calculations are as follows.

(a) Sampling just the target residue side chain

```
pH_protocol.<exe> –database <path_to_rosetta_database>
      –s 2OVO.pdb –no_output
      –pH_mode
      –pka_for_resnos 7 27 –pka_for_chainno A
      –core:weights pH_standard.wts
      –ex1 –ex1:level 3 –ex2 –ex2:level 3
      –extrachi_cutoff 0 –use_input_sc
```

where the pK_a values are to be calculated for residues 7 and 27 from chain A and 2OVO.pdb

(b) Sampling the neighboring side chains along with the target residue

```
pH_protocol.<exe> –database <path_to_rosetta_database>
      –s 2OVO.pdb –no_output
      –pH_mode –pH_neighbor_pack –pka_rad 6.0
      –pka_for_resnos 7 27 –pka_for_chainno A
      –core:weights pH_standard.wts
      –ex1 –ex1:level 3 –ex2 –ex2:level 3
      –extrachi_cutoff 0 –use_input_sc
```

(c) Using RosettaRelax to generate backbone conformational ensemble

```
relax.<exe> –database <path_to_rosetta_database>
      –s 2OVO.pdb –ignore_unrecognized_res –nstruct 50
```

(d) The scores used for pK_a calculations with their respective weights (pH_standard.wts)

Score Type	Weight
fa_atr	0.8
fa_rep	0.44
fa_sol	0.65
fa_intra_rep	0.004
fa_dun	0.56
hbond_lr_bb	1.17
hbond_sr_bb	1.17
hbond_bb_sc	1.17
hbond_sc	1.1
hack_elec	1.0
e_pH	1.0
ref	1.0

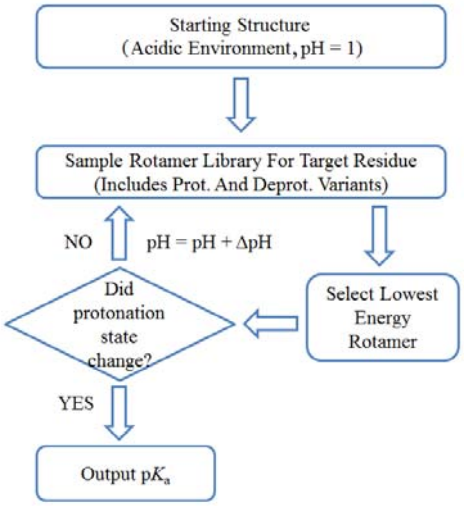

Figure S1 Algorithm for calculating amino acid pK_a values using lowest energy conformers.

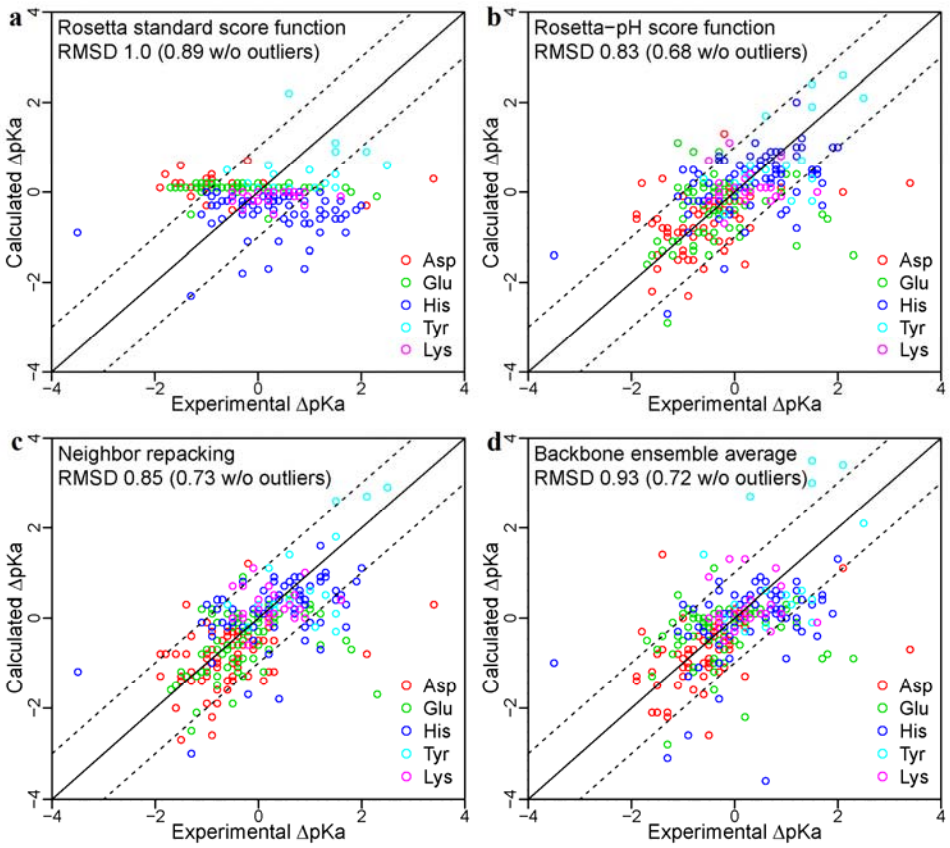

Figure S2 Correlation between predicted and experimental ΔpK_a values calculated using (*a*) the standard Rosetta score function and the Rosetta-pH score function employing (*b*) flexible target side chain, (*c*) flexible neighboring side chains and (*d*) average over a backbone ensemble of 50 structures. ΔpK_a values are evaluated with respect to the reference intrinsic pK_a values (see Figure 1).

Note: The RMSD values based on pK_a and ΔpK_a values are equivalent:

$$\sqrt{\frac{\sum(\Delta pK_a^{pred} - \Delta pK_a^{exp})^2}{n}} = \sqrt{\frac{\sum((pK_a^{pred} - lpK_a) - (pK_a^{exp} - lpK_a))^2}{n}} = \sqrt{\frac{\sum(pK_a^{pred} - pK_a^{exp})^2}{n}}$$

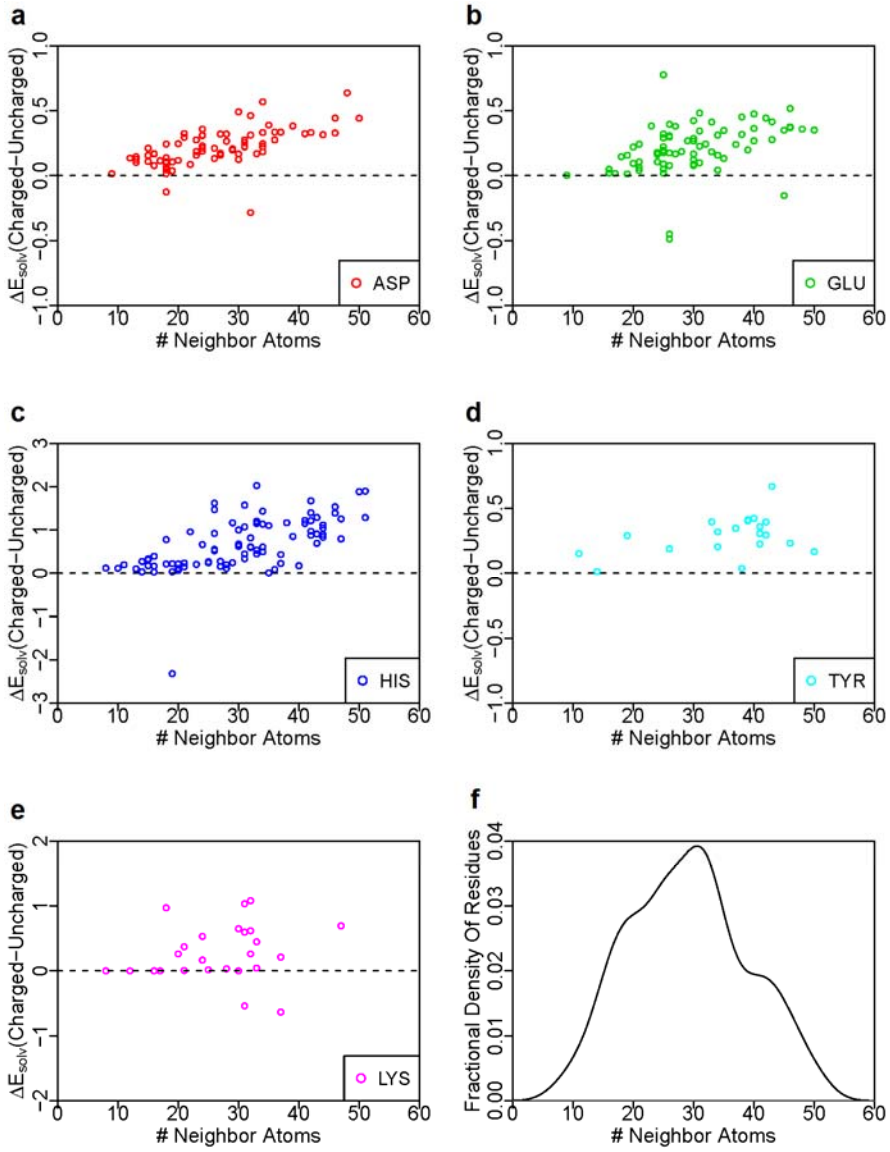

Figure S3 (*a* - *e*) Difference in the solvation energy (ΔE_{solv}) between charged and neutral variants of ionizable residues plotted versus the number of neighbor atoms (heavy atoms within 6 Å of the residue). (*f*) The distribution of fraction of total residues in the dataset as a function of the number of neighbor atoms.

Table S1 Rosetta's residue atom types that determine solvation, Van der Waals and hydrogen bonding parameters for the standard and non-standard amino acid protonation variants.

Residue	Atom types			
	ATOM	**CHARMM**	**Standard protonation variant**	**Non-standard protonation variant**
Asp [†]	$\{O_{\delta 1}\}$ or $\{O_{\delta 2}\}$	{OD1} or {OD2}	OOC	OH
	$\{H_{\delta 1}\}$ or $\{H_{\delta 2}\}$	{HD1} or {HD2}	-	Hpol
Glu [†]	$\{O_{\varepsilon 1}\}$ or $\{O_{\varepsilon 2}\}$	{OE1} or {OE2}	OOC	OH
	$\{H_{\varepsilon 1}\}$ or $\{H_{\varepsilon 2}\}$	{HE1} or {HE2}	-	Hpol
His [‡]	$\{N_{\delta 1}\}$ or $\{N_{\varepsilon 2}\}$	{ND1} or {NE2}	Nhis	Ntrp
	$\{H_{\delta 1}\}$ or $\{H_{\varepsilon 2}\}$	{HD1} or {HE2}	-	Hpol
Tyr	O_{ζ}	OH	OH	OOC
	H_{ζ}	HH	Hpol	-
Lys	N_{ζ}	NZ	Nlys	NH2O
	$H_{\zeta 3}$	3HZ	Hpol	-

[†]For protonated Asp and Glu, sample atom types are shown for both the possible protonation states (with H atoms on $O_{\delta 1}$/ $O_{\delta 2}$ and $O_{\varepsilon 1}$/ $O_{\varepsilon 2}$ atoms).
[‡]For neutral His, the atom types for both the possible tautomers are shown (with H atom on $N_{\delta 1}$ and $N_{\varepsilon 2}$ atoms).

Table S2 Reference scores (aa_{ref}) for the standard and non-standard amino acid side chain protonation variants.

Residue	Reference score	
	Standard protonation variant	**Non-standard protonation variant**
Asp	-0.67	-0.08
Glu	-0.81	-0.68
His	0.56	-0.03
Tyr	0.51	0.97
Lys	-0.65	-0.52

Table S3 Percentage of pK_a predictions within various error ranges from the experimental values while sampling just the target residue.

Cutoff (pH units)	**Asp**	**Glu**	**His**	**Tyr**	**Lys**	**All**
< 0.5	58%	48%	50%	53%	55%	52%
< 1	83%	82%	70%	76%	82%	78%
< 1.5	92%	90%	92%	94%	95%	92%
< 2	96%	93%	96%	100%	100%	96%

Table S4 Proteins used for pK_a predictions with their reference PDB IDs[†], the method used for structure determination, resolution of determined structure (if applicable) and number of residues with experimental pK_a values chosen for calculations.

Protein	PDB	Method	Resolution (Å)	Number of pK_a values
Silver pheasant ovomucoid	2OVO	x-ray	1.50	13
Bovine trypsin inhibitor	4PTI	x-ray	1.50	12
Hen egg white lysozyme	2LZT	x-ray	1.97	17
Turkey egg white lysozyme	135L	x-ray	1.30	6
Human lysozyme	1LZ1	x-ray	1.50	1
Bovine ribonuclease A	3RN3	x-ray	1.45	13
Ribonuclease A analog	3SRN	x-ray	2.00	3
Ribonuclease T1	9RNT	x-ray	1.50	14
	1YGW	NMR	-	14
E.coli ribonuclease H	2RN2	x-ray	1.48	21
S.aureofaciens ribonuclease	1RGG	x-ray	1.20	20
Bromelain inhibitor VI	1BI6	NMR	-	8
Cyclophilin A	2CPL	x-ray	1.63	2
FK506 & rapamycin-binding protein	1FKS	NMR	-	2
Horse metmyoglobin	1YMB	x-ray	1.90	4
Phosphatidylinositol phospholipase	1GYM	x-ray	2.20	4
Staphylococcal nuclease	1STG	x-ray	1.70	3
Streptococcal protein G	1PGA	x-ray	2.07	10
Barnase	1A2P	x-ray	1.50	8
Calbindin D9K	4ICB	x-ray	1.60	8
	2BCA	NMR	-	8
Cobra cardiotoxin V	1KXI	x-ray	2.19	2
Erabutoxin B	3EBX	x-ray	1.40	1
Beta cryptogein	1BEO	x-ray	2.20	3
HIV-1 protease KNI-272 complex	1HPX	x-ray	2.00	3
Human deoxyhaemoglobin	4HHB	x-ray	1.74	11

Protein	PDB	Method	Resolution (Å)	Number of pK_a values
Human oxyhaemoglobin	1HHO	x-ray	2.10	10
Carbonmonoxymyoglobin	2MB5	Neutron diff.	-	9
Human thioredoxin	1ERT	x-ray	1.70	5
Mouse epidermal growth factor	1EPI	NMR	-	6
Ribosomal protein L9	1DIV	x-ray	2.60	6
Fibronectin cell adhesion module III-10	1FNA	x-ray	1.80	4
Mu class glutathione S-transferase	6GST	x-ray	2.20	3
Streptomyces subtilisin inhibitor	3SSI	x-ray	2.30	2
B.amyloliquefaciens subtilisin BPN'	1DUI	x-ray	2.00	1
B. circulans xylanase	1XNB	x-ray	1.49	7

[†]The original citations for individual structures can be found in Refs (1–4).

Table S5 List of experimental predicted pK_a values using 1) standard Rosetta score function 2) Rosetta-pH score function employing varying degrees of conformational flexibility including a single flexible target side chain (site repack), flexible neighboring side chains (neighbor repack) and the average over an ensemble of conformations (ensemble average).

PDB	Residue	Chain	Experimental pK_a	Standard Rosetta pK_a	Site repack pK_a	Neighbor repack pK_a	Ensemble average pK_a
2OVO	Asp-7	A	2.4	4.1	3.3	3.2	2.8
2OVO	Asp-27	A	2.2	4.4	4.2	3.2	3.7
2OVO	Glu-10	A	4.1	4.5	4	3.9	3.9
2OVO	Glu-19	A	3.2	4.5	3.3	3.1	4.3
2OVO	Glu-43	A	4.8	4.5	4.6	4.5	4.5
2OVO	His-52	A	7.5	5.9	7.3	7.2	6.1
2OVO	Tyr-11	A	10.2	10.6	10.4	11.1	9.7
2OVO	Tyr-20	A	11.1	10.1	9.8	9.8	10.3
2OVO	Tyr-31	A	12.5	10.6	12.1	12.9	12.1
2OVO	Lys-13	A	9.9	10.2	10.2	10.4	11.3
2OVO	Lys-29	A	11.1	10.3	10.2	10.4	10.1
2OVO	Lys-34	A	10.1	10	10.3	10.4	10.2
2OVO	Lys-55	A	11.1	10.4	10.5	10.9	10.7
4PTI	Asp-3	A	3.6	4.1	3.6	3.6	3.6
4PTI	Asp-50	A	3.2	4.1	3.2	3.1	3.6
4PTI	Glu-7	A	3.9	4.5	3.9	3.5	4.3
4PTI	Glu-49	A	4	4.5	3.5	3.4	3.5
4PTI	Tyr-10	A	9.4	10.5	10.3	10.3	11.4
4PTI	Tyr-21	A	10	10.1	10.2	10.3	10.3
4PTI	Tyr-23	A	11	10.5	10.6	10.6	10.5
4PTI	Tyr-35	A	10.6	12.2	11.7	11.4	9.8
4PTI	Lys-15	A	10.4	10.4	10.5	10.6	10.5
4PTI	Lys-26	A	10.1	10.4	10.4	10.5	10.4
4PTI	Lys-41	A	10.6	10.2	10.2	10.3	10.4
4PTI	Lys-46	A	9.9	10.4	11.1	11.1	10.4
2LZT	Asp-18	A	2.7	4.1	3.1	2.6	1.8
2LZT	Asp-52	A	3.7	4.1	2.9	2.7	3.3
2LZT	Asp-87	A	2.1	4.1	3.4	2.7	2.6
2LZT	Asp-101	A	4.1	4.2	3.4	3.5	3.9
2LZT	Asp-119	A	3.2	4.1	3	2.9	2.9
2LZT	Glu-7	A	2.9	4.5	3.3	3.2	3.7
2LZT	Glu-35	A	6.2	4.4	3.8	3.7	3.6
2LZT	His-15	A	5.6	5.7	6.1	6.2	6.1
2LZT	Tyr-20	A	10.3	10.1	10.3	10.2	10.4
2LZT	Tyr-23	A	9.8	10.2	9.5	9.8	10
2LZT	Tyr-53	A	12.1	10.9	12.6	12.7	13.4
2LZT	Lys-1	A	10.6	10.2	10.4	10.5	10.6
2LZT	Lys-13	A	10.3	10.2	11.5	11.5	11.7
2LZT	Lys-33	A	10.4	10.2	10.5	10.8	10.1

PDB	Residue	Chain	Experimental pK_a	Standard Rosetta pK_a	Site repack pK_a	Neighbor repack pK_a	Ensemble average pK_a
2LZT	Lys-96	A	10.7	9.9	10.8	10.9	10.8
2LZT	Lys-97	A	10.1	10	10.6	11.1	10.5
2LZT	Lys-116	A	10.2	10.3	9.7	9.8	9.9
3RN3	Asp-38	A	3.5	4.1	3.5	3.1	3.4
3RN3	Asp-53	A	3.9	4.1	3.8	3.5	3.5
3RN3	Asp-83	A	3.5	4.1	2.6	2.6	1.4
3RN3	Asp-121	A	3.1	4.3	2.5	1.4	3
3RN3	Glu-2	A	2.8	4.5	3	2.9	3.1
3RN3	Glu-9	A	4	4.5	4.5	4.4	4.5
3RN3	Glu-49	A	4.7	4.5	4.4	4.3	4.5
3RN3	Glu-86	A	4.1	4.6	5.3	5.3	4.5
3RN3	Glu-111	A	3.5	4.6	4.4	4.6	4.2
3RN3	His-12	A	6.2	5.8	6.6	6.5	5.9
3RN3	His-48	A	6	4.5	5.7	6.1	4.5
3RN3	His-105	A	6.7	5.2	6.5	6.7	6.6
3RN3	His-119	A	6.1	6.1	7	6.1	6.4
2RN2	Asp-10	A	6.1	3.7	4	3.2	5.1
2RN2	Asp-70	A	2.6	4.1	4.3	4.3	5.4
2RN2	Asp-94	A	3.2	4.1	3.6	2.1	3.6
2RN2	Asp-134	A	4.1	4.1	3.8	2.6	4.1
2RN2	Glu-6	A	4.5	4.5	3.6	3.2	3.8
2RN2	Glu-32	A	3.6	4.5	3.1	3.1	3.3
2RN2	Glu-48	A	4.4	4.6	4.2	4.3	4.3
2RN2	Glu-57	A	3.2	4.6	3	2.3	3.6
2RN2	Glu-61	A	3.9	4.5	4.3	4.2	3.9
2RN2	Glu-64	A	4.4	4.5	4.4	4.1	3.9
2RN2	Glu-119	A	4.1	4.5	4	3.5	3.9
2RN2	Glu-129	A	3.6	4.6	3.2	2.9	3.2
2RN2	Glu-131	A	4.3	4.5	4.6	4.5	4.6
2RN2	Glu-135	A	4.3	4.5	3.6	3.5	4.6
2RN2	Glu-147	A	4.2	4.5	4.4	4.2	4.4
2RN2	Glu-154	A	4.4	4.5	4.2	4.1	4.4
2RN2	His-62	A	7	6.2	7.2	7.2	6.6
2RN2	His-83	A	5.5	6.3	6.5	6.5	6.5
2RN2	His-114	A	5	4	3.6	3.3	3.2
2RN2	His-124	A	7.1	6.3	6.7	6.7	6.7
2RN2	His-127	A	7.9	5.7	6.7	6.8	6.2
9RNT	Asp-3	A	3.5	3.6	2.6	2.7	3
9RNT	Asp-15	A	3.5	4.1	3	3	2.9
9RNT	Asp-29	A	4.3	4.1	3.2	2.8	4
9RNT	Asp-49	A	4.2	4.1	4	4	4.1
9RNT	Asp-66	A	3.9	4.1	4.1	3.3	3.6
9RNT	Glu-28	A	5.6	4.4	3.2	3.8	4.3

PDB	Residue	Chain	Experimental pK_a	Standard Rosetta pK_a	Site repack pK_a	Neighbor repack pK_a	Ensemble average pK_a
9RNT	Glu-31	A	5.4	4.5	4.8	4.7	4.6
9RNT	Glu-46	A	3.6	4.5	4.4	3.9	4.4
9RNT	Glu-58	A	4	4.5	3.5	3.2	2.8
9RNT	Glu-82	A	3.3	4.5	3.3	3.2	4.1
9RNT	Glu-102	A	5.3	4.5	4.4	4.2	4.5
9RNT	His-27	A	7	5.8	7	7	6.8
9RNT	His-40	A	7.5	6.3	8.3	7.9	7.1
9RNT	His-92	A	7.3	5	6.9	7	6.3
1RGG	Asp-1	A	3.4	4.1	3.7	3.7	3.9
1RGG	Asp-17	A	3.7	4.1	3.7	3.4	3.8
1RGG	Asp-25	A	4.9	4.1	4.1	4	4.2
1RGG	Asp-33	A	2.4	4.2	1.8	2	1.9
1RGG	Asp-79	A	7.4	4.3	4.2	4.3	3.3
1RGG	Asp-84	A	3	3.7	2.6	2.1	3.1
1RGG	Asp-93	A	3.1	4.1	3.8	3.3	3.2
1RGG	Glu-14	A	5	4.6	4.2	4.9	4.3
1RGG	Glu-41	A	4.1	4.5	4.5	3.1	4.4
1RGG	Glu-54	A	3.4	4.6	3.8	3.7	4.5
1RGG	Glu-74	A	3.5	4.5	4.5	4.5	4.6
1RGG	Glu-78	A	3.1	3.9	1.5	1.9	1.6
1RGG	His-53	A	8.3	6	7.3	7.4	7.6
1RGG	His-85	A	6.4	6.2	6.4	6.4	6.3
1RGG	Tyr-30	A	11.3	10.2	10.7	10.5	10.6
1RGG	Tyr-49	A	10.6	10.1	10.5	10.4	10.6
1RGG	Tyr-52	A	11.5	11.1	11.9	11.8	13.5
1RGG	Tyr-55	A	11.5	10.4	10.3	10.1	10.4
1RGG	Tyr-80	A	11.5	10.9	12.4	12.6	13
1RGG	Tyr-81	A	11.5	10.1	9.8	9.7	9.9
2CPL	His-70	A	5.8	6.1	6.2	6.2	5.9
2CPL	His-126	A	6.3	5.9	6.2	6.4	6.1
1LZ1	His-78	A	7.1	6.2	6.9	6.9	6.5
1YMB	His-81	A	6.6	6.2	7.1	7.1	6.8
1YMB	His-36	A	7.8	5.7	6.7	6.7	6.6
1YMB	His-113	A	5.4	6.2	5.4	5.3	5
1YMB	His-116	A	6.6	6.1	6	6.1	6
1GYM	His-92	A	5.4	6	6.1	6.1	5.9
1GYM	His-82	A	6.9	6	7.1	7.1	6.3
1GYM	His-32	A	7.6	5.8	7.5	7.2	6
1GYM	His-227	A	6.9	5.6	6.6	6.7	2.7
1STG	His-46	A	5.7	6.1	6.1	6.1	6
1STG	His-8	A	6.8	6.2	6.9	7	7.2
1STG	His-124	A	6	6.3	6.2	6.3	5.3
1PGA	Asp-22	A	2.9	4.1	3.1	2.8	2.5

PDB	Residue	Chain	Experimental pK_a	Standard Rosetta pK_a	Site repack pK_a	Neighbor repack pK_a	Ensemble average pK_a
1PGA	Asp-36	A	3.8	4.1	3.8	3.8	3.6
1PGA	Asp-40	A	4	4.1	3.9	3.9	4
1PGA	Asp-46	A	3.6	4.1	3.3	2.9	3
1PGA	Asp-47	A	3.4	4.1	2.3	2.4	2.3
1PGA	Glu-15	A	4.4	4.5	4	3.8	4.4
1PGA	Glu-19	A	3.7	4.5	3.5	3.6	4.4
1PGA	Glu-27	A	4.5	4.5	3.4	4.1	4.4
1PGA	Glu-42	A	4.4	4.5	4.5	4.4	4.5
1PGA	Glu-56	A	4	4.4	3.3	3.2	3.2
1A2P	Asp-8	A	2.9	4.1	2.5	2.6	3.2
1A2P	Asp-12	A	3.8	4.1	2.7	2.7	2.9
1A2P	Asp-22	A	3.3	4.1	3.5	3.5	3.6
1A2P	Asp-44	A	3.4	4.1	3.6	3.6	3.5
1A2P	Asp-75	A	3.1	4.1	1.7	1.8	1.4
1A2P	Asp-86	A	4.2	3.9	2.4	2.8	2.7
1A2P	Glu-29	A	3.8	4.5	4.1	3.8	4.1
1A2P	Glu-60	A	3	4.5	3.1	3.1	4
4ICB	Asp-47	A	3	4.1	2.7	2.7	2.7
4ICB	Glu-4	A	3.8	4.5	4.5	3.7	4.2
4ICB	Glu-5	A	3.4	4.4	4.1	3.2	3.9
4ICB	Glu-11	A	4.7	4.5	4.5	4.4	4.6
4ICB	Glu-17	A	3.6	4.6	5.3	3.7	4
4ICB	Glu-26	A	4.1	4.5	5.2	4.1	4.6
4ICB	Glu-48	A	4.6	4.5	4.1	3.7	4.3
4ICB	Glu-64	A	3.8	4.5	4.6	4.5	4.4
1KXI	Asp-42	A	3.2	4.1	2.8	3	3.3
1KXI	Glu-17	A	4	4.5	4.5	4.4	4.5
1BEO	Asp-21	A	2.5	4.6	2.3	1.3	1.9
1BEO	Asp-30	A	2.5	4.2	3.4	2.7	3.2
1BEO	Asp-72	A	2.6	4.1	3.2	3.2	3.3
1HPX	Asp-29	A	3.2	4.1	2.7	2.7	2.4
1HPX	Asp-30	A	3.9	4.1	2.9	3.1	3.4
1HPX	Asp-60	A	3	4.3	3.4	3.1	2.9
135L	Asp-18	A	2.7	4.1	3	2.8	1.8
135L	Asp-52	A	3.8	3.9	3.8	3.8	3.7
135L	Asp-87	A	2.1	4.1	3.5	3.2	2.7
135L	Asp-119	A	3.4	4.1	3.2	2.6	2.7
135L	Glu-7	A	2.7	4.5	2.8	2.8	3.9
135L	Glu-35	A	6.1	4.5	3.9	3.9	3.5
1DIV	Asp-8	A	3	4.1	3.9	3.9	4
1DIV	Asp-23	A	3.1	4.1	3.5	3.7	3.3
1DIV	Glu-17	A	3.6	4.5	4.4	4.3	4.4
1DIV	Glu-38	A	4	4.5	4.4	4.3	4.5

PDB	Residue	Chain	Experimental pK_a	Standard Rosetta pK_a	Site repack pK_a	Neighbor repack pK_a	Ensemble average pK_a
1DIV	Glu-48	A	4.2	4.5	4.2	4.2	3.7
1DIV	Glu-54	A	4.2	4.6	3.9	4.3	3.4
1XNB	Asp-4	A	3	4.1	3.2	3.2	2.7
1XNB	Asp-11	A	2.5	4.1	2.6	2.6	2.5
1XNB	Asp-106	A	2.7	3.9	3.2	3.6	1.9
1XNB	Asp-119	A	3.2	4.2	2.5	2.4	2.9
1XNB	Asp-121	A	3.6	4.1	3.8	3.7	3.7
1XNB	Glu-78	A	4.6	3.8	4.1	4.7	2.2
1XNB	Glu-172	A	6.7	4.3	3	2.7	3.5
4HHB	His-20	A	6.9	6.2	7.1	7.1	7.1
4HHB	His-45	A	7.2	5.8	5.8	5.8	6.3
4HHB	His-50	A	7.2	4.6	6.5	6.1	6.3
4HHB	His-72	A	6.6	5.9	7.1	7.2	6.8
4HHB	His-89	A	7.2	6.1	6.8	6.7	6.5
4HHB	His-112	A	7.6	6.1	7.3	7.3	6.7
4HHB	His-2	B	6.4	6.2	6.9	6.9	6.6
4HHB	His-77	B	6.7	6.1	6.8	6.8	6.6
4HHB	His-117	B	8.2	5.9	7.3	7.1	6.4
4HHB	His-143	B	6.2	6.1	5.8	5.8	5.4
4HHB	His-146	B	7.1	6.2	7.2	7.2	6.4
1HHO	His-20	A	6.7	6.1	7.4	4.5	7.1
1HHO	His-45	A	7	6.3	6.6	6.5	6.5
1HHO	His-50	A	7.5	6.3	7.1	7.1	6.9
1HHO	His-72	A	6	6.1	7	7.1	7.1
1HHO	His-89	A	7.2	6.3	6.7	6.6	6.8
1HHO	His-112	A	7.5	5.4	5.5	5.6	6.3
1HHO	His-2	B	6.5	6	6.6	6.7	6.5
1HHO	His-77	B	6	6	7.1	7.1	6.9
1HHO	His-117	B	8	5.8	6	6.1	6.7
1HHO	His-146	B	7.9	6.1	6.8	6.8	5.9
2MB5	His-12	A	6.3	6.3	6.7	6.6	6.8
2MB5	His-36	A	8	5.4	6.5	6.6	6.8
2MB5	His-48	A	5.3	6.3	6.8	6.6	6.6
2MB5	His-81	A	6.6	6.2	7.1	7.1	6.8
2MB5	His-97	A	5.6	6.1	6.7	6.7	5.2
2MB5	Tyr-103	A	10.3	10.4	10.6	10.3	12.7
2MB5	His-113	A	5.4	5.6	5.8	5.9	3.7
2MB5	His-116	A	6.5	6.2	6.8	6.7	6.7
2MB5	His-119	A	6.1	5.2	4.6	4.6	6.1
1ERT	Glu-13	A	4.8	4.5	4.6	4.5	4.5
1ERT	Asp-16	A	3.7	4.1	4	4	3.9
1ERT	Asp-20	A	3.6	4.1	3.3	3.1	3.8
1ERT	His-43	A	5.5	6	6.8	6.7	6.5

PDB	Residue	Chain	Experimental pK_a	Standard Rosetta pK_a	Site repack pK_a	Neighbor repack pK_a	Ensemble average pK_a
1ERT	Glu-103	A	4.9	4.5	4.5	4.6	4.6
1FNA	Glu-38	A	3.8	4.5	3.5	3.3	4.3
1FNA	Glu-47	A	3.9	4.5	3	2.5	4.2
1FNA	Asp-67	A	4.2	4.1	4.1	2.8	3.8
1FNA	Asp-80	A	3.4	4.1	3.8	3.8	3
3SRN	His-12	A	6	5.9	6.2	6.1	5.7
3SRN	His-105	A	6.5	4.6	5.4	5.5	6.3
3SRN	His-119	B	6.2	6	5.9	6.1	6.4
6GST	His-83	A	5.2	5.8	6.2	6.2	6
6GST	His-84	A	7.1	6.2	6.8	6.8	6.9
6GST	His-167	A	7.8	5.9	6.8	6.9	6.6
3SSI	His-43	A	3.2	2	1.8	1.6	1.4
3SSI	His-106	A	6	5.8	6.1	6.1	6.1
3EBX	His-6	A	2.8	5.4	4.9	5.1	5.3
1DUI	His-64	A	7.2	5.6	7.2	7.3	6.5
1YGW	Asp-3	A	3.5	4.1	3.5	3.4	3.1
1YGW	Asp-15	A	3.5	4.2	3.7	3.7	2.9
1YGW	Asp-29	A	4.3	4.1	3.9	3.5	4
1YGW	Asp-49	A	4.2	4.1	3.8	3.8	3.9
1YGW	Asp-66	A	3.9	4.1	3.6	3.4	3.3
1YGW	Glu-28	A	5.6	4.5	4.8	4.5	4.8
1YGW	Glu-31	A	5.4	4.5	4.8	4.6	4.7
1YGW	Glu-46	A	3.6	4.5	4.2	4.3	4.5
1YGW	Glu-58	A	4	4.6	4	3.8	3.4
1YGW	Glu-82	A	3.3	4.4	4.3	4.6	4
1YGW	Glu-102	A	5.3	4.5	4.6	4.5	4.2
1YGW	His-27	A	7	5.8	6.5	6.4	6.8
1YGW	His-40	A	7.5	5.7	6.3	6.2	6.9
1YGW	His-92	A	7.3	5.6	5.9	6.2	5.4
1BI6	Glu-5	L	4	4.5	4.8	4.5	5.5
1BI6	Asp-9	H	3	4.2	3.1	2.9	2.7
1BI6	Asp-13	H	3.8	4.1	3.6	3.6	3.6
1BI6	Asp-32	H	3.1	4.4	3.8	3.7	2.9
1BI6	Asp-38	H	3.7	4.1	3.4	3.4	2.8
1BI6	Glu-1	H	3.3	4.6	5.5	3.6	3.8
1BI6	Glu-2	H	3.3	4.6	4.5	4.3	4.9
1BI6	Glu-24	H	4.2	4.5	4.6	4.1	4.1
1FKS	His-94	A	5.8	6.2	6.5	6.5	6.6
1FKS	His-87	A	6.5	6	6.4	6.5	6.7
2BCA	Lys-7	A	10.6	10.4	10.7	11.1	11.7
2BCA	Lys-12	A	11.3	10.4	10.3	10.4	10.5
2BCA	Lys-16	A	11	10.4	10.5	10.7	10.5
2BCA	Lys-29	A	11.2	10.3	10.3	10.6	11.5

PDB	Residue	Chain	Experimental pK_a	Standard Rosetta pK_a	Site repack pK_a	Neighbor repack pK_a	Ensemble average pK_a
2BCA	Lys-41	A	10.9	10.4	10.5	10.6	10.5
2BCA	Lys-55	A	12	10.3	10.4	10.8	10.3
2BCA	Lys-71	A	10.7	10.4	10.4	10.5	10.4
2BCA	Lys-72	A	11.3	10.1	11.2	11.4	11.1
1EPI	Asp-11	A	3.9	4.1	4	4.1	4.1
1EPI	His-22	A	6.8	5.7	7	7	6.9
1EPI	Glu-24	A	4.1	4.5	3.9	3.9	4.4
1EPI	Asp-27	A	4	4.1	3.5	3.5	4.2
1EPI	Asp-40	A	3.6	4.1	3.1	3.2	4.1
1EPI	Asp-46	A	3.8	4.7	5.3	5.2	3.8

SUPPORTING REFERENCES

1. Forsyth, W.R., J.M. Antosiewicz, and A.D. Robertson. 2002. Empirical relationships between protein structure and carboxyl pKa values in proteins. Proteins. 48: 388–403.

2. Edgcomb, S.P., and K.P. Murphy. 2002. Variability in the pKa of histidine side-chains correlates with burial within proteins. Proteins. 49: 1–6.

3. Wisz, M.S., and H.W. Hellinga. 2003. An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. Proteins. 51: 360–377.

4. Li, H., A.D. Robertson, and J.H. Jensen. 2005. Very fast empirical prediction and rationalization of protein pKa values. Proteins. 61: 704–721.

5. Lazaridis, T., and M. Karplus. 1999. Effective energy function for proteins in solution. Proteins. 35: 133–152.

6. Kortemme, T., A.V. Morozov, and D. Baker. 2003. An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein-Protein Complexes. J. Mol. Biol. 326: 1239–1259.

7. Simons, K.T., I. Ruczinski, C. Kooperberg, B.A. Fox, C. Bystroff, et al. 1999. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins. 34: 82–95.

8. Dunbrack, R.L., and F.E. Cohen. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Science. 6: 1661–1681.

9. Kuhlman, B., and D. Baker. 2000. Native protein sequences are close to optimal for their structures. Proc. Natl. Acad. Sci. USA. 97: 10383 –10388.