Supporting Information:

A computational scheme of pK_a values based on the three-dimensional reference interaction site model self-consistent field theory coupled with the linear fitting correction scheme

Ryo Fujiki[1], *Yukako Kasai*[1], *Yuki Seno*[1], *Toru Matsui*[2], *Yasuteru Shigeta*[3], *Norio Yoshida**[1], *Haruyuki Nakano*[1]

1. Department of Chemistry, Graduate School of Science, Kyushu University, 744, Motooka, Nishiku, Fukuoka, 819-0395 Japan

2. Department of Chemistry, Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577 Japan

3. Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577 Japan

Table S1(1). Training set molecules for the LFC parameter fitting and those experimental pK_a values.

Carboxyl	Molecule	pK_a	Ref.	Amine	Molecule	pK_a	Ref.
	CHOCOOH	3.32	1		$Ph(CH_2)_2NH_3^+$	9.83	5
	trans- CH_3CH=$CHCOOH$	4.69			$PhCH_2NH_3^+$	9.34	
	$Ph(OH)_2COOH$	4.48	2		$PhNH_3^+$	4.58	
	H_2C=$CHCH_2COOH$	4.42			$CH_3(CH_2)_3NH_3^+$	10.58	
	$(CH(OH)COOH)_2$	1.14			$CH_3CH_2NH_3^+$	10.67	
	$CHOHCH_3COOH$	3.86			$HO(CH_2)_2NH_3^+$	9.50	
	CH_3COCH_2COOH	3.58			$HONH_3^+$	5.96	
	$CH_3COCOOH$	2.50			NH_4^+	9.21	4
	$CHCl_2COOH$	1.29	3		H_2C=$CHCH_2NH_3^+$	9.49	
	CH_2FCOOH	2.66			$CH_{3(}CH_2)_2NH_3^+$	10.53	
	NO_2CH_2COOH	1.68			Cyclohexylamine	10.64	
	$PhNO_2COOH$	2.45	4		Cyclohexylmethyl amine	10.49	
					Isopropylamine	10.63	
					Methoxyamine	4.60	
					γ -Phenylpropyl amine	10.20	
					neo-Pentylamine	10.21	
					sec-Butylamine	10.56	

Table S1(2). Training set molecules for the LFC parameter fitting and those experimental pK_a values.

Imidazole	Molecule	pK_a	Ref.	Thiol	Molecule	pK_a	Ref.
	2-Methyl-4-hydroxy-aminobenzimidazole	6.65	6		$C_2H_5OCH_2CH_2SH$	9.38	7
	2-Methylbenzimidazole	6.10			$C_2H_5OCOCH_2SH$	7.95	
	2-Methylimidazole	7.75			$C_6H_5CH_2SH$	9.43	
	4-Hydroxy-6-aminobenzimidazole	5.90			CH_2=$CHCH_2SH$	9.96	
	4-Hydroxy benzimidazole	5.30			$HOCH_2CHOHCH_2SH$	9.51	
	4-Methoxy benzimidazole	5.10			n-C_3H_7SH	10.65	
	4-Nitroimidazole	1.50			n-C_4H_9SH	10.66	
	6-Nitrobenzimidazole	3.05			t-$C_5H_{11}SH$	11.21	
	Benzimidazole	5.40			2-Mercaptoethanol	9.50	8
	Imidazole	6.95			2-Mercaptoethylamine	8.60	
	2-Methyl-4-hydroxy-6-nitrobenzimidazole	3.90			Thioglycolic acid	10.31	
	4-Hydroxy-6-nitrobenzimidazole	3.05			Thiophenol	7.8	9
	4-(2-4-dihydroxy phenyl)-imidazole	6.45			o-Aminothiophenol	6.59	
	4-Methyl-imidazole	7.45			3-Mercaptopropionicacid	10.27	
	6-Aminobenzidazole	6.00					
	Histamine	6.00					

Table S1(3). Training set molecules for the LFC parameter fitting and those experimental pK_a values.

Alcohol	Molecule	pK_a	Ref.	Phenol	Molecule	pK_a	Ref.
	CCl_3CH_2OH	11.80			2Cl-4NO_2-phenol	5.42	13
	$CHF_2CF_2CH_2OH$	11.34	10		$C_3H_5CH_2O_2C$-phenol	8.41	
	CH_2=$CHCH_2OH$	15.10			m-CH_3CO-phenol	9.19	
	CH_3CH_2OH	15.90			m-CH_3O-phenol	9.65	
	$CH_3OCH_2CH_2OH$	14.80			m-F-phenol	9.28	
	CH_3OH	15.54			m-$HOCH_2$-phenol	9.83	
	$CHCCH_2OH$	13.55			m-NH_2phenol	9.87	
	$CHCl_2CH_2OH$	12.89			o-OCH-phenol	6.79	
	$HOCH_2CF_2CH_2OH$	11.00			p-Br-phenol	9.34	
	CH_3OCH_2OH	14.80	11		p-$C_2H_5O_2C$-phenol	8.50	14
	$C(CH_2OH)_4$	14.10			p-C_6H_5-phenol	9.51	
	$HOCH_2CHOHCH_2OH$	14.40			p-CH_3O_2C-phenol	8.47	
	C_2H_5OH	16.00			p-CH_3S-phenol	9.53	
	CF_3CH_2OH	12.37			p-CH_3SO_2-phenol	7.83	
	$HOCH_2CH_2OH$	14.77			p-HO-phenol	9.96	
	$CF_3C(CH_3)_2OH$	11.60			p-NC-phenol	7.95	
	$CF_3CH(OH)CH_3$	11.80	12		p-O_2C-phenol	9.39	
					p-$(CH_3)_3N^+$phenol	8.00	

Table S2. The fitted parameters, RMSE, correlation factor r and $G(\mathrm{H}^+)$ for LFC/PCM in each chemical group, using 6-31++G(d,p).

	k^a	C_0	s	RMSE	r	$G(\mathrm{H}^+)^b$
Alcohol	0.238	-61.002	0.325	1.115	0.733	-255.9
Amine	0.380	-98.353	0.519	0.597	0.955	-258.8
Imidazole	0.310	-81.673	0.423	0.647	0.922	-263.2
Thiol	0.260	-66.928	0.354	0.866	0.716	-257.8
Phenol	0.223	-57.650	0.305	0.441	0.925	-258.0
Carboxyl	0.221	-60.102	0.302	0.622	0.853	-271.4
Total				0.721	0.977	

[a] Unit of k is mol/kcal. [b] Unit of $G(\mathrm{H}^+)$ is kcal/mol.

Table S3. The comparison of the computed and experimental pK_a values for amino acids using the PCM.

		pK_a		
Amino acid	Chemical group	LFC/PCM	direct PCM	expt[a]
Asp	Carboxyl	1.97	27.51	3.86
Cys	Thiol	7.10	31.10	8.33
Glu	Carboxyl	4.25	35.08	4.25
His(D/E)[b]	Imidazole	6.23/6.23	29.72/29.74	6.04
Lys	Amine	12.02	34.96	10.53
Tyr	Phenol	9.96	43.85	10.07

[a] Taken from ref. 15. [b] D and E denote the positions where protonation occurs, the delta and epsilon nitrogens, respectively.

References

1. J.F.J. Dippy, S.R.C. Hughes, A. Rozanski, *J. Chem. Soc*. 1959, 2492-2498

2. R.M.C. Dawson, et al., Data for Biochemical Research, Oxford, Clarendon Press, 1959

3. J. March, et al, Advanced Organic Chemistry, 3rd Ed, 1985

4. H.C. Brown, et al., in E.A. Braude, and F.C. Nachod Determination of Organic Structures by Physical Methods, Academic Press, New York, 1955

5. H.K. Hall Jr., *J. Am. Chem. Soc*. 1957, **79**, 5441-5444

6. T.C. Bruice, G.L. Schmir, *J. Am. Chem. Soc*. 1958, **80**, 148-156

7. M.M. Kreevoy, et al. *J. Am. Chem. Soc*. 1960, **82**, 4899-4902.

8. J.T. Edsall, Wyman and Jeffries, Biophysical Chemistry, Academic Press, Inc., New York, 1958

9. J.P. Danehy, Noel, C.J. *J. Am. Chem. Soc*. 1960, **82**, 2511-2515

10. R.N. Haszeldine, *J. Chem. Soc*. 1953, 1748-1757

11. P. Ballinger, F.A. Long, *J. Am. Chem. Soc*. 1960, **82**, 795-798

12. P. Ballinger, F.A. Long, *J. Am. Chem. Soc*. 1959, **81**, 1050-1053

13. V.E. Bower, R.A. Robinson, *J. Phys. Chem*. 1960, **64**, 1078-1079

14. R. Williams, pK_a data compiled by R. Williams. [Online accessed on]. http://www.chem.wisc.edu/areas/organic/index-chem.htm. 2011. [06/12/2018].

15. R. M. C. Dawson, D. C. Elliott, W. H. Elliott, K. M. Jones, *Data for Biochemical Research*. Clarendon Press Oxford: 1969; Vol. 316.