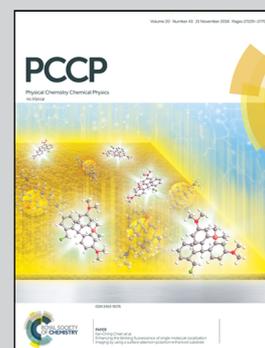


Showcasing research from the Group of Prof. Norio Yoshida and Prof. Haruyuki Nakano at Kyushu University, Japan

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

This work proposes a scheme for quantitatively computing the pK_a of hydrated molecules. It is based on 3D-RISM-SCF theory coupled with the linear fitting correction (LFC) scheme. In LFC/3D-RISM-SCF, pK_a values are evaluated using the Gibbs energy difference between the protonated and unprotonated states and the parameters fitted to the experimental values of training set systems. The computed pK_a values show quantitative agreement with the experimental data.

As featured in:



See Norio Yoshida *et al.*,
Phys. Chem. Chem. Phys.,
2018, 20, 27272.



Cite this: *Phys. Chem. Chem. Phys.*,
2018, 20, 27272

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,^a Yukako Kasai,^a Yuki Seno,^a Toru Matsui,^{ib} Yasuteru Shigeta,^{id} Norio Yoshida^{ib}*^a and Haruyuki Nakano^{id}^a

A scheme for quantitatively computing the acid dissociation constant, pK_a , of hydrated molecules is proposed. It is based on the three-dimensional reference interaction site model self-consistent field (3D-RISM-SCF) theory coupled with the linear fitting correction (LFC) scheme. In LFC/3D-RISM-SCF, pK_a values of target molecules are evaluated using the Gibbs energy difference between the protonated and unprotonated states calculated by 3D-RISM-SCF and the parameters fitted by the LFC scheme to the experimental values of training set systems. The pK_a values computed by LFC/3D-RISM-SCF show quantitative agreement with the experimental data.

Received 10th July 2018,
Accepted 21st August 2018

DOI: 10.1039/c8cp04354j

rsc.li/pccp

1. Introduction

Protonation and deprotonation are fundamental chemical reactions in solution and biological systems. Through these reactions, molecules change their charged states and interactions between surrounding environments. Such changes play an essential role in the solubility of molecules, higher-order structure formation of proteins, molecular recognition, and molecular transportation across membranes. An equilibrium constant of the reaction, an acid dissociation constant, K_a , or its more commonly used logarithmic value, $pK_a = -\log_{10} K_a$, can usually be determined experimentally, by titration, for small molecules. The pK_a values are strongly affected by the surrounding environment such as solvent and protein. Therefore, the pK_a value or the protonation state of an amino acid residue often changes drastically.^{1–6} In practice, the protonation state of dissociative amino acid residues is measured by the neutron diffraction method or the nuclear magnetic resonance method.^{7–10} However, these methods have some disadvantages. The former requires a large crystal of the target protein; in the latter, it is difficult to specify the residue of the signal origin of proteins that have multiple dissociative residues. Therefore, because of the

experimental difficulties, a theoretical method to compute pK_a values has attracted considerable attention.

The pK_a value is closely related to the Gibbs energy difference of the acid dissociation reaction, ΔG , according to the relationship

$$pK_a = \frac{\Delta G}{(\ln 10)RT} \quad (1)$$

where R and T are the gas constant and absolute temperature, respectively, and

$$\Delta G = G(A^-) + G(H^+) - G(HA) \quad (2)$$

where $G(X)$ denotes a Gibbs energy of species X . Here, HA and A^- represent the protonated and unprotonated states of an acid A . A straightforward way to calculate the Gibbs energy of a solvated molecule involves *ab initio* molecular-dynamics-based methods.^{11–14} However, these methods require substantial computational costs, and it is therefore impractical to apply them to complex molecular systems. A more compact method to handle the Gibbs energy of solvated molecules is the hybrid quantum mechanics and molecular mechanics (QM/MM) method, which is commonly used for pK_a evaluation.^{15–21} In this approach, only the reactive moiety is treated by the *ab initio* molecular orbital (MO) or Kohn–Sham density functional theory (KS-DFT) and the remaining parts are treated by classical molecular mechanics. Even more compact methods in computational cost are hybrid methods with the implicit solvation models such as the polarizable continuum model (PCM), and the statistical mechanics integral equation theory of liquids, such as reference interaction site model (RISM), or three-dimensional RISM

^a Department of Chemistry, Graduate School of Science, Kyushu University, 744, Motoooka, Nishiku, Fukuoka, 819-0395, Japan. E-mail: noriwo@chem.kyushu-univ.jp

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

^c Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cp04354j

(3D-RISM) theory.^{22–34} These methods produce qualitatively good solvation free energies within a reasonable computational time. However, for the quantitative evaluation of pK_a , computing the Gibbs energy of the proton, $G(H^+)$, is problematic, because the number of water molecules surrounding the excess proton to be handled by quantum mechanics and the structure of the excess proton–water cluster are unclear. Many theoretical approaches assume that the excess proton exists as the hydronium ion, H_3O^+ , and that such a simple treatment may cause an error in the pK_a value. Therefore, empirical approaches are widely employed to evaluate the pK_a value of amino acid residues.^{35–40} However, because such methods employ empirical parameters, there are reports that the methods sometimes produce unreliable values.^{41–43}

To achieve both low computational cost and high accuracy, Matsui *et al.* proposed a scheme based on the linear relationship between the pK_a and the Gibbs energy difference between HA and A^- .^{24–27} They suggested several types of methods with different approximation levels. In the present paper, we refer to the linear fitting correction (LFC) scheme as a generic term.

In the LFC scheme, pK_a values of target molecules are evaluated using parameters fitted by the least squares method to the experimental values of training set systems. Using this scheme, the computation of the Gibbs energy of a dissociated proton in solution can be circumvented. This scheme has been successfully applied to the evaluation of the pK_a values of amino acids. The results were in good agreement with the experimental observations. The original LFC scheme employs the PCM to take into account the solvent effect on the electronic structure. The PCM and related methods are widely used to investigate the chemical processes in solution.⁴⁴ However, since the solvent environment is regarded as a dielectric continuum characterized by the dielectric constant, it is difficult to reproduce the local molecular interactions, such as hydrogen bonding, and to define the dielectric constant of a heterogeneous environment, such as inside a protein.

In this paper, we propose a new scheme based on the LFC scheme employing the 3D-RISM as a solvent model. Hybrid methods of the 3D-RISM theory and the quantum chemical theory, such as KS-DFT and *ab initio* MO, have been proposed by Kovalenko, Sato, and Hirata. These methods are referred to as KS-DFT/3D-RISM or three-dimensional reference interaction site model self-consistent field (3D-RISM-SCF).^{45,46} 3D-RISM-SCF has been applied to various chemical processes in a solution, including the pK_a shift of drug molecules.^{32,47–54} The method allows us to treat a highly anisotropic solvent environment, such as inside a cavity and channel of a protein. Therefore, employing 3D-RISM-SCF, we expected to establish a method that is applicable to complex biological systems.

We first determined the parameters for the LFC scheme by a least square fitting based on the Gibbs energy of the training set molecules calculated by 3D-RISM-SCF and the corresponding experimental pK_a value. We also examined the basis set dependency on the performance of the scheme. The scheme was applied to amino acids to assess the transferability of the fitted parameters.

2. Method

Linear fitting correction method with empirical parameters

The pK_a value is related to the Gibbs energy difference of the acid dissociation reaction, ΔG , according to eqn (1). In the LFC scheme, eqn (1) is rewritten by introducing the scaling factor s :^{24,25}

$$pK_a = \frac{s\{G(A^-) - G(HA)\}}{(\ln 10)RT} + \frac{sG(H^+)}{(\ln 10)RT} = k\Delta G_0 + C_0 \quad (3)$$

where

$$k = \frac{s}{(\ln 10)RT} \quad (4a)$$

$$\Delta G_0 = G(A^-) - G(HA) \quad (4b)$$

$$C_0 = \frac{sG(H^+)}{(\ln 10)RT} \quad (4c)$$

The scaling factor s should be unity when the calculated Gibbs energy values are identical to exact values, and $k = 0.733 \text{ mol kcal}^{-1}$ when $s = 1$ at 298.15 K. The scaling factor s is an adjustable parameter, which corresponds to the activity coefficient of deprotonation reaction and corrects the systematic error of the computational method. The parameters k and C_0 were determined by the least square fitting to minimize the errors of pK_a values:^{24,25}

$$\varepsilon = \sum_i \{pK_{ai}^{\text{expt}} - (k\Delta G_{0,i} + C_0)\}^2 \quad (5)$$

where pK_{ai}^{expt} is an experimental pK_a value of molecule i and the summation over i is taken for all molecules in the training set that have the same dissociative chemical group and those pK_a values are known. $\Delta G_{0,i}$ is evaluated using *ab initio* MO or KS-DFT with a solvation model such as the PCM. The parameters k and C_0 are determined for each of the dissociative chemical groups, such as carboxyl, amine, alcohol, thiol, phenol, and imidazole.

3D-RISM-SCF theory

In the original LFC methods, the PCM is employed as a solvation model for ΔG_0 calculation. In the present study, we employed 3D-RISM-SCF instead of the PCM to take the solvation effect into account. As the details of the 3D-RISM-SCF method can be found in the literature, we only provide a brief explanation of the theory here.^{45,46}

The Gibbs energy of the solute molecule in the solvent at infinite dilution is defined as the sum of the solute electronic energy (E_0), solvation free energy ($\Delta\mu$), and the kinetic free energy (G_{kin}):

$$G = E_0 + \Delta\mu + G_{\text{kin}} \quad (6)$$

where E_0 is given by

$$E_0 = \langle \Psi | \hat{H}_0 | \Psi \rangle \quad (7)$$

and where \hat{H}_0 and Ψ denote the Hamiltonian of the isolated molecules and the electronic wave function of solute molecules. The kinetic free energy, G_{kin} , includes the vibrational, rotational

and translational energies, which are obtained in a usual quantum mechanics manner after the normal mode analysis. In the present study, we ignore the kinetic term, G_{kin} , because the change in this term due to the deprotonation reaction is rather small and the adjustable parameter can absorb the error emerging from this approximation. However, it is noted that one cannot neglect this term in the case that significant geometry variations occur in the reaction. The solvation free energy is given by

$$\Delta\mu = k_{\text{B}}T \sum_i^{\text{solvent}} \rho_i \int \left[\frac{1}{2} h_i(\mathbf{r})^2 \Theta(-h_i(\mathbf{r})) - c_i(\mathbf{r}) - \frac{1}{2} h_i(\mathbf{r}) c_i(\mathbf{r}) \right] d\mathbf{r} \quad (8)$$

where i runs over the solvent interaction sites. Θ , k_{B} , T , and ρ_i denote the Heaviside step function, the Boltzmann constant, the absolute temperature, and the number density of the solvent site i , respectively. $h_i(\mathbf{r})$ and $c_i(\mathbf{r})$ are total and direct correlation functions, obtained by solving the 3D-RISM equation coupled with the Kovalenko–Hirata closure:⁵⁵

$$h_i(\mathbf{r}) = \sum_j^{\text{solvent}} c_j(\mathbf{r}) * X_{ij}(\mathbf{r}) \quad (9)$$

$$h_i(\mathbf{r}) = \begin{cases} \exp(d_i(\mathbf{r})) - 1 & \text{for } d_i(\mathbf{r}) < 0 \\ -d_i(\mathbf{r}) & \text{for } d_i(\mathbf{r}) \geq 0 \end{cases} \quad (10a)$$

$$d_i(\mathbf{r}) = -\frac{1}{k_{\text{B}}T} u_i(\mathbf{r}) + h_i(\mathbf{r}) - c_i(\mathbf{r}) \quad (10b)$$

where $*$ denotes a convolution integral. $X_{ij}(\mathbf{r})$ is a solvent susceptibility function, obtained by solving the RISM equation for pure solvent systems prior to 3D-RISM-KH calculation. $u_i(\mathbf{r})$ is an interaction potential function between a solute molecule and solvent molecules at position r . In the 3D-RISM-SCF framework, $u_i(\mathbf{r})$ is given by

$$u_i(\mathbf{r}) = 4 \sum_j^{\text{solute}} \varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} + \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right\} + q_i \sum_j^{\text{solute}} \frac{Z_j}{r_{ij}} - q_i \int \frac{|\Psi(r')|^2}{|r-r'|} d\mathbf{r}' \quad (11)$$

where ε_{ij} and σ_{ij} are the Lennard-Jones parameters (with usual meanings), and q_i denotes the point electronic charge on the solvent site i . Z_j is a nuclear charge of atom j .

3. Computational details

In the present study, the parameters for six types of chemical groups were determined, namely, alcohol, amine, imidazole, thiol, phenol, and carboxyl. Table S1 in the ESI† summarizes the training data sets for parameter fitting.

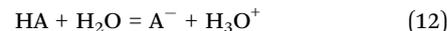
Prior to the Gibbs energy calculation, the structure optimization of protonated (HA) and unprotonated (A^-) states was performed at the B3LYP/6-31++G(d,p) level, in water, with the PCM, for all the training set molecules. For the Gibbs energy calculation, two different sizes of basis sets were employed, 6-31++G(d,p) and 6-31G, to examine the basis set dependency of the parameter fitting.

The parameters used in the 3D-RISM calculation were temperature of 298.15 K and density of solvent water of 1.0 g cm^{-3} . The Lennard-Jones parameters for solute molecules were taken from the general Amber force field (GAFF) parameter set with antechamber software.⁵⁶ The extended simple point charge model (SPC/E) parameter set for the geometrical and potential parameters for the solvent water was employed with modified hydrogen parameters ($\sigma = 1.0 \text{ \AA}$; and $\varepsilon = 0.056 \text{ kcal mol}^{-1}$).^{57,58} The grid spacing for the 3D grid was 0.5 \AA and the number of grid points on each axis was 128.

All calculations were performed with a modified version of the GAMESS program package, for which the 3D-RISM-SCF program has been implemented.^{59–62}

4. Results and discussion

Table 1 summarizes the fitted parameters. The computed and experimental pK_{a} values are compared in Fig. 1, which also depicts the pK_{a} values computed without using the LFC scheme, as well as the LFC values. For the pK_{a} computation without the LFC, we assumed the acid dissociation reaction



and the associated pK_{a} formula

$$\text{pK}_{\text{a}} = \frac{G(\text{A}^-) + G(\text{H}_3\text{O}^+) - G(\text{HA}) - G(\text{H}_2\text{O})}{(\ln 10)RT} \quad (13)$$

where the Gibbs energy of each molecule is calculated by 3D-RISM-SCF. Hereafter, we refer to this treatment as a direct 3D-RISM-SCF scheme.

Fig. 1 shows that the accuracy of the computed pK_{a} value is drastically improved for all the chemical groups by introducing the LFC scheme to 3D-RISM-SCF; we refer to this as the LFC/3D-RISM-SCF scheme. LFC/3D-RISM-SCF shows an excellent score for both the root mean square error (RMSE), 0.709, and the correlation factor, $s = 0.978$. Although the direct 3D-RISM-SCF scheme shows good correlation with the experimental value, $r = 0.912$, the absolute pK_{a} values are significantly overestimated (the RMSE of direct 3D-RISM-SCF is 18.43), indicating that the Gibbs energy of the reaction is overestimated by direct 3D-RISM-SCF. The overestimation of the Gibbs energy of the reaction is suppressed by the scaling factor s , which shows the range 0.43 to 0.67. In addition, the contribution of the Gibbs energy of the proton, $G(\text{H}^+)$, is also well parameterized by C_0 .

Table 1 The fitted parameters, RMSE, correlation factor r , and $G(\text{H}^+)$ in each chemical group, using the 6-31++G(d,p) basis set

	k^a	C_0	s	RMSE	r	$G(\text{H}^+)^b$
Alcohol	0.443	-112.460	0.604	1.175	0.698	-254.0
Amine	0.396	-98.600	0.540	0.469	0.973	-248.9
Imidazole	0.338	-84.960	0.462	0.629	0.927	-251.1
Thiol	0.490	-123.403	0.668	0.821	0.750	-252.0
Phenol	0.317	-78.788	0.432	0.423	0.931	-248.5
Carboxyl	0.319	-81.552	0.435	0.661	0.832	-255.3
Total				0.709	0.978	

^a Unit of k is mol kcal^{-1} . ^b Unit of $G(\text{H}^+)$ is kcal mol^{-1} .

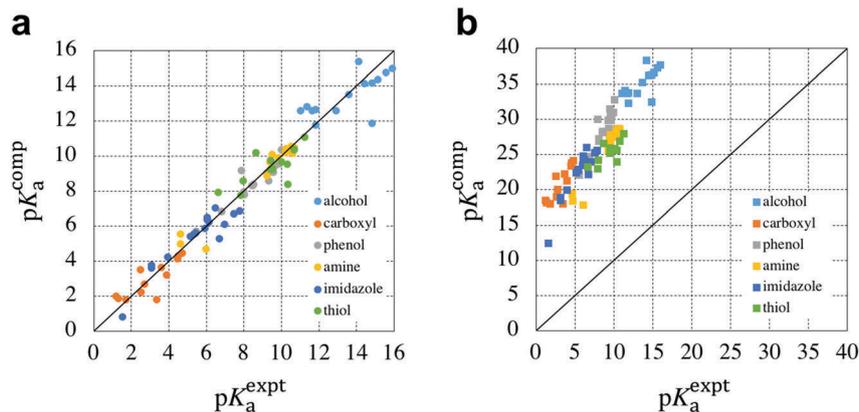


Fig. 1 Comparison of the computed pK_a values with the experimental values, determined using (a) LFC/3D-RISM-SCF and (b) direct 3D-RISM-SCF. The references for the experimental values are given in Table S1 in ESI.†

Table 2 The fitted parameters, RMSE, correlation factor r , and $G(H^+)$ in each chemical group, using the 6-31G basis set

	k^a	C_0	s	RMSE	r	$G(H^+)^b$
Alcohol	0.257	-62.559	0.351	1.154	0.711	-243.3
Amine	0.335	-84.256	0.457	0.524	0.966	-251.3
Imidazole	0.300	-77.250	0.410	0.667	0.917	-257.2
Thiol	0.421	-101.552	0.574	0.934	0.658	-241.0
Phenol	0.235	-57.753	0.321	0.375	0.946	-245.2
Carboxyl	0.258	-66.346	0.351	0.605	0.861	-257.3
Total				0.726	0.977	

^a Unit of k is mol kcal⁻¹. ^b Unit of $G(H^+)$ is kcal mol⁻¹.

Table 2 summarizes the $G(H^+) = C_0/k$ values. The $G(H^+)$ values take in the range -255 to -248 kcal mol⁻¹ is shown. These results are comparable with those obtained in the previous LFC approach by Matsui *et al.*, which ranged from -268 to -246 kcal mol⁻¹, and in the experimental and other theoretical approaches, which ranged from -264 to -259 kcal mol⁻¹.²⁵

In Fig. 2, the computed and the experimental pK_a values are compared (individual panels are used for each chemical group). The alcohol and thiol show relatively large RMSE and small correlation coefficient values. In the case of alcohols, a methoxyethanol (CH₃OCH₂OH) shows remarkable deviation (the experimental pK_a value is 14.8 and the LFC/3D-RISM-SCF value is 12.1). If the methoxyethanol is removed from the training set, then the RMSE and correlation factor are improved (0.88 and 0.87, respectively). This molecule probably has additional factors that affect the fitting parameters other than the chemical group. Hence, it is likely necessary to consider additional parameters or factors to improve accuracy further. In other cases, the thiol group, a mercaptoethylamine, and a thioglycolic acid show relatively large deviations. These molecules have multiple dissociative groups, which might have some influence on the accuracy.

To consider the basis set dependence of the parameters and their accuracy, another basis set, 6-31G, was examined. The results of the fitted parameters are summarized in Table 2, and the experimental and computed pK_a values are compared in Fig. 3. The parameter k for 6-31G takes a value of 0.23–0.42,

while for 6-31++G(d,p) the range in values is 0.3–0.48. Here, with the 6-31G basis set, both the RMSE and correlation values are slightly worse than with the 6-31++G(d,p) basis set; however, the accuracy of the results determined using LFC/3D-RISM-SCF is acceptable. The direct 3D-RISM-SCF results for the thiol group highlight interesting behavior; thiol shows considerably shifted values (compared with other groups). When using a small basis set, the description of the electronic structure is inadequate, which may be the cause of the shift. Surprisingly, such an irregular behavior of a specific chemical group can be compensated by the parameters in the LFC scheme. This result clearly indicates that LFC/3D-RISM-SCF allows us to use the computationally cheaper basis set, thereby providing a significant advantage when the scheme is applied to large molecular systems such as biomolecules.

To assess the transferability of the fitted parameters to the biomolecules, the pK_a calculations for the dissociative amino acids by LFC/3D-RISM-SCF were examined. Table 3 and Fig. 4 respectively compare the experimental and computed pK_a values of several amino acid side chains. Here, we examined an aspartic acid (Asp), glutamic acid (Glu), cysteine (Cys), histidine (His), lysine (Lys), and tyrosine (Tyr). The computed pK_a values by LFC/3D-RISM-SCF show quantitative agreement with the experimental data; in contrast, direct 3D-RISM-SCF shows serious deviation (the RMSE of LFC/3D-RISM-SCF is 0.39, that of direct 3D-RISM-SCF is 18.7). These results indicate that the parameters created by LFC/3D-RISM-SCF have good transferability and that it can be used for the pK_a prediction of proteins.

To assess the solvent model dependencies on the effectiveness of the scheme, we also performed calculations with the PCM for the same training set. In Fig. 4, the computed pK_a values by the LFC and direct schemes are compared with the experimental pK_a values. The fitted parameters, RMSE, and correlation factors are summarized in Table S2 in the ESI.† Although the correlation of the direct PCM values with the experimental values is relatively low, 0.80, and the total RMSE is very high, 27.9, the computed pK_a values by the LFC/PCM scheme show high accuracy and good correlation (the total RMSE and correlation factor are 0.72 and 0.98) as noted by

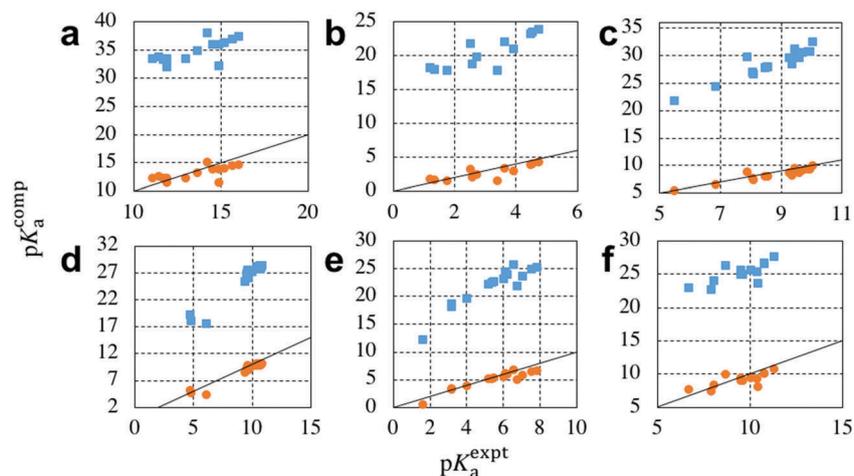


Fig. 2 Comparison of the computed pK_a values with the experimental values. Values for each of the chemical groups are presented in separate panels: (a) alcohol, (b) carboxyl, (c) phenol, (d) amine, (e) imidazole, and (f) thiol. The filled squares and circles denote the pK_a values determined by direct 3D-RISM-SCF and LFC/3D-RISM-SCF, respectively.

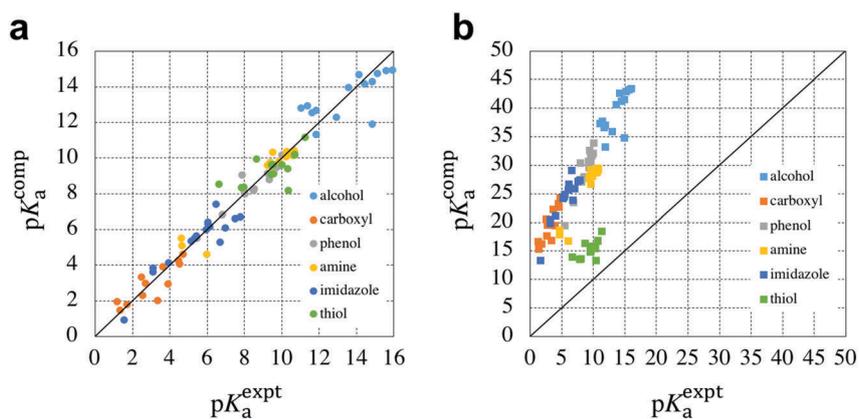


Fig. 3 Comparison of the computed pK_a values using the 6-31G basis set with the experimental values, using (a) LFC/3D-RISM-SCF and (b) direct 3D-RISM-SCF. The references for the experimental values are given in Table S1 in ESI.†

Table 3 Computed and experimental pK_a values of amino acids

Amino acid	Chemical group	pK_a		
		LFC/3D-RISM-SCF	Direct 3D-RISM-SCF	Experimental ^a
Asp	Carboxyl	3.92	22.86	3.86
Cys	Thiol	9.14	25.07	8.33
Glu	Carboxyl	3.94	22.92	4.25
His(D/E) ^b	Imidazole	6.31/6.27	22.92/24.39	6.04
Lys	Amine	10.69	24.29	10.53
Tyr	Phenol	9.64	28.95	10.07

^a Taken from ref. 63. ^b D and E denote the positions where protonation occurs, the delta and epsilon nitrogens, respectively.

Matsui *et al.* (Fig. 5a and b).²⁵ LFC/3D-RISM-SCF shows slightly better values in the RMSE and correlation factor than LFC/PCM. In the case of the application of the fitted parameters to the amino acids, LFC/PCM shows excellent transferability (Fig. 5c and Table S3 in the ESI†). The RMSE for the amino acids by LFC/PCM is 1.03, and that by LFC/3D-RISM-SCF is 0.39.

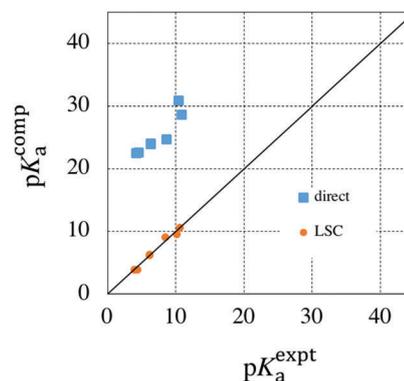


Fig. 4 Comparison of the computed pK_a values for amino acids with values determined experimentally. The filled squares and circles denote the computed pK_a values by direct 3D-RISM-SCF and LFC/3D-RISM-SCF, respectively.

This result indicates that LFC/3D-RISM-SCF has better transferability of the LFC scheme to biomolecules than the PCM.

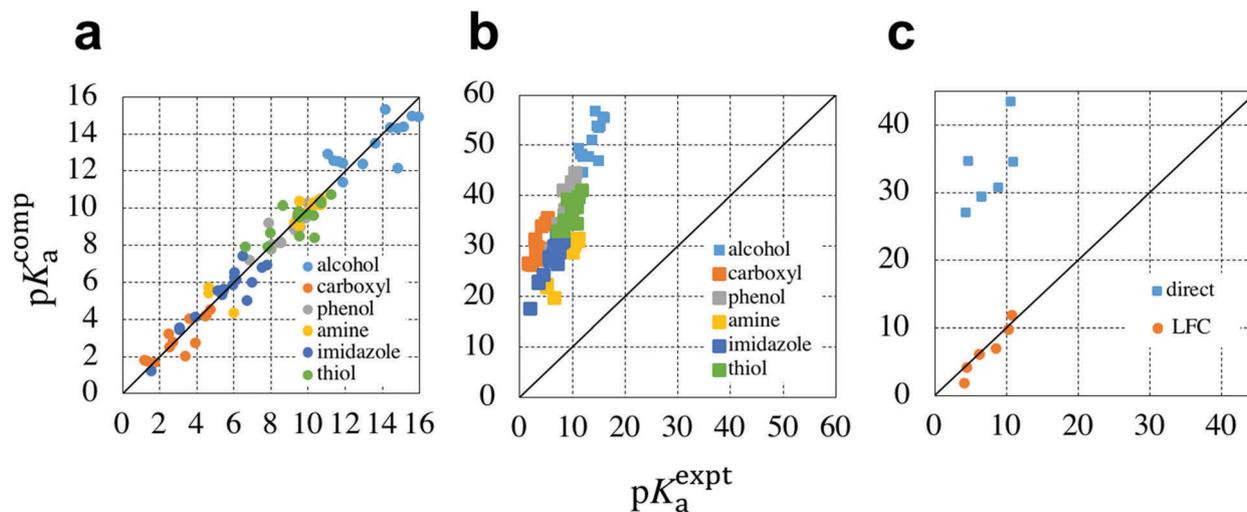


Fig. 5 Comparison of the computed pK_a values with the experimental values. (a) Comparison of the LFC/PCM values with the training set molecules. (b) Comparison of the direct PCM values with the training set molecules. (c) Comparison of LFC/PCM and direct PCM values with the amino acids. The filled squares and filled circles denote the direct and LFC values, respectively.

5. Summary

We have proposed a scheme for computing pK_a values based on 3D-RISM-SCF with the LFC scheme. According to this scheme, the pK_a value is computed by utilizing the linear relationship between the pK_a value and the Gibbs energy difference between the protonated and unprotonated states of target molecules. The parameters were determined by the least square fitting for the experimental values of a training set for each chemical group. The parameters introduced here correspond to the Gibbs energy of the excess proton and the scaling factor. The error of the computed pK_a values arising from the treatment of the excess proton in water and the computational condition such as basis sets for electronic structure calculations are well absorbed by the parameters. It is suggested that, with this scheme, the computationally cheap basis set can be used for pK_a calculations. The parameters were applied to the amino acid molecules which were not included in the training set, and a good performance was found. Furthermore, LFC/3D-RISM-SCF shows better performance than the LFC/PCM scheme, especially in terms of the transferability of the parameters.

These features may allow us to use this scheme for the prediction of pK_a values of amino acids in biological systems. In order to apply the LFC/3D-RISM-SCF scheme to amino acids in proteins, a method taking account of environment other than water, such as surrounding residue and ions, which are not currently considered, is necessary. Previously, we proposed the use of advanced methods of 3D-RISM-SCF, in combination with quantum chemical methods, applicable to the biomolecular systems, which we referred to as the quantum mechanics/molecular mechanics/RISM (QM/MM/RISM) and the fragment molecular orbital/3D-RISM (FMO/3D-RISM) methods.^{61,62} The combinational use of the scheme proposed here, and QM/MM/RISM or FMO/3D-RISM, may be a powerful

tool to tackle the problems related to the protonation and deprotonation of dissociated amino acid residues in biological systems. Such studies with the LFC/3D-RISM-SCF are in progress.

Matsui *et al.* investigated the redox potentials of several half reactions, metal complexes, and physiologically active molecules using the LFC scheme with the PCM.^{64–67} As the present LFC/3D-RISM-SCF scheme outperforms the LFC/PCM scheme, it is expected that extension to the redox potentials can improve the accuracy of the estimation. These issues will be addressed in future work.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by Grants-in-Aid [16H00842, 16K05519, 18K05036] from MEXT, Japan. NY also acknowledges support of TOYOTA Riken Scholar from the Toyota Physical and Chemical Research Institute. The authors are grateful to Prof. Yoshihiro Watanabe (Kyushu University) for helpful discussions.

References

- 1 A. A. Gorfe, P. Ferrara, A. Caflisch, D. N. Marti, H. R. Bosshard and I. Jelesarov, *Proteins*, 2002, **46**, 41–60.
- 2 E. L. Mehler, M. Fuxreiter, I. Simon and E. B. Garcia-Moreno, *Proteins*, 2002, **48**, 283–292.
- 3 D. G. Isom, C. A. Castaneda, B. R. Cannon and B. E. Garcia-Moreno, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 5260–5265.
- 4 J. M. Ho, *Aust. J. Chem.*, 2014, **67**, 1441–1460.
- 5 J. M. Ho, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2859–2868.

- 6 J. Ho and M. L. Coote, *Theor. Chem. Acc.*, 2009, **125**, 3.
- 7 M. Tanokura, M. Tasumi and T. Miyazawa, *Biopolymers*, 1976, **15**, 393–401.
- 8 K. Bartik, C. Redfield and C. M. Dobson, *Biophys. J.*, 1994, **66**, 1180–1184.
- 9 M. D. Joshi, A. Hedberg and L. P. McIntosh, *Protein Sci.*, 1997, **6**, 2667–2670.
- 10 A. Jadhav, V. S. Kalyani, N. Barooah, D. D. Malkhede and J. Mohanty, *ChemPhysChem*, 2015, **16**, 420–427.
- 11 A. K. Tummanapelli and S. Vasudevan, *J. Phys. Chem. B*, 2014, **118**, 13651–13657.
- 12 A. K. Tummanapelli and S. Vasudevan, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6383–6388.
- 13 D. H. Yu, R. B. Du and J. C. Xiao, *J. Comput. Chem.*, 2016, **37**, 1668–1671.
- 14 A. W. Sakti, Y. Nishimura and H. Nakai, *J. Chem. Theory Comput.*, 2018, **14**, 351–356.
- 15 A. Warshel, *J. Phys. Chem.*, 1979, **83**, 1640–1652.
- 16 A. Warshel, *Biochemistry*, 1981, **20**, 3167–3177.
- 17 A. Warshel, F. Sussman and G. King, *Biochemistry*, 1986, **25**, 8368–8372.
- 18 W. L. Jorgensen and J. M. Briggs, *J. Am. Chem. Soc.*, 1989, **111**, 4190–4197.
- 19 K. M. Merz, *J. Am. Chem. Soc.*, 1991, **113**, 3572–3575.
- 20 G. H. Li and Q. Cui, *J. Phys. Chem. B*, 2003, **107**, 14521–14528.
- 21 D. Riccardi, P. Schaefer and Q. Cui, *J. Phys. Chem. B*, 2005, **109**, 17715–17733.
- 22 Y. Fu, L. Liu, R. C. Li, R. Liu and Q. X. Guo, *J. Am. Chem. Soc.*, 2004, **126**, 814–822.
- 23 T. N. Brown and N. Mora-Diez, *J. Phys. Chem. B*, 2006, **110**, 9270–9279.
- 24 T. Matsui, A. Oshiyama and Y. Shigeta, *Chem. Phys. Lett.*, 2011, **502**, 248–252.
- 25 T. Matsui, T. Baba, K. Kamiya and Y. Shigeta, *Phys. Chem. Chem. Phys.*, 2012, **14**, 4181.
- 26 T. Baba, T. Matsui, K. Kamiya, M. Nakano and Y. Shigeta, *Int. J. Quantum Chem.*, 2014, **114**, 1128–1134.
- 27 T. Matsui, Y. Shigeta and K. Morihashi, *J. Chem. Theory Comput.*, 2017, **13**, 4791–4803.
- 28 H. Sato and F. Hirata, *J. Phys. Chem. A*, 1998, **102**, 2603–2608.
- 29 N. Yoshida, R. Ishizuka, H. Sato and F. Hirata, *J. Phys. Chem. B*, 2006, **110**, 8451–8458.
- 30 K. Kido, H. Sato and S. Sakaki, *Int. J. Quantum Chem.*, 2012, **112**, 103–112.
- 31 K. Kido, H. Sato and S. Sakaki, *J. Phys. Chem. B*, 2009, **113**, 10509–10514.
- 32 Y. Seno, N. Yoshida and H. Nakano, *J. Mol. Liq.*, 2016, **217**, 93–98.
- 33 J. Frau, N. Hernandez-Haro and D. Glossman-Mitnik, *Chem. Phys. Lett.*, 2017, **671**, 138–141.
- 34 J. H. Jensen, C. J. Swain and L. Olsen, *J. Phys. Chem. A*, 2017, **121**, 699–707.
- 35 H. Li, A. D. Robertson and J. H. Jensen, *Proteins*, 2005, **61**, 704–721.
- 36 J. H. Jensen, H. Li, A. D. Robertson and P. A. Molina, *J. Phys. Chem. A*, 2005, **109**, 6634–6643.
- 37 J. C. Shelley, A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin and M. Uchimaya, *J. Comput. -Aided Mol. Des.*, 2007, **21**, 681–691.
- 38 M. Meloun and S. Bordovska, *Anal. Bioanal. Chem.*, 2007, **389**, 1267–1281.
- 39 C. Z. Liao and M. C. Nicklaus, *J. Chem. Inf. Model.*, 2009, **49**, 2801–2812.
- 40 M. Rostkowski, M. H. M. Olsson, C. R. Sondergaard and J. H. Jensen, *BMC Struct. Biol.*, 2011, 11.
- 41 T. Kesvatera, B. Jonsson, E. Thulin and S. Linse, *J. Mol. Biol.*, 1996, **259**, 828–839.
- 42 M. Tollinger, J. D. Forman-Kay and L. E. Kay, *J. Am. Chem. Soc.*, 2002, **124**, 5714–5717.
- 43 I. Andre, S. Linse and F. A. A. Mulder, *J. Am. Chem. Soc.*, 2007, **129**, 15805–15813.
- 44 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3093.
- 45 A. Kovalenko and F. Hirata, *J. Chem. Phys.*, 1999, **110**, 10095–10112.
- 46 H. Sato, A. Kovalenko and F. Hirata, *J. Chem. Phys.*, 2000, **112**, 9463–9468.
- 47 A. Kovalenko and F. Hirata, *J. Mol. Liq.*, 2001, **90**, 215–224.
- 48 S. Gusarov, T. Ziegler and A. Kovalenko, 2005 *International Conference on Memes, Nano and Smart Systems, Proceedings*, 2005, pp. 207–208, DOI: 10.1109/Icmens.2005.108.
- 49 S. Gusarov, T. Ziegler and A. Kovalenko, *J. Phys. Chem. A*, 2006, **110**, 6083–6090.
- 50 S. Gusarov, T. A. Fedorova, Y. Y. Dmitriev and A. Kovalenko, *Int. J. Quantum Chem.*, 2009, **109**, 1672–1675.
- 51 J. Hong, N. Yoshida, S.-H. Chong, C. Lee, S. Ham and F. Hirata, *J. Chem. Theory Comput.*, 2012, **8**, 2239–2246.
- 52 S. R. Stoyanov, C. X. Yin, M. R. Gray, J. M. Stryker, S. Gusarov and A. Kovalenko, *Can. J. Chem.*, 2013, **91**, 872–878.
- 53 Y. Tanaka, N. Yoshida and H. Nakano, *J. Comput. Chem.*, 2015, **36**, 1655–1663.
- 54 Y. Tanaka, Y. Kawashima, N. Yoshida and H. Nakano, *J. Comput. Chem.*, 2017, **38**, 2411–2419.
- 55 A. Kovalenko and F. Hirata, *Chem. Phys. Lett.*, 2001, **349**, 496–502.
- 56 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.
- 57 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.
- 58 S. Ten-No, F. Hirata and S. Kato, *J. Chem. Phys.*, 1994, **100**, 7443–7453.
- 59 M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347–1363.
- 60 N. Yoshida and F. Hirata, *J. Comput. Chem.*, 2006, **27**, 453–462.
- 61 N. Yoshida, Y. Kiyota and F. Hirata, *J. Mol. Liq.*, 2011, **159**, 83–92.

- 62 N. Yoshida, *J. Chem. Phys.*, 2014, **140**, 214118.
- 63 R. M. C. Dawson, D. C. Elliott, W. H. Elliott and K. M. Jones, *Data for biochemical research*, Clarendon Press, Oxford, 1969.
- 64 T. Matsui, Y. Kitagawa, M. Okumura and Y. Shigeta, *J. Phys. Chem. A*, 2015, **119**, 369–376.
- 65 T. Matsui, Y. Kitagawa, Y. Shigeta and M. Okumura, *J. Chem. Theory Comput.*, 2013, **9**, 2974–2980.
- 66 T. Matsui, Y. Kitagawa, M. Okumura, Y. Shigeta and S. Sakaki, *J. Comput. Chem.*, 2013, **34**, 21–26.
- 67 I. Kurniawan, K. Kawaguchi, M. Shoji, T. Matsui, Y. Shigeta and H. Nagao, *Bull. Chem. Soc. Jpn.*, 2018, accepted.