Chemistry Letters, 2024, **53**, upad009
https://doi.org/10.1093/chemle/upad009
Advance access publication 4 December 2023

Letter

Assessment of the applicability of the LFC/3D-RISM-SCF scheme for pK_a prediction in methanol solutions

Ryo Fujiki[1,2], Toru Matsui[3], Yasuteru Shigeta[2], Norio Yoshida[1,4,*], Haruyuki Nakano[1]

[1]Department of Chemistry, Graduate School of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

[2]Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan

[3]Department of Chemistry, Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan

[4]Department of Complex Systems Science, Graduate School of Informatics, Nagoya University, Furocho, Chikusa-ku, Nagoya 464-8601, Japan

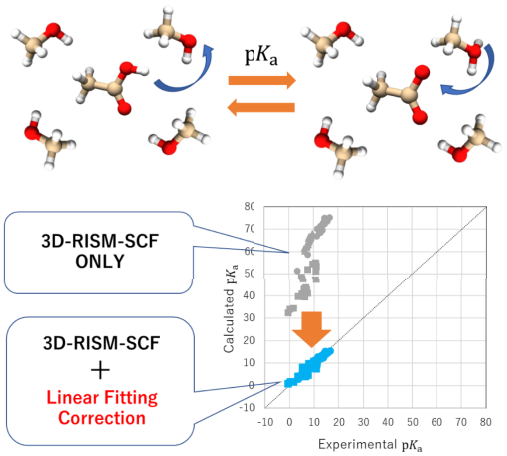
*Corresponding author: Department of Chemistry, Graduate School of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan; Department of Complex Systems Science, Graduate School of Informatics, Nagoya University, Furocho, Chikusa-ku, Nagoya 464-8601, Japan. Email: noriwo@nagoya-u.jp

Abstract

The applicability of the linear fitting correction with the three-dimensional reference interaction-site model self-consistent field (LFC/3D-RISM-SCF) scheme, a pK_a prediction scheme, for methanol solutions was investigated. The correlation between experimental and predicted pK_a values of dissociative molecules with phenol, amine, and carboxyl derivatives was examined. The pK_a values of the LFC/3D-RISM-SCF results showed a good linear correlation with the experimental pK_a. This result demonstrates that the LFC/3D-RISM-SCF method can be applied to a variety of solvents other than water.

Keywords: 3D-RISM-SCF, methanol solution, pK_a.

Graphical Abstract



The applicability of the linear fitting correction with the three-dimensional reference interaction-site model self-consistent field (LFC/3D-RISM-SCF) scheme, a pK_a prediction scheme, for methanol solutions was investigated. The correlation between experimental and predicted pK_a values of dissociative molecules with phenol, amine, and carboxyl functional groups was examined. The pK_a values of the LFC/3D-RISM-SCF results showed a good linear correlation with the experimental pK_a. This result demonstrates that the LFC/3D-RISM-SCF method can be applied to a variety of solvents other than water.

The solvent effect of an organic solvent is an important factor in various fields of chemistry, physics, and biology. For example, it relates to the reaction rate and reaction pathway in organic synthesis, and the solubility and absorption rate of drugs in pharmacy and drug design.[1] Methanol is a simple protic organic solvent that is used as a model solvent for molecules with both hydrophobicity and polarity.

When solvated molecules contain proton-dissociable functional groups, the properties of the molecules change significantly depending on their dissociation state. Therefore, it is necessary to know the correct protonation state to predict the property, function, and structure of solvated molecules. The equilibrium of the deprotonation reaction is greatly affected by solvent effects—the pK_a in water is therefore very different from that in methanol.

[Received on 8 October 2023; revised on 31 October 2023; accepted on 31 October 2023; corrected and typeset on 27 January 2024]

Recently, we proposed a highly accurate pK_a prediction scheme, called the linear fitting correction with the three-dimensional reference interaction-site model self-consistent field (LFC/3D-RISM-SCF) scheme—an extension of the pK_a prediction method originally proposed by Matsui et al.—and successfully applied it to pK_a prediction in an aqueous solution.[2–4] This scheme combines the data-learning technique and quantum chemical computation coupled with the integral equation theory of molecular liquids. It solves the problem common to the computational prediction of pK_a, i.e. the problem of evaluating the Gibbs energy of excess protons, and simultaneously gives a highly accurate evaluation of solute–solvent interactions.

Earlier, one of the authors confirmed the transferability of parameters suitable for pK_a in water toward that in DMSO using quantum chemical calculations with the polarizable continuum method instead of the 3D-RISM-SCF scheme.[5] In the present study, the LFC/3D-RISM-SCF scheme is applied to molecules in a methanol solution. First, new parameters for the methanol solution are determined. Thereafter, the pK_a values obtained from the data sets are compared with experimental data. The applicability of the LFC/3D-RISM-SCF scheme for a methanol solution is then discussed.

The pK_a value is related to the Gibbs energy difference of the acid dissociation reaction, $HA \rightleftharpoons A^- + H^+$, as

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

where

$$\Delta G = G(A^-) + G(H^+) - G(HA). \tag{2}$$

This equation is rewritten by introducing the scaling factor s as,

$$\begin{aligned} pK_a &= \frac{s\{G(A^-) - G(HA)\}}{(\ln 10)RT} + \frac{s\{G(H^+)\}}{(\ln 10)RT} \\ &= k\Delta G_0 + C_0, \end{aligned} \tag{3}$$

with

$$\begin{aligned} k &= \frac{s}{(\ln 10)RT}, \\ \Delta G_0 &= G(A^-) - G(HA), \\ C_0 &= \frac{s\{G(H^+)\}}{(\ln 10)RT}, \end{aligned} \tag{4}$$

where R and T denote the gas constant and absolute temperature, respectively. The scaling factor s is an adjustable parameter, correcting the systematic errors of the computational method such as those originating from basis functions or density functionals. The parameters k and C_0 are determined by least square fitting to minimize the errors of pK_a values,

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

where $pK_{a,i}^{expt}$ 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 already known. $\Delta G_{0,i}$ is evaluated using 3D-RISM-SCF.[6] The parameters k and C_0 are determined for each of the dissociative chemical groups, such as carboxyl and phenol in the present study.

The Gibbs energy of the solvate molecule X is expressed as

$$G(X) = E_{solute} + \Delta\mu, \tag{6}$$

where E_{solute} and $\Delta\mu$ denote the solute electronic energy and the solvation free energy, respectively, which are given by

$$E_{solute} = \langle\Psi|\hat{H}_0|\Psi\rangle, \tag{7}$$

and

$$\Delta\mu = \rho/k_B T \sum_\alpha \int d\boldsymbol{r}[h_\alpha(\boldsymbol{r})^2\Theta(-h_\alpha(\boldsymbol{r}))/2 - c_\alpha(\boldsymbol{r}) - h_\alpha(\boldsymbol{r})c_\alpha(\boldsymbol{r})/2], \tag{8}$$

where $c_\alpha(\boldsymbol{r})$ and $h_\alpha(\boldsymbol{r})$, respectively, are the direct and total correlation function obtained by solving the 3D-RISM equation coupled with the Kovalenko–Hirata closure.[7] ρ, and k_B are the number density of solvent methanol, and the Boltzmann constant. Θ denotes the Heaviside step function.

The parameters for the functional groups (phenol, carboxyl, and amine) were determined based on the training data set taken from ref.[8] The number of molecules included in the training and test set are summarized in Table 1, and the list of molecules is given in the online supplementary material. Prior to the Gibbs energy calculation, structure optimizations of the protonated (HA) and deprotonated (A^-) states were performed at the B3LYP/6-31++G(d,p) level,[9–14] in methanol, with the polarizable continuum model,[15] for all the training set molecules.

The following parameters were used in the 3D-RISM-SCF calculation:[16–18] temperature 298.15 K and density of the solvent methanol 0.79 g cm^{-3}. The Lennard-Jones parameters for solute molecules were taken from the general Amber force field (GAFF) parameter set with antechamber software.[19] The OPLS-UA parameter set for the geometrical and potential parameters for the solvent methanol was employed.[20] The grid spacing for the 3D grid was 0.5 Å and the number of grid points on each axis was 128. All calculations were performed using a modified version of the GAMESS program package, for which the 3D-RISM-SCF program was implemented.[21–25]

The parameters were determined by least squares fitting for the phenol, carboxyl, and amine groups in methanol solution. The LFC parameters are summarized in Table 2. The results indicate a good correlation between the experimental and theoretical pK_a values for the phenol and carboxyl groups, which suggests that the LFC scheme is applicable to nonaqueous solvent systems. The estimated Gibbs energies of protons for phenol, carboxyl, and amine groups are −267, −245, and

Table 1. Number of molecules included in the training and test molecular sets.

	Training set	Test set
Phenol	24	11
Carboxyl	24	10
Amine	24	8

Table 2. LFC parameters for the phenol, carboxyl, and amine groups in methanol solution.

	k^a	C_0	RMSE[b]	R^{2b}	s	$G(H^+)^a$
Phenol	0.383	−102.5	0.69	0.93	0.52	−267.4
Carboxyl	0.260	−68.4	0.37	0.81	0.36	−246.6
Amine	0.333	−82.2	1.50	0.80	0.46	−262.4

[a]The unit of k and $G(H^+)$ are mol kcal^{-1} and kcal mol^{-1}, respectively.
[b]RMSE and R^2 denote the root mean square error from the experimental value and the coefficient of determination, respectively.

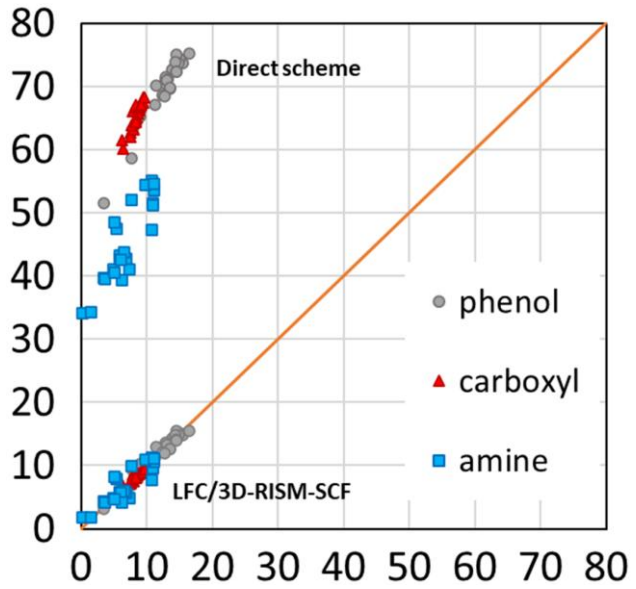


Fig. 1. Computed **pK_a** values for the training molecular set are plotted against the experimental values. The direct and LFC/3D-RISM-SCF schemes are compared.

−262 kcal mol^{-1}, respectively, which are similar to those reported in other computational studies.[26–28] This implies that in this scheme, high accuracy can be achieved by assigning different parameters to each functional group. In the high accuracy limit of the computational value of Gibbs energy, the scaling factor s should be 1. The s values are quite different depending on the functional group, namely 0.52, 0.36, and 0.45 for phenol, carboxyl, and amine groups, respectively. This behavior is different from the previous study for the water solvent.[4] In the case of water solvent, s values of phenol and carboxyl have similar values to each other and the amine has a larger s value compared with those for the phenol and carboxyl. These differences can be attributed to differences in solvent models. Unlike water, methanol molecules have hydrophobic methyl groups and are more anisotropic than water. This is expected to complicate the interactions between solute–solvent interactions. Therefore, further improvements may be possible by introducing the repulsive bridge correction (RBC) or other methods to handle such anisotropic solvent interactions.[29,30]

In Fig. 1, computed pK_a values for the training molecular set for both the direct and the LFC/3D-RISM-SCF schemes are plotted against the experimental values. The result of each scheme seems to have a good linear correlation. However, the LFC/3D-RISM-SCF scheme has smaller errors than the direct scheme, which clearly indicates that the correction of the LFC/3D-RISM-SCF scheme is also effective for the

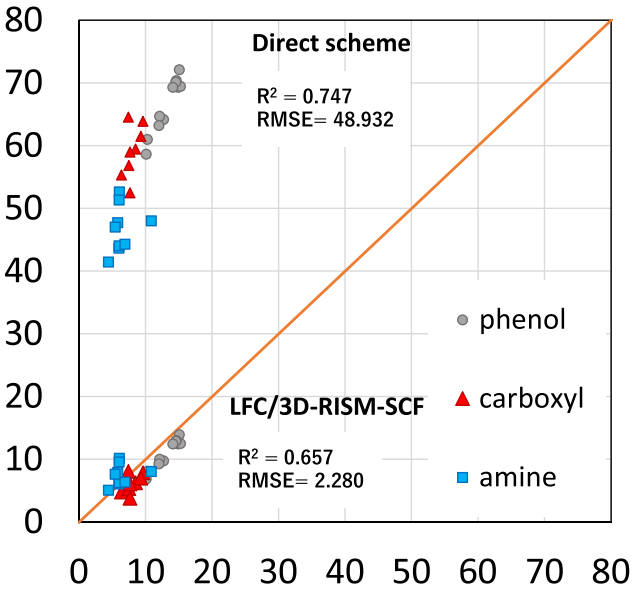


Fig. 2. Computed **pK_a** values for the test molecular set are plotted against the experimental values. The direct and LFC/3D-RISM-SCF schemes are compared.

prediction of pK_a in methanol. As seen in Fig. 1, the pK_a values of the phenol and carboxyl groups are overestimated by about 55 pK_a and those of the amine group by about 35 pK_a. These differences may be attributed to differences in the charges of the molecules participating in the reaction. Namely, the molecules in the amine group have a positive charge in the protonated state, whereas they have a neutral charge in the deprotonated state. On the other hand, the molecules in the phenol and carboxyl groups are charge neutral in the protonated state and have a negative charge in the deprotonated state. In the state with a net charge, strong hydrogen bonds form between the solute and solvents. In the case of the amine group, the oxygen of the hydroxyl group of the solvent forms a hydrogen bond with the excess proton of solute amine, whereas in the case of the phenol and carboxyl groups, the hydrogen of the hydroxyl group of the solvent coordinates with the oxygen of the solute. The difference in the hydrogen bond form is thought to be reflected in the difference in the degree of overestimation. The reason why only amines have different effective Gibbs energy values for excess proton, as mentioned earlier, may also be due to the difference in proton sources as explained above. The LFC/3D-RISM-SCF method was shown to be able to handle such differences because of the molecular nature, as the parameters are determined for each functional group.

For the evaluation of the LFC/3D-RISM-SCF method in methanol, the relationships between calculated and experimental pK_a values are examined for the test molecular set and plotted in Fig. 2. The errors of each group are corrected by fitting parameters, and the correlations in each group are given in the figure. At a glance, the absolute value of pK_a is drastically improved by applying the LFC/3D-RISM-SCF scheme. On the other hand, the correlation of all the groups (R^2: 0.657) is lower than the results from the direct scheme (R^2: 0.747)—values are nonetheless comparable. The correlations of each group are as follows: phenol 0.941, carboxyl 0.310, and amine 0.421. Although the phenol group showed a good correlation with the experimental data, the carboxyl and amine groups showed poor correlations. The poor overall

correlation is considered to be due to the low R^2 values of the carboxyl and amines.

The amine group, in particular, shows a relatively low correlation, insufficient to correct only using fitting parameters. In the determination of the fitting parameters of amine, as the correlation is already poorer than that of other groups, this result of the test molecular set indicates the probability of additional correction parameters or more detailed separation of amine group than current references like aniline, quinoline, and pyridine. This behavior of poor correlation for amines has been reported in previous studies.[3] The development of novel methods to improve this behavior is a future challenge.

In this study, the applicability of the LFC/3D-RISM-SCF scheme in methanol solution was examined. Three molecular groups which have different functional groups were considered: the phenol group, the carboxyl group, and the amine group. The calculated results showed improvement in the pK_a values compared with the direct 3D-RISM-SCF scheme and showed good agreement with experimental values. These findings suggest that LFC/3D-RISM-SCF is also useful for predicting pK_a in methanol. Although this study was conducted on methanol solutions, a study has been conducted on the applicability of the original LFC scheme with PCM to the DMSO solution.[5] These results suggest that the LFC scheme can be applied to a wide range of solutions.

We also found that the accuracy of pK_a prediction varies depending on the proton source. We cannot propose a method to improve this at this time, but it may be possible to improve the accuracy by taking the molecular orientation of the solvent into account. For example, one possibility is to introduce the repulsive bridge correction method into 3D-RISM, or to use the molecular Ornstein-Zernike-SCF method instead of 3D-RISM-SCF.[29–33] Such a study is in progress in the authors' group.

In summary, the results also indicate the extensibility of the LFC/3D-RISM-SCF scheme to other organic solvents and mixed solvents. 3D-RISM-SCF can easily handle multicomponent solvent systems, which are difficult to handle with continuum models. The scheme proposed in this article should be an effective pK_a prediction tool in complex systems.

Supplementary data

Supplementary material is available at *Chemistry Letters* online.

Funding

This work was financially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Nos. 19H02677 and 22H05089). Numerical calculations were partially conducted at the Research Center for Computational Science, Institute for Molecular Science, National Institutes of Natural Sciences (Project 22-IMS-C076), and using MCRP-S at the Center for Computational Sciences, University of Tsukuba. NY also received support from the MEXT Program: Data Creation and Utilization-Type Material Research and Development Project Grant No. JPMXP1122714694.

Conflict of interest statement. None declared.

Data availability

Data supporting the results of this study are available from the corresponding author upon reasonable request.

References

1. T. N. Brown, N. Mora-Diez, *J. Phys. Chem. B.* **2006**, *110*, 9270–9279.
2. T. Matsui, A. Oshiyama, Y. Shigeta, *Chem. Phys. Lett.* **2011**, *502*, 248–252.
3. T. Matsui, Y. Shigeta, K. Morihashi, *J. Chem. Theory Comput.* **2017**, *13*, 4791–4803.
4. R. Fujiki, Y. Kasai, Y. Seno, T. Matsui, Y. Shigeta, N. Yoshida, H. Nakano, *Phys. Chem. Chem. Phys.* **2018**, *20*, 27272.
5. K. Hengphasatporn, T. Matsui, Y. Shigeta, *Chem. Lett.* **2020**, *49*, 307–310.
6. H. Sato, A. Kovalenko, F. Hirata, *J. Chem. Phys.* **2000**, *112*, 9463–9468.
7. A. Kovalenko, F. Hirata, *Chem. Phys. Lett.* **2001**, *349*, 496–502.
8. E. L. M. Miguel, P. L. Silva, J. R. Pliego, *J. Phys. Chem. B.* **2014**, *118*, 5730–5739.
9. A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
10. P. J. Stephens, F. J. Devlin, C. F. Chabolowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623–11627.
11. R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724–728.
12. P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta.* **1973**, *28*, 213–222.
13. W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257–2261.
14. T. Clark, J. Chandrasekhar, G. W. Spiznagel, P. V. R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294–301.
15. J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999–3093.
16. D. Beglov, B. Roux, *J. Phys. Chem. B.* **1997**, *101*, 7821–7826.
17. D. Beglov, B. Roux, *J. Chem. Phys.* **1996**, *104*, 8678–8689.
18. A. Kovalenko, F. Hirata, *Chem. Phys. Lett.* **1998**, *290*, 237–244.
19. J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
20. W. L. Jorgensen, J. D. Madura, C. J. Swenson, *J. Am. Chem. Soc.* **1984**, *106*, 6638–6646.
21. M. W. Schmidt, K. K. Baldridge, 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, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347–1363.
22. N. Yoshida, F. Hirata, *J. Comput. Chem.* **2006**, *27*, 453–462.
23. N. Yoshida, Y. Kiyota, F. Hirata, *J. Mol. Liq.* **2011**, *159*, 83–92.
24. N. Yoshida, *J. Chem. Phys.* **2014**, *140*, 214118.
25. N. Yoshida, *IOP Conf. Series: Mater. Sci. Eng.* **2020**, *773*, 012062.
26. G. J. Tawa, I. A. Topol, S. K. Burt, R. A. Caldwell, A. A. Rashin, *J. Chem. Phys.* **1998**, *109*, 4852–4863.
27. J. J. Fifen, M. Nsangou, Z. Dhaouadi, O. Motapon, N.-E. Jaidane, *J. Chem. Theory Comput.* **2013**, *9*, 1173–1181.
28. F. Rived, M. Rosés, E. Bosch, *Anal. Chim. Acta.* **1998**, *374*, 309–324.
29. A. Kovalenko, F. Hirata, *J. Chem. Phys.* **2000**, *113*, 2793–2805.
30. K. Kido, D. Yokogawa, H. Sato, *J. Chem. Phys.* **2012**, *137*, 024106.
31. N. Yoshida, S. Kato, *J. Chem. Phys.* **2000**, *113*, 4974–4984.
32. N. Yoshida, *Proc. Comput. Sci.* **2011**, *4*, 1214–1221.
33. R. Ishizuka, N. Yoshida, *J. Chem. Phys.* **2013**, *139*, 084119.