Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics

Supporting Information

3D-RISM-SCF Method with Dual Solvent Boxes for Highly Polarized System: Application to 1,6-Anhydrosugar Formation Reaction of Phenyl α- and β-D-Glucosides under Basic Condition

Shinji Aono[a], Takashi Hosoya[b,c], and Shigeyoshi Sakaki[a*]

[a] *Fukui Institute for Fundamental Chemistry, Kyoto University, Nishihiraki-cho, Takano, Sakyo-ku, Kyoto 606-8103, Japan*

[b] *Institute for Integrated Cell-Material Science, Kyoto University, Ushinomiya-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan*

[c] *Present address: Department of Chemistry, University of Natural Resources and Life Science, Vienna, University Research Center Tulln (UFT), Konrad Lorenz Strasse 24, 3430 Tulln an der Donau, Austria*

References

The complete description of reference [65] is, as follows:
GAUSSIAN 09, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, 2010.

The complete description of reference [69] is, as follows:
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. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, J. Comput. Chem., **14**, 1347-1363, (1993).

Details to solve RISM integral equation with dual solvent box protocol

To obtain the solvent spatial distribution, $g_\xi(\mathbf{r})$, in equilibrium with the solute electronic and nuclear structures, the solute-solvent 3D-RISM equation is given, as follows:

$$\tilde{\eta}_\xi^{(x)}(\mathbf{k}) = \tilde{h}_\xi^{(x)}(\mathbf{k}) - \tilde{c}_\xi^{(x)}(\mathbf{k}), \tag{S1}$$

$$\tilde{\eta}_\xi^{(x)}(\mathbf{k}) = \sum_\delta \tilde{c}_\delta^{(x)}(\mathbf{k})\tilde{H}_{\delta\xi}^{VV}(k) - \tilde{c}_\xi^{(x)}(\mathbf{k}), \tag{S2}$$

$$\tilde{H}_{\delta\xi}^{VV}(k) \equiv \tilde{\omega}_{\delta\xi}^{V}(k) + \rho^V \tilde{h}_{\delta\xi}^{VV}(k), \tag{S3}$$

where the superscript (x) represents (f) and (n); see Eqs. (8) and (9) in the main text. The tilde denotes the quantity in reciprocal space. The ω^V is the solvent intra-molecular correlation function and the h^{VV} is the solvent-solvent total correlation function obtained by solving the solvent-solvent 1D-RISM equation. To solve the solute-solvent total correlation function, $h_\xi(\mathbf{r})$, one more other closure relation is needed. In the present calculations, we used the Kovalenko-Hirata (KH) closure,

$$c_\xi^{(x)}(\mathbf{r}) = \begin{cases} \exp\left[-\beta u_\xi(\mathbf{r}) + \eta_\xi^{(x)}(\mathbf{r})\right] - \eta_\xi^{(x)}(\mathbf{r}) - 1 & \text{for } -\beta u_\xi(\mathbf{r}) + \eta_\xi^{(x)}(\mathbf{r}) < 0 \\ -\beta u_\xi(\mathbf{r}) & \text{for } -\beta u_\xi(\mathbf{r}) + \eta_\xi^{(x)}(\mathbf{r}) \geq 0 \end{cases}, \tag{S4}$$

where β is $1/k_BT$, where k_B and T are Boltzmann constant and temperature, respectively. The $u_\xi(\mathbf{r})$ is the interaction potential between the solute molecule and solvent site ξ:

$$u_\xi(\mathbf{r}) = q_\xi^V \left\{ \sum_{\mu\nu} D_{\mu\nu} V_{\mu\nu}^{\mathrm{ES}}(\mathbf{r}) + \sum_b \frac{Z_b}{\mid \mathbf{r} - \mathbf{R}_b \mid} \right\} + u_\xi^{\mathrm{LJ}}(\mathbf{r}), \tag{S5}$$

where Z_b is the nuclear charge on solute site b and $u_\xi^{\mathrm{LJ}}(\mathbf{r})$ is the Lennard-Jones (LJ) interaction between the solute molecule and the solvent site ξ.

From the numerical point of view, Eqs. (S2) and (S4) are not directly solved by the usual procedure applied in the RISM method. Instead of Eqs. (S2) and (S4), we usually remove the long-range Coulomb potential from the direct correlation function and interaction potential, as follows:

$$c_\xi^{s\,(x)}(\mathbf{r}) = c_\xi^{(x)}(\mathbf{r}) - \phi_\xi^l(\mathbf{r}), \tag{S6}$$

$$\phi_\xi^s(\mathbf{r}) = -\beta u_\xi(\mathbf{r}) - \phi_\xi^l(\mathbf{r}), \tag{S7}$$

where ϕ_ξ^l is the long-range Coulomb potential for the solvent site ξ,

$$\phi_\xi^l(\mathbf{r}) = -\beta q_\xi^V \sum_b \frac{(Q_b + Z_b)\,\mathrm{erf}(\lambda \mid \mathbf{r} - \mathbf{R}_b \mid)}{\mid \mathbf{r} - \mathbf{R}_b \mid}. \tag{S8}$$

By using the short-range parts of $c_\xi^{(x)}$ and ϕ_ξ, Eqs. (S2) and (S4) can be converted to the following convenient form,

$$\tilde{\theta}_\xi^{(x)}(\mathbf{k}) = \sum_\delta \tilde{c}_\delta^{s\,(x)}(\mathbf{k}) \tilde{H}_{\delta\xi}^{VV}(k) - c_\xi^{s\,(x)}(\mathbf{k}), \tag{S9}$$

$$c_\xi^{s\,(x)}(\mathbf{r}) = \begin{cases} \exp[\theta_\xi^{(x)}(\mathbf{r}) + \gamma_\xi(\mathbf{r}) + \phi_\xi^s(\mathbf{r})] - (\theta_\xi^{(x)}(\mathbf{r}) + \gamma_\xi(\mathbf{r})) + 1 & \text{for } d_\xi^{(x)}(\mathbf{r}) < 0 \\ \phi_\xi^s(\mathbf{r}) & \text{for } d_\xi^{(x)}(\mathbf{r}) \geq 0 \end{cases}, \tag{S10}$$

where

$$\tilde{\gamma}_\xi(\mathbf{k}) = -4\pi\beta \left\{ \sum_\delta q_\delta^V \tilde{H}_{\delta\xi}^{VV}(k) \right\} \left\{ \sum_{b=1,n_u} \frac{Q_b + Z_b}{k^2} e^{i\mathbf{k}\cdot\mathbf{R}_b - k^2/4\lambda^2} \right\}, \tag{S11}$$

$$d_\xi^{(x)}(\mathbf{r}) = \theta_\xi^{(x)}(\mathbf{r}) + \gamma_\xi(\mathbf{r}) + \phi_\xi^s(\mathbf{r}). \tag{S12}$$

In Eqs. (S9) and (S10), the numerical error coming from the 3D fast Fourier transform (3D-FFT) algorithm becomes negligible, because the damping parameter λ rapidly decreases the short-range function, $c_\xi^{s\,(x)}$, in comparison with the box size of solvents. To solve the $\theta_\xi^{(x)}$ and the $c_\xi^{s\,(x)}$, it is convenient to provide the $\gamma_\xi(\mathbf{r})$ at a cubic grid of $2N_0$ points/axis with the spacing by $L_0/2$. In the neutral system, the $\gamma_\xi(\mathbf{r})$ can be easily evaluated by the 3D-FFT algorithm, because the $\tilde{\gamma}_\xi(\mathbf{k})$ becomes zero value at $\mathbf{k} = \mathbf{0}$. On the other hand, in the charged system, the $\tilde{\gamma}_\xi(\mathbf{k})$ becomes non-zero value at $\mathbf{k} = \mathbf{0}$. To avoid the non-zero value problem of $\tilde{\gamma}_\xi(\mathbf{k})$ at $\mathbf{k} = \mathbf{0}$, the $\tilde{\gamma}_\xi(\mathbf{k} + \Delta\mathbf{k}/2)$ was usually evaluated at the reciprocal grid space with spacing by $\Delta\mathbf{k}$ and converted to the $\gamma_\xi(\mathbf{r})$ by the 3D-FFT algorithm. This means that the $\tilde{\gamma}_\xi(\mathbf{k})$ at $\mathbf{k} = \mathbf{0}$ was interpolated with sine and/or cosine functions.

In the present calculation based on the infinite dilution approximation of one charged solute, we can reduce the numerical error in the above evaluation of $\gamma_\xi(\mathbf{r})$ by dividing the $\gamma_\xi(\mathbf{k})$ in Eq. (S11) into two terms, as follows:

$$\tilde{\gamma}_\xi(\mathbf{k}) = -4\pi\beta \left\{ \sum_\delta q_\delta^V \tilde{H}_{\delta\xi}^{VV}(k) \right\} \left\{ \frac{q_{\mathrm{tot}}^U}{k^2} e^{-k^2/4\lambda^2} + \sum_{b=1,n_u+1} \frac{Q_b + Z_b}{k^2} e^{i\mathbf{k}\cdot\mathbf{R}_b - k^2/4\lambda^2} \right\}, \tag{S13}$$

where q_{tot}^U represents the total charge of solute molecule. Q_{n_u+1} and $\mathbf{R}_{n_u+1}$ are taken to be $-q_{\mathrm{tot}}^U$ and $\mathbf{0}$, because the molecular center of solute, $\bar{\mathbf{R}}_c$, is practically taken as $\mathbf{0}$ in the present 3D-RISM calculation; see Eqs. (10) and (11) in the main text. In Eq. (S13), due to the charge neutrality of system, the second term can be converted by the 3D-FFT algorithm without the non-zero value problem at $\mathbf{k} = \mathbf{0}$. Also, the first term can be converted by the

1D-FFT algorithm with a very fine grid spacing and a very large number of grids, leading to the numerically improved calculations of $c_\xi^{(x)}(\mathbf{r})$ and $h_\xi^{(x)}(\mathbf{r})$.

As a result, Eqs. (8) and (9) in the main text can be evaluated in the convenient forms, as follows:

$$c_\xi(\mathbf{r}) = \mathrm{Sw}(\mathbf{r})c_\xi^{s\,(f)}(\mathbf{r}) + (1-\mathrm{Sw}(\mathbf{r}))c_\xi^{s\,(n)}(\mathbf{r}) - \phi_\xi^l(\mathbf{r}), \tag{S14}$$

$$h_\xi(\mathbf{r}) = \mathrm{Sw}(\mathbf{r})\Big(c_\xi^{s\,(f)}(\mathbf{r}) + \theta_\xi^{(f)}(\mathbf{r})\Big) + (1-\mathrm{Sw}(\mathbf{r}))\Big(c_\xi^{s\,(n)}(\mathbf{r}) + \theta_\xi^{(n)}(\mathbf{r})\Big) + \gamma_\xi(\mathbf{r}). \tag{S15}$$

This dual solvent box protocol can efficiently reduce both CPU time and memory usage without the loss of accuracy, as shown in Supporting Information Tables S1 and S2.

It is noted that the multi solvent box protocol can be easily implemented by newly adding the different size of the solvent box to the dual solvent box protocol but the box length of the smallest one must be necessarily taken to be long enough to sufficiently damp the short-range terms of $c_\xi^{s\,(f)}(\mathbf{r})$ and $\theta_\xi^{(f)}(\mathbf{r})$.

Comparison of free energy profile between 3D-RISM-SCF and PCM methods

As shown in Figures 1 snd S1, the calculated free energies of **RS** and **IS** were not largely different between the 3D-RISM-SCF-RHF/MP2 and PCM/MP2 methods; the free energies of $\mathbf{RS}_\beta$ and $\mathbf{IS}_\alpha$ were calculated to be about 1 to 2 kcal/mol more stable in the 3D-RISM-SCF-RHF/MP2 calculation than in the PCM/MP2 calculation. The calculated free energies of **TS** and **PS** were, however, considerably different between the 3D-RISM-SCF-RHF/MP2 and PCM/MP2 methods; these values calculated by the 3D-RISM-SCF-RHF/MP2 method are about 7 to 10 kcal/mol higer than those calculated by the PCM/MP2 method.

To make clear the origin of this difference in the calculated activation barrier between the 3D-RISM-SCF-RHF/MP2 and the PCM/MP2 methods, the free energy components are compared in Table 1 of the main text. In Table 1, the geometry relaxation energy, E_{relax}, was calculated to be somewhat larger in the 3D-RISM-SCF-RHF/MP2 calculation than the PCM/MP2 one, reflecting the tendency that the solute geometry optimized by the 3D-RISM-SCF-RHF/MP2 method changes more largely than that optimized by the PCM/MP2 method does, when going from gas phase to aqueous phase. However, this geometry relaxation does not considerably influence the free energy profile in aqueous phase. On the other hand, the electronic reorganization energy, E_{reorg}, tends to decrease the activation free energy in aqueous phase but this contribution is apparently smaller than that of the solvation free energy, $\Delta\mu$, which significantly increases the activation free energy and reaction free energy in aqueous phase.

In particular, both two methods provide the solvation free energy, $\Delta\mu$, in the following orders: $\mathbf{RS}_\alpha > \mathbf{IS}_\alpha > \mathbf{PS}_\alpha > \mathbf{TS}_\alpha$ and $\mathbf{IS}_\beta > \mathbf{RS}_\beta > \mathbf{TS}_\beta > \mathbf{PS}_\beta$. In phenyl α-D-glucoside, the activation free energy and the reaction free energy were calculated to be higher by the 3D-RISM-SCF-RHF/MP2 than the PCM/MP2, by 8.1 kcal/mol and 5.8 kcal/mol, respectively. In phenyl β-D-glucoside, the corresponding values were also calculated to be higher by the 3D-RISM-SCF-RHF/MP2 than the PCM/MP2 by 8.2 kcal/mol and 8.5 kcal/mol, respectively. As a result, the activation free energy and the reaction free energy are higher in the 3D-RISM-SCF-RHF/MP2 than in the PCM/MP2 but the tendency of the higher reactivity of the β-anomer becomes somewhat smaller in the 3D-RISM-SCF-RHF/MP2 than in the PCM/MP2. These different features of the free energy profiles between two methods mainly arise from the accuracy of the evaluation of solvation free

energy, $\Delta\mu$; remember that the solvation effect by hydrogen bond is considerably involved in the present reaction and that the RISM-SCF method can describe the solvation effect by hydrogen bond but the PCM does not well.

Comparison of solvation effects between 3D- and 1D-RISM methods.

From the viewpoint of the RISM theory, it is interesting to compare the calculated solvation free energy between the 3D- and 1D-RISM-SCF methods. Because the pure point charge approximation for the whole region gave rise to the failure in the SCF-convergence of the 1D- and 3D-RISM-SCF calculations for this reaction, the $\Delta\mu$ and $\Delta\mu^{\mathrm{LRA}}$ values in both of the 3D-RISM and the 1D-RISM methods with the point charge approximation (hereafter denoted as 3D-RISM(PCA) and 1D-RISM(PCA)) were evaluated by using the converged ESP charges which are self-consistently determined by the present 3D-RISM-SCF-RHF calculation. [1] The $\Delta\mu$ and $\Delta\mu^{\mathrm{LRA}}$ values are different between the 1D-RISM and 3D-RISM(PCA) calculations, but the relative values to $\mathbf{IS}_\beta$ are similar between these two calculations, interestingly; see Table S2 for the $\Delta\mu$ and $\Delta\mu^{\mathrm{LRA}}$ energy changes by the 3D-RISM-SCF and 3D-RISM(PCA) calculations. These results indicates that the approach based on the pure point charge approximation is reliable semi-quantitatively at least and hence this point charge approximation can be reasonably applied to analysis of solvation effect.

Table S2 exhibits four important results: (i) The solvation free energy evaluated by the 3D-RISM-SCF method with the dual solvent boxes, [0.5 Å, 128 points/axis] and [0.25 Å, 128 points/axis], is almost the same as that evaluated by the 3D-RISM-SCF method with only a single fine solvent box, [0.25 Å, 256 points/axis], where the deviation is less than 0.1 kcal/mol. This means that the fine grid spacing is not necessary to describe the solvation structure in the region distant from the solute and that the present dual solvent box protocol practically well works. (ii) The calculation with the normal solvent box [0.5 Å, 128 points/axis] provides moderately different solvation free energy from that with the fine solvent box [0.25 Å, 256 points/axis], where the difference is less than 0.6 kcal/mol. This result suggests that the geometry optimization with a single solvent box, [0.5 Å, 128 points/axis], appropriately works in general but the present method with the dual solvent boxes much better works. (iii) The solvation free energy evaluated by the 1D-RISM(PCA) calculation is about 25 kcal/mol smaller than that evaluated by the 3D-RISM(PCA) calculation. However, the solvation free energy calculated by the 1D-RISM(PCA) changes in almost the same manner as that by the 3D-RISM(PCA) method when going from **RS** to **PS**. (iv) The linear response solvation free energy, $\Delta\mu^{\mathrm{LRA}}$, mostly changes in parallel with the solvation free

energy, $\Delta\mu$, indicating that the decomposition analysis of the linear response solvation free energy in Section 4.4 of the main text presents reasonable results.

In the 1D-RISM-SCF-RHF method, the solvation free energy can be easily decomposed into the contributions from the individual solute sites, as follows:

$$\Delta\mu^{1\mathrm{D}} = \sum_b \Delta\mu_b^{1\mathrm{D}} = \frac{\rho^V}{\beta} \sum_{b,\xi} \int dr 4\pi r^2 \left\{ \frac{1}{2} h_{b\xi}^2(r)\Theta(-h_{b\xi}(r)) - c_{b\xi}(r) - h_{b\xi}(r)c_{b\xi}(r) \right\}. \tag{S16}$$

where the KH closure is used in Eq. (S16) to compare the solvation free energy between the 1D-RISM and 3D-RISM methods. This decomposition of solvation free energy can be straightforwardly performed in other different expressions of solvation free energy such as Gaussian fluctuation (GF) and hyper-netted chain (HNC). [2,3] The changes of the 1D-RISM solvation free energy, $\Delta\mu^{1\mathrm{D}}$, and its components, $\{\Delta\mu_b^{1\mathrm{D}}\}$, are shown in Supporting Information Table S3. As shown in Table S3, the $\Delta\mu^{1\mathrm{D}}$ energies of the reactive sugar moieties and the phenoxyl groups are much more destabilized than the $\Delta\mu^{\mathrm{LRA}}$ but that of the solvent-exposed moiety is more stabilized than the $\Delta\mu^{\mathrm{LRA}}$. Despite of these differences, the energy component, $\{\Delta\mu_b^{1\mathrm{D}}\}$, changes through the reaction in a similar manner to the $\Delta\mu^{\mathrm{LRA}}$, as discussed in Section 4.5 of the main text. This means that the 1D-RISM method is also useful to successfully make analysis of solvation effect with the solute electronic structure converged by the present 3D-RISM-SCF calculation.

[1] : The comparison between the 1D- and the 3D-RISM-SCF methods can be made only with the PCA procedure, because the 1D-RISM-SCF method can not employ the ESP directly evaluated with the solute wave function; remember that the ESP used in the 1D-RISM must be averaged over the orientation around the individual solute sites.

[2] : S. J. Singer and D. Chandler, *Mol. Phys.* 1985, **55**, 621.

[3] : D. Chandler, Y. Singh, and D. M. Richardson, *J. Chem. Phys.* 1984, **81**, 1975.

Solvation around the O1 and O5 atoms in $\mathbf{RS}_\alpha$ and $\mathbf{RS}_\beta$

In Table 3 of the main text, the $\Delta\mu^{\mathrm{LRA}}$ energies of O1 and O5 atoms exhibit the tendency of the larger stabilizing solvation of the β-anomer. The ratios of O1 and O5 ESP charges in $\mathbf{RS}_\beta$ to those in $\mathbf{RS}_\alpha$, $q^{\mathrm{ESP}}_{\mathrm{O1}\beta}/q^{\mathrm{ESP}}_{\mathrm{O1}\alpha}$ and $q^{\mathrm{ESP}}_{\mathrm{O5}\beta}/q^{\mathrm{ESP}}_{\mathrm{O5}\alpha}$, are calculated to be 0.98 and 1.12, respectively. The extent of the larger $\Delta\mu^{\mathrm{LRA}}$ stabilization energies of the O1 and O5 atoms in the β-anomer is not sufficiently explained by this solute charge difference, indicating that the microscopic solvation difference between two anomers must contribute to the larger stabilizing solvation free energy of the β anomer. To understand this feature in view of microscopic solvation structure, we wish to compare here the difference in the solvation around the O1 and O5 atoms between $\mathbf{RS}_\alpha$ and $\mathbf{RS}_\beta$.

The solvent radial distribution around the O1 atom clearly indicates that the Hw site of water tends to more favorably approach the O1 atom of $\mathbf{RS}_\beta$ than that of $\mathbf{RS}_\alpha$, as shown in Figure 5 of the main text. This is because the intra-molecular O2-H$\cdots$O1 hydrogen bond exists in $\mathbf{RS}_\alpha$ and this prevents the Hw site of water from approaching the O1 atom of $\mathbf{RS}_\alpha$; see Figure 1 of the main text. As a result, the electrostatic interaction between the negatively charged O1 atom and the positively charged Hw site more largely stabilizes $\mathbf{RS}_\beta$ than $\mathbf{RS}_\alpha$, as shown in Table 3.

On the other hand, the solvent radial distribution around the O5 atom exhibits the interesting difference between two anomers, as shown in Figure 5 of the main text. In $\mathbf{RS}_\alpha$ and $\mathbf{RS}_\beta$, the O5 atom is similarly exposed to the solvent, indicating that the first solvation shell of water solvent around the O5 atom is not largely different between two anomers. However, Figure 5 indicates that the Ow-interacting solvation, where the Ow site of water is closer to the O5 atom than the Hw site is, increases more largely at the region, about 3.0 Å, of $\mathbf{RS}_\alpha$ than that of $\mathbf{RS}_\beta$. Judging from the distance (3.0 Å) between the O5 atom and the Ow-increasing region, this Ow-increasing region does not correspond to the water solvent interacting with the O1 atom through the Hw-O1 hydrogen bond. To clarify what is this Ow-increasing region around the O5 atom, we compared the solvent spatial solvation structure of the Ow site between $\mathbf{RS}_\alpha$ and $\mathbf{RS}_\beta$, as shown in Supporting Information Figure S3. Interestingly, the Ow site of water in $\mathbf{RS}_\alpha$, whose Hw site interacts with the phenyl moiety in the direction of the π orbital of the phenyl group, is closer to the O5 atom than that in $\mathbf{RS}_\beta$, leading to the larger destabilization of the α-anomer.

As a result, these solvation structures around the O1 and O5 atoms more largely stabilizes $\mathbf{RS}_\beta$ than $\mathbf{RS}_\alpha$ but does not compensate the larger stabilizing solvation free energy of $\mathbf{RS}_\alpha$, which arises mainly from the solvation around the O3 atom, as discussed in the main text.

Reason for larger stabilizing solvation of the phenyl moiety of $\mathbf{IS}_\beta$ and $\mathbf{TS}_\beta$

As shown in Table 3 of the main text, the $\Delta\mu^{\mathrm{LRA}}$ energy of the phenyl moiety clearly becomes lower in $\mathbf{IS}_\beta$ and $\mathbf{TS}_\beta$ than in $\mathbf{IS}_\alpha$ and $\mathbf{TS}_\alpha$, respectively. To make clear the reason for this difference in solvation of the phenyl moiety, we investigated the solvation features of the isolated phenoxyl moiety in aqueous phase. To this aim, we optimized the geometry by the 3D-RISM-SCF-DFT method with the B3LYP functional and calculated the solvation structure around the phenoxyl moiety, as shown in Supporting Information Figure S6.

The spatial distribution function of the solvation free energy, $\Delta\mu_{\mathrm{SDF}}(\mathbf{r})$, indicates that the Hw site of water interacts with the phenyl group in the direction of the π orbital of the phenyl group; see Figure S6. The $\Delta\mu^{\mathrm{LRA}}$ energy of the phenyl moiety, however, indicates that the water molecules around the phenyl moiety entirely contributes not to the stabilization but to the destabilization by about 8 kcal/mol. To inspect the position of the water in the first solvation shell, the solvent distribution functions of the Hw and Ow sites along the π orbital of phenyl group are compared. As shown in Figure S7, the Hw site of water in the first solvation shell is distant from the molecular plane of the phenyl group by about 2.0 Å and the Ow site is distant by about 3.0 Å.

Comparing the solute conformation between the α- and the β-anomers in Figure 1 of the main text, the H atom of the C2-H moiety is close to the molecular plane of the phenyl group by about 3 Å in $\mathbf{IS}_\beta$ and $\mathbf{TS}_\beta$. In other words, the H atom of the C2-H moiety of $\mathbf{IS}_\beta$ and $\mathbf{TS}_\beta$ exists near the first solvation shell of water solvent along the π orbital of the phenyl group. Judging from the distance between the H atom of the C2-H moiety and the molecular plane of the phenyl group, the short-range LJ interaction between the H atom of the C2-H moiety and water solvent contributes to the decrease of the destabilizing solvation of the phenyl group by the Ow site of water. Actually, in **IS** and **TS**, the solvent radial distribution function around the molecular center of the phenyl group becomes lower in the first solvation shell of the β-anomer than that of the α-anomer; see Supporting Information Figure S8. As a result, the H atom of the C2-H moiety of $\mathbf{IS}_\beta$ and that of $\mathbf{TS}_\beta$ can prevent the Ow of water solvent from unfavorably approaching the π orbital region of the phenyl group, leading to the larger stabilizing solvation of the β-anomer.

Table S1: Comparison of 3D-RISM-SCF computational cost between single and dual solvent box protocols.[a]

	3D-RISM-SCF[**1n**][b]		3D-RISM-SCF[**2**][c]		3D-RISM-SCF[**1f**][d]	
Box	Single box with normal grid		Dual boxes		Single box with fine grid	
Quality	Low		Similar to [**1f**][e]		High	
	1 CPU	6 CPUs	1 CPU	6 CPUs	1 CPU	6 CPUs
CPU time (minute) in gas phase energy terms						
One-electron Calc.	0.0	0.0	0.0	0.0	0.0	0.0
Two-electron Calc.	0.0	0.0	0.0	0.0	0.0	0.0
Fock Calc.	1.5	0.3	1.5	0.3	1.5	0.3
RHF Calc.	27.0	4.5	27.0	4.5	27.0	4.5
CPU time (minute) in gas phase gradient terms						
One-electron Calc.	0.0	0.0	0.0	0.0	0.0	0.0
Two-electron Calc.	5.5	0.9	5.5	0.9	5.5	0.9
CPU time (minute) in 3D-RISM calculation /(one RISM-cycle)						
ESP Calc.	8.9	1.5	21.5	3.6	71.0	11.7
Hessian Calc. (NR/PC)	12.3	2.4	13.1/17.2[f]	2.7/3.6	185.6	32.5
Solve 3D-RISM (KH)	7.2	1.8	7.3/11.8[f]	1.8/3.0	120.1	27.2
Additional Fock Calc.	9.0	1.5	53.8	9.1	73.3	12.2
CPU time (minute) in additional gradient terms						
Rys term Calc.	89.9	15.0	208.1	39.2	781.1	129.3
PCA term Calc.	2.4	0.5	2.4	0.5	2.4	0.5
Total /(one optimization step)						
CPU time (minute)	796.8	188.1	1564.5	365.1	3744.6	823.4
CPU memory (GB)	2.0	8.1	2.4	10.5	9.6	40.7

[a] $\mathbf{RS}_\alpha$ was calculated by the 3D-RISM-SCF-DFT(B3LYP) with 6-31G(d) basis sets.

[b] With a single box, [0.5 Å,128 points/axis].

[c] With dual boxes, [0.5 Å,128 points/axis] and [0.25 Å,128 points/axis].

[d] With a single box, [0.25 Å,256 points/axis].

[e] The quality of 3D-RISM-SCF[**2**] is close to that of 3D-RISM-SCF[**1f**]; see Table S2.

[f] The former value is for a large box and the latter is for a small box.

Table S2: Dependence of solvation free energy on computational levels of RISM theory[a].

	$\mathbf{RS}_\alpha$	$\mathbf{IS}_\alpha$	$\mathbf{TS}_\alpha$	$\mathbf{PS}_\alpha$	$\mathbf{RS}_\beta$	$\mathbf{IS}_\beta$	$\mathbf{TS}_\beta$	$\mathbf{PS}_\beta$
(A) Solvation free energy, $\Delta\mu$								
3D-RISM-SCF[**2**][b]	-49.8	-48.7	-32.2	-37.4	-46.6	-52.9	-33.7	-31.5
	(3.1)[f]	(4.2)	(20.7)	(15.5)	(6.3)	(0.0)	(19.2)	(21.4)
3D-RISM-SCF[**1f**][c]	-49.8	-48.7	-32.2	-37.4	-46.6	-52.9	-33.7	-31.5
	(3.1)	(4.2)	(20.7)	(15.5)	(6.3)	(0.0)	(19.2)	(21.4)
3D-RISM-SCF[**1n**][d]	-50.1	-49.3	-32.3	-37.6	-47.0	-53.0	-33.7	-32.0
	(2.9)	(3.7)	(20.7)	(15.4)	(6.0)	(0.0)	(19.3)	(21.0)
3D-RISM[**2**](PCA)[b,e]	-56.6	-55.2	-38.7	-43.2	-53.4	-59.1	-39.5	-36.7
	(2.5)	(3.9)	(20.4)	(15.9)	(5.7)	(0.0)	(19.6)	(22.4)
1D-RISM(PCA)[e]	-30.0	-29.5	-14.3	-19.6	-29.0	-32.6	-13.4	-10.1
	(2.6)	(3.1)	(18.3)	(13.0)	(3.6)	(0.0)	(19.2)	(22.5)
(B) Linear response solvation free energy, $\Delta\mu^{\mathrm{LRA}}$								
3D-RISM-SCF[**2**][b]	-108.7	-106.9	-91.3	-96.9	-105.7	-111.7	-93.2	-91.6
	(3.0)	(4.8)	(20.4)	(14.8)	(6.0)	(0.0)	(18.5)	(20.1)
3D-RISM-SCF[**1f**][c]	-108.7	-106.9	-91.3	-96.9	-105.7	-111.7	-93.2	-91.6
	(3.0)	(4.8)	(20.4)	(14.8)	(6.0)	(0.0)	(18.5)	(20.1)
3D-RISM-SCF[**1n**][d]	-109.0	-107.5	-91.6	-97.3	-106.1	-111.7	-93.2	-92.1
	(2.7)	(4.2)	(20.1)	(14.4)	(5.6)	(0.0)	(18.5)	(19.6)
3D-RISM[**2**](PCA)[b,e]	-112.4	-110.6	-95.1	-100.3	-109.5	-114.7	-96.8	-95.1
	(2.3)	(4.1)	(19.6)	(14.4)	(5.2)	(0.0)	(17.9)	(19.6)
1D-RISM(PCA)[e]	-126.1	-124.6	-110.3	-114.5	-124.2	-129.4	-111.0	-107.3
	(3.3)	(4.8)	(19.1)	(14.9)	(5.2)	(0.0)	(18.4)	(22.1)

Unit is kcal/mol.

[a]KH Closure was employed to solve the solute-solvent RISM integral equation.

[b]With dual boxes, [0.5 Å,128 points/axis] and [0.25 Å,128 points/axis].

[c]With a single box, [0.25 Å,256 points/axis].

[d]With a single box, [0.5 Å,128 points/axis].

[e]Calculated with point charges evaluated by 3D-RISM-SCF[**2**]; see also page S8-S9.

[f]In parentheses are differences from $\mathbf{IS}_\beta$.

Table S3: Changes of 1D-RISM solvation free energy components, $\Delta\mu_b^{1D}$, at several important sites b.

site b	**RS**$_\alpha$	**IS**$_\alpha$	**TS**$_\alpha$	**PS**$_\alpha$	**RS**$_\beta$	**IS**$_\beta$	**TS**$_\beta$	**PS**$_\beta$
total	-30.0	-29.5	-14.3	-19.6	-29.0	-32.6	-13.4	-10.0
Sugar[a]	-44.0	-39.6	22.3	36.8	-39.7	-45.8	10.7	46.3
Component								
O5	-35.4	-29.6	-8.9	-24.9	-42.6	-37.9	-19.1	-24.5
O2-H	-23.6	-23.4	-11.4	-15.0	-25.6	**-122.4**[e]	**-90.3**	**-22.1**
O3-H	-22.9	**-110.0**	-13.1	-17.8	-16.1	-26.8	-23.2	-23.3
O4-H	**-105.9**	-26.0	-21.4	-15.2	**-104.4**	-19.3	-13.7	-17.9
O6-H	-24.7	-14.4	**-78.9**	**-24.8**	-24.4	-9.9	-5.6	-3.9
Others[b]	168.5	163.8	156.1	134.5	173.4	170.5	162.5	130.3
Phenoxyl[c]	14.1	10.0	-36.6	-56.4	10.8	13.2	-24.1	-56.4
Component								
O1	-25.6	-30.7	-75.1	-101.1	-28.7	-16.2	-55.4	-97.1
Phenyl[d]	39.7	40.7	38.5	44.8	39.5	29.4	31.3	40.7

Unit is kcal/mol.

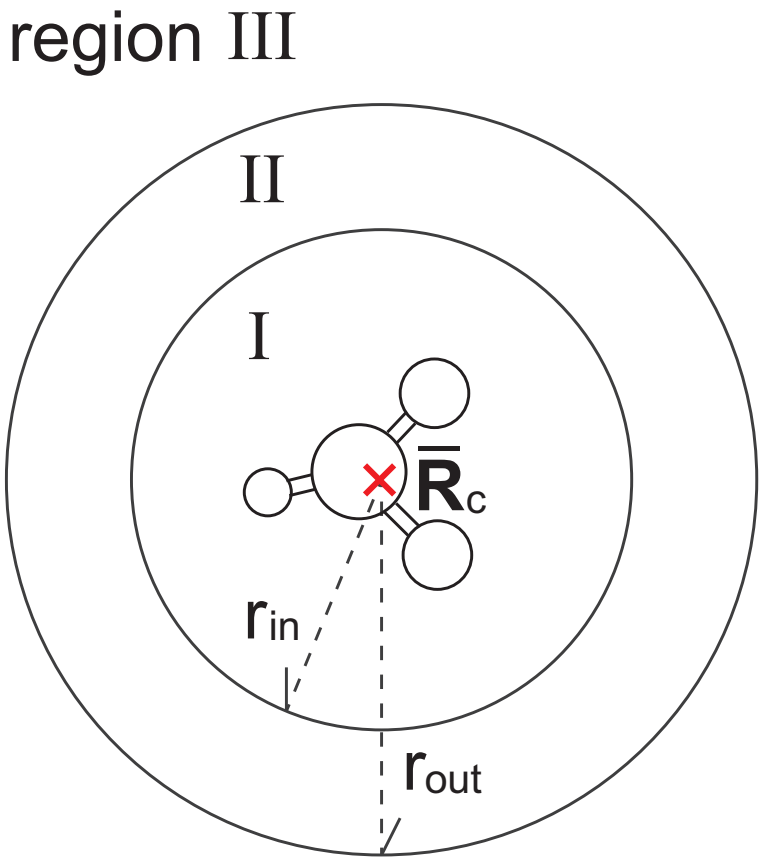
[a] All atoms on the sugar group except for the O1 atom.

[b] All remaining moieties consisting of the sugar ring.

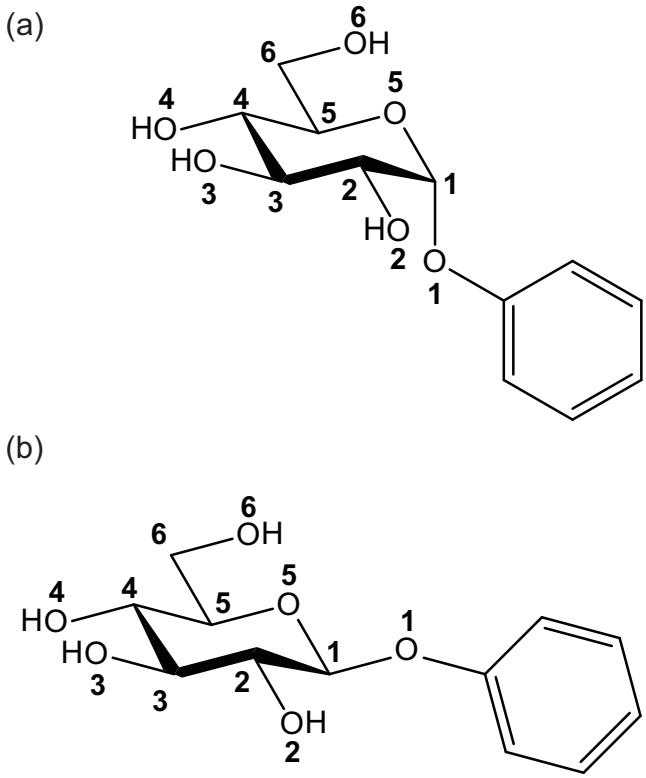
[c] All atoms on the phenoxyl group.

[d] All atoms on the phenyl group.

[e] Bold type represents the oxyanion (or oxygen) without neighboring H.



Scheme S1: Three divided regions around solute. Regions I, Ⅱ and Ⅲ are inner, switching and outer regions, respectively. $\bar{\mathbf{R}}_c$ is the center of solute.



Scheme S2: Structures of phenyl α- and β-D-glucosides. (a) Phenyl α-D-glucoside. (b) Phenyl β-D-glucoside.

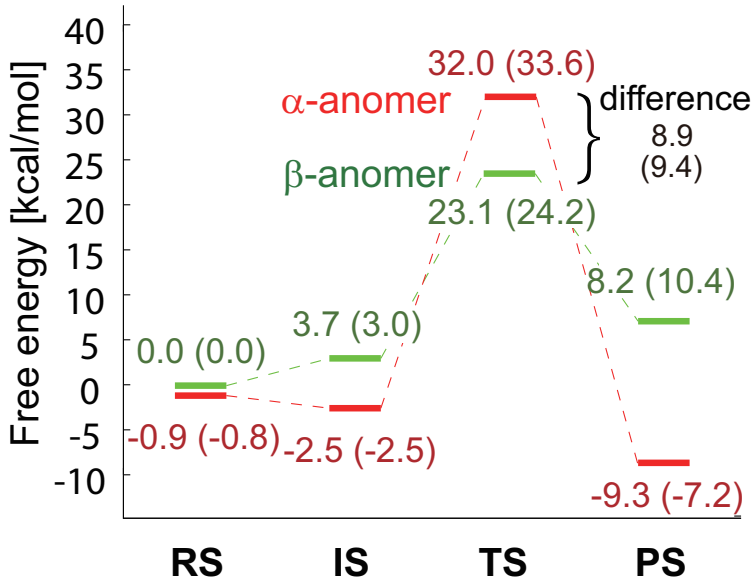
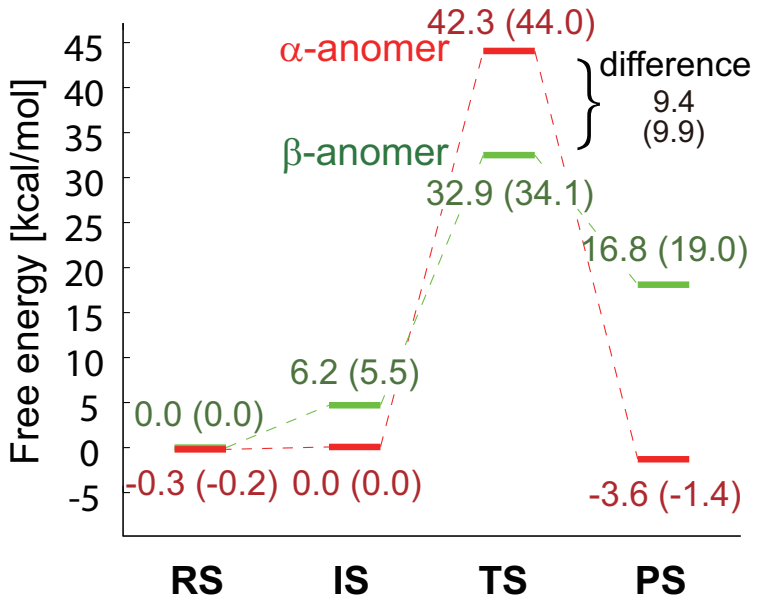


Figure S1: PCM/MP2-calculated free energy profiles of rate determining step of levoglucosan formation at 373.15 K. Red and green lines represent phenyl α- and β-D-glucoside, respectively. In parentheses are free energy changes without zero point energy, thermal energy, and entropy.

(a) 3D-RISM-SCF-RHF/SCS-MP2-calculated free energy in aqueous phase.



(b) SCS-MP2-calculated free energy in gas phase.

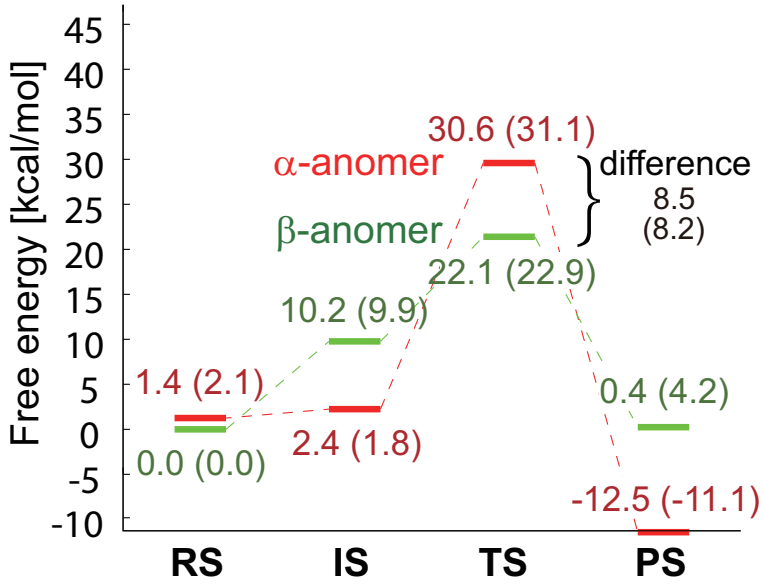


Figure S2: Free energy profiles of rate determining step of levoglucosan formation reaction in aqueous and gas phases at 373.15 K. Red and green lines represent phenyl α- and β-D-glucosides, respectively. In parentheses are free energy changes without zero point energy, thermal energy, and entropy. Calculated with (a) the 3D-RISM-SCF-DFT/B3LYP-optimized geometry and (b) the gas phase geometry.

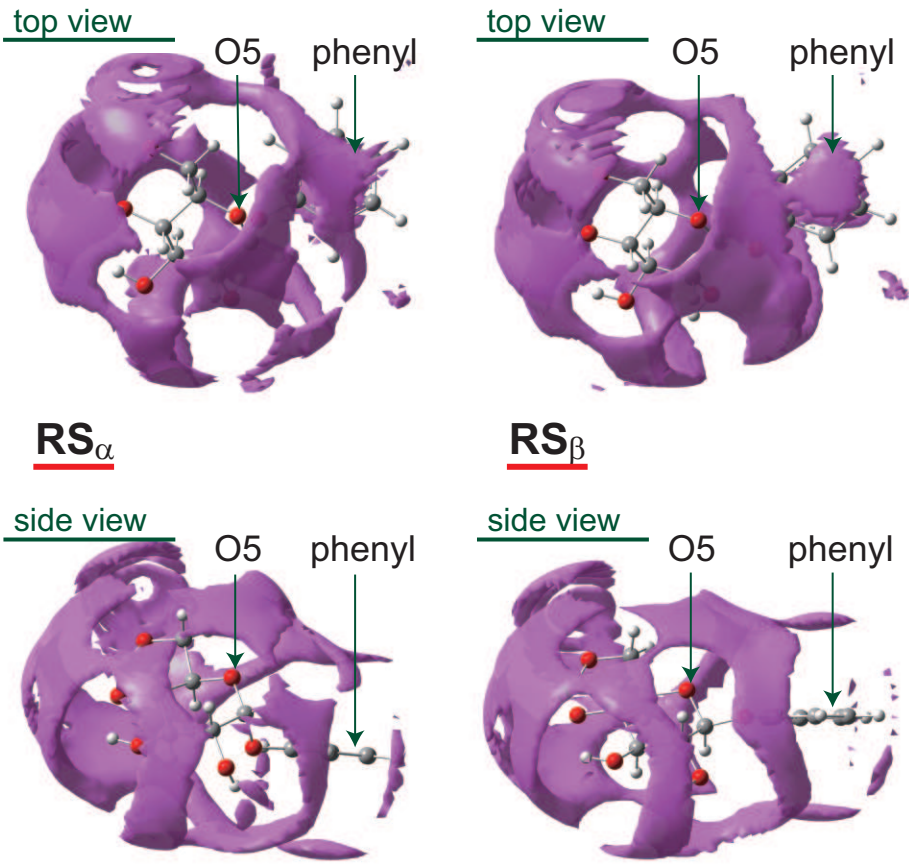


Figure S3: Spatial distribution function of the solvent Ow site in $\mathbf{RS}_\alpha$ and $\mathbf{RS}_\beta$. Left is $\mathbf{RS}_\alpha$. Right is $\mathbf{RS}_\beta$. Surface is plotted at $g_{\mathrm{Ow}}(\mathbf{r}) = 2.5$.

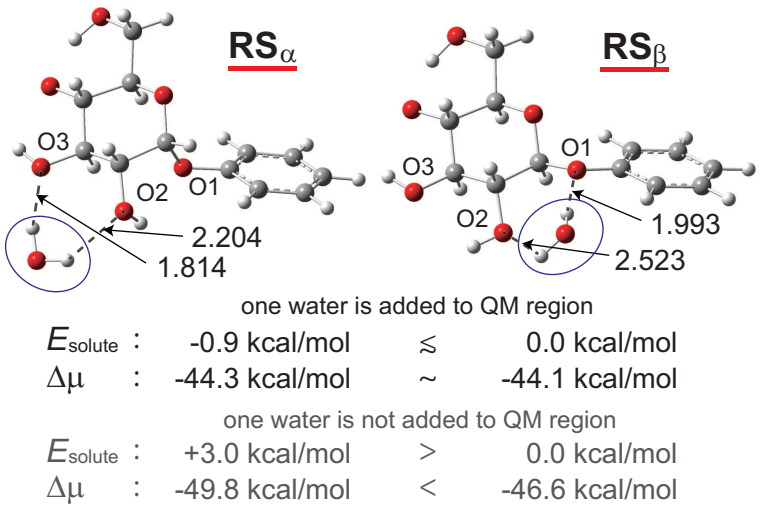


Figure S4: Hydrogen bonding structures and 3D-RISM-SCF-RHF/MP2-calculated free energy components of bridging water. Left is phenyl α-D-glucoside. Right is phenyl β-D-glucoside. Geometry of water was optimized by the 3D-RISM-SCF-DFT/B3LYP method with fixed geometry of phenyl α- and β-D-glucosides optimized in Figure 3. The E_{solute} of $\mathbf{RS}_{\beta}$ is taken as standard. Unit is Å.

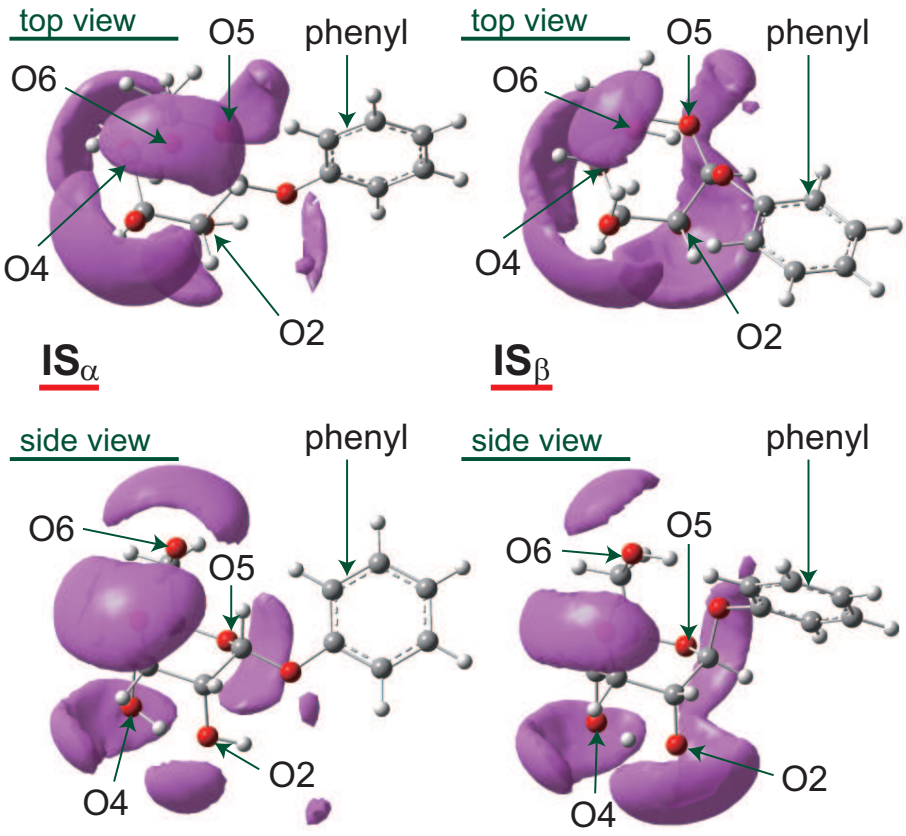
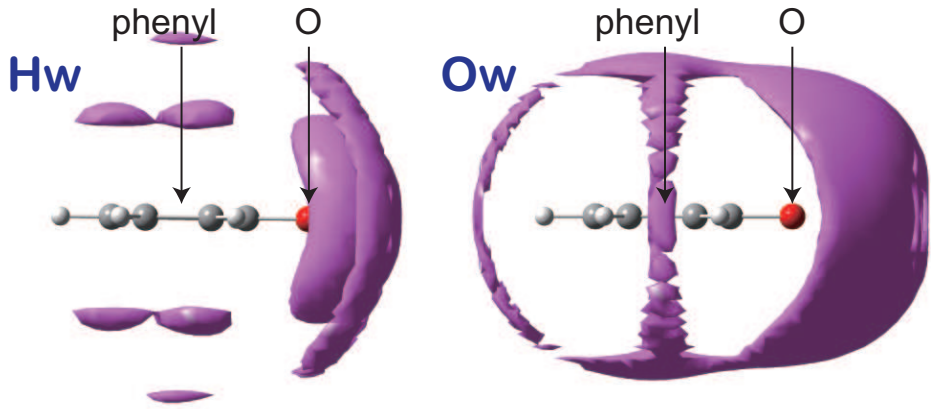


Figure S5: Spatial distribution function of the solvent Hw site in $\mathbf{IS}_\alpha$ and $\mathbf{IS}_\beta$. Left is $\mathbf{IS}_\alpha$. Right is $\mathbf{IS}_\beta$. Surface is plotted at $g_{\mathrm{Hw}}(\mathbf{r}) = 2.0$.

(a) Solvent spatial distribution function, $g_{\xi}(\mathbf{r})$.



(b) Spatial distribution of linear response solvation free energy, $\Delta\mu_{\mathrm{SDF}}(\mathbf{r})$.

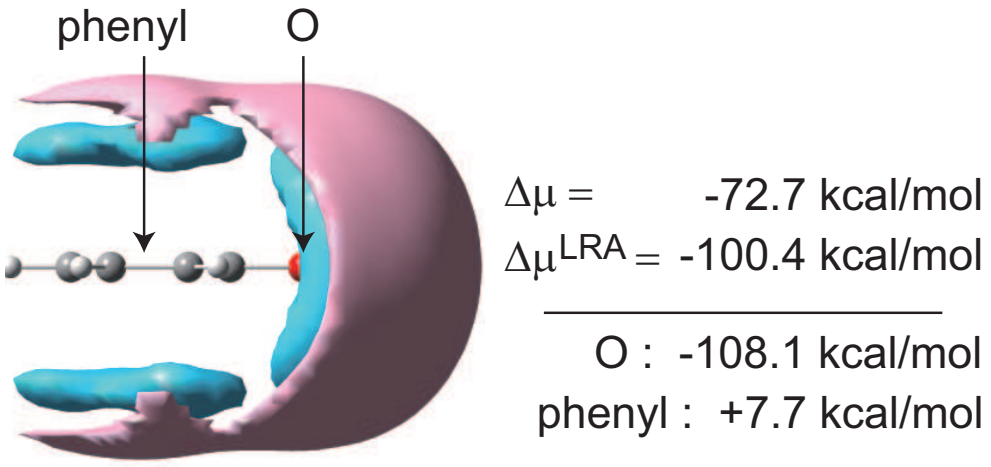


Figure S6: Solvation of the isolated phenoxyl moiety in aqueous phase. (a) Left is $g_{\mathrm{Hw}}(\mathbf{r})$=1.5. Right is $g_{\mathrm{Ow}}(\mathbf{r})$=2.4. (b) Blue and purple surfaces correspond to the stabilization and destabilization regions, respectively. Surface is plotted for $|\,\Delta\mu_{\mathrm{SDF}}(\mathbf{r})\,| \geq 1.0$ kcal/(mol·Å^3).

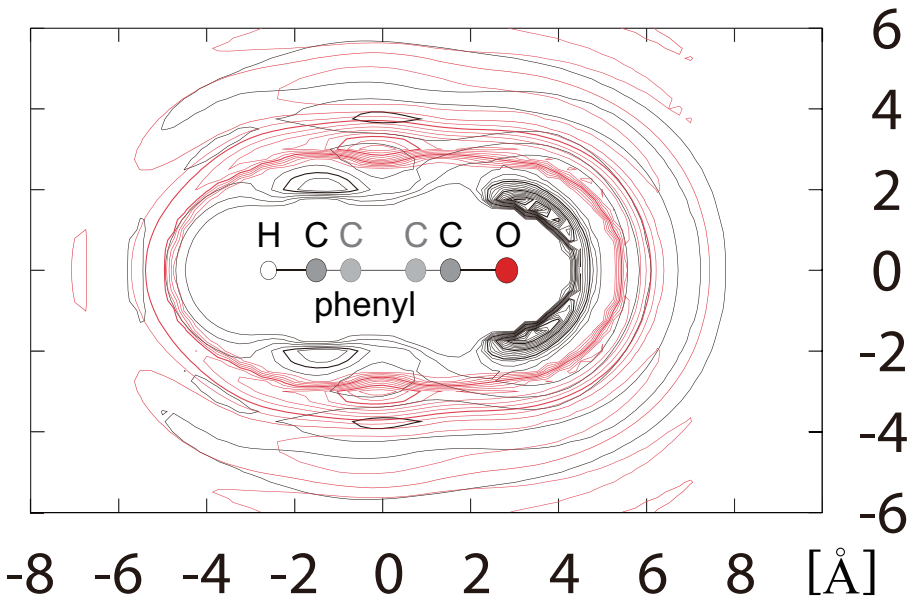


Figure S7: Solvation of the isolated phenoxyl moiety along its π orbital direction. Black and red lines represent the Hw and Ow sites of water, respectively. Spacing of contour lines is 0.3.

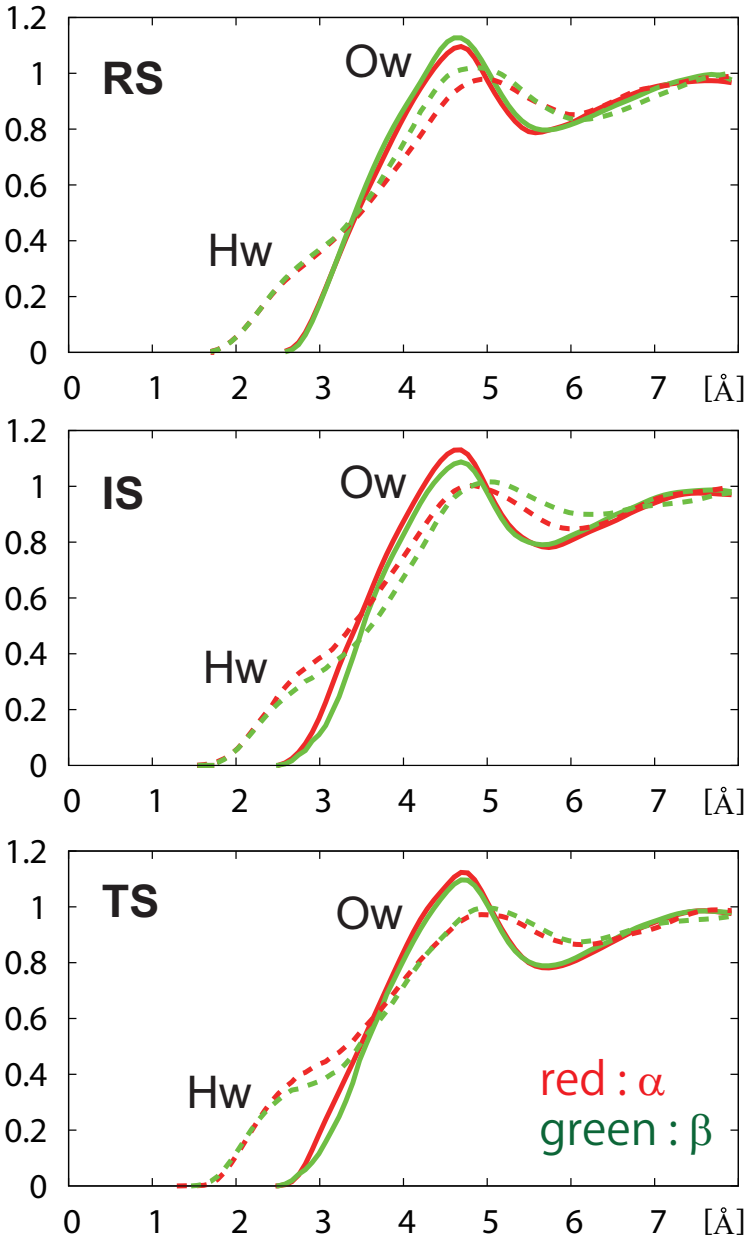


Figure S8: Solvent radial distribution function around the molecular center of the phenyl group in RS, IS, and TS. Solid and dashed lines represent the Ow and Hw sites of water, respectively. Red and green lines represent phenyl α- and β-D-glucosides, respectively.