DOCTRAL THESIS

Computational methods for configurational entropy using internal and Cartesian coordinates

Simon Hikiri

Yokohama City University
Graduate School of Medical Life Science
Department of Medical Life Science

2016
Computational methods for configurational entropy using internal and Cartesian coordinates

Simon Hikiri

Yokohama City University
Graduate School of Medical Life Science
Department of Medical Life Science

2016

Members of the Judging Committee:
Professor Hideo Takahashi (chair)
Professor Akinori Kidera
Professor Mamoru Sato
Professor Kentaro Tomii
Computational methods for configurational entropy using internal and Cartesian coordinates

Simon Hikiri,† Takashi Yoshidome,†‡ and Mitsunori Ikehuchi*†

†Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suchiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
‡Present address: Department of Applied Physics, Graduate School of Engineering, Tohoku University, 6-6-05, Aoba, Aramaki, Aoba-ku, Sendai 980-8579, Japan
E-mail: ike@tsurumi.yokohama-cu.ac.jp

Abstract

The configurational entropy of solute molecules is a crucially important quantity to study various biophysical processes. Consequently, it is necessary to establish an efficient quantitative computational method to calculate configurational entropy as accurately as possible. In the present paper, we investigate the quantitative performance of the quasi-harmonic and related computational methods, including widely used methods implemented in popular molecular dynamics (MD) software packages, compared with the Clausius method, which is capable of accurately computing the change of the configurational entropy upon temperature change. Notably, we focused on the choice of the coordinate systems (i.e., internal or Cartesian coordinates). The Boltzmann-quasi-harmonic (BQH) method using internal coordinates outperformed all the six methods examined here. The introduction of improper torsions in the BQH method improves its performance, and anharmonicity of proper torsions in proteins is identified to be the origin of the superior performance of the BQH method. In contrast, widely used methods implemented in MD packages show rather poor performance. In addition, the
enhanced sampling of replica-exchange MD simulations was found to be efficient for the convergent behavior of entropy calculations. Also in folding/unfolding transitions of a small protein, Chignolin, the BQH method was reasonably accurate. However, the independent term without the correlation term in the BQH method was most accurate for the folding entropy among the methods considered in this study, because the QH approximation of the correlation term in the BQH method was no longer valid for the divergent unfolded structures.
Introduction

The entropy of solute molecules, $S_{\text{Solute}}$, is an important factor to consider when studying a variety of biophysical processes such as protein folding and ligand binding. In theoretical studies of biological processes, the free energy is often decomposed into the solute energy, entropy and solvation free energy. The calculation of the solute energy is straightforward by way of standard force fields, and various efficient methods to calculate solvation free energies for biological molecules have been proposed. For the solute entropy, standard computational methods implemented in software packages such as AMBER, CHARMM and GROMACS have been routinely applied, but these methods should be assessed in terms of their accuracy. Because the importance of the solute entropy has been highlighted recently in the molecular recognition of proteins, a highly quantitative computational method to describe the solute entropy accurately is required. To this end, our concern in the present paper is computational methods using molecular dynamics (MD) simulations.

$S_{\text{Solute}}$ for a solute molecule composed of $N$ atoms is expressed using the Boltzmann-Shannon equation:

$$S_{\text{Solute}} = -k_B \int \rho(P^{3N}, Q^{3N}) \ln \left( (2\pi \bar{h})^{3N} \rho(P^{3N}, Q^{3N}) \right) dP^{3N} dQ^{3N},$$

where $k_B$ and $\bar{h}$ are the Boltzmann constant and the reduced Planck constant ($\bar{h} = h/2\pi$), respectively, and $\rho(P^{3N}, Q^{3N})$ is the probability distribution function (PDF) of Cartesian coordinates ($Q^{3N}$) and conjugate momenta ($P^{3N}$). Although $S_{\text{Solute}}$ is, in principle, calculated from $\rho(P^{3N}, Q^{3N})$, the computation of $\rho(P^{3N}, Q^{3N})$ from MD trajectories is quite difficult because of the massive $6N$-dimensional phase space. Therefore, several methods to calculate $S_{\text{Solute}}$ approximately have been proposed so far.

An earlier proposed method is based on normal mode analysis (NMA). In this method, $S_{\text{Solute}}$ is expressed as a sum of the entropies of harmonic oscillators with NMA-derived frequencies except for the six zero frequencies, which correspond to the translational and
rotational degrees of freedom of the solute molecule. The configurational entropy, $S_{\text{Solute Conf}}$, is defined by the subtraction of the translational and rotational entropies of a solute molecule from $S_{\text{Solute}}$. It should be noted that the configurational entropy in this paper includes both spatial and momentum contributions. In NMA, the configurational entropy corresponds to the entropy associated to the intra-molecular vibrations of a molecule.

In NMA, a harmonic potential energy surface is assumed in the vicinity of a single local minimum. However, in intricate molecules such as proteins, multiple local minima of the potential energy surface exist, even in the equilibrium dynamics at room temperature. Therefore, the method based on NMA would underestimate $S_{\text{Solute Conf}}$ because it neglects the transitions between the local minima.

To overcome this limitation of NMA, the quasi-harmonic (QH) method has been proposed. In the QH method, MD simulations, in which structures visit multiple potential minima, are employed for conformational sampling and configurational entropies are estimated using the variance-covariance matrix calculated from the MD trajectories. Because MD simulations are widely utilized for analysis of protein dynamics, the QH method is the current standard to calculate configurational entropies. In most of the MD packages such as CHARMM, AMBER and GROMACS, the QH method is implemented for the calculation of configurational entropies. The formulation of the configurational entropy in the QH method is identical to NMA except for the calculation of modes and frequencies. In the QH method, the modes are obtained through principal component analysis (PCA) of the mass-weighted variance-covariance matrix from the MD trajectories after the overall translations and rotations of the solute molecules are removed by using least-square fits of snapshots of the trajectories. The MD program packages employ the entropy formulations for quantum harmonic oscillators, though MD is performed on the basis of classical mechanics.

Instead of Cartesian coordinates used in most of the current MD packages, internal coordinates, which represent solute conformations using a set of bond lengths, angles, and torsion angles, were utilized to calculate configurational entropy. Historically, the proposal
of the QH method using internal coordinates\textsuperscript{22} preceded that of Cartesian coordinates. In the internal coordinate system, the coordinates representing the overall translations and rotations are naturally separated from the coordinates that describe conformations of a molecule. Therefore, the degrees of freedom of a solute molecule are reduced to $3N - 6$ in an internal coordinate system. In the QH method, the PDF of internal coordinates is approximated to be a multivariate Gaussian distribution. Using this approximation, the configurational entropy is calculated through the determinant of the variance-covariance matrix of internal coordinates, with correlations between the coordinates up to second order considered. The limitation of the QH method is that the PDF is assumed to be a unimodal Gaussian distribution. However, torsional angle distributions are frequently multimodal, corresponding to local potential minima such as trans and gauche conformations.

Di Nola \textit{et al.} proposed an improvement of the QH method, referred to as the Boltzmann quasi-harmonic (BQH) method.\textsuperscript{27} In the BQH method, the authors focused on the diagonal term factored out from the variance-covariance matrix, leading to a decomposition of the QH entropy into the diagonal and correlation terms. Then, Di Nola \textit{et al.} proposed to replace the QH formulation of entropies for the diagonal term with the sum of the exact Boltzmann-Shannon equation for each coordinate. In the BQH method, non-Gaussian PDFs of torsion angles are treated adequately and the correlation term remains identical to that used in the QH method.

Beyond the QH and related methods, the nonparametric entropy methods, which estimate the configurational entropy without the analytical approximation to PDF such as the QH approximation, have been proposed.\textsuperscript{28,29} The category of the nonparametric methods includes mining minima,\textsuperscript{30} nearest neighbor,\textsuperscript{31} mutual information expansion,\textsuperscript{32} maximum information spanning trees\textsuperscript{33} and the hypothetical scanning molecular dynamics.\textsuperscript{34} Although these methods have provided valuable insights into the configurational entropy of small systems, the computational cost becomes problematic for large systems such as proteins.\textsuperscript{28} Therefore, the hybrid method, in which the QH approximation and the nonparametric method are
combined, could be a balanced choice between accuracy and the computational cost. The BQH method is one of the hybrid methods.

Because many computational methods for configurational entropy have been proposed, an assessment of their quantitative performances is required to establish an accurate procedure to calculate configurational entropies. Recently, Harpole and Sharp proposed an assessment method that is based on the Clausius equation. The Clausius equation gives the entropy change of a system as a function of temperature change. When a temperature-independent implicit solvent model is employed, the entropy change of a system becomes the same as the configurational entropy of the solute molecule. Therefore, the Clausius method is capable of calculating the configurational-entropy change as the temperature changes without ignoring the high-order correlations and anharmonicity of PDF. The Clausius method assumes that the heat capacity is independent of temperature in the target range of temperature. However, even for the system with the temperature-dependent heat capacity, a sufficient small step of a temperature change in the Clausius method allows us to accurately estimate the entropy change. Harpole and Sharp tested the quantitative performance of the QH and BQH methods and found that the BQH method is superior to the QH method. However, the QH methods using Cartesian coordinates implemented in most MD packages have yet to be assessed. It should be noted that because they focused on the spatial contribution of the configurational entropy, only the potential energies were used and the kinetic energies were ignored for the calculations using the Clausius method.

Based on the above background, the present paper addresses five outstanding issues. One of five issues is which coordinate system provides better quantitative performance. Most MD packages employ the QH method using Cartesian coordinates for entropy calculations. In free-energy analysis such as MM-PBSA and MM-GBSA, the entropy is routinely estimated using the QH method with Cartesian coordinates combination. Chang et al. pointed out that the QH method with Cartesian coordinates introduces substantial errors of several kilocalories per mole. However, they used the entropy formulation based on the classical
statistical mechanics, which differs from the entropy formulations for quantum harmonic oscillators employed in most MD packages. Therefore, a test of the quantitative performance of these methods is inevitable. To this end, the Clausius method proposed by Harpole and Sharp is extended in the current study to include both the potential and kinetic energies because the kinetic and potential components are inseparable in the entropy of quantum harmonic oscillators used in MD packages.

The second issue addressed in this paper is the treatment of torsion angles. In the BQH method, the treatment of torsion angles with multimodal PDFs was improved. However, treatment of correlation between coordinates in the BQH method is identical to the QH method. That is, correlations between coordinates are included up to the second order. However, three-body correlations between torsion angles are found frequently in protein dynamics. To address this issue, we consider the rotation of a methyl group as an example (Figure 1). Because the three hydrogen atoms rotate simultaneously, the three torsion angles (referred to as proper torsions) that describe the positions of the three hydrogen atoms are strongly correlated (Figure 1 (a)). Because the QH and BQH methods include correlations up to the second order, ignoring the three-body correlation would introduce a systematic error. To overcome this problem, we introduce improper torsions in the BQH method. Using improper torsions, the three hydrogen atoms in a methyl group are described by a proper torsion and two improper torsions. The rotation of the methyl group is represented by a single proper torsion, and the distortion of the methyl group from the standard structure is described by improper torsions (Figure 1 (b)). Consequently, the effect of the three-body correlation is eliminated by introducing improper torsions, leading to a possible improvement of the quantitative performance of the BQH method.

The third issue to be addressed in this paper is the convergence of the entropy calculation. To attain convergence, sufficient sampling of conformations is required in MD simulations. However, in conventional MD simulations, structures are frequently trapped in local potential minima, and this might be an obstacle for a strict test of the computational
Figure 1: Two definitions of the three torsions for the hydrogen atoms in a methyl group. In panel (a), the positions of three hydrogen atoms are determined using proper torsions only. Proper torsions 1–3 are defined as C\textsubscript{\alpha}–C\textsubscript{\beta}–C\textsubscript{\gamma}–H\textsubscript{1}, C\textsubscript{\alpha}–C\textsubscript{\beta}–C\textsubscript{\gamma}–H\textsubscript{2} and C\textsubscript{\alpha}–C\textsubscript{\beta}–C\textsubscript{\gamma}–H\textsubscript{3}. In panel (b), two proper torsions are replaced by two improper torsions. Proper torsion 1 is the same as in panel (a). Improper torsions 1 and 2 are defined as the dihedral angles between the red and green planes.

methods for the calculation of reliable configurational entropies. To overcome this problem, we employ temperature replica exchange MD (REMD) simulations.\textsuperscript{40,41} In such simulations, the structures do not become trapped in local potential minima because they can proceed through their high-temperature structures. We compare the convergent behavior of entropy calculations between REMD and conventional MD simulations.

The fourth issue is a test of the entropy calculation for isothermal processes, which are relevant to various biomolecular phenomena, e.g. protein folding and ligand binding. We performed folding/unfolding simulations for a small protein, Chignolin, using the REMD method. Then, we evaluated the entropy changes upon isothermal folding transitions. In this case, the entropy changes upon folding were compared with the temperature derivatives of the free energy changes. This is a severe test because of a diversity of structures in the unfolded state.

The fifth issue is the contribution of the correlation term in the configurational entropy. Li and Br"ushweiler showed that the independent term (the diagonal term or one-dimensional term) dominantly contributed to the temperature-dependent configurational entropy in the
alanine dipeptide.\textsuperscript{42} For a small protein, Villin Head Piece, the independent term was also dominant, although the complete convergence was not achieved using the conventional MD. However, Killian \textit{et al.} pointed out that the contribution of the correlation could not be ignored for ligand binding.\textsuperscript{43} Also, in the BQH calculation, the correlation contributions were not so small.\textsuperscript{35} Here, using the enhanced sampling of REMD, the contribution of the correlation term to the configurational-entropy change upon temperature change was investigated for molecules of various sizes including proteins up to sixty residues. Also, we discussed the contribution of the correlation term in the configurational entropy for the folding transition of Chignolin.

\section*{Theory and Methods}

\subsection*{Formulation of configurational entropy}

In the present paper, we focus on the configurational entropy, $S_{\text{Conf}}^{\text{Solute}}$. First, we formulate the configurational entropy using Cartesian coordinates used in most current MD packages, and we then describe the formulation using internal coordinates. Finally, for an assessment of these methods, the Clausius method is explained.

\subsection*{The QH method using Cartesian coordinates}

In the QH method using Cartesian coordinates, the mass-weighted variance-covariance matrix is first calculated from MD trajectories in which the overall translations and rotations of the solute molecule are removed by using least-square fits of mass-weighted coordinates. Then, the matrix is diagonalized to obtain eigenvalues $\lambda_i$, from which the frequencies of the modes are evaluated as $\omega_i = \sqrt{k_B T / \lambda_i}$. Six eigenvalues are zero, which correspond to the overall translational and rotational degrees of freedom of the molecule.
In the case of classical harmonic oscillators, \( S_{\text{Conf}}^{\text{Solute}} \) is expressed as:

\[
S_{\text{Conf}}^{\text{Solute}} \approx \sum_{i} \frac{3N-6}{2} k_B \ln \left( \frac{e^2}{\alpha_i^2} \right),
\]

\[\alpha_i = \frac{\hbar \omega_i}{k_B T}.\]  

Hereafter, the expression of eq. (2) is referred to as \( S_{\text{CC}}^{\text{CC}} \). In \( S_{\text{CC}}^{\text{CC}} \), both spatial and momentum contributions to the configurational entropy are included.\(^{23,24}\) Eq. (2) can be derived from the harmonic potential, Boltzmann distribution and Boltzmann-Shannon entropy eq. (1).

In MD simulation packages, quantum harmonic oscillators are often used instead of classical harmonic oscillators despite the fact that MD simulations are performed on the basis of classical mechanics. In the AMBER software package, the expression for the entropy of quantum harmonic oscillators\(^{24}\) is employed as:

\[
S_{\text{Conf}}^{\text{Solute}} \approx \sum_{i} \frac{3N-6}{2} k_B \left[ \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln \left( 1 - e^{-\alpha_i} \right) \right].
\]

Hereafter, we refer to this expression as \( S_{\text{QM}}^{\text{CC}} \).

In the GROMACS and CHARMM program packages, an approximation proposed by Schlitter\(^{25}\) is used:

\[
S_{\text{Conf}}^{\text{Solute}} \approx \sum_{i} \frac{3N-6}{2} k_B \ln \left( 1 + \frac{e^2}{\alpha_i^2} \right).
\]

Hereafter, we refer to this approximation as \( S_{\text{QMS}}^{\text{CC}} \).

In the CHARMM program package, and in addition to eq. (4), the approximation proposed by Andricioaei and Karplus\(^{23}\) is also implemented:

\[
S_{\text{Conf}}^{\text{Solute}} \approx \sum_{i} k_B \left[ 1 + \frac{1}{2} \ln \left( \frac{1}{\alpha_i^2} + \frac{1}{12} \right) \right].
\]
This expression is referred to as $S_{QMK}^{CC}$.

The QH and BQH methods using internal coordinates

Instead of Cartesian coordinates, internal coordinates can also be applied to the QH method. Here, we employ classical statistical mechanics. In classical statistical mechanics, $\rho(P^{3N}, Q^{3N})$ is expressed as the multiplication of the momentum ($\rho(P^{3N})$) and configurational ($\rho(Q^{3N})$) PDFs. Thus, $S_{\text{Solute}}$ is decomposed into the corresponding components, $S_{\text{Solute}}^P$ and $S_{\text{Solute}}^Q$, respectively. Because $S_{\text{Solute}}^P$ can be calculated analytically, we focus on $S_{\text{Solute}}^Q$. We first perform the transformation of the integration variables in $S_{\text{Solute}}^Q$ from Cartesian coordinates $\{Q^{3N}\}$ into internal coordinates $\{q^{3N-6}\}$ as:

$$S_{\text{Solute}}^Q = -k_B \int dQ^{3N} \rho(Q^{3N}) \ln \rho(Q^{3N})$$

$$= -k_B \int d\tilde{q}^{3N-6} \tilde{\rho}(\tilde{q}^{3N-6}) \ln \tilde{\rho}(\tilde{q}^{3N-6}) + k_B \ln 8\pi^2 V$$

$$+ k_B \int d\tilde{q}^{3N-6} \tilde{\rho}(\tilde{q}^{3N-6}) \ln J(\tilde{q}^{3N-6}),$$

(6)

where $V$ is the volume of the system, $J(\tilde{q}^{3N-6})$ is the Jacobian whose functional form is described in ref. 44, and $\tilde{\rho}(\tilde{q}^{3N-6}) \equiv 8\pi^2 V J(\tilde{q}^{3N-6}) \rho(Q^{3N})$. The second term arises from the translation and rotation of a solute molecule. Hereafter, we denote the first term of eq. (6) as $S_{\text{Conf},q}^{\text{Solute}}$.

In the QH method, $\tilde{\rho}(\tilde{q}^{3N-6})$ is assumed to be the multivariate Gaussian distribution $\tilde{\rho}_G(\tilde{q}^{3N-6})$.

$$\tilde{\rho}_G(\tilde{q}^{3N-6}) = \frac{1}{(\sqrt{2\pi})^{3N-6}|\sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\tilde{q}^{3N-6} - \langle q^{3N-6} \rangle)^T \sigma^{-1} (\tilde{q}^{3N-6} - \langle q^{3N-6} \rangle) \right]$$

(7)

Here, $\langle q^{3N-6} \rangle$ is the ensemble average of $q^{3N-6}$ and $\sigma$ is the variance-covariance matrix in which each element is given by $\sigma_{ij} \equiv \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle$, where $q_i$ is the $i$-th internal
variable \((i=1, 2, 3, \cdots, 3N-6)\). Using eq. (7), \(S_{\text{Conf},q}^{\text{Solute,First}}\) becomes:

\[
S_{\text{Conf},q}^{\text{Solute,First}} \approx \frac{1}{2} k_B \sum_i^{3N-6} \ln[3N-6] |\sigma_i| \\
= \frac{1}{2} k_B \sum_i^{3N-6} \ln[2\pi e \sigma_{ii}] + \frac{1}{2} k_B \ln|C|, \tag{8}
\]

where \(C\) is a matrix in which each element is the correlation coefficient, \(C_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii} \sigma_{jj}}\). Hereafter, we refer to the first term as the independent term because it is the component without all the correlations. Because correlations between the variables are involved in the second term, it is referred to as the correlation term.

In the BQH method,\(^{27} \frac{k_B}{2} \ln[2\pi e \sigma_{ii}]\) in the independent term of eq. (8) is replaced by the exact Boltzmann entropy for the \(i\)-th variable, \(q_i\):

\[
\frac{1}{2} k_B \sum_i^{3N-6} \ln[2\pi e \sigma_{ii}] \rightarrow -k_B \sum_i^{3N-6} \int dq_i \tilde{\rho}(q_i) \ln \tilde{\rho}(q_i). \tag{9}
\]

As for \(S_P^{\text{Solute}}\), we employ the following expression:

\[
S_P^{\text{Solute}} \equiv \frac{3N-6}{2} k_B \ln T + \text{constant}, \tag{10}
\]

where the \textit{constant} is independent of \(T\), and the explicit expression is given in any standard statistical mechanics textbook, for example, eq. (5–19) on p. 86 of ref 24. The expression is identical to that derived using Cartesian coordinates because of the equipartition theorem.\(^{45}\) In the expression, the contribution of the translation and rotation of a solute molecule is not taken into consideration.

In summary, the expressions of the configurational entropy formulated using the internal
coordinate are:

\[S_{\text{Conf}}^{\text{Solute}} = S_{P}^{\text{Solute}} + S_{\text{Conf},Q}^{\text{Solute}},\]

\[S_{\text{Conf},Q}^{\text{Solute}} \approx S_{\text{Ind}} + \frac{1}{2} k_B \ln |C| + k_B \int dq^{3N-6} \rho(q^{3N-6}) \ln J(q^{3N-6}),\]  

(11)

\[S_{\text{Ind}} \equiv \left\{ \begin{array}{ll}
\frac{1}{2} k_B \sum_{i} \ln \left[ 2 \pi \epsilon \sigma_{ii} \right] & (\text{QH}), \\
-k_B \sum_{i} \int dq_i \tilde{\rho}(q_i) \ln \tilde{\rho}(q_i) & (\text{BQH}).
\end{array} \right.\]  

(12)

It is noted that the contributions of the translation and rotation of a solute molecule are not taken into account. We denote \(S_{\text{Conf}}^{\text{Solute}}\) calculated through the QH and BQH methods as \(S_{\text{QH}}^{\text{IC}}\) and \(S_{\text{BQH}}^{\text{IC}}\), respectively.

**Clausius method**

To test the quantitative performance of the six expressions of \(S_{\text{Conf}}^{\text{Solute}}\) given above, we employ the Clausius method proposed by Harpole and Sharp.\(^{35}\) It is based on the Clausius equation in which the change of the entropy in a system upon temperature change, \(\Delta S_{\text{Total}}\), is:

\[\Delta S_{\text{Total}} = \int_{T_0}^{T_1} \frac{C}{T} dT,\]  

(13)

where \(C\) is the heat capacity of the system and \(T_0\) and \(T_1\) (\(> T_0\)) are the initial and final temperatures, respectively. When temperature-independent implicit solvent models such as the generalized Born (GB) model with fixed dielectric constants are employed, \(\Delta S_{\text{Total}}\) becomes identical to the change of \(S_{\text{Solute}}^{\text{IC}}\) (eq. (1)) upon temperature change, \(\Delta S_{\text{Solute}}^{\text{IC}}\).
Because $C$ can be constant when $T_1 - T_0$ is small, $\Delta S^{\text{Solute}}$ is given by:

$$
\Delta S^{\text{Solute}} = C \ln \left( \frac{T_1}{T_0} \right)
= \frac{\Delta E}{T_1 - T_0} \ln \left( \frac{T_1}{T_0} \right)
= \frac{\Delta E_{\text{Kinetic}} + \Delta E_{\text{Potential}}}{T_1 - T_0} \ln \left( \frac{T_1}{T_0} \right),
$$

(14)

where $\Delta E_{\text{Kinetic}}$ and $\Delta E_{\text{Potential}}$ are the changes of the kinetic and potential energies of a solute molecule upon temperature change, respectively, and $\Delta E$ represents their sum. It is noted that while Harpole and Sharp considered only $\Delta E_{\text{Potential}}$, we consider both $\Delta E_{\text{Kinetic}}$ and $\Delta E_{\text{Potential}}$.

By removing the contribution of the translation and rotation of a solute molecule from eq. (14), the expression becomes:

$$
\Delta S^{\text{Solute}}_{\text{Conf}} = \left( 3N - \frac{6}{2} k_B + \frac{\Delta E_{\text{Potential}}}{T_1 - T_0} \right) \ln \left( \frac{T_1}{T_0} \right).
$$

(15)

It is noted that the contribution of the potential energy arising from the translational and rotational energies of the solute molecule is zero. Because approximations other than $C = \text{constant}$ are not involved in eq. (15), we can employ eq. (15) to test the quantitative performance of $\Delta S^{\text{Solute}}_{\text{Conf}}$ calculated using the six expressions above. Hereafter, the expression of eq. (15) is referred to as $\Delta S_{\text{Clausius}}$.

**Computational details**

**REMD simulations**

REMD\textsuperscript{40,41} simulations were performed using AMBER12\textsuperscript{14} to calculate $\Delta S^{\text{Solute}}_{\text{Conf}}$ for the small molecules butane, cyclohexane, alanine dipeptide and oseltamivir and for the proteins trp cage\textsuperscript{46} and protein A\textsuperscript{47} (Table 1). All simulations were performed under the $NVT$ ensemble using the Langevin thermostat and a collision frequency $\gamma$ of 5.0 ps$^{-1}$. The time step was
set at 0.5 fs because we applied no constraints such as hydrogen SHAKE\textsuperscript{48} when performing the simulations. The simulation times for all molecules were 200 ns per replica and replica exchanges were attempted every 10 ps. Solute conformations were sampled every 1 ps. We used the GB model\textsuperscript{10} for the solvent model. The force fields, the number of replicas and the ranges of the temperatures for solutes are listed in Table 1.

<table>
<thead>
<tr>
<th>Molecules</th>
<th>PDB ID</th>
<th># of atoms</th>
<th>Force field</th>
<th># of replicas</th>
<th>Temperature range [K]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Butane</td>
<td>na</td>
<td>14</td>
<td>GAFF</td>
<td>32</td>
<td>260 – 600</td>
</tr>
<tr>
<td>Cyclohexane</td>
<td>na</td>
<td>18</td>
<td>GAFF</td>
<td>32</td>
<td>260 – 600</td>
</tr>
<tr>
<td>Alanine dipeptide</td>
<td>na</td>
<td>22</td>
<td>AMBER99SB</td>
<td>32</td>
<td>260 – 600</td>
</tr>
<tr>
<td>Oseltamivir</td>
<td>na</td>
<td>44</td>
<td>GAFF</td>
<td>32</td>
<td>260 – 600</td>
</tr>
<tr>
<td>Trp cage</td>
<td>1L2Y</td>
<td>304</td>
<td>AMBER99SB</td>
<td>12</td>
<td>255 – 310</td>
</tr>
<tr>
<td>Protein A</td>
<td>1BDD</td>
<td>941</td>
<td>AMBER99SB</td>
<td>12</td>
<td>255 – 310</td>
</tr>
</tbody>
</table>

**MD simulations**

For comparison, we also performed 200 ns of conventional MD simulations for alanine dipeptide and oseltamivir under the same conditions as that of the REMD simulations. In the conventional MD simulations, temperatures were set at 280 K and 300 K.

**Folding/unfolding simulations**

In addition to the temperature-dependent entropy calculations, we also test the entropy calculations for isothermal processes. To this end, we performed folding/unfolding simulations for a small protein, Chignolin (PDB ID: 1UAO), using the REMD method. The conditions of the folding/unfolding simulations were the same as those of other proteins, except for the temperature range of 315–400 K.

**Computational of configurational entropy**

For the calculation of $\Delta S_{\text{Clausius}}$, $\Delta E_{\text{Potential}}$ was calculated using the trajectory data sampled at $T_0$ and $T_1$. $T_0$ was 280 K and $T_1$ was set at several temperatures. The same $T_0$ and $T_1$
were used for the calculation of $\Delta S_{\text{Solute}}^{\text{Conf}}$ using the other six methods.

For $S_{\text{IC}}^\text{CL}$, $S_{\text{IC}}^\text{QM}$, $S_{\text{IC}}^\text{QMS}$ and $S_{\text{IC}}^\text{QMK}$, we first removed the overall translations and rotations of the solutes from the trajectory data using the least square fit, and then $\omega_i$ was obtained through a PCA of the mass-weighted variance-covariance matrix. After diagonalization of the matrix, $3N - 6$ eigenvalues except for the six eigenvalues that are essentially zero were used for the calculations.

For $S_{\text{IC}}^\text{QH}$ and $S_{\text{IC}}^\text{BQH}$, the transformation from Cartesian coordinates to internal coordinates was applied to the MD trajectories. Correlations involving torsion angles $q_i$ were calculated using variables $z = e^{iq_i}$ in accordance with the study by Harpole and Sharp. This treatment avoids the problem of torsion angle discontinuity. An Intel math kernel library was used for the diagonalization of $C$.

For an estimate of statistical errors, we employed a cumulative average by dividing the trajectory data into 20 segments, each of which is 10 ns. The cumulative average of a quantity $A$ is represented by $\bar{A}$. Cumulative error was evaluated at 95 % confidence of $\Delta S_{\text{Conf}}^{\text{Solute}}$.

**Assessment of computational methods for configurational entropy**

The quantitative performance of computational methods for entropy was tested by the absolute error $\sigma_{\text{AbsE}}$ and the deviation rate $R_{\text{dev}}$ which are defined by:

$$\sigma_{\text{AbsE}} = \left| \Delta S_{\text{Solute}}^{\text{Conf,X}} - \Delta S_{\text{Clausius}} \right|,$$

$$R_{\text{dev}} = \frac{\sigma_{\text{AbsE}}}{\Delta S_{\text{Clausius}}} \times 100 \, (\%),$$

where $S_{\text{Conf,X}}^{\text{Solute}}$ is one of $S_{\text{CL}}^\text{CC}$, $S_{\text{QM}}^\text{CC}$, $S_{\text{QMS}}^\text{CC}$, $S_{\text{QMK}}^\text{CC}$, $S_{\text{IC}}^\text{QH}$, $S_{\text{IC}}^\text{BQH}$ and $S_{\text{Ind}}^\text{IC}$.

For folding/unfolding simulations of Chignolin, the configurational entropies were compared with the thermodynamic entropy calculated from the temperature derivatives of free
energy changes. The free energy changes upon folding were expressed as,

$$
\Delta G^F_{UF}(T) = -k_B T \log \frac{P_F(T)}{P_U(T)},
$$

(18)

where $\Delta G^F_{UF}(T)$ is the free energy change from the unfolded to folded state at temperature $T$, and $P_F(T)$ and $P_U(T)$ are the probabilities of the unfolded and folded structures at temperature $T$, respectively. From the free energy changes, the entropy changes were evaluated as,

$$
\Delta S^F_{UF}(T) = -\frac{\partial \Delta G^F_{UF}}{\partial T}.
$$

(19)

Because the temperature-independent implicit solvent model was used in this study, the thermodynamic entropy in eq. (19) corresponds to the configurational entropy of the protein. $\Delta S^F_{UF}(T)$ was calculated from the polynomial fit of $\Delta G^F_{UF}(T)$. 

Results and Discussion

Convergent behaviors of configurational entropies

First, the convergent behaviors of the configurational entropies were examined. In the cases of oseltamivir and alanine dipeptide, the convergent behavior of \( \Delta S_{\text{Solute Conf}} \) obtained through the REMD simulations was significantly better for all cases than that of the conventional MD simulations (Figures 2 (a)-(d) and S1). \( \Delta S_{\text{Solute Conf}} \) were sufficiently converged at 100 ns when the REMD simulations were employed. By contrast, entropies calculated using conventional MD simulations were not converged even after 200 ns. Therefore, REMD simulations are required for reliable calculations of configurational entropies.

Time evolutions of proper torsions provide a clue for why rapid convergence was achieved using REMD simulations (Figures 2 (e), (f), and S2). In conventional MD simulations, transitions between local minima were observed only a few times. In contrast, many transitions between local minima occurred in the REMD simulations, indicating sufficient sampling for proper torsions.

Using the trajectories from the REMD simulations, we also analyzed the convergent behaviors of \( \Delta S_{\text{Solute Conf}} \) for the other solutes. It was found that \( \Delta S_{\text{Solute Conf}} \) were sufficiently converged at 200 ns (Figure S3). Thus, we show the results obtained through the 200 ns REMD simulation hereafter.

Assessment of computational methods for configurational entropy

\( \Delta S_{\text{Solute Conf}} \) calculated by the six methods were compared to \( \Delta S_{\text{Clausius}} \) (Figure 3). Though only the results at \( T_1 = 300 \text{ K} \) are shown in Figure 3, the results for the other \( T_1 \) were given in Figure S4.

First, the BQH method outperformed the other methods for the small molecules and proteins. This is caused by the improvement of introducing improper torsions, as discussed in the next section. Regarding the choice of the coordinate system employed in the entropy
Figure 2: The convergent behaviors of $\Delta S_{\text{Solute}}^{\text{Conf}}$ for oseltamivir using the REMD (blue) and MD (red) simulations. Temperatures are set at $T_0 = 280$ K and $T_1 = 300$ K. Panels (a)–(d) represent the time evolutions of $\Delta S_{\text{Clausius}}$ (a), $\Delta S_{\text{CL}}^{\text{CC}}$ (b), $\Delta S_{\text{QH}}^{\text{IC}}$ (c), and $\Delta S_{\text{BQH}}^{\text{IC}}$ (d), respectively. In panels (e) and (f), time evolutions of a proper torsion angle of oseltamivir in the REMD (e) and MD (f) simulations at 300 K are shown. The error bars indicate 95% confidence interval.
calculations, internal coordinates were found to be superior to Cartesian coordinates, particularly for proteins. In small molecules, the accuracy of $\Delta S_{\text{CC}}$ was comparable to those of $\Delta S_{\text{QH}}$ and $\Delta S_{\text{BQH}}$, suggesting the QH approximation works well, even in Cartesian coordinates for small molecules. However, because protein fluctuations are much more complicated, the QH approximation in Cartesian coordinates becomes no longer valid. Consequently, internal coordinates are appropriate for the QH and related methods. Additionally, details are discussed subsequently when we address the validity of the QH approximation in internal coordinates.

As shown in Figure 3, the QH methods using quantum harmonic oscillators exhibit poor quantitative performance when they were compared to $\Delta S_{\text{Clausius}}$. This result is understandable because $\Delta S_{\text{Clausius}}$ is founded on classical statistical mechanics. However, in our opinion, the employment of quantum harmonic oscillators for entropy calculations in the current context is inadequate for the following reasons. First, the MD simulations were performed on the basis of classical mechanics. Thermodynamic quantities such as free energies have thus far been successfully calculated on the basis of classical statistical mechanics using classical MD simulations. Second, the modes obtained using the QH approximation are not true harmonic oscillations. Because proteins undergo transitions among local free-energy minima at room temperature, protein dynamics are not simple oscillators. The QH approximation should be considered as a Gaussian approximation of PDFs rather than a harmonic-oscillator approximation of protein dynamics. Third, the QH approximation in Cartesian coordinates is invalid for proteins, as discussed above. Consequently, the QH methods employing the entropy formulations of quantum harmonic oscillators in Cartesian coordinate system, which are implemented in most MD software packages, cannot be recommended for the calculation of configurational entropy.
Figure 3: Accuracy of the configurational entropies computed using the six QH and BQH methods relative to the Clausius method. The values of $\Delta S^{\text{IC}}_{QH}$ (green), $\Delta S^{\text{IC}}_{BQH}$ (blue), $\Delta S^{\text{CC}}_{CL}$ (orange), $\Delta S^{\text{CC}}_{QM}$ (magenta), $\Delta S^{\text{CC}}_{QMS}$ (cyan) and $\Delta S^{\text{CC}}_{QMK}$ (purple) are plotted against those of $\Delta S^{\text{Clausius}}$ for small molecules (a) and proteins (b). Figures (c) and (d) represent the values of $\sigma_{\text{AbsE}}$ (top) and $R_{\text{dev}}$ (bottom) obtained using the six methods for the small molecules (c) and proteins (d). Temperatures are set at $T_0 = 280$ K and $T_1 = 300$ K.
Improvement of the BQH method by improper torsions

We now examine how introducing improper torsions improved the BQH method. Though the BQH method can appropriately treat multimodal PDFs for torsion angles, the treatment of the correlations between coordinates in the BQH method is identical to the QH method; that is, the correlation is considered up to the second order. However, as described in the Introduction, three-body (or more) correlations in torsion angles are frequently observed in proteins. To reduce the effects of multi-body correlations, we introduced improper torsion angles to replace some of the proper torsion angles.

For comparison, the BQH entropy, for which all torsions were treated as proper torsions, was also calculated (See Figure 4). As a result, the BQH entropy only with proper torsions clearly worsened $R_{\text{dev}}$ to 7–10% relative to the case in which improper torsions were introduced. Therefore, the use of improper torsions is inevitable to quantitatively calculate $S_{\text{Conf}}^{\text{Solute}}$ with the BQH method (Figures 3 and S4).

Figure 4: Improvement of the BQH method by introducing improper torsions. $R_{\text{dev}}$ for the BQH method treating all torsion angles as proper torsions are compared with $R_{\text{dev}}$ when improper torsions were introduced. Temperatures are set at $T_0 = 280$ K and $T_1 = 300$ K.

Origin of the superiority of the BQH method

Through the decomposition of $\Delta S_{\text{BQH}}^{\text{Conf}}$ into components, we investigated the reason why the BQH method outperformed the other methods. First, it was decomposed into the kinetic, independent, correlation and Jacobian terms using eq. (11). Figure 5 (a) shows the ratio
\( R_C \) defined by:

\[
R_C = \frac{\Delta S^{\text{Solute}}_i}{\Delta S^{\text{IC}}_{\text{BQH}}} \times 100 \text{ (%)},
\]

where \( i \) is the kinetic, independent, correlation or Jacobian term. In Figure 5 (a), \( R_C \) of the Jacobian term is not shown because it was very small. Instead, the values of the Jacobian term are shown in Figure S5. Although the Jacobian term was very small in the coordinate system employed here, it depends on the choice of the coordinate system. The kinetic and independent terms mainly contributed to \( \Delta S^{\text{Solute}}_{\text{Conf}} \). The kinetic term was calculated analytically; therefore, the accuracy of the independent term dominated the quantitative performance of the methods. Because the independent term was accurately calculated using the Boltzmann-Shannon equation in the BQH method, this method outperformed the other methods.

Contributions of the correlation term to \( \Delta S^{\text{Solute}}_{\text{Conf}} \) was small relative to the kinetic and independent terms. Particularly for small molecules without rings, i.e., butane and alanine dipeptide, correlation terms were very small (0.07 % for butane, 0.56 % for alanine dipeptide), consistent with the previous study on alanine dipeptide.\(^{42}\) For small molecules with rings, the correlation terms were slightly observed. By contrast, in Protein A, the correlation terms relatively more contributed to the entropy changes, possibly because of the well packed three-dimensional structures of the protein. Interestingly, in the ligand binding to proteins, it was pointed out that the contribution of the correlation terms could not be ignored.\(^{43}\)

To examine whether the correlation terms estimated by the BQH approach contributes to the accuracy of the entropy calculation, the independent and kinetic terms without the correlation term of the BQH entropy were compared to \( \Delta S^{\text{IC}}_{\text{BQH}} \) (Figure 5(b)). In all the model systems except for cyclohexane, the correlation term of the BQH entropy contributed to the accuracy \( \Delta S^{\text{Solute}}_{\text{Conf}} \). Because the ring structure of cyclohexane exhibited structural transitions between stable conformations, higher-order correlations might cancel the second-order cor-
relations evaluated by the BQH method. Therefore, only the second-order correlations of the BQH method might be insufficient for cyclohexane. By contrast, in proteins, i.e., Trp cage and Protein A, the correlation term of the BQH method greatly contributed to the accuracy of $\Delta S^\text{Conf}_\text{Conf}$, suggesting that the harmonic approximation employed in the correlation term of the BQH method works properly for the well-packed three-dimensional structures of proteins.

Figure 5: Contributions of kinetic, independent and correlation terms in $\Delta S_{\text{IC}}^\text{BQH}$. In panel (a), the ratios $R_C$ of the kinetic (white), independent (black) and correlation (mesh) terms in $\Delta S_{BQH}^\text{IC}$ are shown. In panel (b), $R_{\text{dev}}$ for $\Delta S_{BQH}^\text{IC}$ (white) and only the independent and kinetic terms without the correlation term of $\Delta S_{BQH}^\text{IC}$ (black) are shown. Temperatures were set at $T_0 = 280$ K and $T_1 = 300$ K.

Further, the independent term was decomposed into the bond-length, bond-angle, proper and improper torsion angle terms. We examined the anharmonicity of each coordinate using the technique by Harpole and Sharp.\textsuperscript{35} In this case, the degree of anharmonicity was analyzed using the following ratio $R_{i}^{\text{anh}}$:

$$R_{i}^{\text{anh}} = \frac{\Delta S_{i}^{\text{Ind}}}{\Delta S_{i}^{\text{1D, harm}}},$$  \hspace{1cm} (21)

where $\Delta S_{i}^{\text{Ind}}$ is the cumulative average of the temperature change of $S_{i}^{\text{Ind}} \equiv -k_B \int dq_i \rho(q_i) \ln \rho(q_i)$. 

24
The degree of anharmonicity for internal coordinates in protein A. $R_{\text{anh}}$ for bond lengths (a), bond angles (b), proper torsion angles (c), and improper torsions (d) are plotted. Dashed lines represent an exactly harmonic behavior ($R_{\text{anh}} = 1$). Temperatures were set at $T_0 = 280$ K and $T_1 = 300$ K.

$\Delta S_{\text{harm}}^{1\text{D}}$ is the entropy change of the harmonic oscillator with temperature and is given by $1/2k_B\ln(T_1/T_0)$. If $R_{\text{anh}}^i$ is close to 1, the $i$-th internal coordinate exhibits harmonic behavior.

The results of protein A and the other molecules are shown in Figures 6, S6 and S7.

The behaviors of the bond lengths and angles were nearly harmonic (Figures 6 (a), (b)). In contrast, highly anharmonic behaviors were observed for the proper torsions (Figure 6 (c)), indicating that the BQH method is suitable for proper torsions. This is the reason for the superior performance of the BQH method. It was also found that the behavior of the improper torsions was nearly harmonic (Figure 6 (d)). Thus, the introduction of improper torsions improves the performance of the BQH method. The results for the other solute molecules were similar to the case for protein A (Figures S6 and S7).

Here, the three proper torsions with the largest $R_{\text{anh}}^i$ are shown as an example in Figure 7. All three torsions were located in the side chains of residues. The PDFs of these proper torsions were multi-modal distributions that were quite different from the Gaussian distributions (Figure 7 (d), (e), and (f)). Furthermore, the relative heights of the peaks in the
Figure 7: The three proper torsions that have the largest $R_{\text{anh}}$ in protein A. Proper torsions are located in the side chains of Asn53 ($R_{\text{anh}}=11.02$), Phe31 ($R_{\text{anh}}=9.18$) and Arg28 ($R_{\text{anh}}=9.07$), and are indicated by arrows. The PDFs of three torsion angles at 280 K (green) and 300 K (red) are shown in panels (d), (e), and (f).

PDFs changed with temperature, indicating that the populations of the side-chain rotamers shifted as the temperature changed. These configurational changes were far from harmonic behavior, particularly in Cartesian coordinates. Therefore, the BQH method using internal coordinates is efficient for the calculation of configurational entropy.

**Entropy changes for folding/unfolding transitions**

The final test is the entropy calculations for isothermal processes, which are relevant to various biomolecular phenomena such as protein folding and ligand binding. We successfully performed folding/unfolding REMD simulations for a small protein, Chignolin. The root mean square deviation (RMSD) of 2.5 Å from the experimental structure was used as the threshold between the folded and unfolded states. As shown in Figure 8 (a), the melting curve indicates that the folding/unfolding transitions took place at the melting temperature $T_m$ of 305–315 K. The experimental $T_m$ is 311–315 K, which is in good agreement with the
simulation $T_m$. The probabilities of the folded and unfolded structures at each temperature ($P_F(T)$ and $P_U(T)$, respectively) were well converged after 150 ns (Figure S8). From the polynomial fit of the free energy differences, $\Delta G^F_U(T)$ (Figure 8 (b)), calculated from $P_F(T)$ and $P_U(T)$, we estimated the thermodynamic entropy changes $\Delta S^F_U(T)$ from the unfolded to folded state (Figure 8 (e)). The experimental unfolding free energy and entropy are 0.26–0.45 kcal/mol (298 K) and 19.7–24.3 cal/mol/K ($T_m$), respectively. The unfolding free energy and entropy calculated from the simulation were $0.13 \pm 0.1$ kcal/mol (300 K) and $16.4 \pm 1.1$ cal/mol/K ($T_m \sim 315$ K), respectively, which are somewhat smaller than the experimental data. Because we employed the temperature-independent implicit solvent model in this study, the thermodynamic entropy corresponds to the configurational entropy. Therefore, the discrepancy to the experimental data is possibly due to hydration effects.

For each of the folded and unfolded states, the configurational entropies from PDF were calculated using six methods. Because both folded and unfolded structures were sufficiently sampled around $T_m$, the configurational entropies around $T_m$ are statistically reliable. Here, we also compared the independent term without the correlation term of the BQH method, $\Delta S_{\text{Ind}}$. As shown in Figure 8 (c) and (d), at $T_m$, $\Delta S_{\text{ICBQH}}$ and $\Delta S_{\text{Ind}}$ with improper torsions were in reasonable agreement with the thermodynamic entropy changes. Without improper torsions, the calculation results were significantly worsened (Figure 8 (e)). The correlation term of $S_{\text{ICBQH}}$ worsened the discrepancy to the thermodynamic at $T_m$ and the discrepancy became larger at temperatures higher and lower than $T_m$. As a consequence, $\Delta S_{\text{Ind}}$ was the most accurate method for the entropy calculation in this folding/unfolding simulation. The small or opposite contribution of the correlation term may be due to the diversity of unfolded structures and the large conformational changes between folded and folded structures.

However, omitting the correlation term in entropy calculation may not be extended to other processes such as the ligand binding. Compared with the folding transitions, conformational changes upon ligand binding and the diversity of unbound structures of ligands and receptors may not be very large. Instead, the well-packed three-dimensional structures
of receptors have correlated motions. Therefore, it is possible that the correlation term of the configurational entropy plays important roles.

Figure 8: Folding and unfolding REMD simulations for Chignolin. Panels (a) and (b) show the temperature dependences of $P_F(T)$ and $\Delta G_{FU}^F(T)$, respectively. Panels (c) and (d) represent $\sigma_{AbsE}$ and $R_{dev}$, respectively, obtained using seven methods at $T = 305$ K. In panel (e), the temperature dependences of the isothermal configurational-entropy changes for the folding transition of Chignolin are compared among eight methods. In the panels, colors represent $\Delta S_{FU}^F$ (reference) (red), $\Delta S_{CC}^{CL}$ (orange), $\Delta S_{CC}^{QM}$ (magenta), $\Delta S_{CC}^{QMS}$ (cyan), $\Delta S_{QM}$ (purple), $\Delta S_{QH}^{IC}$ (green), $\Delta S_{BQH}^{IC}$ (blue), and $\Delta S_{Ind}$ of BQH (black) with improper torsions, and $\Delta S_{Ind}$ of BQH (gray) without improper torsions. The error bars indicate 95% confidence interval. In Figure S9, $T$-weighted entropies are shown.
Conclusion

In the present paper, the quantitative performance of the QH and related computational methods for configurational entropy, $S_{\text{Conf}}^\text{Solute}$, was investigated using the Clausius method as a reference. The Clausius method is capable of accurately computing the change of $S_{\text{Conf}}^\text{Solute}$ upon temperature change, $\Delta S_{\text{Conf}}^\text{Solute}$, despite the fact that the calculation is limited to the temperature change and the temperature-independent implicit solvent model. The six QH and related computational methods, including widely used methods implemented in popular MD program packages, were tested to examine their accuracies for a set of solutes including small molecules to proteins. The BQH method outperformed all the methods examined. The introduction of improper torsions reduced the effects of multi-body correlations, improving the BQH method. The kinetic and independent terms of the internal coordinates dominate $\Delta S_{\text{Conf}}^\text{Solute}$. Anharmonicity of proper torsions, observed in Protein A, was a reason of the superior performance of the BQH method to the QH method. Contributions of the correlation term in the BQH method were dependent on models to some extent. In small molecules, the correlation terms were very small. By contrast, in Protein A, relatively more contributions of the correlation terms were observed, possibly because of the well-packed three-dimensional structure of the protein.

Internal coordinates should be used to calculate configurational entropy because the QH approximation in Cartesian coordinates is invalid, particularly for proteins. Many methods implemented in popular MD program packages for the calculation of configurational entropy employ the QH approximations coupled with Cartesian coordinates and quantum harmonic oscillators. These approximations were found to give poor performance.

For the convergence of the entropy calculations, REMD simulations are efficient to avoid becoming trapped in local energy minima. In conventional MD simulations, the rate of transitions in torsion angles is too slow to obtain sufficient sampling. In the current simulations, no constraints, such as SHAKE, were applied. The quantitative performance of the QH and related methods using the REMD simulations with the SHAKE algorithm should be
assessed in future.

For isothermal processes, the configurational-entropy changes upon folding/unfolding transitions were evaluated. Using REMD methods, we successfully performed MD simulations for folding/unfolding transitions of a small protein, Chignolin. At melting temperature, the BQH-entropy change was in reasonable agreement with the thermodynamic-entropy change calculated from the temperature derivative of the free-energy changes upon folding. The introduction of improper torsions clearly improved the performance. However, the independent term without the correlation term in the BQH method was more accurate. This is because the QH approximation employed for the calculation of the correlation term in BQH method was no longer valid for the divergent unfolded structures. In fact, in the wide range of temperature around the melting temperature, the independent term in the BQH method was in good agreement with the thermodynamic entropy.

The hydration entropy is an important issue for entropy calculation. In this study, the implicit solvent model was employed for the quantitative assessment of the computational methods for configurational entropies. Efficient computational methods for hydration entropy have been proposed so far. Thus, the combination of the BQH and related methods with the computational methods for hydration entropy will allow us to calculate the total entropy changes upon biophysically relevant processes. The assessment of such combinations is also an important issue of future studies.

Supporting Information Available

Convergent behaviors of $\Delta S_{\text{Solute}}^{\text{Conf}}$ for alanine dipeptide using the REMD and MD simulations (Figure S1), time evolutions of the $\phi$ and $\psi$ angles of alanine dipeptide obtained from the REMD and MD simulations at 300K (Figure S2), convergent behaviors of $\Delta S_{\text{Solute}}^{\text{Conf}}$ using the trajectories obtained through the REMD simulations for butane, cyclohexane, alanine dipeptide, oseltamivir, trp cage, and protein A (Figure S3), temperature dependences of
configurational entropies using the six QH and BQH methods (Figure S4), temperature
dependences of the Jacobian term of the QH and BQH methods with internal coordinate
systems (Figure S5), harmonic behaviors of bond lengths, bond angles and improper torsions
(Figure S6), anharmonic behaviors of proper torsion angles (Figure S7), convergent behaviors
of \( P_F(T) \) and \( P_U(T) \) of chignolin at each temperature (Figure S8), and temperature-weighted
folding entropies for Chignolin (Figure S9). This material is available free of charge via the
Internet at http://pubs.acs.org/.

Acknowledgement

We thank Prof. Akinori Kidera and Drs. Sotaro Fuchigami, Tsutomu Yamane and Toru Eki-
moto for helpful discussions. This work was financially supported by the Platform for Drug
Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture,
Sports, Science and Technology (MEXT) and the Japan Agency for Medical Research and
Development (AMED) to M.I.; Innovative Drug Discovery Infrastructure through Functional
Control of Biomolecular Systems, Priority Issue 1 in Post-K Supercomputer Development
(Project ID:hp150269 and hp160223) from MEXT to M.I.; Grants-in-Aid for Scientific Re-
search (No. 25291036) to M.I; RIKEN Dynamic Structural Biology Project to M.I.. This
research used computational resources at the Tsurumi Campus in Yokohama City University.

References


(13) Yoshidome, T; Ekimoto, T; Matubayasi, N; Harano, Y; Kinoshita, M; Ikeguchi, M *J. Chem. Phys.* 2015, 142, 175101.

32


Graphical TOC Entry

for Table of Contents use only
Supporting information for:
Computational methods for configurational entropy using internal and Cartesian coordinates

Simon Hikiri,† Takashi Yoshidome,†‡ and Mitsunori Ikeguchi*,†

†Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suchiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
‡Present address: Department of Applied Physics, Graduate School of Engineering, Tohoku University, 6-6-05, Aoba, Aramaki, Aoba-ku, Sendai 980-8579, Japan

E-mail: ike@tsurumi.yokohama-cu.ac.jp
Figure S1: Convergent behaviors of $\Delta S_{\text{Solute}}^{\text{Conf}}$ for alanine dipeptide using the REMD (blue) and MD (red) simulations. Temperatures were set at $T_0 = 280$ K and $T_1 = 300$ K. Panels (a)–(d) represent the time evolutions of $\Delta S_{\text{Cl}}$ (a), $\Delta S_{\text{CL}}^{\text{CC}}$ (b), $\Delta S_{\text{QH}}^{\text{IC}}$ (c), and $\Delta S_{\text{BQH}}^{\text{IC}}$ (d), respectively. The error bars indicate 95% confidence interval.
Figure S2: Time evolutions of the $\phi$ ((a) and (b)) and $\psi$ ((c) and (d)) angles of alanine dipeptide obtained from the REMD ((a) and (c)) and MD simulations ((b) and (d)) at 300K.
Figure S3: Convergent behaviors of $\Delta S_{\text{Conf}}^{\text{Solute}}$ using the trajectories obtained through the REMD simulations for butane (a), cyclohexane (b), alanine dipeptide (c), oseltamivir (d), trp cage (e), and protein A (f). In the panels, the colors represent $\Delta S_{\text{Clausius}}$ (red), $\Delta S_{\text{CL}}^{\text{CC}}$ (orange), $\Delta S_{\text{QH}}^{\text{IC}}$ (green), and $\Delta S_{\text{BQH}}^{\text{IC}}$ (blue). Temperature were set at $T_0 = 280$ K and $T_1 = 300$ K. The error bars indicate 95 % confidence interval.
Figure S4: Temperature dependences of configurational entropies using the six QH and BQH methods. The values of $\Delta S_{\text{Clausius}}$ (red), $\Delta S_{\text{IC QH}}$ (green), $\Delta S_{\text{IC BQH}}$ (blue), $\Delta S_{\text{CC CL}}$ (orange), $\Delta S_{\text{CC QM}}$ (magenta), $\Delta S_{\text{CC QMS}}$ (cyan) and $\Delta S_{\text{CC QMK}}$ (purple) for butane (a), cyclohexane (b), alanine dipeptide (c), oseltamivir (d), trp cage (e) and protein A (f) were plotted against $T$. The reference temperature was set at $T_0 = 280 \text{ K}$. The error bars indicate 95 % confidence interval.
Figure S5: Temperature dependences of the Jacobian term of the QH and BQH methods with internal coordinate systems. The changes in the Jacobian term with temperature increase ($\Delta S_{\text{Jacobian}}$) for butane (a), cyclohexane (b), alanine dipeptide (c), oseltamivir (d), trp cage (e) and protein A (f) were plotted against $T$. The reference temperature was set at $T_0 = 280$ K. The error bars indicate 95% confidence interval.
Figure S6: Harmonic behaviors of bond lengths, bond angles and improper torsions. The values of $R^{anh}$ for bond lengths (blue), bond angles (green), and improper torsions (red) for butane (a), cyclohexane (b), alanine dipeptide (c), oseltamivir (d) and trp cage (e) are plotted. The dashed line indicates exactly harmonic behavior ($R^{anh} = 1$). Temperatures are set at $T_0 = 280$ K and $T_1 = 300$ K.
Figure S7: Anharmonic behaviors of proper torsion angles. The value of $R^{\text{anh}}$ for proper torsions for butane (a), cyclohexane (b), alanine dipeptide (c), oseltamivir (d) and trp cage (e) are plotted. The dashed line indicates exactly harmonic behavior ($R^{\text{anh}} = 1$). Temperatures are set at $T_0 = 280$ K and $T_1 = 300$ K.
Figure S8: Convergent behaviors of $P_f(T)$ (blue) and $P_u(T)$ (red) of chignolin at each temperature. The error bars indicate 95% confidence interval.
Figure S9: Temperature-weighted folding entropies for Chignolin. Panel (a) represents $\sigma_{\text{AbsE}}$ obtained using seven methods at $T = 305$ K. In panel (b), the temperature-weighted isothermal entropy changes for the folding transition of Chignolin are shown. In the panel, colors represent $-T\Delta S^\text{F}_U$ (reference) (red), $-T\Delta S^\text{CC}_\text{CL}$ (orange), $-T\Delta S^\text{CC}_\text{QM}$ (magenta), $-T\Delta S^\text{CC}_\text{QMS}$ (cyan), $-T\Delta S^\text{IC}_\text{QH}$ (green), $-T\Delta S^\text{IC}_\text{BQH}$ (blue), and $-T\Delta S^\text{Ind}_{\text{BQH}}$ (black) with improper torsions, and $-T\Delta S^\text{Ind}_{\text{BQH}}$ (gray) without improper torsions. The error bars indicate 95% confidence interval.