

Supporting Information I for the Movable Type Method Applied to Protein-Ligand Binding

*Zheng Zheng, Melek N. Ucisik and Kenneth M. Merz, Jr. **

Department of Chemistry and the Quantum Theory Project, 2328 New Physics Building, P.O.
Box 118435, University of Florida, Gainesville, Florida 32611-8435

Here we describe a very simple methane-butane two-body system as an example that illustrates, in detail, the binding free energy calculation using the "movable type" (MT) method with the KECSA energy function. The binding of two-molecules in solution can be written as:

$$\Delta G_b^s = \Delta G_b^g + \Delta G_{solv}^{1-2} - \Delta G_{solv}^1 - \Delta G_{solv}^2 \quad (1i)$$

where methane and butane are indicated as molecule 1 and 2, respectively.

Using $\Delta\Delta G_{solv} = \Delta G_{solv}^{1-2} - \Delta G_{solv}^1 - \Delta G_{solv}^2$, Equation 1i simplifies to:

$$\Delta G_b^s = \Delta G_b^g - \Delta\Delta G_{solv} \quad (2i)$$

We use the canonical ensemble, so we are computing Helmholtz free energies in Equation 2i.

Because both methane and butane have no polar atoms, and their accessible surface areas have a negligible change after association, we can ignore computing $\Delta\Delta G_{solv}$ further simplifying this example.

For the gas-phase binding free energy calculation, which is the core of the problem, the binding free energy is given as a ratio of partition functions:

$$\Delta G_b^g \approx \Delta A_b^g = -RT \ln \left[\frac{Z_{1-2}}{Z_2 Z_1} \right] = -RT \ln \left[\frac{\int e^{-\beta E_{1-2}(r)} dr}{\int e^{-\beta E_2(r)} dr \int e^{-\beta E_1(r)} dr} \right] = -RT \ln \left[\frac{F_{1-2} \langle e^{-\beta E_{1-2}(r)} \rangle}{F_2 \langle e^{-\beta E_2(r)} \rangle F_1 \langle e^{-\beta E_1(r)} \rangle} \right] \quad (3i)$$

As is shown in Equations 5 and 6 in the article, the degrees of freedom (DoFs) for protein-ligand systems can be approximated as:

$$\frac{F_{PL}}{F_P F_L} = \frac{F_{boundP}^{external} F_{boundL}^{external} F_{boundP}^{internal} F_{boundL}^{internal}}{F_{freeP}^{external} F_{freeP}^{internal} \times F_{freeL}^{external} F_{freeL}^{internal}} = \frac{F_{boundL}^{external}}{F_{freeL}^{external}} = \frac{a\pi^2 V_{pocket}}{8\pi^2 C} = \frac{aV_{pocket}}{8C} \quad (4i)$$

For small-molecule systems, like the methane-butane case in our current example, if butane is treated as the "receptor" and methane as the "ligand", then V_{pocket} (translational DoF for the bound ligands) is approximately equal to C (translational DoF of the free ligand), because the smaller the "receptor" is, the weaker the constraints on the "ligand". A small "receptor" like butane, when compared to any protein molecule, has a greatly reduced constraint on the translational movement of a "ligand" like methane, making $V_{pocket} \approx C$. Moreover, for the rotational DoF, the volume factor "a" is also approximately equal to 8, for small molecules because the small receptor affords more accessible volume than found in protein-ligand systems. Hence for two-body small molecule systems:

$$\frac{F_{1-2}}{F_2 F_1} \approx 1 \quad (5i)$$

yielding:

$$\Delta G_b^g \approx -RT \ln \left[\frac{\langle e^{-\beta E_{1-2}(r)} \rangle}{\langle e^{-\beta E_2(r)} \rangle \langle e^{-\beta E_1(r)} \rangle} \right] \quad (6i)$$

Using the MT method, molecular energies are decomposed into atom pair interaction energies, including bond, angle, torsion and long-range interactions sampled along atom pairwise distances multiplied by the corresponding probabilities, as shown in Equation 7i:

$$\langle e^{-\beta E} \rangle = \sum_i \prod_j Q_{ij} e^{-\beta E_{ij}} \quad (7i)$$

On the left hand side of Equation 7i, E is the microstate energy within an ensemble average, and on the right hand side each "j" indicates each atom pair in the molecule.

In the KECSA implementation of the MT method, only heavy atoms are utilized in the calculation, making methane molecule one of the simplest cases:

$$E_{\text{methane}} = E_{\text{sp}^3 \text{ Carbon}} \quad (8i)$$

$$\langle e^{E_{\text{methane}}} \rangle = e^{E_{\text{sp}^3 \text{ Carbon}}} \quad (9i)$$

The atom pairwise energy is zero in methane, making the energy of methane equal to the energy of a single sp^3 carbon atom. In the MT method, it is not necessary to carry out a single atom energy calculation, because they cancel out when using Equation 6i for the free energy calculation.

Butane has four sp^3 carbon atoms in a chain, making three sp^3 carbon- sp^3 carbon bonds, two sp^3 carbon- sp^3 carbon angles, and one sp^3 carbon- sp^3 carbon torsion. Sampling along the distance for all interactions and gathering all possible combinations generates all the necessary energy terms in Equation 10i.

$$\left\langle e^{E_{\text{butane}}} \right\rangle = \sum_{\alpha}^{\text{Bond Distance Range}} \prod_a^3 Q_{a\alpha} e^{-\beta E_{a\alpha}} \times \sum_{\beta}^{\text{Angle Distance Range}} \prod_b^2 Q_{b\beta} e^{-\beta E_{b\beta}} \times \sum_{\gamma}^{\text{Torsion Distance Range}} \prod_c^1 Q_{c\gamma} e^{-\beta E_{c\gamma}} \quad (10i)$$

Using bond energy terms as an example (as used in the article) each of the three sp^3 carbon- sp^3 carbon bond partition functions can be modeled as

$$z_k^{\text{bond}} = \begin{bmatrix} z_k^{\text{bond}}(r_1) \\ z_k^{\text{bond}}(r_2) \\ \vdots \\ z_k^{\text{bond}}(r_a) \\ \vdots \\ z_k^{\text{bond}}(r_n) \end{bmatrix} = \begin{bmatrix} e^{-\beta E_k^{\text{bond}}(r_1)} \\ e^{-\beta E_k^{\text{bond}}(r_2)} \\ \vdots \\ e^{-\beta E_k^{\text{bond}}(r_a)} \\ \vdots \\ e^{-\beta E_k^{\text{bond}}(r_n)} \end{bmatrix} \quad (11i)$$

k indicates each sp^3 carbon- sp^3 carbon bond and 1 through n are the distance increments. Each increment is set to 0.005\AA . The bond energy is modeled as a harmonic oscillator:

$$z_k^{\text{bond}}(r_a) = e^{-79.98(r_a - 1.53)^2 + 4.195} \quad (12i)$$

-79.98 is the energy unit used in KECSA and will be ultimately converted into kcal/mol. 1.53\AA is the distance at which the sp^3 carbon- sp^3 carbon bond energy reaches its minimum and 4.195 is a constant that adjusts the energy baseline to zero.

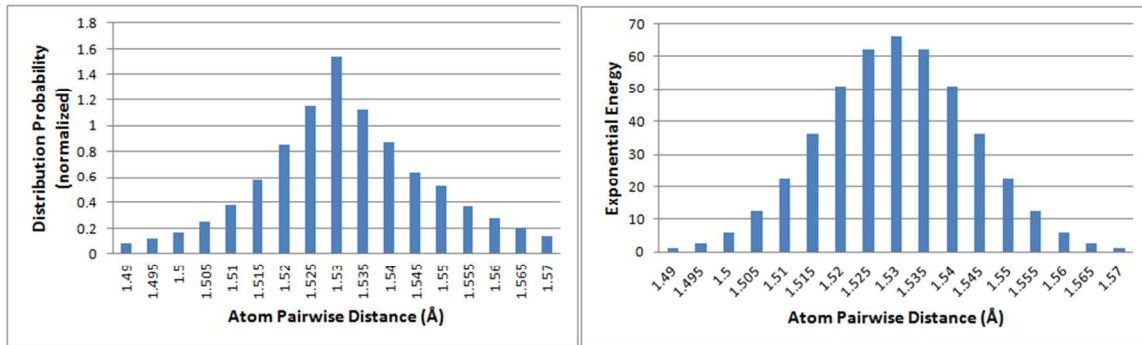


Figure 1. sp^3 carbon- sp^3 carbon bond probability distribution and exponential energy vs. atom pairwise distance.

Sampling of bond energies includes all distances that derive non-zero partition function values, which range from 1.72 Å to 1.99Å in the sp^3 carbon- sp^3 carbon bond interaction case. The product over all three bond-linked sp^3 carbon- sp^3 carbon pairs derives the total bond partition function in butane (Equation 13i).

$$\begin{aligned}
Z_{\text{butane}}^{\text{bond}} &= z_1^{\text{bond}} \otimes z_2^{\text{bond}} \otimes z_3^{\text{bond}} \\
&= \begin{bmatrix} e^{-\beta E_1^{\text{bond}}(r_1)} \\ \vdots \\ e^{-\beta E_1^{\text{bond}}(r_a)} \\ \vdots \\ e^{-\beta E_1^{\text{bond}}(r_n)} \end{bmatrix} \begin{bmatrix} e^{-\beta E_2^{\text{bond}}(r_1)} & \dots & e^{-\beta E_2^{\text{bond}}(r_a)} & \dots & e^{-\beta E_2^{\text{bond}}(r_n)} \end{bmatrix} \begin{bmatrix} e^{-\beta E_3^{\text{bond}}(r_1)} & \dots & e^{-\beta E_3^{\text{bond}}(r_a)} & \dots & e^{-\beta E_3^{\text{bond}}(r_n)} \end{bmatrix} \\
&= \begin{bmatrix} e^{-\beta E_1^{\text{bond}}(r_1)} e^{-\beta E_2^{\text{bond}}(r_1)} e^{-\beta E_3^{\text{bond}}(r_1)} & e^{-\beta E_1^{\text{bond}}(r_1)} e^{-\beta E_2^{\text{bond}}(r_1)} e^{-\beta E_3^{\text{bond}}(r_2)} & \dots & e^{-\beta E_1^{\text{bond}}(r_1)} e^{-\beta E_2^{\text{bond}}(r_n)} e^{-\beta E_3^{\text{bond}}(r_n)} \\ e^{-\beta E_1^{\text{bond}}(r_2)} e^{-\beta E_2^{\text{bond}}(r_1)} e^{-\beta E_3^{\text{bond}}(r_1)} & e^{-\beta E_1^{\text{bond}}(r_2)} e^{-\beta E_2^{\text{bond}}(r_1)} e^{-\beta E_3^{\text{bond}}(r_2)} & \dots & e^{-\beta E_1^{\text{bond}}(r_2)} e^{-\beta E_2^{\text{bond}}(r_n)} e^{-\beta E_3^{\text{bond}}(r_n)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta E_1^{\text{bond}}(r_n)} e^{-\beta E_2^{\text{bond}}(r_1)} e^{-\beta E_3^{\text{bond}}(r_1)} & e^{-\beta E_1^{\text{bond}}(r_n)} e^{-\beta E_2^{\text{bond}}(r_1)} e^{-\beta E_3^{\text{bond}}(r_2)} & \dots & e^{-\beta E_1^{\text{bond}}(r_n)} e^{-\beta E_2^{\text{bond}}(r_n)} e^{-\beta E_3^{\text{bond}}(r_n)} \end{bmatrix}
\end{aligned} \tag{13i}$$

r_1 through r_n in Equation 13i indicates the sp^3 carbon- sp^3 carbon bond distance range from 1.72 Å to 1.99Å

The angle and torsion partition functions can be built in a similar fashion:

$$\begin{aligned}
\mathbf{Z}_{\text{butane}}^{\text{angle}} &= \mathbf{z}_1^{\text{angle}} \otimes \mathbf{z}_2^{\text{angle}} \\
&= \begin{bmatrix} e^{-\beta E_1^{\text{angle}}(r_1)} \\ \vdots \\ e^{-\beta E_1^{\text{angle}}(r_a)} \\ \vdots \\ e^{-\beta E_1^{\text{angle}}(r_n)} \end{bmatrix} \begin{bmatrix} e^{-\beta E_2^{\text{angle}}(r_1)} & \dots & e^{-\beta E_2^{\text{angle}}(r_a)} & \dots & e^{-\beta E_2^{\text{angle}}(r_n)} \end{bmatrix} \\
&= \begin{bmatrix} e^{-\beta E_1^{\text{angle}}(r_1)} e^{-\beta E_2^{\text{angle}}(r_1)} & e^{-\beta E_1^{\text{angle}}(r_1)} e^{-\beta E_2^{\text{angle}}(r_2)} & \dots & e^{-\beta E_1^{\text{angle}}(r_1)} e^{-\beta E_2^{\text{angle}}(r_n)} \\ e^{-\beta E_1^{\text{angle}}(r_2)} e^{-\beta E_2^{\text{angle}}(r_1)} & e^{-\beta E_1^{\text{angle}}(r_2)} e^{-\beta E_2^{\text{angle}}(r_2)} & \dots & e^{-\beta E_1^{\text{angle}}(r_2)} e^{-\beta E_2^{\text{angle}}(r_n)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta E_1^{\text{angle}}(r_n)} e^{-\beta E_2^{\text{angle}}(r_1)} & e^{-\beta E_1^{\text{angle}}(r_n)} e^{-\beta E_2^{\text{angle}}(r_2)} & \dots & e^{-\beta E_1^{\text{angle}}(r_n)} e^{-\beta E_2^{\text{angle}}(r_n)} \end{bmatrix}
\end{aligned} \tag{14i}$$

$$\begin{aligned}
\mathbf{Z}_{\text{butane}}^{\text{torsion}} &= \mathbf{z}_1^{\text{torsion}} \\
&= \begin{bmatrix} e^{-\beta E_1^{\text{torsion}}(r_1)} \\ \vdots \\ e^{-\beta E_1^{\text{torsion}}(r_a)} \\ \vdots \\ e^{-\beta E_1^{\text{torsion}}(r_n)} \end{bmatrix}
\end{aligned} \tag{15i}$$

In the model, bonds, angles, torsions and long-range interactions are all formulated using the atom pairwise distance, because (1) the units of the variables all agree with each other in the matrix calculation and (2) databases with the lowest dimension, *i.e.* one dimension for an atom pairwise distance based database, is much easier to manipulate than databases with unmatched dimensions, *e.g.* an angle-based database contains two dimensions with three atoms, and the torsion-based database has three dimensions with four atoms. To avoid this all angles and torsions are all represented by atom pairwise distances, which is described below.

Angles are converted into distances using the law of cosines. In Figure 2, the angle α can be described using the distance x together with d_1 and d_2 using Equation 16i:

$$\alpha = \arccos\left(\frac{d_1^2 + d_2^2 - x^2}{2d_1d_2}\right) \quad (16i)$$

and the angle energy terms are formulated as:

$$E^{angle}(\alpha) = \varepsilon(\alpha - \theta_0)^2 = E^{angle}(r) = \varepsilon\left(\arccos\left(\frac{d_1^2 + d_2^2 - x^2}{2d_1d_2}\right) - \theta_0\right)^2 \quad (17i)$$

Depending on the center atoms that connect the two atoms having an angle interaction, different parameters are assigned to ε , d_1 , d_2 and θ_0 . In butane, the sp^3 carbon-(sp^3 carbon)- sp^3 carbon angle partition function is modeled as:

$$e^{E^{angle}(r)} = e^{-29.90\left(\arccos\left(\frac{4.68-x^2}{4.68}\right)-1.935\right)^2 - 1.327} \quad (18i)$$

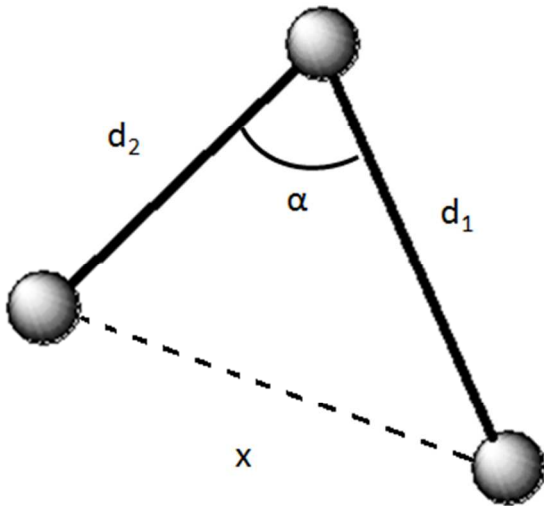


Figure 2. Representation of the angle α using the pairwise distance x .

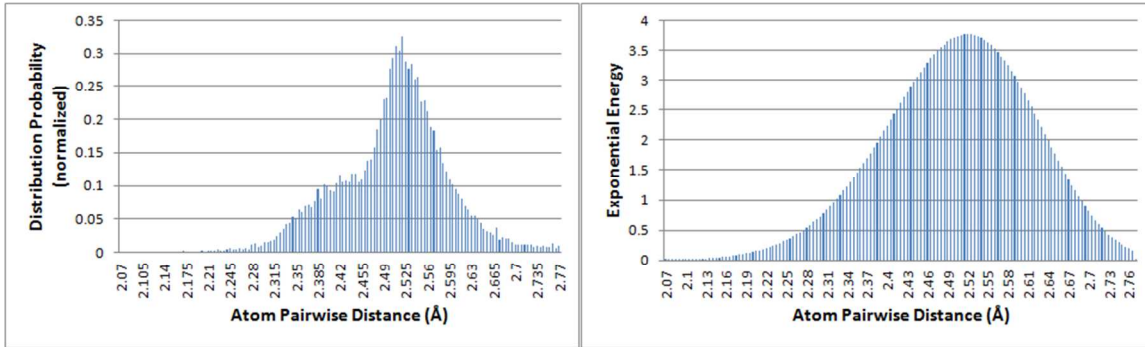


Figure 3. sp^3 carbon-(sp^3 carbon)- sp^3 carbon angle probability distribution and the exponential energy vs. atom pairwise distance.

As shown in Figure 4, the torsion angle α is modeled using an atom pairwise distance x with the help of bond lengths d_1 d_2 d_3 and angles β_1 and β_2 .

$$\alpha = \arccos \left(\frac{d_1^2 \cdot \sin^2(\pi - \beta_1) + d_3^2 \cos^2(\pi - \beta_2) + (d_3 \cos(\pi - \beta_2) + d_2 + d_1 \cos(\pi - \beta_1))^2 - x^2}{2d_1 \sin(\pi - \beta_1) \sqrt{d_3^2 - d_3^2 \cos^2(\pi - \beta_2)}} \right) \quad (19i)$$

and the torsion partition function is modeled as:

$$e^{E^{Torsion}(\alpha)} = \varepsilon_1 (1 + \sin(p_1 \alpha)) + \varepsilon_2 (1 + \sin(p_2 \alpha)) \quad (20i)$$

When used in the sp^3 carbon-(sp^3 carbon)- sp^3 carbon)- sp^3 carbon torsion case, the parameters are given in Equation 21i:

$$e^{E^{Torsion}(\alpha)} = e^{-0.0841(1 + \sin(6.07\alpha)) + 0.0945(1 + \sin(-4.61\alpha))} \quad (21i)$$

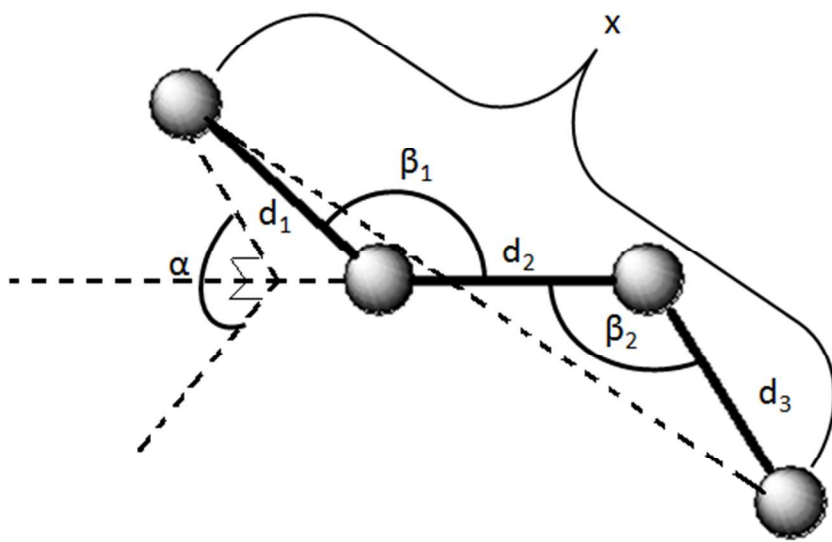


Figure 4. Representation of torsion angle α with the atom pairwise distance x .

The probability distribution and exponential energy vs. atom pairwise distance for the sp^3 carbon-(sp^3 carbon- sp^3 carbon)- sp^3 carbon torsion are shown in Figure 5.

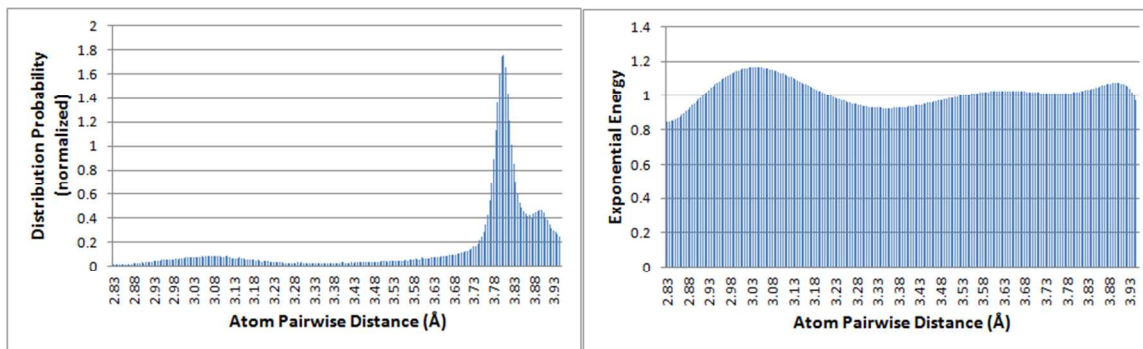


Figure 5. sp^3 carbon-(sp^3 carbon- sp^3 carbon)- sp^3 carbon torsion probability distribution and exponential energy vs. atom pairwise distance.

Hence, the partition functions representing 3 bonds, 2 angles and 1 torsion can be calculated using Equation 13i through 15i.

$$\mathbf{Z}_{\text{butane}} = \mathbf{Z}_{\text{butane}}^{\text{bond}} \otimes \mathbf{Z}_{\text{butane}}^{\text{angle}} \otimes \mathbf{Z}_{\text{butane}}^{\text{torsion}} \quad (22i)$$

Probability matrixes are modeled similarly using:

$$\mathbf{Q}_{\text{butane}} = \mathbf{Q}_{\text{butane}}^{\text{bond}} \otimes \mathbf{Q}_{\text{butane}}^{\text{angle}} \otimes \mathbf{Q}_{\text{butane}}^{\text{torsion}} \otimes \mathbf{Q}_{\text{butane}}^{\text{long-range}} \quad (23i)$$

with probability values gathered from available structural data. The ensemble averaged partition function of butane can then be derived with Equation 24i, including all interaction matrixes defined by \square_{butane} together with four sp^3 carbon single atom energies.

$$\langle e^{E_{\text{butane}}} \rangle = \mathbf{Q}_{\text{butane}} \cdot \mathbf{Z}_{\text{butane}} \cdot e^{4E_{\text{sp}^3 \text{ Carbon}}} \quad (24i)$$

The methane-butane complex partition function is, thereby, modeled as:

$$\mathbf{C}_{1-2} = \mathbf{Q}_{1-2} \cdot \mathbf{Z}_{1-2} \cdot e^{E_{\text{sp}^3 \text{ Carbon}}} \cdot e^{4E_{\text{sp}^3 \text{ Carbon}}} \quad (25i)$$

Methane has only one sp^3 carbon single atom energy partition function and no interaction matrix, while butane has four sp^3 carbon single atom energy partition functions, probability matrix \square_{butane} and an atom pairwise interaction partition function matrix \square_{butane} . The complex probability matrix \square_{1-2} and atom pairwise interaction partition function matrix \square_{1-2} are modeled as:

$$\mathbf{Z}_{1-2} = \mathbf{Z}_{\text{butane}}^{\text{bond}} \otimes \mathbf{Z}_{\text{butane}}^{\text{angle}} \otimes \mathbf{Z}_{\text{butane}}^{\text{torsion}} \otimes \mathbf{Z}_{1-2}^{\text{long-range}} \quad (26i)$$

$$\mathbf{Q}_{1-2} = \mathbf{Q}_{\text{butane}}^{\text{bond}} \otimes \mathbf{Q}_{\text{butane}}^{\text{angle}} \otimes \mathbf{Q}_{\text{butane}}^{\text{torsion}} \otimes \mathbf{Q}_{\text{butane}}^{\text{long-range}} \otimes \mathbf{Q}_{1-2}^{\text{long-range}} \quad (27i)$$

Long-range interactions are separated into van der Waals-electrostatic terms and hydrogen bond terms, the details of how they are modeled and their parameterization are given in the relevant literature.¹

$$\mathbf{Z}_{1-2}^{\text{long-range}} = \mathbf{Z}_{1-2}^{\text{vdw-elec}} \otimes \mathbf{Z}_{1-2}^{\text{H-bond}} \quad (28i)$$

$$\mathbf{Q}_{1-2}^{\text{long-range}} = \mathbf{Q}_{1-2}^{\text{vdw-elec}} \otimes \mathbf{Q}_{1-2}^{\text{H-bond}} \quad (29i)$$

With the necessary components assembled, the gas-phase binding free energy can be derived using Equation 30i.

$$\begin{aligned}
 \Delta G_b^g &= -RT \ln \left[\frac{\langle e^{-\beta E_{1-2}(r)} \rangle}{\langle e^{-\beta E_{\text{butane}}(r)} \rangle \langle e^{-\beta E_{\text{methane}}(r)} \rangle} \right] \\
 &= -RT \ln \left[\frac{\mathbf{Q}_{1-2} \cdot \mathbf{Z}_{1-2} \cdot e^{E_{\text{sp}^3 \text{ Carbon}}} \cdot e^{4E_{\text{sp}^3 \text{ Carbon}}}}{\mathbf{Q}_{\text{butane}} \cdot \mathbf{Z}_{\text{butane}} \cdot e^{4E_{\text{sp}^3 \text{ Carbon}}} \cdot e^{E_{\text{sp}^3 \text{ Carbon}}}} \right] \\
 &= -RT \ln \left[\frac{\mathbf{Q}_{1-2} \cdot \mathbf{Z}_{1-2} \cdot e^{E_{\text{sp}^3 \text{ Carbon}}} \cdot e^{4E_{\text{sp}^3 \text{ Carbon}}}}{\mathbf{Q}_{\text{butane}} \cdot \mathbf{Z}_{\text{butane}} \cdot e^{4E_{\text{sp}^3 \text{ Carbon}}} \cdot e^{E_{\text{sp}^3 \text{ Carbon}}}} \right] \\
 &= -RT \ln \left[\frac{\mathbf{Q}_{1-2} \cdot \mathbf{Z}_{1-2}}{\mathbf{Q}_{\text{butane}} \cdot \mathbf{Z}_{\text{butane}}} \right]
 \end{aligned} \tag{30i}$$

Reference

1. Zheng, Z.; Merz Jr, K. M., Development of the Knowledge-Based and Empirical Combined Scoring Algorithm (KECSA) To Score Protein–Ligand Interactions. *J. Chem. Inf. Model.* **2013**, 53, 1073-1083.