JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Free Energy-Based Conformational Search Algorithm Using the Movable Type Sampling Method

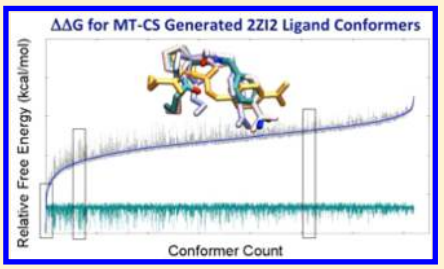
Li-Li Pan,[†] Zheng Zheng,[†] Ting Wang,[†] and Kenneth M. Merz, Jr.*[,†,‡]

[†]Department of Chemistry, Michigan State University, 578 South Shaw Lane, East Lansing, Michigan 48824, United States

[‡]Institute for Cyber Enabled Research, Michigan State University, 567 Wilson Road, Room 1440, East Lansing, Michigan 48824, United States

Supporting Information

ABSTRACT: In this article, we extend the movable type (MT) sampling method to molecular conformational searches (MT-CS) on the free energy surface of the molecule in question. Differing from traditional systematic and stochastic searching algorithms, this method uses Boltzmann energy information to facilitate the selection of the best conformations. The generated ensembles provided good coverage of the available conformational space including available crystal structures. Furthermore, our approach directly provides the solvation free energies and the relative gas and aqueous phase free energies for all generated conformers. The method is validated by a thorough analysis of thrombin ligands as well as against structures extracted from both the Protein Data Bank (PDB) and the Cambridge Structural Database (CSD). An in-depth comparison between OMEGA and MT-CS is presented to illustrate the differences between the two conformational searching strategies, i.e., energy-based versus free energy-based searching. These studies demonstrate that our MT-based ligand conformational search algorithm is a powerful approach to delineate the conformational ensembles of molecular species on free energy surfaces.



1. INTRODUCTION

The conformational search problem has been of interest over the past several decades for the prediction of the conformations of small molecules.[1−3] The molecular conformers generated using these methods are useful for virtual screening,[4−6] 3D-QSAR,[7] pharmacophore modeling,[8−10] protein−ligand binding,[2,11] etc. In light of the importance of conformational searching, a number of algorithms have been developed over the past 20 years.[1,3,12−32] These algorithms efficiently generate realistic molecular conformers according to predefined sets of criteria.

In general, a conformational searching algorithm uses either a systematic or stochastic searching strategy.[33] The systematic search approach exhaustively explores regular variations of the torsion angles.[15,16,19,25] Each systematically generated conformation has its single-point energy calculated typically with a force field. One drawback of systematic searching is that the number of variables dramatically increases with the number of atoms,[33,34] thereby increasing the cost of conformation generation. In a stochastic search algorithm,[12,13] conformations are randomly generated and then minimized using an energy function. Structure generation is terminated once the desired number of structures is generated, the program reaches a predefined stopping point, or all conformers are sampled. Currently, genetic algorithms,[21,25,35−39] Monte Carlo simulated annealing,[18,20,39,40] and distance geometry[15,31,41] methods are three widely used stochastic searching methods. The stochastic search method offers significant computational efficiency when compared to that of traditional systematic search methods. However, the global minimum energy conformer cannot be guaranteed to be identified using stochastic searching methods. Moreover, each conformer generated has to be energy-minimized using classical or quantum mechanical (QM) methods, which adds additional computational expense to the procedure.[42,43]

Hence, locating the key minima (preferably the free energy minima) efficiently while covering a variety of energy states in the output conformations is a challenging task for any conformational searching method. Inspired by a novel energy sampling method, the movable type (MT) sampling method[44,45] devised in our group, we have developed a conformation generation algorithm, the MT conformational search algorithm (MT-CS). The terminology of MT is traced back to an ancient Chinese printing method later pioneered by Gutenberg in the western world. In short, the MT method calculates the partition function of a molecular system within a defined ensemble volume using Monte Carlo integration. To achieve this, MT collects and combines different atom pairwise distances according to a prebuilt pair potential database of atom pairwise energies as a function of distance (r) in terms of bonds, angles, dihedrals, and nonbonded interactions, thereby generating an ensemble that can be used to estimate the free energy for any given structure.[44] For simplicity and due to the limited distance variation, the bond and angle distances for all atom pair types are fixed to their corresponding energy

Received: July 16, 2015
Published: November 10, 2015

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

minimum values. On the basis of their relative strengths, different torsion and nonbonded interactions are given different weights in the selection of distance values during the course of a MT-CS calculation. The MT-CS algorithm translates all collected atom pairwise distances into Cartesian coordinates to yield the final conformations. A standard MT sampling calculation[44] is then performed against each unique conformer to collect an energy ensemble for each conformer. Overall, using the MT-CS approach relevant conformers can be generated rapidly (0.002 s per conformer) and, importantly, on the gas phase free energy surface.

Besides exploring the gas phase free energy surface for a given molecule, we also want to obtain the solvation free energy as well to understand conformational preferences in aqueous solution.[46−50] The further addition of the solvation free energy using the MT method[51] offers another level of sophistication in our conformer generation modeling. Hence, the MT-based conformational search approach offers notable differences over traditional conformer generation approaches.

Our study focuses on the molecular conformation generation of drug-like small molecules on the corresponding free energy (gas and aqueous phases) surface. The MT-CS approach is validated against a variety of drug-like compound structures via the examination of the match between the generated conformations with crystal structure conformations, either in the free or protein-bound state. Two validation benchmarks were studied herein. (1) Validation of the energy/free energy surface calculations was accomplished using nine thrombin inhibitors[52,53] selected from the work of Klebe's laboratory.[54] For the purpose of a more thorough analysis, one of the thrombin inhibitors was selected and analyzed by OMEGA which was then followed by *ab initio* optimization[55] in order to compare with MT-CS following the same quantum-based optimization procedure. Discussion focuses on the difference in the performance of the two conformational searching strategies: energy (using OMEGA) versus free energy-based (using MT-CS). (2) Validation of the robustness of the conformer generation of MT-CS was carried out by examining another 299 organic compounds from the Cambridge Structural Database (CSD)[56,57] and 104 ligands from the Protein Data Bank (PDB). The major difference between PDB and CSD compounds is that CSD compounds are unbound ligands whereas PDB compounds are all protein-bound. In all cases, the gas phase free energy surface was generated and then the individual solvation free energies of each conformer were added to arrive at the predicted global minimum.

2. METHODS

2.1. The Movable Type Sampling Algorithm. The MT sampling method utilizes Monte Carlo integration to estimate the canonical partition function. A molecular energy is sampled by the sum of all of the atom pairwise energies within the canonical ensemble. The MT method uses random matching of distances, yet it avoids random walk sampling with reference to a chemically reasonable initial structure. Importantly, the initial or seed structure provides the atom pairwise contact distances from which MT starts its sampling, thus creating a blurred region surrounding each initial structure.

Against each initial structure, the MT algorithm creates a vector of energies as a function of r for every detected atom pairwise contact, including a series of distance increments/decrements centering at the reference distance from the initial structure. The sampling is performed using a strategy of matrix multiplication through all of the atom pairwise energy vectors in the molecular system, allowing random matching of the atom pairwise energies as a function of distance (r) within a defined volume for each of the initial structures.

When the pair potential sampling is limited to a very narrow range (±0.1 Å for bonds/angles and ±0.5 Å for dihedral and nonbonded contacts deviated from the initial structure), this procedure calculates a component, similar to the vibrational partition function, of the canonical partition function given the pairwise degrees of freedom of each initial structure, yet the MT method calculates the partition function using energy functions for bonds, angles, torsions, and noncovalent interactions instead of the harmonic oscillator approximation used in a vibrational normal-mode analysis. A database is first constructed containing interaction energies between all classes of atom pairs found in the chemical space under investigation as a function of distance. The molecular energy is then calculated simply by assembling atom pairwise energies selected from the database. This database can be generated from any pairwise potential, but, herein, we use the statistical potential KECSA developed in our group.[45]

Formally, the MT sampling method generates extremely unwieldy matrices describing the molecular interactions, but in order to keep the sampling size at a manageable level and to enhance computational speed, we introduced a scramble/tiling strategy for the matrix multiplications used in the atom pairwise energy assembly. First, the vector of each atom pairwise interaction as a function of r is selected from the energy database and the energy cutoff or blurring distance is then applied to this vector. The resultant vector is then scrambled according to the distance distributions. An example of a scrambled atom pairwise energy vector is illustrated in eq 1.

$$\begin{aligned} \mathcal{Z}_\alpha &= \begin{bmatrix} \vdots \\ z_\alpha(r_0 - 2\Delta r) \\ z_\alpha(r_0 - \Delta r) \\ z_\alpha(r_0) \\ z_\alpha(r_0 + \Delta r) \\ z_\alpha(r_0 + 2\Delta r) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ e^{-\beta E_\alpha(r_0 - 2\Delta r)} \\ e^{-\beta E_\alpha(r_0 - \Delta r)} \\ e^{-\beta E_\alpha(r_0)} \\ e^{-\beta E_\alpha(r_0 + \Delta r)} \\ e^{-\beta E_\alpha(r_0 + 2\Delta r)} \\ \vdots \end{bmatrix} \\ \Rightarrow \text{scramble}(\mathcal{Z}_\alpha) &= \begin{bmatrix} z_\alpha(r_0 + a\Delta r) \\ z_\alpha(r_0 + b\Delta r) \\ z_\alpha(r_0 - \Delta r) \\ \vdots \\ z_\alpha(r_0) \end{bmatrix} \end{aligned} \tag{1}$$

where r_0 is the reference distance between atom pair α from the initial structure, Δr is the (constant) step size for the sampling of the pairwise distance, and a and b are arbitrary integers of the multiples of Δr associated with r_0.

The vector goes through multiple different random scramble steps and is then tiled into a fixed size matrix, the Z-matrix (shown in eq 2). In the MT calculation, identically sized Z-matrices are established for all atom pairwise energies in the molecular system under investigation. The molecular energy assembly is performed by pointwise multiplication through all of the Z-matrices in the molecular system (shown in eq 3).

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

$$Z_\alpha = \underbrace{\left.\begin{bmatrix} \text{scramble}(\mathcal{Z}_\alpha)_1 & \text{scramble}(\mathcal{Z}_\alpha)_{i+1} & \cdots & \text{scramble}(\mathcal{Z}_\alpha)_{n-i+1} \\ \text{scramble}(\mathcal{Z}_\alpha)_2 & \text{scramble}(\mathcal{Z}_\alpha)_{i+2} & & \text{scramble}(\mathcal{Z}_\alpha)_{n-i+2} \\ \vdots & \vdots & \ddots & \vdots \\ \text{scramble}(\mathcal{Z}_\alpha)_i & \text{scramble}(\mathcal{Z}_\alpha)_{2i} & \cdots & \text{scramble}(\mathcal{Z}_\alpha)_n \end{bmatrix}\right\} m \text{ rows}}_{n \text{ columns}} \quad (2)$$

$$Z_M = Z_1 \cdot Z_2 \cdot \ldots \cdot Z_n \quad (3)$$

The random disordered permutations to each Z-matrix are meant to maximize the variety of energy combinations at different distances, and the fixed-size matrix multiplication is to maintain a computationally acceptable sampling size. The row number m of the Z-matrices is defined as the least multiplier of all of the atom pairwise vector sizes in the molecular system under study, in order to fit all atom pairwise energy vectors with different distance distribution ranges, whereas the column number n is a user defined number defining the sampling size. Hence, the final Z-matrix after assembly includes $m \times n$ molecular conformational energies.

In the Z-matrix, atom pairwise energies at different distances are assembled to simultaneously represent different energy states in the canonical ensemble resulting from sampling against a single seed or initial structure. Due to the very narrow sampling range set for each atom pairwise distance in this study, the ensemble volume is approximated as independent of the sampled energies. Please see ref 44 for a detailed description of the evaluation of the sampling volume. The free energy is then calculated by incorporating all the Z-matrices into

$$A = -RT \ln[Q_M] \approx -RT \ln[V_M \langle e^{-\beta E_M(\tau)} \rangle] = -RT \ln\left[V_M \frac{\sum_i^N Z_M^i}{N \times m \times n}\right] \quad (4)$$

where $E_M(\tau)$ is the molecular energy as a function of the geometric variable τ, V_M indicates the sampling volume, and N is the number of initial structures.

The free energy is computed directly from the NVT ensemble, thus avoiding issues related to the additivity of the free energy.[54] In particular, we assemble the interaction energies using eq 3 and then place this into eq 4 to directly compute the free energy, thereby affording the ability to avoid issues related to the decomposition of the free energy into entropic and enthalpic components. This is a real advantage of the MT method relative to other conformational sampling approaches.

The MT energy sampling method incorporates our newly developed implicit water model for the solvation free energy calculation.[51] In this model, the water molecules are modeled as isotropic rigid balls with van der Waals radii of 1.6 Å and placed into isometric solute-surrounding solvent layers. The solute−solvent interaction sampling starts from the solute's water accessible surface until 8 Å away from the solute's van der Waals surface using an increment of 0.005 Å per layer. The number of water molecules was limited by comparing their maximum cross-sectional areas with the solvent accessible surface area at each solvent layer for each atom in the solute. The number of water molecules (N_w) accessible to each atom at distance R away from the atomic center of mass is rounded down using the maximum cross-sectional area (S_w) of water and the atomic solvent accessible surface area (S_a) in the solvent layer at distance R.

$$N_w(r) = \text{floor}\left(\frac{S_a(r)}{S_w}\right) \quad (5)$$

The maximum cross-sectional areas (S_w) of a water molecule is calculated as

$$S_w = \int_{\pi/2-\theta}^{\pi/2} 2\pi(R_a + R_w)R_w \sin\left(\frac{\pi-\theta}{2}\right) d\left(\frac{\pi}{2} - \theta\right) = 2\pi(R_a + R_w)R_w \cos\left(\frac{\pi-\theta}{2}\right) \quad (6)$$

where R_w and R_a are the van der Waals radii for water and the atom in the solute molecule, respectively.

The Boltzmann factor matrix for the kth solute atom−water ($\mathcal{Z}_k^{A-S}$) interaction is defined as a Boltzmann-weighted solute atom−water energy multiplied by the number of accessible water molecules at the different distances. Multiplication of the Z-matrices for all solute atom−water interactions composes the final solute molecule−water Z-matrix ($\mathcal{Z}_{total}^{L-S}$), which when multiplied by the Z-matrix for the intrasolute molecular interactions ($\mathcal{Z}_{total}^{L}$) derives the final Z-matrix for the solute−solvent complex system ($\mathcal{Z}_{total}^{LS}$). Multiplication of the final Z-matrix with its corresponding normalized Q-matrix generates the Boltzmann-weighted energy ensemble ($\mathcal{C}_{total}^{LS}$). With the energy ensembles for the solute molecule ($\mathcal{C}_{total}^{L}$) and solute−solvent complex ($\mathcal{C}_{total}^{LS}$), the solvation free energy is calculated using eq 10.

$$\mathcal{Z}_k^{A-S} = \begin{bmatrix} e^{-\beta E_k^{A-S}(r_1)N_w(r_1)} & e^{-\beta E_k^{A-S}(r_{i+1})N_w(r_{i+1})} & \cdots & e^{-\beta E_k^{A-S}(r_{n-i+1})N_w(r_{n-i+1})} \\ e^{-\beta E_k^{A-S}(r_2)N_w(r_2)} & e^{-\beta E_k^{A-S}(r_{i+2})N_w(r_{i+2})} & \cdots & e^{-\beta E_k^{A-S}(r_{n-i+2})N_w(r_{n-i+2})} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta E_k^{A-S}(r_i)N_w(r_i)} & e^{-\beta E_k^{A-S}(r_j)N_w(r_j)} & \cdots & e^{-\beta E_k^{A-S}(r_n)N_w(r_n)} \end{bmatrix} \quad (7)$$

$$\mathcal{Z}_{total}^{L-S} = \text{disorder}(\mathcal{Z}_1^{A-S}) \cdot \text{disorder}(\mathcal{Z}_2^{A-S}) \cdot \ldots \cdot \text{disorder}(\mathcal{Z}_n^{A-S}) \quad (8)$$

$$\mathcal{C}_{total}^{LS} = \overline{\mathcal{Q}}_{total}^{LS} \cdot \mathcal{Z}_{total}^{LS} = \overline{\mathcal{Q}}_{total}^{L-S} \cdot \overline{\mathcal{Q}}_{total}^{L} \cdot \mathcal{Z}_{total}^{L-S} \cdot \mathcal{Z}_{total}^{L} \quad (9)$$

$$\Delta G_{solv}^{L} \approx -RT \ln\left[\frac{Z_{LS}}{Z_L}\right] = -RT \ln\left[\frac{\langle e^{-\beta E_{LS}(r)} \rangle}{\langle e^{-\beta E_L(r)} \rangle}\right] = -RT \ln\left[\frac{\text{sum}(\mathcal{C}_{total}^{LS})}{\text{sum}(\mathcal{C}_{total}^{L})}\right] = -RT \ln\left[\frac{\text{sum}(\overline{\mathcal{Q}}_{total}^{LS}\mathcal{Z}_{total}^{LS})}{\text{sum}(\overline{\mathcal{Q}}_{total}^{L}\mathcal{Z}_{total}^{L})}\right] \quad (10)$$

The generation of initial structures plays an essential role in the determination of the free energy using the MT sampling method by adding more components to the MT simulation of the canonical partition function. On the other hand, by sampling the localized partition function of each conformer, the MT method provides the ability to carry out a conformational search on the free energy landscape.

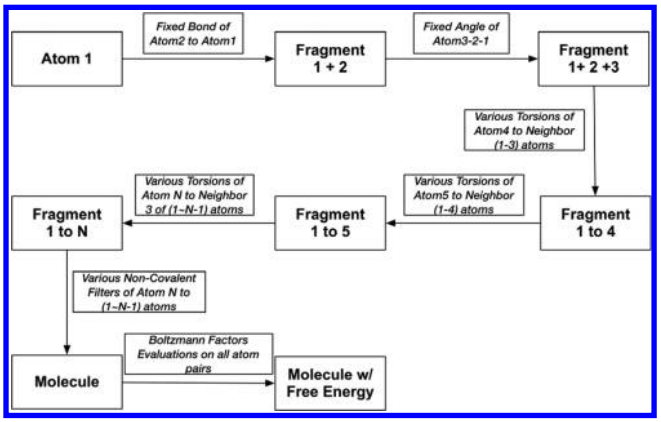
2.2. Generating Conformations. *2.2.1. Database Preparation for Bonds, Angles, Torsion Angles, and Noncovalent Interactions.* There is one database used by the MT sampling algorithm: the pairwise distance-dependent Boltzmann ener-

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

gies. All bonds and angles were fixed at the distances associated with their minimum energies (i.e., strongest bond or angle preference), whereas for the torsion angles, distances were selected based on the Boltzmann energy curves as a function of distance. Noncovalent interactions were evaluated after conformation generation via the torsion angles, and ring structures were treated as fixed units.

2.2.2. Torsion Angle Driven 3D Structure Generation. The MT conformational search program requires an initial *mol2* input file with complete atom connection information. Torsion angle parameters for every rotatable bond were used to generate the different conformations according to their local minimum energies. The structure generation started with the first atom placed at a fixed position, and the molecule was built from there by adding one atom (or a fixed ring unit) to the system at a time. The molecular construction details are given in Scheme 1. In this way, an ensemble of molecule conformers

Scheme 1. Conformer Generation Workflow



was generated that represented the conformational space favored by the torsion angles. The torsion energies were then recorded for each conformation. Simultaneously, all bond and angle energies were also recorded for future reference.

2.2.3. Filtering Conformers and Free Energy Determination. The initial ensemble of conformations was further screened to eliminate those with internal clashes. Each member of the final conformer ensemble had its associated free energy computed using the full set of Boltzmann factors (eq 4). The conformation that had the lowest gas phase free energy was selected as the reference (zero free energy) to obtain the relative gas phase free energies for the remaining conformations in the ensemble.

Each member of the resultant ensemble was then solvated and had its solvation free energy determined using KMTISM.[51] Adding these free energies to the gas phase free energies yields the relative aqueous phase free energy. Note that both the gas and aqueous phase free energies are relative values, whereas the solvation free energy is an absolute value.

2.3. Small Molecules from the CSD and PDB. We selected 299 molecules from the CSD, which represented a culled subset (299 out of 492) of molecules examined in a validation study of OMEGA.[19] The CSD ID codes for the selected molecules are given in the Supporting Information. The 299 molecules were selected to have *R*-factors smaller than 10.0, with at least four rotatable bonds and a heavy atom count of larger than 10. Furthermore, we also examined 104 molecules out of 197 PDB ligands explored for validation previously. For each molecule, thousands to tens of thousands of conformers were generated depending on the exact nature of the individual molecules. The resultant conformations were then analyzed relative to the original CSD and PDB structures.

2.4. Thrombin Ligands. Besides the 299 CSD and 104 PDB structures, nine thrombin ligands from Klebe's work were examined in more detail (PDB ID's 2ZFP, 2ZGB, 2ZC9, 2ZI2, 2ZIQ, 2ZHQ, 2ZGX, 2ZNK, and 2ZDA). Moreover, we picked the 2ZI2 thrombin ligand (3c) out of the nine total ligands to compare to an OMEGA conformational search and to further calculate the conformer energies at the M06-2X/6-31G* level of theory (using Gaussian 09[55]). The resultant global minimum and local energy minima were compared to the conformations generated by MT conformational search both structurally and thermodynamically. The absolute free energies for the DFT-optimized structures were obtained from a vibrational analysis carried out using the Gaussian 09 package.

We also carried out MM-PB/GBSA calculations on the nine thrombin ligands to obtain their solvation free energies. The nine ligand structures stripped from their respective PDB files were optimized at the M06-2X/6-31G* in Gaussian09 prior to a 5 ns MD simulation with AMBER12 using the general AMBER force field (GAFF).[66] Trajectory data points were collected every 1 ps, resulting in 5000 frames for later analysis. These 5000 frames were put into the MM-PBSA.py program[58] in AMBERTools for post-MD analysis. The solvation free energies using both the linearized Poisson−Boltzmann (MM-PB/SA)[59,60] and generalized Born (MM-GB/SA)[61] models were calculated and stored for later comparison to MT results.

The MT conformational sampling code was written in Python (version 2.7), and the Visual Molecular Dynamics (VMD)[62,63] program was used for structure analysis. GNUPlot[64] was used for all plots.

3. RESULTS AND DISCUSSION

We generated over 1 458 000 total conformers for the 299 CSD, 104 PDB, and nine thrombin structures according to the criteria delineated in the Methods section. The conformational ensembles were compared to their original crystal structures and, in the case of one of the thrombin inhibitors, to an OMEGA conformational search followed by M06-2X/6-31G* minimization.

3.1. Structure Validation of the Thrombin Ligands. The thrombin ligands represented a typical challenge for conformational searching programs. In generating these conformational ensembles, one must overcome several challenges including chirality, ring formation, and intramolecular hydrogen bonding. Below, we discuss, in detail, structural RMSDs, free energies, and the global minimum structure generated using MT sampling for 3c (PDB ID 2ZI2) as a representative example.

3.1.1. The Conformational Ensemble. We generated over 78 000 conformers for the nine thrombin inhibitors that were structurally and thermodynamically characterized by Klebe and co-workers (Figure 1; we have retained their original numbering).[54]

The small molecules in the thrombin ligand set have between 8 and 11 rotatable bonds that need to be considered in the MT conformational search. Every single torsion angle (rotatable bond) has 2−4 local minimum energy values from which we can set the torsion angle. Thus, this could lead to 2×2^8 to 2×4^{11} total conformations, which is reduced to 1800 to 31 000 because of nonphysical close contacts, which were eliminated.

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

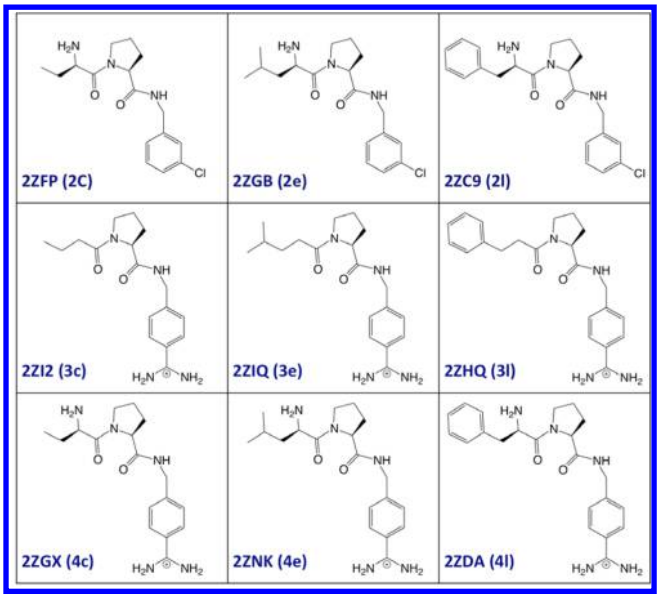


Figure 1. Thrombin ligands from the work of Klebe and co-workers.[54] For 2c, 2e, 2l, 4c, 4e, and 4l, the $-NH_2$ group at low pH values would become $-NH_3^+$.

Our choice of the nonbonded distance criteria (~2.8 Å) eliminates unrealistic MT conformations.

Chirality considerations also reduce the conformational space examined, and this is also true for the thrombin ligands. These ligands all have a carbon atom with the S configuration, which connects the amide to the five-member ring (Figure 1). This S-chirality eliminates all of the conformations with R stereochemistry. Furthermore, six ligands (2c, 2e, 2l, 4c, 4e, and 4l) have a carbon atom with R stereochemistry where the amino (ammonium) group binds (Figure 1), which eliminates all conformations with S stereochemistry. Overall, through various cutoffs and stereochemical considerations, the resultant sizes of the thrombin inhibitors ensembles were of manageable size (Table 1).

Table 1. MT Conformational Ensembles and the Associated RMSDs (in Å) for the Thrombin Ligands of Figure 1

	conformer amount	mean RMSD	min. RMSD	first aq. free energy RMSD	first gas free energy RMSD
2c	2010	1.521	0.660	1.613	2.165
2e	2132	1.604	0.746	1.128	2.248
2l	6643	1.803	0.738	1.575	2.156
3c	7201	1.723	0.866	2.300	1.985
3e	18 728	1.821	0.990	1.578	2.628
3l	31 747	2.094	0.945	2.794	2.644
4c	1826	1.663	0.904	1.231	1.656
4e	1950	1.660	0.940	1.804	1.767
4l	6079	1.909	0.906	1.985	2.452

3c (2ZI2) is a typical ligand among the nine thrombin ligands. It has one five-member ring with one S-chiral carbon atom, one potential intramolecular hydrogen bond, and one benzamidine functional group. The 10 rotatable bonds have 3–5 energy maxima, which result in between $(3 \times 2)^{10}$ and $(5 \times 2)^{10}$ unique conformations (each choice for a rotatable bond generates two positions for the targeted atom), but this is reduced to 7201 ensemble members by our filters (Table 1). All conformations were generated within 2 min, including their gas phase free energies and RMSDs. The subsequent solvation calculations take ~3 s per conformation on a laptop computer.

3.1.2. Structure Validation. Table 1 shows the RMSD results for the conformations generated for the nine ligands compared to their corresponding protein-bound PDB structures. Here, we provide not only the minimum RMSDs in the conformational ensembles but also the RMSDs for the lowest free energy structures, both in the gas and aqueous phases. From the table, we can see that the MT conformational search approach yields geometries with average RMSDs between 1.521 and 2.094 Å and minimum RMSDs between 0.660 and 0.990 Å. Figure 2 summarizes the minimum RMSD

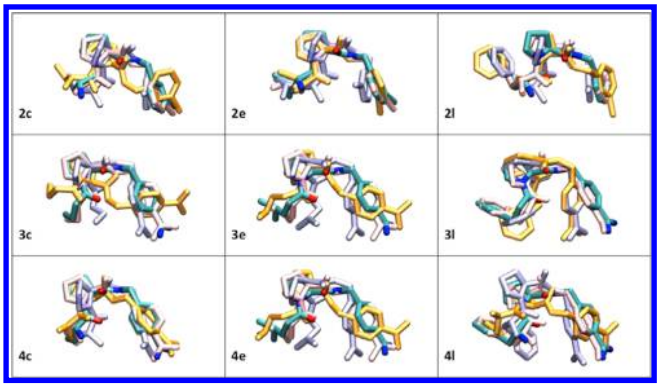


Figure 2. Geometry comparisons for the nine thrombin ligands. Multicolor, PDB structures; pink, minimum RMSD conformation; blue, lowest gas phase free energy conformation; orange, lowest aqueous phase free energy conformation. All conformations are aligned to their corresponding PDB structures.

conformations (pink), the lowest gas phase free energy conformations (blue), and the lowest aqueous phase free energy conformations (orange), with all structures aligned to the corresponding PDB crystal structure (multicolored). The RMSDs for the aqueous and gas phase lowest free energy conformers are also listed in the table, with their values ranging from 1.128 to 2.794 Å for the former and from 1.656 to 2.644 Å for the latter. The increased RMSDs for the gas and aqueous phase structures arise because these are free structures rather that protein-bound. The detailed RMSD curves are shown in Supporting Information Figure S1.

Figure 3 shows the RMSD curves for 3c associated with its three key conformers. From this figure, we observe that the RMSD range for 3c is 0.866–2.679 Å. The minimum RMSD conformation has a 0.866 Å RMSD value, and the minimum gas phase structure is 1.985 Å, with the main geometric differences being found in the five-member ring. The aqueous phase structure has a 2.300 Å RMSD and differs mostly in the region connecting the five- and six-membered rings. The lower right corner of this figure gives an overlay of the four different structures.

3.1.3. Free Energies. We obtained the gas phase free energies initially for every single conformer when doing the conformation generation. Subsequently, we obtained their relevant solvation free energies via postconformational search calculations on the corresponding geometries, as shown in Supporting Information Figure S2. We report calculations both on the protonated ($R-NH_3^+$) and unprotonated ($R-NH_2$) amine groups, and these results are shown in Table 2. It is not surprising that the resultant solvation free energies are more negative for the structures (2c, 2e, 2l, 4c, 4e, and 4l) with

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853–5864

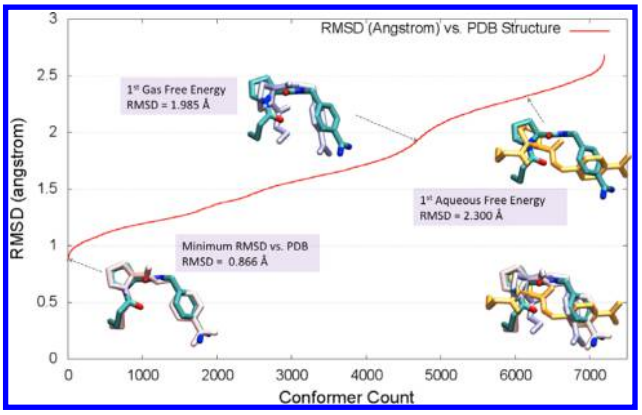


Figure 3. RMSD ranges for thrombin ligand 3c. The purple shadowed text identifies the three key conformations and their RMSD values. Here, the multicolored molecule is the PDB structure and the single colored molecules are MT generated conformations: purple, minimum RMSD conformation; blue, lowest gas phase free energy conformation; orange, lowest aqueous phase free energy conformation. All conformations are aligned to the PDB structure.

Table 2. Solvation Free Energies for the Thrombin Ligands, Including MT Solvation Ranges, MT Averages, and MM-PBSA Averages

	MT average		MM-PBSA average	
free energy (kcal/mol)	N3[a]	N4[b]	N3[a]	N4[b]
2c	−10.74	−39.02	−21.72	−74.98
2e	−10.08	−35.58	−23.68	−73.94
2l	−11.45	−45.24	−25.01	−74.20
3c	−13.71	N/A	−63.73	N/A
3e	−12.91	N/A	−63.79	N/A
3l	−14.49	N/A	−65.95	N/A
4c	−18.49	−86.25	−70.53	−168.33
4e	−17.71	−82.51	−67.73	−167.02
4l	−18.95	−91.21	−71.27	−163.88

[a]N3 indicates the $-NH_2$ functional group for 2c, 2e, 2l, 4c, 4e, and 4l. N3 is not applicable for 3c, 3e, and 3l since this functional group is absent. [b]N4 indicates the $-NH_3^+$ functional group for 2c, 2e, 2l, 4c, 4e, and 4l. N4 is not applicable for 3c, 3e, and 3l since this functional group is absent.

the $-NH_3^+$ functional group than those with the $-NH_2$ functional group. Moreover, 4c, 4e, and 4l are better solvated than 2c, 2e, and 2l because the former are singly charged (benzamidine groups) and the latter set are neutral (Figure 1). The negative solvation free energies for 3c, 3e, and 3l arise due to the positive charge on the benzamidine moiety.

We also carried out MM-PBSA calculations to further analyze our solvation free energy results. The MM-PBSA results are also shown in Table 2. The MM-PBSA values are more negative than our average values, which was the trend we observed in a detailed analysis of the MT, MM-PB/SA, and MM-GB/SA solvation methods.[51] In this article, we further compared our MT solvation free energy results with the MM-PBSA results, with resultant Pearson's correlation coefficients of 0.846 for the N3 sets and 0.956 for the N4 sets. The plot is shown in Supporting Information Figure S3. Hence, we demonstrate that, although we could not directly compare our solvation free energies of the thrombin inhibitors with the MM-PBSA results because of the latter's tendency to give more negative solvation free energies, the trends in solvation free energy match reasonably well.

To further analyze the nine ligands, we also computed the absolute free energies from the conformational ensembles coupled with eq 4. Using the MT approach, this is possible to do, and, for example, on going from 2c → 2e, the gas phase free energy change arises from 2c + 2 methane C atoms → 2e. Moreover, this allows us to obtain the free energy changes among the nine thrombin ligands, as shown in Figure 4, in both

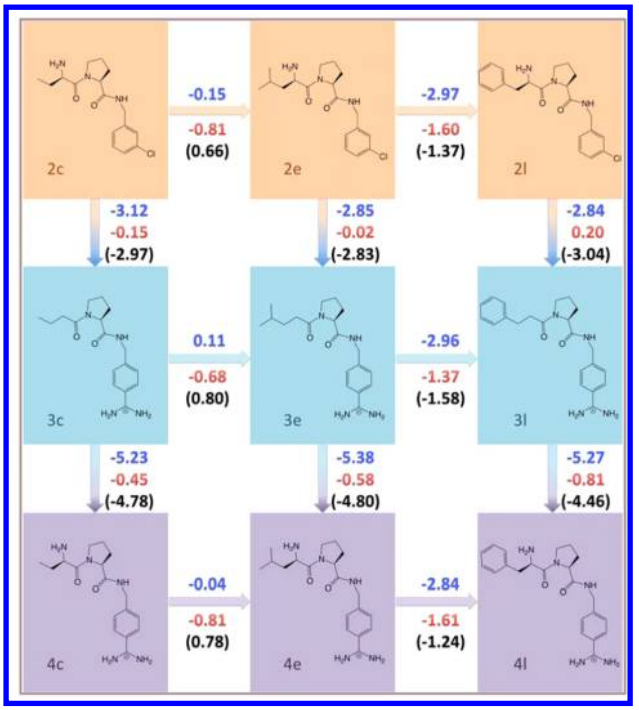


Figure 4. Free energy perturbations (FEPs) among the nine thrombin ligands calculated by MT. Colored values are the free energy changes (in kcal/mol) under aqueous phase (blue) and gas phase (red) conditions. The values in black parentheses are the changes in solvation free energies.

solution and the gas phase. Here, we find that the free energy differences are larger in solution than in the gas phase due to changes in the solvation free energies. Again, using 2c → 2e as an example, we gain free energy from C−C bond formation but pay an entropic penalty when 2 free methane carbon atoms attach to 2c to form 2e. The solvation free energy is slightly unfavorable (0.66 kcal/mol) as well. Using the data from Figure 4, similar thermodynamic analyses can be performed.

We also plot the aqueous phase free energies as well as the gas phase free energies for the nine ligands in Supporting Information Figure S4. Herein, we discuss only 3c as a representative example, focusing on the analysis of the relative free energy changes in different phases (gas phase, aqueous phase, and the solvation free energy change from the gas to aqueous phase). In Figure 5, in order to include all three free energy distributions in one figure with the proper scaling, we fix the individual lowest free energy values in all three distributions at 0 kcal/mol and convert all of the conformational free energies to relative values ($\Delta\Delta G$(aq), $\Delta\Delta G$(gas), and $\Delta\Delta G$(sol)) referenced to their individual lowest free energy conformations. In this case, we are looking at the free energy changes in each phase instead of comparing the free energy values between phases with respect to each conformation.

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

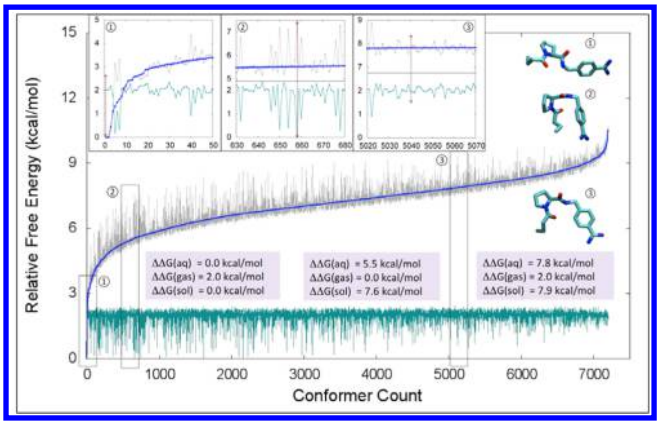


Figure 5. Relative free energies vs conformation count for thrombin ligand 3c. Blue curve, aqueous phase free energy; gray curve, solvation free energy; cyan curve, gas phase free energy. The structures are as follows: (1) lowest aqueous phase free energy conformation, (2) lowest gas phase free energy conformation, and (3) minimum RMSD conformation compared with the PDB structure. The small windows are enlarged versions of the free energy for every 50 conformers, where conformers 1−3 are included. The pink shadowed text, from left to right, represent the relative aqueous phase free energy $\Delta\Delta G$(aq), relative gas phase free energy $\Delta\Delta G$(gas), and relative solvation free energy $\Delta\Delta G$(sol) for the three conformers (1−3), respectively. Note that $\Delta\Delta G$(aq), $\Delta\Delta G$(gas), and $\Delta\Delta G$(sol) do not have direct relationship with each other. They are only relative to their lowest aqueous, gas, and solvation free energy conformations, respectively.

Figure 5 shows the aqueous phase relative free energy curve (blue, lowest to highest free energy) with their corresponding gas phase relative free energies (cyan) and solvation relative free energies (gray). Prior to conversion to relative values, all free energies, including the aqueous phase, gas phase, and solvation free energies, are all originally determined as absolute values with the gas phase G(gas) and solvation G(sol) terms summing to the aqueous phase G(aq) = G(gas) + G(sol) term.

In this relative aqueous phase free energy curve, the range is 0.00−10.51 kcal/mol; interestingly, the relative solvation free energy curve tracks the aqueous phase free energy curve, with the relative solvation free energy in the range of 0.00−10.74 kcal/mol and the relative gas phase free energy in the range of 0.00−2.44 kcal/mol, indicating that the aqueous phase conformer sequence is driven by the solvation free energy. For example, the first ranked structure in the aqueous phase is the top-ranked conformation in terms of its solvation free energy. Thus, in the case of 3c (in fact, also for the remaining eight thrombin ligands; see Supporting Information Figure S4), the gas phase free energy only slightly modifies the relative stability of the conformations contained in the ensemble. This resultant plot has an S shape, where the first 50 conformers (Figure 5.1) have a sharp free energy increase (0.00 to 3.41 kcal/mol in the aqueous phase), whereas the middle two sets of 50 conformers (Figure 5.2 and 5.3) have a modest free energy increase (5.48 to 5.57 kcal/mol and 7.82 to 7.84 kcal/mol, respectively). Potential structural similarities are excluded, where each of the middle two sets of 50 conformers has different aqueous and gas phase free energies, as shown in the figure. Overall, the top-ranked aqueous phase conformation is favored because of both its favorable solvation free energy (0.0 kcal/mol) and gas phase free energy (2.0 kcal/mol). The top-ranked gas phase conformation has 5.5 kcal/mol aqueous phase free energy and a 0.0 kcal/mol gas phase free energy with a solvation free energy of 7.6 kcal/mol. The minimum RMSD conformation has a less favorable solvation free energy (7.9 kcal/mol), but it has a relatively unfavorable gas phase free energy (7.8 kcal/mol), leading to its position in the aqueous phase free energy plot.

3.2. DFT Validation. We also compared the MT-generated conformations with DFT-optimized local minimum energy structures for 3c. First, we used OMEGA[19,65] to independently generate 341 conformations for 3c and then geometry-optimized these conformations at the M06-2X/6-31G* level of theory using Gaussian 09. From this, we obtained eight low-

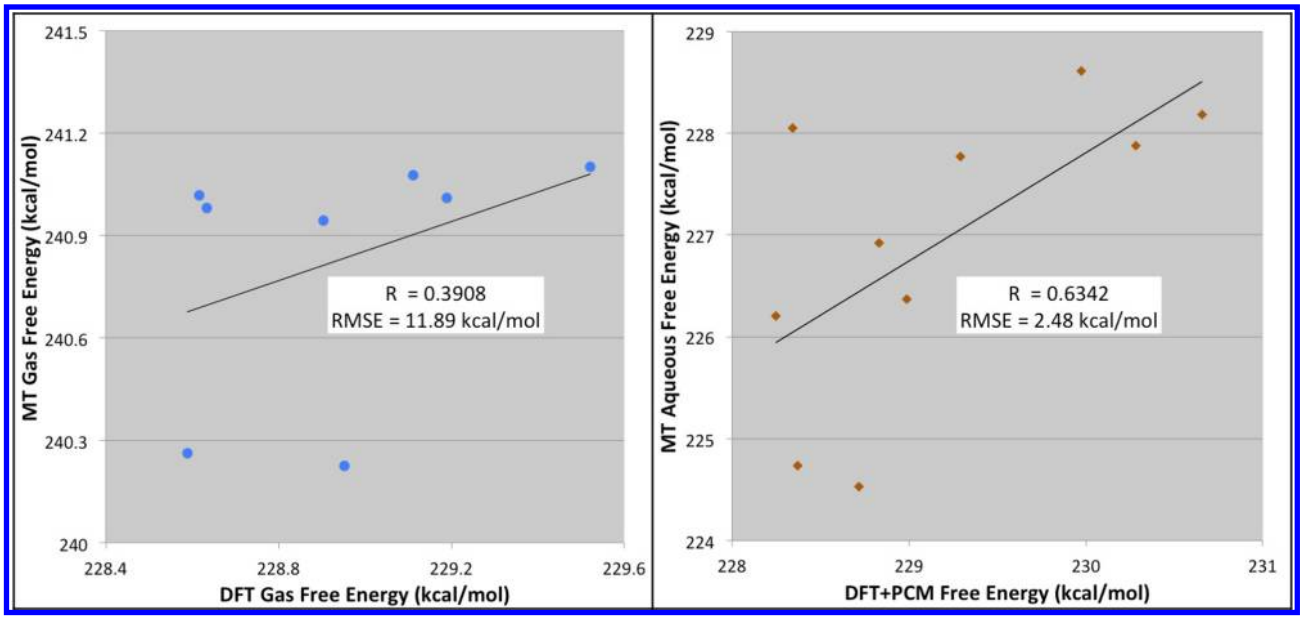


Figure 6. Comparison of MT-CS and DFT calculated free energy values for the eight gas phase (left) and 10 PCM model (right) DFT conformations. The dots in each panel represent the positions of these conformations, and the line represents a linear fit. Pearson's R values are given on the plots. The root-mean-squared error (RMSE) is also included in the graph.

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

energy structures in the gas phase and 10 using the PCM model. We then used a vibrational analysis (as implemented in Gaussian 09) to obtain the absolute free energies for the DFT-generated conformations for further comparison to MT. We then used our MT method to estimate the absolute free energy values for all of the MT-generated conformations using eq 4 for a head-to-head absolute free energy comparison. Figure 6 shows the comparison of the MT vs DFT derived free energies, in both the gas and aqueous phases. The resultant DFT free energies range from 228.589 to 229.521 kcal/mol for the gas phase and 228.248 to 230.655 kcal/mol for the PCM model, whereas the MT free energy for these conformations ranges from 240.226 to 241.102 kcal/mol for the gas phase and 224.526 to 228.609 kcal/mol for solution. Details are shown in Supporting Information Tables S1 and S2. The values of Pearson's correlation are shown in Figure 6. The aqueous phase gives the best match between the MT and G09 absolute free energies with a correlation of 0.6342, indicating that the MT and DFT absolute free energies at the same structure are in qualitative agreement. The gas phase gives a poorer match, with a correlation of 0.3908 between the MT and DFT results.

The 18 DFT-determined absolute free energy values were then compared to the MT absolute free energies, and the comparison is shown in Figure 7. We find that the eight gas phase and 10 PCM model structures have absolute free energy values that are on the same scale as our MT-generated absolute free energy conformational curves. Table 3 shows the RMSD results for the 18 lowest energy structures from the DFT calculations when compared to their closest partner in terms of free energy. In the table, we list RMSD values for the M06-2X/6-31G* conformation vs the nearest (by free energy) MT conformation. We also visually compare the geometries in Supporting Information Figure S5. A geometric comparison of the DFT and MT conformations with the PDB structure is given in Supporting Information Table S1 as well. We find that the RMSDs between the DFT conformations and the closest in free energy MT conformations range from 0.754 to 2.599 Å, whereas the lowest MT conformation vs lowest DFT one had RMSDs of 1.509 Å for the gas phase and 1.656 Å for the aqueous phase.

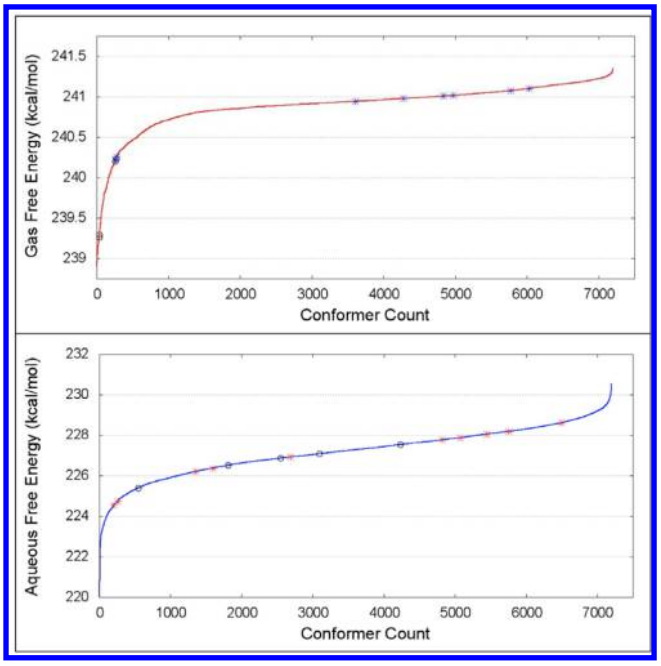


Figure 7. Free energy for DFT-optimized 3c conformers vs MT-calculated conformers in both the gas (upper) and aqueous (lower) phases. The curves represent the free energy values for MT-CS-generated conformers, and the points on the curve represent the positions of the eight gas phase (upper) and 10 PCM model (lower) DFT conformations on the MT free energy curves. Note that the free energies are absolute values. Black circles represent MT-calculated free energy positions of the DFT energy-minimized lowest five gas/aqueous MT conformers.

Table 3. Geometric Comparisons of M06-2X/6-31G* (DFT)-Generated Conformations and MT-Generated Conformations[a]

PCM phase rank	DFT vs MT	gas phase rank	DFT vs MT
1	1.613	1	1.977
2	2.003	2	1.480
3	1.713	3	1.535
4	2.495	4	1.905
5	1.639	5	1.351
6	0.482	6	1.226
7	1.292	7	1.211
8	2.599	8	1.528
9	1.487		
10	0.754		

[a]RMSD values in Å.

To further study the relationship of energy-minimized conformations to their corresponding MT free energies, we took the five lowest gas and aqueous phase free energy conformations from our MT studies and then optimized them at the M06-2X/6-31G(d) level of theory. We then took these 10 DFT energy-minimized MT (DFT-MT) conformations and calculated their new gas and aqueous phase MT-derived free energies. The results are shown also in Figure 7 as black circles. We find that the five gas phase conformations generated three novel conformations in the gas phase and that the five aqueous ones produced five in the aqueous phase. The aligned conformations are shown in Supporting Information Figure S6.

From Figure 7, we observe that in the gas phase $^{2}/_{3}$ of our MT-DFT optimized conformations have lower free energies than the eight OMEGA-DFT-optimized conformations, whereas the remainder ($^{1}/_{3}$) have a free energy values close to the OMEGA-DFT lowest free energy conformation. In the aqueous phase, the MT-DFT lowest free energy conformation has a higher free energy value than the OMEGA-DFT one, whereas the remainder ($^{4}/_{5}$) were distributed among the remaining 10 OMEGA-DFT conformations. Detailed values are shown in Table 4 with their related absolute MT free energies. In this table, the trends of the absolute MT free energy values are consistent with those of the DFT electronic energy values in both the gas and aqueous phases. Note that in the gas phase numbers 3−5 are nearly geometrically identical, so they have close MT free energies and DFT electronic energies.

What is more, the gas phase DFT energy-minimized conformations (eight total) did not yield low free energy conformations along the MT free energy curve (Figure 7, top panel). Therefore, in order to analyze this further, we examined the individual contributions of the entropy and enthalpy. In the DFT determination of the Gibbs free energy, the thermal energy and entropy at 298.15 K are each computed, allowing us to investigate the contributions of enthalpy and entropy. Hence, we performed another series of linear correlations

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

Table 4. Calculated Relative DFT Energies and Absolute MT Free Energies (in kcal/mol) for the Top Five Lowest Free Energy Gas and Aqueous MT Conformations Further Optimized Using Gas and Aqueous Phase (PCM) DFT Calculations[a]

gas phase			aqueous phase		
ranking	relative DFT electronic energy (kcal/mol)	absolute MT gas phase free energy (kcal/mol)	ranking	relative DFT electronic energy (kcal/mol)	absolute MT aqueous free energy (kcal/mol)
1	0.000	239.262	1	0.000	225.393
2	1.031	239.294	2	1.069	226.863
3	3.421	240.240	3	3.084	227.533
4	3.428	240.240	4	3.254	227.077
5	3.446	240.203	5	3.374	226.511

[a]Note here that the DFT energies are relative to the lowest energy conformation.

comparing *H* vs *G* and *TS* vs *G*, which is shown in Figure 8. Detailed *H*, *S*, and *G* values are shown in Supporting Information Table S3. From this, we observe that the *TS* vs *G* graph (right) has a high Pearson's *R* value of 0.9155, whereas the *H* vs *G* graph (left) has a lower Pearson's *R* value of 0.5912, implying that that the *TS* term has a higher influence than the *H* term in determining the energy landscape of the 3c conformations. That means that the 3c molecule has an entropy-driven free energy landscape.

It is this latter observation that, in part, explains why the OMEGA/DFT-derived conformations appear well above the minimum of the MT-derived free energy landscape: the DFT conformations are obtained via energy minimization calculations, not an optimization on the free energy surface. From the latter analysis, we find that the preferred low free energy conformations for 3c arise due to entropy effects, and MT conformation generation is done on the free energy surface (both entropy and enthalpy (energy)), so it is not surprising that MT yields lower free energy conformations.

3.3. CSD and PDB Structure Validation. We also completed CSD and PDB crystal structure validation with our MT conformational search program. Table 5 gives the general details including conformation counts and the observed RMSD ranges as well as their median values. From this table, we observe that the conformation counts, not surprisingly, vary significantly depending on the number of rotatable bonds found in the molecule being studied (from 2 to 228919 of CSD and from 2 to 32593 of PDB total conformations). However, those with smaller conformation counts do not necessarily have a better match with the crystal structures since the diversity of conformations explored is simultaneously reduced. The detailed RMSD information together with the associated conformation count for every CSD and PDB structure is listed in Supporting Information.

Figure 9 shows the RMSD ranges compared to the crystal structures for all of the 299 CSD molecules, and Figure 10 shows this information for the PDB molecules. For the CSD RMSDs, as indicated by the blue line in the figure, we find that the minimum RMSDs differ quite a bit from one another, ranging from 0.014 to 1.720 Å, with a median value of 0.388 Å, which overall indicates that at least one MT conformation has a high structural similarity with the CSD structure. The lowest free energy conformations in the gas phase reasonably match the CSD structures, with 239 out of 299 structures having RMSDs < 2.0 Å. The median values are 1.270 Å. In the Supporting Information, we give the detailed RMSDs and free energies for all 299 structures. Meanwhile, for the 104 PDB molecules, we obtained minimum RMSDs ranging from 0.009 to 1.663 Å, with a median value of 0.253 Å. The lowest free energy conformations in the gas phase qualitatively match the PDB structures for 82 out of the 104 structures, with RMSDs < 2.0 Å. The median values are 1.342 Å. Using these metrics (RMSDs < 2.0 Å), we have ~78% match between the lowest free energy MT conformations for both CSD and PDB structures in comparison to their related crystal structures. Detailed values are shown in Table 5.

One difference between the CSD and PDB structures is that, although both are crystal structures, the CSD is for single molecules in a crystal, whereas the PDB is for protein-bound small molecules. One example of some of the issues that can arise in structure validation is illustrated by the molecule shown

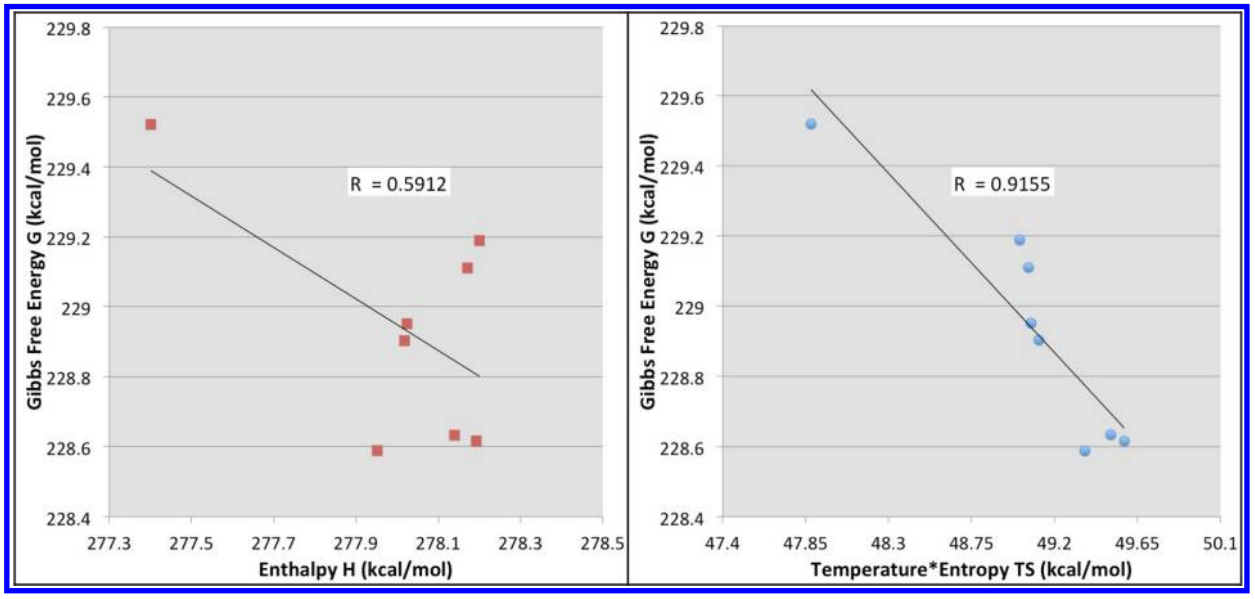


Figure 8. Linear correlation comparisons of *H* vs *G* (left) and *TS* vs *G* (right). Pearson's *R* values are attached on the graphs.

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

Table 5. MT Conformation Counts and RMSD Values (in Å) for 299 CSD and 104 PDB Structures

		conformer count	average RMSD	min. RMSD	first gas RMSD
CSD	general range	2−228919	0.102−3.068	0.014−1.720	0.014−3.872
	median value	84	1.295	0.388	1.270
PDB	general range	2−32593	0.166−3.512	0.009−1.663	0.069−4.107
	median value	142	1.376	0.253	1.314

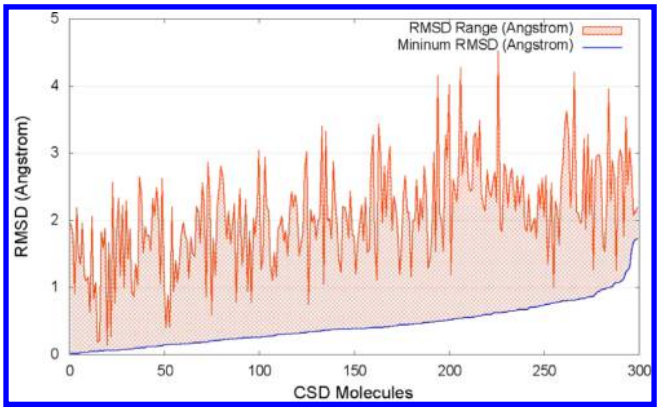


Figure 9. Minimum RMSD values and RMSD ranges for all CSD molecules. Blue curve, minimum RMSDs of the 299 CSD structures, red cross-hatched regions, RMSD ranges of the 299 CSD structures.

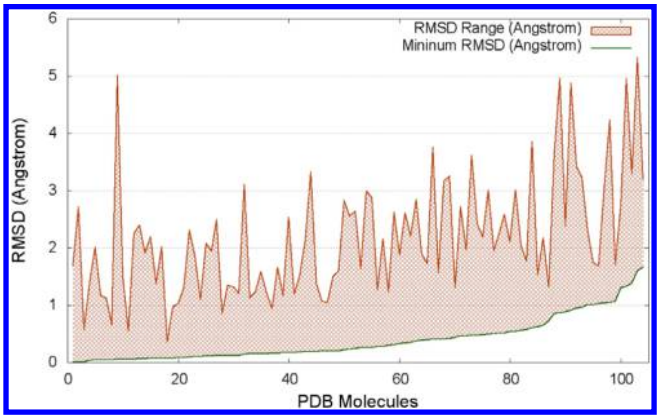


Figure 10. Minimum RMSD values and RMSD ranges for all PDB molecules. Green curve, minimum RMSDs of the 104 PDB structures, orange cross-hatched regions, RMSD ranges of the 104 PDB structures.

in Figure 11 (CSD ID BAJZOB10). The minimum free energy structure, in both the gas and aqueous phases, has a high preference to form two intramolecular hydrogen bonds, whereas in the CSD crystal structure, this is not observed, but what is preferred are intermolecular hydrogen bonds forming in the crystalline lattice. Since we are not modeling a crystal but a molecule in solution, this observation is not surprising. Nevertheless, we still obtained minimum RMSDs of 0.806 and 1.002 Å for both the gas and aqueous phase low free energy conformations. The geometry comparisons are shown in Figure 11, where the complete free energy profile is also included.

4. CONCLUSIONS

In this article, we introduce a novel conformational search program based on the MT sampling method, which directly gives conformational free energies in the gas and aqueous phases. We explain our conformation generation strategies and

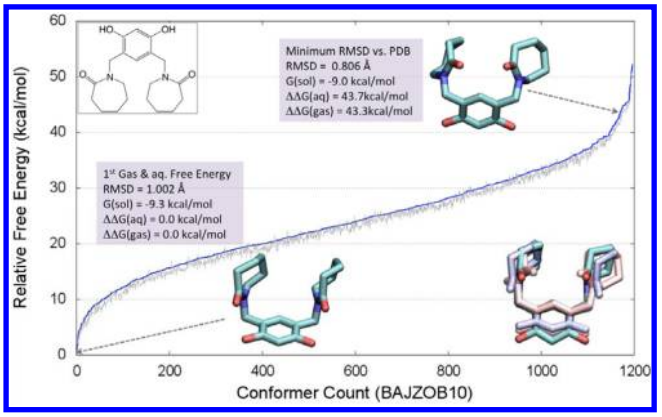


Figure 11. Relative free energies vs conformer count for CSD molecule BAJZOB10. Blue curve, aqueous phase free energy; gray curve, gas phase free energy. The single multicolored structures are the minimum RMSD conformation vs the CSD and lowest gas/aqueous phase conformation with the requisite information given in the text boxes. The overlapped structures at the bottom right represent the minimum RMSD conformer (pink) and lowest gas/aqueous phase free energy conformer (ice blue) aligned to the CSD crystal structure (multicolored). The sequence of the conformations is based on the order of the relative aqueous phase free energy.

validated the MT approach via comparisons to the crystal structures of nine thrombin inhibitors found in the PDB, 299 CSD molecules, and 104 PDB molecules found in the OMEGA validation effort.[19] The highlight of our approach is that we use the MT method to focus our torsional space search to relevant regions, providing solvation free energies and gas and aqueous phase relative free energies for all conformations generated. We compared our results to experiment by examining RMSDs of our conformational ensembles and individual structures from the crystal structures. We also compared our solvation free energies to those computed using MM-PBSA. We also compared the best conformations obtained using DFT with the MT-CS-generated conformations. Overall, the MT sampling approach rapidly and efficiently generates conformational ensembles of small molecules on the free energy surface, opening up new avenues to understand the complex behavior of molecules both in solution and when bound to receptors.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00930.

Plots for RMSDs, solvation, and relative gas and aqueous phase free energies of the nine thrombin ligands as well as selected MT vs MM-PBSA solvation free energy comparisons. Tables with RMSDs, DFT, and MT free energies for DFT-optimized structures (for both gas and PCM models) as well as the gas phases' enthalpy, entropy, and Gibbs free energy at room temperature calculated at the M06-2X/6-31G* level in Gaussian 09.

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

The figures of these DFT-validated conformers are also included (PDF)

CSD ID codes for the 299 selected molecules (TXT, TXT)

AUTHOR INFORMATION

Corresponding Author

*Phone: (517) 355-9715. Fax: (517) 353-7248. E-mail: kmerz1@gmail.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank members of the Merz group for meaningful discussions on this topic. The authors also thank the high performance computing centers at the University of Florida (UFHPC) and Michigan State University (HPCC) for providing and maintaining computational resources.

REFERENCES

(1) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic 3-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Model.* **1994**, *34*, 1000−1008.

(2) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7*, 938−950.

(3) Ebejer, J. P.; Morris, G. M.; Deane, C. M. Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* **2012**, *52*, 1146−1158.

(4) Stahura, F. L.; Bajorath, M. New methodologies for ligand-based virtual screening. *Curr. Pharm. Des.* **2005**, *11*, 1189−1202.

(5) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.

(6) Hahn, M. Three-dimensional shape-based searching of conformationally flexible compounds. *J. Chem. Inf. Model.* **1997**, *37*, 80−86.

(7) Verma, J.; Khedkar, V. M.; Coutinho, E. C. 3D-QSAR in Drug Design - A Review. *Curr. Top. Med. Chem.* **2010**, *10*, 95−115.

(8) Schwab, C. H. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technol.* **2010**, *7*, e245−253.

(9) Schwab, C. H. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technol.* **2010**, *7*, e245−e253.

(10) Kristam, R.; Gillet, V. J.; Lewis, R. A.; Thorner, D. Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. *J. Chem. Inf. Model.* **2005**, *45*, 461−476.

(11) Makino, S.; Kuntz, I. D. Automated flexible ligand docking method and its application for database search. *J. Comput. Chem.* **1997**, *18*, 1812−1825.

(12) *Catalyst*, release 4.7; Accelrys Software Inc.: San Diego, CA, 2014.

(13) *Macromodel*, version 10.4; Schrödinger, Inc.: New York, 2014.

(14) Oda, A.; Yamaotsu, N.; Hirono, S.; Takano, Y.; Fukuyoshi, S.; Nakagaki, R.; Takahashi, O. Evaluations of the conformational search accuracy of CAMDAS using experimental three-dimensional structures of protein-ligand complexes. *J. Phys.: Conf. Ser.* **2013**, *454*, 012028.

(15) *Molecular Operating Environment (MOE)*, version 2013.08; Chemical Computing Group Inc.: Montreal, QC, Canada, 2013.

(16) O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. Confab - Systematic generation of diverse low-energy conformers. *J. Cheminf.* **2011**, *3*, 8.

(17) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534−546.

(18) Miteva, M. A.; Guyon, F.; Tuffery, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*, W622−627.

(19) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.

(20) Sperandio, O.; Souaille, M.; Delfaud, F.; Miteva, M. A.; Villoutreix, B. O. MED-3DMC: a new tool to generate 3D conformation ensembles of small molecules with a Monte Carlo sampling of the conformational space. *Eur. J. Med. Chem.* **2009**, *44*, 1405−1409.

(21) Liu, X. F.; Bai, F.; Ouyang, S. S.; Wang, X. C.; Li, H. L.; Jiang, H. L. Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC Bioinf.* **2009**, *10*, 101.

(22) Lagorce, D.; Pencheva, T.; Villoutreix, B. O.; Miteva, M. A. DG-AMMOS: a new tool to generate 3d conformation of small molecules using distance geometry and automated molecular mechanics optimization for in silico screening. *BMC Chem. Biol.* **2009**, *9*, 6.

(23) Sauton, N.; Lagorce, D.; Villoutreix, B. O.; Miteva, M. A. MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinf.* **2008**, *9*, 184.

(24) Chen, I. J.; Foloppe, N. Conformational sampling of druglike molecules with MOE and catalyst: Implications for pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **2008**, *48*, 1773−1791.

(25) Li, J.; Ehlers, T.; Sutter, J.; Varma-O'Brien, S.; Kirchmair, J. CAESAR: A new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J. Chem. Inf. Model.* **2007**, *47*, 1923−1932.

(26) Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tuffery, P. Frog: a FRee Online druG 3D conformation generator. *Nucleic Acids Res.* **2007**, *35*, W568−W572.

(27) Sadowski, J.; Bostrom, J. MIMUMBA revisited: Torsion angle rules for conformer generation derived from X-ray structures. *J. Chem. Inf. Model.* **2006**, *46*, 2305−2309.

(28) Hare, B. J.; Walters, W. P.; Caron, P. R.; Bemis, G. W. CORES: An automated method for generating three-dimensional models of protein/ligand complexes. *J. Med. Chem.* **2004**, *47*, 4731−4740.

(29) Balducci, R.; Pearlman, R. S. CONFORT: A rational conformation analysis tool. *Abstr. Pap. Am. Chem. Soc.* **1999**, *217*, U644−U645.

(30) Landrum, G. *RDKit: Open-source cheminformatics*. http://www.rdkit.org.

(31) *Rubicon*, version 4.7.1; Daylight Chemical Information Systems, Inc: Aliso Viejo, CA. http://www.daylight.com.

(32) *CORINA*; Molecular Networks GmbH: Erlangen, Germany. http://www.molecular-networks.com.

(33) Leach, A. R. *Molecular Modeling: Principles and Applications*, 2nd ed.; Prentice Hall: Harlow, England, 2001; Chapter 9, p 457.

(34) Leach, A. R. A survey of methods for searching the conformational space of small and medium-size molecules. *Rev. Comput. Chem.* **1991**, *2*, 1−55.

(35) Vainio, M. J.; Johnson, M. S. Generating conformer ensembles using a multiobjective genetic algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462−2474.

(36) Pavlov, T.; Todorov, M.; Stoyanova, G.; Schmieder, P.; Aladjov, H.; Serafimova, R.; Mekenyan, O. Conformational coverage by a genetic algorithm: Saturation of conformational space. *J. Chem. Inf. Model.* **2007**, *47*, 851−863.

(37) Mekenyan, O.; Dimitrov, D.; Nikolova, N.; Karabunarliev, S. Conformational coverage by a genetic algorithm. *J. Chem. Inf. Model.* **1999**, *39*, 997−1016.

(38) Chang, G.; Guida, W. C.; Still, W. C. An Internal Coordinate Monte-Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.

(39) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. Applications of Simulated Annealing to the Conformational-Analysis of Flexible Molecules. *J. Comput. Chem.* **1991**, *12*, 342−349.

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864

(40) Grebner, C.; Becker, J.; Stepanenko, S.; Engels, B. Efficiency of Tabu-Search-Based Conformational Search Algorithms. *J. Comput. Chem.* **2011**, *32*, 2245−2253.
(41) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. The Theory and Practice of Distance Geometry. *Bull. Math. Biol.* **1983**, *45*, 665−720.
(42) Fu, Z.; Li, X.; Merz, K. M. Conformational Analysis of Free and Bound Retinoic Acid. *J. Chem. Theory Comput.* **2012**, *8*, 1436−1448.
(43) Fu, Z.; Li, X.; Merz, K. M. Accurate Assessment of the Strain Energy in a Protein-Bound Drug Using QM/MM X-ray Refinement and Converged Quantum Chemistry. *J. Comput. Chem.* **2011**, *32*, 2587−2597.
(44) Zheng, Z.; Ucisik, M. N.; Merz, K. M. The Movable Type Method Applied to Protein-Ligand Binding. *J. Chem. Theory Comput.* **2013**, *9*, 5526−5538.
(45) Zheng, Z.; Merz, K. M., Jr. Development of the knowledge-based and empirical combined scoring algorithm (KECSA) to score protein-ligand interactions. *J. Chem. Inf. Model.* **2013**, *53*, 1073−83.
(46) Tomasi, J.; Persico, M. Molecular-Interactions in Solution - an Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027−2094.
(47) Cramer, C. J.; Truhlar, D. G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.
(48) Orozco, M.; Luque, F. J. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **2000**, *100*, 4187−4225.
(49) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3093.
(50) Cramer, C. J.; Truhlar, D. G. A universal approach to solvation modeling. *Acc. Chem. Res.* **2008**, *41*, 760−768.
(51) Zheng, Z.; Wang, T.; Li, P. F.; Merz, K. M. KECSA-Movable Type Implicit Solvation Model (KMTISM). *J. Chem. Theory Comput.* **2015**, *11*, 667−682.
(52) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(53) Battle, G.; Kleywegt, G.; Velankar, S.; Oldfield, T.; Gore, S.; Mir, S.; Dana, J. Chemistry-related resources at the Protein Data Bank in Europe. *Abstr. Pap. Am. Chem. Soc.* **2013**, *245*.
(54) Baum, B.; Muley, L.; Smolinski, M.; Heine, A.; Hangauer, D.; Klebe, G. Non-additivity of functional group contributions in protein-ligand binding: a comprehensive study by crystallography and isothermal titration calorimetry. *J. Mol. Biol.* **2010**, *397*, 1042−54.
(55) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
(56) Henderson, S.; Bruno, I. J. Cambridge Structural Database: Moving with the times. *Abstr. Pap. Am. Chem. Soc.* **2013**, *245*.
(57) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.
(58) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314−3321.
(59) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(60) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(61) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(62) *Visual Molecular Dynamics (VMD)*, version 1.9.2; University of Illinois at Urbana−Champaign: Urbana, IL, 2015. http://www.ks.uiuc.edu/Research/vmd/.
(63) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(64) *gnuplot* 4.6.6, 2014. http://gnuplot.info.
(65) Hawkins, P. C. D. Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. *OMEGA* 2.5.1.4; OpenEye Scientific Software: Santa Fe, NM.
(66) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2005**, *26*, 114−114.

DOI: 10.1021/acs.jctc.5b00930
J. Chem. Theory Comput. 2015, 11, 5853−5864