ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

# From Thermodynamics to Kinetics: Enhanced Sampling of Rare Events

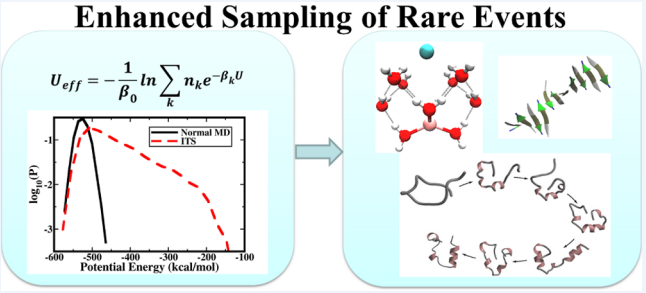
Lijiang Yang,[†,‡] Cheng-Wen Liu,[†] Qiang Shao,[§] Jun Zhang,[†,‡] and Yi Qin Gao*[,†,‡]

[†]Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, and [‡]Biodynamic Optical Imaging Center, Peking University, Beijing 100871, China

[§]Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

Supporting Information

**CONSPECTUS:** Despite great advances in molecular dynamics simulations, there remain large gaps between the simulations and experimental observations in terms of the time and length scales that can be approached. Developing fast and accurate algorithms and methods is of ultimate importance to bridge these gaps. In this Account, we briefly summarize recent efforts in such directions. In particular, we focus on integrated tempering sampling. The efficiency of this sampling method has been demonstrated by applications to a range of chemical and biological problems: protein folding, molecular cluster structure searches, and chemical reactions. The combination of integrated tempering sampling and a trajectory sampling method allows the calculation of rate constants and reaction pathways without predefined collective coordinates.




## 1. INTRODUCTION

With the fast development of computer hardware and computational methods, great progress has been made in molecular simulations of complex systems, which are important in many areas including chemistry, biology, physics, and material sciences. A major obstacle in understanding the structure and dynamics of complex systems, including macromolecules and large molecular clusters, relates to the exploration of complex energy landscapes. Since fast degrees of freedom, such as vibrations, limit the integration time step in a simulation to femtoseconds or shorter, even a microsecond simulation requires the propagation of equations for many degrees of freedom over billions of steps, leading to a very high computational cost.[1] A number of methods have been developed over the last several decades to overcome the computational bottleneck of molecular simulations. These methods are mainly designed to alleviate difficulties related to the large time and length scales of complex systems and to enhance the sampling in the most relevant phase space. In this sense, the development of coarse-grained models represents a viable strategy for enhanced sampling but is not the focus of this Account. Instead, we review enhanced sampling techniques that allow an accelerated search in the configuration and trajectory space and thus fast thermodynamics and/or kinetics calculations. Largely as a result of such efforts, significant progress has been made over the past several decades, and molecular dynamics (MD) simulations are now routinely used to provide structural, thermodynamic, and even dynamical information on complex molecular systems.

The high energy barriers between different structures/conformations mean that transitions between them, such as activated chemical reactions and concerted structural changes, appear as rare events. Between the transitions, the stable structures are repeatedly sampled and, very often, the sampling of the reactant structures becomes redundant. Therefore, when one calculates thermodynamic properties, it is convenient to make use of equilibrium statistical mechanics to enhance the sampling of rare events by systematically modifying the potential energy surface, e.g., by decreasing the energy barrier to enhance the sampling over the transition regions. Such methods include but are not limited to the widely used umbrella sampling,[2] J-walking,[3] adaptive umbrella sampling,[4] metadynamics,[5] accelerated MD simulation,[6,7] conformational flooding,[8] conformational space annealing,[9] hyperdynamics simulation,[10] potential smoothing methods,[11] and adaptive biasing force methods.[12] To effectively guide the simulations, some methods, such as the popular umbrella sampling and metadynamics simulations and their many variants, use predefined collective coordinates. Alternatively, especially when it is difficult to identify proper collective coordinates, generalized ensemble methods can be used to generate a uniform distribution of energy (or temperature) in an MD simulation. A number of such methods have been tested and used, the most popular of which include replica-exchange molecular dynamics (REMD),[13] parallel tempering,[14] multi-

Received: July 28, 2014
Published: March 17, 2015





DOI: 10.1021/ar500267n
Acc. Chem. Res. 2015, 48, 947−955

canonical simulation,[15] the Wang–Landau algorithm,[16] statistical temperature sampling,[17] simulated tempering,[18] and the enveloping distribution sampling method,[19] among many others. Other recent developments of free energy calculation methods include the random walk in an orthogonal space[20] and diffusion-map-directed molecular dynamics simulations.[21]

In studies of structural and conformational changes of large molecules (e.g., protein folding), REMD is probably the most widely used enhanced sampling method. REMD enhances the sampling of rare events by simulations using replicas at both low and high temperatures. The equilibration in the entire system makes it difficult to enhance the sampling over only a selected collection of degrees of freedom. Such problems have been investigated by Berne and co-workers, who developed the replica exchange with solute tempering (REST) method.[22] The applicability of the REST method to a broad range of molecular systems needs to be further tested.

Instead of running parallel simulations at many different temperatures, the integrated tempering enhanced sampling (ITS) method[23] generates a generalized non-Boltzmann distribution covering a large energy range, similar to the enveloping distribution sampling method.[19] The method allows fast sampling of the configuration space and efficient calculations of thermodynamic properties based on a fast evaluation of the partition function. In this Account, we focus mainly on the theoretical background and numerical implementation of the ITS method and give a few examples of its applications. Possible future development and applications of this method are also discussed.

## 2. INTEGRATED TEMPERING SAMPLING AND SELECTIVE INTEGRATED TEMPERING SAMPLING: ENHANCED SAMPLING IN ENERGY AND CONFIGURATION SPACES

In ITS, the sampled energy range is broadened through the introduction of a sum-over-temperature non-Boltzmann distribution factor (see the Supporting Information for further details). The distribution corresponding to the distribution factor is obtained either through a Monte Carlo procedure or by using the effective force calculated from the following effective potential energy:

$$U_{\text{eff}} = -\frac{1}{\beta_0}\ln\sum_k n_k e^{-\beta_k U} \tag{1}$$

where $U$ is the potential energy of the system under study, $\beta_0 = 1/k_B T_0$ (where $k_B$ is the Boltzmann constant and $T_0$ is the temperature of the system), $\beta_k$ denotes a series of temperatures covering both low and high temperatures around $T_0$, and $n_k$ denotes weighting factors obtained through an iterative procedure.[23,24] The biased force $F_{\text{eff}}$ used in the simulations is

$$F_{\text{eff}} = -\frac{\partial U_{\text{eff}}}{\partial r} = \frac{\sum_k \beta_k n_k e^{-\beta_k U}}{\beta_0 \sum_k n_k e^{-\beta_k U}} F \tag{2}$$

where $F$ is the force in the original system. Unlike methods using collective coordinates such as umbrella sampling and metadynamics, the ITS and REMD methods are both implemented in the absence of predefined reaction coordinates. In contrast to the REMD method, the ITS method avoids multiple parallel calculations and exchange operations between parallel trajectories and thus requires fewer computational resources. It also circumvents the problem of re-equilibration for the kinetic energy arising from the exchange events in REMD. In addition, since the enhanced sampling is performed in the energy space, unlike the REMD method, it is convenient in the implementation of ITS to divide the system into subspaces and to enhance the sampling for a preselected subsystem; this approach is called selective integrated tempering sampling (SITS). For example, in explicit-solvent simulations of protein folding, the protein atoms can be targeted for enhanced sampling, and a large variety of protein configurations (e.g., both folded and unfolded) can be sampled while the solvent is kept at the near-room-temperature configurations.[25] The differentiated sampling of such a system is conveniently achieved by introducing an effective potential in the form

$$U_{\text{eff}} = E_w - \frac{1}{\beta_0}\ln\sum_k n_k e^{-\beta_k(E_p+E_{pw})} \tag{3}$$

In the above equation, the total energy $U$ is divided into multiple components:

$$U = E_p + E_w + E_{pw} \tag{4}$$

where $E_p$, $E_w$, and $E_{pw}$ are the internal energy of the part of interest (protein), the internal energy of all water molecules, and the energy of interactions between the protein and water, respectively.

From ITS or SITS simulations, a converged calculation yields a biased distribution function in the configuration space, $\rho_{\text{eff}}(r) \propto e^{-\beta_0 U_{\text{eff}}(r)}$. The desired distribution function at the temperature $T_0$ is easily calculated as

$$\rho(r) = \rho_{\text{eff}}(r) e^{-\beta_0[U(r)-U_{\text{eff}}(r)]} \tag{5}$$

where $\rho_{\text{eff}}(r)$ is calculated from the enhanced sampling simulation and $U_{\text{eff}}(r)$ is defined as in eq 1 or 3. Other thermodynamic properties can be readily obtained once $\rho(r)$ is known.[25,26]

In a few tested examples, ITS and SITS were shown to be highly efficient.[25,26] For example, SITS was applied to study the conformational transition of alanine dipeptide in aqueous solution. The potentials of mean force (PMFs) for the main-chain $\psi$ torsion angle calculated using SITS (blue ○), umbrella sampling (black +), and REMD (red △) are shown in Figure 1a. Apparently, the PMF obtained using a 40 ns SITS

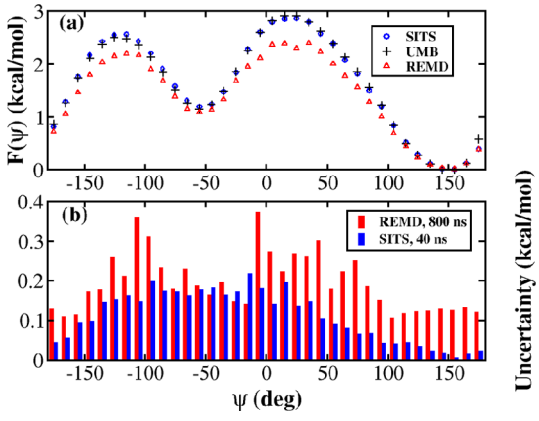



**Figure 1.** PMFs in $\psi$ for alanine dipeptide in explicit-solvent simulations. (a) PMFs in $\psi$ obtained from enhanced sampling simulations. (b) Computed uncertainties for nonoverlapping 2 ns blocks.



DOI: 10.1021/ar500267n
Acc. Chem. Res. 2015, 48, 947–955

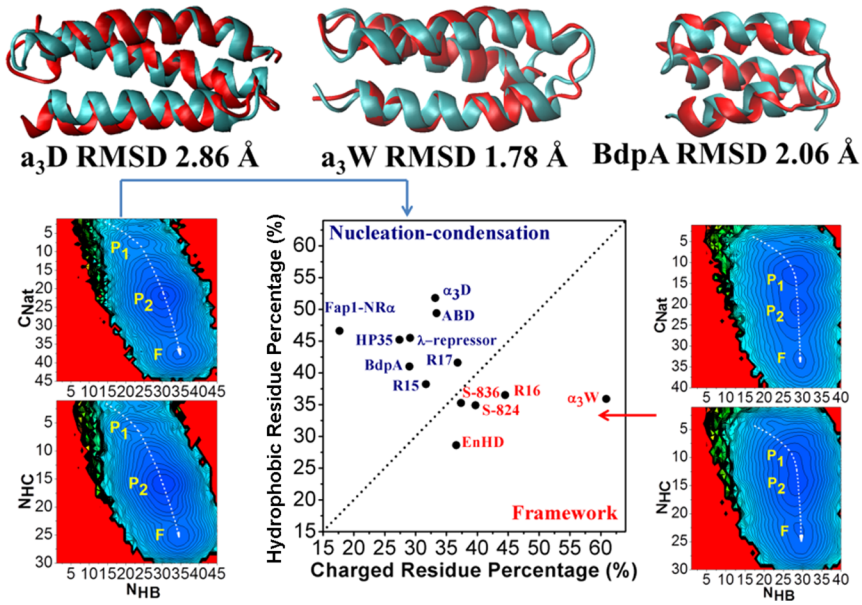



**Figure 2.** (top) Comparison of the simulated best structures (blue) and the NMR structures (red) of $\alpha_3$D, $\alpha_3$W, and BdpA. (bottom) Folding mechanism classification based on the average percentages of hydrophobic and charged residues, and the two-dimensional free energy landscapes of (left) $\alpha_3$D and (right) $\alpha_3$W indicating nucleation−condensation and framework folding mechanisms, respectively. The dashed lines in the free energy landscapes represent the folding pathways connecting the distinct states on the configuration space (partially folded ($P_1$ and $P_2$) and folded (F) states).

simulation is in good agreement with that calculated using an umbrella sampling simulation. The umbrella sampling simulation used the $\psi$ angle as the reaction coordinate, whereas in the SITS calculation, no reaction coordinate was used or needed. Furthermore, the error bars calculated for the SITS results are smaller than those obtained for the much longer REMD simulation (800 ns), further showing the high efficiency of SITS. In a study of a simple peptide model system,[26] we compared ITS with several other popular enhanced sampling methods to show that ITS is more efficient in configuration sampling and free energy calculations for this model system (see the Supporting Information for further details).

## 3. APPLICATIONS

In the following we present examples of applications of ITS/SITS to complex problems such as protein folding, molecular cluster structure searching, and chemical reactions.

### 3.1. Protein and Polypeptide Folding

Proteins fold into native structures to carry out their biological functions. Understanding the protein folding mechanism is of fundamental importance to biology. As a result of complicated inter-residue interactions, protein folding is typically a slow process with a complex free energy landscape. The sampling of protein conformations by means of MD simulations is a difficult problem. The application of ITS allows multiple reversible folding and unfolding events to be observed for a variety of polypeptides and small proteins by microsecond-long all-atom simulations.[27−31] The equilibrium sampling of the protein structures allows not only a quantitative assessment of the force fields used but also a detailed analysis of the protein-folding free energy landscape. The calculated population distribution of low-free-energy structures and comprehensive pictures of the low-free-energy protein folding pathway were compared directly with experimental results.[31,32]

In a series of studies, the folding of $\beta$-hairpins and its temperature dependence were investigated by conducting ITS molecular dynamics (ITS-MD) simulations. Through a comparative study of $\beta$-structured polypeptides, including peptide 1, the C-terminal $\beta$-hairpin from the B1 domain of protein G (GB 1p), TRPZIP2, and TRPZIP4, we investigated the important role of the turn conformational propensity in dictating the folding pathway of the $\beta$-hairpin. It was found that in the presence of a strong turn-promoting sequence, $\beta$-hairpin folding is accelerated and turn formation is an easy process compared with packing of the cross-strand hydrophobic core and the assembly of backbone hydrogen bonds; otherwise, turn formation becomes the rate-limiting step, and the folding process is initiated by the packing of the hydrophobic core.[27,28] This observation is consistent with the results of circular dichroism (CD)/infrared temperature-jump studies[33] and NMR/CD spectroscopy studies.[34] In addition, the simulations predicted that the stabilities of the backbone hydrogen bonds of the four polypeptides depend differently on the temperature.[27] Using the same simulation methodology, we evaluated the cold denaturation tendency of five $\beta$-structured polypeptides (MrH1, MrH4a, MrH3a, MrH3b, and MrH4b)[35] and again obtained results that were in good agreement with the experimental results.[36]

ITS-MD simulations were also performed to systematically investigate the folding mechanism of $\alpha$-helix bundle proteins, including the B domain of protein A from *Staphylococcus aureus* (BdpA, 46 residues), $\alpha_3$D (73 resides), and $\alpha_3$W (67 residues) (Figure 2, top panel).[29,30] For each protein under study, the all-atom ITS-MD simulations again yielded multiple folding and unfolding events. In all of the folding events, the proteins reached native-like structures with small root-mean-square-deviation (RMSD) values, which had been difficult to obtain.[37,38] The subsequent free energy landscape analysis showed that the folding mechanisms of the helical bundles are


DOI: 10.1021/ar500267n
Acc. Chem. Res. 2015, 48, 947−955

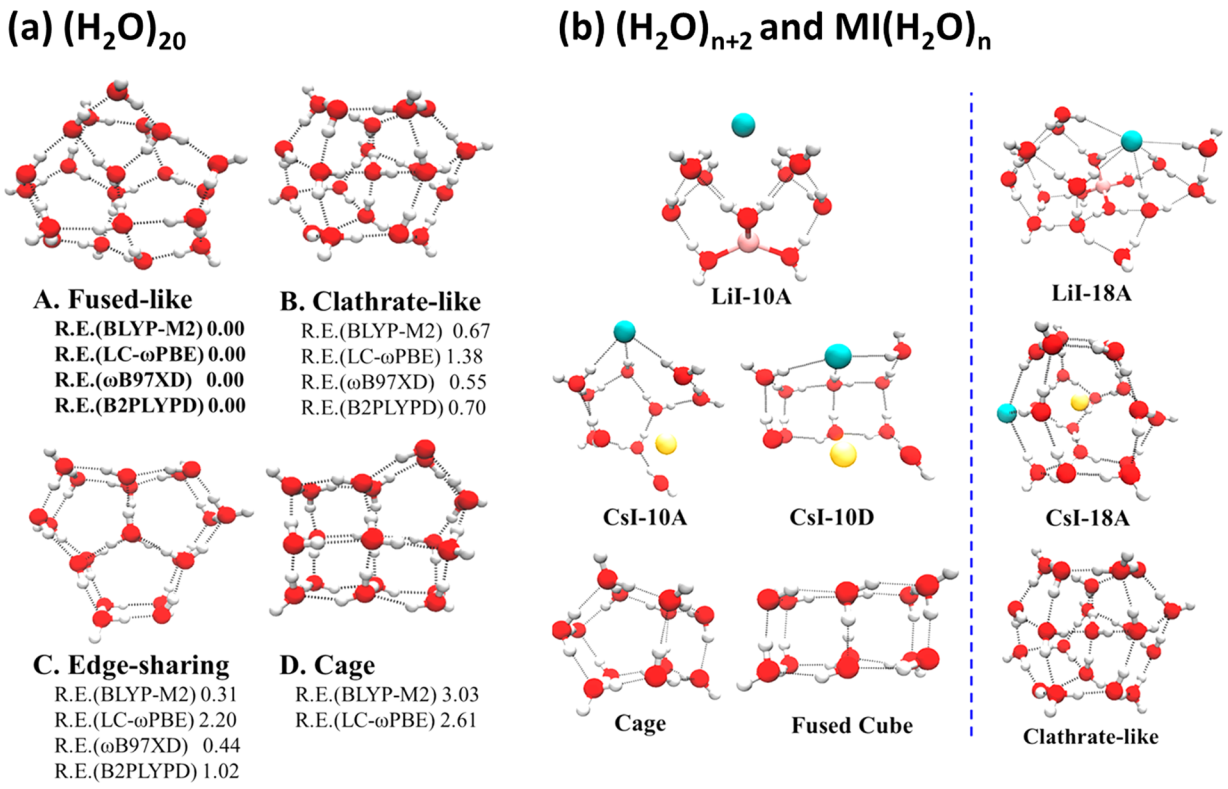



**Figure 3.** (a) Low-energy structures of the $(H_2O)_{20}$ cluster and (b) comparison of $MI(H_2O)_n$ clusters with $(H_2O)_{n+2}$ clusters ($n$ = 10, 18; M = Li, Cs). The relative energies calculated using different functionals are presented in (a).

heavily sequence-dependent although their native structures have similar topologies (Figure 2, bottom panel).

In another application, in collaboration with the experimental group, we combined ITS-MD simulations with an ion mobility spectrometry (IMS) experiment to decipher the gas-phase conformations of various proteins/polypeptides. The tested polypeptides included protonated TRPZIP1 and its derivatives,[39] alkali-metal ion ($Na^+$, $K^+$, $Cs^+$) adducts of gramicidin A (GA),[40] and bradykinin fragments 1−5 (RPPGF) and 1−8 (RPPGFSPF).[41] For each polypeptide, the calculated ion−neutral collision cross section and its variance with the protonation state and the binding of metal ions were in good agreement with the IMS measurements. These studies showed that the combined ITS-MD and IMS approach provides a straightforward and efficient method for measuring the intrinsic conformational preferences of proteins.

### 3.2. Searches for Stable Structures of Molecular Clusters

Another area in which the ITS method finds its application is the search for stable structures of molecular clusters and the following calculations of their thermodynamic properties. Molecular clusters are normally characterized by complex energy landscapes, with many local minima separated by relatively high energy barriers. As a result, although a large number of water cluster structures have been identified by previous studies, it remains a great challenge to identify new stable isomers. The difficulty relates to effectively crossing high energy barriers and at the same time locating energy minima. Such difficulties can be largely alleviated by ITS-MD simulations, which have been shown to be efficient in sampling/identifying low-energy structures of clusters of varied chemical compositions and sizes.

In recent studies, ITS-MD simulations in combination with density functional theory (DFT) calculations allowed us to sample more efficiently the structures of water and water−salt clusters. To benchmark the method, we first searched for low-energy structures of pure water clusters that had been extensively studied. In these studies, classical ITS-MD simulations were performed, followed by energy minimization using the low-energy structures generated in the ITS-MD simulations. The structures were then further optimized by conducting quantum-chemical calculations employing dispersion-corrected density functionals such as BLYP-M2,[42] ωB97XD,[43] and B2PLYPD.[44] Using this approach, we obtained not only all of the low-energy structures of $(H_2O)_{12}$ and $(H_2O)_{20}$ found by previous studies[45,46] but also new structures, for example, a new lowest-energy structure for $(H_2O)_{20}$ (structure A in Figure 3a).

The ITS-MD−DFT protocol was then used to investigate the structure and intermolecular interactions in salt−water clusters, such as $LiI(H_2O)_n$, $CsI(H_2O)_n$, and $NaCl(H_2O)_n$. The calculations clarified the features common to these different clusters; e.g., anions but not cations tend to reside at the cluster surface. Furthermore, the structures of $CsI(H_2O)_{10}$ resemble those of $(H_2O)_{12}$, replacing two water molecules with $Cs^+$ and $I^-$. $CsI(H_2O)_{18}$ clusters tend to contain a clathrate-like motif, with $Cs^+$ remaining at the center of a cage formed by 17 water molecules and the $I^-$ ion, resembling well the clathrate-like structure of $(H_2O)_{20}$ (Figure 3b). In contrast, the stable four-coordinate $Li^+$ dictates the structure of $LiI(H_2O)_n$ clusters. Interestingly, we found that the number of water molecules needed to form solvent-separated ion pairs varies with the salt pair, being 5 and >20 for clusters containing LiI, and CsI, respectively (Figure 4). Our ongoing study has shown that this number is around 10 for NaCl. These results indicate that the


DOI: 10.1021/ar500267n
Acc. Chem. Res. 2015, 48, 947−955

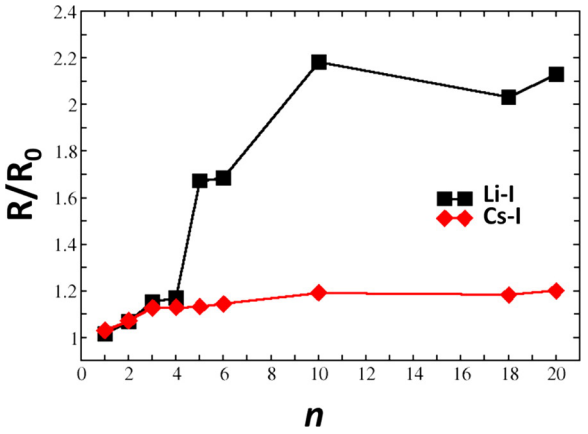



**Figure 4.** Cation−anion (Li−I and Cs−I) distance $R$ (normalized by the bare ion distance $R_0$) as a function of the number of water molecules $n$. The LiI and CsI data were taken from Li et al.[49] and Liu et al.[47]

hydration of salt ions is *ion-pair*-specific and that the structures of salt−water clusters are determined by the competition among ion−ion, ion−water, and water−water interactions. Furthermore, the ITS simulations provided thermodynamic information, such as the temperature dependence of the distribution functions of the water coordination number and cation−anion distance.[47] The latter information is important for our understanding of the ion-specific Hofmeister series.[48]

### 3.3. Quantum-Mechanical/Molecular-Mechanical (QM/MM) Calculation of Solution Chemical Reactions

SITS can be naturally introduced into QM/MM calculations, in which only a small part of the simulation system is treated using quantum mechanics and the rest is treated by classical molecular mechanics. For such calculations, since the events of interest (e.g., chemical reactions) normally occur within the quantum region, it is desirable to explore the molecular configurations of the quantum-mechanically treated subsystem. One can easily make use of the SITS scheme to enhance the sampling over the QM region while keeping the MM part less perturbed by introducing the effective potential

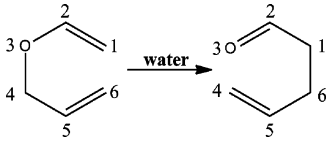
$$U_{\text{eff}} = E_{\text{MM}} - \frac{1}{\beta_0} \ln \sum_k n_k \mathrm{e}^{-\beta_k (E_{\text{QM}} + E_{\text{QM/MM}})} \tag{6}$$

where $E_{\text{MM}}$ is the self-energy of the MM region (e.g., the solvent), $E_{\text{QM}}$ is the self-energy of the QM region (e.g., the reacting molecule), and $E_{\text{QM/MM}}$ is the energy of the interaction between the QM and MM regions.

Such a scheme was applied to gain an understanding of the solvation and substituent effects in the Claisen rearrangement (Scheme 1). Although the Claisen rearrangement[50] was the earliest recorded [3,3] sigmatropic reaction, the water-acceleration effect on this reaction has drawn wide interest, with the mechanism remaining unclear. To reveal more details concerning the solvation effect of the aliphatic Claisen rearrangement, a SITS-QM/MM simulation protocol was designed in which the solute (i.e., the reactant) was treated quantum-mechanically employing the self-consistent-charge density-functional tight-binding (SCC-DFTB) method developed by Elstner et al.[51] Higher-level QM methods such as DFT and the coupled-cluster singles and doubles model can also be employed but would require more computational resources. All of the solvent molecules were treated with MM. SITS was integrated into such a hybrid simulation method to specifically achieve an efficient sampling of the QM-treated reactant. In the simulation, the reactant, allyl vinyl ether (AVE), reached a conformational equilibrium between compact and extended conformations. The conformational equilibrium constant was calculated over a range of temperatures. It was found that there was a shift in the equilibrium toward the compact conformation under aqueous conditions compared with the conformational distribution of the same molecule in methanol or toluene. At 300 K, the probability that AVE adopts the compact conformation is about 6% in water but only about 2% in toluene. The compact conformation (Figure 5a,b), which is the proper configuration for the subsequent chemical reaction, was found to be polarized with respect to the extended conformation according to Mulliken population analysis of the QM/MM-calculated data.

SITS-QM/MM simulations also allowed us to observe directly the transition of AVE to 4-pentenal (Scheme 1 and Figure 5c). Therefore, SITS allowed the chemical transition process over high barriers to be realized in silico without any intrinsic reaction coordinates. SITS herein exhibited excellent capability and adaptability for sampling of chemical events.

**Scheme 1. Aliphatic Claisen Rearrangement Reaction and the Indices of Heavy Atoms of the Reactant (Allyl Vinyl Ether) and the Product (4-Pentenal)**




## 4. EFFICIENT KINETICS CALCULATIONS: ENHANCED SAMPLING OF REACTIVE TRAJECTORIES

One of the common problems in applying enhanced sampling methods over energy or configuration space is that dynamics as well as kinetics information on the original system is lost as a result of the use of effective Hamiltonians. In many cases, especially in studies of reaction mechanisms, it is desirable to obtain information on pathways of transitions on a complex and many-dimensional energy surface. A few methods have been developed to search for transition paths effectively, including the nudged elastic band method,[52] the string method,[53] the weighted-ensemble method,[54] milestoning,[55] and the forward flux sampling method.[56] However, it is generally difficult to obtain kinetics information for chemical reactions directly using the above-mentioned methods, although the calculation of free energies along the minimum-energy or minimum-free-energy paths does allow evaluation of the rate constant with the help of transition state theory.

One effective method of sampling reactive trajectories is transition path sampling (TPS), which was proposed by Chandler and co-workers.[57] In TPS, an initial trajectory is obtained to connect the reactant and product states employing methods such as targeted molecular dynamics or high-temperature simulations. Such a trajectory can be "unreal" but is used to generate an ensemble of trajectories employing a shooting and shifting Monte Carlo procedure. To further improve the computational efficiency and especially to calculate the rate constant, a few variants of the method have been proposed, including transition interface sampling[58] and the related partial path transition interface sampling.[59]



DOI: 10.1021/ar500267n
Acc. Chem. Res. 2015, 48, 947−955

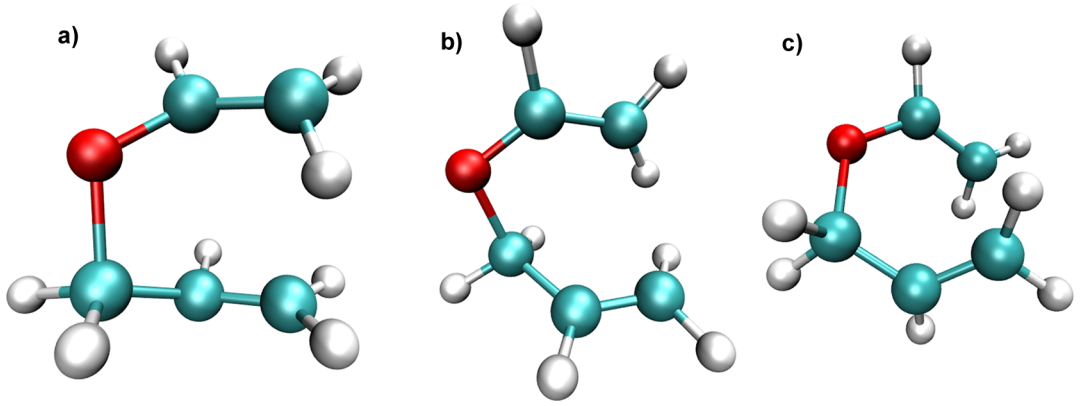



**Figure 5.** (a, b) Typical structures of the compact conformation defined in the QM/MM study, including (a) a quasi-six-membered ring and (b) a quasi-eight-membered ring. (c) Snapshot of the chairlike structure of the transition state.

Although TPS is efficient in generating successful transition paths around a known pathway without the requirement of predefined reaction coordinates, the ratio of the number of successful trajectories to the total number of trajectories is needed for the accurate evaluation of rate constants. Such a flux calculation has to be done separately in TPS and related methods. It is thus desirable to further develop methods that easily sample all of the important reaction pathways and perform rate constant calculations without *any* information on preselected reaction coordinates. Inspired by TPS, we introduced a combined approach that takes advantage of both ITS-MD and TPS (shooting) methods.[60] In this method, we first carry out ITS-MD simulations to identify the active phase space of the reaction of interest. Path sampling is then performed on the original potential energy surface starting from the phase space points sampled by the ITS-MD simulation (and thus having a known Boltzmann distribution), which automatically gives an enhanced sampling of rare events such as chemical reactions over high energy barriers. Since the statistical weight of each trajectory can be calculated using eq 5, the rate constant for the original system can be obtained directly as the ratio of the reactive and total trajectories:

$$\text{rate} = \frac{\sum_{\text{succ traj}} e^{f(v)/k_B T}}{t \sum_{\text{all traj}} e^{f(v)/k_B T}} \tag{7}$$

Furthermore, since trajectory shooting is performed using an ensemble of phase space points biased toward successful transitions, this reactive trajectory sampling method increases the efficiency in rate constant calculations and reaction pathway searching. Compared with traditional TPS, there are three major advantages: (1) the initial trajectories are automatically generated by efficient ITS-MD simulations; (2) a thorough sampling of the phase space avoids the entrapment of trajectories in particular pathway(s), allowing the search of multiple pathways separated by high barriers; and (3) both the reactive and nonreactive trajectories are calculated directly, avoiding the need for a predetermined reaction coordinate and the calculation of the reactive flux.

The above method has been applied to a model system with multiple minima on its free energy surface[24] as well as the conformational change of methyl maltose.[60] For the latter application, first, ITS-MD simulations were conducted and a couple thousand phase space points were selected to shoot trajectories on the original (unbiased) potential energy surface. The logarithm of the visiting probability as a function of two torsional angles was calculated from the successful transition path ensemble and is shown in Figure 6. All three transition

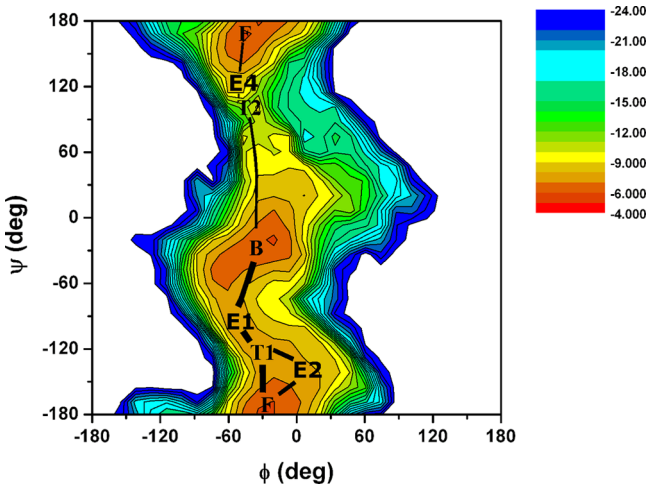



**Figure 6.** Logarithm of probabilities of states over $\varphi$ and $\psi$ calculated from successful transition path ensembles.

pathways found in a previous TPS study[61] were also captured in this study. The most favorable transition path ensemble is E1, which goes through a rotation about $\psi$ in the negative direction and passes through transition state $T_1$. A less favorable transition path ensemble is E2, which passes through metastable region G instead of $T_1$. The third transition path ensemble is E4, which passes through transition state $T_2$. By means of this approach, in a total simulation time of approximately 3.2 $\mu$s, 10729 successful transition trajectories were obtained. If normal MD simulations has been used, only ~20 transitions would have been observed. Thus, there is a remarkable improvement in the transition sampling efficiency when the current method is used.

The SITS-based trajectory sampling method could also be applied to the QM/MM simulation of the Claisen rearrangement mentioned above. The calculated rate constants are in good agreement with experimental results (see the Supporting Information).



DOI: 10.1021/ar500267n
*Acc. Chem. Res.* 2015, 48, 947−955

## 5. SUMMARY

In this Account, we have summarized the ideas of enhanced sampling methods based on the integrated tempering sampling technique in both phase and trajectory spaces. The existing applications of ITS have shown its efficiency in configuration searching and thermodynamic calculations for problems ranging from protein folding/unfolding to molecular cluster structure sampling and chemical reactions. In particular, this method can be used to selectively enhance the sampling of a selected subsystem (SITS), which makes it a potentially useful tool in many biomolecular simulations in complex environments. Meanwhile, enhanced sampling of reactive trajectories and rate constant calculations can be achieved with low computational cost when ITS is used in combination with the transition-path shooting technique. These methods are expected to have broad applications for a large variety of complex systems.

As illustrated in the previous discussions, significant progress has been made over previous decades in developing efficient and reliable sampling methods for studies of complex systems. Many of these methods have been successfully applied to a range of chemical and biological problems. However, even with the help of these methods, many calculations are still too demanding. It is therefore important to look into the further development of efficient simulation methods, including new sampling algorithms. The ITS method described in this Account certainly has its own limitations and needs further improvement. For example, it is an advantage that ITS enhances the sampling of complex systems without a priori reaction coordinates. However, such a feature of the method could also be a weakness. By effectively combining the ability of ITS to quickly scan a large energy space and the abilities of other methods to accelerate the sampling along defined collective coordinates, one could further improve the sampling efficiency. Along this direction, in a recent study ITS was combined with umbrella sampling to obtain better and faster convergence in free energy calculations.[62] Another difficulty of ITS is that it is difficult to achieve convergence of the weighting factors $n_k$ in eq 1 for large systems. In a recently proposed procedure,[63] a set of $n_k$ was estimated from the precalculated average potential energies at a series of temperatures, which is again not effective enough for large systems. New algorithms with self-adaptive abilities are needed.

In summary, it remains a challenge to develop new methods to be used in studies of complex problems such as large-scale protein conformational changes, protein−ligand binding and protein aggregation, and, more generally, phase transitions. Sampling algorithms in trajectory space are also needed to further reduce the computational cost in obtaining large numbers of transition trajectories for kinetics calculations as well as for the quantitative study of nonequilibrium processes. It appears necessary to choose and combine different approaches that specifically fit the system of interest and its physical processes. Desired sampling methods should preserve both thermodynamic and kinetic information. For many methods, their abilities to reproduce correct thermodynamic information are still questionable. It is therefore important to test these methods using different complex and real systems.

## ■ ASSOCIATED CONTENT

### Supporting Information

Description of the basic concept of ITS, comparisons of the sampling efficiencies of ITS and several popular enhanced sampling methods, and enhanced sampling of reactive trajectories for chemical reactions. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: gaoyq@pku.edu.cn.


### Notes

The authors declare no competing financial interest.

### Biographies

**Lijiang Yang** received his B.S. (1997) from Nankai University and Ph.D. (2002) from the same institute (with Tianlun Chen). He worked at the University of California, Irvine (with R. Luo) and Texas A&M University (with Y. Q. Gao) as a postdoctoral researcher before joining Peking University as an associate research professor. His current research focuses on enhanced sampling method development with applications to biological systems.

**Cheng-Wen Liu** received his B.S. (2010) from Southwest University. He is now pursuing his Ph.D. at Peking University with Y. Q. Gao. He is interested in understanding the solvation of ions and chemical reactions at interface.

**Qiang Shao** received his B.S. (1999) from Nanjing University, M.S. (2004) from the College of William & Mary, and Ph.D. (2009) from Texas A&M University (with Y. Q. Gao). He was a postdoctoral researcher at Peking University and then joined Shanghai Institute of Materia Medica, Chinese Academy of Sciences (CAS) in 2011 as an associate professor. His research focuses on applying enhanced sampling methods to protein folding and protein−drug interactions.

**Jun Zhang** received his B.S. (2012) from Peking University. He is now pursuing his Ph.D. at Peking University with Y. Q. Gao. He is interested in developing and applying enhanced sampling methods in QM/MM simulations and computational biology.

**Yi Qin Gao** received his B.S. (1993) from Sichuan University, M.S. (1996) from the Institute of Chemistry, CAS, and Ph.D. (2001) from Caltech (with Rudolph A. Marcus). He did postdoctoral research with Martin Karplus at Harvard University and then started his independent career at Texas A&M University in 2004. His research group moved to Peking University in 2010. His laboratory is working in the field of theoretical and computational chemistry, developing efficient computational methods and statistical mechanics tools to study solvation effects in biomolecule structure formation and chemical reactions.

## ■ ACKNOWLEDGMENTS


This work was supported by the Ministry of Science and Technology of China through the 973 Program [Grants 2012CB917304 (to Y.Q.G.) and 2014CB910400 (to Q.S.)] and the National Natural Science Foundation of China [Grants 21233002 and 21125311 (to Y.Q.G.) and 21373016 (to L.Y.)]


## ■ REFERENCES


(1) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular simulation: A computational microscope for molecular biology. *Annu. Rev. Biophys* **2012**, *41*, 429−52.




DOI: 10.1021/ar500267n
*Acc. Chem. Res.* 2015, 48, 947−955

(2) Torrie, G. M.; Valleau, J. P. Non-Physical Sampling Distributions in Monte Carlo Free-Energy Estimation—Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(3) Frantz, D. D.; Freeman, D. L.; Doll, J. D. Reducing Quasi-Ergodic Behavior in Monte Carlo Simulations by J-Walking: Applications to Atomic Clusters. *J. Chem. Phys.* **1990**, *93*, 2769−2784.
(4) Bartels, C.; Karplus, M. Probability distributions for complex systems: Adaptive umbrella sampling of the potential energy. *J. Phys. Chem. B* **1998**, *102*, 865−880.
(5) Bussi, G.; Laio, A.; Parrinello, M. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* **2006**, *96*, No. 090601.
(6) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(7) Gao, Y. Q.; Yang, L. J. On the enhanced sampling over energy barriers in molecular dynamics simulations. *J. Chem. Phys.* **2006**, *125*, No. 114103.
(8) Grubmuller, H. Predicting Slow Structural Transitions in Macromolecular Systems—Conformational Flooding. *Phys. Rev. E* **1995**, *52*, 2893−2906.
(9) Lee, J.; Scheraga, H. A.; Rackovsky, S. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *J. Comput. Chem.* **1997**, *18*, 1222−1232.
(10) Voter, A. F. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **1997**, *78*, 3908−3911.
(11) Piela, L.; Kostrowicki, J.; Scheraga, H. A. The Multiple-Minima Problem in the Conformational-Analysis of Molecules—Deformation of the Potential-Energy Hypersurface by the Diffusion Equation Method. *J. Chem. Phys.* **1989**, *93*, 3339−3346.
(12) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169−9183.
(13) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(14) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(15) Berg, B. A.; Neuhaus, T. Multicanonical Algorithms for First Order Phase Transitions. *Phys. Lett. B* **1991**, *267*, 249−253.
(16) Wang, F. G.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(17) Kim, J.; Straub, J. E.; Keyes, T. Statistical-temperature Monte Carlo and molecular dynamics algorithms. *Phys. Rev. Lett.* **2006**, *97*, No. 050601.
(18) Marinari, E.; Parisi, G. Simulated Tempering—A New Monte Carlo Scheme. *Europhys. Lett.* **1992**, *19*, 451−458.
(19) Lin, Z.; van Gunsteren, W. F. Enhanced conformational sampling using enveloping distribution sampling. *J. Chem. Phys.* **2013**, *139*, No. 144105.
(20) Zheng, L. Q.; Chen, M. G.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227−20232.
(21) Zheng, W. W.; Rohrdanz, M. A.; Clementi, C. Rapid Exploration of Configuration Space with Diffusion-Map-Directed Molecular Dynamics. *J. Phys. Chem. B* **2013**, *117*, 12769−12776.
(22) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749−13754.
(23) Gao, Y. Q. An integrate-over-temperature approach for enhanced sampling. *J. Chem. Phys.* **2008**, *128*, No. 064105.
(24) Gao, Y. Q. Self-adaptive enhanced sampling in the energy and trajectory spaces: Accelerated thermodynamics and kinetic calculations. *J. Chem. Phys.* **2008**, *128*, No. 134111.
(25) Yang, L. J.; Gao, Y. Q. A selective integrated tempering method. *J. Chem. Phys.* **2009**, *131*, No. 214109.
(26) Yang, L. J.; Shao, Q.; Gao, Y. Q. Comparison between integrated and parallel tempering methods in enhanced sampling simulations. *J. Chem. Phys.* **2009**, *130*, No. 124111.
(27) Shao, Q.; Gao, Y. Q. Temperature Dependence of Hydrogen-Bond Stability in β-Hairpin Structures. *J. Chem. Theory Comput* **2010**, *6*, 3750−3760.
(28) Shao, Q.; Wei, H.; Gao, Y. Q. Effects of turn stability and side-chain hydrophobicity on the folding of β-structures. *J. Mol. Biol.* **2010**, *402*, 595−609.
(29) Shao, Q.; Gao, Y. Q. The relative helix and hydrogen bond stability in the B domain of protein A as revealed by integrated tempering sampling molecular dynamics simulation. *J. Chem. Phys.* **2011**, *135*, No. 135102.
(30) Shao, Q. Probing sequence dependence of folding pathway of α-helix bundle proteins through free energy landscape analysis. *J. Phys. Chem. B* **2014**, *118*, 5891−5900.
(31) Markiewicz, B. N.; Yang, L. J.; Culik, R. M.; Gao, Y. Q.; Gai, F. How Quickly Can a β-Hairpin Fold from Its Transition State? *J. Phys. Chem. B* **2014**, *118*, 3317−3325.
(32) Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L. C.; Knoester, J. Melting of a β-Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations. *J. Phys. Chem. B* **2010**, *114*, 10913−10924.
(33) Du, D.; Zhu, Y.; Huang, C. Y.; Gai, F. Understanding the key factors that control the rate of β-hairpin folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15915−15920.
(34) Olsen, K. A.; Fesinmeyer, R. M.; Stewart, J. M.; Andersen, N. H. Hairpin folding rates reflect mutations within and remote from the turn region. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15483−15487.
(35) Shao, Q.; Shi, J.; Zhu, W. Molecular dynamics simulation indicating cold denaturation of β-hairpins. *J. Chem. Phys.* **2013**, *138*, No. 085102.
(36) Dyer, R. B.; Maness, S. J.; Franzen, S.; Fesinmeyer, R. M.; Olsen, K. A.; Andersen, N. H. Hairpin folding dynamics: The cold-denatured state is predisposed for rapid refolding. *Biochemistry* **2005**, *44*, 10406−10415.
(37) Jang, S.; Kim, E.; Shin, S.; Pak, Y. Ab initio folding of helix bundle proteins using molecular dynamics simulations. *J. Am. Chem. Soc.* **2003**, *125*, 14841−14846.
(38) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. Atomically detailed folding simulation of the B domain of staphylococcal protein A from random structures. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14812−14816.
(39) Chen, L. X.; Shao, Q.; Gao, Y. Q.; Russell, D. H. Molecular Dynamics and Ion Mobility Spectrometry Study of Model β-Hairpin Peptide, Trpzip1. *J. Phys. Chem. A* **2011**, *115*, 4427−4435.
(40) Chen, L. X.; Gao, Y. Q.; Russell, D. H. How Alkali Metal Ion Binding Alters the Conformation Preferences of Gramicidin A: A Molecular Dynamics and Ion Mobility Study. *J. Phys. Chem. A* **2012**, *116*, 689−696.
(41) Fernandez-Lima, F. A.; Wei, H.; Gao, Y. Q.; Russell, D. H. On the structure elucidation using ion mobility spectrometry and molecular dynamics. *J. Phys. Chem. A* **2009**, *113*, 8221−8234.
(42) Song, Y.; Akin-Ojo, O.; Wang, F. Correcting for dispersion interaction and beyond in density functional theory through force matching. *J. Chem. Phys.* **2010**, *133*, No. 174115.
(43) Chai, J.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom−atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(44) Schwabe, T.; Grimme, S. Double-hybrid density functionals with long-range dispersion corrections: Higher accuracy and extended applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397−3406.
(45) Lenz, A.; Ojamae, L. On the stability of dense versus cage-shaped water clusters: Quantum-chemical investigations of zero-point energies, free energies, basis-set effects and IR spectra of $(H_2O)_{12}$ and $(H_2O)_{20}$. *Chem. Phys. Lett.* **2006**, *418*, 361−367.
(46) Furtado, J. P.; Rahalkar, A. P.; Shanker, S.; Bandyopadhyay, P.; Gadre, S. R. Facilitating Minima Search for Large Water Clusters at the MP2 Level via Molecular Tailoring. *J. Phys. Chem. Lett.* **2012**, *3*, 2253−2258.
(47) Liu, C.-W.; Wang, F.; Yang, L.; Li, X.-Z.; Zheng, W.-J.; Gao, Y. Q. Stable Salt−Water Cluster Structures Reflect the Delicate


DOI: 10.1021/ar500267n
*Acc. Chem. Res.* 2015, 48, 947−955

Competition between Ion−Water and Water−Water Interactions. *J. Phys. Chem. B* **2014**, *118*, 743−751.
(48) Xie, W. J.; Gao, Y. Q. A Simple Theory for the Hofmeister Series. *J. Phys. Chem. Lett.* **2013**, *4*, 4247−4252.
(49) Li, R.-Z.; Liu, C.-W.; Gao, Y. Q.; Jiang, H.; Xu, H.-G.; Zheng, W.-J. Microsolvation of LiI and CsI in Water: Anion Photoelectron Spectroscopy and ab Initio Calculations. *J. Am. Chem. Soc.* **2013**, *135*, 5190.
(50) Gajewski, J. J.; Jurayj, J.; Kimbrough, D. R.; Gande, M. E.; Ganem, B.; Carpenter, B. K. On the Mechanism of Rearrangement of Chorismic Acid and Related Compounds. *J. Am. Chem. Soc.* **1987**, *109*, 1170−1186.
(51) Elstner, M.; Porezag, G.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(52) Mills, G.; Jonsson, H. Quantum and Thermal Effects in $H_2$ Dissociative Adsorption—Evaluation of Free-Energy Barriers in Multidimensional Quantum Systems. *Phys. Rev. Lett.* **1994**, *72*, 1124−1127.
(53) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, No. 052301.
(54) Huber, G. A.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **1996**, *70*, 97−110.
(55) West, A. M. A.; Elber, R.; Shalloway, D. Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide. *J. Chem. Phys.* **2007**, *126*, No. 145104.
(56) Allen, R. J.; Frenkel, D.; ten Wolde, P. R. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *J. Chem. Phys.* **2006**, *124*, No. 024102.
(57) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(58) Moroni, D.; van Erp, T. S.; Bolhuis, P. G. Investigating rare events by transition interface sampling. *Physica A* **2004**, *340*, 395−401.
(59) Moroni, D.; Bolhuis, P. G.; van Erp, T. S. Rate constants for diffusive processes by partial path sampling. *J. Chem. Phys.* **2004**, *120*, 4055−4065.
(60) Fu, X. B.; Yang, L. J.; Gao, Y. Q. Selective sampling of transition paths. *J. Chem. Phys.* **2007**, *127*, No. 154106.
(61) Dimelow, R. J.; Bryce, R. A.; Masters, A. J.; Hillier, I. H.; Burton, N. A. Exploring reaction pathways with transition path and umbrella sampling: Application to methyl maltoside. *J. Chem. Phys.* **2006**, *124*, No. 114113.
(62) Yang, M.; Yang, L.; Gao, Y. Q.; Hu, H. Combine umbrella sampling with integrated tempering method for efficient and accurate calculation of free energy changes of complex energy surface. *J. Chem. Phys.* **2014**, *141*, No. 044108.
(63) Zhao, P.; Yang, L. J.; Gao, Y. Q.; Lu, Z. Y. Facile implementation of integrated tempering sampling method to enhance the sampling over a broad range of temperatures. *Chem. Phys.* **2013**, *415*, 98−105.



DOI: 10.1021/ar500267n
*Acc. Chem. Res.* 2015, 48, 947−955