UNIVERSITY OF CALIFORNIA, SAN DIEGO

Enhanced Sampling Development

for

Accessing Long Time Scale Protein Dynamics

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Levi C.T. Pierce

Committee in charge:

Professor J. Andrew McCammon, Chair
Professor Michael Galperin
Professor Ghorisankar Ghosh
Professor Mike Gilson
Professor Elizabeth A. Komives
Professor John Weare

2012

The dissertation of Levi C.T. Pierce is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2012

EPIGRAPH

Computers are incredibly fast, accurate, and stupid. Human beings are incredibly slow, inaccurate, and brilliant. Together they are powerful beyond imagination.

Albert Einstein

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Most importantly I thank my parents, David and Martha Pierce, who provided for all my needs, including encouragement and restraint, from my earliest years. My graduate research is a culmination of my father's attention to detail and incredible work ethic and my mother's creative spirit. I also thank Kristin Mangalindan who has given me great joy, love, and comfort providing a wonderful balance in my life outside of the lab.

I especially thank J. Andrew McCammon for his patient guidance in my graduate studies and his willingness to let me succeed or fail in directions that I chose for myself. From my masters work I thank Dr. Terry Clark and Yuh-Hwa Wang who shaped both my computational skills as well as my understanding of the biological sciences.

The McCammon group is a continually evolving group of brilliant minds which I have had the great pleasure of being a part of during my graduate studies. I thank Dr. Phineus Markwick for his guidance and mentorship in my initial graduate studies. I thank Dr. Ross Walker and Dr. Romelia Salomon-Ferrer for their help in my more recent work on the GPU. I would specially like to thank Sara Nichols whom has helped me develop as an independent researcher throughout my graduate years and been an exceptionally wonderful mentor in the lab. I thank Cesar Augusto F. de Oliveira, Bill Sinko, and Pat Blachly for their great ideas, conversations, and friendship. I also must thank Patti Craft who has been helpful in so many different ways throughout my time at UCSD.

Chapter II contains material from "Accelerating chemical reactions: Exploring reactive free-energy surfaces using accelerated *ab initio* molecular dynamics" published in 2011 in *Journal of Chemical Physics* (Volume 134, pages174107-1 to 174107-12), authored by Levi C.T. Pierce, Phineus R. L. Markwick, J. Andrew McCammon, and Nikos L. Doltsinis. All material has been reproduced with the consent of all other authors.

Chapter III contains material from "Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam" published in 2011 in *Journal of Physical Chemistry Letters* (Volume 2, pages 158-164), authored by, Phineus R. L. Markwick, Levi C.T. Pierce, David B. Goodin, and J. Andrew McCammon. All material has been reproduced with the consent of all other authors.

Chapter IV contains material from "Routine Access to Millisecond time scales with accelerated molecular dynamics" (submitted) to *Journal of chemical theory and computation,* authored by Levi C.T. Pierce, Romelia Salomon-Ferrer, Cesar Augusto F. de Oliveira, J. Andrew McCammon**,** and Ross C. Walker. All material has been reproduced with the consent of all other authors.

VITA

1981 Born Salt Lake City, Utah

2000-2005 Bachelor of Science Computer Engineering, University of Kansas

2005-2008 Masters of Science Computer Engineering, University of Kansas

2008-2012 Doctor of Philosophy Chemistry, University of California San Diego

PUBLICATIONS

Pierce, L.C.T.; Salomon-Ferrer, R.; de Oliveira C.A.; McCammon, J.A.; Walker, R.C.; Routine Access to Millisecond Time scales with Accelerated Molecular Dynamics. in press

Friedman, A.J.; Durrant, J.D.; Pierce, L.C.; McCammon, J.A. The Molecular Dynamics of Trypanosoma brucei UDP-Galactose-4'-Epimerase: A Drug Target for African Sleeping Sickness. *Chemical Biology & Drug Design.* 2012 May 2; doi: 10.1111/j.1747-0285.2012.01392.x

Lindert, S.; Kekenes-Huskey, P.; Huber, G.; Pierce, L.C.T.; McCammon, J.A.; Dynamics and Calcium Association to the N-Terminal Regulatory Domain of Human Cardiac Troponin C: A Multi-Scale Computational Study. *J Chem Phys B* 2012 Feb. 14. doi: 10.1021/jp212173f

Sinko, W.; de Oliveira, C.A.; Pierce, L.C.T.; McCammon, J.A.; Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations. *J Chem Theory Comput.* 2012 Jan. 10;8(1):17-23. doi: 10.1021/ct200615k

Pierce, L.C.T.; Markwick, P.R.L; McCammon, J.A.; Doltsinis, N.L. Accelerating chemical reactions: exploring reactive free-energy surfaces using accelerated *ab initio* molecular dynamics. *J Chem Phys.* 2011 May 7;134(17):174107. doi: 10.1063/1.3581093

Bucher, D.; Pierce, L.C.T.; Markwick, P.R.L; McCammon, J.A. On the Use of Accelerated Molecular Dynamics to Enhance Configurational Sampling in *ab initio* Simulations. *J Chem Theory Comput.* 2011 Apr 12;7(4):890-897. Epub 2011 Mar 4. doi: 10.1021/ct100605v *Cover Feature*

Markwick P.R.L; Pierce, L.C.T.; Goodin D.B.; McCammon, J.A.; Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam. *J Phys Chem Lett.* 2011 Feb 3;2(3):158-164. Epub 2011 Jan 7. doi: 10.1021/jz101462n

FIELDS OF STUDY

Major Field: Physical Chemistry

Studies in computational methods for enhanced sampling of protein dynamics.

Professor J. Andrew McCammon

ABSTRACT OF THE DISERTATION

Enhanced Sampling Development

for

Accessing Long Time Scale Protein Dynamics

by

Levi C.T. Pierce

Doctor of Philosophy in Chemistry

University of California, San Diego, 2012

Professor J. Andrew McCammon, Chair

Computational modeling has played a great role in solving many questions in biochemical and biomedical research. However, many biologically relevant processes occur on long time scales, which are inaccessible to conventional modeling techniques. Accessing these long time scales has been a great challenge for computational scientists, which has led to the development of numerous methods for enhanced sampling. In this work the well established accelerated molecular dynamics (aMD) method is implemented into several different codes in both a classical mechanics framework as well as a quantum mechanics framework in order to access events occurring on time scales otherwise inaccessible with conventional molecular dynamics codes.

Recently there has been a great interest in developing codes to run on graphics processing units (GPUs), which, have been shown to be well suited for conventional molecular dynamics. The research presented in this dissertation shows that the combination of the highly parallelized, inexpensive GPU and the efficient enhanced sampling method, aMD, allow access to events occurring on the millisecond time scale. Importantly, this development is made available to the general scientific community with the release of the Amber12 simulation package.

I Accessing events on various time scales with simulation

Physical and biological events occur on a range of time scales spanning 12 orders of magnitude, from femtoseconds (10^{-15} s), less than the time it takes for a chemical bond to vibrate, to milliseconds (10^{-3}s), the time it takes for some proteins to fold. Numerous other interesting events fall between these two extremes such as proton transfer occurring on a picosecond (10^{-12}s) time scale, side chain flips and rotations occurring on the nanosecond (10^{-9}s) time scale, and large conformational changes occurring on the microsecond (10^{-6}s) time scale. While numerous experimental techniques can access these various time scales an atomistic representation is most often lacking or difficult to acquire. Computational modeling and simulation can be used to provide an all atom description, which can be used to reproduce experimental observables.

Ultimately, one would like to not only validate but also predict experimental results but we are limited in both a temporal and spatial regime due to the computational cost of running long simulations and the complexity of the simulation being conducted. Currently, conventional MD (cMD) carried out on large super computers for several weeks can access time scales up to several microseconds for moderately sized proteins (1500 residues). One of the great recent advancements in the field of computational science has been David E. Shaw's specialized super computer, Anton, which has allowed one of the first millisecond long simulations (1,000,000ns) to be carried out on BPTI[1]. While the DE Shaw Research group has graciously provided time on an Anton machine, access is still limited to only a handful

of research groups. Therefore, the challenge of accessing biological events on long time scales using conventional resources is still a very active area of research.

Large conformational changes occurring within a protein are often associated with long time scale dynamics neatly characterized by SAXS and NMR experiments. These large conformational changes are typically characterized by infrequent transition events, the protein moves from one-energy minima to another, usually involving a high-energy barrier. Using transition-state theory one can compute rate constants directly for these transition processes and begin to build a coarse network of transition probabilities. If one knows *a priori* the reaction pathway there are numerous methods available to drive the system along the pathway such as meta-dynamics, adaptive biasing force, and umbrella sampling. However, one generally has no idea what pathway a biomolecule might take when transitioning from one state to another and therefore, some variant of conventional MD must be applied to explore the energy landscape. Furthermore, a pathway cannot be defined if only one conformation of a protein is known *a priori* and therefore, methods that assume an initial and final state are not applicable.

The essential problem of cMD is our protein of interest remains trapped in a metastable state for a long time effectively showing no exploration of conformational space in the conventional timespan of a typical simulation. In order to increase the transition rate of escape out of these metastable states several forms of enhancing transition rates have been developed, namely, parallel-replica dynamics[2], hyperdynamics[3], and temperature-accelerated dynamics[4]. The focus of this work is based on hyperdynamics first proposed by Arthur F. Voter, which, constructs an

auxiliary system in such a way that the dynamics are faster than cMD, while enforcing that the modified system maps onto the unmodified system by a suitable renormalization of time[3].

In this approach, the auxiliary system is obtained by adding a nonnegative bias potential $\Delta V_b(\mathbf{r})$ to the potential of the original system $V(\mathbf{r})$ so that the height of the barriers between different states is effectively reduced. Following the approach of Voter, Donald Hamelberg applied this technique to a biomolecule system within the framework of the Amber simulation package, calling it accelerated molecular dynamics (aMD). Accelerated MD modifies the energy landscape by adding a boost potential, $\Delta V(r)$, to the original potential energy surface when $V(r)$ is below a pre-defined energy level E [5], as

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \frac{(E-V(r))^2}{\alpha + (E-V(r))} & V(r) < E \end{cases} \quad \text{(I.1)}$$

where α modulates the depth and the local roughness of the energy basins on the modified potential. In principle, this approach also allows the correct canonical average of an observable, calculated from configurations sampled on the modified potential energy surface, to be fully recovered from the accelerated MD simulations. In order to simultaneously enhance the sampling of internal and diffusive degrees of freedom a dual boosting approach was employed, based on separate torsional and total boost potentials as [6]

$$V(r) = V_o(r) + V_t(r) \quad \text{(I.2)}$$

$$V^*(r) = \{V_o(r) + [V_t(r) + \Delta V_t(r)]\} + \Delta V_T(r) \quad \text{(I.3)}$$

where $V_o(r)$ is the original potential, $V_t(r)$ is the total potential of the torsional terms, $\Delta V_t(r)$ and $\Delta V_T(r)$ are the boost potentials applied to the torsional terms $V_t(r)$ and the total potential energy $V_T(r)$, respectively.

Since the first implementation by Hamelberg[5] the aMD method has been successfully applied to study numerous biological systems[7-9]. While efficient conformational space exploration can be carried out without any prior knowledge of the energy landscape, obtaining time-correlated observables is currently not possible with a single aMD simulation. Another limitation of the method is it is often difficult to recover statistics on the unbiased energy surface due to large perturbations in the bias, which give rise to large weights associated with very few of the snapshots from the simulation. Given these limitations the aMD method is well suited for generating a suite of starting conformations which can be coupled with more rigorous free energy methods such as ABF to provide an overall accurate description of the free energy landscape.[10] In this work several new implementations are described including an implementation in an *ab initio* framework and a fast and efficient classical MD version, which has been optimized for highly parallelized computing and graphics processing units (GPUs).

The advancement of computational science on conventional graphic processing units (GPUs) has allowed researchers efficient and inexpensive access to 10s to 100s of microseconds of simulation time on just a single desktop computer[9,11,12]. Several groups have been utilizing the power of GPUs, which has been driven by the introduction of a simple application programming interface called compute unified device architecture (CUDA). Using CUDA M. J. Harvey and G. De Fabritiis released

commercial software, ACEMD in 2009, which was one of the first codes to run explicit water molecular dynamics simulations on the GPU[11]. In the same year Vijay Pande and the OPENMM consortium released an open source CUDA cMD code, which was linked into the GROMACS engine[12]. More recently John E. Stone and Klaus Schulten have developed a GPU implementation for NAMD and perhaps more importantly several analysis tools within VMD which take advantage of the GPU[13,14].

One of the fastest GPU implementations, developed by Scott Le Grand and Ross Walker, is built into the framework of the Amber molecular dynamics package[15,16]. The Amber implementation runs the entire simulation on the GPU whereas other codes such as NAMD only use the GPU to compute non-bonded interactions. One important question to ask is how long will this technology be around because a great deal of time is needed to port scientific codes to the GPU? With the gaming industry ultimately driving the demand for GPU technology and not the computational science community these devices will be around for the next five years at least and will be used to push the bounds of science.

Chapter 2 presents the implementation of the aMD method within the Carr-Parrinello molecular dynamics code (CPMD). In this work it is demonstrated that access to events that cannot be captured using conventional CPMD and would normally take a long time to simulate, such as proton transfer and the rearrangement of covalent bonds can be enhanced using aMD. Using two examples, the double proton transfer reaction in formic acid dimer (FAD) and the hypothetical adiabatic ring-opening and subsequent rearrangement reactions in methylenecyclopropane (MCP), it is demonstrated that *ab initio* aMD can be readily employed to efficiently explore the

reactive potential energy surface (PES), allowing the prediction of chemical reactions and the identification of meta-stable states.

Chapter 3 presents an aMD implementation that allows the simulation to adaptively change the acceleration as it explores phase space. Adaptive aMD (Ad-aMD) is an efficient and robust conformational space-sampling algorithm that is particularly well suited to proteins with highly structured potential energy surfaces exhibiting complex, large-scale collective conformational transitions. Ad-aMD simulations of substrate-free P450cam reveal that this system exists in equilibrium between fully and partially open conformational states. The mechanism for substrate binding depends on the size of the ligand. Larger ligands enter the P450cam binding pocket, and the resulting substrate-bound system is trapped in an open conformation via a population shift mechanism. Small ligands, which fully enter the binding pocket, cause an induced-fit mechanism, resulting in the formation of an energetically stable closed conformational state. These results are corroborated by recent experimental studies and potentially provide detailed insight into the functional dynamics and conformational behavior of the entire cytochrome-P450 super-family.

Chapter 4 presents the synthesis of the parallel power of conventional gaming graphics cards with the enhanced sampling method of aMD. This work combines the enhanced sampling method, accelerated molecular dynamics (aMD) with the inherent power (as implemented in Amber) of graphics processor units (GPUs) and is applied to the study of Bovine Pancreatic Trypsin Inhibitor (BPTI). A 500ns aMD simulation is compared to a previous millisecond unbiased brute force MD simulation carried out on BPTI showing the same conformational space is sampled by both approaches. The

correct relative populations defined by the χ_1, χ_2, and χ_3 dihedral angles of the disulfide bond C14-C38 are observed and improved agreement with observed chemical shift differences from prior experimental work is obtained. To our knowledge this represents the first implementation of aMD on GPUs and also the longest aMD simulation of a biomolecule run to date. Our implementation will be made available to the community with the release of the Amber software suite (v12) enabling researchers routine access to ms events sampled from dynamics simulations using off the shelf hardware.

I.A References

(1) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Science 2010, 330, 341.

(2) Voter, A. F. Physical Review B 1998, 57, R13985.

(3) Voter, A. F. Physical Review Letters 1997, 78, 3908.

(4) Sorensen, M. R.; Voter, A. F. J Chem Phys 2000, 112, 9599.

(5) Hamelberg, D.; Mongan, J.; McCammon, J. A. J Chem Phys 2004, 120, 11919.

(6) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. J Chem Phys 2007, 127, 155102.

(7) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. PLoS Comput Biol 2009, 5, e1000325.

(8) de Oliveira, C. s. A. F.; Grant, B. J.; Zhou, M.; McCammon, J. A. PLoS Comput Biol 2011, 7, e1002178.

(9) Cervantes, C. F.; Markwick, P. R.; Sue, S. C.; McCammon, J. A.; Dyson, H. J.; Komives, E. A. Biochemistry 2009, 48, 8023.

(10) Wereszczynski, J.; McCammon, J. A. Proc Natl Acad Sci U S A 2012.

(11) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. Journal of Chemical Theory and Computation 2009, 5, 1632.

(12) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. Journal of Computational Chemistry 2009, 30, 864.

(13) Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. J Mol Graph Model 2010, 29, 116.

(14) Levine, B. G.; Stone, J. E.; Kohlmeyer, A. J. Comput. Phys. 2011, 230, 3556.

(15) Andreas W. Götz, M. J. W., Dong Xu, Duncan Poole, Scott Le Grand, and Ross C. Walker.

(16) Romelia Salomon-Ferrer, A. W. G., Duncan Poole, Scott Le Grand, and Ross C. Walker.

II Accelerating chemical reactions: Exploring reactive free-energy surfaces using accelerated *ab initio* molecular dynamics

II.A Abstract

A biased potential molecular dynamics simulation approach, accelerated molecular dynamics (aMD), has been implemented in the framework of *ab initio* molecular dynamics for the study of chemical reactions. Using two examples, the double proton transfer reaction in formic acid dimer (FAD) and the hypothetical adiabatic ring-opening and subsequent rearrangement reactions in methylenecyclopropane (MCP), it is demonstrated that *ab initio* aMD can be readily employed to efficiently explore the reactive potential energy surface (PES), allowing the prediction of chemical reactions and the identification of meta-stable states. An adaptive variant of the aMD method is developed, which additionally affords an accurate representation of both the free energy surface and mechanism associated with the chemical reaction of interest and can also provide an estimate of the reaction rate.

II.B Introduction

The study of chemical reactions using quantum chemistry calculations is a well-established and active field of research. The general protocol for such studies involves the determination of local energy minima (reactant and product states) and saddle points (transition states) on the potential energy surface (PES). An optimal reaction pathway, the minimum energy path (MEP), connecting the reactant, transition and product states is then determined along with the associated energy profile.[1-9]

Despite the fact that these methods are well established, the determination of transition states can be very difficult and computationally expensive. In general, prior knowledge of the PES is required including at the very least an accurate representation of the reactant and product states as well as any putative intermediates. Furthermore, these strategies often fail for systems with a large number of degrees of freedom, when entropic effects are important and the *free* energy surface needs to be explored.

In principle, *ab initio* molecular dynamics (AIMD) simulations[10] are ideally suited to exploring free energy surfaces of complex systems, however, the potential of AIMD to study chemical reactions is hindered by the fact that these processes occur on time-scales that are significantly longer than those accessible using standard *ab initio* MD methodologies. Chemical reactions, i.e. the cleavage or formation of one or more covalent bonds, and conformational rearrangements within a molecule occur when a system migrates from one local energy minimum to another. The energy barriers that separate the reactant and product states usually vary from 5-20 kcal/mol and therefore chemical reactions occur on time-scales ranging from hundreds of nanoseconds to milliseconds.

Despite the sustained and rapid increase in available computational power and the development of efficient simulation algorithms, AIMD simulations of even small isolated molecules are generally limited to time-scales of hundreds of picoseconds. In the last two decades considerable progress has been made in the development of more sophisticated methods to explore both the configurational and reaction space of molecular systems more efficiently,[11,12] allowing the study of slow molecular motions and rare events. In general, these methods can be divided into two groups: The first

involves the identification of transition pathways between known initial (reactant) and final (product) states or for a known reaction coordinate. Such methods include targeted molecular dynamics,[13,14] (and constrained dynamics in general[15]) transition path sampling[16,17] and essential molecular dynamics.[18] The second group contains those methods that efficiently explore large areas of the PES without necessarily requiring knowledge of the reaction coordinate and without imposing constraints, allowing for the rapid identification of thermodynamically dominant regions. These methods include replica exchange MD methods,[19,20] meta-dynamics,[21] hyper dynamics,[22] and accelerated molecular dynamics (aMD).[23] The principle behind aMD is to add a continuous, non-negative bias to the actual potential energy surface which raises the low energy regions on the potential energy landscape, decreasing the magnitude of the energy barriers and accelerating the exchange between low energy configurational states, while still maintaining the essential details of the underlying PES. One of the most favorable characteristics of this method is that it yields a canonical average of an observable, so that thermodynamic and other equilibrium properties can be accurately determined.

In the framework of classical molecular dynamics, aMD has already been successfully employed to study the slow time-scale dynamics of poly-peptides[24,25] and proteins, such as H-Ras,[26] ubiquitin[27] and IκBα.[28] The enhanced conformational space sampling afforded by aMD in these studies was shown to significantly improve the theoretical prediction of experimental NMR observables, such as residual dipolar couplings,[27,28] scalar J-couplings[29] and chemical shifts,[30] which are sensitive to dynamic averaging on the micro- to millisecond time-scale. As a robust free energy

sampling method, classical aMD has also been successfully combined with molecular modeling approaches to study the conformational behavior of natively unstructured proteins.[31] aMD used in conjunction with classical MD is however severely limited in its applicability due to the limitations of standard force fields, in particular their inability to describe chemical reactions.

In this paper, we therefore explore the possibility of using the accelerated molecular dynamics approach in the framework of *ab initio* molecular dynamics, specifically for the study of chemical reactions. We have employed standard aMD and developed a novel variant of aMD, *adaptive aMD* (Ad-aMD) in the framework of Car-Parrinello molecular dynamics (CP-MD) to study two chemical reactions: The well-known double proton transfer (DPT) event in formic acid dimer (FAD) and the (hypothetical) adiabatic ring-opening and rearrangement reactions of methylenecyclopropane (MCP). The DPT reaction of FAD and the cyclic rearrangement of MCP are depicted diagrammatically in Fig. 1(a) and 2(a), respectively. Using these two test systems, we show that the *ab initio* aMD approach and its adaptive variant can be used to predict chemical reactions, to obtain accurate free energy statistics for the chemical reaction of interest and also to afford a meaningful estimate of the associated reaction rate.

II.C Methods

In the following work, we have employed three *ab initio* aMD protocols: '*Standard ab inito aMD*' to explore the reactive PES, '*Adaptive ab initio aMD*' (*Ad-*

aMD) to obtain an accurate estimate of the free energy of the chemical reaction, and a specific protocol of Ad-aMD used to obtain a reaction rate constant. These three protocols are discussed in detail below.

II.C.1 Standard *ab initio* aMD

Following the work of Hamelberg *et al.*,[23] in the standard aMD formalism, a continuous non-negative bias potential, $\Delta V(\mathbf{r})$ is defined such that when the true (underlying) potential of the system, $V(\mathbf{r})$, is below a certain, pre-defined threshold 'boost' energy, E_b, the simulation is performed on a modified potential, $V^*(\mathbf{r}) = V(\mathbf{r}) + \Delta V(\mathbf{r})$, but when $V(\mathbf{r}) > E_b$, the simulation is performed on the true potential [$V^*(\mathbf{r}) = V(\mathbf{r})$]. The modified potential, $V^*(\mathbf{r})$ is related to the true potential, $V(\mathbf{r})$, bias potential, $\Delta V(\mathbf{r})$ and boost energy, E_b, by: [23]

$$V^*(r) = V(r) \quad , \qquad V(r) \geq E_b \tag{II.1}$$

$$V^*(r) = V(r) + \Delta V(r) \quad , \qquad V(r) < E_b$$

and the bias potential, $\Delta V(\mathbf{r})$ is defined as:

$$\Delta V(r) = \frac{\left(E_b - V(r)\right)^2}{\alpha + E_b - V(r)} \tag{II.2}$$

In the framework of CP-MD, the true potential, $V(\mathbf{r})$, is defined as the density functional energy. The application of this bias potential, results in raising the potential energy wells and thereby a flattening of the PES, thus enhancing the exchange rate between low energy states. The extent of acceleration (i.e. how aggressively we enhance the configurational space sampling) is determined by the choice of the boost energy, E_b and the acceleration parameter, α. Configurational space sampling can be enhanced by either increasing the boost energy, or decreasing α. A schematic representation of the standard aMD protocol is shown in Fig. 3(a). It is important to notice that in the 'standard *ab initio* aMD' protocol, both acceleration parameters, E_b, and α, are constants.

The aMD method also yields correct canonical averages of an observable so that thermodynamic and other equilibrium properties of the system can be accurately determined. The corrected canonical ensemble average of the system is obtained by simply re-weighting each point in the configuration space on the modified potential by the strength of the Boltzmann factor of the bias potential, $exp(\beta\Delta \mathrm{V}[\mathbf{r}(t_i)])$, at that particular point. When the system is on the normal potential, the bias is zero. The reader is referred to reference 23 for more details.

II.C.2 *ab initio* Adapative aMD (Ad-aMD)

In the standard aMD methodology, the system evolves on a modified PES which, due to the application of the bias potential, allows for enhanced configurational space sampling. The standard aMD protocol provides an approximate, albeit

somewhat crude representation of the true underlying PES. However, as is demonstrated in the Results section, this protocol does not allow one to specifically focus on a particular rare event or chemical reaction of importance. Indeed, in most cases, the system samples a large amount of high-energy configurational space, much of which is of no particular interest. In order to obtain accurate free energy statistics for a particular reaction, it is necessary to observe this specific process multiple times over. One means of selectively accelerating a particular event is adaptive aMD (Ad-aMD). In the Ad-aMD protocol, one defines a specific reaction coordinate, ϕ, that accurately describes the particular reaction of interest, and accelerates the system along the pre-defined reaction coordinate. Using the approximate representation of the true PES, $V(\phi)$ obtained from the standard aMD simulation, a specific desired modified potential, $V^*(\phi)$ is defined. Ideally the modified potential reflects all the characteristics of the underlying true PES, however, an exact reproduction of all the characteristics is not necessary to obtain accurate free energy statistics (as demonstrated below in the case of MCP). In the Ad-aMD protocol, one of the acceleration parameters, α, is held fixed and the other, E_b is adapted as a function of the reaction coordinate using the approximate representation of the true potential energy, $V(\phi)$, to obtain the desired, pre-defined modified potential, $V^*(\phi)$. The functional form for $E_b(\phi)$ is obtained by simply re-arranging equation (2) as:

$$E_b(\phi) = \frac{(V^*(\phi) - V(\phi) \pm \sqrt{\left(V^*(\phi) - V(\phi)\right)^2 + 4\alpha\left(V^*(\phi) - V(\phi)\right)}}{2} + V(\phi) \quad \text{(II.3)}$$

As can be seen, there are two solutions for $E_b(\phi)$. In the present work, we always use the solution associated with the positive.

In the Ad-aMD simulation, when the pre-defined modified potential, $V^*(\phi)$ lies above the true underlying potential, $V(\phi)$, the system evolves on the modified potential energy surface, $V^*(\phi)$ and the associated adaptive bias potential is given as:

$$\Delta V(\phi) = \frac{\left(E_b(\phi) - V(\phi)\right)^2}{\alpha + E_b(\phi) - V(\phi)} \tag{II.4}$$

However, when the pre-defined modified potential lies below the true underlying potential, the acceleration is switched off and the system evolves on the true PES, $V(\phi)$. In this way, during the Ad-aMD simulation, the system is only accelerated across a certain region of the reaction coordinate, which defines the reaction sub-space of interest. The desired modified potential, $V^*(\phi)$ can be defined and positioned relative to the true potential, $V(\phi)$ at will in a physically meaningful fashion. In the present examples (FAD and MCP), a simple polynomial function was used to describe the modified potential and an initial estimate of $E_b(\phi)$ was generated using the results of the standard aMD trajectories. Due to thermal fluctuations and local molecular distortions induced by the application of the bias potential, the initial representation of the true underlying potential across the reaction coordinate, $V(\phi)$ obtained from the standard aMD simulation is not particularly accurate. As a result of this, the initial

estimate of the adaptive boost potential, $E_b(\phi)$, which is calculated using equation (3) does not immediately have the desired effect of the fixed pre-defined modified potential, $V^*(\phi)$. In order to overcome this problem, both the underlying potential, $V(\phi)$ and the adaptive boost potential, $E_b(\phi)$ are updated in a history-dependent fashion during the Ad-aMD simulation: The reaction coordinate, ϕ, is divided into bins, and during the Ad-aMD simulation, the true potential energy, $V(\phi)$ in each bin is averaged.

The enhanced sampling of the reaction space provides an increasingly accurate representation of the true underlying potential, $V(\phi)$. In the present case, $E_b(\phi)$ is updated every 500 MD steps using the improved potential energy statistics for $V(\phi)$, however the updating frequency is an adjustable parameter. The adaptive aMD approach can be regarded as a generalized extension of the *local boost method* (LBM) developed by Wang et. al.[32] For the sake of consistency, we have chosen to use the same functional form of the bias potential in both standard and adaptive aMD protocols. However, it should be recognized that having explicitly defined the desired modified potential, any functional form of the bias potential could be implemented. A diagrammatic representation of adaptive aMD is shown in Figure II.3(b). Specific details of the modified potential employed for the FAD and MCP systems are described in the Results section.

A free energy profile for the chemical reaction can be readily calculated from the Ad-aMD trajectory using the canonical ensemble free energy re-weighting protocol: The reaction coordinate that describes the chemical reaction of interest is divided into bins and the structures collected across the Ad-aMD trajectory are

allocated to their respective bin, ϕ_j. The effective population statistic for each structure, i, is given by $exp(\beta\Delta V[\phi(t_i)])$. For each bin, the population statistics are summed across the entire Ad-aMD trajectory to give an effective total population in that bin, $pop(\phi_j)$. The free energy profile, $\Delta G(\phi_j)$, is then given as:

$$\Delta G\left(\phi_j\right) = -RT\ln\left[\frac{pop\left(\phi_j\right)}{pop\left(\phi_{\max}\right)}\right] \quad , \tag{II.5}$$

where $pop(\phi_{max})$ is the effective total population of the most populated bin.

II.C.3 Using adaptive aMD to estimate reaction rates

In both standard and adaptive aMD the system evolves on a modified potential at an accelerated rate with a non-linear time-scale of Δt^*, given as: [23]

$$\Delta t_i^* = \Delta t \exp\left(\beta\Delta V\left[r\left(t_i\right)\right]\right) \quad , \tag{II.6}$$

where Δt is the actual time-step of the simulation on the modified potential. In principle therefore, it is possible to estimate the time-scale of events observed during the aMD simulations as:

$$t^* = \sum_i^N \Delta t_i^* = \Delta t \sum_i^N \exp\left(\beta \Delta V\left[r\left(t_i\right)\right]\right) \tag{II.7}$$

$$t^* = t\left\langle \exp\left(\beta \Delta V\left[r\left(t_i\right)\right]\right)\right\rangle \ ,$$

where N is the total number of molecular dynamics steps performed during the whole simulation, and $\langle exp(\beta \Delta V[\boldsymbol{r}(t_i)]) \rangle$ is the boost factor.

However, according to transition state theory (TST),[33,34] the above equation is only correct if the boost energy, E_b, lies below the entire transition state region. In the standard aMD protocol, it is rarely possible to achieve a sufficient level of acceleration to observe the desired transition while fulfilling this criterion. However, the same is not true for adaptive aMD. Once the free energy profile as a function of the chosen reaction coordinate is accurately determined, the adaptive boost energy and acceleration parameter, α, can be readily adjusted, to generate a modified potential with a sufficiently small energy barrier to observe the reaction of interest, while ensuring that the boost energy lies below the entire transition state region. This adaptive aMD protocol is shown schematically in Fig. II.3(c). Under these conditions, equation (I.7) fulfills the TST criteria and a meaningful estimate of the reaction rate can be obtained.

II.C.4 Computational Details

All molecular dynamics simulations were carried out at T=300K using an in-house modified version of the CPMD 3.12 package.[35] In the case of FAD, the system was placed in the center of a periodically repeating cubic box of side length L=25 a.u.

and the Becke (B) exchange[36] and Lee, Yang, Parr (LYP) correlation[37] functional was employed. For MCP, the system was placed in a periodically repeating cubic box of side length L=20 a.u. and the gradient corrected PBE density functional[38] was employed. For both systems, a fictitious electron mass of 400 a.u. was ascribed to the electronic degrees of freedom and the coupled equations of motion were solved using the velocity Verlet algorithm with a time-step of 4 a.u. For FAD, core electrons were treated using the norm-conserving pseudo-potentials of Troullier and Martins[39] and the valence orbitals were expanded in a plane-wave basis up to an energy cut-off of 70 Ry.

For MCP, ultrasoft pseudopotentials[40] were employed and the valence orbitals were expanded in a plane-wave basis set up to an energy cut-off of 25 Ry. Standard CP-MD simulations were performed using a Nosé-Hoover thermostat[41] on the ions. In the case of the *ab initio* aMD and adaptive aMD simulations, a Nosé-Hoover thermostat was applied to both the electronic and nuclear degrees of freedom. All *ab initio* aMD and adaptive aMD simulations presented in this work were performed within the framework of CP-MD, however, we would like to point out that the biased potential simulation methods described here could equally well be implemented in the framework of Born-Oppenheimer MD.

The general protocol followed in this work starts with a standard CP-MD simulation. This simulation provides an estimate for the true underlying density functional energy of the system in the thermodynamically stable state at 300K, V_0, and provides the initial coordinates for the biased potential simulations. Standard aMD simulations are then performed at increasing levels of acceleration in order to explore the configurational space of the system, to identify potential chemical reactions and

meta-stable states and to acquire an approximate, albeit rather crude representation of the underlying PES. Using this information, a suitable reaction coordinate is defined and an adaptive aMD simulation is performed providing an accurate representation of the free energy surface for the reaction. Once an accurate description of the free energy surface is available, the adaptive aMD parameters can be selected in order to obtain an estimate of the reaction rate.

II.D Results

II.D.1 Exploring the configurational space of FAD using *ab initio* aMD

In order to explore the configurational space of FAD, a series of *ab initio* aMD simulations across a variety of acceleration levels were performed. In all simulations, the acceleration parameter, α, was fixed at 0.016 a.u (10.04 kcal/mol) and the level of acceleration was controlled by varying the boost energy, E_b. The optimal acceleration level for enhanced configurational space sampling was observed when (E_b-V_0) was set between 0.018 a.u. (11.3 kcal/mol) and 0.024 a.u. (15.1 kcal/mol). Within this acceleration regime, FAD rapidly exchanges between numerous 'states', which, are depicted in Fig. II.1. The variation in the true, underlying potential energy obtained over a 50,000 step segment from a representative aMD trajectory is depicted in Fig. II.4a and the different 'states' that are visited are indicated. The system clearly spends a large amount of time exploring high energy regions of configurational space compared to the two thermodynamically stable dimer states (shown in Fig. II.1(a)).

These high energy 'states' include an extended dimer configuration (Fig. II.1(b)), where both hydrogen bonds are longer than 2.5Å and the system comes close to complete dissociation and a twisted configuration (Fig. II.1(d)), in which one hydrogen bond is maintained and the angle subtended between the two monomer planes increases by more than 90^o. Indeed, in one aMD simulation a complete 360^o rotation of one of the monomeric units was observed. In addition to these motions, we also observe anti-correlated rotation of the two monomeric units in the dimer plane and a 'wagging' motion (tilting of the monomers out of the plane of the dimer). All these motions can clearly be identified as extended collective inter-monomer vibrational modes that are readily observed in standard CP-MD simulations of FAD. Around step No. 42200 (Fig. II.4(b)), we observe a double proton transfer (DPT) event.

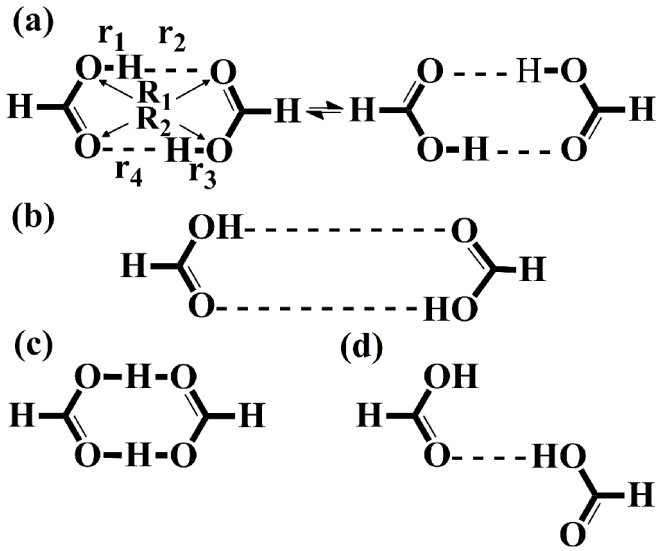

Figure II 1 (a) Diagrammatic representation of the DPT reaction in formic acid dimer. The specific interatomic distances required to define the reaction co- ordinate ρ1 are labeled (see text for more details). (b) The extended formic acid dimer system. (c) The transition state structure associated with DPT. (d) The twisted formic acid dimer.

Thermal fluctuations combined with the presence of some structural distortions due to the application of the bias potential result in a rather noisy and somewhat crude representation of the of the underlying PES shown in Fig. II.4(a). Nevertheless, particularly when averaging the potential energy statistics across the aMD trajectory, it is possible to obtain an approximate, albeit rather crude estimate of the variations in the potential energy of the system as it evolves from one configurational state to another. For example, the thermodynamically stable dimer state at 300K (Fig. II.1(a)) has an approximate potential energy of -77.629 a.u. (+/- 0.04 a.u.), which is very similar to the average potential energy obtained from the standard CP-MD simulations (-77.630 a.u.) and is referred to from here on as V_0. The potential energy for an extended dimer conformation, as the system approaches dissociation, is seen to lie on average approximately 0.015 - 0.020 a.u. (9.5 - 12.5 kcal/mol) above V_0 and the associated potential energy barrier for DPT is approximately 0.009 - 0.012 a.u. (5.7 – 7.5 kcal/mol). These values obviously only provide an approximate representation of the true underlying PES. For the sake of comparison, the true potential energy barrier to DPT on the minimum energy path is 5.4 kcal/mol. We would like to point out that the aim of the standard aMD simulations is not to obtain accurate energetic statistics, but rather to enhance the configurational space sampling and to identify local energy minima and transition states on the PES.

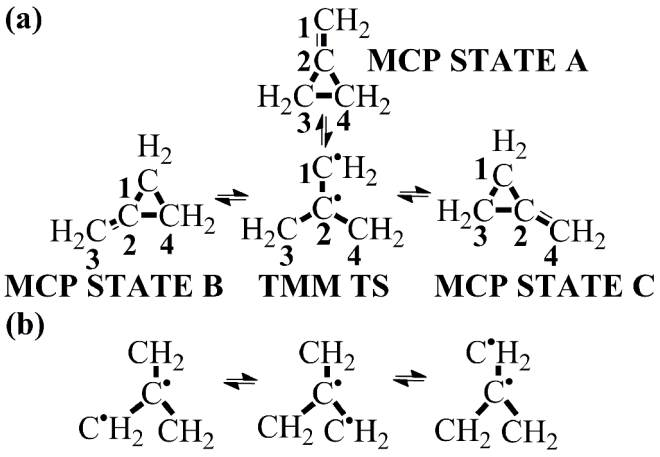

Figure II 2 (a) Diagrammatic representation of the ring opening and rearrangement reactions of methylene-cyclopropane. (b) Resonance structures of the biradical TMM intermediate.

Despite the relatively low potential energy barrier, the frequency of DPT events in standard aMD simulations is rather low. On average, we observed one DPT event every 250,000 steps. The bias potential obviously has a much stronger effect on the flexible H-bonds, and hence the inter-monomer collective vibrational degrees of freedom. Increasing the boost energy to larger values does not afford an increase in the frequency of observed DPT events, but generally results in even larger amplitude fluctuations of the two monomeric units, eventually resulting in dissociation of the dimer. Interestingly, we never observe the formation of a zwitterionic state.

II.D.2 Obtaining an accurate free energy profile for DPT using Ad-aMD

The standard aMD simulations described above identified the existence of DPT events on slow time-scales and also provided an approximate estimate for the

associated potential energy barrier for this reaction. In order to obtain an accurate estimate of the free energy barrier for this process, it is necessary to define a suitable reaction coordinate for the DPT event and perform adaptive aMD. Miura *et al.* have previously described two reaction coordinates to accurately describe the symmetrized hydrogenic movement (ρ_1) and inter-monomer vibrations (ρ_2) in FAD:[42]

$$\rho_1 = r_1 - r_2 + r_3 - r_4 \tag{II.8}$$

$$\rho_2 = R_1 + R_2 \tag{II.9}$$

where the relevant inter-atomic distances, r_1, r_2, r_3, r_4, R_1 and R_2 are defined in Fig. II.1(a). The single reaction coordinate, ρ_1, is sufficient to describe the DPT event in this system. Using this reaction coordinate and the approximate representation of the underlying PES obtained from the standard aMD trajectory with respect to ρ_1, a desired modified potential was constructed. Ideally, the desired modified potential, $V^*(\rho_1)$ should fulfill certain criteria: First, $V^*(\rho_1)$ must lie above $V(\rho_1)$ across the entire reaction space of interest. Secondly, the functional form of $V^*(\rho_1)$ should approximately mimic the true underlying PES, at least in terms of the approximate positions of the local energy minima and maxima with respect to the reaction coordinate. In order to achieve efficient configurational space sampling, the activation energy barriers on the modified potential should be approximately 0.5 kcal/mol.

In a series of initial test Ad-aMD simulations, we found that if the activation energy barrier is defined to be much larger than this value, the frequency of observed reaction events decreases substantially. Alternatively, if the activation energy barrier is set to much smaller values, the modified potential becomes too flat, resulting in the

observation of random motion and the system spends a large amount of time exploring high energy, unrealistic regions of the configurational space during the Ad-aMD simulation. Finally, the modified potential should lie below the true potential outside the reaction space of interest. In Fig. II.5, we show how the modified potential, $V^*(\rho_1)$ was constructed for FAD.

The approximate PES obtained from the standard aMD simulation is represented by the dashed black line. The PES is clearly symmetric about ρ_1=0 Å, with energy minima at ρ_1= +/- 1.4 Å. The upper estimate of the activation energy barrier (located at ρ_1=0 Å) lies 7.5 kcal/mol above the thermodynamically stable state (ρ_1= +/- 1.4 Å) and the reaction space of interest lies in the region {-2.8 Å < ρ_1 < +2.8 Å}.

Following the criteria defined above, a series of points describing the desired modified potential were defined manually (Fig. II.5, blue squares): At ρ_1=0 Å, the modified PES was fixed 8.5 kcal/mol above the true potential energy of the thermodynamically stable state, V_0, and therefore at least 1 kcal/mol above the estimated true PES at the transition state. The modified potential energy minima were placed at ρ_1=+/- 1.4 Å, which coincides exactly with the minima on the true underlying PES and were fixed at a value such that the activation energy barrier on the modified potential was 0.5 kcal/mol. Two further points were defined at ρ_1= +/- 2.8 Å that intercept the underlying PES. A few interstitial points were then manually added, maintaining the symmetric properties of the modified PES and these points were fitted

to a simple polynomial function. The resulting desired modified potential (shown in red in Fig. II.5) is given by:

$$V^{*}(\rho_1) = (8.5 - V_0) - 0.517(\rho_1)^2 + 0.154(\rho_1)^4 - 0.01(\rho_1)^6 \quad , \qquad \text{(II.10)}$$

where ρ_1 is in units of angstroms and the polynomial coefficients are in units of kcal/mol.

The adaptive boost potential $E_b(\rho_1)$ necessary to produce the desired modified potential was found to converge exceedingly quickly (within 50,000 steps). Fig. II.6 depicts the enhanced configurational space sampling obtained from a 250,000-step (unweighted) adaptive aMD simulation of FAD projected onto the two reaction coordinates, ρ_1 and ρ_2, compared to a 250,000-step (~25 ps) standard CP-MD simulation. Using the adaptive aMD protocol, a DPT event is observed once every 10,000 steps on average. This represents a 25-fold increase in the frequency of observed DPT events compared to the standard aMD simulations discussed above. After performing the free energy re-weighting protocol to obtain the correct Boltzmann canonical ensemble distribution, the resulting free energy profile as a function of ρ_1 is shown in Fig. II.7. The free energy barrier for DPT is 6.5 kcal/mol, which is in excellent agreement with the value of 6.4 kcal/mol found in a previous study on FAD using the dynamic distance constraint method with the same density functional.[43]

The structural dynamic changes of the system across the reaction coordinate are also in very good agreement with previous observations.[14,43] For example, even before performing the free-energy weighting protocol (equation 6), it is clearly apparent that DPT only occurs when the two monomeric units are contracted (giving a low ρ_2 value of 4.8 Å) and that the system becomes considerably more planar on approaching the transition state. The transition state itself is well-defined at $\{\rho_1,\rho_2\}=\{0.0,4.8\}$ Å and has a very similar structure to the transition state obtained from the MEP.

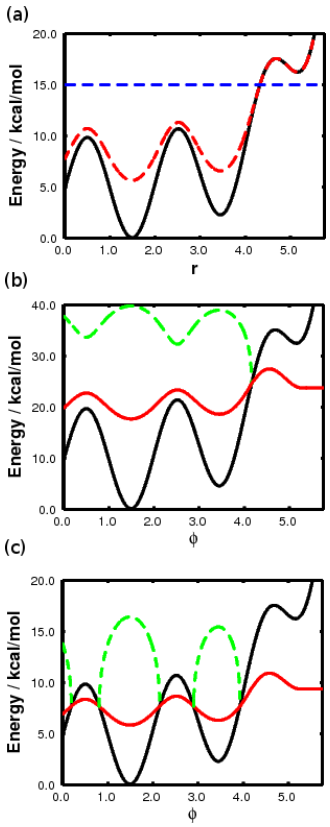

Figure II 3 (a) Diagrammatic representation of the "standard aMD" protocol. The true underlying potential is shown in black. Using a fixed boost energy, *Eb*, at 15.0 kcal/mol (dashed blue line) and an acceleration parameter, α = 25 kcal/mol, the resulting accelerated (modified) potential is shown in red. Both acceleration parameters are kept fixed across the entire configurational space. (b) The adaptive aMD protocol: The modified potential (red line) is defined as a function of a specific reaction coordinate (φ) and positioned above the true underlying potential energy surface (black line) across the entire reaction space of interest $\{0 < \varphi < 4.1\}$. Keeping the α parameter fixed at 25-kcal/mol, the boost energy (dashed green line) is adapted in a history-dependent fashion to achieve the desired acceleration potential across the reaction coordinate. When $\varphi > 4.1$, the acceleration is switched off and the system evolves on the true underlying potential. (c) Diagrammatic representation of the adaptive aMD protocol used to estimate reaction rates. The true underlying potential and the desired modified potential are shown in black and red, respectively. In direct comparison to Fig. II.3(b), the desired modified potential has been shifted such that it lies below the true underlying potential across the entire transition state region. The adaptive boost energy (dashed green line is only applied to the system when the defined modified potential lies above the true potential. In this way, the acceleration is switched off whenever the system approaches the transition state region and an accurate estimate of the reaction rate can be obtained.

The adaptive aMD simulations discussed above provide a very accurate representation of both the enthalpy and free energy statistics associated with the DPT reaction. Using this information, a second adaptive aMD simulation of FAD was performed in order to obtain an estimate for the associated DPT reaction rate, as described in the Methods section. The desired modified PES was lowered such that it intersected the true underlying potential at ρ_1 = +/- 0.2Å and a second set of adaptive aMD simulations were performed. In comparison to the previous aMD simulations, the DPT event in this adaptive aMD protocol was observed considerably less frequently as the system is not being accelerated over the transition state region.

The estimated rate constant for DPT was found to be approximately 0.077 ns^{-1} which is slightly slower than the TST predicted rate constant ($k_BT/h*exp[\beta\Delta G/RT]$ = 0.108 ns^{-1}, using the value for ΔG determined in the previous Ad-aMD simulation with V^* above the true barrier, and assuming a transmission coefficient, κ=1). This suggests that the transmission coefficient (which is notoriously difficult to calculate) for the DPT reaction is 0.7. We would like to stress that the accuracy of the free energy barrier and the associated rate constant is, of course, also dependent on the accuracy of the poly-electronic wave-function, i.e. on the quantum chemical electronic structure method employed. The biased potential methods employed in this study do not improve or correct for any inherent errors in the energy function described by the specific density functional.

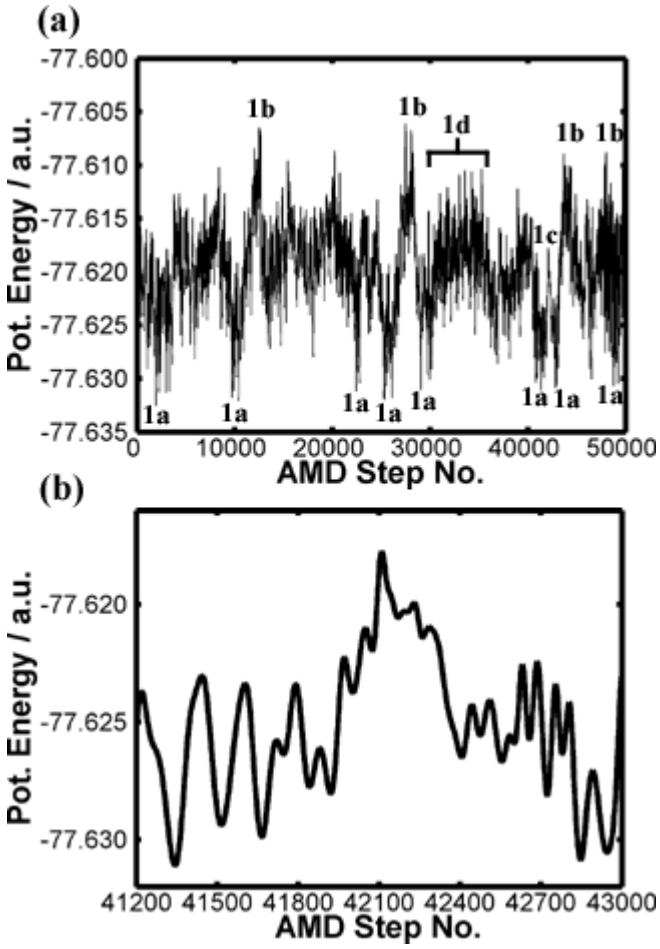

Figure II 4 (a) Variation of the underlying potential energy of FAD during the standard aMD simulation. The labels across the trajectory refer to the different states and motions depicted in Fig. II.1. (b) Variation of the potential energy of FAD across the binary proton transfer event. The estimated energy barrier to DPT is ~0.009–0.012 a.u. (5.7–7.5 kcal/mol).

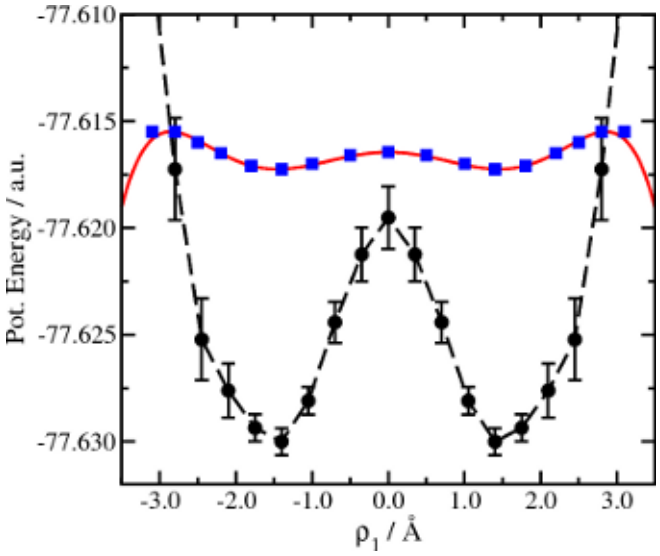

Figure II 5 Construction of the desired modified potential: The estimated variation of the potential energy of FAD as a function of the reaction coordinate ρ1 obtained from the standard aMD (cf. Fig. II.4) is shown by the dashed black line. A series of points (blue squares) describing the desired modified potential were defined manually and fit to a simple polynomial function [red line, Eq. (10)]. The desired modified potential lies above the true potential across the entire reaction space of interest, –2.8 Å < ρ1 < 2.8 Å.

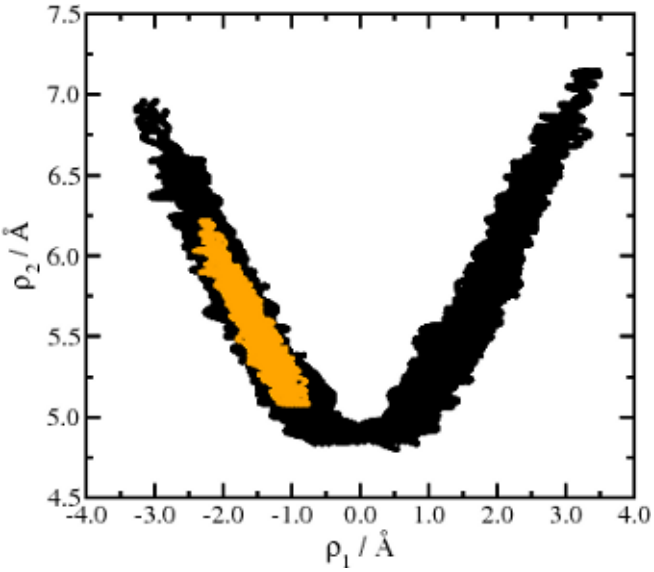

Figure II 6 Comparison of the conformational space sampling observed for formic acid dimer by standard CP-MD (orange) and adaptive aMD (black). In the (unweighted) adaptive aMD trajectory, the system exhibits multiple DPT events across the transition state located at {ρ1, ρ2} = {0.0, 4.8}.

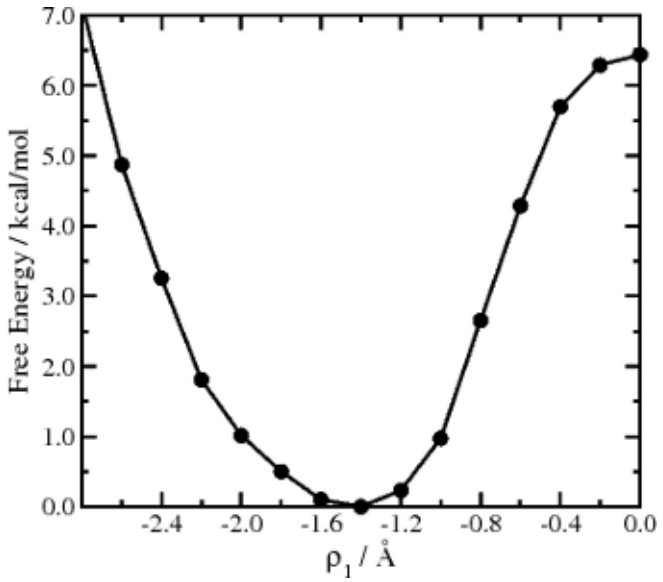

Figure II 7 The free energy of formic acid dimer as a function of the reaction coordinate ρ1 obtained from the adaptive aMD simulation.

II.D.3 Cyclic rearrangement of methylene-cyclopropane (MCP)

Before presenting the results for MCP, we would like to point out that this system was chosen as a purely hypothetical study in order to show that the aMD and adaptive aMD approaches can be successfully applied to study transitions across extremely large energy barriers. To the best knowledge of the authors, the ring opening and subsequent rearrangement reactions of MCP, similar to many other cyclic organic systems, proceed *via* a non-adiabatic mechanism which involves transitions between the ground state and both a singlet excited state and a *triplet* tri-methyl-methane (TMM) bi-radical state.[44] The simulations presented in this work have been performed under strict adiabatic (ground electronic state) conditions.

Analogous to the study of FAD described previously, the first step in the investigation into the ring-opening and subsequent rearrangement reactions of MCP was to perform a standard CP-MD simulation of the system in the gas phase at 300K, which afforded an estimate of the average density functional energy of MCP in the stable thermodynamic state. A series of standard aMD simulations were then performed in which the boost energy, E_b, was fixed 60 kcal/mol above this potential energy minimum and the acceleration parameter, α, was systematically decreased in steps of 10 kcal/mol from 60 kcal/mol to 10 kcal/mol.

This approach is slightly different to that performed in the case of FAD, where α was held fixed and the level of acceleration was defined by increasing the boost energy. As discussed in the Methods section, the level of acceleration is controlled by both the boost energy E_b and α. In the case of MCP, it is clear that the ring-opening mechanism involves a transition over an exceedingly large energy barrier. In such a scenario, it was found that the configurational space sampling could best be controlled by fixing (E_b-V_0) at a large value and using the parameter α to control the level of acceleration. Even at a moderate acceleration level $\{(E_b\text{-}V_0), \alpha\}=\{60.0$ kcal/mol, 60.0 kcal/mol$\}$, we observed cleavage of the C3-C4 bond (see Fig. 2(a): MCP STATE A) resulting in a partial opening of the cyclopropane ring.

The C3-C2-C4 angle, θ, was seen to increase from 66^o (equivalent to the thermodynamically stable geometry) to as much as 90^o and the underlying (true) potential energy of the system increased by up to 20 kcal/mol. Under more aggressive acceleration conditions (lower values of α), the ring-opening mechanism was even

more pronounced, however, at such aggressive acceleration levels, we also observed some intra-molecular distortions, particularly involving C-H bond cleavage events. In order to overcome this problem, a second set of standard aMD simulations were performed where the acceleration (and associated force modification) was only applied to the heavy C-atom frame of the molecule. The (partial) standard aMD simulations identified two chemical reactions:

The first involved ring-opening and subsequent ring-closure *via* a trigonal-planar TMM bi-radical intermediate and the underlying potential energy barrier for this chemical reaction was found to be approximately 40 kcal/mol. The second chemical reaction involved cleavage of the C2-C3 bond, leading to the formation of an exceedingly high energy vinyl-radical system.

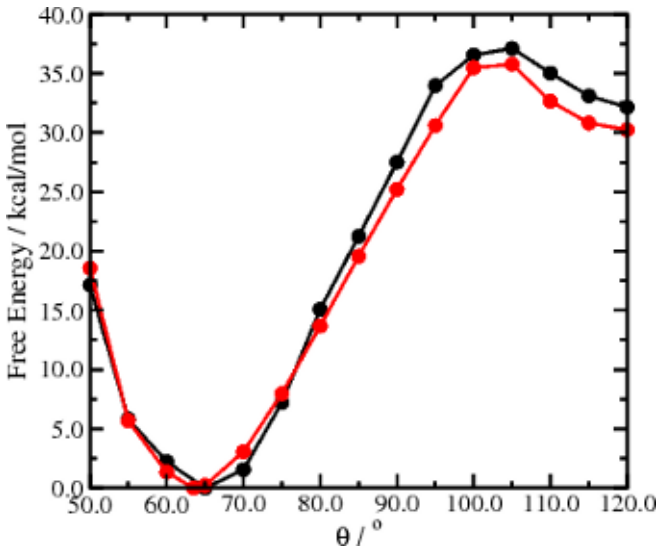

Figure II 8 Comparison of the conformational space sampling observed in MCP by standard CP-MD (orange) and adaptive aMD (black). In the (unweighted) adaptive aMD trajectory, the system rapidly exchanges between all three iso- merit states of MCP located at ~{θ 1 , θ 2 } = {65, 150}˚, {150,150}˚, and {150, 65}˚. The TMM metastable intermediate is located at {120, 120}˚.

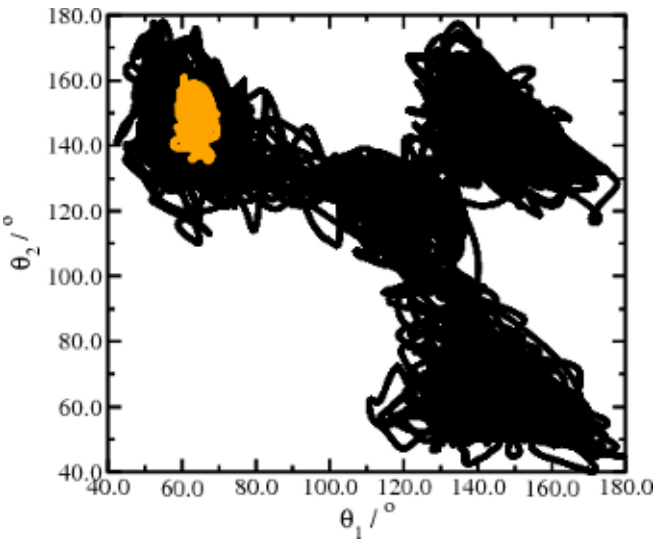

Figure II 9 The free-energy profile for ring opening of MCP as a function of the C–C–C angle, θ obtained from the adaptive aMD method (black line), and the standard constrained CP-MD approach (red line). The closed ring system has a free-energy minimum at 66° and the transition state is located approximately at θ = 105°.

In order to study the lower energy chemical reaction, the ring opening and subsequent rearrangement of the MCP system via the bi-radical TMM intermediate in more detail, a series of adaptive aMD simulations were performed. Making use of the observed planar symmetry of the C-atom frame in the standard aMD simulations, the reaction coordinate for the adaptive aMD simulations was defined as the smallest angle, θ, subtended between two peripheral atoms over the central C-atom (C2) forming the heavy atom frame: In this way, acceleration was constantly applied to the system as it evolved between the different isomeric states, each of which, while being chemically identical, differs in respect to the specific C-atoms that form the cyclopropane ring.

The desired modified potential as a function of θ, was therefore defined within the boundary conditions ($0^0 < \theta < 120^0$). Similar to the case of FAD, an approximate representation of the underlying PES, $V(\theta)$, was obtained by averaging the potential energy statistics from the standard aMD trajectory with respect to θ. The modified potential, $V^*(\theta)$, was then constructed by manually defining an initial set of points. The underlying PES obtained from the standard CP-MD and aMD simulations had a single energy minimum at $\theta = 66^o$. The upper estimate of the activation energy barrier for ring opening and subsequent ring-closure, located broadly at $\theta=120^o$, was found to be approximately 40 kcal/mol above the thermodynamically stable state. For small values of θ, particularly, $\theta < 40^o$, the system was energetically highly unstable.

Based on this potential energy function, the transition state for the modified potential was positioned at $\theta=120^o$ and fixed 41 kcal/mol above the true potential energy of the thermodynamically stable state, V_0 (located at $\theta=66^o$), more than 1 kcal/mol above the true PES at the assumed transition state ($\theta=120^o$). The modified potential energy minimum was placed at $\theta=66^o$, which coincides with the minimum on the true underlying PES and was fixed at a value such that the activation energy barrier on the modified potential was 0.5 kcal/mol. A third point for the modified potential was defined at $\theta=40^o$, which intercepted the true underlying PES. Similar to the case of FAD, a few interstitial points were manually added and these points were fitted to a simple polynomial function. The resulting predefined fixed modified potential is given by:

$$V^{*}(\theta) = (41.522 - V_0) + 1.21\theta - 4.27\theta^2 + 3.30\theta^3 - 0.7352\theta^4 \quad , \qquad \text{(II.11)}$$

where the angle θ is in radians and the polynomial coefficients are given in kcal/mol. Two C-C-C angles $\{\theta_1, \theta_2\}$ are sufficient to accurately describe the configurational space of the MCP system. Fig. II.8 depicts the enhanced configurational space sampling obtained from a 500,000-step (unweighted) adaptive aMD simulation of MCP projected onto the two internal degrees of freedom, $\{\theta_1, \theta_2\}$, compared to a 500,000-step (~ 50 ps) standard CP-MD simulation.

While the standard CP-MD simulation samples a very limited region of configurational space centered approximately around $\{\theta_1, \theta_2\} = \{65^o, 150^o\}$, the adaptive aMD trajectory readily samples all three isomeric states of MCP and the inter-conversion between these states proceeds *via* a meta-stable trigonal planar bi-radical TMM state located at $\{\theta_1, \theta_2\} = \{120^o, 120^o\}$. The associated free energy profile for ring-opening as a function of θ is depicted in Fig. 9 and shows a transition state located at $\theta = 105^o$. The free energy for ring-opening is found to be approximately 36 kcal/mol and the meta-stable TMM intermediate state lies approximately 32.5 kcal/mol above the thermodynamically stable energy minimum located at $\{\theta = 66^o\}$.

In order to assess the accuracy of this result, we calculated the free energy profile for the ring-opening of MCP using an established method:[15,45,46] A series of CP-MD simulations were performed in which the C3-C2-C4 angle that defines the propane ring was constrained and the associated free energy profile was obtained by thermodynamic integration of the average constraint force over this reaction

coordinate (θ), including a correction for the metric effects on the associated Lagrangian. The free energy profile obtained from this constrained CP-MD approach is shown with the present adaptive aMD result in Fig. II.9. A remarkable agreement between the two different free-energy methods is observed, confirming the fact that *ab initio* adaptive aMD is a robust protocol for determining accurate free energy statistics, even in the case when the associated free energy barriers are extremely large. Under the strict constraints of this (hypothetical) adiabatic study, the associated reaction path for ring-opening and subsequent ring-closure of MCP possesses two transition states.

This free energy profile can be well understood when one considers that as the system approaches trigonal-planar symmetry, the resonance hybridization effect (Figure II.2(b)) results in energetic stabilization of the bi-radical TMM intermediate. Despite the fact that the functional form of the pre-defined modified potential, $V^{*}(\theta)$, employed in the adaptive aMD simulations does not resemble the true free energy profile for this reaction, it is important to recognize that the free energy weighting procedure still provides extremely accurate free energy statistics. This observation underlines the fact that it is not necessary to define a modified potential that contains all the characteristics of the true underlying energy landscape: Accurate free energy statistics are consistently produced *via* the subsequent free energy weighting protocol. As was previously mentioned in the Methods section, the pre-defined modified potential only has to mimic the true underlying PES accurately enough to ensure that it lies above the true potential across the entire reaction space of interest, thereby ensuring that the system is constantly under the influence of the acceleration potential, resulting in more efficient configurational space sampling.

The adaptive aMD approach does not only provide an accurate estimation of the free energy profile associated with the ring-opening and subsequent re-arrangement of MCP, but also affords a very accurate description of the associated reaction mechanism. The transition state identified in the adaptive aMD simulations is almost identical to that observed in the constrained CP-MD protocol: As the cylcopropane ring opens, the two C-C bonds that originally formed the cyclopropane ring increase in length from 1.5 Å to 1.8 Å, as does the C=C ethylene bond. Both ring-opening and ring-closure events on the (hypothetical) adiabatic PES involve a dis-rotatory motion of the two CH_2 groups. These structural geometric changes are observed in both the adaptive aMD simulations and the constrained CP-MD analysis.

II.E Discussion

The results presented above demonstrate how the application of the accelerated molecular dynamics approach in the framework of *ab initio* molecular dynamics can be used to study chemical reactions and to obtain an accurate representation of both the free energy surface and the associated reaction mechanism. It is important to note that the two bias potential methods (standard aMD and adaptive aMD) serve different purposes: The primary purpose of the standard aMD simulations is to explore the configuration and reactive free energy surface of the system, to identify stable and meta-stable states and to predict chemical reactions of interest. For example, the standard aMD simulations performed on FAD readily identified the dissociated state (Fig. II.1(b)) and the twisted state (Fig. II.1(d)) and predicted the DPT chemical reaction via the transition state (Fig II.1.(c)). In the case of MCP, the standard aMD

simulations identified two chemical reactions: Ring-opening and subsequent ring-closure via a TMM intermediate and cleavage of the C2-C3 bond leading to the formation of a high energy vinyl-radical state. For both systems it was possible to obtain an approximate representation of the underlying PES. The primary purpose of the adaptive aMD method is to obtain accurate free energy statistics for a specific chemical reaction of interest and, when appropriate, an estimate of the associated reaction rate.

In general, one faces the following dilemma concerning enhanced configurational space sampling and obtaining accurate free energy statistics: Ideally, we would like to explore large areas of the PES of a system as efficiently as possible and at the same time extract accurate free energy statistics. However, these two aims are often mutually exclusive, as the more widely we explore the PES, the less accurate the free energy statistics will be in the narrow regions of interest. The different roles of the two bias potential methods presented in this paper reflect this mutual exclusivity. While the standard aMD approach is much more efficient at exploring the total PES of the system, identifying a large number of high energy 'states' or activated processes, the associated free energy statistics are not very accurate.

Having identified a particular chemical reaction of interest, the application of the bias potential can then be adapted in order to focus more on the specific associated reacting degrees of freedom, enhancing the specific sampling of that particular process and thereby affording more accurate free energy statistics. This forms the basis of adaptive aMD. We would like to point out that a similar argument can be made for other contemporary enhanced configurational space sampling methods. For example,

in the constrained CP-MD approach, having defined a reaction coordinate, it is possible to obtain an accurate free energy profile with respect to that specific reaction coordinate, but only a very small part of the total PES of the system is sampled.

As discussed above, the primary aim of the standard aMD approach is to efficiently explore the PES of a system and to identify activated processes and chemical reactions. The efficiency of the enhanced configurational space sampling in aMD simulations depends strongly on the choice of the acceleration parameters, E_b (or more specifically [E_b-V_0]) and α. In practice, the 'optimal' choice of these aMD parameters is extremely system-specific and depends on many factors including the size of the system and the nature of the underlying potential energy landscape. Nevertheless, based on the work presented here, it is possible to construct some general guidelines.

The boost energy, E_b, represents the potential energy ceiling: Any activated process that involves a transition over an energy barrier greater than [E_b-V_0] will not generally be observed in the standard aMD simulations. While initial estimates of the energy barriers (and hence the magnitude of [E_b-V_0]) can be made using our understanding of different classes of chemical reactions, in the case where one has no prior knowledge of the underlying PES for a given system, a suitable choice for [E_b-V_0] in general is 10-15 kcal/mol, which, according to TST, will allow the prediction of chemical reactions occurring on the micro- to millisecond time-scale. Nevertheless, when approaching extremely rare events or reactions mediated by non-adiabatic mechanisms (such as that observed in MCP) that involve transitions over exceedingly

large energy barriers (usually in excess of 30 kcal/mol), a significantly larger initial estimate of [E_b-V_0] is required.

Based on the work presented here, we have identified that the 'optimal' value for the acceleration parameter, α, generally lies between *0.7**[E_b-V_0] and *0.1**[E_b-V_0]. Obviously, when approaching a system for which one has no prior understanding of the underlying PES, an initial series of test aMD simulations across a variety of acceleration levels is necessary in order to identify the 'optimal' acceleration parameters. For example, in the case of FAD, we performed six initial short aMD simulations using a fixed value of α (0.016 a.u.) and varying the value of [E_b-V_0] between 0.016 a.u. and 0.026 a.u. in steps of 0.002 a.u. We would like to note that even when performing very short initial test aMD simulations (such as 100,000 steps which is the equivalent of ~10 ps), the appropriate acceleration level can be readily identified: If the acceleration level is too low, ostensibly no enhanced configurational space sampling is observed. In contrast, applying too high an acceleration level, results in a molecular explosion or dissociation of the system.

The initial test aMD simulations also provide valuable qualitative information about the associated activation energies for different processes occurring in the system. It should be recognized that the efficiency of the standard *ab initio* aMD approach to predict chemical reactions is dependent on the size and inherent flexibility of the system under study. aMD amplifies the motion of *all* the vibrational degrees of freedom in the system, including low energy, large amplitude molecular motions, and not just those degrees of freedom associated with the chemical reaction. In some cases, the acceleration of these low energy vibrational fluctuations (such as the inter-

monomer breathing mode of FAD) brings the system towards the chemical transition state, thereby facilitating the observation of the chemical reaction of interest. However, whether a particular chemical reaction is sufficiently accelerated by aMD also depends on the number of alternative dynamical pathways. Therefore, the efficiency of the aMD method to predict chemical reactions may be compromised to some extent when applied to larger, more flexible systems than those reported in this paper.

In addition to their ability to predict chemical reactions of interest, the standard aMD simulations also provide an approximate representation of the underlying variation in the PES for these processes (as demonstrated specifically in the Results section for FAD). As such, these simulations allow us to formulate a suitable reaction coordinate for the specific reaction of interest, which is then used in the adaptive aMD simulations to obtain accurate free energy statistics. The initial definition of the magnitude and functional form of the boost energy (for a fixed value of α) employed in the adaptive aMD simulations can be obtained using the approximate representation of the underlying PES. It is therefore clear that defining the initial acceleration parameters in the adaptive aMD approach is much simpler than in the case of standard aMD, where no information about the PES is available.

The adaptive aMD method presented here was developed to show that it is possible to obtain accurate free energy statistics and reaction rates for chemical reactions in the general framework of *ab initio* aMD. However, it is clear that having predicted a chemical reaction and formulated a suitable reaction coordinate from the standard aMD simulations, alternative methods to obtain the associated free-energy

profile, such as meta-dynamics or constrained MD can be employed and in some cases may be more favorable. An objective comparison of the efficiency and accuracy of enhanced configurational space sampling and free energy methods is difficult to perform as certain methods are more suited to some systems than to others. In terms of exploring the PES and predicting activated processes and chemical reactions, the standard aMD approach has significant advantages over other methods (such as meta-dynamics and constrained CP-MD) as it does not require any *a priori* understanding of the underlying PES, nor does it require the specific prior definition of a reaction coordinate or a set of collective variables. As such, we consider that the standard aMD approach could be particularly useful when studying larger, more complex systems where the specific construction of an appropriate reaction coordinate or a suitable definition of the appropriate collective variables might not be readily apparent.

In terms of obtaining accurate free energy statistics, the relative efficiency of the different methods is not so clear: A direct comparative analysis of the adaptive aMD and constrained MD approaches applied to the systems presented in this work suggests that the two methods are equally efficient. We would like to stress that this observation is based on just two systems and as such, it is inappropriate to draw general conclusions about the relative efficiency of methods. In particular, we note once again that the test systems studied here are small and possess a limited number of internal degrees of freedom. The relative efficiency of the two methods when applied to larger and inherently more flexible systems remains unclear. Nevertheless, the Ad-aMD approach does not require a correction for metric effects on the associated Lagrangian, which is often necessary and tedious in the constrained CP-MD method.

Furthermore, the Ad-aMD approach affords an estimate of the reaction rate constant, *including a meaningful estimate of the associated transmission coefficient*, which is notoriously difficult to obtain using any standard method for enhanced configuration space sampling. In general therefore, we consider that the *ab initio* aMD methods presented here can be readily used alone, or to complement existing methods to study chemical reactions and reactive free energy surfaces.

II.F Conclusions

We have demonstrated that the accelerated molecular dynamics approach, which has previously been successfully applied to the study of conformational transitions in poly-peptides and proteins, can also be implemented in the framework of *ab initio* molecular dynamics to study chemical reactions. A two-step procedure is proposed: in step 1, the standard aMD approach is applied, which allows one to efficiently explore the reactive subspace of molecular systems, identifying a variety of of high energy states and predicting potential chemical reactions. In step2, a new variant, adaptive aMD, is introduced, which allows one to focus on a particular chemical reaction or rare event of interest, providing a detailed description of the reaction mechanism, an accurate representation of the associated free energy statistics and an approximate estimate of the associated reaction kinetics.

In comparison to other methods*, ab initio* aMD allows for the exploration of the reaction space of a molecular system without any *a priori* understanding of the underlying free energy surface, nor does it involve the specific definition of a pre-defined reaction coordinate. The biased potential *ab initio* MD simulations presented

here have been performed in the framework of Car-Parrinello MD, but the application of this method in the framework of Born-Oppenheimer MD is equally viable. Our present study of chemical reactions in simple systems, the DPT event in FAD and the ring-opening and subsequent rearrangement of MCP have demonstrated that both *ab initio* aMD and adaptive *ab initio* aMD methods are readily applicable to the study of chemical reactions involving transitions over intermediate to large free energy barriers up to 25-kcal/mol. The study of chemical reactions involving transitions over substantially larger energy barriers is possible (as demonstrated by the example of MCP in this work) but is somewhat more challenging. However, we would like to point out that such high energy processes are usually mediated by more complex non-adiabatic processes that require more elaborate methods.[47,48]

Most chemical reactions of interest, particularly bio-chemical reactions, generally occur on micro- to millisecond time-scales and have associated free energy barriers ranging up to a magnitude of 15-kcal/mol. We believe that standard *ab initio* aMD is a powerful and general tool for searching for any reactive pathways in an unbiased fashion. It can be readily applied to highly complex systems, retaining its predictive power. In fact, we are presently implementing these *ab initio* aMD methods in a QM/MM manifold in order to study enzymatic reactions in a realistic bio-molecular environment. Accurate free energy landscapes are readily obtained using the new, adaptive aMD method, i.e. by selectively accelerating the reactive modes identified in the standard aMD simulation. Generalization of Ad-aMD to many-dimensional reaction coordinates in more complex systems is straightforward once they have been identified. In conclusion, we consider that the *ab initio* aMD methods

presented here can be readily used alone, or to complement existing methods to study chemical reactions and reactive free energy surfaces.

II.G Acknowledgements

NLD would like to thank DFG for supporting this project within FOR 618, which included funding for a research visit by PRLM to Bochum. PRLM gratefully acknowledges HHMI for financial support. Work at UCSD is supported in part by the NSF, NIH, CTBP, NBCR and the NSF Supercomputer Centers.

This chapter contains material from "Accelerating chemical reactions: Exploring reactive free-energy surfaces using accelerated *ab initio* molecular dynamics" published in 2011 in *Journal of Chemical Physics* (Volume 134, pages174107-1 to 174107-12), authored by Levi C.T. Pierce, Phineus R. L. Markwick, J. Andrew McCammon, and Nikos L. Doltsinis. All material has been reproduced with the consent of all other authors.

II.H References

(1) H.B. Schlegel, J. Comput. Chem. **24**, 1514 (2003).

(2) C.J. Cerjan and W.H. Miller, J. Chem. Phys. **75**, 2800 (1981).

(3) B. Paizs, G. Fogarasi and P. Pulay, J. Chem. Phys. **109**, 6571 (1998).

(4) J. M. Bofill, J. Comput. Chem. **15**, 1 (1994).

(5) K. Fukui, Acc. Chem. Res. **14**, 363 (1981).

(6) S. Fischer and M. Karplus, Chem. Phys. Lett. **194**, 252 (1992).

(7) S.R. Billeter, A.J. Turner and W. Thiel, Phys. Chem. Chem. Phys. **2**, 2177 (2000).

(8) G. Henkelman and H. Jonsson, J. Chem. Phys. **111**, 7010 (1999) and **113**, 9978, (2000).

(9) D.G. Truhlar and B. C. Garrett, Annu. Rev. Phys. Chem. **35**, 159 (1984).

(10) R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

(11) R. Elber, Curr. Opin. Struct. Biol. **15**, 151 (2005).

(12) B. J. Berne and J.E. Straub, Curr. Opin. Struct. Biol. **7**, 181 (1997).

(13) J. Schlitter, M. Engels and P. Kruger, J. Mol. Graph, **12**, 84 (1994).

(14) P.R.L. Markwick, N.L. Doltsinis and D. Marx, J. Chem. Phys. **122**, 054112 (2005).

(15) N. L. Doltsinis in *Computational Nanoscience: Do it yourself!*, edited by J. Grotendorst *et. al*., (FZ Jülich, 2006), www.fz-juelich.de/nic-series/volume31/doltsinis2.pdf

(16) C. Dellago, P.G. Bolhuis, F.S. Csajka and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).

(17) P.G. Bolhius, D. Chandler, C. Dellago and P.L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).

(18) A. Amadei, A.B. Linssen and H.J. Berendsen, Proteins **17**, 412 (1993).

(19) A. Mitsutake, Y. Sugita and Y.J. Okamoto, J. Chem. Phys. **118**, 6664 (2003).

(20) A.F. Voter, Phys. Rev. B **57**, R13985 (1998).

(21) A. Laio, M. Parrinello, *Computing Free Energies and Accelerating Rare Events with Metadynamics.* in Computer Simulations in Condensed Matter: From Materials to Chemical Biology, ed. M. Ferrario, G. Ciccotti, K. Binder (Springer Verlag, Berlin Heidelberg, 2006**),** Vol. 1, pp 315-347.

(22) A.F. Voter, J. Chem. Phys. **106**, 4665 (1997).

(23) D. Hamelberg, J. Mongan and J.A. McCammon, J. Chem. Phys. **120**, 11919 (2004).

(24) D. Hamelberg, C.A. De Oliveira and J.A. McCammon, J. Chem. Phys. **127**, 155102 (2007).

(25) L. Yang, M.P. Grubb and Y.Q. Gao, J. Chem. Phys. **126**, 125102 (2007).

(26) B.J. Grant, A.A. Gorfe and J.A. McCammon, PLOS Comput. Biol. e1000325 (2009).

(27) P.R.L. Markwick, G. Bouvignies, L. Salmon, J.A. McCammon, M. Nilges and M. Blackledge, J. Am. Chem. Soc. **131**, 16968 (2009).

(28) C.F. Cervantes, P.R.L. Markwick, S.C. Sue, J.A. McCammon, H.J. Dyson and E.A. Komives, Biochemistry, **48**, 8023 (2009).

(29) P.R.L. Markwick, S.A. Showalter, G. Bouvignies, R. Bruschweiler and M. Blackledge, J. Biomol. NMR, **45**, 17 (2009).

(30) P.R.L. Markwick, C.F. Cervantes, B.L. Abel, E.A. Komives, M. Blackledge and J.A. McCammon, J. Am. Chem. Soc. **132**, 1220 (2010).

(31) M.D. Mukrasch, P.R.L. Markwick, J. Biernat, M. Bergen, P. Bernado, C. Griesinger, E. Mandelkow, M. Zweckstetter and M. Blackledge, J. Am. Chem. Soc. **129**, 5235 (2007).

(32) J.-C. Wang, S. Pal and K.A. Fichthorn, Phys. Rev. B **63**, 085403 (2001).

(33) K. Laidler and C. King, J. Phys. Chem. 87, 2657 (1983).

(34) D.G. Truhlar, B.C. Garrett and S.J. Kippenstein, J. Phys. Chem. **100**, 12771 (1996).

(35) J. Hutter et al., MPI für Festkörperforschung, Stuttgart, and IBM Zurich Research Laboratory; www.cpmd.org

(36) A.D. Becke, Phys. Rev. A **38**, 3098 (1988).

(37) C.T. Lee, W.T. Yang and R.G. Parr, Phys. Rev. B **37**, 785 (1988).

(38) J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

(39) N. Troullier and J.L. Martins, Phys. Rev. B **43**, 1993 (1991).

(40) D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990)

(41) S. J. Nose, J. Chem. Phys. **81**, 511 (1984).

(42) S. Miura, M.E. Tuckerman and M. L. Klein, J. Chem. Phys. **109**, 5290 (1998).

(43) C. Burisch, P.R.L. Markwick, N.L. Doltsinis and J. Schlitter, J. Chem. Theor. Comput. **4**, 164 (2008).

(44) B. Ma and H.F Schaefer III, Chem. Phys. **1**, 31 (1996).

(45) G. Ciccotti, M. Ferrario, J.T. Hynes and R. Kapral, Chem. Phys. **129**, 241 (1989).

(46) A. Curioni, M. Sprik, W. Andreoni, H. Schiffer, J. Hutter and M. Parrinello, J. Am. Chem. Soc. **119**, 7218 (1997).

(47) N.L. Doltsinis and D. Marx, Phys. Rev. Lett. **88**, 166402 (2002).

(48) E. Tapavicza, I. Tavernelli and U. Rothlisberger, Phys. Rev. Letts. **98**, 023001 (2007).

III Adaptive accelerated Molecular Dynamics

III.A Abstract

An extended accelerated molecular dynamics (aMD) methodology called adaptive aMD is presented. Adaptive aMD (Ad-aMD) is an efficient and robust conformational space-sampling algorithm that is particularly well suited to proteins with highly structured potential energy surfaces exhibiting complex, large-scale collective conformational transitions. Ad-aMD simulations of substrate- free P450cam reveal that this system exists in equilibrium between fully and partially open conformational states. The mechanism for substrate binding depends on the size of the ligand. Larger ligands enter the P450cam binding pocket, and the resulting substrate-bound system is trapped in an open conformation via a population shift mechanism. Small ligands, which fully enter the binding pocket, cause an induced-fit mechanism, resulting in the formation of an energetically stable closed conformational state. These results are corroborated by recent experimental studies and potentially provide detailed insight into the functional dynamics and conformational behavior of the entire cytochrome-P450 super- family.

III.B Introduction

The function of bio-macromolecules is determined by both their 3D structure and dynamics.[1,2] Proteins are inherently flexible systems displaying a broad range of

dynamics over a hierarchy of time scales. Many biologically important processes, such as enzyme catalysis,[3] ligand binding, and signal transduction,[4] occur on the microsecond- millisecond time scale.[5] The study of such slow time scale dynamics remains a challenge to experimentalists and theoreticians alike. Despite the sustained and rapid increase in available computational power and the development of efficient simulation algorithms, MD simulations of large proteins and biomachines are generally limited to time scales of tens to hundreds of nanoseconds. Considerable progress has been made in the development of more sophisticated methods to sample the conformational space of proteins more efficiently,[6,7] allowing the study of functionally important slow molecular motions.

In general, these methods can be divided into two groups. The first involves the identification of transition pathways between known initial and final states. Such methods include transition path sampling[8] and targeted molecular dynamics.[9] The second group contains those methods that efficiently sample low-energy molecular conformations, allowing the rapid identification of thermodynamically dominant regions on the potential energy surface (PES). These methods include replica exchange MD,[10] meta-dynamics,[11] and accelerated molecular dynamics (aMD).[12] The principle behind aMD is to add a continuous non-negative bias potential to the actual PES, which raises the low-energy regions on the potential energy landscape, decreasing the magnitude of the energy barriers and accelerating the exchange between low-energy conformational states while still maintaining the essential details of the underlying potential energy landscape. One of the favorable characteristics of

this method is that it yields a canonical average of an observable, so that thermodynamic and other equilibrium properties can be determined.

aMD has already been successfully employed to study slow time scale dynamics in small proteins, such as ubiquitin[13] and IκBR.[14] The enhanced conformational space sampling by aMD in these studies was shown to significantly improve the theoretical prediction of experimental NMR observables, such as residual dipolar couplings,[13,14] scalar J couplings,[13,15] and chemical shifts[16] that are sensitive to dynamic averaging on the micro- to millisecond time scale. As a robust free- energy sampling method, aMD has also been successfully combined with molecular modeling approaches to study the conformational behavior of natively unstructured proteins.[17]

Despite these initial successes, certain aspects of the aMD methodology, including both efficiency and versatility, need to be improved in order to study more complex dynamic behavior in large biomolecular systems. In light of this, we have developed an extended aMD methodology called adaptive aMD (Ad-aMD).

III.C Methods

The principal idea behind Ad-aMD is to use the information that is obtained about the potential energy landscape of the system during an aMD simulation to optimize the acceleration parameters in order to sample the conformational space more efficiently. By learning from the simulation itself, the acceleration parameters are adapted to create an optimal modified "history-dependent" PES. Ad-aMD provides efficient and enhanced conformational sampling for systems exhibiting a highly structured potential energy landscape. History-dependent adaption of the acceleration

level during the course of the simulation allows the system to rapidly traverse exceedingly large energy barriers, identifying complex, collective conformational transitions while still maintaining the integrity of the underlying PES. In this paper, we introduce the Ad-aMD method and apply it to the study of the molecular plasticity and functional dynamics of P450cam from Pseudomonas putida.

P450cam (CYP101) is a member of the cytochrome-P450 superfamily, a large and diverse group of heme mono-oxygen ases that activate O2 for oxygen insertion into a wide variety of substrates. Previous X-ray crystallographic studies have shown that P450cam can be trapped in a range of conforma- tional states. While camphor-bound P450cam adopts a "closed" conformation,[18,19] a variety of "open" conformations have been observed in response to binding large tethered adamantane probes.[20-22] Substrate-free P450cam has long been regarded to exist in the closed state following the report of a substrate-free structure obtained by soaking dithiothreitol (DTT) out of the active site of crystals, affording a conformation very similar to the camphor-bound form.[23] Small-angle X-ray scattering[24] and hydrostatic pressure[25,26] experiments have also supported the view that substrate-free P450cam exists in a closed conformation. However, these studies have recently been brought into question following the observation of an open conformation of P450cam in the absence of substrate.[27]

The details of accelerated molecular dynamics have been discussed previously in the literature,[12] and we merely provide a brief summary here. In the standard aMD formalism, a continuous non-negative bias potential, $\Delta V(r)$, is defined such that when the true underlying potential of the system, $V(r)$, lies below a certain threshold boost energy, E_b, the simulation is performed on a modified potential, $V^*(r) = V(r) + \Delta V(r)$,

but when V(r) > E_b, the simulation is performed on the true potential [V*(r) = V(r)]. The modified potential is related to the true potential, bias potential and boost energy by[12]

$$V^*(r) = V(r) \quad , \qquad V(r) \geq E_b \tag{III.1}$$

$$V^*(r) = V(r) + \Delta V(r) \quad , \qquad V(r) \; < E_b$$

and the bias potential, $\Delta V(\mathbf{r})$ is defined as:

$$\Delta V(r) = \frac{\left(E_b - V(r)\right)^2}{\alpha + E_b - V(r)} \tag{III.2}$$

The application of the bias potential results in raising and flattening of the potential energy landscape, thereby enhancing the escape rate between low-energy conformational states, and the extent of acceleration is determined by the choice of the acceleration parameters E_b and α. In the standard aMD protocol, the parameters E_b and α are kept constant.

A more efficient extension to the accelerated molecular dynamics approach is Ad-aMD. In this approach, one of the acceleration parameters (α) is held fixed, and the boost potential, E_b, is adapted in a history-dependent fashion during the aMD simulation using the population statistics on the modified potential. In order to achieve this, it is necessary to project the trajectory onto a suitable predefined conformational subspace. In the present application, the P450cam system was projected onto the conformational subspace defined by the two lowest principal components obtained from a PCA analysis performed on a collection of available X-ray crystal structures of

P450cam cocrystallized with different substrates. The adaptive boost potential, $E_b(p,q)$, is then given by

$$E_b(p,q) = E_b(0) + \sum_i a * \exp - \left(\frac{(p - \langle p_{i-1} \rangle)^2}{2c^2} + \frac{(q - \langle q_{i-1} \rangle)^2}{2c^2} \right), \qquad \text{(III.3)}$$

where p and q are the projected principal components and the index, i, refers to the number of adaptive Gaussians added to the boost potential during the simulation. The simulation is initiated as a standard aMD simulation with a boost energy of $E_b(p,q) = E_b(0)$ (the base-boost potential). After 500,000 MD steps (the equivalent of 500 ps), the trajectory is projected into the principal component space (p,q), and the average PC- projection coordinates ($\langle p_{i-1} \rangle, \langle q_{i-1} \rangle$) are calculated.

An adaptive 2D-Gaussian boost potential centered at these coordinates is then added to the boost potential, $E_b(p,q)$, and the simulation is performed for another 500,000 MD steps. The resulting 500,000 structures are projected into the principal component space, and the average PC-projection coordinates are used to define the center of the next adaptive 2D-Gaussian boost potential. In this way, the boost potential is adapted every 500,000 MD steps in a history-dependent fashion. The parameters a and c define the magnitude and width of the adaptive Gaussian boost potentials, respectively. In the present work, "dual boost" Ad-aMD simulations were performed, in which two adaptive acceleration potentials were applied to the P450cam system. The first acceleration potential was applied to the torsional potential only, and a second, weaker acceleration was applied across the entire potential. A schematic

representation of the Ad-aMD method is shown in Fig. III.1.

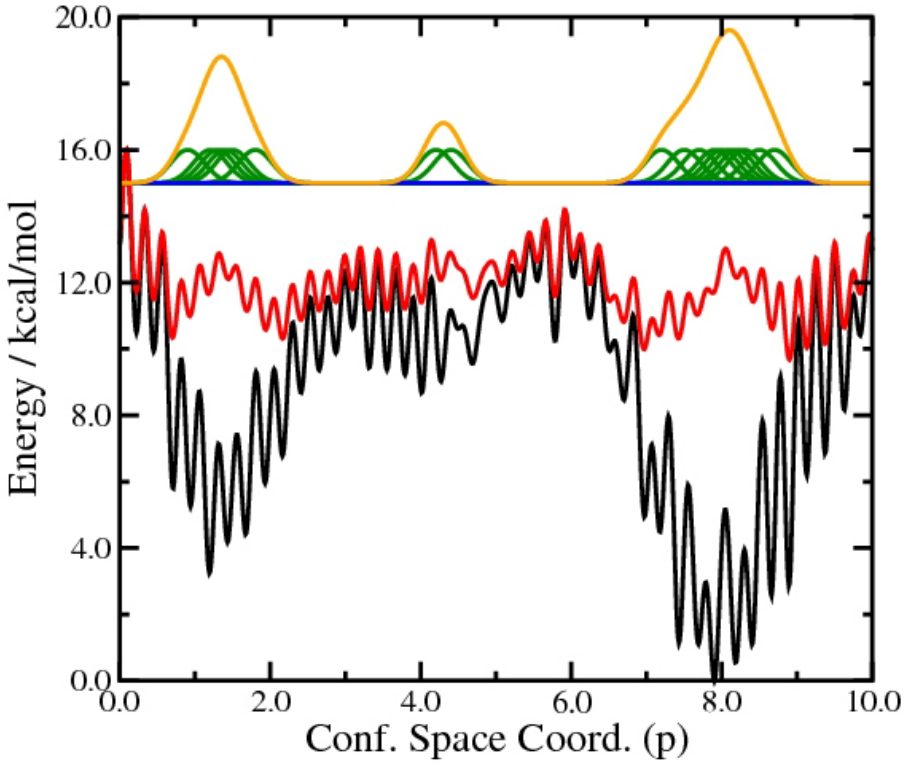

Figure III 1 Schematic representation of the adaptive aMD method. The rugged and highly structured true PES of the protein as a function of the configurational space coordinate, p, is shown in black. The base- boost potential, Eb(0) is fixed at 15 kcal/mol above the potential energy minimum (blue line). The history-dependent adaptive Gaussian boost potentials are shown in green. The final adaptive boost potential, $E_b(p)$, is shown in orange, and the resulting modified potential energy surface on which the system evolves during the Ad-AMD simulation is represented by the red line using the fixed acceleration parameter α=12 kcal/mol.

All MD, aMD, and Ad-aMD simulations presented in this work were performed on a 404 residue construct of substrate-free P450cam using an in-house modified version of the AMBER10 sander simulation suite.[28] The ff99SB force field[29] was employed for the solute residues (Leu11-Val414), with the exception of a nonstandard Cys357-heme residue for which a force field was generated in-house, and the TIP4P water force field was used for the solvent molecules. A comparative

analysis of previous dual-boost aMD simulation studies[13,14,30] identified that for torsional acceleration, the optimal value of [E_b(dih) - V0(dih)] (where V0(dih) is the average torsional potential energy obtained from a standard MD simulation) is approximately equal to 3-5 kcal/mol times the number of solute residues, and the associated acceleration parameter, R(dih), is equal to 1/5 of this value. For the background total acceleration, [Eb(tot)-V0(tot)] and R(tot) should both be equal to 0.16 kcal/mol times the number of atoms in the simulation cell (NASC).31 In light of this, the acceleration parameters employed for all standard aMD simulations in this work are {[Eb(dih)-V0(dih)], α(dih); [Eb(tot)-V0(tot)], α(tot)}={1400, 280; 0.16NASC, 0.16NASC} kcal/mol. The acceleration param eters used for all of the Ad-aMD simulations in this work are {[Eb(dih)(0) - V0(dih)], α(dih); [Eb(tot)(0) - V0(tot)], α(tot)}= {700, 280; 0.08NASC, 0.16NASC} kcal/mol. Notice that for both torsional and total acceleration terms, the respective α values are held constant and are the same as those used for the standard aMD simulations. However, in each case, the base-boost potential, Eb(0) has been substantially lowered.

The strength of the adaptive Gaussian bias potentials, a, was set to 10.0 kcal/mol for the torsional acceleration and 0.01- [Eb(tot)(0) - V(tot)] for the total background acceleration. The width of the adaptive Gaussian potentials, c=1.80 Å, was defined such that the full width of the adaptive Gaussian bias potential in the {PC1, PC2} projection space encompassed the entire PC-projection space sampled by a standard 5 ns MD simulation of substrate-free P450cam. The reader is referred to the SI for a discussion of the specific choice of Ad-aMD parameters used, as well as for full simulation details.

The five X-ray crystal structures of P450cam (PDB Ids: 5CP4, 1RE9, 1RF9, 3P6T, 3P6X) projected onto their own principal components {PC1, PC2} are depicted in Figure 1. The closed P450cam-camphor conformation is located at {-25.82, -6.87}, and the most open conformation (P450cam-- AdaC3-C8-Dans) is located at {21.01, -5.77}. For each of these X-ray crystal structures, the substrate was removed, and a standard 5 ns classical MD (CMD) simulation was performed. In all five CMD simulations, the system was found to be stable, affording backbone RMSDs to the respective X-ray crystal structure of 1.0-1.2 Å. The results of these initial simulations suggest that P450cam is stable in both closed and open conformations in the absence of the respective substrate. The conformational space sampling of a standard 25 ns CMD simulation for the closed conformation projected onto the principal component space is shown in Figure III.2 (black circles).

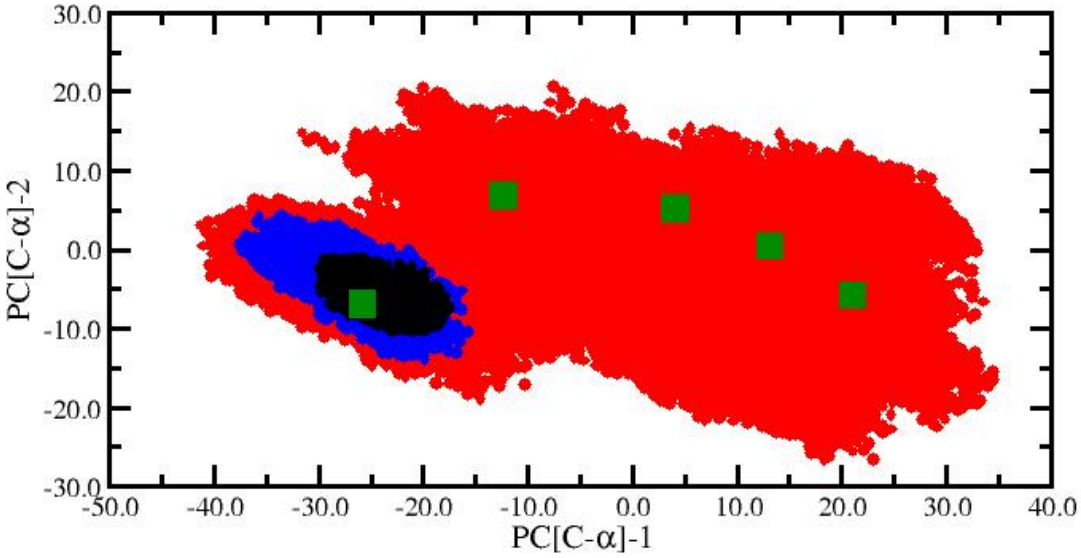

Figure III 2 The conformational dynamics of substrate-free P450cam projected into the principal component space {PC1,PC2} obtained from a PCA analysis of five X-ray crystal structures (green squares). The black circles represent the conformational space sampling afforded by a standard 25 ns classical MD simulation. The blue circles represent the conformational space sampling afforded by a standard 25,000,000 step AMD simulation. The red circles represent the conformational space sampling obtained from a 25,000,000 step Ad-AMD simulation. All simulations were initiated in the closed conformation located at {PC1, PC2} = {-25.82,-6.87}.

III.D Results

A 25,000,000 step standard aMD simulation initiated in the closed conformation was performed, and the (unweighted) conformational space sampling is depicted in blue (Figure III.2). The aMD trajectory clearly samples more conformational space than the standard CMD simulation, identifying a large number of substates as the system oscillates back and forth about the native closed conformation. However, even under these acceleration conditions, the system never exits the closed state. In comparison, a 25,000,000 step Ad-aMD simulation initiated

in the closed state (shown in red in Figure III.2) reveals the true molecular plasticity of substrate-free P450cam. In the first half of the Ad-aMD simulation, the system resides broadly in the closed conformation, and a considerable amount of adaptive bias is applied before the system exits the closed state and then rapidly samples a large number of open conformational states. The conformational space sampling observed in the Ad-aMD simulation not only encompasses all five X-ray crystal structures but also identifies new extended open conformations. An ensemble of structures collected over the Ad-aMD simulation is shown in Figure III.3.

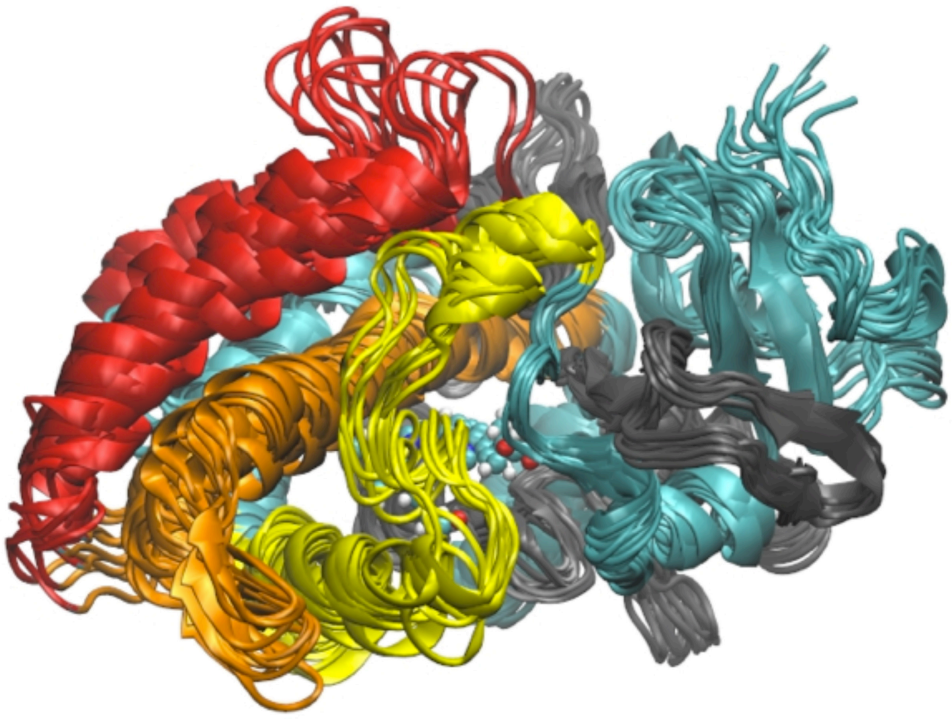

Figure III 3 Ensemble of structures for substrate free P450cam extracted from the Ad-aMD Simulation superposed by performing a backbone RMSD fit to residues 295-405 (gray). The Ad-aMD simulation reveals a complex collective motion of substrate-free P450cam involving predominantly the F and G helices and the FG loop (red), as well as the H and I helices and the H-I loop (orange) and the B and C helices and the B-C loop (yellow). The remainder of the protein is shown in cyan.

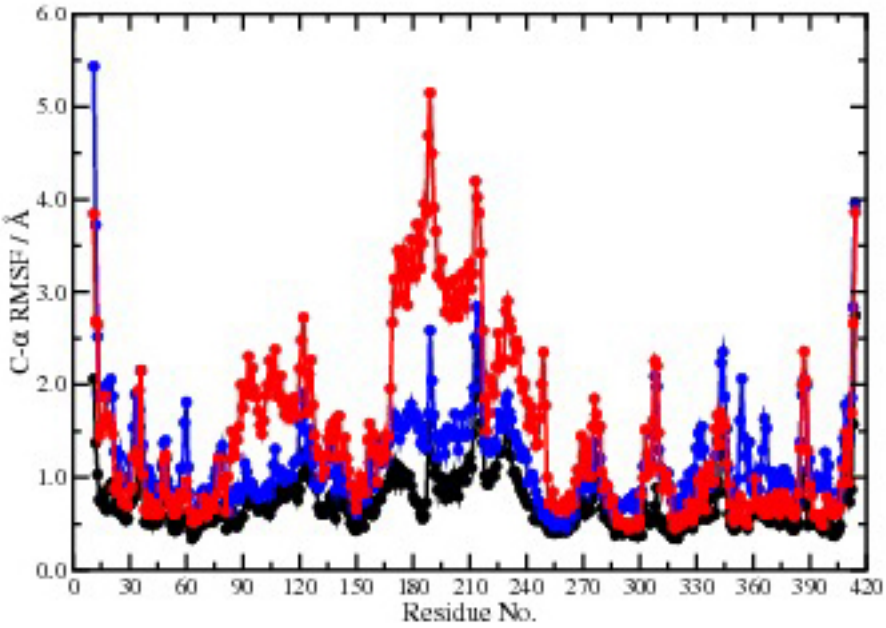

Figure III 4 C-α atom root-mean-square fluctuation (RMSF) obtained from the standard 25 ns MD simulation (black), the 25,000,000 step AMD simulation (blue), and the 25,000,000 step Ad-AMD simulation (red) of substrate-free P450cam initiated in the closed state. In each case, the structures collected across the ensemble were first superposed by performing a backbone RMSD fit to residues 295-405. A significant increase in the RMSFs from the Ad-AMD simulation is observed in the B and C helices and the B-C loop (residues 89-120) and the F, G, H, and I helices and the F-G and H-I loops (residues 171-267).

The Ad-aMD simulation reveals a complex collective motion of substrate-free P450cam involving predominantly the F and G helices and the F-G loop, as well as the H, I, B, and C helices and the B-C loop, which is seen to be considerably more flexible in the open conformational states. As such, over one-third of the residues in the protein are specifically involved in the functional dynamics of the system, a result, which is quantitatively represented in Figure III.4.

As a robust conformational space sampling protocol, the Ad-aMD approach is obviously much more efficient than standard aMD. Interestingly, the magnitude of the bias potential applied to the system in the Ad-aMD method is considerably lower than

that applied during the standard aMD simulation. The enhanced efficiency of Ad-aMD arises from the fact that, rather than raising and flattening the entire PES, the adaptive bias potential is selectively applied only to the lowest-energy regions on the underlying PES. As the perturbation applied to the system is, on average, much smaller, the resulting trajectory affords a much higher level of structural integrity. During the Ad-aMD simulation, the system was seen to come to within 1.3 Å of the backbone RMSD for all five X-ray crystal structures, a remarkable result considering that the backbone RMSD afforded by the CMD simulations initiated in these different conformations was found to be 1.0-1.2 Å.

A qualitative representation of the free-energy surface can be readily obtained from the Ad-aMD simulation by analyzing the magnitude of the adaptive boost energy across the conformational space, as shown in Figure III.5.

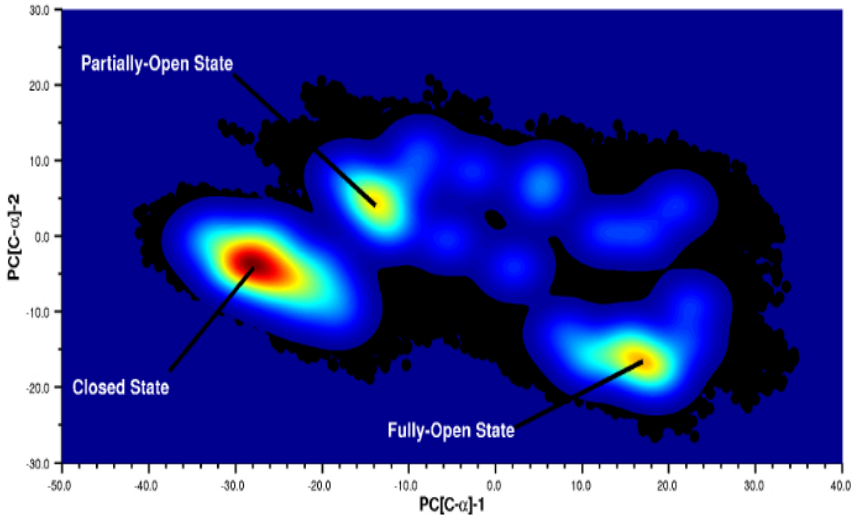

Figure III 5 Variation in the adaptive boost potential, $E_b(p,q)$ in the projected principal component space {PC1,PC2}obtained from the Ad- AMD simulation initiated in the closed state. The lowest boost potential (the base-boost potential) is shown in blue, and the largest boost potential, affording the most aggressive acceleration, is shown in red. The black circles represent the conformational space sampling obtained from the 25,000,000 step Ad-AMD simulation.

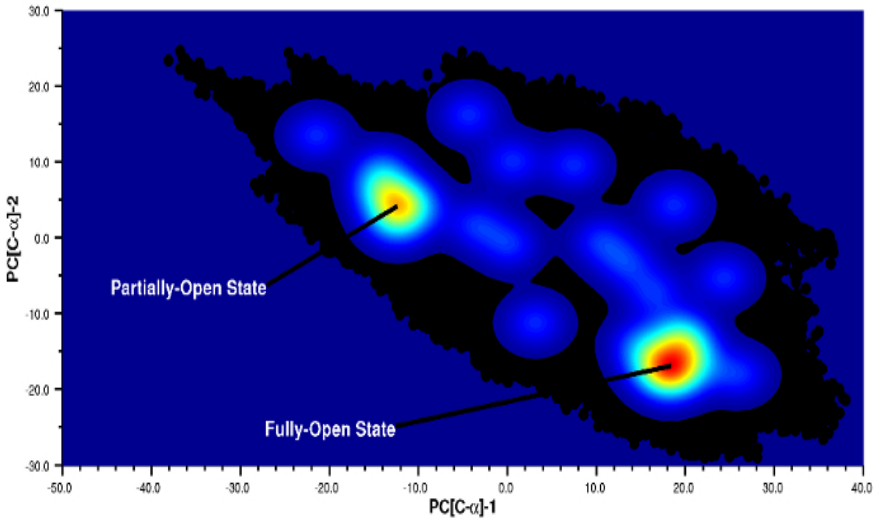

Figure III 6 Variation in the adaptive boost potential, $E_b(p,q)$ in the projected principal component space {PC1,PC2} obtained from a representative Ad-AMD simulation of substrate-free P450cam initiated in the open state (using the X-ray crystal structure coordinates from P450cam-AdaC3-C8-Dans). The lowest boost potential (the base-boost potential) is shown in blue, and the largest boost potential, affording the most aggressive acceleration, is shown in red. The black circles represent the conformational space sampling obtained from the 25,000,000 step adaptive AMD simulation. The adaption of the modified potential was terminated after 15,000,000 steps (see text). Notably, in comparison to the conformational space sampling obtained from the Ad-AMD simulation initiated in the closed state (Figures III.2 and III.5), the closed conformation is never visited during the simulation.

III.E Conclusions

Obviously the lowest free-energy conformations are associated with those regions where it was necessary to apply the largest adaptive boost potential. We clearly observe three local free-energy minima. The lowest energy state appears to be the closed conformation, broadly located at {PC1,PC2}= {-28.0,-5.0}. In addition, we observe two other local free- energy minima associated with a fully open {18.0,-18.0} and partially open {-14.0,4.0} conformational state. However, the free-energy statistics alone do not provide the complete story. We noticed during the Ad-aMD simulation that once the system had exited the closed state, it rapidly moved between the fully and partially open states but never returned to the closed conformation. This was a surprising result as we expected to see the system freely exchange between the closed and open states on the adaptive modified potential.

We subsequently performed four more 25 000 000 step Ad-aMD simulations, this time initiating the simulations in the open conformational states using the atomic coordinates from the available X-ray crystal structures. In all four of these Ad-aMD simulations, within 15 000 000 steps, the system sampled the entire open configurational space and exchanged so rapidly between the (already sampled) fully and partially open states that the addition of further adaptive Gaussian boost potentials was no longer appropriate. For the remainder of the simulation, the adaption was therefore switched off. The Ad-aMD simulations readily identified the two local free-energy minima associated with the fully open and partially open conformational states but never visited the closed conformation, as shown in Figure III.6.

It should be noted that, unlike several other enhanced conformational space

sampling methods in which the system is driven along a predefined reaction coordinate, whether the system would naturally follow that reaction coordinate or not, Ad-aMD will not force a system to undergo an unnatural transition to a kinetically inaccessible state. The results of all of the Ad-aMD simulations strongly suggest that while the closed conformation is energetically stable, it is kinetically inaccessible to substrate-free P450cam. Therefore, we propose that substrate-free P450cam exists in equilibrium between a fully open and partially open state, located at {PC1,PC2} = {18.0,-18.0} and {-14.0,4.0}, respectively (Figure III.6). The lowest-energy, and hence most populated, conformation is the fully open state. As such, P450cam can accommodate a variety of substrates of differing size. Larger ligands can enter the binding pocket when P450cam exists in the fully open state and the resulting substrate- bound system is trapped in an open conformation via a population shift mechanism. However, when a small ligand, such as camphor or DTT, fully enters the binding pocket, it triggers an induced-fit mechanism, resulting in the protein "zipping up" around the bound substrate, leading to the formation of an energetically stable closed conformational state. These results are corroborated by recent experimental studies[27] and potentially provide detailed insight into the functional dynamics and conformational behavior of the entire cytochrome-P450 superfamily. Indeed, several other members of the P450 superfamily, most notably P450 EryK32 and PikC,[33] have also been shown to coexist in open and closed states in the absence of substrate.

In conclusion, we have presented a novel variant of the accelerated molecular dynamics method called adaptive aMD. Ad-aMD is an extremely efficient and robust conformational space sampling algorithm which also affords a qualitatively accurate

description of the free-energy surface. A series of Ad-aMD simulations performed on substrate-free P450cam revealed that this system exists in equilibrium between a fully open and partially open conformational state. The mechanism for substrate binding to P450cam depends on the size of the ligand. Larger ligands enter the P450cam binding pocket, and the resulting substrate-bound system is trapped in an open conformation via a population shift mechanism. Small ligands, such as camphor or DTT, which fully enter the binding pocket, cause an induced-fit mechanism, resulting in the formation of an energetically stable closed conformational state.

III.F Supporting Information

III.F.1 Computational Details

All classical MD (CMD), accelerated molecular dynamics (aMD) and adaptive accelerated molecular dynamics (Ad-aMD) simulations were performed on a construct of substrate-free P450cam comprising residues LEU(11) to VAL(414) [using the same residue numbering employed in the X-ray crystal structure PDB IDs: 5CP4, 1RE9, 1RF9, 3P6T and 3P6X]. In some X-ray crystal structures, particularly those associated with the more 'open' conformational states, some atomic coordinates were undetermined: For example, residues 90-94 and 104-106 in the P450cam-AdaC3-C8-Dans system. These residues were modeled using the structural information from the (complete) closed conformation (PDB ID 5CP4). All simulations were performed on substrate-free P450cam, so the associated substrate was removed. The AMBER ff99SB force-field was employed for all solute residues with the exception of residue

CYS(357) [once again using the same residue numbering associated with the X-ray crystal structure PDB IDs) which is chemically bound to the heme group. A force-field was generated in-house for this non-standard CYS-heme residue following the protocol described by A. Oda *et al.* [*J. Comput. Chem.* **2005**, *26*(8), 818-826].

Each substrate-free P450cam system was placed at the center of a cubic box containing pre- equilibrated solvent water molecules which, were treated using the TIP4P force-field. The dimensions of the solvent box were defined such that the distance between the surface of the protein and the edge of the box was at least 12.0 Angstroms and Na+ counter-ions were added to enforce neutrality of the simulation cell. After performing standard energy minimization and equilibration procedures, a 5-ns CMD simulation was performed on each conformation of substrate-free P450cam at 300K under ambient pressure conditions. These simulations were performed under periodic boundary conditions using a time-step of 1.0-fs. The temperature of the system was controlled using a Langevin thermostat and a Berendsen weak-coupling pressure-stat was employed. Electrostatic interactions were treated using the Particle Mesh Ewald (PME) method with a direct space sum cutoff of 10.0 Angstrom. As discussed in the paper, these initial 5-ns CMD simulations confirmed that the different conformational states (both open and closed) of substrate-free P450cam were stable in the absence of the respective substrate. The simulations also afforded an estimate of the average dihedral angle energy, V(dih), and total potential energy, V(tot), for each system, which were used to define the specific acceleration parameters in the aMD and Ad-aMD simulations. All aMD and Ad-aMD simulations were performed under exactly the same physical conditions as described above for the standard CMD

simulations. All simulations were performed using an in-house modified version of the AMBER10 sander simulation module.

III.F.2 Principal Component Analysis (PCA)

Atomic coordinates for the available P450cam X-ray crystal structures (PDB IDs: 5CP4, 1RE9, 1RF9, 3P6T and 3P6X) were used to perform a principal component analysis (PCA). Iterated rounds of structural superposition were used to identify the most structurally invariant region, which was identified to consist of residues 295 to 405 [using the X-ray crystal structure PDB residue numbering]. The five X-ray crystal structures were then superposed using backbone (N,Cα,C') root-mean-square fitting to this structurally invariant 'core' and a PCA was performed using the coordinates of the backbone C-α atoms for residues (15-89, 95-103, 107-405). The PCA was therefore performed on the whole P450cam construct excluding the flexible N- and C-terminal tails and residues 90-94 and 104- 106, whose atomic coordinates were absent in some of the X-ray crystal structures and subsequently modeled (see above). The two lowest principal component eigenvectors {PC1, PC2} which accounted for 80% of the structural covariance were then used to generate the two-dimensional PC-projection plots (Figures III.3 and III.5 in the manuscript). The {PC1,PC2} projection plot allows us to compare the efficiency of 'essential' conformational space sampling obtained from the CMD, standard aMD and Ad-aMD simulations. The {PC1,PC2} conformational sub-space was also used as the framework for the adaptive aMD simulations (see manuscript).

III.F.3 Acceleration Parameters: Standard aMD

The acceleration level (ie. how much the potential energy surface is raised and flattened) in an aMD simulation is determined by the acceleration parameters E_b and α. In the standard aMD protocol, these two acceleration parameters are constant. Generally speaking, more aggressive acceleration can be achieved by either increasing the magnitude of E_b or decreasing the magnitude of the acceleration parameter, α. However, if the boost energy, E_b, is too large and the acceleration parameter, α, is too small, the modified potential energy surface becomes iso-energetic, resulting in a random walk through phase-space, causing the system to spend a large proportion of time sampling energetically unfavorable conformational space (and can eventually result in unfolding). The 'optimal' acceleration parameters [those that efficiently enhance conformational space sampling without generating instabilities in the trajectory and a random walk] are extremely system-specific and depend on numerous factors including the size of the system and the nature of the underlying potential energy surface. Previous aMD studies on proto-typical systems (most notably a recent very detailed study on ubiquitin [*J. Am. Chem. Soc.* **2009**, *131*(46), 16968-75]) have identified that for torsional acceleration, the optimal value of [E_b(dih)-V(dih)] in kcal/mol (where V(dih) is the average torsional potential energy obtained from a standard MD simulation) is approximately equal to 3 to 5 times the number of (solute) residues in the system and the associated acceleration parameter, α(dih), should be set to one fifth of this value. For the background, total acceleration, [E_b(tot)-V(tot)] and α(tot) should both be equal to 0.16 kcal/mol times the number of atoms in the simulation cell (NASC) [*J. Chem. Phys.* **2007**, *127*, article no. 155102]. Noticeably, in

this 'dual boost' aMD protocol, the total background acceleration is much weaker than the torsional acceleration in order to prevent a breakdown in the local structuring of the solvent water molecules around the solute. In light of this discussion, the aMD simulations presented in this manuscript were performed using the acceleration parameters:

Torsional aMD

$$[E_{dih} - V(dih)] = 1400\ kcal/mol_{\ b}$$

$$\alpha_{dih} = 280\ kcal/mol$$

Total aMD

$$[E\ tot\ -\ V(tot)] = 0.16 * (\mathrm{NASC})\ kcal/mol\ b$$

$$\alpha\ tot\ = 0.16 * (NASC)\ kcal/mol$$

Notice that the particular construct of substrate-free P450cam in this study has 404 residues, so [Eb(tot)-V(tot)] is approximately equal to 3.5*(No. solute residues). In a series of initial test aMD simulations, we found that using a more aggressive acceleration level resulted in unstable trajectories and partial unfolding for the 'open' states.

III.F.4 Acceleration Parameters: Adaptive aMD

Similar to the standard aMD protocol, we would like to point out that the choice of acceleration parameters for the adaptive aMD method is also highly system-specific. It should also be recognized that the specific choice of these parameters will determine the balance between how efficiently one samples the conformational space and the integrity of the resulting trajectory, including the effective resolution of the

qualitative free energy surface. For example, let us consider the base-boost potential, $E_b(0)$: If $E_b(0)$ is set to a low value (for example $[E_b(0) - V] \sim 0.0$ kcal/mol, which is the equivalent of performing a standard CMD simulation), the extent of conformational space sampling between the adaptive steps in the Ad-aMD protocol will be small and a very long Ad-aMD trajectory would be required for exhaustive conformational space sampling. The integrity of the resulting trajectory and the effective resolution of the qualitative free energy surface would however be very good: The system will basically get stuck in every conformational sub-state and micro-state, which will then be represented in the qualitative free energy surface as the adaptive Gaussian bias potentials are applied in a history- dependent fashion. On the other hand, if $E_b(0)$ is set to a very high value, the efficiency of the conformational space sampling between adaptive steps will be very fast and, particularly in those regions of the PES that are not highly structured, the system will rapidly evolve from one conformational sub-state to another, but the integrity of the resulting trajectory and the resolution of the effective qualitative free energy surface will be much reduced. A similar argument can be made for the height (a) and width (c) of the adaptive Gaussian bias-potentials: The application of strong, broad adaptive Gaussian bias-potentials during the Ad-aMD simulation will significantly enhance the efficiency of conformational space sampling, but the integrity of the resulting Ad-aMD trajectory and the resolution of the qualitative free energy surface will be significantly reduced. In the present substrate-free P450cam 'dual boost' Ad-aMD simulations, the following Ad-aMD parameters were employed:

Torsional Ad-aMD

$$[E_{dih}(0) - V(dih)] = 700\ kcal/mol\ b$$

$$\alpha_{dih} = 280\ kcal/mol\ a = 10.0\ kcal/mol\ c = 1.80 \quad Angstroms$$

Total Ad-aMD

$$[Etot\ (0) - V(tot)] = 0.08*(NASC)\ kcal/mol$$

$$b\ \alpha tot = 0.16*(NASC) kcal/mol$$

$$a = 0.01*[Etot\ (0) - V(tot)]\ b$$

$$c = 1.80 \quad Angstroms$$

Notice that for both torsional and total acceleration terms, the respective α values are held constant and are the same as those used for the standard aMD simulations (see above). However, in each case, the base-boost potential, $E_b(0)$ has been substantially lowered such that $[E_b(0)\text{-}V]$ is equal to half the value used in the standard aMD simulations, $[E_b\text{-}V]$. In a series of initial Ad-aMD simulations, we found that a substantial increase in the conformational space sampling could be achieved if the strength of the adaptive Gaussian bias-potentials (*a*) was set to be between 1% and 5% of $[E_b(0)\text{-}V]$. The width of the adaptive Gaussian potentials, (*c*), was defined such that the full width of the adaptive Gaussian bias- potential in the {PC1,PC2} projection space encompassed the entire PC-projection space sampled by the standard 5-ns CMD simulation of substrate-free P450cam. For the sake of simplicity, the magnitude of the adaptive Gaussian bias-potentials (i.e. the height parameter (*a*) and width parameter (*c*)) were kept constant during the Ad-aMD simulations.

III.F.5 The relative efficiency of aMD and Ad-aMD

As we discuss in the manuscript, the Ad-aMD method is particularly well suited to proteins with highly structured potential energy surfaces. This statement

refers to the relative efficiency of the standard aMD and Ad-aMD approaches: If we look at the successful applications of the standard aMD approach, we find that the systems studied fall into two general categories. The first category consists of relatively small, usually globular, proteins such as GB3 or ubiquitin, with a single, well defined native fold. The second group consists of highly flexible systems, such as poly-peptides or natively unstructured protein (NUP) constructs. If we consider the potential energy surface (PES) for these systems, we can understand why the standard aMD approach has proven to be so successful for these types of systems. In the case of small globular proteins, the PES has a well-defined single minimum. Following Frauenfelder's seminal work on protein dynamics, the system undergoes rare excursions away from the native fold, exploring high energy regions of the configurational space, possibly visiting some meta-stable states (conformational sub-states and micro-states associated with the native fold), before returning to the energy minimum. The application of the bias potential using the standard aMD approach enhances the frequency of these excursions, thereby affording efficient configurational space sampling. In the case of poly-peptides or NUPs, the PES looks more like an egg-box: These systems possess a large number of local energy minima separated by relatively small energy barriers.

Once again, the application of the bias potential in the standard aMD protocol raises and flattens the underlying PES, lowering the magnitude of these small energy barriers and enhancing the exchange rate between the numerous local potential energy minima. Many biological systems (including P450cam) possess a very different type of PES. These are systems that can exist in several distinct conformational states. Each

conformational state can be considered to be a low potential energy well on the PES and each possesses a hierarchy of associated sub-states and micro-states. This is exactly what we mean when we refer to a 'highly structured PES' and we depict this diagrammatically in schematic 1 in the manuscript. For these types of systems, the standard aMD approach is not particularly efficient: By raising and flattening the entire PES, the system, starting in a well-defined conformational state, starts to explore the associated sub-states and micro-states within the broad potential energy well that defines the conformational state. However, even on the modified (accelerated) potential, the energy statistics dictate that as the system moves away from the low energy conformational state (i.e. as it starts to climb the rugged potential energy well), the probability of the system returning to the conformational state from which it came is always greater than the probability that it will exit the broad potential energy well and search out a different conformational state. As a result, one generally sees that the system oscillates back and forth within the potential energy well, but does not leave the conformational state.

Simply increasing the acceleration level (i.e. raising and flattening the modified PES even further) might increase the probability that the system will exit the conformational state, but this will also render large regions of the PES iso-energetic, resulting in a 'random walk' through configurational space, which is certainly not conducive to efficient conformational space sampling. In the Ad-aMD approach, we define a conformational sub-space using one or more collective coordinates, which allow us to differentiate one conformational state from another. Within the framework of this conformational sub-space, we can then selectively apply a history-dependent

adaptive boost potential which specifically destabilizes the low potential energy wells (i.e. the conformational states) on the PES without rendering large regions of the PES iso-energetic, allowing the system to efficiently exchange between different conformational states.

One could perform Ad-aMD on any system, but we believe that the enhanced efficiency of the Ad-aMD approach compared to the standard aMD protocol will be most significant for those systems that possess a highly structured PES for the reasons that we have outlined above. Along the same lines, we would like to point out that, given a long enough trajectory, the standard aMD approach will sample the same conformational space as the Ad-aMD method.

III.J Acknowledgements

P.R.L.M. gratefully acknowledges HHMI for financial support. Work at UCSD is supported in part by the NSF, NIH, CTBP, NBCR, and the NSF Supercomputer Centers.

This chapter contains material from “Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam” published in 2011 in *Journal of Physical Chemistry Letters* (Volume 2, pages 158-164), authored by Phineus R. L. Markwick, Levi C.T. Pierce, David B. Goodin, and J. Andrew McCammon. All material has been reproduced with the consent of all other authors.

III.K References

(1) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. Science 1991, 254, 1598–1603.

(2) Benkovic, S. J.; Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. Science 2003, 301, 1196–1202.

(3) Tousignant, A.; Pelletier, J. N. Protein Motions Promote Catalysis. Chem. Biol. 2004, 11, 1037–1042.

(4) Rousseau, F.; Schymkowitz, J. A Systems Biology Perspective on Protein Structural Dynamics and Signal Transduction. Curr. Opin. Struct. Biol. 2005, 15, 23–30.

(5) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. Enzyme Dynamics During Catalysis. Science 2002, 295, 1520–1523.

(6) Elber, R. Long-time scale Simulation Methods. Curr. Opin. Struct. Biol. 2005, 15, 151–156.

(7) Berne, B. J.; Straub, J. E. Novel Methods of Simulating Phase Space in the Simulation of Biological Systems. Curr. Opin. Struct. Biol. 1997, 7, 181–189.

(8) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transi- tion Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. Annu. Rev. Phys. Chem. 2002, 53, 291– 318.

(9) Schlitter, J.; Engels, M.; Kruger, P. J. Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions. J. Mol. Graph. 1994, 12, 84–89.

(10) Mitsutake, A.; Sugita, Y.; Okamoto, Y. J. Replica-Exchange Multicanonical and Multicanonical Replica-Exchange Monte Carlo Simulations of Peptides. I. Formulation and Benchmark Test. J. Chem. Phys. 2003, 118, 6664–6675.

(11) Laio, A.; Parrinello, M. Computing Free Energies and Accelerating Rare Events with Metadynamics. In Computer Simula- tions in Condensed Matter: From Materials to Chemical Biology; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer Verlag: Berlin, Heidelberg, Germany, 2006; Vol. 1, pp 315-347.

(12) 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.

(13) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. Toward a Unified Representa- tion of Protein Structural Dynamics in Solution. J. Am. Chem. Soc. 2009, 131, 16968–16975.

(14) Cervantes, C. F.; Markwick, P. R. L.; Sue, S. C.; McCammon, J. A.; Dyson, H. J.; Komives, E. A. Functional Dynamics of the Folded Ankyrin Repeats of IκBR Revealed by Nuclear Mag- netic Resonance. Biochemistry 2009, 48, 8023.

(15) Markwick, P. R. L.; Showalter, S. A.; Bouvignies, G.; Bruschweiler, R.; Blackledge, M. Structural Dynamics of Protein Backbone j Angles: Extended Molecular Dynamics Simulations Versus Experimental 3J Scalar Couplings. J. Biomol. NMR 2009, 45, 17–21.

(16) Markwick, P. R. L.; Cervantes, C. F.; Abel, B. L.; Komives, E. A.; Blackledge, M.; McCammon, J. A. Enhanced Conformational Space Sampling Improves the Prediction of Chemical Shifts in Proteins. J. Am. Chem. Soc. 2010, 132, 1220–1221.

(17) Mukrasch, M. D.; Markwick, P. R. L.; Biernat, J.; Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified from Residual Dipolar Couplings and Molecular Simulation. J. Am. Chem. Soc. 2007, 129, 5235–5243.

(18) Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. The 2.6-Å Crystal Structure of Pseudomonas Putida Cytochrome P-450. J. Biol. Chem. 1985, 260, 16122–16130.

(19) Poulos, T. L.; Finzel, B. C.; Howard, A. J. High-Resolution Crystal Structure of Cytochrome P450cam. J. Mol. Biol. 1987,
195, 687–700.

(20) Dmochowski, I. J.; Crane, B. R.; Wilker, J. J.; Winkler, J. R.;
Gray, H. B. Optical Detection of Cytochrome P450 by Sensi- tizer-Linked Substrates. Proc. Natl. Acad. Sci. U.S.A. 1999, 96, 12987–12990.

(21) Dunn, A. R.; Dmochowski, I. J.; Bilwes, A. M.; Gray, H. B.; Crane, B. R. Probing the Open State of Cytochrome P450cam with Ruthenium-Linker Substrates. Proc. Natl. Acad. Sci. U.S. A. 2001, 98, 12420–12425.

(22) Hays, A. M. A.; Dunn, A. R.; Chiu, R.; Gray, H. B.; Stout, C. D.; Goodin, D. B. Conformational States of Cytochrome P450cam Revealed by Trapping of Synthetic

Molecular Wires. J. Mol. Biol. 2004, 344, 455–469.

(23) Poulos, T. L.; Finzel, B. C.; Howard, A. J. Crystal Structure of Substrate-Free Pseudomonas Putida Cytochrome P-450. Bio- chemistry 1986, 25, 5314–5322.

(24) Lewis, B. A.; Sligar, S. G. Structural Studies of Cytochrome P-450 Using Small Angle X-Ray Scattering. J. Biol. Chem. 1983, 258, 3599–3601.

(25) Di Primo, C.; Deprez, E.; Hoa, G. H.; Douzou, P. Antagonistic Effects of Hydrostatic Pressure and Osmotic Pressure on Cytochrome P-450cam Spin Transition. Biophys. J. 1995, 68, 2056–2061.

(26) Fischer, M. T.; Scarlata, S. F.; Sligar, S. G. High-Pressure Investigations of Cytochrome P-450 Spin and Substrate Binding Equilibria. Arch. Biochem. Biophys. 1985, 240, 456– 463.

(27) Y-T. Lee, Y.-T.; R.F. Wilson, R. F.; Rupniewski, I.; Goodin, D. B. P450cam Visits an Open Conformation in the Absence of Substrate. Biochemistry 2010, 49, 3412–3419.

(28) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Ross, W. C.; et al. AMBER10; University of California: San Francisco, CA, 2008.

(29) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Param eters. Proteins: Struct., Funct., Bioinf. 2006, 65, 712–725.

(30) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. Ras Conforma- tional Switching: Simulating Nucleotide Dependent Confor- mational Transitions with Accelerated Molecular Dynamics. PLOS Comput. Biol. 2009, 5, e1000325.

(31) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. Sampling of Slow Diffusive Conformational Transitions with Accelerated Molecular Dynamics. J. Chem. Phys. 2007, 127, 155102.

(32) Savino, C.; Montemiglio, L. C.; Sciara, G.; Miele, A. E.; Kendrew, S. G.; Jemth, P.; Gianni, S.; Vallone, B. Investigating the Structural Plasticity of a Cytochrome P450: Three Dimen- sional Structures of P450 EryK and Binding to its Physiologi- cal Substrate. J. Biol. Chem. 2009, 284, 29170–29179.

(33) Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. The Structural Basis for Substrate Anchoring, Active Site Selectivity and Product Formation by P450 PikC from Streptomyces Venezuelae. J. Biol. Chem. 2006, 281, 26289–26297.

IV Accessing Millisecond Time scales

IV.A Abstract

In this work we combine the enhanced sampling method, accelerated molecular dynamics (aMD) with the inherent power (as implemented in Amber) of graphics processor units (GPUs) and apply the implementation to Bovine Pancreatic Trypsin Inhibitor (BPTI). A 500ns aMD simulation is compared to a previous millisecond unbiased brute force MD simulation carried out on BPTI showing the same conformational space is sampled by both approaches. We observe the correct relative populations defined by the χ_1, χ_2, and χ_3 dihedral angles of the disulfide bond C14-C38, furthermore, we obtain improved agreement with observed chemical shift differences from prior experimental work. To our knowledge this represents the first implementation of aMD on GPUs and also the longest aMD simulation of a biomolecule run to date. Our implementation will be made available to the community with the next release of the Amber software suite (v12) enabling researchers routine access to ms events sampled from dynamics simulations using off the shelf hardware.

IV.B Introduction

Conventional molecular dynamics allows one to access time scales on the order of 10 to 100s of nanoseconds, however, many biological processes of interest occur on longer time scales of up to milliseconds or more[1-6]. Efforts to explore these long time scales have lead to the development of several advanced sampling technique

such as conformational flooding[7,8], hyperdynamics[9,10], metadynamics[11-13], and the adaptive biasing force method[13-15]. Inspired by Voter, accelerated molecular dynamics (aMD), is an additional enhanced conformational sampling method, that provides access to time scales beyond those obtainable with conventional MD (cMD)[9]. Here, we emphasize one of the great advantages of aMD which is no prior knowledge of the potential energy landscape needs to be known and, consequently, no reaction coordinate needs to be defined prior to running the simulation.

In addition to advances in algorithms to achieve dynamics on longer time scales recent efforts by D. E. Shaw Research have focused on building a specialized computer, Anton, with the sole purpose of simulating protein dynamics[16]. With this great engineering advancement simulation time scales have been pushed to range 100s of microseconds to 10s of milliseconds using unbiased brute force cMD[17,18]. While time on an Anton machine has been graciously granted to the scientific community, access is still limited and many researchers are stuck waiting in queues on crowded super computers or local clusters. Recently, the advancement of computational science on conventional graphic processing units (GPU) has allowed researchers efficient and inexpensive access to 10s to 100s of microseconds of simulation time on just a single desktop computer[18-20]. By combining the advanced sampling method of aMD and the inherent power of the GPU, we present the synthesis of a tool that allows researchers access to inexpensive efficient exploration of long time scale events. It should be noted at the outset, however, that details of the time evolution are not readily available with aMD.

IV.C Implementation

In its original form, the aMD method modifies the potential energy landscape by raising energy minima that lie below a defined threshold level, while leaving those areas lying above the threshold unmodified. As a result, barriers separating adjacent energy basins are reduced, providing the simulation access to conformational space that cannot be easily accessed in a cMD simulation. Historically, the implementation of aMD was first done by Hamelberg *et al.* within the framework of the sander module in the AMBER 7 package and used to study several small peptide and protein systems[21,22]. More recently, Wang *et al.* ported the aMD method to NaMD taking advantage of the inherent scalability of this MD package[23]. Following this approach, we have ported aMD to the three main MD engines included in Amber 12: the CPU versions sander and pmemd, and the GPU version pmemd.cuda. The performance enhancements for cMD on GPUs obtained with pmemd.cuda alone have been remarkable and can be found on the Amber website[24] and in the following publication[25]. For this work we have used our implementation of aMD in the pmemd.cuda MD engine and will refer to it simply as aMD.

Bovine Pancreatic Trypsin Inhibitor (BPTI) is a small protein with 58 residues, that has been extensively studied experimentally, being the first subject of NMR experiments to characterize individual hydration water molecules in proteins[26], and also the first protein to be simulated with molecular dynamics[27]. More recently D.E. Shaw Research reported a remarkable 1.03ms MD simulation of BPTI in explicit water[28]. Using our aMD implementation in pmemd.cuda, we have performed a 500ns aMD simulation of BPTI, maintaining the same conditions as the simulation on Anton.

Using the 1.03ms simulation of BPTI acquired from D.E. Shaw Research throughout our analysis, we have shown both methodologies sample the same conformational space. Additionally, we performed a second 500ns cMD simulation to use as a measure of the amount of sampling attainable by regular MD with the same computational effort To the best of our knowledge our aMD implementation is the first to support GPU acceleration and the work presented here represents the longest single aMD simulation of a biomolecule run to date. Both Amber simulations were completed in two weeks on individual desktop computers containing a single $500 GTX580 GPU. Details of the simulations can be found in the supplementary material.

IV.D Results

IV.D.1 Structural Analysis

In the Anton millisecond simulation, five long-lived structural states were identified that persisted for 6-26μs with deviations of up to 3.5Å from the crystal structure, 5PTI, were observed[28,29]. Using these same five structures and the crystal structure as references, we calculated RMSD values (heavy backbone atoms) along our 500ns aMD simulation, shown in Figure IV.1. RMSD results show we sample structures less than 1.0Å away for four of the five long-lived states and 2.0Å for the third state. The RMSD with respect to the X-RAY structure shows that the protein moves away from the crystal structure 10ns into the simulation and achieves a maximum RMSD of 2.9 Å before coming back to within 0.45Å emphasizing that the system is sampling states both far from, and near to, the crystal structure as was seen

in the long 1ms cMD simulation. Three more large transitions are observed in the simulation around 300ns, 340ns, and 450ns while none are observed in the 500ns cMD simulation.

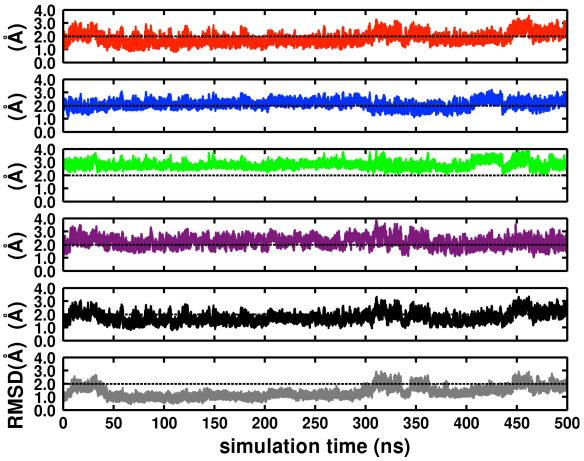

Figure IV 1 RMSD of the aMD trajectory with respect to each of the long-lived structures identified from the 1ms cMD BPTI simulation. The colors red, blue, green, purple, and black correspond to the colors of the kinetic clusters identified in the 1ms cMD BPTI paper. The grey color represents RMSD with respect to the crystal structure 5PTI. The dashed black line is set at 2.0Å in all the plots as a reference.

IV.D.2 Explored Conformational Space

To characterize the conformational states explored in more detail, principal component analysis was carried out on the 500ns aMD simulation using Bio3D[30]. Figure 2a displays the two-dimensional representation of the structural dataset as a projection of the Boltzmann reweighted distribution onto the subspace defined by the first and second principal component vectors (PC1 and PC2) built from the C-alpha atoms spanning residues 4 to 54. In this analysis, PC1 and PC2 describe 40% and 16%, respectively, of the total variance of the motions in the simulation (Movies S2-S3). The five long-lived structures and the crystal structure were then projected into

this space. The 500ns cMD control simulation and the 1ms cMD simulation were also projected into the subspace defined by the aMD simulation (Figures IV.2b and IV.2c). It is clear from this figure that the 500ns cMD control simulation does not explore the amount of space that the aMD simulation does and is trapped in the basin localized around the crystal structure. The aMD simulation has a very broad basin around the crystal structure (-2.5, 0.0) as one might expect. This is in contrast to the 1ms cMD simulation, which heavily samples the basin around (-2.5, -5.0) away from the crystal structure. As explained later this is what is ultimately responsible for the observed populations in the Anton 1ms cMD simulation differing substantially from experiment.

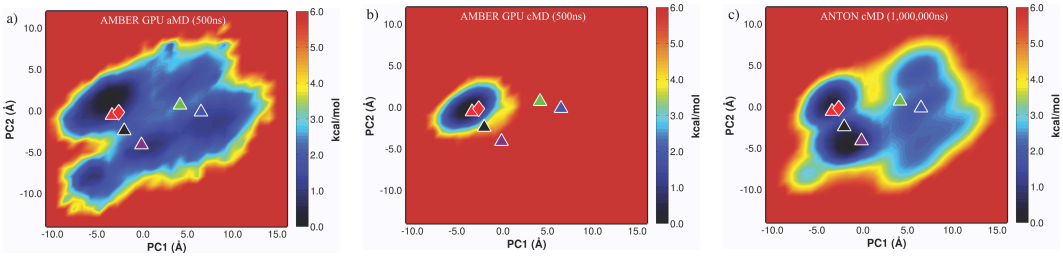

Figure IV 2 The free energy principal component projection of (a) 500ns aMD simulation (b) 500ns cMD simulation and (c) 1ms simulation onto (PC1, PC2) defined by the aMD simulation. The long-lived structures are projected onto the free energy surface and are labeled as red, blue, green, purple, and black triangles. The crystal structure, 5PTI, is demarked by the red diamond (see also movie S1).

IV.D.3 NMR Observables

The recent analysis carried out by Xue *et al.* examined in detail the χ_1, χ_2, and χ_3 dihedral angles associated with the disulfide bond formed between Cysteine 14 and Cysteine 38 during the course of the 1ms cMD BPTI simulation[31]. In this communication a similar analysis was performed. Figure IV.3a shows the Boltzmann

reweighted χ_1(C14) vs. χ_1(C38) free energy surface explored throughout the 500ns aMD simulation. In contrast, the 500ns control cMD simulation never visits the minima explored by the aMD simulation and again remains trapped in a state close to the crystal structure (Figure IV.3b), as was seen in the PC projections. Comparing the free energy surface explored by the long 1ms cMD simulation (Figure IV.3c) to that of our aMD simulation, it is clear the aMD simulation explores the same states as the unbiased simulation The χ_2, and χ_3 dihedral angles associated with Cys14 and Cys38 can be used to further describe isomerization configurations of the disulfide bond: the major state M which consists of three substates (M1, M2, and M3) and the two minor or excited states m_{C38}, and m_{C14}[31]. A detailed description of these states is included in the supplementary material (Figure IV.4), which can be compared to those analyzed by Xue *et al*. for the 1ms cMD simulation. In contrast with the 1ms cMD simulation, which predicts the excited state m_{C14} to be the most populated, we find the major state M to be the most populated (61%) after Boltzmann reweighting our distribution. We also find agreement with the observed experimental excited state populations for m_{C14} and m_{C38} (Table IV.2).

Table IV 1 ^{15}N Chemical Shift Differences between the conformational states M1(M),m_{C14}, and m_{C38}.

	$\Delta\delta_{M1,mC14}$(ppm)			$\Delta\delta_{M1,mC38}$(ppm)		
res.	exptl[32]	1ms cMD[33]	500ns aMD	exptl[26]	1ms cMD[33]	500ns aMD
C14	3.6	1.8	3.4	-0.4	0.6	0.1
K15	4.7	-1.4	0.2	-0.5	-1.5	-0.1
C38	|0.8|	-2.2	-0.1	-1.7	-1.9	-3.2
R39	|1.2|	-0.5	-2.3	-3.7	-2.6	-3.7

Structures from states M, m_{C38}, and m_{C14} were extracted from the trajectory by using the lowest energy structure in the M state as a reference to select an ensemble of structures with similar energy from each of the three basins for performing the chemical shift analysis. Using the SHIFTX2 software package[34] chemical shift differences were computed between the ensembles representing the different substates (Table IV.1). A rms deviation of 1.8 ppm was obtained from the aMD simulation (computed for the shifts with known sign) compared to 2.7 ppm computed from the 1ms cMD simulation. In general, good agreement is achieved with values calculated from the 1ms cMD simulation and the experimental values[32,35]. We would like to highlight the fact that one cannot compute these values using only 500ns of cMD since the simulation never explores the excited states during the course of such a short simulation (Figure IV.3b).

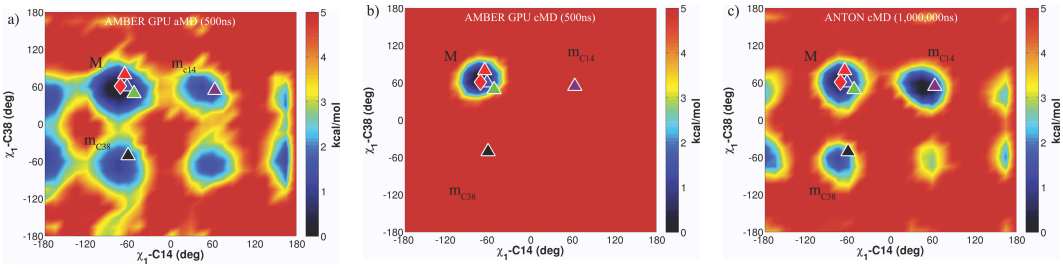

Figure IV 3 The χ_1-C14 vs. χ_1-C38 dihedral angle free energy surfaces of (a) 500ns aMD simulation (b) 500ns cMD simulation and (c) 1ms cMD simulation. The long-lived stable states are plotted onto the free energy surface and are labeled as red, blue, green, purple, and black triangles. The crystal structure, 5PTI, is demarked by the red diamond. The major state M is labeled along with the minor states m_{C14} and m_{C38}.

IV.D.4 Water Occupancy

Examination of the water occupancy throughout the aMD simulation correctly identifies the four long-lived waters in agreement with experiment[36]. The longest-lived water, W122, is identified by the 1ms cMD simulation as having a lifetime of 14 μs. During the course of the aMD simulation, several exchange events are captured at this location and an interesting "revolving door" mechanism is identified, whereby the disulfide bond rotates around and pushes the water out of the pocket (highlighted in movie S4). The longest binding events of water molecules at this site occur when in the "crystallographic" basin of the aMD simulation, consistent with the 1ms cMD simulation.

IV.E Conclusion

This work shows that using conventional off the shelf GPU hardware combined with an enhanced sampling algorithm, events taking place on the millisecond time scale can be effectively, and correctly, sampled with dynamics simulations orders of magnitude shorter (2000X) than those time scales. The

implementation is validated in this work by comparison with a long unbiased cMD simulation and experimental information. Structurally, we show the long-lived states identified by the 1ms cMD simulation have been sampled both in terms of RMSD and coverage in PC space. We demonstrate that access to these states cannot be obtained with the same length of simulation carried out using conventional MD. Important structural waters were preserved and found to exhibit the same occupancies as those experimentally. Most importantly, chemical shift differences computed from the aMD simulation were found to be in better agreement with experiment compared to those calculated from the long 1ms cMD simulation. Furthermore, we show that the correct relative dihedral populations were obtained from the aMD simulation compared to the 1ms cMD simulation. We conclude by emphasizing never before has the aMD method been benchmarked against a long cMD simulation and we commend D.E. Shaw Research for providing their data to the community.

IV.F Supporting Information

IV.F.1 Computational Details

We prepared the simulation of BPTI using the MD software package AMBER11, maintaining a protocol that mimics that of Shaw and coworkers[1]. The structure was selected from the first alternative structure of the joint neutron/X-ray refined structure with PDB ID 5PTI, with deuterium atoms changed to hydrogen atoms[2]. The system was solvated in a cubic box with sides measuring 52 Å containing 6 chloride ions and 4215 water molecules, modeled using the AMBER

ff99SB-ILDN force field[3,4]. The water was modeled with the 4-particle TIP4P-Ew force field[5], which was previously shown to better describe the rotational motion of proteins[6] than the related 3-particle water model, TIP3P[7]. We applied SHAKE to constrain all bonds involving hydrogen atoms[8]. We minimized the resulting structure to remove any clashes. We applied harmonic positional restraints of strength 10 kcal/mol/Å^2 to the protein backbone atoms, kept the pressure at 1 atm and increased the temperature from 10K to 300K as a linear function of time over the course of 1.2ns, using Berendsen temperature and pressure control algorithms with relaxation times of 0.5 picoseconds for both the barostat and the thermostat[9]. We removed the restraints and performed a 6-ns simulation at constant isotropic pressure of 1 atm and temperature of 300K. We used a 10-Å cutoff radius for range-limited interactions, with Particle Mesh Ewald electrostatics for long-range interactions[10]. We also note that in this study we did not modify the omega torsions of the peptide backbone as suggested by Doshi and Hamelberg in order to maintain as closely the parameters used in the Shaw study[11].

The production simulation of BPTI using cMD on the GPU was carried out using NVT conditions. A Langevin thermostat was used to maintain the temperature at 300K with a collision frequency of 2ps. The simulation time step was 2fs and snap shots were saved every 2ps. An average dihedral energy and total potential energy was computed from 50ns of cMD and used as a reference for the aMD simulations. The aMD simulations were carried out using the exact same conditions as described above starting from the equilibrated simulation used for the 50ns cMD simulation.

IV.F.2 Details of aMD

Accelerated MD modifies the energy landscape by adding a boost potential, $\Delta V(r)$, to the original potential energy surface when $V(r)$ is below a pre-defined energy level E [12] , as

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \frac{(E-V(r))^2}{\alpha + (E-V(r))} & V(r) < E \end{cases} \quad \text{(IV.1)}$$

where α modulates the depth and the local roughness of the energy basins on the modified potential. In principle, this approach also allows the correct canonical average of an observable, calculated from configurations sampled on the modified potential energy surface, to be fully recovered from the accelerated MD simulations. In order to simultaneously enhance the sampling of internal and diffusive degrees of freedom a dual boosting approach was employed, based on separate torsional and total boost potentials as [13]

$$V(r) = V_o(r) + V_t(r) \quad \text{(IV.2)}$$

$$V^*(r) = \{V_o(r) + [V_t(r) + \Delta V_t(r)]\} + \Delta V_T(r) \quad \text{(IV.3)}$$

where $V_o(r)$ is the original potential, $V_t(r)$ is the total potential of the torsional terms, $\Delta V_t(r)$ and $\Delta V_T(r)$ are the boost potentials applied to the torsional terms $V_t(r)$ and the total potential energy $V_T(r)$, respectively.

IV.F.3 Selection of boost parameters

Selection of the boost parameters E and α for dihedral boost and the total boost is based on the average dihedral and total potential energy obtained at the end of the 50ns simulation. These values correspond to 619 kcal/mol for the dihedral energy and -49106 kcal/mol for the total potential energy. In our simulation we have a total of 58 protein residues and 18,226 total atoms. Previous dual boost simulations have suggested that the torsional boost parameter, E(dih), should be set equal to the average dihedral energy, V_o(dih), obtained from the cMD simulation plus 4 kcal/mol times the number of solute residues and the alpha(dih) paramter is then set to 1/5 of this value[14,15]. So for our system we set the following for our dihedral boost parameters:

E(dih)=619 kcal mol^{-1} + (4kcal mol^{-1} $residue^{-1}$ * 58 solute residues) = 851 kcal mol^{-1}

Alpha(dih)=(1/5)*(4kcal mol^{-1} $residues^{-1}$ * 58 solute residues) = 46.4kcal mol^{-1}

For the total boost parameter, E(tot), the value should be equal to 0.16 * total number of atoms plus the average total potential energy, V_o(tot), obtained from the cMD simulation and the alpha(tot) should simply be 0.16*total number of atoms[15,16]. So for our system we set the following:

E(tot)= -49106 kcal mol^{-1} + (0.16kcal*mol^{-1} $atom^{-1}$ * 18,226 atoms) = -46190 kcal mol^{-1}

Alpha(tot)= (0.16kcal mol^{-1} $atom^{-1}$ * 18,226 atoms) = 2916 kcal mol^{-1}

Within our input file for running the simulation we then include these four parameters.

IV.F.4 Reweighting

We collect the bias potentials for both total and dihedral boost for every frame and reweight each frame accordingly. In this work several two-dimensional free energy profiles were reweighted and here we describe the procedure as applied to the PCA plots:

A two-dimensional (2D) histogram was constructed with a bin width and height of 1Å. An indicator function, δ_{kij}, denotes whether the ($PC1$, $PC2$) coordinates fall in the bin ($PC1_i$, $PC2_j$) for frame k,

$$\delta_{kij} = \begin{cases} 1 & (PC1, PC2)_k \in (PC1_i, PC2_j), \\ 0 & other\ wise. \end{cases} \quad \text{(IV.4)}$$

The reweighted histogram at bin ($PC1_i$, $PC2_j$) is then given by

$$H_{ij} = \sum_{k=1}^{K} (\delta_{kij} * \exp(\beta \Delta V_k)), \quad \text{(IV.5)}$$

where K is the total number of snapshots, and ΔV_k is the boost potential at frame k. In this work the exponential was approximated using a 10th order Maclaurin series expansion but several other reweighting schemes are also applicable[17]. The histogram of cMD simulations can be obtained using the same equation by setting the boost potential to zero. The PC1-PC2 free energy profile is then determined by

$$W_{ij} = -k_B T \ln H_{ij} + W_o, \quad \text{(IV.6)}$$

where W_o is an arbitrary constant. In this work, W_o was chosen such that the minimum of the free energy profiles was set to zero.

Movie S1. Shows how the simulation explores the χ_1 (C14) vs χ_1 (C38) space as well as the PC1 vs PC2 space throughout time. The scales for the plots are the same as those found in figures 2a and 3a.

Movie S2-S3. Shows the range of motion of BPTI along PC1 (cyan) and PC2 (red).

Movie S4. Shows an average isomap of water occupancy computed from the full trajectory superimposed on top of the simulation. The 4 long-lived waters W122, W111, W112, and W113 identified in the 1ms cMD simulation persist in the same region identified in the crystal structure 5PTI during the aMD simulation. Representing the waters that exchange with different colors highlights the exchange occurring at site W122. VMD was used to perform the analysis and the water selection was based on the same one used for the analysis of the waters in the 1ms cMD simulation.

IV.F.5 Population Analysis

Table IV 2 Conformational species of BPTI classified according to the isomeric state of the C14-C38 disulfide bridge

Conformational States	Exptl pop(%)	1ms cMD pop(%)	500ns aMD pop(%)
M(Grey et al. [18])	~95	34	61
m_{C14}(Grey et al. [18])	~1	50	2.6
m_{C38}(Grey et al. [18])	~4	6	7.9

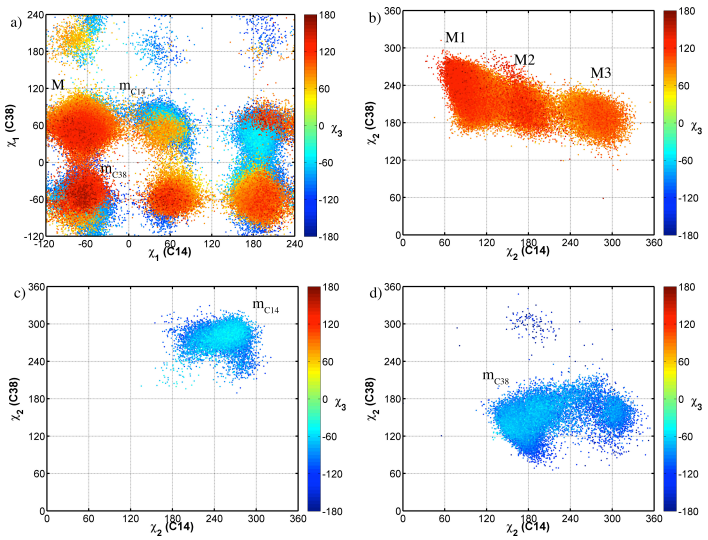

Figure IV 4 (a) Scatter plot of χ_1(C14) vs. χ_1(C38) derived from the 500ns aMD simulation of BPTI. The data are colored according to χ_3 values; the clusters that contain conformational species M, m_{C14}, and m_{C38} are labeled in the plot. (b) Scatter plot of χ_2(C14) vs χ_2(C38) for conformational species M, where M are defined as follows: $-120° < \chi_1(C14) < 0°$; $0° < \chi_1(C38) < 120°$; $0<\chi_3<180°$. The cluster structure observed in this plot serves as a basis for separating the microstates M1, M2, and M3. (c) Scatter plot of χ_2(C14) vs χ_2(C38) for conformational species m_{C14}, where m_{C14} are defined as follows: $0° < \chi_1(C14) < 120°$; $0° < \chi_1(C38) < 120°$; $-180° < \chi_3 < 0°$. (d) Scatter plot of χ_2(C14) vs χ_2(C38) for conformational species m_{C38}, where m_{C38} are defined as follows: $-120° < \chi_1(C14) < 0°$; $-120° < \chi_1(C38) < 0°$; $-180° < \chi_3 < 0°$. For the purpose of plotting, torsional angles χ_1(C14)-χ_1(C38) are defined over the interval from -120° to 240° (0° to 360°). In the main text, the standard range is used for both angles, -180° to 180°. The corresponding plot for the 1ms cMD simulation can be found in the supplementary material of Xue et al. [19].

IV.G Acknowledgement

This research used resources of the Keeneland Computing Facility at the Georgia Institute of Technology, which is supported by the National Science Foundation under Contract OCI-0910735. This work was funded in part by the National Science Foundation through the Scientific Software Innovations Institutes program - NSF SI2-SSE (NSF1047875) and subsequent grants to R.C.W and also by the University of California (UC Lab 09-LR-06-117792) grant to R.C.W. Computer

time was provided by the San Diego Supercomputer Center through National Science Foundation award TGMCB090110 to R.C.W. The work was also supported by a CUDA fellowship to R.C.W. from NVIDIA. The J.A.M group is supported by NSF, NIH, HHMI, NBCR and CTBP.

This chapter contains material from "Routine access to millisecond time scales with accelerated molecular dynamics" (submitted) to *Journal of Chemical Theory and Computation*, authored by Levi C.T. Pierce, Romelia Salomon-Ferrer, Cesar Augusto F. de Oliveira, J. Andrew McCammon**,** and Ross C. Walker. All material has been reproduced with the consent of all other authors.

Article II. IV.H References

(1) Kubelka, J.; Chiu, T. K.; Davies, D. R.; Eaton, W. A.; Hofrichter, J. *J Mol Biol* 2006, *359*, 546.

(2) Schaeffer, R. D.; Fersht, A.; Daggett, V. *Curr Opin Struct Biol* 2008, *18*, 4.

(3) Freddolino, P. L.; Schulten, K. *Biophys J* 2009, *97*, 2338.

(4) Gilson, M. K.; Zhou, H. X. *Annu Rev Biophys Biomol Struct* 2007, *36*, 21.

(5) Lindahl, E.; Sansom, M. S. *Curr Opin Struct Biol* 2008, *18*, 425.

(6) Khalili-Araghi, F.; Gumbart, J.; Wen, P. C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K. *Curr Opin Struct Biol* 2009, *19*, 128.

(7) H., G. *Phys. Rev. E* 1995, *52*.

(8) Lange, O. F.; Schäfer, L. V.; Grubmüller, H. *Journal of Computational Chemistry* 2006, *27*, 1693.

(9) Voter, A. F. *Physical Review Letters* 1997, *78*, 3908.

(10) Voter, A. F. *J Chem Phys* 1997, *106*, 4665.

(11) Bussi, G.; Laio, A.; Parrinello, M. *Physical Review Letters* 2006, *96*, 090601.

(12) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. *Curr Opin Struct Biol* 2010, *20*, 148.

(13) Darve, E.; Pohorille, A. *J Chem Phys* 2001, *115*, 9169.

(14) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. *J Chem Phys* 2008, *128*, 144120.

(15) Henin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. *Journal of Chemical Theory and Computation* 2009, *6*, 35.

(16) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Istv; #225; Kolossv, n.; ry; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. In *Proceedings of the 34th annual international symposium on Computer architecture*; ACM: San Diego, California, USA, 2007, p 1.

(17) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. *Proceedings of the National Academy of Sciences* 2011, *108*, 18684.

(18) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. *Journal of the American Chemical Society* 2011, *133*, 9181.

(19) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. *Journal of Chemical Theory and Computation* 2009, *5*, 1632.

(20) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. *Journal of Computational Chemistry* 2009, *30*, 864.

(21) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J Chem Phys* 2004, *120*, 11919.

(22) Hamelberg, D.; de Oliveira, C. A.; McCammon, J. A. *J Chem Phys* 2007, *127*, 155102.

(23) Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A. *Comput Sci Discov* 2011, *4*.

(24) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B.P. Roberts, B. Wang, S. Hayik,

A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman *University of California, San Francisco* 2010.

(25) Götz, A.W.; Williamson, M.J.; Xu, D.; Poole, D.; Grand, S.L.; Walker, R.C. *Journal of Chemical Theory and Computation,* 2012, in review.

(26) Otting, G.; Liepinsh, E.; Wuethrich, K. *Journal of the American Chemical Society* 1991, *113*, 4363.

(27) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* 1977, *267*, 585.

(28) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* 2010, *330*, 341.

(29) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J Mol Biol* 1984, *180*, 301.

(30) Grant, B. J.; Rodrigues, A. P.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. *Bioinformatics* 2006, *22*, 2695.

(31) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. *Journal of the American Chemical Society* 2012.

(32) Grey, M. J.; Wang, C.; Palmer, A. G., 3rd *Journal of the American Chemical Society* 2003, *125*, 14324.

(33) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. *Journal of the American Chemical Society* 2012, *134*, 2555.

(34) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. *J Biomol NMR* 2011, *50*, 43.

(35) Berndt, K. D.; Beunink, J.; Schroeder, W.; Wuethrich, K. *Biochemistry* 1993, *32*, 4564.

(36) Persson, E.; Halle, B. *Journal of the American Chemical Society* 2008, *130*, 1774.

IV.I Suplemental References

(1) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341.

(2) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J Mol Biol* **1984**, *180*, 301.

(3) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.

(4) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950.

(5) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J Chem Phys* **2004**, *120*, 9665.

(6) Wong, V.; Case, D. A. *J Phys Chem B* **2008**, *112*, 6013.

(7) W. L. Jorgensen, J. C., J. D., Madura, R. W. Impey, M. L. Klein *AIP* **1983**, *79*, 10.

(8) Lippert, R. A.; Bowers, K. J.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Shaw, D. E. *J Chem Phys* **2007**, *126*, 046101.

(9) al., H. J. C. B. e. *J. Chem. Phys.* **1984**, *81*, 7.

(10) Darden, T., D. York, and L. Pedersen *AIP* **1993**, *98*, 4.

(11) Doshi, U.; Hamelberg, D. *J Phys Chem B* **2009**, *113*, 16590.

(12) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J Chem Phys* **2004**, *120*, 11919.

(13) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. *J Chem Phys* **2007**, *127*, 155102.

(14) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. *PLoS Comput Biol* **2009**, *5*, e1000325.

(15) de Oliveira, C. s. A. F.; Grant, B. J.; Zhou, M.; McCammon, J. A. *PLoS Comput Biol* **2011**, *7*, e1002178.

(16) Hamelberg, D.; de Oliveira, C. A.; McCammon, J. A. *J Chem Phys* **2007**, *127*, 155102.

(17) Shen, T.; Hamelberg, D. *J Chem Phys* **2008**, *129*, 034103.

(18) Grey, M. J.; Wang, C.; Palmer, A. G., 3rd *Journal of the American Chemical Society* **2003**, *125*, 14324.

(19) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. *Journal of the American Chemical Society* **2012**.