Understanding the mechanisms of enzymatic catalysis requires a detailed understanding of the complex interplay of structure and dynamics of large systems that is a challenge for both experimental and computational approaches. QM/MM methods have been extensively used to study these reactions, but the difficulties arising from the hybrid treatment of the system are well documented. More importantly, the computational demands of QM/MM simulations mean that the dynamics of the reaction can only be considered on a timescale of nanoseconds even though the conformational changes needed to react the catalytically active state happen on a much slower timescale. Here we demonstrate an alternative approach that uses transition state force fields (TSFFs) derived by the quantum-guided molecular mechanics (Q2MM) method that provides a consistent treatment of the entire system at the classical molecular mechanics level and allows simulations at the microsecond timescale. Application of this approach the second hydride transfer transition state of HMG-CoA reductase from Pseudomonas mevalonii (PmHMGR) identified three remote residues, R396 E399 and L407, (15-27 Å away from the active site) that have a remote dynamic effect on enzyme activity. The predictions were subsequently validated experimentally via site-directed mutagenesis. These results show that microsecond timescale MD simulations of transition states are possible and can predict rather than just rationalize remote allosteric residues.
Microsecond Timescale Simulations at the Transition State of PmHMGR Predict Remote Allosteric Residues

Authors: Taylor R. Quinn¹, Calvin N. Steussy², Brandon E. Haines³, Jinping Lei⁴, Wei Wang⁴, Fu Kit Sheong⁴, Cynthia V. Stauffacher², Xuhui Huang⁴, Per-Ola Norrby⁵,⁶, Paul Helquist¹, Olaf Wiest¹⁶*

Affiliations:
¹Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA.
²Department of Biological Sciences and Purdue Center for Cancer Research, Purdue University, West Lafayette, IN 47907, USA.
³Department of Chemistry, Westmont College, Santa Barbara, CA 93108, USA.
⁴Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.
⁵Data Science and Modelling, Pharmaceutical Sciences, R&D, AstraZeneca, Gothenburg, Pepparedsleden 1, SE-431 83 Mölndal, Sweden.
⁶Lab of Computational Chemistry and Drug Design, School of Chemical Biology and Biotechnology, Peking University, Shenzhen Graduate School, Shenzhen, China.
*Correspondence to: owiest@nd.edu.

Understanding the mechanisms of enzymatic catalysis requires a detailed understanding of the complex interplay of structure and dynamics of large systems that is a challenge for both experimental and computational approaches.¹ QM/MM methods have been extensively used to study these reactions, but the difficulties arising from the hybrid treatment of the system are well documented.²,³ More importantly, the computational demands of QM/MM simulations mean that the dynamics of the reaction can only be considered on a timescale of nanoseconds even though the conformational changes needed to react the catalytically active state happen on a much slower timescale.⁴ Here we demonstrate an alternative approach that uses transition state force fields (TSFFs) derived by the quantum-guided molecular mechanics (Q2MM) method⁵ that provides a consistent treatment of the entire system at the classical molecular mechanics level and allows simulations at the microsecond timescale. Application of this approach to the second hydride transfer transition state of HMG-CoA reductase from Pseudomonas mevalonii (PmHMGR) identified three remote residues, R396 E399 and L407, (15-27 Å away from the active site) that have a remote dynamic effect on enzyme activity. The predictions were subsequently validated experimentally via site-directed mutagenesis. These results show that microsecond timescale MD simulations of transition states are possible and can predict rather than just rationalize remote allosteric residues.

A typical study of an enzymatic reaction mechanism will start from a crystal structure and introduce an initial guess structure of the transition state, which is then optimized using QM/MM, combining treatment of the active site using a computationally affordable electronic structure method with a classical force field treatment of the remainder of the system. The surrounding protein environment is in a local minimum corresponding to the initial crystal structure that may or may not be similar to the conformation of the protein at the transition state. Sufficient conformational sampling is thus needed to ensure that the enzyme is in a conformation competent for catalysis. Depending on the movements involved to adopt the catalytically active conformation, this requires timescales from nanoseconds (for side chain movements) to microseconds (for loop and helix motions). This problem, together with the well-known problems resulting from the
boundary region that needs to be treated simultaneously by a QM and a MM method, is well appreciated\textsuperscript{2,3} and different QM/MM-MD approaches have been developed,\textsuperscript{6} but they are typically iterative approximations to the problem that are limited to protein dynamics on the low nsec timescale.\textsuperscript{7} Although protein dynamics at the transition state is not commonly a direct factor in enzyme catalysis,\textsuperscript{8} the equilibrium structure of a protein (1), often elucidated by x-ray crystallography, has to adopt a reactive conformation that enables catalysis (2), as shown in Figure 1. It should be noted that this is not a single conformation, but rather an ensemble along coordinates orthogonal (blue) to the reaction coordinate (red). In the presence of the transition state of a reaction that in typical computational studies of enzyme catalysis is introduced into the enzyme structure in conformation (1) corresponds to a higher energy conformation (3) that has to undergo a conformational change to the reactive conformation able to stabilize the transition state (4) by both electrostatic and van der Waals forces. Adequate treatment of the long- and short-range electrostatics requires both sampling of the conformational space that maximizes interaction with the substrates in the active site, and short-range electrostatics require capturing how small distortions in the transition state affect the energetics of the reaction. These conformational changes can involve side chain or larger conformational changes on the \textmu s timescale that are not accessible by QM/MM-MD approaches.

\textbf{Figure 1.} Schematic view of computational studies of enzyme catalysis. An initial guess for the transition state is introduced in a crystal structure of an enzyme (1), leading to a high-energy structure (3) that is optimized to (4), which is assumed to represent the reactive complex (2). Conformational changes from (3) to (4) are likely to occur on the \textmu s timescale.

This problem is widely recognized and a number of methods have been developed to approximate the transition state in a more efficient manner. Transition state “mimics”, systems in a pseudo-intermediate state,\textsuperscript{9} have been used to approximate the reactant state at a cohesive MM level that allows for long simulations, but these mimics capture distortions in the active site only at a very approximate level. The empirical valence bond (EVB) method\textsuperscript{10} and the related ACE
method have been very successful in studying catalysis in enzymes and small molecules, respectively, through the parameterization of mixed ground state force fields. Both methods assume that the transition state (TS) is a weighted average of ground states. In particular for charges, this is not always the case as can be seen by considering charge distribution in many transition states that are more polar than either the reactant or the product of a reaction.

The use of transition state force fields (TSFFs) is a promising alternative to QM/MM methods because they treat the entire system at a consistent level of theory and allow long time-scale MD simulations. TSFFs have been shown to be highly accurate compared to high-level DFT calculations and experimental data for a wide range of small-molecule reactions, but have only been sparingly used for the study of enzyme reactivity and never for the study of enzyme dynamics or mechanism. Conceptually, this approach is related to the EVB method. The key difference is that rather than using a mix of the reactant and product ground state force field (FF) and adjusting the parameters using empirical information to represent the transition state, the TSFF approach reparameterizes a FF at the transition state using data from electronic structure methods. Because this includes geometric and electronic features that might not be represented well by either the reactant or the product, the resulting TSFF is expected to be more accurate. The calculation of the [2x2] interaction matrix is not necessary, making TSFFs as fast as a traditional FF, and truly predictive in that it does not use any experimental information in the parameterization. The increased speed of the classical FF then allows long simulation times to allow the complete protein to better sample the protein in a reactive configuration as discussed above.

It should be noted that TSFFs use different energy functions for the starting material and the transition state and are therefore not suitable for the calculation of absolute activation energies. Rather, they focus on the key question of how the structure of the protein changes from the non-reactive crystal structure to the reactive conformation to catalyze a reaction, e.g. by changing the direction and magnitude of dipoles that stabilize active site interactions as well as longer-range interactions.

We developed the Quantum-Guided Molecular Mechanics (Q2MM) method for the automated fitting of TSFFs to high-level electronic structure calculations. The general approach used by Q2MM is shown in Figure 2A, where the parameters of a small number of atoms representing the active site of the enzyme are fitted to the reference data from electronic structure calculations by minimizing the penalty function $\chi^2$. In addition to geometric data such as bond lengths angles or dihedrals, Q2MM also fits to the QM Hessian elements, with special treatment of the reaction coordinate to give it a positive curvature in the TSFF, to properly account for energy costs of small distortions in the active site. Because only a few structures are calculated for the training set, larger active site models than in a typical QM/MM study can be treated at a higher level of theory. The TSFF, in combination with standard force field parameters for the remainder of the protein, is then used in MD simulations to study the crucial question of how the enzyme responds to the presence of the transition state and thus catalyzes the reaction.
Figure 2. A) Mechanism of PmHMGR. Green is the GS, blue is the INT2, and TS2 is described by the arrows in INT2. B) Residues included in the parameterization of the TS2 TSFF. C) The optimized TSFF (green carbons) is overlaid with the QM optimized structure (magenta).

To demonstrate this novel approach to enzymatic reaction mechanisms, we studied HMG-CoA reductase from *Pseudomonas mevalonii* (PmHMGR) which uses two equivalents of NADH to convert HMG-CoA to mevalonate via sequential hydride transfer steps in a single active site.²⁰ This obligate homodimer is the point of feedback control for polyisoprenoid biosynthesis²¹ and its human homolog is the target of the widely used statin drugs. More importantly, it has an extraordinarily complex reaction mechanism (Fig. S1) that despite decades of study continues to provide novel insights into enzyme catalysis.²²,²³ Based on our previous QM/MM study of PmHMGR at the ONIOM-(B3LYP/6-31g(d,p):AMBER) level of theory,²² a TSFF was created for the second hydride transfer from the NADH to mevaldehyde by Q2MM.⁵ Figure 2B shows the active site model for which the FF parameters were fitted to the results from the Q2M calculations. Using the functional form of the AMBER99SB force field,²⁴ the relevant atoms and interactions in the active site were reparameterized for the TS (Fig. S3) while the standard parameters for the remainder of the enzyme were kept. The Q2MM code, which is publically available,²⁵ was interfaced with a modified version of the *nmode* module in Amber 8²⁶ to obtain the MM Hessian values for comparison to the QM reference Hessian values generated from a frequency calculation in G09.²⁷ For optimization of the Q2MM penalty function, literature weight factors for bonds (100 Å⁻¹), angles (2 degree⁻¹) and dihedrals (1 degree⁻¹) corresponding to the inverse tolerance of each type of data were used.²⁸ Literature weight factors for the Hessian elements were converted to units if kcal mol Å² to correspond to Amber force field units. Those weights are 0.0 for self-interactions, 0.31 for 1-4 interactions and 0.031 for all others. The resulting TSFF shows excellent agreement between the active site structure from the QM/MM²² and TSFF calculations (Fig. 2C).
The TSFF was used to perform 10 µs of adaptive sampling followed by 3-5 µs of MD simulation in an NVT ensemble. For comparison, the ternary complex of the starting material, NADH, and PmHMGR (constructed from non-productive ternary complex pdb code 1QAX, 2.8 Å resolution) as well as the intermediate immediately preceding the hydride transfer (Fig. 1A, GS and INT2) were also calculated using standard AMBER99SB parameters. The root mean squared deviation (RMSD) per residue was extracted from the trajectories for each state and the RMSF difference between the ground state and the transition state was color-coded onto the structure of the protein (Fig. 4A).

**Figure 3.** Contribution (tICA coefficient) of the selected residues for the slowest dynamics of the system in the ground state (GS, black) and transition state (TS2, red). The contributions of T374, R396, E399 and L407 are 0.009, 0.014, 0.010 and 0.015, respectively, in the GS state, while they are 0.008, 0.018, 0.015 and 0.019, respectively, in TS2. The big differences between the ground state and transition state observed in HMG-CoA and NADH are coming from those selected heavy atoms around the transferring hydrogen that were treated by TSFF in the TS2 state. The “HMG-CoA” and “NADH” labels in the x-axis denote the heavy atoms of the substrate and cofactor respectively. “small/large domain” denotes the Cα atoms of the residues that contact with HMG-CoA/NADH in the small and large domain, and “flap domain” denotes the Cα atoms of the residues (374 to 428) in the hinge region and flap domain.

The biggest differences in RMSD between the GS and TS2 are observed in the second α-helix on the flap domain, which is significantly more flexible in the ground state despite being between 15 at 27 Å away from the active site. To demonstrate that transition state MD simulations of PmHMGR need to be at microsecond timescale in order to capture the allosteric role of remote residues on enzymatic functions, we performed time-lagged Independent Component Analysis (tICA) analysis based on MD simulations to identify residues that contribute the most to the slowest dynamics of the system in the ground state and transition state, respectively. tICA is a variation of the linear variational approach that transforms the input coordinates (such as Euclidean distances, torsion angles) into collective variables (time structure-based independent components, tICs) sorted by “slowness”.

In agreement with the RMSF analysis, we found that the flap domain
is more flexible in the ground state than in the transition state. The largest differences in the per-residue contributions (Fig. 3, Fig. S4-6) between the ground state and transition state are observed in the flap domain and substrates HMG-CoA and NADH. For example, the remote residues R396, E399 and L407 have significantly larger tICA values in the transition state than in the ground state, indicating that they are important to the allosteric effect during PmHMGR’s enzymatic catalysis. This is noteworthy because the flap domain has been postulated to be involved in catalysis, but there has been no previous suggestion of allostery of remote residues. In comparison, T374 is located at a similar distance on the flap domain but does not show a discernible difference in the tICA value between the ground- and transition state.

To test the hypothesis that the changes in the flexibility of the remote residues between ground- and transition states have a functional role in enzyme catalysis, the four residues in second α-helix on the flap domain discussed above were chosen for experimental mutagenesis studied based on their decreased RMSF in TS2, their contributions in the tICA analysis, and their representative positions in the flap domain. As shown in Figure 4 A,B, L407 is the center of a solvent-exposed hydrophobic patch of residues on the last two α-helices approximately 15 Å away from the active site. It was mutated to a serine in order to disrupt the electronics in the area while keeping the surface area of the residue similar. R396 is located on the first loop of the flap domain, ~22 Å away from the active site and uses the side chain to hydrogen bond to residues on the last two helices. It was replaced by an alanine residue. E399 is positioned at the end of the second helix, ~27 Å away from the active site. It is an excellent test case for the computational predictions because it is not engaged in any significant non-bonded interactions in the crystal structure but exhibited very high flexibility in the GS compared to the TS and has a large contribution in the tICA. As a negative control, we studied the alanine mutant of T374, a residue that showed virtually no difference in the tICA values between the GS and TS2 states.

**Figure 4.** A) RMSD difference between the ground state and the transition state. Yellow/red indicates areas where the RMSD of the GS over the trajectory is larger than in TS2 B) Location of mutated residues highlighted in green residues and neighboring interactions. C) Results of the
kinetics with respect to the wild type (WT) D) Partial sequence alignment of four Class II HMGRs. Sequences were aligned using T-COFFEE. Conserved glutamate highlighted in blue, other conserved residues are highlighted in red.

The effect of the four point mutations on the relative maximum rate of the conversion of mevalonate, CoA and two equivalents of NAD$^+$ to HMG-CoA were determined. In agreement with the simulations, the activity of the L407S decreased by 57% (Fig 4C) compared to the wild type. The R396A mutant experiences little change, possibly because the flexible loop region is more tolerant of mutations. The E399A mutant experienced the greatest decline in activity with 69%, despite being the furthest from the active site and, based on the crystal structure, engaging only in interactions with the solvent. These results cannot be predicted based on the crystal structure or rationalized using short-timescale simulations but are in line with the results from long-timescale MD simulations at the transition state. They also rationalize the surprising observation that this glutamate residue is conserved in all Class II HMGRs (Fig. 3D), which is hard to explain based on the available crystal structures. As predicted by the RMSF and tICA analysis of the trajectories, the T374A mutant is virtually indistinguishable from the wild type despite also being part of the flap domain and at a similar distance from the active site.

This work demonstrates that TSFFs allow conformational sampling at the transitions state of enzyme catalyzed reactions on the μsec timescale, 2-3 orders of magnitude longer than what can be reached using traditional QM/MM methods. Computational enzymology in this time regime provides experimentally verifiable predictions on properties such as dynamic allosteric effects on catalysis, a topic of intense interest in biophysics and drug design rather than the more typical computational rationalization of experimental observations. Given the role of dynamics on enzyme catalysis, e.g. by slow conformational changes to reach the reactive conformation, we expect that HMGR is a typical case rather than an exception in requiring long-timescale simulations at the transition state. The general availability of the Q2MM code will allow future applications of TSFFs to computational enzymology on the μsec timescale.

Acknowledgments: We thank Marcus Arieno for preliminary simulations and Moumita Sen for the T374A mutant data. This work was supported by the National Institutes of Health (1R01GM111645 and T32GM075762) and the Hong Kong Research Grant Council (AoE/P-705/16).

Author contributions: T.R.Q., J.L., W.W. and F.K.S. ran the simulations, T.R.Q. and C.N.S performed the mutagenesis studies, B.E.H. and P.O.N. generated the TSFF and C.V.S., P.H. and OW conceived and designed the studies. All authors analyzed data and contributed to the manuscript.

Competing interests: The authors declare no competing interests.

References
1 Bottaro, S. & Lindorff-Larsen, K. Biophysical experimentalists and biomolecular simulations: A perfect match. Science 361, 355-360 (2018).
2 Senn, H. M. & Thiel, W. QM/MM Methods for Biomolecular Systems. Angew, Chem. Intl. Ed. 48, 1198-1229 (2009).
3 Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). Angew. Chem. Intl. Ed. 53, 10020–10031 (2014).
Henzler-Wildman, K. A. et al. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-916 (2007).

Rosales, A. R. et al. Application of Q2MM to Predictions in Stereoselective Synthesis. *Chem. Comm.* **54**, 8294-8311 (2018).

Hu, H. & Yang, W. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **59**, 573-601 (2008).

Lu, X., Ovchinnikov, V., Demapan, D., Roston, D. & Cui, Q. Regulation and Plasticity of Catalysis in Enzymes: Insights from Analysis of Mechanochemical Coupling in Myosin. *Biochemistry* **56**, 1482-1497 (2017).

Jindal, G. & Warshel, A. Misunderstanding the preorganization concept can lead to confusions about the origin of enzyme catalysis. *Proteins Struct. Funct. Bioinformatics* **85**, 2157-2161 (2017).

van der Kamp, M. W. et al. Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nature Comm.* **9**, 1177 (2018).

Åqvist, J. & Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* **93**, 2523-2544 (1993).

Weill, N., Corbeil, C. R., De Schutter, J. W. & Moitessier, N. Toward a computational tool predicting the stereochemical outcome of asymmetric reactions: Development of the molecular mechanics - based program ACE and application to asymmetric epoxidation reactions. *J. Comp. Chem.* **32**, 2878-2889 (2011).

Patrascu, M. B. P., J.; Pinus, S.; Bezanson, M.; Norrby, P. O.; Moitessier, N. From Desktop to Benchtop – A Paradigm Shift in Asymmetric Synthesis. *ChemRxiv* https://doi.org/10.26434/chemrxiv.9758558.v1 (2019).

Wong, K. F., Selzer, T., Benkovic, S. J. & Hammes-Schiffer, S. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proc. Nat. Acad. Sci.* **102**, 6807-6812 (2005).

Lee, M., Bai, C., Feliks, M., Alhadeff, R. & Warshel, A. On the control of the proton current in the voltage-gated proton channel Hv1. *Proc. Nat. Acad. Sci.* **115**, 10321-10326 (2018).

Jensen, F. & Norrby, P.-O. Transition states from empirical force fields. *Theor. Chem. Acc.* **109**, 1-7 (2003).

Hansen, E. R., A.R.; Tutkowski, B.M.; Norrby, P.O.; Wiest, O. Prediction of Stereochemistry using Q2MM. *Acc. Chem. Res.* **49**, 996-1005 (2016).

Ryberg, P. H., S. M.; Kongsted, J.; Norrby, P. O.; Olsen, L.; Ryde, U. . Transition-state docking of flunitrazepam and progesterone in cytochrome P450. *J. Chem. Theory Comput.* **4**, 673-681 (2008).

Ryberg, P. O., L.; Norrby, P. O.; Ryde, U. General transition-state force field for cytochrome p450 Hydroxylation. . *J. Chem. Theory Comput.* **3**, 1765-1773 (2007).

Rosales, A. R. et al. CatVS: Virtual screening of asymmetric catalysts. *Nature Catalysis* **2**, 41-45 (2018).

Lawrence, C. M., Rodwell, V. W. & Stauffacher, C. V. Crystal structure of Pseudomonas mevalonii HMG-CoA reductase at 3.0 angstrom resolution. *Science* **268**, 1758-1762 (1995).

Goldstein, J. L. & Brown, M. S. Regulation of the mevalonate pathway. *Nature* **343**, 425-430 (1990).
Haines, B. E., Steussy, C. N., Stauffacher, C. V. & Wiest, O. Molecular Modeling of the Reaction Pathway and Hydride Transfer Reactions of HMG-CoA Reductase. *Biochemistry* **51**, 7983-7995 (2012).

Haines, B. E., Wiest, O. & Stauffacher, C. V. The Increasingly Complex Mechanism of HMG-CoA Reductase. *Acc. Chem. Res.* **46**, 2416-2426 (2013).

Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinformatics* **65**, 712-725 (2006).

Wilding, E. I. *et al.* Essentaility, Expression, and Characterization of the Class II 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase of *Staphylococcus aureus*. *J. Bact.* **182**, 5147-5152 (2000).

Case, D. A. D., T.A.; Cheatham, III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R. E.; Luo, R. M., K.M.; Wang, B. Pearlman, D.A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; & Mongan, J. H., V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J.W.; Ross, W.S.; Kollman, P.A. . Amber 8. *University of California, San Francisco* (2004).

Gaussian 09, Revision A1 (Wallingford, CT, 2009).

Maple, J. R. *et al.* Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comp. Chem.* **15**, 162-182 (1994).

Sheong, F. K., Silva, D.-A., Meng, L., Zhao, Y. & Huang, X. Automatic state partitioning for multibody systems (APM): an efficient algorithm for constructing Markov state models to elucidate conformational dynamics of multibody systems. *J. Chem. Theor. Comp.* **11**, 17-27 (2014).

Tabernero, L., Bochar, D. A., Rodwell, V. W. & Stauffacher, C. V. Substrate-induced closure of the flap domain in the ternary complex structures provides insights into the mechanism of catalysis by 3-hydroxy-3-methylglutaryl–CoA reductase. *Proc. Nat. Acad. Sci.* **96**, 7167-7171 (1999).

Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135-144 (2014).

Hedl, M., Tabernero, L., Stauffacher, C. V. & Rodwell, V. W. Class II 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductases. *J. Bacteriol.* **186**, 1927–1932 (2004).

Goodey, N. M. & Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nature Chem. Bio.* **4**, 474 (2008).
Supplementary Information for

Microsecond Timescale Simulations at the Transition State of *Pm*HMGR

Predict Remote Allosteric Residues

Taylor R. Quinn, Calvin N. Steussy, Brandon E. Haines, Jinping Lei, Wei Wang, Fu Kit Sheong, Cynthia V. Stauffacher, Xuhui Huang, Per-Ola Norrby, Paul Helquist, Olaf Wiest*

Correspondence to: owiest@nd.edu

This PDF file includes:

Supplementary Information
Figs. S1 to S6

Other Supplementary Materials for this manuscript include the following:

Movies S1: HMGR with color-coded RMSF and residues chosen for mutations
Data S1: Pymol Session with color-coded RMSF
Data S2: AMBER-formatted parameter file for TS2

Contents

1. Parameterization of TS2 S2
2. Molecular Dynamics S4
3. Time-lagged Independent Component Analysis (tICA)
   3.1. Adaptive Sampling Simulations S5
   3.2. Selecting Kinetically Slow Variables S5
   3.3. Projecting the MD samplings to the ground state energy landscape S6
   3.4. Identifying the important remote residues for the reductive reaction by tICA S7
4. Cloning Expression, Purification and Kinetics of HMGR Mutants S8
5. References
1. Parameterization of the TSFF

The QM reference data used to parameterize the TSFF for the second hydride transfer of PmHMGR are generated from a QM/MM calculation of a truncated enzyme active site model (“Theozyme” Figure S1). The truncated QM/MM models of the enzyme were built from the full enzyme transition state structures previously determined\(^1\) by deleting all residues beyond 3.0 Å of the active site followed by re-optimization to the TS at the ONIOM(M06/6-31g(d,p):AMBER) level of theory as implemented in the G09 suite of programs.\(^2\) The TSFF was developed for atoms in the active site (Figure S2) using Q2MM\(^3\) in analogy to the procedure described for the development of small-molecule TSFFs.\(^4\)

The QM region of the QM/MM calculations is comprised of the nicotinamide and ribose rings of NAD, the thiol group of CoA, mevaldehyde up to the terminal carboxylate group, and the side chains up to C\(\alpha\) of catalytically relevant residues E83, K267, D283, and H381. \((\text{Figure S1})\)\(^1\) Tao’s iterative method for updating the partial MM charges of the atoms in the QM region\(^5\) was used for the ONIOM mechanical embedding scheme.\(^6\) Shrinking the model size from >12,000 atoms to ~1,300 atoms makes the Hessian calculation required for the Q2MM method manageable while maintaining the protein environment around the TS.

*Figure S1.* The ONIOM model of the second hydride transfer used as reference data for the parameterization of the TSFF. The QM region is depicted with the ball and stick representation, while the MM region is depicted in the line representation with hydrogens omitted for clarity.

The TSFF was developed using Q2MM, the most recent and extensively revised version of which can be found on github.com/q2mm. It was interfaced with a modified version of the nmode module in Amber 8 to obtain the MM Hessian values for comparison to the QM reference Hessian values generated from a frequency calculation in G09. For optimization of the Q2MM objective function, literature weight factors for bonds (100 Å\(^{-1}\)), angles (2 degree\(^{-1}\)) and dihedrals (1 degree\(^{-1}\)) were implemented corresponding to the inverse tolerance of each type of data.\(^7\) Literature weight factors for the Hessian elements\(^7\) were converted to units kcal\(^{-1}\) mol Å\(^2\) to correspond with the units of the force constants in the AMBER99SB force field. They are 0.0 for self-interactions, 0.31 for 1-4 interactions and 0.031 for all others.
New atom types were created for the hydride acceptor carbon (AC), hydride (HT), hydride donor carbon (DC), the nicotinamide ring because of hybridization changes during the reaction and for Glu83 because of its proposed role of stabilizing the TS. (Figure S2) Overall, this creates 150 parameters that need to be reparametrized in the TSFF. The initial values for force constant parameters were set by analogy to existing values wherever possible, and the initial equilibrium value parameters were set to the value from the QM/MM calculations. The remainder of atoms in the substrates, NAD, and protein residues were assigned atom types from GAFF8, published parameters,8,9 and AMBER99SB,10 respectively. The X-AC-HT-X and X-HT-DC-X dihedrals were excluded from the parameterization automatically by zeroing the force constant because they contain angles larger than 150°. Partial charges for the QM reference structure were generated by fitting them to the HF/6-31g* electrostatic potential using the RESP protocol in the AMBER antechamber program and were held fixed throughout the parameterization of the other parameters. The use of implicit solvation in the MM calculations was necessary to avoid unphysical equilibrium values. Large portions of the QM reference structure do not need to be parameterized. The corresponding atomic coordinates are frozen and the associated data points are removed from the penalty function by zeroing out the relevant weight factors. Freezing the peripheral atoms is particularly important for the QM/MM model where the MM region consists of highly fragmented polypeptide chains.

Equilibrium value parameters were carefully monitored to prevent them from deviating from the QM values beyond thresholds of ±0.05 Å for bonds and ±5° for angles. Various tethering functions were used over several rounds of optimizing the force constant and equilibrium values to avoid unphysical local minima. The final TSFF was then generated by gradually releasing the tethering functions and is reported in the Supporting Information. The final optimization was performed by tightening the convergence criterion to 0.001%. As expected based on the positive curvature added to the reaction coordinate in the Hessian matrix, the bond force constants for the forming (AC-HT) and breaking (DC-HT) bonds are very large. Optimization of the TS using the TSFF in Amber shows excellent agreement (Figure 2B) with the QM/MM TS it was fit to reproduce.
2. Molecular Dynamics

The ground state (GS) and intermediate (INT2) states involved in the overall mechanism of the conversion of HMGCoA by two equivalents of NADH, catalyzed by HMGR (Figure S3) were built manually from the non-productive ternary complex HMG-CoA/NAD+ (pdb code 1QAX). The transition state (TS2) was built via the methods described below. The AMBER99SB force field was used for the GS and INT2 systems and for TS2 outside of the TSFF-parameterized active site.

Molecular dynamics (MD) simulations were run primarily using the pmemd module in Amber16. A time step of 1 fs was used to integrate the equations of motion for all transition state simulations and 2 fs for the ground state and intermediate state systems. The systems were minimized in 1000 step increments that gradually reduced restraints on the atoms, then were heated to 300K over 30 ps followed by a full temperature equilibration for 10 ps. NPT equilibration occurred over 2 ns before running initial adaptive sampling simulations of each state for 1 µs.

Systems were setup with the Particle Mesh Ewald method for treating long-range electrostatic interactions, a 10 Å cutoff for nonbonded van der Waals interactions and periodic boundary conditions. The SHAKE algorithm was used to constrain the hydrogen atom bonds except for hydrogens in residues directly related to the transition state. The temperature was maintained using the Langevin Thermostat and pressure in the NPT maintained using a Berendsen barostat. The TS simulations were run in independent 100 ns trajectories and the GS and INT2 were run in 200 ns trajectories. Analysis was done using the cpptraj module in Amber16.
3. Time-lagged Independent Component Analysis (tICA)

3.1. Adaptive Sampling Simulations

To adequately sample the conformational space of *Pseudomonas mevalonii* 3-hydroxyl-3-methylglutaryl coenzyme A reductase (*Pm*HMGR), we adopted an adaptive sampling strategy, selecting starting structures from the above short unbiased MD trajectories. Starting structures for the adaptive sampling were chosen from the unbiased MD trajectories above by a K-centers clustering algorithm. The root-mean-square fluctuation (RMSF) of the Cα atoms of the hinge region (residues 374-377) and flap domain (residues 377-428) was used as the distance function for the clustering, and the conformation that closest to the cluster center was selected from each cluster and used as the starting structures. Then, the selected structures were re-solvated into 124,000 explicit TIP3P water molecules using a truncated octahedron box with a box length of 180 Å through ADDTOBOX program in AMBER14. Subsequent restrained equilibration simulations and unrestrained production MD simulations were carried out following the above MD simulation protocol. Finally, a total of 55, 35, and 42 200-ns production MD simulations were collected for the data analysis of the ground state (GS), intermediate state (INT2) and transition state (TS2), respectively.

3.2. Selecting Kinetically Slow Variables

The time-lagged Independent Component Analysis (tICA) is an efficient approach to elucidate the slowest motions for protein dynamics. The optimal slow tICs can be scaled to define a kinetic distance metric subspace, on which the fast and slowly mixed states can be separated and the geometric analysis can be used to unravel rare event transitions. Thus, tICA has been widely used to reduce the dimension of MD simulation data, and the optimal slow tICs can describe the slowest dynamics underline functional conformational changes of interest.

Approximately 26,500 pairwise distances between the following pairs of atoms were selected as the input features for tICA:

a) heavy atoms of substrate and cofactor - Cα atoms of hinge region and flap domain
b) heavy atoms of substrate and cofactor - Cα atoms of small domain and large domain residues that contact with substrate and cofactor in the crystal structure (PDBID:1QAX)
c) Cα atoms of hinge region and flap domain - Cα atoms of hinge region and flap domain
d) Cα atoms of hinge region and flap domain - Cα atoms of small domain and large domain residues that contact with substrate and cofactor in the crystal structure (PDBID:1QAX)

When *Pm*HMGR catalyzes the reduction of HMGCoA, the hinge region and flap domain were suggested to undergo conformational changes to facilitate the cofactor (NADH) exchange. In addition, protein residues in contact with substrate HMG-CoA and cofactor (e.g. residue 614-719 in the small domain, residue 1-108 and 220-375 in the large domain may also be important for the reduction reaction. To elucidate the slowest dynamics of the system and identify remote allosteric residues, we chose all relevant pairwise distance combinations as listed in item a) to d) above. Figure S4 shows the implied timescales resulting from the tICA for GS, INT2 and TS2, demonstrating that the relevant movements occur on a timescale of hundreds of ns and are therefore not accessible through shorter timescale MD simulations.
Figure S4. The implied timescales (ITS) for tICA in the ground state (GS, blue), intermediate state (INT2, green) and transition state (TS2, red) respectively. The slowest dynamic (the top line) of the system can be obtained by tICA but not from shorter MD simulations.

As shown in Figure S5, the projections of free energy landscape on the two slowest tICs (of the GS ensemble) yields smooth landscape and clearly indicate the difference of conformational dynamics among GS, INT2 and TS2. We note that we chose a relatively large number of pairwise distances in the tICA analysis that may induce statistical noise. To address this issue, we further validated our model by only selecting a subset of distance pairs that mostly contribute to the difference of conformational dynamics between GS, INT2 and TS2 based on our chemical intuition. In particular, we chose a subset of only approximately 2700 distance pairs to perform the tICA analysis by including the following pairs of atoms:

e) heavy atoms around the transferring hydrogen in HMG-CoA/NADH in the TS2 state - Cα atoms of hinge region and flap domain

c) Cα atoms of hinge region and flap domain - Cα atoms of hinge region and flap domain

As shown in Figure S6b, the results from this new model (~2700 distances) are consistent with our original model (~26,500 distances), and yield the same set of residues that have significant contributions to the conformational dynamics.

3.3. Projecting the MD samplings to the ground state energy landscape

To compare the conformational space in the GS, INT2, and TS2 states, we performed the following free energy landscape projections based on our tICA: (1) projecting the MD conformations of GS onto its own slowest two tICs; (2) projecting the TSFF samplings onto the slowest two tICs of GS state; (3) projecting the MD conformations of INT2 onto the slowest two tICs of GS state (Figure S5).
Figure S5. The projection of the TSFF MD samplings (TS2, red), the MD conformations of the intermediate state (INT2, green) and ground state (GS, blue) onto the slowest two tICs of the ground state.

3.4. Identifying the important relevant remote residues by tICA

We employed tICA to identify the important remote residues for the reductive reaction catalyzed by PmHMGR. In tICA, each component (tIC) is defined by the linear combinations of the input features (pairwise distance in our work), so the feature with maximum coefficient should have maximum contribution to this tIC. We calculated the contribution of each remote residue for the slowest dynamics in the GS, INT2, and TS2 states respectively by summing the corresponding normalized tICA coefficient values of the atom pairs which include the selected atoms of this residue.
**Figure S6.** Contributions (tICA coefficient) of the selected residues for the slowest dynamics of the system in the ground state (GS, blue), intermediate state (INT2, green) and transition state (TS2, red). (a) The contributions of R396 (1), E399 (2) and L407 (3) are: 0.014, 0.010 and 0.015 respectively in the GS state, 0.021, 0.011 and 0.014 in the INT2 state, while they are 0.018, 0.015 and 0.019 respectively in the TS2 state. The large differences between the ground state and transition state that observed in HMG-CoA and NADH originate from those selected heavy atoms around the transferring hydrogen that were treated by TSFF in the TS2 state. The “HMG-CoA” and “NADH” labels in the x-axis denote the heavy atoms of the substrate and cofactor respectively, “small/large domain” denotes the Cα atoms of the residues that contact with HMG-CoA/NADH in the small and large domain, and “flap domain” denotes the Cα atoms of the residues (374 to 428) in the hinge region and flap domain. (b) The contributions calculated from the atom pair subset that only selected from the hinge region/flap domain and some TSFF treated atoms in TS2 state. “TSFF atoms” denotes the heavy atoms around the transferring hydrogen in HMG-CoA/NADH that were treated by TSFF in the TS2 state.
4. Cloning Expression, Purification and Kinetics of HMGR Mutants

The mutants for this study were developed from the PKK-177-3 vector developed in the laboratory of Victor Rodwell for the expression of HMGR from Pseudomonas mevalonii. This vector contains the expression sequence for the protein under a tac promoter and an Ampicillin marker. There are no additional affinity tag residues added to the protein product. The mutants were generated using the New England Biolabs Q5 Site-Directed Mutagenesis system utilizing specific DNA alterations encoded in the primers, supplied by Integrated DNA Technologies. The result of the PCR reaction was the entire PKK-177-3 vector coding for the wild type or one of the mutant HMGR proteins. These were transformed into DH5a and the DNA sequence confirmed by the Purdue Genomics center before being transformed into the BL21 E coli expression system.

The induction and purification protocol was adapted from Rodwell and coworkers. Briefly, the cells were grown in LB broth containing 100 mg/L of Ampicillin to an OD600 of about 0.6. At that point 0.5 mM of B-D-1-thiogalactopyranoside (IPTG) was introduced and the cells moved to 16C incubator for an overnight induction. The cells were harvested by centrifugation at 10,000 RPM in a JA-12 Beckman rotor. The cell pellet was then resuspended in 20 ml of PEG wash buffer (phosphate 10 mM, EDTA 1 mM and glycerol 10% w/v, all pH values adjusted to 7.3) and the recommended concentration of the HALT protease inhibitor cocktail (Thermofisher). This slurry was stored at -80C.

These cells were subsequently lysed with three passages through a french press. This was followed by centrifugation in a JA-25.5 Beckman rotor at 15,000 RPM for 15 minutes. The supernatant was then transferred to a Ti70.1 Beckman ultracentrifuge rotor and pelleted at 50,000g for one hour. The resulting supernatant was isolated and then precipitated with 2/3 volume of saturated ammonium sulfate adjusted to a pH of 7.0. The resulting pellet was isolated and resuspended in 20 ml of PEG buffer plus HALT protease inhibitors then stored at -80C.

After thawing this ammonium sulfate-lysate was added to 500 ml of PEG buffer and passed over a 10 ml DEAE ion exchange column. After washing with additional PEG buffer the material was eluted using a gradient of potassium chloride from 0 to 150 mM. Fractions were collected and those containing a band of the appropriate molecular weight for HMGR were combined and precipitated with ammonium sulfate. The resulting pellets were resuspended in PEG buffer to a concentration of approximately 10 mg/ml. A final gel showed greater that 95% purity for the wild type and the three mutants.

The kinetic experiments were run in a buffer solution (phosphate 25 mM, NaCl 50 mM, pH 8.7) and combined 2000 μM of mevalonate, 600 μM of CoA and 600 μM of NAD+, resulting in the production of two NADH and HMG-CoA. These were followed at 340 nm in a BioTEK Synergy H1 plate reader for two minutes and the initial stable rate reported as the Vmax of the reaction in milli-absorption units per minute. All reactions were repeated 8 times and averaged.

5. References
1 Haines, B. E., Steussy, C. N., Stauffacher, C. V. & Wiest, O. Molecular Modeling of the Reaction Pathway and Hydride Transfer Reactions of HMG-CoA Reductase. Biochemistry 51, 7983-7995 (2012).
2 Gaussian 09, Revision A1 (Wallingford, CT, 2009).
3 www.github.com/q2mm.
4 Rosales, A. R. et al. CatVS: Virtual screening of asymmetric catalysts. Nature Catalysis 2, 41-45 (2018).
Tao, P. et al. Matrix metalloproteinase 2 inhibition: combined quantum mechanics and molecular mechanics studies of the inhibition mechanism of (4-phenoxyphenylsulfonyl) methylthiirane and its oxirane analogue. Biochemistry 48, 9839-9847 (2009).

Vreven, T. et al. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. J. Chem. Theor. Comp. 2, 815-826 (2006).

Norrby, P. O. & Liljefors, T. Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data. J. Comp. Chem. 19, 1146-1166 (1998).

Walker, R. C., de Souza, M. M., Mercer, I. P., Gould, I. R. & Klug, D. R. Large and fast relaxations inside a protein: Calculation and measurement of reorganization energies in alcohol dehydrogenase. J. Phys. Chem. B 106, 11658-11665 (2002).

Pavelites, J. J., Gao, J., Bash, P. A. & Mackerell Jr, A. D. A molecular mechanics force field for NAD+ NADH, and the pyrophosphate groups of nucleotides. J. Comp. Chem. 18, 221-239 (1997).

Hornak, V. et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins Struct. Funct. Bioinformatics 65, 712-725 (2006).

Tabernero, L., Bochar, D. A., Rodwell, V. W. & Stauffacher, C. V. Substrate-induced closure of the flap domain in the ternary complex structures provides insights into the mechanism of catalysis by 3-hydroxy-3-methylglutaryl-CoA reductase. Proc. Nat. Acad. Sci. 96, 7167-7171 (1999).

Haines, B. E., Wiest, O. & Stauffacher, C. V. The Increasingly Complex Mechanism of HMG-CoA Reductase. Acc. Chem. Res. 46, 2416-2426 (2013).

Bowman, G. R., Ensign, D. L. & Pande, V. S. Enhanced modeling via network theory: adaptive sampling of Markov state models. J. Chem. Theor. Comp. 6, 787-794 (2010).

Sheong, F. K., Silva, D.-A., Meng, L., Zhao, Y. & Huang, X. Automatic state partitioning for multibody systems (APM): an efficient algorithm for constructing Markov state models to elucidate conformational dynamics of multibody systems. J. Chem. Theor. Comp. 11, 17-27 (2014).

Price, D. J. & Brooks III, C. L. A modified TIP3P water potential for simulation with Ewald summation. J. Chem. Phys. 121, 10096-10103 (2004).

Naritomi, Y. & Fuchigami, S. Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. J. Chem. Phys. 139, 12B605_601 (2013).

Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. Curr. Opin. Struct. Biol. 25, 135-144 (2014).

Noé, F. & Clementi, C. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. Curr. Opin. Struct. Biol. 43, 141-147 (2017).

Steussy, C. N. et al. A Novel Role for Coenzyme A during Hydride Transfer in 3-Hydroxy-3-methylglutaryl-coenzyme A Reductase. Biochemistry 52, 5195-5205 (2013).

Beach, M. J. & Rodwell, V. W. Cloning, sequencing, and overexpression of mvaA, which encodes Pseudomonas mevalonii 3-hydroxy-3-methylglutaryl coenzyme A reductase. J. Bact. 171, 2994-3001 (1989).

Rodwell, V. W. et al. 3-Hydroxy-3-methylglutaryl-CoA reductase. Met. Enzym. 324, 259-280 (2000).
22 Wilding, E. I. *et al.* Essentiality, Expression, and Characterization of the Class II 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase of *Staphylococcus aureus*. *J. Bact.* **182**, 5147-5152 (2000).
| File Name                 | Size      | Actions                      |
|--------------------------|-----------|------------------------------|
| bfact_Large.mp4          | 31.71 MiB | view on ChemRxiv, download file |
| HMGR_bfacts.pse          | 4.14 MiB  | view on ChemRxiv, download file |
| ts2-wat.prmtop           | 61.30 MiB | view on ChemRxiv, download file |