

Microsecond Timescale Simulations at the Transition State of *Pm*HMGR Predict Remote Allosteric Residues

Authors: Taylor R. Quinn¹, Calvin N. Steussy², Brandon E. Haines³, Jinping Lei⁴, Wei Wang⁴, Fu Kit Sheong,⁴ Cynthia V. Stauffacher², Xuhui Huang⁴, Per-Ola Norrby^{1,5}, Paul Helquist¹, Olaf Wiest^{1,6*}

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 the second hydride transfer transition state of HMG-CoA reductase from *Pseudomonas mevalonii* (*Pm*HMGR) 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^{2,3} and different QM/MM-MD approaches have been developed,⁶ but they are typically iterative approximations to the problem that are limited to protein dynamics on the low nsec timescale.⁷ Although protein dynamics at the transition state is not commonly a direct factor in enzyme catalysis,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 μ sec timescale that are not accessible by QM/MM-MD approaches.

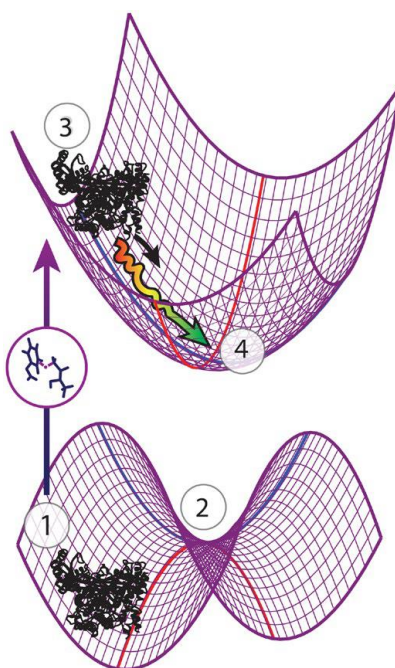


Figure 1. Schematic view of computational studies of enzyme catalysis. An initial guess for the transition state is introduced in a crystal structure of an enzyme (1), leading to a high-energy structure (3) that is optimized to (4), which is assumed to represent the reactive complex (2). Conformational changes from (3) to (4) are likely to occur on the μ sec 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,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) method¹⁰ and the related ACE

method^{11,12} have been very successful in studying catalysis in enzymes^{13,14} 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 case¹⁵ 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,^{5,16} but have only been sparingly used for the study of enzyme reactivity^{17,18} 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^{5,16} 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 χ^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,^{5,19} 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.

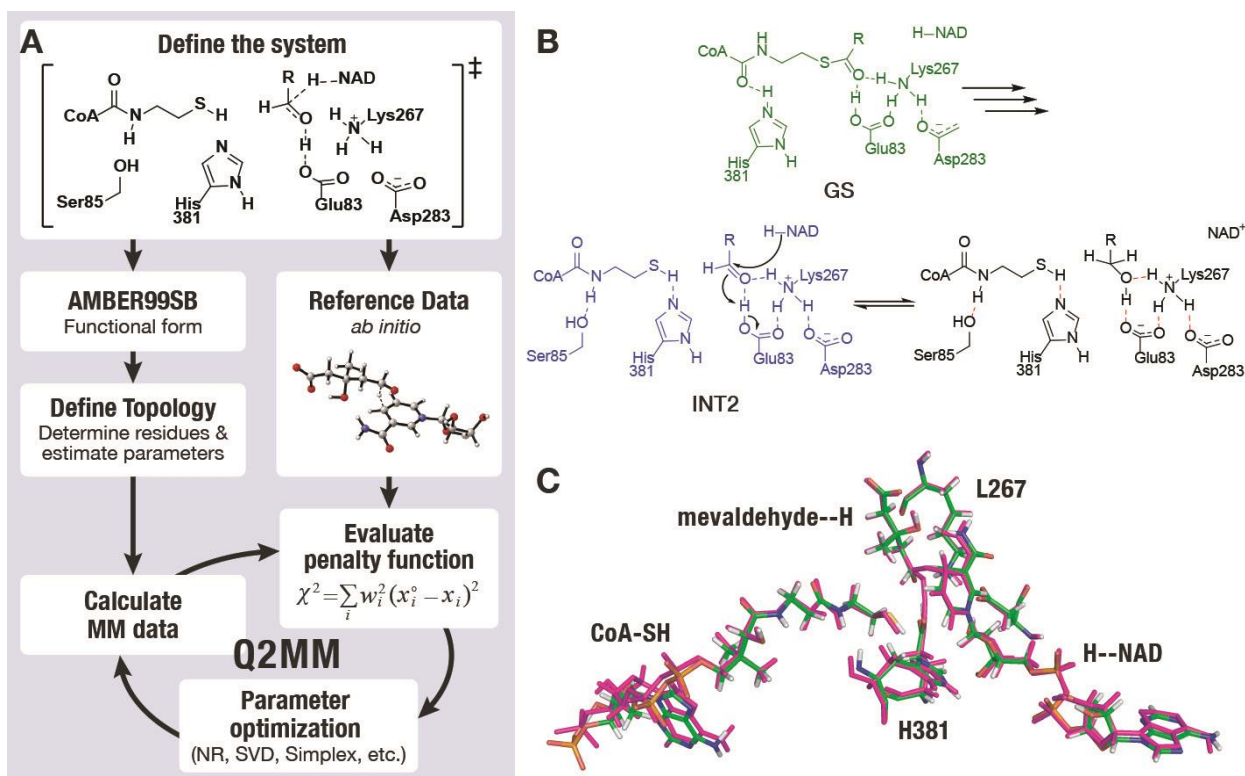


Figure 2. A) Mechanism of *PmHMGR*. Green is the GS, blue is the INT2, and TS2 is described by the arrows in INT2. B) Residues included in the parameterization of the TS2 TSFF. C) The optimized TSFF (green carbons) is overlaid with the QM optimized structure (magenta).

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 publicly 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 of kcal⁻¹ 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 μ s of adaptive sampling²⁹ followed by 3-5 μ s of MD simulation in an NVT ensemble. For comparison, the ternary complex of the starting material, NADH, and *Pm*HMGR (constructed from non-productive ternary complex pdb code 1QAX, 2.8 Å resolution)³⁰ 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).

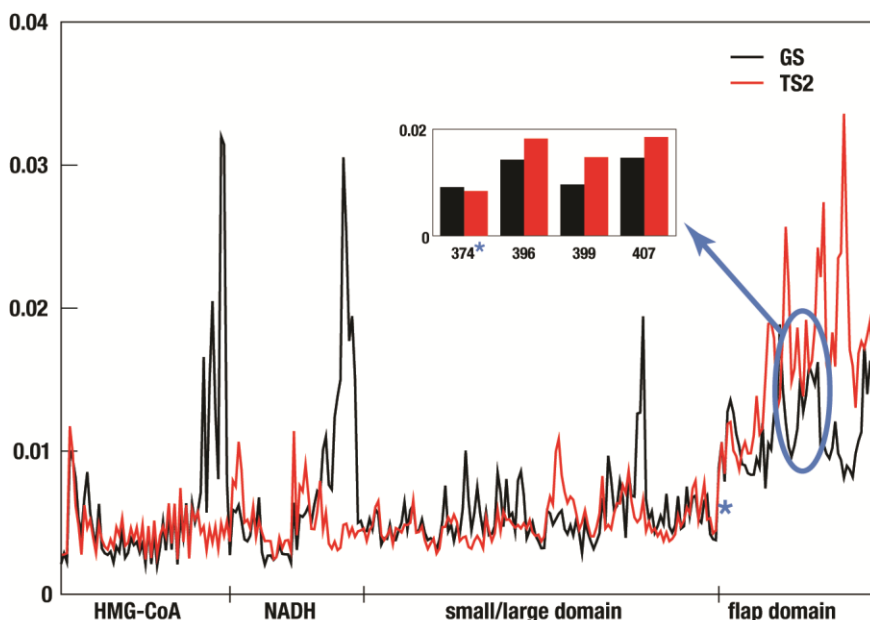


Figure 3. Contribution (tICA coefficient) of the selected residues for the slowest dynamics of the system in the ground state (GS, black) and transition state (TS2, red). The contributions of T374, R396, E399 and L407 are 0.009, 0.014, 0.010 and 0.015, respectively, in the GS state, while they are 0.008, 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 C α atoms of the residues that contact with HMG-CoA/NADH in the small and large domain, and “flap domain” denotes the C α 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 *Pm*HMGR 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”.³¹ 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 *Pm*HMGR's enzymatic catalysis. This is noteworthy because the flap domain has been postulated to be involved in catalysis,²³ 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.

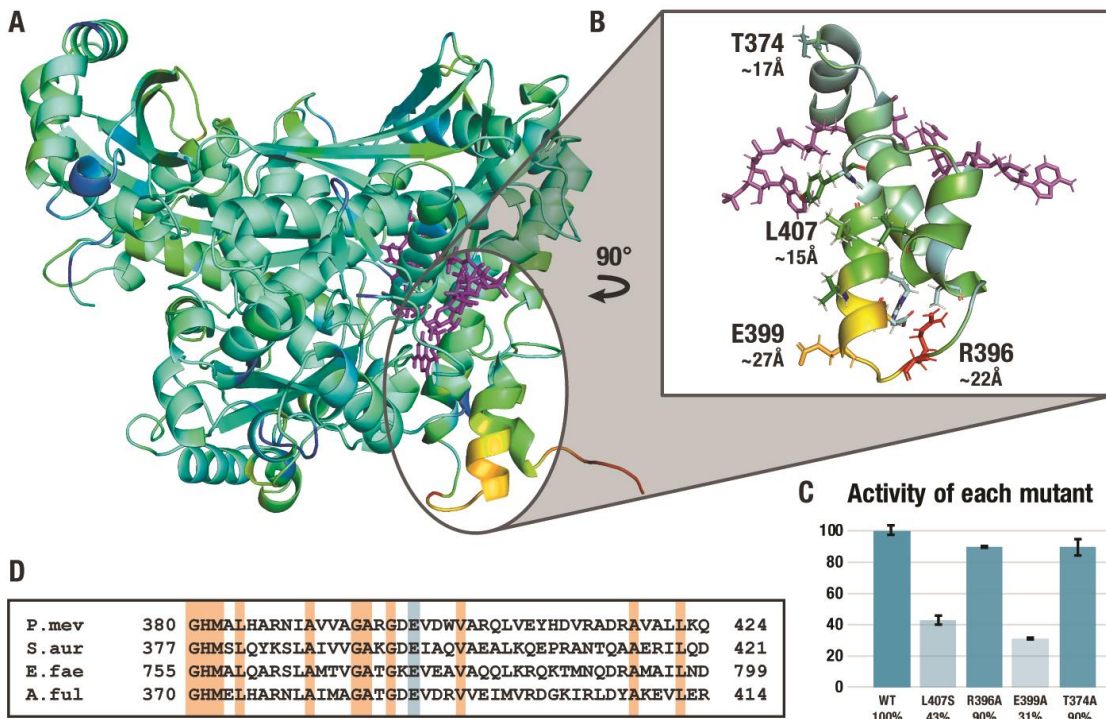


Figure 4. A) RMSD difference between the ground state and the transition state. Yellow/red indicates areas where the RMSD of the GS over the trajectory is larger than in TS2 B) Location of mutated residues highlighted in green residues and neighboring interactions. C) Results of the

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),³² 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 transition state of enzyme catalyzed reactions on the μ sec 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 design³³ rather than the more typical computational rationalization of experimental observations. Given the role of dynamics on enzyme catalysis,⁴ 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 code²⁵ will allow future applications of TSFFs to computational enzymology on the μ sec 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 experimentals 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).

- 4 Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-916 (2007).
- 5 Rosales, A. R. *et al.* Application of Q2MM to Predictions in Stereoselective Synthesis. *Chem. Comm.* **54**, 8294-8311 (2018).
- 6 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).
- 7 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).
- 8 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).
- 9 van der Kamp, M. W. *et al.* Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nature Comm.* **9**, 1177 (2018).
- 10 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).
- 11 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).
- 12 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).
- 13 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).
- 14 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).
- 15 Jensen, F. & Norrby, P.-O. Transition states from empirical force fields. *Theor. Chem. Acc.* **109**, 1-7 (2003).
- 16 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).
- 17 Rydberg, 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).
- 18 Rydberg, 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).
- 19 Rosales, A. R. *et al.* CatVS: Virtual screening of asymmetric catalysts. *Nature Catalysis* **2**, 41-45 (2018).
- 20 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).
- 21 Goldstein, J. L. & Brown, M. S. Regulation of the mevalonate pathway. *Nature* **343**, 425-430 (1990).

- 22 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).
- 23 Haines, B. E., Wiest, O. & Stauffacher, C. V. The Increasingly Complex Mechanism of HMG-CoA Reductase. *Acc. Chem. Res.* **46**, 2416-2426 (2013).
- 24 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).
- 25 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).
- 26 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).
- 27 Gaussian 09, Revision A1 (Wallingford, CT, 2009).
- 28 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).
- 29 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).
- 30 Taberner, 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).
- 31 Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135-144 (2014).
- 32 Hedl, M., Taberner, L., Stauffacher, C. V. & Rodwell, V. W. Class II 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductases. *J. Bacteriol.* **186**, 1927-1932 (2004).
- 33 Goodey, N. M. & Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nature Chem. Bio.* **4**, 474 (2008).