Electronic Supplementary Information for

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

Taylor R. Quinn[1,2], Calvin N. Steussy[3], Brandon E. Haines[4], Jinping Lei[5,6], Wei Wang[5,] Fu Kit Sheong[5], Cynthia V. Stauffacher[3], Xuhui Huang[5], Per-Ola Norrby[1,7], Paul Helquist[1], Olaf Wiest[1,8*]
[1]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA.
[2] Early TDE Discovery, Early Oncology, Oncology R&D, AstraZeneca, Boston, United States
[3] Department of Biological Sciences and Purdue Center for Cancer Research, Purdue University, West Lafayette, IN 47907, USA.
[4] Department of Chemistry, Westmont College, Santa Barbara, CA 93108, USA.
[5] Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.
[6] School of Pharmaceutical Sciences School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China
[7] Data Science and Modelling, Pharmaceutical Sciences, R&D, AstraZeneca Gothenburg, Pepparedsleden 1, SE-431 83 Mölndal, Sweden.
[6] Lab of Computational Chemistry and Drug Design, School of Chemical Biology and Biotechnology, Peking University, Shenzhen Graduate School, Shenzhen, China.

Corresponding Author: Olaf Wiest
**Email:** owiest@nd.edu
**ORCID ID:** 0000-0001-9316-7720

**This PDF file includes additional information on:**



**Additional ESI files**

HMGR with color-coded RMSF and residues chosen for mutations in mp4 format
Pymol Session of transition state model with color-coded RMSF.
AMBER-formatted parameter file for TS2 in prmtop format (packed as .tar file).

**1. Parameterization of the TSFF**

The QM reference data used to parameterize the TSFF for the second hydride transfer of *Pm*HMGR 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 Cα of catalytically relevant residues E83, K267, D283, and H381. (**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.

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 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.(8) 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,(9, 10) and AMBER99SB,(8) 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).(11) The transition state (TS2) was built via the methods described below. The AMBER99SB(8) 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 using explicit solvation with TIP3P waters(12) and 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,(13) selecting starting structures from the above short unbiased MD trajectories.(14) Starting structures for the adaptive sampling were chosen from the unbiased MD trajectories above by a K-centers(14) clustering algorithm. The root-mean-square fluctuation (RMSF) of the $C_\alpha$ 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(12) water molecules using a truncated octahedral box with a box length of180 Å through ADDTOBOX program in AMBER.(15) 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.(16, 17) 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.(18) 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 - $C_\alpha$ atoms of hinge region and flap domain
b) heavy atoms of substrate and cofactor - $C_\alpha$ atoms of small domain and large domain residues that contact with substrate and cofactor in the crystal structure (PDBID:1QAX)
c) $C_\alpha$ atoms of hinge region and flap domain - $C_\alpha$ atoms of hinge region and flap domain
d) $C_\alpha$ atoms of hinge region and flap domain - $C_\alpha$ 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.(19) 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.

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 - $C_\alpha$ atoms of hinge region and flap domain*

*c) $C_\alpha$ atoms of hinge region and flap domain - $C_\alpha$ 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)**.

### 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 *Pm*HMGR. 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 residuefor 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.

## 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(20) 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.(21) 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.(22) 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.

**Fig. 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.

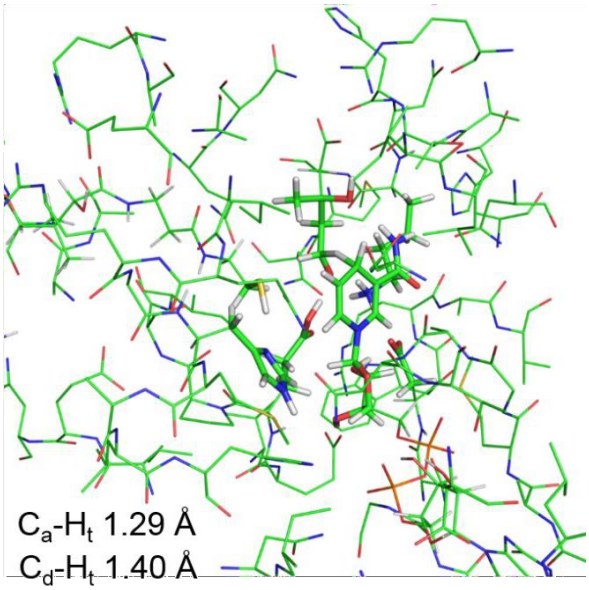


**Fig. S2.** New atom types defined for parameterization of the TSFF

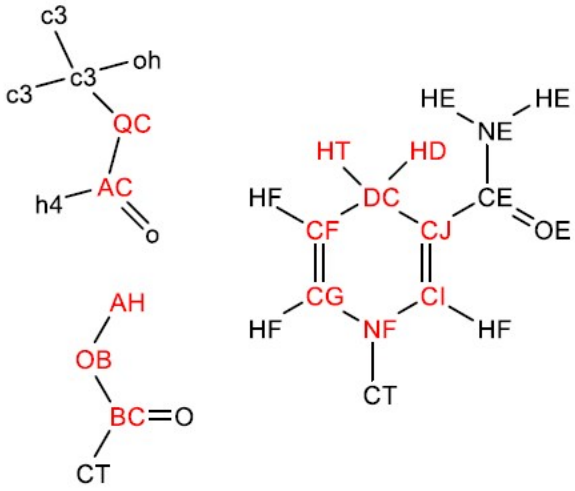


**Fig. S3.** Overall Mechanism of the reduction of HMGCoA by NADH, catalyzed by HMGR[12]

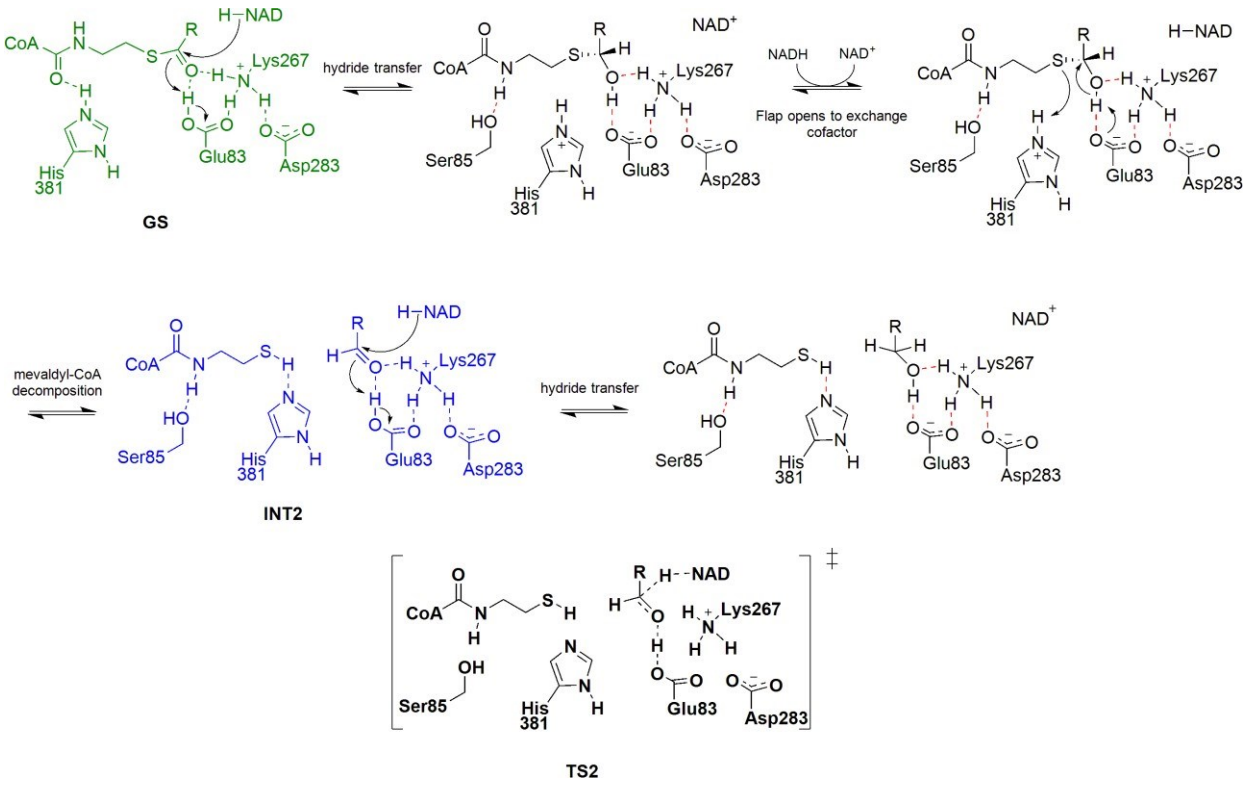


**Fig. S4.** The implied timescales (ITS) for tICA in the ground state (GS, blue), intermediate state (INT2, green) and transition state (TS2, red) respectively. The slowest dynamic (the top line) of the system can be obtained by tICA but not from shorter MD simulations.

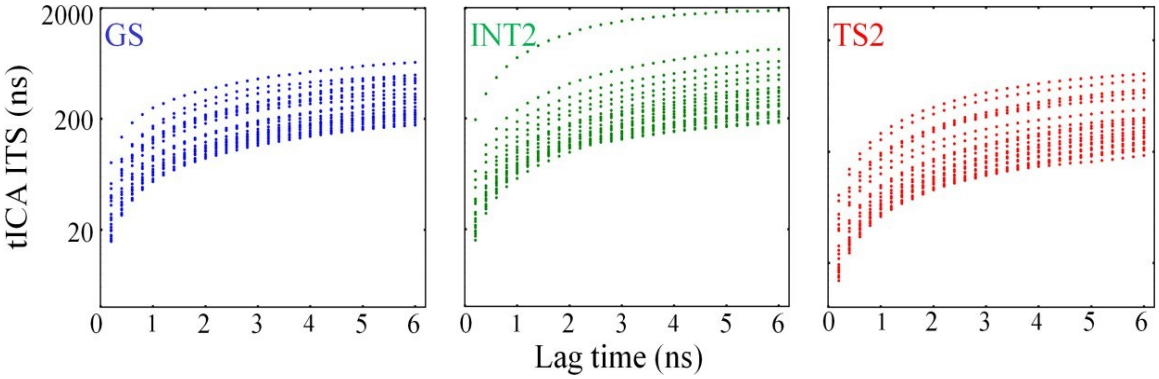


**Fig. S5.** The projection of the TSFF MD samplings (TS2, red), the MD conformations of the intermediate state (INT2, green) and ground state (GS, blue) onto the slowest two tICs of the ground state

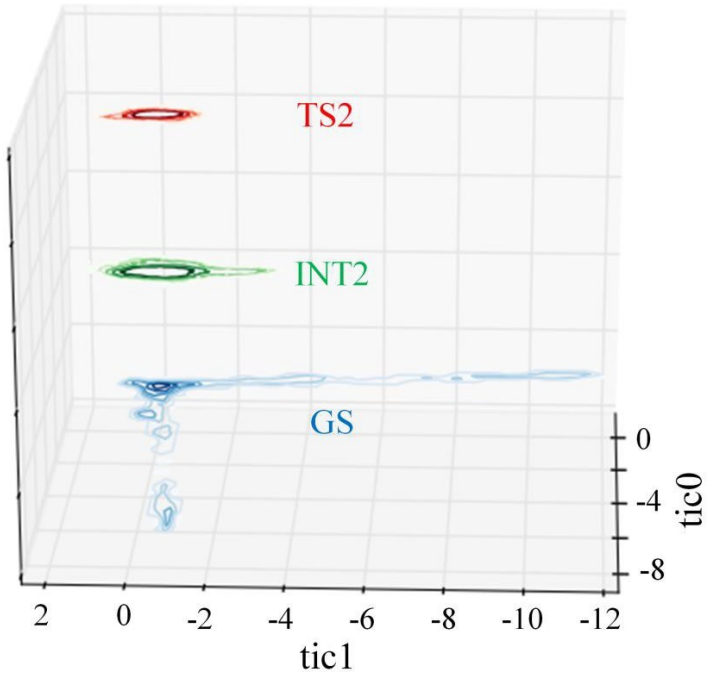
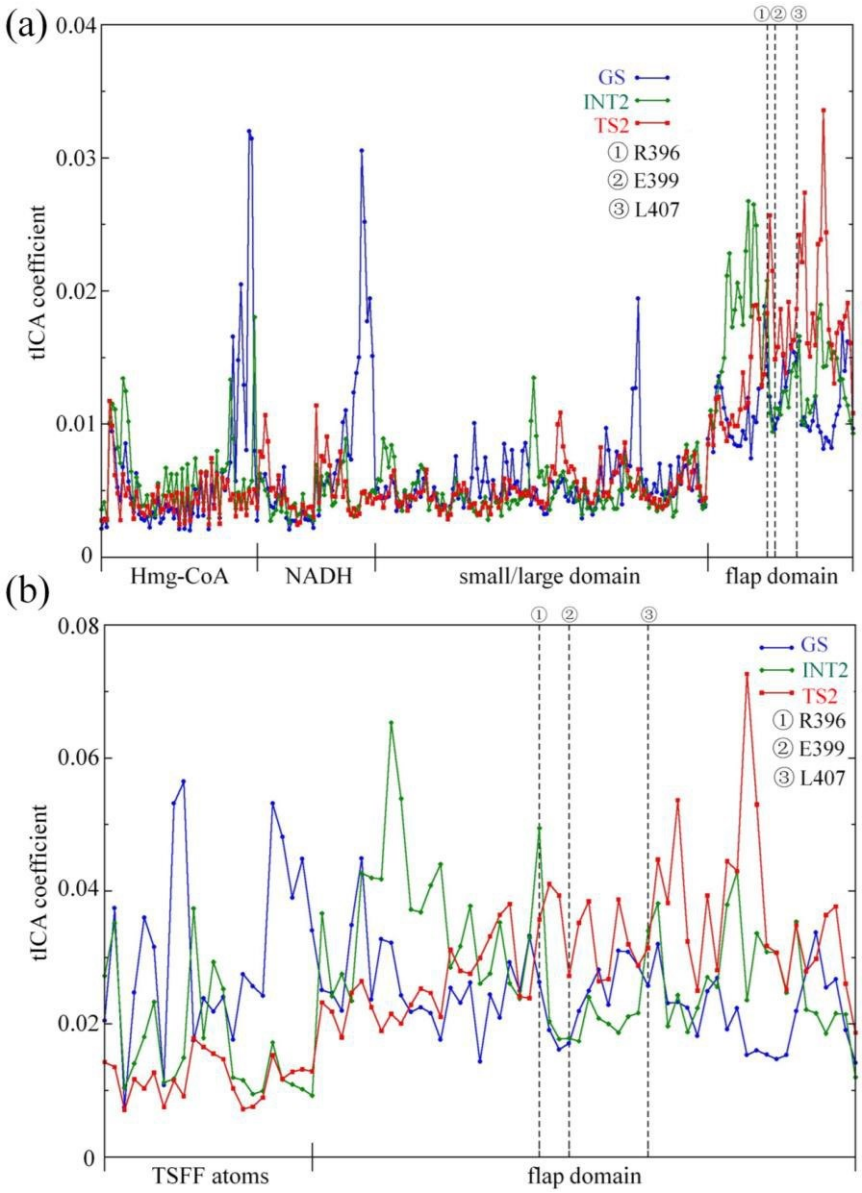


**Fig. S6.** Contributions (tICA coefficient) of the selected residues for the slowest dynamics of the system in the ground state (GS, blue), intermediate state (INT2, green) and transition state (TS2, red). (a) The contributions of R396 (①), E399 (②) and L407 (③) are: 0.014, 0.010 and 0.015 respectively 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



**SI References**


1. B. E. Haines, C. N. Steussy, C. V. Stauffacher, O. Wiest, Molecular Modeling of the Reaction Pathway and Hydride Transfer Reactions of HMG-CoA Reductase. *Biochemistry* **51**, 7983-7995 (2012).
2. M. J. Frisch, Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., ;, Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J., J. A. ;, Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., Fox, D. J. (2009) Gaussian 09, Revision A1. (Wallingford, CT).
3. github.com/q2mm.
4. A. R. Rosales *et al.*, Application of Q2MM to Predictions in Stereoselective Synthesis. *Chem. Comm.* **54**, 8294-8311 (2018).
5. P. Tao *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).
6. T. Vreven *et al.*, Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theor. Comp* **2**, 815-826 (2006).
7. P. O. Norrby, T. Liljefors, Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data. *J. Comp. Chem.* **19**, 1146-1166 (1998).
8. V. Hornak *et al.*, Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinformatics* **65**, 712-725 (2006).
9. R. C. Walker, M. M. de Souza, I. P. Mercer, I. R. Gould, D. R. Klug, Large and fast relaxations inside a protein: Calculation and measurement of reorganization energies in alcohol dehydrogenase. *J. Phys. Chem. B* **106**, 11658-11665 (2002).
10. J. J. Pavelites, J. Gao, P. A. Bash, A. D. Mackerell Jr, A molecular mechanics force field for NAD+ NADH, and the pyrophosphate groups of nucleotides. *J. Comp. Chem.* **18**, 221-239 (1997).
11. L. Tabernero, D. A. Bochar, V. W. Rodwell, C. V. Stauffacher, 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).
12. D. J. Price, C. L. Brooks III, A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* **121**, 10096-10103 (2004).
13. G. R. Bowman, D. L. Ensign, V. S. Pande, Enhanced modeling via network theory: adaptive sampling of Markov state models. *J. Chem. Theor. Comp.* **6**, 787-794 (2010).
14. F. K. Sheong, D.-A. Silva, L. Meng, Y. Zhao, X. Huang, 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).
15. D. A. Case *et al.*, Amber 14. (2014).
16. Y. Naritomi, S. Fuchigami, 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).
17. J. D. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135-144 (2014).
18. F. Noé, C. Clementi, Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Curr. Opin. Struct. Biol.* **43**, 141-147 (2017).

19. C. N. Steussy *et al.*, A Novel Role for Coenzyme A during Hydride Transfer in 3-Hydroxy-3-methylglutaryl-coenzyme A Reductase. *Biochemistry* **52**, 5195-5205 (2013).
20. M. J. Beach, V. W. Rodwell, Cloning, sequencing, and overexpression of mvaA, which encodes Pseudomonas mevalonii 3-hydroxy-3-methylglutaryl coenzyme A reductase. *J. Bact.* **171**, 2994-3001 (1989).
21. V. W. Rodwell *et al.*, 3-Hydroxy-3-methylglutaryl-CoA reductase. *Met. Enzym.* **324**, 259-280 (2000).
22. E. I. Wilding *et al.*, Essentiality, Expression, and Characterization of the Class II 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase ofStaphylococcus aureus. *J. Bact.* **182**, 5147-5152 (2000).