MD Simulations Reveal Complex Water Paths in Squalene Hopene Cyclase – Tunnel Obstructing Mutations Increase the Flow of Water in the Active Site

Camilla Gustafsson[a], Serguei Vassiliev[b], Charlotte Kürten[c], Per-Olof Syrén[a,c] and Tore Brinck[a]

[a] Applied Physical Chemistry, School of Chemical Science and Engineering, KTH Royal Institute of Technology, Teknikringen 36, 100 44 Stockholm, Sweden

[b] Department of Biological Sciences, Brock University, Mackenzie Chown F 234, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1, Canada

[c] Science for Life Laboratory, Stockholm – School of Biotechnology, Division of Proteomics and Nanobiotechnology, KTH Royal Institute of Technology, Tomtebodavägen 23a, 171 65 Solna, Sweden

Supporting Information

In the PDB-file 2SQC, residue number 376 is a CYS, but should natively be an ASP. This residue was changed back to an ASP. The hydrogens were removed, and added once more by the Reduce tool in AMBER. [57] The protonation states for the amino acids in the active site were determined by the program propKa. The pH was set to 6, and at that pH the aspartate should be protonated, and the histidine deprotonated according to the Henderson-Hasselbalch equation. The protonation state of the His 451 is debated, and it was protonated (HIP).

Both 1UMP and 2SQC were used for the MD analysis. Structures used for each simulation are listed in Table S1.

Table S1. PDB files used for the different simulations.

Model	Structure
Empty protein	2SQC
Substrate in channel (3 different locations)	2SQC
Substrate in active site (substrate and protonated substrate)	1UMP
Intermediate in active site	1UMP
Product in active site	1UMP
Substrate in active site (mutated S38W) (substrate, protonated substrate)	1UMP
Substrate in active site (mutated S168F)	1UMP
Substrate in active site (mutated S168W)	1UMP
Substrate in active site (mutated F605W)	1UMP
Substrate in active site (mutated V440F)	1UMP
Product in active site (higher temperature)	1UMP
Product in channel	1UMP

All simulations were performed in the NTP ensemble (constant number of particles, temperature and pressure) with the AMBER14 package. [34]

Minimization was performed on the CPU, the sequential steps (heating, equilibrium and production) was performed on the GPU on Ubuntu 14.04.

The optimal temperature for SHC is between 303 and 330 K. Previous studies have shown that MD simulations for POPC bilayers are feasible at temperatures up to 330 K. [58] Heating was performed slowly in two sequential runs, first between 0-100 K over 5 ps with a random seed for the Langevin thermostat. In the second heating between 100-330 K performed over 100 ps, an anisotropic Berendsen weak-coupling barostat was used for pressure equilibration, and a Langevin thermostat for the temperature.

Since periodic boundary conditions were used, a 5 ns equilibration phase with a barostat was performed before the production MD was initiated.

The temperature in the production simulations was 330 K controlled by a Langevin thermostat. This temperature is above the phase transition of POPC lipids, and the lipids were at equilibrium in the liquid crystalline state. [59] The pressure was 1 bar with anisotropic Berendsen pressure coupling with constant pressure using periodic boundary conditions. SHAKE was applied to constrain covalent bonds with hydrogen atoms, the cutoff for nonbonded interactions was 10.0 Å. The length of the production simulations were between 150-300 ns.

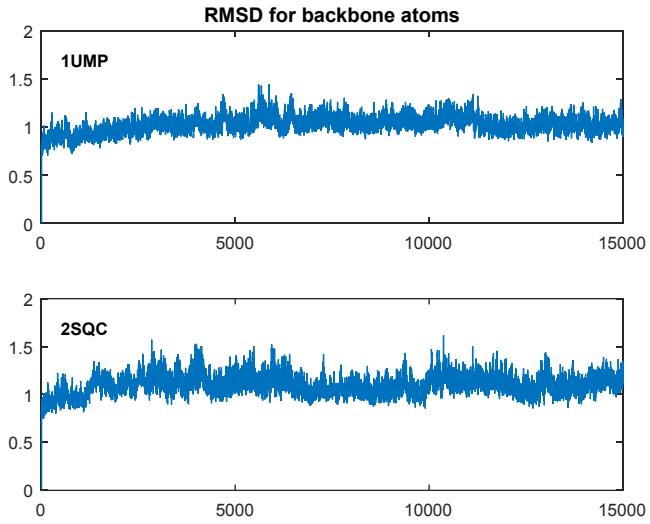

Figure S1. RMSD for backbone atoms in 1UMP and 2SQC over 150 ns trajectory.

Asp376 was chosen as initial starting point for the CAVER analysis.

Table S2. B-factors for the residues in the DXDD motif for the protein variants. The analysis shown represents the last 100 ns of the 150 ns trajectory.

Residue	wt	S38W	S168F	S168W	F605W	V440F
374	8.39	7.72	7.73	9.79	9.09	7.07
376	8.33	7.24	7.83	9.26	8.58	8.04
377	13.61	11.91	12.61	12.51	14.55	13.92

In addition to the bond length analysis determined for the 150 ns long simulations (last 100 ns used), see Figure S2 for definition, the average bond length distance for each of the five shorter (20 ns) parallel trajectories for the wt and S38W can be found in Table S3.

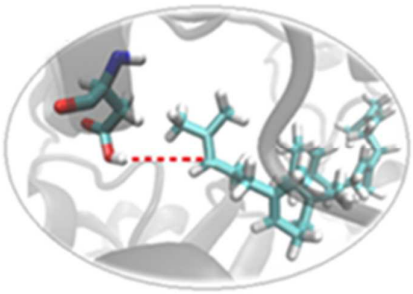

Figure S2. The relevant protonation distance between the catalytic residue Asp376 and the proton accepting carbon in the substrate is indicated with a red dotted line.

Table S3. Average proton transfer distance between the catalytic residue and the substrate for the wt and S38W for the five parallel trajectories of 20 ns (Traj1-5). The last two columns show the average for the five parallel trajectories and average for the long trajectories (150 ns), respectively.

Variant	Traj1	Traj2	Traj3	Traj4	Traj5	Average (20 ns)	Average (150 ns)
wt	3.4	4.2	4.0	5.2	5.4	4.4	4.1
S38W	5.9	6.0	5.8	5.5	5.6	5.8	4.8

This expanded investigation of the sample space showed that the distance tended to be generally longer between the catalytic residue and the substrate in the S38W mutant than in the wt and confirmed the results from the initial single trajectories of 150 ns.

Streamline tracking

To implement this technique, time-averaged 3D direction of motion data is extracted from the dynamics simulation by determination of a diffusion tensor. Directionality of diffusion is represented by a tensor which is described with nine components, each one associated with a pair of axes xx, yy, zz, xy, yx, xz, zx, yz. To calculate the tensor field the simulation system volume is divided into *n* small cubic elements (voxels) with a volume of about 1 $Å^3$ each. For all voxels containing a water molecule at time *t* the position of this water molecule at time (t+Δt) is determined and the tensor elements are calculated according to the Einstein relation:

$$T^{\alpha\beta} = \frac{\langle[\alpha(t+\Delta t)-\alpha(t)]\rangle \cdot \langle[\beta(t+\Delta t)-\beta(t)]\rangle}{2\Delta t} \qquad \alpha,\beta = \{x,y,z\}$$

To obtain all 9 tensor elements α and β are sequentially substituted with x, y, and z. Tensor elements are averaged over the time window of the MD run and the diffusion tensor is diagonalized to find the direction of fastest motion. In this work Δt was 10 ps.

Streamlines are started in voxels which exhibit the highest linear anisotropy of water diffusion: $c(x,y,z) = (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2 + \lambda_3)$, where λ are eigenvalues or the diffusion tensor. The value of c threshold in this study was c>0.5 Neighboring voxels with a high diffusion rate favoring a particular direction are then connected by tracking algorithm to give streamlines. Streamlines thus represent areas where water molecules have moved the most in a directional manner in the MD trajectory.