# 1 Selection of Variants

To guide the selection of mutations for directed evolution, a multiple sequence alignment was performed between the human and mouse wild-type protein. This MSA revealed 47 positions that differed between the human and mouse sequences, suggesting potential sites for mutagenesis. In order to prioritize residues for mutagenesis, the distances from the alpha carbon of Cys/Sec49 (active site residue) and rest of the 46 positions were calculated. Residues were then grouped into distance criteria of <10 Å, 10-15 Å, 15-20 Å, 20-25 Å, 25-30 Å, and 30-35 Å. This distance-based approach allowed us to select residues within different proximity ranges to Cys/Sec49. Grouping residues into distance bins provides a systematic and stochastic way to explore the impact of mutations at varying distances from the target residue (Cys/Sec49). We have mutated all these residues one by one slowly changing the human into mouse and the mouse into human structure, in order to understand the changes in the activation barrier in going from proximal to distal mutations in both the structures.

# 2 Computational Model Preparation

The initial structure used to make the EVB model for mouse cysteine wild type was (PDB ID - 7FC2) and for human selenocysteine wild type it was generated with Alphafold. The parametrization includes in preparing a simulation that has missing force field parameters along with standard parameters and library files. In this case, it is Selenium (U). After equilibration, continuing with the FEP simulations which constitute the core part of the EVB methodology. For creating the selenium parameters (selenol, selenolate ion and selenenic acid), FFLD in Maestro was used. We used the charges provided by Maestro; the hydrogen and solvent were added using the Q program [1]. The TIP3P water parameters were used in combination with the other protein parameters not present in the standard OPLS library in Q.

## 2.1 Free Energy Calculation Using Q

Free energy perturbation (FEP) calculations with Q involve running a set of consecutive input files which have the mapping parameter $\lambda$ ranging in a way (usually [1, 0] to [0, 1] between two states). Qfep is a program which reads the energy files generated by Qdyn and calculates the total change in free energy for the complete perturbation from state A ($\varepsilon_1$) to state B ($\varepsilon_2$). Zwanzig's formula as shown below calculates the difference in free energy between the two states:

$$\Delta G = \sum \Delta g = \sum -R \cdot T \cdot \ln \left\langle e^{-\left(\frac{\Delta V_{\text{eff}}}{R \cdot T}\right)} \right\rangle_A \qquad (1)$$

Here, $\Delta V_{\text{eff}}$ is defined as the difference in $V_{\text{eff}}$ between two adjacent perturbation steps.

Qfep also calculates free energy functions, or potentials of mean force, using the perturbation formula. The reaction coordinate $X$ is defined as the energy gap between the states $X = \Delta V = \varepsilon_1 - \varepsilon_2$ and is divided into intervals $X_m$ (bins). The first term in the equation represents the free energy difference between the initial state $\varepsilon_1$ and the mapping potential $V_i$:

$$\Delta G(X_m) = \Delta G(\lambda_i) - R \cdot T \cdot \ln \left\langle e^{-\left(\frac{E_g(X_m) - V_i(X_m)}{R \cdot T}\right)} \right\rangle_i \tag{2}$$

The second term represents the free energy difference between the mapping potential $V_i$ and the ground state potential $E_g$. The average in this term is taken over those configurations where $X$ belongs to $X_m$.

$$\Delta G(\lambda_i) = -R \cdot T \cdot \ln \left( \sum_{n=0}^{i-1} \left\langle e^{-\left(\frac{V_{n+1} - V_n}{R \cdot T}\right)} \right\rangle_n \right) \tag{3}$$

$E_g$ is the solution to the secular determinant. The system is then represented by an $n \times n$ EVB Hamiltonian.

$$\left[ \begin{array}{cc} H_{11} & H_{12} \\ H_{21} & H_{22} \end{array} \right]$$

Here, $H_{11}$ and $H_{22}$ are the energies of the two valence states which are calculated using classical force fields. For a two-state representation, the solution becomes:

$$E_g = \frac{1}{2} \cdot (\epsilon_1 + \epsilon_2) - \frac{1}{2} \sqrt{(\epsilon_1 - \epsilon_2)^2 + 4 \cdot H_{12}^2} \tag{4}$$

where $H_{ij}$ or $H_{12}$ is the off-diagonal matrix element representing the quantum mechanical coupling of the states. $H_{ij} \neq 0$ results in the mixing of states $i$ and $j$. In Qfep, the off-diagonal element $H_{ij}$ is a function of the form:

$$H_{ij} = A_{ij} \cdot e^{-(\mu(r_{ij} - r_0) + \eta(r_{ij} - r_0)^2)} \tag{5}$$

The EVB method allows calibration of simulated reference reactions to experimental data obtained from gas-phase or solution experiments. The two EVB parameters $H_{ij}$ (mostly $A_{ij}$) and $\Delta\alpha$ are adjusted in a way until the calculated profile and the experimental data coincide. The $\Delta\alpha$ parameter determines the $\Delta G^\circ$ level, and $H_{ij}$ regulates the degree of mixing of the states at the transition state, i.e., the $\Delta G^{\ddagger}$ level.

## 2.2 EVB simulations

Spherical boundary conditions [2] were applied to the system using Q program [1] with a 50 Å diameter water sphere surrounding the protein.

After equilibration of the system, EVB free energy perturbation (FEP) calculations were done for the step 1 (until formation of selenolate ion). The FEP

protocol involves a gradual change of the (mapping) potential energy by the coupling parameter. For each of the FEP calculations, 51 discrete windows were considered, each window of 10 ps at 2 fs time step, that gave a total of 1.02 ns of sampling of each free energy profile. These calculations were replicated 100 times starting from the minimized structure for each of the systems, wild type and mutants. The EVB gas phase shift $\alpha$ and the off-diagonal $H_{ij}$ were determined iteratively for all systems keeping human selenocysteine as our reference, the barrier was determined according to the QM/MM calculations carried out by Morokuma [3] for the first step of reaction which was 16.4 kcal/mol.

# References

[1] John Marelius, Karin Kolmodin, Isabella Feierberg, and Johan Åqvist. Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems, 1999.

[2] Gregory King and Arieh Warshel. A surface constrained all-atom solvent model for effective simulations of polar solutions. *The Journal of Chemical Physics*, 91:3647–3661, 1989.

[3] Rajeev Prabhakar, Thom Vreven, Michael J. Frisch, Keiji Morokuma, and Djamaladdin G. Musaev. Is the protein surrounding the active site critical for hydrogen peroxide reduction by selenoprotein glutathione peroxidase? an oniom study. *Journal of Physical Chemistry B*, 110:13608–13613, 7 2006.