

Computational analysis of the evolution of glutathione peroxidase 6 (GPX6) activation free energy

Nayanika Das,[†] Vijay Baladhye,[‡] and Jordi Villà-Freixa^{*,†,¶}

[†]*Computational Biochemistry and Biophysics Lab, Research Group on Bioinformatics and Bioimaging (BP²), Department of Biosciences, Universitat de Vic - Universitat Central de Catalunya, 08500 Vic, Spain*

[‡]*Savitribai Phule Pune University, Pune, India*

[¶]*IRIS-CC*

E-mail: jordi.villa@uvic.cat

Abstract

Outstanding success in computational protein design has been achieved in recent years by combining machine learning approaches with physicochemical properties analysis of the explored variants. Despite these efforts, however, less success has been obtained in designing good computational protocols for optimization of enzyme activity. Here we propose the use of the Empirical Valence Bond (EVB) method to evaluate free energies of activation of enzyme variants to obtain reasonable mutational pathways leading to a functionally optimized protein. In particular, we propose a method that explores a lower free energy difference pathway for the directed evolution of enzymes based on EVB activation free energies. To test the idea, we study the hypothetical connectivity of the selenocysteine-containing human glutathione peroxidase 6 protein (GPX6) into its ortholog cysteine-containing mouse GPX6. We show how it is possible

to find a mutational pathway connecting the two protein sequences and structures that provide the empirical barriers for the first step of the GPX6-catalyzed reaction. Moreover, this protocol offers the potential for addressing complex issues such as epistasis in enzyme engineering, further enhancing its utility in enzyme optimization.

Introduction

Selenium (Se), in the form of selenocysteine (Sec, U—the 21st amino acid), occurs in 25 proteins in the human proteome. Insertion of Sec into a protein is much more complicated than the other 20 amino acids because a UGA stop codon must be recoded as a sense codon for Sec.[?] The complexity of this process signifies that Sec must fulfill a chemical function that exerts biological pressure on the genome to maintain the Sec-insertion machinery.[?] ? The view of Sec as a sophisticated innovation implies that for each occurrence of Sec in an enzyme, there is a unique and specific reason for the use of Se to enhance the enzymatic reaction relative to that of S.[?] This view also implies that since Se “speeds reactions,” Sec should have widely substituted for Cys in enzymes, which clearly has not occurred. Specific reasons for the usage of Sec might include the enhanced nucleophilic character of Se relative to S, or another might be the much lower pKa of a selenol relative to that of a thiol.[?]

Besides the fundamental role organoselenides have in organic catalysis, particularly in oxidations of substrates by H_2O_2 , selenium-mediated redox reactions are key steps in biological processes related to oxidative stress control and cell signaling.[?] Glutathione peroxidase from class 1 (GPX1, EC 1.11.1.9) is a selenoprotein that protects cells from oxidative damage by catalyzing the reduction of H_2O_2 .

Although the catalytic triad of glutathione peroxidase (GPX) has been well recognized, there has been little evidence for the relevance of the interactions among the triad amino acids, i.e., selenocysteine (U), glutamine (Q), and tryptophan (W). Hence, the mechanism has been studied in various aspects. A catalytic mechanism proposed for GPX3 by Prabhakar et al. based on DFT calculations has the resting state of Sec as selenol. In the first part of

this reaction, hydrogen peroxide coordinates to the active site of the enzyme, and the proton transfer is happening to the oxygen atom of the Gln residue.^{7 7} Another study demonstrated that the ionized selenolate state of Sec in GPX1 is the favorable form for the reduction of hydroperoxide substrates where they modeled the first reaction step of the redox cycle of based on the proposed GPX1- *Sec*⁻-Arg177 mechanism.⁷

Mechanism of GPX

A catalytic mechanism proposed for GPX3 by Morokuma et al. involves the formation of a selenolate anion, followed by bond cleavage in hydrogen peroxide, leading to the formation of selenenic acid, with a computed barrier of 18.0 kcal/mol.⁷ Flohe et al. propose an alternate mechanism involving charge separation steps.⁷

Directed Evolution and Epistasis in Enzyme Design

Directed evolution mimics natural selection to optimize enzyme traits. Computational design adapts these principles to predict mutation pathways, potentially enhancing enzyme catalytic performance. By exploring mutational pathways with EVB, we investigate the activation energy variations in GPX6 homologs to assess catalytic activity changes due to selenocysteine/cysteine substitution.^{7 7}

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 the 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.

Combinations of Variants

The approach was more targeted towards leveraging the calculated free energy changes of individual mutants within specific distance bins relative to Cys/Sec49. Mutants exhibiting the lowest ΔG values within each distance bin (10 Å, 15 Å, 20 Å, 25 Å, 30 Å, and 35 Å) were selected and combined. This strategy aimed to identify a "mutant pathway" for directed evolution, progressively incorporating beneficial mutations based on their proximity to the active site and their predicted energetic contributions.

Empirical Valence Bond Model based on QM/MM Calculations

The EVB approach constructs potential energy surfaces for elementary chemical reaction steps.[?] Using a two-state model, the EVB Hamiltonian provides insights into thermodynamic activation parameters based on classical force fields and coupling terms derived from DFT or QM/MM calibrations.[?]

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 preparing a simulation that has missing force field parameters along with standard parameters and library files. In this case, it is Selenium (U). After equili-

bration, 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.[?] The TIP3P water parameters were used in combination with the other protein parameters not present in the standard OPLS library in Q.

Free Energy Calculation Using Q

Free energy perturbation (FEP) calculations were carried out using the Q program.[?] The free energy differences between the Sec and Cys variants were computed by evaluating the perturbations between these two states. The resulting values were used to predict the activation energy barriers for the different mutant pathways. These calculations allow for the exploration of the potential energy surface of the enzyme reactions and to calculate the activation free energy profiles.

Conclusion

In summary, by combining mutagenesis strategies based on proximity analysis, free energy calculations, and empirical valence bond simulations, we demonstrate an approach for identifying mutational pathways that optimize enzyme activity. This framework could be extended to other enzyme systems for more precise and efficient enzyme engineering, potentially enabling the design of enhanced biocatalysts.