ChemComm

COMMUNICATION

Check for updates

Fast rational enzyme design by computational non-equilibrium alchemical transformations

Carlos Castillo-Orellana and Esteban Vöhringer-Martinez [ID] *

Custom-designed enzymes offer potential for sustainable fine chemical production, but traditional experimental methods used for their design are often inefficient and labor-intensive. Here, we propose a computational workflow that predicts changes in activation free energy barriers caused by mutations. This approach uses non-equilibrium alchemical free energy calculation with *ab initio* derived force fields to predict how mutations affect the rate-limiting step in enzyme kinetics. We applied the methodology to two enzymes that catalyze the hydride transfer from NADPH to their respective substrates, achieving results closely matching experimental data with minimal errors of only a few kJ mol^{-1}. Additionally, its low computational requirements make it perfect for high-throughput analyses, aiding in rational enzyme design.

Custom-designed enzymes hold significant potential to revolutionize the environmentally friendly production of fine chemicals, as well as to support the development of a circular economy. The computational design of catalytic sites and ligand-binding pockets within protein frameworks is increasingly resulting in initial, though usually modest, activity.[1] These *de-novo* enzymes can undergo substantial enhancement through directed evolution, which can boost their activity by several orders of magnitude in the most successful examples.[2] The active sites need further refinement to accurately position catalytic residues, stabilizing the most catalytically active conformations or transition states through electrostatic pre-organization.

In experimental settings, enhancing the active site often involves site-directed mutagenesis, which is time-consuming and labor-intensive, often resulting in unsuccessful mutations or promising ones that need sequencing for validation, further increasing time and costs. Computational strategies typically rely on quantum mechanics molecular mechanics (QM/MM) or highly parametrized techniques like empirical valence bond (EVB) theory and reactive force fields, demanding substantial

computational power and human input to confirm and validate a reaction mechanism. Nonetheless, after determining the reaction mechanism for the wild-type and identifying the rate-limiting transition state, it is not essential to trace the entire reaction path to examine the impact of a mutation as we explain below, provided that the reactant and transition state stay unchanged.

Fig. 1 describes a wild type (wt) or mutant (mut) enzyme catalyzed reaction from the reactant state (RS) or Michaelis complex to the transition state (TS) and finally to the product state (PS), highlighting the activation Gibbs free energies. Rational enzyme design aims to predict the difference in activation barriers ($\Delta\Delta G^{\ddagger}$) between these variants. As Gibbs
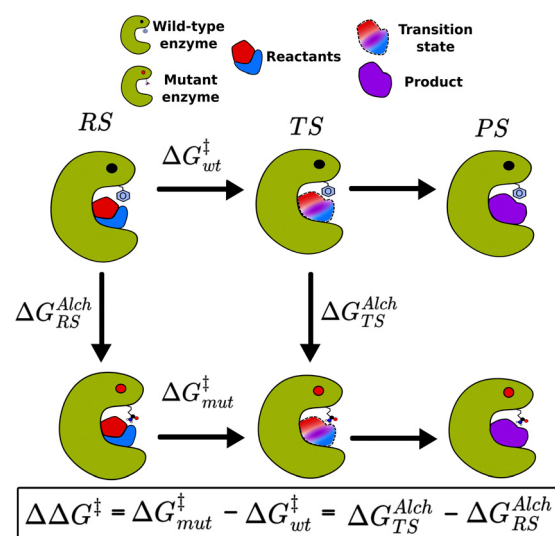

Fig. 1 Enzyme catalyzed reactions from the Michaelis-complex or reactant state (RS) *via* the transitions state to form the products in the wild type enzyme (top) and the mutant (bottom) with the respective activation Gibbs free energies. The alchemical transformations from the wild type enzyme to the variant in the RS or TS state form a thermodynamic cycle which provides the difference in activation free energy between the mutant and wild type as the difference in the alchemical transformation free energies.

Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, Edmundo Larenas 129, Concepción, Chile. E-mail: evohringer@udec.cl

free energy is a state function, this difference can also be calculated by determining the change in free energy when converting alchemically the wild type protein into the mutant, while it is bound to either the transition state or reactant, ΔG_{TS}^{Alch} or ΔG_{RS}^{Alch}. Consequently, $\Delta\Delta G^{\ddagger}$ relevant for rational enzyme design is also accessible as the difference between these values as shown in the equation in Fig. 1.

Research shows that using alchemical transformations of amino acids in proteins within free energy calculations effectively predicts protein stability at the force field level, eliminating the need for costly QM/MM simulations.[3] We propose using this technique to predict the difference in activation free energy between mutant and wild-type proteins in both the reactant and transition states. However, the method's precision relies heavily on accurately describing intermolecular interactions within the catalytic cavity of both states.

To describe these intermolecular interactions accurately, we developed a workflow summarized in Fig. 2 that generates non-polarizable force fields for both states. With the bespoke force field, rational enzyme design becomes easily accessible with non-equilibrium alchemical free energy calculations.

For the reactant state, the focus was on nonbonded interactions due to their impact on the alchemical transformation of amino acid side chains in the active site. Bonded interactions are usually accessible from general force fields like CGENF or GAFF. Our previous publications demonstrated that non-bonded force field parameters derived from the polarized electron density using Atom-In-Molecules (AIM) methods describe intermolecular interactions accurately.[7] We validated their accuracy for amino acid side chains with *ab initio* energy decomposition analysis and in the condensed phase predicting binding free energies in host–guest systems and protein–ligand interactions, as well as thermophysical properties of neutral organic liquids. Here, we used the previously validated minimal

basis iterative stockholder (MBIS) method[8] to calculate atomic charges and Lennard-Jones parameters, as detailed in the SI.

To create a custom force field for the transition state, we first determine the reaction mechanism and the minimum free energy path (MFEP) in the wild-type enzyme using the adaptive string method.[9] An initial guess of the transition state structure from the MFEP is further optimized at the QM level to locate the transition state on the potential energy surface, identified by a specific imaginary vibrational frequency. We then derive nonbonded and bonded force field parameters for this transition state structure. Bonded parameters are extracted from the Hessian matrix, adjusting the unique negative eigenvalue using the Q2MM protocol.[4] Covalent bond and angle parameters are calculated using a modified Seminario method,[10] while dihedral angles are maintained from the general force fields like CGENFF. Nonbonded parameters are obtained from the polarized electron density, as detailed in the SI.

Additionally, we also determine nonbonded force field parameters of the side chain of the mutating residue in the catalytic pocket using the AIM method *via* a QM/MM single-point calculation (see the SI). Finally, we combine the bespoke force fields of reactant and transition state with the CHARMM36m biomolecular force field[11] to achieve a comprehensive atomistic representation of the interactions of both states.

The hybrid system, containing both wild-type (wt) and variant residues, is constructed using the RS and TS structures with the PMX software.[6] After equilibrating both states for the wt and mutant enzyme, non-equilibrium alchemical transformations are performed in both directions. The alchemical free energies are then estimated using PMX. Further details on the validation of the alchemical free energies and non-equilibrium alchemical transformation are described in the SI.

To evaluate our protocol, we applied it to the dihydrofolate reductase (DHFR) enzyme that catalyzes the hydride transfer from nicotinamide adenine dinucleotide phosphate (NADPH) to dihydrofolate, producing tetrahydrofolate. DHFR has been extensively studied for over 40 years due to its role in various metabolic pathways and its significance as a target for anti-cancer and antibiotic drugs. Consequently, high-accuracy data on its protein structure, enzymatic properties, and phylogenetics are readily available. Additionally, numerous computational studies of DHFR, including conformational dynamics and mechanistic insights, provide reliable data, making it an ideal candidate for validating our workflow.

We applied the adaptive string method to determine the minimum free energy path (MFEP) and representative structures for the reactant and transition states. Fig. 3 shows the free energy profile obtained at the DFTB3(mod)/CHARMM36m level of theory (see SI), revealing an activation free energy barrier of 70.3 kJ mol^{-1}, which aligns with previously reported QM/MM studies.[12]

After characterizing the TS and RS, we derived bespoke force fields that describe atomic interactions in each state. This methodology assumes that the newly derived parameters are transferable across different variants. We studied three variants (I14V, I14A, and I14G), for which QM/MM and experimental
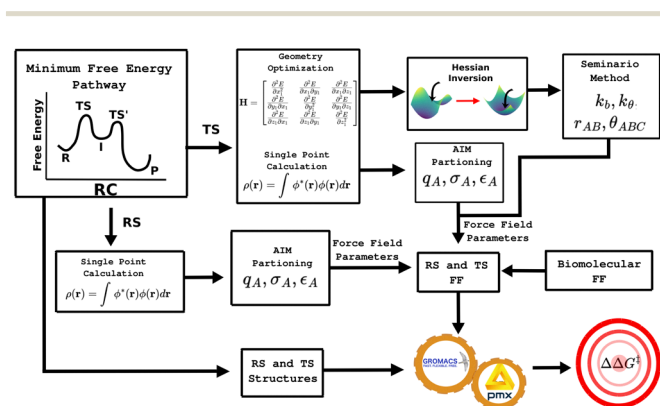

Fig. 2 The non-equilibrium alchemical transformation protocol to compute $\Delta\Delta G^{\ddagger}$ induced by protein mutation. The protocol starts extracting the transition state and the reactant structure from the MFEP. For the reactant state, non-covalent force field parameters are derived from the electron density of the substrates, invoking the AIM method. For the transition state the same process is performed in addition to the parametrization of covalent terms with an adapted Q2MM protocol.[4] The force field parameters and structures are used as inputs for GROMACS[5] and PMX[6] to calculate the difference in activation free energy.
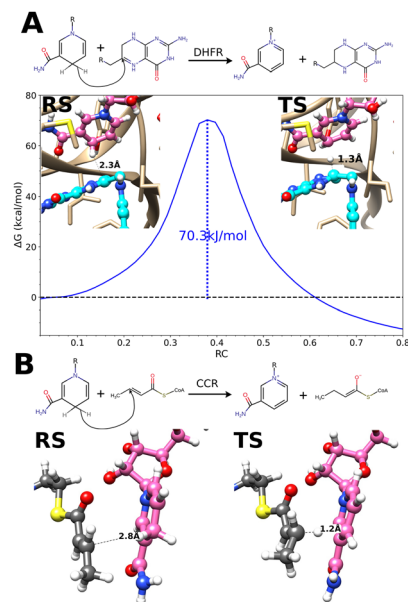
Fig. 3 (A) The hydride transfer reaction between NADPH and dihydrofolate catalyzed by DHFR is depicted together with the minimum free energy path derived from adaptive string simulations. The insets correspond to reactant (RS) and transition state (TS) structures emphasizing the donor–acceptor distance. (NADPH is illustrated in pink and tetrahydrofolate in cyan). (B) The hydride transfer reaction between NADPH and crotonyl-CoA catalyzed by CCr is shown together with the structures from the reactant and transition state with characteristic donor–acceptor distances at the bottom (the rest of the enzyme is occluded for clarity).

activation free energy values have been reported. The volume occupied by the side chain of the catalytic relevant residue 14 is systematically reduced from isoleucine to valine, alanine, or glycine, reducing the catalytic rate constant by at least four orders of magnitude compared to the wildt-ype. This slowdown corresponds to an increase of 16.7 kJ mol^{-1} in the activation energy barrier for the smallest residue, glycine, as illustrated in Table 1. Our method $\left(\Delta\Delta G_{MM}^{\ddagger}\right)$ exhibits the same trend, with

Table 1 Activation free energy difference in kJ mol^{-1} for five DHFR and three CCr variants. The experimental data (exp) were calculated with the Eyring equation using experimental catalytic rate constants. For DHFR, Quantum mechanics-molecular mechanics (QM/MM) free energy differences are compared to our method using molecular mechanics (MM) and MM with optimized non-bonded force field parameters for the mutated side chain (MM$_{opt}$), as for Ccr (errors of the mean were calculated from three independent simulations)

DHFR Mut	$\Delta\Delta G_{Exp}^{\ddagger}$	$\Delta\Delta G_{QM/MM}^{\ddagger}$	$\Delta\Delta G_{MM}^{\ddagger}$	$\Delta\Delta G_{MM_{opt}}^{\ddagger}$
I14G	16.7[12]	18.0	10.5 ± 0.9	11.7 ± 0.3
I14A	8.8[12]	12.6	4.6 ± 0.9	7.1 ± 0.7
I14V	4.6[12]	7.1	0.5 ± 0.6	0.2 ± 0.2
S148A	0.2[13]	—	0.5 ± 0.1	0.9 ± 0.1
D122N	7.5[14]	5.9 ± 0.3	2.4 ± 0.3	2.7 ± 0.1

Ccr Mut	$\Delta\Delta G_{Exp}^{\ddagger}$ [15]		$\Delta\Delta G_{MM}^{\ddagger}$	$\Delta\Delta G_{MM_{opt}}^{\ddagger}$
E171A	6.6		5.2 ± 1.0	7.4 ± 1.0
H365N	7.5		7.2 ± 1.0	8.8 ± 1.0
F170A	6.1		4.6 ± 1.0	2.6 ± 1.0

errors between 4.1 and 6.2 kJ mol^{-1}, where the largest error corresponds to I14G, which is the most perturbative mutation. In addition, we studied the effect of distal mutations of varying nature, including changes in net charge, polarity and side-chain size (S148A, D122N and Table S1). We observed the same perturbation dependent error with larger deviations from variants involving a change in net charge (D122N).

Doron et al.'s QM/MM data[12] outperforms our method, with errors between 1.3 and 3.8 kJ mol^{-1} effectively reflecting side chain size trends. However, the computational cost of umbrella sampling simulations with semi-empirical methods is much higher than our MM approach, which only took 48 hours on a standard 8-core CPU workstation with one GPU.

Recently, we've pointed out the limitations of the CHARMM36m biomolecular force field in accurately representing the ab initio derived Pauli repulsion and dispersion energies between side chains at close distance.[7] These issues are mostly resolved when Lennard-Jones parameters and atomic charges are determined using an atom-in-molecules strategy with the MBIS partitioning framework. Steric and dispersion interactions mainly affect the non-polar side chains mutated at residue 14, leading us to develop specific nonbonded parameters for each mutation. As shown in the right column of Table 1 $\left(\Delta\Delta G_{MM_{opt}}^{\ddagger}\right)$, this approach reduces the error to between 1.7 and 5.0 kJ mol^{-1} and outperforms the QM/MM method for the I14A variant. Furthermore, the optimized nonbonded parameters improve the prediction of I14G and I14A for which the change in the steric and dispersion energies is more pronounced. This improvement is less pronounced in charged systems (e.g. D122 variants).

To assess our approach's effectiveness further, we applied it to the rate limiting step of crotonyl-CoA carboxylase/reductase (CCR), the fastest CO$_2$ fixing enzyme. CCR differs from DHFR as it is a multi-domain enzyme with a multi step mechanism that involves conformational changes. Recently, we demonstrated that the reduction of crotonyl-CoA by NADPH is the rate limiting step (see Fig. 3(B)), resulting in a reactive enolate that binds CO$_2$ to produce ethylmalonyl-CoA.[15,16] We previously measured the catalytic rate constant for CO$_2$ fixation for the native enzyme and three variants: E171A, H365N, and F170A. Table 1 summarizes the calculated activation free energy differences for each variant from the experimental data.

Prediction of $\Delta\Delta G^{\ddagger}$ in Ccr is more complex than DHFR due to mutations that change protonation states (H365N), alter side chain polarity (E171A), or reduce size (F170A). Using the TS and RS structures from prior work,[16] we applied our method to derive force field parameters for both states (see SI).

In our previous studies, we demonstrated that Glu171's proximity to His365 stabilizes its protonated form. Consequently, the E171A mutation alters the protonation state of His365 from double to single protonated, thus preserving the system's overall charge. Table 1 shows the activation free energy difference derived from our non-equilibrium alchemical transformations at the MM level with a mere 1.4 kJ mol^{-1} deviation from the experimental benchmark and 0.8 kJ mol^{-1}

error when optimizing the nonbonded force field parameters of the mutated side chain (MM_{opt}). The H365N mutation poses a significant challenge due to the alteration of the system's net charge. To circumvent artifacts arising from the particle mesh Ewald method used for long-range electrostatics with periodic boundary conditions, we employ a double system-single box strategy, wherein a tripeptide (GNG) is incorporated into the box and undergoes transformation in the opposite direction of the CCR enzyme (H → N). For this mutation, Table 1 reveals an error of just 0.3 kJ mol^{-1} for the standard force field and 1.3 kJ mol^{-1} for the optimized side chain version. The F170A mutation parallels those in DHFR discussed earlier, as the side chain volume decreases, shifting from aromatic to aliphatic interactions. This results in a more significant system perturbation, possibly accounting for the slightly larger error of 1.5 kJ mol^{-1} and 3.5 kJ mol^{-1} for the MM and MM_{opt} force fields, respectively. It's crucial to note that despite CCR involving charged residues where non-covalent interactions are more potent than in DHFR systems, the errors remain below the threshold of chemical accuracy (4.184 kJ mol^{-1}) for each mutation.

In summary, we present a computational method that is both affordable and accurate for rational enzyme design, aimed at efficiently calculating the changes in activation free energy barriers resulting from electrostatic, dispersion and steric perturbations induced by mutations. Our physics-based technique is grounded in non-equilibrium alchemical transformations, providing a fast, automated, and reliable alternative for free energy calculations. To perform these calculations, the system's interactions are represented using bespoke force fields suitable for both transition and reactant states. Nonbonded interaction parameters are derived using an AIM method, while bonded interactions rely on the Seminario method. The low computational cost of molecular mechanics allows high-throughput analysis of multiple enzyme mutations for one reaction, assuming a conserved transition state structure, thereby facilitating efficient enzyme design.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the SI. See DOI: **https://doi.org/10.1039/d5cc03075g**.

The data supporting this article have also been included in the following Zenodo link for download: **https://doi.org/10.5281/zenodo.15554552**.

Notes and references

1 D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik and D. Baker, *Nature*, 2008, **453**, 190–195.
2 F. H. Arnold, *Angew. Chem., Int. Ed.*, 2018, **57**, 4143–4148.
3 M. Aldeghi, B. L. de Groot and V. Gapsys, *Computational Methods in Protein Evolution*, Springer, 2019, pp. 19–47.
4 A. R. Rosales, T. R. Quinn, J. Wahlers, A. Tomberg, X. Zhang, P. Helquist, O. Wiest and P.-O. Norrby, *Chem. Commun.*, 2018, **54**, 8294–8311.
5 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
6 V. Gapsys, S. Michielssens, D. Seeliger and B. L. De Groot, *J. Comput. Chem.*, 2015, **36**, 348–354.
7 C. Castillo-Orellana, F. Heidar-Zadeh and E. Vöhringer-Martinez, *J. Chem. Theory Comput.*, 2025, **21**, 2043–2054.
8 T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier and P. W. Ayers, *J. Chem. Theory Comput.*, 2016, **12**, 3894–3912.
9 K. Zinovjev and I. Tuñón, *J. Phys. Chem. A*, 2017, **121**, 9764–9772.
10 A. E. A. Allen, M. C. Payne and D. J. Cole, *J. Chem. Theory Comput.*, 2018, **14**, 274–281.
11 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell, *Nat. Methods*, 2017, **14**, 71–73.
12 D. Doron, V. Stojković, L. Gakhar, A. Vardi-Kilshtain, A. Kohen and D. T. Major, *J. Phys. Chem. B*, 2015, **119**, 906–916.
13 K. F. Wong, T. Selzer, S. J. Benkovic and S. Hammes-Schiffer, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6807–6812.
14 A. R. Mhashal, Y. Pshetitsky, R. Eitan, C. M. Cheatum, A. Kohen and D. T. Major, *J. Phys. Chem. B*, 2018, **122**, 8006–8017.
15 G. M. M. Stoffel, D. A. Saez, H. DeMirci, B. Vögeli, Y. Rao, J. Zarzycki, Y. Yoshikuni, S. Wakatsuki, E. Vöhringer-Martinez and T. J. Erb, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 13964–13969.
16 R. Recabarren, M. Tinzl, D. A. Saez, A. Gomez, T. J. Erb and E. Vöhringer-Martinez, *ACS Catal.*, 2023, **13**, 6230–6241.