Supporting Information: Structural and Catalytic Consequences of Active-Site vs. Distal Mutations in Human Dehalogenase: Insights from Molecular Dynamics Simulations

Soumyajit Karmakar, Biman Giri, and Sabyashachi Mishra*

Department of Chemistry, Indian Institute of Technology Kharagpur, 721302, Kharagpur, West Bengal, India

E-mail: mishra@chem.iitkgp.ac.in

Phone: +91-3222-282328

Running header

Table S1: Extra-point (EP) parameters applied to I-Tyr analog in the present work.

Parameters	I-Tyr
Mass (m)	0.00
Radius (R)	1.00 Å
ϵ	0.00
r_{eq} (X-EP)	2.15 Å
k_r (X-EP)	600.0 kcal·mol^{-1}·Å^{-2}
θ_{eq} (C-X-EP)	180.0°
k_θ (C-X-EP)	150.0 kcal·mol^{-1}·rad^{-2}
γ (C-C-X-EP)	0.00
V_n (C-C-X-EP)	0.00

Table S2: Comparison of predicted pKa values for acidic, neutral, and basic residues using PROPKA3 and H++ ($\mathrm{pK}_{1/2}$).

Res	ID	PROPKA	H++	Res	ID	PROPKA	H++
ASP	117	5.06	4.93	ASP	142	3.37	1.68
ASP	144	3.92	3.51	ASP	172	3.93	3.27
ASP	186	2.54	0.47	ASP	276	3.17	2.52
ASP	283	3.75	4.32	GLU	72	3.35	3.43
GLU	83	4.58	3.43	GLU	85	4.42	3.01
GLU	92	4.67	4.83	GLU	95	4.01	4.80
GLU	109	4.60	5.24	GLU	114	4.73	3.94
GLU	133	5.41	3.20	GLU	154	3.79	0.30
GLU	155	4.38	4.40	GLU	156	3.18	0.24
GLU	157	5.51	1.45	GLU	158	4.87	5.62
GLU	183	4.68	3.03	GLU	214	4.15	0.94
GLU	257	3.88	< 0	GLU	271	3.96	3.78
CYS	221	12.61	> 12	CYS	243	12.59	> 12
TYR	81	10.38	11.73	TYR	94	10.60	11.20
TYR	161	15.50	> 12	TYR	184	14.77	> 12
TYR	211	13.35	11.26	TYR	212	11.05	9.41
TYR	267	11.63	> 12	HIS	73	5.58	6.47
HIS	78	4.51	5.82	HIS	80	6.16	6.37
HIS	131	5.95	5.95	HIS	147	6.85	6.91
HIS	167	6.23	4.90	HIS	199	5.74	3.65
HIS	210	4.09	1.05	HIS	256	5.43	4.68
HIS	290	5.94	< 0	LYS	84	10.59	10.56
LYS	88	10.81	10.71	LYS	99	10.33	10.54
LYS	141	10.25	10.63	LYS	146	8.99	> 12
LYS	148	10.62	10.46	LYS	151	10.41	10.75
LYS	163	10.31	9.57	LYS	174	10.30	9.54
LYS	175	9.30	10.48	LYS	182	10.34	8.75
LYS	196	10.78	11.17	LYS	206	10.30	10.09
LYS	207	10.16	9.57	LYS	208	9.84	9.88
LYS	258	10.10	9.10	LYS	270	10.52	10.91
LYS	278	11.01	10.30	LYS	280	9.06	9.94
ARG	89	12.52	> 12	ARG	100	14.65	8.55
ARG	101	12.56	> 12	ARG	104	11.43	10.55
ARG	121	12.33	> 12	ARG	150	14.19	> 12
ARG	164	13.25	> 12	ARG	168	12.55	> 12
ARG	177	12.48	> 12	ARG	246	13.57	> 12
ARG	248	12.17	11.92	ARG	253	13.59	> 12
ARG	279	12.14	> 12				

Theoretical Background of Wako–Saitô–Muñoz–Eaton (WSME) Model In the WSME model, each residue is assigned a binary folding state: 1 for a folded–like conformation and 0 for an unfolded–like conformation. Each of the 2^N possible microstates corresponds to a particular pattern of ordered and disordered residues. A contiguous folded segment of residues from m to n is associated with a free energy[S1,S2]

$$F_{m,n} = G^{\text{stab}}_{m,n} - TS_{m,n}, \tag{S1}$$

where $S_{m,n}$ is the entropic penalty and $G^{\text{stab}}_{m,n}$ is the stabilization due to van der Waals, electrostatic, and solvation terms[S1,S2].

$$G^{\text{stab}}_{m,n} = E_{\text{vdW}} + E_{\text{elec}} + G_{\text{solv}}. \tag{S2}$$

E_{vdW} is estimated from the number of heavy-atom contacts within 6 Å (46 J mol^{-1}) per contact[S3]. The electrostatic stabilization is modeled using the Debye–Hückel potential[S4,S5],

$$E_{\text{elec}} = K_{\text{Coulomb}} \sum_{i,j} \frac{q_i q_j}{\epsilon_{\text{eff}}\, r_{ij}} \exp(-\kappa r_{ij}), \tag{S3}$$

where q_i and q_j are the charges of residues i and j, r_{ij} is their separation, $\epsilon_{\text{eff}} = 29$ is the effective dielectric constant[S5,S6], and κ is the inverse Debye length determined by the solvent ionic strength ($I = 0.15$ M). The solvation free energy is given by,

$$G_{\text{solv}} = x^{\text{contact}}_{m,n}\; C^{\text{contact}}_{p} \left[(T - T_{\text{ref}}) - T \ln(T/T_{\text{ref}})\right]. \tag{S4}$$

Here, $x^{\text{contact}}_{m,n}$ is the number of native contacts in the segment and C^{contact}_{p} is the heat capacity change per native contact, taken as 0.358 $Jmole^{-1}K^{-1}$ [S2] and $T_{\text{ref}} = 385$ K[S5]. Each contiguous ordered segment contributes a statistical weight $z_{m,n} = \exp[-\beta F_{m,n}]$, with $\beta = (RT)^{-1}$. The exact partition function sums over all 2^N microscopic configurations that can be expressed as a sum over collections of non–overlapping ordered segments,

$$Z_N = \sum_{\mathcal{C} \in \mathcal{S}_N} \prod_{(m,n)\in\mathcal{C}} z_{m,n}, \tag{S5}$$

Since direct evaluation of Z_N is prohibitively expensive for large N, we used approximate schemes, which include the single–sequence approximation (SSA), where only configurations containing a single contiguous ordered segment are retained[S7,S8], i.e.,

$$Z_{\text{SSA}} = \sum_{1 \le m \le n \le N} z_{m,n},$$

the double–sequence approximation (DSA) includes two non–overlapping ordered segments,

$$Z_{\text{DSA}} = \sum_{1 \le m_1 \le n_1 < m_2 \le n_2 \le N} z_{m_1,n_1}\, z_{m_2,n_2},$$

and DSA with inter-segment loop (DSAw/L) of length $\ell = m_2 - n_2 + 1$,

$$Z_{\mathrm{DSAw/L}} = \sum_{\substack{1 \le m_1 \le n_1 < m_2 \le n_2 \le N \\ \text{interacting}}} \exp\Big[-\beta F_{(m_1,n_1)\cup(m_2,n_2)}\Big]\; z_L^{\ell}. \tag{S6}$$

Here, z_L is the entropic penalty for the disordered loop of length ℓ. Only pairs of segments that form at least one native interaction across the loop contribute. The corresponding partition function reads The total partition function is evaluated as

$$Z = Z_{\mathrm{SSA}} + Z_{\mathrm{DSA}} + Z_{\mathrm{DSAw/L}}, \tag{S7}$$

from which, the folding free energy is obtained as $\Delta G_{\mathrm{fold}} = -RT \ln Z$, and residue–level folding probabilities are evaluated from normalized statistical weights of all configurations. The parameters used in this model are tabulted below.

Table S3: WSME model parameters used to study protein folding in the present work.

Parameters	Values
Cutoff used for heavy-atom contacts (r_c)	6 Å
Average mean–field energy (ξ)	46 J mol^{-1}
Effective dielectric constant (ϵ_{eff})	29
Coulumb constant (K_{Coulomb})	1389 kJ Å mol^{-1}
Heat capacity change per native contact (C_p^{contact})	0.358 J mol^{-1} K^{-1}
Reference temperature (T_{ref})	385 K

Table S4: The residue-wise nomenclature of the secondary structural elements of the hIYD enzyme as used in the present work (PDB ID: 4TTC).

α-helices			**Loops**		
No.	Element	ID range	No.	Element	ID range
1	α1	82–99	loop I	α1–α2 loop	100–111
2	α2	112–124	loop II	α2–α3 loop	125–128
3	α3	129–132	loop III	α4–α5 loop	164–166
4	α4	143–163	loop IV	α5–α6 loop	178–183
5	α5	167–177	L-Dynamic	Dynamic loop	198–212
6	α6	184–188	loop V	α7–α8 loop	232–242
7	α7	213–231			
8	α8	243–251			

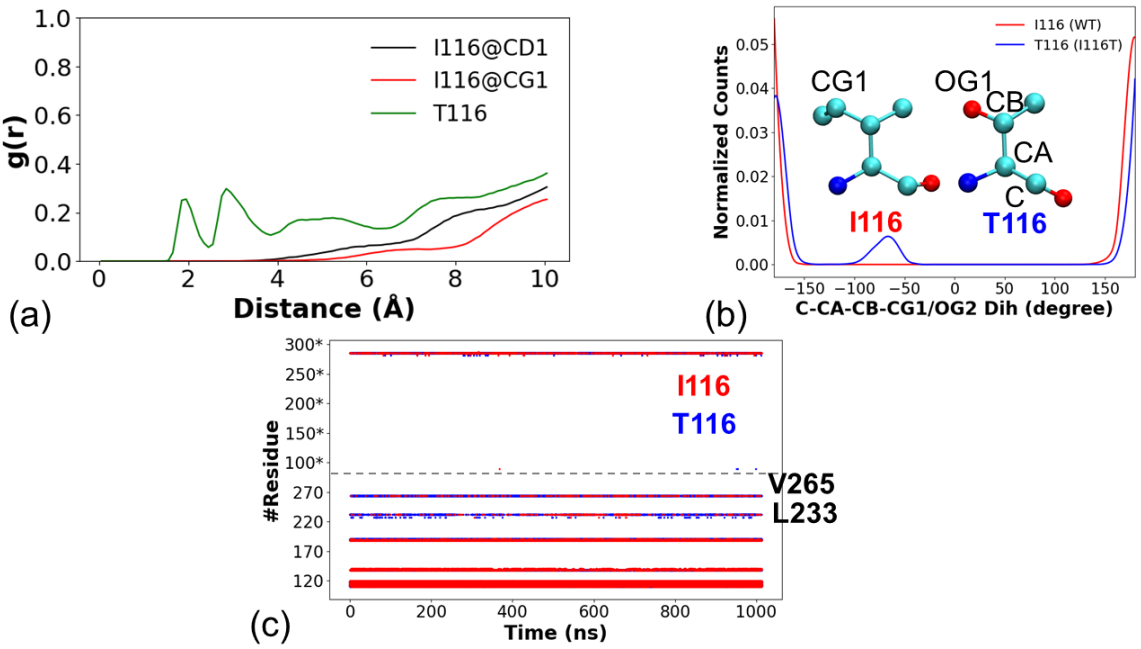

Figure S1: (a) Radial distribution function (g(r)) of water around I116 (wt, black/red for CD1 and CG1) and T116 (mutant, green). The mutant shows pronounced hydration peaks at 2–3 Å, indicating direct water contacts absent in the hydrophobic wt environment. (b) Sidechain dihedral angle distribution of I116 (wt, red) versus T116 (mutant, blue). The hydroxyl–bearing threonine sidechain samples broader conformations compared to the locked rotamer of isoleucine, introducing local flexibility. (c) Contact map of I116 (red) versus T116 (blue) over the MD trajectory. The mutant exhibits altered and more transient residue contacts, notably with V233 and L265 (labels), consistent with increased hydration and flexibility.

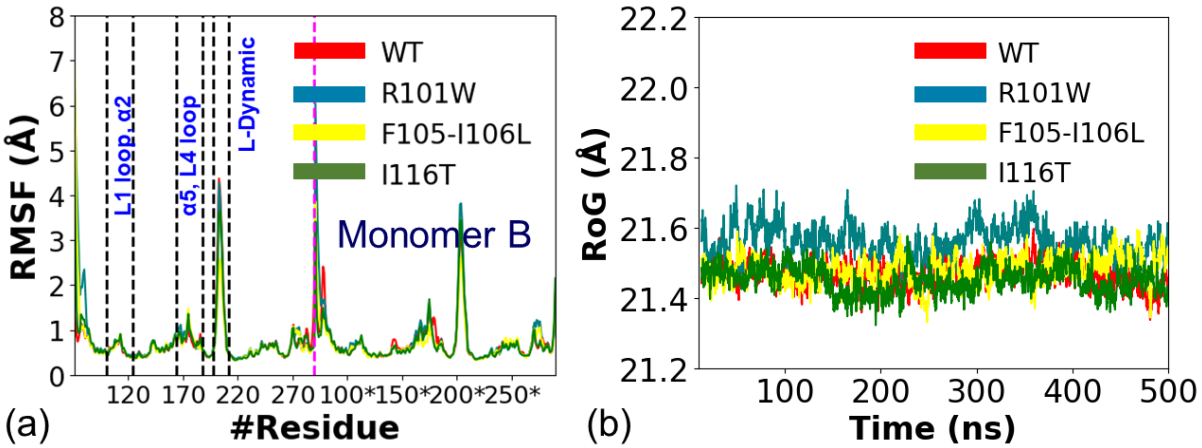

Figure S2: (a) The backbone (C_α, C, and N) RMSF of the wild-type and three mutant systems, (b) The time-evolution of the radius of gyration of the four systems.

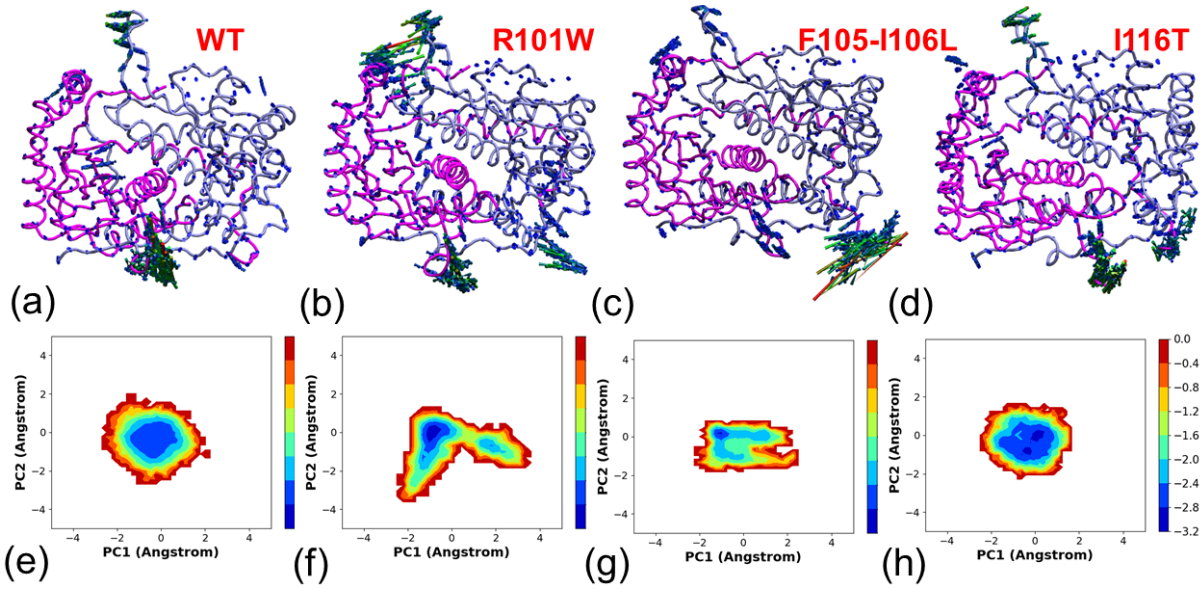

Figure S3: The principal component analysis of wild-type and mutants (R101W, F105–I106L, and I116T). (a-d) The atomic displacements C_α atoms along the first four principal components (PC1 to PC4) (having $>$ 5 % contribution) derived from the filtered trajectory. The arrows represent the direction and magnitude of motion along conformations, highlighting dominant collective movements associated with the principal modes. Both monomers are colored in magenta and iceblue, respectively. (e-h) The free energy landscape along the first two principal components for wild-type and mutants.

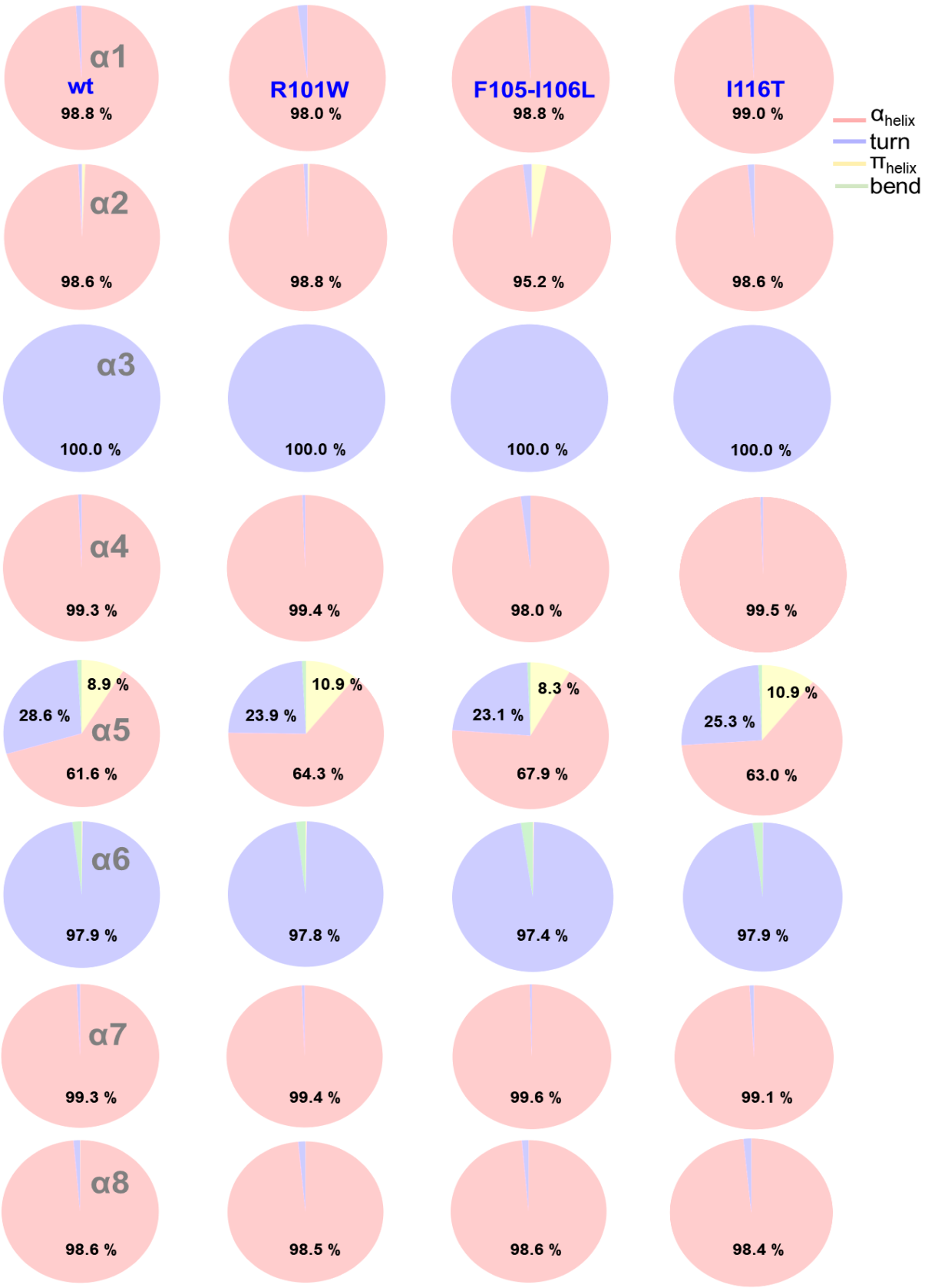

Figure S4: Average secondary structure composition of α1-α8 helices from MD simulations of wt and mutants, showing the mean fractional occupancy of each structure type across all frames.

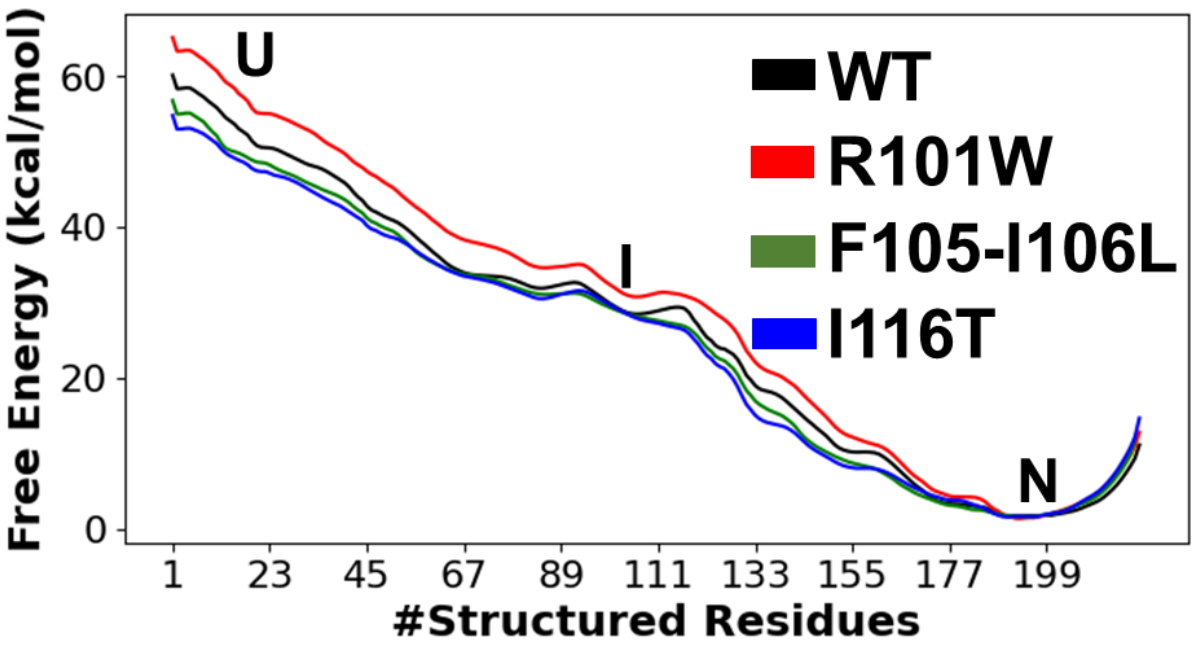

Figure S5: Predicted one–dimensional free–energy profile as a function of reaction coordinate, the number of structured residues. The four systems are shown in black (wt), red (R101W), green (F105–I106L), and blue (I116T). The unfolded (U), intermediate (I), and native states (N) are denoted.

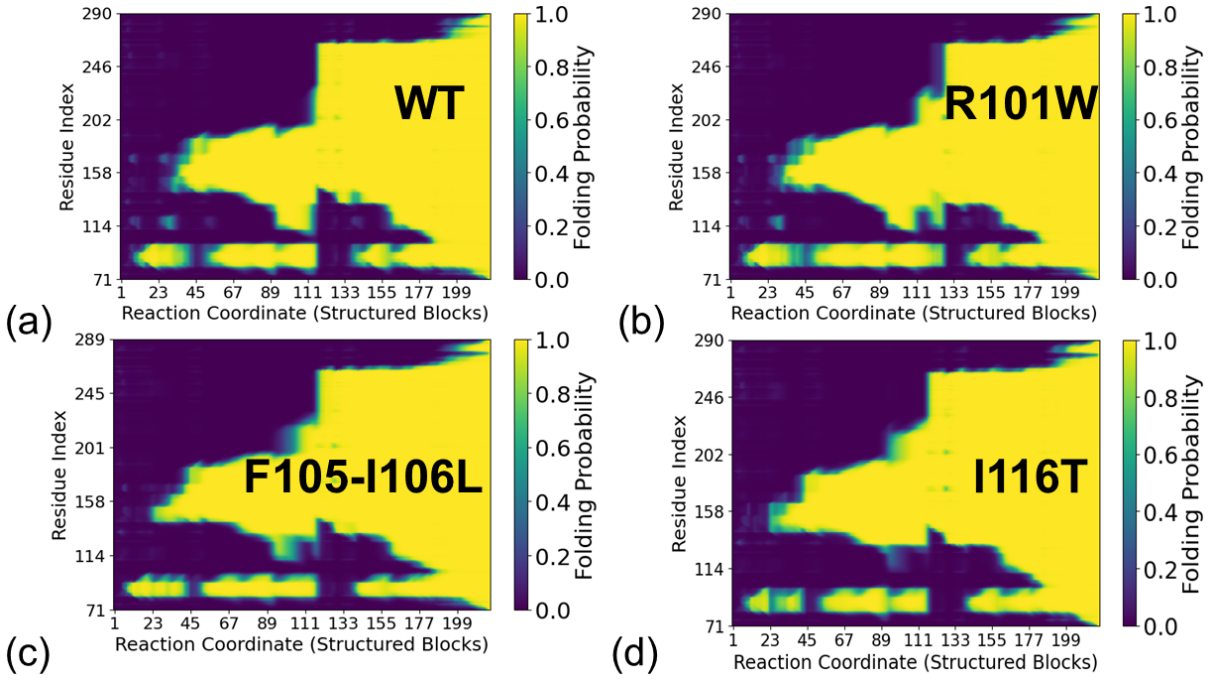

Figure S6: Folding probability curves of (a) wt, (b) R101W, (c) F105–I106L, and (d) I116T mutants.

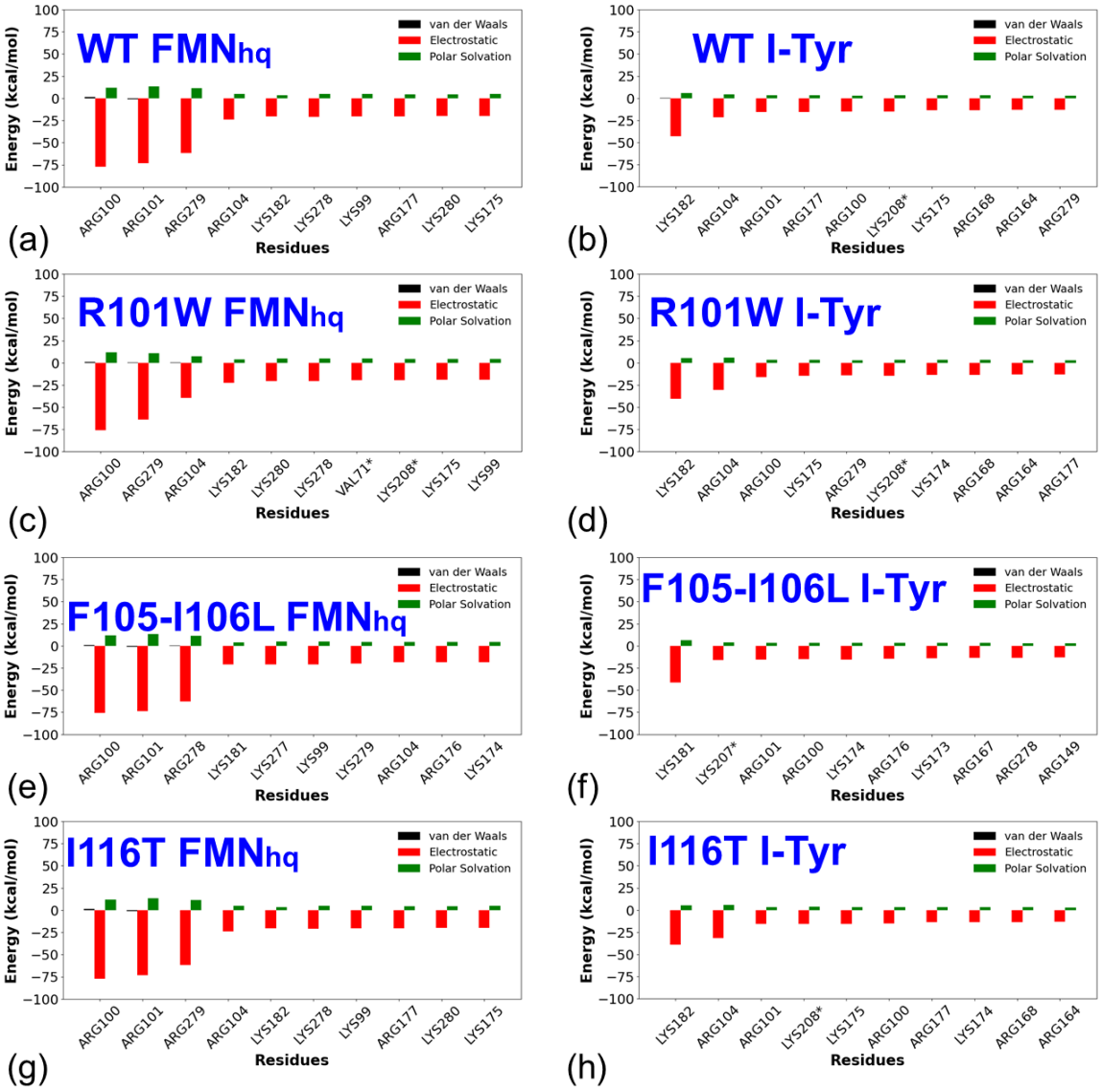

Figure S7: Residue-wise decomposition of binding free energies for $\mathrm{FMN_{hq}}$ and I-Tyr in (a)–(b) wt, (c)–(d) R101W, (e)–(f) F105–I106L, and (g)–(h) I116T hIYD. The left column represents $\mathrm{FMN_{hq}}$ and the right column represents I-Tyr.

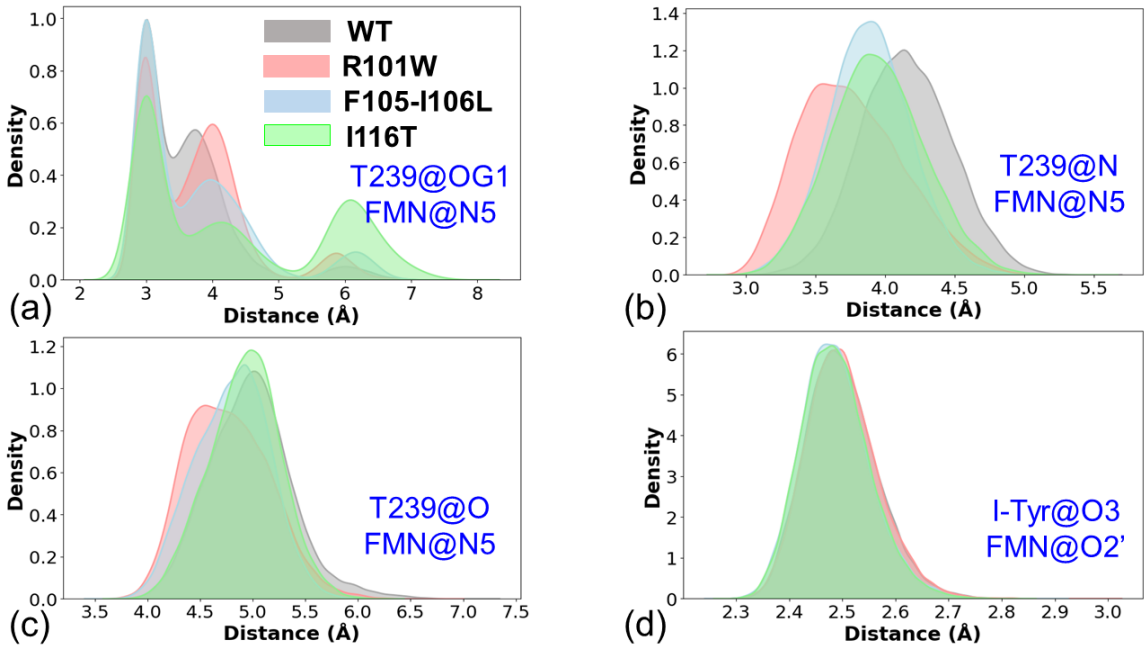

Figure S8: The distance distribution of interactions ((a)OG1, (b)N, and (c)O atoms) between T239 and N5 atom of $\mathrm{FMN_{hq}}$ in wild-type and mutants (R101W, F105–I106L, and I116T). (d)The distance distribution of recognition interaction between $\mathrm{FMN_{hq}}$ state's 2′-OH and phenolate oxygen of I-Tyr.

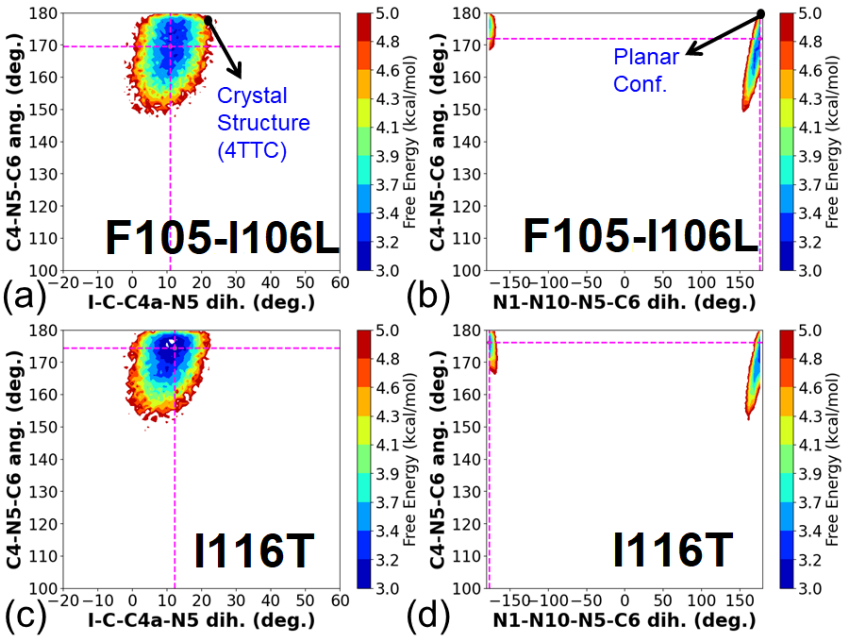

Figure S9: The free energy profile of C4-N5-C6 angle and I-C-C4a-N5 dihedral angle of $\mathrm{FMN_{hq}}$ and I-Tyr in the (a) F105–I106L, (c) I116T mutants. Free energy profile of C4-N5-C6 angle and N1-N10-N5-C6 dihedral angle in the (b) F105–I106L, (d) I116T mutants. The crystal structure values of these dihedrals are 22^0 and 177.2^0, respectively, as shown by a black dot in the figure. The isoalloxazine planar conformation as present in the crystal structure (PDB ID: 4TTC), is also shown.

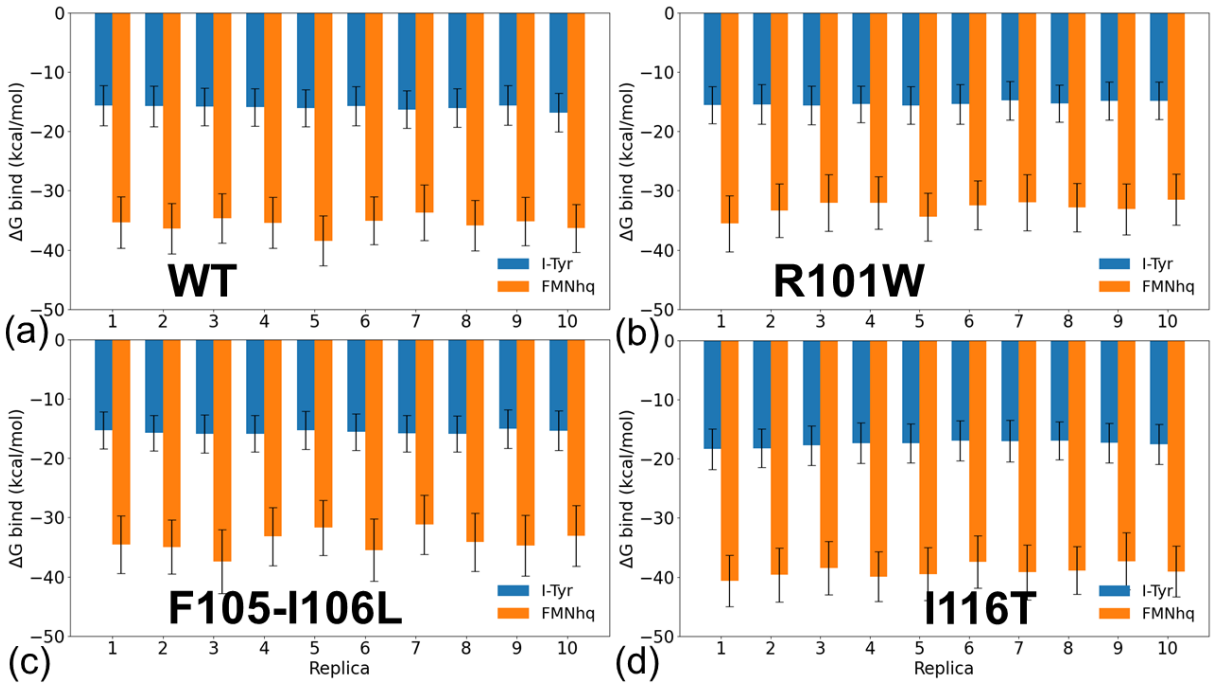

Figure S10: MM–PBSA binding free energy (ΔG_{bind}) estimates for I-Tyr (blue) and $\mathrm{FMN_{hq}}$ (orange) across ten independent replicas for the wild-type protein and the R101W, F105–I106L, and I116T variants. Bars represent mean values, and error bars indicate standard deviations, illustrating the reproducibility and variability of the MM–PBSA calculations across replicates.

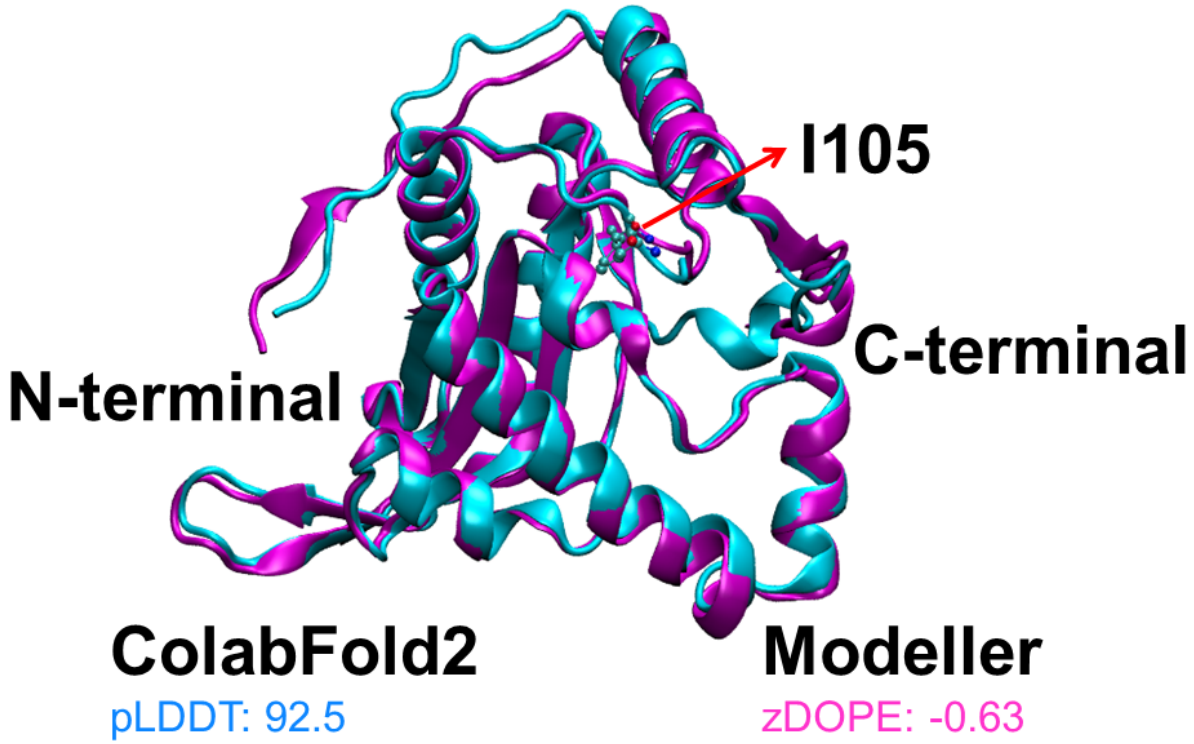

Figure S11: Homology modeling of the deletion mutant (F105–I106L) using Modeller produced a high–quality structural model (magenta) based on template PDB 3tnz chain A, with 90 % sequence identity and strong model quality metrics (GA341 = 1.0, zDOPE = -0.63, TSVMOD RMSD = 0.874 Å). Independent structure prediction with ColabFold2 (cyan) is supported by the model, yielding a high mean pLDDT score of 92.5, indicative of very high local confidence. Structural comparison between the Modeller and ColabFold2 models shows close agreement, with a backbone RMSD of 0.6 Å over the overlapping region (residues 82–275), demonstrating strong convergence between template–based and deep–learning–based approaches. Together, these results indicate that the deletion mutant adopts a well–defined and robust fold, and that the modeled structure is suitable for downstream structural and functional analyses.

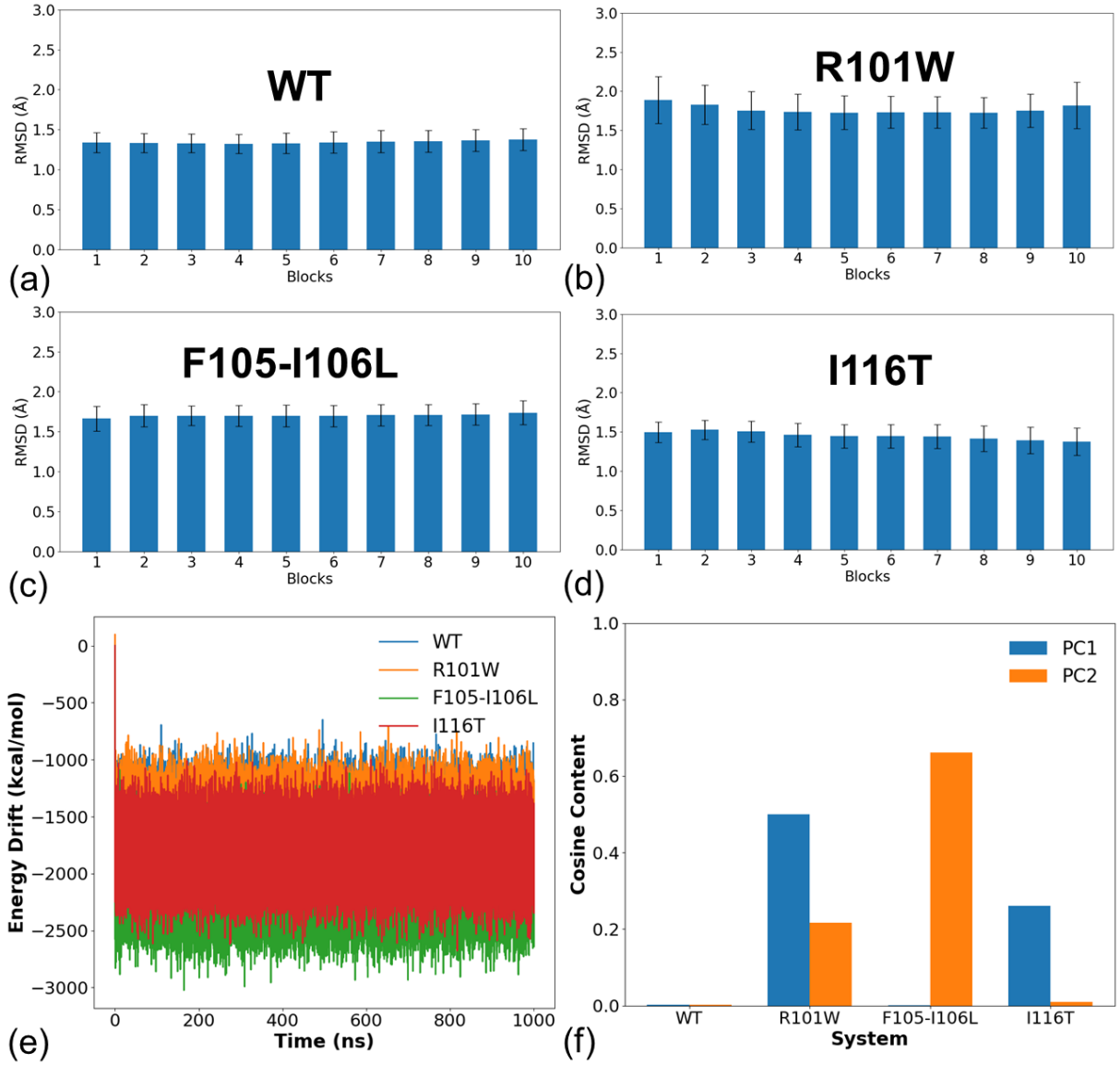

Figure S12: Block-averaged backbone RMSD analysis for (a) WT, (b) R101W, (c) F105–I106L, and (d) I116T systems. The RMSD values (Å) are shown for trajectory blocks, with error bars representing the standard deviation within each block. This analysis highlights the structural stability and convergence behavior of the WT and mutant systems over the course of the simulations. (e) Energy drift as a function of simulation time for WT, R101W, F105–I106L, and I116T systems, demonstrating stable energy behavior throughout the trajectories. (f) Cosine content of the first two principal components (PC1 and PC2) for all systems, used to assess the convergence and physical relevance of the principal component analysis. Low cosine content (< 0.7) indicates well–converged collective motions along the corresponding principal component.

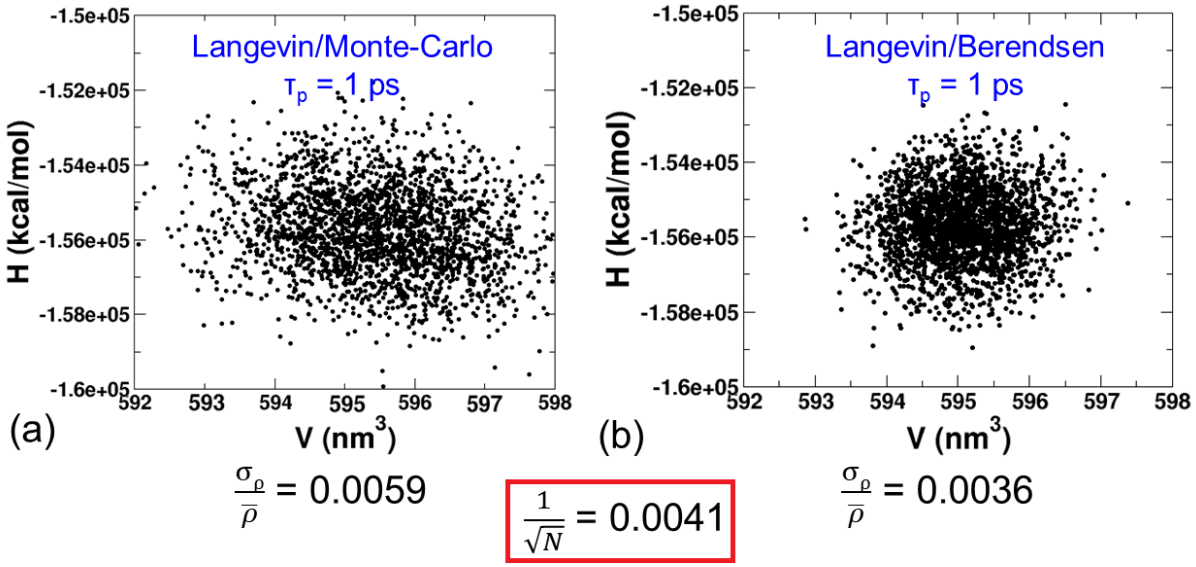

Figure S13: Joint enthalpy–volume (H–V) distributions obtained from NPT simulations using (a) a Langevin thermostat with Monte–Carlo barostat and (b) a Langevin thermostat with Berendsen barostat (τ_p = 1 ps). Both simulations sample stable and well–defined volume and enthalpy distributions. The relative density fluctuation is 0.0059 for the Monte–Carlo barostat and 0.0036 for the Berendsen barostat; the expected finite–size statistical fluctuation $1/\sqrt{N}$ = 0.0041 is shown for comparison.

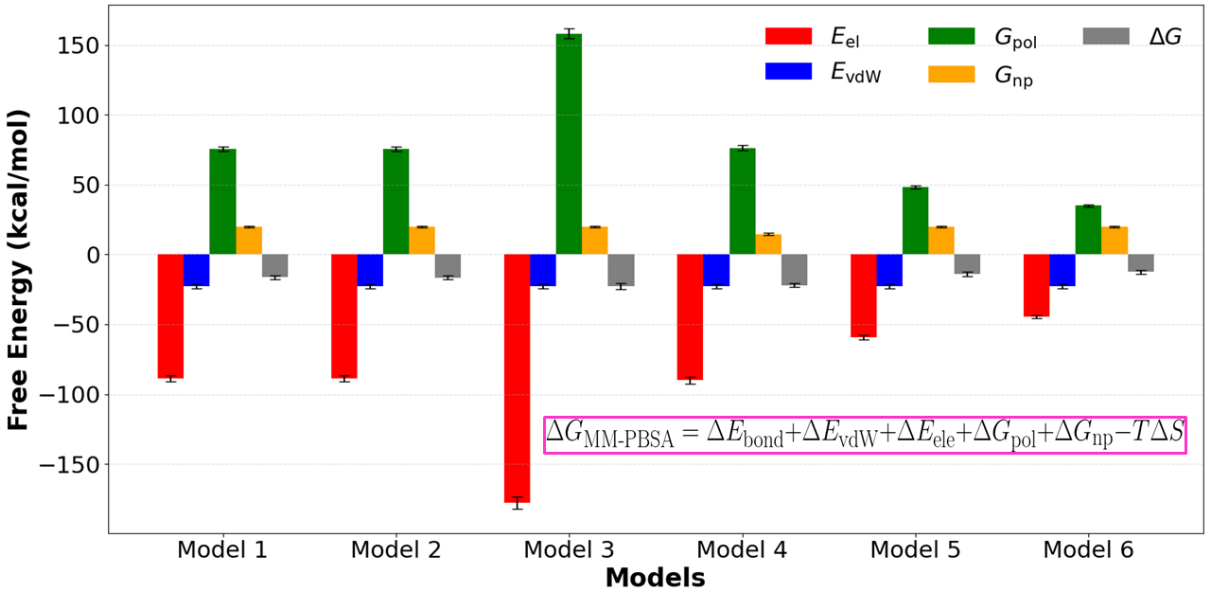

Figure S14: (a) Sensitivity analysis of MM–PBSA binding free energies for the R101W mutant obtained by varying the internal dielectric constant, ionic strength, and PB radii while using the same MD trajectory. The results demonstrate strong sensitivity to the dielectric constant and PB radii, weak dependence on ionic strength, and substantial cancellation between electrostatic and polar solvation contributions.

Systematic sensitivity analysis of the MM–PBSA binding free energy: We tested 6 models. Using an internal dielectric constant of 4.0, AMBER radii, and an ionic strength of 0.15 M (Model 1), the binding free energy was $\Delta G = -16.4 \pm 1.5$ kcal/mol, which remained essentially unchanged upon reducing the ionic strength to 0.10 M (Model 2, $\Delta G = -16.5 \pm 1.5$ kcal/mol), indicating a weak sensitivity to bulk electrostatic screening. In contrast, lowering the internal dielectric constant to 2.0 (Model 3) resulted in a substantially more favorable binding free energy ($\Delta G = -22.7 \pm 2.2$ kcal/mol), driven by a pronounced enhancement of the electrostatic interaction energy (E_{el}) that was only partially compensated by an increased polar solvation penalty (G_{pol}). Changing the PB radii from the AMBER set to the precomputed values while maintaining $\varepsilon = 4.0$ and an ionic strength of 0.15 M (Model 4) led to a noticeable stabilization of the complex ($\Delta G = -21.9 \pm 1.6$ kcal/mol), reflecting a moderate sensitivity to solvation boundary definitions. This also required the removal of the extra-point charge from the simulation, for which no precomputed parameters are available. Increasing the internal dielectric constant to 6.0 (Model 5) and 8.0 (Model 6) progressively weakened electrostatic stabilization and yielded less favorable binding free energies of $\Delta G = -14.0 \pm 1.4$ kcal/mol and $\Delta G = -12.6 \pm 1.5$ kcal/mol, respectively. Across all models, van der Waals and nonpolar solvation contributions remained comparatively invariant. The observed sensitivity trends are fully consistent with the benchmarking MM-PBSA studies of Ryde et. als (*Expert Opin. Drug Discov.*, 2015, 10, 5, 449-461), which show the dominant role of electrostatics, the partial cancellation between E_{el} and G_{pol}, and the strong dependence of predicted binding affinities on dielectric and solvation model parameters.

Table S5: $\Delta\Delta G$ of binding of I-Tyr with R101W mutant relative to that with wt-hIYD (i.e., $\Delta G(\text{R101W}) - \Delta G(\text{WT})$) for six different models of MM-PBSA.

Model	**Dielectric (ε)**	**Ionic Strength (M)**	**PB Radii Set**	$\Delta\Delta G$ **(kcal/mol)**
Model 1	4.0	0.15	AMBER	$+0.34$
Model 2	4.0	0.10	AMBER	$+0.42$
Model 3	2.0	0.15	AMBER	-0.55
Model 4	4.0	0.15	PB (precomputed)	-1.15
Model 5	6.0	0.15	AMBER	$+0.61$
Model 6	8.0	0.15	AMBER	$+0.78$

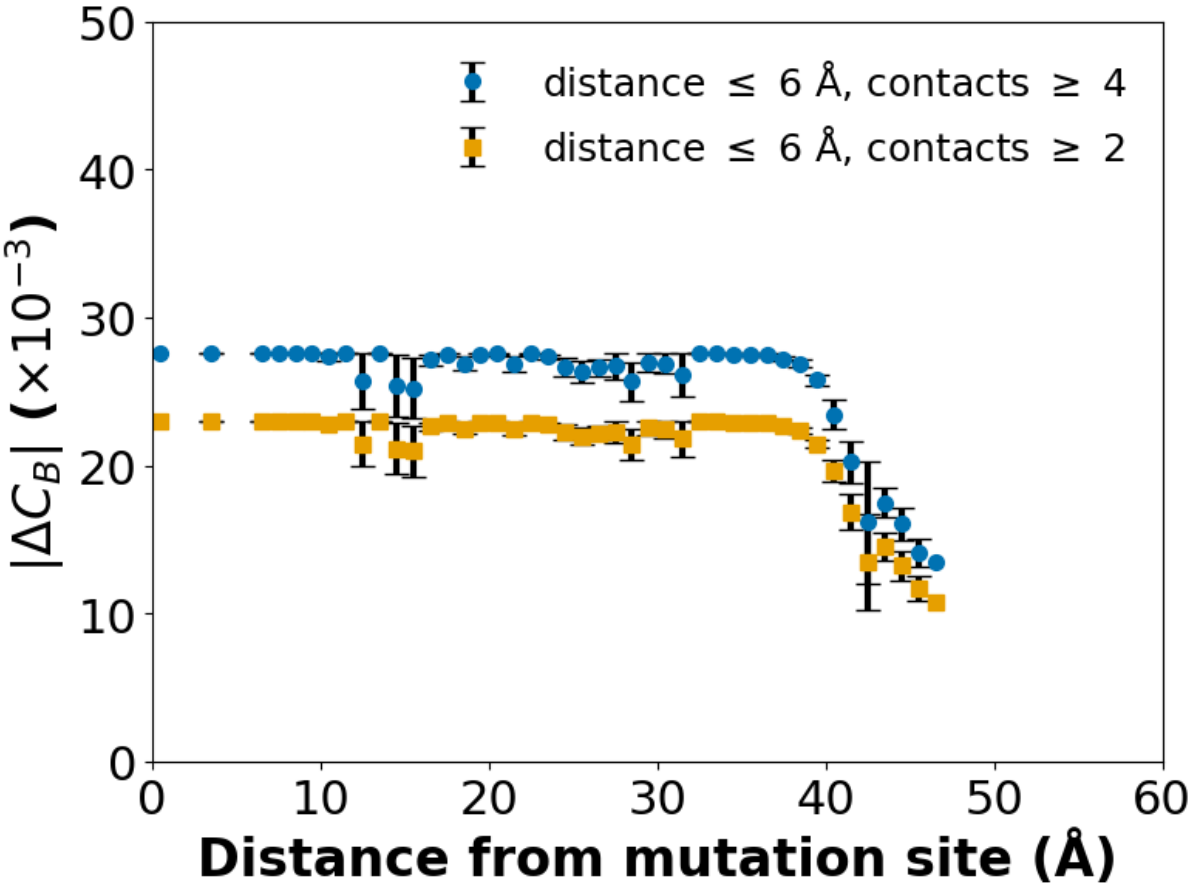

Figure S15: Distance dependence of changes in betweenness centrality, ΔCB (R101W mutant - wt), as a function of the distance from the mutation site. Residue interaction networks were constructed from 1 μs MD trajectories using a distance cutoff of 6 Å. Blue circles correspond to residue pairs defined by a stringent contact criterion ($\geq$ 4 contacts), while orange squares show results obtained with a relaxed criterion ($\geq$ 2 contacts). Error bars represent the standard error of the mean.

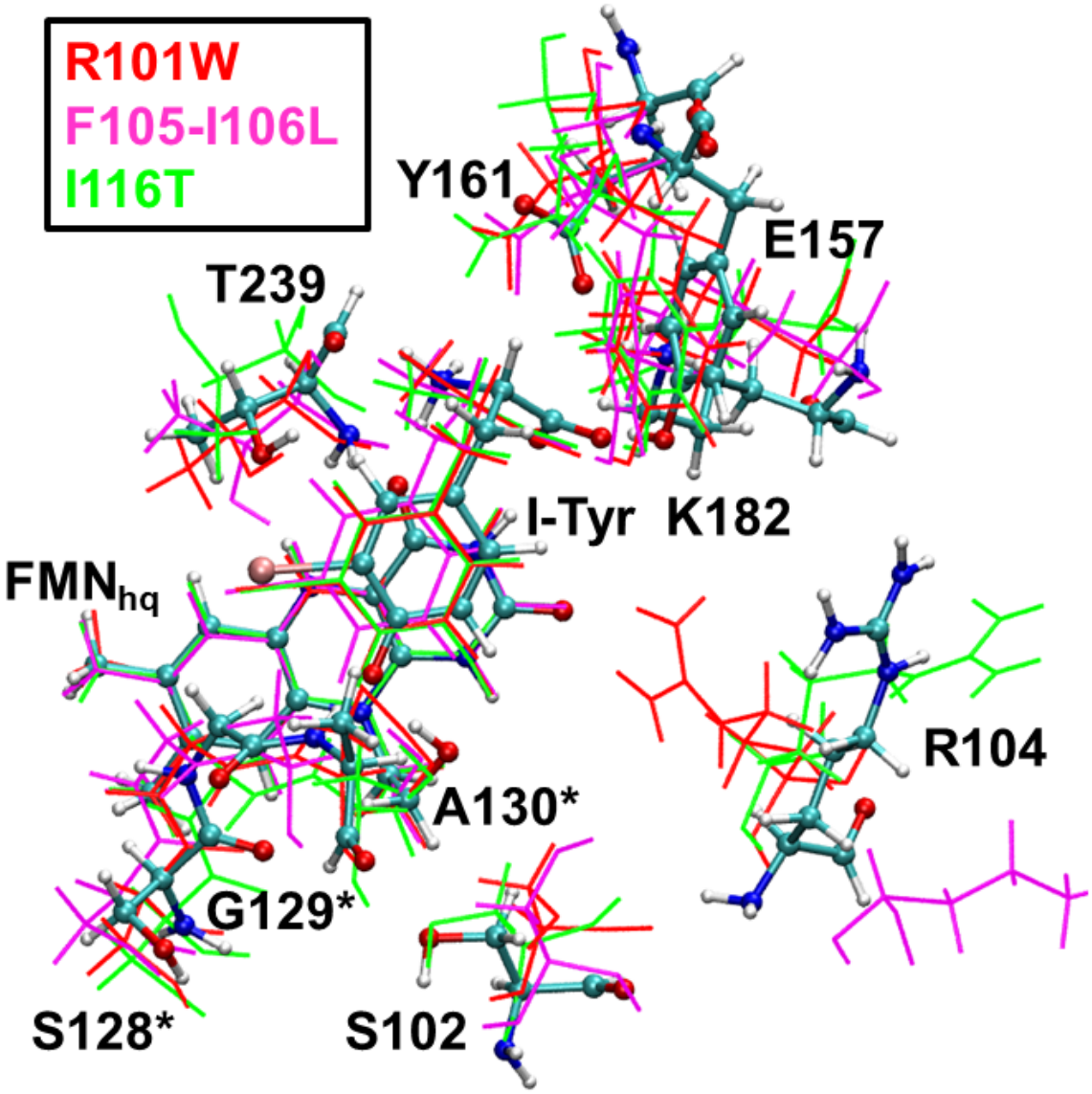

Figure S16: Superposition of the active–site QM–cluster structures for the wt enzyme and the three mutant systems. The flavin mononucleotide in its hydroquinone form ($\mathrm{FMN_{hq}}$), I–Tyr, and catalytically relevant residues (Y161, E157, K182, R104, S102, S128*, G129*, A130*, and T239) are shown. The mutant structures are color–coded as R101W (red), F105–I106L (magenta), and I116T (green), overlaid on the wt reference (shown in sticks).

Reference

(S1) Wako, H., and Saitô, N. (1978) Statistical mechanical theory of the protein conformation. II. Folding pathway for protein. *J. Phys. Soc. Jpn. 44*, 1939–1945.

(S2) Naganathan, A. N. (2013) A rapid, ensemble and free energy based method for engineering protein stabilities. *J. Phys. Chem. B 117*, 4956–4964.

(S3) Rajasekaran, N., Suresh, S., Gopi, S., Raman, K., and Naganathan, A. N. (2017) A general mechanism for the propagation of mutational effects in proteins. *Biochemistry 56*, 294–305.

(S4) Muñoz, V., and Eaton, W. A. (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Nat. Acad. Sci. 96*, 11311–11316.

(S5) Naganathan, A. N. (2012) Predictions from an Ising-like statistical mechanical model on the dynamic and thermodynamic effects of protein surface electrostatics. *J. Chem. Theo. Comput. 8*, 4646–4656.

(S6) Gopi, S., Aranganathan, A., and Naganathan, A. N. (2019) Thermodynamics and folding landscapes of large proteins from a statistical mechanical model. *Curr. Res. Struct. Biol. 1*, 6–12.

(S7) Ooka, K., Liu, R., and Arai, M. (2022) The Wako-Saitô-Muñoz-Eaton model for predicting protein folding and dynamics. *Molecules 27*, 4460.

(S8) Henry, E. R., and Eaton, W. A. (2004) Combinatorial modeling of protein folding kinetics: free energy profiles and rates. *Chem. Phys. 307*, 163–185.