Supporting Information

Emergence of a Low-Dimensional Energetic Landscape Governing Calcium Binding in Troponin C

Abdul Basit

School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

Corresponding Author:
Abdul Basit
meetabasit@gmail.com

S1 Minimal Structural Model

S1.1 Physical Basis of Structural Descriptors

Whether the structural descriptors capture intrinsic physical determinants of calcium binding, we examined their sensitivity to the spatial definition of the coordination environment. Two geometric parameters were varied: the steric cutoff (6–10 Å) and the coordination cutoff (4–7 Å). As summarized in Table S1, performance was invariant to the coordination cutoff for a fixed steric radius. Specifically, all coordination cutoffs tested at a steric cutoff of 6 Å yielded identical performance ($R_p = 0.83$), whereas steric cutoffs of 7 and 9 Å produced $R_p = 0.77$, and steric cutoffs of 8 and 10 Å yielded $R_p = 0.78$. The highest performance was obtained for the 6 Å steric cutoff, independent of the coordination cutoff examined. Increasing the steric cutoff beyond 6 Å led to lower but consistent performance, suggesting that inclusion of more distal residues contributes limited additional information to the local calcium-binding environment.

Table S1: Cutoff sensitivity analysis of the steric–coordination. Pearson correlation coefficients (R_p) remained highly across a broad range of steric and coordination cutoffs, indicating strong robustness of the model to parameter variation.

Steric Cutoff (Å)	Coordination Cutoff (Å)	R_p
6.0	4.0	0.83
6.0	5.0	0.83
6.0	6.0	0.83
6.0	7.0	0.83
7.0	4.0	0.77
7.0	5.0	0.77
7.0	6.0	0.77
7.0	7.0	0.77
8.0	4.0	0.78
8.0	5.0	0.78
8.0	6.0	0.78
8.0	7.0	0.78
9.0	4.0	0.77
9.0	5.0	0.77
9.0	6.0	0.77
9.0	7.0	0.77
10.0	4.0	0.78
10.0	5.0	0.78
10.0	6.0	0.78
10.0	7.0	0.78

S1.2 Structural feature augmentation and model selection

Whether additional geometric descriptors improve performance, we augmented the base structural model with four additional features: local packing density, coordination number, radial position, and backbone curvature. The base model comprised three structural descriptors: inverse Ca–ligand distance (d^{-1}), steric perturbation energy (E_{steric}), and mutation-induced local volume change (ΔV). All possible combinations of the additional descriptors were appended to this base set and evaluated using leave-one-out cross-validation (LOOCV). The results are summarized in Table S2. Across all tested combinations, performance remained stable, with R_p values ranging from 0.82 to 0.84. The best-performing model included local packing density and backbone curvature in addition to the base descriptors, yielding $R_p = 0.84$ and RMSE = 0.46. Descriptor combinations involving coordination number or radial-position features produced comparable performance with only minor variations in RMSE and correlation metrics. The base three-feature model already achieved strong performance ($R_p = 0.83$), and inclusion of additional descriptors resulted in only marginal improvements. The limited performance variation across descriptor combinations indicates that the dominant determinants of mutational energetics are already captured by the minimal base descriptor set, supporting a low-dimensional representation linking geometric perturbations to $\Delta\Delta G$.

Table S2: Top-performing structural descriptor combinations ranked by Pearson correlation coefficient (R_p). Descriptor labels were simplified for clarity: inverse Ca^{2+} distance (d^{-1}), steric perturbation energy (E_{steric}), residue volume change (ΔV), local packing density (ρ_{local}), coordination number (CN), radial position (r_{centroid}), and backbone angle (θ_{bb}).

Descriptor Combination	N_f	RMSE	R_p
$d^{-1} + E_{\text{steric}} + \Delta V + \rho_{\text{local}} + \theta_{\text{bb}}$	5	0.46	0.84
$d^{-1} + E_{\text{steric}} + \Delta V + \text{CN} + \theta_{\text{bb}}$	5	0.47	0.84
$d^{-1} + E_{\text{steric}} + \Delta V$	3	0.47	0.83
$d^{-1} + E_{\text{steric}} + \Delta V + \text{CN}$	4	0.47	0.83
$d^{-1} + E_{\text{steric}} + \Delta V + \theta_{\text{bb}}$	4	0.48	0.82
$d^{-1} + E_{\text{steric}} + \Delta V + \rho_{\text{local}}$	4	0.48	0.82
$d^{-1} + E_{\text{steric}} + \Delta V + \rho_{\text{local}} + r_{\text{centroid}}$	5	0.48	0.82
$d^{-1} + E_{\text{steric}} + \Delta V + r_{\text{centroid}}$	4	0.49	0.82
$d^{-1} + E_{\text{steric}} + \Delta V + \text{CN} + r_{\text{centroid}}$	5	0.49	0.82
$d^{-1} + E_{\text{steric}} + \Delta V + \rho_{\text{local}} + \text{CN} + \theta_{\text{bb}}$	6	0.49	0.82

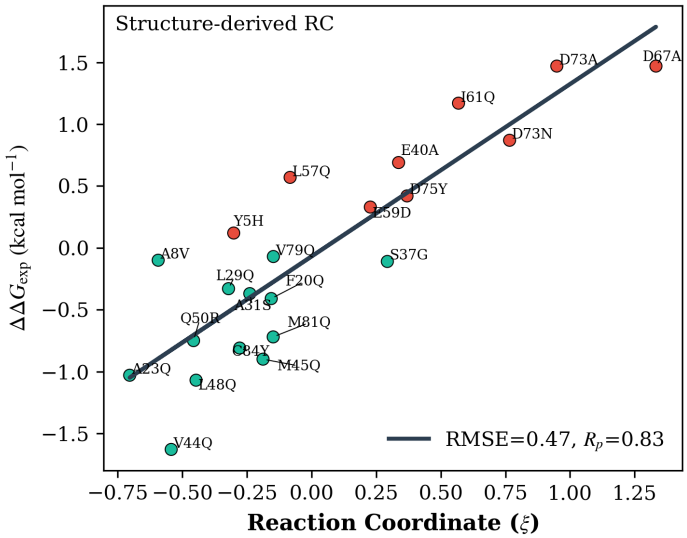

Figure S1: Correlation between the ξ and experimental $\Delta\Delta G_{\text{exp}}$. The ξ is constructed exclusively from structure-derived descriptors, including inverse distance to the metal center (d^{-1}), steric packing perturbation (E_{steric}), and residue volume change (ΔV). Each point corresponds to a mutant, colored by the sign of $\Delta\Delta G$ (red: destabilizing; green: stabilizing). The solid line represents the least-squares linear fit between ξ and $\Delta\Delta G_{\text{exp}}$

S2 Minimal Electrostatic Model

To evaluate descriptor redundancy and feature-selection robustness, a nested leave-one-out cross-validation (nested LOOCV) framework was applied across the electrostatic descriptor space. Pairwise Pearson correlations between descriptors are summarized in Table S3. Strong anticorrelation was observed between the electrostatic calcium interaction term (E_{Ca}) and charge–potential coupling ($dq\phi$) ($R_p = -0.94$), indicating substantial shared variance. Similarly, Born solvation energy exhibited strong correlation with E_{Ca} ($R_p = 0.82$), suggesting partial electrostatic redundancy within the descriptor manifold. Despite these correlations, nested feature selection demonstrated stable convergence toward a minimal descriptor set. Across 22 LOOCV folds, the descriptor pair (E_{Ca}, ϕ^2) was selected in 14 folds, while the augmented subset $(E_{\text{Ca}}, \phi^2, E_{\perp})$ appeared in 6 folds. All other subsets occurred only once. The most frequently selected subset consisted of:

$$(E_{\text{Ca}}, \phi^2)$$

The quadratic electrostatic response term (ϕ^2) was selected in all folds (selection fraction = 1.00), while the calcium interaction energy (E_{Ca}) was retained in 91% of folds. In contrast, descriptors such as $dq\phi$ and $E_{\parallel}$ were never selected, indicating limited independent predictive contribution after accounting for dominant electrostatic effects. The resulting metrics were:

$$\mathrm{RMSE} = 0.64, \qquad R_p = 0.65$$

It indicate that electrostatic information is concentrated within a compact and stable low-dimensional descriptor manifold rather than distributed across many weakly independent variables.

Table S3: Pairwise Pearson correlation coefficients between electrostatic descriptors.

	$dq\phi$	E_{Ca}	Born	ϕ^2	E_{mag}	$E_{\parallel}$	$E_{\perp}$
$dq\phi$	1.00	-0.94	-0.75	-0.38	0.15	-0.06	-0.23
E_{Ca}	-0.94	1.00	0.82	0.33	-0.14	0.16	0.26
Born	-0.75	0.82	1.00	0.15	-0.03	0.18	0.30
ϕ^2	-0.38	0.33	0.15	1.00	-0.13	-0.08	0.11
E_{mag}	0.15	-0.14	-0.03	-0.13	1.00	0.23	0.07
$E_{\parallel}$	-0.06	0.16	0.18	-0.08	0.23	1.00	0.21
$E_{\perp}$	-0.23	0.26	0.30	0.11	0.07	0.21	1.00

Table S4: Nested LOOCV feature-selection stability statistics.

Feature	Selection Count	Selection Fraction
ϕ^2	22	1.00
E_{Ca}	20	0.91
$E_{\perp}$	8	0.36
Born	2	0.09
E_{mag}	1	0.05
$dq\phi$	0	0.00
$E_{\parallel}$	0	0.00

Table S5: Most frequently selected feature subsets during nested LOOCV.

Feature Subset	Frequency
$(E_{\mathrm{Ca}}, \phi^2)$	14
$(E_{\mathrm{Ca}}, E_{\perp}, \phi^2)$	6
$(E_{\perp}, \mathrm{Born}, \phi^2)$	1
$(E_{\mathrm{mag}}, E_{\perp}, \mathrm{Born}, \phi^2)$	1

S2.1 Parameter Robustness Analysis

To identify the optimal parameter space, a systematic scan was performed over the electrostatic scaling parameter (κ), internal dielectric constant (ϵ_{in}), screening parameter (λ), and effective radius (r_{eff}). The resulting models were evaluated using two statistical metrics: root mean square error (RMSE) and Pearson correlation coefficient (R_p). The top-performing parameter combinations were ranked independently according to minimum RMSE and maximum R_p. The analysis revealed a consistent optimum near $\kappa = 0.05$ and

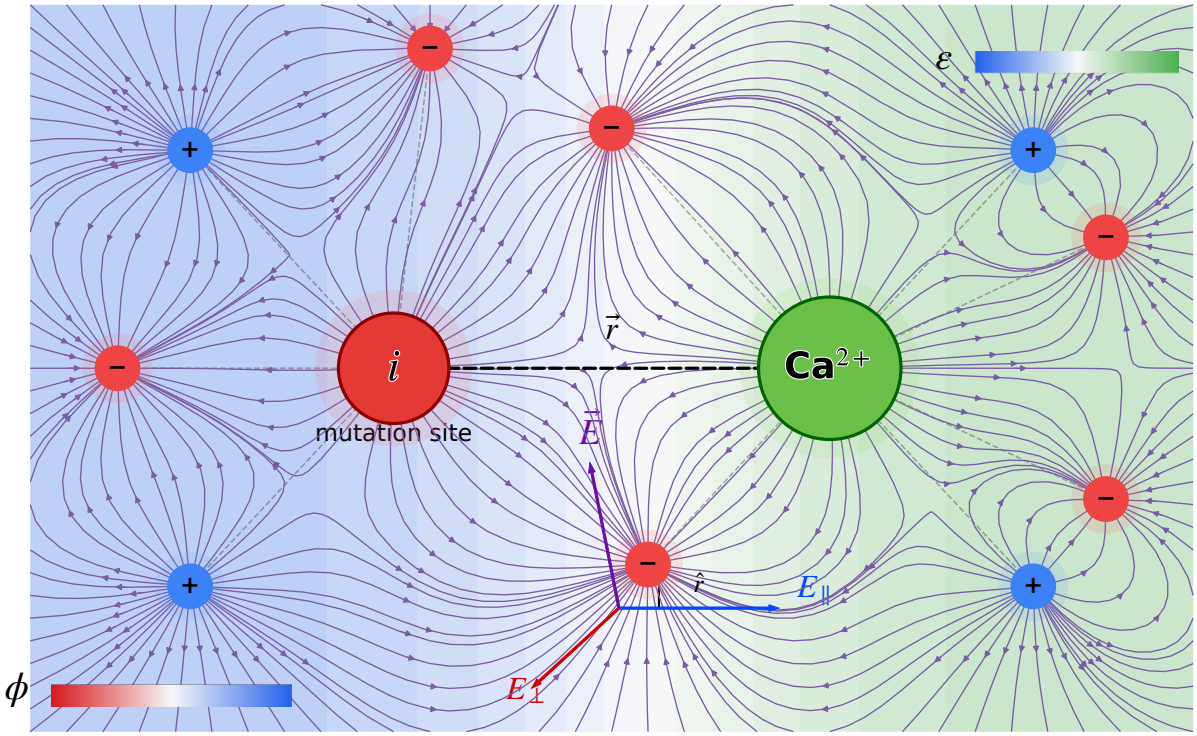

Figure S2: Electrostatic representation of mutation-induced perturbations near the Ca^{2+}-binding environment. The mutation site (i) and the Ca^{2+} ion are separated by the vector $\hat{r}$ defining the mutation-to-ion direction. Surrounding charged residues generate a heterogeneous electrostatic and dielectric environment represented by the local electric field $\mathbf{E}$ and the spatially varying dielectric response $\epsilon(r)$. Purple streamlines illustrate the organization of the electrostatic field surrounding the ion, while the background color map represents the electrostatic potential ϕ. Positive and negative effective residue charges are indicated by blue and red spheres, respectively. The electric field is decomposed into directional components parallel ($E_\parallel$) and perpendicular ($E_\perp$) to the mutation-to-ion axis. The shaded dielectric region surrounding the ion represents the transition between low-dielectric protein interior and high-dielectric solvent environment.

$\lambda = 4.0$, with ϵ_{in} values between 8.0 and 10.0. Variations in r_{eff} produced only marginal differences in predictive performance. The best overall parameter combination achieved an RMSE of 0.526 and an R_p value of 0.801 using

$$\kappa = 0.05, \quad \epsilon_{\text{in}} = 10.0, \quad \lambda = 4.0, \quad r_{\text{eff}} = 2.0$$

Table S6: Top 10 parameter combinations ranked by maximum Pearson correlation coefficient (R_p).

κ	ϵ_{in}	λ	r_{eff}	RMSE	R_p
0.05	10.0	4.0	2.0	0.526	0.801
0.05	10.0	4.0	2.5	0.526	0.801
0.05	10.0	4.0	3.0	0.526	0.801
0.05	10.0	4.0	1.5	0.526	0.801
0.05	8.0	4.0	2.5	0.526	0.800
0.05	8.0	4.0	2.0	0.526	0.800
0.05	8.0	4.0	3.0	0.526	0.800
0.05	8.0	4.0	1.5	0.526	0.800
0.05	6.0	4.0	2.0	0.527	0.800
0.05	6.0	4.0	3.0	0.527	0.800

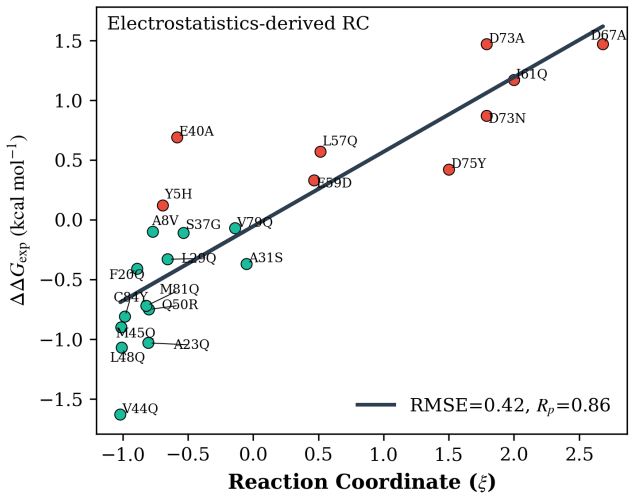

Figure S3: Correlation between the electrostatic reaction coordinate (ξ) and experimental binding free energy changes ($\Delta\Delta G_{\text{exp}}$). The solid line represents the best linear fit. Each point corresponds to a single mutation, labeled by residue identity.

S3 Minimal Dynamical Model

S3.1 Parameter Robustness Analysis

To evaluate the robustness of the elastic-network dynamical framework, a parameter scan was performed across multiple hyperparameter combinations, including network cutoff distance, correlation threshold, mode number, and local interaction radius. Performance remained stable across parameter ranges. The highest-performing parameter combinations consistently produced:

$$R_p \approx 0.85, \qquad \text{RMSE} \approx 0.44$$

with only minimal variation across neighboring parameter regimes. In particular, the optimal region centered around a network cutoff of 9 Å, 20 low-frequency modes, and correlation thresholds between 0.05 and 0.15. Importantly, varying the local interaction radius between 8–14 Å produced negligible changes in predictive performance, indicating that the extracted dynamical signal is intrinsically stable.

Table S7: Top 10 parameter combinations ranked by Pearson correlation coefficient (R_p).

Cutoff (Å)	Corr. Thr.	Modes	Radius (Å)	RMSE	R^2	R_p
9.0	0.05	20	8.0	0.44	0.85	
9.0	0.05	20	10.0	0.44	0.85	
9.0	0.05	20	12.0	0.44	0.85	
9.0	0.05	20	14.0	0.44	0.85	
9.0	0.10	20	8.0	0.44	0.85	
9.0	0.10	20	10.0	0.44	0.85	
9.0	0.10	20	12.0	0.44	0.85	
9.0	0.10	20	14.0	0.44	0.85	
9.0	0.15	20	8.0	0.44	0.85	
9.0	0.15	20	10.0	0.44	0.85	

S3.2 Dynamical Feature Combination Search

To identify the minimal dynamical descriptor manifold governing mutational free-energy changes, an exhaustive combinatorial feature search was performed across all ANM-derived dynamical descriptors. The evaluated descriptor space included communication efficiency (η), collectivity ($f_{\text{collective}}$), perturbation-response sensitivity (χ), configurational entropy (f_{entropy}), local flexibility (f_{local}), mutation correlation (mut_corr), asymmetry metrics (A), and residue fluctuation amplitudes (σ^2). The highest-performing dynamical model consisted of the following five descriptors:

$$(\eta, f_{\text{collective}}, \chi, f_{\text{entropy}}, \text{sigma}^2)$$

This model achieved:

$$\text{RMSE} = 0.44, \qquad R_p = 0.85$$

A closely related four-feature model excluding squared fluctuations yielded nearly identical performance:

$$(\eta, f_{\text{collective}}, \chi, f_{\text{entropy}})$$

with unchanged accuracy. This observation indicates that the dynamical signal is already captured and that additional descriptors contribute only marginal refinements. The top-performing models consistently retained communication efficiency, collectivity, perturbation-response sensitivity, and entropy-related descriptors. This convergence suggests that mutational free-energy perturbations are strongly coupled to large-scale co-

operative dynamics rather than localized residue fluctuations alone. The persistence of similar descriptor subsets across the top-ranking models further indicates that the manifold is stable and low-dimensional.

Table S8: Top 10 dynamical feature combinations ranked by Pearson correlation coefficient (R_p).

Feature Combination	N_f	RMSE	R_p
$\eta + \kappa + \chi + S + \sigma^2$	5	0.44	0.85
$\eta + \kappa + \chi + S$	4	0.44	0.85
$\eta + \kappa + \chi + S + L$	5	0.45	0.84
$\eta + \kappa + \chi + S + C$	5	0.46	0.84
$\eta + \kappa + A + \chi + S$	5	0.47	0.83
$\eta + \kappa + \chi$	3	0.48	0.83
$\eta + \chi + S + \sigma^2$	4	0.49	0.82
$\kappa + \chi + S + \sigma^2$	4	0.49	0.81
$\eta + \kappa + S$	3	0.50	0.81
$\eta + \chi + \sigma^2$	3	0.51	0.80

Here, η denotes communication efficiency, κ collective fluctuation amplitude, χ perturbation-response sensitivity, S spectral entropy, σ^2 mean-square fluctuation, A PRS asymmetry, C mutual correlation, and L local interaction density.

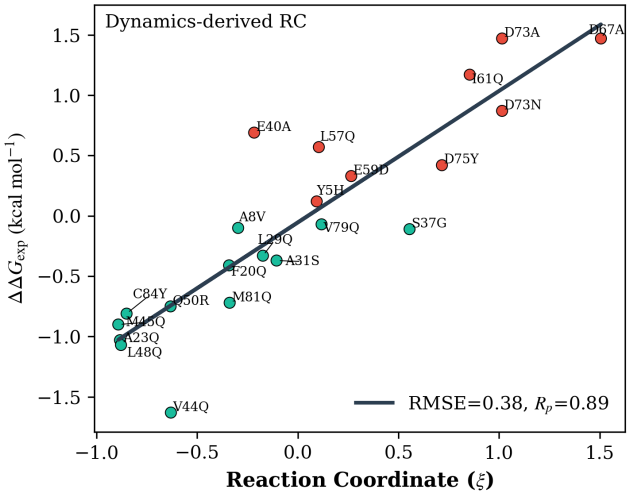

Figure S4: Correlation between the dynamical reaction coordinate (ξ) derived from elastic network model analysis and experimental binding free energy changes ($\Delta\Delta G_{\text{exp}}$). The solid line denotes the best linear fit. Each point corresponds to an individual mutation and is labeled by residue identity.

Table S9: Top-performing unified multiscale feature combinations ranked by Pearson correlation coefficient (R_p).

Unified Feature Combination	RMSE	R_p
$\Delta V + E_{\text{steric}} + E_{\text{Ca}} + \phi^2 + E_{\perp}$	0.38	0.90
$\sigma^2 + E_{\text{steric}} + E_{\text{Ca}} + \phi^2 + E_{\perp}$	0.38	0.89
$\Delta V + E_{\text{steric}} + E_{\text{Ca}} + \phi^2 + \chi$	0.39	0.89
$\Delta V + E_{\text{steric}} + E_{\text{Ca}} + E_{\perp} + \chi$	0.39	0.89
$\sigma^2 + E_{\text{steric}} + E_{\text{Ca}} + \phi^2 + \chi$	0.39	0.89
$\Delta V + E_{\text{steric}} + \phi^2 + E_{\perp} + \chi$	0.40	0.88
$\Delta V + E_{\text{Ca}} + \phi^2 + E_{\perp} + \chi$	0.40	0.88
$\sigma^2 + E_{\text{Ca}} + \phi^2 + E_{\perp} + \chi$	0.41	0.88
$\Delta V + E_{\text{steric}} + E_{\text{Ca}} + \phi^2$	0.41	0.87
$E_{\text{steric}} + E_{\text{Ca}} + \phi^2 + E_{\perp} + \chi$	0.41	0.87

S4 Unified Model

S4.1 Sector-Resolved Reaction Coordinate

To dissect the contribution of distinct physical interaction sectors to mutational energetics, the global reaction coordinate was decomposed into sector-specific latent coordinates corresponding to structural, electrostatic, and dynamical interactions. For each sector s, the associated feature vector

$$\mathbf{x}_{i,s} \in \mathbb{R}^{d_s}$$

was standardized as

$$\tilde{\mathbf{x}}_{i,s} = \frac{\mathbf{x}_{i,s} - \boldsymbol{\mu}_s}{\boldsymbol{\sigma}_s},$$

where $\boldsymbol{\mu}_s$ and $\boldsymbol{\sigma}_s$ denote the sector-specific mean and standard deviation vectors, respectively. Sector-resolved reaction coordinates were then constructed using the regression coefficients obtained from the global supervised model. For each sector s, the latent coordinate was defined as

$$\xi_{i,s} = \mathbf{w}_s^{\top} \tilde{\mathbf{x}}_{i,s},$$

where $\mathbf{w}_s$ denotes the subset of learned regression coefficients corresponding to features within sector s. This formulation projects each mutation onto the thermodynamically predictive direction within the corresponding interaction sector. The embedding of mutation i is thus given by

$$\Xi_i = \left(\xi_{i,\text{struct}}, \xi_{i,\text{elec}}, \xi_{i,\text{dyn}}\right),$$

defining a three-dimensional latent representation that integrates structural, electrostatic, and dynamical perturbation responses. Experimental mutation free energies ($\Delta\Delta G_{\text{exp}}$) were projected onto this manifold to visualize the energetic organization of mutational perturbations across sectors. Unlike a single global coordinate, this decomposition preserves mechanistic interpretability by retaining physically distinct latent contributions as-

Table S10: Highly correlated feature pairs within the unified multiscale descriptor space.

Feature Pair	Pearson Correlation (R_p)
$dq\phi$ vs E_{Ca}	-0.94
d^{-1} vs E_{Ca}	-0.91
E_{Ca} vs Born	0.82
E_{steric} vs ΔV	0.76
χ vs η	0.76
$dq\phi$ vs Born	-0.75
σ^2 vs χ	-0.73
η vs $f_{\mathrm{collective}}$	0.68
η vs χ	0.67
$f_{\mathrm{collective}}$ vs χ	0.61
d^{-1} vs η	-0.62
d^{-1} vs χ	-0.58
E_{Ca} vs η	0.60
E_{Ca} vs χ	0.54
σ^2 vs η	-0.55
local density vs coordination number	0.84
f_{local} vs σ^2	0.81

Table S11: Performance comparison of regression models for predicting $\Delta\Delta G$. Metrics are reported from cross-validation.

Model	RMSE	R_p
Polynomial (2^{nd} order)	0.36	0.91
Linear	0.37	0.90
Elastic Net	0.40	0.88
Ridge	0.40	0.88
Lasso	0.41	0.88
SVR (RBF kernel)	0.44	0.86
Gradient Boosting	0.52	0.79
Random Forest	0.53	0.78

sociated with structural rearrangement, electrostatic redistribution, and dynamical modulation. The resulting manifold enables direct assessment of how different interaction sectors contribute to mutation-induced free-energy changes.

S4.2 Residual Discriminant Coordinate

The global reaction coordinate is defined as

$$\xi = Xw \tag{1}$$

which provides a one-dimensional description of the thermodynamic response. The corresponding linear model

$$\hat{y} = a\xi + b \tag{2}$$

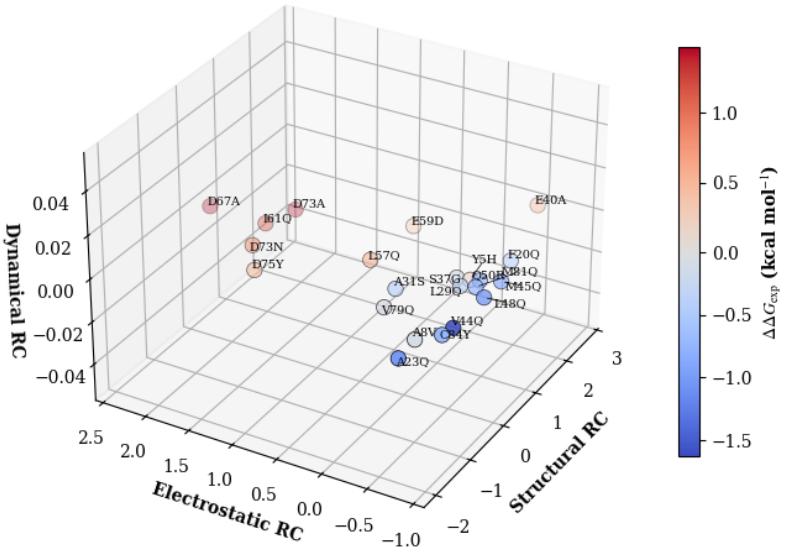

Figure S5: **Sector-resolved reaction coordinate of mutational perturbations.** Mutations are embedded in a three-dimensional latent space constructed from structural, electrostatic, and dynamical interaction sectors. For each sector, descriptors were standardized and projected onto the corresponding sector-restricted regression direction, yielding latent coordinates $\Xi_i = (\xi_{i,\mathrm{struct}}, \xi_{i,\mathrm{elec}}, \xi_{i,\mathrm{dyn}})$, where $\xi_{i,s} = \mathbf{w}_s^\top \tilde{\mathbf{x}}_{i,s}$. The three axes therefore represent the structural, electrostatic, and dynamical components of the thermodynamically predictive reaction coordinate. Each point corresponds to a single mutation and is colored by the experimentally measured mutation free energy, $\Delta\Delta G_{\mathrm{exp}}$ (kcal mol^{-1}). Labeled mutations denote representative stabilizing and destabilizing variants across the manifold. The embedding reveals partial energetic stratification across the sector-resolved coordinates, indicating that mutation-induced free-energy perturbations are organized along a low-dimensional manifold reflecting coordinated contributions from multiple physical interaction sectors.

captures the dominant variation in $\Delta\Delta G$. The residual

$$r = y - \hat{y} \tag{3}$$

represents the component of the thermodynamic response not explained by the global coordinate. To identify systematic structure in these residuals, the joint space (ξ, r) is partitioned into discrete regimes using k-means clustering,

$$\ell = \mathrm{KMeans}(\xi, r) \tag{4}$$

which groups mutations exhibiting similar deviations from the linear model. A secondary coordinate is then constructed in the original feature space. After standardization of

the features (X_s), linear discriminant analysis is applied to the cluster labels, yielding a direction v_{LDA} that maximally separates the identified regimes. The residual discriminant coordinate is defined as

$$\mathrm{RDC} = X_s, v_{\mathrm{LDA}} \tag{5}$$

This coordinate captures structured thermodynamic variation beyond the primary reaction coordinate ξ, enabling a two-dimensional representation of the free-energy response as

$$\Delta\Delta G = G(\xi, \mathrm{RDC}) \tag{6}$$