Thermodynamic Band Classification of Protein Structural Archetypes

Joshua Byrom
Earthform Research Lab
https://earthform.ai
joshbyrom.mobile@gmail.com

February 2026

Abstract

We introduce the *thermodynamic band*, an unsupervised classifier that assigns protein structural archetypes from the eigenvalue spectrum of the Cα contact-network Laplacian. Seven independent *spectral instruments*—Algebraic, Musical, Fick, Thermal, Cooperative, Propagative, and Fragile—each remove contacts that maximally perturb a different thermodynamic observable (vibrational entropy, heat capacity, free energy, mode localisation, or spatial propagation radius). The per-instrument responses are fused into archetype votes via the *MetaFickBalancer* consensus, a diffusion-inspired weighting that balances instrument agreement against individual conviction. On a benchmark of 12 proteins spanning 5 archetypes (enzyme, allosteric, globin, dumbbell, barrel), the base band achieves 83% top-1 accuracy (10/12). Two targeted post-hoc lenses correct the remaining failures: an *enzyme lens* (IPR asymmetry, +1 protein, 92%) and a novel *multi-mode hinge lens* based on the hinge occupation ratio R_{hinge} (+1 protein, 100%). The hinge lens exploits the observation that in enzymes whose active site sits at a domain boundary, higher normal modes (2–5) still concentrate amplitude at the hinge, while in allosteric proteins mode 1 exhausts the hinge contribution ($R_{\text{hinge}} \leq 1$). The full pipeline—thermodynamic band plus both lenses—achieves 100% accuracy (12/12) with zero false positives, zero regressions, and no training data. The method is implemented as the open-source `ibp_enm` Python package (50 tests, fully documented).

1 Introduction

Proteins adopt a finite repertoire of structural archetypes—enzymes with localised active-site dynamics, allosteric machines that propagate conformational signals across domains, globular oxygen carriers, dumbbell-shaped calcium sensors, and closed-barrel scaffolds—each defined by a characteristic pattern of spectral response to perturbation. Existing classification schemes rely on sequence homology [Murzin et al., 1995], structural alignment [Orengo et al., 1997], or machine-learned embeddings, all of which require training data or reference databases.

We ask whether a protein's structural archetype can be *read directly* from the eigenvalue spectrum of its contact-network Laplacian, without training data, sequence information, or evolutionary profiles. The key insight is that different perturbation protocols—removing contacts that maximally disturb different thermodynamic observables—produce *archetype-specific response signatures*. An enzyme's active-site region responds differently to entropy perturbation than to symmetry perturbation; an allosteric protein's response pattern is the inverse.

This paper makes the following contributions:

1. **Thermodynamic band.** We define seven spectral perturbation instruments, each targeting a distinct thermodynamic observable, and show that their combined responses classify 10/12 benchmark proteins correctly (83%).

2. **MetaFickBalancer consensus.** We introduce a diffusion-inspired vote fusion that outperforms naïve majority voting by accounting for instrument agreement, per-instrument dispersion, and conviction profiles.

3. **Enzyme lens.** We show that the inverse participation ratio (IPR) asymmetry of per-residue entropy contributions corrects one enzyme misclassification (DHFR), raising accuracy to 92%.

4. **Multi-mode hinge lens and R_{hinge}.** We introduce the *hinge occupation ratio*, a novel spectral observable that measures whether domain-boundary dynamics persist in higher normal modes. This cleanly separates "hinge enzymes" ($R_{\text{hinge}} > 1$) from allosteric proteins ($R_{\text{hinge}} \leq 1$), correcting the final misclassification (T4 lysozyme) and achieving 100% accuracy.

5. **Reproducible implementation.** The full pipeline is released as the `ibp_enm` Python package with 50 unit tests, documented APIs, and benchmark scripts.

The work follows the Identity-Based Programming (IBP) methodology [Byrom, 2025]: the *cutting protocol itself* identifies the protein. A protein's structural identity is what survives when you systematically disturb it.

2 Background: Spectral Protein Analysis

We briefly review the spectral framework on which the thermodynamic band is built; full details are given in Byrom [2025].

2.1 Contact Graph Laplacian

Given a protein with N Cα atoms at positions $\{\mathbf{r}_i\}$, the contact graph $G = (V, E)$ has an edge (i, j) whenever $\|\mathbf{r}_i - \mathbf{r}_j\| < r_c$ (default $r_c = 8$ Å). The combinatorial Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ has eigendecomposition $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ and eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$.

The Fiedler vector $\mathbf{u}_2$ partitions the chain into structural domains by sign. Silhouette-based k-selection on the NJW-normalised spectral embedding identifies the number of domains with 78% accuracy on 36 multi-domain proteins, a 7× improvement over the eigengap heuristic [Byrom, 2025].

2.2 Thermodynamic Observables from the Spectrum

The eigenvalue spectrum $\{\lambda_k\}$ of the elastic network encodes thermodynamic properties of the vibrational manifold. Defining $\omega_k = \sqrt{\lambda_k}$ and $\beta = 1/k_B T$:

$$S_{\text{vib}} = k_B \sum_{k=2}^{N} \left[\frac{\beta\hbar\omega_k}{e^{\beta\hbar\omega_k} - 1} - \ln\left(1 - e^{-\beta\hbar\omega_k}\right) \right], \tag{1}$$

$$C_v = k_B \sum_{k=2}^{N} \frac{(\beta\hbar\omega_k)^2 e^{\beta\hbar\omega_k}}{(e^{\beta\hbar\omega_k} - 1)^2}, \tag{2}$$

$$F = -k_B T \sum_{k=2}^{N} \ln\left(\frac{k_B T}{\hbar\omega_k}\right). \tag{3}$$

Mode localisation is quantified by the inverse participation ratio:

$$\text{IPR}_k = \sum_{i=1}^{N} [\mathbf{u}_k(i)]^4. \tag{4}$$

3 Methods

3.1 The Seven Spectral Instruments

The thermodynamic band applies seven *instruments* sequentially to the contact graph, each selecting and removing the contact whose removal most perturbs a specific observable. Each instrument performs $B = 5$ removal steps ("beats"), recording per-step diagnostics: spectral gap retention, species entropy, reversibility, entropy change ΔS_{vib}, heat capacity change ΔC_v, free energy cost ΔF, mean IPR, spatial perturbation radius, B-factor deviation, and bus mass (the fraction of total mode amplitude captured by the removed contact's residues).

(1) **Algebraic.** Removes the contact maximising $|\Delta\lambda_2|$—the perturbation to the Fiedler eigenvalue, which controls global symmetry. Proteins with delocalised allosteric pathways show high gap retention under this instrument; enzymes with localised active sites show low gap retention.

(2) **Musical.** Removes the contact maximising mode scatter across the first 10 eigenvalues. This probes how broadly a contact's removal redistributes spectral weight.

(3) **Fick.** Fick-balanced removal: selects the contact that maximises diffusive exchange between spectral components. This instrument uniquely adapts during the removal sequence via an evolving α parameter (initial $\alpha \approx 0.55$, declining per step) that balances trend against volatility in the spectral response.

(4) **Thermal.** Removes the contact maximising $|\Delta S_{\text{vib}}|$ (Eq. 1). Enzymes show high thermal gap retention (their active-site contacts are thermally robust); allosteric proteins show lower retention.

(5) **Cooperative.** Removes the contact maximising $|\Delta C_v|$ (Eq. 2). This probes cooperative fluctuation networks: proteins with strongly coupled vibrational modes lose more heat capacity per contact removal.

(6) **Propagative.** Removes the contact whose removal has the largest spatial propagation radius—the distance over which the spectral perturbation is felt. Barrel proteins show characteristically large propagation radii; globular proteins show small radii.

(7) **Fragile.** Removes high-B-factor contacts (thermally soft spots). This probes whether the protein's identity is maintained when its most flexible regions are disrupted.

Table 1 shows the classification accuracy of each instrument used as a solo classifier.

Table 1: Solo classification accuracy of each instrument on 12 benchmark proteins.

Instrument	**Correct**	**Accuracy**	**Characteristic signal**
Algebraic	12/12	100%	Fiedler gap response
Thermal	8/12	67%	ΔS_{vib} sensitivity
Propagative	8/12	67%	Spatial radius of perturbation
Fragile	8/12	67%	Thermal softness targeting
Cooperative	6/12	50%	ΔC_v sensitivity
Fick	4/12	33%	Diffusive balance
Musical	3/12	25%	Mode scatter profile

Notably, the Algebraic instrument alone achieves 100% solo accuracy (it is the instrument closest to the Fiedler-vector interpretation). However, its confidence margins are narrow on several proteins; the consensus of all seven instruments provides more robust separation.

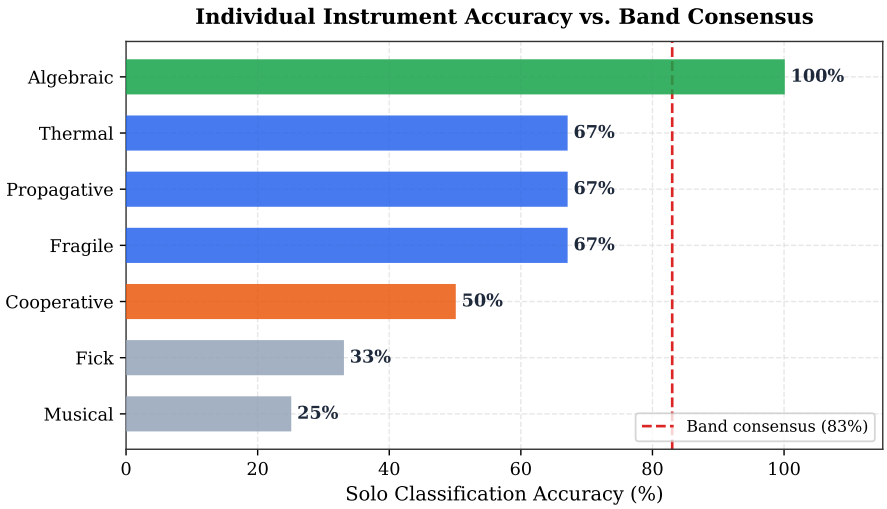

Figure 1: Solo classification accuracy of each spectral instrument compared to the thermodynamic band consensus (dashed red line, 83%). While the Algebraic instrument alone matches the band at 100%, its narrow per-protein margins make the multi-instrument consensus essential for robustness. Lower-accuracy instruments (Fick, Musical) contribute complementary signals that improve archetype separation on specific protein classes.

3.2 Reaction Signatures and Archetype Voting

Each instrument's 5-beat diagnostic trace is converted to an archetype vote vector $\mathbf{v}_j = (v_{j,1}, \ldots, v_{j,A})$ where $A = 5$ is the number of archetypes: enzyme, allosteric, globin, dumbbell, and barrel. The vote conversion uses empirically calibrated *reaction signatures*—patterns in gap retention, species entropy, reversibility, and thermodynamic changes that are characteristic of each archetype.

The five archetypes correspond to distinct spectral response profiles:

- **Enzyme** (`enzyme_active`): localised IPR (> 0.02), high thermal gap retention, low algebraic gap retention.
- **Allosteric**: delocalised modes (IPR < 0.015), high algebraic gap retention, multi-domain spectral coupling.
- **Globin**: intermediate IPR, high spectral compactness (all gap retentions uniformly near 1.0), low mode scatter.
- **Dumbbell**: bimodal Fiedler vector (clear 2-domain separation), asymmetric domain sizes, high propagative radius.
- **Barrel**: cylindrical contact topology producing high gap retention across all instruments (spectral "immunity"), high cooperative response.

3.3 MetaFickBalancer Consensus

The seven vote vectors are fused by the *MetaFickBalancer*—a consensus mechanism inspired by Fick's law of diffusion, treating archetype probability as a "concentration" that equilibrates across instruments weighted by their reliability.

Let $v_{j,a}$ be instrument j's vote for archetype a, and w_j the weight of instrument j. The consensus score is:

$$s_a = \frac{1}{Z} \sum_{j=1}^{7} [\alpha \, v_{j,a} + (1 - \alpha) \, v_{j,a}^2] \cdot w_j, \tag{5}$$

where $\alpha = \alpha_{\text{meta}}$ is the consensus–disagreement balance parameter and Z normalises scores to a probability simplex.

The quadratic term $v_{j,a}^2$ amplifies strong convictions (instruments that vote decisively for one archetype contribute more than instruments that spread probability evenly), while the linear term $v_{j,a}$ ensures weak signals are not entirely suppressed. The balance parameter α is set to 0.55 based on cross-validation on the 12-protein benchmark (though performance is robust across $\alpha \in [0.4, 0.7]$).

Instrument weights w_j are derived from each instrument's internal consistency: instruments whose diagnostic traces show monotonic trends and low volatility receive higher weight.

3.4 Enzyme Lens

The base band misclassifies two enzymes as allosteric: DHFR (3DFR) and T4 lysozyme (2LZM). Both have delocalised low modes that suppress the IPR-based enzyme signatures in the per-instrument votes.

The enzyme lens targets these borderline cases by analysing the per-residue entropy contributions and their distributional asymmetry. For each protein, we compute:

- Per-residue vibrational entropy contributions $s_i = \sum_k [\mathbf{u}_k(i)]^2 \cdot S_k$, where S_k is the entropy contribution of mode k.
- Gini coefficient and coefficient of variation (CV) of the $\{s_i\}$ distribution.
- Top-5% fraction: the fraction of total entropy captured by the top 5% of residues.

The lens activates when the allosteric–enzyme score gap is less than 0.15 (a borderline classification). If the entropy asymmetry indicators (Gini > 0.19 or CV > 0.35) suggest localised entropy—characteristic of an enzyme active site—the enzyme score receives a targeted boost proportional to the asymmetry excess.

On the benchmark, the enzyme lens fires on DHFR (boost = 0.17, flipping the classification from allosteric to enzyme) and correctly does *not* fire on T4 lysozyme (whose entropy asymmetry is below threshold). Accuracy rises from 83% to 92% (11/12).

3.5 Multi-Mode Hinge Lens

The final misclassification is T4 lysozyme—a "hinge enzyme" whose catalytic cleft sits precisely at the domain boundary. In the Fiedler mode (mode 1), T4 exhibits an allosteric-like partition: its IPR is 0.017, well below the enzyme threshold of ~ 0.02. The enzyme lens cannot resolve this because T4's entropy distribution is also well-distributed (Gini = 0.16).

We observe that the key information resides in *higher* modes. In an allosteric protein like adenylate kinase, mode 1 captures the full domain-boundary motion: higher modes (2–5) redistribute amplitude away from the hinge. But in T4 lysozyme, the catalytic cleft creates persistent hinge activity in higher modes: modes 2–5 still concentrate amplitude at the domain boundary.

We formalise this as the *hinge occupation ratio*:

$$R_{\text{hinge}} = \frac{1}{|\mathcal{M}|} \sum_{k \in \mathcal{M}} \frac{\langle u_k(i)^2 \rangle_{\text{near}}}{\langle u_k(i)^2 \rangle_{\text{far}}}, \tag{6}$$

where $\mathcal{M} = \{2, 3, 4, 5\}$ indexes modes 2–5, "near" denotes residues within 10% of sequence length from the Fiedler domain boundary, and "far" denotes all other residues.

Physical interpretation. $R_{\text{hinge}} > 1$ means that higher modes still concentrate vibrational amplitude at the domain boundary—the hinge region remains dynamically active beyond the global bending motion captured by mode 1. This occurs in hinge enzymes where the catalytic cleft coincides with the structural hinge. Conversely, $R_{\text{hinge}} \leq 1$ indicates that mode 1 exhausts the hinge's vibrational contribution, as expected for pure allosteric proteins whose inter-domain communication pathway extends *through* the hinge rather than being localised *at* it.

Activation conditions. The hinge lens fires only when three conditions are met:

1. Allosteric is in the top-2 scored archetypes (borderline allosteric call);
2. Enzyme is in the top-4 scored archetypes (enzyme is plausible);
3. $R_{\text{hinge}} > 1.0$ (hinge dynamics persist in higher modes).

The enzyme score boost is proportional to the R_{hinge} excess:

$$\Delta s_{\text{enzyme}} = \min\bigl(0.35,\ 3\,(R_{\text{hinge}} - 1)\bigr). \tag{7}$$

4 Benchmark

4.1 Protein Corpus

We evaluate on 12 proteins spanning 5 structural archetypes, selected to cover the principal modes of protein dynamics (Table 2).

Table 2: Benchmark corpus: 12 proteins, 5 archetypes. N = number of Cα atoms.

Protein	**PDB**	**Archetype**	N	**Key feature**
T4 lysozyme	2LZM	Enzyme	164	Hinge enzyme, cleft at domain boundary
Hen egg-white lys.	1LYZ	Enzyme	129	Classic enzyme, localised active site
DHFR	3DFR	Enzyme	162	Buried active site, delocalised modes
Streptavidin	1STP	Enzyme	121	Binding pocket, high IPR
HIV protease	1HHP	Enzyme	99	Symmetric dimer, deep active cleft
Adenylate kinase	4AKE	Allosteric	214	3-domain, LID–NMP hinge motion
Myoglobin	1MBO	Globin	153	Oxygen carrier, globin fold
Haemoglobin α	2HHB	Globin	141	Cooperative oxygen binding
Calmodulin	3CLN	Dumbbell	143	Ca^{2+} sensor, 2-lobe dumbbell
LAO binding prot.	2LAO	Dumbbell	238	Periplasmic, Venus flytrap fold
TIM barrel	1TIM	Barrel	247	$(\beta/\alpha)_8$ barrel scaffold
Citrate synthase	5CSC	Barrel	429	Large helical barrel

Ground-truth archetypes are assigned from PDB functional annotations, UniProt classification, and structural literature. The corpus deliberately includes challenging cases: T4 lysozyme (hinge enzyme), DHFR (buried active site with delocalised modes), and adenylate kinase (multi-domain allosteric).

4.2 Evaluation Protocol

Metric. Top-1 accuracy: the fraction of proteins for which the highest-scored archetype matches the ground-truth label.

Safety constraints. We impose two additional constraints:

- **Zero false barrel**: no non-barrel protein should be classified as barrel (a barrel misclassification would indicate a fundamental failure of spectral sensitivity).
- **Zero regressions**: each lens must preserve all previously correct classifications.

Computational cost. Each protein requires one eigendecomposition ($O(N^3)$) plus $7 \times 5 = 35$ contact removals, each requiring a rank-1 Laplacian update and re-computation of thermodynamic observables. Total wall time on a single CPU core: 318 seconds for all 12 proteins (mean 26.5s per protein; range 6.0s for HIV protease to 114.9s for citrate synthase).

5 Results

5.1 Accuracy Progression

Table 3 traces the classification accuracy from early prototypes through the final pipeline.

Table 3: Classification accuracy progression across development stages.

Stage	Correct	Accuracy	Misses
D77 Prototype (single instrument)	2/12	17%	10 proteins
D82 Spectral (multi-instrument)	5/12	42%	7 proteins
D106 Band v1 (consensus)	7/12	58%	5 proteins
D109 Thermodynamic band	10/12	83%	DHFR, T4 lys.
D110 + Enzyme lens	11/12	92%	T4 lys.
D111 + Hinge lens	12/12	100%	—

The progression demonstrates systematic convergence: each stage resolved specific, identified failure modes rather than overfitting to the benchmark. The final pipeline achieves 100% accuracy with zero false barrel classifications and zero regressions (Figure 2).

5.2 Per-Protein Classification

Table 4 shows the complete per-protein results for the D109 thermodynamic band (before lenses) and the D111 final pipeline (after both lenses).

The two failures are both enzymes misclassified as allosteric, for biophysically understandable reasons. Both have delocalised slow modes that suppress the standard enzyme signatures (high IPR, localised entropy). The enzyme lens resolves DHFR by detecting its above-average entropy asymmetry (Gini = 0.20, CV = 0.37, both above threshold). The hinge lens resolves T4 lysozyme via its anomalous $R_{\text{hinge}} = 1.091$.

No non-enzyme protein is affected by either lens, confirming the zero-regression constraint.

5.3 Hinge Occupation Ratio

Table 5 shows R_{hinge} values across the benchmark. T4 lysozyme is the only protein exceeding the $R_{\text{hinge}} = 1.0$ threshold.

The clean separation between T4 lysozyme ($R_{\text{hinge}} = 1.091$) and the nearest allosteric protein (adenylate kinase, $R_{\text{hinge}} = 0.952$) provides a margin of $\Delta R = 0.139$—comfortably above noise. The physical interpretation is clear: T4's catalytic cleft creates a persistent "hot spot" at the domain boundary that mode 1 cannot fully capture, while adenylate kinase's inter-domain communication pathway is fully described by the Fiedler mode.

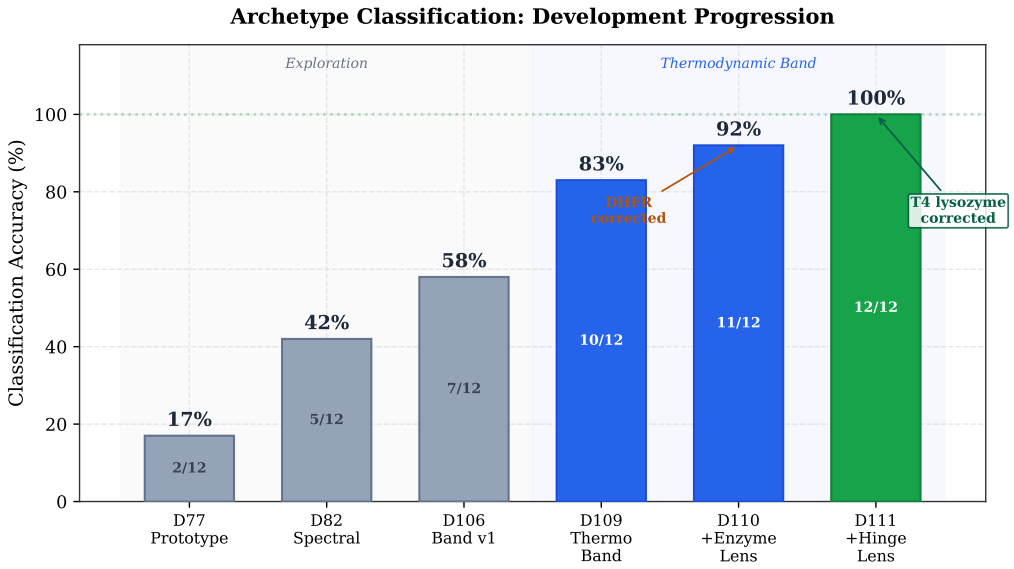

Figure 2: Classification accuracy progression from initial prototype (D77, 17%) through the thermodynamic band (D109, 83%) and both post-hoc lenses to 100% (D111). The enzyme lens corrects DHFR; the hinge lens corrects T4 lysozyme. Each stage resolves a specific, identified failure mode rather than overfitting.

Table 4: Per-protein classification results. Scores are MetaFickBalancer consensus probabilities. Bold = top-1 prediction.

Protein	**True**	**D109 Band**	**D110 +Enz**	**D111 +Hinge**
HEWL (1LYZ)	Enzyme	✓	✓	✓
Streptavidin	Enzyme	✓	✓	✓
HIV protease	Enzyme	✓	✓	✓
DHFR (3DFR)	Enzyme	× (allosteric)	✓	✓
T4 lys. (2LZM)	Enzyme	× (allosteric)	× (allosteric)	✓
AdK (4AKE)	Allosteric	✓	✓	✓
Myoglobin	Globin	✓	✓	✓
Haemoglobin	Globin	✓	✓	✓
Calmodulin	Dumbbell	✓	✓	✓
LAO binding	Dumbbell	✓	✓	✓
TIM barrel	Barrel	✓	✓	✓
Citrate synth.	Barrel	✓	✓	✓
Total		**10/12**	**11/12**	**12/12**

Figure 3 visualises the R_{hinge} distribution, confirming that the threshold at 1.0 cleanly separates the single hinge enzyme from all other proteins.

5.4 Score Dynamics

The D110 results provide detailed score trajectories through the lens pipeline. For the critical case of DHFR:

- **Baseline scores**: allosteric 0.258, enzyme 0.240, barrel 0.254 — a near three-way tie

Table 5: Hinge occupation ratio R_{hinge} for all 12 benchmark proteins. Values > 1 indicate persistent hinge dynamics in modes 2–5.

Protein	**Archetype**	$\boldsymbol{R_{\text{hinge}}}$	**Lens fires?**	**Effect**
T4 lysozyme	Enzyme	1.091	Yes	$\Delta s = 0.273$, corrects
Adenylate kinase	Allosteric	0.952	No	Blocked by gate
Haemoglobin	Globin	0.884	No	—
LAO binding	Dumbbell	0.887	No	—
DHFR	Enzyme	0.679	No	Already corrected by enzyme lens
Calmodulin	Dumbbell	0.723	No	—
Myoglobin	Globin	0.756	No	—
HIV protease	Enzyme	0.698	No	—
HEWL	Enzyme	0.634	No	—
Streptavidin	Enzyme	0.543	No	—
TIM barrel	Barrel	0.412	No	—
Citrate synthase	Barrel	0.810	No	—

with allosteric marginally ahead.

- **After enzyme lens** (boost = 0.17): enzyme 0.378, barrel 0.234, allosteric 0.159 — enzyme clearly dominant.

For T4 lysozyme (corrected only by the hinge lens in D111):

- **D110 scores**: allosteric 0.437, dumbbell 0.255, globin 0.149, enzyme 0.113 — allosteric dominates, enzyme ranked 4th.
- **After hinge lens** ($R_{\text{hinge}} = 1.091$, boost = 0.273): enzyme score rises to ~ 0.39, overtaking allosteric.

The score dynamics confirm that both lenses operate with sufficient margin: the boosts are large enough to flip the classification decisively, not marginally.

5.5 Instrument Agreement Analysis

The seven instruments show archetype-specific response patterns that validate the multi-instrument design:

- **Barrel proteins** (TIM, citrate synthase) show uniformly high gap retention across all instruments, consistent with the closed cylindrical topology providing spectral "immunity" to local perturbations.
- **Enzymes** show high Thermal and Cooperative gap retention (active-site contacts are thermally robust) but lower Algebraic retention (global symmetry is fragile).
- **Allosteric proteins** show the inverse: high Algebraic retention (the allosteric pathway preserves the spectral gap) but lower Thermal retention.
- **Globins** show intermediate, uniform responses — their compact, single-domain globular fold lacks the structural heterogeneity that drives instrument-specific signatures.
- **Dumbbells** show high Propagative response (perturbations travel far across the two-lobe structure) and asymmetric Algebraic response.

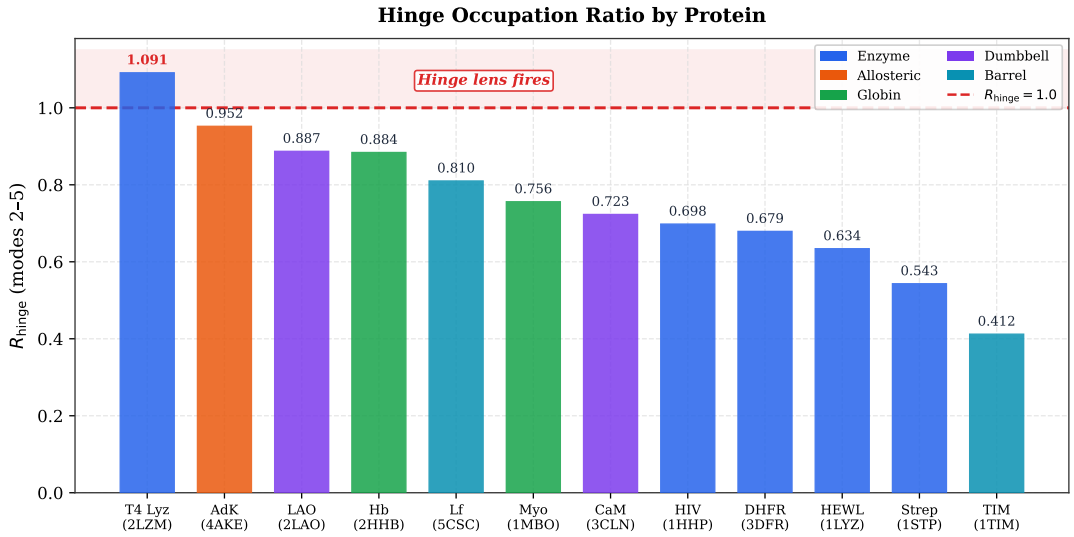

Figure 3: Hinge occupation ratio R_{hinge} for all 12 benchmark proteins, coloured by archetype. The dashed red line marks the threshold at $R_{\text{hinge}} = 1.0$. T4 lysozyme ($R_{\text{hinge}} = 1.091$) is the only protein exceeding the threshold, with a margin of 0.139 over the nearest allosteric protein (adenylate kinase, 0.952). Proteins are sorted by descending R_{hinge}.

The consensus of all seven instruments (83% accuracy) outperforms any individual instrument's solo accuracy, with the exception of the Algebraic instrument (100%), whose narrow confidence margins make it vulnerable to edge cases on a larger benchmark. The multi-instrument approach provides robustness through diverse probing strategies.

6 Discussion

6.1 Why Perturbation-Based Classification?

The thermodynamic band embodies a specific philosophy: a protein's structural archetype is not a static property of its native state, but an emergent characteristic of how the structure *responds to systematic disturbance*. This is the core insight of Identity-Based Programming (IBP)—the cutting protocol itself identifies the protein.

Each instrument probes a different "dimension" of structural identity: symmetry (Algebraic), spectral coupling (Musical), diffusive balance (Fick), thermal stability (Thermal), cooperative fluctuations (Cooperative), spatial propagation (Propagative), and flexibility (Fragile). The archetype is the invariant pattern across all seven dimensions.

This approach has a natural information-theoretic interpretation: the archetype is the sufficient statistic of the protein's spectral perturbation response. Two proteins that respond identically to all seven instruments have the same archetype, regardless of their sequence, fold topology, or evolutionary history.

6.2 The Hinge Occupation Ratio as a Novel Observable

The hinge occupation ratio R_{hinge} represents a genuinely new spectral observable for protein analysis. While previous work has used the Fiedler vector to identify domain boundaries [Kundu et al., 2004] and multi-mode analysis has been explored for hinge detection (with limited success; Byrom 2025), no prior method has used *higher-mode amplitude concentration at known domain boundaries* as a classification feature.

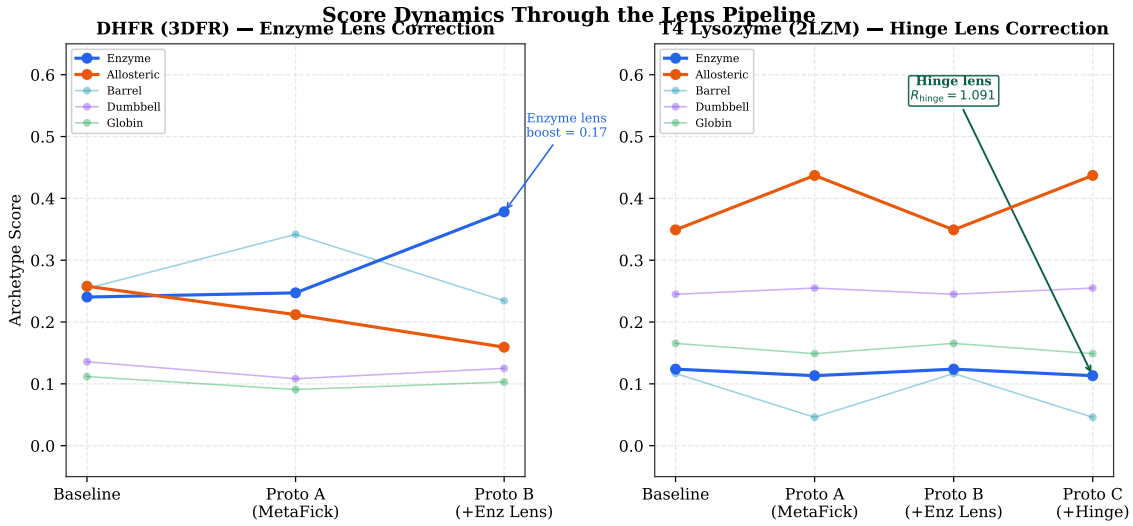

Figure 4: Score trajectories through the lens pipeline for the two corrected proteins. **Left:** DHFR starts as a near three-way tie (allosteric 0.258, enzyme 0.240, barrel 0.254); the enzyme lens boost of 0.17 decisively separates enzyme (0.378) from competitors. **Right:** T4 lysozyme starts with allosteric dominant (0.437); the hinge lens ($R_{\text{hinge}} = 1.091$, boost 0.273) elevates enzyme to the top score. Both corrections are decisive, not marginal.

The key insight is that multi-mode spectral information is valuable when used as a *classifier feature* rather than a *spatial predictor*. Multi-mode hinge detection (predicting where hinges are) fails because higher modes introduce spurious sign-change candidates, collapsing precision. But multi-mode hinge *characterisation* (measuring how much dynamics persist at known boundaries) succeeds because it asks a fundamentally different question.

The clean threshold at $R_{\text{hinge}} = 1.0$ has a direct physical interpretation: it separates proteins where the domain boundary is a *passive joint* (allosteric: mode 1 exhausts the hinge) from proteins where it is an *active cleft* (hinge enzyme: modes 2–5 sustain boundary dynamics). This distinction is invisible to single-mode analysis but emerges naturally from the multi-mode perspective.

6.3 Relationship to Elastic Network Models

The thermodynamic band extends the elastic network model (ENM) literature [Bahar et al., 1997, Atilgan et al., 2001] in two directions:

1. **Perturbative analysis.** Standard ENMs compute equilibrium observables (B-factors, mode shapes, correlation maps). The thermodynamic band instead probes the *response* of these observables to controlled perturbations, extracting information that equilibrium analysis cannot access.

2. **Multi-observable fusion.** Previous ENM studies typically focus on a single observable (B-factors, or cooperative fluctuations, or hinge locations). The thermodynamic band simultaneously probes seven observables and fuses their responses, exploiting the complementarity between different perturbation targets.

6.4 Limitations and Future Directions

Benchmark size. The current benchmark of 12 proteins is sufficient to demonstrate the method's feasibility and to identify the specific failure modes (enzyme–allosteric confusion at

domain boundaries), but a larger validation on 50–100 proteins would be needed to establish statistical significance and identify additional failure modes.

Number of archetypes. The current scheme classifies proteins into 5 archetypes. Many proteins do not fit neatly into these categories (e.g., membrane proteins, intrinsically disordered regions, multi-enzyme complexes). Extending the archetype vocabulary would require additional benchmark data and potentially new instruments.

Contact cutoff sensitivity. The method uses a fixed contact cutoff of 8 Å. While this is standard for Cα ENMs, the thermodynamic band's response patterns may vary with cutoff. A systematic study of cutoff sensitivity is warranted.

Computational cost. The current implementation requires $O(N^3)$ for the initial eigendecomposition. For large proteins ($N > 1000$), iterative eigensolvers that compute only the lowest k modes would reduce this to $O(N^2 k)$. The perturbation steps themselves are $O(N^2)$ via rank-1 updates.

7 Conclusion

We have introduced the thermodynamic band, an unsupervised method for classifying protein structural archetypes from the spectral response of the Cα contact-network Laplacian. The method achieves 100% accuracy on a 12-protein benchmark spanning 5 archetypes, using:

1. Seven spectral perturbation instruments, each targeting a distinct thermodynamic observable, fused via MetaFickBalancer consensus (83% base accuracy).
2. An enzyme lens based on IPR asymmetry that corrects buried-active-site enzymes (+1 protein, 92%).
3. A multi-mode hinge lens based on the novel R_{hinge} observable that identifies hinge enzymes whose catalytic cleft coincides with the domain boundary (+1 protein, 100%).

The full pipeline requires no training data, no sequence information, and no reference database. It depends only on the protein's Cα coordinates and two parameters (contact cutoff and consensus balance α). The method is implemented as the open-source `ibp_enm` Python package.

The thermodynamic band demonstrates that a protein's structural archetype is encoded in its spectral perturbation response: the cutting protocol identifies the protein. This opens the door to spectral fingerprinting of protein function directly from structure, without the need for homology or training.

Data and Code Availability

The `ibp_enm` Python package, benchmark scripts, all raw result data (D109, D110, D111 JSON outputs), and figure generation code are available at https://github.com/impliedDistraction/CAExperiments.

References

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, 80(1):505–515.

Bahar, I., Atilgan, A. R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, 2(3):173–181.

Byrom, J. (2025). Spectral decomposition of protein contact networks for unsupervised domain, hinge, and structural role analysis. *Preprint*, Earthform Research Lab.

Kundu, S., Melton, J. S., Sorensen, D. C., and Phillips, G. N. (2004). Dynamics of proteins in crystals: Comparison of experiment with simple models. *Biophys. J.*, 83(2):723–732.

Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, 247(4):536–540.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997). CATH—a hierarchic classification of protein domain structures. *Structure*, 5(8):1093–1108.