Allosteric Regulation — Spectral Communication in Proteins via the Latent Framework

Dr. Tamás Nagy

tnagyphd@gmail.com

draft • 2026-04-08

Abstract

Allostery couples distant sites in a macromolecule: ligand binding at one pocket reshapes dynamics and thermodynamics elsewhere. Classical narratives invoke structural pathways, entropy shifts, and population shifts among pre-existing states, but a portable quantitative language that connects atomistic detail to low-dimensional collective coordinates remains incomplete.

This paper models allosteric communication using Latent coordinates derived from elastic network models (ENMs) of protein structures. Normal modes furnish an orthogonal basis; the Latent Number ρ measures how compressible allosteric response vectors become in that basis, while the effective dimension N^* counts how many modes are needed to explain cross-site correlations above noise. The framework targets the spectral facet of allostery: which low-frequency motions participate in coupling, and how efficiently perturbations transmit across domains.

The results are complementary to mutational studies: genetics reveals which residues matter; Latent spectral analysis explains how those residues might be mechanically coupled at linear order.

The Platonic artifact elysium/fields/bio_allostery/platonic.py supplies **36 machine-checked lemmas** (real-arithmetic templates grouped under six biological headings below); they scaffold inequalities in the Latent formalism but do **not** by themselves constitute a full ENM calculus on PDB structures. A reproducible synthetic harness (numerical_validation.py) runs **16 regression checks** on three toy Kirchhoff–GNM topologies (globular random cloud, two-domain hinge, ring-like scaffold); all pass for the default seeds and cutoffs, with **topology-dependent** scalars as tabulated in §4. The Latent compression number ρ and coupling-specific N^* are defined in §2.4; the table in §4 reports the distinct spacing ratio $r = \lambda_2/\lambda_1$ and an eigenvalue-mass mode count tied to $\varepsilon = 0.9$.

1. Introduction

1.1 Allostery as a compression problem

High-resolution structures and molecular dynamics trajectories provide enormous state spaces, yet biologists summarize allostery with a handful of collective motions. That summary is a Latent hypothesis: a small basis explains most of the coupling between sites. The Latent framework makes this hypothesis measurable via ρ and N^*.

1.2 Elastic networks as a controlled testbed

Full MD simulations are expensive and force-field dependent. Gaussian ENMs provide a reproducible linearization: a Hessian of pairwise springs defines normal modes independent of solvent in the simplest treatments [3,4]. While ENMs omit anharmonicity, they are ideal for scaffolding spectral diagnostics and for isolating linearized communication from solvent friction.

1.3 Contributions

We encode ENM Latents in a Platonic proof artifact (36 real-arithmetic templates; see §3 scope), outline spectral–coupling inequalities at that formalization level, and benchmark three **synthetic** topology classes. Cross-domain links connect to neural coding efficiency and hub–spoke network models of signal routing.

1.4 Relation to structural biology workflows

Structural biologists already inspect correlated motions and difference-distance matrices. The Latent layer does not replace those views; it compresses them into two scalars (ρ, N^*) that travel across systems. A reviewer can compare an enzyme family's allosteric "dimensionality" without eyeballing hundreds of mode shapes.

2. Mathematical Framework

2.1 Coordinates and Hessian

Let positions $r_i \in \mathbb{R}^3$ for residues $i = 1, \ldots, n$ be given for a reference structure. Displacement coordinates $q \in \mathbb{R}^{3n}$ stack Cartesian perturbations. The ENM potential is

$$U(q) = \tfrac{1}{2} q^\top H q,$$

with Hessian H assembled from pairwise springs between nodes within a cutoff distance. Mass-weighting and rigid-body projection follow standard protocols documented in the code artifact.

2.2 Normal modes as Latent basis

Eigenvectors $u^{(\ell)}$ of H (after projection) diagonalize the harmonic energy. Latent coordinates $z_\ell = u^{(\ell)} \cdot q$ evolve independently under pure ENM dynamics. Allosteric perturbations are modeled as localized forces f applied at binding-site atoms; the linear response solves $Hq = f$.

2.3 Coupling tensor

For observables A, B linear in q (e.g., inter-residue distances), define the bilinear coupling

$$\mathcal{C}_{A,B} \ = \ \mathbb{E}\big[(A - \bar{A})(B - \bar{B})\big]$$

under a thermal ellipsoid defined by H and temperature T. Expanding A, B in modes yields coefficients α_ℓ, β_ℓ. The Latent coupling energy $\sum_\ell \alpha_\ell \beta_\ell \lambda_\ell^{-1}$ is the ENM analogue of mode-resolved allosteric strength.

2.4 Latent Number and effective dimension

Let v be the vector of modal contributions to a chosen coupling functional. Define ρ as the ratio of baseline variance to residual variance after optimal k-term truncation of v in decreasing magnitude order; N^* is the minimal k achieving error ε. This parallels other Latent papers while respecting the spectral weights λ_ℓ^{-1}.

Benchmark diagnostics (§4). The packaged regression script uses two **additional** scalars for reproducibility: a **spacing ratio** $r := \lambda_2/\lambda_1$ of the lowest nonzero Kirchhoff eigenvalues, and a **mode-count proxy** N^* from cumulative eigenvalue mass, $\sum_{\ell \leq k} \lambda_\ell / \sum_\ell \lambda_\ell \geq 0.9$ with $\varepsilon = 0.9$ fixed. These are not identical to coupling-vector ρ and coupling-truncation N^* above; they summarize the spectrum of the toy networks only.

2.5 Observables and probes

We standardize on two observable pairs: (i) a binding-site distance and a distal catalytic distance; (ii) a pair of dihedral proxies constructed from backbone atoms. These are linearized observables in the small-displacement regime, matching the ENM validity domain. Nonlinear observables require Jacobian extensions deferred to future work.

2.6 Thermal calibration

Temperature enters only through the fluctuation covariance, proportional on the mechanical subspace to a **generalized inverse** of H (Moore–Penrose pseudoinverse H^+ after rigid-body projection; for real symmetric H this coincides with the inverse on the range). Benchmark tables fix temperature relative to the spring scale so that **dimensionless ratios** (e.g. $r = \lambda_2/\lambda_1$) are independent of overall stiffness rescaling.

3. Formal Proof Chain

Scope. Section headings below are a **conceptual roadmap** aligned with platonic.py. The file proves short universally quantified real inequalities (via the Platonic kernel) whose names echo ENM biology; they should be read as **consistency templates**, not as a substitute for full structural proofs on protein-sized operators. Full ENM-specific statements (interlacing on contact deletions, pathwise bounds on atomistic graphs, etc.) remain to be formalized at that fidelity.

Group A — Normal mode properties (6 theorems). Symmetry and spectrum of H under uniform spring constants; invariance under global rigid motions after projection; interlacing when springs are removed; monotonicity of lowest nonzero frequency under stiffening; bounds on participation ratios; completeness of the mode basis in projected space.

Group B — Allosteric coupling (6 theorems). Bilinearity and Cauchy–Schwarz bounds for $\mathcal{C}_{A,B}$; positivity for diagonal observables; domain-wise additivity under block Hessians; stability under small perturbations of spring sets; scaling with temperature; consistency with linear response theory.

Group C — Signal transmission (6 theorems). Pathwise bounds on induced displacements along residue chains; decay laws with graph distance on ENM contact graphs; comparison between endpoint forcing and distributed loads; energy partition lemmas across modes; maximum coupling principles on trees; robustness to single-bond deletions.

Group D — Drug design hooks (6 theorems). Modeling ligands as localized springs shifts H; eigenvalue perturbation formulas; conditions for mode localization at binding sites; bounds on induced shifts in N^*; sensitivity of ρ to localized stiffening; composition rules for multiple ligands.

Group E — Communication efficiency (6 theorems). Definitions of intra- vs inter-domain tensors on partitions; dominance theorems when domain contact surface is small; inequalities relating ρ to domain-averaged coupling; sparsity patterns in mode-space projections; concentration of coupling in top modes; tradeoffs between stiffness and reachability.

Group F — Cross-domain (6 theorems). Neural coding morphisms mapping mode stacks to population vectors; hub–spoke graph models mirroring domain connectors; invariance lemmas under orthogonal mixes of modes; functoriality to fitness landscapes when mutational effects project onto ENM coordinates; consistency with information-theoretic channel capacities in linear Gaussian channels; composition with phylogenetic covariance models on sequence families.

Proof dependencies. Groups A–B are foundational. Group C builds on B with graph-distance structure. Group D is a perturbative layer atop A. Group E combines B with partitions. Group F is purely functorial, importing packaged inequalities from earlier groups.

4. Numerical Validation

We built Kirchhoff matrices for three **synthetic** topology classes in numerical_validation.py: a random globular cloud ($N = 50$), a two-domain assembly with a hinge ($N = 100$), and a ring-like scaffold ($N = 80$). For each, we diagonalized the network, recorded the spacing ratio $r = \lambda_2/\lambda_1$, the mode-count proxy N^*/N at 90% cumulative eigenvalue mass, and a coarse coupling from the active site to a distal site (sites 0 and $N - 1$ for the globular and two-domain models; sites 0 and $N/2$ for the ring scaffold) built from the first ten modes.

System class	$r = \lambda_2/\lambda_1$	N^*/N (90% eval. mass)	10-mode coupling $\|c\|$ (see text)
Globular	4.56	90%	3.4×10^{-5}
Multi-domain	15.5	91%	0.011
Ion channel	1.00	91%	0.025

Values: numerical_validation.py, default seed=42 and cutoffs as in code (2026-04-10 run).

Interpretation. Large r indicates a wide separation between the first two nonzero modes; the near-unity ion-channel value reflects near-degeneracy of the ring Laplacian's lowest modes. The high N^*/N values show that **90% eigenvalue mass** is a conservative tail in these dense toy graphs—not a claim that allostery itself requires half the modes. A separate check reports the fraction of the **unsigned** endpoint-coupling sum carried by the three lowest modes; on the globular instance this exceeds 20%, while the hinge and ring geometries violate a naive "three modes suffice" story—underscoring that this fraction is **not** monotone in topology.

Test harness (16/16). The sixteen checks comprise: three eigenvalue-positivity tests; three compression-threshold tests ($N^*/N < 1$); three positive-coupling tests; three "low-mode" fraction tests (against a 20% nominal floor—passed under the harness definition but not uniformly interpretable across topologies); one multi-domain intra- vs inter-domain comparison (declared pass if values agree within 5%); and three spectral-gap positivity tests. They are **smoke tests** for the script, not claims about experimental structures.

Robustness. Cutoff distance sweeps $\{8, 10, 12\}$ Å were not re-audited in this revision; expect r and N^*/N to move with graph density. PDB-based pipelines are future work.

5. Cross-Domain Connections

Neural coding. Population responses to structured stimuli sometimes concentrate variance in few principal components. The same ρ diagnostic compares neural and protein systems once observables are chosen analogously.

Network hubs. Inter-domain interfaces act like hub edges in graph abstractions of ENMs. Hub–spoke theorems in Group F formalize why interface mutants disproportionately alter allostery.

6. Discussion

The Latent ENM story is spectral: allostery, at linear order, is which modes participate and with what weights. The definitions of ρ and coupling-specific N^* in §2.4 aim to quantify that participation; the §4 table uses **different** spectral summaries (r and eigenvalue-mass N^*) tied to the released harness. On the toy two-domain geometry, intra- and inter-domain couplings sampled by the script are **comparable** (within a few percent), so "intra always dominates inter" is **not** a reported empirical law here—only a **conditional** theme in Group E when contact surface is small.

Limitations are clear: ENMs omit solvent-mediated pathways, electrostatics, and explicit conformational switching. Theorems about H are not claims about full MD.

Future work will merge MD snapshots with ENM Latents via weighted bases, incorporate anharmonic corrections along slow modes, and map sequence variation to shifts in N^* across orthologs.

Non-claims. We do not predict free energies of binding or kinetics; the paper addresses linear response geometry only.

Practical reading for medicinal chemists. When N^*/N is already large, fragment linking strategies that further soften interfaces may hurt specificity by delocalizing coupling; conversely, when ρ is high, localized stiffening may produce outsized long-range shifts—exactly the regime where careful SAR around a hinge pays off.

Ethics and data. The §4 numerical benchmark uses **synthetic** coordinates only (no PDB entries in that harness). Any future PDB-based pipeline would rely on public-domain structures; no patient data appear. Conclusions are mechanistic, not clinical.

During the preparation of this work the author used large language models to assist with manuscript drafting and organization. The author reviewed and edited the content and takes full responsibility for the publication.

References

1. Monod, J., Wyman, J. & Changeux, J.-P. "On the nature of allosteric transitions." *J. Mol. Biol.*, 1965.
2. McCammon, J. A. et al. "Dynamics of folded proteins." *Nature*, 1977.
3. Tirion, M. M. "Large amplitude elastic motions in proteins." *Phys. Rev. Lett.*, 1996.
4. Bahar, I., Lezon, T. R., Yang, L.-W. & Eyal, E. "Global dynamics of proteins." *Bioinformatics*, 2010.
5. Cover, T. M. & Thomas, J. A. *Elements of Information Theory.* Wiley, 2006.
6. Halabi, N. et al. "Protein sectors: evolutionary units of three-dimensional structure." *Cell*, 2009.
7. del Sol, A., Tsai, C.-J., Ma, B. & Nussinov, R. "The origin of allosteric functional modulation." *Structure*, 2009.
8. Motlagh, H. N. et al. "The ensemble nature of allostery." *Nature*, 2014.
9. Atilgan, A. R. et al. "Anisotropy of fluctuation dynamics of proteins." *Biophys. J.*, 2001.
10. Zhang, Z. & Wriggers, W. "Local feature analysis: a statistical theory for constitutional and allosteric effects." *Structure*, 2015.
11. Eckmann, J.-P. & Ruelle, D. "Ergodic theory of chaos and strange attractors." *Rev. Mod. Phys.*, 1985.
12. Go, N., Noguti, T. & Nishikawa, T. "Dynamics of a small globular protein." *Proc. Natl. Acad. Sci. USA*, 1983.
13. Levy, R. M., Srinivasan, A. R., Olson, W. K. & McCammon, J. A. "Quasi-harmonic method for studying very low frequency modes in proteins." *Biopolymers*, 1984.
14. Merris, R. "Laplacian matrices of graphs: a survey." *Linear Algebra Appl.*, 197–198, 143–176, 1994.