Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Smart Medicine
DOI: 10.54254/2753-8818/2026.HZ37271

Predicting Molecular Dynamics Contact Persistence from Static Structures for Weighted Gaussian Network Models

Minghan Lyu

School of Software, Shandong University, Jinan, China
202300300382@mail.sdu.edu.cn

Abstract. Uniform Gaussian network models (GNMs) are inexpensive and interpretable, yet they assign the same coupling to contacts that behave quite differently in molecular dynamics (MD). This work examined a narrower question: can the fraction of MD frames in which a contact survives be inferred from one static structure and then used as a GNM weight? Contact-persistence and Cα root-mean-square fluctuation (RMSF) labels were reconstructed from three ATLAS replicas for 200 protein chains. Evaluation used an Evolutionary Classification of Protein Domains (ECOD) component-disjoint split of 132/38/30 proteins. On the 30-protein test set, the validation-locked histogram gradient boosting (HGB) model reached a protein-macro mean absolute error of 0.0856 and a Spearman correlation of 0.8878. Most of that predictability came from geometry: the margin over distance-isotonic regression was statistically supported but small. With predicted persistence mapped to nonnegative E8 couplings, RMSF Spearman correlation rose from 0.7170 for uniform GNM to 0.7229. The component-level gain was 0.00536 (95% bootstrap confidence interval, 0.00126-0.00962; Holm-adjusted sign-flip $p = 0.0173$). A continuous-persistence oracle reached 0.7657. Thus, persistence is a useful and auditable edge target, but the present mapping recovers only a modest, protein-dependent improvement rather than a new state of the art.

Keywords: protein flexibility, Gaussian network model, molecular dynamics, contact persistence, interpretable machine learning

1. Introduction

Proteins function as conformational ensembles, not as single static structures. Atomistic molecular dynamics (MD) can resolve local fluctuations and collective motion, but its sampling and storage costs become restrictive when many proteins or design candidates must be screened. A Gaussian network model (GNM) offers a transparent analytic alternative: it converts a Cα contact graph into fluctuation estimates [1, 2]. The usual uniform coupling is also its obvious weak point, because contacts that share a geometric cutoff need not remain equally stable during MD.

Heterogeneous and distance-dependent elastic networks have addressed this mismatch in several ways [3-5]. Learned contact maintenance [6] and ensemble-derived couplings [7, 8] show that edge-specific information can alter network behavior, although those approaches either simplify stability to a class label or require dynamics for the target protein. Replicated trajectories in ATLAS [9],

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ37271

together with newer MD collections [10] and direct flexibility predictors [11-15], now make a cross-protein test possible. The unresolved issue is not whether flexibility can be learned at all, but whether an interpretable dynamic edge statistic adds anything after a strong geometric baseline.

This paper uses contact persistence as that intermediate statistic. For a residue pair in a static E10 candidate graph, it is the fraction of retained MD frames in which the Cα distance remains at most 8 Å. The quantity is continuous, bounded, and measurable across independent replicas. It leads directly to three questions:

1. Can continuous contact persistence be predicted across structurally separated proteins more accurately than by geometry-only baselines?

2. Does true or predicted persistence contain information that improves a fixed analytic GNM relative to uniform or distance-only couplings?

3. Are the resulting effects robust enough, statistically and computationally, to support a useful empirical conclusion?

This work answers them on a frozen 200-chain ATLAS subset with an ECOD-component-disjoint train/validation/test protocol. The persistence model was locked on validation data, GNM scales were fitted only on training proteins, and confirmatory resampling used ECOD components rather than edge or residue rows. The benchmark was designed this way because a large edge table does not create a large number of independent proteins.

The work is an empirical benchmark, not a new network architecture. It reconstructs replicated persistence and RMSF labels, compares geometry-only, tabular, message-passing, direct neural, and analytic baselines, and separates the oracle opportunity from the gain actually recovered by prediction. That separation is central to the result: geometry explains most persistence variation, predicted persistence gives a small mean improvement over uniform GNM, and the oracle gap is much larger.

2. Literature review

2.1. Classical and heterogeneous elastic networks

Uniform elastic networks begin with a deliberately simple premise: native contacts share one coupling. Tirion showed that this approximation could describe large-amplitude motion in an atomic network [1]; Bahar et al. then formulated the Cα GNM used for residue-level fluctuations [2]. Later studies fitted contact parameters, changed interaction ranges, or calibrated ENMs against MD [3-5]. This work uses of persistence borrows the idea of unequal contacts, but not a literal stiffness interpretation. The label records occupancy under the fixed 8 Å protocol.

2.2. Learning contacts and ensemble-derived couplings

Putz and Brock used learned maintained-versus-breaking contact labels to modify an ENM [6]. This work also learns at the edge level, but retain the fraction of frames in contact and its three-replica variation instead of imposing a binary class. References [7, 8] derive richer weights from ensembles or positional covariance, which requires dynamics from the protein under study. This work held-out chain contributes only one static structure; MD provides labels and the oracle, not predictor input.

DOI: 10.54254/2753-8818/2026.HZ37271

2.3. Molecular-dynamics datasets and direct flexibility prediction

Each ATLAS protein has three standardized replicas, making disagreement in persistence and RMSF directly measurable [9]. mdCATH offers broader size and temperature coverage [10], but its domain definitions, simulations, and preprocessing differ from the frozen ATLAS contract. This work did not pool the datasets. Cross-dataset validation therefore requires a separately designed and locked experiment.

Recent work spans cryo-EM-informed flexibility prediction, flexibility-conditioned design, fast flexibility prediction, and confidence-weighted Gaussian networks [11-15]. Because these studies differ in inputs, targets, data versions, and splits, their published scores do not support a direct ranking against ours. This work therefore includes a direct RMSF predictor evaluated on the same frozen split.

3. Materials and methods

3.1. Frozen ATLAS-paper200 cohort and split

ATLAS v1 contains 1,390 protein chains, each with three independent all-atom MD replicas of approximately 100 ns [9]. This work did not analyze the full release. Before formal evaluation, This work frozes 200 chains (50 from each length quartile) to keep the computation tractable. All claims below refer to this cohort.

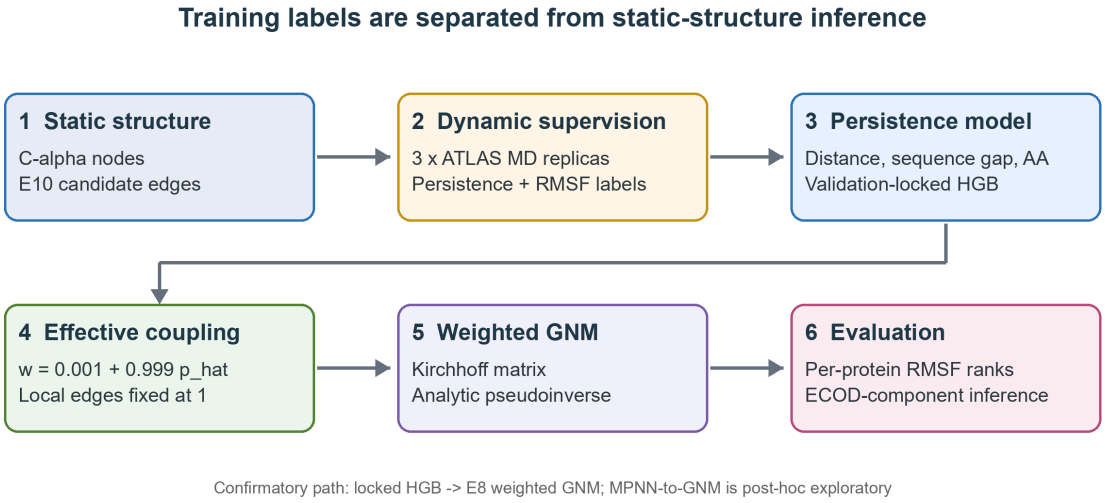

Figure 1. Study design. Replicated molecular dynamics supplies edge-persistence and root-mean-square-fluctuation supervision during model development. Inference for a new protein begins from one static structure, predicts edge persistence, maps it to nonnegative couplings, and solves an analytic weighted Gaussian network model

For each replica, the frozen processing contract retained 1,000 frames at times of at least 100 ps. The chain, not the frame, was the analysis unit. ECOD-based connected components were formed before splitting: 132 proteins (71 components) for training, 38 (38 components) for validation, and 30 (28 components) for testing. Components do not cross subsets. This work used validation data to choose the model and lock global scales, then opened the test split. Incomplete CATH annotations

DOI: 10.54254/2753-8818/2026.HZ37271

are reported only as context, not as a second confirmatory benchmark. Figure 1 shows the development and inference workflow.

3.2. Contact-persistence and RMSF labels

Each residue was represented by its Cα atom. Static pairs within 10 Å formed the E10 candidate universe, while E6 and E8 membership was retained for cutoff sensitivity checks. This work learned only nonlocal edges satisfying $|i - j| \geq 4$. Edges satisfying $|i - j| \leq 3$ later received a fixed unit GNM weight. For a separate coverage audit, this work also measured the event-weighted recall of dynamic nonlocal Cα contacts (at most 8 Å and lasting at least 100 ps) in a wider static E12 graph; E12 was not used for prediction or GNM evaluation.

For candidate pair (i, j) in replica r, persistence is the fraction of retained frames in which the Cα distance does not exceed 8 Å:

$$p_{ij}^{(r)} = \frac{1}{T_r} \sum_{t=1}^{T_r} \mathrm{I}[d_{ij}^{(r,t)} \leq 8\ \text{Å}]. \quad (1)$$

In Equation (1), T_r is the retained-frame count and $d_{ij}^{(r,t)}$ is the Cα distance between residues i and j at frame t. The indicator is one when that distance is at most 8 Å. This work averaged the label across the three replicas but retained the individual values and their dispersion for audit. The E10 graph captured a median 0.9876 of dynamic contact events in the frozen coverage check. The predictor therefore estimates persistence only for static E10 candidates; it does not discover contacts that appear outside that graph.

This work reconstructed Cα RMSF separately for each replica after residue mapping, alignment, and periodic-boundary handling, and then averaged the replicas. Agreement with the official ATLAS RMSF served as an implementation check. Across 200 chains, the median reconstructed-versus-official Pearson correlation was 0.999894 and median normalized MAE was 0.007652. Replica agreement was lower, as expected: median Spearman correlation was 0.956885 for persistence and 0.897314 for RMSF. Every chain passed the formal data gate. Predefined quality subsets were kept for sensitivity analysis instead of removing inconvenient test cases after evaluation.

3.3. Persistence predictors and model locking

Every persistence predictor saw only information available from a static structure. Tabular models used Cα distance, log-transformed sequence separation, and a symmetric amino-acid-pair representation. The neural graph added 21-category residue identity, normalized sequence position, 16 radial-basis distance encodings over 0-10 Å, and log sequence separation.

The benchmark compared a global mean, distance-binned mean, distance isotonic regression, an edge multilayer perceptron (EdgeMLP), histogram gradient boosting (HGB), and a two-layer message-passing neural network (MPNN). The HGB model used 200 boosting iterations, learning rate 0.05, at most 15 leaf nodes, minimum 10 samples per leaf, and L2 regularization 0.001. Neural models used hidden dimension 64, two message-passing layers, dropout 0.1, AdamW with learning rate 0.001 and weight decay 0.0001, a maximum of 100 epochs, and validation early stopping with patience 15. Neural seeds were 17, 42, and 2026 where reported.

Protein-macro validation MAE determined the formal model before test access. HGB with distance, sequence-gap, and amino-acid features reached 0.084182 and was locked. Results from additional neural seeds and the later MPNN-to-GNM route remain supplementary or post hoc; they do not replace that decision.

DOI: 10.54254/2753-8818/2026.HZ37271

3.4. Persistence-weighted GNM

For the formal weighted GNM, each nonlocal E8 edge received the nonnegative effective coupling in Equation (2):

$$w_{ij} = 0.001 + 0.999\,\mathrm{clip}(\hat{p}_{ij}, 0, 1). \tag{2}$$

The predicted persistence $\hat{p}_{ij}$ is clipped to the unit interval; the 0.001 floor prevents a retained edge from receiving zero weight. Local edges satisfying $|i - j| \leq 3$ keep unit weight.

For an edge with coupling w_{ij}, the off-diagonal Kirchhoff entry is its negative and each diagonal entry is the sum of incident weights. The unscaled residue fluctuation is then:

$$u_i = \sqrt{(\Gamma^{+})_{ii}}. \tag{3}$$

In Equation (3), u_i is the unscaled fluctuation for residue i, and the superscript plus on Γ^{+} denotes the Moore-Penrose pseudoinverse. Each method and cutoff used one multiplicative scale fitted with equal protein weight across the 132 training chains. Per-protein rescaling on the test set was not allowed. The locked E8 scales were applied unchanged; E6 and E10 remained sensitivity analyses and could not displace the pre-specified E8 result.

The GNM baselines used uniform, inverse-distance, and inverse-square-distance couplings. A continuous oracle substituted MD-derived persistence for $\hat{p}_{ij}$. This work also evaluated a direct two-layer MPNN-to-RMSF model and an RMSF-only differentiable GNM ablation. Together, these comparisons separate edge-target quality from the behavior of the analytic mapping.

3.5. Metrics, statistical tests, and runtime

Persistence was evaluated primarily by protein-macro MAE, with protein-macro Spearman and Pearson correlations, Brier score, and event-level average precision as secondary measures. Because highly persistent edges are common, AUPRC values were interpreted together with prevalence and were not used as the main conclusion.

RMSF was evaluated primarily by the mean within-protein Spearman correlation. Pearson correlation, normalized MAE, and win rate were secondary. Three comparisons were specified and locked before test opening: continuous-oracle GNM versus the validation-selected uniform GNM; locked HGB persistence versus distance isotonic regression; and E8 HGB-persistence GNM versus the same uniform GNM. Effect estimates and 95% confidence intervals used 10,000 bootstrap resamples at the ECOD-component level. Two-sided paired sign-flip tests also used 10,000 permutations and were adjusted across the three comparisons with Holm's procedure.

Each protein was timed 30 times after five warm-up runs in one CPU process. The clock started with canonical tables and structure data already in memory and covered prediction plus the GNM solve. File access, parsing, and MD label generation were excluded. These are in-memory execution times, not estimates of speedup over MD.

DOI: 10.54254/2753-8818/2026.HZ37271

4. Results

4.1. Label reliability and benchmark integrity

The frozen gate excluded none of the 200 chains. Median event coverage was 0.987598 for the E10 candidate graph and 0.996079 in the wider E12 coverage audit. Persistence also reproduced more closely across replicas than RMSF rank (median Spearman 0.956885 versus 0.897314). Reconstructed and official RMSF agreed closely (median Pearson 0.999894), serving as a consistency check on RMSF label reconstruction; this comparison does not independently validate the ATLAS force field or simulation time scale. Figure 2 and Figure 3 show the quality distributions and the component-disjoint split.

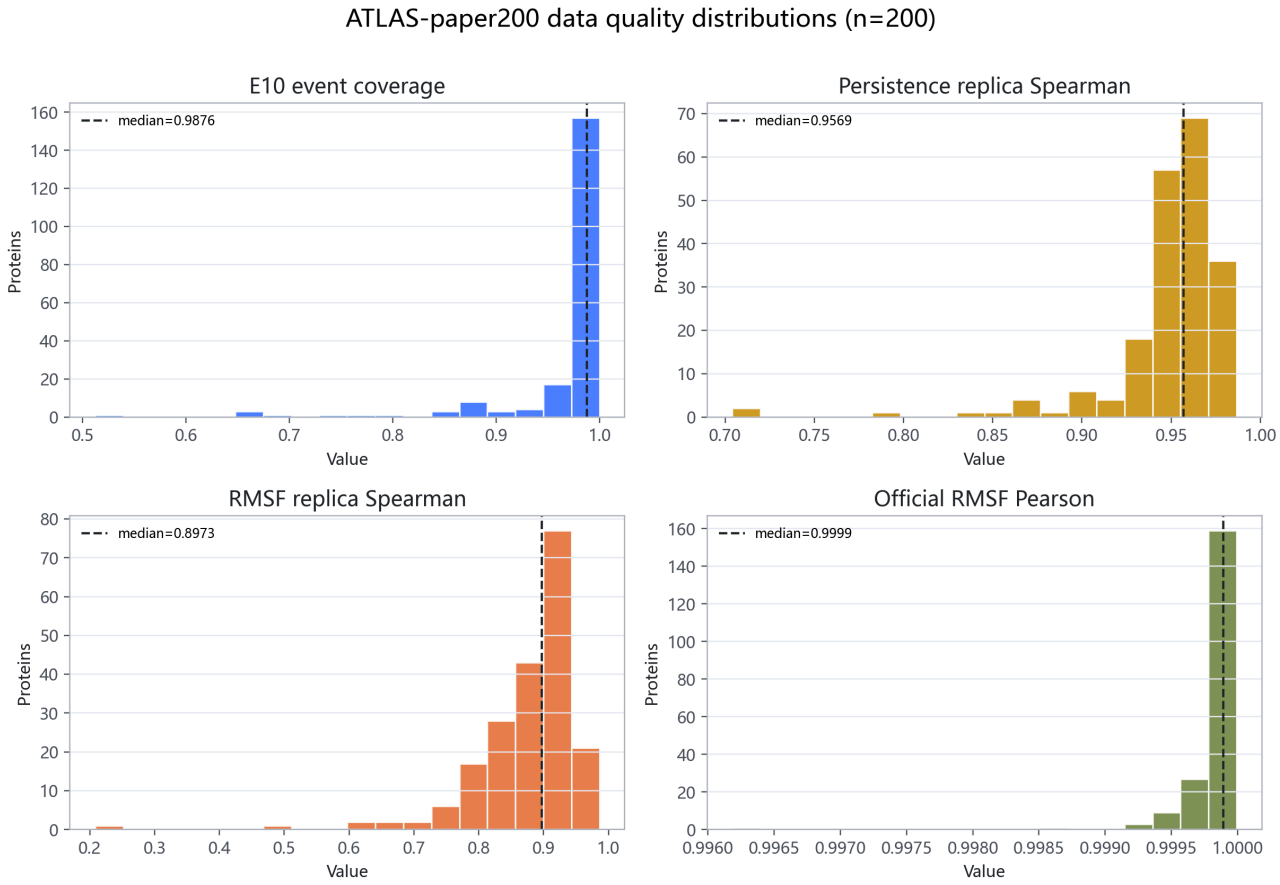

Figure 2. Label and data-quality distributions for the frozen 200-chain cohort. Dashed lines mark medians, and the long tails motivate predefined quality sensitivity analyses rather than post hoc exclusion

DOI: 10.54254/2753-8818/2026.HZ37271

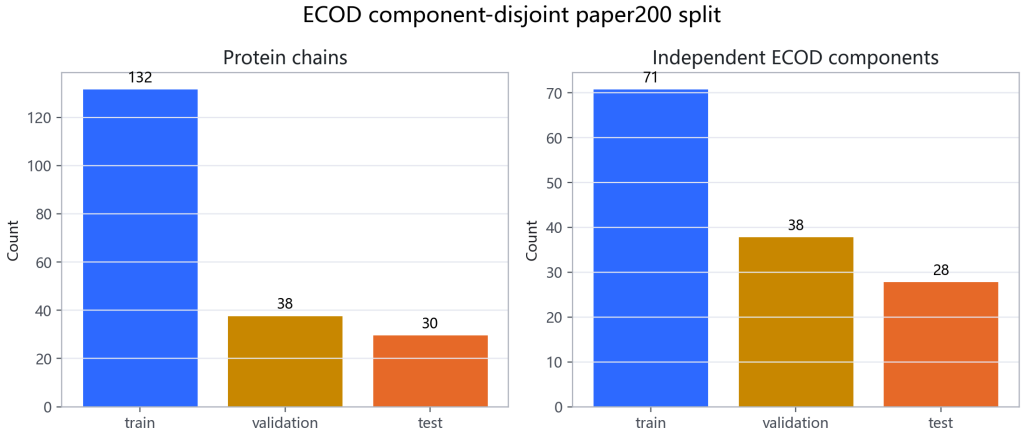

Figure 3. Protein and independent-component counts in the ECOD-component-disjoint split. Confirmatory inference uses the 28 test components rather than treating all edge or residue rows as independent

The 30 test proteins belong to 28 ECOD components (Table 1), so edge-level sample size would give a misleading impression of precision. The same caution applies to scope: these 200 chains are a frozen subset of the 1,390-chain ATLAS v1 release.

Table 1. Frozen cohort, split, and label-quality summary

Quantity	Train	Validation	Test	Full cohort / note
Protein chains	132	38	30	200 total
ECOD components	71	38	28	zero cross-split overlap
ATLAS replicas per chain	3	3	3	1,000 retained frames per replica
Median E10 event coverage	-	-	-	0.987598
Median persistence replica Spearman	-	-	-	0.956885
Median RMSF replica Spearman	-	-	-	0.897314
Median official RMSF Pearson	-	-	-	0.999894

4.2. Contact persistence is predictable, but geometry is the dominant signal

The locked HGB model reached a test protein-macro MAE of 0.085623, Spearman correlation of 0.887829, and Pearson correlation of 0.930927 (Table 2). Its Brier score was 0.025810; AUPRC was 0.964136 at a persistence threshold of 0.5 and 0.942615 at 0.8. Persistent contacts are common in the static graph, however, so those classification values look stronger than the underlying class balance warrants. This work therefore base the interpretation on continuous MAE and rank correlation.

The difficult comparison was not HGB versus a global mean, but HGB versus distance alone. Distance bins reached MAE 0.094342 and Spearman 0.871899; isotonic regression reached 0.089221 and 0.873508. Against isotonic, locked HGB improved component-level oriented MAE by 0.003721 (95% CI 0.002042-0.005650; Holm-adjusted sign-flip $p = 0.000300$). HGB changes features and model form together, so the increment cannot be attributed to either. Cα separation already supplied most contact ordering.

DOI: 10.54254/2753-8818/2026.HZ37271

The matched seed-42 ablation exposed dependence on distance (Figure 4): removing it raised MPNN MAE to 0.367398 and lowered Spearman to 0.385789. Removing amino-acid identity changed little and slightly improved the test point estimate. This result is specific to the present features and sample, not evidence that chemistry is irrelevant. Message passing modestly outperformed EdgeMLP at the same seed.

Table 2. Contact-persistence prediction on the 30-protein test set

Persistence predictor	Test MAE (lower)	Test Spearman (higher)	Interpretation
Global mean	0.410852	undefined	prevalence-only reference
Distance bins	0.094342	0.871899	strong geometry baseline
Distance isotonic	0.089221	0.873508	confirmatory distance baseline
EdgeMLP	0.088525	0.870364	no message passing
HGB, locked	0.085623	0.887829	formal persistence model
MPNN ensemble	0.086052	0.891459	supplementary neural model
MPNN without distance	0.367398	0.385789	distance ablation

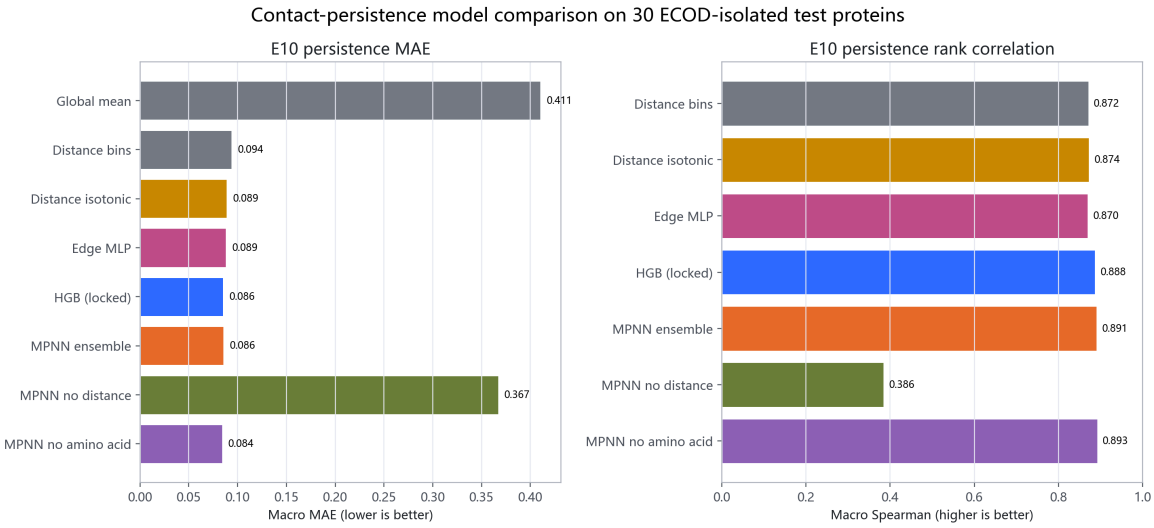

Figure 4. Persistence-prediction performance on the 30 ECOD-isolated test proteins. Distance-only baselines are strong; removing distance causes a large failure, whereas amino-acid removal has little effect

4.3. Persistence weighting gives a small GNM improvement and a larger oracle opportunity

Uniform GNM reached a test protein-macro RMSF Spearman correlation of 0.716999. The validation-locked HGB model, evaluated through the pre-specified E8 GNM, yielded 0.722930 (Table 3). At the ECOD-component level, the mean gain was 0.005363 (95% bootstrap interval 0.001260-0.009616; Holm-adjusted sign-flip $p = 0.017298$). Twenty test proteins improved and ten did not; the protein-level mean was 0.005931. The average effect is supported, but neither large nor universal.

Two chains show the spread behind the mean. For 1xeo_A, Spearman rose from 0.676753 to 0.717429 (+0.040676); for 5a5y_A it fell from 0.687272 to 0.674293 (-0.012979). These post-

DOI: 10.54254/2753-8818/2026.HZ37271

evaluation examples have opposite signs; they do not define a subgroup. The count of 20 improvements and 10 declines is the evidence for heterogeneity.

The continuous-persistence oracle reached Spearman 0.765700, a component-level gain of 0.045023 over uniform GNM (95% CI 0.030135-0.060934; Holm-adjusted $p = 0.000300$). The +0.045023 effect is 8.4 times the locked HGB effect within the same target and linear map. It is not an upper bound for other RMSF methods. Because predictor, map, and target were not varied separately, their contributions to the gap remain unresolved.

On the prespecified Spearman endpoint, uniform GNM scored 0.716999; inverse-distance and inverse-square scored 0.654842 and 0.485948. HGB-GNM had the higher Spearman point estimate than direct MPNN-to-RMSF (0.722930 versus 0.693517), while normalized MAE favored the direct model (0.547586 versus 0.673239). This contrast was not among the three locked comparisons, so both metrics are reported descriptively.

Table 3. Root-mean-square-fluctuation prediction on the 30-protein test set

RMSF method	Spearman	Pearson	NMAE (lower)	Role
Uniform GNM	0.716999	0.722423	0.675352	formal simple baseline
Inverse-distance GNM	0.654842	0.685218	0.679126	geometry-weighted baseline
Inverse-square GNM	0.485948	0.568463	0.666128	geometry-weighted baseline
HGB-persistence GNM, E8	0.722930	0.726965	0.673239	formal primary result
HGB-persistence GNM, E10	0.728645	0.730911	0.675352	cutoff sensitivity only
Continuous-persistence oracle	0.765700	0.774202	0.658783	upper-bound reference
Direct MPNN-to-RMSF	0.693517	0.715341	0.547586	direct neural baseline
RMSF-only differentiable GNM	0.258511	0.373802	0.649566	supervision ablation

4.4. Sensitivity, exploratory neural weighting, and runtime

All predefined quality filters retained a positive HGB-GNM minus uniform effect. With E10 coverage of at least 0.8 or 0.9, or RMSF-replica Spearman of at least 0.7, component gains were 0.00540-0.00592 and no bootstrap interval crossed zero. Low-quality tails alone do not explain the mean, but the filters do not imply benefit for every protein.

Cutoff choice moved Spearman more than the quality filters. With the same HGB predictions, E6 gave 0.564002 and E10 gave 0.728645; E6 may be too sparse. E8 remains formal because it was fixed before test opening, whereas E10 is a sensitivity result. Later MPNN-to-GNM runs averaged 0.730141 ± 0.000918 across three seeds and reached 0.730367 as an ensemble. Both are post hoc exploratory, not replacements for the primary HGB analysis.

Typical in-memory CPU runs were short: HGB-GNM took a median 22.0 ms for 76 residues, 27.5 ms for 149, 40.7 ms for 227, and 103.6 ms for 453. On the 2,128-residue stress test it required 2.91 s, mostly for the GNM pseudoinverse. The MPNN ensemble was slower for short and medium chains and reached about 3.06 s on the longest. These timings support practical static-structure inference. They omit I/O, structure parsing, and historical MD generation, so they are not an MD speedup measurement.

Figure 5 shows the full RMSF rank comparison, and Figure 6 reports the three component-level confirmatory effects with uncertainty intervals.

DOI: 10.54254/2753-8818/2026.HZ37271

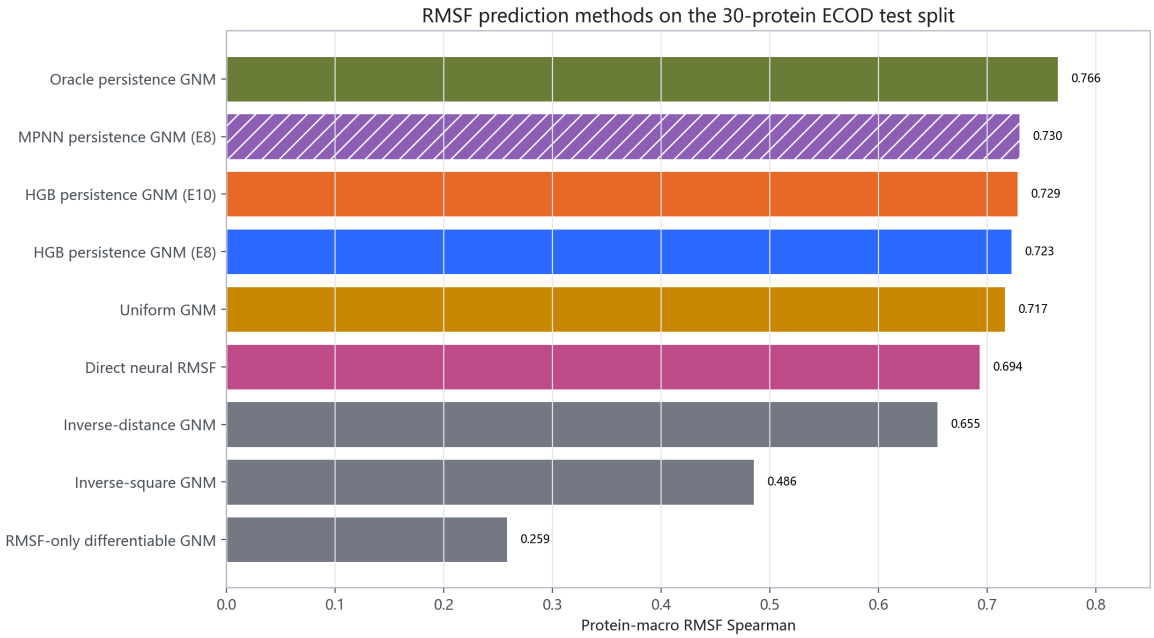

Figure 5. Protein-macro RMSF Spearman correlation on the frozen test split. The hatched MPNN-to-GNM result is post hoc exploratory; E10 is a cutoff sensitivity result rather than the pre-specified E8 primary result

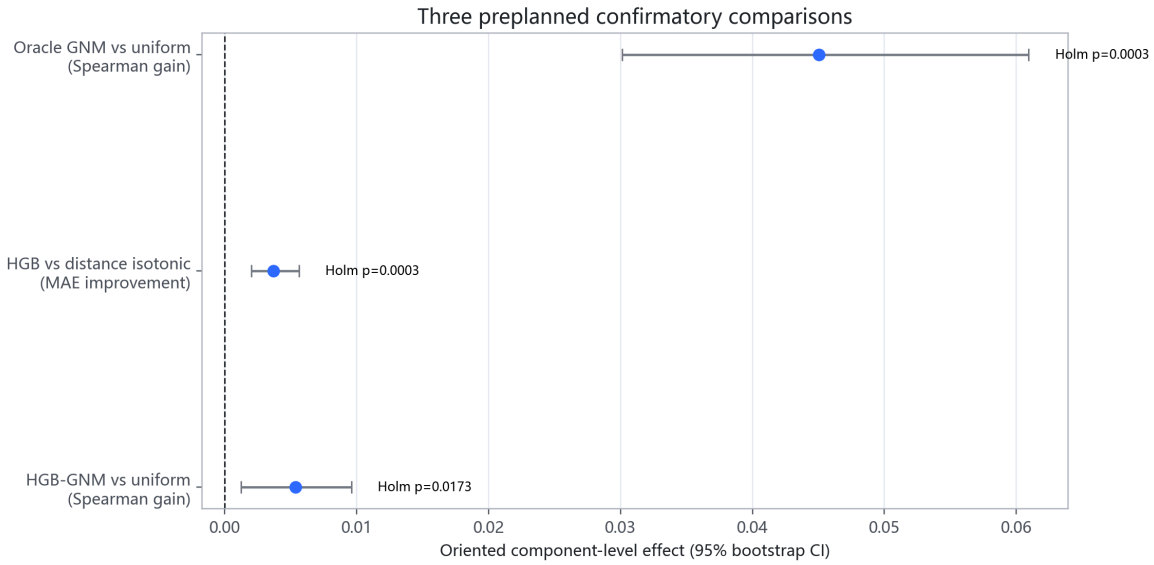

Figure 6. ECOD-component-level effects for the three pre-specified comparisons. Points show mean oriented effects, bars show 95% component-bootstrap intervals, and labels show Holm-adjusted paired sign-flip probabilities

5. Discussion

5.1. What the benchmark establishes

Across the persistence comparisons, Cα separation was the recurring source of signal. Distance-isotonic regression reached a test Spearman correlation of 0.873508, while removing distance from

DOI: 10.54254/2753-8818/2026.HZ37271

the matched MPNN reduced it to 0.385789. The locked HGB model reached 0.887829. Learning therefore refined an already strong distance-based ordering of contacts, with a comparatively small margin beyond geometry.

The locked and oracle couplings changed RMSF ranking on very different scales. HGB-to-E8 gave a +0.005363 component-level mean change in RMSF Spearman, whereas substituting MD-derived persistence gave +0.045023. The label is mechanically informative in this formulation, although the predicted-weight network recovers little of its oracle effect. These results do not separate errors in persistence prediction from limitations of the linear coupling map or target definition.

The component analysis supports an average effect, not a consistent protein-level benefit. Its positive direction was unchanged in the predefined quality subsets, but 10 of the 30 test proteins performed worse than under uniform GNM. This work interprets this as a small cohort-average shift within the frozen paper200 protocol. The unresolved negative cases prevent extending that conclusion to an arbitrary protein.

5.2. Relation to previous ENM and learning work

The strength of uniform GNM is consistent with the classical view that topology is an effective prior for low-frequency fluctuations [1, 2]. On the same E8/RMSF test, inverse-distance and inverse-square couplings scored 0.654842 and 0.485948. Their poor performance here does not test all distance-dependent ENMs [4]. The 0.765700 oracle shows that MD-derived edge information can alter residue ranking under this map. That direction is consistent with target-calibrated and ensemble-derived ENM studies [3, 5, 7, 8], although persistence remains only one summary of the trajectories.

Putz and Brock used learned maintained-versus-breaking contact labels to modify an ENM [6]. This work instead retains contact fraction and replica spread. At inference, a held-out protein supplies one static structure; methods in [7, 8] derive weights from target dynamics, whereas ours does not use those dynamics as predictor input. This convenience came with a small realized effect: +0.005363 component-level RMSF Spearman, versus +0.045023 from MD persistence in the same map.

Scores in [11-15] use different data releases, targets, normalizations, inputs, and splits and cannot be ranked with ours. The controlled comparison is same-split direct RMSF MPNN: its Spearman was 0.693517 versus 0.722930 for HGB-GNM, while normalized MAE was 0.547586 versus 0.673239. On these point estimates, edge weighting gave the higher within-protein rank correlation, whereas direct prediction gave the lower normalized error; neither establishes a cross-study winner.

5.3. Why is the realized gain small

Three stages could hide mechanically important errors. Persistent contacts dominate E10, so overall MAE 0.085623 and Spearman 0.887829 can coexist with mistakes on rarer transient edges. Clipped predictions then enter one linear, nonnegative map, and RMSF comes from the full Kirchhoff pseudoinverse. An error at a sparse inter-domain edge may therefore outweigh several errors in a redundant cluster; this work did not test this mechanism. The 100 ns replicas and 8 Å contact definition also capture one protocol-dependent view of dynamics, not a universal mechanical parameter.

A follow-up should connect each E8 edge error to its effect on modes or residue ranks. A newly locked test could compare mechanically weighted losses and monotone maps, with MPNN-to-GNM

DOI: 10.54254/2753-8818/2026.HZ37271

specified in advance. 1xeo_A (+0.040676) and 5a5y_A (-0.012979) provide opposite cases for structural inspection, but the aggregate analysis cannot locate the source of either change.

5.4. Limitations

Paper200 contains 200 frozen ATLAS v1 chains, 50 per length quartile, not all 1,390. ECOD components do not cross the 132/38/30 split. Incomplete CATH annotations are metadata; no CATH-specific retraining or external CATH test was run.

Persistence here is the three-replica mean fraction of frames in which a Cα pair is within 8 Å. Force field, trajectory length, sampling, alignment, or contact rule would change it. Prediction was limited to static E10 candidates; their median event coverage was 0.987598, but contacts outside E10 were never targets. This is a protocol-defined statistic, not a microscopic spring constant.

Methods [11-15] were not rerun on the frozen split, so no state-of-the-art ranking follows. MPNN-to-GNM was evaluated after the locked HGB test and remains post hoc exploratory. Runtime omits loading, parsing, and MD label generation; even the 2,128-residue stress test is an in-memory numerical check, not an MD speedup comparison.

6. Conclusion

The frozen test gives a narrow answer. HGB predicted persistence with Spearman 0.887829, mostly from distance. Its E8 weights moved mean RMSF Spearman from 0.716999 to 0.722930: +0.005363 at component level (95% CI 0.001260-0.009616; Holm-adjusted p = 0.017298), with 20 proteins improving and 10 declining. MD persistence in the same map reached 0.765700. The label contains more usable information than the current predictor and map recover. For paper200, persistence weighting is an auditable but modest extension of uniform GNM, not a general replacement. That claim needs a newly locked external cohort and matched inputs, labels, and splits.

References

[1] Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical Review Letters, 77*(9), 1905-1908. https://doi.org/10.1103/PhysRevLett.77.1905

[2] Bahar, I., Atilgan, A. R., & Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design, 2*(3), 173-181. https://doi.org/10.1016/S1359-0278(97)00024-2

[3] Erman, B. (2006). The Gaussian network model: Precise prediction of residue fluctuations and application to binding problems. *Biophysical Journal, 91*(10), 3589-3599. https://doi.org/10.1529/biophysj.106.090803

[4] Yang, L., Song, G., & Jernigan, R. L. (2009). Protein elastic network models and the ranges of cooperativity. *Proceedings of the National Academy of Sciences, 106*(30), 12347-12352. https://doi.org/10.1073/pnas.0902159106

[5] Orellana, L., Rueda, M., Ferrer-Costa, C., Lopez-Blanco, J. R., Chacón, P., & Orozco, M. (2010). Approaching elastic network models to molecular dynamics flexibility. *Journal of Chemical Theory and Computation, 6*(9), 2910-2923. https://doi.org/10.1021/ct100208e

[6] Putz, I., & Brock, O. (2017). Elastic network model of learned maintained contacts to predict protein motion. *PLOS ONE, 12*(8), e0183889. https://doi.org/10.1371/journal.pone.0183889

[7] Yu, C. C., Raj, N., & Chu, J. W. (2022). Edge weights in a protein elastic network reorganize collective motions and render long-range sensitivity responses. *The Journal of Chemical Physics, 156*(24), 245105. https://doi.org/10.1063/5.0095107

[8] Yu, C. C., Raj, N., & Chu, J. W. (2023). Statistical learning of protein elastic network from positional covariance matrix. *Computational and Structural Biotechnology Journal, 21*, 2524-2535. https://doi.org/10.1016/j.csbj.2023.03.033

[9] Vander Meersche, Y., Cretin, G., Gheeraert, A., Gelly, J. C., & Galochkina, T. (2024). ATLAS: Protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Research, 52*(D1), D384-D392. https://doi.org/10.1093/nar/gkad1084

DOI: 10.54254/2753-8818/2026.HZ37271

[10] Mirarchi, A., Giorgino, T., & De Fabritiis, G. (2024). mdCATH: A large-scale MD dataset for data-driven computational biophysics. *Scientific Data, 11*(1), 1299. https://doi.org/10.1038/s41597-024-04140-z

[11] Song, X., Bao, L., Feng, C., Huang, Q., Zhang, F., Gao, X., & Han, R. (2024). Accurate prediction of protein structural flexibility by deep learning integrating intricate atomic structures and cryo-electron microscopy density information. *Nature Communications, 15*, 5538. https://doi.org/10.1038/s41467-024-49858-x

[12] Kouba, P., Planas-Iglesias, J., Damborsky, J., Sedlar, J., Mazurenko, S., & Sivic, J. (2025). Learning to engineer protein flexibility. *International Conference on Learning Representations (ICLR 2025)*. https://openreview.net/forum?id=L238BAx0wP

[13] Viliuga, V., Seute, L., Wolf, N., Wagner, S., Elofsson, A., Stühmer, J., & Gräter, F. (2025). Flexibility-conditioned protein structure design with flow matching. *Proceedings of the 42nd International Conference on Machine Learning, 267*, 61513-61533. https://proceedings.mlr.press/v267/viliuga25a.html

[14] Pražnikar, J. (2026). Fast prediction of protein flexibility. *Bioinformatics, 42*(4), btag175. https://doi.org/10.1093/bioinformatics/btag175

[15] Ferraz, M., Alesiani, F., & Christiansen, H. (2026). Confidence-weighted elastic Gaussian networks to predict protein flexibility. *GenBio 2026*. https://openreview.net/forum?id=uLpTqlBhuC