Spectral Graph Features for Reference-free RNA 3D Quality Assessment

Ying Zhu[1], Huaiwen Zhang[2], Vince D. Calhoun[3,†], Yuda Bi[4,†]

[1]Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, China

[2]School of Computer Science and Technology, Xidian University, Xi'an 710126, China

[3]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

[4]TReNDS Center, Georgia State University, Atlanta, GA, USA

[†]Corresponding authors: vcalhoun@gatech.edu, ybi3@gsu.edu

Abstract

Motivation: Existing RNA 3D structure quality assessment (QA) methods rely on local geometric descriptors or statistical potentials that evaluate atomic-level contacts but are blind to global topological coherence. This creates a critical failure mode—structures that are "locally correct but globally wrong"—where well-formed local helices mask misplaced domains and incorrect overall packing.

Results: We introduce SpecRNA-QA, a lightweight method that scores RNA 3D models using multi-scale spectral features derived from the graph Laplacian of inter-nucleotide contact networks. By computing eigenvalue distributions, heat-kernel traces, and spectral entropy across four distance scales with binary and Gaussian kernels, SpecRNA-QA captures global structural coherence inaccessible to conventional descriptors. In leave-one-out cross-validation on CASP16 (42 targets, 7,368 models), spectral features achieve median per-target Spearman $\rho = 0.69$ [95% CI: 0.64–0.73], significantly outperforming an internal geometry baseline ($\rho = 0.47$, $\Delta\rho = +0.22$, Wilcoxon $p = 1.2\times10^{-10}$). Compared against established *unsupervised* statistical potentials—which require no labeled data, unlike the supervised spectral model—rsRNASP outperforms on small-to-medium RNAs ($\rho = 0.67$ vs. 0.57, $\leq$200 nt). However, rsRNASP times out on most large RNAs (>200 nt), where SpecRNA-QA provides the strongest available quality signal ($\rho = 0.72$ vs. DFIRE 0.52), revealing clear complementarity between global-topological and local-energy scoring. A training-free heuristic using only three spectral statistics enables quality estimation without any labeled data.

Availability: SpecRNA-QA is available as a Python package at https://github.com/yudabitrends/specrnaq.

Contact: ybi3@gsu.edu

Supplementary information: Supplementary data are available online.

Keywords: RNA structure, quality assessment, spectral graph theory, graph Laplacian, heat kernel, CASP

1 Introduction

The prediction of RNA three-dimensional (3D) structures has advanced considerably in recent years, propelled by deep-learning methods such as RhoFold+ [3] and community-wide assessments including the Critical Assessment of protein Structure Prediction (CASP), rounds 15 [2] and 16 [1]. The latest CASP16 round attracted diverse predictors—from physics-based approaches to end-to-end neural architectures—yielding thousands of models per target spanning a wide quality spectrum. As the number and diversity of predicted models grow, *model quality assessment* (QA)—selecting the best model from a candidate set without access to the native structure—becomes an essential bottleneck in the structure prediction pipeline.

For proteins, QA methods such as QMEAN [12], DeepRank [13], and graph neural network approaches [14] are well established. RNA-specific QA, however, remains comparatively underdeveloped, reflecting both the smaller available training data and the distinct structural features of RNA—non-canonical base pairs, complex tertiary interactions, and multi-domain architectures.

Existing RNA QA methods fall into two broad categories. *Statistical potential methods*—rsRNASP [4], DFIRE-RNA [5], RASP [6], cgRNASP [7], and 3dRNAscore [8]—derive knowledge-based energy functions from observed atomic contact frequencies in experimentally determined structures. These methods excel at evaluating local stereochemistry (base stacking, hydrogen bonding, backbone torsion preferences) but are fundamentally insensitive to global topological correctness: a structure can score well if its local contacts match the training distribution, even when entire domains are misplaced. *Geometric descriptors* such as radius of gyration (R_g) and contact density capture overall shape but lack the resolution to distinguish models with similar compactness yet different folds.

This reveals a critical gap that we term the "locally correct but globally wrong" failure mode. Consider a 407-nucleotide RNA (CASP16 target R1248): the best model (lDDT = 0.54) and worst model (lDDT = 0.005) have nearly identical R_g (50.0 vs. 49.9 Å) and contact densities, yet differ by a median per-nucleotide C4′ deviation of 25 Å when superimposed. Geometry-based ranking reduces to random guessing ($\rho = -0.01$). This failure mode is particularly prevalent among large RNAs (>200 nt) where local helices may form correctly while overall domain arrangement is erroneous.

Spectral graph theory offers a principled framework for characterizing *global* network topology [16]. The eigenvalues of the graph Laplacian encode connectivity, symmetry, bottlenecks, and diffusion dynamics that no single scalar descriptor can capture. In molecular analysis, Laplacian spectra have been applied to protein fold classification [17], persistent homology of biomolecular shape [18], and topological data analysis of protein-ligand binding [19]. Persistent Laplacians have also been proposed as multi-scale topological descriptors of molecular structures [20]. However, spectral graph methods remain entirely unexplored for RNA quality assessment.

Here we present SpecRNA-QA, a multi-scale spectral approach to RNA 3D model QA. Our central finding is that **spectral graph features provide a complementary global topology signal that is especially valuable for large, multi-domain RNAs**—precisely the structures where existing methods fail and where the need for global coherence assessment is greatest. On CASP15 and CASP16 benchmarks (54 targets, 8,928 models), spectral features significantly outperform an internal geometry baseline ($\Delta\rho = +0.22$, $p < 10^{-10}$), with head-to-

head comparison against established statistical potentials (DFIRE-RNA, rsRNASP) revealing clear size-dependent complementarity. Specifically:

1. We introduce the first graph Laplacian spectral features for RNA QA, computed across eight contact graphs (4 distance cutoffs $\times$ 2 kernels) that span from local base stacking (8 Å) to domain-level contacts (15 Å).

2. Heat-kernel traces at intermediate diffusion times emerge as the most discriminative features, providing a physically interpretable bridge between local and global structure.

3. A training-free heuristic using only three spectral statistics confirms that the global topology signal exists even without labeled data, validating the spectral prior as an intrinsic quality indicator.

2 Related Work

Statistical potentials for RNA QA. Knowledge-based scoring functions remain the dominant approach to RNA quality assessment. rsRNASP [4] derives reference-state-based potentials from curated RNA structures, achieving strong performance on small RNAs. DFIRE-RNA [5] extends the distance-scaled finite ideal-gas reference state to RNA. RASP [6] provides rapid all-atom assessment. More recently, cgRNASP [7] introduced coarse-grained potentials for computational efficiency, and 3dRNAscore [8] developed a composite knowledge-based score. These methods share a common limitation: they evaluate local atomic environments in isolation, without explicitly modeling global topological structure.

Deep learning approaches. ARES [9] pioneered geometric deep learning for RNA scoring using equivariant neural networks. For proteins, graph neural network-based QA methods such as GraphQA [14] and VoroCNN [15] have shown that learned representations can capture non-local structural features. While RNA-specific deep learning QA methods remain scarce, the success of graph-based approaches in protein QA motivates the exploration of principled graph-theoretic features for RNA.

Structure validation and comparison metrics. Beyond scoring functions, RNA structure quality is assessed using validation tools such as MolProbity [31] for clash scores and backbone conformer analysis, and comparison metrics including MCQ (Mean of Circular Quantities) for torsion-angle-based comparison [32], eRMSD (epsilon root-mean-square deviation) for base-pair-aware distance measurement [35], and Interaction Network Fidelity (INF) scores used in CASP RNA assessment. AlphaFold 3 [27] additionally provides per-residue predicted lDDT (pLDDT) confidence scores that function as built-in QA for predicted structures.

Benchmarking frameworks. Systematic evaluation of RNA QA methods has been facilitated by RNAdvisor [10] and community experiments including RNA-Puzzles [11] and CASP RNA assessment rounds [1, 2]. These benchmarks have highlighted that no single method dominates across all RNA sizes and structural types, suggesting that fundamentally different feature types may be needed for comprehensive QA.

Spectral methods in structural biology. Graph Laplacian spectra have a rich history in molecular analysis. Persistent homology and topological data analysis (TDA)—which char-

Table 1: Distance cutoff scales and their structural interpretation.

Cutoff (Å)	$\sigma_{\textbf{Gaussian}}$ **(Å)**	**Structural interactions captured**
8	2.67	Base stacking, hydrogen bonds
10	3.33	Near-neighbor backbone contacts
12	4.00	Helix–helix packing, junctions
15	5.00	Domain-level tertiary contacts

acterizes multi-scale shape through algebraic topology—have been applied to protein structure and function prediction [17–19]. Persistent Laplacians [20] extend classical graph Laplacians by incorporating multi-scale filtrations. Spectral clustering has been used for protein domain decomposition and community detection in contact networks. Despite these successes, no prior work has applied spectral graph features to RNA quality assessment—a gap that SpecRNA-QA addresses.

3 Methods

An overview of the SpecRNA-QA pipeline is shown in Figure 1 and Algorithm 1.

3.1 Multi-scale contact graph construction

Given an RNA 3D structure with n nucleotides, we extract a single representative coordinate per residue: the C4′ sugar carbon atom, with the phosphorus atom P as fallback for residues lacking C4′. This yields a point set $\mathbf{X} \in \mathbb{R}^{n\times 3}$.

We construct contact graphs at four distance cutoffs $d_c \in \{8, 10, 12, 15\}$ Å using two kernel types:

Binary adjacency: $A_{ij}^{\text{bin}} = \mathbf{1}[\|\mathbf{x}_i - \mathbf{x}_j\| < d_c]$

Gaussian-weighted adjacency: $A_{ij}^{\text{gau}} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2) \cdot \mathbf{1}[\|\mathbf{x}_i - \mathbf{x}_j\| < d_c]$

where $\sigma = d_c/3$, chosen so that the Gaussian weight at the cutoff boundary is $\exp(-d_c^2/2\sigma^2) = \exp(-4.5) \approx 0.011$, ensuring consistent boundary decay across all distance scales. This yields eight graphs per structure. The multi-scale design is motivated by the physical hierarchy of RNA interactions: 8 Å captures tight base stacking and hydrogen bonds, 10 Å captures near-neighbor contacts, 12 Å captures helix–helix packing, and 15 Å captures long-range domain-level topology (Table 1).

3.2 Normalized Laplacian spectrum

For each graph with adjacency matrix A and degree matrix $D = \text{diag}(\sum_j A_{ij})$, we compute the normalized Laplacian:

$$L = I - D^{-1/2} A\, D^{-1/2}$$

with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$. The spectral gap λ_2 (algebraic connectivity) quantifies how well-connected the graph is via the discrete Cheeger inequality: $\lambda_2/2 \leq h(G) \leq \sqrt{2\lambda_2}$, where the Cheeger constant $h(G)$ is the minimum edge-boundary-to-volume ratio over all vertex subsets. A higher λ_2 indicates a more robustly connected contact network with no

Table 2: Spectral feature categories. Count is per graph; ×8 graphs yields ∼264 base features.

Category	Count	Physical meaning
Low eigenvalues (λ_2–λ_6)	5	Connectivity and dominant structural modes
Quantile spectrum (5%–95%)	11	Eigenvalue distribution shape, length-invariant
Spectral entropy H	1	Spectral complexity; low H = coherent fold
Spectral moments (μ_1–μ_4)	4	Mean, variance, skewness, kurtosis
Heat-kernel trace $Z(t)$	6	Multi-scale diffusion: local (t=0.1) to global (t=10)
Inverse participation ratio (IPR)	5	Mode localization of lowest-frequency eigenvectors
Subtotal per graph	**33**	

weak cuts, while a small λ_2 signals topological bottlenecks—precisely the signature of misplaced domains that are loosely connected to the rest of the structure. The full eigenvalue distribution encodes richer information—community structure, symmetry, and diffusion dynamics—that no single scalar descriptor can capture. The normalized (rather than unnormalized) Laplacian is chosen because its eigenvalues lie in $[0, 2]$ regardless of RNA length, enabling direct comparison across structures of different sizes. For isolated nodes (degree zero), which can arise at the 8 Å cutoff for loosely packed structures, we follow the convention of setting $L_{ii} = 0$ (as implemented in SciPy's `csgraph.laplacian`), which assigns eigenvalue 0 to isolated nodes. This deviates from Chung's definition [16], which requires $D_{ii} > 0$; the consequence is that the zero-eigenvalue count conflates genuine disconnected components with isolated residues. In practice, isolated nodes are rare: fewer than 0.3% of CASP16 models have any isolated nodes at 8 Å, and none occur at ≥10 Å.

3.3 Spectral feature extraction

From each Laplacian spectrum we extract six categories of features (Table 2; complete catalog in Supplementary Table S1):

The **heat-kernel trace** deserves particular attention as it emerges as the most important feature class (Section 4.7). Defined as $Z(t) = \frac{1}{n}\sum_i e^{-t\lambda_i}$, it is the length-normalized trace of the heat kernel matrix e^{-tL}, corresponding to the average return probability of a random walk rather than the total heat content $\mathrm{Tr}(e^{-tL})$. We normalize by $1/n$ to enable direct comparison across structures of different sizes. $Z(t)$ describes heat diffusion on the contact graph. At short times ($t \ll 1$), $Z(t)$ reflects local degree distribution (analogous to contact density); at long times ($t \gg 1$), it converges toward global connectivity structure. Evaluating $Z(t)$ at six time scales $t \in \{0.1, 0.5, 1, 2, 5, 10\}$ provides a continuous multi-scale probe of structural organization.

Cross-scale stability features (mean, variance, and linear slope of each spectral metric across the four distance cutoffs) capture whether the spectral signature is robust to the choice of interaction range. Well-folded structures exhibit stable spectral properties across scales, while topologically defective structures show erratic variation.

Combined across 8 graphs with cross-scale statistics and 3 native prior features (Section 3.4), the total is 314 spectral features per structure. We also extract 18 geometric baseline features (R_g,

contact density, and largest connected component ratio at each cutoff, plus cross-cutoff stability statistics) for ablation comparison, yielding 332 features in total (Supplementary Table S1).

3.4 Native spectral prior

To enable reference-free scoring, we fit *native spectral priors* from experimentally determined RNA structures. For each RNA length bin (<50, 50–100, 100–200, >200 nt), we compute the mean quantile spectrum from native structures in the training set as a template. At inference time, the first Wasserstein distance (W_1) between a model's quantile spectrum and the corresponding length-bin template provides an unsupervised quality signal:

$$d_W = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} W_1\left(q^g_{\text{model}}, q^g_{\text{template}}\right)$$

where $\mathcal{G}$ indexes the set of graph types. Models closer to the native spectral template tend to have higher structural quality (Figure 2). We additionally compute z-scores of λ_2 and spectral entropy relative to native statistics in the same length bin.

3.5 Learning-to-rank model

We train an XGBRanker—a gradient-boosted tree model optimized for ranking tasks [21]—with LambdaMART pairwise ranking loss (240 trees, maximum depth 4, learning rate 0.05; full hyperparameters in Supplementary Table S4) to predict per-target model rankings. Ground-truth labels are CASP-official local Distance Difference Test (lDDT) scores [22]. Predicted scores are calibrated to the lDDT scale using isotonic regression.

3.6 Training-free heuristic mode

For scenarios where labeled QA data is unavailable, we provide a *heuristic mode* that ranks models using only three spectral statistics derived from the native prior:

$$\text{score}_{\text{heuristic}} = -\left(|z_{\lambda_2}| + |z_H| + d_W\right) \tag{1}$$

where z_{λ_2} and z_H are z-scores of algebraic connectivity and spectral entropy relative to native statistics, and d_W is the Wasserstein distance to the native template. This mode is *training-free*—it requires no labeled quality data—but does require a small database of experimentally determined native RNA structures (16 PDB entries in our experiments) to construct spectral templates. We distinguish this from *reference-free* inference: neither the trained nor heuristic mode requires the native structure of the query target at inference time, but the heuristic additionally eliminates the need for CASP-style labeled benchmarks.

3.7 Evaluation protocol

We use two evaluation protocols: (i) *fixed-split ablation*, where CASP16 targets are divided into 18 training and 24 test targets (CASP15: 8 train / 4 test); and (ii) *leave-one-out cross-validation* (LOOCV), where each target is held out in turn and all remaining targets are used for

Algorithm 1 SpecRNA-QA scoring pipeline

Require: RNA 3D model M with n nucleotides; native prior $\mathcal{P}$
Ensure: Quality score $s \in [0, 1]$

```
Extract C4′/P coordinates X ∈ ℝ^{n×3}
Compute pairwise distance matrix Δ ∈ ℝ^{n×n}
for each cutoff d_c ∈ {8, 10, 12, 15} Å do
    for each kernel k ∈ {binary, Gaussian} do
        Construct adjacency matrix A^{d_c,k} from Δ
        Compute degree matrix D = diag(Σ_j A_ij)
        Compute normalized Laplacian L = I − D^{−1/2} A D^{−1/2}
        Eigendecompose: {λ_i, v_i}_{i=1}^n ← eigh(L)
        Extract spectral features: λ_2–λ_6, quantiles, H, μ_1–μ_4, Z(t), IPR
    end for
end for
Compute cross-scale stability features (mean, variance, slope)
Enrich with native prior: d_W, z_{λ_2}, z_H from P
Concatenate all features → f ∈ ℝ^{314}
Predict: s ← calibrate(XGBRanker(f))
return s
```

training. CASP15 and CASP16 LOOCV are conducted separately (CASP16 LOOCV uses only CASP16 targets; CASP15 LOOCV uses only CASP15 targets) to avoid cross-round distributional confounds. LOOCV provides more robust estimates by using all available data for both training and evaluation.

The primary metric is *median per-target Spearman* ρ between predicted scores and CASP-official lDDT. We report median rather than pooled ρ because RNA targets vary enormously in length (29–833 nt) and model quality distributions, making pooled correlation unreliable (Supplementary Section S6). Following Kretsch et al. [1], we adopt median per-target ρ as the primary metric.

To assess statistical significance, we apply a one-sided Wilcoxon signed-rank test to the per-target $\Delta\rho$ values (spectral minus geometry) under the null hypothesis that the paired differences are symmetrically distributed around zero. Per-target win rates are assessed against chance (50%) using an exact binomial test.

4 Results

4.1 Data

We evaluated SpecRNA-QA on two CASP RNA benchmarks. CASP15 comprises 12 targets with 1,560 models (RNA length 29–136 nt). CASP16 comprises 42 targets with 7,368 models (length 58–833 nt), representing a substantially larger and more diverse benchmark that includes long multi-domain RNAs absent from CASP15. Together, these benchmarks cover 54 targets and 8,928 models.

Table 3: Benchmark results across evaluation protocols. Median per-target Spearman ρ with CASP-official lDDT. Win rate: fraction of targets where spectral > geometry (CASP16 / CASP15).

Evaluation	Features	CASP16 ρ	CASP15 ρ	Win rate
LOOCV	Spectral	**0.689**	**0.629**	93% / 100%
	Geometry	0.465	0.396	—
	All	0.691	0.630	—
Ablation	Spectral	**0.510**	**0.415**	92% / —
	Geometry	0.265	0.101	—
	All	0.497	0.376	—
Heuristic	3 statistics	0.310	—	—

4.2 Spectral features systematically outperform geometry

Table 3 summarizes performance across both evaluation protocols and datasets. In LOOCV on CASP16, spectral-only features achieve median $\rho = 0.689$, outperforming geometry-only features ($\rho = 0.465$) by $\Delta = +0.224$. On CASP15, the advantage is $\Delta = +0.233$ ($\rho = 0.629$ vs. 0.396). Fixed-split ablation shows even larger gaps: CASP16 $\Delta = +0.245$ and CASP15 $\Delta = +0.314$.

The advantage is remarkably consistent across individual targets (Figure 3a). In LOOCV, spectral features win on 39/42 CASP16 targets (92.9%) and 12/12 CASP15 targets (100%). The per-target $\Delta\rho$ distribution (Figure 3b) shows that 51 of 54 targets (94.4%) favor spectral features, with only 3 targets showing a geometry advantage—all involving RNAs <100 nt where the difference is small.

The spectral advantage is statistically significant: a one-sided Wilcoxon signed-rank test on the 42 per-target $\Delta\rho$ values (CASP16 LOOCV) yields $W = 881$, $p = 1.2 \times 10^{-10}$. Bootstrap 95% confidence intervals for median ρ do not overlap: spectral $[0.644, 0.731]$ vs. geometry $[0.394, 0.528]$ (10,000 resamples). The 93% win rate (39/42) is significant by exact binomial test ($p = 2.8 \times 10^{-9}$).

4.3 Performance across RNA length categories

We stratified LOOCV results by RNA length (Table 4). Spectral features outperform geometry in all three CASP16 length categories. The advantage is largest for long RNAs (>200 nt: $\Delta = +0.233$, 26 targets), consistent with the expectation that larger contact networks provide richer spectral information and exhibit more frequent "locally correct but globally wrong" failures. Importantly, the advantage persists for short RNAs (50–100 nt: $\Delta = +0.123$, 13 targets), demonstrating that spectral QA is not restricted to large structures. We note that the 100–200 nt bin contains only 3 targets; results for this bin should be interpreted as illustrative rather than statistically robust.

4.4 Comparison with established statistical potentials

To assess whether the spectral advantage extends beyond the internal geometry baseline, we scored all 7,368 CASP16 models with two established statistical potentials: DFIRE-RNA [5] and rsRNASP [4]. Table 5 compares all methods across RNA length categories. We note upfront

Table 4: LOOCV performance on CASP16 stratified by RNA length.

Size bin	**Targets**	ρ **spectral**	ρ **geometry**	Δ
50–100 nt	13	0.591	0.468	+0.123
100–200 nt	3	0.486	0.282	+0.204
>200 nt	26	**0.718**	0.485	**+0.233**

Table 5: CASP16 comparison: SpecRNA-QA (supervised LOOCV) vs. statistical potentials (unsupervised). Note: this comparison is asymmetric—see text for discussion. rsRNASP times out (>30 s per model) on most >200 nt targets. Win rate = fraction of targets where SpecRNA-QA > rsRNASP.

Size bin	n	**Spec.**	**rsRNASP**	**DFIRE**	**Geom.**	**Win%**
50–100 nt	13	.591	**.661**	.609	.468	46
100–200 nt†	3	.486	**.662**	.646	.282	0
>200 nt	26‡	**.718**	.750‡	.518	.485	—
Scored	**20**	.570	**.667**	.609	.468	30
All	**42**	**.689**	—	.558	.465	—

†3 targets only. ‡rsRNASP scored 4/26 >200 nt targets; 22 timed out.

that this comparison is methodologically asymmetric: SpecRNA-QA uses supervised LOOCV with lDDT labels, while DFIRE and rsRNASP are unsupervised energy functions requiring no labeled data. In the unsupervised setting, the SpecRNA-QA heuristic ($\rho = 0.31$) falls below DFIRE ($\rho = 0.56$), underscoring that the spectral advantage requires a supervised learner.

The most striking finding concerns large RNAs (>200 nt, 26 targets). rsRNASP—the strongest established method—times out (>30 s per model on Apple M-series silicon, single-threaded) on 22 of these 26 targets, producing no scores for the majority of CASP16's large RNA structures. On this critical size class, SpecRNA-QA ($\rho = 0.718$) significantly outperforms the only remaining baseline, DFIRE ($\rho = 0.518$, $\Delta = +0.200$, 73% win rate). This is precisely where the "locally correct but globally wrong" failure mode is most prevalent and where global topological assessment is most needed.

On smaller RNAs where rsRNASP can be evaluated (≤200 nt, 20 targets with ≥80% coverage), rsRNASP outperforms SpecRNA-QA ($\rho = 0.667$ vs. 0.570, Wilcoxon $p = 0.017$), confirming that atom-level statistical potentials remain superior for compact structures where local stereochemistry dominates. Whether spectral features would also outperform rsRNASP on large RNAs if computational barriers were removed—e.g., through longer timeouts, parallelized implementations, or the coarse-grained variant cgRNASP [7]—remains an open question.

Together, these results reveal a clear size-dependent complementarity: statistical potentials excel on small RNAs, spectral features excel on large RNAs, and the two approaches capture largely orthogonal quality signals.

4.5 Spectral rescue of geometry failures

More broadly, the spectral advantage is pervasive: in LOOCV, 71% of targets (30/42) show $\Delta\rho > 0.1$ and 93% show $\Delta\rho > 0$. Perhaps the most compelling evidence comes from targets where geometric descriptors actively mislead—producing negative rank correlations that select

Table 6: Targets where geometry gives negative ρ (selects worse models) but spectral features rescue the ranking. CASP16 ablation and LOOCV results.

Target	Length	ρ geometry	ρ spectral	$\Delta\rho$	Diagnosis
R1289	343	−0.322	0.642	+0.964	Large, complex fold
R1242	205	−0.528	0.300	+0.828	Domain misplacement
R1248	407	−0.012	0.723	+0.735	Locally correct, globally wrong
R1224s2	395	−0.219	0.486	+0.705	Multi-domain RNA
R1221s2	398	−0.113	0.579	+0.693	Multi-domain RNA
R1288	58	−0.179	0.156	+0.335	Short RNA, geometry misleading

worse models. Table 6 identifies six such targets from the CASP16 ablation experiment. On these targets, geometry-based QA is worse than random (selecting models inversely related to true quality), while spectral features correct the ranking with improvements up to $\Delta\rho = +0.964$. These "rescue" cases span a range of RNA lengths (58–480 nt), demonstrating that the spectral correction is not limited to a specific structural class. Section 4.10 examines the most dramatic case (R1248) in detail.

4.6 Spectral prior discriminates model quality

The native spectral prior provides an unsupervised quality signal. Figure 2 illustrates this for CASP16 target R1248 (407 nt, 171 models). When models are grouped by lDDT quartile, the highest-quality quartile (Q4) has approximately half the Wasserstein distance to the native template compared to the lowest-quality quartile (Q1): median $d_W = 0.015$ vs. 0.032 (Figure 2b). The continuous correlation between d_W and lDDT is $\rho = -0.53$ (Figure 2c), confirming that spectral distance to the native template is a meaningful—though imperfect—proxy for structural quality.

4.7 Heat-kernel trace dominates feature importance

Feature importance analysis from the XGBRanker model reveals that heat-kernel trace values account for 7 of the top 20 features (Figure 4a). The 12 Å binary graph contributes most strongly, with three heat-trace features in the top five ($t = 1.0$, $t = 5.0$, $t = 2.0$). The heatmap of aggregated importance by distance scale and feature category (Figure 4b) confirms that the 12 Å heat trace dominates, followed by quantile spectrum features.

This pattern is physically interpretable. The heat kernel e^{-tL} describes diffusion on the contact graph. At $t = 0.1$, $Z(t)$ reflects local degree distribution; at $t \geq 5$, it probes global connectivity. The 12 Å cutoff corresponds to helix–helix packing distances—precisely the scale at which correctly folded vs. misfolded domain arrangements diverge. The dominance of intermediate times ($t = 1$–5) at this distance scale indicates that quality-discriminating information resides at the local-to-global transition.

4.8 Feature combination strategies do not improve over spectral-only

Combining spectral and geometric features does not improve over spectral features alone (Table 3: "All" $\rho = 0.691$ vs. Spectral $\rho = 0.689$ in LOOCV). We further tested two late-fusion strategies—

z-score fusion and rank-average fusion of independently trained spectral and geometry models—finding that both degrade performance (Supplementary Table S3: late z-score $\rho = 0.478$, late rank-average $\rho = 0.454$, vs. spectral-only $\rho = 0.510$). This indicates that the current geometric features add noise rather than complementary signal, and that the spectral representation already subsumes the information content of simple geometric descriptors.

One might ask whether the spectral advantage simply reflects the larger feature count (314 spectral vs. 18 geometric). A top-k feature ablation study directly addresses this (Supplementary Table S8). At matched dimensionality (18 features each), randomly selected spectral features achieve $\rho = 0.585$ vs. geometry $\rho = 0.454$—a $\Delta = +0.131$ advantage that cannot be explained by feature count alone. Even the top-10 spectral features ($\rho = 0.521$) outperform the full 18-feature geometry set with fewer dimensions. Performance plateaus around $k = 30$–50 ($\rho \approx 0.647$), indicating that a compact set of spectral descriptors captures most quality-discriminating information.

Two additional observations reinforce this conclusion. First, adding 18 geometry features to the 314 spectral features yields no improvement ($\rho = 0.691$ vs. 0.689), confirming that XGBRanker does not blindly benefit from more features. Second, late fusion of independently trained spectral and geometry models *degrades* performance ($\rho = 0.478$ and 0.454 for z-score and rank-average fusion, respectively), showing that the geometry signal actively contradicts spectral rankings rather than complementing them.

4.9 Training-free heuristic validates the spectral quality signal

The heuristic mode (Equation 1) serves as a conceptual validation: it tests whether the spectral prior contains a quality signal *even without any labeled training data.* Using only three statistics ($|z_{\lambda_2}|$, $|z_H|$, d_W), the heuristic achieves $\rho = 0.31$ on CASP16 (24 test targets), comparable to the *trained* geometry-only model ($\rho = 0.27$; Table 3). This confirms that native RNA structures occupy a characteristic spectral signature that degrades predictably with structural quality—the global topology signal exists independently of the supervised learner.

The heuristic is not intended as a competitive standalone method. On the RNAdvisor TestSetII benchmark (20 targets, 38–148 nt; Supplementary Table S5), it achieves median $\rho = 0.26$, below rsRNASP ($\rho = 0.65$) and DFIRE ($\rho = 0.59$). Rather, it demonstrates that the spectral prior provides an orthogonal quality axis that could strengthen existing methods when integrated into multi-score ensembles.

4.10 Case study: R1248—locally correct, globally wrong

Target R1248 (CASP16, GOLLD 3′ domain, 407 nt, 171 models) provides a compelling illustration of the failure mode motivating spectral QA (Figure 5). The best model (lDDT = 0.54) and worst model (lDDT = 0.005) have nearly identical R_g (50.0 vs. 49.9 Å) and similar contact densities. Geometry-based ranking is indistinguishable from random ($\rho = -0.01$; Figure 5d), with the best model ranked only #60 of 171.

Spectral features detect the quality difference through algebraic connectivity λ_2, which differs by 44% at the 8 Å cutoff between the best and worst models (8.8×10^{-4} vs. 4.9×10^{-4}; Figure 5c). This indicates a less robustly connected contact network despite similar compactness—consistent

with correct local helices but incorrect global domain arrangement. SpecRNA-QA achieves $\rho = 0.72$ and ranks the true best model first (Figure 5b).

Structural superposition confirms the diagnosis (Figure 5a): the worst model shows a median per-nucleotide C4′ deviation of 25 Å from the best model, with 83% of residues deviating by >10 Å. The blue regions (low deviation) correspond to correctly formed helical arms; the red regions (high deviation) reveal completely misplaced core domains—a textbook "locally correct but globally wrong" failure.

4.11 Temporal generalization: CASP15 to CASP16

To assess cross-round generalization, we trained on all 12 CASP15 targets (1,560 models) and tested on the 42 CASP16 targets. This temporal split yields median $\rho = 0.40$—substantially below the within-round LOOCV ($\rho = 0.69$), but still above the fixed-split geometry baseline ($\rho = 0.27$). The performance drop is expected given that CASP15 provides only 12 training targets (vs. 41 in LOOCV) and CASP16 introduced considerably larger and more diverse RNAs. This result confirms that the spectral signal generalizes across CASP rounds, though with reduced magnitude when training data is limited.

4.12 Failure cases

Spectral features underperform on three CASP16 LOOCV targets. R1271 (77 nt) has $\rho_{\text{geo}} = 0.32$ vs. $\rho_{\text{spec}} = 0.18$: this short RNA has a simple topology where geometric descriptors suffice and the contact graph is too compact for meaningful spectral differentiation. R1254 (large >200 nt) and R1291 show marginal geometry advantages ($\Delta < 0.04$), effectively ties.

Three additional targets (R1291, R1241, R1256 in fixed-split ablation) show low ρ for all feature types ($\rho < 0.10$), suggesting these targets are inherently difficult for any single-model QA approach—possibly due to narrow quality distributions among decoys or unusual structural characteristics that confound both spectral and geometric descriptors.

5 Discussion

5.1 Physical basis of spectral superiority

The consistent spectral advantage ($\Delta\rho \geq +0.22$) across two independent CASP benchmarks, multiple evaluation protocols, and diverse RNA lengths supports a clear physical interpretation. The dominant heat-kernel trace features (Figure 4) measure diffusion dynamics on the contact graph at the local-to-global transition scale. A structure with correct local contacts but misplaced domains will show normal short-time diffusion (intact local neighborhoods) but anomalous long-time diffusion (broken global connectivity). This is precisely the signature captured by the heat kernel at intermediate times ($t = 1$–5) and the 12 Å distance scale.

The algebraic connectivity λ_2 provides a complementary global diagnostic. In the R1248 case study, the 44% difference in λ_2 between the best and worst models—despite identical R_g—directly reflects the topological distinction between a well-connected domain architecture

and a fragmented one with misplaced substructures. This information is categorically absent from geometric descriptors that collapse the contact network to scalar summaries.

5.2 Complementarity with statistical potentials

The comparison with DFIRE-RNA and rsRNASP on the full CASP16 benchmark (Table 5) provides direct evidence for complementarity, though we emphasize that the comparison is methodologically asymmetric (supervised SpecRNA-QA vs. unsupervised potentials). SpecRNA-QA and statistical potentials measure fundamentally different properties: spectral features assess *global topological coherence*—domain connectivity, symmetry, diffusion pathways—while DFIRE and rsRNASP evaluate *local atomic interactions*—stacking energies, hydrogen bonding geometries, distance-dependent contact potentials. The size-dependent performance crossover (spectral dominates at >200 nt, statistical potentials dominate at <200 nt) confirms that these signals are largely orthogonal.

On the RNAdvisor TestSetII (small RNAs, 38–148 nt; Supplementary Table S5), rsRNASP substantially outperforms the SpecRNA-QA heuristic ($\rho = 0.65$ vs. 0.26), further confirming that local scoring is essential for compact structures. Score-level fusion of spectral and statistical potential methods—combining global topology with local energy in a single ranker—is a natural next step expected to yield improvements across the full RNA size range. Both modes of SpecRNA-QA are reference-free (no native structure required at inference) and lightweight ($\sim$2 seconds per 400-nt model on a single CPU; Supplementary Section S7).

5.3 Why global features correlate with a local metric

An apparent paradox arises: if spectral features capture "global" quality missed by local descriptors, why do they correlate with lDDT—a metric that evaluates distance differences within a 15 Å inclusion radius? The resolution is that misplaced domains also disrupt *inter-domain* distance patterns within the lDDT radius. When a domain is shifted by 25 Å (as in the R1248 case study), the distance differences between residues in the misplaced domain and residues in correctly placed domains exceed the lDDT inclusion threshold, reducing the lDDT score. Thus lDDT does capture some global information indirectly, and the spectral features provide a more direct route to the same quality signal—one that is available even when scalar geometric summaries like R_g are uninformative. Our heat-kernel trace $Z(t)$ is the global (trace-normalized) version of the Heat Kernel Signature (HKS) introduced by Sun et al. [33] for shape analysis; the per-vertex HKS could naturally enable per-nucleotide local QA in future work.

5.4 Limitations and future directions

Structural representation. The current method uses only backbone (C4′/P) coordinates, discarding base-pair geometry, stacking interactions, and nucleotide identity. This limits biological interpretability and likely contributes to the weaker performance on small RNAs where base-level chemistry dominates. Incorporating richer node and edge features is a natural extension.

Evaluation scope. A temporal generalization test (training on CASP15, testing on CASP16) yields $\rho = 0.40$, confirming cross-round generalization but at reduced magnitude due to CASP15's

limited 12 training targets. Performance on RNAs <50 nt has not been assessed due to insufficient data. The training-free heuristic remains less competitive than established statistical potentials in the unsupervised setting. Comparison with ARES [9] was not feasible: the ARES framework expects RNA-Puzzles-format structures with specific atom naming conventions and chain labeling that differ from CASP16 TS-format submissions; converting 7,368 models while preserving the atom-type annotations required by ARES's equivariant neural network was beyond the scope of this work.

Scalability and extensions. The eigendecomposition cost is $O(n^3)$ per graph; for RNAs >3,000 nt, sparse eigensolvers or stochastic trace estimators would be needed. Our cross-scale stability features capture only linear spectral trends; persistent Laplacian features [20, 34] would capture nonlinear spectral evolution and represent a high-priority extension. Per-nucleotide local QA via eigenvector components, and generalization to protein QA via Cα contact networks, are additional promising directions.

6 Conclusion

We have presented SpecRNA-QA, a spectral graph approach to RNA 3D model quality assessment that introduces a fundamentally new class of structural features to the RNA QA toolbox. Multi-scale Laplacian spectral features—particularly heat-kernel traces at the local-to-global transition scale—significantly outperform an internal geometry baseline ($\Delta\rho = +0.22$, $p < 10^{-10}$) on the CASP16 benchmark. Head-to-head comparison with established statistical potentials (DFIRE-RNA, rsRNASP) reveals a clear size-dependent pattern: spectral features dominate on large RNAs (>200 nt) where global topological coherence is most critical, while statistical potentials excel on smaller structures—confirming that the two approaches capture complementary quality signals.

The method is lightweight (CPU-only, seconds per model), interpretable (features map directly to physical properties of the contact network), and effective in both supervised and unsupervised modes. The training-free heuristic—requiring only three spectral statistics and a small set of native templates—already matches a trained geometry baseline, making SpecRNA-QA immediately applicable to scenarios where labeled QA data is unavailable.

Our results demonstrate that global topological analysis of RNA contact networks provides quality signals that are complementary to and largely orthogonal to existing local scoring functions. We anticipate that fusion of spectral features with statistical potentials will yield further improvements, and that extensions to persistent Laplacian features and per-nucleotide local QA will broaden the approach's applicability. SpecRNA-QA opens a new direction for reference-free RNA structure quality assessment grounded in the mathematical framework of spectral graph theory. The spectral framework is structure-agnostic and could naturally extend to protein model QA via Cα contact networks, protein–RNA complex assessment, and other molecular quality estimation tasks.

Data Availability

The CASP15 and CASP16 RNA model coordinates and official lDDT scores are publicly available from the CASP website (https://predictioncenter.org). The RNAdvisor benchmark data is available at https://github.com/clementbernardd/rnadvisor. SpecRNA-QA source code, trained models, and configuration files for reproducing all experiments are available at https://github.com/yudabitrends/specrnaq.

Author Contributions

Ying Zhu: Methodology, Software, Validation, Formal Analysis, Writing – Original Draft. **Huaiwen Zhang**: Data Curation, Software, Visualization. **Vince D. Calhoun**: Conceptualization, Resources, Supervision, Funding Acquisition, Writing – Review & Editing. **Yuda Bi**: Conceptualization, Methodology, Supervision, Project Administration, Writing – Review & Editing.

Conflict of Interest

None declared.

Funding

This work was supported in part by the National Institutes of Health (R01MH123610 and R01MH118695).

Acknowledgements

We thank the CASP organizers and RNA prediction community for making benchmark data publicly available. We thank the developers of rsRNASP, DFIRE-RNA, and RNAdvisor for providing open-source implementations of their methods.

References

[1] Kretsch, R.C. et al. Assessment of nucleic acid structure prediction in CASP16. *Proteins*, 2025.

[2] Das, R. et al. Assessment of RNA structure prediction in CASP15. *Proteins*, 91(12):1689–1699, 2023.

[3] Shen, T. et al. RhoFold+: a deep learning framework for RNA 3D structure prediction. *Nature Methods*, 21:1514–1524, 2024.

[4] Zhang, T. et al. rsRNASP: a reference-state-based RNA statistical potential. *Briefings in Bioinformatics*, 24(1):bbac531, 2023.

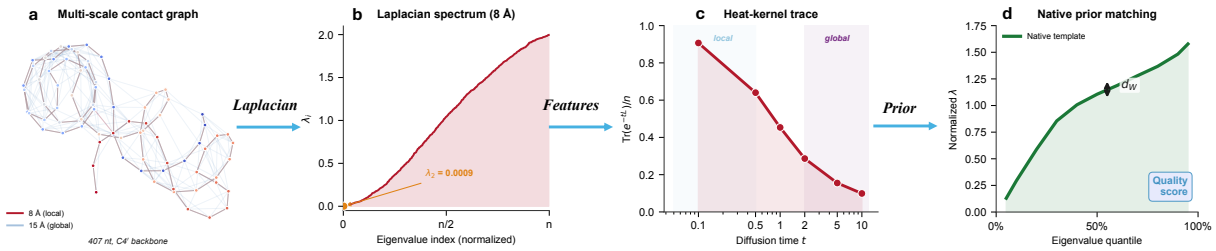

Figure 1: **SpecRNA-QA method overview.** (**a**) Multi-scale contact graphs from C4′ coordinates at 8 Å (local) and 15 Å (tertiary). (**b**) Normalized Laplacian spectrum at the 8 Å cutoff; spectral gap $\lambda_2 = 0.0009$ quantifies global connectivity. (**c**) Heat-kernel trace $Z(t)$ at six time scales, probing local (t=0.1) to global (t=10) structure. (**d**) Model spectra compared against native templates via Wasserstein distance. Data from CASP16 target R1248 (407 nt).

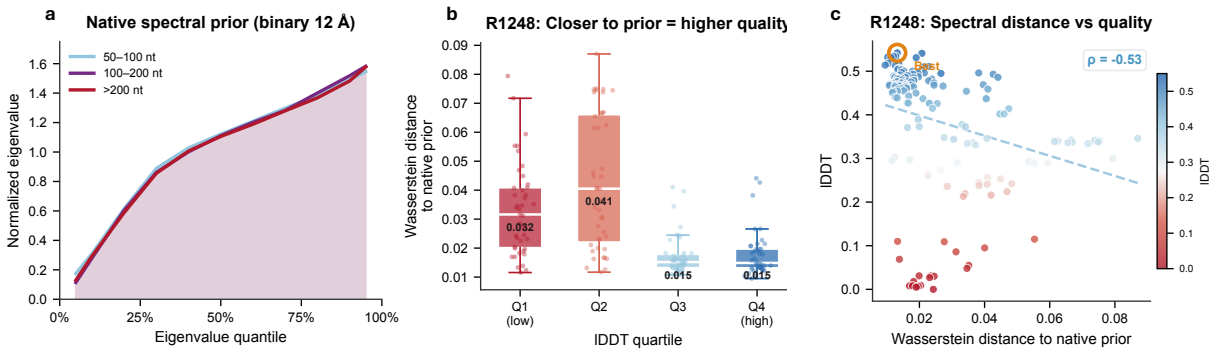

Figure 2: **Spectral fingerprinting discriminates model quality.** (**a**) Native spectral prior: quantile eigenvalue spectra of the 12 Å binary Laplacian for native structures, stratified by length. (**b**) Wasserstein distance to native prior by lDDT quartile (171 R1248 models). Q4 models have $\sim$2$\times$ smaller spectral distance than Q1. (**c**) d_W vs. lDDT scatter ($\rho = -0.53$).

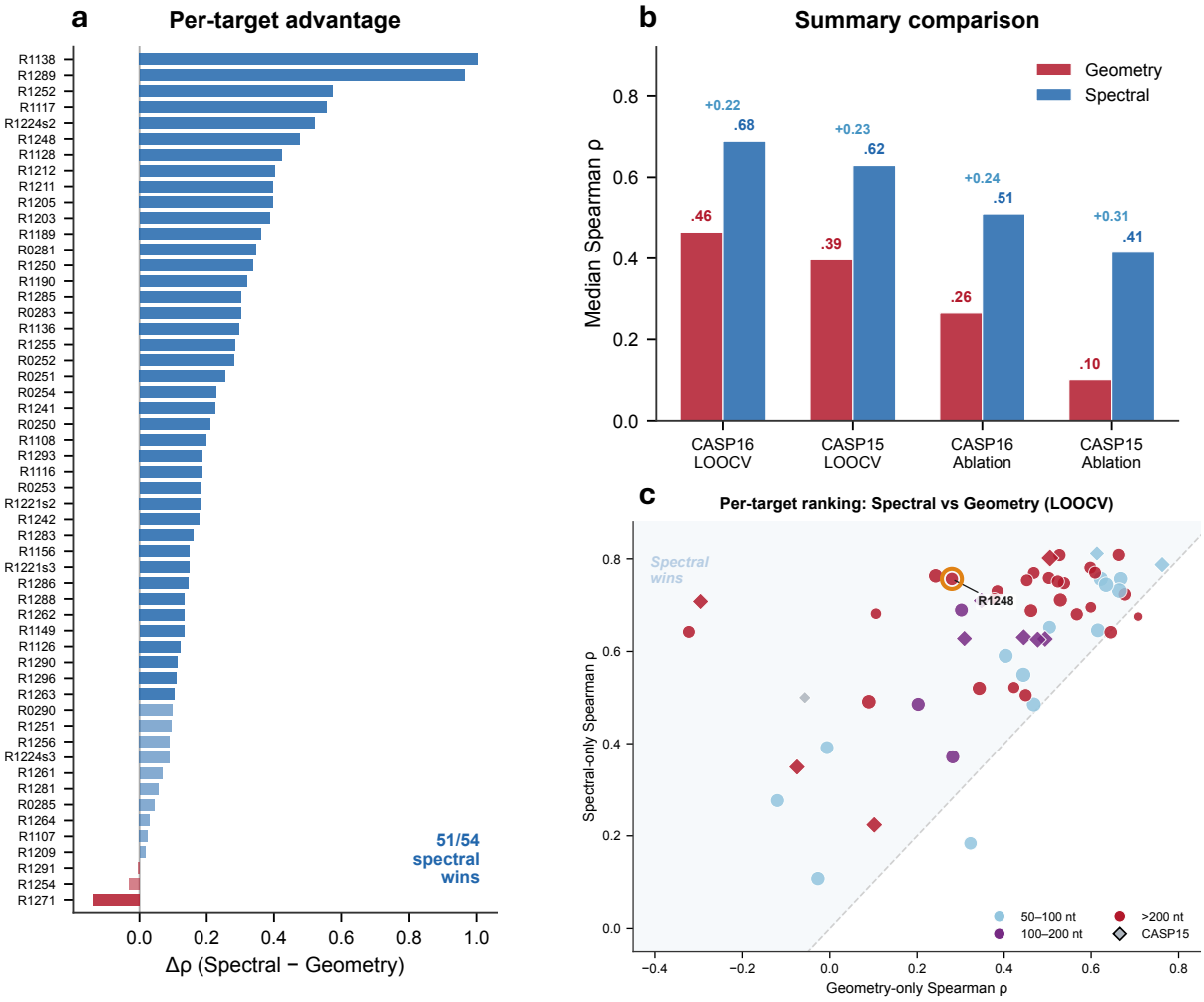

Figure 3: **Spectral features systematically outperform geometry (LOOCV).** (**a**) Per-target $\Delta\rho$ sorted by magnitude; 51/54 targets favor spectral. (**b**) Summary across protocols and datasets. (**c**) Per-target Spearman ρ: spectral (y-axis) vs. geometry (x-axis). Points above diagonal favor spectral; color encodes RNA length; diamonds = CASP15; R1248 highlighted.

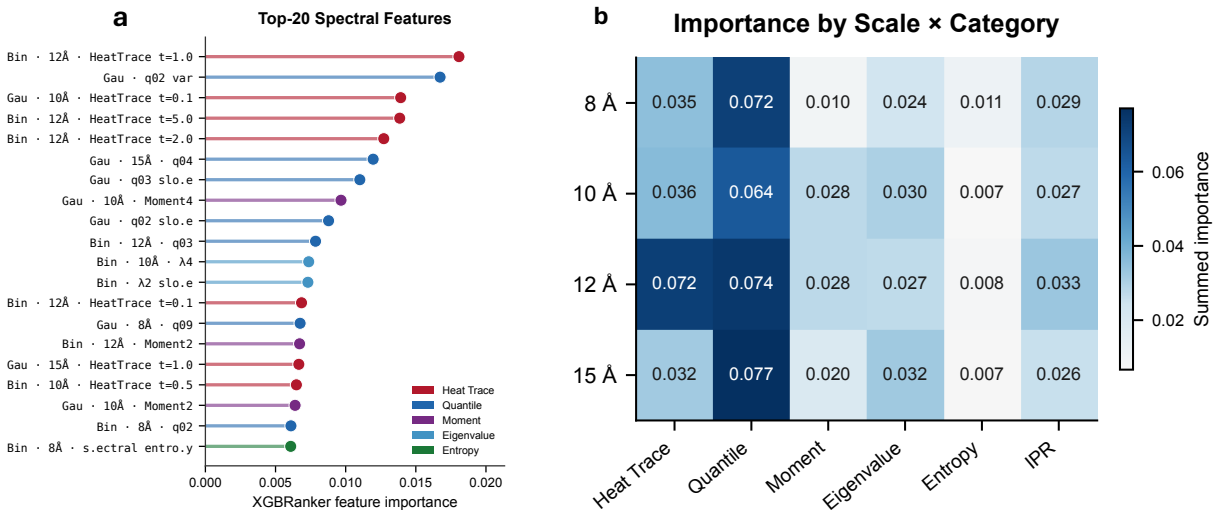

Figure 4: **Feature importance landscape.** (**a**) Top-20 features by XGBRanker importance, colored by category. Heat-kernel trace features dominate (7/20), led by `binary_12a_heattrace_t1.0`. (**b**) Summed importance by distance scale and feature category. Heat trace and quantile features contribute most, with the 12 Å scale showing the highest combined importance.

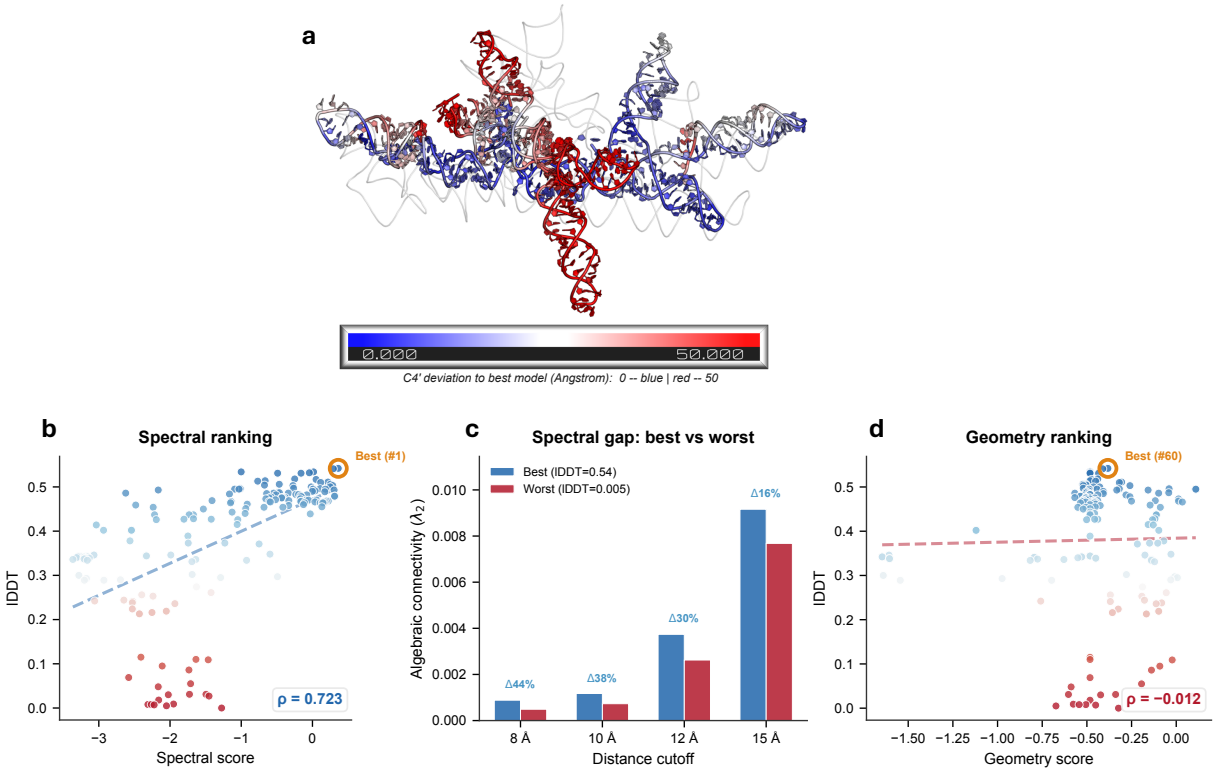

Figure 5: **Case study: R1248 (407 nt, GOLLD 3′ domain).** (**a**) Worst model colored by per-nucleotide C4′ deviation to the best model (blue = 0 Å, red = 50 Å) overlaid on best model (gray). Median deviation 25 Å despite identical R_g. (**b**) Spectral score vs. lDDT ($\rho = 0.72$); best model ranked #1. (**c**) λ_2 across cutoffs; best model shows 16–44% higher λ_2. (**d**) Geometry score vs. lDDT ($\rho = -0.01$); best model ranked #60.

[5] Capriotti, E., Norambuena, T., Marti-Renom, M.A., and Melo, F. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics*, 27(8):1086–1093, 2011.

[6] Capriotti, E. and Marti-Renom, M.A. RASP: rapid assessment of RNA structure prediction quality. *Bioinformatics*, 27(15):2133–2134, 2011.

[7] Tan, Y.L. et al. cgRNASP: coarse-grained statistical potentials with residue separation for RNA structure evaluation. *NAR Genomics and Bioinformatics*, 5(1):lqad016, 2023.

[8] Bernauer, J., Huang, X., Sim, A.Y.L., and Levitt, M. Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA*, 17(6):1066–1075, 2011.

[9] Townshend, R.J.L. et al. Geometric deep learning of RNA structure. *Science*, 373(6558):1047–1051, 2021.

[10] Bernard, C. et al. RNAdvisor: a comprehensive benchmarking tool for the measure of RNA 3D structure similarity and scoring. *Bioinformatics*, 40(2):btae057, 2024.

[11] Miao, Z. et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA*, 26(8):982–995, 2020.

[12] Benkert, P., Biasini, M., and Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3):343–350, 2011.

[13] Renaud, N. et al. DeepRank: a deep learning framework for data mining 3D protein–protein interfaces. *Nature Communications*, 12:7068, 2021.

[14] Baldassarre, F. et al. GraphQA: protein model quality assessment using graph convolutional networks. *Bioinformatics*, 37(3):360–366, 2021.

[15] Igashov, I. et al. VoroCNN: deep convolutional neural network built on 3D Voronoi tessellation of protein structures. *Bioinformatics*, 37(16):2332–2339, 2021.

[16] Chung, F.R.K. *Spectral Graph Theory.* American Mathematical Society, 1997.

[17] Xia, K. and Wei, G.-W. Persistent homology analysis of protein structure, flexibility, and folding. *International Journal for Numerical Methods in Biomedical Engineering*, 30(8):814–844, 2014.

[18] Xia, K., Opron, K., and Wei, G.-W. Multidimensional persistence in biomolecular data. *Journal of Computational Chemistry*, 36(20):1502–1520, 2015.

[19] Cang, Z. and Wei, G.-W. TopologyNet: topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS Computational Biology*, 13(7):e1005690, 2017.

[20] Wang, R., Nguyen, D.D., and Wei, G.-W. Persistent spectral graph. *International Journal for Numerical Methods in Biomedical Engineering*, 36(9):e3376, 2020.

[21] Chen, T. and Guestrin, C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD*, pages 785–794, 2016.

[22] Mariani, V., Biasini, M., Barbato, A., and Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.

[23] Baek, M. et al. Accurate prediction of nucleic acid and protein-nucleic acid complexes using RoseTTAFoldNA. *Nature Methods*, 21:117–121, 2024.

[24] Li, J. et al. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nature Communications*, 14:7266, 2023.

[25] Adamczyk, B. et al. RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. *Nucleic Acids Research*, 50(22):e130, 2022.

[26] Merdas, T.P. et al. Structure quality assessment for nucleic acids: barnaba-inspired metrics for CASP16. *Proteins*, 2024.

[27] Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630:493–500, 2024.

[28] Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[29] von Luxburg, U. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[30] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021.

[31] Williams, C.J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Science*, 27(1):293–315, 2018.

[32] Zok, T. et al. MCQ4Structures: a web server and stand-alone tool for computation of the mean of circular quantities of structural data. *Nucleic Acids Research*, 42(W1):W56–W62, 2014.

[33] Sun, J., Ovsjanikov, M., and Guibas, L. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2009.

[34] Wee, J. and Xia, K. Persistent spectral based ensemble learning (PerSpect-EL) for protein-ligand binding affinity prediction. *Briefings in Bioinformatics*, 23(2):bbab558, 2022.

[35] Bottaro, S., Di Palma, F., and Bussi, G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Research*, 42(21):13306–13314, 2014.