Supplementary Material for "Frequency-Domain Latent Attention Gating for Cross-Domain Token Aggregation"

Kewei Li[1,2], Rongying Zhang[3], Xueli Wang[1,2], Xiwen Gong[4], Zhongjian Wang[5], Lan Huang[1,2], Ruochi Zhang[1,2,#], Fengfeng Zhou[1,2,#]

[1]College of Computer Science and Technology, Jilin University, Changchun 130012, China
[2]Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China
[3]Institute for Quantitative and Computational Biology, University of California, Los Angeles, CA 90024, United States of America
[4]Greenwich High School, Greenwich, CT 06830, United States of America
[5]BCPM Data Limited, Chengdu 610041, China

[#]Correspondence may be addressed to Drs. Ruochi Zhang (zrc720@gmail.com), and Fengfeng Zhou (FengfengZhou@gmail.com or ffzhou@jlu.edu.cn).

June 6, 2026

This supplementary document collects the analyses referenced but not expanded in the main paper. These materials are intended as supporting evidence for the clean-data results and ablation-based interpretation in the main paper, rather than as additional primary benchmarks. Section S1 lists the extended ablation tables, Section S2 reports the protein-side AMP token-knockout diagnostic that motivated the method, and the final Section S3 summarizes the protein-side FLaG mechanism probes.

S1 Extended Ablation Tables

The remaining module-level ablations were run under a lighter 5-seed protocol and are reported here as supporting evidence. The fully aligned 10-seed comparison in the main paper is the basis for model selection; the tables below are intended to clarify component roles rather than reopen that decision.

S2 Preliminary AMP Single-Residue Knockout Diagnostic

This section reports the residue-level hidden-state knockout diagnostic that originally motivated FLaG. We use the ESM2-8M setting only, because the remaining protein-side probes are all defined on that setting as well. For a peptide with final-layer token representations $\mathbf{H}^{(L)} = [\mathbf{h}_1, \ldots, \mathbf{h}_T]$, we mask one residue position p at a time in the final hidden layer, rerun the readout, and record:

$$\Delta_p = \left| \hat{y}^{(-p)} - \hat{y} \right|$$

Here, $\hat{y}$ is the unmodified prediction and $\hat{y}^{(-p)}$ is the prediction after masking position p. The displayed curves aggregate Δ_p over 10 aligned seeds and compare the resulting response profile across mean pooling, max pooling, attention pooling, and FLaG. This is a diagnostic rather than a mechanistic interpretability claim: its purpose is to show whether the final aggregation rule preserves or flattens residue-level perturbation structure after the encoder has already produced late-layer token representations.

Table S1: Protein-side module ablations with ESM2-8M over 5 seeds. Values are test RMSE; lower is better. Bold indicates the best value in each dataset column; underline indicates the second best.

Variant	*E. coli*	*S. aureus*
mean	0.576	0.569
latent attn.	0.572	0.567
FFT+latent	<u>0.557</u>	<u>0.549</u>
FFT+gate	**0.552**	0.555
FLaG w/o residual	0.559	0.557
FLaG (tp_mean)	0.570	0.560
FLaG (tp_attn)	0.563	0.564
FLaG (tp_max)	0.564	**0.545**

Table S2: Image/text ablations over 5 seeds. Values are accuracy (%). Bold indicates the best value in each column; underline indicates the second best.

Variant	CIFAR-100	IMDB
mean	76.93	94.02
attn	76.60	93.72
FFT+latent	<u>77.05</u>	94.00
FFT+gate	**77.15**	94.03
FLaG w/o residual	<u>77.05</u>	93.89
FLaG (tp_mean)	76.97	93.87
FLaG (tp_attn)	76.68	<u>94.04</u>
FLaG (tp_max)	76.90	**94.14**

The currently available 10-sample visualization is a fixed descriptive subset used for probe-to-probe continuity. Split into one five-peptide panel per bacterium in Figs. S1 and S2, it matches the main-text summary. Mean pooling is comparatively flat, while max pooling, attention pooling, and FLaGretain more differentiated response profiles. The contrast is also visible numerically: the average within-peptide position-wise response range is 0.0219 for mean pooling versus 0.0972 for FLaGon *E. coli*, and 0.0235 versus 0.0936 on *S. aureus*. We therefore interpret this diagnostic as evidence that the aggregation rule materially shapes which residue-level perturbations survive to the sequence-level prediction, not as evidence that the highlighted residues correspond to biological functional sites or causal mechanisms.

S3 Protein-Side FLaG Mechanism Probes

This section reports four protein-side probes on the ESM2-8M AMP setting: sequence-frequency band knock-out, gate spectral effect, latent-query readout, and structure-proxy stratification. The goal is not to claim a resolved biological mechanism, but to characterize which sequence-frequency components FLaGdepends on, how the learned gate reshapes those components, how the latent queries route information across frequency bins, and whether the resulting summaries co-vary with a lightweight structure proxy. Unless otherwise stated, all summaries are aggregated over 10 aligned seeds on both AMP tasks. Because extending the full per-sample analysis to all peptides is computationally expensive, we randomly sampled 30 peptides from each of *E. coli* and *S. aureus* for these mechanism probes. The main text shows representative cases from this subset, whereas the figures below provide the complete 30-peptide galleries for each bacterium.

S3.1 Sequence-Frequency Band Knockout

For this probe, we follow the prior DCT-based spectral knockout protocol rather than reusing FLaG's internal rFFT computation directly. For a chosen encoder layer ℓ, let $\mathbf{H}_i^{(\ell)} \in \mathbb{R}^{T_i \times D}$ denote the hidden states of peptide i. The hook applies an orthonormal DCT along the sequence dimension, partitions the resulting T_i sequence-frequency coefficients into eight prism bands, removes one band with a notch mask,

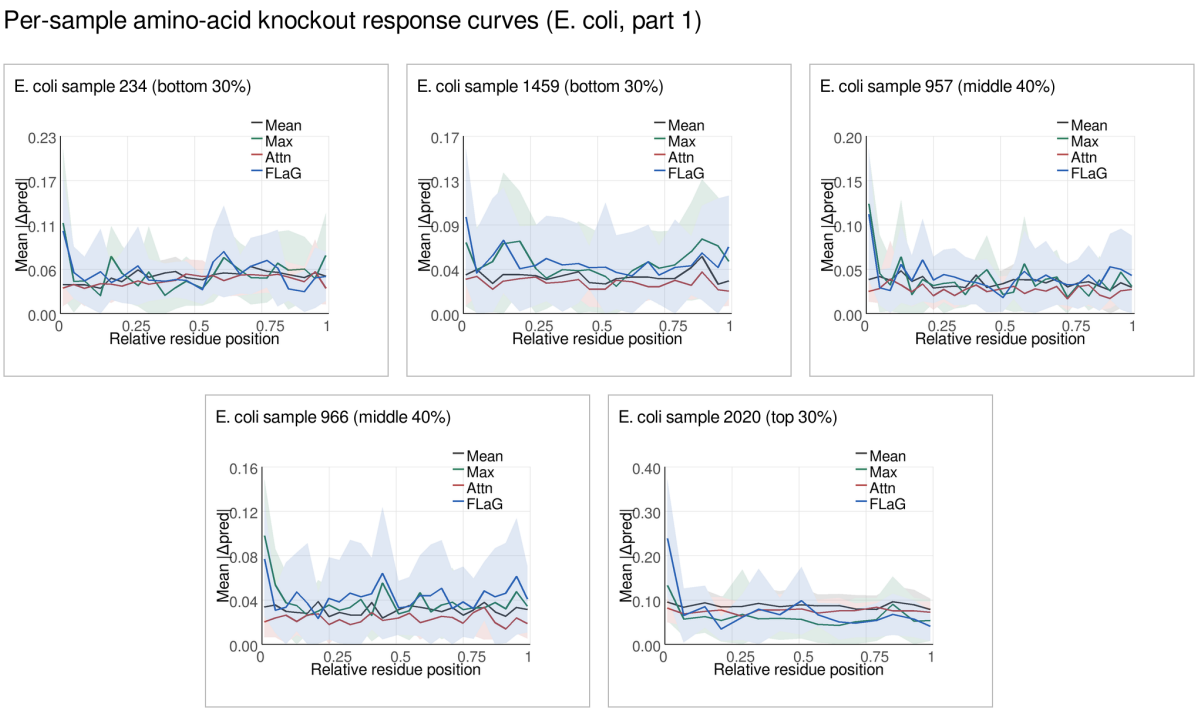

Figure S1: Current *E. coli* single-residue knockout diagnostic on ESM2-8M. Each panel shows the mean absolute sequence-level prediction change over 10 seeds after masking one residue position at a time in the final hidden layer. The five peptides were chosen so that the same cases can be followed across the other protein-side probes.

Table S3: Summary of the band-knockout probe on the two ESM2-8M AMP tasks. Larger MSE increase means stronger reliance on that perturbed band.

Dataset	Mean MSE increase, band 0	Mean MSE increase, second strongest band	Peak layer	Peak layer/band MSE increase
E. coli	1.027	0.409 (band 7)	3	1.920 (layer 3, band 0)
S. aureus	1.245	0.093 (band 1)	3	2.436 (layer 3, band 0)

and reconstructs perturbed hidden states by the inverse DCT. If $\hat{y}_i$ is the baseline prediction and $\hat{y}^{\mathrm{ko}}_{i,\ell,b}$ is the prediction after knocking out band b at layer ℓ, we record the per-peptide response as:

$$\Delta\mathrm{MSE}_{i,\ell,b} = \left(\hat{y}^{\mathrm{ko}}_{i,\ell,b} - y_i\right)^2 - \left(\hat{y}_i - y_i\right)^2$$

The heatmaps report $\Delta\mathrm{MSE}_{i,\ell,b}$ after averaging over the 10 aligned seeds.

Table S3 shows a stable high-level pattern across the two AMP tasks. The band-knockout sensitivity is dominated by the lowest sequence-frequency band in both datasets, and the strongest perturbation effect appears in the middle-to-late layers, peaking at layer 3 in both cases. This pattern is consistent with FLaGrelying primarily on coarse sequence-scale spectral structure rather than on evenly distributed band usage.

S3.2 Gate Spectral Effect

This probe is run during standard inference and uses the intermediate tensors from the FLaG forward pass. Let $\mathbf{F}_i$ denote the pre-gate frequency-token matrix for peptide i, let $\mathbf{g}_i$ denote the learned gate, and let $\tilde{\mathbf{F}}_i = \mathbf{F}_i \odot (1 + \mathbf{g}_i)$ denote the gated frequency-token matrix. Writing $\mathbf{F}_{i,f}$ and $\tilde{\mathbf{F}}_{i,f}$ for the tokens at frequency bin f and $\mathcal{B}_b$ for the set of bins assigned to band b, the implementation first computes the per-bin energies:

$$e^{\mathrm{pre}}_{i,f} = \|\mathbf{F}_{i,f}\|_2^2, \qquad e^{\mathrm{post}}_{i,f} = \|\tilde{\mathbf{F}}_{i,f}\|_2^2$$

The implementation then stores the band-wise means:

$$E^{\mathrm{pre}}_{i,b} = \frac{1}{|\mathcal{B}_b|} \sum_{f \in \mathcal{B}_b} e^{\mathrm{pre}}_{i,f}, \qquad E^{\mathrm{post}}_{i,b} = \frac{1}{|\mathcal{B}_b|} \sum_{f \in \mathcal{B}_b} e^{\mathrm{post}}_{i,f}$$

The galleries therefore show, for each peptide, the blue pre-gate bars $E^{\mathrm{pre}}_{i,b}$ and the orange post-gate bars $E^{\mathrm{post}}_{i,b}$ across the eight sequence-frequency bands. In the aggregated analysis across 10 seeds, the post-gate band energies $E^{\mathrm{post}}_{i,b}$ are consistently higher than the corresponding pre-gate energies $E^{\mathrm{pre}}_{i,b}$ across all eight sequence-frequency bands on both datasets. At the same time, the lowest-frequency bands remain the most energetic before and after gating. This pattern indicates that the gate broadly amplifies the spectrum while preserving the original low-frequency dominance, rather than redistributing energy toward a narrowly selected band.

S3.3 Latent-Query Readout

This probe inspects the latent side of FLaG through the raw cross-attention weights inside the multi-head attention operation $\mathrm{MHA}(\mathbf{Q}, \mathbf{F}, \mathbf{F})$. Let H denote the number of attention heads, and let $\alpha^{(s,h)}_{i,q,f}$ denote the attention weight assigned, for seed s and attention head h, from latent query q to frequency bin f for peptide i. The per-sample galleries visualize:

$$\bar{\alpha}_{i,q,f} = \frac{1}{10H} \sum_{s=1}^{10} \sum_{h=1}^{H} \alpha^{(s,h)}_{i,q,f}$$

These are the raw cross-attention weights averaged first over heads and then over the 10 aligned seeds. These figures therefore extend the same quantity shown in the main text from representative cases to the complete 30-peptide galleries for each bacterium.

Table S4: Structure-proxy stratification of FLaG band sensitivity using helix-propensity buckets on the 30-peptide manifest subset. "Mean absolute band sensitivity" averages |ΔMSE| over all layer/band combinations within each bucket. "Band 0 sensitivity" reports the mean MSE increase for the lowest sequence-frequency band only.

Dataset	Helix-propensity bucket	Mean absolute band sensitivity	Band 0 sensitivity
E. coli	bottom 30%	0.081	0.475
	middle 40%	0.165	0.783
	top 30%	0.237	1.488
S. aureus	bottom 30%	0.047	0.275
	middle 40%	0.128	0.797
	top 30%	0.166	1.060

The raw-score heatmaps confirm a weak-but-visible pattern rather than a flat one. Across the selected peptides, most latent queries within a sample still follow broadly similar cross-attention shapes, but the amplitude of the band-wise fluctuation varies noticeably across samples. In particular, peptides such as *E. coli* sample 1459 and *S. aureus* sample 210 show clearer local oscillations than the more uniform cases, which is why we interpret this probe primarily as evidence of sample-level heterogeneity on the latent side.

S3.4 Structure-Proxy Stratification

This probe moves from single-peptide views to the full 30-peptide subset for each bacterium. For a peptide sequence s_i, we define the helix-propensity proxy as:

$$h_i = \frac{1}{|s_i|} \sum_{a \in s_i} \mathbf{1}[a \in \{E, A, L, M, Q, K, R, H\}]$$

Within each bacterium, the 30 sampled peptides are ranked by h_i and split into bottom-30%, middle-40%, and top-30% buckets. Let $\mathcal{L}$ denote the encoder layers and let $\mathcal{B}$ denote the eight sequence-frequency bands. The per-peptide mean absolute band sensitivity shown in the first and third panels of Fig. S9 is:

$$S_i = \frac{1}{|\mathcal{L}|\,|\mathcal{B}|} \sum_{\ell \in \mathcal{L}} \sum_{b \in \mathcal{B}} |\Delta\mathrm{MSE}_{i,\ell,b}|$$

The supplementary table also reports the band-0-only summary:

$$S_i^{(0)} = \frac{1}{|\mathcal{L}|} \sum_{\ell \in \mathcal{L}} \Delta\mathrm{MSE}_{i,\ell,0}$$

Here, band 0 is the lowest sequence-frequency band. Using the gate statistic from the previous subsection, the per-peptide gate-weight spread shown in the second and fourth panels of Fig. S9 is:

$$R_i = \max_b w_{i,b} - \min_b w_{i,b}$$

The band weight $w_{i,b}$ is defined as:

$$w_{i,b} = \frac{1}{|\mathcal{B}_b|} \sum_{f \in \mathcal{B}_b} \frac{e_{i,f}^{\mathrm{post}}}{e_{i,f}^{\mathrm{pre}} + \varepsilon}$$

Each bar in Fig. S9 and each table entry in Table S4 is the corresponding bucket-wise mean over peptides assigned to that helix-propensity bucket.

Table S4 and Fig. S9 extend the same picture from another angle. Under this simple structure proxy, peptides in the higher helix-propensity bucket exhibit stronger average band sensitivity and markedly larger band-0 reliance in both datasets. The gate effect shifts more mildly and does not show a comparably stable monotonic trend across both bacteria. Accordingly, we treat this analysis as descriptive support for a protein-side inductive-bias story rather than as evidence of a resolved biological mechanism.

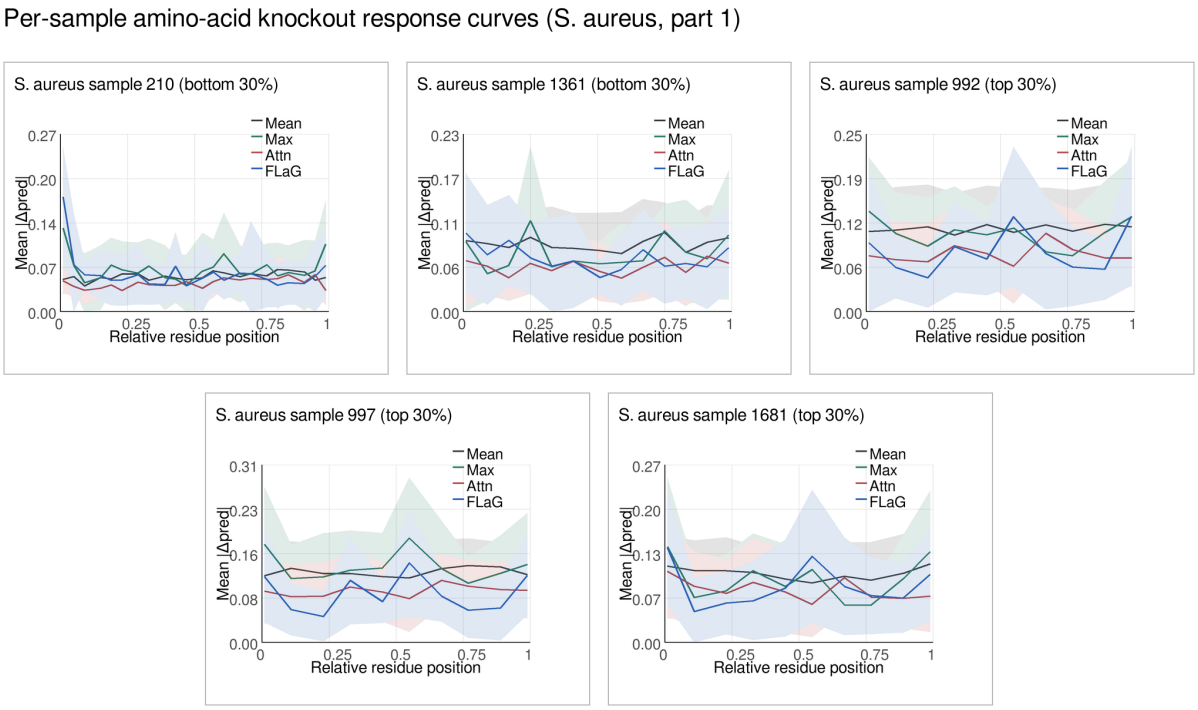

Figure S2: Current *S. aureus* single-residue knockout diagnostic on ESM2-8M. The same flattening contrast between mean pooling and the more structured alternatives remains visible across the five selected peptides.

Figure S3: Full *E. coli* band-knockout gallery. The 30 per-sample heatmaps are shown in a single 6×5 layout. Each panel visualizes $\Delta\mathrm{MSE}_{i,\ell,b}$ over layer ℓ and band b, averaged over the 10 aligned seeds and formatted consistently with the main-text representative panels.

Figure S4: Full *S. aureus* band-knockout gallery. The 30 per-sample heatmaps are shown in a single 6 × 5 layout. Each panel visualizes $\Delta\mathrm{MSE}_{i,\ell,b}$ over layer ℓ and band b, averaged over the 10 aligned seeds and formatted consistently with the main-text representative panels.

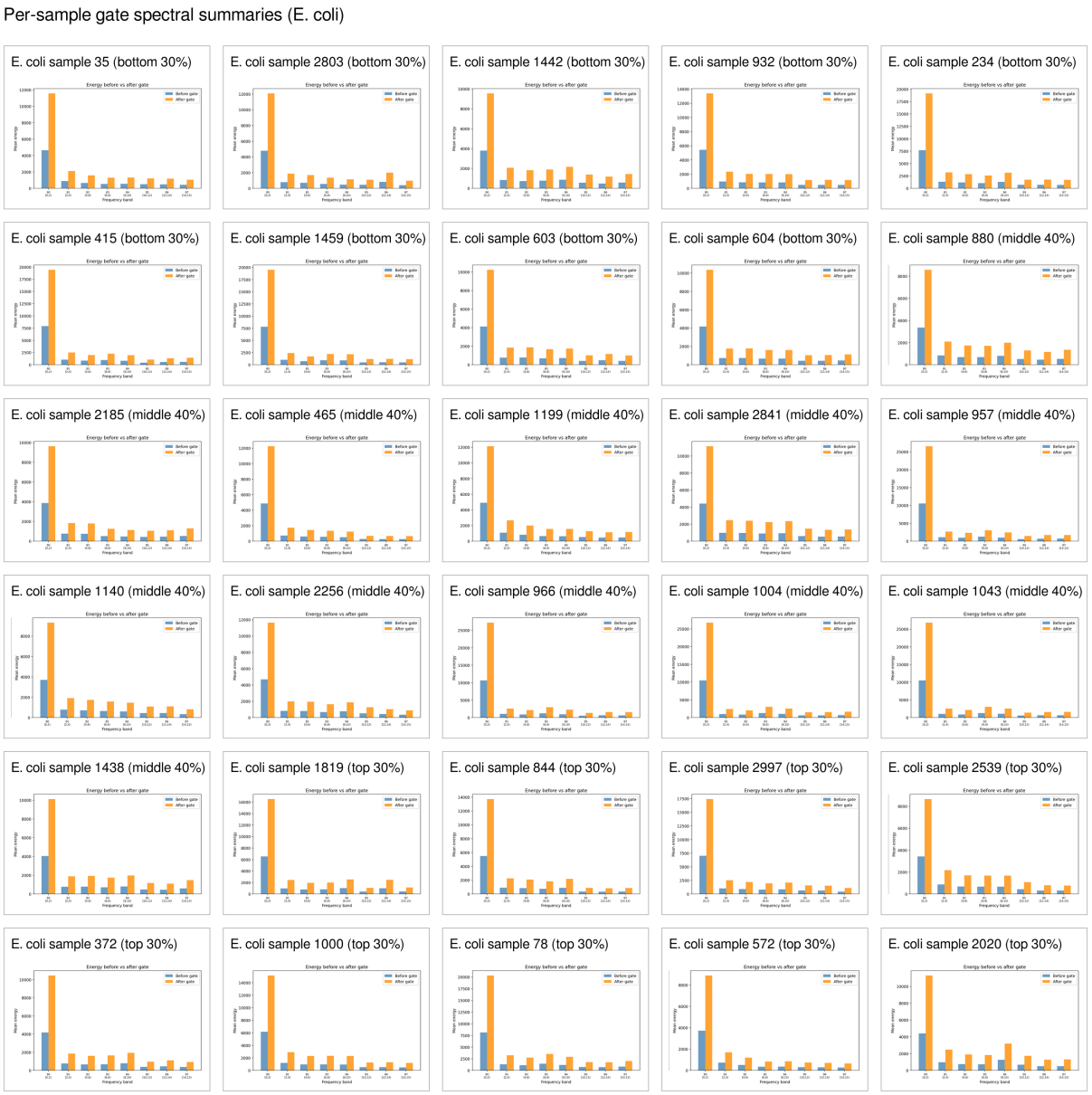

Figure S5: Full *E. coli* gate-energy gallery. The 30 per-sample summaries are shown in a single 6×5 layout. Within each panel, the blue bars show $E_{i,b}^{\mathrm{pre}}$ and the orange bars show $E_{i,b}^{\mathrm{post}}$ across the eight sequence-frequency bands, using the same view as in the main text.

Figure S6: Full *S. aureus* gate-energy gallery. The 30 per-sample summaries are shown in a single 6 × 5 layout. Within each panel, the blue bars show $E_{i,b}^{\mathrm{pre}}$ and the orange bars show $E_{i,b}^{\mathrm{post}}$ across the eight sequence-frequency bands, using the same view as in the main text.

Per-sample raw cross-attention score heatmaps (E. coli)

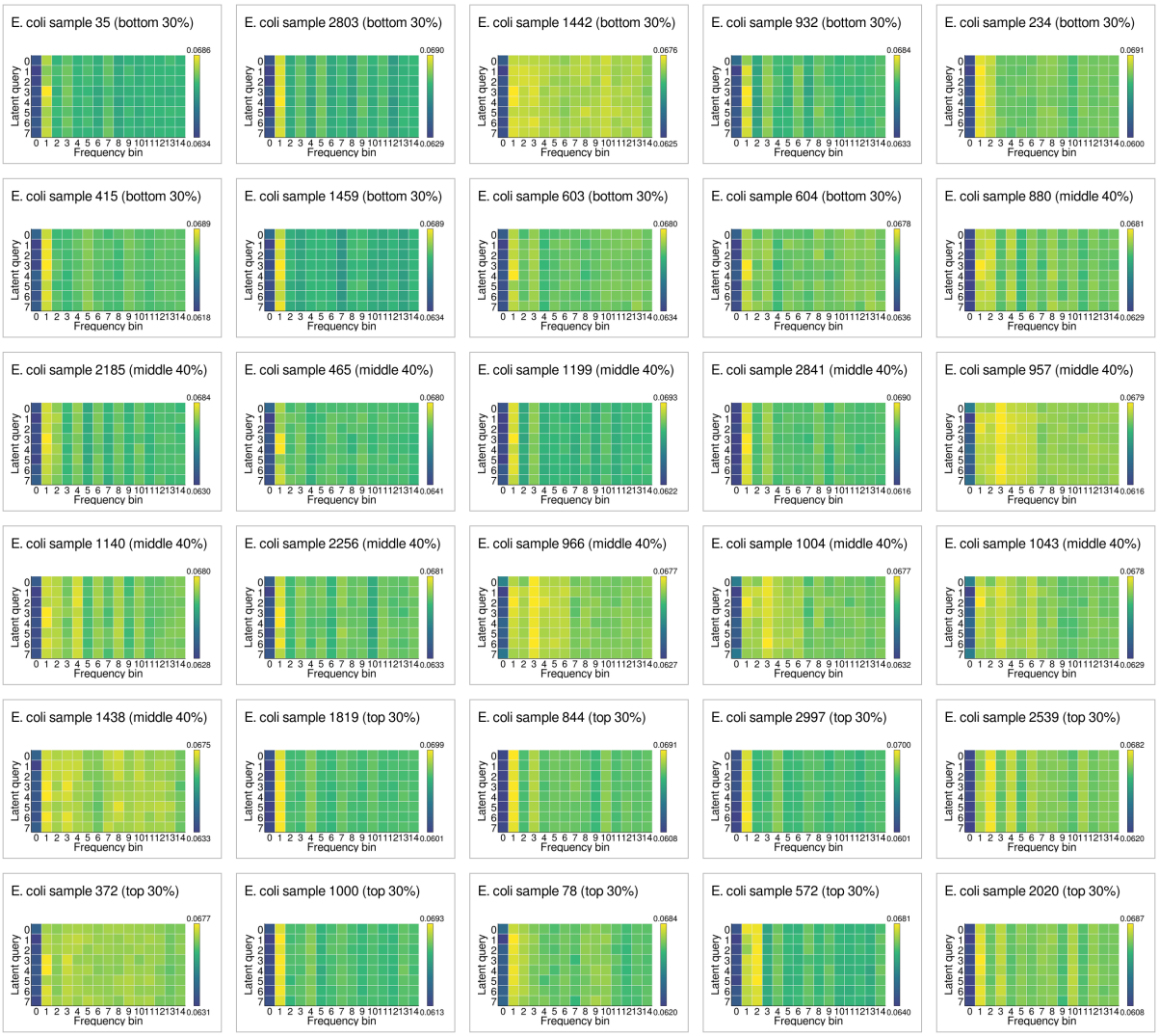

Figure S7: Full *E. coli* latent-query raw cross-attention gallery. The 30 per-sample heatmaps are shown in a single 6×5 layout; each panel visualizes $\bar{\alpha}_{i,q,f}$ with frequency bin on the horizontal axis and latent query on the vertical axis, using the same formatting as in the main text.

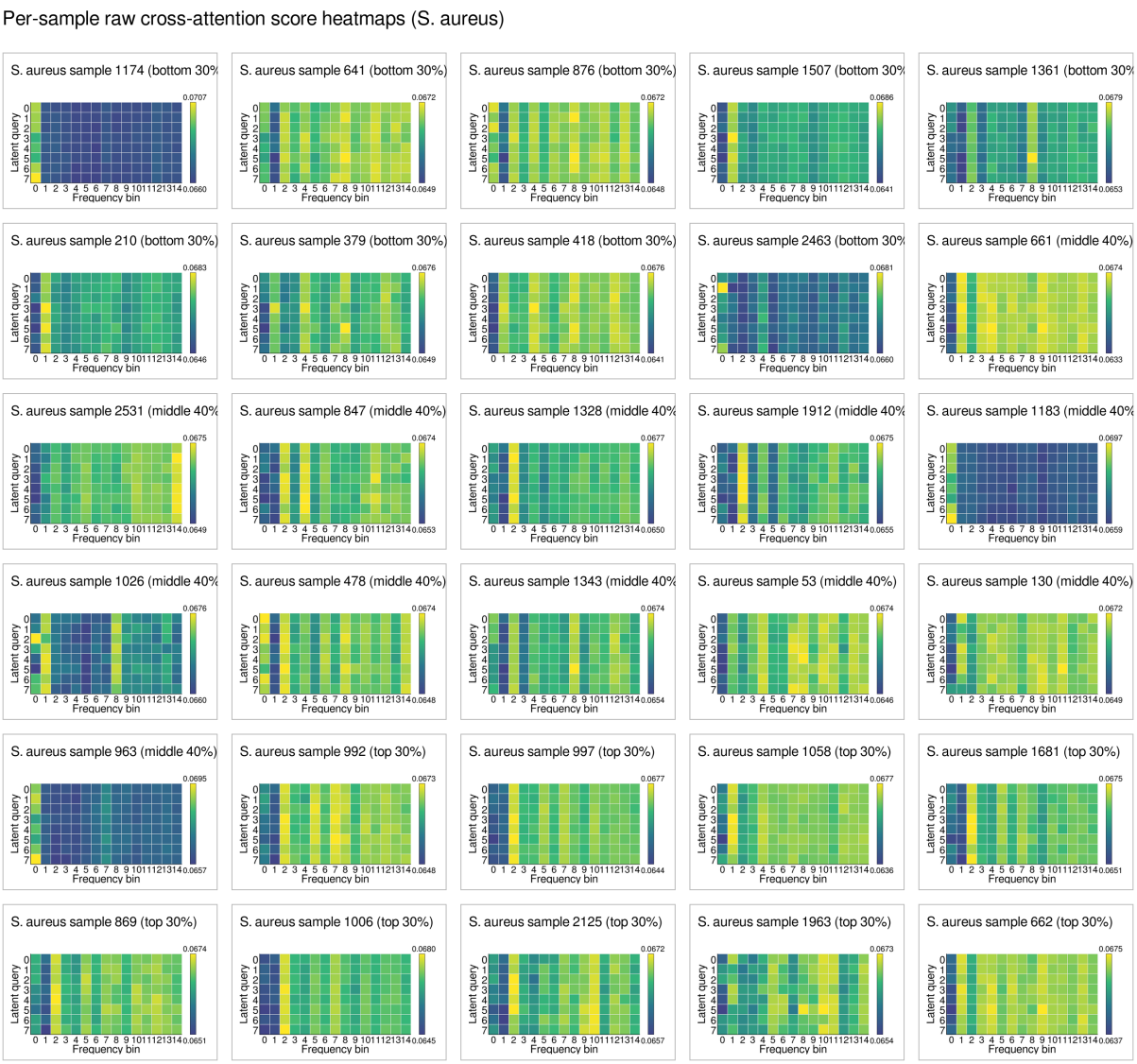

Figure S8: Full *S. aureus* latent-query raw cross-attention gallery. The 30 per-sample heatmaps are shown in a single 6×5 layout; each panel visualizes $\bar{\alpha}_{i,q,f}$ with frequency bin on the horizontal axis and latent query on the vertical axis, using the same formatting as in the main text.

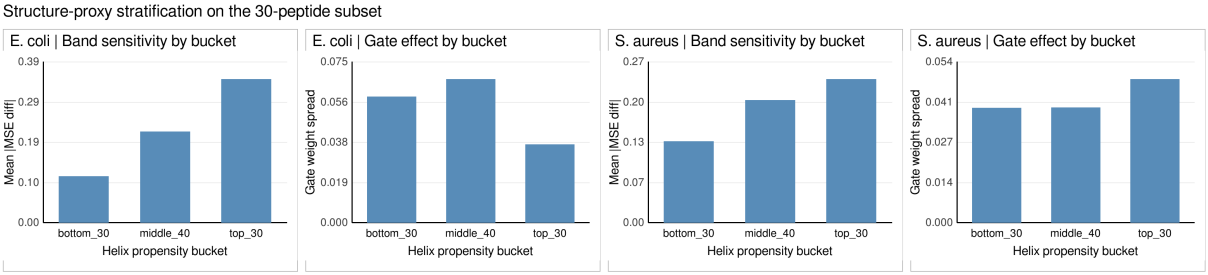

Figure S9: Structure-proxy stratification on the 30-peptide subset. From left to right, the four panels show the bucket-wise means of S_i for *E. coli*, R_i for *E. coli*, S_i for *S. aureus*, and R_i for *S. aureus*. Higher helix-propensity buckets exhibit stronger band sensitivity in both datasets, while the gate trend is milder and more dataset-dependent.