
Pre-Structural Embeddings for Protein-RNA Binding Affinity Prediction

Josef Hanke

Centre for Misfolding Diseases
University of Cambridge

Sebastian Pujalte Ojeda

Centre for Misfolding Diseases
University of Cambridge

Shengyu Zhang

Centre for Misfolding Diseases
University of Cambridge

Werngard Czechitzky

Medicinal Chemistry, R&ED
AstraZeneca, Gothenburg

Leonardo De Maria

Medicinal Chemistry, R&ED
AstraZeneca, Gothenburg

Michele Vendruscolo*

Centre for Misfolding Diseases
University of Cambridge

Abstract

Biomolecular foundation models such as AlphaFold 3 and Boltz-2 are typically used through their decoded structural outputs. However, many biologically important molecules, including RNA and intrinsically disordered proteins, populate conformational ensembles that cannot be fully represented by a single static structure. We propose that intermediate representations from these models contain rich pre-structural embeddings useful for downstream tasks involving conformationally flexible biomolecules. Including predicted-structure information is expensive and provides little value in this context. We validate this representation strategy on protein-RNA binding affinity prediction, a task where both partners are structurally heterogeneous and structure-based methods generalize poorly outside related examples. We introduce ZeroFold, a transformer-based model that integrates protein and RNA pre-structural embeddings from Boltz-2 through cross-modal attention to predict binding affinity directly from sequence. To ensure rigorous evaluation, we construct PRADB, a curated benchmark of 2,582 unique protein-RNA pairs with experimentally measured affinities, split at a strict 40% sequence identity threshold for both modalities. On the held-out test set, ZeroFold achieves a Spearman correlation of 0.64, approaching the empirical ceiling imposed by inter-replicate experimental variability [Landrum and Riniker, 2024, Kramer et al., 2012]. Under progressively stricter evaluations that remove training-set overlap with competing methods, ZeroFold performs consistently better than the leading structure-based predictor (CoPRA) and sequence-based predictor (DeePNAP), with the advantage widening as overlap is reduced. Ablations replacing pre-structural embeddings with language-model features or Boltz-2 distogram-derived structural features substantially degrade performance. These results support pre-structural embeddings as an optimal representation strategy for learning the interactions of flexible biomolecules.

*Correspondence: mv245@cam.ac.uk

1 Introduction

Protein-RNA interactions are fundamental to cellular regulation and are involved in key regulatory functions across cell types and tissues [Glinos et al., 2022, Liu et al., 2024]. Their misregulation has been implicated in a wide range of diseases, [Tao et al., 2024, Vaghf et al., 2022, Van Nostrand et al., 2020, Corley et al., 2020, Gebauer et al., 2021]. Protein-RNA interactions remain challenging to predict as their complexes typically involve large shallow surface areas that lack the hydrophobic pockets characteristic of traditionally druggable targets [Seufert et al., 2022], and the binding frequently involves disordered protein regions.

Several computational strategies have been developed to address this problem, broadly categorised into two approaches. Structure-based methods, such as PRdeltaGPred [Hong et al., 2023], PredPRBA [Deng et al., 2019], as PRA-Pred [Harini et al., 2024] and the state-of-the-art CoPRA [Han et al., 2025] extract features at the binding interface from experimentally resolved structures for training affinity predictors. Sequence-based models, such as PNAB [Yang and Deng, 2019] and recently DeePNAP [Pandey et al., 2024] avoid this structural dependency but perform variably across test sets [Han et al., 2025]. A central limitation across all these models is the scarcity of high-quality training data, with existing datasets often either noisy and heterogeneous at scale or clean but limited in size.

A key insight motivating this work is that the intermediate (pre-structural) embeddings from structure prediction algorithms encode richer information about binding than the final predicted structures themselves. Extracting these embeddings before the structure decoding step avoids the information loss incurred when a conformational ensemble is reduced to a small set of decoded structures [Brotzakis et al., 2025, Zhang et al., 2025], which is particularly consequential for flexible molecules such as RNA. This principle has recently been validated for protein-ligand binding affinity prediction, [Zhang et al., 2025, Schnapka et al., 2026]. Similar approaches have been successful across a range of other tasks, such as predicting conformational protein ensembles [Lewis et al., 2025], binding pockets [Gazizov et al., 2026], and protein-protein interactions [Shi et al., 2026].

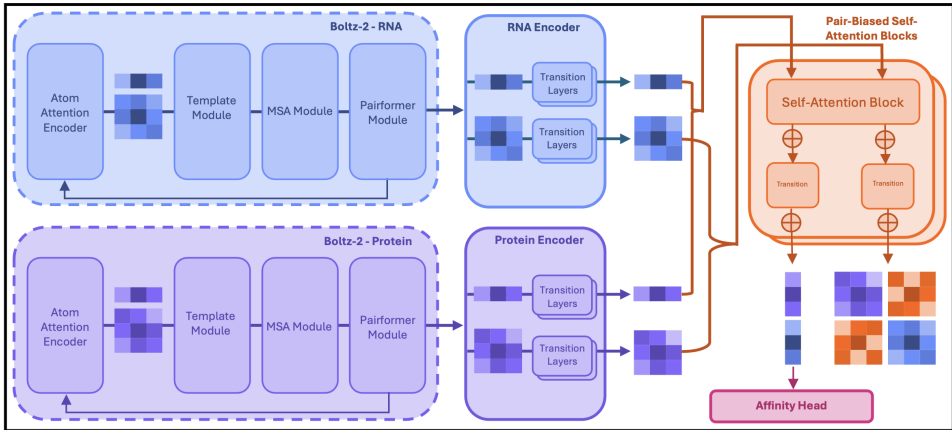


Figure 1: **Schematic of ZeroFold.** ZeroFold is a transformer-based predictor of protein-RNA binding affinity. Pre-structural embeddings are extracted from the Boltz-2 trunk before its structure-decoding step. Encoders refine both streams followed by a stack of pair-biased self-attention blocks. Tokens are pooled and passed to an affinity head that outputs a scalar pK_D . Dotted boxes denote components with fixed parameters (the Boltz-2 trunk); solid boxes denote trainable components.

Our contributions.

We assembled **PRADB**, the largest ML-ready dataset for this task, comprising 2,582 unique protein-RNA pairs with 3,939 experimentally measured affinities, with a rigorous benchmark split at a strict 40% sequence identity threshold for both protein and RNA sequences.

We introduce **ZeroFold** (Fig. 1), the first sequence-based protein-RNA affinity predictor built on pre-structural embeddings, which outperforms both the leading structure-based predictor (CoPRA) and the leading sequence-based predictor (DeePNAP).

Through **controlled ablations**, we demonstrate that pre-structural embeddings outperform both language-model and decoded structural features. Further, we find that incorporating decoded structures provides no measurable benefit and may be detrimental, consistent with the conformational flexibility of RNA being poorly captured by decoded structures.

2 Methods

Dataset. To train and evaluate ZeroFold, we assembled PRADB, a curated dataset of 2,582 unique protein-RNA pairs with experimentally measured binding affinities (pK_D) drawn from four complementary sources: ProNAB [Harini et al., 2022], BioLiP2 [Zhang et al., 2024], the UTexas Aptamer Database [Askari et al., 2024], and PDBBind v2020 [Liu et al., 2017] (see Figure A.1). The 2,582 unique pairs derive from 502 distinct proteins and 1,392 distinct RNA sequences, with protein sequence lengths spanning a wide range that encompasses both compact RNA-binding domains and larger multidomain assemblies, and RNA sequence lengths similarly covering a broad spectrum from short aptamers to longer structured RNAs. Further details are in Appendix C.

ZeroFold Model. ZeroFold comprises four main components: the RNA encoder, the protein encoder, the pair-biased self-attention blocks, and the affinity prediction head. ZeroFold employs Boltz-2 [Passaro et al., 2025] to extract **representations** for both protein and nucleic acid chains. These are the single-sequence and pair representations, which encode the per-residue (or per-nucleotide) and residue-residue (or nucleotide-nucleotide) spatial and evolutionary relationships, respectively. They are extracted from the final layer of the Boltz-2 trunk (`recycling_steps` = 3), just before the structure module decoding. Both **RNA and protein encoders** map their Boltz-2 trunk representations into refined single and pair feature streams. They are each passed through two successive transition layers with dropout and residual connections. Encoded protein and RNA single representations are concatenated into a joint token sequence, while the pair representations are combined into a block-diagonal tensor with the off-diagonal blocks being initialised to zero. The joint single and pair streams are refined by stacked **pair-biased self-attention blocks** [Zhang et al., 2025]. The affinity head then applies masked mean pooling separately over the final protein and RNA single-token representations, concatenates the two pooled vectors, and passes them through a layer-normalised MLP with residual blocks to output a scalar pK_D prediction. Full details are in Appendix D.

Clustering and sample weighting. To prevent data leakage, protein and nucleic acid sequences were clustered independently with MMseqs2 `easy-cluster` [Kallenborn et al., 2025] at 40% sequence identity and 80% coverage, and train/validation/test splits were assigned at the cluster level. Following AlphaFold-Multimer [Evans et al., 2021], training samples were weighted inversely to the product of their protein and RNA cluster sizes, down-weighting densely sampled families and up-weighting rare pairs. See Appendix C for details.

Training and ablations. All models were trained for 100 epochs on a single NVIDIA GH200 120GB GPU using the Adam optimizer [Kingma and Ba, 2014], using identical splits. Each epoch comprised 900 samples from the training set weighted by cluster representation. For each ablation, the same 40-trial Bayesian hyperparameter sweeps with Hyperband early termination were performed, and the hyperparameters that achieved the best validation SCC were selected. The sequence-LM representation ablation used ESM-2 t33 650M UR50D protein embeddings [Lin et al., 2023] and RiNALMo giga v1 RNA embeddings [Penić et al., 2025]. To generate distograms, we discretised intra-chain C_α (protein) and C'_4 (RNA) pairwise distances from five Boltz-2 predicted structures into 64 bins over 0–50Å and averaged the per-sample one-hot encodings, giving a soft distogram as the pair input z , with the Boltz-2 per-residue embedding.

3 Results

ZeroFold performance The final model performed strongly on the test set, achieving a mean absolute error (MAE) of 1.22, a root-mean-square error (RMSE) of 1.47, a Pearson correlation coefficient (PCC) of 0.63, and a Spearman correlation coefficient (SCC) of 0.64. Replicate measurements, where present, agree only moderately (PCC 0.68 and 0.58 on the train- and test-set pairs), indicating substantial label noise and an approximate upper bound on model performance (Appendix Figure A.1E)[Kramer et al., 2012, Landrum and Riniker, 2024].

In addition to its predictive performance, ZeroFold offers a substantial computational advantage by precomputing structural embeddings for proteins and RNAs. For screening libraries of N protein molecules against M RNA molecules, the model requires only $N + M$ expensive embedding passes followed by cheap affinity head predictions, in comparison to the $N \times M$ expensive steps in conventional structure-based models (Appendix Figure A.2), saving time and cost in large-scale protein–RNA screening.

Comparison with SOTA predictors We compared the performance of ZeroFold to the leading structure- and sequence- based protein-RNA affinity predictors, CoPRA [Han et al., 2025] and DeePNAP [Pandey et al., 2024]. Structure-based predictors, can leverage interaction information encoded in 3D structures, but only when structures are available. In this study, predicted complex structures were first generated using Boltz-2 and subsequently used as input for CoPRA. However, both CoPRA and DeePNAP apply much less stringent sequence similarity thresholds (70% and 100% respectively) on the proteins in their test sets, and none on the RNA side. This potentially inflates the reported performance due to data leakage.

To compare ZeroFold and these models, performance was evaluated on four progressively stricter subsets of the ZeroFold test set (Table 1), from setting A, where everything was included, regardless of overlap with the training data of CoPRA or DeepNAP, to Setting D where a stricter 40% protein-sequence-identity threshold was applied. A set where both protein and RNA sequences had $< 40\%$ identity with the corresponding training sets was omitted because it left too few samples.

Table 1: **ZeroFold comparison with CoPRA and DeePNAP.** Comparison on the full test set (A) and progressively stricter subsets (B–D), imposing increasingly stringent constraints on protein sequence overlap with the respective training datasets.

Subset	CoPRA comparison					DeePNAP comparison				
	n	PCC		SCC		n	PCC		SCC	
		CoPRA	ZeroFold	CoPRA	ZeroFold		DeePNAP	ZeroFold	DeePNAP	ZeroFold
A) All	297	0.50	0.63 $\pm$ 0.01	0.54	0.64 $\pm$ 0.01	297	0.46	0.63 $\pm$ 0.01	0.35	0.64 $\pm$ 0.01
B) Protein not in reference	280	0.47	0.67 $\pm$ 0.00	0.52	0.67 $\pm$ 0.01	48	0.48	0.58 $\pm$ 0.04	0.42	0.59 $\pm$ 0.05
C) Max protein identity $< 70\%$	162	0.46	0.71 $\pm$ 0.01	0.55	0.69 $\pm$ 0.01	26	0.44	0.58 $\pm$ 0.04	0.45	0.59 $\pm$ 0.02
D) Max protein identity $< 40\%$	97	0.22	0.54 $\pm$ 0.01	0.23	0.49 $\pm$ 0.05	23	0.41	0.50 $\pm$ 0.06	0.34	0.50 $\pm$ 0.04

As expected, the performance of CoPRA and DeePNAP decreased as the evaluation criteria became more stringent, with PCC falling from 0.50 and 0.46 in A to 0.22 and 0.41 in D, reflecting the difficulty of predicting affinities without closely related sequences. In contrast, the performance of ZeroFold remained largely consistent across these subsets, with PCC around 0.6. Notably, ZeroFold outperformed both even in the least restrictive setting, where the other models’ training sets contained identical or highly similar protein sequences, and it does so without experimentally determined structural inputs. Bootstrap confidence intervals for all entries are given in Appendix B.

Ablation study on sequence embeddings To quantify the contribution of pre-structural embeddings and to justify encoding the protein and RNA separately, four alternative representations were compared in Table 2. First, each protein–RNA pair was embedded in a joint Boltz-2 pass, so that pair features span both chains. Second, the full Boltz-2 pipeline was run, including diffusion, over the default five structures to incorporate explicit structural representations. Third, the structure was discarded entirely, encoding protein and RNA sequences with ESM-2 [Lin et al., 2023] and RiNALMo [Penić et al., 2025] respectively. Fourth, to test the contribution of the pair features, only the Boltz-2 single representation was retained and the pair input was set to zero.

In Table 2, pre-structural embeddings are the strongest representation (A; PCC 0.628 ± 0.015). Encoding the protein and RNA jointly gives no benefit and is consistently slightly worse (C; 0.578 ± 0.009), supporting our decision to encode the two chains separately. Adding explicit predicted structure to either pre-structural variant lowers performance, markedly for the separately encoded pair (B; 0.478 ± 0.104 , the most variable condition) and mildly for the joint one (D; 0.564 ± 0.025). This indicates that the geometric signal in the distogram is already encoded within the trunk features; we attribute the residual advantage of the pre-structural representation to ensemble-level information retained before decoding, though we do not test this directly. Frozen language-model embeddings perform markedly worse than the trunk representation (E; PCC 0.521 ± 0.060 versus 0.628 ± 0.015 for A); replacing them with a one-hot encoding alongside the same distogram does not close the gap

Table 2: **Representation ablation.** Held-out test-set performance for different input representations. The joint encoder takes trunk embeddings from a single Boltz-2 pass over the protein–RNA complex. Sequence Language Model takes ESM-2 and RiNALMo token embeddings. The distogram representations are derived from inter-residue distance matrices computed from Boltz-2 predicted coordinates. The single-only condition retains the Boltz-2 single representation s and sets the pair input z to zero. Mean $\pm$ 1 s.d. across the top-3 of 5 seeds selected by validation SCC.

Representation	PCC	SCC
A - Pre-structural (PS)	0.63 $\pm$ 0.01	0.64 $\pm$ 0.01
B - PS + Distogram	0.48 $\pm$ 0.10	0.50 $\pm$ 0.08
C - Joint encoding	0.58 $\pm$ 0.01	0.60 $\pm$ 0.01
D - Joint + Distogram	0.56 $\pm$ 0.03	0.57 $\pm$ 0.02
E - Sequence Language Model	0.52 $\pm$ 0.06	0.50 $\pm$ 0.08
F - Sequence + Distogram	0.54 $\pm$ 0.04	0.59 $\pm$ 0.02
G - PS single only (no pair representation)	0.49 $\pm$ 0.02	0.52 $\pm$ 0.03

(F; 0.535 ± 0.037), indicating that neither evolutionary nor decoded structural features substitute for the trunk representation. Retaining only the single representation and zeroing the pair input also lowers performance substantially (G; PCC 0.493 ± 0.021 , SCC 0.524 ± 0.033), to a level comparable to the language-model baseline, indicating that the pair features carry much of the trunk’s predictive signal.

Discussion and limitations Several limitations remain. The available quantitative protein-RNA affinity data are modest in scale relative to other areas of machine learning. Thus, prediction accuracy is likely to vary across protein classes, RNA types, and affinity regimes. Because Boltz-2 was trained on public structural data, some PRADB test pairs may have been seen during its pretraining, and we have not excluded this overlap. It does not confound the comparison with CoPRA, which inherits the same structures from Boltz-2, but it could favour ZeroFold over baselines that do not use Boltz-2 features, such as DeePNAP. Attention analysis could provide additional evidence that the advantage of pre-structural embeddings arises from ensemble encoding rather than from general pretraining.

Pre-structural embeddings are thus an informative transfer strategy for conformationally heterogeneous partners. We hypothesise that this is because trunk representations retain ensemble-level information attenuated by structure decoding [Brotzakis et al., 2025, Passaro et al., 2025]. This is especially consequential for protein-RNA systems, where RNA flexibility, context dependence, and binding-induced rearrangements limit the utility of static structural models [Ganser et al., 2019]. Crucially, the approach requires no resolved structure at inference time, making it directly applicable to the large majority of protein-RNA pairs for which structural data do not exist.

A central finding is that the advantage of ZeroFold over both structure-based and sequence-based baselines widens as evaluation conditions become fairer. Against CoPRA, its lead grows as test subsets are progressively filtered to remove sequences with training-set overlap. Against DeePNAP, ZeroFold attains higher correlations on every subset (A–D), although with only 23–48 pairs remaining on the stricter subsets, these differences are not statistically significant (Appendix B). Together, these comparisons suggest that the gains of ZeroFold reflect genuine generalisation rather than favourable dataset composition.

In conclusion, we introduced ZeroFold, a sequence-based transformer for protein–RNA binding affinity prediction built on pre-structural embeddings from Boltz-2, together with PRADB, a curated benchmark split at a strict 40% sequence identity threshold for both modalities. ZeroFold outperformed leading structure- and sequence-based methods under progressively stricter evaluation, and controlled ablations showed that pre-structural embeddings were more informative than language-model and decoded structural features. These results suggest that intermediate representations from biomolecular foundation models are a more effective strategy than predicted static structures for transferring information to tasks modelling interactions involving conformationally flexible biomolecules such as RNA.

Acknowledgments and Disclosure of Funding

This work was supported by AstraZeneca through a PhD studentship. Part of this work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/T022159/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).

The authors acknowledge the use of resources provided by the Isambard-AI National AI Research Resource (AIRR). Isambard-AI is operated by the University of Bristol and is funded by the UK Government's Department for Science, Innovation and Technology (DSIT) via UK Research and Innovation; and the Science and Technology Facilities Council [ST/AIRR/I-A-I/1023] [McIntosh-Smith et al., 2024].

References

- Ali Askari, Sumedha Kota, Hailey Ferrell, Shriya Swamy, Kayla S Goodman, Christine C Okoro, Isaiah C Spruell Crenshaw, Daniela K Hernandez, Taylor E Oliphant, Akshata A Badrayani, et al. Utexas aptamer database: the collection and long-term preservation of aptamer sequence information. *Nucleic Acids Research*, 52(D1):D351–D359, 2024.
- Z Faidon Brotzakis, Shengyu Zhang, Mhd Hussein Murtada, and Michele Vendruscolo. Alphafold prediction of structural ensembles of disordered proteins. *Nature communications*, 16(1):1632, 2025.
- Meredith Corley, Margaret C Burns, and Gene W Yeo. How rna-binding proteins interact with rna: molecules and mechanisms. *Molecular cell*, 78(1):9–29, 2020.
- Lei Deng, Wenyi Yang, and Hui Liu. Predprba: prediction of protein-rna binding affinity using gradient boosted regression trees. *Frontiers in genetics*, 10:637, 2019.
- Richard Evans, Michael O’neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Židek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *bioRxiv*, pages 2021–10, 2021.
- Laura R Ganser, Megan L Kelly, Daniel Herschlag, and Hashim M Al-Hashimi. The roles of structural dynamics in the cellular functions of rnas. *Nature reviews Molecular cell biology*, 20(8):474–489, 2019.
- Artem Gazizov, Anna Lian, Casper Goverde, Jody Mou, Sergey Ovchinnikov, and Nicholas F Polizzi. Af2bind: predicting small-molecule binding sites using the pair representation of alphafold2. *Nature Methods*, pages 1–10, 2026.
- Fátima Gebauer, Thomas Schwarzl, Juan Valcárcel, and Matthias W Hentze. Rna-binding proteins in human genetic disease. *Nature Reviews Genetics*, 22(3):185–198, 2021.
- Dafni A Glinos, Garrett Garborcauskas, Paul Hoffman, Nava Ehsan, Lihua Jiang, Alper Gokden, Xiaoguang Dai, François Aguet, Kathleen L Brown, Kiran Garimella, et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature*, 608(7922):353–359, 2022.
- Rong Han, Xiaohong Liu, Tong Pan, Jing Xu, Xiaoyu Wang, Wuyang Lan, Zhenyu Li, Zixuan Wang, Jiangning Song, Guangyu Wang, et al. Copra: Bridging cross-domain pretrained sequence models with complex structures for protein-rna binding affinity prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 246–254, 2025.
- K Harini, M Sekijima, and M Michael Gromiha. Pra-pred: Structure-based prediction of protein-rna binding affinity. *International Journal of Biological Macromolecules*, 259:129490, 2024.
- Kannan Harini, Ambuj Srivastava, Arulsamy Kulandaisamy, and M Michael Gromiha. Pronab: database for binding affinities of protein–nucleic acid complexes and their mutants. *Nucleic acids research*, 50(D1):D1528–D1534, 2022.

- Xu Hong, Xiaoxue Tong, Juan Xie, Pinyu Liu, Xudong Liu, Qi Song, Sen Liu, and Shiyong Liu. An updated dataset and a structure-based prediction model for protein–rna binding affinity. *Proteins: Structure, Function, and Bioinformatics*, 91(9):1245–1253, 2023.
- Felix Kallenborn, Alejandro Chacon, Christian Hundt, Hassan Sirelkhatim, Kieran Didi, Sooyoung Cha, Christian Dallago, Milot Mirdita, Bertil Schmidt, and Martin Steinegger. Gpu-accelerated homology search with mmseqs2. *Nature Methods*, 22(10):2024–2027, 2025.
- Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2014.
- Christian Kramer, Tuomo Kallioikoski, Peter Gedeck, and Anna Vulpetti. The experimental uncertainty of heterogeneous public k i data. *Journal of medicinal chemistry*, 55(11):5165–5173, 2012.
- Gregory A Landrum and Sereina Riniker. Combining ic50 or k i values from different sources is a source of significant noise. *Journal of chemical information and modeling*, 64(5):1560–1567, 2024.
- Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew YK Foong, Victor García Satorras, Osama Abdin, Bastiaan S Veeling, Iryna Zaporozhets, et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761):eadv9817, 2025.
- Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of machine learning research*, 18(185):1–52, 2018.
- Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- Shurong Liu, Junhong Huang, Jie Zhou, Siyan Chen, Wujian Zheng, Chang Liu, Qiao Lin, Ping Zhang, Di Wu, Simeng He, et al. Nap-seq reveals multiple classes of structured noncoding rnas with regulatory functions. *Nature Communications*, 15(1):2425, 2024.
- Zhihai Liu, Minyi Su, Li Han, Jie Liu, Qifan Yang, Yan Li, and Renxiao Wang. Forging the basis for developing protein–ligand interaction scoring functions. *Accounts of chemical research*, 50(2):302–309, 2017.
- Simon McIntosh-Smith, Sadaf R. Alam, and Christopher Woods. Isambard-ai: a leadership-class supercomputer optimised specifically for artificial intelligence. 2024. URL <https://doi.org/10.48550/arXiv.2410.11199>.
- Uddeshya Pandey, Sasi M Behara, Siddhant Sharma, Rachit S Patil, Souparnika Nambiar, Debasish Koner, and Hussain Bhukya. Deepnap: A deep learning method to predict protein–nucleic acid binding affinity from their sequences. *Journal of Chemical Information and Modeling*, 64(6):1806–1815, 2024.
- Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, et al. Boltz-2: Towards accurate and efficient binding affinity prediction. *BioRxiv*, 2025.
- Rafael Josip Penić, Tin Vlašić, Roland G Huber, Yue Wan, and Mile Šikić. Rinalmo: general-purpose rna language models can generalize well on structure prediction tasks. *Nature Communications*, 16(1):5671, 2025.
- Vincent Schnapka, Tatiana I Morozova, Samiran Sen, and Massimiliano Bonomi. Atomic resolution ensembles of intrinsically disordered proteins with alphafold. *Nature Communications*, 2026.
- Lisa Seufert, Thomas Benzinger, Michael Ignarski, and Roman-Ulrich Müller. Rna-binding proteins and their role in kidney disease. *Nature Reviews Nephrology*, 18(3):153–170, 2022.
- Liang Shi, Zuobai Zhang, Huiyu Cai, Santiago Miret, Zhi Yang, and Jian Tang. Towards all-atom foundation models for biomolecular binding affinity prediction. In *The Fourteenth International Conference on Learning Representations*, 2026.

- Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.
- Yining Tao, Qi Zhang, Haoyu Wang, Xiyu Yang, and Haoran Mu. Alternative splicing and related rna binding proteins in human health and disease. *Signal transduction and targeted therapy*, 9(1): 26, 2024.
- Atena Vaghf, Behzad Khansarinejad, Ehsanollah Ghaznavi-Rad, and Mahdieh Mondanizadeh. The role of micrnas in diseases and related signaling pathways. *Molecular Biology Reports*, 49(7): 6789–6801, 2022.
- Eric L Van Nostrand, Peter Freese, Gabriel A Pratt, Xiaofeng Wang, Xintao Wei, Rui Xiao, Steven M Blue, Jia-Yu Chen, Neal AL Cody, Daniel Dominguez, et al. A large-scale binding and functional map of human rna-binding proteins. *Nature*, 583(7818):711–719, 2020.
- Wenyi Yang and Lei Deng. Pnab: prediction of protein-nucleic acid binding affinity using heterogeneous ensemble models. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 58–63. IEEE, 2019.
- Chengxin Zhang, Xi Zhang, Lydia Freddolino, and Yang Zhang. Biolip2: an updated structure database for biologically relevant ligand–protein interactions. *Nucleic acids research*, 52(D1): D404–D412, 2024.
- Shengyu Zhang, Donghui Huo, Robert I Horne, Yumeng Qi, Sebastian Pujalte Ojeda, Aixia Yan, and Michele Vendruscolo. Sequence-based virtual screening using transformers. *Nature communications*, 16(1):6925, 2025.

A Supplementary Figures

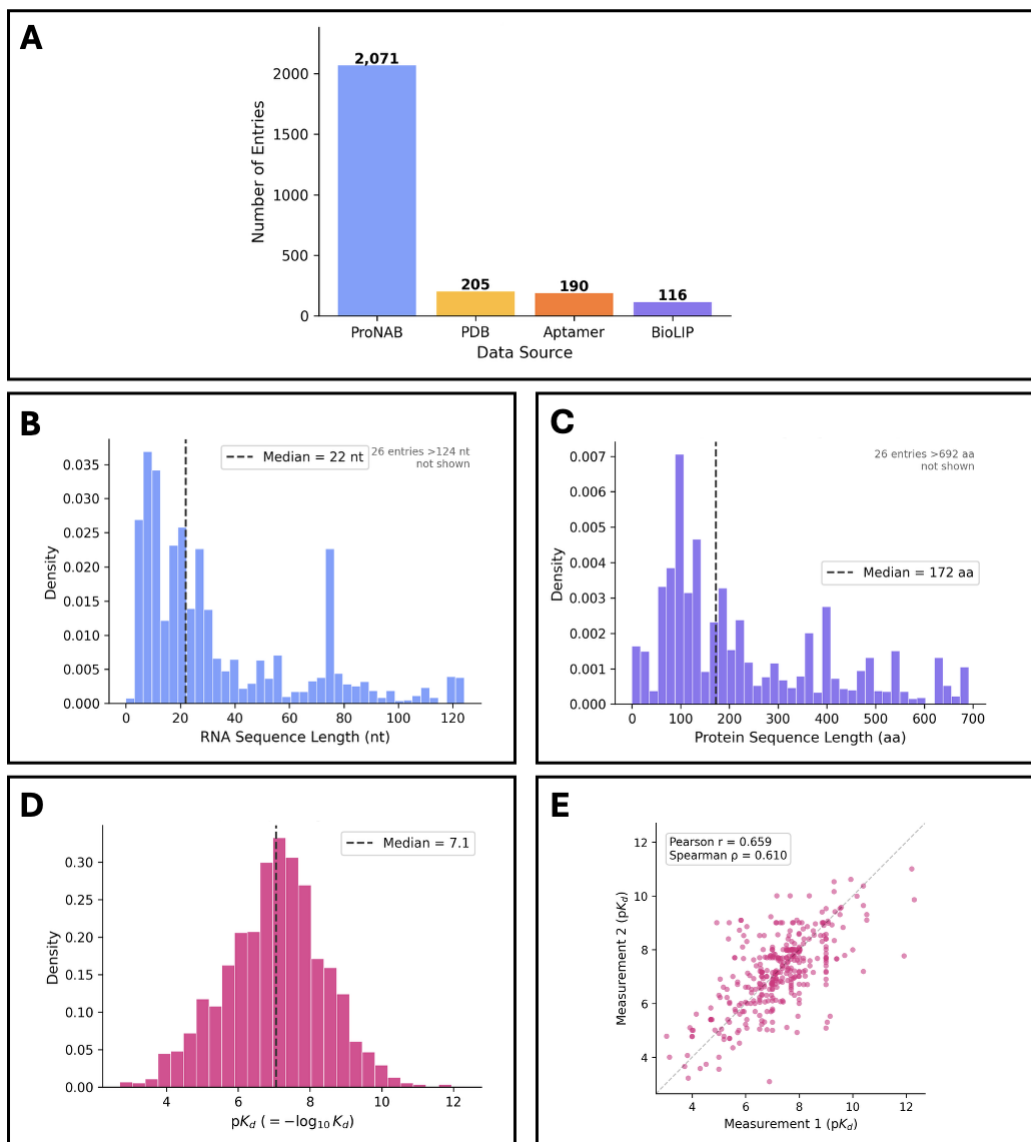


Figure A.1: Composition and characteristics of PRADB. (A) Breakdown of the 2,582 unique protein-RNA pairs in PRADB by source database, illustrating the dominance of ProNAB and the complementary coverage provided by BioLiP2, the UTexas Aptamer Database, and PDBbind v2020. (B) Distribution of RNA sequence lengths across 1,392 unique RNA sequences, reflecting the broad range of RNA classes represented in the dataset. (C) Distribution of protein sequence lengths across 502 unique proteins, highlighting the diversity of RNA-binding protein families included. (D) Distribution of experimentally measured binding affinities (pK_d) across all 2,582 protein-RNA pairs, showing the range and central tendency of affinity values used for model training and evaluation. (E) Agreement between independent replicate affinity measurements for the 351 protein-RNA pairs with two or more experimental values, illustrating the degree of experimental noise that sets a practical upper bound on achievable model performance.

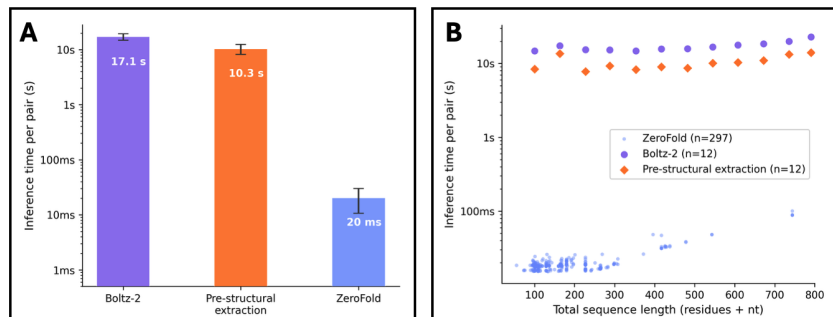


Figure A.2: **Computational efficiency of ZeroFold compared with Boltz-2.** (A) Mean inference time per protein-RNA pair on a log scale for full Boltz-2 structure prediction ($n = 12$ pairs; `recycling_steps` = 3, `sampling_steps` = 200, one diffusion sample), Boltz-2 trunk-only extraction of pre-structural embeddings, and the trained ZeroFold affinity model ($n = 297$ test pairs). Error bars show one standard deviation across samples. Both pipelines require the trunk forward pass; the structure-based pipeline additionally performs diffusion-based structure decoding, whereas ZeroFold passes pre-structural embeddings directly to its affinity model. (B) Inference time as a function of total sequence length (protein residues + nucleotides), showing that the separation between the two regimes holds across the length range of the benchmark. Model loading is excluded from all measurements, as is Boltz-2 MSA preprocessing. Trunk extraction is performed once per chain rather than per pair, so its cost amortises over a screen as $N + M$ passes rather than $N \times M$.

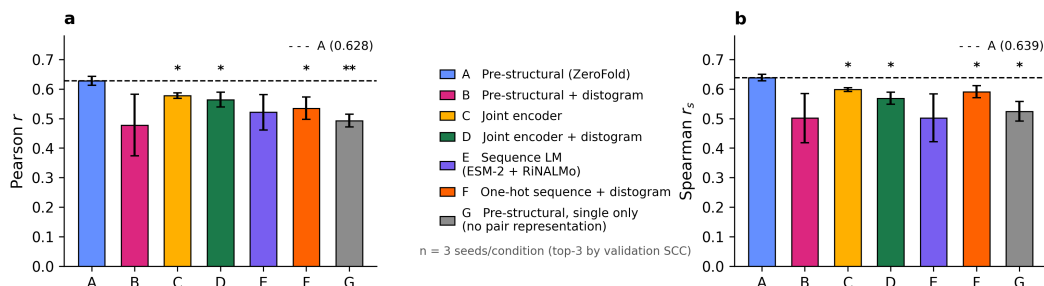


Figure A.3: **Representation ablation for affinity prediction.** Held-out test-set Pearson (a) and Spearman (b) correlations for seven input representations (A–G, as in Table 2). Bars show mean $\pm$ 1 s.d. across the top-3 of 5 seeds selected by validation Spearman; dashed line marks condition A. * and ** denote $p < 0.05$ and $p < 0.01$ respectively by Welch’s two-sample t-test against condition A ($n = 3$ per condition, not corrected for multiple comparisons).

B Uncertainty Estimation for the Benchmark Comparison

The $\pm$ values reported in Table 1 describe variation across training seeds: how much the correlation moves if ZeroFold is retrained from a different initialisation. They do not describe the uncertainty of the correlation estimate itself, which is governed by the number of test pairs available. These are distinct quantities, and they diverge sharply on the stricter subsets: subset D retains 97 pairs in the CoPRA comparison and 23 in the DeePNAP comparison, so a correlation computed there is far less precisely determined than one computed on the full 297-pair test set, however stable training may be.

To quantify this, we estimated 95% percentile bootstrap confidence intervals for every entry of Table 1 (Tables A.1 and A.2). For each of 10,000 replicates, we resampled the test pairs of a subset with replacement to its original size, recomputed each correlation, and took the 2.5th and 97.5th percentiles of the resulting distribution. The resampled indices are drawn once per replicate and applied to *all* models simultaneously (a paired bootstrap), so that every replicate compares the models on an identical set of pairs. This removes the shared sampling variation from the comparison and permits a confidence interval on the difference between models, reported as Δ . ZeroFold correlations are computed separately for each of the three seeds and then averaged, matching the convention

of Table 1; CoPRA predictions are averaged over its five cross-validation folds, as in the main comparison.

The Δ column is the informative one. Because the individual intervals widen substantially on the small subsets, they may overlap even where the paired difference is consistent across resamples; a Δ interval excluding zero indicates that ZeroFold outperforms the baseline on that subset at the 5% level, whereas one spanning zero indicates that the subset is too small to resolve the difference. We report both so that the strength of each comparison can be judged directly rather than inferred from point estimates alone.

Table A.1: **Bootstrap confidence intervals on the CoPRA comparison subsets.** Point estimates with 95% percentile bootstrap confidence intervals (10,000 paired resamples of the test pairs with replacement; the same resampled indices are applied to both models within each replicate). ZeroFold values are the mean across the three seeds, matching Table 1. Δ is the ZeroFold – CoPRA difference; an interval excluding zero indicates a significant advantage on that subset.

Subset	n	CoPRA	ZeroFold	Δ
<i>Pearson correlation</i>				
A) All	297	0.50 [0.42, 0.57]	0.63 [0.57, 0.69]	0.14 [0.05, 0.22]
B) Protein not in reference	280	0.47 [0.40, 0.54]	0.67 [0.62, 0.72]	0.20 [0.13, 0.27]
C) Max protein identity < 70%	162	0.46 [0.35, 0.57]	0.71 [0.65, 0.77]	0.25 [0.15, 0.35]
D) Max protein identity < 40%	97	0.22 [0.01, 0.48]	0.54 [0.37, 0.69]	0.33 [0.15, 0.48]
<i>Spearman correlation</i>				
A) All	297	0.54 [0.46, 0.62]	0.64 [0.57, 0.70]	0.10 [0.02, 0.18]
B) Protein not in reference	280	0.52 [0.43, 0.60]	0.67 [0.61, 0.73]	0.15 [0.08, 0.23]
C) Max protein identity < 70%	162	0.55 [0.42, 0.66]	0.69 [0.59, 0.76]	0.14 [0.04, 0.24]
D) Max protein identity < 40%	97	0.23 [0.03, 0.44]	0.49 [0.34, 0.62]	0.26 [0.04, 0.47]

Table A.2: **Bootstrap confidence intervals on the DeePNAP comparison subsets.** As Table A.1, for the subsets defined against the ProNAB reference. Both models are evaluated on the same pairs within each subset, restricted to those DeePNAP can score; ZeroFold values therefore differ from Table 1 by up to 0.01, where ZeroFold is evaluated on the full test set.

Subset	n	DeePNAP	ZeroFold	Δ
<i>Pearson correlation</i>				
A) All	287	0.46 [0.35, 0.56]	0.63 [0.56, 0.69]	0.17 [−0.08, 0.26]
B) Protein not in reference	48	0.48 [0.23, 0.69]	0.58 [0.38, 0.75]	0.10 [−0.06, 0.27]
C) Max protein identity < 70%	26	0.44 [0.13, 0.75]	0.57 [0.35, 0.78]	0.13 [−0.05, 0.31]
D) Max protein identity < 40%	23	0.41 [0.05, 0.77]	0.49 [0.25, 0.72]	0.08 [−0.12, 0.31]
<i>Spearman correlation</i>				
A) All	287	0.35 [0.22, 0.46]	0.63 [0.56, 0.69]	0.29 [0.18, 0.39]
B) Protein not in reference	48	0.42 [0.14, 0.65]	0.60 [0.40, 0.74]	0.18 [−0.01, 0.37]
C) Max protein identity < 70%	26	0.45 [0.07, 0.73]	0.58 [0.26, 0.76]	0.14 [−0.21, 0.47]
D) Max protein identity < 40%	23	0.34 [0.02, 0.71]	0.47 [0.11, 0.71]	0.13 [−0.29, 0.49]

Against CoPRA, the Δ interval excludes zero on every subset and for both correlation measures, including the most stringent subset D (Δ SCC 0.26 [0.04, 0.47], $n = 97$). Against DeePNAP, the advantage is resolved on the full comparison set (Δ SCC 0.29 [0.18, 0.39], $n = 287$) but not on subsets B–D, where 23–48 pairs remain and every interval spans zero. ZeroFold attains the higher point estimate throughout, but we do not claim a statistically resolved advantage over DeePNAP under the stricter identity thresholds; establishing one would require a larger set of test pairs whose proteins are absent from the ProNAB reference.

C Dataset Details

To train and evaluate ZeroFold, we assembled PRADB, a curated dataset of 2,582 unique protein-RNA pairs with experimentally measured binding affinities (pK_D) drawn from four complementary sources: ProNAB [Harini et al., 2022], BioLiP2 [Zhang et al., 2024], the UTexas Aptamer Database [Askari et al., 2024], and PDBBind v2020 [Liu et al., 2017] (Figure A.1A). ProNAB constitutes the

largest single contribution, reflecting its broad coverage of RNA-binding protein families, while the remaining three databases provide complementary coverage of structurally characterised complexes and aptamer interactions. The per-source counts given below precede cross-source deduplication; the breakdown of the 2,582 unique pairs is shown in Figure A.1A.

ProNAB. ProNAB (academic license) is a manually curated database of over 20,000 experimentally measured protein–nucleic acid binding affinities, including 5,323 protein-RNA complexes [Harini et al., 2022]. After retaining entries with valid K_D measurements and filtering for natural amino acids and unmodified nucleotide bases, we obtained 3,588 protein-RNA pairs. Replicate measurements for identical sequence pairs often differed substantially (Appendix Figure A.1E), likely reflecting assay- and condition-dependent experimental variability. Because no single standard condition could be defined consistently across the database, all valid measurements were retained.

BioLiP2. The next largest source of affinity data was BioLiP2 [Zhang et al., 2024] (BSD license). This database contains biologically relevant protein-ligand interactions derived from the Protein Data Bank (PDB), annotated with binding affinity measurements collected from external databases and manual literature curation. BioLiP2 contains 425 protein-RNA complexes with reported binding affinity data.

UTexas Aptamer Database. The aptamer database curated by the University of Texas compiles binding affinity measurements reported in the literature between 1990 and 2022 [Askari et al., 2024]. Within this resource, 258 protein-RNA complexes were associated with labelled binding affinity values.

PDBBind v2020 . The final dataset used was PDBBind v2020 (under the PDBBind v2020 Data Use Agreement), which was considered the most reliable source of affinity data due to its high level of curation and the availability of experimentally determined 3D structural information alongside binding affinity annotations [Liu et al., 2017]. This dataset contains 239 protein-RNA complexes with reported binding affinities.

PRADB. The union of the four previously described databases was taken to produce a redundant set of 3,939 protein-RNA pairs. For pairs associated with multiple affinity measurements, the value reported in PDBBind v2020 was preferentially selected due to its higher level of curation; if no PDBBind v2020 measurement was available, the median affinity across the remaining sources was used. After resolving these redundancies, the final Protein-RNA Affinity DataBase (PRADB) comprised 2,582 unique protein-RNA sequence pairs. The distributions of protein sequence lengths, RNA sequence lengths, and affinity values are shown in Appendix Figure A.1B–D.

The distribution of experimentally measured affinities (pK_D) is approximately unimodal and centred around 7.1, with a standard deviation of 1.38, reflecting a dataset that is neither dominated by very low-affinity nor very high-affinity binders and spans a range well suited to training a regression model (Figure A.1D).

Clustering and sample weighting. To prevent data leakage, protein and nucleic acid sequences were clustered independently with MMseqs2 `easy-cluster` [Kallenborn et al., 2025] at 40% sequence identity and 80% coverage, and train/validation/test splits were assigned at the cluster level. We audited the splits with bidirectional MMseqs2 `easy-search`, removing 23 validation/test entries above the 40% threshold to training sequences, and removed 20 training pairs with RNA longer than 1,500 nt, yielding 2,056 training, 209 validation, and 297 test pairs ($\approx 8:1:1$). Because both tools rely on k -mer seeding, which is insensitive for the short RNAs in this dataset (median 22 nt) and causes them to be split into predominantly singleton clusters, we additionally verified the splits by exhaustive Smith–Waterman alignment of every evaluation sequence against every training sequence; no validation or test protein or RNA exceeds 40% local identity at 80% mutual coverage to any training sequence.

Following AlphaFold-Multimer [Evans et al., 2021], training samples were weighted inversely to the product of their protein and RNA cluster sizes (Appendix C), down-weighting densely sampled families and up-weighting rare pairs:

$$w_{ij} = \frac{1}{|PC_i|} \cdot \frac{1}{|RC_j|} \quad (1)$$

where PC_i is the protein cluster of pair i and RC_j is the RNA cluster of pair j . Weights were normalised to sum to unity across the training set. This down-weights complexes from densely

sampled families (such as ribosomal proteins paired with rRNA) and up-weights unique sequence pairs, encouraging the model to learn from the full diversity of the dataset rather than fitting dominant clusters.

D ZeroFold Model

ZeroFold comprises four main components: the RNA encoder, the protein encoder, the pair-biased self-attention blocks, and the affinity prediction head.

RNA/Protein input representations. ZeroFold employs Boltz-2 [Passaro et al., 2025] to extract representations for both protein and nucleic acid chains. These representations were extracted from the final layer of the Boltz-2 trunk (`recycling_steps` = 3), just before the structure module decoding. The protein and nucleic acid representations each consist of two components. First, we extracted the single sequence representation $\{s_i^{\text{single}}\}$, where $s_i^{\text{single}} \in \mathbb{R}^{c_{\text{single}}}$, $c_{\text{single}} = 384$, and $i \in \{1 \dots N_{\text{res}}\}$, which is the per-residue (or per-nucleotide) embedding from the final trunk layer. Second, we extracted the pair representations $\{z_{ij}^{\text{pair}}\}$, where $z_{ij}^{\text{pair}} \in \mathbb{R}^{c_{\text{pair}}}$, $c_{\text{pair}} = 128$ and $i, j \in \{1 \dots N_{\text{res}}\}$, which are the pairwise embeddings encoding residue–residue (or nucleotide–nucleotide) spatial and evolutionary relationships.

RNA-protein encoder. Each encoder maps its Boltz-2 trunk representations into refined single and pair feature streams. The single representation s_i^{single} and pair representation z_{ij}^{pair} are each passed through two successive transition layers with dropout and residual connections (omitted where a layer changes width). Here, a transition layer denotes an AlphaFold-style position-wise two-layer MLP: layer normalisation, four-fold channel expansion, ReLU, and projection to the model’s internal width ($c_s = 384$, $c_z = 128$). The RNA encoder is identical except that a learned 32-dimensional nucleic acid type embedding (RNA vs. DNA) is concatenated to its single representation beforehand, allowing it to condition on nucleic acid type. Full dimensions and hyperparameters are given in Appendix E.

Pair-biased self-attention module and affinity head Encoded protein and RNA single representations are concatenated into a joint token sequence, so that self-attention operates across both modalities, while the pair representations are combined into a block-diagonal tensor with the off-diagonal blocks initialised to zero. The joint streams are refined by stacked pair-biased self-attention blocks [Zhang et al., 2025]: within each block, the pair representation is projected to a per-head bias added to the attention logits, the attention output updates the single stream, and the logits are projected through a two-layer MLP to update the pair stream. Each block is followed by transition layers, with residual connections and dropout throughout. The affinity head then applies masked mean pooling separately over the final protein and RNA single-token representations, concatenates the two pooled vectors, and passes them through a layer-normalised MLP with residual blocks to output a scalar pK_{D} prediction.

Training and evaluation. A Bayesian hyperparameter sweep [Snoek et al., 2012] with Hyperband early termination [Li et al., 2018] tuned the model using the PRADB training/validation split, optimising validation SCC. The search space included model depth, number of attention heads, hidden dimensions, affinity-head depth/width, dropout rates, maximum learning rate, cosine learning-rate schedule parameters, weight decay, and gradient clipping.

The final model was trained for 100 epochs, with each epoch comprising 900 samples from the training set weighted by cluster representation (see above). All ZeroFold variants and downstream experiments were trained on a single NVIDIA GH200 120GB GPU using the Adam optimizer [Kingma and Ba, 2014].

Ablations. The sequence-LM representation used ESM-2 t33 650M UR50D protein embeddings [Lin et al., 2023] and RiNALMo giga v1 RNA embeddings [Penić et al., 2025], using `sequence_lm_s_only` features with `c_single_in` = 1280 for both encoders. These were used as the single input s to the encoder, with the pair input z set to a zero tensor of the model’s internal pair dimension.

To generate the distograms, we ran the full Boltz-2 pipeline including diffusion and, for each of the five predicted structures, extracted the C_{α} atom (backbone offset 1) of every protein residue and the C'_4 atom (backbone offset 5) of every nucleotide. Intra-chain pairwise distances between these

representative atoms were computed from the predicted coordinates, discretised into 64 linear bins spanning 0–50 Å with distances beyond the range assigned to the final bin, and the resulting per-structure one-hot encodings were averaged over the five samples to give a soft distogram $[L, L, 64]$ as the pair input z . We use a wider range than the 2–22 Å employed by the AlphaFold2 and Boltz-2 distogram heads because those are calibrated for interface-proximity prediction; applied to full-chain matrices, whose median separation is 30 Å, that range would place most pairs in the overflow bin. For the combined conditions, the soft distogram was concatenated with the Boltz-2 pair representation along the feature axis ($c_{\text{pair},\text{in}} = 192$). The single input s was the Boltz-2 per-residue embedding of dimension 384.

Ablations were run using the same training, validation and test splits as ZeroFold. The same Bayesian hyperparameter sweep with Hyperband early termination was run for each ablation (40 trials, except 12 trials for the single-only condition G, reduced for computational reasons), and the hyperparameters achieving the best validation SCC were selected. Models were then trained on five random seeds using these fixed hyperparameters, and the top three seeds by validation SCC were evaluated on the test set.

All ZeroFold variants and downstream experiments were trained on a single NVIDIA GH200 120GB GPU using the Adam optimizer [Kingma and Ba, 2014].

E Hyperparameters

Table A.3 gives the final ZeroFold configuration, selected by a 40-trial Bayesian sweep with Hyperband early termination optimising validation SCC.

Table A.3: **ZeroFold hyperparameters** (condition A). Architecture and optimiser settings selected by validation SCC.

Parameter	Value
<i>Architecture</i>	
Cross-modal blocks (no_blocks)	8
Attention heads (no_heads)	32
Attention hidden dimension (c_hidden)	128
Single representation width c_s	384
Pair representation width c_z	128
Affinity-head blocks (aff_no_blocks)	6
Affinity-head width (aff_c_hidden)	384
<i>Regularisation</i>	
Encoder and trunk dropout	0.0886
Affinity-head dropout	0.2276
Weight decay	5.02×10^{-7}
Gradient clipping	1.232
<i>Optimisation</i>	
Optimiser	Adam
Maximum learning rate	5.619×10^{-4}
Cosine schedule T_0	9,470
Cosine η_{min} ratio	0.0645
Batch size	1
Maximum epochs	100
Batches per epoch	900
Early-stopping patience	25
Random seed	2,823,573

Each ablation received its own independent sweep over the same search space (40 trials; 12 for condition G, reduced for computational reasons); the selected settings differ from ZeroFold and from each other (Table A.4).

Checkpoint selection. Models were trained for up to 100 epochs with early stopping (patience 25) on validation Spearman correlation. The checkpoint achieving the highest validation Spearman was the one evaluated on the held-out test set.

Table A.4: **Selected hyperparameters for the joint-encoder and single-only ablations.** Remaining conditions used the same search space; all sweep configurations are released with the code.

	Blocks	Heads	c_hidden	Max LR	Dropout
C — Joint encoder	4	16	128	9.25×10^{-4}	0.264
D — Joint + distogram	4	16	64	8.62×10^{-4}	0.058
G — PS single only	6	32	64	3.33×10^{-4}	0.286