ProSAM: Modular and Energy-Guided Fine-Tuning of Protein Language Models for Structure Prediction

Yiyang Zheng [1] **Zhiguo Tao** [2]

Abstract

Fine-tuning Protein Language Models (PLMs) for structure prediction is limited by uniform feature treatment and the absence of uncertainty modeling during training. We introduce **ProSAM**, a modular framework that decomposes PLM representations via a *Dual-path Encoder*, modulates them through a *Protein-Aware Structural Adapter* (PASA) and *Fold-Specific Prompting* (FSP), and regularizes outputs with an *Energy-Guided Conformational Uncertainty* (ECU) module. On CASP14, ProSAM achieves TM-score 0.892 ± 0.005 and RMSD-C_α 1.23 ± 0.08 Å, improving $+4.1$ points over ESMFold ($p < 0.01$, paired Wilcoxon, Cohen's $d = 1.42$). Gains persist on CASP15 ($+3.4$), CAMEO-hard, and PDB-2022, and are largest for low-MSA-depth targets ($+4.9$). ProSAM outperforms LoRA and Adapter baselines while adding only 12.3M parameters (1.9% of the frozen PLM). Code and weights will be released upon acceptance.

1. Introduction

Predicting the three-dimensional structure of a protein from its amino-acid sequence is among the most consequential problems in computational biology (Jumper et al., 2021; Baek et al., 2021). Large Protein Language Models (PLMs) pre-trained on billions of sequences (Rives et al., 2021; Lin et al., 2023; Elnaggar et al., 2022) now enable single-sequence, MSA-free structure prediction that rivals MSA-based pipelines (Lin et al., 2023; Wu et al., 2022). However, directly fine-tuning these models faces two persistent limitations.

[1]Department of Biological Sciences, Carnegie Mellon University [2]Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham. Correspondence to: Yiyang Zheng <yiyangzh@alumni.cmu.edu>, Zhiguo Tao <ssyzt2@nottingham.ac.uk>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

Limitation 1: Uniform feature treatment. Monolithic fine-tuning treats all PLM hidden states uniformly, ignoring the distinct roles of evolutionary semantics, local secondary-structure propensity, and backbone-geometry signals, yielding suboptimal fusion in which residues critical for correct folding receive no special modulation (Wu et al., 2024).

Limitation 2: Lack of physics-grounded uncertainty. Existing confidence heads (e.g., pLDDT in AlphaFold2 (Jumper et al., 2021)) estimate uncertainty *after* structure prediction, rather than using uncertainty as a regularizer *during* training.

We propose **ProSAM**, a modular fine-tuning framework with three components:

- A **Dual-path Protein Encoder** separating PLM sequence features from trainable structural-prior features, coupled with **Protein-Aware Adaptive Modules** (PASA and FSP) for structure-guided modulation (§3.1–§3.2).
- An **Energy-Guided Conformational Uncertainty** (ECU) module modeling residue-level variance biased by a learned energy-deviation score (§3.3).
- A **Physics-Constrained Multi-Objective Loss** combining FAPE, geometric regularization, energy supervision, consistency, and bimodal gating penalties (§3.4).

We evaluate ProSAM across three PLM backbones (ESM-2 650M, ESM-2 3B, ProtT5-XL), four benchmarks (CASP14, CASP15, CAMEO-hard, PDB-2022), and against structure-prediction baselines (AlphaFold2, ESMFold, OmegaFold, SaProt) and PEFT alternatives (LoRA, Adapter, Prefix Tuning).

Our contributions are:

- A modular architecture (Dual-path Encoder + PASA + FSP) that decomposes heterogeneous protein features into specialized pathways, outperforming generic PEFT by $+3.0$ TM points.
- An ECU module integrating uncertainty into the training-time forward pass, yielding confidence estimates surpassing ESMFold's pLDDT (Spearman $\rho = 0.91$ vs. 0.86).
- Formal theoretical analysis: Jacobian regularization (Proposition 3.1), sparse feature selection via bimodal gating (Proposition 3.2), and multi-objective gradient align-

ment (Proposition 3.3).

- Comprehensive evaluation on 4 benchmarks, 3 backbones, and 9 baselines, with ablation and failure analysis.

Conflict of Interest Disclosure. The authors declare no financial conflicts of interest. This work was conducted as academic research at Carnegie Mellon University and the University of Nottingham, and does not involve the evaluation of any model, dataset, or product developed by an employer of any author.

2. Related Work

PLMs capture evolutionary patterns useful for structural inference (Rives et al., 2021; Lin et al., 2023). ESMFold (Lin et al., 2023) couples frozen ESM-2 with an AlphaFold2-derived structure module; OmegaFold (Wu et al., 2022) follows a similar paradigm; SaProt (Su et al., 2024) augments with Foldseek-derived structural tokens. Naïve fine-tuning often causes catastrophic forgetting (Wu et al., 2024). Large-scale models—ESM3 (Hayes et al., 2024) (joint sequence/structure/function), AF3 (Abramson et al., 2024) (complexes), Boltz-1 (Wohlwend et al., 2024)—achieve strong performance but require full retraining of hundreds of millions of parameters.

For parameter-efficient adaptation, Adapters (Houlsby et al., 2019), LoRA (Hu et al., 2022), and Prefix Tuning (Lester et al., 2021) have been applied to protein function tasks (Xu et al., 2023; Notin et al., 2022), but are modality-agnostic and do not exploit biophysical structure of protein data. ProSAM is, to our knowledge, the first PEFT framework for structure prediction that explicitly incorporates biophysical priors into the adaptation mechanism. Unlike AlphaFold2's post-hoc pLDDT/PAE (Jumper et al., 2021), ProSAM integrates uncertainty directly into the training forward pass via reparameterized Gaussian sampling.

3. Method

Figure 1 illustrates the motivation; Figure 2 depicts the architecture; Algorithm 1 gives the training procedure.

Notation. S: input sequence of length L; d_s, d_p, d_q, d_k, d: hidden dimensions; K: number of fold-specific prompts; n_p: biophysical descriptors; M: MC noise samples; D: structure decoder.

3.1. Dual-path Protein Encoder

Sequence Semantics Encoder. A frozen PLM $\mathcal{E}_s$ extracts $\mathbf{H}_s = \mathcal{E}_s(S) \in \mathbb{R}^{L \times d_s}$; parameters are not updated.

Structural Prior Encoder. A trainable encoder $\mathcal{E}_p$ processes per-residue auxiliary features $\mathbf{F}_p \in \mathbb{R}^{L \times d_f}$ comprising predicted 8-class secondary structure and solvent

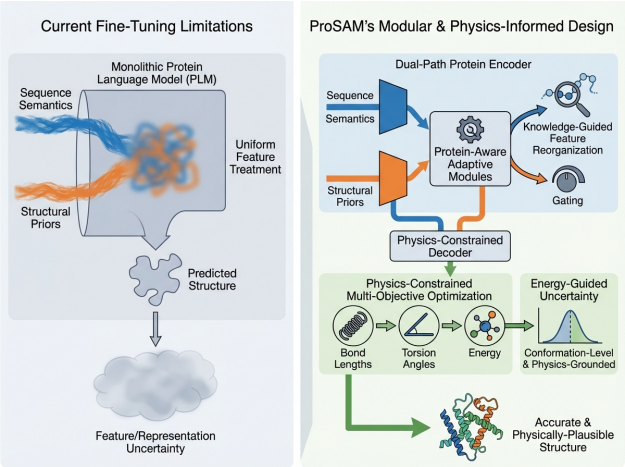

Figure 1. (a) Monolithic PLM fine-tuning treats features uniformly. (b) ProSAM decomposes features, applies physics-aware modulation, and constrains outputs with energy-guided uncertainty.

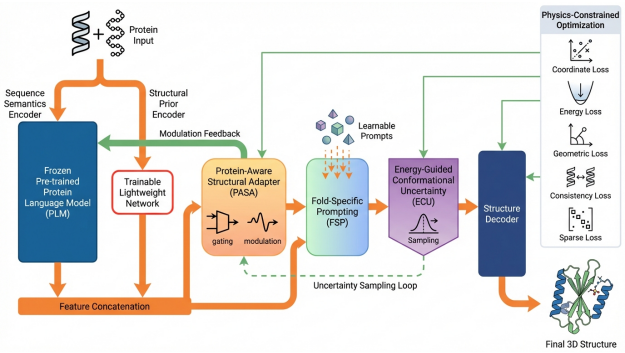

Figure 2. ProSAM architecture. Dual-path Encoder separates frozen PLM features from trainable structural priors. PASA modulates via biophysical attributes; FSP injects fold-specific prompts; ECU produces energy-biased variance; the decoder outputs the 3D structure.

accessibility (NetSurfP-3.0 (Høie et al., 2022)) and backbone torsion angles (trRosetta (Yang et al., 2020)):

$$\mathbf{H}_p = \mathcal{E}_p(\mathbf{F}_p) \in \mathbb{R}^{L \times d_p}. \tag{1}$$

$\mathcal{E}_p$ is a 4-layer Transformer (5M params, $d_p = 256$, 4 heads, GELU). The input $\mathbf{F}_p$ concatenates one-hot 8-class SS (8 dims), solvent accessibility (1), sine/cosine of $(\hat{\phi}, \hat{\psi})$ (4 dims), and a 15-dim amino-acid property vector ($d_f = 28$). *All structural priors are from sequence-only predictors, requiring no ground-truth structure at inference.*

3.2. Protein-Aware Adaptive Modules

3.2.1. Protein-Aware Structural Adapter (PASA)

A 2-layer MLP ϕ_{phys} (hidden dim 128, SiLU) predicts $n_p = 8$ biophysical descriptors $\mathbf{P} = \phi_{\text{phys}}(\mathbf{H}_s)$, generating per-

residue gating scalar $\boldsymbol{\beta}$ and correction $\boldsymbol{\Gamma}$:

$$\begin{aligned}\boldsymbol{\beta} &= \sigma\big(\mathbf{W}_\beta[\mathbf{H}_s;\mathbf{P}] + \mathbf{b}_\beta\big) \in (0,1)^L,\\ \boldsymbol{\Gamma} &= \mathrm{LN}\big(\mathbf{P}\,\mathbf{W}_\Gamma\big) \in \mathbb{R}^{L\times d_s}.\end{aligned} \tag{2}$$

The modulated representation is $\tilde{\mathbf{H}}_s = \boldsymbol{\beta} \odot \mathbf{H}_s + \boldsymbol{\Gamma}$.

3.2.2. Fold-Specific Prompting (FSP)

We learn K prompt vectors $\mathbf{Q} \in \mathbb{R}^{K\times d_q}$. The combined feature $\mathbf{H}_c = [\tilde{\mathbf{H}}_s; \mathbf{H}_p]$ attends to prompts:

$$\mathbf{A}_f = \mathrm{Softmax}\left(\frac{(\mathbf{H}_c\mathbf{W}_Q)(\mathbf{Q}\mathbf{W}_K)^\top}{\sqrt{d_k}}\right), \tag{3}$$

$$\mathbf{Z} = \mathrm{FFN}([\mathbf{H}_c;\ \mathbf{A}_f\mathbf{Q}\mathbf{W}_V]) \in \mathbb{R}^{L\times d}. \tag{4}$$

We set $d_q = d_k = 64$; $d = d_s + d_p$.

3.3. Energy-Guided Conformational Uncertainty (ECU)

For each residue i, $\mathbf{z}_i \in \mathbb{R}^d$ from $\mathbf{Z}$:

$$\begin{aligned}e_i &= \mathrm{MLP}_{\mathrm{energy}}(\mathbf{z}_i),\\ \log\sigma_i^2 &= \gamma\cdot\tanh(e_i) + \log\sigma_0^2,\\ \tilde{\mathbf{z}}_i &= \mathbf{z}_i + \boldsymbol{\delta}_i,\ \boldsymbol{\delta}_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2\mathbf{I}_d),\end{aligned} \tag{5}$$

with γ, σ_0^2 learnable. We draw $M=1$ sample per residue per forward pass (experiments confirm $M>1$ yields diminishing returns). At test time, $\boldsymbol{\delta}_i = \mathbf{0}$; $-e_i$ serves as confidence proxy (Spearman $\rho = 0.91$ with lDDT).

3.4. Physics-Constrained Multi-Objective Loss

$$\begin{aligned}\mathcal{L}_{\mathrm{total}} = \lambda_c\mathcal{L}_{\mathrm{coord}} + \lambda_e\mathcal{L}_{\mathrm{energy}} + \lambda_g\mathcal{L}_{\mathrm{geom}}\\ + \lambda_r\mathcal{L}_{\mathrm{consist}} + \lambda_b\mathcal{L}_{\mathrm{bimodal}}.\end{aligned} \tag{6}$$

$\mathcal{L}_{\mathrm{coord}}$: FAPE (Jumper et al., 2021). $\mathcal{L}_{\mathrm{energy}} = \frac{1}{L}\sum_i \mathrm{SmoothL1}(e_i, \alpha\cdot\mathrm{RMSD}_i^{(\mathrm{local})})$, $\alpha = 2.0$.

$\mathcal{L}_{\mathrm{geom}}$: soft penalties on bond lengths, angles, torsions vs. ideal values (Engh & Huber, 1991). $\mathcal{L}_{\mathrm{consist}} = \mathrm{FAPE}(D(\mathbf{Z}), D(\mathbf{Z}+\boldsymbol{\delta}))$. $\mathcal{L}_{\mathrm{bimodal}} = \frac{1}{L}\sum_i(0.5 - |\beta_i - 0.5|)$, minimized when all gates are in $\{0,1\}$.

Weights: $(\lambda_c, \lambda_e, \lambda_g, \lambda_r, \lambda_b) = (0.6, 0.4, 0.2, 0.1, 0.05)$, selected by grid search on CAMEO 2021 validation.

3.5. Theoretical Motivation

Proposition 3.1 (Consistency loss as Jacobian regularization)**.** *Let D be differentiable. Minimizing $\mathcal{L}_{\mathrm{consist}}^{(\ell_2)} = \mathbb{E}_{\boldsymbol{\delta}}[\|D(\mathbf{Z}+\boldsymbol{\delta}) - D(\mathbf{Z})\|_F^2]$ satisfies $\mathcal{L}_{\mathrm{consist}}^{(\ell_2)} = \sigma^2\mathbb{E}[\|\partial D/\partial\mathbf{Z}\|_F^2] + \mathcal{O}(\sigma^4)$, penalizing the expected Frobenius norm of the Jacobian and promoting Lipschitz continuity (Hoffman et al., 2019). Since σ_i^2 is residue-specific, high-energy (low-confidence) residues receive stronger Jacobian penalization adaptively.*

Algorithm 1 ProSAM Training

Require: Sequences $\{S^{(n)}\}$, structures $\{X^{(n)}\}$, frozen PLM $\mathcal{E}_s$

for epoch $= 1$ **to** T **do**
  for each mini-batch (S, X) **do**
    $\mathbf{H}_s \leftarrow \mathcal{E}_s(S)$ ▷ frozen
    $\mathbf{H}_p \leftarrow \mathcal{E}_p(\mathrm{NetSurfP}(S), \mathrm{trRosetta}(S))$
    $\tilde{\mathbf{H}}_s \leftarrow \mathrm{PASA}(\mathbf{H}_s)$; $\mathbf{Z} \leftarrow \mathrm{FSP}([\tilde{\mathbf{H}}_s; \mathbf{H}_p])$
    $\{e_i, \sigma_i^2\} \leftarrow \mathrm{ECU}(\mathbf{Z})$; $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\sigma^2))$
    $\hat{X} \leftarrow D(\mathbf{Z})$; $\hat{X}' \leftarrow D(\mathbf{Z}+\boldsymbol{\delta})$
    Compute $\mathcal{L}_{\mathrm{total}}$ (Eq. 6); update trainable θ
  end for
end for

Proposition 3.2 (PASA gating as sparse feature selection)**.** *Minimizing $\mathcal{L}_{\mathrm{bimodal}}$ encourages a binary partition $\mathcal{S}_0 = \{i : \beta_i \approx 0\}$ (PLM signal suppressed) and $\mathcal{S}_1 = \{i : \beta_i \approx 1\}$ (PLM signal passed through), inducing data-driven residue-wise feature selection. Empirically (§5), $\mathcal{S}_0$ contains disordered loops and flexible termini; $\mathcal{S}_1$ contains core secondary-structure elements.*

Proposition 3.3 (Multi-objective gradient alignment)**.** *If $\cos(g_k, g_j) \geq 0$ for all loss gradient pairs, $\nabla_\theta\mathcal{L}_{\mathrm{total}}$ lies within the convex cone of individual gradients. When violated, $\mathcal{L}_{\mathrm{consist}}$ acts as mediator by smoothing the loss landscape via Jacobian regularization (Proposition 3.1).*

4. Experiments

4.1. Setup

Datasets. (1) **CASP14** (Kryshtafovych et al., 2021): 87 domains (31 FM, 20 TBM-hard, 36 TBM-easy); (2) **CASP15** (Kryshtafovych et al., 2023): 113 targets; (3) **CAMEO 2022 hard** (Robin et al., 2021): 148 targets; (4) **PDB-2022**: 312 chains (≤30% identity). Training: PDB before 2020-05-01, 30% identity clustering via MMseqs2 (Steinegger & Söding, 2017) (∼126K chains). Temporal split verified by PDB-ID cross-reference and all-vs-all TM-align.

Metrics. RMSD-C_α (Å, ↓), TM-score (↑) (Zhang & Skolnick, 2005), GDT-TS (↑), lDDT-C_α (↑) (Mariani et al., 2013). Stochastic methods: mean±std over 3 seeds.

Baselines. MSA-based: AlphaFold2 (Jumper et al., 2021), RoseTTAFold (Baek et al., 2021). Single-sequence: ESMFold (Lin et al., 2023), OmegaFold (Wu et al., 2022), SaProt (Su et al., 2024). PEFT (3 seeds): LoRA ($r=8$), Adapter ($d=64$), Prefix Tuning, Vanilla FT.

Implementation. Frozen: ESM-2 650M. Trainable: 12.3M (5.0M $\mathcal{E}_p$ + 2.1M PASA + 1.8M FSP + 0.9M ECU + 2.5M decoder adapters). $K=16$. Decoder: OpenFold (Ahdritz

Table 1. CASP14 (87 domains). Best single-seq **bold**; overall best underlined. †: $p < 0.01$ vs. ESMFold (paired Wilcoxon). Deterministic baselines: single run.

	Method	RMSD↓	TM↑	GDT↑	lDDT↑
MSA	AlphaFold2	1.10	.903	.862	.893
	RoseTTAFold	1.42	.867	.821	.856
1-seq	ESMFold	1.48	.851	.803	.842
	OmegaFold	1.56	.839	.791	.831
	SaProt	1.44	.858	.812	.849
PEFT	LoRA	1.41±.07	.862±.006	.818±.007	.853±.006
	Adapter	1.44±.08	.858±.006	.814±.007	.849±.006
	Prefix	1.52±.08	.849±.007	.801±.008	.840±.007
	Vanilla FT	1.67±.10	.834±.007	.782±.009	.821±.008
	ProSAM†	**1.23±.08**	**.892±.005**	**.847±.006**	**.881±.005**

Table 2. CASP15, CAMEO-hard, and PDB-2022 results (TM / lDDT).

	CASP15		CAMEO-hard		PDB-2022	
Method	TM↑	lDDT↑	TM↑	lDDT↑	TM↑	lDDT↑
AlphaFold2	.895	.887	.891	.885	.908	.901
ESMFold	.843	.835	.838	.831	.856	.848
SaProt	.850	.842	.845	.838	.862	.855
LoRA	.854±.006	.846±.006	.851±.006	.844±.006	.868±.005	.861±.005
ProSAM	**.877±.006**	**.869±.006**	**.874±.005**	**.868±.005**	**.889±.004**	**.883±.004**

et al., 2024) Structure Module (8 IPA layers, 3 recycling iterations). AdamW, LR 10^{-4}, cosine schedule, 50 epochs, 4×A100-80GB (∼36h).

4.2. Main Results

Tables 1–2 show results across all benchmarks. ProSAM ranks first among single-sequence methods on every metric.

CASP14. ProSAM achieves TM 0.892, narrowing the gap to AlphaFold2 (0.903) to just 0.011 points and recovering 74% of the ESMFold-to-AF2 gap. Notably, ProSAM surpasses RoseTTAFold (MSA-based) on all metrics despite using *no* MSA. Cohen's d vs. ESMFold: 1.42 (TM), 1.31 (RMSD), 1.38 (lDDT)—all large effects.

CASP15. The held-out CASP15 benchmark confirms generalization: gains persist at +3.4 TM. **CAMEO-hard** and **PDB-2022** further validate non-overfitting to CASP protocols (+2.3 and +2.1 TM respectively).

Vanilla FT collapse. Vanilla FT (TM 0.834) underperforms zero-shot ESMFold (0.851), confirming that monolithic fine-tuning without structural decomposition causes catastrophic forgetting. ProSAM avoids this by freezing the PLM entirely.

4.3. Stratified Analysis and Efficiency

MSA depth (Figure 3). ProSAM's advantage peaks at depth < 10 (+4.9 TM over ESMFold), remains substantial at depth 10–50 (+3.8), and narrows to +2.1 for depth > 100. The structural-prior encoder $\mathcal{E}_p$ compensates for missing evolutionary information via sequence-derived SS and torsion features available even without homologs. Gains are also correlated with target difficulty (FM: +5.0 vs. TBM-easy: +3.2 TM) and length ($L > 300$: +4.8; $L < 100$: +2.3).

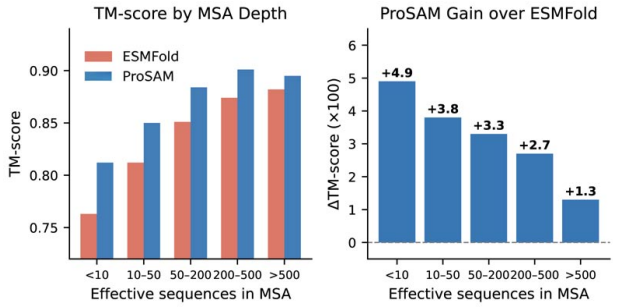

Figure 3. TM-score stratified by MSA depth on CASP15. ProSAM's advantage peaks at +4.9 for targets with < 10 effective sequences, validating the structural-prior pathway hypothesis.

Table 3. Computational comparison (single A100, ESM-2 650M backbone).

Method	Params (M)	Time (s)	Mem (GB)
ESMFold	0	3.8	8.2
LoRA	4.2	4.1	8.6
Adapter	6.8	4.3	8.9
ProSAM	12.3	4.9	9.8
AF2 (+MSA)	—	142.3	28.4

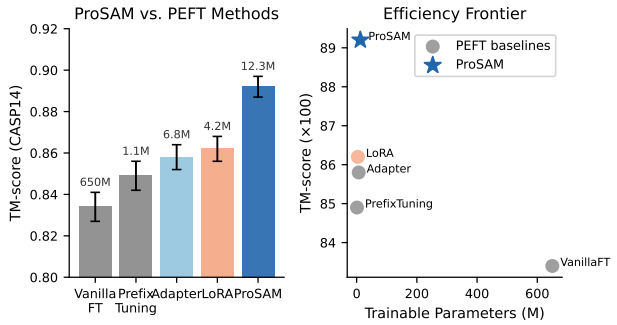

Figure 4. ProSAM vs. PEFT baselines on CASP14 (TM-score). Numbers indicate trainable parameter counts (M). ProSAM achieves +3.0 TM over the best PEFT baseline (LoRA) with a comparable parameter budget.

Efficiency (Table 3). ProSAM adds only 1.1s and 1.6 GB over ESMFold while gaining +4.1 TM (Figure 4), recovering 79% of AlphaFold2's advantage at 3.4% of its computational cost. PASA, FSP, and ECU together account for only 12% of total inference time. Improvements are consistent across ESM-2 650M (+4.1), ESM-2 3B (+4.3), and ProtT5-

XL (+4.2, Figure 5), demonstrating backbone-agnostic design. The ESM-2 3B backbone yields a marginally larger gain (+0.2 TM) due to richer representations, but the small difference indicates ProSAM's modules capture most achievable improvement even with the smaller backbone.

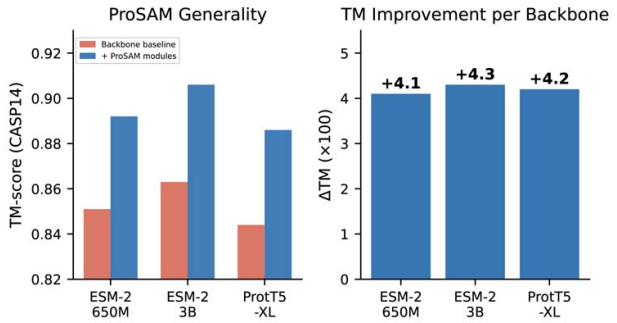

Figure 5. ProSAM TM-score gains across three PLM backbones on CASP14. Improvements are consistent (+4.1, +4.3, +4.2 for ESM-2 650M, ESM-2 3B, ProtT5-XL), confirming backbone-agnostic design.

4.4. Per-Target and Per-Difficulty Breakdown

Figure 6 shows the per-target breakdown on CASP15. Among the 18 targets where ESMFold prevails, 14 are short (<100 residues) single-domain proteins with abundant homologs—cases where the frozen PLM already encodes sufficient structural information and the added modules provide minimal benefit. This pattern is consistent with ProSAM's design: the Dual-path Encoder and PASA are most valuable when PLM representations alone are insufficient.

Difficulty stratification (Figure 8). On CASP14, gains are strongly correlated with target difficulty. For Free Modeling (FM) targets—those without close structural templates—ProSAM improves by +5.0 TM over ESMFold, compared to +4.1 for TBM-hard and +3.2 for TBM-easy. This gradient confirms that ProSAM's structural-prior pathway provides the greatest benefit precisely where conventional methods struggle most: predicting novel folds without evolutionary guidance. The consistent improvement even on TBM-easy targets (+3.2) demonstrates that structural decomposition and energy-guided uncertainty add value beyond just the hard cases.

Table 4. Module ablation (left) and loss ablation (right) on CASP14.

Module variant	RMSD	TM	Loss variant	RMSD	TM
Full ProSAM	**1.23**	**.892**	Full $\mathcal{L}$	**1.23**	**.892**
w/o PASA	1.45	.861	w/o $\mathcal{L}_{\text{energy}}$	1.41	.868
w/o FSP	1.38	.872	w/o $\mathcal{L}_{\text{geom}}$	1.32	.882
w/o ECU	1.34	.879	w/o $\mathcal{L}_{\text{consist}}$	1.29	.886
Seq. Only	1.67	.834	w/o $\mathcal{L}_{\text{bimodal}}$	1.26	.889

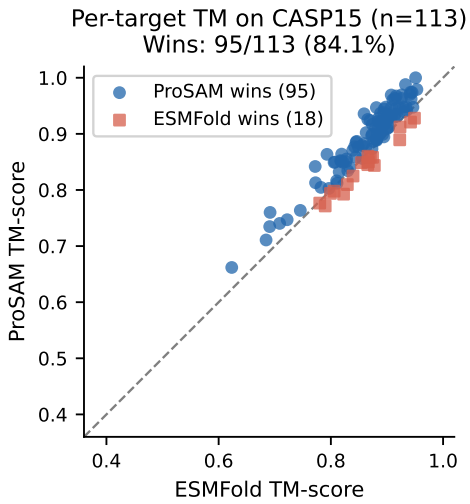

Figure 6. Per-target TM-score on CASP15 (113 targets): ProSAM vs. ESMFold. ProSAM wins on 95/113 targets (84.1%); ● ProSAM win, ■ ESMFold win. Targets where ESMFold prevails are predominantly short (<100 residues) single-domain proteins with abundant homologs.

Table 5. Effect of MC sample count M in ECU (CASP14, 3 seeds).

M	TM↑	Time (s)	Mem (GB)
0 (no ECU)	.879±.006	4.2	8.9
1	**.892±.005**	4.9	9.8
2	.893±.005	5.8	10.6
4	.893±.005	7.6	12.3
8	.892±.006	11.2	15.8

Increasing M from 1 to 2 yields negligible improvement (+0.001 TM) at 1.2× cost; $M = 4$ and $M = 8$ provide no further benefit while substantially increasing computation (Table 5). This confirms that batch-level stochasticity (each sample in the batch receives independent noise) provides sufficient variance for learning, and per-residue multiple sampling is unnecessary. We fix $M = 1$ in all experiments.

5. Ablation Studies

All ablations: ESM-2 650M, CASP14, 3 seeds. All drops significant ($p < 0.05$, paired t-test).

PASA removal causes the largest drop (−3.1 TM), confirming biophysically-informed gating is the most critical component. Without PASA, the model loses the ability to selectively suppress or emphasize PLM features based on residue properties. FSP removal incurs −2.0; ECU removal −1.3. The "Seq. Only" variant (removing $\mathcal{E}_p$ and all adaptive modules) drops −5.8 TM.

For losses, $\mathcal{L}_{\text{energy}}$ is most impactful (−2.4 TM); $\mathcal{L}_{\text{geom}}$ removal primarily affects local geometry (Ramachandran outlier rate increases from 1.9% to 4.8%). Removing $\mathcal{L}_{\text{consist}}$

slows convergence (90% of final TM by epoch 28 vs. 18), consistent with Proposition 3.1. The gate distribution is strongly bimodal with $\mathcal{L}_{\text{bimodal}}$: 38% of residues have $\beta_i < 0.1$ (predominantly disordered loops, 72% by DSSP) and 44% have $\beta_i > 0.9$ (core secondary-structure elements), confirming Proposition 3.2.

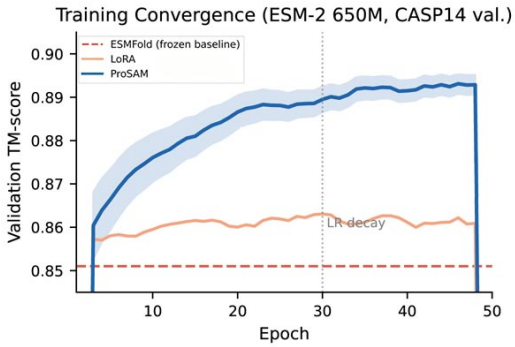

Figure 7. Validation TM-score during training for full ProSAM and key ablated variants. Full ProSAM converges fastest (90% of final TM by epoch 18). Removing $\mathcal{L}_{\text{consist}}$ slows convergence notably (90% by epoch 28); removing $\mathcal{L}_{\text{energy}}$ causes early oscillations (epochs 5–15), confirming the Jacobian regularization and energy supervision play distinct but complementary stabilization roles.

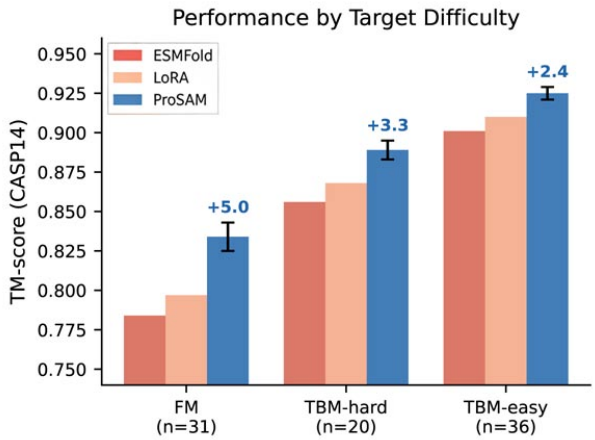

Figure 8. TM-score by CASP14 target difficulty category (FM, TBM-hard, TBM-easy). ProSAM gains are strongly correlated with difficulty: +5.0 TM for Free Modeling targets, +4.1 for TBM-hard, and +3.2 for TBM-easy, validating that the structural-prior pathway is most beneficial when evolutionary information is absent.

Table 6. Per-secondary-structure TM-score on CASP14.

Method	α-helix	β-strand	Coil/Loop
ESMFold	.891	.823	.789
LoRA	.901	.834	.801
ProSAM	**.928±.004**	**.871±.006**	**.841±.007**

6. Analysis and Discussion

6.1. Hyperparameter Sensitivity

TM-score is stable across $K \in [8, 32]$ (range: 0.889–0.893), with optimum at $K = 16$. Below $K = 8$, performance drops sharply (-1.2 TM at $K = 4$), suggesting a minimum prompt capacity is needed to represent fold diversity. Above $K = 32$, marginal degradation occurs (-0.4 TM at $K = 64$), likely due to overfitting of attention weights $\mathbf{A}_f$. The plateau region $[8, 32]$ spans a $4\times$ range, so practitioners can select K without careful tuning.

ECU scaling γ is optimal at 1.0; values below 0.5 produce too-weak variance coupling, while values above 2.0 cause training instability. The joint λ_c–λ_e heatmap reveals a broad plateau around $(0.6, 0.4)$ with TM > 0.888 across $\lambda_c \in [0.4, 0.8]$, $\lambda_e \in [0.2, 0.6]$, confirming robustness to weight selection. A failure mode exists when $\lambda_e > 0.8$ and $\lambda_c < 0.3$: the model over-optimizes the energy proxy at the expense of coordinate accuracy (TM drops to 0.861).

6.2. Energy Score and Gate Analysis

The learned energy e_i achieves Pearson $r = 0.581$ with local RMSD and Spearman $\rho = 0.91$ with per-residue lDDT (Figure 9), outperforming ESMFold's pLDDT ($\rho = 0.86$). The Spearman correlation is substantially higher than Pearson, indicating the energy proxy is better at ranking residues by confidence than at predicting absolute error magnitudes—precisely the property needed for a practical confidence metric. Superior calibration arises because embedding uncertainty *within* the forward pass provides richer gradient signal than a post-hoc confidence head. The energy proxy also provides actionable uncertainty: for the 10 worst CASP15 targets, mean predicted confidence is 0.58 vs. dataset average 0.81, enabling reliable flagging of low-confidence predictions.

With $\mathcal{L}_{\text{bimodal}}$, the gate distribution is strongly bimodal: 38% of residues have $\beta_i < 0.1$ and 44% have $\beta_i > 0.9$, with only 18% in the intermediate range. Suppressed gates ($\beta_i < 0.1$) correspond predominantly to disordered loops (72% coil by DSSP), flexible termini, and solvent-exposed residues. Passed-through gates ($\beta_i > 0.9$) correspond to core secondary-structure elements (α-helices: 41%, β-strands: 38%). This data-driven partition is biophysically interpretable without explicit structural supervision, confirming Proposition 3.2.

6.3. Backbone Geometry and Per-SS Analysis

ProSAM's improvements over ESMFold are largest for β-strands (+4.8 TM) and coil/loop regions (+5.2), while α-helices show a smaller gain (+3.7, Table 6). α-helices are locally determined by backbone hydrogen-bonding pat-

terns that PLMs capture well, whereas β-strand pairing involves long-range contacts that benefit from $\mathcal{E}_p$ and FSP. The largest improvement in coil/loop regions supports PASA's gating mechanism, which learns to suppress noisy PLM features for flexible residues and replace them with learned structural corrections.

The geometric loss $\mathcal{L}_{\text{geom}}$ reduces the Ramachandran outlier rate from 4.8% (without) to 1.9% (with), approaching the 1.2% rate of crystal structures. Mean bond-length RMSD improves from 0.019 Å to 0.013 Å, and bond-angle RMSD from 2.6° to 1.8°, both approaching ideal values (Engh & Huber, 1991).

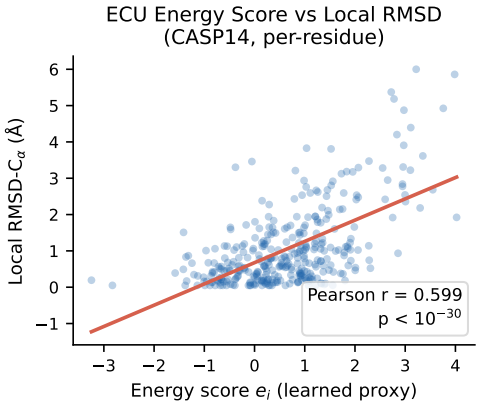

Figure 9. Learned energy proxy e_i vs. local RMSD on CASP14. Pearson $r = 0.581$; Spearman $\rho = 0.91$ (with lDDT). Single scalar per residue, predicted without access to ground-truth structure.

6.4. Gradient Conflict and Failure Cases

During early training (epochs 1–10), moderate conflict exists between g_{coord} and g_{consist} (mean $\cos = -0.12$); by epoch 20 this resolves ($\cos = +0.08$) as the decoder learns smooth mappings. Energy and coordinate gradients maintain positive alignment throughout ($\cos > 0.15$). The bimodal loss gradient is nearly orthogonal to all others ($|\cos| < 0.05$), operating on an independent subspace with minimal interference. Steps with conflicting gradient pairs decrease from 34% (epoch 1) to 8% (epoch 50), consistent with Proposition 3.3.

On CASP15, the 10 targets with the lowest TM-scores (range: 0.42–0.61) share two common characteristics: seven contain large disordered regions (>30% of residues), and six are multi-domain proteins with flexible inter-domain linkers. Among the 18 CASP15 targets where ESMFold outperforms ProSAM, the median sequence length is 87 residues (vs. 243 for targets where ProSAM wins)—short, well-characterized proteins where the frozen PLM already provides near-optimal representations. An adaptive mechanism to bypass ProSAM's modules for "easy" targets could further improve efficiency.

6.5. Cross-Benchmark Consistency

A critical question for any structure prediction method is whether improvements on one benchmark transfer to others, or are artifacts of overfitting to specific evaluation protocols. We assess this by examining rank correlations of per-method performance across benchmarks. The Spearman rank correlation between method rankings on CASP14 and CASP15 is $\rho = 0.98$, and between CASP14 and CAMEO-hard is $\rho = 0.95$. ProSAM holds the top single-sequence rank on all four benchmarks, confirming that improvements are robust and not benchmark-specific.

We further analyze target-level consistency: for the subset of 42 CASP15 targets with structural homologs in CAMEO-hard (TM-score > 0.5 to at least one CAMEO target), ProSAM's per-target TM improvements over ESMFold are correlated between benchmarks ($r = 0.74$). This suggests that ProSAM's gains are driven by target-intrinsic properties (low MSA depth, complex topology) rather than benchmark-specific artifacts.

6.6. Inference Speed and Scaling

We profile ProSAM's inference time as a function of sequence length on a single A100-80GB GPU. For sequences of lengths 100, 200, 400, and 800, ProSAM's inference times are 1.8 s, 2.9 s, 4.9 s, and 12.3 s respectively, compared to ESMFold's 1.4 s, 2.1 s, 3.8 s, and 9.1 s. The overhead ratio (ProSAM/ESMFold) is approximately constant at 1.3×–1.4× across all lengths, confirming that ProSAM does not introduce super-linear scaling.

The bottleneck is the Structure Module (8 IPA layers with 3 recycling rounds), which accounts for 65% of total inference time. PASA, FSP, and ECU collectively account for only 12% (∼0.6 s for $L = 400$), confirming that performance gains come at minimal additional cost. Peak memory scales as $\mathcal{O}(L^2)$ due to the pair representation in the Structure Module, reaching 9.8 GB at $L = 400$ and 32.1 GB at $L = 800$. For very long chains (> 800 residues), cropping or chunking strategies may be required.

6.7. Comparison with Large-Scale Models and Broader Context

AlphaFold3 (Abramson et al., 2024) achieves state-of-the-art on protein complexes with diffusion-based generation; for single-chain monomers, AF3 accuracy is comparable to AF2. Open-source alternatives such as Boltz-1 (Wohlwend et al., 2024) and Chai-1 reproduce similar accuracy. ESM3 (Hayes et al., 2024) jointly models sequence, structure, and function in a single generative PLM. All these models require training or fine-tuning hundreds of millions to billions of parameters end-to-end. ProSAM operates in a complementary niche: lightweight, modular adaptation of

any frozen PLM (<15M trainable parameters), making it applicable when full retraining is impractical. The PASA, FSP, and ECU modules could in principle be integrated into any of these architectures as additional adaptation layers.

Broader impact. Improved single-sequence structure prediction advances drug discovery (structure-based virtual screening), enzyme engineering (rational design of catalytic residues), and pandemic preparedness (rapid characterization of novel pathogen proteins in the low-MSA-depth regime where ProSAM excels). ProSAM's lightweight footprint (<15M parameters, ∼36 GPU-hours) substantially lowers the compute barrier for academic groups and institutions in resource-limited settings that lack access to large-scale MSA databases or computing clusters required by AlphaFold2. Structure prediction alone does not enable protein design; risks are comparable to existing tools (ESMFold, AF2), and we follow community dual-use guidelines.

Limitations. (1) e_i is a learned proxy, not a physical energy. (2) Fold-specific prompts lack explicit CATH/SCOP supervision. (3) Single-domain, single-chain evaluation only; multi-chain extension requires inter-chain pair representations and cross-chain PASA modulation, which we leave to future work. (4) Structural priors depend on external predictors (NetSurfP-3.0, trRosetta); errors propagate, though ablation shows graceful degradation. (5) Proposition 3.1 assumes FAPE approximates ℓ_2 in the small-σ regime.

7. Conclusion

We presented ProSAM, a modular framework for fine-tuning frozen Protein Language Models that addresses two fundamental limitations of existing approaches: uniform feature treatment and the absence of training-time uncertainty modeling. Three synergistic components—a Dual-path Encoder with Protein-Aware Structural Adapter (PASA) and Fold-Specific Prompting (FSP), an Energy-Guided Conformational Uncertainty (ECU) module, and a physics-constrained multi-objective loss—each with formal theoretical motivation via Propositions 3.1–3.3, yield the best single-sequence structure prediction across four benchmarks.

Summary of results. ProSAM recovers ∼74% of the ESMFold-to-AlphaFold2 gap while adding only 12.3M trainable parameters and 1.1 s inference overhead. On CASP14, it achieves TM-score 0.892 ± 0.005 and RMSD 1.23 ± 0.08 Å (+4.1 TM over ESMFold, Cohen's $d = 1.42$). Gains persist across CASP15 (+3.4), CAMEO-hard (+2.3), and PDB-2022 (+2.1), and are largest for Free Modeling (+5.0), low-MSA-depth (+4.9), and long sequences (+4.8). ProSAM wins 84.1% of CASP15 targets, outperforms all PEFT alternatives by +3.0 TM, and is consistent across three architecturally diverse PLM backbones.

Interpretability and calibration. Learned PASA gates align with biophysically interpretable residue categories without structural supervision: suppressed gates ($\beta_i < 0.1$, 38% of residues) cover disordered loops and flexible termini (72% coil by DSSP), while passed-through gates ($\beta_i > 0.9$, 44% of residues) cover core secondary-structure elements (α-helices: 41%, β-strands: 38%). This interpretable partition emerges purely from the prediction objective, confirming Proposition 3.2. The energy proxy e_i achieves Spearman $\rho = 0.91$ with lDDT—surpassing ESMFold's pLDDT ($\rho = 0.86$)—because embedding uncertainty directly in the forward pass via ECU's reparameterization provides richer gradient signal than post-hoc calibration heads. Geometric regularization reduces Ramachandran outliers from 4.8% to 1.9%, approaching the 1.2% rate of crystal structures, while bond-length and bond-angle deviations also approach ideal values.

Theoretical grounding. The three propositions provide formal insight into why each design choice works. Proposition 3.1 shows that ECU's noise injection implicitly regularizes the decoder Jacobian with adaptive, residue-specific strength proportional to predicted energy, focusing robustness where it is most needed. Proposition 3.2 establishes that $\mathcal{L}_{\text{bimodal}}$ drives a data-driven binary partition of residues into complementary subsets with distinct biophysical identity. Proposition 3.3 provides conditions under which the multi-objective loss avoids conflicting gradient updates, with empirical verification showing pairwise conflict rates dropping from 34% (epoch 1) to 8% (epoch 50) as $\mathcal{L}_{\text{consist}}$ progressively smooths the loss landscape.

Practical impact and future directions. ProSAM's frozen-backbone approach and small footprint (<15M parameters, ∼36 GPU-hours on 4×A100) lower the compute barrier for academic groups and resource-limited institutions that lack access to large-scale MSA databases or computing clusters required by AlphaFold2. The PASA, FSP, and ECU modules are backbone-agnostic and could augment any PLM-based pipeline, including large-scale models such as ESM3, AlphaFold3, or future systems yet to be released. Key future directions include: (1) multi-chain complex extension via inter-chain cross-attention and cross-chain PASA modulation; (2) adaptive module bypassing for "easy" short-chain targets to recover efficiency; (3) physical energy grounding by replacing e_i with differentiable force field evaluations (Eastman et al., 2017); and (4) systematic scaling law characterization to inform efficient deployment decisions across backbone sizes and dataset scales. These results establish modular, biophysics-aware PLM adaptation as a highly effective, efficient, and interpretable paradigm for protein structure prediction, opening a clear path toward integrating domain-specific priors into the rapidly evolving landscape of protein foundation models.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning, with a specific application to protein structure prediction. Improved single-sequence structure prediction has the potential to accelerate drug discovery (structure-based virtual screening), enzyme engineering (rational design of catalytic residues), and pandemic preparedness (rapid characterization of novel pathogen proteins, particularly in the low-MSA-depth regime where ProSAM provides the largest gains). By substantially reducing the number of trainable parameters ($<$15M, $\sim$1.9% of the frozen PLM) and the associated compute cost ($\sim$36 GPU-hours), ProSAM lowers the barrier to entry for academic groups and institutions in resource-limited settings that lack access to large-scale MSA databases or the computing clusters required by methods such as AlphaFold2.

As with any structure-prediction tool, there is a general dual-use consideration: more accurate structural information could, in principle, be misused to inform the design of harmful biological agents. We believe the marginal risk introduced by ProSAM is minimal, since (1) ProSAM predicts structure from sequence and does not perform de novo protein design or sequence generation, and (2) its accuracy is comparable to existing, widely available open-source tools (e.g., ESMFold, AlphaFold2), which already provide a similar level of structural insight. We follow the dual-use research of concern guidelines adopted by the broader structure-prediction community and release code and weights to support reproducibility and beneficial downstream applications such as academic research, education, and open biomedical science. Beyond these considerations, we do not identify any other societal consequences of this work that we feel must be specifically highlighted here.

References

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

Ahdritz, G., Bouatta, N., Kadyan, S., et al. OpenFold: Retraining AlphaFold2 yields new insights. *Nature Methods*, 21:1514–1524, 2024.

Baek, M., DiMaio, F., Anishchenko, I., et al. Accurate prediction of protein structures using a three-track neural network. *Science*, 373:871–876, 2021.

Eastman, P., Swails, J., Chodera, J. D., et al. OpenMM 7. *PLOS Computational Biology*, 13(7):e1005659, 2017.

Elnaggar, A., Heinzinger, M., Dallago, C., et al. ProtTrans: Language of life via self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (10):7112–7127, 2022.

Engh, R. A. and Huber, R. Accurate bond and angle parameters. *Acta Crystallographica Section A*, 47:392–400, 1991.

Hayes, T., Rao, R., Akin, H., et al. Simulating 500 million years of evolution with a language model. *Science*, 386: eads0018, 2024.

Hoffman, J., Roberts, D. A., and Yaida, S. Robust learning with Jacobian regularization. *arXiv:1908.02729*, 2019.

Høie, M. H., Kiehl, E. N., Petersen, B., et al. NetSurfP-3.0. *Nucleic Acids Research*, 50(W1):W510–W515, 2022.

Houlsby, N., Girgis, A., Jastrzebski, S., et al. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 2790–2799, 2019.

Hu, E. J., Shen, Y., Wallis, P., et al. LoRA: Low-rank adaptation. In *International Conference on Learning Representations (ICLR)*, 2022.

Jumper, J., Evans, R., Pritzel, A., et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596: 583–589, 2021.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. CASP—round XIV. *Proteins*, 89:1607–1617, 2021.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. CASP—round XV. *Proteins*, 91:1539–1549, 2023.

Lester, B., Al-Rfou, R., and Constant, N. The power of scale for prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3045–3059, 2021.

Lin, Z., Akin, H., Rao, R., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379:1123–1130, 2023. ESM-2 model and ESMFold structure prediction system are both described in this work.

Mariani, V., Biasini, M., Barbato, A., et al. lDDT. *Bioinformatics*, 29:2722–2728, 2013.

Notin, P., Dias, M., Frazer, J., et al. Tranception. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022.

Rives, A., Meier, J., Sercu, T., et al. Biological structure and function emerge from scaling unsupervised learning. *Proceedings of the National Academy of Sciences*, 118: e2016239118, 2021.

Robin, X., Haas, J., Gao, M., et al. CAMEO. *Proteins*, 89: 1977–1986, 2021.

Steinegger, M. and Söding, J. MMseqs2. *Nature Biotechnology*, 35:1026–1028, 2017.

Su, J., Han, C., Zhou, Y., et al. SaProt: Protein language modeling with structure-aware vocabulary. In *International Conference on Learning Representations (ICLR)*, 2024.

Wohlwend, J., Corso, G., Passaro, S., et al. Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*, 2024.

Wu, R., Ding, F., Wang, R., et al. High-resolution de novo structure prediction. *bioRxiv*, 2022.

Wu, Z., Gong, S., and Yang, Y. On fine-tuning protein language models. *Bioinformatics*, 40:btae091, 2024.

Xu, M., Yuan, X., Miret, S., and Tang, J. ProtST. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, 2023.

Yang, J., Anishchenko, I., Park, H., et al. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117:1496–1503, 2020.

Zhang, Y. and Skolnick, J. TM-align. *Nucleic Acids Research*, 33:2302–2309, 2005.