Supplementary Information

1 Details of hyperparameters in LMProtein training

Table S1. Fine-tuning and training hyperparameters used in LMProtein.

Hyperparameter name	Description	Value / Range	Notes
batch_size	Number of protein sequences per batch	30	Selected empirically to balance GPU memory (12 GB) and convergence stability
learning_rate	Initial learning rate for Adam optimizer	1×10^{-4}	Chosen after small-scale trials (1e-5 - 5e-4)
optimizer	Optimization algorithm	Adam	Standard for transformer fine-tuning
weight_decay	Weight decay coefficient	0.001	To prevent overfitting
dropout	Dropout ratio in CNN/LSTM layers	0.5	Determined via ablation tests
epochs	Maximum training epochs	50 - 60	With early stopping at plateau
activation	Nonlinear function	ReLU	For CNN and MLP layers
loss_function	Multi-task loss combining CE and MSE	-	See Section 2.3 of main text

During model fine-tuning, several candidate batch sizes (16, 32, 64) were tested in preliminary runs. Batch = 30 was finally chosen as it achieved stable gradient updates and full GPU utilization on NVIDIA RTX 3060 (12 GB) without overflow.

Learning rate, dropout, and weight decay were adjusted to minimize validation loss and overfitting risk.

2 Training and convergence characteristics

Training and validation loss curves of LMProtein across 40 epochs.
Both curves show smooth convergence, with early stopping triggered at epoch 37.

This indicates that the selected hyperparameters (batch = 30, lr = 1e-4, dropout = 0.5) provide stable optimization performance.

3 Architecture and parameter statistics

Table S2. Summary of major modules and parameter sizes in LMProtein.

Module	Layer / Structure	Hidden size	Parameters
Embedding	ESM-2 pretrained model (t33 650M)	1280 per residue	650 M
CNN-1	1D Conv (kernel = 129)	1280 → 32	5.3 M
CNN-2	1D Conv (kernel = 257)	1280 → 32	10.6 M
BatchNorm + Dropout	Applied to concatenated 1344-dim features	-	-
LSTM (2 layers)	Hidden size = 1024	8.4 M	
MLP (2 layers)	1280 → 512 → 1	0.7 M	

Total trainable parameters ≈ 675 M (including fine-tuned layers).
All computations were performed under PyTorch 2.2 environment with CUDA 12.2.

4 Evaluation metrics

LMProtein adopts a mixed set of metrics for classification and regression tasks:

Task	Metric	Formula / Definition	Purpose
SS3 / SS8	Accuracy, Precision, Recall, F1	Standard multi-label evaluation	Classification
Φ, Ψ	MAE (°)	mean(	y_pred - y_true
Fluorescence / Stability	Spearman’s ρ and MSE	ρ = rank correlation	Correlation of predicted vs. experimental data

This combination ensures consistent comparison with prior works such as NetSurfP-2.0 and SPOT-1D-LM.

5 Supplementary Figures

Figure S1. Distribution of SS8 classes across training and test sets (TS115, CB513, CASP12, CASP14_FM).

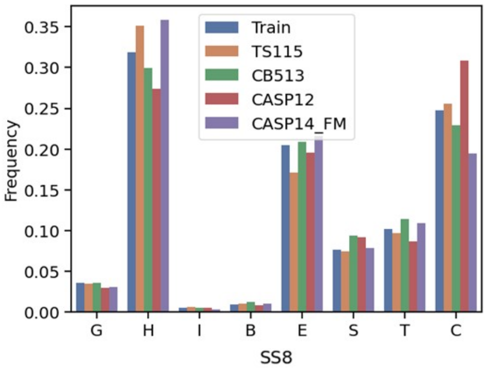
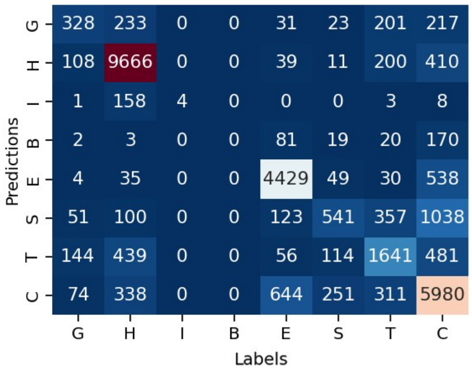


Figure S2. Matrix of SS8 prediction results on the combined test datasets.

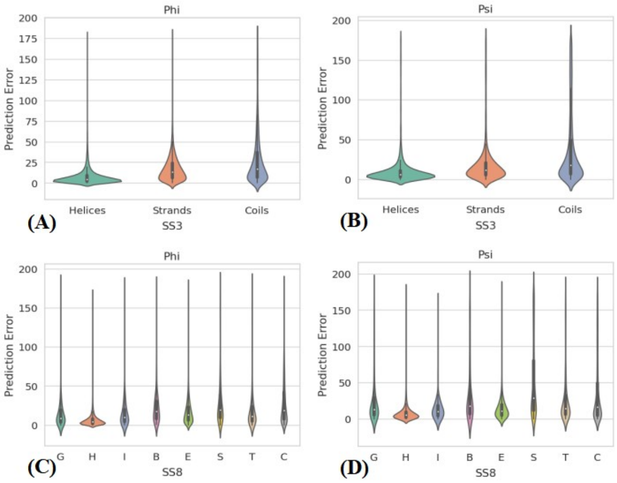


Predictions \ Labels	G	H	I	B	E	S	T	C
G	328	233	0	0	31	23	201	217
H	108	9666	0	0	39	11	200	410
I	1	158	4	0	0	0	3	8
B	2	3	0	0	81	19	20	170
E	4	35	0	0	4429	49	30	538
S	51	100	0	0	123	541	357	1038
T	144	439	0	0	56	114	1641	481
C	74	338	0	0	644	251	311	5980

The distribution in (1) reveals substantial class imbalance, particularly the scarcity of π-helix (I) and β-bridge (B) types compared with α-helix (H) and β-strand (E).

The confusion matrix in (2) shows that LMProtein achieves high accuracy for frequent classes (H, E, C), while rare classes (I, B) are often misclassified, indicating that the imbalance remains a key limitation in SS8 prediction tasks.

Figure S3.Performance for Phi and Psi prediction.



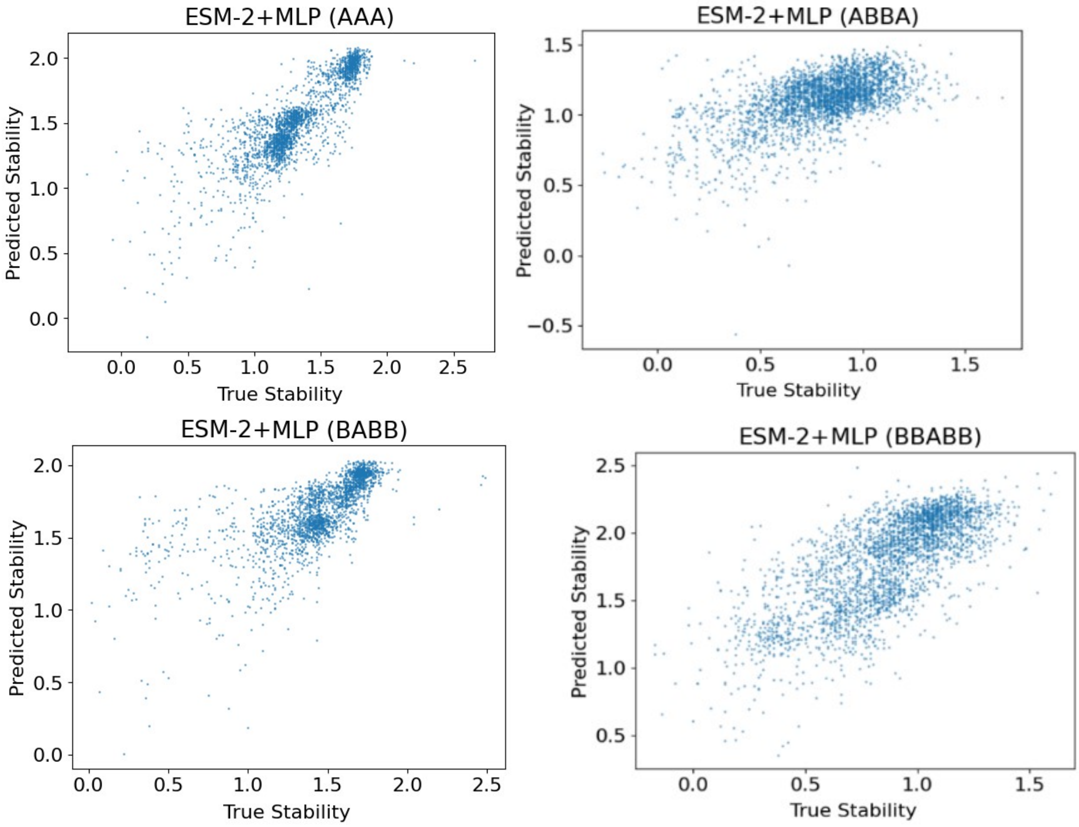
Violin plots illustrating the prediction error distributions of backbone dihedral angles Φ (A, C) and Ψ (B, D) for different secondary structure types.

(A–B) show the error distributions under SS3 classification (Helices, Strands, Coils), and (C–D) show the same under SS8 classification.

The plots indicate that prediction errors are smallest for helices and largest for coils, consistent with the higher structural regularity of α-helices and β-strands compared with irregular coil regions.

Similar error patterns are observed for both Φ and Ψ, suggesting that LMProtein captures local structural constraints consistently across residue-level tasks.

Figure S4.



Predicted versus true stability scores of LMProtein on different protein fold topologies using the ESM-2+MLP model.

(A) ααα, (B) αββα, (C) βαββ, and (D) ββαββ topologies are shown, corresponding to increasing fold complexity.

The scatter plots reveal that ααα and αββα topologies exhibit the strongest linear correlations (Spearman's ρ = 0.84 and 0.79, respectively), indicating higher prediction reliability for structurally regular folds.

In contrast, more complex topologies (βαββ and ββαββ) show greater dispersion due to increased conformational variability.

These results complement Figure 7 in the main text by providing a more detailed analysis of topology-specific performance.

6 Hardware and reproducibility

Experiments were conducted on a single NVIDIA RTX 3060 (12 GB) GPU.
Training time for a full run of LMProtein (40 epochs) averaged 8 hours, depending on task type.
All source code and preprocessed datasets are publicly available at:

https://github.com/hluo421/LMProtein

This repository includes scripts for environment setup, dataset processing, model fine-tuning, and evaluation.