LaSalle Universidad
DIRECCIÓN DE INVESTIGACIÓN

Fine-tuning Transformers for Neoantigen Detection

Universidad La Salle
Ph.D(c). Vicente Machaca Arceda

Content

Content

Immunotherapy to Treat Cancer

Immunotherapy is a type of cancer treatment that helps your immune system fight cancer [1].

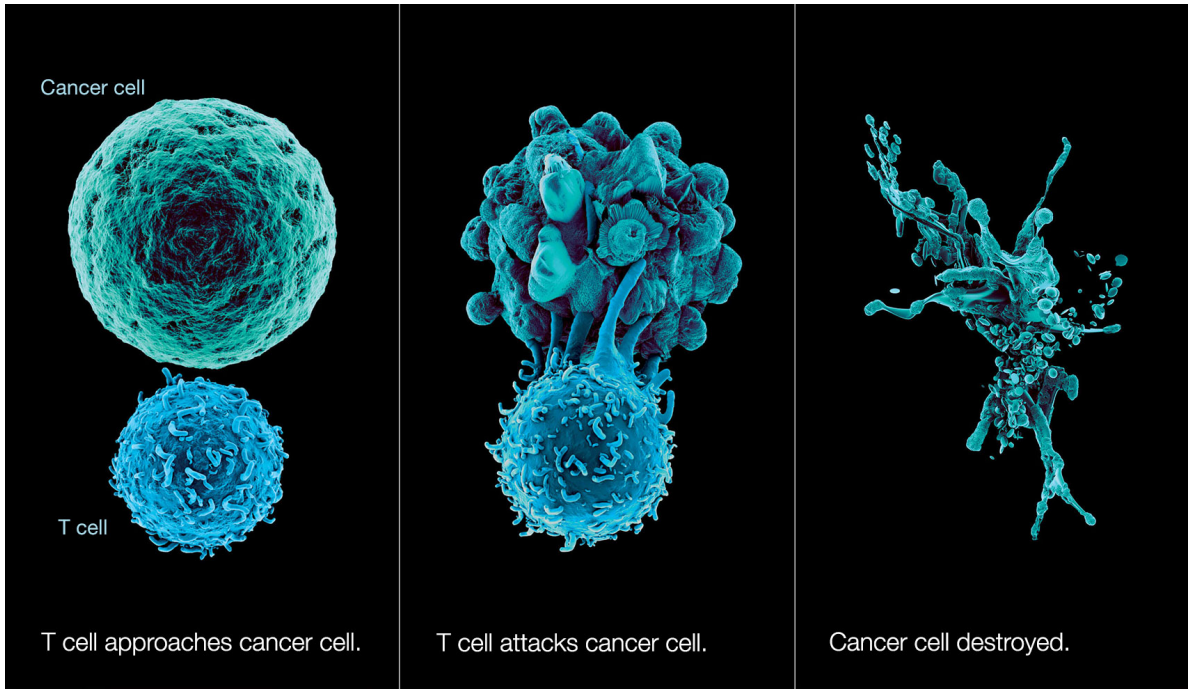


Figure: Example of how a T cell attack a cancer cell [2].

Neoantigen

A new protein that forms on cancer cells when certain mutations occur in tumor DNA. Neoantigens used in vaccines and other types of immunotherapy are being studied in the treatment of many types of cancer [3, 4].

Currently, there is a lot of methods to detect neoantigens; however, only a small number of them manage to stimulate the immune system [5, 6].

Immunotherapy for Cancer

Personalized Vaccines

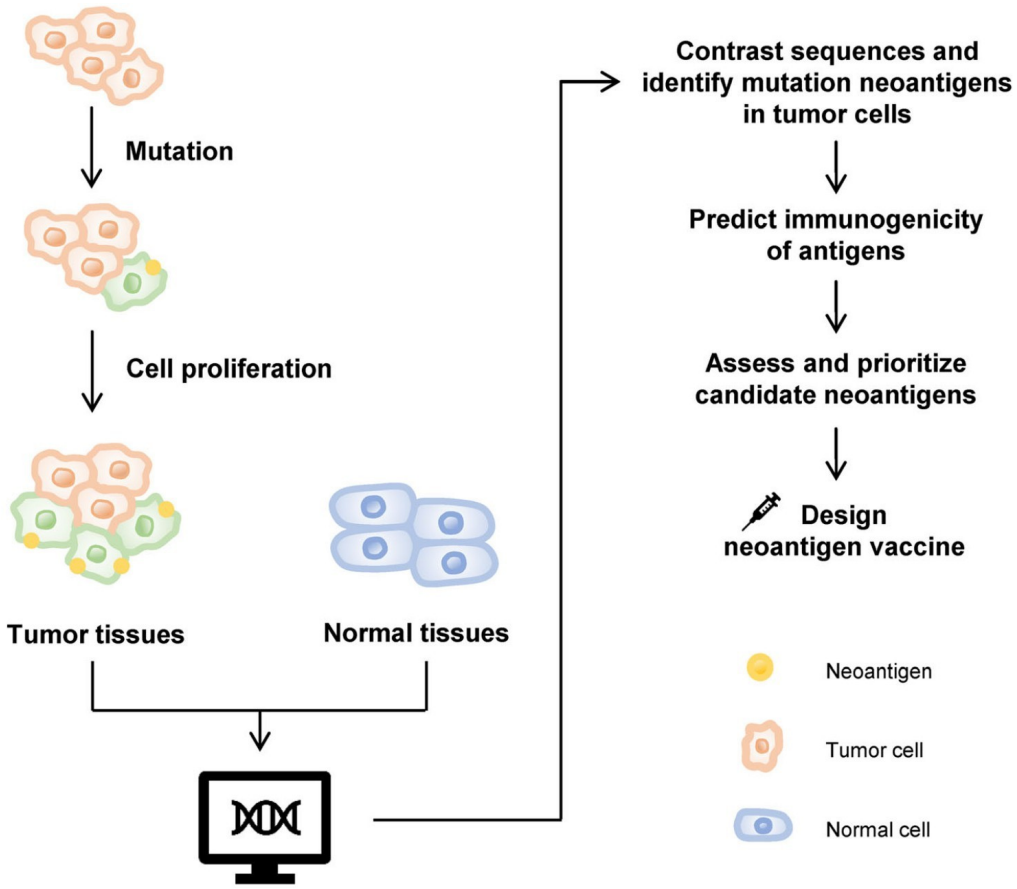


Figure: Personalized vaccines process for Cancer [7].

Immunotherapy for Cancer

Personalized Vaccines

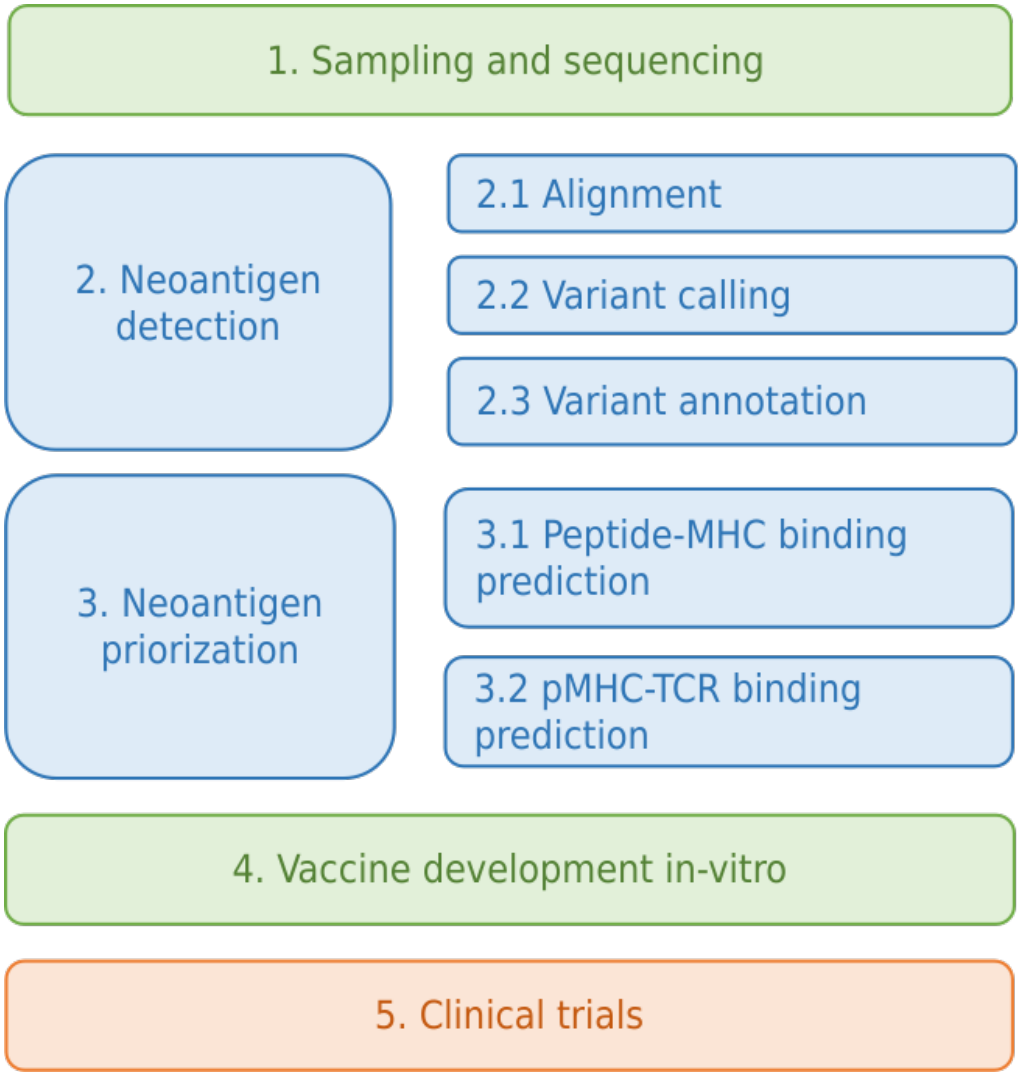


Figure: Personalized vaccines process for Cancer.

pMHC binding prediction

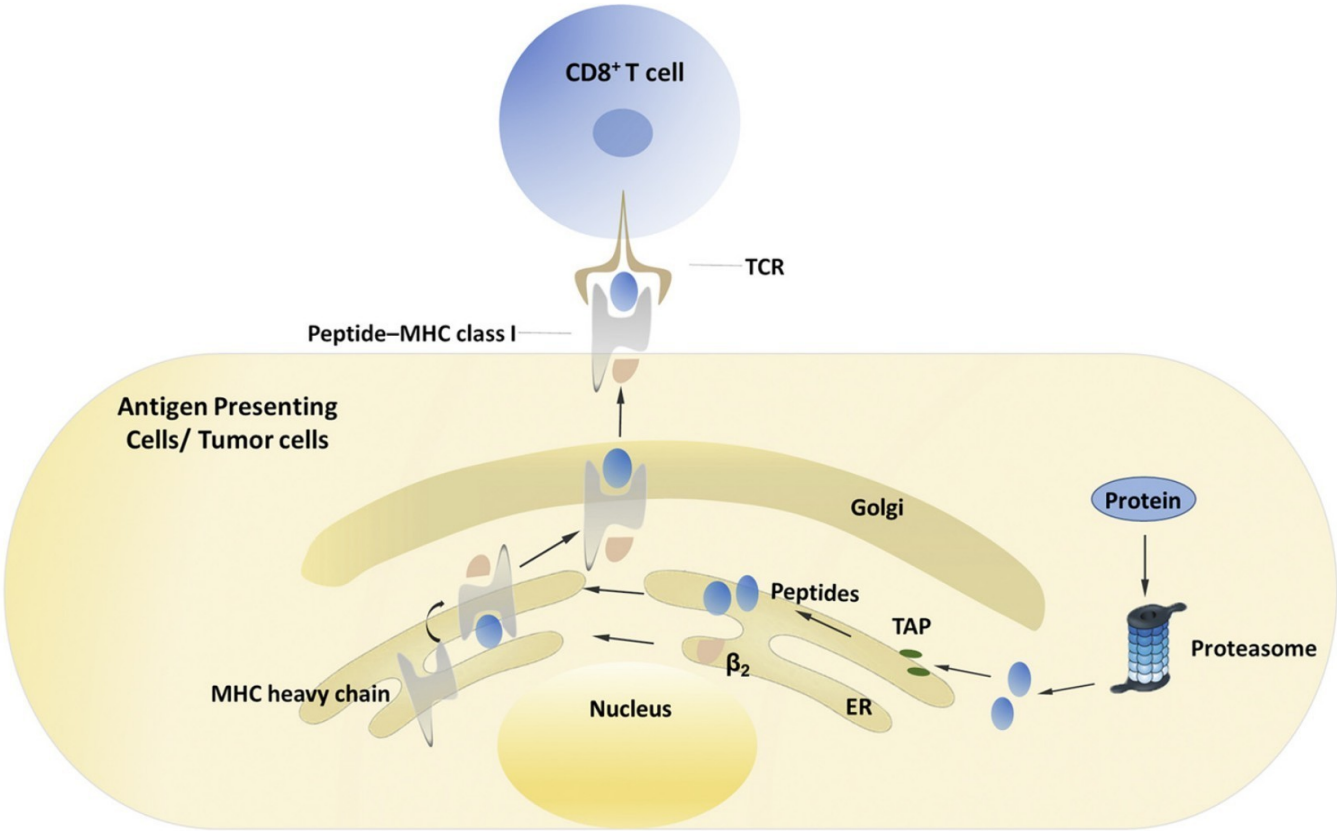


Figure: pMHC presentation process in MHC class I [8].

Content

Problem
Peptide-MHC Binding Prediction

Less than 5% of detected neoantigens (peptides binded to MHC) succeed in activating the immune system [9].

This is a **binary classification problem**. A peptide could be represented like: $p = \{A, ..., Q\}$ and a MHC like: $q = \{A, N, ..., Q, E\}$. Finally, we need to know the probability of affinity between p and q (pMHC)

Problem

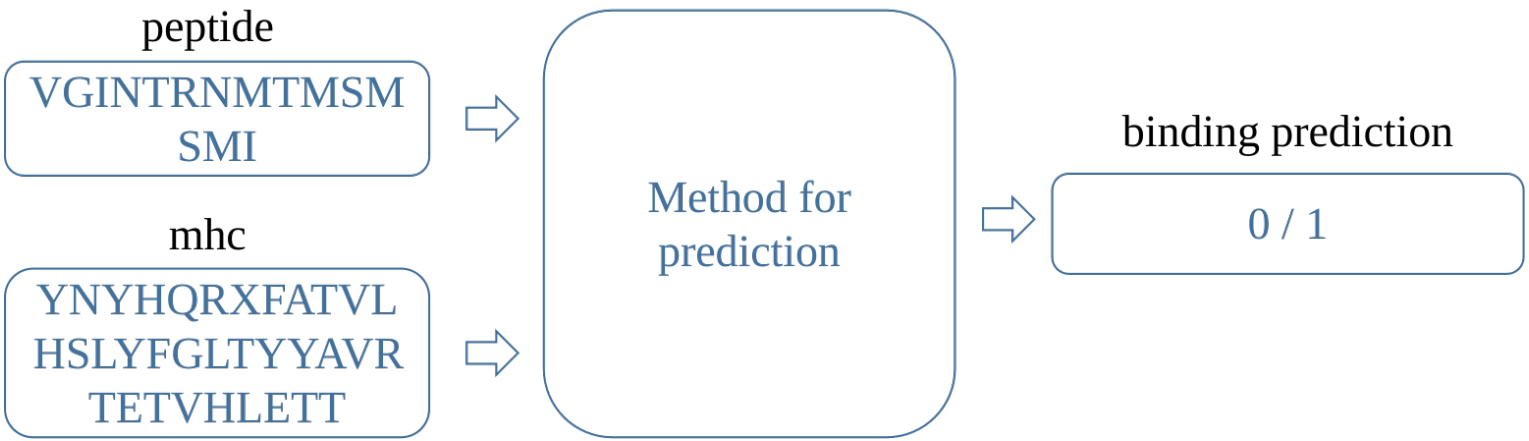


Figure: pMHC binding prediction problem.

Content

Proposal

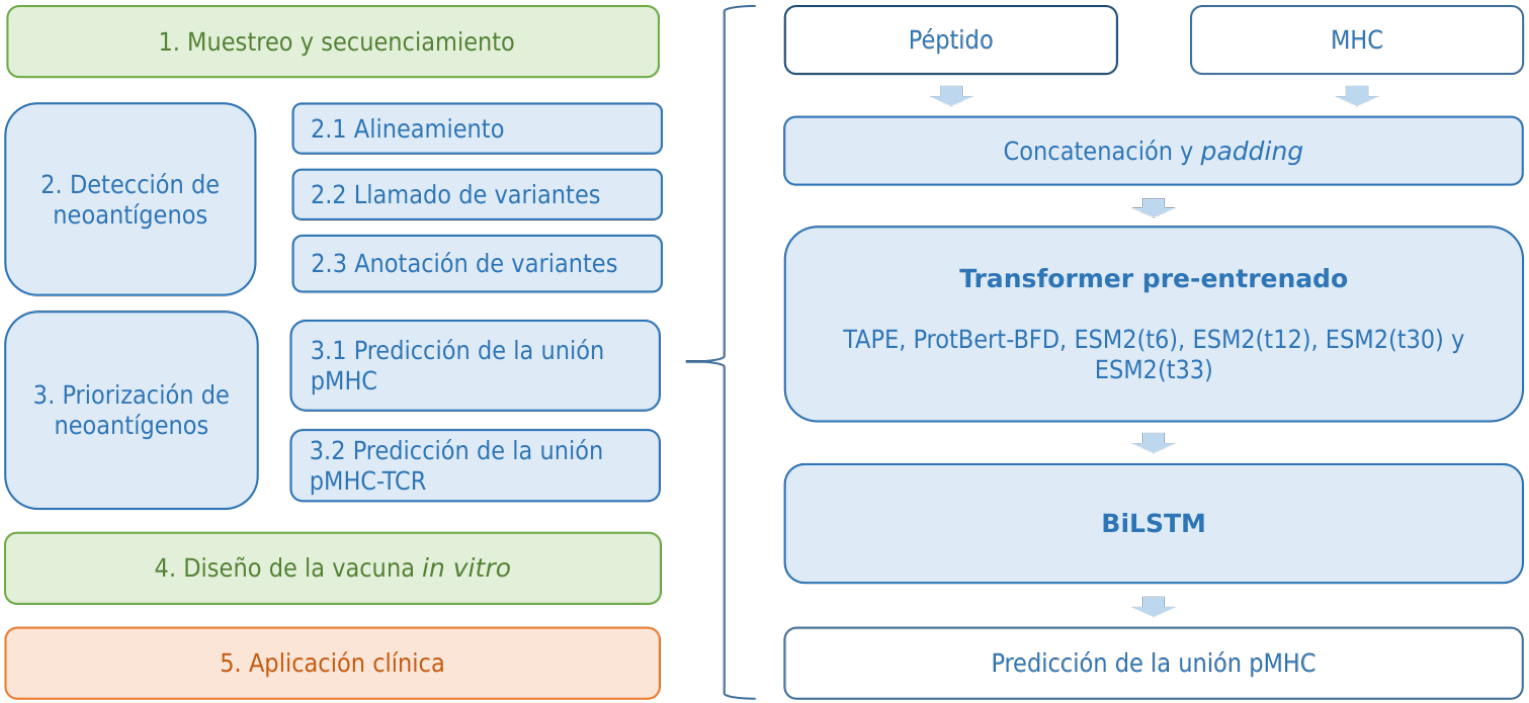


Figure: Proposal for pMHC binding prediction.

Content

Databases

Training: 539,019; Validation: 179,673; and Testing: 172,580.

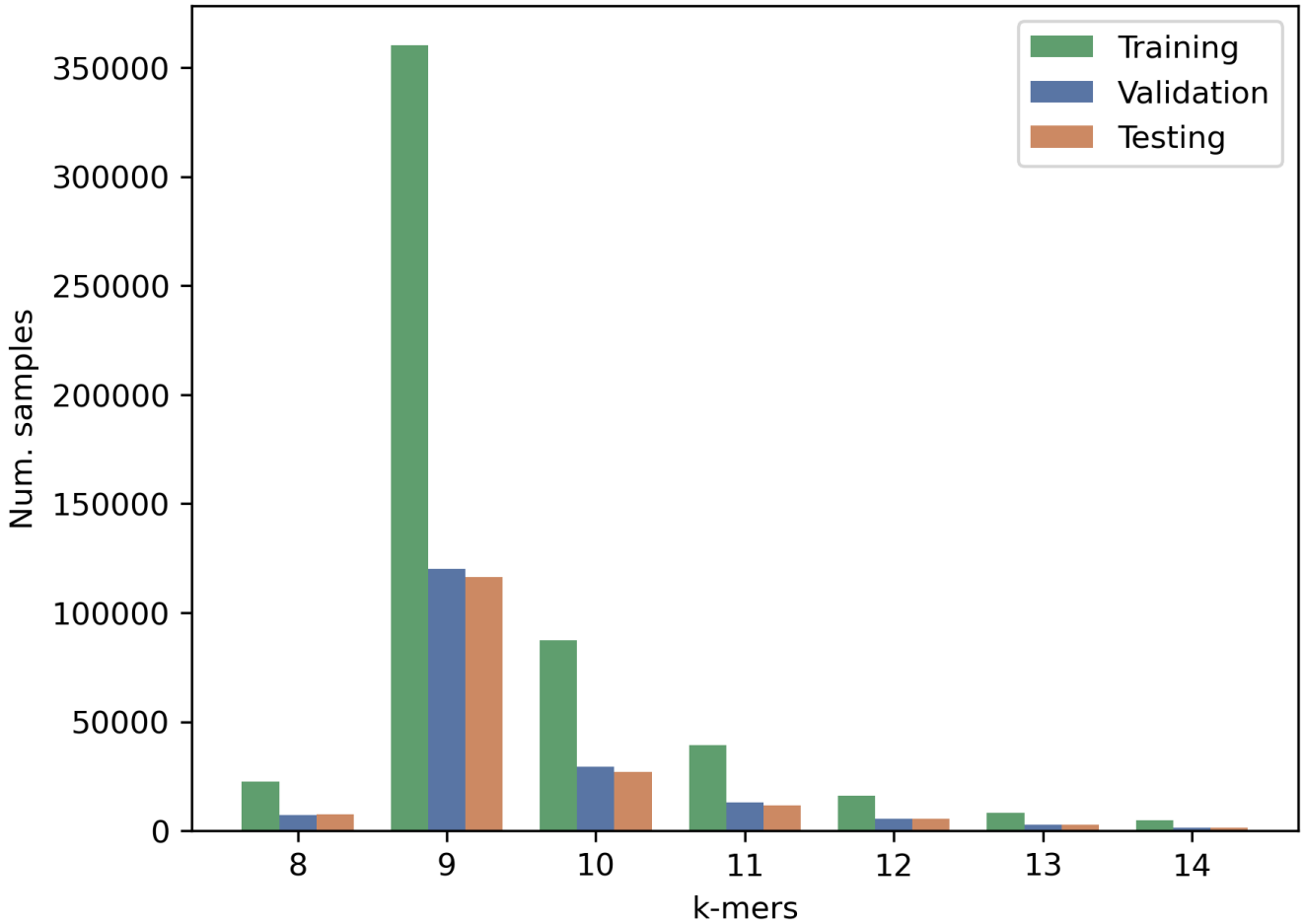


Figure: Number of samples per k-mer.

Content

Pre-trained models

Table: Differences between TAPE, ProtBert-DFB, and ESM2. HS: *Hidden size*; AH: *Attention heads*.

Model	BD	Samples	Layers	HS	AH	Params.
TAPE	Pfam	30M	12	768	12	92M
ProtBert-BFD	BFD	2122M	30	1024	16	420M
ESM2(t6)	Uniref50	60M	6	320	20	8M
ESM2(t12)	Uniref50	60M	12	480	20	35M
ESM2(t30)	Uniref50	60M	30	640	20	150M
ESM2(t33)	Uniref50	60M	33	1280	20	650M

Content

Results
(Training for 3 epochs)

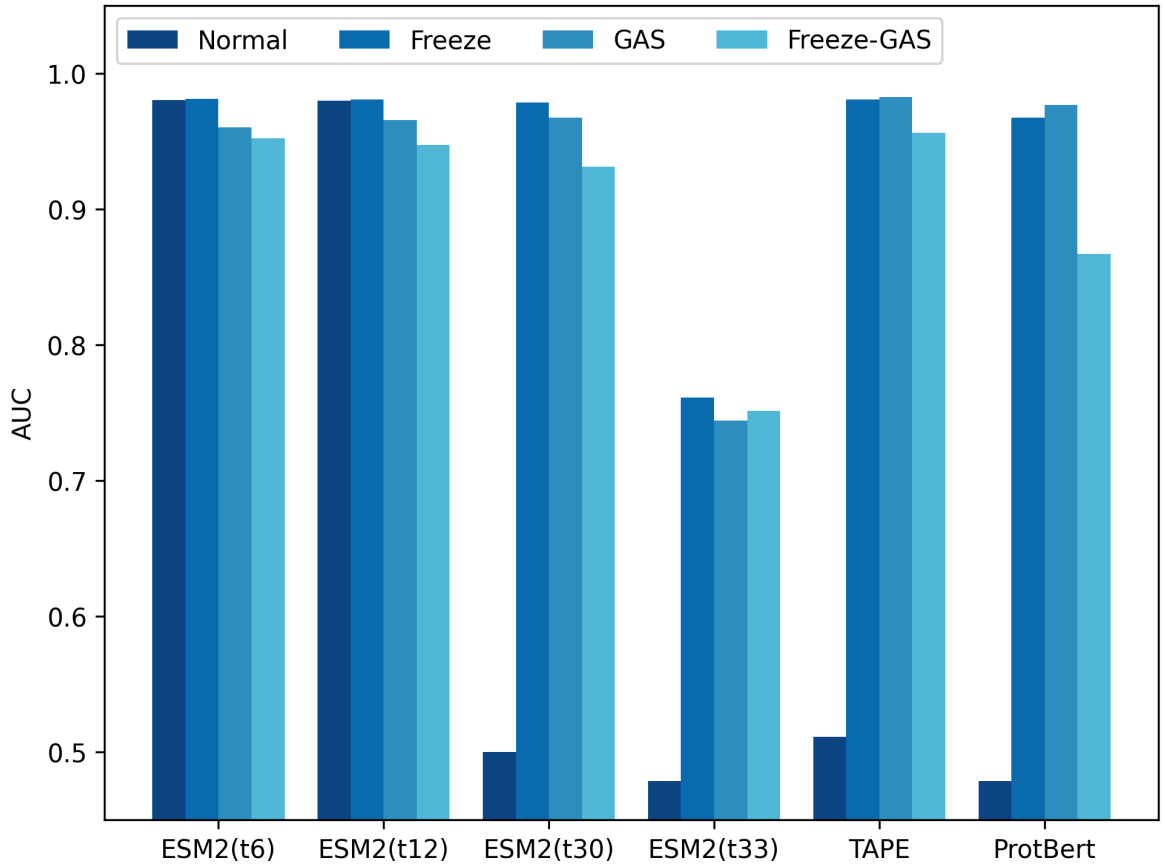
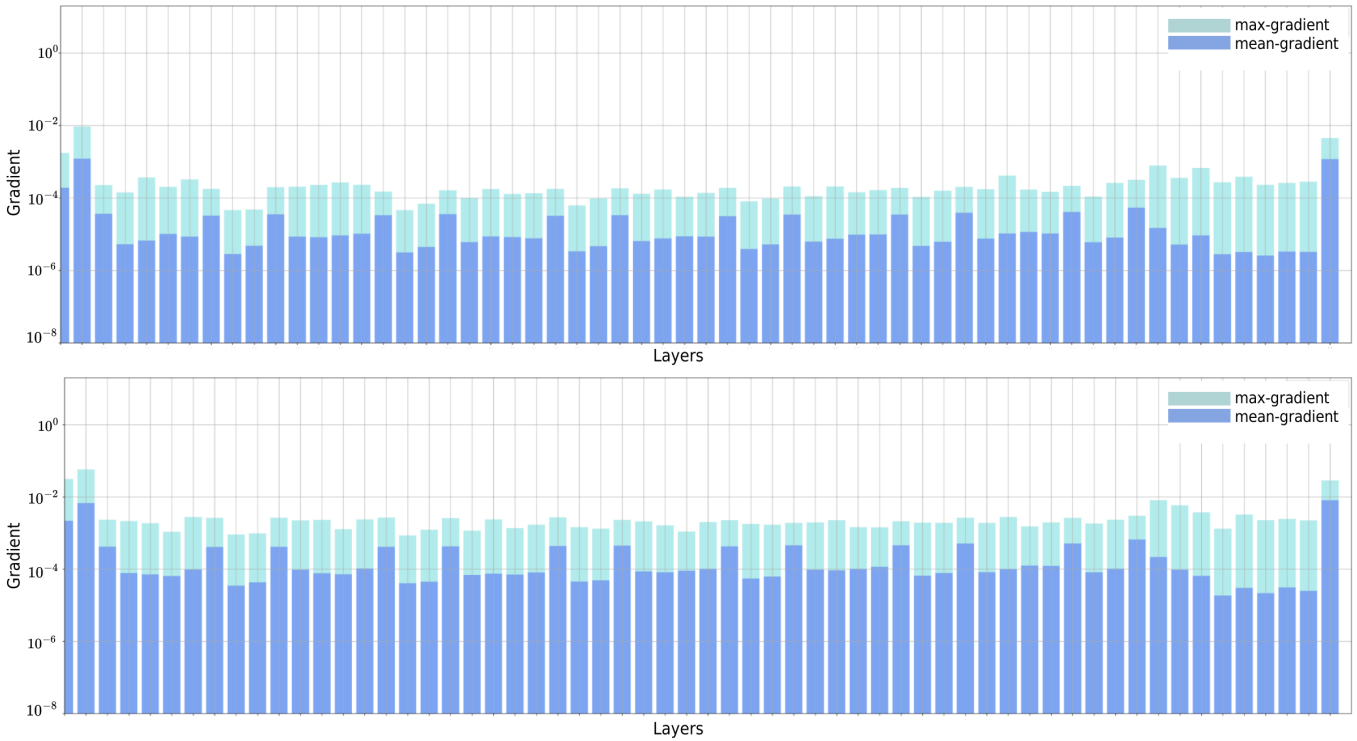


Figure: Comparative analysis of Area Under the Curve (AUC) in Transformer model architectures using various training methodologies.

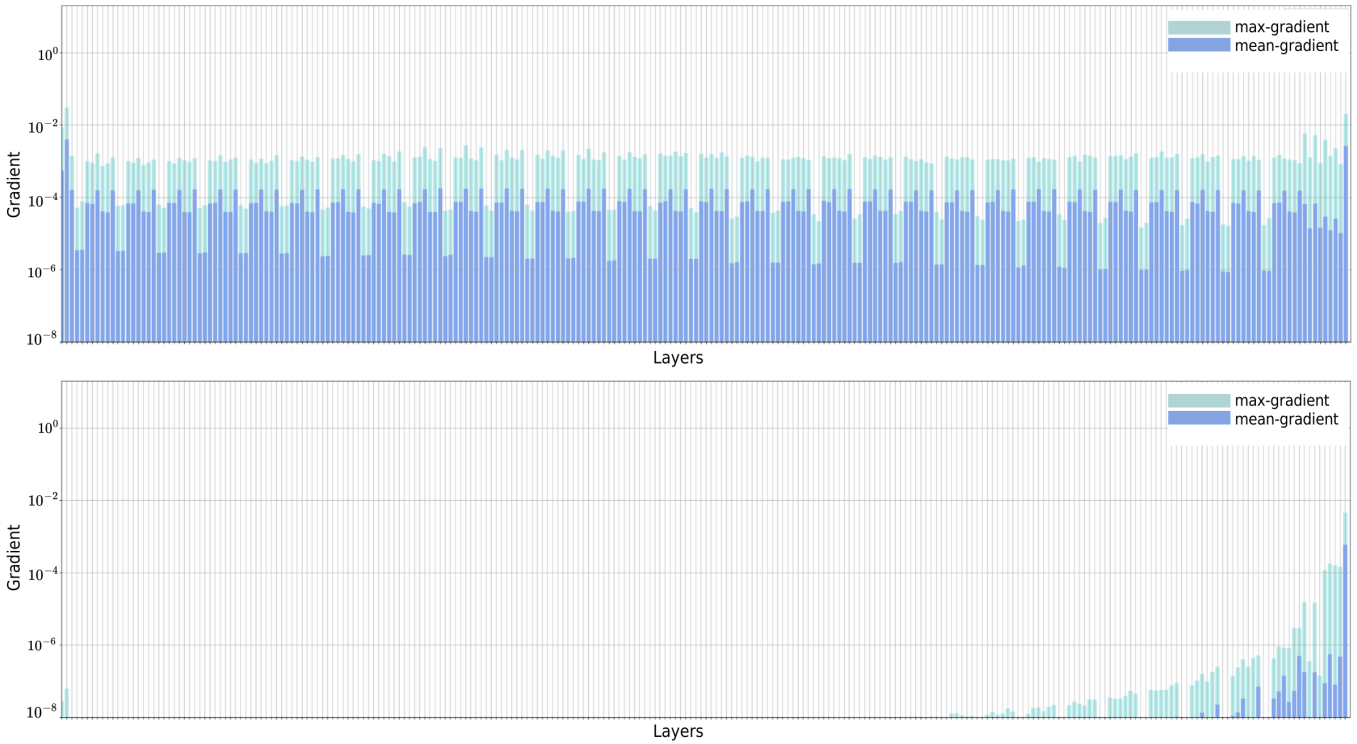
Results
Vanish gradient problem

Gradients for ESM2(t6)



Results
Vanish gradient problem

Gradients for ESM2(t30)



Results

Comparison (Training for 30 epochs)

	Accuracy	Precision	Recall	F1-score	AUC	MCC
ESM2(t6)-Normal	0.9390	0.9333	**0.9453**	0.9392	0.9797	0.8780
ESM2(t6)-Freeze	**0.9401**	**0.9398**	0.9402	**0.9400**	**0.9830**	**0.8802**
ESM2(t6)-GAS	0.9366	0.9322	0.9413	0.9368	0.9818	0.8732
ESM2(t6)-Freeze-GAS	0.9354	0.9326	0.9383	0.9355	0.9813	0.8708
ESM2(t30)-Normal	-	-	-	-	-	-
ESM2(t30)-Freeze	**0.9393**	0.9304	**0.9493**	**0.9397**	0.9787	**0.8787**
ESM2(t30)-GAS	0.9346	**0.9337**	0.9352	0.9345	0.9808	0.8691
ESM2(t30)-Freeze-GAS	0.9363	0.9319	0.9411	0.9365	**0.9818**	0.8726
TAPE-Normal	-	-	-	-	-	-
TAPE-Freeze	0.9395	**0.9404**	0.9382	0.9393	0.9815	0.8790
TAPE-GAS	**0.9415**	0.9352	**0.9484**	**0.9418**	**0.9841**	**0.8831**
TAPE-Freeze-GAS	0.9359	0.9297	0.9428	0.9362	0.9820	0.8719

Results
Comparison with state-of-art tools

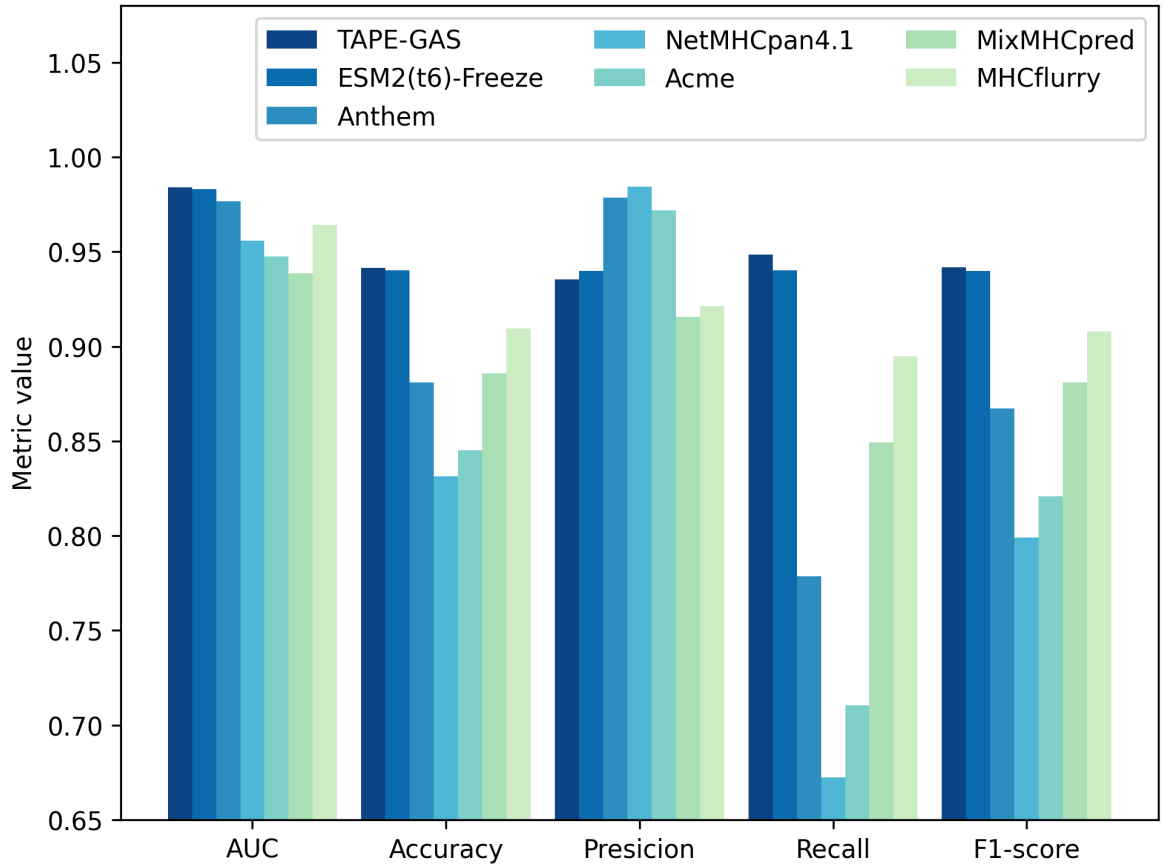


Figure: The AUC values for TAPE-GAS and ESM2(t6) trained for 30 epochs, in comparison to state-of-the-art methods.

Results
Comparison with state-of-art tools

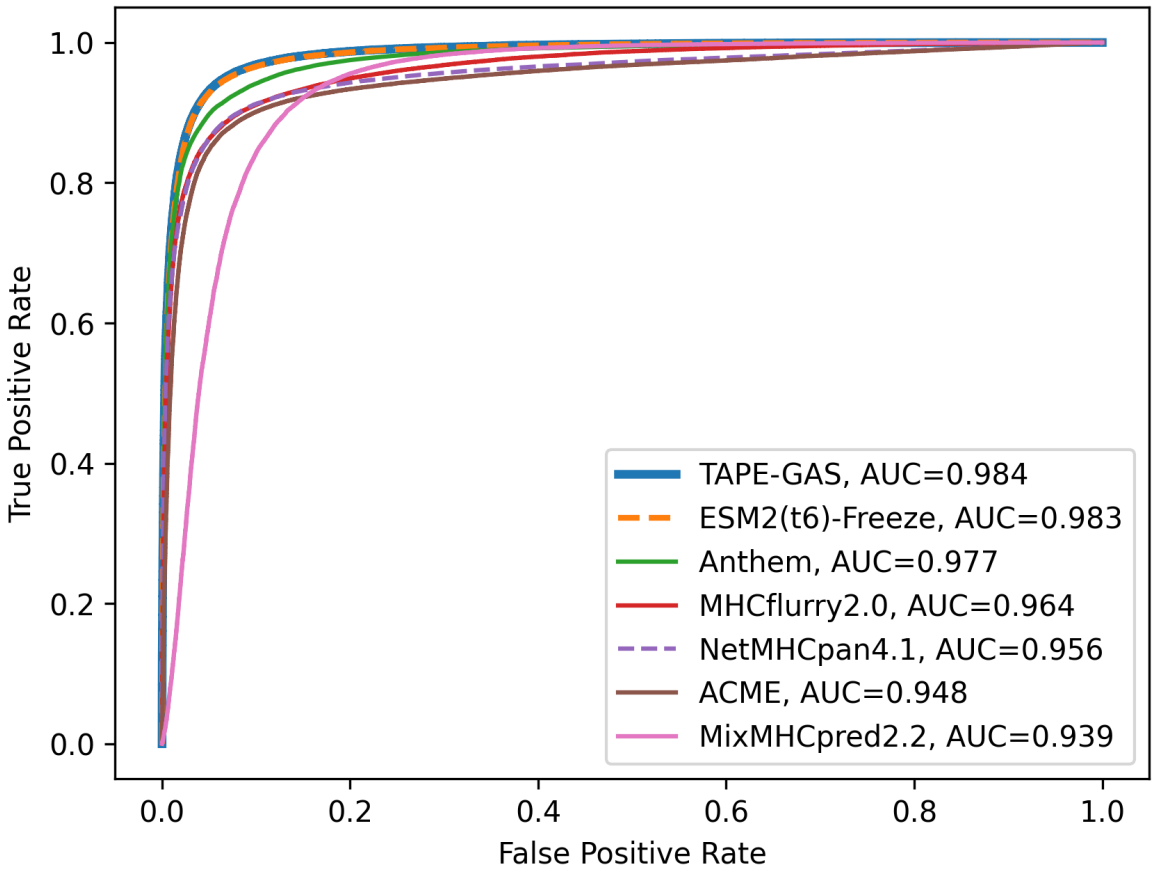


Figure: ROC curves for TAPE-GAS and ESM2(t6) trained for 30 epochs, in comparison to state-of-the-art methods.

Results
Comparison with state-of-art tools

Table: Performance evaluation of Transformer models TAPE-GAS and ESM2(t6)-Freeze, trained for 30 epochs, against Anthem, NetMHCpan4.1, ACME, MixMHCpred2.2, and MhcFlurry2.0.

	Accuracy	**Precision**	**Recall**	**F1-score**	**AUC**	**MCC**
TAPE-GAS	**0.9415**	0.9352	**0.9484**	**0.9418**	**0.9841**	**0.8831**
ESM2(t6)-Freeze	**0.9401**	0.9398	**0.9402**	**0.9400**	**0.9830**	**0.8802**
Anthem	0.8811	**0.9786**	0.7787	0.8673	0.9768	0.7785
NetMHCpan4.1	0.8312	**0.9844**	0.6724	0.7991	0.9557	0.6982
ACME	0.8452	0.9717	0.7105	0.8208	0.9476	0.7165
MixMHCpred2.2	0.8857	0.9155	0.8493	0.8811	0.9386	0.7733
MhcFlurry2.0	0.9093	0.9211	0.8948	0.9078	0.9642	0.8189

Content

Discussion

Fine-tuning ESM2 models

The most favorable results were obtained with the smallest model, **ESM2(t6)**. we believe is not sufficiently large for ESM2(t33), a model boasting 650 million parameters.

Another potential reason could be attributed to the use of **Rotary Position Embedding (RoPE)** used instead of absolute positional encoding.

Discussion

Layer Freezing and GAS

This approach involves locking the Transformer model while updating only the BiLSTM parameters. This method is generally well-suited to accelerate the training process, even though it may lead to a slight sacrifice in performance.

Surprisingly, **for ESM2 models, this methodology yielded the best results, while for TAPE and ProtBert-BFD, it yielded the expected outcomes**.

Discussion

TAPE, ProtBert-BFD and ESM2

ProtBert-BFD got the worst result despite this model were pre-trained with the largest dataset BFD with 2122M samples. We believe, this result is caused by the noisy information and sequence mistakes in BFD dataset.

TAPE achieved the best results. TAPE models were pre-trained using the Pfam dataset, it is derived from UniProtKB and **selectively includes sequences belonging to Reference Proteomes rather than the entire UniProtKB**

ESM2(t6) achieved results that closely rival TAPE. ESM2(t6) comprises only 8 million parameters, compared to 92 million parameters of TAPE.

References I

[1] Cancer.net,
"Qué es la inmunoterapia," 2022.

[2] NortShore,
"Immunotherapy," 2022.

[3] NCI,
"National cancer institute dictionary," 2022.

[4] Elizabeth S Borden, Kenneth H Buetow, Melissa A Wilson, and Karen Taraszka Hastings,
"Cancer neoantigens: Challenges and future directions for prediction, prioritization, and validation,"
Frontiers in Oncology, vol. 12, 2022.

References II

[5] Ina Chen, Michael Chen, Peter Goedegebuure, and William Gillanders,
"Challenges targeting cancer neoantigens in 2021: a systematic literature review,"
Expert Review of Vaccines, vol. 20, no. 7, pp. 827–837, 2021.

[6] Qing Hao, Ping Wei, Yang Shu, Yi-Guan Zhang, Heng Xu, and Jun-Ning Zhao,
"Improvement of neoantigen identification through convolution neural network,"
Frontiers in immunology, vol. 12, 2021.

[7] Miao Peng, Yongzhen Mo, Yian Wang, Pan Wu, Yijie Zhang, Fang Xiong, Can Guo, Xu Wu, Yong Li, Xiaoling Li, et al.,
"Neoantigen vaccine: an emerging tumor immunotherapy,"
Molecular cancer, vol. 18, no. 1, pp. 1–14, 2019.

References III

[8] Xiaomei Zhang, Yue Qi, Qi Zhang, and Wei Liu,
"Application of mass spectrometry-based mhc immunopeptidome profiling in neoantigen identification for tumor immunotherapy,"
Biomedicine & Pharmacotherapy, vol. 120, pp. 109542, 2019.

[9] L Mattos, M Vazquez, F Finotello, R Lepore, E Porta, J Hundal, P Amengual-Rigo, CKY Ng, A Valencia, J Carrillo, et al.,
"Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the esmo precision medicine working group,"
Annals of oncology, vol. 31, no. 8, pp. 978–990, 2020.

References IV

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin,
"Attention is all you need,"
Advances in neural information processing systems, vol. 30, 2017.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova,
"Bert: Pre-training of deep bidirectional transformers for language understanding,"
arXiv preprint arXiv:1810.04805, 2018.

[12] Yaqi Zhang, Gancheng Zhu, Kewei Li, Fei Li, Lan Huang, Meiyu Duan, and Fengfeng Zhou,
"Hlab: learning the bilstm features from the protbert-encoded proteins for the class i hla-peptide binding prediction,"
Briefings in Bioinformatics, 2022.

References V

[13] Fuxu Wang, Haoyan Wang, Lizhuang Wang, Haoyu Lu, Shizheng Qiu, Tianyi Zang, Xinjun Zhang, and Yang Hu, "Mhcroberta: pan-specific peptide–mhc class i binding prediction through transfer learning with label-agnostic protein sequences," *Briefings in Bioinformatics*, vol. 23, no. 3, pp. bbab595, 2022.

[14] Yanyi Chu, Yan Zhang, Qiankun Wang, Lingfeng Zhang, Xuhong Wang, Yanjing Wang, Dennis Russell Salahub, Qin Xu, Jianmin Wang, Xue Jiang, et al., "A transformer-based model to predict peptide–hla class i binding and optimize mutated peptides for vaccine design," *Nature Machine Intelligence*, vol. 4, no. 3, pp. 300–311, 2022.

References VI

[15] Hans-Christof Gasser, Georges Bedran, Bo Ren, David Goodlett, Javier Alfaro, and Ajitha Rajan,
"Interpreting bert architecture predictions for peptide presentation by mhc class i proteins,"
arXiv preprint arXiv:2111.07137, 2021.

[16] Jun Cheng, Kaïdre Bendjama, Karola Rittner, and Brandon Malone,
"Bertmhc: improved mhc–peptide class ii interaction prediction with transformer and multiple instance learning,"
Bioinformatics, vol. 37, no. 22, pp. 4172–4179, 2021.

References VII

[17] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, et al.,
"Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences,"
Proceedings of the National Academy of Sciences, vol. 118, no. 15, 2021.

[18] 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, vol. 379, no. 6637, pp. 1123–1130, 2023.

Questions?