Supplementary Material

Model architecture

The input for the model is the AA as sequences. Each AA is represented as a separate token that is represented by a numerical vector(embedding), inspired by Natural Language Processing (NLP) techniques. Four options were tested for the AA representation: 1. Random - The AA were represented through 10-dimensions embedding vectors. The embedding is initialized randomly and learned in the training process. 2. Kidera - The AA were represented through 10-dimensions embedding vectors, initiated as their Kidera factors values, and learned in the training process. The Kidera factors are a set of properties of each AA, We assumed that starting from these representations would give an advantage for the embedding learning. 3. Kidera + biochemical - The AA were represented through 10-dimensions embedding vectors, initiated as their Kidera factors values, and learned in the training process. The biochemical properties of each AA are concatenated to the embedding output of each AA. 4. ESM-2/ESM-IF1 - each AA is represented in a dimension of 1280/512, respectively, of the pre-trained ESM models. ESM is an embedding representation that was learned by Large Language Model (LLM) on an extensive dataset, with impressive performances. And predict an embedding depending on the sequence.

The output is fed into BiLSTM or GCN. BiLSTM models have the ability to capture contextual information from both past and future tokens in a sequence. Using GCN we can represent the protein as a graph, and capture the 3D structure. Both, BiLSTM and GCN, allow the use of the whole protein to predict for each residue if it is within or outside an epitope, simultaneously for all residues. 1. BiLSTM - BiLSTM model with 2 layers. The BiLSTM produces an output for each token in the input in a dimension of 200 for the linear epitopes model or 20 for the conformational epitopes model. A dropout of 0.2/0.25 for conformational/linear epitopes is used. Each output is fed into the same MLP and reduced the dimension by half. 2. GCN - GCN model with 2 hidden layers. The output dimension of the first layer is 16, and 10 of the second layer. A node's features normalization is applied in both layers. A Relu activation and dropout of 0.2 is used. An output is produced for each node in the graph (i.e. for each AA).

The output of the BiLSTM/GCN is fed into MLP to reduce the output dimension of each AA to 1, and a Sigmoid function is applied.

We used Adam optimization, with a learning rate and L2-regularization of 0.001 for the conformational epitopes, and a learning rate of 0.001 and L2-regularization of 10^-6 for the linear epitopes.

Supplementary Tables

	AUC	BAC	MCC	PR-AUC
Kidera+bio up to 1023	0.76 ± 0.0041	0.69 ± 0.0037	0.14 ± 0.0045	0.12 ± 0.0037
kidera +bio	0.8 ±0.0039	0.72 ±0.004	0.14 ±0.002	0.12 ±0.0024
embedding	0.8±0.0023	0.73±0.004	0.14±0.002	0.12±0.004
embedding up to 1023	0.76±0.0042	0.69±0.004	0.14±0.0024	0.12±0.0049
kidera only	0.8±0.0045	0.72±0.0051	0.14±0.0032	0.11±0.0037
kidera up to 1023	0.76±0.0034	0.7±0.0037	0.14±0.0045	0.12±0.0037
ESM-2	0.77±0.001	0.7±0.0	0.14±0.0	0.12±0.004

Table S1: CALIBER Linear epitope prediction performance of BiLSTM models on the Validation dataset. Comparing the different embedding methods: aa embedding, ESM-2, Kidera embedding, Kidera embedding + biochemical properties.

Embedding	Model	AUC	BAC	MCC	PR-AUC
Random	BiLSTM	0.67 ± 0.0113	0.62 ± 0.0081	0.15 ± 0.0102	0.19 ± 0.0162
Kidera	BiLSTM	0.69 ± 0.0093	0.64 ± 0.0072	0.16 ± 0.0092	0.21 ± 0.0128
Kidera+bio	BiLSTM	0.71 ± 0.0101	0.65 ± 0.0078	0.18 ± 0.0099	0.23 ± 0.0121
ESM-2	BiLSTM	0.85 ± 0.0055	0.77 ± 0.0054	0.35 ± 0.0103	0.46 ± 0.0116
ESM-2	GCN	0.84 ± 0.0049	0.77 ± 0.0047	0.35 ± 0.0087	0.45 ± 0.0101
ESM-2	Boosting	0.85 ± 0.0019	0.78 ± 0.002	0.37 ± 0.0024	0.46 ± 0.0055
ESM-IF1	BiLSTM	0.85 ± 0.003	0.77 ± 0.0024	0.36 ± 0.0051	0.47 ± 0.0049
ESM-IF1	GCN	0.84 ± 0.0059	0.76 ± 0.0053	0.34 ± 0.0102	0.44 ± 0.0112
ESM-IF1	Boosting	0.84 ± 0.0007	0.76 ± 0.0	0.34 ± 0.0013	0.44 ± 0.0018

Table S2: CALIBER Conformational epitope prediction performance of BiLSTM, GCN, and boosting models on the Validation dataset. Comparing the different embedding methods: ESM-2 and ESM-IF1.

ESM	Model	AUC	BAC	MCC	PR-AUC
ESM-IF1+RSA	BiLSTM	0.769	0.69	0.26	0.34
ESM-IF1+RSA	GCN	0.768	0.69	0.26	0.32
ESM-IF1+RSA	Boosting	0.785	0.68	0.23	0.37
ESM-2+RSA	BiLSTM	0.762	0.69	0.27	0.31
ESM-2+RSA	GCN	0.749	0.69	0.25	0.29
ESM-2+RSA	Boosting	0.768	0.68	0.23	0.33

Table S3: CALIBER Conformational epitope prediction performance of BiLSTM, GCN, and boosting models on the Test dataset. Comparing the different embedding methods: ESM-2 and ESM-IF1 concatenated to the RSA.

Model	Encoding	Epitope	Threshold
BiLSTM	Random	Linear	0.1
BiLSTM	ESM-2	Linear	0.12
BiLSTM	ESM-2	Conformational	0.29
BiLSTM	ESM-IF	Conformational	0.21
BiLSTM	ESM-2	Both	0.15
GCN	ESM-2	Conformational	0.28
GCN	ESM-IF1	Conformational	0.21
Boosting	ESM-IF1	Conformational	0.3
Boosting	ESM-2	Conformational	0.3

Table S4: The selected threshold for each model, as it is calculated on the website.

RSA	CALIBER	BepiPred-3.0
0-0.2	0.729	0.683
0.2-0.4	0.705	0.676
0.4-0.6	0.684	0.65
0.6-0.8	0.701	0.699
0.8-1.0	0.711	0.708

Table S5: Performance comparison of CALIBER and BepiPred-3.0 on the blind test set Epitope3D, on different ranges of RSA.

Length	CALIBER	BepiPred-3.0
0-100	0.697	0.709
100-200	0.74	0.628
200-300	0.713	0.634
300-500	0.579	0.579
500-2000	0.76	0.745

Table S6: Performance comparison of CALIBER and BepiPred-3.0 on the blind test set Epitope3D, on different ranges of chain length.