

Viral protein language model

A I Y

0.1 ... 0.2 ... 0.01

Output
grammaticality
 $p(x_i | \mathbf{x}_{[N] \setminus \{i\}})$

Semantic
embedding $\mathbf{z}$

BiLSTM
layers

Input context
 $\mathbf{x}_{[N] \setminus \{i\}}$

V L S _ K A A

UMAP 2



UMAP 1

Analyze
semantic
landscape

VLSAKAA
VLSIKAA
VLSYKAA

Model viral
fitness



Use fitness and
semantic change
to predict escape