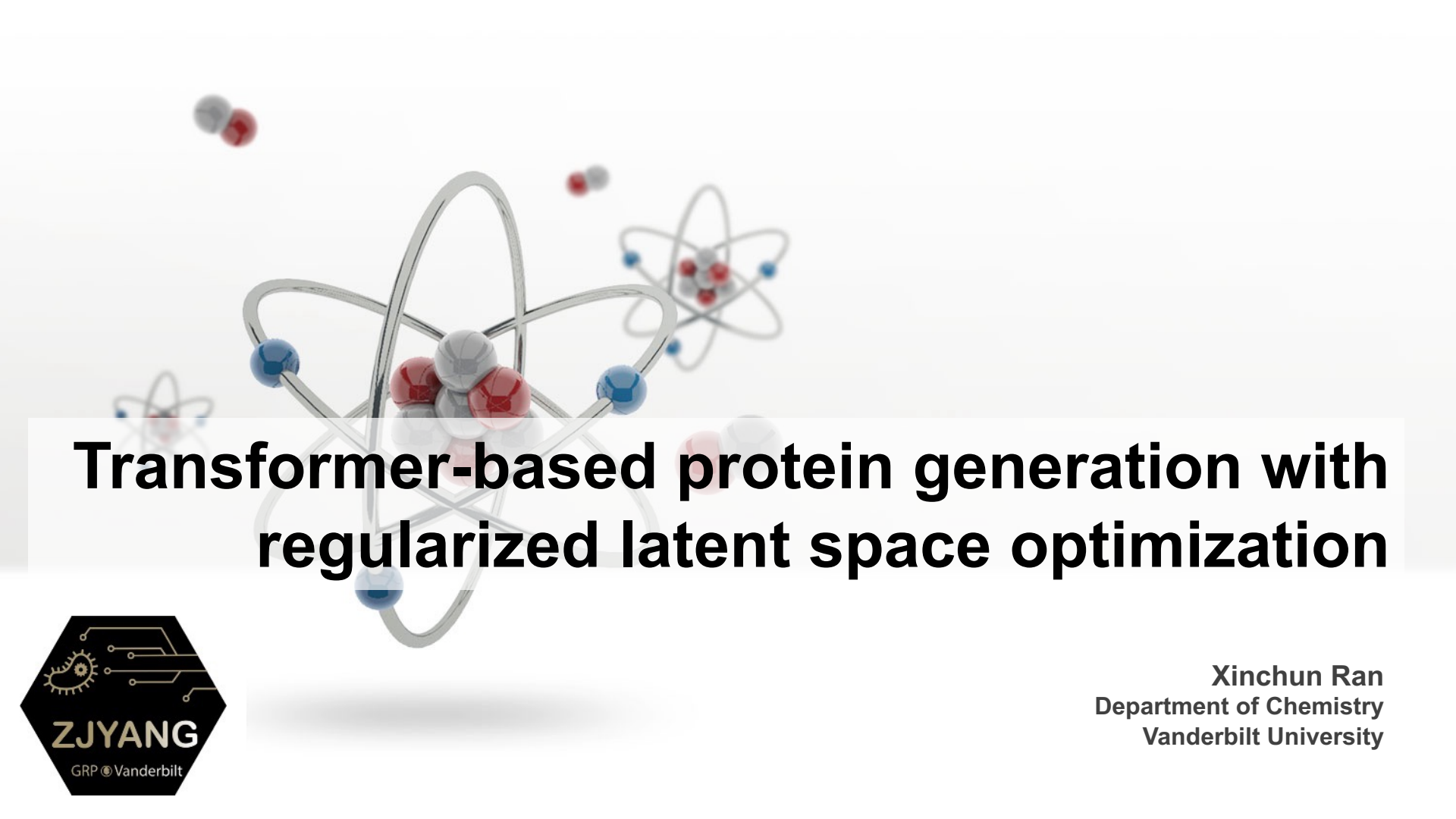
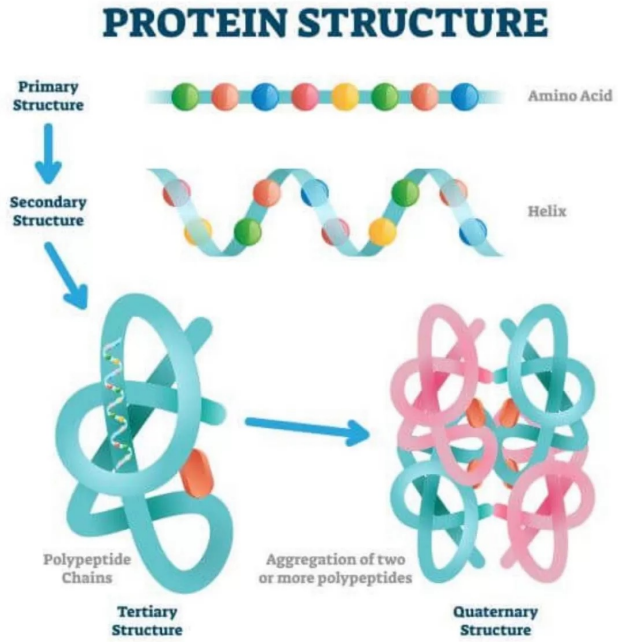


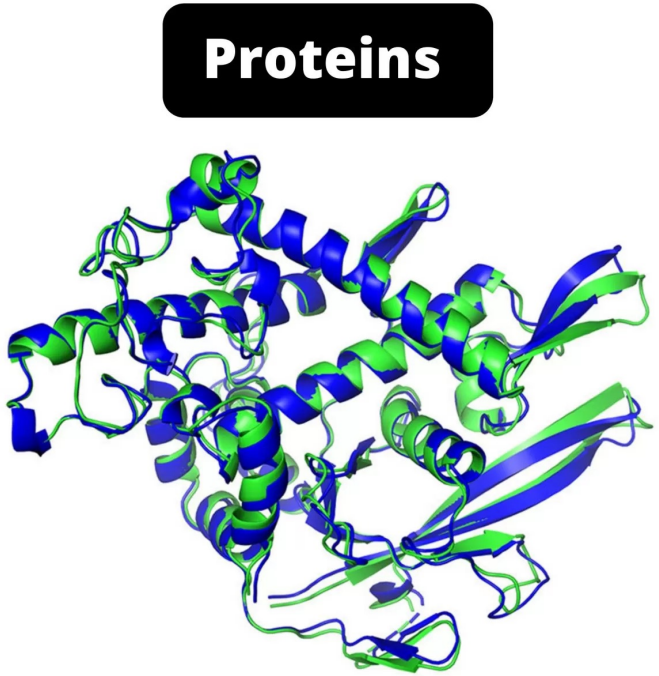
Transformer-based protein generation with regularized latent space optimization

Xinchun Ran
Department of Chemistry
Vanderbilt University

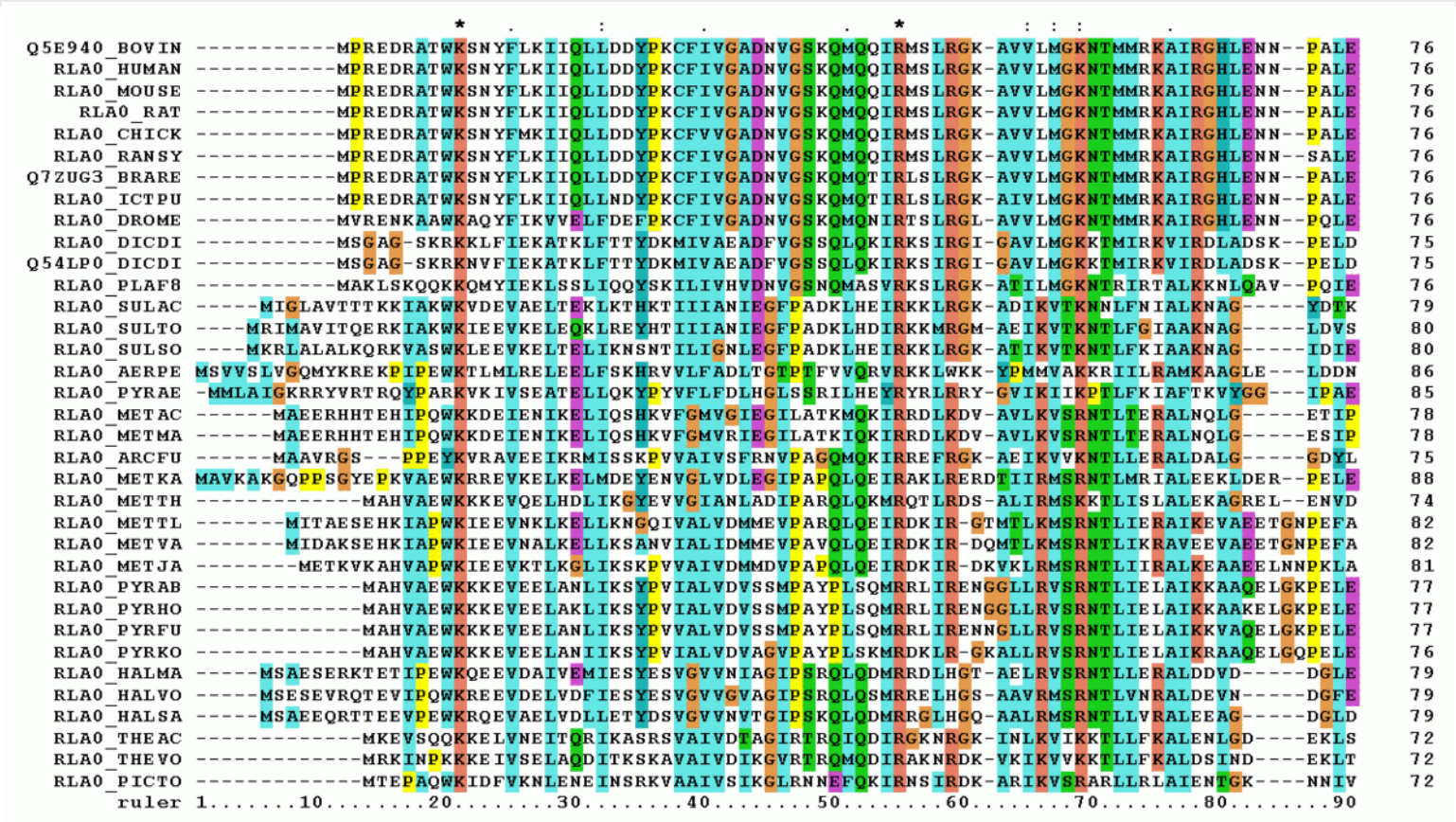
Background: Protein Sequences

The amino acid sequence of a polypeptide chain determines the final 3D structure of the protein

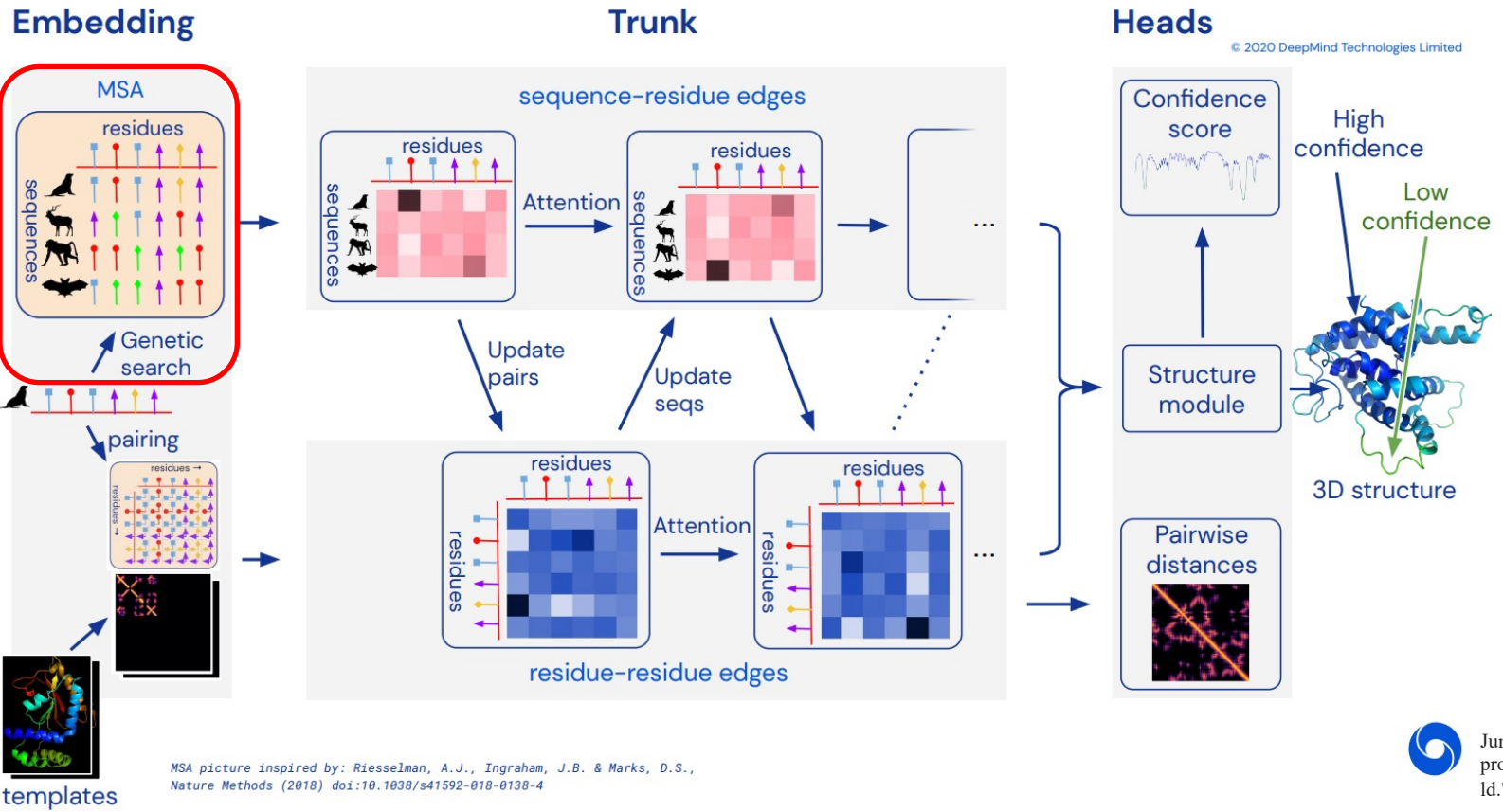




Background: Evolutionary information (MSA)

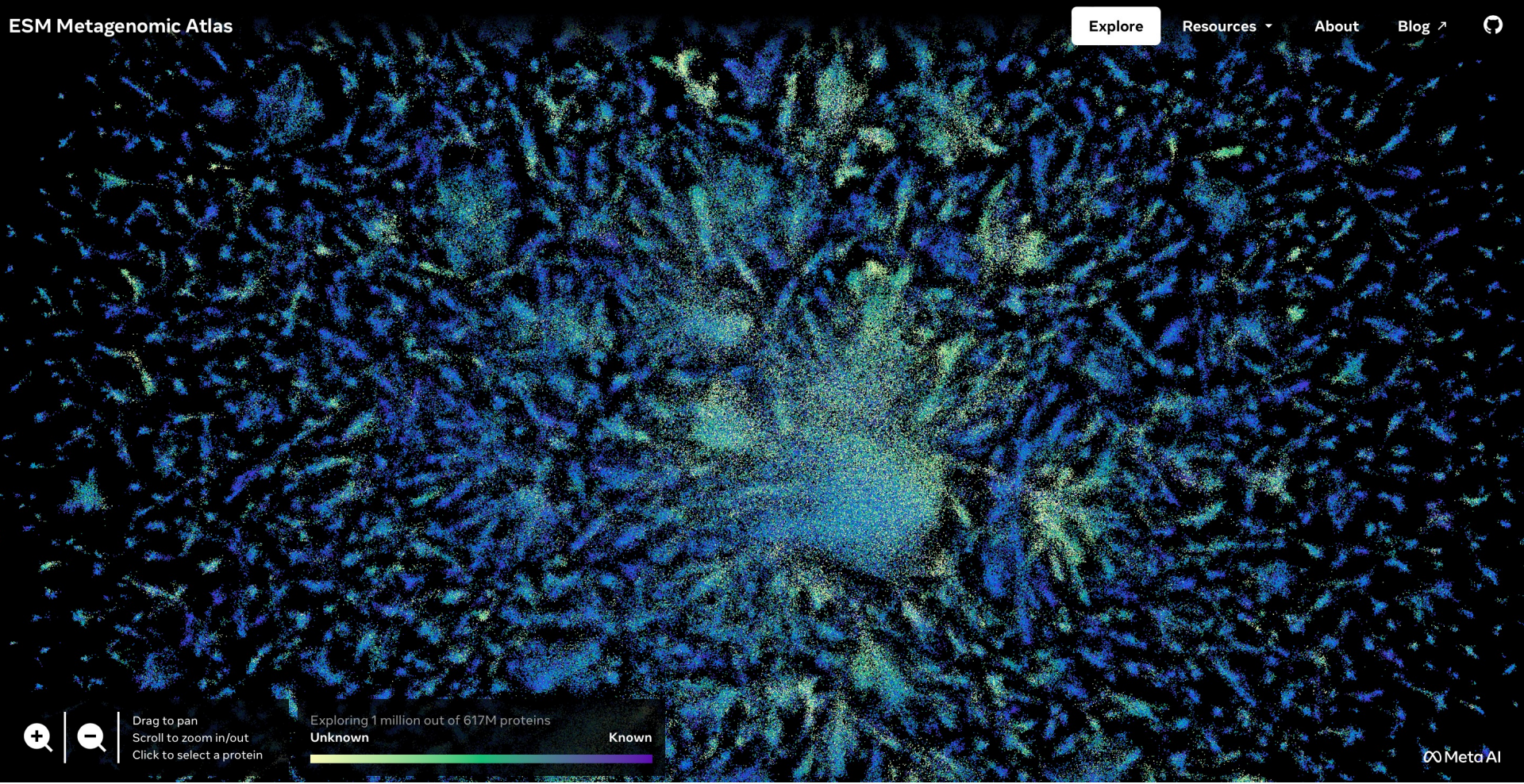


Background: Alphafold2 solutions?



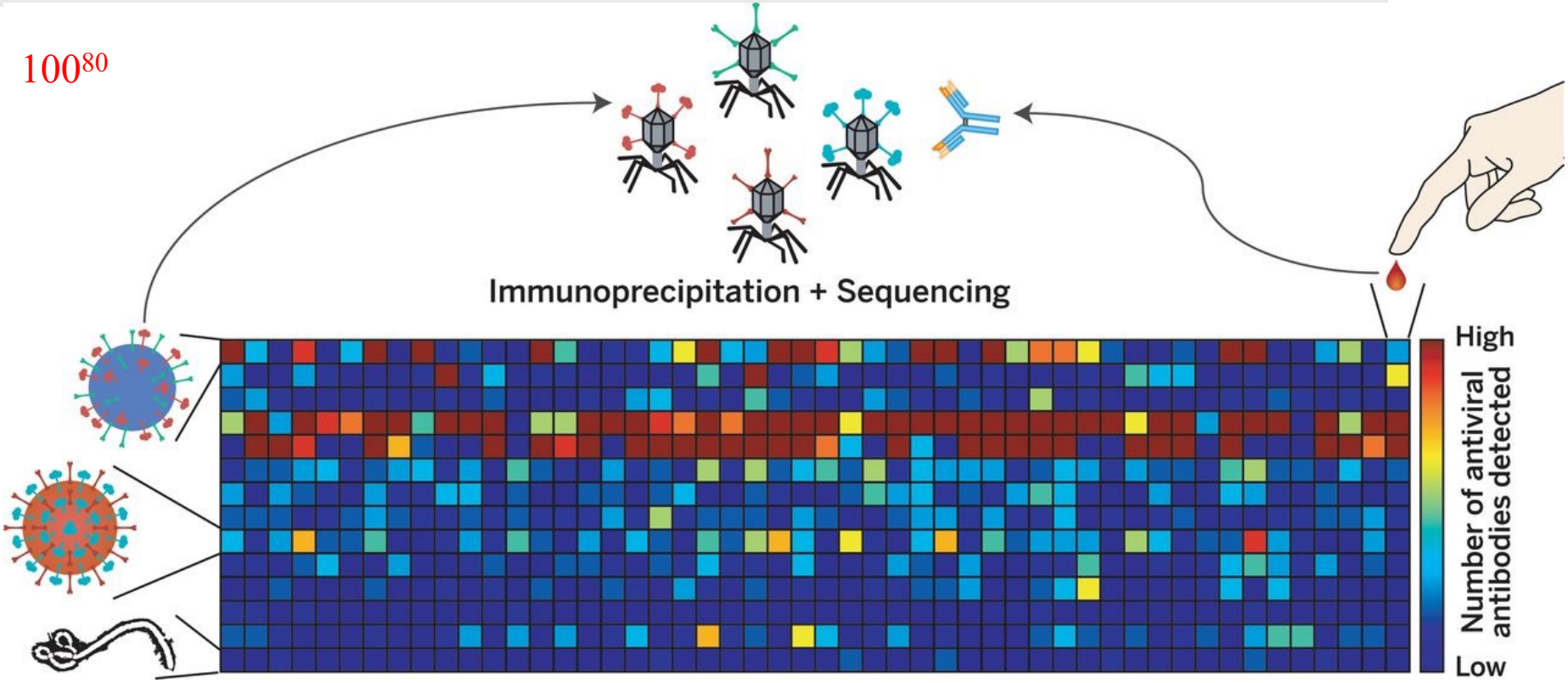
Not single sequence structure prediction

Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." *Nature* 596.7873 (2021): 583-589.

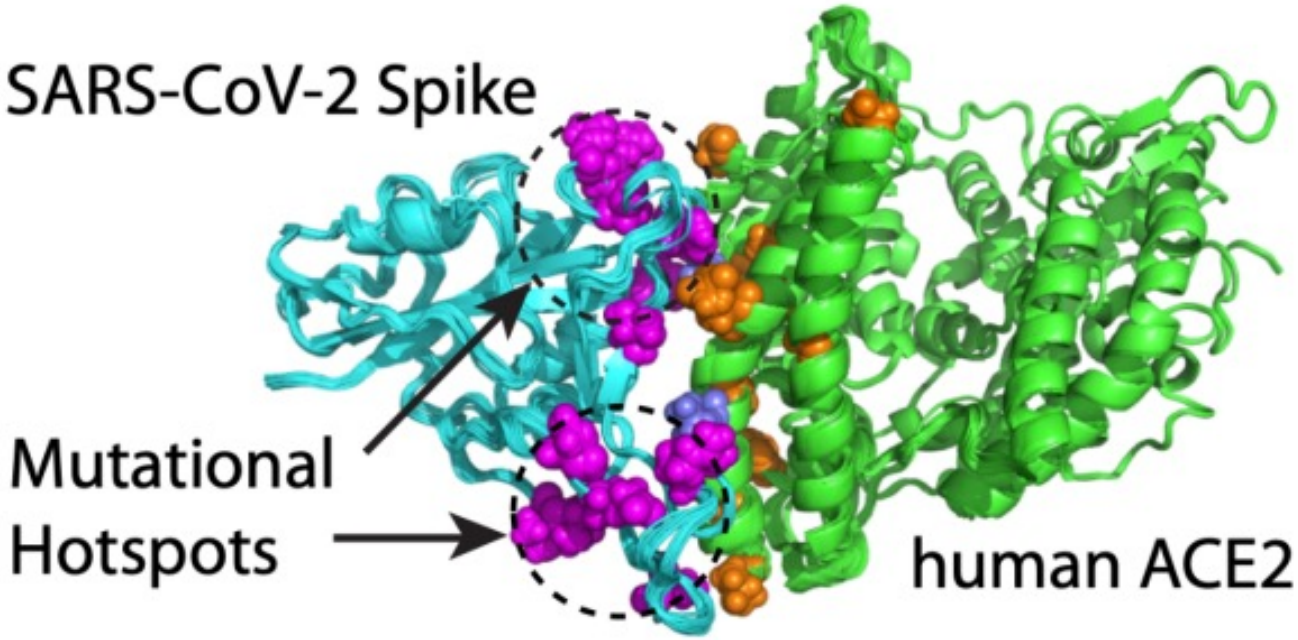

ESM Metagenomic Atlas
Explore
Resources
About
Blog
Drag to pan
Scroll to zoom in/out
Click to select a protein
Exploring 1 million out of 617M proteins
Unknown
Known
Meta AI

Real proteins sequence space

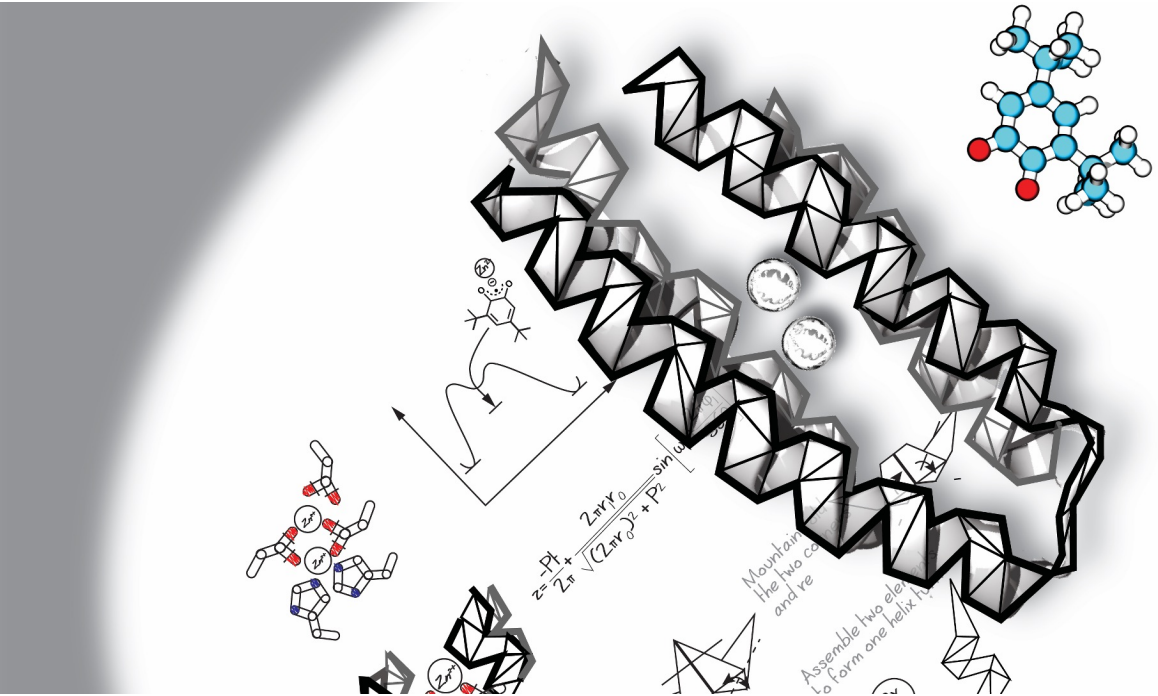
100^{80}



Traverse entire protein sequences space



Solutions: Protein Language model (PLM)



How to search the entire sequence space combination?

How can we guide the model goes to our desired folds/functions?