Published on Mar 10, 2025 by Arcadia Science **DOI**: 10.57844/arcadia-f52b-1451

Paired residue prediction dependencies in ESM2

During a quick analysis of the ESM2 model for masked token prediction, we noticed that amino acid probability distributions of residues affect each other in a pattern that mirrors a protein's 3D contact map. But less so for the larger model sizes. Our question to you is, why?

Contributors (A-Z)

Brae M. Bigge, Keith Cheveralls, Megan L. Hochstrasser, Evan Kiefl

Version 1 · Mar 31, 2025

This is a notebook pub stub!

We're experimenting with a new publishing format that we call a "**notebook pub**" **[1]**. Instead of coding and documenting our analysis in Python notebooks and then writing up a pub that contains all the same information with links out to GitHub, **we're turning the analysis into the pub**. We've developed a notebook pub template that renders the final content (narrative, code, tables, and figures) as a webpage and makes all the underlying code fully available. This means the entire product is completely reproducible. And we encourage you to reproduce it! Check out answers to FAQs on all of this, instructions on reproducing the pub, and info on how you can contribute.

In the future, we hope to host notebook pubs directly on PubPub. Until that's possible, we'll create stubs like this with key metadata like the DOI, author roles, citation information, and an external link to the pub itself.

View the notebook

The **full notebook pub** is available here.

The **source code** to generate it is available in this GitHub repo (DOI: 10.5281/zenodo.15002836).

References

1 Avasthi P, Bigge BM, Hochstrasser ML, Kiefl E, Roth R, Sabbagh U, York R. (2025). Closing the divide between analysis and publication: The notebook pub. https://doi.org/10.57844/ARCADIA-CA21-23BB