

# 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, Daniel Burns, Keith Cheveralls, Megan L. Hochstrasser, Evan Kiefl

*Version 3 · Dec 03, 2025*

## Purpose

This notebook details a quick analysis we performed on the ESM2 (Evolutionary Scale Modeling) models [1] for masked token prediction. We stumbled upon a counterintuitive result related to the effect that masking one residue has on the distribution of another.

Before jumping into our results, let's first establish what masked token prediction is, how it works, and our consequent motivation for this analysis.

The **full notebook pub** is available [here](#).

The **source code** to generate it is available in [this GitHub repo](#) (DOI: [10.5281/zenodo.15002836](https://doi.org/10.5281/zenodo.15002836)).

---

## References

- <sup>1</sup> Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, dos Santos Costa A, Fazel-Zarandi M, Sercu T, Candido S, Rives A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. <https://doi.org/10.1126/science.ade2574>
-