

## PEINT (PROTEIN EVOLUTION IN TIME)

Tech ID: 34416 / UC Case 2026-075-0

### PATENT STATUS

Patent Pending

### BRIEF DESCRIPTION

UC Berkeley researchers have developed a sophisticated computer-implemented framework that leverages transformer architectures to model the evolution of biological sequences over time. Unlike traditional phylogenetic models that often assume sites evolve independently, this framework utilizes a coupled encoder-decoder transformer to parameterize the conditional probability of a target sequence given multiple unaligned sequences. By capturing complex interactions and dependencies across different sites within a protein or genomic sequence, the model estimates the transition likelihood for each position. This estimation allows for a high-fidelity simulation of evolutionary trajectories. This approach enables a deeper understanding of how proteins change across different timescales and environmental pressures.

### SUGGESTED USES

»

Pathogen Tracking and Prediction: Modeling the future mutational landscape of viruses and bacteria to predict emerging strains and potential outbreaks.

»

Therapeutic and Vaccine Design: Identifying highly conserved or co-evolving sites to develop robust vaccines that remain effective against future evolutionary variants.

»

Enzyme Engineering: Simulating evolutionary pathways to discover novel mutations that enhance protein stability or catalytic activity for industrial applications.

»

Ancestral Sequence Reconstruction: Accurate computational inference of ancient proteins to study the origins of specific biological functions.

»

Drug Resistance Mapping: Predicting how cancer cells or pathogens might evolve in response to specific treatments, facilitating the design of more resilient therapies.

### ADVANTAGES

»

Captures Site Interactions: Successfully models "epistasis"—the interaction between different sites in a sequence—which is often ignored by simpler, site-independent models.

»

Handles Unaligned Sequences: Capable of processing unaligned biological sequences, reducing the heavy computational burden and potential errors associated with Multiple Sequence Alignment (MSA).

»

Continuous-Time Modeling: Integrates branch lengths (\$t\_k\$) directly into the transformer's probability estimations, allowing for modeling across arbitrary evolutionary distances.

»

Scalability and Speed: Leverages the parallel processing strengths of transformer architectures to analyze large-scale biological datasets more efficiently than traditional Markov Chain Monte Carlo methods.

»

High-Resolution Probabilistic Output: Provides precise likelihood estimates for specific transitions, offering a granular view of the evolutionary "fitness landscape."

### RELATED MATERIALS

### CONTACT

Laleh Shayesteh  
[lalehs@berkeley.edu](mailto:lalehs@berkeley.edu)  
tel: 510-642-4537.



### INVENTORS

» Song, Yun S.

### OTHER INFORMATION

#### CATEGORIZED AS

» **Biotechnology**

» [Bioinformatics](#)

» [Genomics](#)

» [Proteomics](#)

» **Computer**

» [Software](#)

» **Research Tools**

» [Bioinformatics](#)

» [Other](#)

#### RELATED CASES

2026-075-0