Comparisons of Bacterial Antiviral Homologs Found Across the Tree of Life

Phelps Tin (College 2026), Shirley Yuan, Yuri Silva, Andrew Santiago-Frangos

How Widespread Is the Bacterial Antiviral Protein Tiamat?

The Tiamat protein is an intriguing antiviral protein found in bacteria. This project explores its potential conservation across different domains of life. Tiamat proteins are crucial in bacterial defense mechanisms, countering viral phage infections. Given their importance, understanding Tiamat's structural and functional conservation across various life forms can provide significant insights into the evolution of antiviral mechanisms.

To contextualize the significance of the Tiamat protein, it's helpful to draw a parallel with Argonaute proteins, which are central to RNA interference and gene regulation. Now deemed promising candidates for directed nucleic acid manipulation and genome editing, Argonautes are highly conserved from archaea to humans. This project hypothesizes that Tiamat proteins may also exhibit a conservation pattern, suggesting a fundamental role in defense.

Workflow
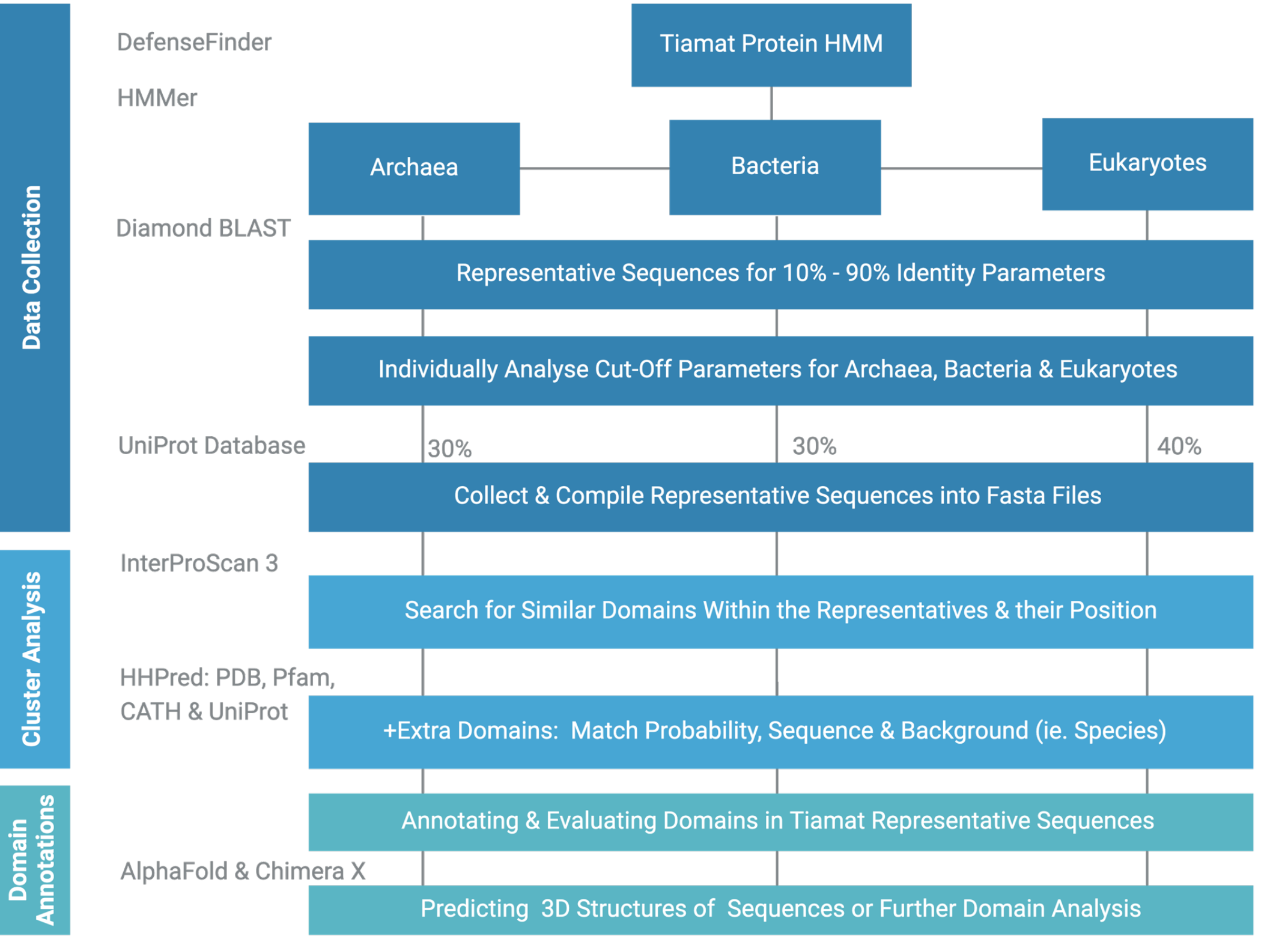
Determining Appropriate Percentage Identity Parameters for Clustering
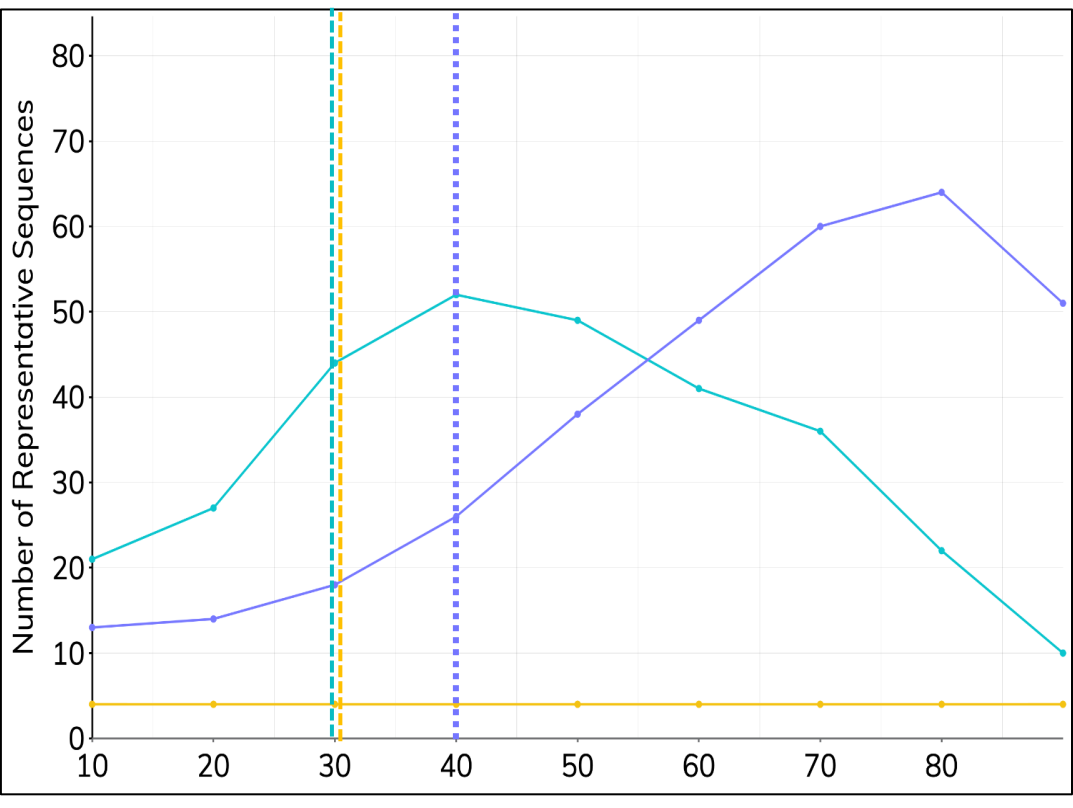

The number of representative sequences reflects the number of clusters. As the %Identity increases, fewer sequences are similar enough to form clusters. Clusters with only one sequence were filtered.

To balance representation and annotation practicality, the parameters of 30%, 30%, and 40% were chosen for Bacteria, Archaea, and Eukaryotes, respectively.

## Prominent Domains In Bacterial Tiamat Representative Sequences
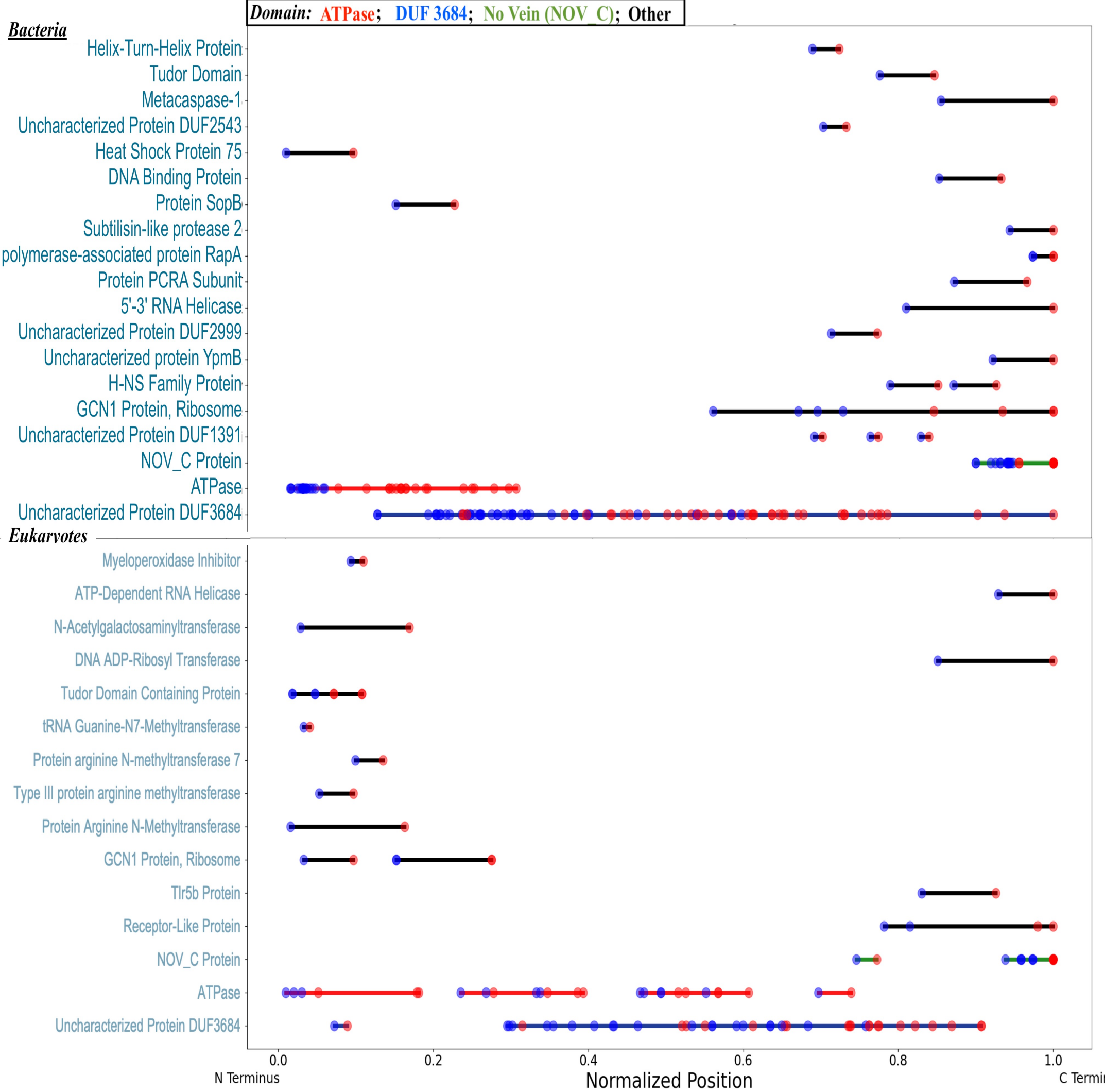
The above figures show the positions of different prominent Tiamat Protein functional domains in terms of a normalized position. Because amino acids differ between the sequences, the plot against normalized position displays a more evident trend of positioning between the 3 most common protein domains: ATPase, DUF3864, and NOV_C No vein.
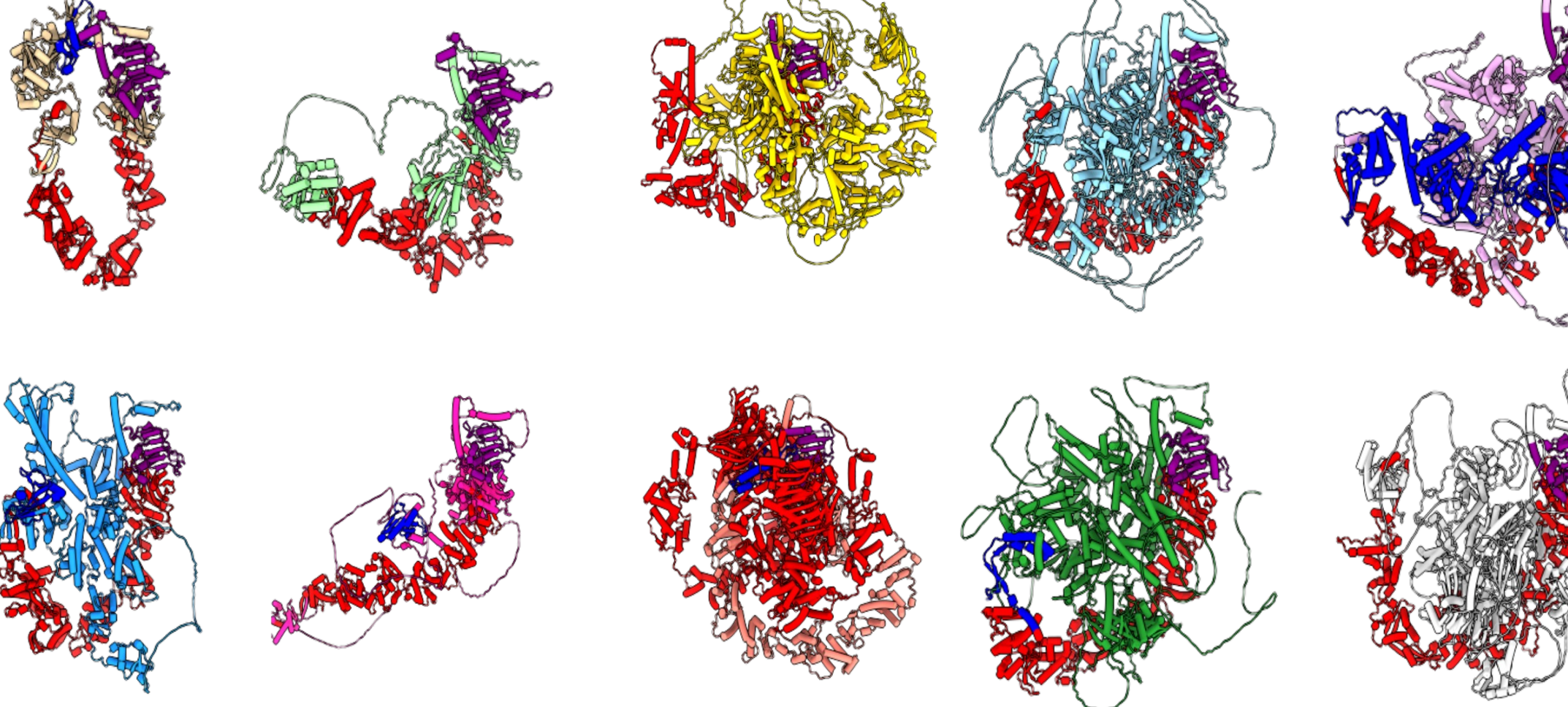

Predicted by AlphaFold and annotated with ChimeraX, the figure includes 3D models of the 10 Eukaryotic representative sequences. The purple residues are aligned and represent the ATPase domain. DUF3684 is red; No Vein is blue.

Distribution of Domains Along Sequences
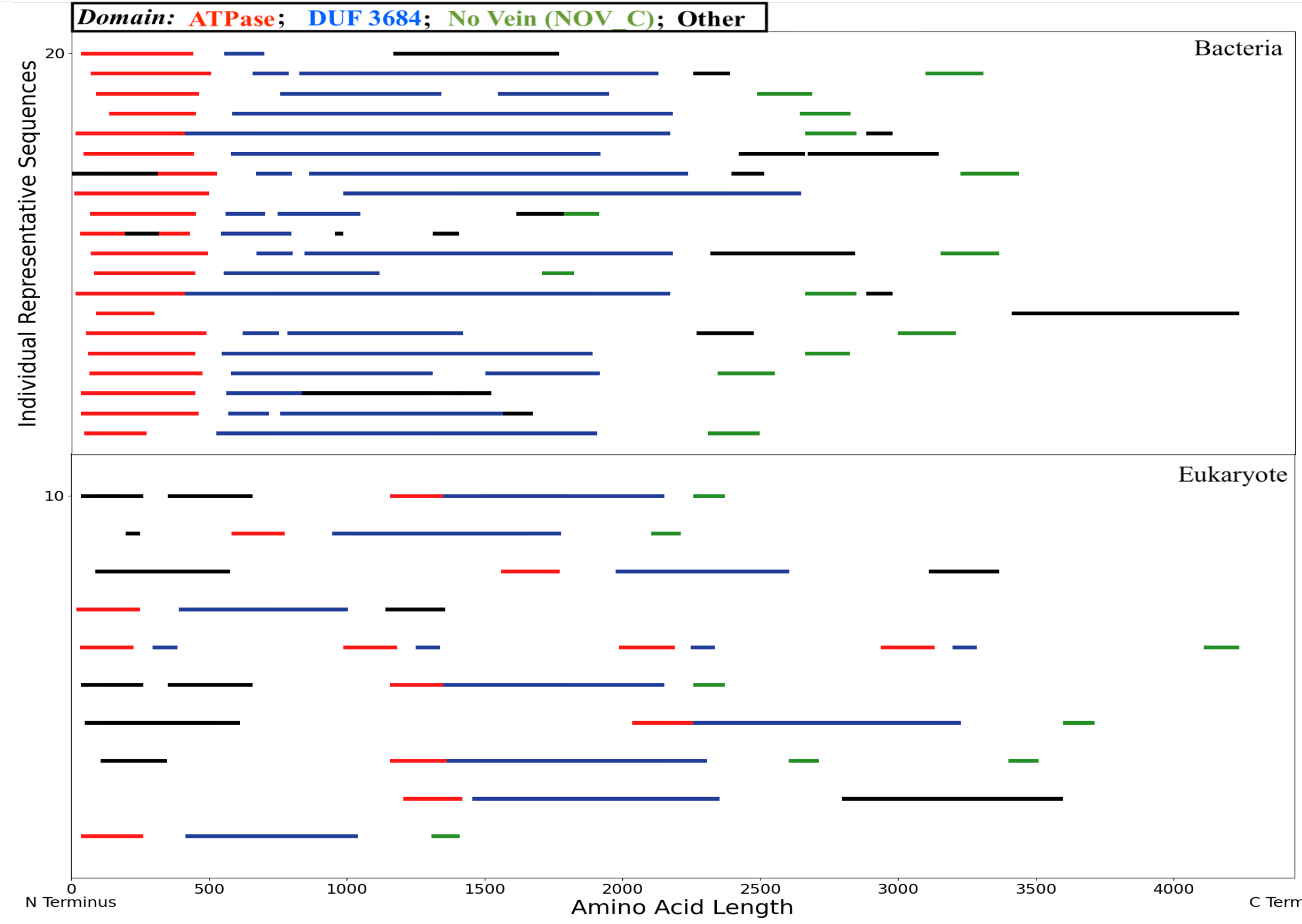
The graphs show the length and location of amino acids of the most shared domains of each representative sequence in Bacteria and Eukaryotes. A non-normalized position shows sequence length variation. The pattern of 3 domains: with ATPase to the N Terminus, DUF3864 spanning the midsection, and No Vein close to the C Terminus, reveals a conserved structure across Bacteria and Eukaryotes, suggesting significance to Tiamat's function and mechanism of action.
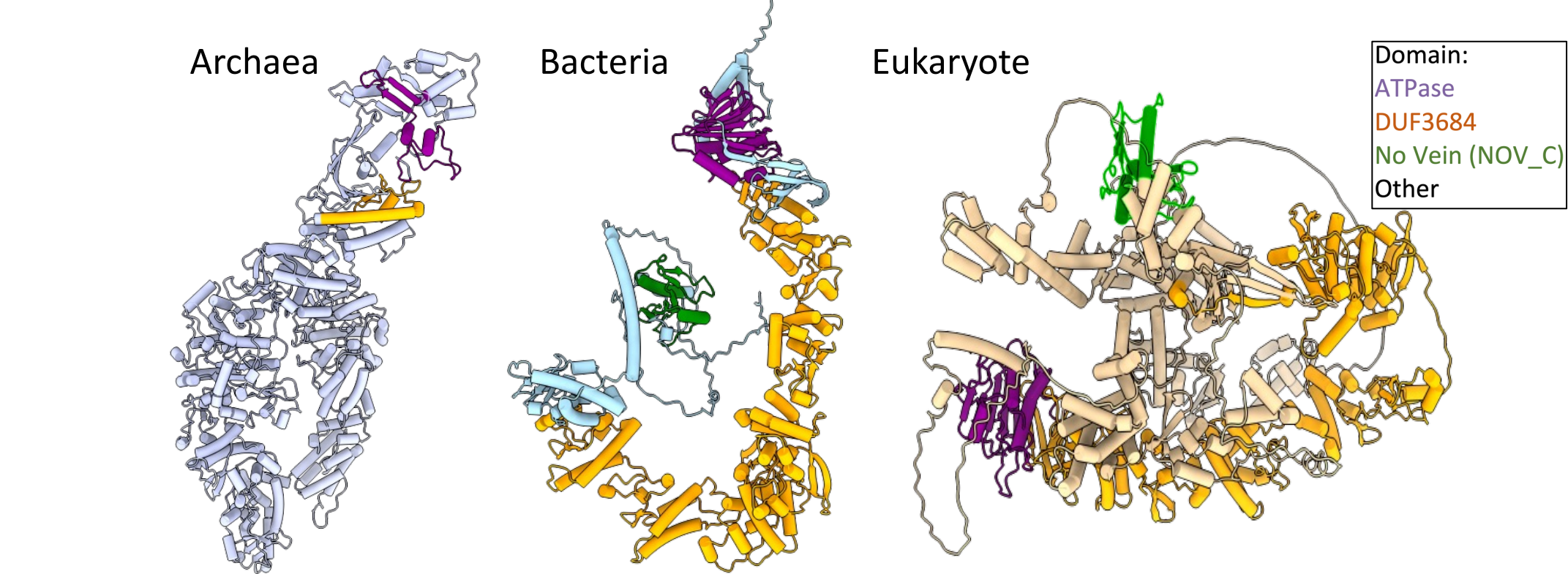
The figure above shows the 3D structures of the representative sequences with the largest cluster size in each phylogenetic domain. The 3 domains are highlighted

Conclusions

This workflow uses computational tools to provide a detailed analysis of Tiamat sequences. From these models and annotations, we can investigate the prevalence of Tiamat sequences across the Tree of Life and explore their mechanisms of action. The subsequent steps include analyzing the domains' functions to hypothesize the overall protein's function. Following the theoretical understanding, we begin using biochemical assays to test our hypotheses.

References & Acknowledgements