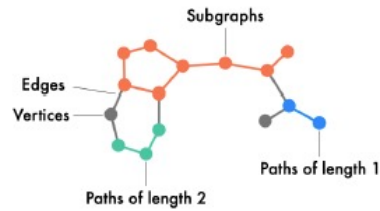


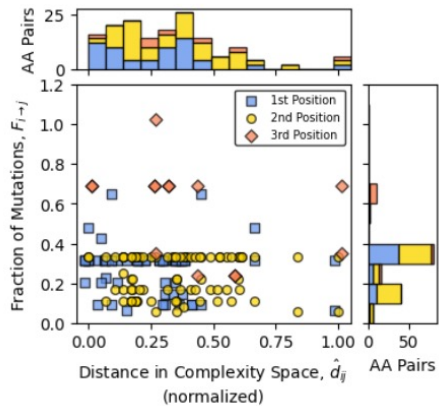
# Molecular Complexity Constrained Early Amino Acid Recruitment into the Genetic code

Chemical graph theory complexity

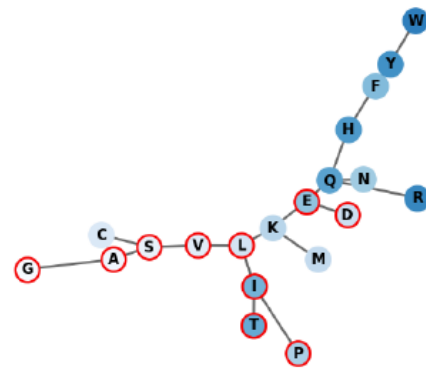


Representation of the two main frameworks used to quantify molecular complexity: A) graph-theoretic and B) information-theoretic approaches

Chemical information theory complexity



Normalized molecular complexity distance versus the fraction of possible single-point mutations between amino acid pairs. Colors and symbols indicate codon position: first (blue squares), second (yellow circles), and third (orange diamonds).



Minimum spanning tree (MST) based on 16 complexity metrics. Amino acids are colored from early (light) to later (dark) according to a proposed evolutionary timeline. Thick red edges indicate amino acids with prebiotic evidence. Edge lengths reflect original distances between connected nodes only.

Not all amino acid substitutions are equally likely. We compared the **20 proteinogenic amino acids** in a **multidimensional molecular complexity space**, and found that amino acids with similar structural complexity are more often connected by single-point mutations in the genetic code.

This relationship is especially strong for mutations occurring at the third codon position, where genetic changes are most readily tolerated.

The pattern suggests that **the genetic code is organized to favor substitutions that preserve overall molecular complexity**, reducing the likelihood of disruptive changes to protein structure and function.

These findings support the idea that **molecular complexity was not only a consequence of evolution but also a constraint that helped shape the pathways through which the genetic code evolved.**