**Introduction**

Biomolecules have enormous amount of topological information, which tightly connects with their biological functions. In this work, we reveal the topology-function relationship of biomolecules using the persistent homology analysis (PHA). We use fullerene molecules as an example and find out that our PHA is able to quantitatively predict fullerene stability. We introduce PHA to extract molecular topological fingerprints (MTFs) based on the persistence of molecular topological invariants, and utilize MTFs for biomolecular data analysis, including characterization, identification and analysis (CIA). We construct both all-atom and coarse-grained representations of MTFs. We further employ MTFs to characterize protein topological evolution during protein folding and quantitatively predict the protein folding stability. An excellent consistence between our persistent homology prediction and molecular dynamics simulation is found. Finally, we propose a multiscale, multiresolution and multidimensional persistent homology to match the resolution with the scale of interest so as to create a topological microscopy for the underlying data.

**Persistent homology theory**

Homology utilizes a topological space with an algebraic group representation to characterize topological features, such as isolated components, circles, holes and void. Persistent homology further embeds homology to match the resolution with the scale of interest so as to create a topological microscopy for the underlying data.

**Homology utilizes a topological space with an algebraic group representation to characterize topological features, such as isolated components, circles, holes and void.**

**Fundamental theorem of finitely generated abelian groups:**

\[ \mathbb{Z} \cong \mathbb{Z}^n \times \mathbb{Z}/2^m \times \mathbb{Z}/3^k \times \mathbb{Z}/p^l \]

(1)

- Betti number can be simply calculated by
  \[ \beta_i = \text{rank} H_i = \text{rank} Z_i - \text{rank} B_i \]

(2)

- **Filtration process:**
  \[ \varphi : K^0 \subseteq K^1 \subseteq \cdots \subseteq K^k = K \]

(3)

- **Persistence:** The p-persistent i-th homology group \( K_i \) is
  \[ H^p_i = \mathbb{Z}/(\text{Im} \varphi^p) \cap \mathbb{Z}/(\text{Ker} \varphi^p) \]

(4)

**Fullerene stability analysis**

Fullerene molecules have carbon-cage structures, which contain only pentagonal and hexagonal rings. Using the Victorita-Rips complex, the persistence of Betti numbers (i.e., ranks of homology groups), including \( \beta_0, \beta_1 \) and \( \beta_2 \), provides the information regarding length of the bond, width of the pentagon and hexagon ring, and size of the central void.

**Cryo-EM data analysis**

Cryo-electron microscopy (Cryo-EM) data is usually suffered from low signal to noise ratio (SNR). A geometric flow based denoising process is used to reveal the intrinsic topological invariants.

**Molecular topological fingerprints**

Persistent homology is used to extract MTFs based on barcode representation. MTFs are utilized for protein CIA.

**Protein folding analysis**

We use steered molecular dynamics to simulate the unfolding process of PDB 1UIQ. To analyze its topological evolution, PINs are calculated for every configurations and then stacked together in sequence to deliver a multidimensional persistence diagram.

**References**

1. Kelin Xia and Guo-Wei Wei, “Persistent homology analysis of protein structure, flexibility and folding”, IJNMBE, 30, 814-844 (2014)
2. Kelin Xia, Xin Feng, Yiying Tong and Guo-Wei Wei, “Persistent homology for the quantitative prediction of fullerene stability”, IJC, 36, 408-422 (2015)

**Acknowledgement**

This work was partially supported by NSF grants DMS-1140352 and HHMI 500283. NH Investigator R01GM060208, MCB Center for Mathematical Molecular Biosciences initiative and NTU SUG-M408142.310.