A Clustering Algorithm for Molecular Structures: Application on the Met-Enkephalin Peptide

Ruxi Xiang1, M. Jennifer Van1, Mahmoud Namazi2,3, Estela Blaisten-Barojas4,5,*, and Amarda Shehu1,6,*

1Dept. of Computer Science, 2Dept. of Mathematical Sciences, 3Dept. of Electrical and Computer Engineering, 4Computational Materials Science Center, and 5School of Physics, Astronomy, and Computational Sciences, 6Dept. of Bioengineering

George Mason University, Fairfax, VA 22030

[blaisten or amarda]@gmu.edu

Abstract

- Met-enk is an endogenous peptide that mediates pain and dependence on opioids [1].
- It flexes its structure to bind different opioid receptors.
- Wet-laboratory techniques have revealed a few structural states of met-enk [2].

- Research Objective: provide a comprehensive view of the structure space of met-enk through a variety of computational methods.
- Project team of two faculty and three undergraduate student researchers.
- My role: design of an algorithm to cluster computed structures of met-enk and so automate detection of thermodynamically-stable structural states.

Methodology

- Clustering algorithm adapts the known SPICKER algorithm used to cluster computed decoys in de novo protein structure prediction, [3]
- Original implementation uses a computationally-expensive dissimilarity measures obtained after optimal superimposition of two structures.
- Our adaptation is to integrate a fast, novel angular-based distance function.

References


Methodology Continued

Flowchart of algorithm

Distance function

\[ D(V_1, V_2) = \frac{1}{n} \sum_{i=1}^{n} |V_{1i} - V_{2i}| \]

Results

Clustering of Wet-lab Structures

Cluster 1 Cluster 2
15 5
Cluster 1 Cluster 2 Cluster 3
57 23 6

Clustering of Structures obtained via Basin Hopping and Rosetta energy function [5]

Cluster 1 Cluster 2 Cluster 3
4006 1629 708

Cluster 4 Cluster 5
847 843

Conclusion

- Clusters highlight met-enk populates diverse states
- Algorithm can cluster >10,000 structures in < 1 minute of CPU time.
- Algorithm is generally applicable for analysis of molecular structure data.
- Correspondence to be established to determine which wet-lab states are reproduced and which ones are novel and discovered in silico.