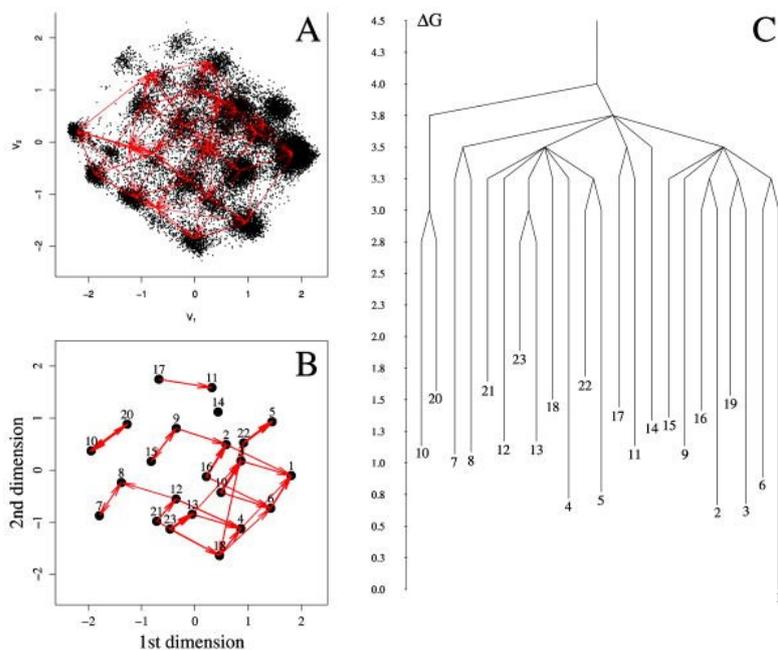# Προτεινόμενο Θέμα Διπλωματικής Εργασίας

glykos@mbg.duth.gr

Title : **"Automating the multidimensional clustering of data derived from a Principal Component Analysis of molecular dynamics trajectories using HDBSCAN."**

Aim : To write a short (or not so short) script in Python that will (a) read multidimensional data derived from a PCA analysis of a molecular dynamics trajectory (as produced by the programs carma/grcarma), (b) will use HDBSCAN[1] to cluster the data, and, (c) based on that clustering, prepare PDB files of the major clusters suitable for visualization with molecular graphics packages. Time permitting, a graph of the transitions between the clusters that HDBSCAN identified would be nice, like this one[2] :

---

1   https://medium.com/data-science/a-gentle-introduction-to-hdbscan-and-density-based-clustering-5fd79329c1e8
2   https://pubs.aip.org/aip/jcp/article/128/24/245102/934522/Construction-of-the-free-energy-landscape-of