

Homework Assignment

You will be given a set (~ 130) enzyme active sites to be clustered.

You will implement:

- A similarity metric between active sites

 - There's no “right” answer, but you should think hard about choosing something biologically meaningful.

- A partitioning algorithm to cluster the set of active sites

- A hierarchical algorithm to cluster the set of active sites

Homework Assignment

I will provide you with Python code which performs the I/O and contains some useful classes.

Answer the following questions

1. Explain your similarity metric, and why it makes sense biologically.
2. Explain your choice of partitioning algorithm.
3. Explain your choice of hierarchical algorithm.
4. Were your clustering results the same using both algorithms?
5. Did your clusterings have any biological meaning?

Turn in your Python code and question answers to me via email.

Due Date: 12:01am 5-5-04