

# Algorithms in Bioinformatics I, WS2002/3

## Assignment sheet # 14

Daniel Huson

January 27, 2003

### Finding the proper fold for two proteins.

The protein sequence for mysterin is:

```
>mysterin
ANDIQLNQKLSVDAKGAKQPIDTRAVNERYPHAVLFGGTCGGTIISP
TWILTAGHCTLFNDGRGVLGATNNSDVSGVYRFTKRLIIHPLFSVGPY
WLNAAEFNLKQVAARWDFLLAELEEPLPLDGKIMAAAKLDDQPDLPAG
LDVGYAGYGTDDHHGGTMRSEMAMELSVQSNEVCSKLEQFEAKDMLC
AKGRPPRYDSACNGDSGSLVDNNGRLVGVASWVENDAFECRNGNLVVF
SRVSSVREWIRQVTNI
```

The protein sequence for enigmin is:

```
>enigmin
DDTNAPLCLCDEPGILGRKEFVSDATKTIIEKAVEEVAKEGGVSGRGFSLF
SHHPVFRECGKYE CRTV RPEHSRCYNFPPIHFKSECPVSTRDCEPVFGYT
AAGEFRVIVQAPRAGFRQC VVWQH KCRYGSNNCGFNCRCTQQRSVVRLVTFNL
EKNGLFCETFR TCCG CPCR SF
```

Both sequences are contained in the file:

[www-ab.informatik.uni-tuebingen.de/teaching/ws02/abi1/assign/prot.fasta](http://www-ab.informatik.uni-tuebingen.de/teaching/ws02/abi1/assign/prot.fasta)

Your assignment is to evaluate several programs in finding the proper fold for these proteins.

## 1 Run two threading programs and determine the fold (4 points)

Run at least two threading programs to see if you can determine the proper fold.

For example, use

- 123D, or
- TOPITS.

The output from these programs should list (and is often hotlinked to) the coordinate files of the structures listed as hits. Try to look at the structures of the hits, or look at their sequence entries in SwissProt to see what family they belong to. Do you see hits that cluster into one family? Do the hits correspond to similar structures? Note that multiple hits in one family are more reliable than hits in a variety of families. Assign each protein to a family based on threading analysis.

## 2 Run additional analyzes (4 points)

Run at least two other analysis from the list below for the two proteins, for example:

- BLAST version 2,
- BLOCKS, or
- PhD.

How do the results from the threading vs. sequence/secondary structure-based searches compare?  
To what families would you assign mysterin and enigmin?

## 3 Multiple alignment (2 points)

Construct a multiple sequence alignment of either protein with several homologs.

**Due by 10am, Monday, Feb 3, 2003.**

(Source: <http://instruct1.cit.cornell.edu/courses/biobm732>)