

Handout on Pattern Discovery in Protein Sequences

Katerina Kechris

November 27, 2001

Review Article

Brazma, Jonassen, Eidhammer & Gilbert (1998) *Journal of Computational Biology* 4:279-305.

PRATT

Jonassen, Collins, & Higgins (1995) *Protein Science* 4:1587-1595.

<http://www.ebi.ac.uk/pratt/>

TEIRESIAS

Rigoutsos & Floratos (1998) *Bioinformatics* 14:56-67.

<http://www.research.ibm.com/bioinformatics/teiresias.html>

SPLASH

Califano (2000) *Bioinformatics* 16:341-357.

<http://www.research.ibm.com/splash/>

EMOTIF

Nevill-Manning, Wu & Brutlag (1998) *PNAS* 95: 5865-5871

<http://motif.stanford.edu/emotif/>

EM ALGORITHM

Lawrence & Reilly (1990) *Proteins* 7:41-51.

MEME

Bailey & Elkan (1995) *Machine Learning Journal* 21:51-83.

<http://meme.sdsc.edu/meme/website/>

GIBBS MOTIF SAMPLER

Lawrence, Altschul, Bogouski, Liu, Neuwald & Wooten (1993) *Science* 262: 208-214.

Liu, Neuwald & Lawrence (1995) *Journal of the American Statistical Association* 432: 1156-1170.

Neuwald, Liu & Lawrence (1995) *Protein Science* 4: 1618-1632.

<http://bayesweb.wadsworth.org/gibbs>

PROTEIN DATABASES

Links to various databases:

<http://www.ebi.ac.uk/interpro/databases.html>

The sequences for the myoglobin example can be obtained from:

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>

Scop database:

<http://scop.mrc-lmb.cam.ac.uk/scop/>

Pfam database:

Protein family alignments based on HMM's. The alignments can be saved in Fasta format with no gaps for input into the motif algorithms.

<http://www.sanger.ac.uk/Software/Pfam/>