

Hidden Markov models provide a probabilistic model of protein domain or family data. Databases of Hidden Markov Models can be extremely useful tools for the analysis of sequence data. While the Pfam database is unquestionably the most popular collection of HMM's in existence today, there are many other useful collections that are currently available, and some of them might be more suitable for a particular need. In addition, there are several versions of Pfam, and it is useful to know which version to use, and why.

In this tutorial we will examine the varieties of software methods that are currently available, such as:

Profile HMM methods:

HMMER

S.R. Eddy, WashU St. Louis.

SAM

R. Hughey, K. Karplus, A. Krogh, D. Haussler and others, UC Santa Cruz.

META-MEME

B. Grundy, UC San Diego.

SPSPfam - Optimized version of HMMPfam

G. Montry, Southwest Parallel software

DeCypherHMM - accelerated HMM implementation

TimeLogic Corporation

Wise2

E. Birney

Other available profile searching implementations:

PFTOOLS

P. Bucher, ISREC Geneva.

Implementation of "generalized profiles", which are essentially profile HMMs.

GCG PROFILES

M. Gribskov.

Part of the GCG software package.

Selected PSSM search implementations:

PSI-BLAST

S. Altschul et al., NCBI.

MEME

T. Bailey, C. Elkan, W. Grundy.

PROBE

A. Neuwald, J. Liu, C. Lawrence.

We will study many different HMM databases, such as the various flavors of Pfam, TIGRfam, SMART, Superfamily, Tlfam, COGfam, KINfam, and related resources such as the Conserved Domain Database. Specific tips and pointers for optimizing results from each one will be shared. Finally, we will look at methods for the development of customized collections of HMMs, with the associated benefits and pitfalls.