

Prediction of coiled-coil domains.

Mauro Delorenzi and Terry Speed

Genetics and Bioinformatics, The Walter & Eliza Hall Institute,
Parkville Melbourne, VIC 3050 Australia.

Email: delorenzi@wehi.edu.au

The selection of proteins with certain characteristics from genomic sequences is a central goal of computational biology and data-mining.

Proteins with coiled-coil domains are of interest for molecular biologists studying a variety of processes such as protein transport and membrane fusions and the infection of cells by parasites, [1], [2]. Coiled-coils are also motifs that mediate oligomerization and protein-protein interactions in many systems, f.ex. transcription factors [3].

We tested the performance of the two most commonly used tools (Coils and Paircoil) and our own new program Marcoil for the prediction of coiled-coil domains from a protein sequence.

Marcoil is based on a hidden Markov model (HMM) and was trained in two different ways: the traditional Maximum Likelihood (ML) method and the Conditional Maximum Likelihood method (CML, [4]).

In the context of genomic data mining, the Sensitivity and Precision with which coiled-coil domains can be identified are the relevant features of diagnostic quality.

{Sensitivity = $TP / (TP + FN)$; Precision = $TP / (TP + FP)$ }.

A comparison of the Sensitivity versus Precision curves on our testset suggests that the ML and CML methods optimise performance in different segments of the curve, and that the HMM has a slight advantage over the traditional recognition methods, which use a fixed-length window and position specific scoring matrices (with correlation terms in Paircoil).

While all programs easily identify the subgroup of typical long coiled-coil structures and usually miss the short and hydrophobic coiled-coil domains, their performance differs on hydrophilic domains of intermediate length. The advantage of the HMM method probably derives from the flexibility it permits in domain lengths.

References:

[1] Skehel J.J., Wiley D.C. (1998)

Coiled Coils in Both Intracellular Vesicle and Viral Membrane Fusion. *Cell* 95, 871-874.

[2] Chan D.C., Kim P.S. (1998)

HIV Entry and Its Inhibition. *Cell* 93, 681-684.

[3] Glover J.N.M., Harrison S.C. (1995)

Crystal Structure of the heterodimeric bZIP transcription factor cFos-cJun bound to DNA. *Nature* 373, 681-684.

[4] Krogh A, Riis SK, (1999)

Hidden Neural Networks. *Neural Computation* 11, 541-563.