Related Programs

Here is a brief list of other programs for sequence pattern searches I am aware of, some of them using PROSITE as their source of protein patterns. This list is certainly not complete or exhaustive and restricts itself to non-commercial products. Most of the commercial products such as GCG or PC/Gene also offer routines for pattern matching, and some of them can also use the PROSITE database.

Pattern Searching

Prosearch, by Lee Kolakowski.
Protein pattern searching with the PROSITE database. Unix, VMS, DOS. Requires the "awk" program which is available as public domain software for most computer systems.

Scrutineer, by Peter Sibbald (Sibbald et al., 1991).
A sophisticated program for sequence pattern searches in protein databases. Allows the identification of patterns whose complexity goes well beyond the PROSITE syntax. PROSITE patterns can be used by Scrutineer in a converted form. VMS, Unix.

Overseer, by Peter Sibbald (Sibbald et al., 1992).
Similar to Scrutineer, but for DNA sequences. VMS.

Signal Scan, by Dan Prestidge (Prestridge, 1991).
A program to identify known eukaryotic signals in DNA sequences (can use TFD). Unix, VMS.

Motif and Pattern, by Kerry Cockwell and Ian Giles (Cockwell & Giles, 1989).
Two programs for simple pattern searching. Cannot use PROSITE. Unix, VMS, DOS.

Promot, by Michael Sternberg (Sternberg, 1991).
Pattern searching with the PROSITE database. VMS.

Block Searching

The only other programs besides MacPattern for searching protein sequences with site-specific scoring matrices derived from the BLOCKS database I currently know of are

PATMAT, by James Wallace and Steven Henikoff (1992).
Available from ncbi.nlm.nih.gov. UNIX, DOS.

PATMAT is very flexible and also allows pattern searches and conventional sequence similarity searches against protein sequences and translated nucleotide sequences.

BlockSearch, by Rainer Fuchs (1993a).
A derivative of MacPattern for block searches. VMS, OpenVMS, UNIX, (maybe DOS).

Statistical Analysis

A comprehensive package for statistical analyses of protein sequences is

SAPS, by Volker Brendel (Brendel et al., 1992).
Available for UNIX systems from V. Brendel (volker@gnomic.stanford.edu).

Others

The TFD database, by David Ghosh.
Contains all sorts of information about transcription factors on the protein and DNA level. A relational database which is distributed as ASCII table dumps and in ASN.1 format. Certain tables of the TFD database can be used by Signal Scan and MacPattern (after running the filter program tfd2prosite on them). A special electronic discussion forum was recently established for TFD. For more information contact ghosh@ncbi.nlm.nih.gov.

plsearch, by Randall and Temple Smith.
A complex program to identify unknown sequence pattern in groups of protein sequences.

Most programs mentioned here can be obtained from the EMBL File Server (NetServ@EMBL-Heidelberg.DE) or ftp.embl-heidelberg.de