Motif search from a FASTA file

# Required software and scripts

Jalview – to view and align FASTA sequences  
Perl – required to run motif search scripts, get Strawberry Perl 64-bit  
ScanProsite ps\_scan – Perl script for motif searching, make sure to download the win32 zip file

# Things you need to do motif search

A collection of your protein sequences in a FASTA file  
Your motif in a specially formatted dat file (see next section)

# Motif and corresponding dat file

A motif pattern looks like this: {C}(12)-C-{C}(1,12)-C-{C}(1,12)-C-C-{C}(1,12)-C-{C}(1,12)-C-{C}(12).  
Which corresponds to this: \*\*\*\*\*\*\*\*\*\*\*\*C^^^^^C^^^^^CC^^^^^C^^^^^C\*\*\*\*\*\*\*\*\*\*\*\*  
(\*) is anything but cysteine  
(C) is cysteine  
(^^^^^) is anything but cysteine for 1 to 12 residues in length  
For example this sequence satisfies the above criteria: SSNSSSFEDVNPCPICLHPVNEEAYLDCCFHKFCYSCIEHWAMLVAKRH

A collection of motif in a dat file will look like this:  
//

ID 6C\_HLP; PATTERN.

AC M06001;

PA {C}(12)-C-{C}(1,12)-C-{C}(1,12)-C-C-{C}(1,12)-C-{C}(1,12)-C-{C}(12).

//

ID 6C\_HLP; PATTERN.

AC M06002;

PA {C}(12)-C-{C}(1,12)-C-{C}(1,12)-C-{C}(1,12)-C-C-{C}(1,12)-C-{C}(12).

//

ID 6C\_HLP; PATTERN.

AC M06003;

PA {C}(12)-C-C-{C}(1,12)-C-{C}(1,12)-C-{C}(1,12)-C-{C}(1,12)-C-{C}(12).

//

# How to search

Run Perl (command line), navigate to the folder with ps\_scan contents, your FASTA file, and your motif dat file.  
Execute the motif search command:   
ps\_scan.pl -d motifs.dat my\_sequences.fa -o fasta >result.fa

ps\_scan.pl is the script, motifs.dat contain your motifs, my\_sequences.fa is the FASTA file of all your sequences, result.fa will be the result file in FASTA created by the program after it finishes executing