PSIPRED RELEASE NOTES

PSIPRED Version 2.3

By David Jones, January 2002

Here are some very brief notes on using the PSIPRED V2 software.

PSIPRED is supplied in source code form - it must be compiled before it can be used. The code should compile on any ANSI C compiler e.g. the GNU C compiler.

Please see the LICENSE file for the license terms for the software. Basically it's free to anyone (including commercial users) as long as you don't want to sell the software or, for example, store the results obtained with it in a database and then try to sell the database. If you do wish to sell the software or use it in a commercial product, then please contact Inpharmatica Ltd. (http://www.inpharmatica.co.uk).

PSIPRED is run via a tcsh shell script called "runpsipred" - this is a very simple script which you should be able to convert to Perl or whatever scripting language you like.

If your sequence does not have any homologues in the current data banks, then it is possible to run PSIPRED on a single sequence. In this case, PSIPRED is run via a tcsh shell script called "runpsi_single". Unfortunately, like every other secondary structure prediction method, PSIPRED does not perform well on single sequences. Any secondary structure prediction based on a single sequence should be considered as unreliable.

Before running PSIPRED, please check the runpsipred and runpsi_single scripts to see if the path variables are set to wherever you have installed the program and data files. The default is to assume that the program is installed in the current directory - this is probably NOT what you want!

INSTALLATION

Firstly compile the software:

```
tcsh% cd to-wherever-you-untarred-PSIPRED

tcsh% cd src

tcsh% make
```

```
tcsh% make install

The executables will be placed in the PSIPRED bin directory.

You must also install the PSI-BLAST and Impala software from the
NCBI toolkit, and also install appropriate sequence data banks.

The NCBI toolkit can be obtained from URL ftp://ftp.ncbi.nih.gov

PSI-BLAST executables can be obtained from ftp://ftp.ncbi.nih.gov/blast


EXAMPLE USAGE
=============

In this example the target sequence is called "example.fasta":

tcsh% runpsipred example.fasta

Running PSI-BLAST with sequence example.fasta ...
Predicting secondary structure...
Pass1 ...
Pass2 ...
Cleaning up ...
Final output file: example.horiz
Finished.

That's it - you can then look at the output:

tcsh% more example.horiz


SPECIAL OPTIONS
===============

The psipass2 program has several special options which you can use if you wish.

For example, the default command is as follows:

psipass2 weights_p2.dat 1 0.98 1.09 output.ss2 input.ss > output.horiz

Arguments 2,3 & 4 are as follows:

Argument 2: No of filter iterations
This controls the amount of "smoothing" that is carried out on the final
prediction. The recommended setting is 1, but it may be worth trying
higher values to increase the level of smoothing.
```

Argument 3&4: Helix/Strand Decision constants
These options control the bias for helix (Arg3) and strand (Arg4) predictions. The default values are close to 1.0, but if you know your protein is, for example, mostly comprised of beta strands then you can increase the bias towards beta strand prediction. For example:

```
psipass2 weights_p2.dat 1 1.0 1.5 output.ss2 input.ss > output.horiz
```

increases the bias towards beta strand prediction by approximately 50%.

SEQUENCE DATA BANK

It is important to ensure than the sequence data bank used with PSI-BLAST has been filtered to remove low-complexity regions, transmembrane regions, and coiled-coil segments. If this is not done, then it is essential that the PSI-BLAST output for the target sequence is checked by-eye to ensure that no spurious sequences have been included in the PSI-BLAST alignment. A program called "pfilt" is included which will filter FASTA files before using the formatdb command to generate the encoded BLAST data bank files.

For example:

```
tcsh% pfilt nr.fasta > filtnr

tcsh% formatdb -t filtnr -i filtnr

tcsh% cp filtnr.p?? $BLASTDB
```

(note that the above command assumes you have already set the BLASTDB environment variable to the directory where you usually keep your BLAST data banks)

CHANGES FROM THE ORIGINAL PSIPRED

The following is a quick summary of the main changes since the original PSIPRED.

1. The program now makes use of PSI-BLAST binary checkpoint files (using the Impala program makemat) to reduce loss of precision when parsing the original ASCII position specific matrices.

2. By default the 1st pass uses an average of 4 different neural network

weight sets - this improves prediction accuracy slightly.

3. In addition to the normal horizontal summary output format, the program now also produces a full table of results which shows the individual coil, helix, strand network outputs.

4. A one-line header is output at the start of the output files to allow THREADER 3 (and other programs) to automatically recognise a PSIPRED prediction.