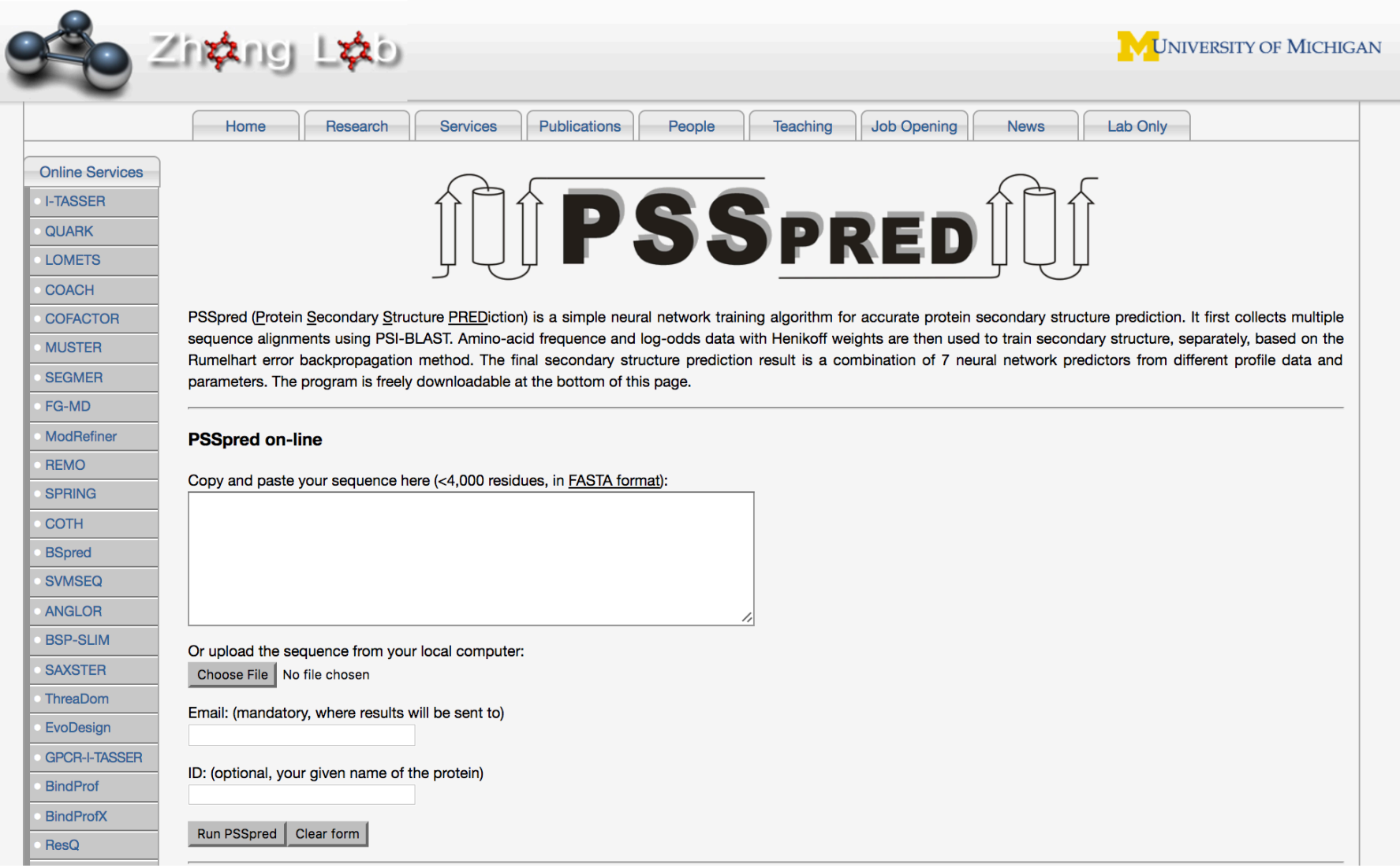
Machine Learning in Structure Biology (Practicals)

Yang Zhang

Department of Computational Medicine and Bioinformatics,
Department of Biological Chemistry

University of Michigan

PSSpred: http://zhanglab.ccmb.med.umich.edu/PSSpred/



Practical-I: Protein secondary structure prediction

1, please generate secondary structure by on-line server of PSSpred for the following protein sequence:

MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRVKHLKTEAEMKASEDLKK
HGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQLEL
GAMNKAFRKDIAAKYKELGYQG

2, download the standalone PSSpred program and generate secondary structure prediction for the following protein sequence:

AILLPVEGAQLSELRQIPAEGGPVLHLRLDSPQFSQFGEIYFSEVLPRRVKAWKRHSLTQLFAVPVGCI
HVVLYDGREKSPTSGRLAQVTLGRPDNYRLLRIPPQVWYGFAATGDTPALVANCTDIPHRQGESERAP
QDAPFIPFSWAGADLSGT

Output of PSSpred

PSSpred results for T29162

```
seq:    1 LPDYTNIGLYEGDWVWLKKFPGDRHIAIRPATKMAFSKIRELRHENVALYLGLFLAGGAG  60
SS:       CCCCCCCCEEECCCEEEEEEECCCCCHHHHHHHHHHHHHHHHCCCCCCEEEEEEECCCCC
conf:     998763179829189999956997278999999999999837871213677798447887

seq:   61 GPAAPGEGVLAVVSEHCARGSLQDLLAQRDIKLDWMFKSSLLLDLIKGIRYLHHRGVAHG 120
SS:       CCCCCCCCCCEEEEEECCCCCCHHHHHHHCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCC
conf:     765578995899996499999789997169999989999999999999999984898778

seq:  121 RLKSRNCVVDGRFVLKVTDHGHGRLLEAQRVLPEPPSAEDQLWTAPELLRDPVLERRGTL 180
SS:       CCCCCCEEEECCCEEEEEECCCCCCEEEECCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCC
conf:     887531366198079996377121551564224666543555688788618988889885

seq:  181 AGDVFSLGIIMQEVVCRSAPYAMLELTPEEVVKRVQSPPPLCRPSVSIDQAPMECIQLMK 240
SS:       CHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCHHHHHHHH
conf:     222457889999999789999998899899999998588889999998789999999999

seq:  241 QCWAEQPELRPSMDRTFELFKSINKGRKMNIID 273
SS:       HHHHCCHHHCCCHHHHHHHHHHHHHCCCCCCCC
conf:     999629444999999999999965126899899
```

SS: C- coil, H-helix, E-strand

Confidence score 1-9

- Download PSSpred prediction in FASTA format: seq.SS
- Download PSSpred prediction in I-TASSER format: seq.dat
- Download the original confidence file: seq.dat.ss

Original output file

If you use PSSpred in your publication, please cite:
Y. Zhang, http://zhanglab.ccmb.med.umich.edu/PSSpred

'seq.dat.ss' file

273			coil	helix	beta
1	L	C	0.960	0.022	0.014
2	P	C	0.923	0.030	0.048
3	D	C	0.900	0.035	0.075
4	Y	C	0.845	0.050	0.120
5	T	C	0.789	0.042	0.185
6	N	C	0.658	0.026	0.316
7	I	C	0.534	0.009	0.469
8	G	E	0.149	0.006	0.839
9	L	E	0.053	0.005	0.939
10	Y	E	0.071	0.011	0.913
11	E	C	0.599	0.012	0.382
12	G	C	0.940	0.009	0.050
13	D	C	0.531	0.009	0.447
14	W	E	0.108	0.003	0.904
15	V	E	0.056	0.002	0.950
16	W	E	0.028	0.002	0.977
17	L	E	0.025	0.002	0.978
18	K	E	0.024	0.002	0.973
19	K	E	0.050	0.003	0.949
20	F	E	0.234	0.005	0.764
21	P	C	0.812	0.009	0.182
22	G	C	0.946	0.022	0.030

Three steps to install PSSpred

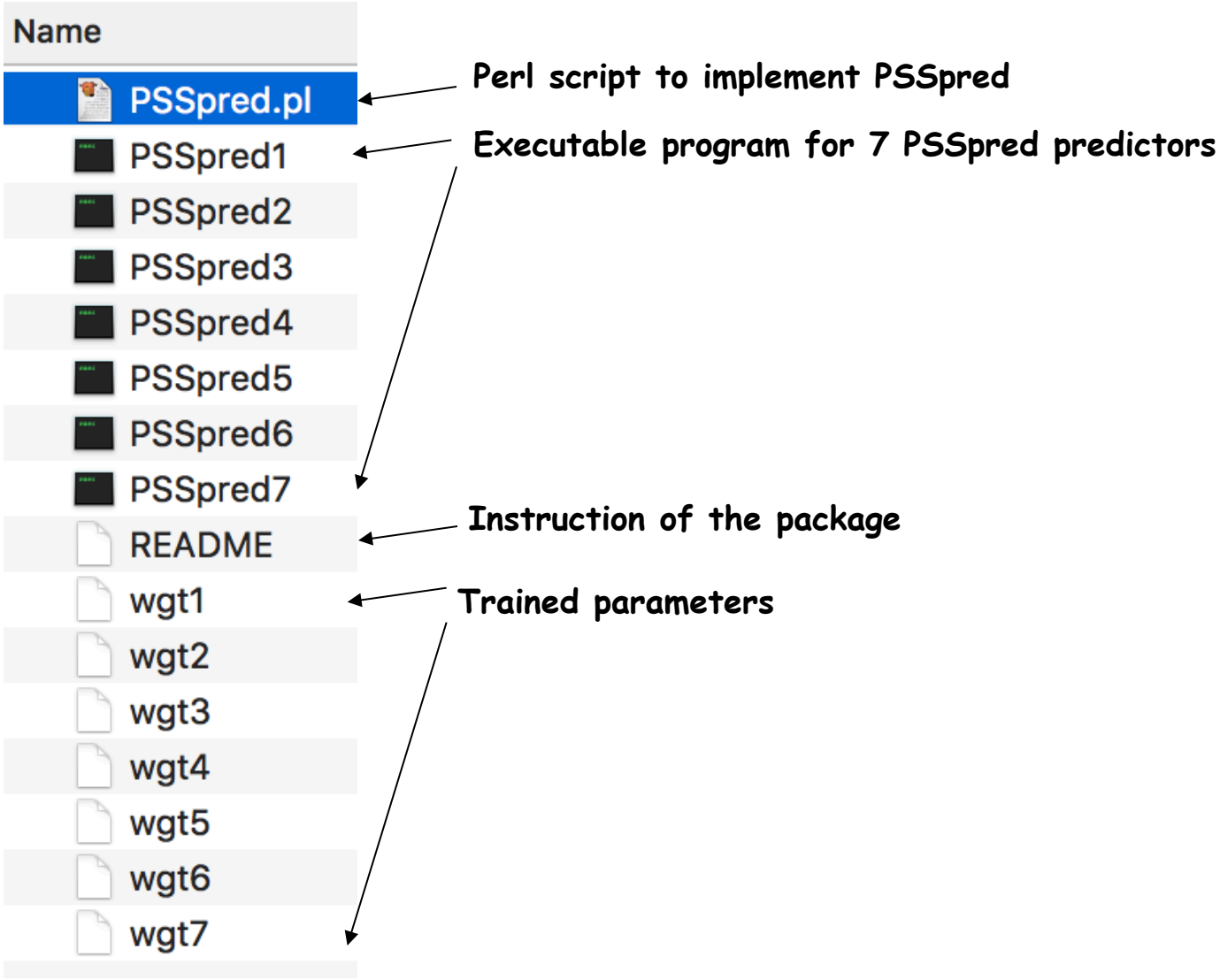
http://zhanglab.ccmb.med.umich.edu/PSSpred/

- Click PSSpred_v3.tar.gz to download the current version of the PSSpred package (the programs can only implemented in 64 bit Linux machines).
    - 1, unpack the PSSpred files by "tar -zxvf PSSpred_v3.tar.gz".
    - 2, Run PSSpred by "PSSpred.pl seq.fasta". An instruction can be found at the head of the enclosed "PSSpred.pl" file.
    - 3, PSSpred needs following external files. If you do not have them installed in your computer, you can download from here:
        - Click nr.tar.gz to download non-redundant sequence database.
        - Click blastv2.6.tar.gz to download NCBI PSI-BLAST programs.

Considering the limitation of speed, please also download the nr database from:

ftp://ftp.ncbi.nlm.nih.gov/blast/db/

What are included in 'PSSpred_v3.tar.gz'



What will you see in the readme file?

1, how to install PSSpred?

a, download 'PSSpred_v3.tar.gz' at
http://zhanglab.ccmb.med.umich.edu/PSSpred/

b, unpack the PSSpred files by "tar -zxvf PSSpred_v3.tar.gz"

c, download and install non-redundant sequence file at
http://zhanglab.ccmb.med.umich.edu/cgi-bin/download_ftp.cgi?ID=nr.tar.gz

d, download and install psi-blast program at
http://zhanglab.ccmb.med.umich.edu/PSSpred/blastv2.6.tar.gz

e, change the path ($blastdir, $db, $PSSpreddir) in 'PSSpred.pl'

Important

2, how to run 'PSSpred.pl' in Linux system?

>PSSpred.pl seq.txt

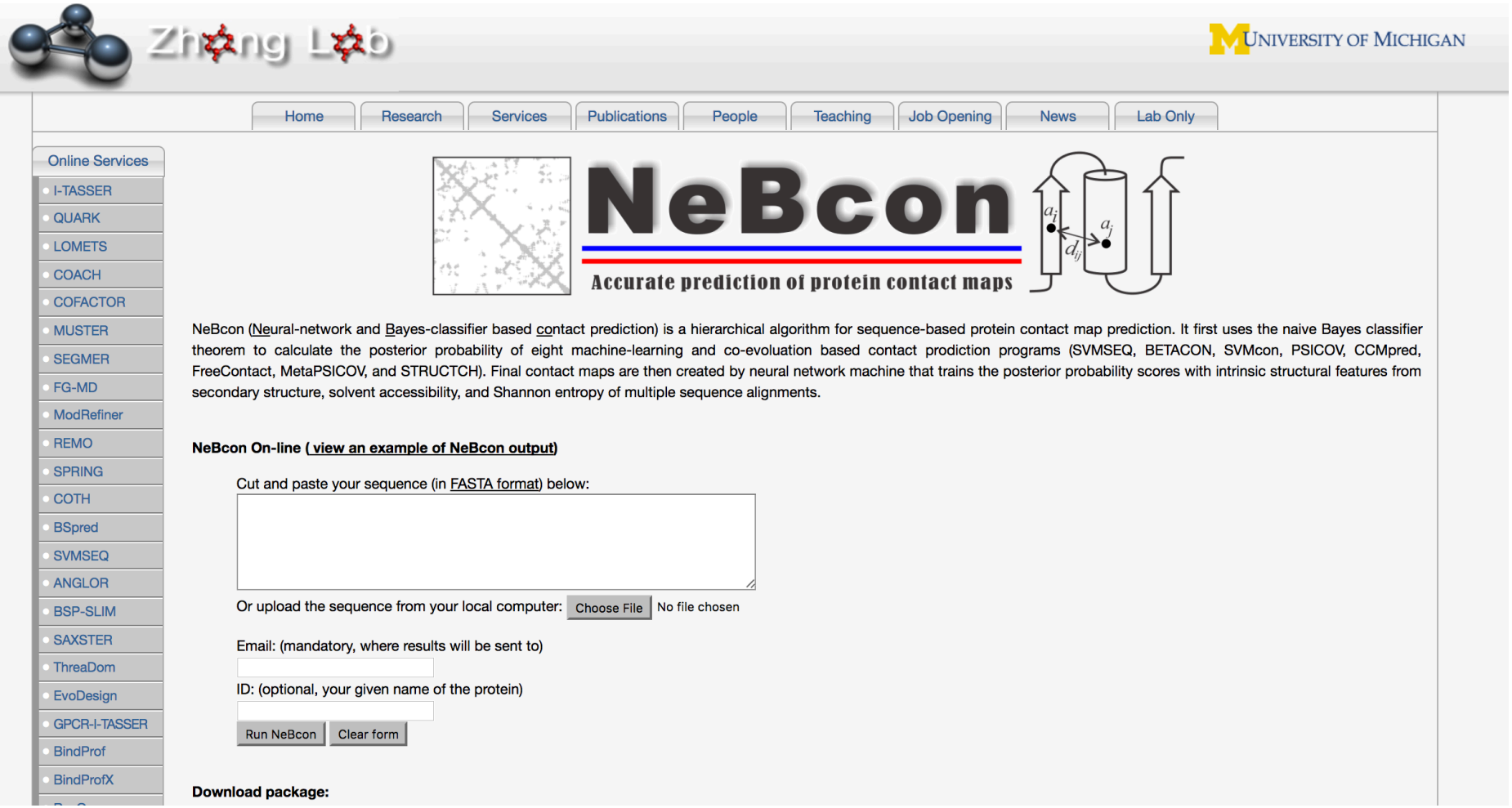
Note:

a, seq.txt is fasta file at current directory (the only input file)

b, output file:
seq.dat
seq.dat.ss

c, PSSpred.pl is split into three steps:
Step 1: prepare and run PSI-BLAST
Step 2: prepare mtx, pssm.txt, profw, freqccw, freqccwG
Step 3: run PSSpred and output files

NeBcon: http://zhanglab.ccmb.med.umich.edu/NeBcon/



Example output: http://zhanglab.ccmb.med.umich.edu/NeBcon/EXAMPLE/

Practical-II: Protein contact prediction

1, please generate contact prediction by on-line server of NeBcon for the following protein sequence:

MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRVKHLKTEAEMKASEDLKK
HGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQLEL
GAMNKAFRKDIAAKYKELGYQG

2, download the standalone NeBcon program and generate contact prediction for the following protein sequence:

AILLPVEGAQLSELRQIPAEGGPVLHLRLDSPQFSQFGEIYFSEVLPRRVKAWKRHSLTQLFAVPVGCI
HVVLYDGREKSPTSGRLAQVTLGRPDNYRLLRIPPQVWYGFAATGDTPALVANCTDIPHRQGESERAP
QDAPFIPFSWAGADLSGT