

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/>

 **Zhang Lab** 

Home Research Services Publications People Teaching Job Opening News Lab Only

Online Services

- I-TASSER
- QUARK
- LOMETS
- COACH
- COFACTOR
- MUSTER
- SEGMER
- FG-MD
- ModRefiner
- REMO
- SPRING
- COTH
- BSpred
- SVMSEQ
- ANGLOR
- BSP-SLIM
- SAXSTER
- ThreaDom
- EvoDesign
- GPCR-I-TASSER
- BindProf
- BindProfX
- ResQ



PSSpred (Protein Secondary Structure PREDiction) is a simple neural network training algorithm for accurate protein secondary structure prediction. It first collects multiple sequence alignments using PSI-BLAST. Amino-acid frequency and log-odds data with Henikoff weights are then used to train secondary structure, separately, based on the Rumelhart error backpropagation method. The final secondary structure prediction result is a combination of 7 neural network predictors from different profile data and parameters. The program is freely downloadable at the bottom of this page.

PSSpred on-line

Copy and paste your sequence here (<4,000 residues, in FASTA format):

Or upload the sequence from your local computer:
 Choose File No file chosen

Email: (mandatory, where results will be sent to)

ID: (optional, your given name of the protein)

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:

AILLPVEGAQLSELRQIPAEGGPVLHLRLDSPSQFSQFGEIYFSEVLPRRVKAWKRHSLTQLFAVPVGCI
HVVLVYDGREKSPTSGRLAQVTLGRPDNYRLLRIPPQVWYGFAATGDTPALVANCTDIPHQRQGESERAP
QDAPFIPFSWAGADLSGT

Output of PSSpred

PSSpred results for T29162

```
seq:   1 LPDYTNIGLYEGDWVWLKKFPGDRHIAIRPATKMAFSKIRELRHENVALYLGLFLAGGAG 60
SS:     CCCCCCCEEECCEEEEEECCCCCHHHHHHHHHHHCCCCCEEEEEEECCCC
conf: 99876317982918999995699727899999999999837871213677798447887
```

```
seq: 61 GPAAPGEGVLAVVSEHCARGSLQDLLAQRDIKLDWMFKSSLLDLIKGIRYLHHRGVAHG 120
SS: CCCCCCCCCEEEEEECCCCCHHHHHHHCCCCCHHHHHHHHHHHHHHHHHCCCC
conf: 7655789958999964999978999716999989999999999984898778
```

```
seq: 121 RLKSRNCVVDGRFVLKVTDHGHRLLAEQRVLPEPPSAEDQLWTAPELLRDPVLERRGTL 180
SS: CCCCCCCEEECCCEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCC
conf: 887531366198079996377121551564224666543555688788618988889885
```

```
seq: 181 AGDVFSLGIIMQEVVCRSAPYAMLELTPEEVVKRVQSPPLCRPSVSIDQAPMECIQLMK 240
SS: CHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHCCCCCCCCCCCCCCCCHHHHHHHH
conf: 22245788999999978999998998999999858888999998789999999999
```

```
seq: 241 QCWAEQPELRPMSMDRTFELFKSINKGRKMNIID 273
SS: HHHHCCCCHHCCCCHHHHHHHHHHHHHHCCCCCCCC
conf: 9996294449999999999965126899899
```

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

'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'

Name
PSSpred.pl
PSSpred1
PSSpred2
PSSpred3
PSSpred4
PSSpred5
PSSpred6
PSSpred7
README
wgt1
wgt2
wgt3
wgt4
wgt5
wgt6
wgt7

Perl script to implement PSSpred

Executable program for 7 PSSpred predictors

Instruction of the package

Trained parameters

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/>

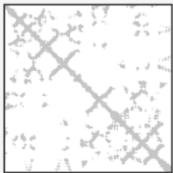
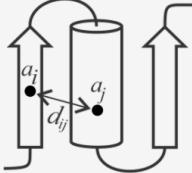




[Home](#) [Research](#) [Services](#) [Publications](#) [People](#) [Teaching](#) [Job Opening](#) [News](#) [Lab Only](#)

Online Services

- I-TASSER
- QUARK
- LOMETS
- COACH
- COFACTOR
- MUSTER
- SEGMER
- FG-MD
- ModRefiner
- REMO
- SPRING
- COTH
- BSpred
- SVMSEQ
- ANGLOR
- BSP-SLIM
- SAXSTER
- ThreaDom
- EvoDesign
- GPCR-I-TASSER
- BindProf
- BindProfX

 **NeBcon** 
Accurate prediction of protein contact maps

NeBcon (Neural-network and Bayes-classifier based contact prediction) is a hierarchical algorithm for sequence-based protein contact map prediction. It first uses the naive Bayes classifier theorem to calculate the posterior probability of eight machine-learning and co-evolution based contact prediction programs (SVMSEQ, BETACON, SVMcon, PSICOV, CCPred, FreeContact, MetaPSICOV, and STRUCTCH). Final contact maps are then created by neural network machine that trains the posterior probability scores with intrinsic structural features from secondary structure, solvent accessibility, and Shannon entropy of multiple sequence alignments.

NeBcon On-line ([view an example of NeBcon output](#))

Cut and paste your sequence (in [FASTA format](#)) below:

Or upload the sequence from your local computer: No file chosen

Email: (mandatory, where results will be sent to)

ID: (optional, your given name of the protein)

Download package:

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:

AILLPVEGAQLSELRQIPAEGGPVLHLRLDSPSQFSQFGEIYFSEVLPRRVKAWKRHSLTQLFAVPVGCI
HVVLYDGREKSPTSGRLAQVTLGRPDNYRLLRIPPQVWYGFAATGDTPALVANCTDIPHQRQGESERAP
QDAPFIPFSWAGADLSGT