

Secondary structure prediction

Secondary structure prediction

Amino acid sequence -> Secondary structure

Alpha helix

Beta strand

Disordered/coil

70% accuracy 1991, 81% accuracy in 2009

Secondary structure prediction

Limits:

Limited to globular proteins

Not for membrane proteins

Secondary structure prediction

Applications

Site directed mutagenesis

Locate functionally important residues

Find structural units / domains

Secondary structure prediction

Techniques

Linear statistics

Physicochemical properties

Linear discrimination

Machine learning

Neural Networks

K-nearest neighbours

Evolutionary trees

Residue substitution matrices

Using evolutionary information = Multiple sequence alignments.

Secondary structure prediction

Jnet

Cuff and Barton (2000)

Neural Network

Training set: 480 proteins (non homologous)

Construction of MSA for each using BLAST

Secondary structure prediction

Jnet

Neural network

N_i Neuron i , N_j neuron j
 W_{ij} weight from N_i to N_j



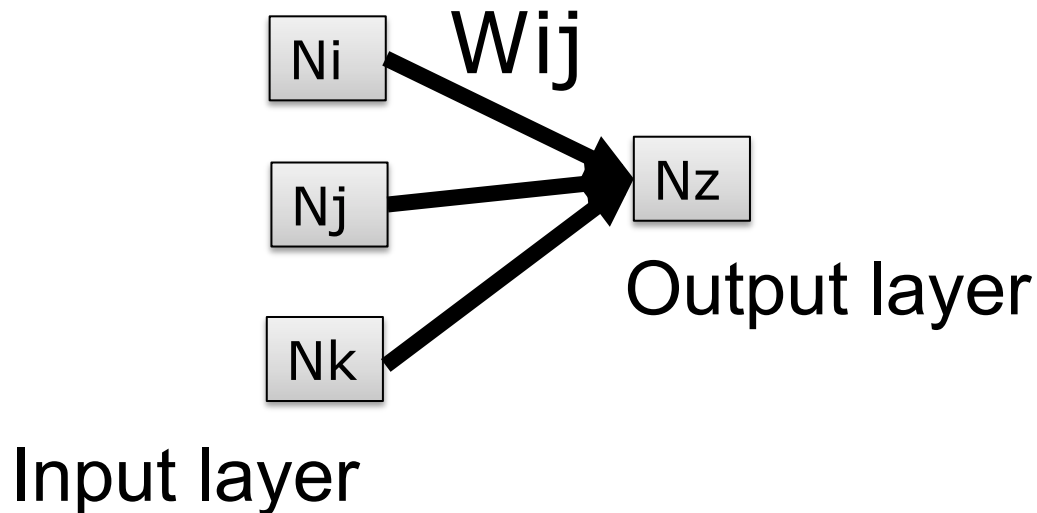
Signal forward propagation
Output from $N_i * \text{Weight } N_i \text{ to } N_j$
Input to N_j is $I_j = O_i * W_{ij}$

Secondary structure prediction

Jnet

Neural network

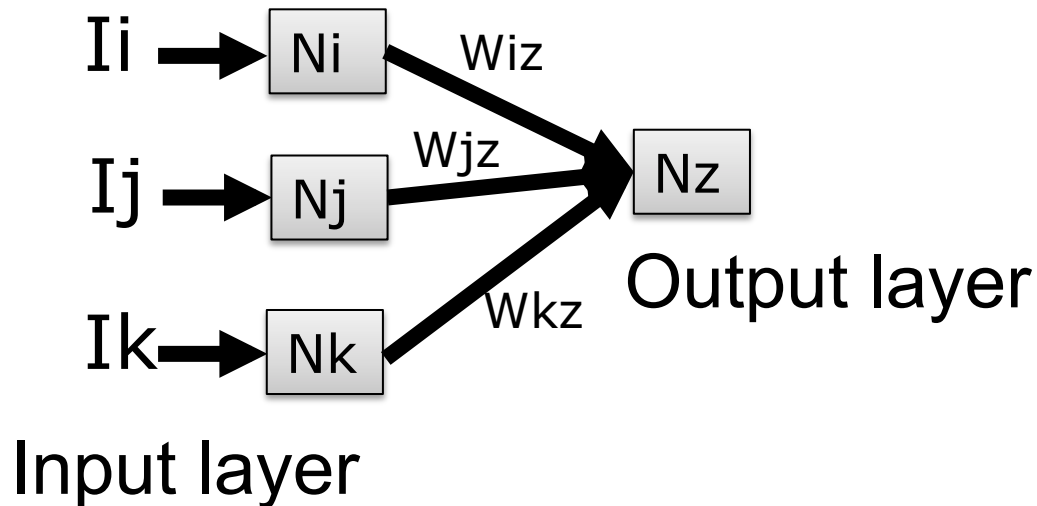
N_i Neuron i , N_j neuron j
 W_{ij} weight from N_i to N_j



Secondary structure prediction Jnet

Neural network

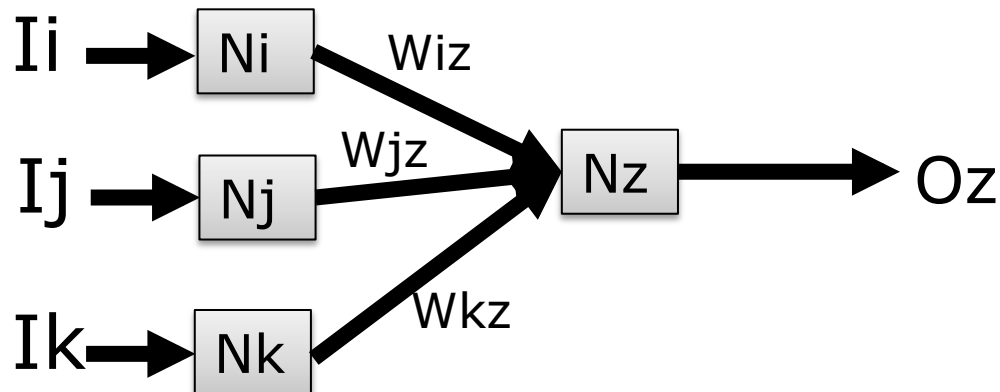
The network receives input values



Secondary structure prediction Jnet

Neural network

Signal forward propagation



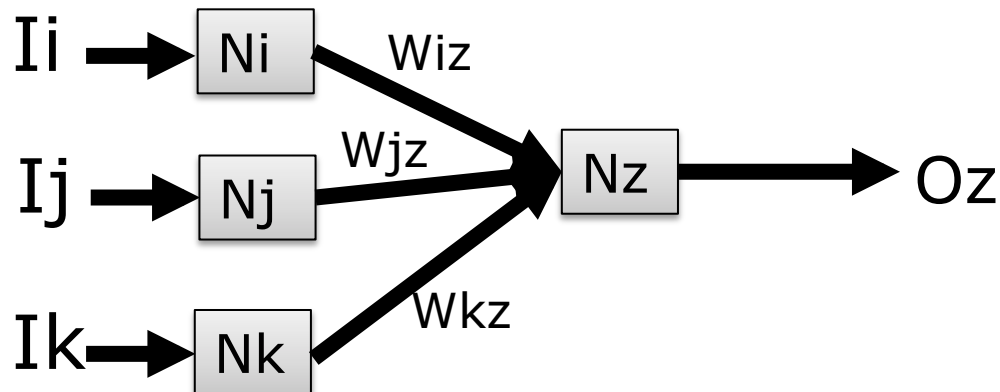
Sum of outputs from N_i , N_j , N_k

Secondary structure prediction

Jnet

Neural network

Compute the error



Desired value is 1, O_z is 0.8

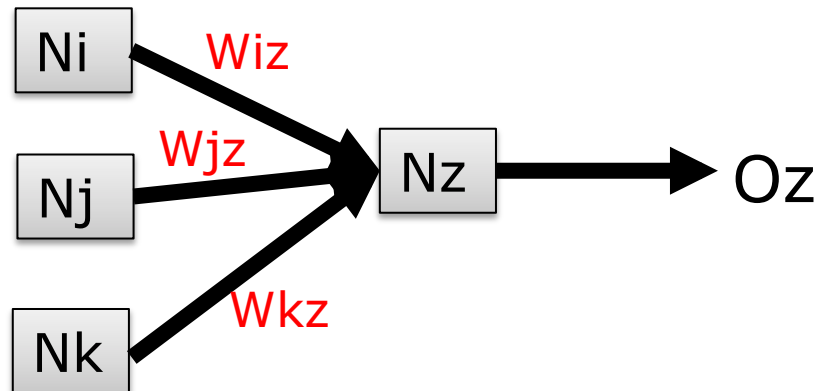
Error is $= O_z - \text{desired value} = 1 - 0.8 = 0.2$

Secondary structure prediction

Jnet

Neural network

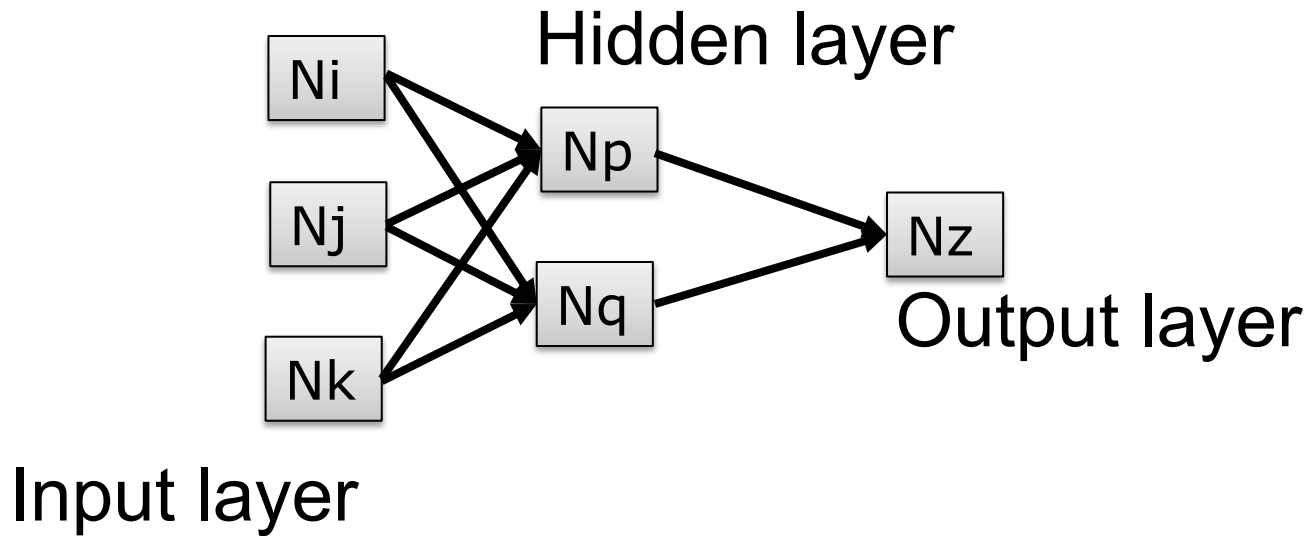
Error backpropagation



Weights are modified so that the result is a bit closer to what we wanted

Secondary structure prediction Jnet

Neural network



Secondary structure prediction

Jnet

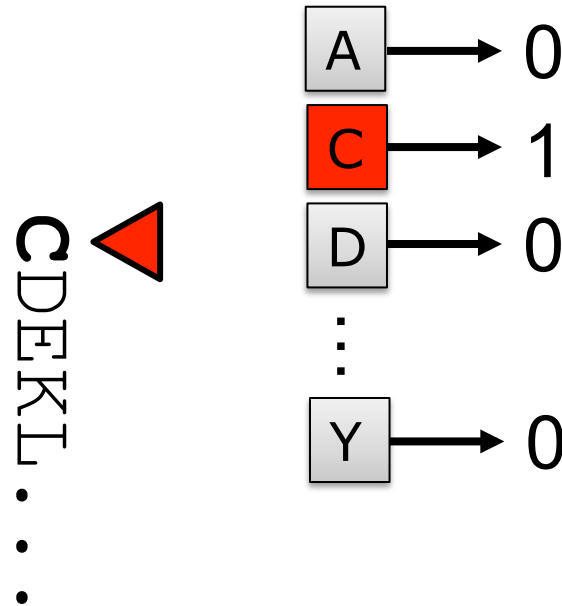
Neural network



Input layer: read a sequence CTEIL...

Secondary structure prediction Jnet

Neural network

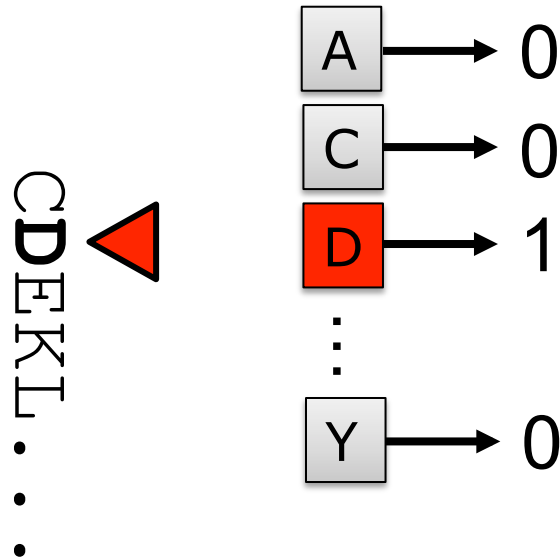


Input layer: read a sequence CDEKL...

Secondary structure prediction

Jnet

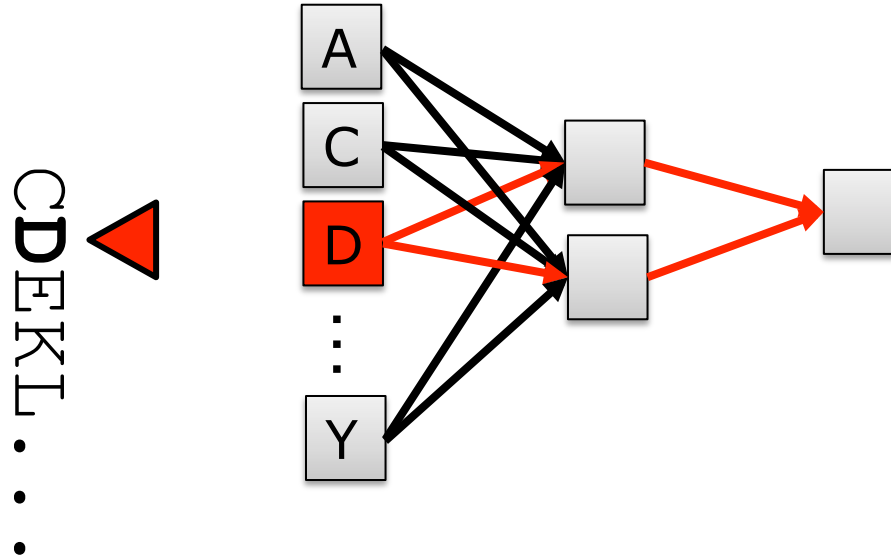
Neural network



Input layer: read a sequence CDEKLI...

Secondary structure prediction Jnet

Neural network

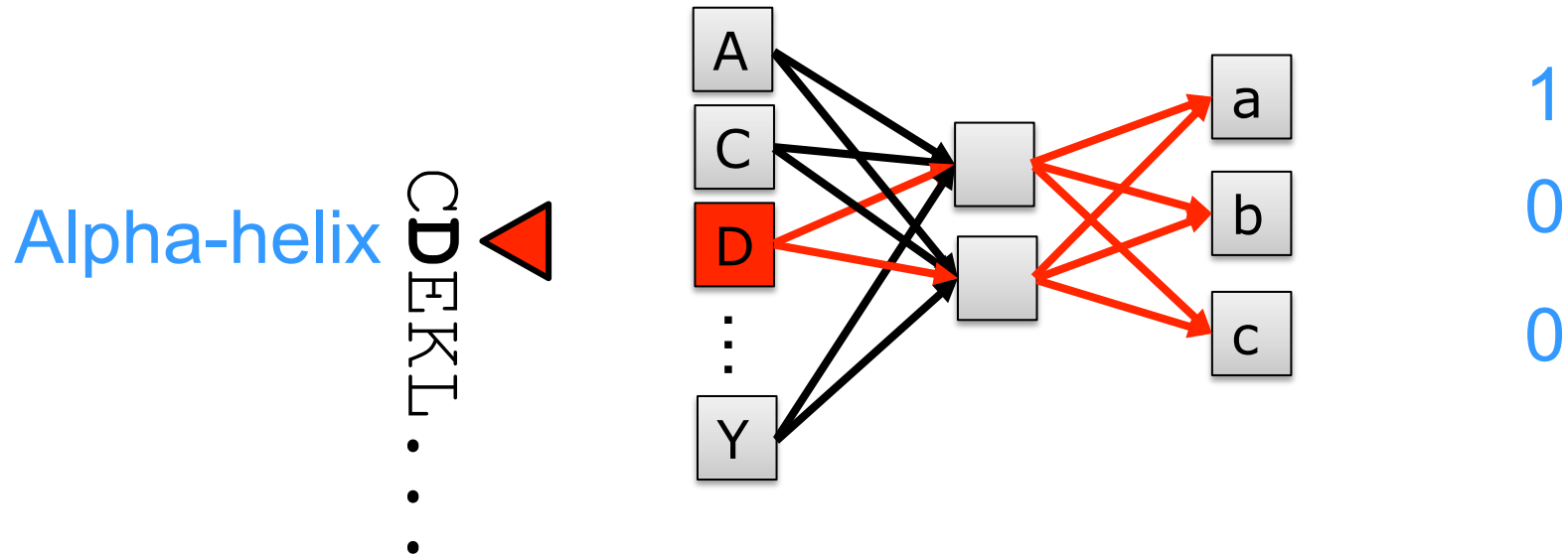


Input layer: read a sequence CDEKL...

Secondary structure prediction

Jnet

Neural network



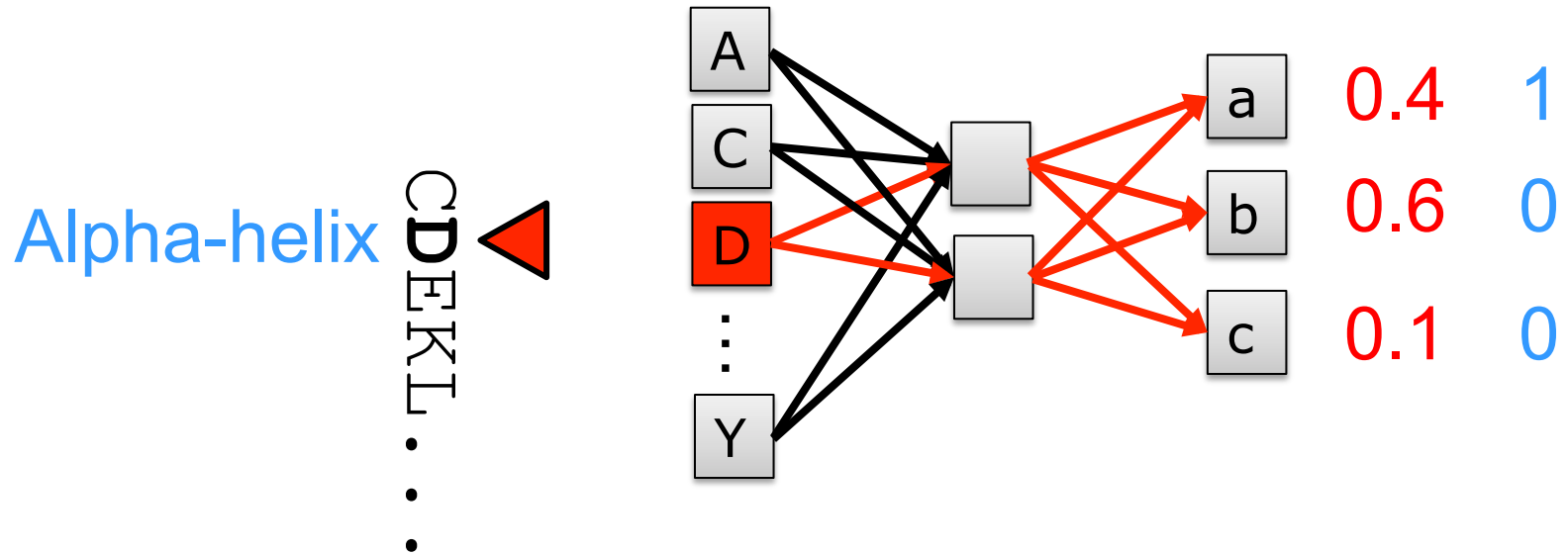
Output layer: structure

Desired output: known structure

Secondary structure prediction

Jnet

Neural network

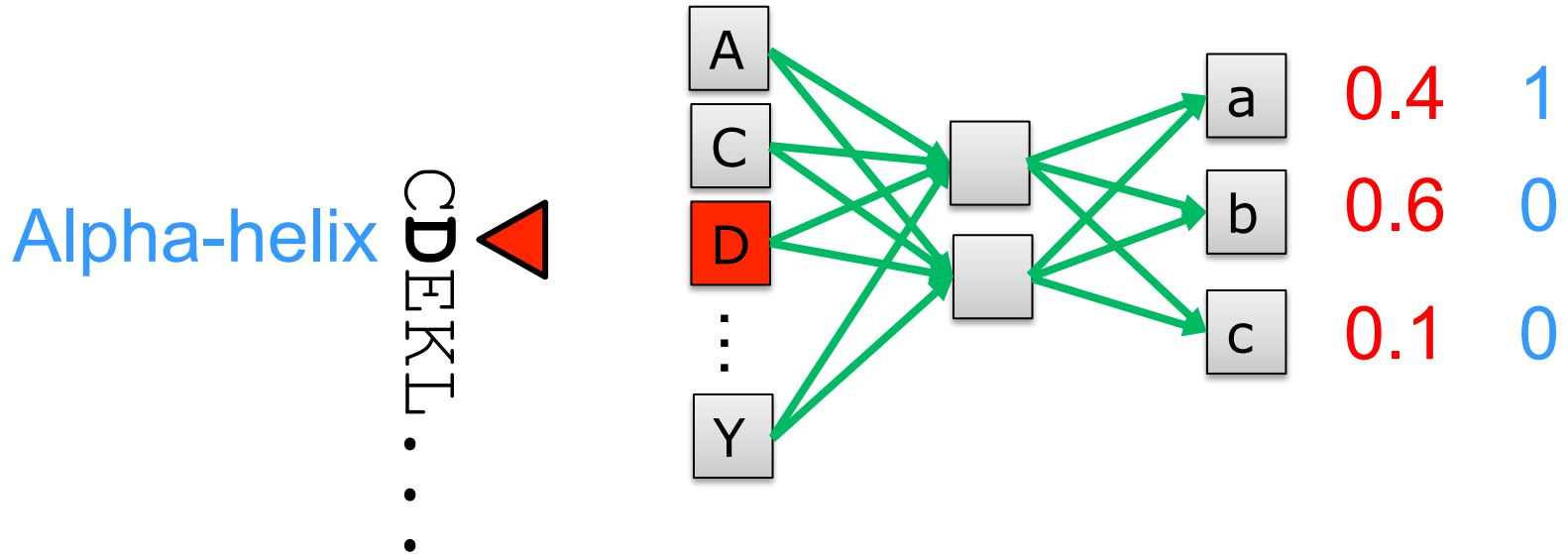


Output layer: structure

Desired output: known structure

Secondary structure prediction Jnet

Neural network

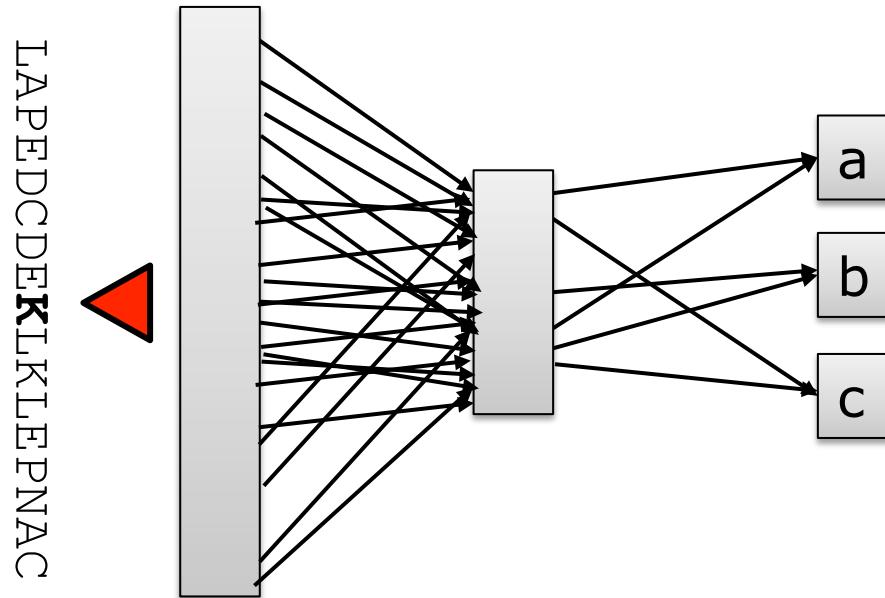


Error backpropagation = weights are modified

Secondary structure prediction Jnet

Jnet architecture

Sequence to structure network



Input layer = window of 17 residues

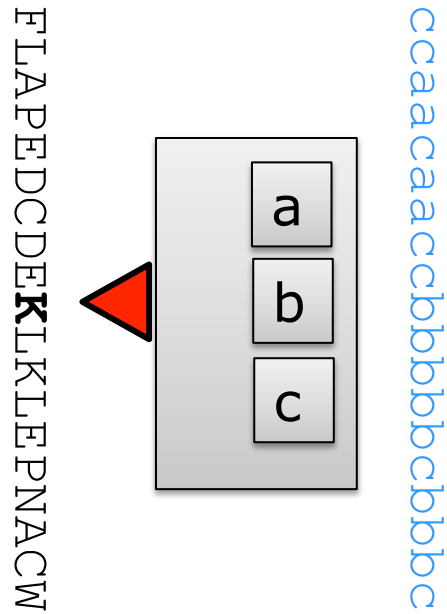
Hidden layer = 9 neurons

Output layer = 3 neurons

Secondary structure prediction

Jnet

Jnet architecture

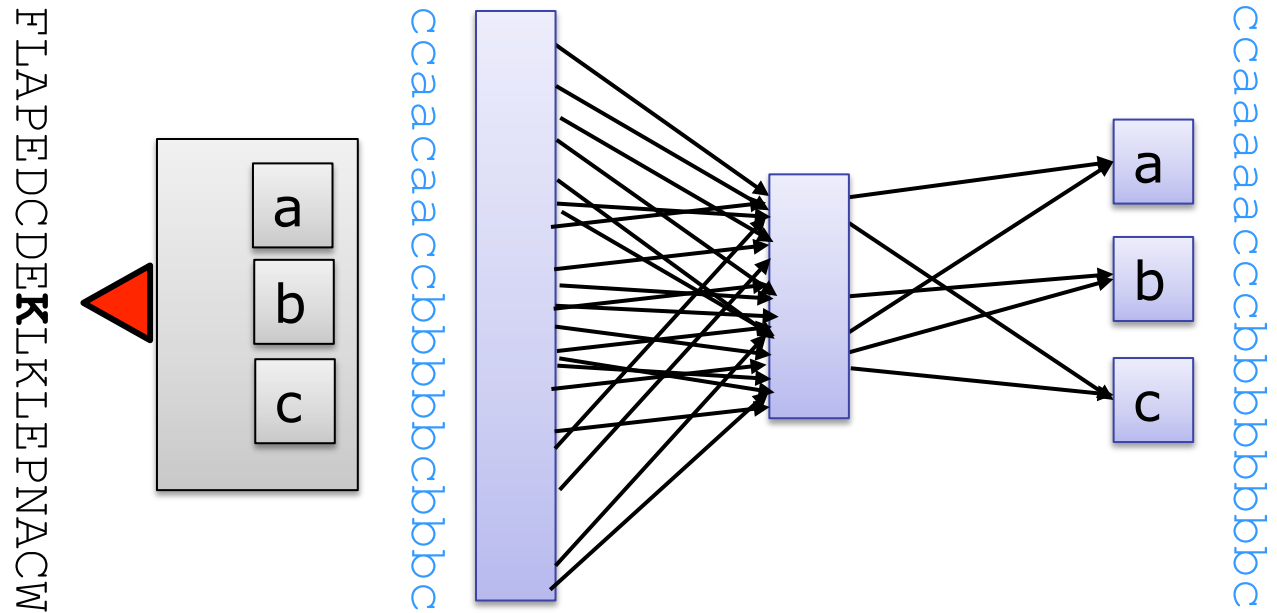


Secondary structure prediction

Jnet

Jnet architecture

Structure to structure network



Input layer = window of 19 residues
Hidden layer = 9 neurons
Output layer = 3 neurons

Secondary structure prediction



Geoff Barton,
University of
Dundee

Jpred 4 Incorporating Jnet

A Protein Secondary Structure Prediction Server

[Home](#)

[About](#)

[News](#)

[F.A.Q.](#)

[Help & Tutorials](#)

[Contact](#)

[Publications](#)

Input sequence(?)

```
MQVWPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSKWVPCLEFSKVGfVYRENHRSPGYDGRYWTMWKLPfMGCTD  
ATQVLKELEEAKKAYPDAFVRIIGFDNVRQVQLISFIAYKPPGC
```

[Advanced options \(click to show/hide\)](#)

[Make Prediction](#)

[Reset Form](#)

Primary citation: Drozdetskiy A, Cole C, Procter J & Barton GJ. Paper in preparation
Previous: Cole C, Barber JD & Barton GJ. *Nucleic Acids Res.* 2008. 35 (suppl. 2) W197-W201 [[link](#)]
More citations: [link](#).



Cole et al (2008) *Nucleic Acids Research*

Secondary structure prediction

Jpred

Uses algorithm Jnet2.0

Three state prediction

Alpha, beta, coil

Accuracy 81.5% (2008)

But if no homolog (orphan sequence) 65.9%!

PSIBLAST PSSM matrix

HMMer profiles (instead of aa frequencies)

Multiple neural networks

100 hidden layer units

Secondary structure prediction

Jpred

First, search against PDB sequences using BLAST (but only for warning)

PSIBLAST search of UniRef90, 3 iterations, Alignment of hits (filtered at 75% id)

Profiles from alignment (PSSM and HMMer)

Profiles are input to JNet

Alternative: user provides alignment (faster)

Secondary structure prediction

Advanced Jpred4 usage

Jpred 4
Incorporating Jnet

A Protein Secondary Structure Prediction Server

Home

About

News

F.A.Q.

Help & Tutorials

Contact

Publications

Input sequence^(?)

```
MQVWPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSKWVPCLEFSKVGfVYRENHRSPGYDGRYWTMWKLPFMFGCTD
ATQVLKELEEAKKAYPDAFVRIIGFDNVRQVQLISFIAYKPPGC
```

Advanced options (click to show/hide)

...or upload a file^(?)

Datei auswählen Keine ausgewählt

Select type of input^(?)

Single Sequence (click to select format):

Multiple Alignment (click to select format):

Skip searching PDB before prediction^(?)

Check to skip

Email address (optional)^(?)

email@domain

Query name (optional)^(?)

TestName_17

Make Prediction

Reset Form

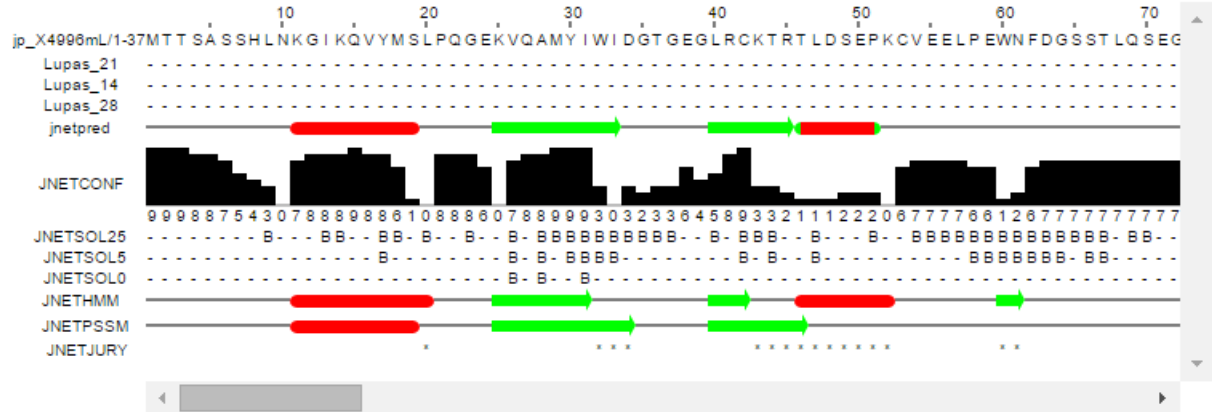
Secondary structure prediction

Jpred output

Results

After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days.

- [View results summary in SVG - displayed below:](#)



- [View full results in HTML](#)
- [View simple results in HTML](#)
- [View results in PS](#)
- [View results in PDF](#)
- [View results in Jalview](#) (Links to a separate page with the Jalview Java applet)
- [View full multiple sequence alignment with gaps and insertions](#)
- [View full multiple sequence alignment without gaps and insertions](#)
- [View everything in a results directory](#)
- [Get all \(but PS\) files in TAR.GZ archive](#)

This Jpred prediction was made with following.

Jnet version: 2.3.1

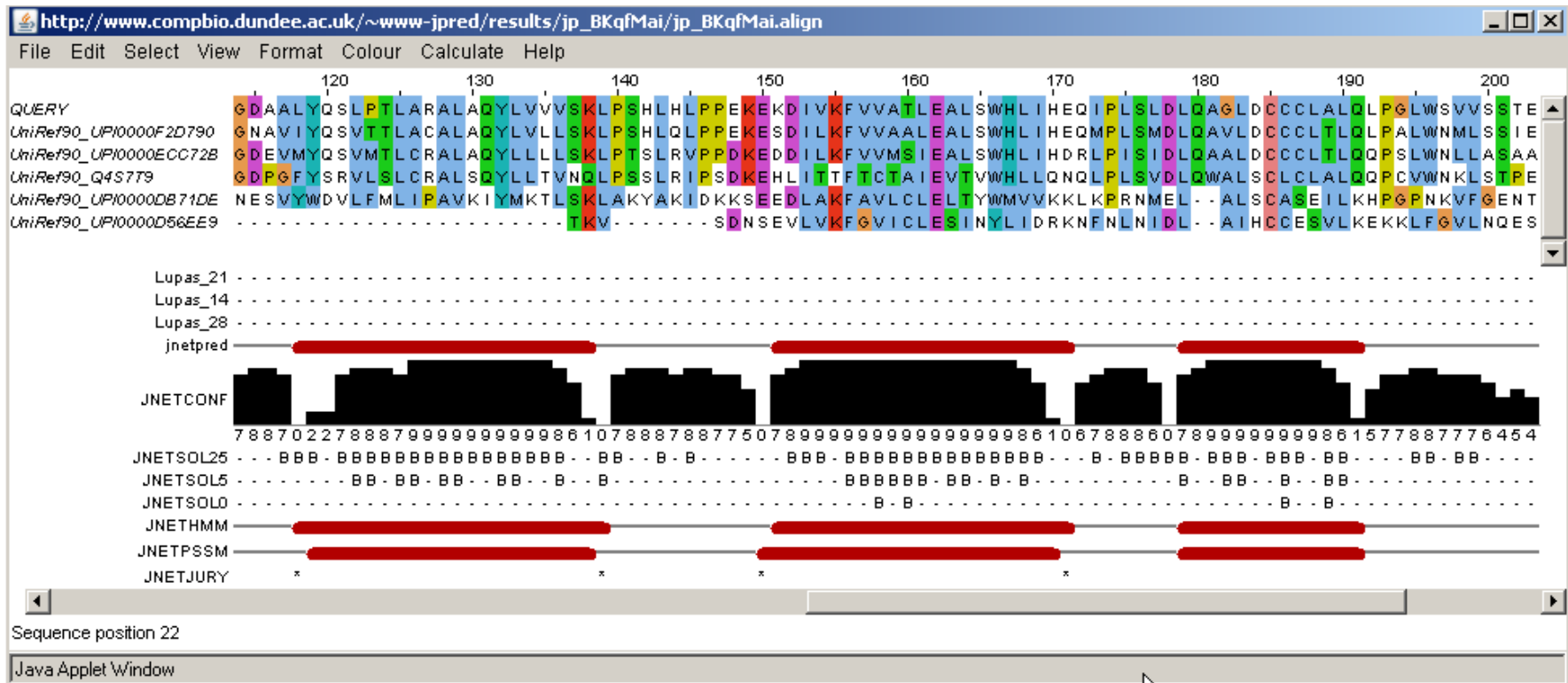
UniRef90 release: 2014_07, 09-Jul-2014



Secondary structure prediction

JPred

Jpred output / Jalview



Secondary structure prediction

JPred

Jpred output / view all

```
QUERY : KDWYVHLVKSQCWTRSDSALLEGAEVNRIPAEDMNAFMMNSEFNLSLLAPCLSLGMSEISGGQKSALFEAAREVTLARVSGTVQQI
UniRef90 UPI0000F2D790 : QDWYLSLVKFCCTKSDSALLEGAEVNRIPPGELTPFMLSKEFNLCLLAPCLSLGVREISSGQSSSLFETARVTLDRVASLVQQI
UniRef90 UPI0000ECC72B : KDWYMSLVRSQCCIKSADSALLEGAEVNRIPQPDLSNFMMNSKEFNLSLLAPCLSLGMNEISRDKSSLFEAARRVTLHDLSATVLNI
UniRef90 Q4S7T9 : KEWYVALVKSQCCVHGDVSLLETTELLTKLPPADLLSVMSCKEFNLRLLCPCLSLGVQRLVVRGQGSLLLETALHVTLEQLAGATGLI
UniRef90 UPI0000DB71DE : RNWFLSQVKLRCCNSNMYNLSEAAQLLKVLDFEDCLGILSSKEFNIAILKPCIIILGVRVVEKCKQKSPLYSAAKQCLLDHIHYVIDLI
UniRef90 UPI0000D56EE9 : KNWYLAQIREKCNK-----ETAQLLTRLDYEELCNFLRNADFSKVVFKECLSV-----PTLEDSERFDSRFLTQIYDL:

: 1-----11-----21-----31-----41-----51-----61-----71-----81-----
OrigSeq : KDWYVHLVKSQCWTRSDSALLEGAEVNRIPAEDMNAFMMNSEFNLSLLAPCLSLGMSEISGGQKSALFEAAREVTLARVSGTVQQI

Jnet : --HHHHHHHHHH-----HHHHHHHH-----HHHHHHHH-----HHHHHHHHHHHHHHHH-----HHHHHHHHHHHHHHHH--
jhmm : --HHHHHHHHHH-----HHHHHHHH-----HHH-HHHH-----HHHHHHHHHHHHHHHH-----HHHHHHHHHHHHHHHH--
jpssm : --HHHHHHHHHH-----HHHHHH-----HHHH-HHHH-----HHHHHHHHHH-----HHHH-----HHHHHHHHHHHHHHHH--

Lupas 14 : -----
Lupas 21 : -----
Lupas 28 : -----

Jnet_25 : --BB--BBB--BB-----BBBB-BB-BB--B-B--BB-BB-B--BBBBBBB-BBBBBB--B-----BB-BB--BBB--BB-BB--I
Jnet_5 : -----B--B-----B--BB-BB--B--B--BB-B--B--BB--BB-B-B--B--B-----B--B--B--B--B-----
Jnet_0 : -----B-----B-----B-----B-----B-----
Jnet Rel : 82799999998514677750012145327754430067503787424331321200101000777632799999999999998722!
```

Secondary structure prediction

JPred

Jpred output / PDF output

```
QUERY          1 10 20 30 40
UniRef90_UPI0000F2D790 K D W Y V H L V K S Q C W T R S D S A L L E E G A E L V N R I P P A E D M N A F M M
UniRef90_UPI0000ECC72B Q D W Y L S L V K F S Q C C T K S D S A L L E E G A E L V N R I P P G E L T P F M L
UniRef90_Q4S7T9       K D W Y M S L V R S Q C C I K S D S A L L E E G A E L L N R I P P Q P D L N S F M N
UniRef90_UPI0000DB71DE K E W Y V A L V K S Q C C V H G D V S L L E E T T E L L L T K L P P A D L L S V M S
UniRef90_UPI0000D56EE9 R N W F L S Q V K L R C C N S N N Y N L S E E A A Q L L K V L D F E D C L G L L S
UniRef90_UPI0000D56EE9 K N W Y L A Q I R E K C N S K E E T A Q L L T R L D Y E E L C N F L R
```

lupas_21
lupas_14
lupas_28
jnet
conf
sol25
sol5
sol0
jhmm
jpssm



```
QUERY          50 60 70 80
UniRef90_UPI0000F2D790 N S E F N L S L L A P C L S L G M S E I S G G Q K S A L F E E A A R E V T L A R V
UniRef90_UPI0000ECC72B S K E F N L C L L A P C L S L G V R E E I S S G Q S S S L F E E T A R R S V T L D R V
UniRef90_Q4S7T9       S K E F N L S L L A P C L S L G M N E I S S R D Q K S S L F E E A A R R V T L E H L
UniRef90_UPI0000DB71DE C K E F N L R L L C P C L S L G V Q R L V R G Q G S L L L E T A L H V T L E Q L
UniRef90_UPI0000D56EE9 S K E F N I A I L K P C I T L G V R S V E K C Q K S P L Y S A A K Q C L L D H I
UniRef90_UPI0000D56EE9 N A D F S K V V F K E C L S V P T L E D S E R F D S R F L
```

Exercise 1/3

Jalview 2D prediction

Starting Jalview

Open Firefox with JRE (from ZDV)

Go to <http://www.jalview.org>

Click the pink arrow “Launch Jalview Desktop”

You can close all the demo windows that appear

Exercise 1/3

Jalview 2D prediction

Load an alignment

Use MR1_fasta.txt

This is an alignment of a fragment of the mineralocorticoid receptor

Open it from File > Input alignment > From file
(Hint: You can load it directly as an URL, e.g.

https://cbdm.uni-mainz.de/files/2015/02/MR1_fasta.txt)

The alignment has its own Menu tabs

Try Colour > Clustalx

to see conservation

Exercise 2/3

Jalview 2D prediction

Web service -> Secondary Structure Prediction -> Jnet
secondary str pred

No selection (or all sequences selected) = Jnet runs on top
sequence using the alignment (**fast**)

One sequence (or region) selected = Jnet runs on that
sequence using homologs (**slow**)

Some sequences selected = Jnet runs on top one using
homologs (**slow**)

Try with no sequences selected

Exercise 2/3

Jalview 2D prediction

If this doesn't work you can run directly MR1_fasta.txt on jpred4.

Use the advanced option

Upload a file option

Select type of input = Multiple alignment (use format FASTA)

Tick the skip PDB search option

There is an option to view output in Jalview

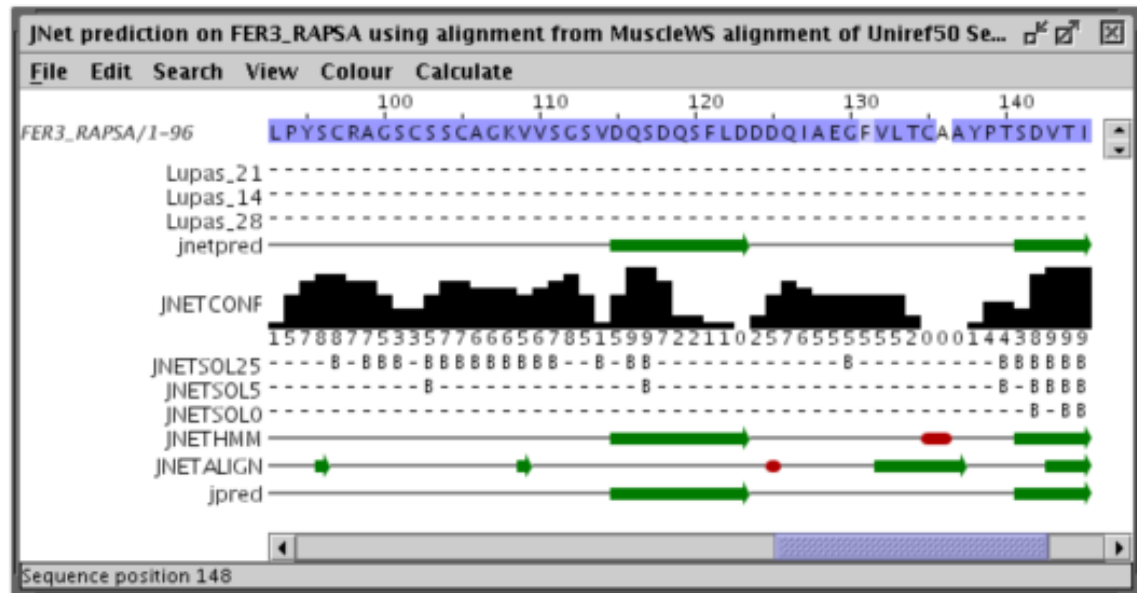
Exercise 3/3

Jalview 2D prediction

Annotations:

- Lupas_21, Lupas_14, Lupas_28

Coiled-coil predictions for the sequence. 21, 14 and 28 are windows used.



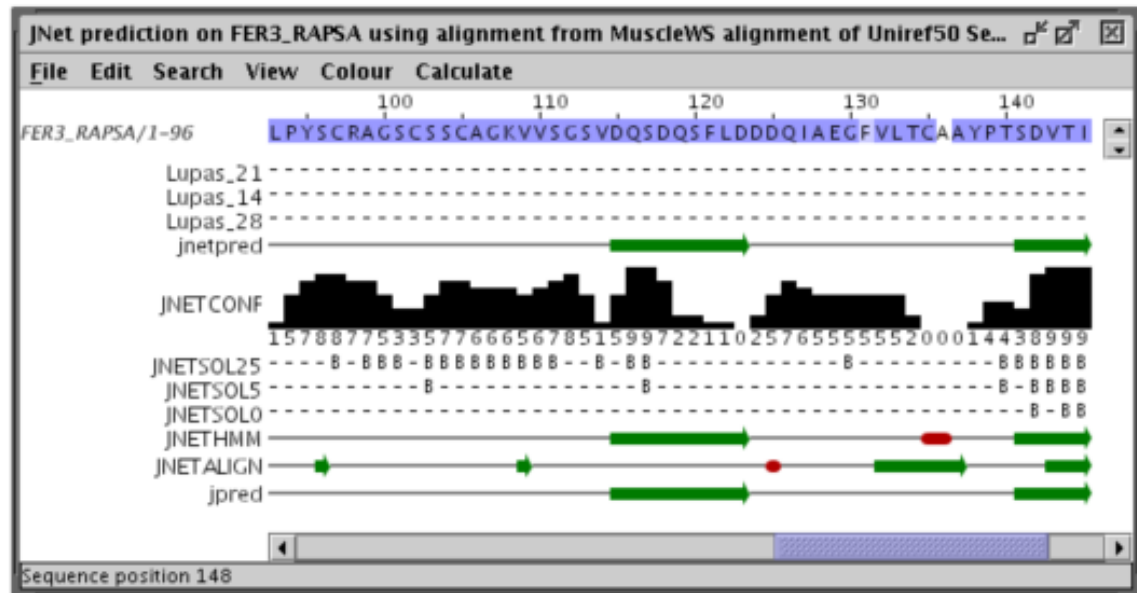
Exercise 3/3

Jalview 2D prediction

Annotations:

- JNETHMM, JNETALIGN: predictions using diff profiles
- Jnetpred: Consensus prediction.

Beta sheets: green arrows. Alpha helices: red tubes.



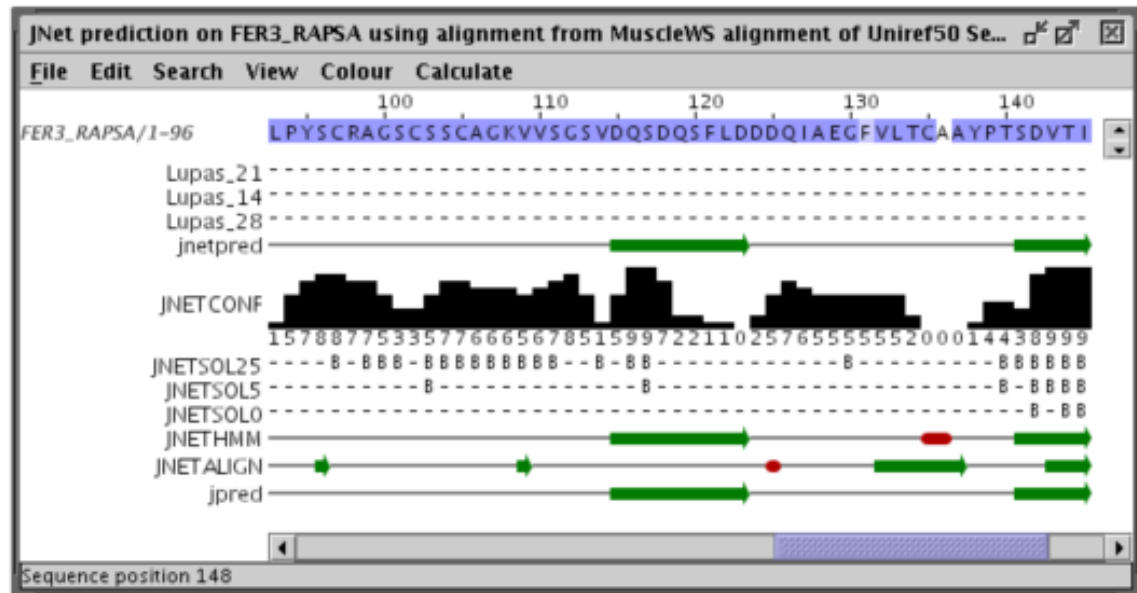
Exercise 3/3

Jalview 2D prediction

Annotations:

- JNETCONF

Confidence in the prediction.



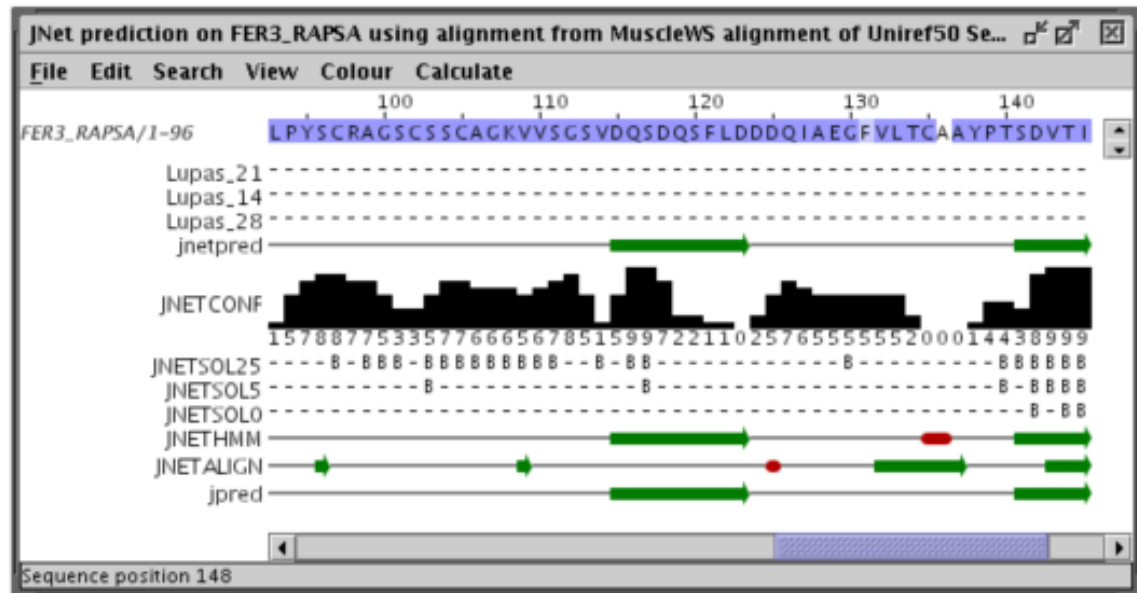
Exercise 3/3

Jalview 2D prediction

Annotations:

- JNETSOL25, JNETSOL5, JNETSOL0

Solvent accessibility predictions - binary predictions of 25%, 5% or 0% solvent accessibility.



Exercise 3/3

2D prediction of known 3D

Obtain the sequence of the human glutamine synthetase.

Run BLAST with the human sequence against:

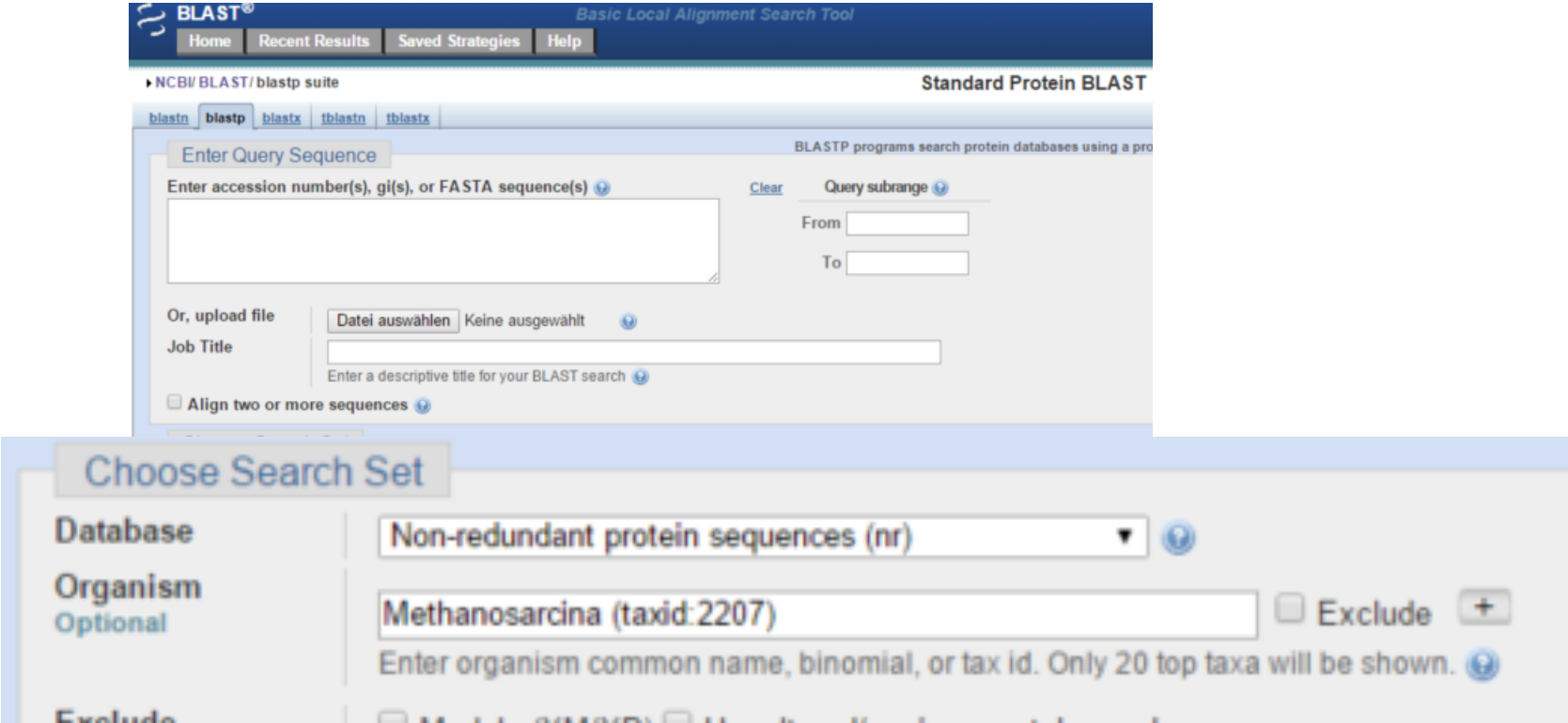
- 1) the archaea *Methanosarcina*
- 2) the bacteria *Escherichia coli*
- 3) the fungi *Pseudozyma Antarctica*

Get the best homolog, align the sequences (including the human protein, on top) and use the input in Jalview.

Exercise 3/3

2D prediction of known 3D

NCBI BLAST against single species is **faster!**



Exercise 3/3

2D prediction of known 3D

Obtain the sequence of the human glutamine synthetase.

Run BLAST with the human sequence against:

- 1) the archaea *Methanosarcina*
- 2) the bacteria *Escherichia coli*
- 3) the fungi *Pseudozyma Antarctica*

Get the best homolog from each, align the sequences and use the input in Jalview. Put the human protein on top.

Exercise 3/3

2D prediction of known 3D

Load the alignment in Jalview and run web prediction.

Alternative. Run the alignment in the Jpred4 server. (Hint: You could run the human sequence alone but that will search for homologs and will take very long)

Compare the prediction with the known 3D of the human protein (open it in Chimera, File > Fetch by ID > PDB 2QC8)

Exercise 3/3

2D prediction of known 3D

We need to hide all chains except one.

Select one of the chains (ctrl + click on a residue, then arrow up). Invert selection (press arrow right).

Actions > Ribbon > Hide

Actions > Atoms/bonds > Hide

Select the chain and focus on it

Actions > Focus

Exercise 3/3

2D prediction of known 3D

Compare the output of jpred/jalview 2D pref with the 3D structure of this protein.

For example, locate a predicted helix or beta-strand in Jalview. Find out the start and end positions hovering over the human sequence with the mouse (the numbers on top of the alignment are different from the amino acid positions in each sequence).

Color the corresponding residues in the 3D view using
Select > Atom specifier

And ranges: e.g. :113-126 (predicted as helix)

Actions > color > red

Apply color some helices red and strands in green.

Do you see differences? Where are they?

Would you say that the 2D prediction was reasonable?