Ab initio
Protein Structure Prediction

Protein Structure Prediction

- Secondary Structure Prediction
- Ab initio Structure prediction

Secondary Structure Prediction

- *Given a protein sequence $a_1a_2\ldots a_N$, secondary structure prediction aims at defining the state of each amino acid ai as being either H (helix), E (extended=strand), or O (other) (Some methods have 4 states: H, E, T for turns, and O for other).*
- *The quality of secondary structure prediction is measured with a "3-state accuracy" score, or Q_3. Q_3 is the percent of residues that match "reality" (X-ray structure).*

Quality of Secondary Structure Prediction

Determine Secondary Structure positions in known protein structures using DSSP or STRIDE:

1. Kabsch and Sander. Dictionary of Secondary Structure in Proteins: pattern recognition of hydrogen-bonded and geometrical features. Biopolymer 22: 2571-2637 (1983) (DSSP)
2. Frischman and Argos. Knowledge-based secondary structure assignments. Proteins, 23:566-571 (1995) (STRIDE)

Limitations of Q_3

ALHEASGPSVILFGSDVTVPPASNAEQAK Amino acid sequence

hhhhhooooeeeeoooeeeooooohhhhh Actual Secondary Structure

ohhhooooeeeeoooooeeeooohhhhhh Q3=22/29=76%

(useful prediction)

hhhhhoooohhhhooohhhooooohhhhh Q3=22/29=76%

(terrible prediction)

- Q3 for random prediction is 33%

- Secondary structure assignment in real proteins is uncertain to about 10%; Therefore, a “perfect” prediction would have Q3=90%.

Early methods for Secondary Structure Prediction

- *Chou and Fasman*

  (Chou and Fasman. Prediction of protein conformation. Biochemistry, 13: 211-245, 1974)

- *GOR*

  (Garnier, Osguthorpe and Robson. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol., 120:97- 120, 1978)

Chou and Fasman

- *Start by computing amino acids propensities to belong to a given type of secondary structure:*

$$\frac{P(i/Helix)}{P(i)} \qquad \frac{P(i/Beta)}{P(i)} \qquad \frac{P(i/Turn)}{P(i)}$$

Propensities > 1 mean that the residue type I is likely to be found in the Corresponding secondary structure type.

Chou and Fasman

Amino Acid	α-Helix	β-Sheet	Turn
Ala	1.29	0.90	0.78
Cys	1.11	0.74	0.80
Leu	1.30	1.02	0.59
Met	1.47	0.97	0.39
Glu	1.44	0.75	1.00
Gln	1.27	0.80	0.97
His	1.22	1.08	0.69
Lys	1.23	0.77	0.96
Val	0.91	1.49	0.47
Ile	0.97	1.45	0.51
Phe	1.07	1.32	0.58
Tyr	0.72	1.25	1.05
Trp	0.99	1.14	0.75
Thr	0.82	1.21	1.03
Gly	0.56	0.92	1.64
Ser	0.82	0.95	1.33
Asp	1.04	0.72	1.41
Asn	0.90	0.76	1.23
Pro	0.52	0.64	1.91
Arg	0.96	0.99	0.88

Favors
α-Helix

Favors
β-strand

Favors
turn

Chou and Fasman

Predicting helices:

- find nucleation site: 4 out of 6 contiguous residues with P(α)>1
- extension: extend helix in both directions until a set of 4 contiguous residues has an average P(α) < 1 (breaker)
- if average P(α) over whole region is >1, it is predicted to be helical

Predicting strands:

- find nucleation site: 3 out of 5 contiguous residues with P(β)>1
- extension: extend strand in both directions until a set of 4 contiguous residues has an average P(β) < 1 (breaker)
- if average P(β) over whole region is >1, it is predicted to be a strand

Chou and Fasman

Position-specific parameters for turn:

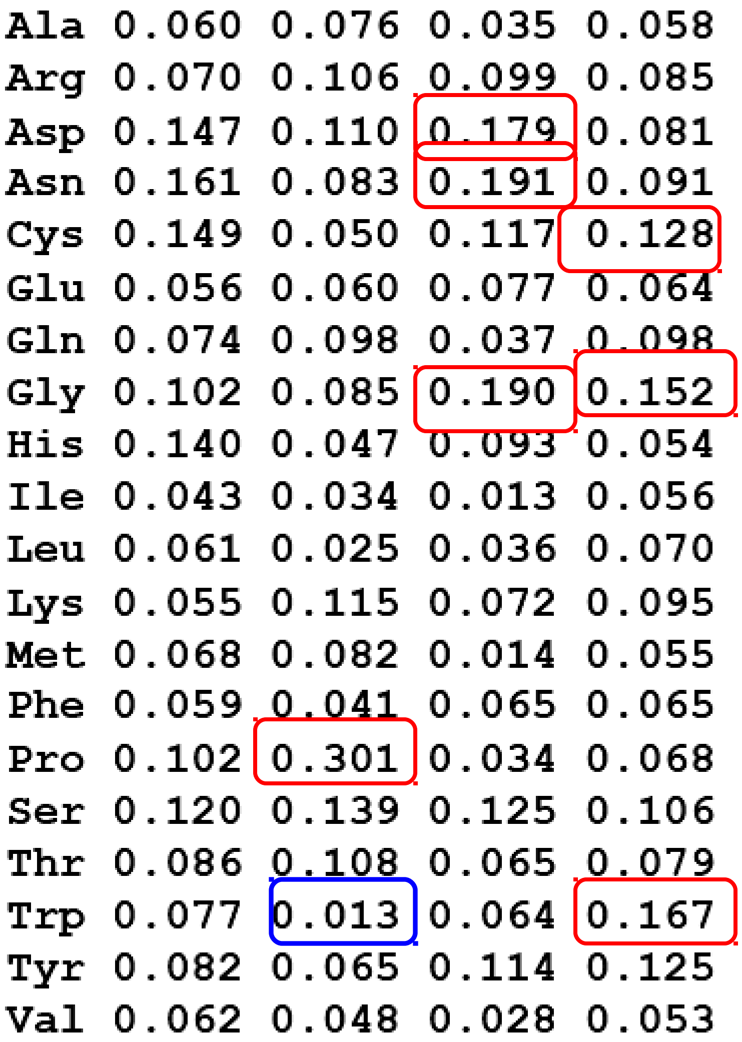
Each position has distinct amino acid preferences.

Examples:

-At position 2, Pro is highly preferred; Trp is disfavored

-At position 3, Asp, Asn and Gly are preferred

-At position 4, Trp, Gly and Cys preferred



	f(i)	f(i+1)	f(i+2)	f(i+3)
Ala	0.060	0.076	0.035	0.058
Arg	0.070	0.106	0.099	0.085
Asp	0.147	0.110	0.179	0.081
Asn	0.161	0.083	0.191	0.091
Cys	0.149	0.050	0.117	0.128
Glu	0.056	0.060	0.077	0.064
Gln	0.074	0.098	0.037	0.098
Gly	0.102	0.085	0.190	0.152
His	0.140	0.047	0.093	0.054
Ile	0.043	0.034	0.013	0.056
Leu	0.061	0.025	0.036	0.070
Lys	0.055	0.115	0.072	0.095
Met	0.068	0.082	0.014	0.055
Phe	0.059	0.041	0.065	0.065
Pro	0.102	0.301	0.034	0.068
Ser	0.120	0.139	0.125	0.106
Thr	0.086	0.108	0.065	0.079
Trp	0.077	0.013	0.064	0.167
Tyr	0.082	0.065	0.114	0.125
Val	0.062	0.048	0.028	0.053

Chou and Fasman

Predicting turns:

- for each tetrapeptide starting at residue i, compute:
    - P_{Turn} (average propensity over all 4 residues)
    - $F = f(i)*f(i+1)*f(i+2)*f(i+3)$

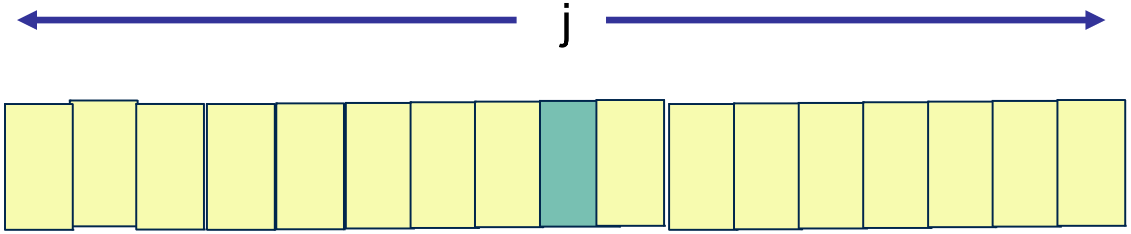
- if $P_{Turn} > P\alpha$ and $P_{Turn} > P\beta$ and $P_{Turn} > 1$ and $F > 0.000075$ tetrapeptide is considered a turn.

Chou and Fasman prediction:

http://fasta.bioch.virginia.edu/fasta_www/chofas.htm

The GOR method

Position-dependent propensities for helix, sheet or turn is calculated for each amino acid. For each position j in the sequence, eight residues on either side are considered.



A helix propensity table contains information about propensity for residues at 17 positions when the conformation of residue j is helical. The helix propensity tables have 20 x 17 entries.
Build similar tables for strands and turns.

GOR simplification:
The predicted state of AAj is calculated as the sum of the position-dependent propensities of all residues around AAj.

GOR can be used at : http://abs.cit.nih.gov/gor/ (current version is GOR IV)

Accuracy

- Both Chou and Fasman and GOR have been assessed and their accuracy is estimated to be Q3=60-65%.

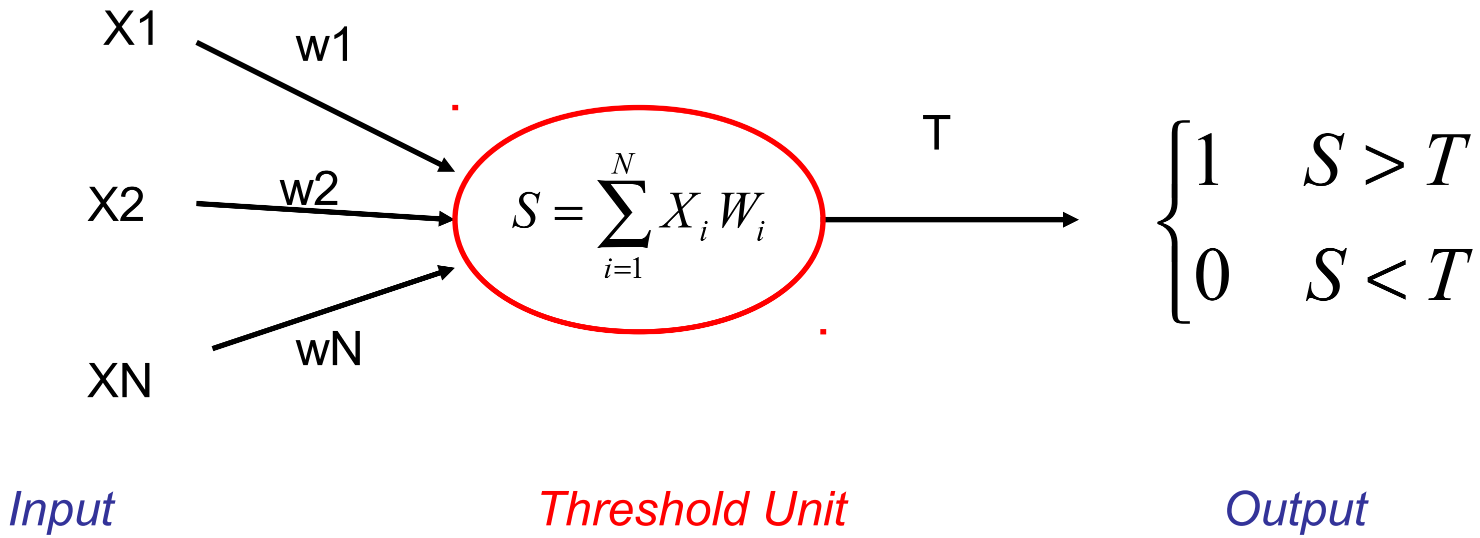
(initially, higher scores were reported, but the experiments set to measure Q3 were flawed, as the test cases included proteins used to derive the propensities!)

Neural networks

The most successful methods for predicting secondary structure are based on neural networks. The overall idea is that neural networks can be trained to recognize amino acid patterns in known secondary structure units, and to use these patterns to distinguish between the different types of secondary structure.

Neural networks classify “input vectors” or “examples” into categories (2 or more).
They are loosely based on biological neurons.

The perceptron


X1
w1
X2
w2
XN
wN
$S = \sum_{i=1}^{N} X_i W_i$
T
$\begin{cases} 1 & S > T \\ 0 & S < T \end{cases}$
Input
Threshold Unit
Output

The perceptron classifies the input vector X into two categories.

If the weights and threshold T are not known in advance, the perceptron must be trained. Ideally, the perceptron must be trained to return the correct answer on all training examples, and perform well on examples it has never seen.

The training set must contain both type of data (i.e. with “1” and “0” output).

The perceptron

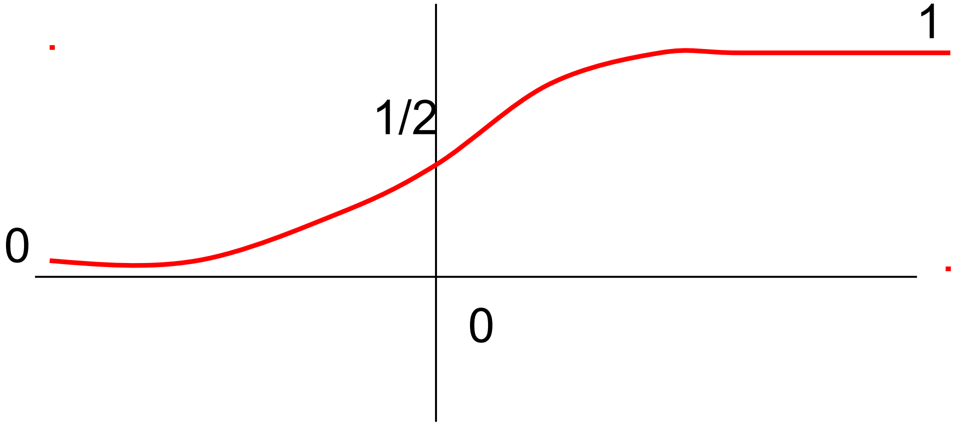
Notes:

- The input is a vector X and the weights can be stored in another vector W.

- the perceptron computes the dot product S = X.W

- the output F is a function of S: it is often set discrete (i.e. 1 or 0), in which case the function is the step function.
For continuous output, often use a sigmoid:

$$F(X) = \frac{1}{1+e^{-X}}$$



- Not all perceptrons can be trained ! (famous example: XOR)

The perceptron

Training a perceptron:

Find the weights W that minimizes the error function:

$$\boxed{E = \sum_{i=1}^{P} \left(F(X^i.W) - t(X^i) \right)^2}$$

P: number of training data
Xⁱ: training vectors
F(W.Xⁱ): output of the perceptron
t(Xⁱ) : target value for Xⁱ

Use steepest descent:

- compute gradient:

$$\boxed{\nabla E = \left(\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \frac{\partial E}{\partial w_3}, \ldots, \frac{\partial E}{\partial w_N} \right)}$$

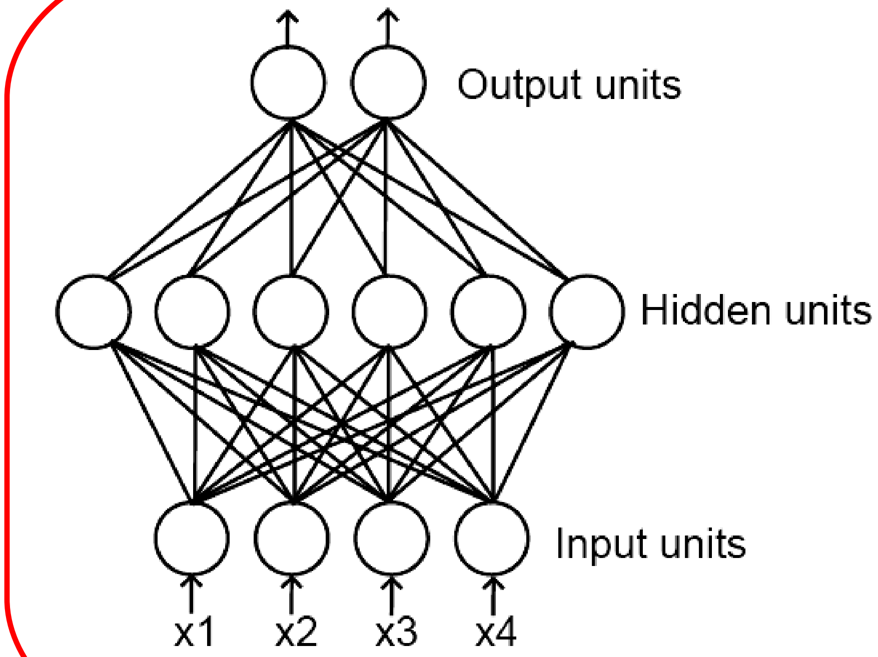
- update weight vector:

$$\boxed{W_{new} = W_{old} - \varepsilon \nabla E}$$

- iterate

(e: learning rate)

Neural Network



A complete neural network is a set of perceptrons interconnected such that the outputs of some units becomes the inputs of other units. Many topologies are possible!

Neural networks are trained just like perceptron, by minimizing an error function:

$$E = \sum_{i=1}^{Ndata} \left(NN(X^i) - t(X^i) \right)^2$$

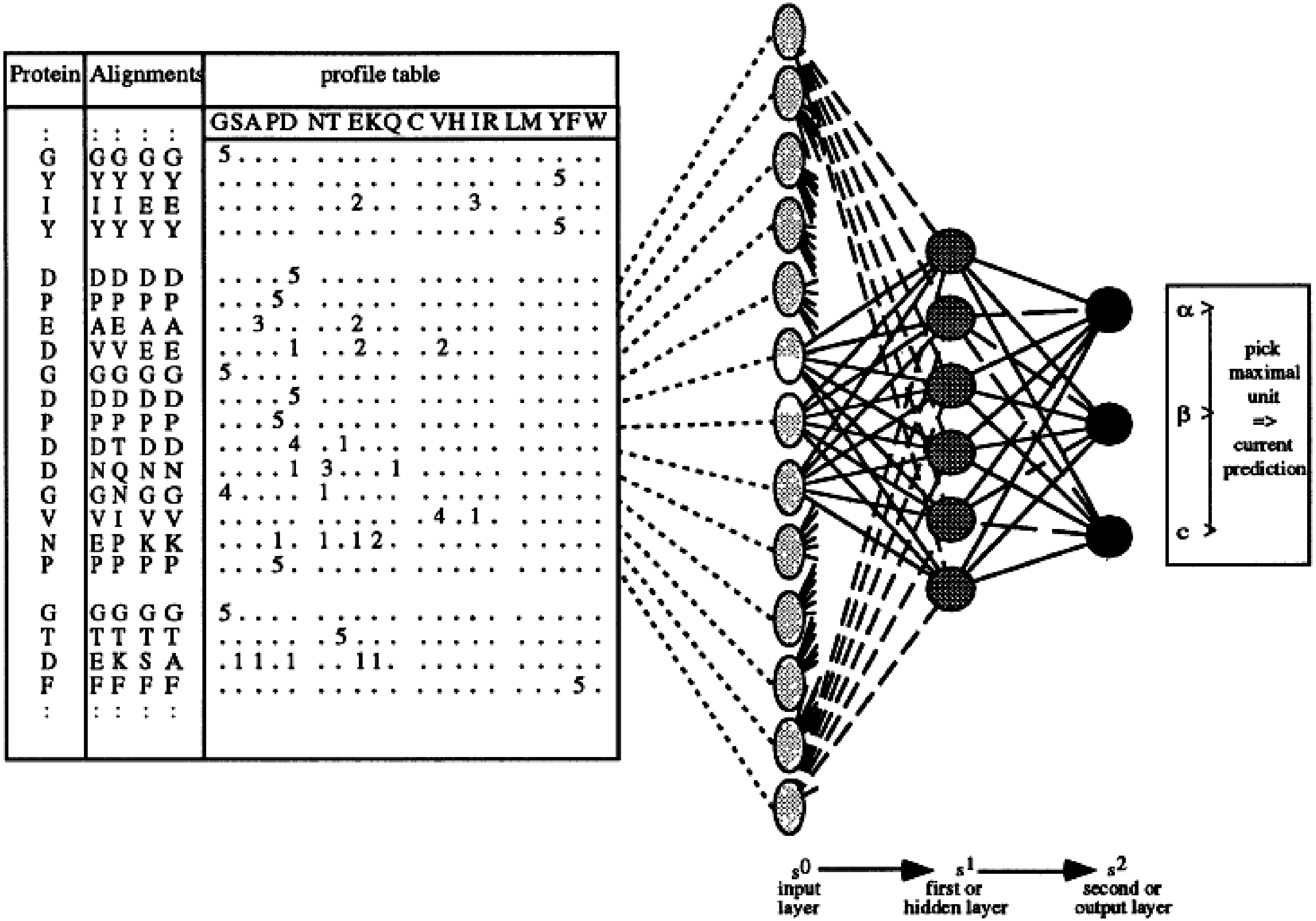
Neural networks and Secondary Structure prediction

Experience from Chou and Fasman and GOR has shown that:

- In predicting the conformation of a residue, it is important to consider a window around it.
- Helices and strands occur in stretches
- It is important to consider multiple sequences

PHD: Secondary structure prediction using NN

Biophysics: Rost and Sander   *Proc. Natl. Acad. Sci. USA 90 (1993)*   7559



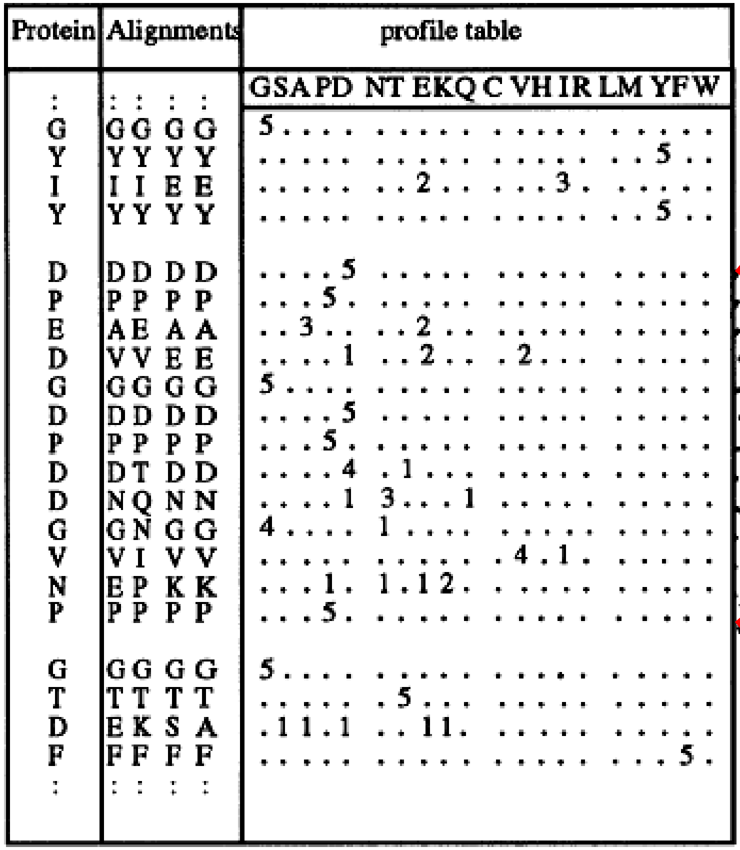
Protein	Alignments	profile table																			
		G	S	A	P	D	N	T	E	K	Q	C	V	H	I	R	L	M	Y	F	W
:	: : : :																				
G	G G G G	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y	Y Y Y Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.
I	I I E E	.	.	.	.	.	.	.	2	.	.	.	.	.	3	.	.	.	.	.	.
Y	Y Y Y Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.
D	D D D D	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P	P P P P	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E	A E A A	.	.	3	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.
D	V V E E	.	.	.	.	1	.	.	2	.	.	.	2	.	.	.	.	.	.	.	.
G	G G G G	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	D D D D	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P	P P P P	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	D T D D	.	.	.	.	4	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
D	N Q N N	.	.	.	.	1	3	.	.	.	1	.	.	.	.	.	.	.	.	.	.
G	G N G G	4	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V	V I V V	.	.	.	.	.	.	.	.	.	.	.	4	.	1	.	.	.	.	.	.
N	E P K K	.	.	.	1	.	1	.	1	2	.	.	.	.	.	.	.	.	.	.	.
P	P P P P	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G	G G G G	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T	T T T T	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.
D	E K S A	.	1	1	.	1	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.
F	F F F F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.
:	: : : :																				

α
β
c
pick maximal unit => current prediction
s^0 input layer
s^1 first or hidden layer
s^2 second or output layer

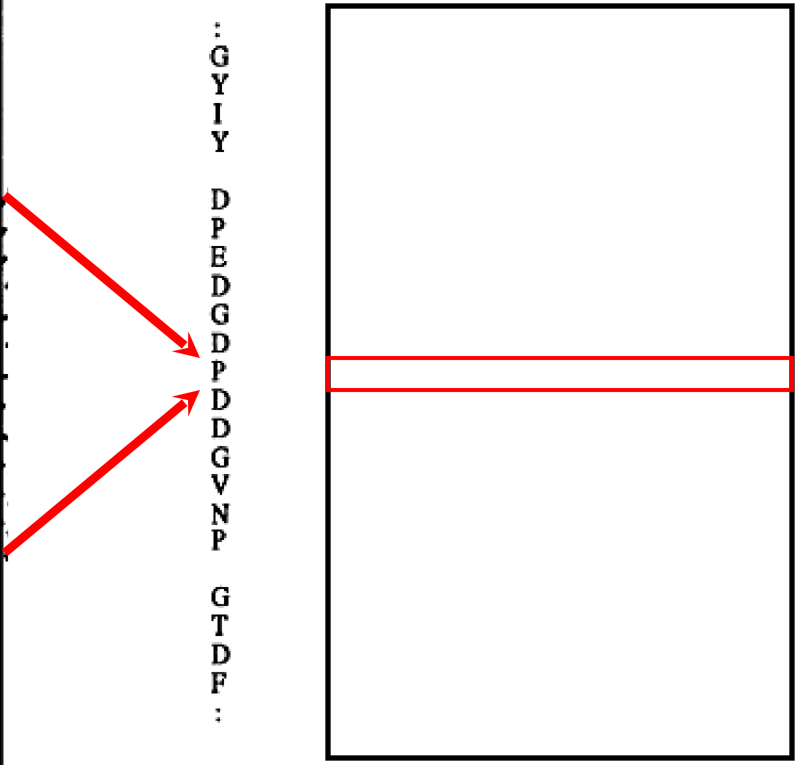
PHD: Input

For each residue, consider a window of size 13:

13x20=260 values

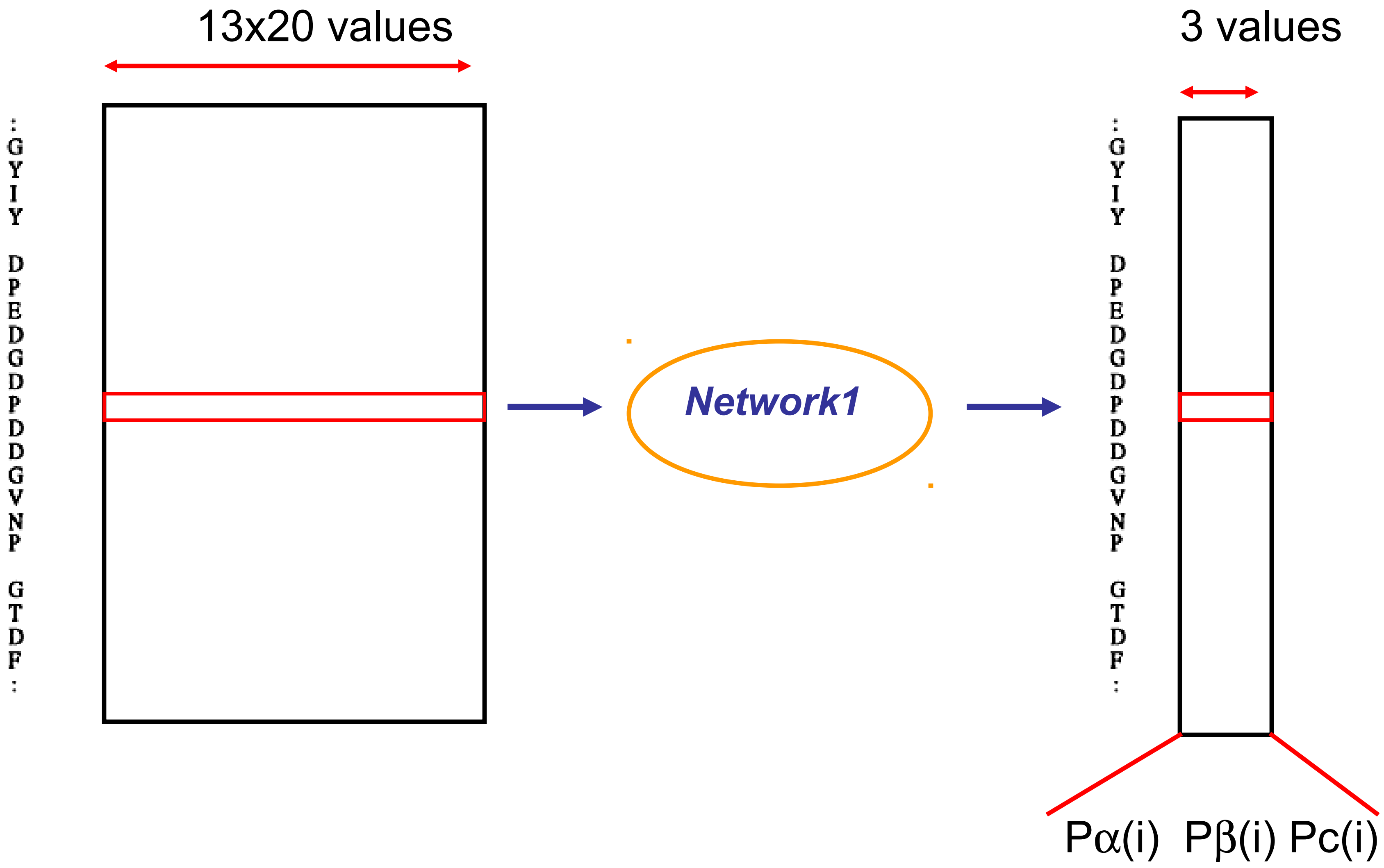


Protein	Alignments	profile table																			
		G	S	A	P	D	N	T	E	K	Q	C	V	H	I	R	L	M	Y	F	W
:	: : : :																				
G	G G G G	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y	Y Y Y Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.
I	I I E E	.	.	.	.	.	.	.	2	.	.	.	.	.	3	.	.	.	.	.	.
Y	Y Y Y Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.
D	D D D D	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P	P P P P	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E	A E A A	.	.	3	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.
D	V V E E	.	.	.	.	1	.	.	2	.	.	.	2	.	.	.	.	.	.	.	.
G	G G G G	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	D D D D	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P	P P P P	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	D T D D	.	.	.	.	4	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
D	N Q N N	.	.	.	.	1	3	.	.	.	1	.	.	.	.	.	.	.	.	.	.
G	G N G G	4	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V	V I V V	.	.	.	.	.	.	.	.	.	.	.	4	.	1	.	.	.	.	.	.
N	E P K K	.	.	.	1	.	1	.	1	2	.	.	.	.	.	.	.	.	.	.	.
P	P P P P	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G	G G G G	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T	T T T T	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.
D	E K S A	.	1	1	.	1	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.
F	F F F F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.
:	: : : :																				



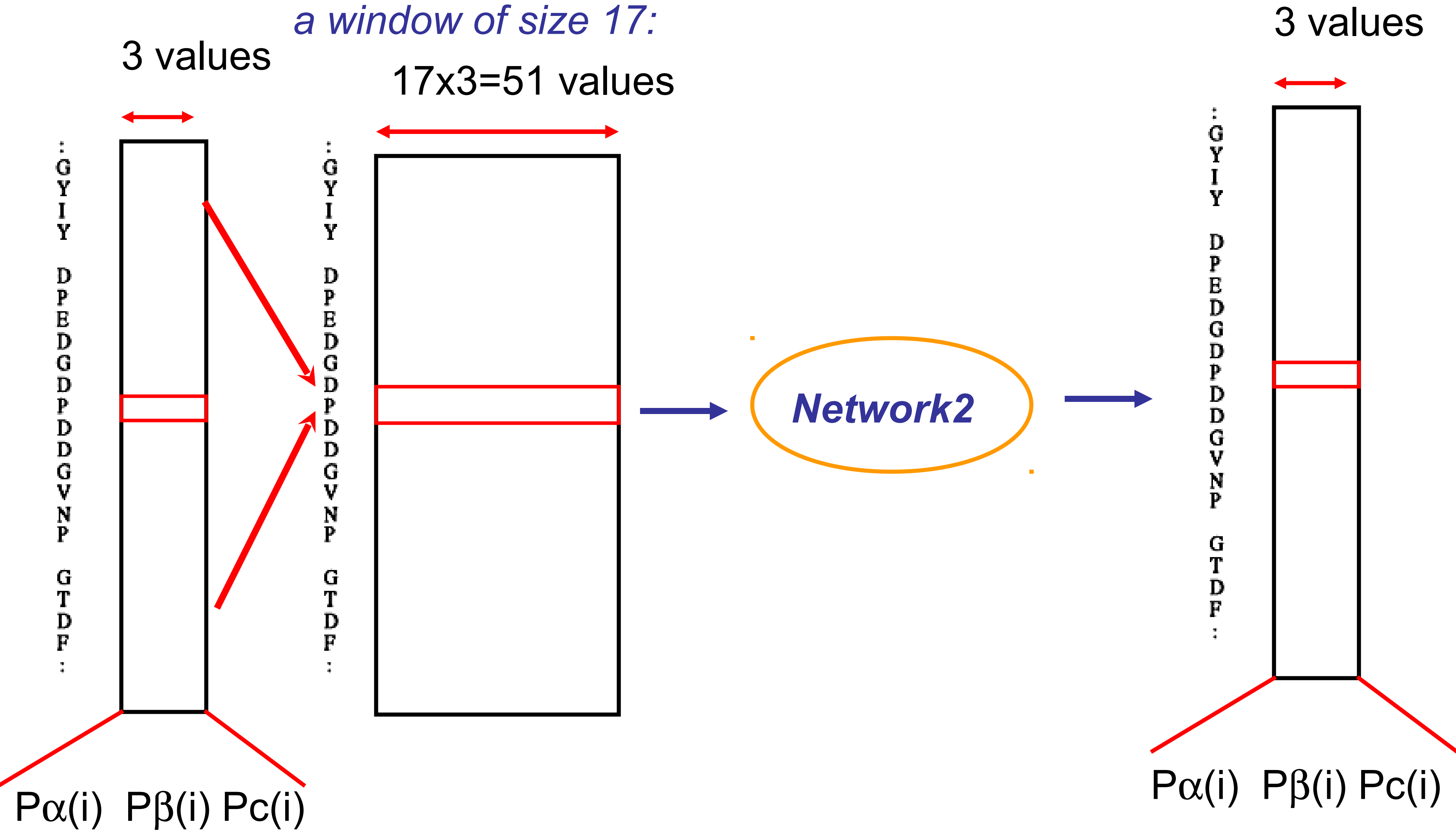
PHD: Network 1

Sequence → Structure



PHD: Network 2
Structure → Structure

For each residue, consider a window of size 17:



PHD

- Sequence-Structure network: for each amino acid aj, a window of 13 residues aj-6…aj…aj+6 is considered. The corresponding rows of the sequence profile are fed into the neural network, and the output is 3 probabilities for aj: P(aj,alpha), P(aj, beta) and P(aj,other)

- Structure-Structure network: For each aj, PHD considers now a window of 17 residues; the probabilities P(ak,alpha), P(ak,beta) and P(ak,other) for k in [j-8,j+8] are fed into the second layer neural network, which again produces probabilities that residue aj is in each of the 3 possible conformation

- Jury system: PHD has trained several neural networks with different training sets; all neural networks are applied to the test sequence, and results are averaged

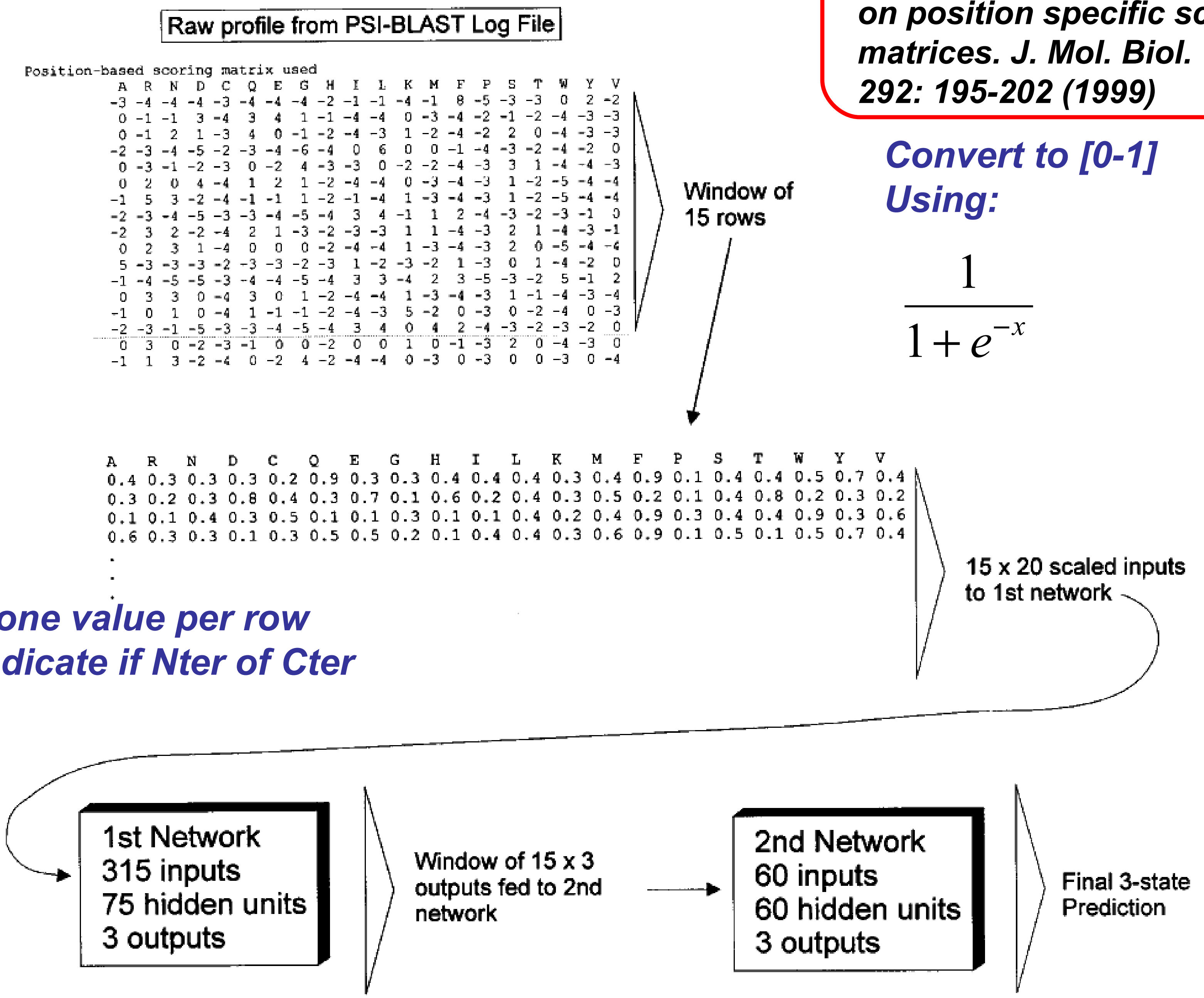
- Prediction: For each position, the secondary structure with the highest average score is output as the prediction

PSIPRED

Jones. Protein secondary structure prediction based on position specific scoring matrices. J. Mol. Biol. 292: 195-202 (1999)

Convert to [0-1] Using:

$$\frac{1}{1+e^{-x}}$$



Raw profile from PSI-BLAST Log File

```
Position-based scoring matrix used
  A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V
 -3 -4 -4 -4 -3 -4 -4 -4 -2 -1 -1 -4 -1  8 -5 -3 -3  0  2 -2
  0 -1 -1  3 -4  3  4  1 -1 -4 -4  0 -3 -4 -2 -1 -2 -4 -3 -3
  0 -1  2  1 -3  4  0 -1 -2 -4 -3  1 -2 -4 -2  2  0 -4 -3 -3
 -2 -3 -4 -5 -2 -3 -4 -6 -4  0  6  0  0 -1 -4 -3 -2 -4 -2  0
  0 -3 -1 -2 -3  0 -2  4 -3 -3  0 -2 -2 -4 -3  3  1 -4 -4 -3
  0  2  0  4 -4  1  2  1 -2 -4 -4  0 -3 -4 -3  1 -2 -5 -4 -4
 -1  5  3 -2 -4 -1 -1  1 -2 -1 -4  1 -3 -4 -3  1 -2 -5 -4 -4
 -2 -3 -4 -5 -3 -3 -4 -5 -4  3  4 -1  1  2 -4 -3 -2 -3 -1  0
 -2  3  2 -2 -4  2  1 -3 -2 -3 -3  1  1 -4 -3  2  1 -4 -3 -1
  0  2  3  1 -4  0  0  0 -2 -4 -4  1 -3 -4 -3  2  0 -5 -4 -4
  5 -3 -3 -3 -2 -3 -3 -2 -3  1 -2 -3 -2  1 -3  0  1 -4 -2  0
 -1 -4 -5 -5 -3 -4 -4 -5 -4  3  3 -4  2  3 -5 -3 -2  5 -1  2
  0  3  3  0 -4  3  0  1 -2 -4 -4  1 -3 -4 -3  1 -1 -4 -3 -4
 -1  0  1  0 -4  1 -1 -1 -2 -4 -3  5 -2  0 -3  0 -2 -4  0 -3
 -2 -3 -1 -5 -3 -3 -4 -5 -4  3  4  0  4  2 -4 -3 -2 -3 -2  0
  0  3  0 -2 -3 -1  0  0 -2  0  0  1  0 -1 -3  2  0 -4 -3  0
 -1  1  3 -2 -4  0 -2  4 -2 -4 -4  0 -3  0 -3  0  0 -3  0 -4
```

Window of 15 rows

```
A   R   N   D   C   Q   E   G   H   I   L   K   M   F   P   S   T   W   Y   V
0.4 0.3 0.3 0.3 0.2 0.9 0.3 0.3 0.4 0.4 0.4 0.3 0.4 0.9 0.1 0.4 0.4 0.5 0.7 0.4
0.3 0.2 0.3 0.8 0.4 0.3 0.7 0.1 0.6 0.2 0.4 0.3 0.5 0.2 0.1 0.4 0.8 0.2 0.3 0.2
0.1 0.1 0.4 0.3 0.5 0.1 0.1 0.3 0.1 0.1 0.4 0.2 0.4 0.9 0.3 0.4 0.4 0.9 0.3 0.6
0.6 0.3 0.3 0.1 0.3 0.5 0.5 0.2 0.1 0.4 0.4 0.3 0.6 0.9 0.1 0.5 0.1 0.5 0.7 0.4
.
.
.
```

15 x 20 scaled inputs to 1st network

1st Network
315 inputs
75 hidden units
3 outputs

Window of 15 x 3 outputs fed to 2nd network

2nd Network
60 inputs
60 hidden units
3 outputs

Final 3-state Prediction

Add one value per row to indicate if Nter of Cter

Performances (monitored at CASP)

CASP	YEAR	# of Targets	<Q3>	Group
CASP1	1994	6	63	Rost and Sander
CASP2	1996	24	70	Rost
CASP3	1998	18	75	Jones
CASP4	2000	28	80	Jones

Secondary Structure Prediction

-*Available servers*:

- JPRED : http://www.compbio.dundee.ac.uk/~www-jpred/
- PHD: http://cubic.bioc.columbia.edu/predictprotein/
- PSIPRED: http://bioinf.cs.ucl.ac.uk/psipred/
- NNPREDICT: http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html
- Chou and Fassman: http://fasta.bioch.virginia.edu/fasta_www/chofas.htm

-*Interesting paper:*

- *Rost and Eyrich. EVA: Large-scale analysis of secondary structure prediction. Proteins 5:192-199 (2001)*

Protein Structure Prediction

- *One popular model for protein folding assumes a sequence of events:*
  - Hydrophobic collapse
  - Local interactions stabilize secondary structures
  - Secondary structures interact to form motifs
  - Motifs aggregate to form tertiary structure

Protein Structure Prediction

A physics-based approach:

- find conformation of protein corresponding to a thermodynamics minimum (free energy minimum)

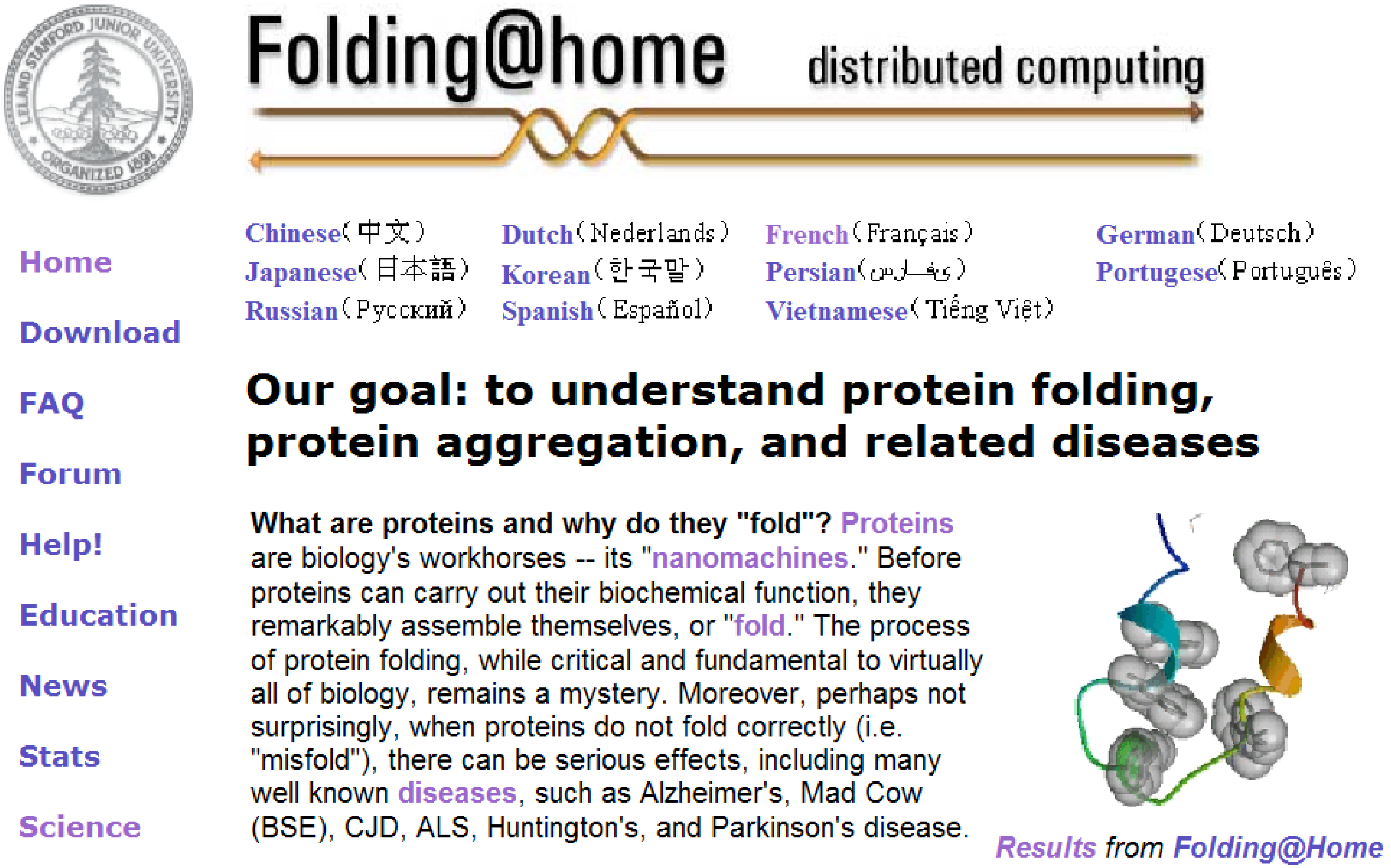
- cannot minimize internal energy alone! Needs to include solvent

- simulate folding…a very long process!

Folding time are in the ms to second time range
Folding simulations at best run 1 ns in one day…

The Folding @ Home initiative

(Vijay Pande, Stanford University)



Folding@home distributed computing

Chinese（中文） Dutch（Nederlands） French（Français） German（Deutsch）
Japanese（日本語） Korean（한국말） Persian（فارسی） Portugese（Português）
Russian（Русский） Spanish（Español） Vietnamese（Tiếng Việt）

Home

Download

FAQ

Forum

Help!

Education

News

Stats

Science

Our goal: to understand protein folding, protein aggregation, and related diseases

What are proteins and why do they "fold"? Proteins are biology's workhorses -- its "nanomachines." Before proteins can carry out their biochemical function, they remarkably assemble themselves, or "fold." The process of protein folding, while critical and fundamental to virtually all of biology, remains a mystery. Moreover, perhaps not surprisingly, when proteins do not fold correctly (i.e. "misfold"), there can be serious effects, including many well known diseases, such as Alzheimer's, Mad Cow (BSE), CJD, ALS, Huntington's, and Parkinson's disease.

Results from Folding@Home

http://folding.stanford.edu/

The Folding @ Home initiative

What does Folding@Home do? Folding@Home is a distributed computing project which studies protein folding, misfolding, aggregation, and related diseases. We use novel computational methods and large scale distributed computing, to simulate timescales thousands to millions of times longer than previously achieved. This has allowed us to simulate folding for the first time, and to now direct our approach to examine folding related disease.

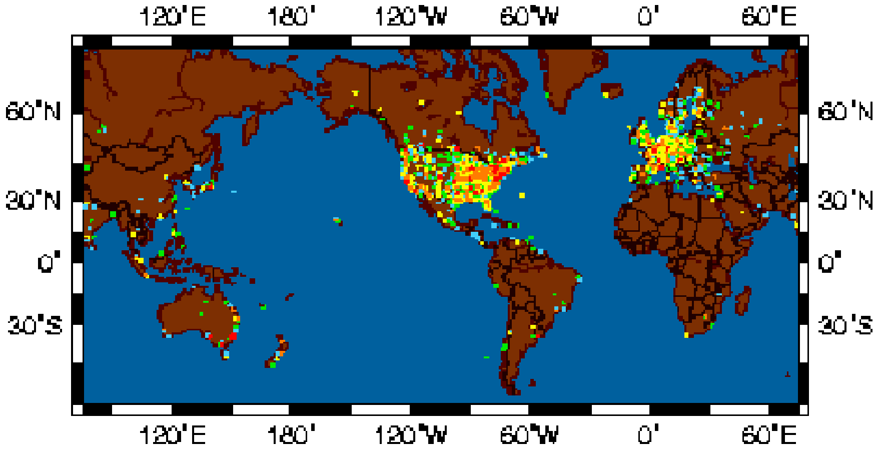
F@H exhibit

See Prof. Pande's lecture on F@H at Xerox PARC

How can you help? You can help our project by downloading and running our client software. Our algorithms are designed such that for every computer that joins the project, we get a commensurate increase in simulation speed.

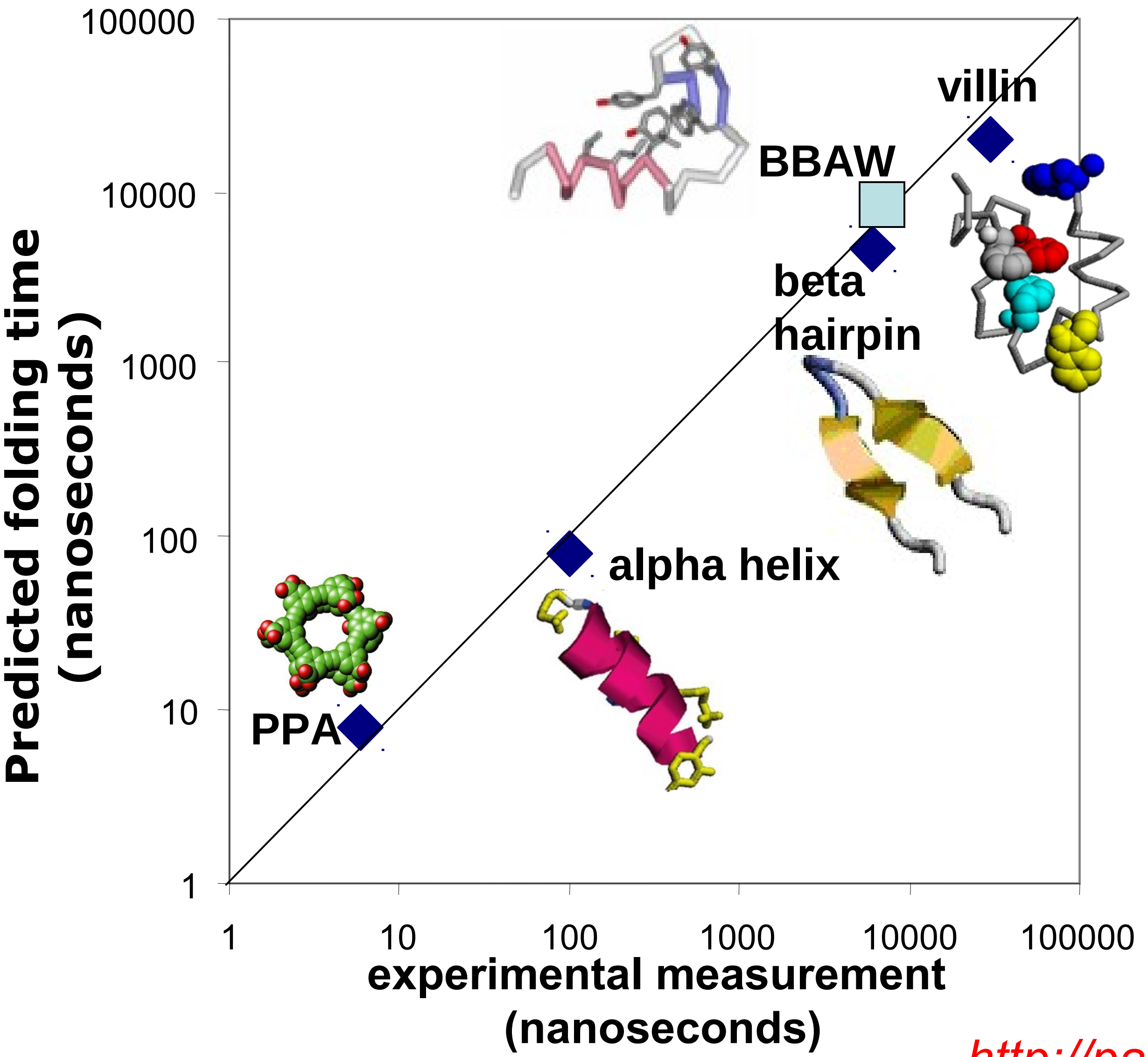
One can also help by donating funds to the project, via Stanford University.

What have we done so far? We have had several successes. You can read about them on our Science page, Results section, or go directly to our press and papers page.



*Since October 1, 2000, over 1,000,000 CPUs **throughout the world** have participated in Folding@Home. Each additional CPU gives us an added boost in performance, allowing us to tackle more difficult problems or solve existing research faster or more accurately.*

Folding @ Home: Results



Experiments:

villin:
Raleigh, et al,
SUNY, Stony Brook

BBAW:
Gruebele, et al, UIUC

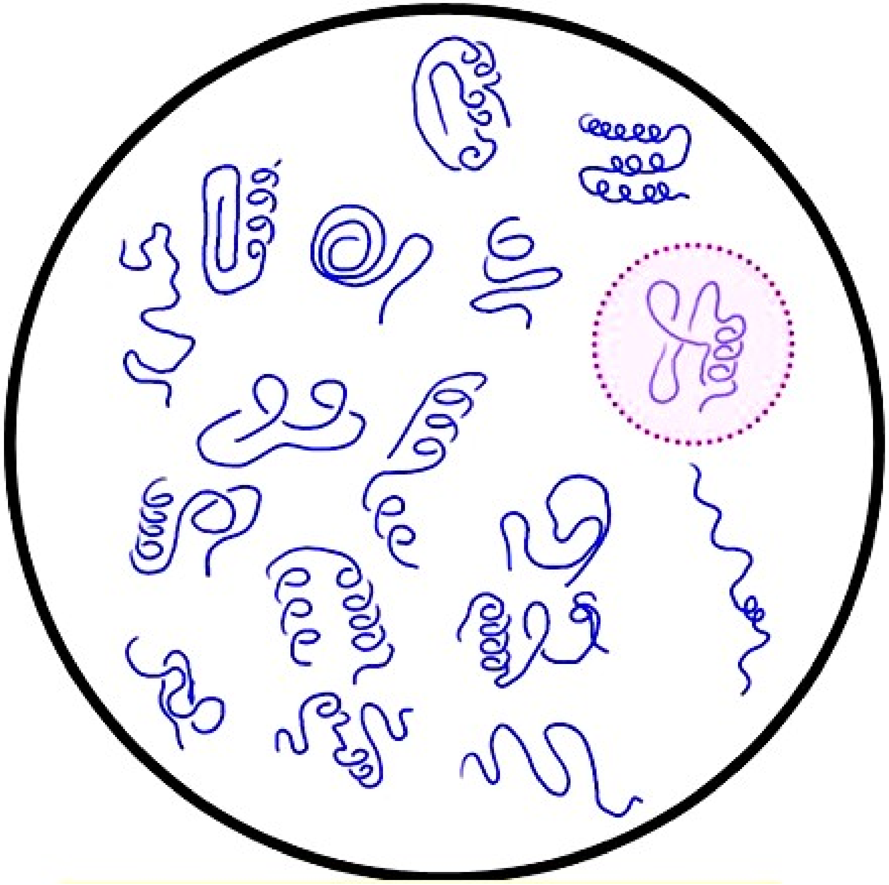
beta hairpin:
Eaton, et al, NIH

alpha helix:
Eaton, et al, NIH

PPA:
Gruebele, et al, UIUC

http://pande.stanford.edu/

Protein Structure Prediction



DECOYS:

Generate a large number of possible shapes

DISCRIMINATION:

Select the correct, native-like fold

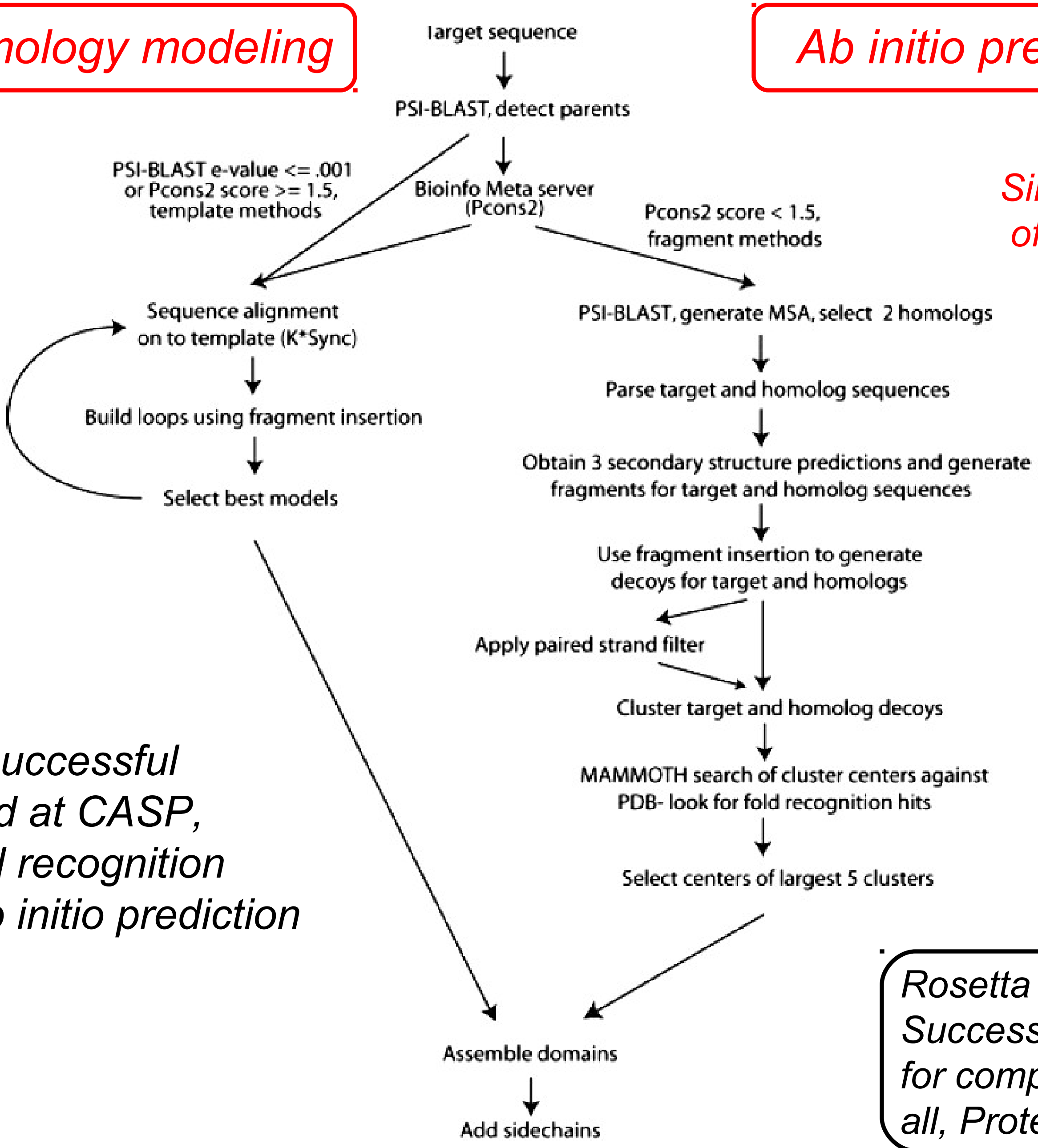
Need good decoy structures

Need a good energy function

ROSETTA at CASP (David Baker)

Homology modeling

Ab initio prediction



Target sequence

PSI-BLAST, detect parents

PSI-BLAST e-value <= .001
or Pcons2 score >= 1.5,
template methods

Bioinfo Meta server
(Pcons2)

Pcons2 score < 1.5,
fragment methods

Sequence alignment
on to template (K*Sync)

Build loops using fragment insertion

Select best models

PSI-BLAST, generate MSA, select 2 homologs

Parse target and homolog sequences

Obtain 3 secondary structure predictions and generate
fragments for target and homolog sequences

Use fragment insertion to generate
decoys for target and homologs

Apply paired strand filter

Cluster target and homolog decoys

MAMMOTH search of cluster centers against
PDB- look for fold recognition hits

Select centers of largest 5 clusters

Assemble domains

Add sidechains

Simultaneous modeling of the target and 2 homologs

Secondary structure prediction

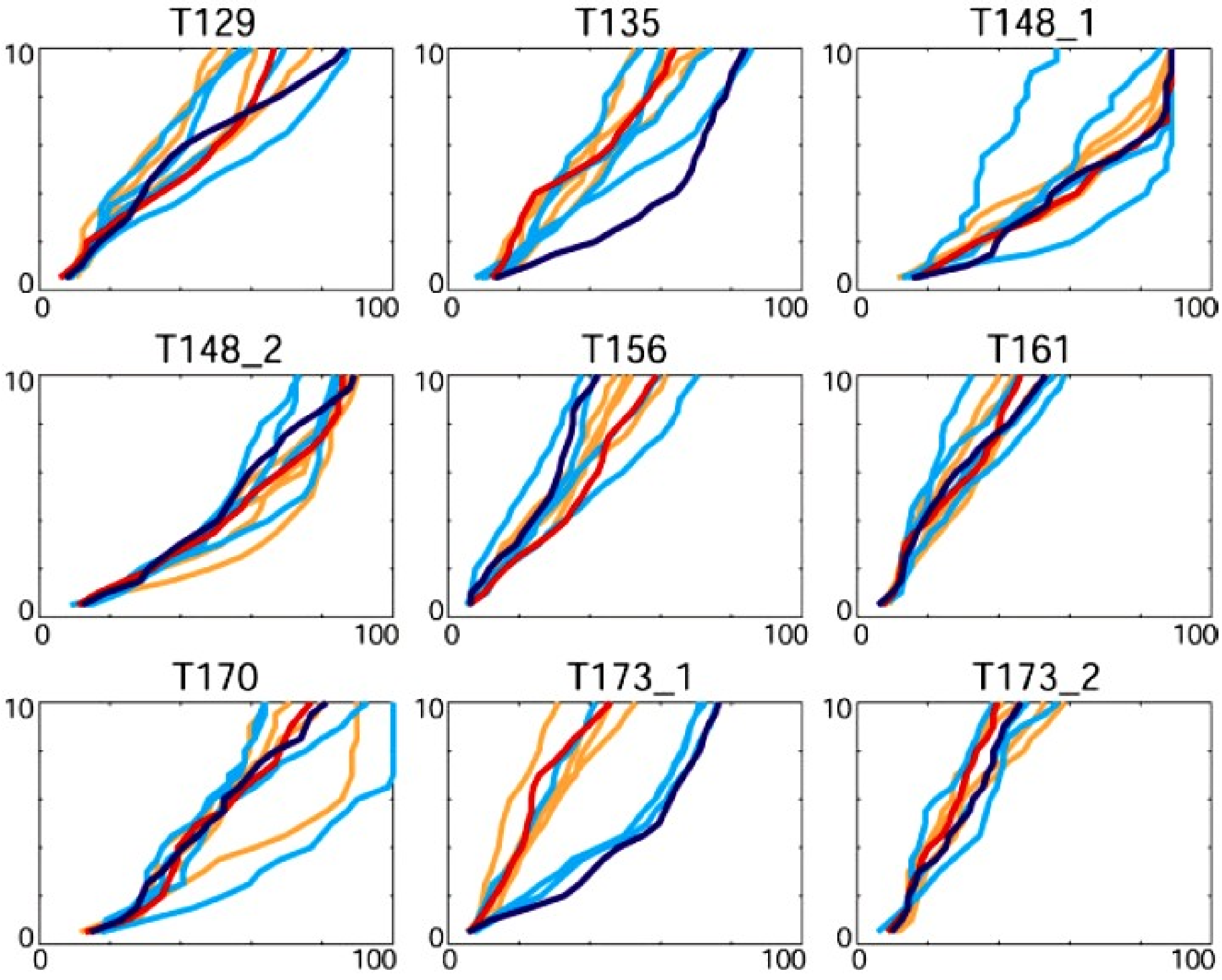
Fragment based approach to generate decoys

Select 5 decoys For prediction

Most successful Method at CASP, for fold recognition and ab initio prediction

Rosetta predictions in CASP5: Successes, failures, and prospect for complete automation. Baker et all, Proteins, 53:457-468 (2003)

ROSETTA results at CASP5



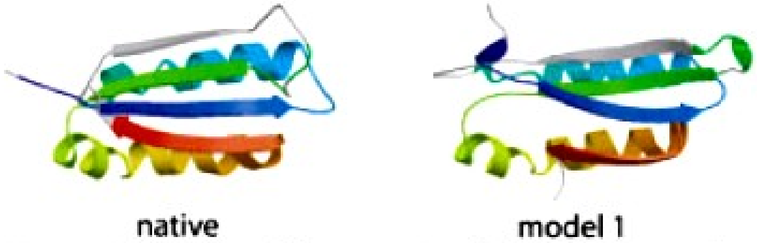
cRMS (model – experimental structure) cutoff (Å)

% of the full target protein

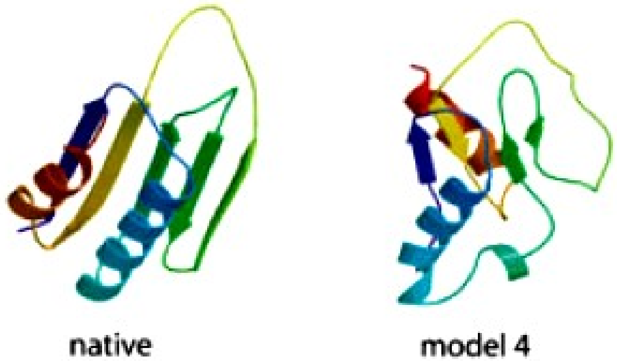
Blue: "human"

Orange: "automatic Server"

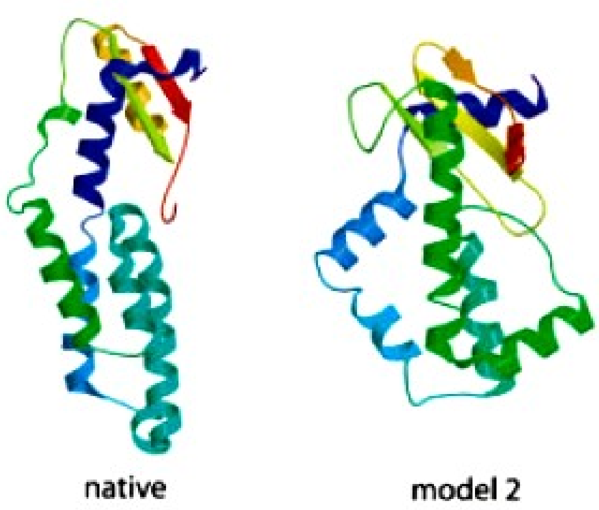
ROSETTA results at CASP5



T135: Boiling stable protein (full chain 1-108)



T149: yjiA (C-terminal domain, 206-318)



T161: HI1480 (full chain, 1-156)

		# of residues with cRMS below 4Å/6Å		
Name	Length	human	Automatic	Best decoy
T135	106	83/98	54/64	94/105
T149	116	52/71	44/62	76/92
T161	154	45/83	57/79	55/95

Rosetta predictions in CASP5: Successes, failures, and prospect for complete automation. Baker et all, Proteins, 53:457-468 (2003)