BMC WorkShop

Protein Structure Prediction

Sequence-Structure alignment

(template selection)

Marc A. Marti-Renom & Damien Devos

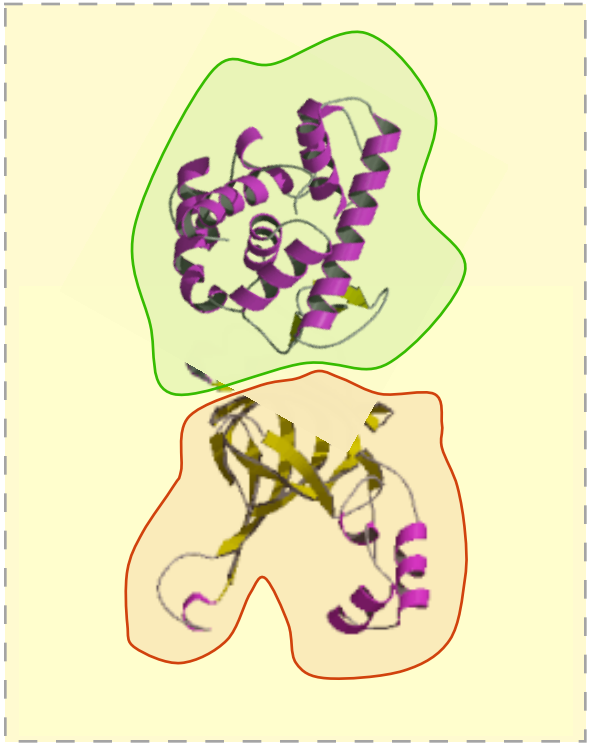
Department of Biopharmaceutical Sciences, UCSF

June 17th and 18th, 2004

Domains (?)

Domain boundaries from sequence

VERY DIFFICULT!!!!



MENFEIWVEKYRPRTLDEVVGQDEVIQRLKGYVERKNIPHLLFSGPPGTGKTATAIALARDLFGENWRDN
FIEMNASDERGIDVVRHKIKEFARTAPIGGAPFKIIFLDEADALTADAQAALRRTMEMYSKSCRFILSCN
YVSRIIEPIQSRCAVFRFKPVPKEAMKKRLLEICEKEGVKITEDGLEALIYISGGDFRKAINALQGAAAI
GEVVDADTIYQITATARPEEMTELIQTALKGNFMEARELLDRLMVEYGMSGEDIVAQLFREIISMPIKDS
LKVQLIDKLGEVDFRLTEGANERIQLDAYLAYLSTLAKK

Domain boundaries from sequence (SnapDragon)

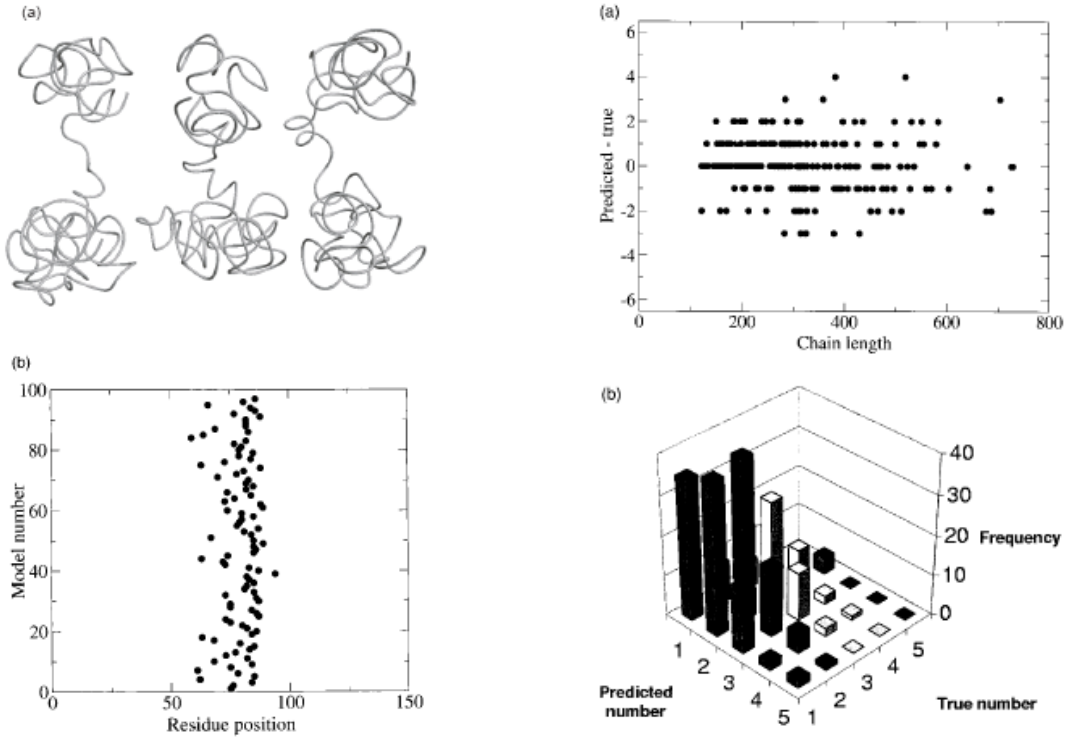
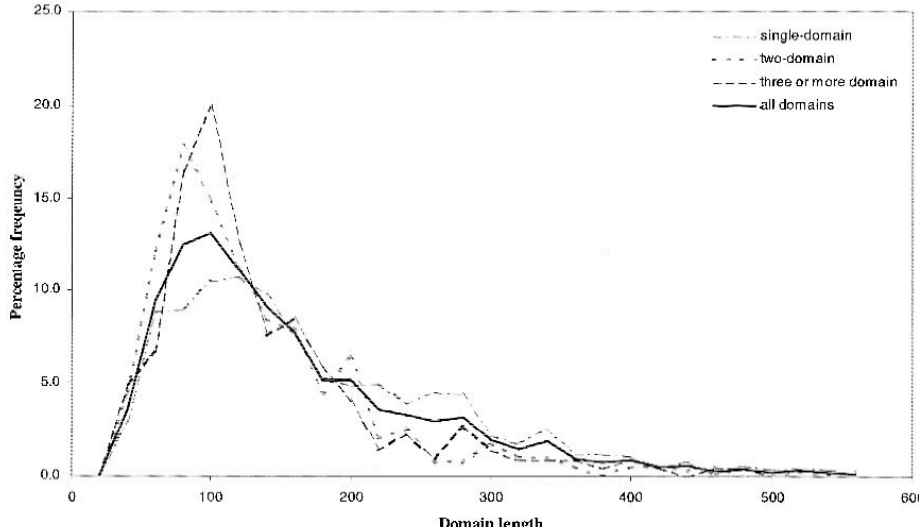


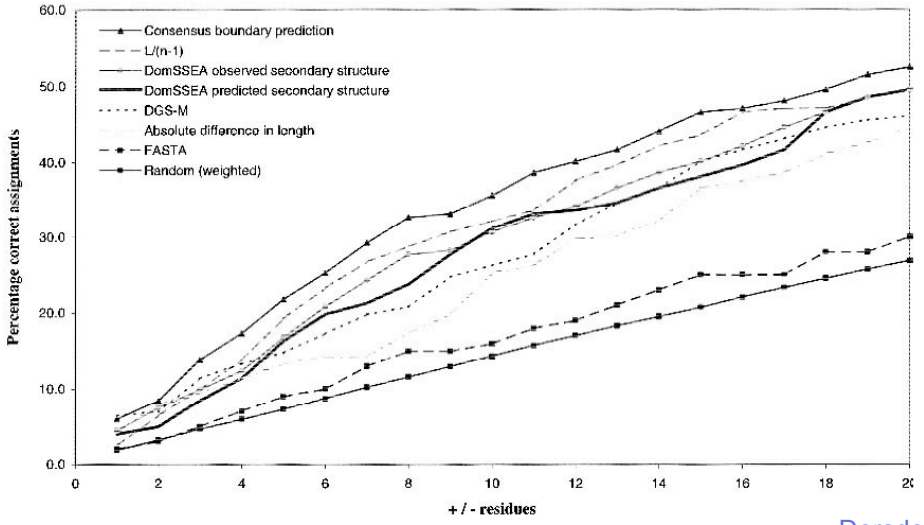
Table 2. Average accuracy percentages of linker prediction over 57 proteins

		Continuous set	Discontinuous set	Full set
Randomised background Z-score >2	Coverage	63.3	43.6	54.8
	Success	27.2	31.1	28.9
Self-normalised Z-score >1	Coverage	64.7	39.5	53.5
	Success	26.6	31.7	28.9
Self-normalised Z-score >2	Coverage	48.7	24.3	38.7
	Success	41.3	28.3	29.9

Georgea and Heringa (2002) *J. Mol. Biol.* **316** pp839

Domain boundaries from sequence (DomSSEA)



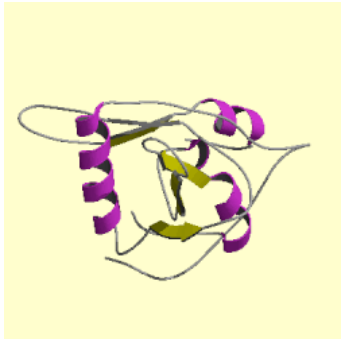


Dersden *et al.* (2003) *Proty. Science* **11** pp2014

Prediction of Secondary Structure (PSI-PRED)

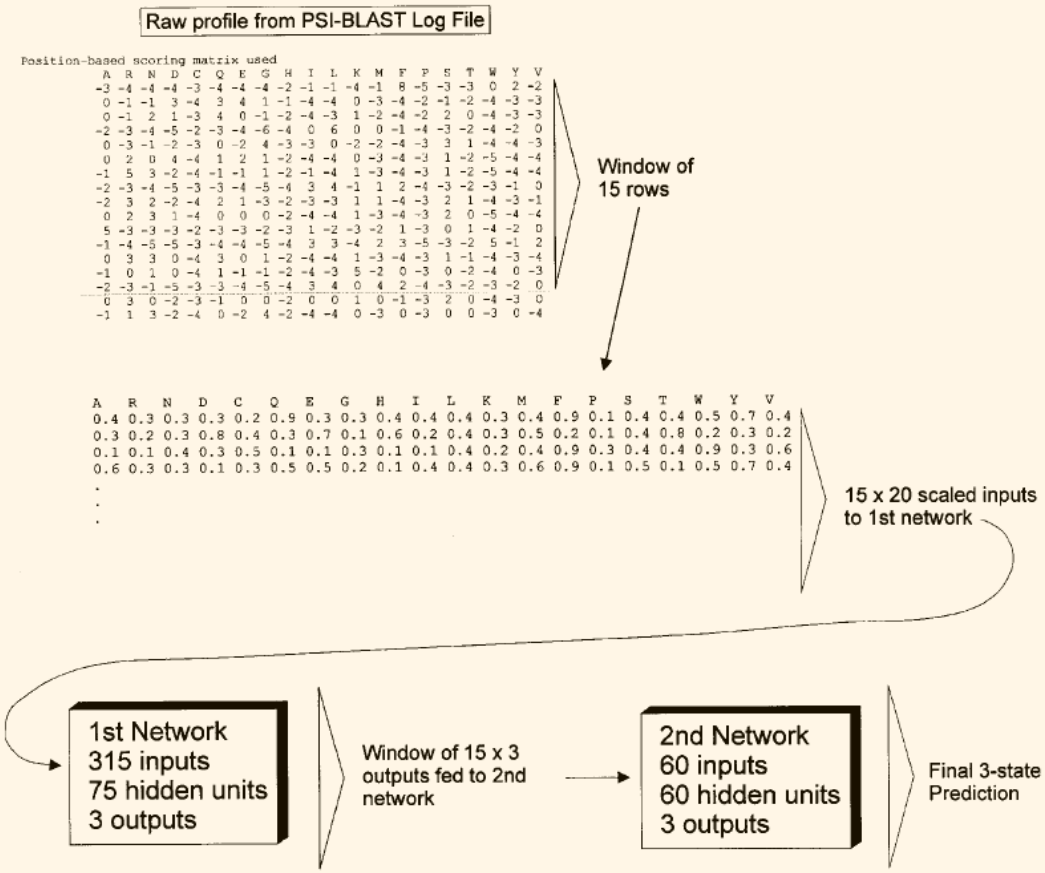
```
>gi42541361
MDIRSVSSLRGLLCLPPSWPRR
```

- Neural Network



✓ **Very simple idea**
✓ **Simple scoring**

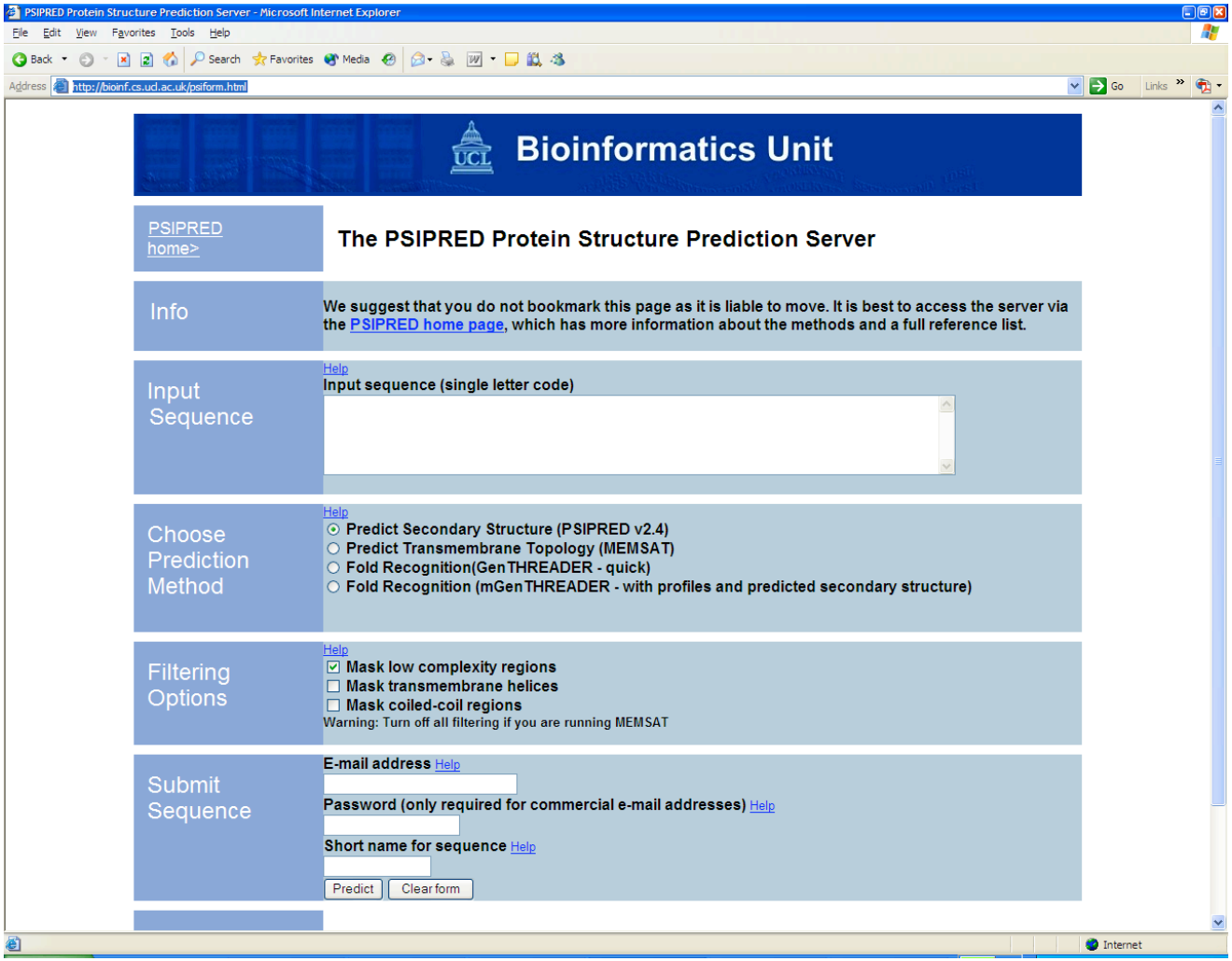
Obscure optimizer



Jones DT. (1999) *J. Mol. Biol.* **292** pp195

Prediction of Secondary Structure (PSI-PRED)

http://bioinf.cs.ucl.ac.uk/psiform.html



Why the alignment is so important?

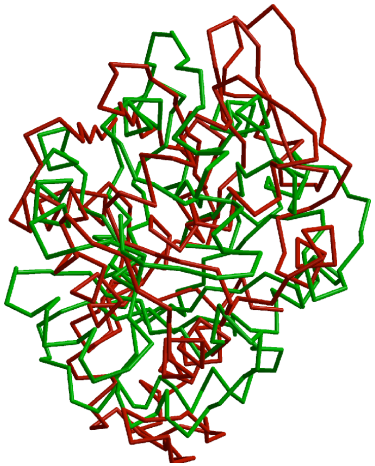
Typical errors in comparative models

MODEL

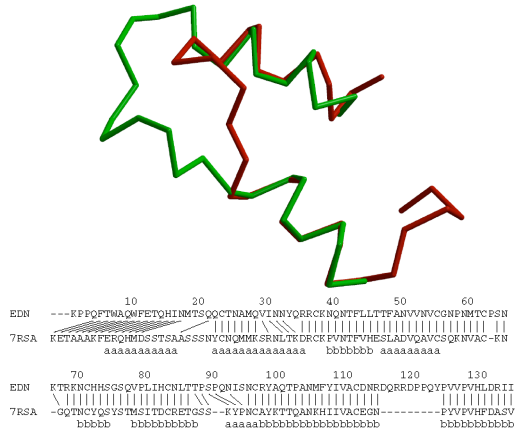
X-RAY

TEMPLATE

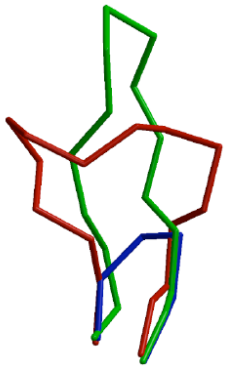
Incorrect template



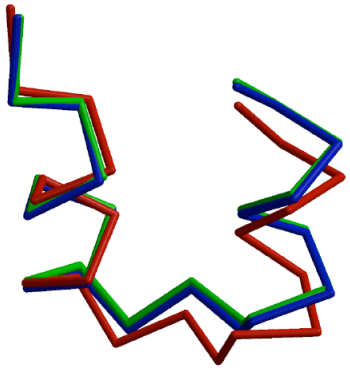
Misalignment



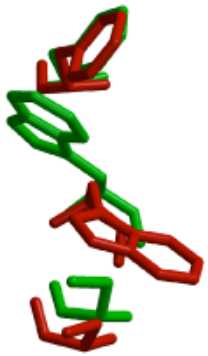
Region without a template



Distortion/shifts in aligned regions

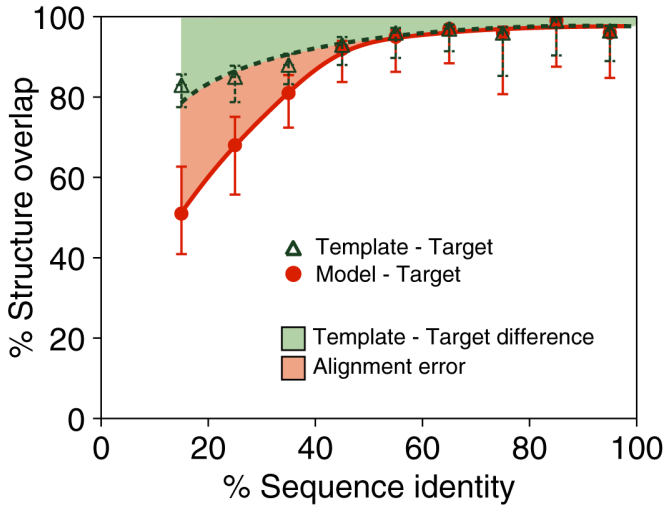


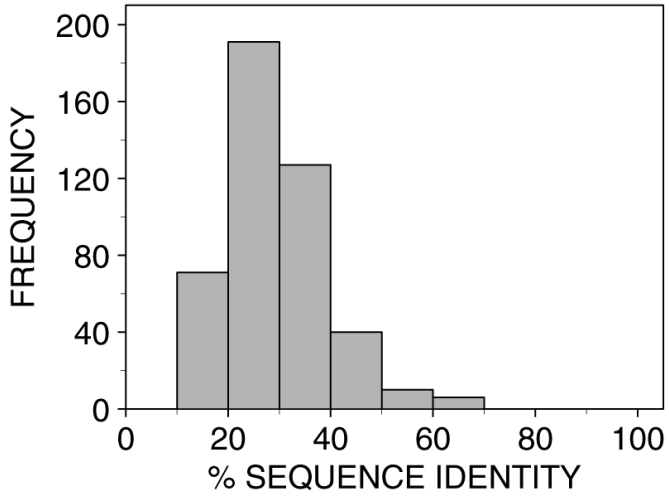
Sidechain packing



Marti-Renom et al. Annu.Rev.Biophys.Biomol.Struct. 29, 291-325, 2000.

Alignment errors are frequent and large





R. Sánchez & A. Šali, *Proc. Natl. Acad. Sci. USA* **95**, 13597, 1998.

Minimizing errors in sequence-structure alignment

- Threading.
- Complex gap penalty functions.
- Multiple sequence profiles.
- Iterative process (model assessment)

Threading

General overview (Threading)

- Matches sequences to 3D structures
  - Requires a scoring function to asses the fit of a sequence to a given fold
  - Scoring functions deried from known structures and include atom contact and solvation terms evaluated in a pairwise fashion
  - May include secondary structure terms, multiple alignments…

- Threading servers available using several different approaches
  - Fold recognition server at Imperial College, UK
    http://www.sbg.bio.ic.ac.uk/~3dpssm/
  - ProteinPredict server at EMBL
    http://www.embl-heidelberg.de/predictprotein/predictprotein.html
  - Protein sequence-structure threading at NCBI
    http://www.ncbi.nlm.nih.gov/Structure/RESEARCH/threading.shtml

Template comparison methods

- Uses 3D “templates” for searching structural databases
  - active site or binding site templates generated to reflect functionally important structural signatures
- Available software/servers
  - Template Search and Superposition (TESS), Thornton Group
    http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html
    Wallace AC; Borkakoti N; Thornton JM. (1997) *Protein Science* **6** pp2308
  - “Fuzzy Functional Forms” , Skolnick - commercial availability
    Fetrow, Js and Skolnick, J (1998) *J. Mo. Biol* **281** pp949
  - Spatial Arrangements of Side-chain and Main-chain (SPASM), Kleywegt, Univ. of Uppsala
    http://portray.bmc.uu.se/cgi-bin/dennis/spasm.pl
    Kleywegt GJ (1999). *J. Mol. Biol.* **285** pp1887

Sequence-Structure alignments

As any other bioinformatics problem…

- **Representation**
- **Scoring**
- **Optimizer**

Empirical energy functions (PMF)

Idea: **energy leads to structure, thus it should be possible to infer energy from many known structures**

To be used in: model refinement and assessment

Properties needed:

- Deep minimum at correct state (native)
- Smooth
- Simple

Types:

- Contact potential
- Distance potentials
- Surface potentials

Approximations/Limitations in PMFs

Database size.

PMF versus Energy (additive/higher order terms).

Reference state.

Physical origin.

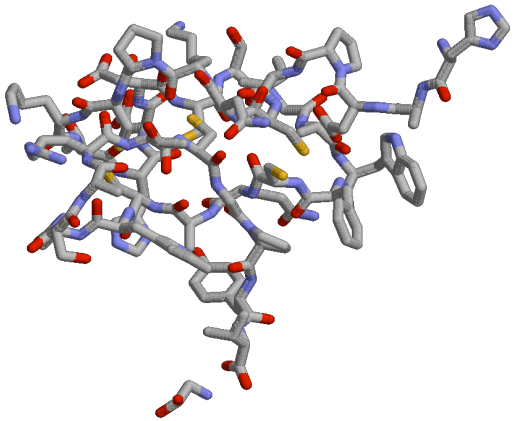
Finkelstein et al. (1995) *Proteins* **23**, pp142

Representation

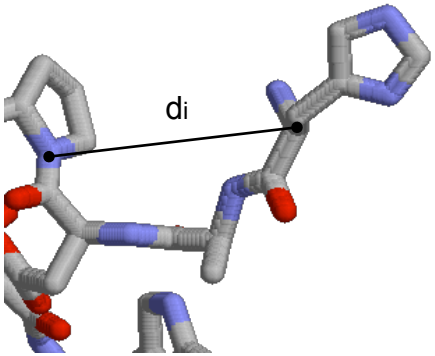
Sequence/Structures

```
>gi42541361
MDIRSVSSLRGLLCLPPSWPRR
```

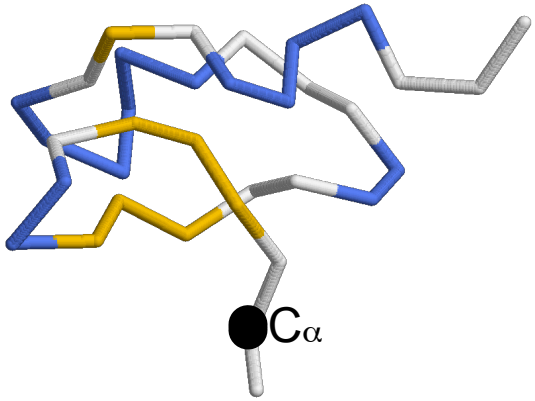
Primary sequence



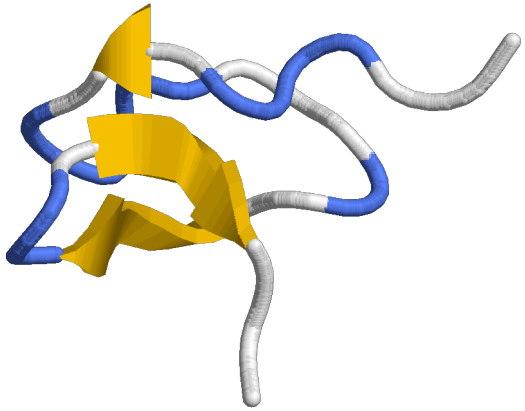
All atoms and coordinates



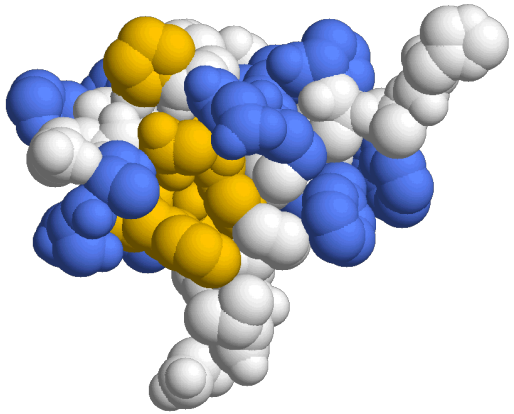
Distance space



Reduced atoms representation



Secondary Structure

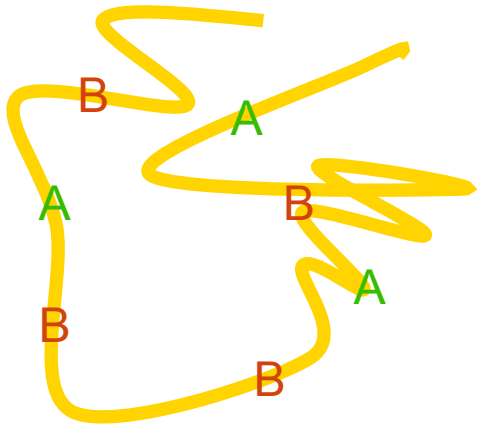


Accessible surface

Statistical Potential… inspiration

$$K = \frac{[AB]}{[A]\cdot[B]}$$

$$\Delta G = -RT\ln(K) = -RT\ln\frac{[AB]}{[A]\cdot[B]}$$



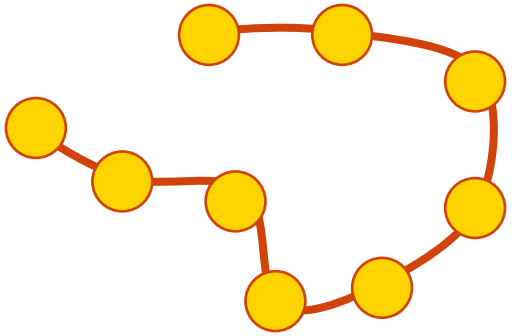
From statistical physics, we know that energy difference between two states (ΔE) and the ratio of their occupancies (N_1:N_2) are related [9]:

$$\Delta E = -kT \ln\left(\frac{N_1}{N_2}\right) \qquad (1)$$

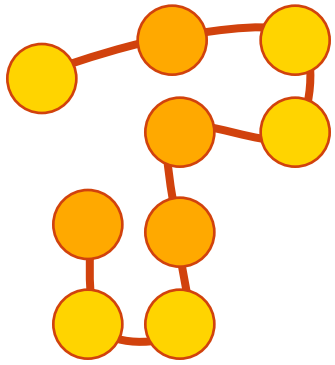
in which T is the absolute temperature and k is the Boltzmann's constant. As we are interested in an interaction energy between two amino acid side chains, it would seem natural to define N_1 as the number of interactions between these two residues types in a group of real protein structures, a number which is readily available from simple database analysis. But this number must be compared with the number of interactions in some other system, N_2, to obtain the energy difference between them.

Tanaka and Sheraga (1975) *PNAS*, **72** pp3802
A. Godzik, (1996) *Structure* **15** pp363

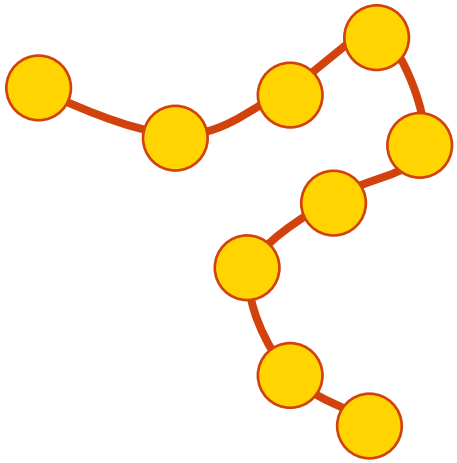
Statistical Potential… interaction types



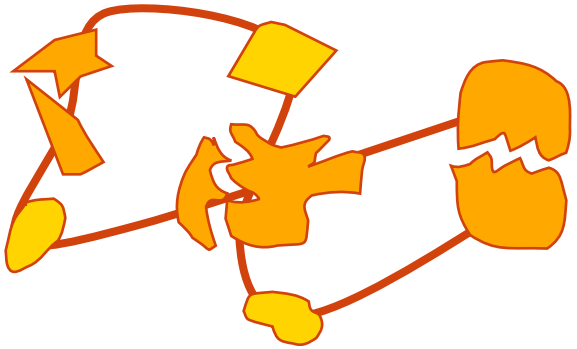
Neutral interactions



Hydrophobic interactions



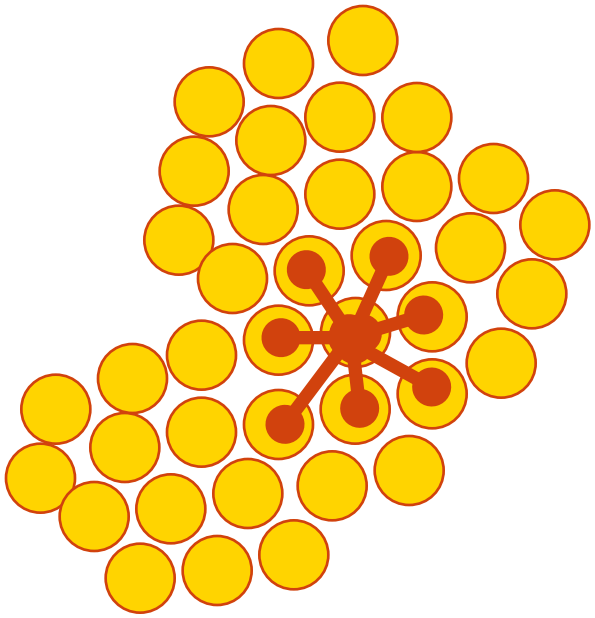
Compact interactions

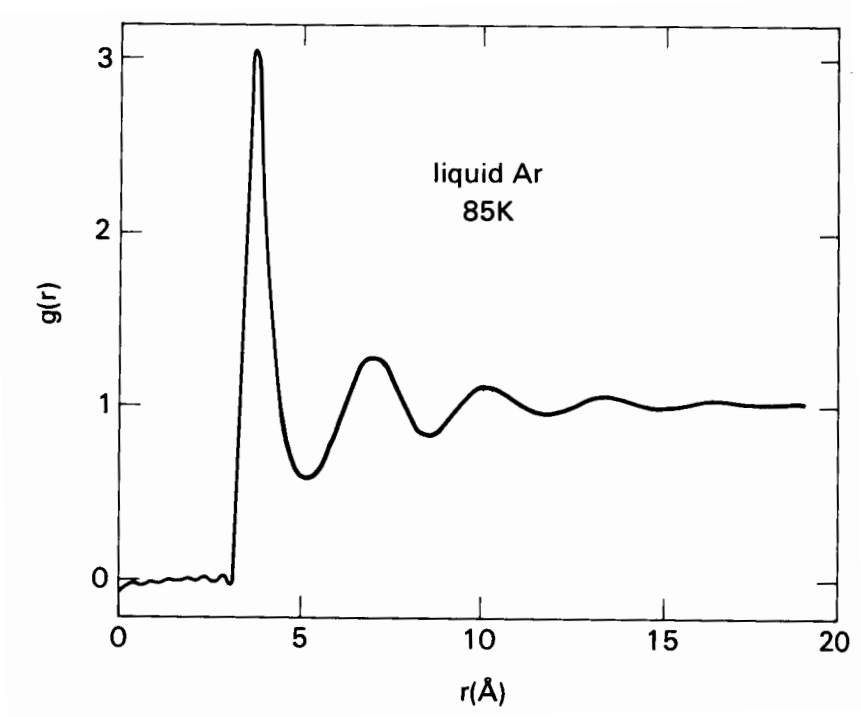


Specific interactions

A. Godzik, (1996) *Structure* **15** pp363

Statistical Potential… reference state

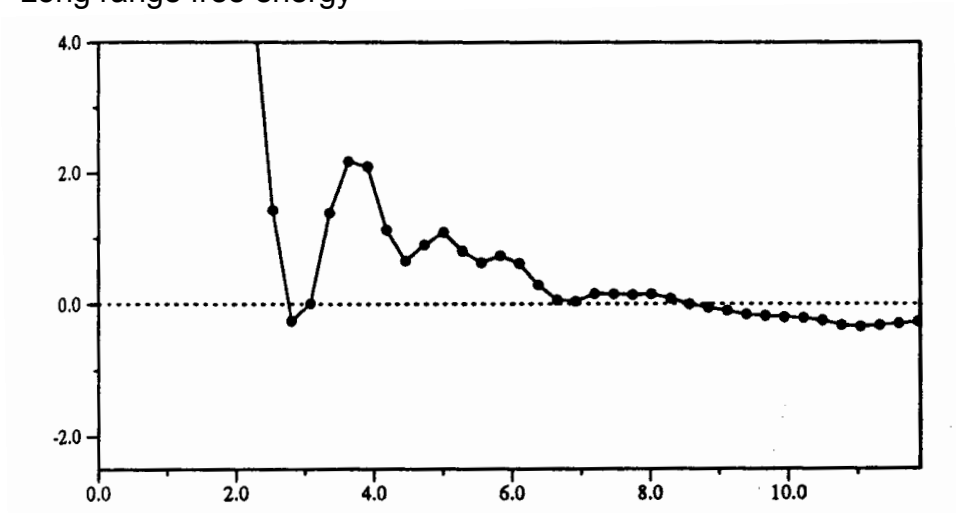




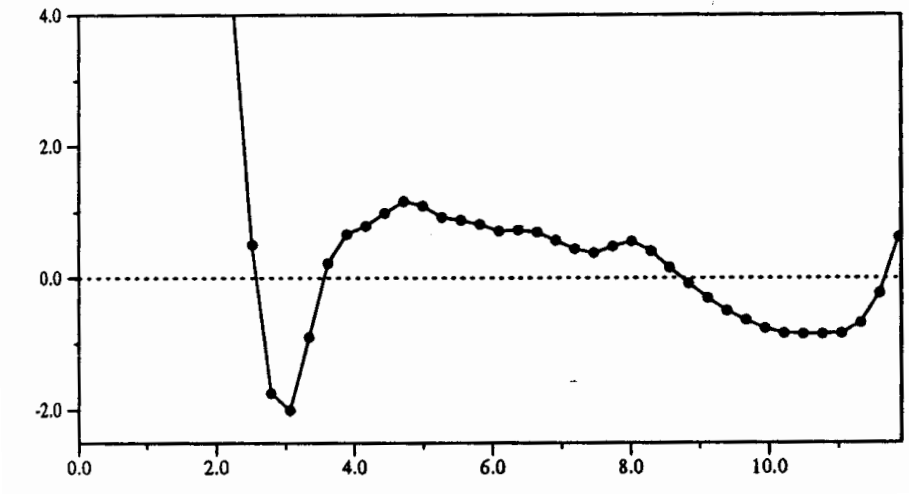
Theory of simple liquids 2nd edition JP Hansen and IR McDonald, Academic Press.

Statistical Potential… Hydrogen Bonds

Long range free energy



Short range free energy



Free energy of the protein backbone hydrogen bond N · · · O compiled from a database of 289 X-ray structures

$$\rho_{NO}(r) = \sum_{ij} \delta(r - r_{ij})$$

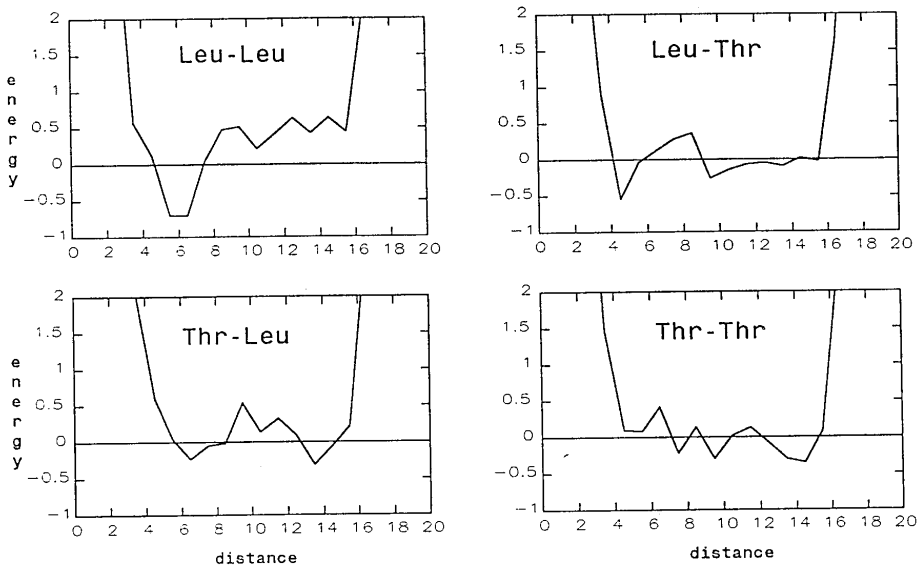
$$g_{NO}(r) = \frac{\rho_{NO}(r)}{\rho^2}$$

$$w_{NO}(r) = -kT \ln\left(g_{NO}(r)\right)$$

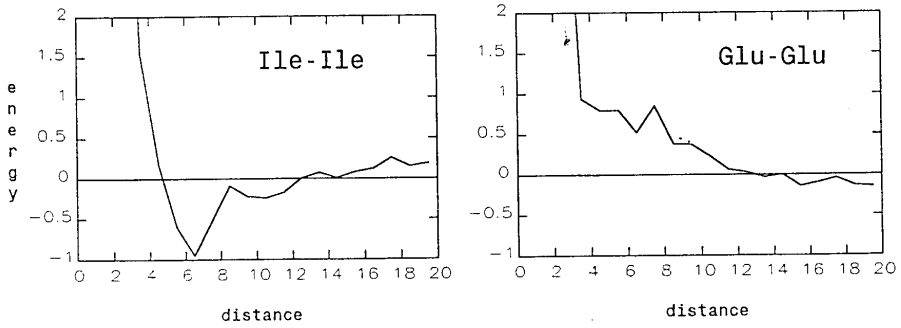
Sippl (1996). *JMB* **260** pp644

Statistical Potential… Distance Potentials

Long range free energy



Short range free energy



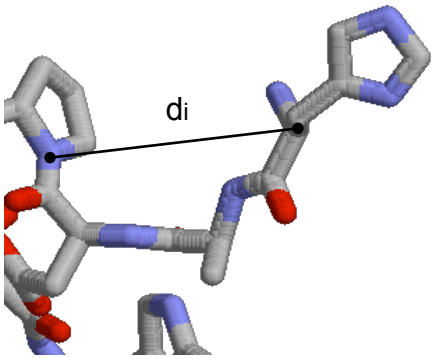
Sippl (1993). *JCAM* **7** pp473

Scoring

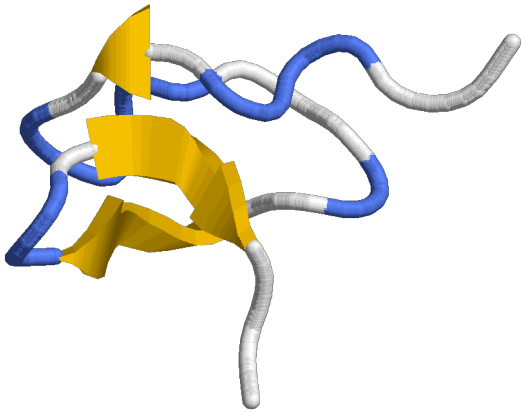
Raw scores of an alignment

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	**9**	-1	-1	-3	0	-3	-3	-3	-4	-3	-3	-3	-3	-1	-1	-1	-1	-2	-2	-2
S	-1	**4**	1	-1	1	0	1	0	0	0	-1	-1	0	-1	-2	-2	-2	-2	-2	-3
T	-1	1	**4**	1	-1	1	0	1	0	0	0	-1	0	-1	-2	-2	-2	-2	-2	-3
P	-3	-1	1	**7**	-1	-2	-1	-1	-1	-1	-2	-2	-1	-2	-3	-3	-2	-4	-3	-4
A	0	1	-1	-1	**4**	0	-1	-2	-1	-1	-2	-1	-1	-1	-1	-1	-2	-2	-2	-3
G	-3	0	1	-2	0	**6**	-2	-1	-2	-2	-2	-2	-2	-3	-4	-4	0	-3	-3	-2
N	-3	1	0	-2	-2	0	**6**	1	0	0	-1	0	0	-2	-3	-3	-3	-3	-2	-4
D	-3	0	1	-1	-2	-1	1	**6**	2	0	-1	-2	-1	-3	-3	-4	-3	-3	-3	-4
E	-4	0	0	-1	-1	-2	0	2	**5**	2	0	0	1	-2	-3	-3	-3	-3	-2	-3
Q	-3	0	0	-1	-1	-2	0	0	2	**5**	0	1	1	0	-3	-2	-2	-3	-1	-2
H	-3	-1	0	-2	-2	-2	1	1	0	0	**8**	0	-1	-2	-3	-3	-2	-1	2	-2
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	**5**	2	-1	-3	-2	-3	-3	-2	-3
K	-3	0	0	-1	-1	-2	0	-1	1	1	-1	2	**5**	-1	-3	-2	-3	-3	-2	-3
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	**5**	1	2	-2	0	-1	-1
I	-1	-2	-2	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	**4**	2	1	0	-1	-3
L	-1	-2	-2	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	**4**	3	0	-1	-2
V	-1	-2	-2	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	**4**	-1	-1	-3
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	**6**	3	1
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	**7**	2
W	-2	-3	-3	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	**11**

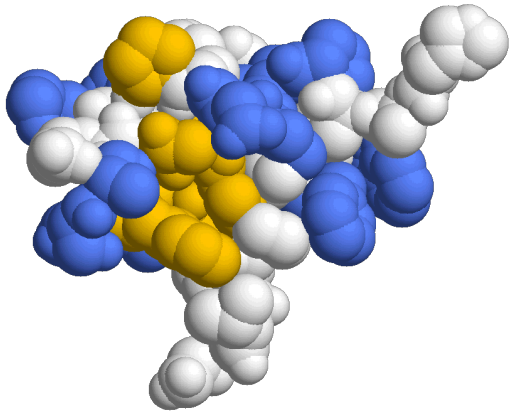
Aminoacid substitutions



Distance space



Secondary Structure (H,B,C)

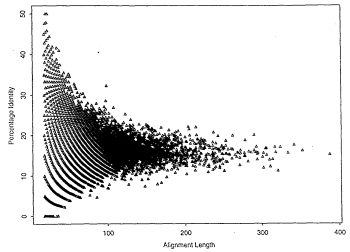


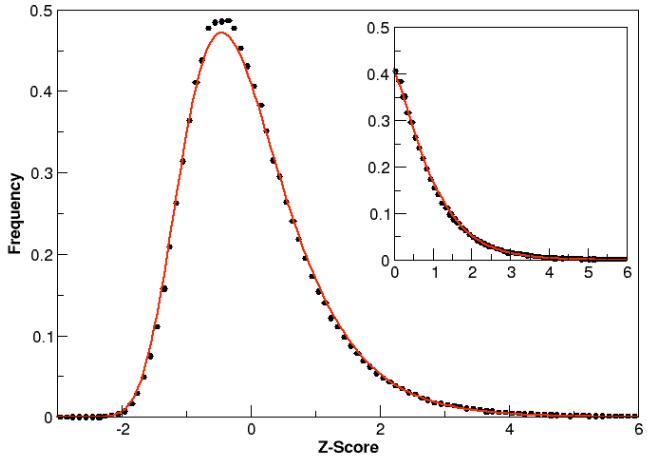
Accessible surface (B,A [%])

Significance of an alignment (score)

Probability that the optimal alignment of two random sequences/structures of the same length and composition as the aligned sequences/structures have at least as good a score as the evaluated alignment.

Empirical





Sometimes approximated by Z-score (normal distribution).

Analytic

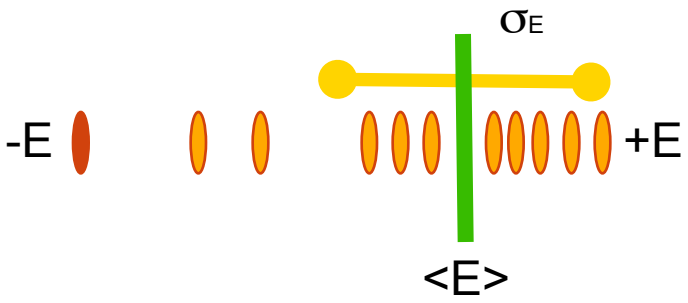
$$P(s) = e^{-\lambda (s-\mu)}$$

$$P(s \geq x) = 1 - \exp\left(e^{-\lambda (s-\mu)} \right)$$

Karlin and Altschul, 1990 *PNAS* **87**, pp2264

Significance of an alignment (score)

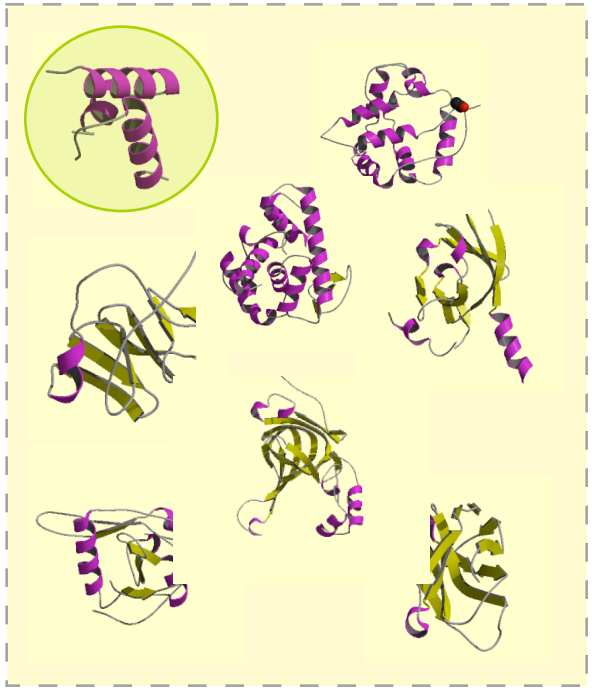
Energy Z-score the model with respect the energy of random models (or rest of decoys).



$$Zscore = \frac{\left(\langle E \rangle - E_m\right)}{\sigma_E}$$

Significance of an alignment (background)

Structural space



Sequence space

MKLLIVLTCISLCSCICTVVQRCASNKPHVLEDPCKVQH

HLSVNQCVLLPQCCPKSCKICTHLISIEVVLTCRAVDKM

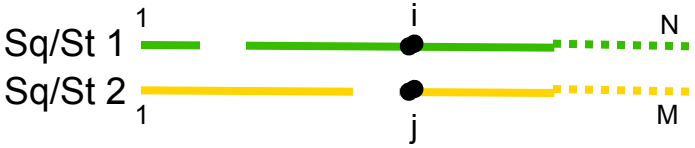
MHVNCVEQCSLQDCIKIAPRVLKTCILCVLKPCLTSVSH

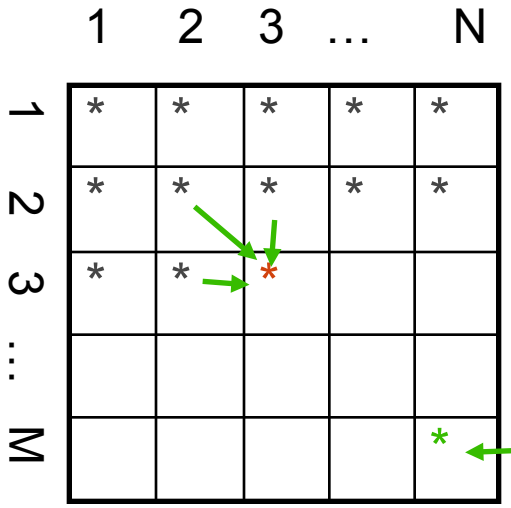
VHLVQPTSCCCKKNCICHVEIRSLDILTKSVQLACLVPM

⋮

MQCCRVQKICDLLAVELCKLHISTPSCKILCVVTSVPHN

Global dynamic programming alignment



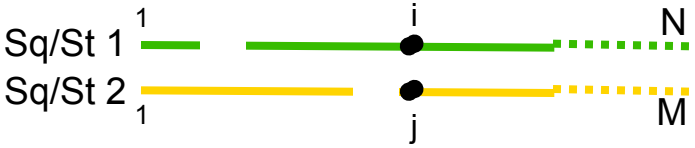


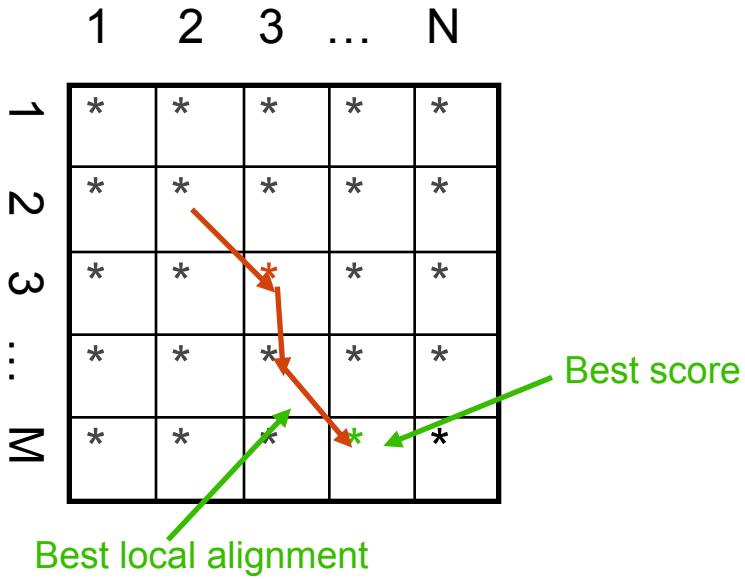
$$D_{i,j} = \min \begin{cases} D_{i,j-1} + \text{Score}_{(\Delta, rj)} \\ D_{i-1,j-1} + \text{Score}_{(ri, rj)} \\ D_{i-1,j} + \text{Score}_{(ri, \Delta)} \end{cases}$$

Backtracking to get the best alignment

Needleman and Wunsch (1970) *J. Mol Biol*, **3** pp443

Local dynamic programming alignment





$$D_{i,j} = \min \begin{cases} D_{i,j-1} + \text{Score}(\Delta, rj) \\ D_{i-1,j-1} + \text{Score}(ri, rj) \\ D_{i-1,j} + \text{Score}(ri, \Delta) \\ 0 \end{cases}$$

Backtracking to get the best alignment

Smith and Waterman (1981) *J. Mol Biol*, **147** pp195

Applications of PMFs

Model assessment.

Ab initio folding simulations.

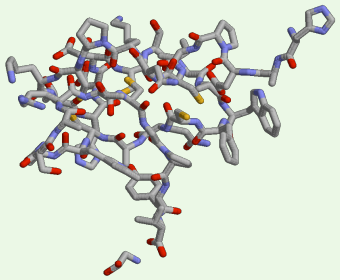
Sequence-structure matching (threading).

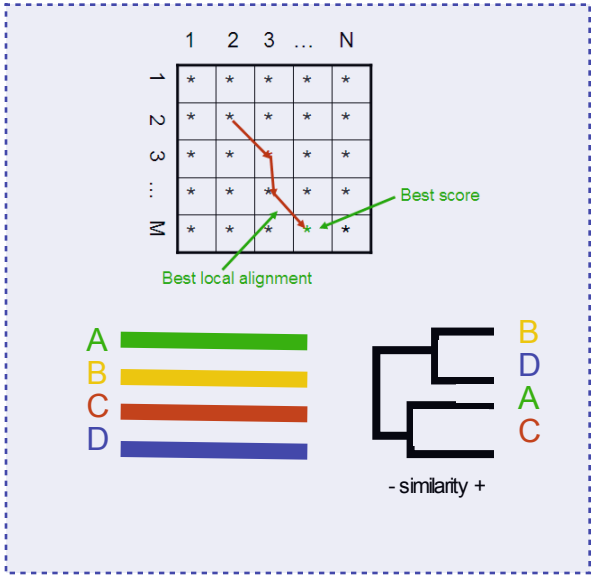
Comparative protein structure modeling (loops, sidechains, …).

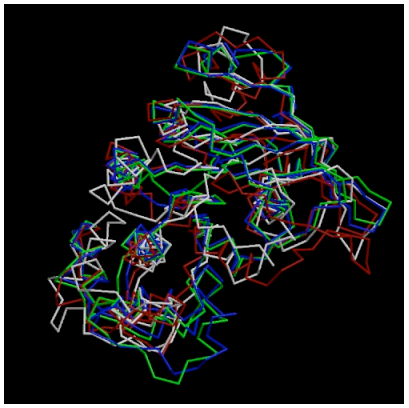
Secondary structure prediction, *etc.*

Finkelstein et al. (1995) *Proteins* **23**, pp142

Sequence-Structural alignment by properties conservation (SALIGN-MODELLER)

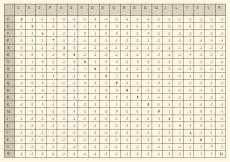


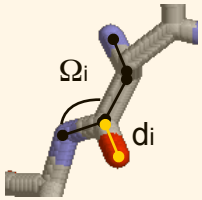


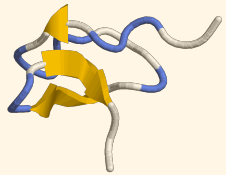


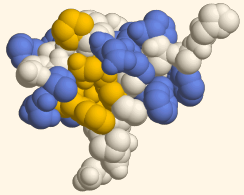
✓ Uses all available structural information
✓ Provides the optimal alignment

Computationally expensive









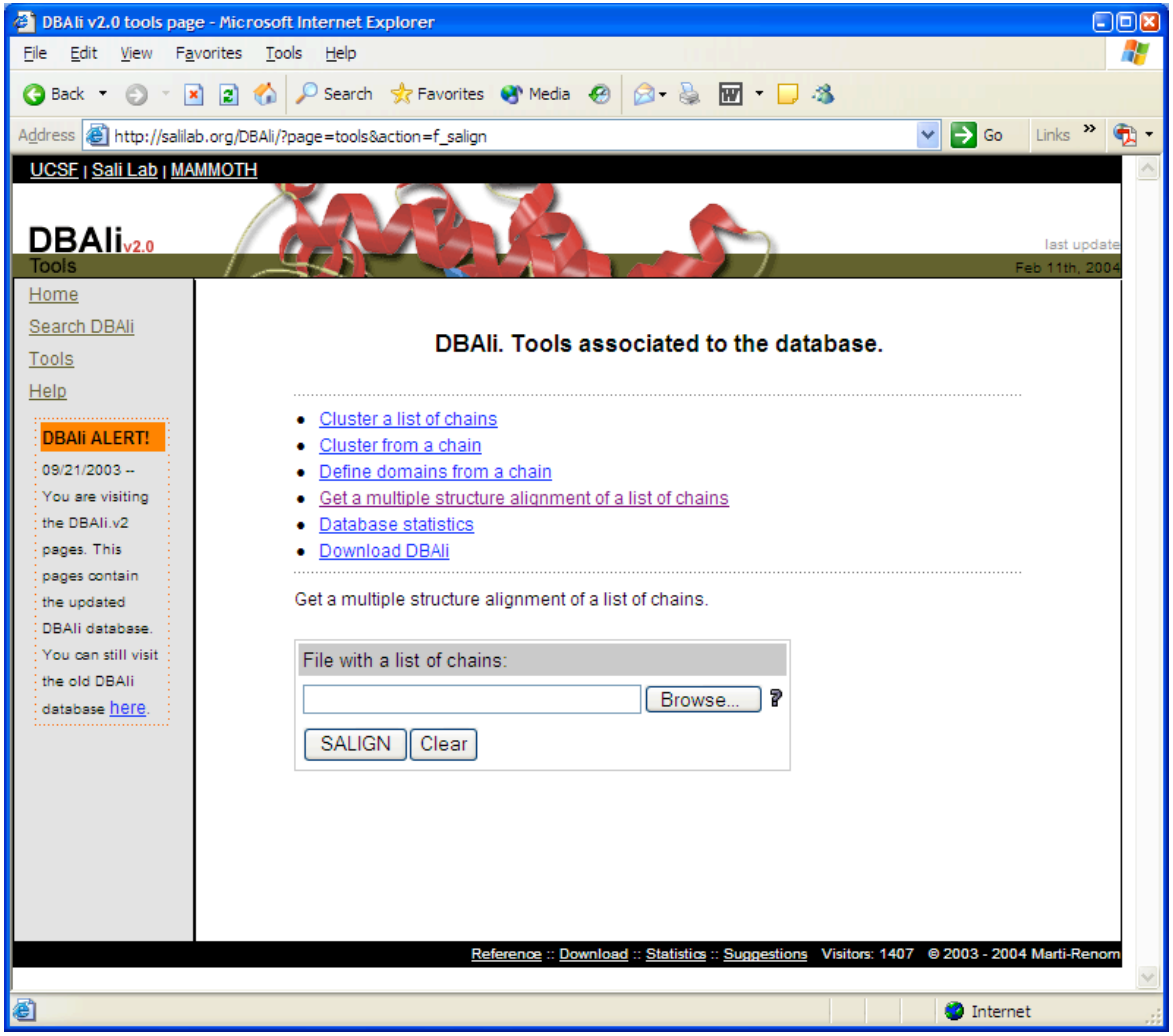
$$RMSD(x,y) = \sqrt{\left(\frac{1}{N}\right)\sum_{i=1}^{N}\left(\left\|\mathbf{x}(i)-\mathbf{y}(i)\right\|^{2}\right)}$$

$R_{i,j}$ $D_{,i(3),j(3)}$ $S_{i,j}$ $B_{i,j}$ $I_{i,j}$

Madhusudhan et al. *in preparation*

Structural alignment by properties conservation (SALIGN-MODELLER)

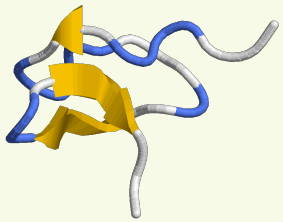
http://www.salilab.org/dbali/

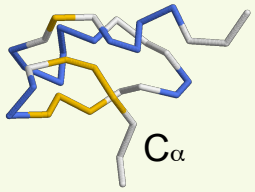


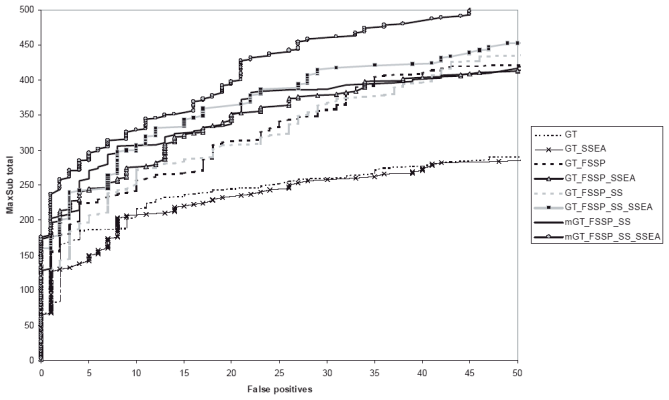
Threading (mGenThreader)

>gi42541361
MDIRSVSSLRGLLCLPPSWPRR

- Neural Network

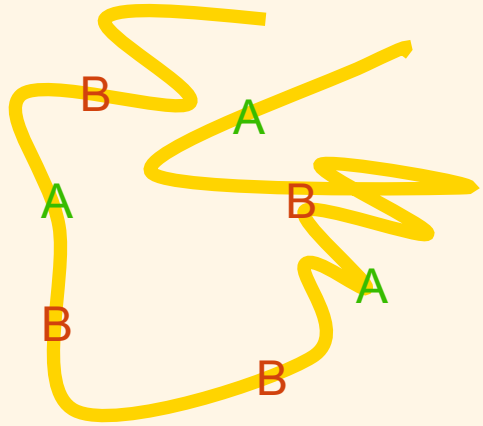


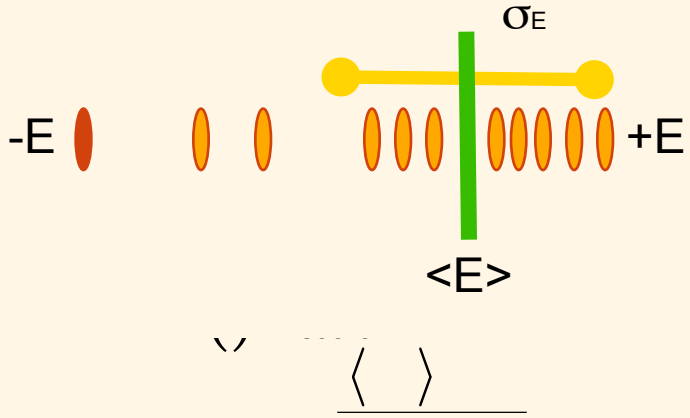




✓ **Good row and significance scoring**

Obscure optimizer

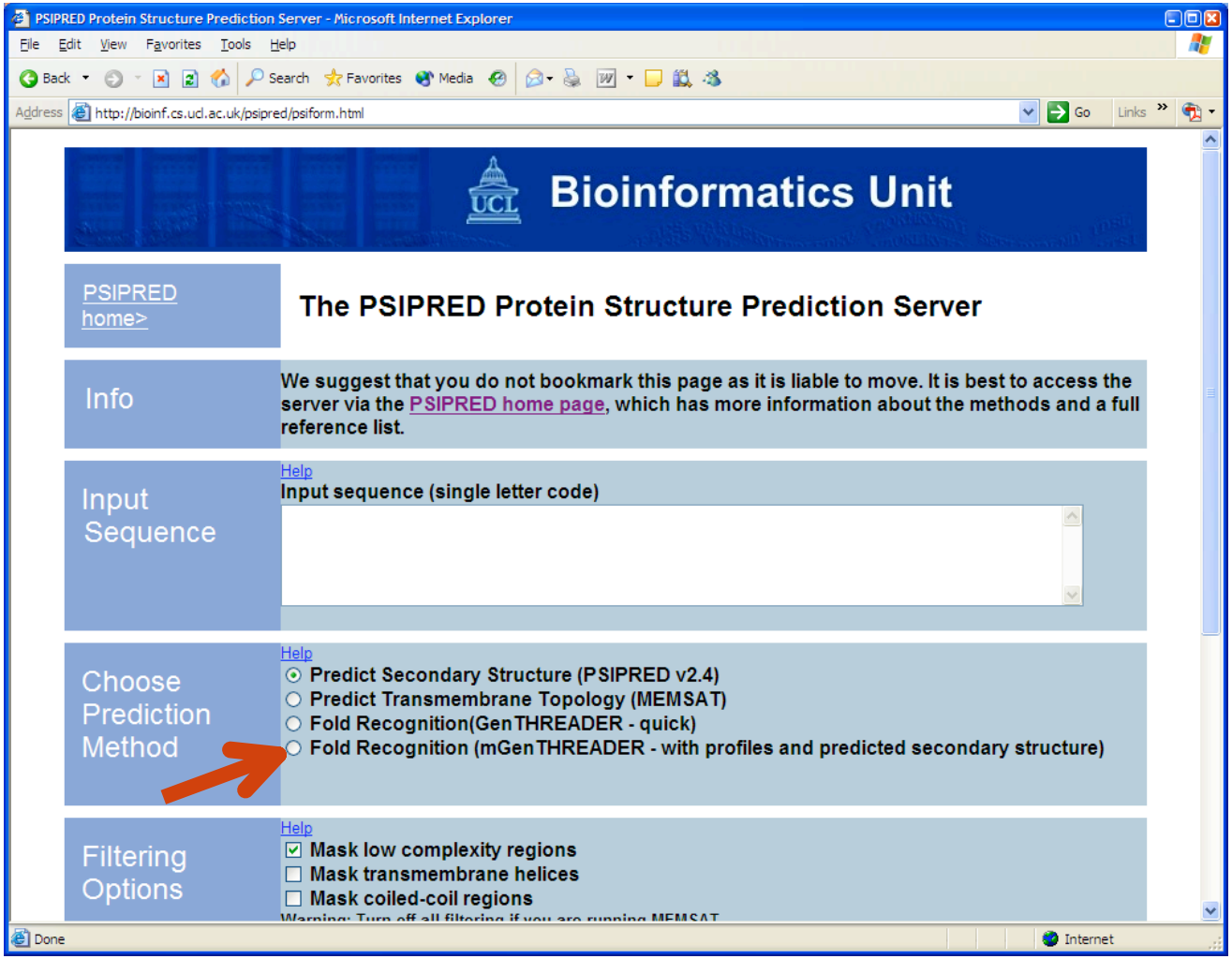




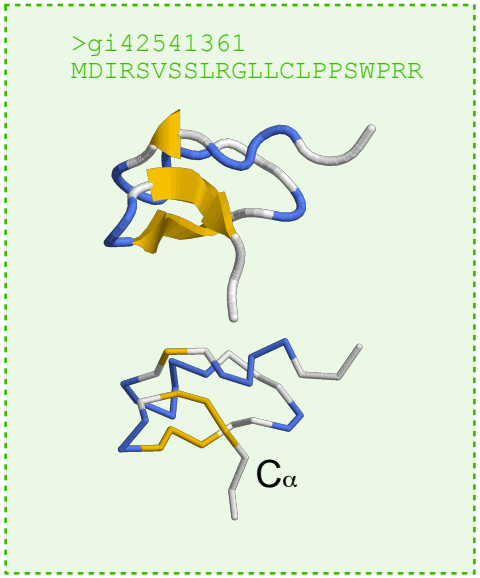
McGuffin LJ, Jones DT. (2003) *Bioinformatics*, **19**, pp874

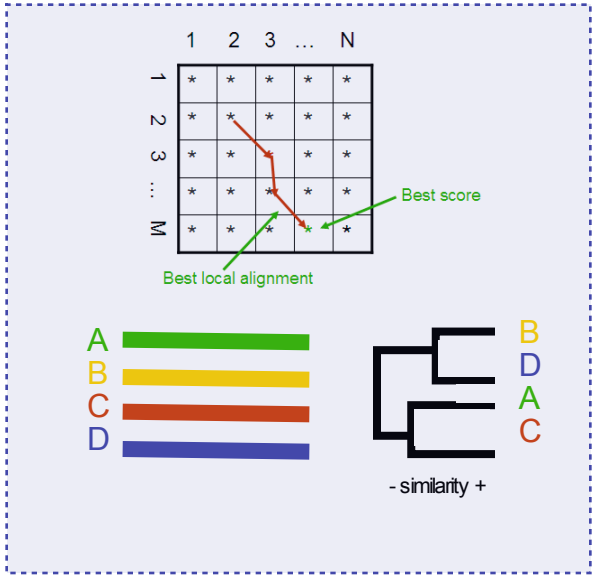
Threading (mGenThreader)

http://bioinf.cs.ucl.ac.uk/psiform.html



Remote homology detection (FUGUE)

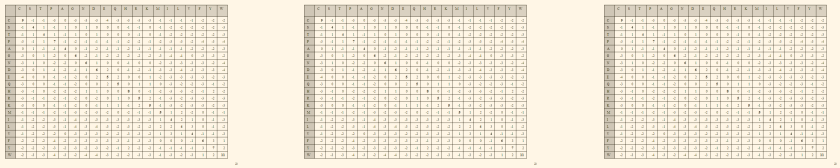




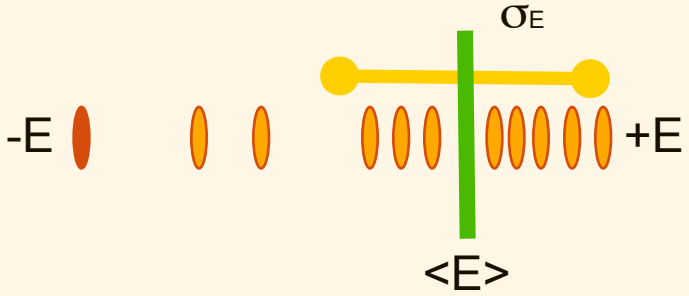
FUGUE

- ✓ Uses most of the structural information
- ✓ Easy to access either locally and on the web
- ✓ Good row and significance scoring

Does not uses multiple sequence information



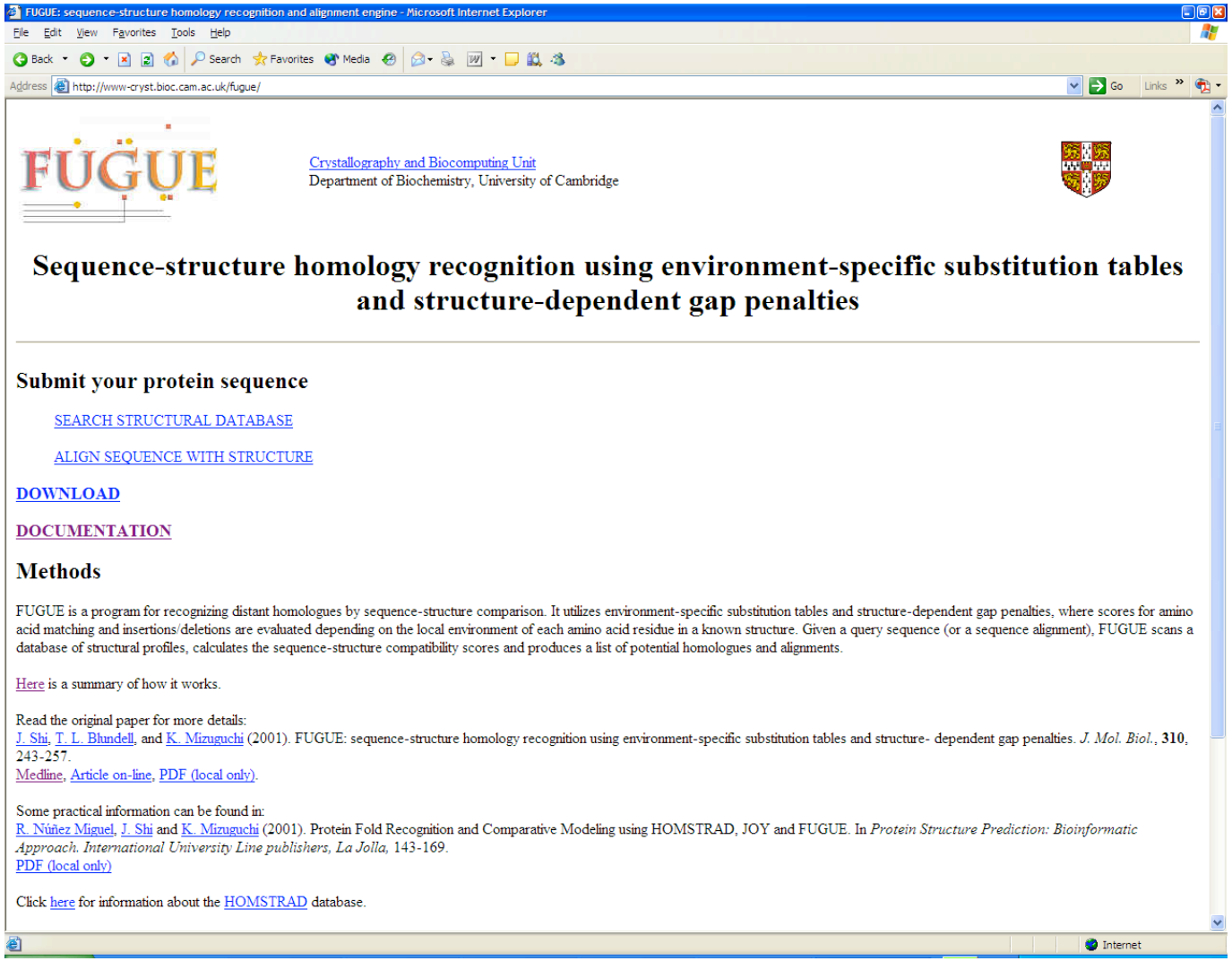
R_{ij}^{H} R_{ij}^{B} R_{ij}^{C}



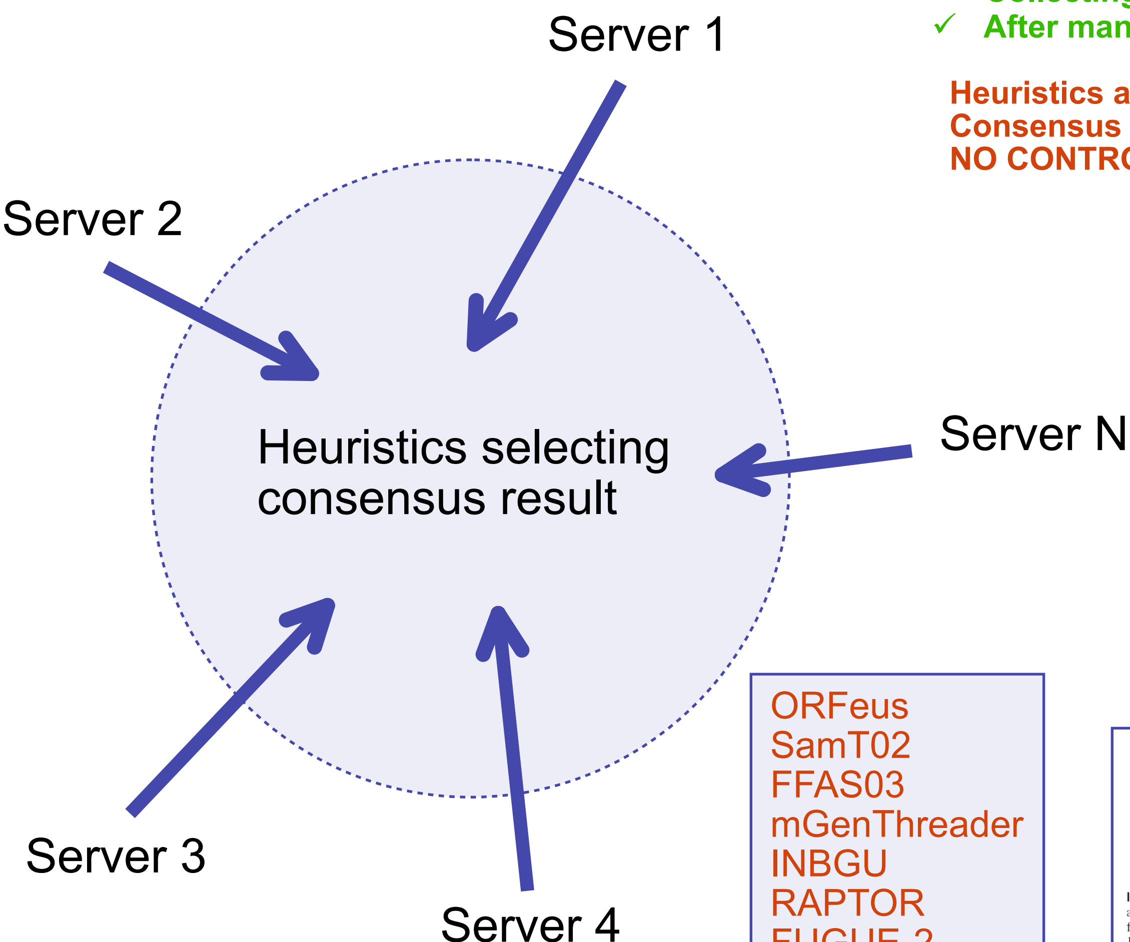
$$Zscore = \frac{\left(\langle E \rangle - E_m\right)}{\sigma_E}$$

Remote homology detection (FUGUE)

http://www-cryst.bioc.cam.ac.uk/fugue/



Meta-Servers (3D-Jury)



- ✓ Collecting several results
- ✓ After manual analysis… good results

Heuristics and complicated scoring
Consensus results
NO CONTROL OF DATA GENERATION or SERVERS!

ORFeus
SamT02
FFAS03
mGenThreader
INBGU
RAPTOR
FUGUE-2
3D-PSSM

Cell, Vol. 113, 701–702, June 13, 2003, Copyright ©2003 by Cell Press

Letter to the Editor

mRNA Cap-1 Methyltransferase in the SARS Genome

The 3D jury system has predicted the methyltransferase fold for the nsp13 protein of the SARS coronavirus. Based on the conservation of a characteristic tetrad of residues, the mRNA cap-1 methyltransferase function has been assigned to this protein, which has potential implications for antiviral therapy.

The latest outbreak of the severe acute respiratory syndrome (SARS) epidemic has led to thousands of potentially lethally infected patients and hundreds of deaths. These numbers are likely to rise, and the spreading disease is already causing major medical and economical concerns. Meanwhile, the SARS coronavirus identified as the pathogen responsible for the disaster has been isolated, and its genome sequenced (Marra et al., 2003; Rota et al., 2003).

We have applied the 3D jury meta predictor (Ginalski et al., 2003) to annotate the structure and function of proteins encoded by the viral positive-strand ssRNA. Novel fold recognition methods utilize the global network of independent structure prediction servers. Detection of patterns of structural similarity between di-…

suggest that the virus also requires the AdoMet-dependent cap-0 methyltransferase. Both functions can be inhibited by carbocyclic analogs of adenosine, such as Neplanocin A or 3-deazaneplanocin A, which interfere with the AdoMet-AdoHcy metabolism of the host cell (De Clercq, 1998; Bray et al., 2002). Those compounds could complement other therapeutic strategies aimed at blocking enzymatic functions such as the RNA-dependent RNA polymerase, the protease, or the helicase encoded by the SARS virus.

Marcin von Grotthuss, Lucjan S. Wyrwicz, and Leszek Rychlewski*
BioinfoBank Institute
Limanowskiego 24A
60-744 Poznan
Poland

*Correspondence: leszek@bioinfo.pl

Figure 1. 3D Model of the nsp13 Domain of the SARS Coronavirus pp1ab Polyprotein

LETTERS

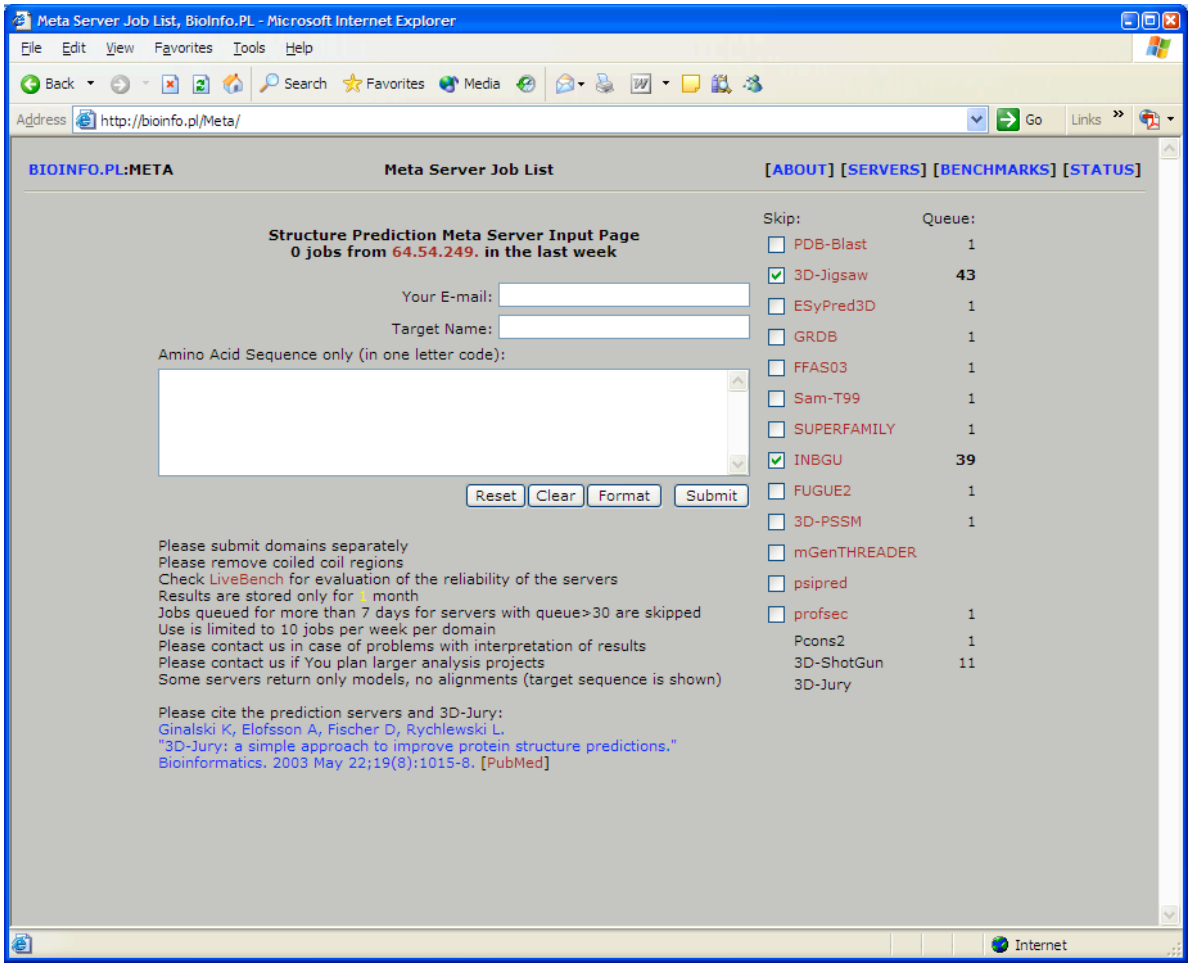
How Unique Is the Rice Transcriptome?

IN THE REPORT "COLLECTION, MAPPING, AND annotation of over 28,000 cDNA clones from *japonica* rice" (S. Kikuchi *et al.*, 18 July, p. 376), the Rice Full-Length cDNA Project Team provides a detailed description of the rice transcriptome. The authors claim that 36% of the tested rice transcripts are not

Ginalski K, et al. (2003) *Bioinformatics* **19** pp1015

Meta-Servers (3D-Jury)

http://bioinfo.pl/Meta/



Complex gap penalty functions

MODELLER SALIGN (ALIGN2D)

Madusudhan M.S. *et al.* in preparation

Profile-Profile alignments

MODELLER SALIGN ('PROFILE')

Marti-Renom *et al.* (2004) Protein sciences. **13**:1071

Experiment (in silico)

- Benchmarking the best alignment methods.
- New alignment method.
- Projected gains.

Methods: Reference set

CE alignments with
- < 40% sequence identity
- > 100 EqPos
- > 50% EqPos
- > 90% coverage for one chain

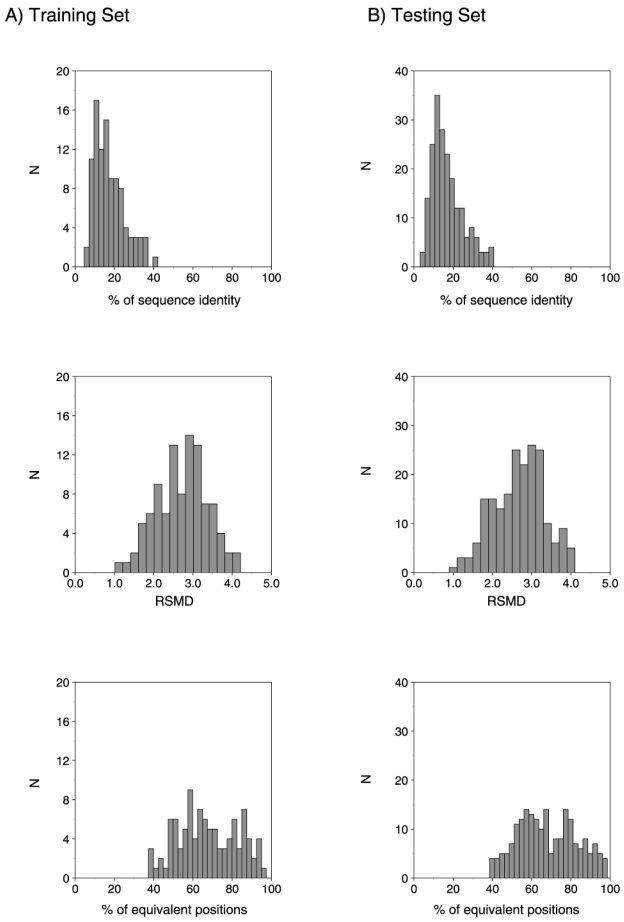
387

Filter: MAMMOTH alignments with
- > 50% EqPos

300

100 Training set

200 Testing set



Methods: Evaluated methods

Sequence A: AGHLAHTRCELKLPTCRGNMSSRFC

Sequence B: AGHLRHTRRCLRLPTAGNARFC

Seq.-Seq.

ALIGN: DP pairwise method

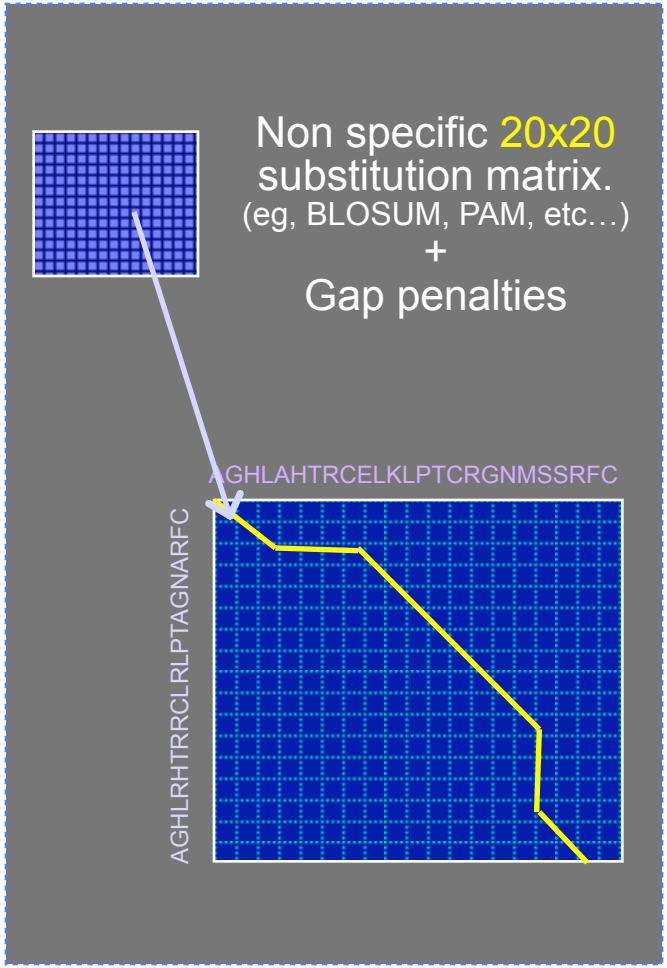
BLAST2SEQ: Local method

Prof.-Seq.

PSI-BLAST: Local search method that uses multiple sequence information for one of the sequences.

Prof.-Prof.

ALIGN4D: DP pairwise method that uses multiple sequence information for both sequences.



Methods: Evaluated methods

Sequence A: AGHLAHTRCELKLPTCRGNMSSRFC

Sequence B: AGHLRHTRRCLRLPTAGNARFC

Seq.-Seq.

ALIGN: DP pairwise method

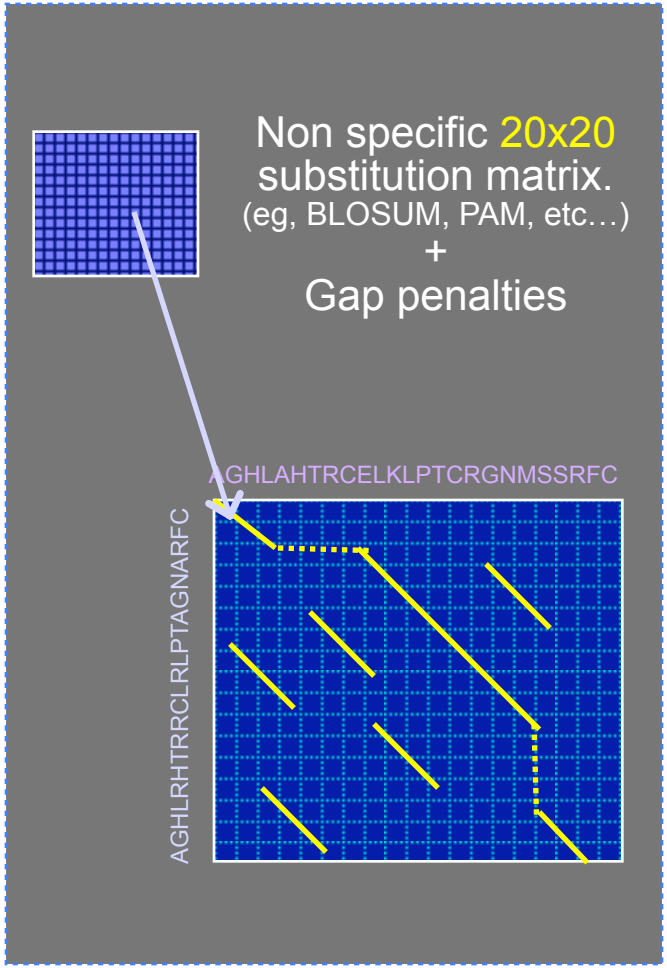
BLAST2SEQ: Local method

Prof.-Seq.

PSI-BLAST: Local search method that uses multiple sequence information for one of the sequences.

Prof.-Prof.

ALIGN4D: DP pairwise method that uses multiple sequence information for both sequences.



Methods: Evaluated methods

Sequence A: AGHLAHTRCELKLPTCRGNMSSRFC

Sequence B: AGHLRHTRRCLRLPTAGNARFC

Seq.-Seq.

ALIGN: DP pairwise method

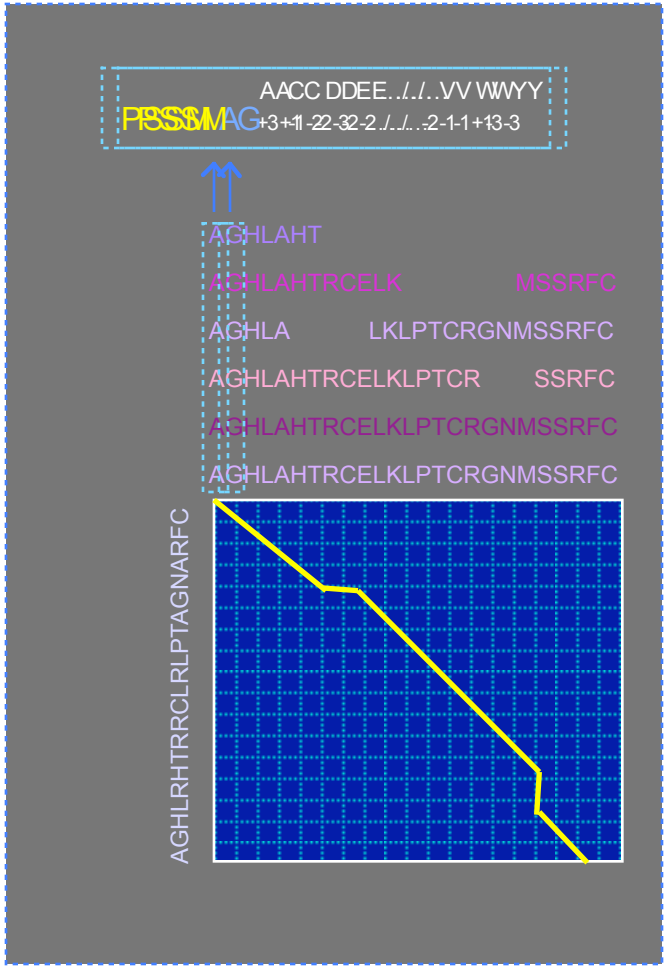
BLAST2SEQ: Local method

Prof.-Seq.

PSI-BLAST: Local search method that uses multiple sequence information for one of the sequences.

Prof.-Prof.

ALIGN4D: DP pairwise method that uses multiple sequence information for both sequences.



Methods. Evaluated methods.

Sequence A: AGHLAHTRCELKLPTCRGNMSSRFC

Sequence B: AGHLRHTRRCLRLPTAGNARFC

Seq.-Seq.

ALIGN: DP pairwise method

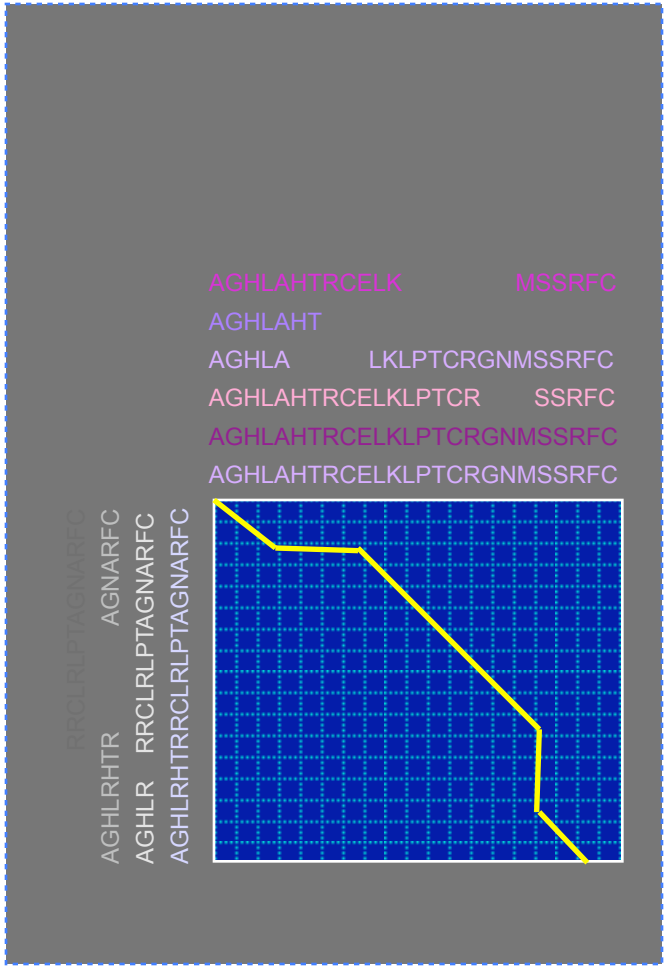
BLAST2SEQ: Local method

Prof.-Seq.

PSI-BLAST: Local search method that uses multiple sequence information for one of the sequences.

Prof.-Prof.

ALIGN4D: DP pairwise method that uses multiple sequence information for both sequences.



Methods. SALIGN.

ALIGN4D protocol	Profile	Comparison	Open	Extension
CC_{PBP}	PSI-BLAST	Correlation Coefficient	-300	0
CC_{HH}	Henikoff-Henikoff	Correlation Coefficient	-300	0
CC_{HS}	H-H + similarity weight	Correlation Coefficient	-150	0
ED_{PBP}	PSI-BLAST	Euclidian Distance	-450	-30
ED_{HH}	Henikoff-Henikoff	Euclidian Distance	-550	0
ED_{HS}	H-H + similarity weight	Euclidian Distance	-450	-10
DP_{PBP}	PSI-BLAST	Dot Product	-250	-30
DP_{HH}	Henikoff-Henikoff	Dot Product	-550	0
DP_{HS}	H-H + similarity weight	Dot Product	-100	-30
JS_{HH}	Henikoff-Henikoff	Jansen-Shannon Distance	-150	0
JS_{HS}	H-H + similarity weight	Jansen-Shannon Distance	-250	0

Methods: Coverage and accuracy

High coverage

Low accuracy

High accuracy

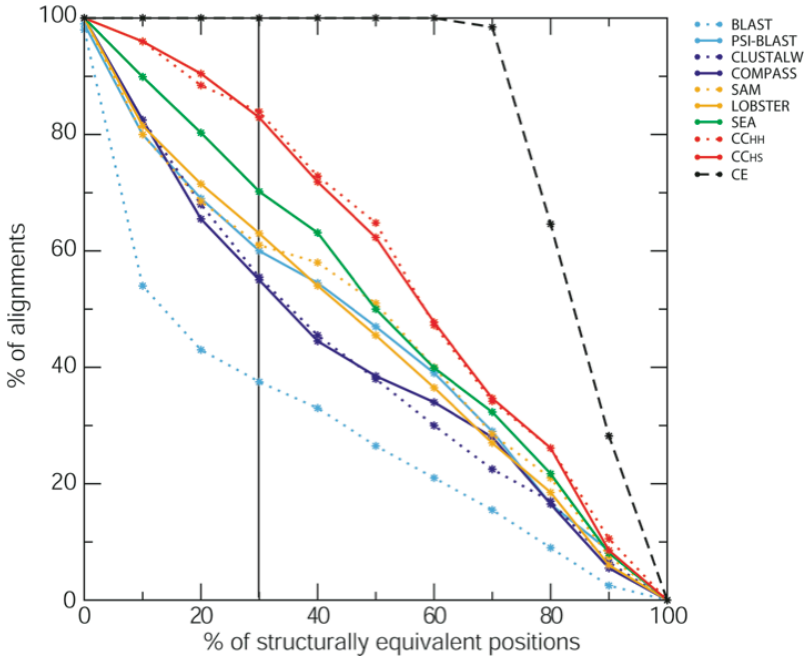
Low coverage

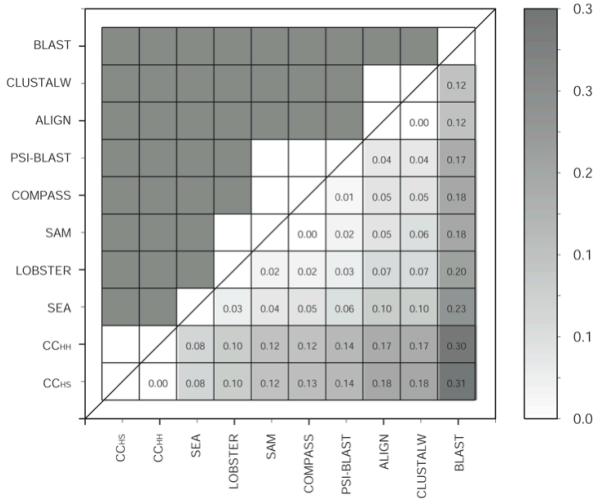
High accuracy

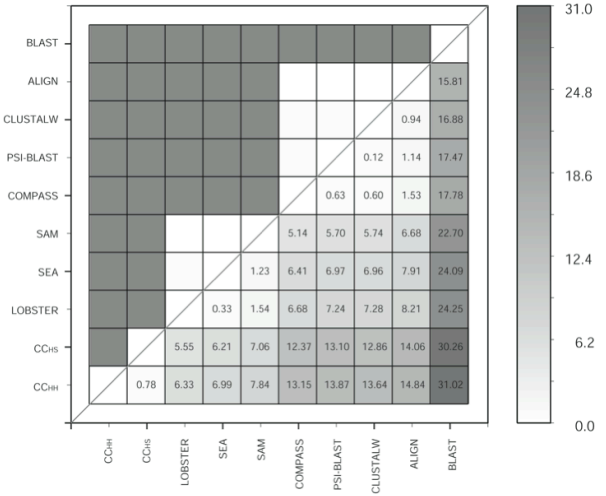
Low accuracy

Results: Comparison of alignment dependent measures

Method	CE overlap [%]	Shift score	RMSD [A]	Structure overlap [%]
CE	100 ± 0	1.00 ± 0.00	2.7 ± 0.6	59.8 ± 12.9
BLAST	26 ± 29	0.32 ± 0.33	5.6 ± 3.7	20.6 ± 23.7
PSI-BLAST	43 ± 31	0.48 ± 0.35	6.5 ± 3.9	30.3 ± 24.9
SAM	48 ± 26	0.50 ± 0.34	9.2 ± 4.7	28.9 ± 24.8
LOBSTER	50 ± 27	0.51 ± 0.32	9.1 ± 4.9	31.1 ± 25.2
SEA	49 ± 27	0.53 ± 0.29	8.4 ± 4.4	33.4 ± 24.3
ALIGN	42 ± 25	0.44 ± 0.28	10.6 ± 5.0	25.7 ± 24.1
CLUSTALW	43 ± 27	0.44 ± 0.31	10.2 ± 4.9	26.4 ± 24.3
COMPASS	43 ± 32	0.49 ± 0.35	4.8 ± 3.2	32.3 ± 24.7
CC_{HH}	56 ± 23	0.61 ± 0.24	7.8 ± 4.2	36.7 ± 22.9
CC_{HS}	56 ± 24	0.62 ± 0.24	7.8 ± 4.2	36.5 ± 23.2



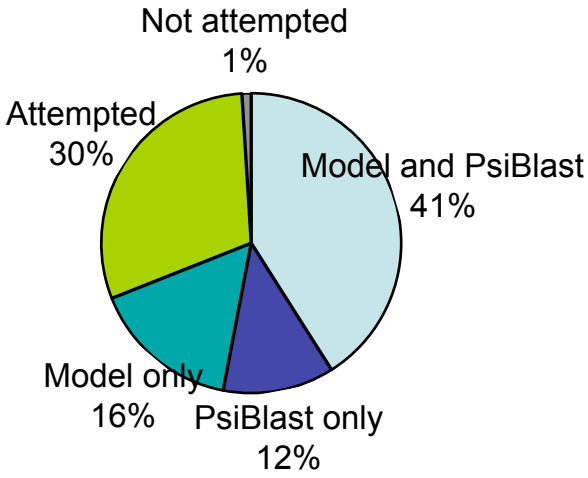




Results. Turn over.

Mycoplasma genitalium MODPIPE Models

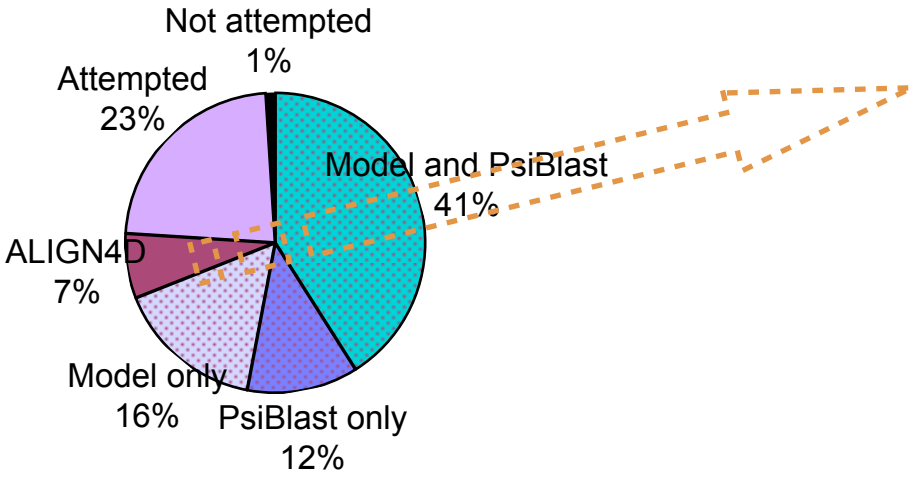
Number of ORFs	479
Average ORF length	364



Results. Turn over.

Mycoplasma genitalium MODPIPE Models

Number of ORFs	479
Average ORF length	364



~ 34 extra accurate models for M. g. genome.

~ 100,000 models for TrEMBL-SP "genome".

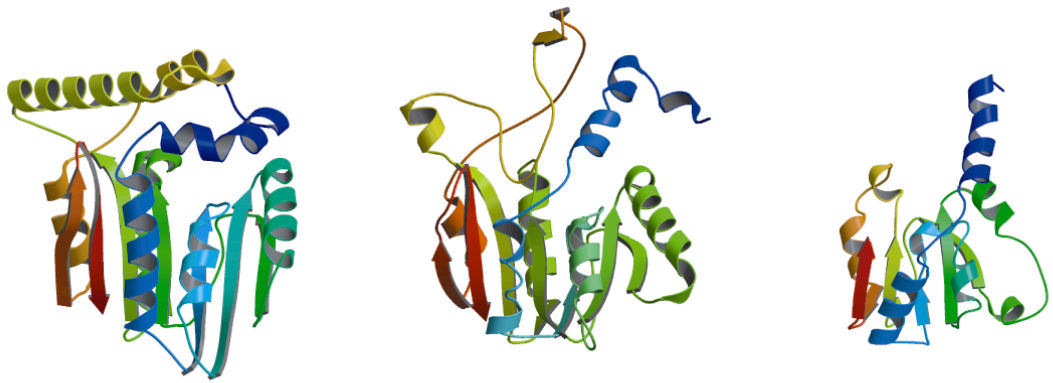
Examples: T0092 model

- Target T0092 at CASP4:
- Hypothetical protein HI0319
- Haemophilus influenzae
- Parent: 1d2cA (Methyltransferase)
- ALIGN4D alignment at 8.4% seq id.

Method	RMSD Å	% of EqPos
ALIGN4D CC_{PBP}	5.9	67.84
PSI-BLAST	4.9	31.72
Best predictions at CASP4	6.0	65.20

Data from CASP4, Asilomar, CA, December 2000.

B) Target T0092



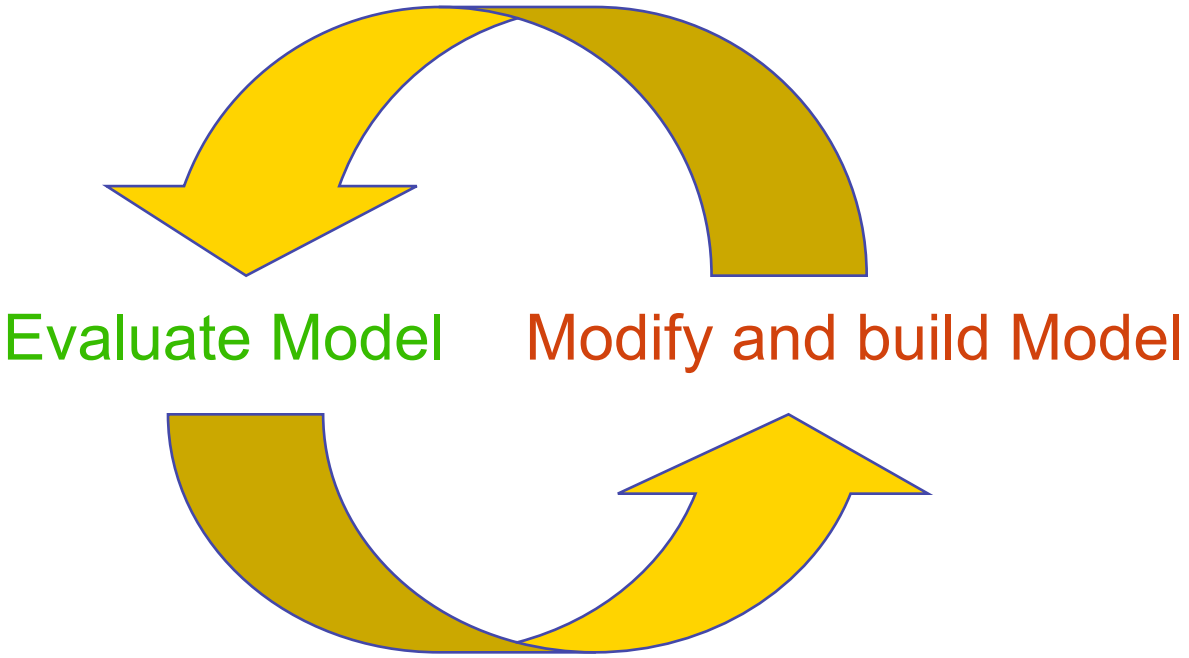
X-Ray structure | ALIGN4D (CC_{PBP}) model | Psi-Blast model

Iterative process

MOULDER

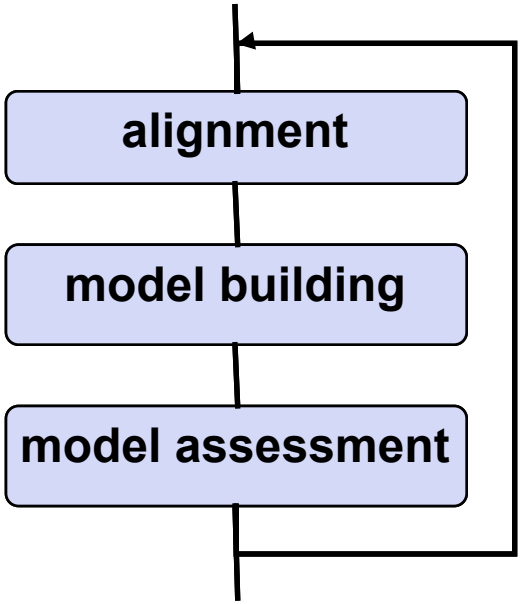
John, B. and Sali, A. (2003) Nucleic Acids Research. **31**:1982-1992

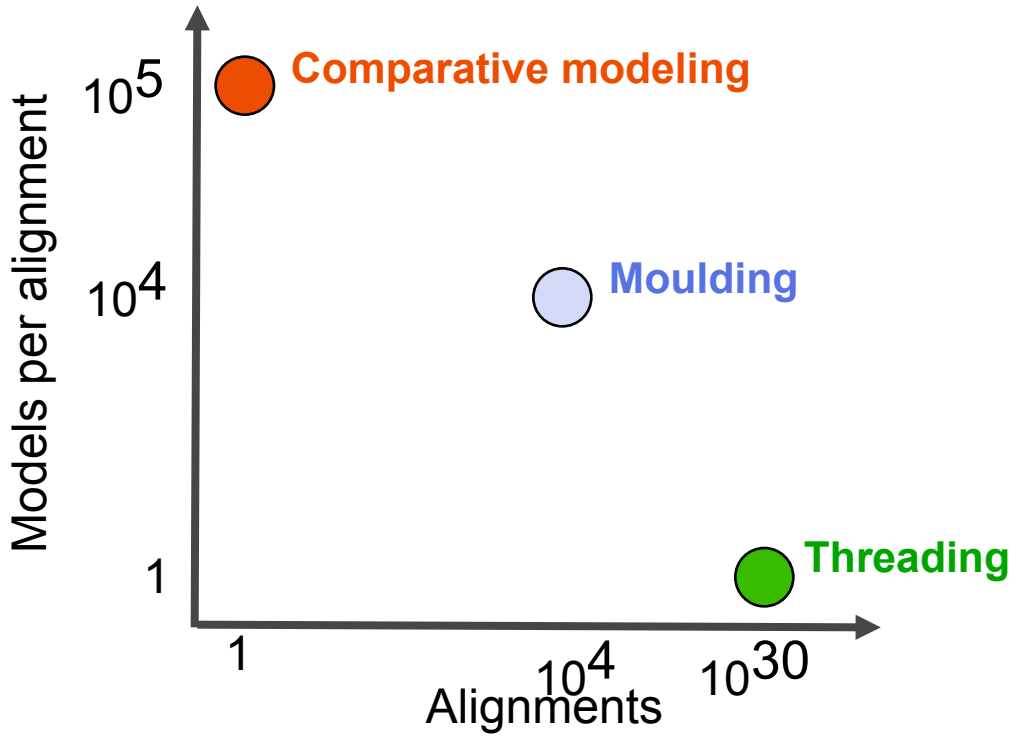
Iterative process… better models(?)



Moulding: iterative alignment, model building, model assessment

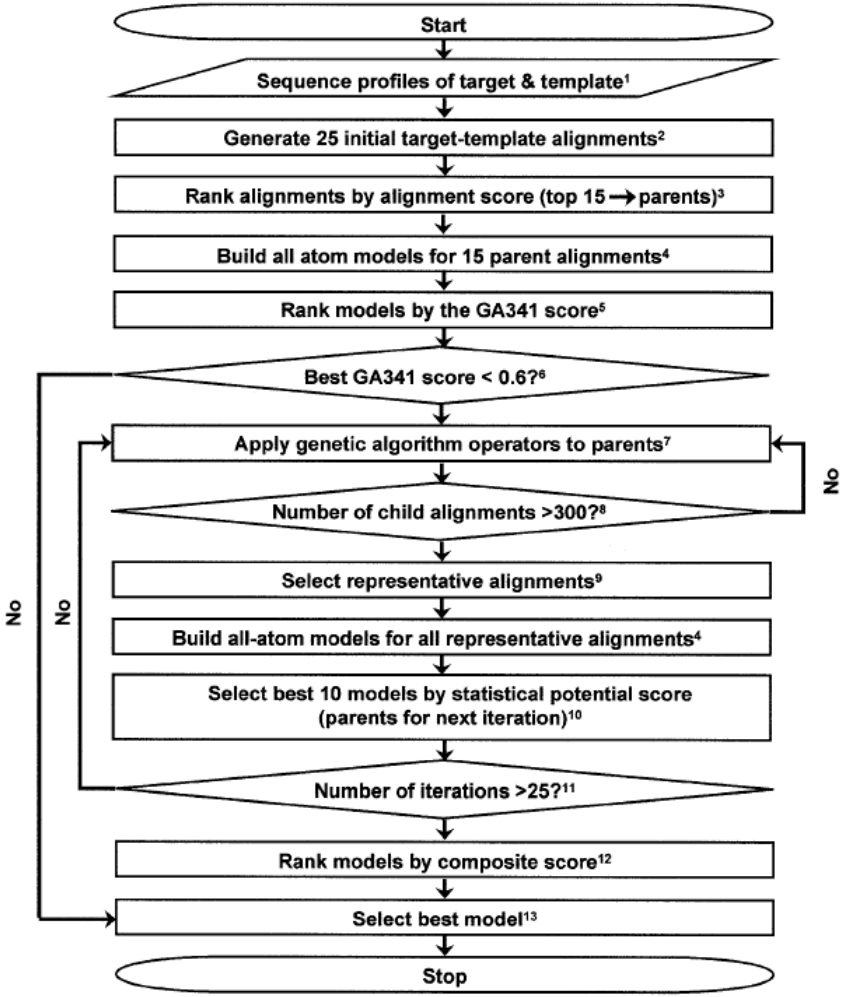
B. John, A. Sali. *Nucl. Acids Res.*, **31**, 1982-1992, 2003.

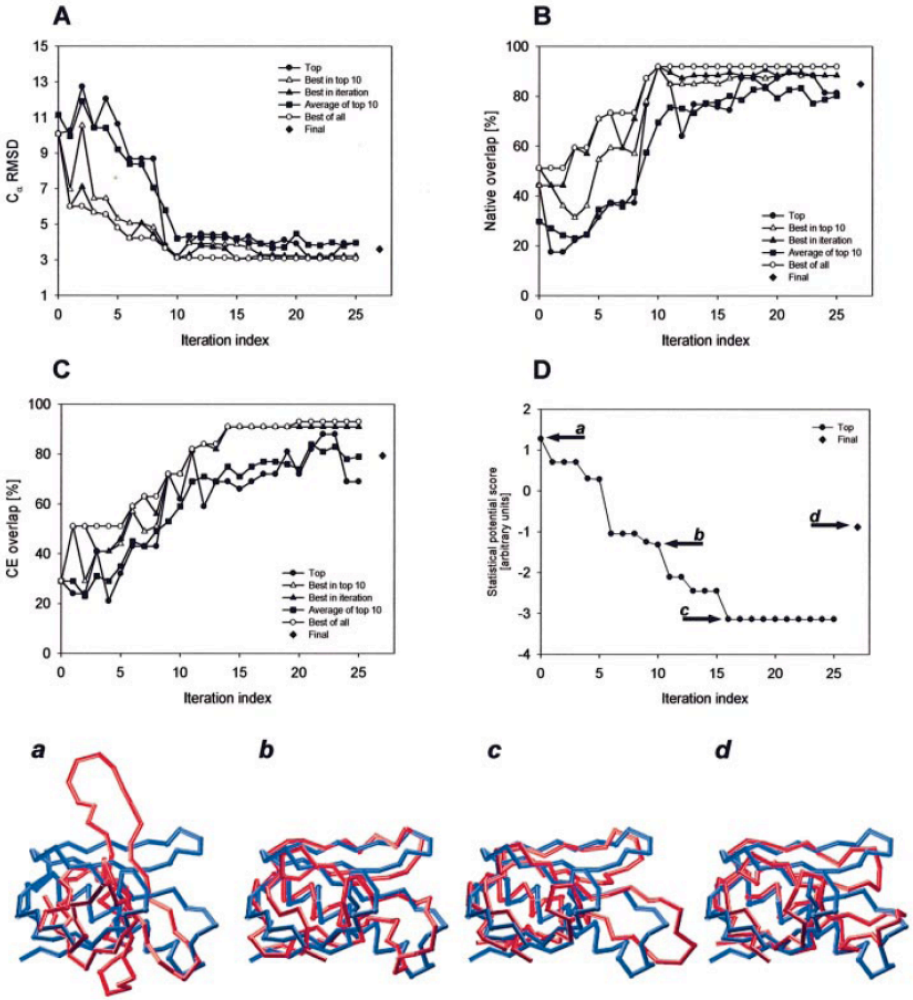




Iterative process… MOULDER

more in model evaluation





John & Sali (2003). *NAR* **31** pp3982