Comparative Protein Structure Prediction

Bioinformatics Master, Universidad de Valencia

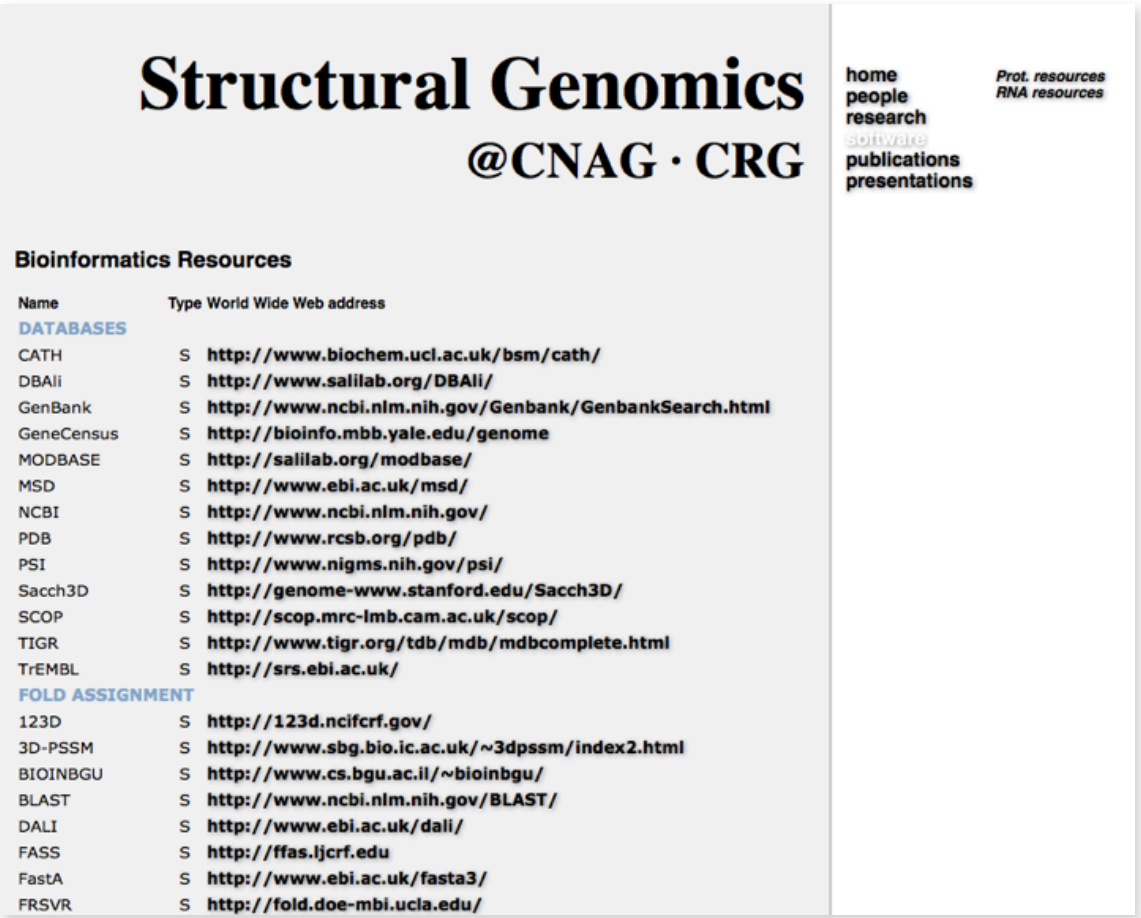
20 Abril, 2015

Francisco Martínez-Jiménez - fmartinez@pcb.ub.es

Objective

TO LEARN HOW-TO MODEL A 3D-STRUCTURE FROM A SEQUENCE AND A KNOWN STRUCTURE

DISCLAIMER!



Structural Genomics
@CNAG · CRG

home
people
research
software
publications
presentations

Prot. resources
RNA resources

Bioinformatics Resources

Name	Type	World Wide Web address
DATABASES		
CATH	S	http://www.biochem.ucl.ac.uk/bsm/cath/
DBAli	S	http://www.salilab.org/DBAli/
GenBank	S	http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html
GeneCensus	S	http://bioinfo.mbb.yale.edu/genome
MODBASE	S	http://salilab.org/modbase/
MSD	S	http://www.ebi.ac.uk/msd/
NCBI	S	http://www.ncbi.nlm.nih.gov/
PDB	S	http://www.rcsb.org/pdb/
PSI	S	http://www.nigms.nih.gov/psi/
Sacch3D	S	http://genome-www.stanford.edu/Sacch3D/
SCOP	S	http://scop.mrc-lmb.cam.ac.uk/scop/
TIGR	S	http://www.tigr.org/tdb/mdb/mdbcomplete.html
TrEMBL	S	http://srs.ebi.ac.uk/
FOLD ASSIGNMENT		
123D	S	http://123d.ncifcrf.gov/
3D-PSSM	S	http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html
BIOINBGU	S	http://www.cs.bgu.ac.il/~bioinbgu/
BLAST	S	http://www.ncbi.nlm.nih.gov/BLAST/
DALI	S	http://www.ebi.ac.uk/dali/
FASS	S	http://ffas.ljcrf.edu
FastA	S	http://www.ebi.ac.uk/fasta3/
FRSVR	S	http://fold.doe-mbi.ucla.edu/

http://sgt.cnag.cat/www/software/?prot_resources

Nomenclature

Homology: Sharing a common ancestor, may have similar or dissimilar functions

Similarity: Score that quantifies the degree of relationship between two sequences.

Identity: Fraction of identical aminoacids between two aligned sequences (case of similarity).

Target: Sequence corresponding to the protein to be modeled.

Template: 3D structure/s to be used during protein structure prediction.

Model: Predicted 3D structure of the target sequence.

Restraint: a measure or condition that keeps something under control.

Nomenclature

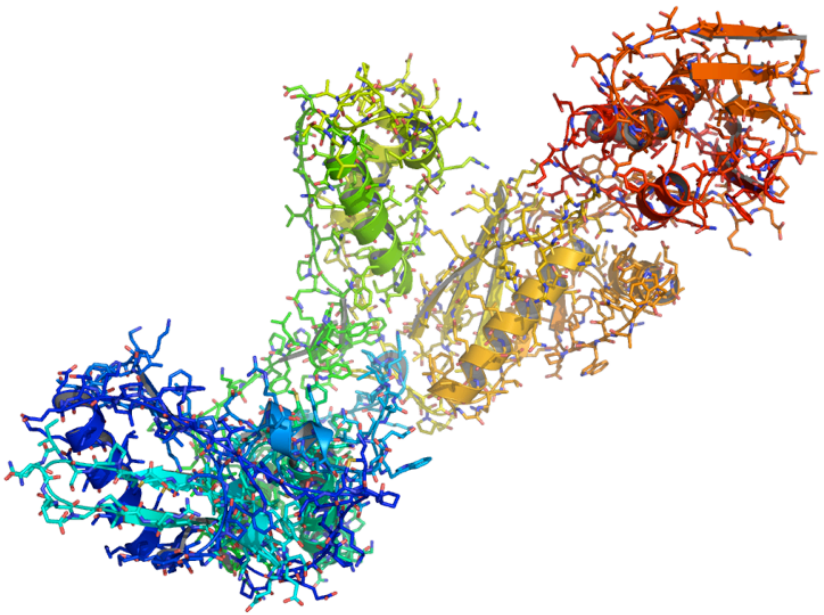
Fold: Three dimensional conformation of a protein sequence (usually at domain level).

Domain: Structurally globular part of a protein, which may independently fold.

Secondary Structure: Regular sub-domain structures composed by alpha-helices, beta-sheets and coils (or loops).

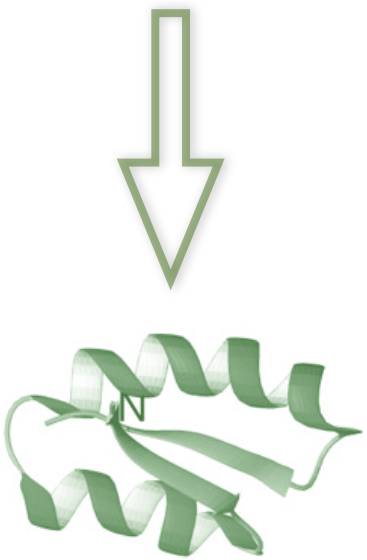
Backbone: Protein structure skeleton composed by the carbon, nitrogen and oxygen atoms.

Side-Chain: Specific atoms identifying each of the 20 residues types.



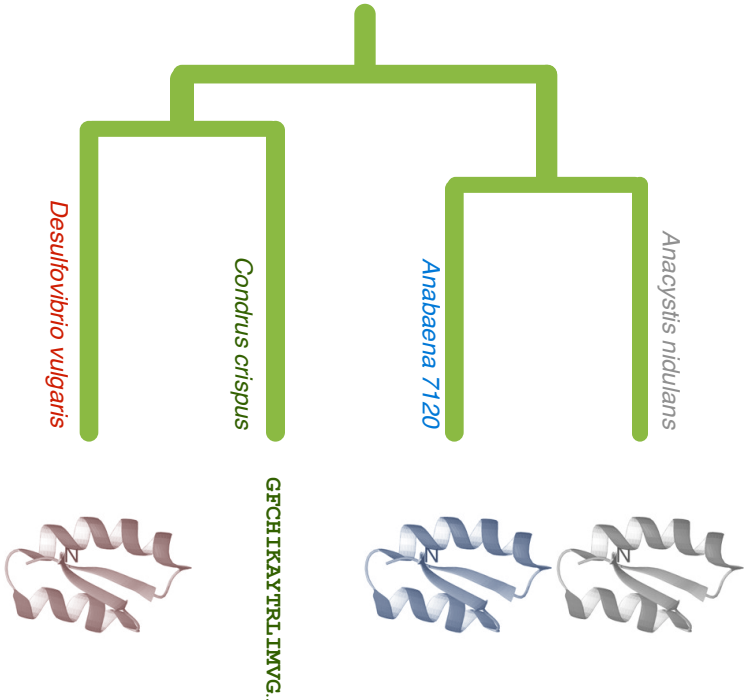
Principles of protein structure

GFCHIKAYTRLIMVG...



Folding (physics)

Ab initio prediction



Evolution (rules)

Homology modeling

Threading

Moulding

Comparative Modeling

D. Baker & A. Sali. Science 294, 93, 2001.

Comparative Modeling

Intro to comparative protein structure prediction

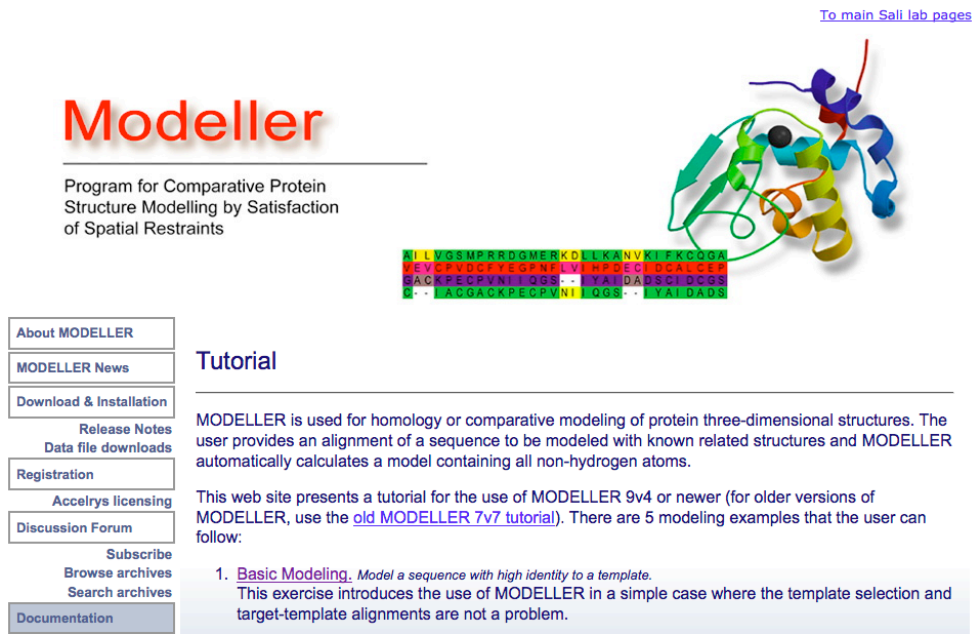
Template Search

Target – Template Alignment

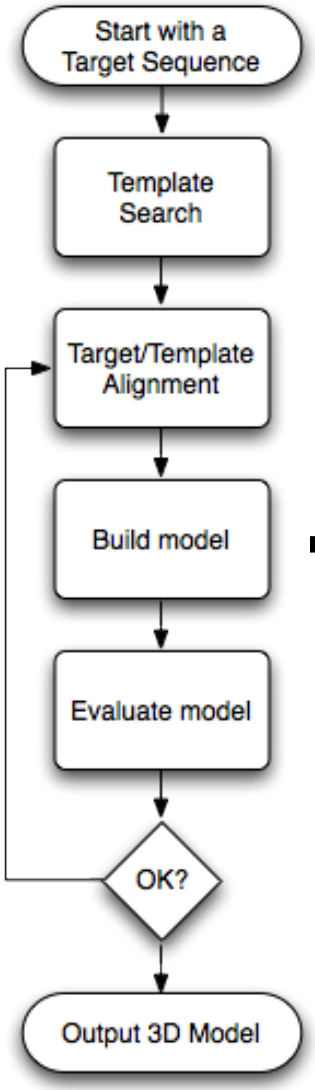
Model Building

Model Evaluation

http://www.salilab.org/modeller/tutorial/



Comparative modeling by satisfaction of spatial restraints
MODELLER

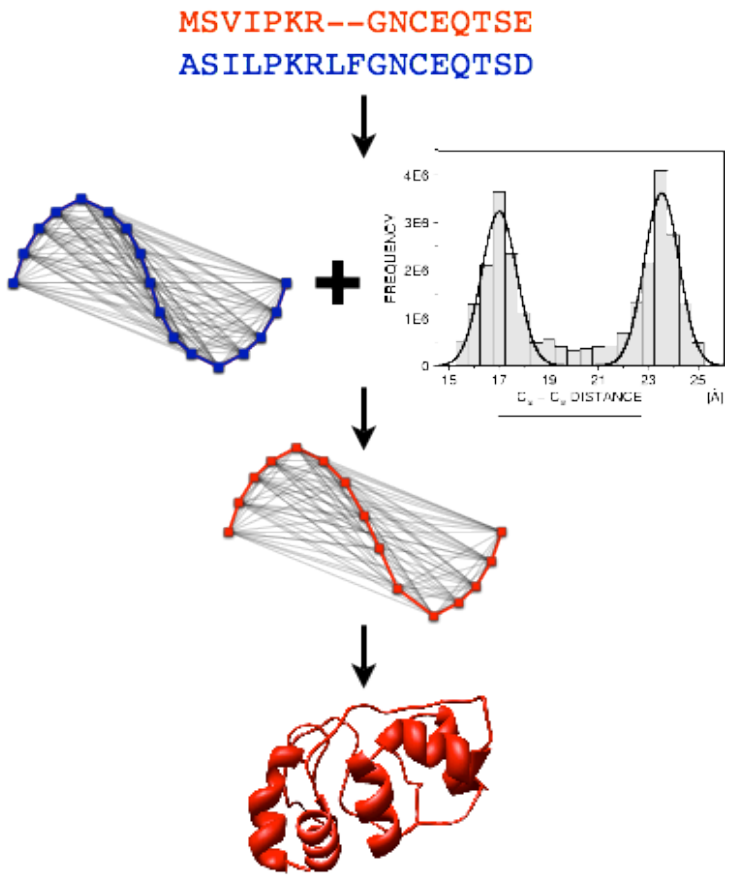


Given an alignment...

extract spatial features from the template(s) and statistics from known structures

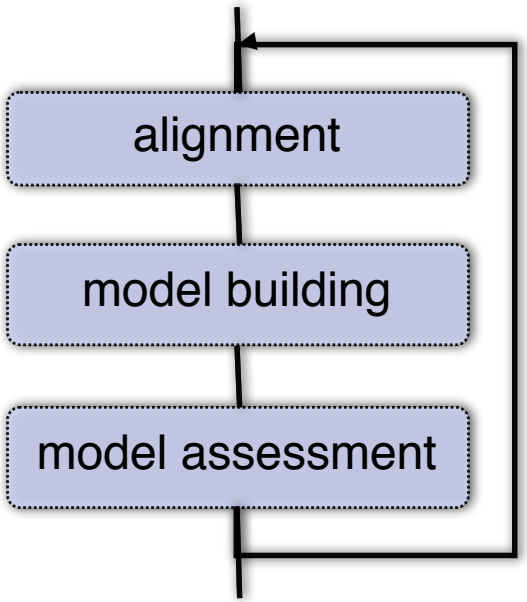
apply these features as restraints on your target sequence

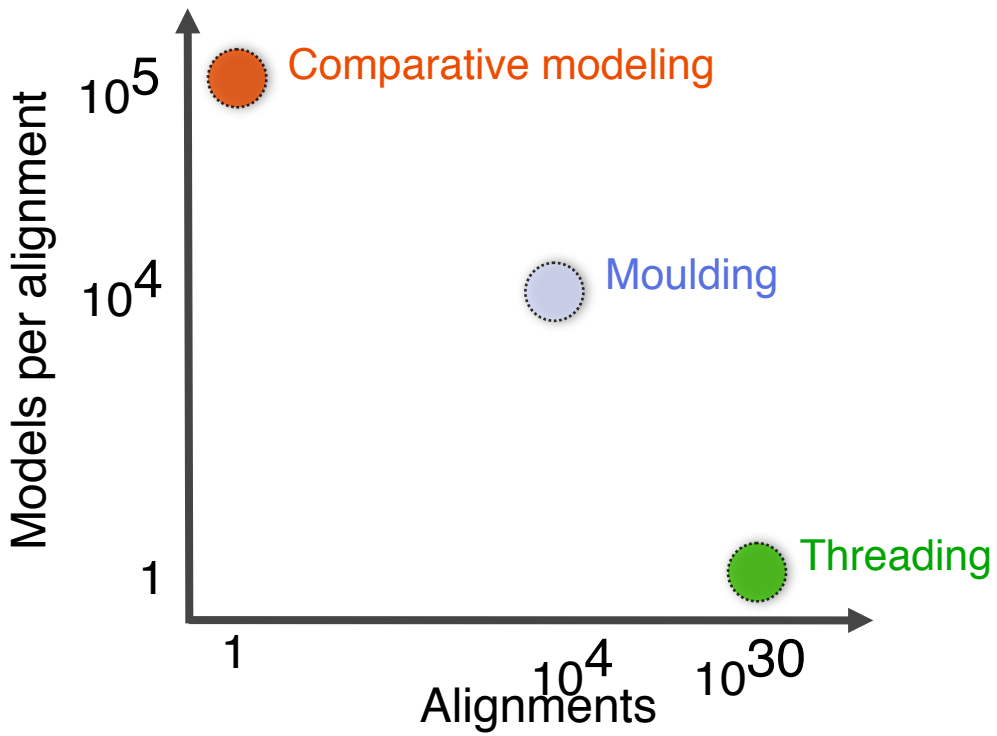
optimize to find the best solution for the restraints to produce your 3D model



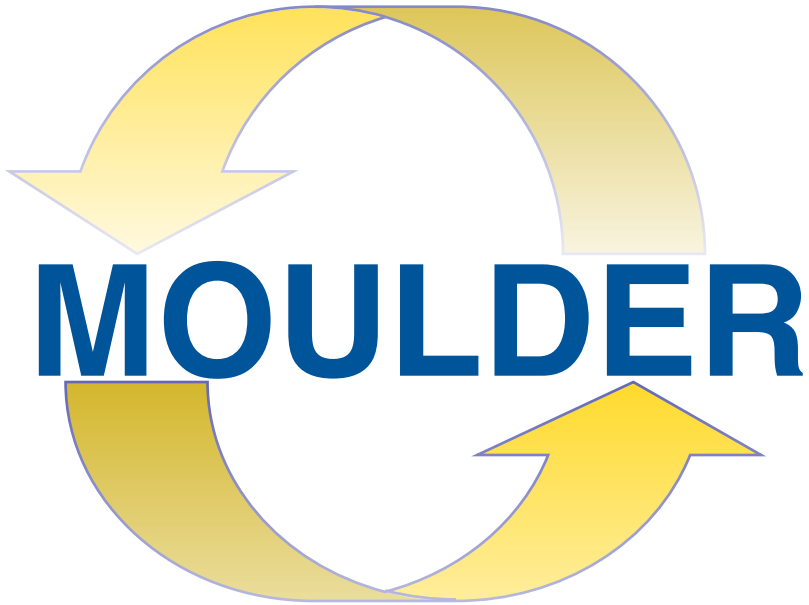
A. Šali & T. Blundell. J. Mol. Biol. 234, 779, 1993.
J.P. Overington & A. Šali. Prot. Sci. 3, 1582, 1994.
A. Fiser, R. Do & A. Šali, Prot. Sci., 9, 1753, 2000.

Types of homology modeling



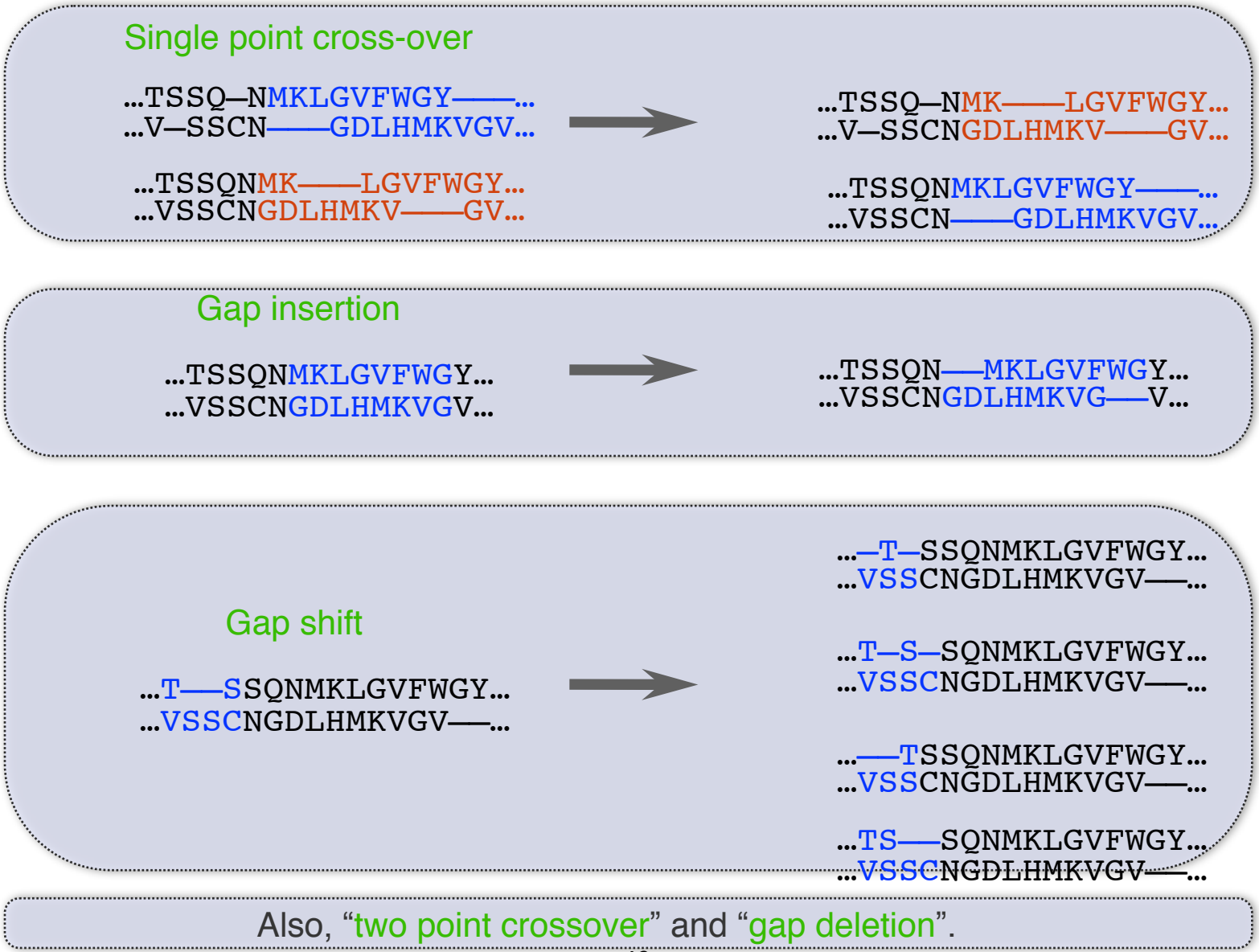


Moulding: iterative alignment, model building, model assessment



John, Sali (2003). NAR pp31 3982

Genetic algorithm operators

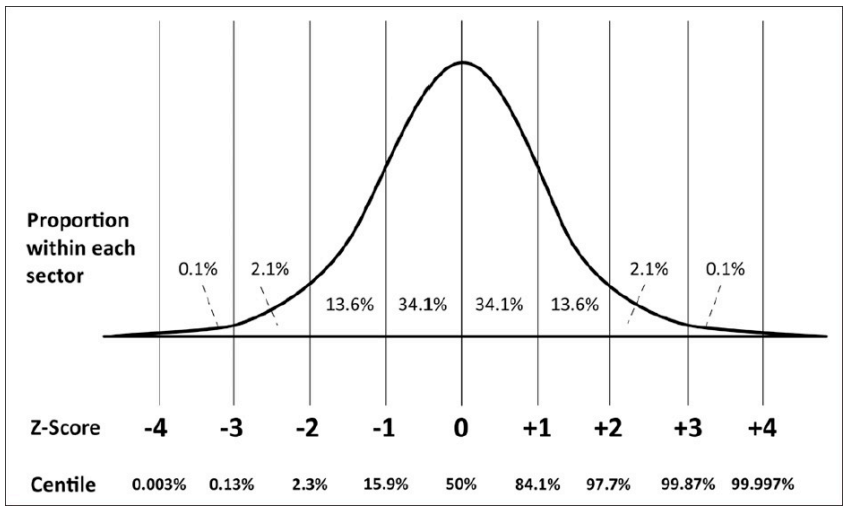


Composite model assessment score

- Weighted linear combination of several structural properties:

MINIMIZE!!

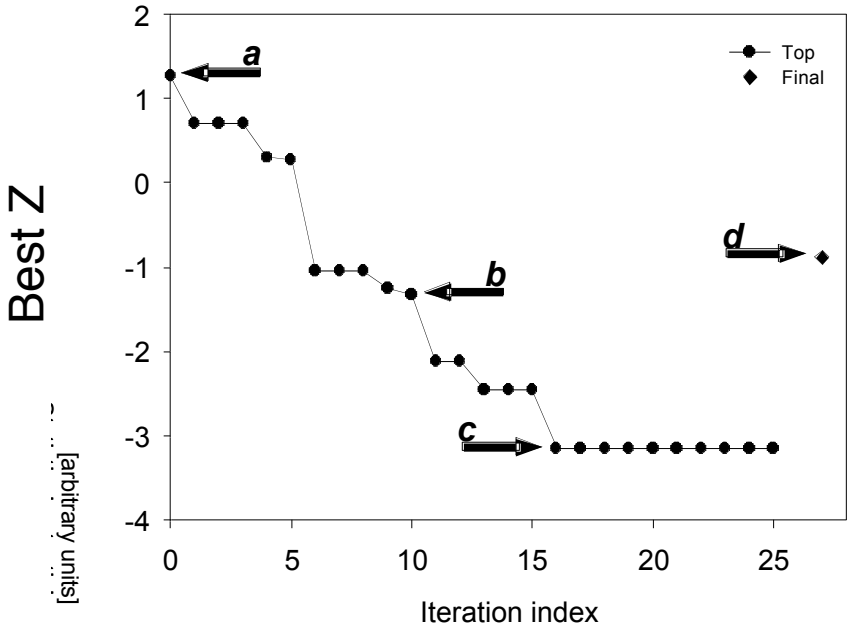
$$Z = 0.17\,Z(P_P) + 0.02\,Z(P_S) + 0.10\,Z(S_C) + 0.26\,Z(H_a) + 0.45\,(A_S)$$



John, Sali (2003). NAR pp31 3982

Application to a difficult modeling case

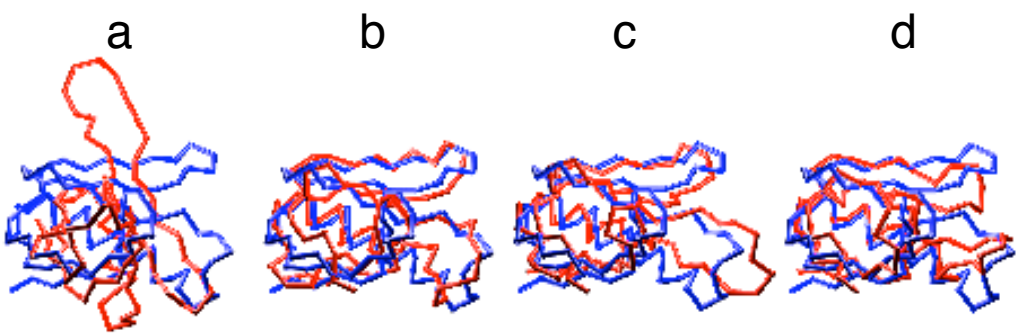
1BOV-1LTS



Sequence identity 4.4%

Initial model Cα RMSD 10.1Å

Final model Cα RMSD 3.6Å



Benchmark with the “very difficult” test set

D. Fischer threading test set of 68 structural pairs (a subset of 19)

Target -template	Sequence identity [%]	Coverage [% aa]	Initial prediction		Final prediction		Best prediction	
			Cα RMSD [Å]	CE overlap [%]	Cα RMSD [Å]	CE overlap [%]	Cα RMSD [Å]	CE overlap [%]
1ATR-1ATN	13.8	94.3	19.2	20.2	18.8	20.2	17.1	24.6
1BOV-1LTS	4.4	83.5	10.1	29.4	3.6	79.4	3.1	92.6
1CAU-1CAU	18.8	96.7	11.7	15.6	10.0	27.4	7.6	47.4
1COL-1CPC	11.2	81.4	8.6	44.0	5.6	58.6	4.8	59.3
1LFB-1HOM	17.6	75.0	1.2	100.0	1.2	100.0	1.1	100.0
1NSB-2SIM	10.1	89.2	13.2	20.2	13.2	20.1	12.3	26.8
1RNH-1HRH	26.6	91.2	13.0	21.2	4.8	35.4	3.5	57.5
1YCC-2MTA	14.5	55.1	3.4	72.4	5.3	58.4	3.1	75.0
2AYH-1SAC	8.8	78.4	5.8	33.8	5.5	48.0	4.8	64.9
2CCY-1BBH	21.3	97.0	4.1	52.4	3.1	73.0	2.6	77.0
2PLV-1BBT	20.2	91.4	7.3	58.9	7.3	58.9	6.2	60.7
2POR-2OMF	13.2	97.3	18.3	11.3	11.4	14.7	10.5	25.9
2RHE-1CID	21.2	61.6	9.2	33.7	7.5	51.1	4.4	71.1
2RHE-3HLA	2.4	96.0	8.1	16.5	7.6	9.4	6.7	43.5
3ADK-1GKY	19.5	100.0	13.8	26.6	11.5	37.7	7.7	48.1
3HHR-1TEN	18.4	98.9	7.3	60.9	6.0	66.7	4.9	79.3
4FGF-81IB	14.1	98.6	11.3	24.0	9.3	30.6	5.4	41.2
6XIA-3RUB	8.7	44.1	10.5	14.5	10.1	11.0	9.0	34.3
9RNT-2SAR	13.1	88.5	5.8	41.7	5.1	51.2	4.8	69.0
AVERAGE	14.2	85.2	9.6	36.7	7.7	44.8	6.3	57.8

Threading / Fold recognition

It is estimated there are only around 1000 to 10.000 stable folds in nature

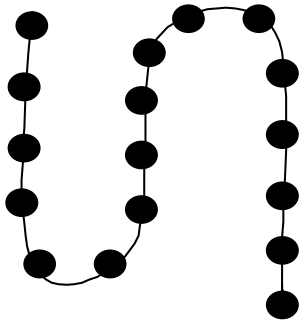
Fold recognition is essentially finding the best fit of a sequence to a set of candidate folds

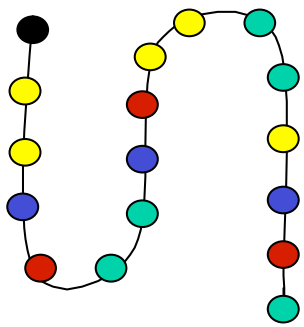
Find the best way to “mount” the residue sequence of one protein on a known structure taken from another protein

for each fold estimate the probability that the sequence can have that fold

Finding the best “fold”

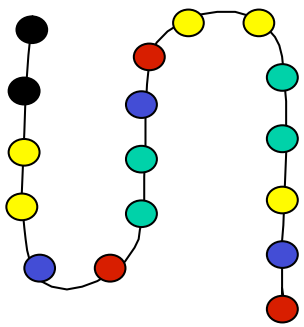
- H bond donor
- H bond acceptor
- Glycin
- Hydrophobic





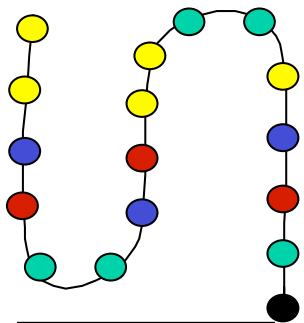
S=-2

Z= -1



S=5

Z=1.5



S=20

Z=5

Threading software

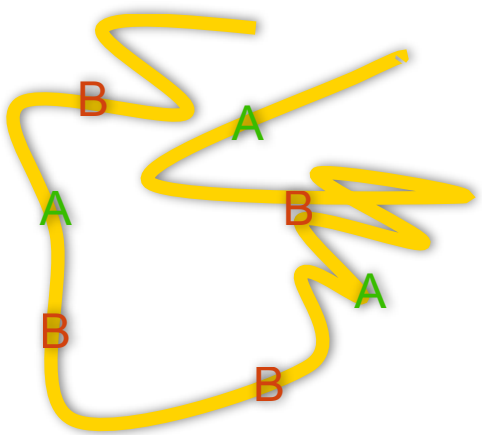
TOPITS: Heuristic Threader, part of larger structure prediction system.

3DPSSM: Integrated system, does its own MSA and secondary structure predictions and then threading. (http://ww.sbg.bio.ic.ac.uk/~3dpssm/index2.html)

GenThreader: Similar to 3DPSSM (http://bioinf.cs.ucl.ac.uk/psipred/)

Model Assessment

(Potentials of Mean Force)

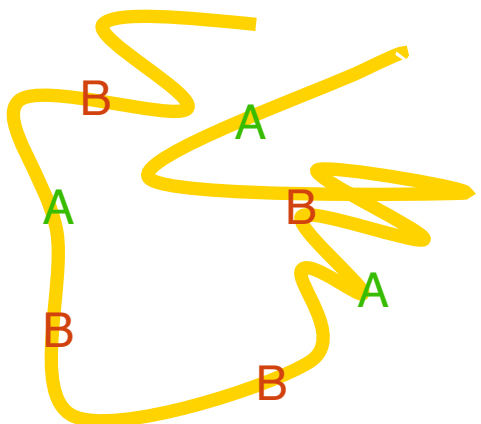


“The native structure generally has the lowest free energy of all states under the native conditions”

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]:

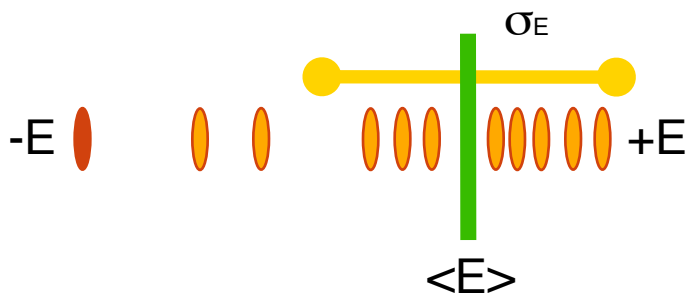
$$\Delta E = -kT \ln\left(\frac{N_1}{N_2}\right) \qquad (1)$$

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*
Sippl, (1990) J.Mo.Biol. 213 pp859
*Godzik, (1996) Structure **15** pp363*

Significance of an alignment (score)

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



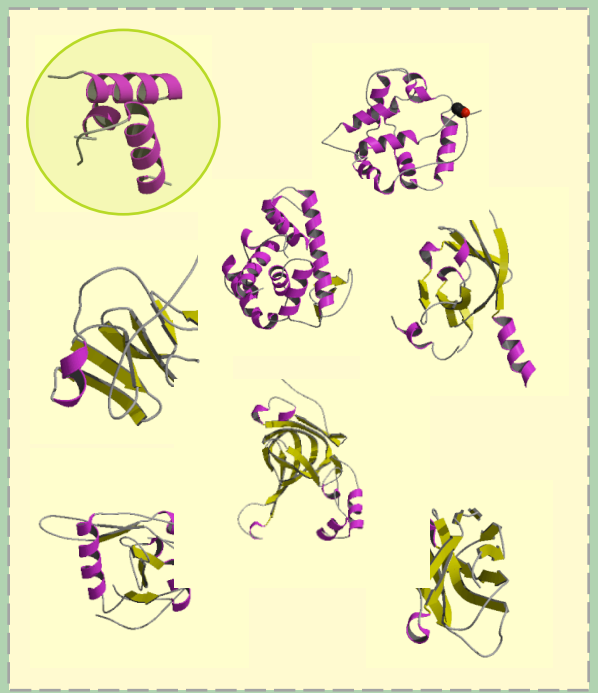
$$Zscore = \frac{(\langle E \rangle - E_m)}{\sigma_E}$$

ProSa

http://www.came.sbg.ac.at/prosa.php

Deriving

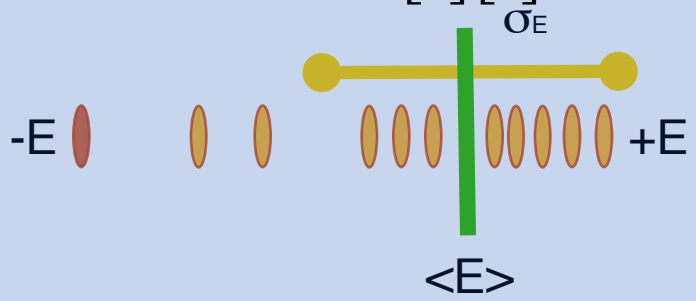
Structural space



Scoring

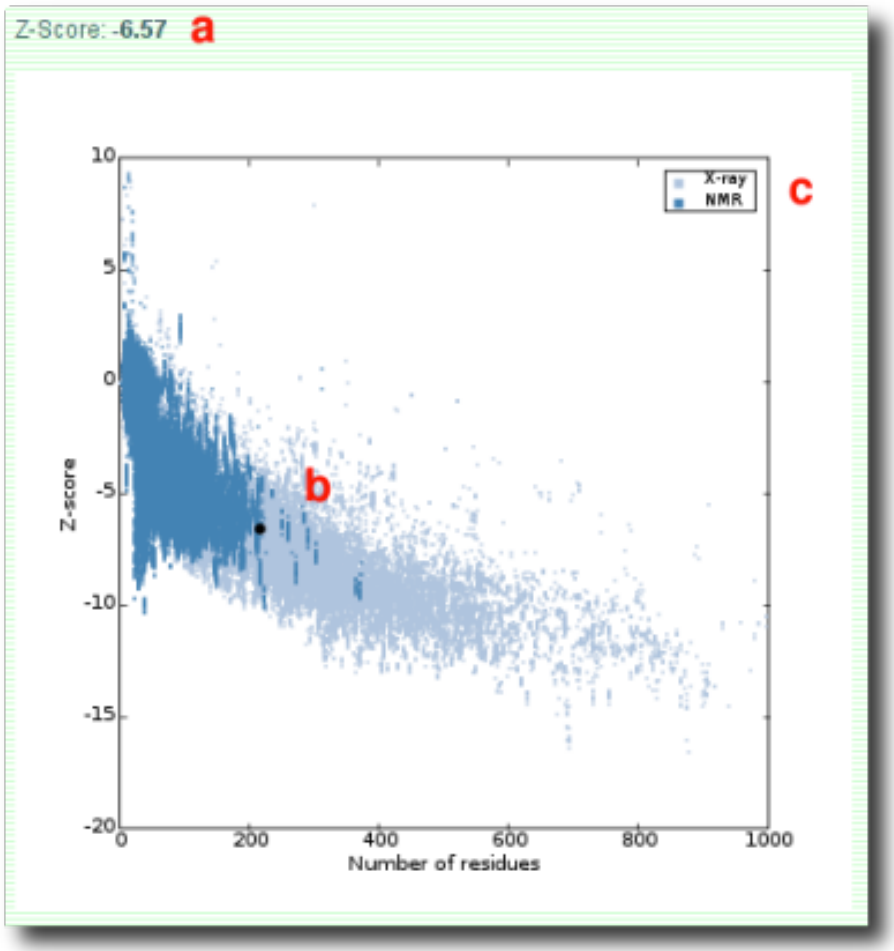
$$K = \frac{[AB]}{[A]\cdot[B]}$$

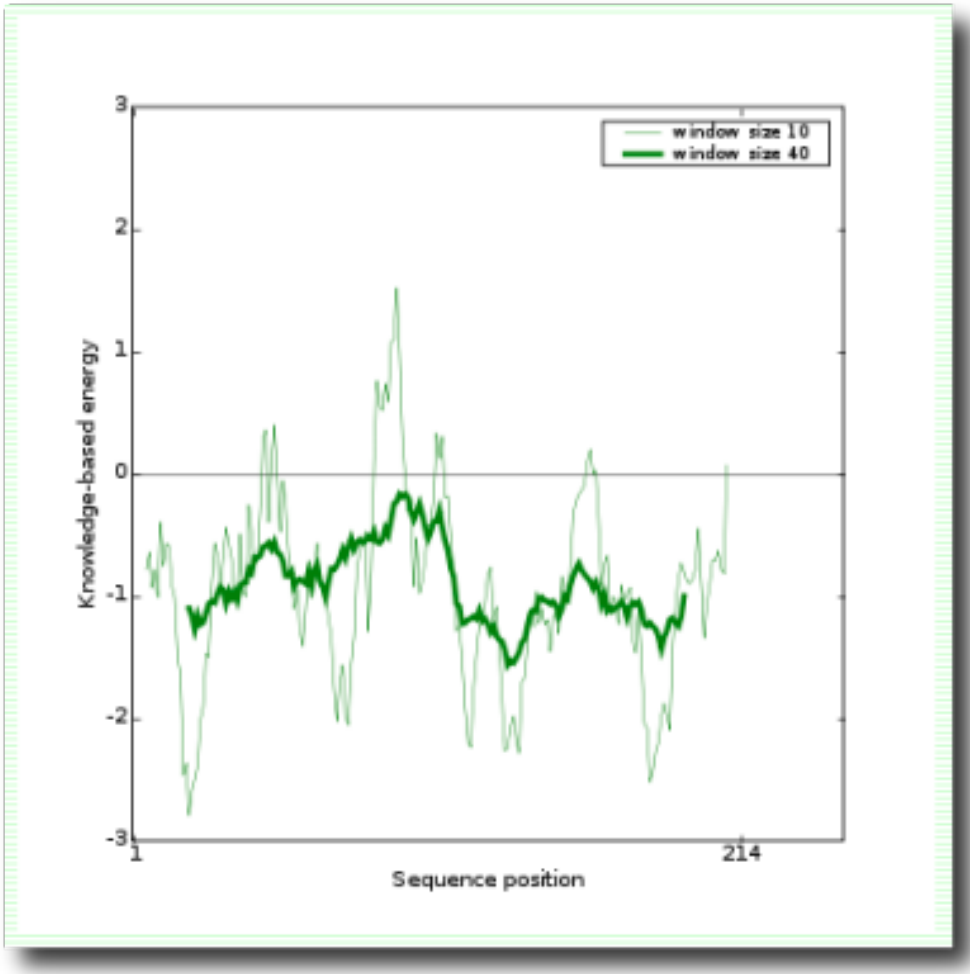
$$\Delta G = -RT\ln(K) = -RT\ln\frac{[AB]}{[A]\cdot[B]}$$



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

Understanding ProSa evaluation score



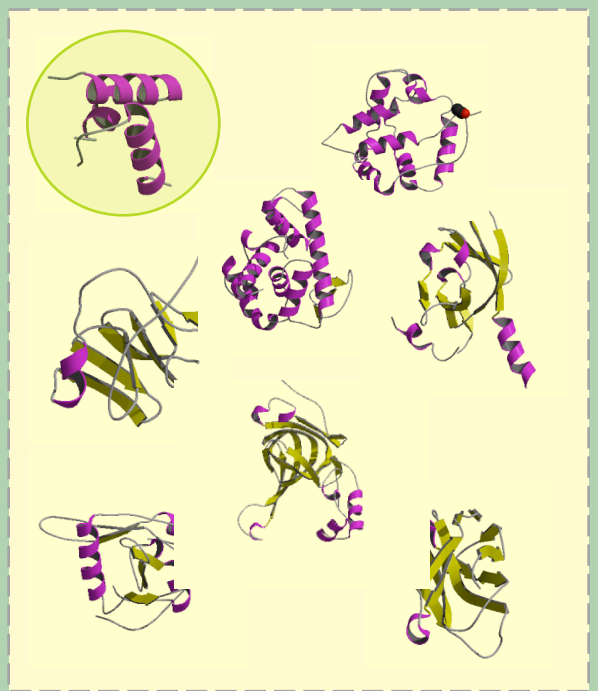


DOPE (MODELLER)

http://www.salilab.org/modeller/

Deriving

Structural space



Scoring

Pseudo-Energy with respect a ideal spherical protein as a reference state

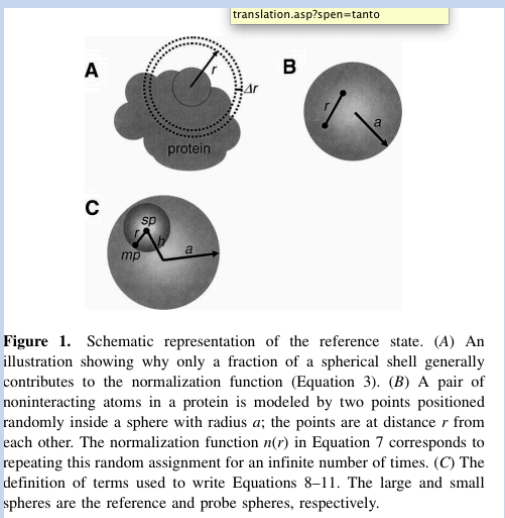


Figure 1. Schematic representation of the reference state. (*A*) An illustration showing why only a fraction of a spherical shell generally contributes to the normalization function (Equation 3). (*B*) A pair of noninteracting atoms in a protein is modeled by two points positioned randomly inside a sphere with radius a; the points are at distance r from each other. The normalization function $n(r)$ in Equation 7 corresponds to repeating this random assignment for an infinite number of times. (*C*) The definition of terms used to write Equations 8–11. The large and small spheres are the reference and probe spheres, respectively.

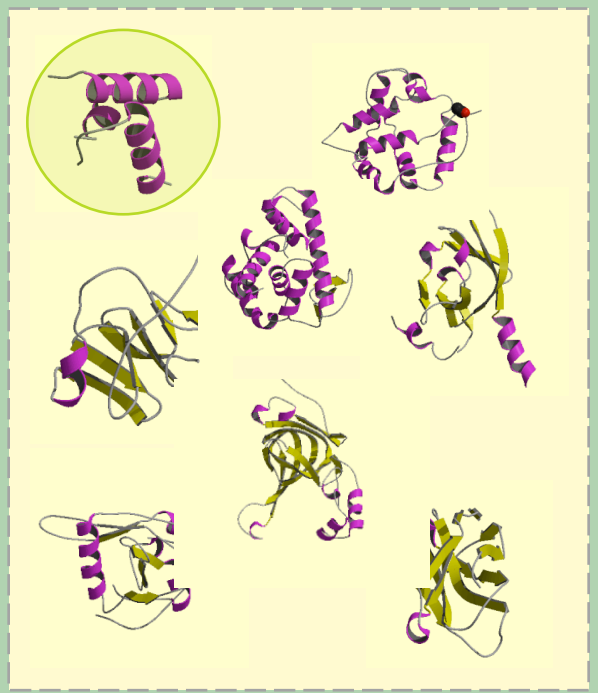
Shen, Min-yi, and Andrej Sali. "Statistical potential for assessment and prediction of protein structures." *Protein science* 15.11 (2006): 2507-2524.

DFIRE/DFIRE2

http://sparks.informatics.iupui.edu/yueyang/DFIRE/dDFIRE-service

Deriving

Structural space



Scoring

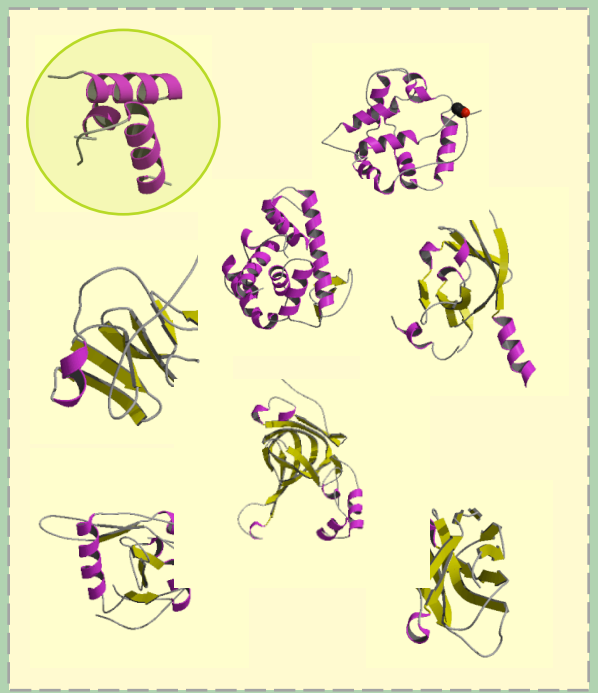
Pseudo-Energy with respect a ideal gas-phase reference state

Verify3D

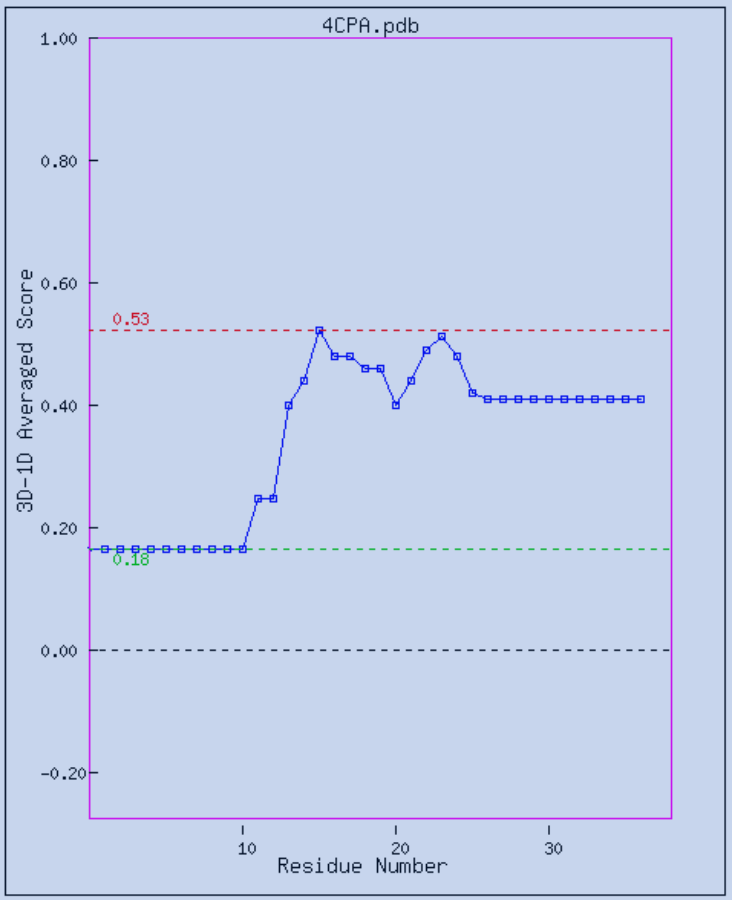
http://nihserver.mbi.ucla.edu/Verify_3D/

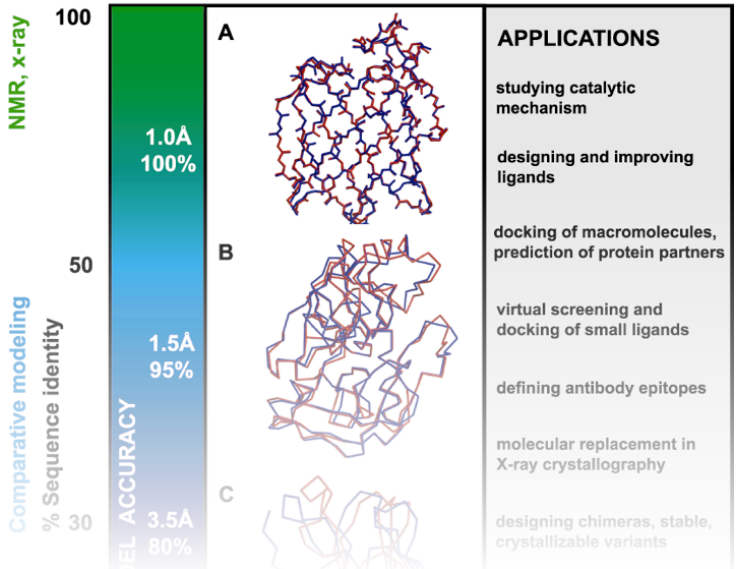
Deriving

Structural space



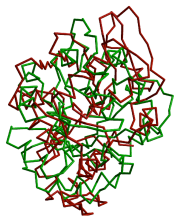
Scoring



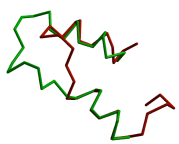


Accuracy and applicability of comparative models

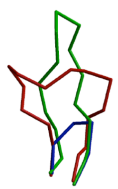
Comparative modeling by satisfaction of spatial restraints Types of errors and their impact



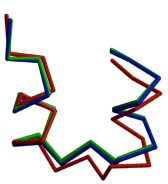
Wrong fold



Miss alignments

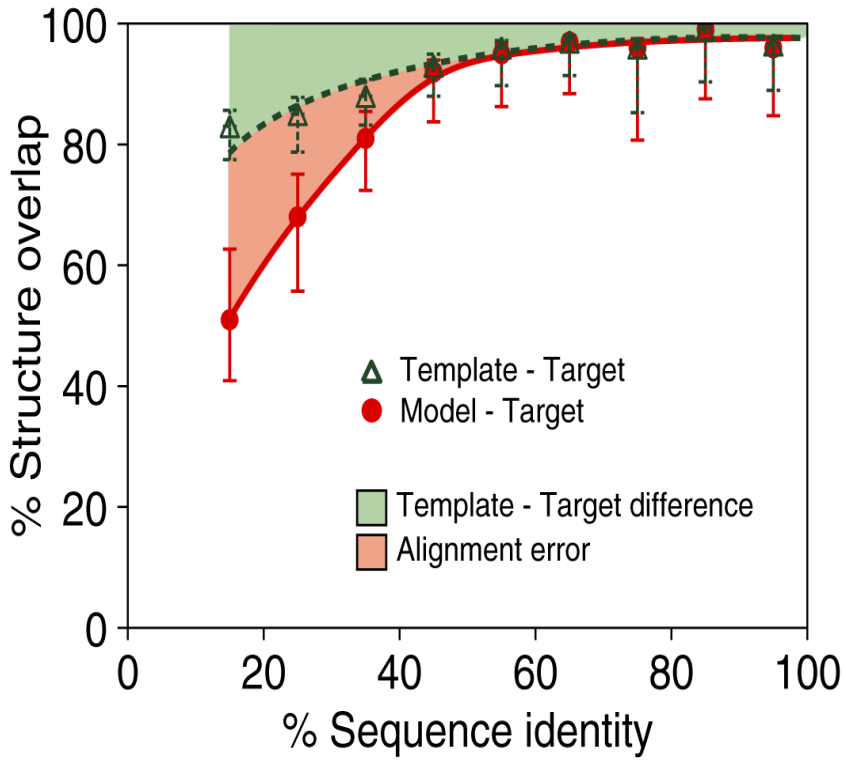


Loop regions



Rigid body distortions

Side-chain packing

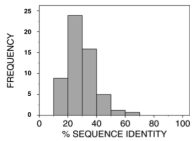


Template - Model - Real Structure

Marti-Renom etal. Ann Rev Biophys Biomol Struct (2000) 29, 291

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}.$$

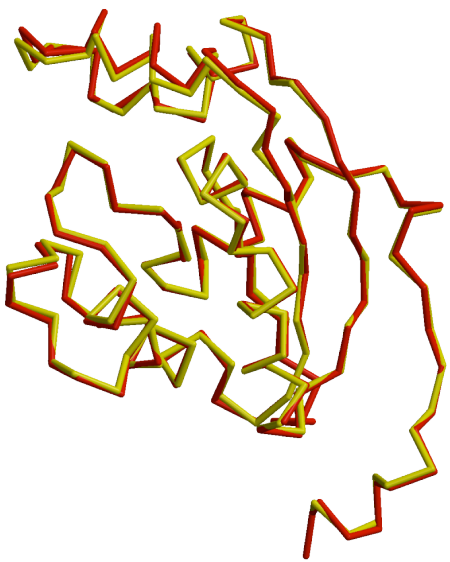
Model Accuracy



HIGH ACCURACY

NM23 Seq id 77%

Cα equiv 147/148
RMSD 0.41Å

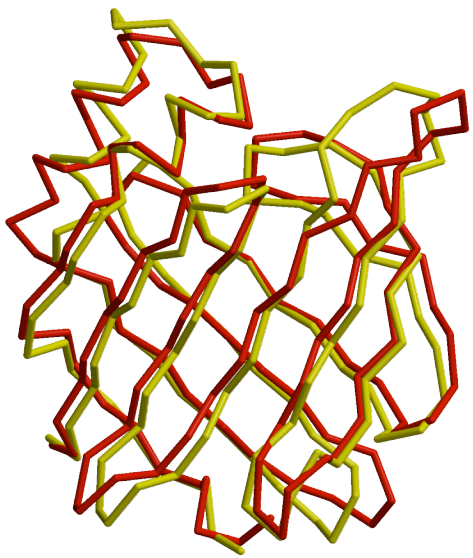


Sidechains
Core backbone
Loops

MEDIUM ACCURACY

CRABP Seq id 41%

Cα equiv 122/137
RMSD 1.34Å

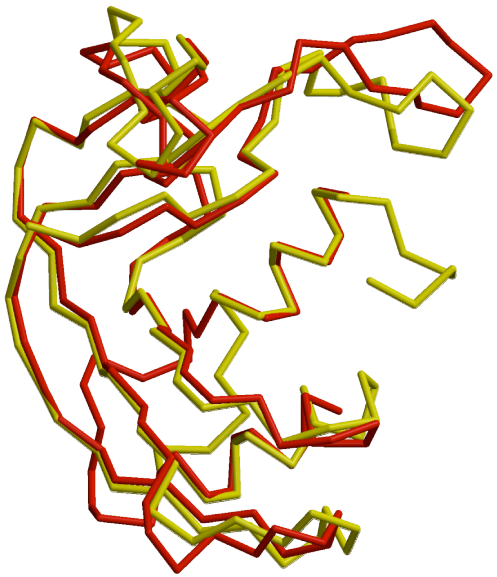


Sidechains
Core backbone
Loops
Alignment

LOW ACCURACY

EDN Seq id 33%

Cα equiv 90/134
RMSD 1.17Å

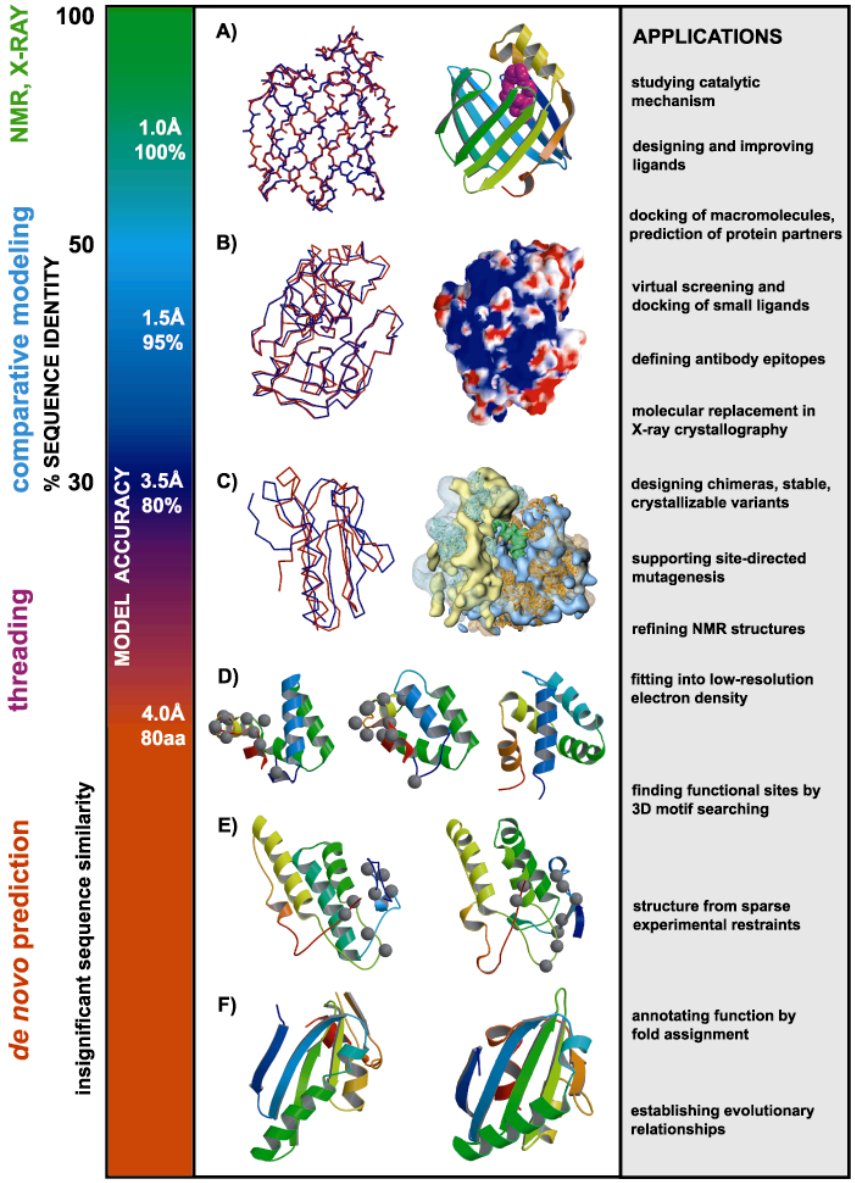


Sidechains
Core backbone
Loops
Alignment
Fold assignment

X-RAY / MODEL

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

Utility of protein structure models, despite errors



D. Baker & A. Sali. Science 294, 93, 2001.

Comparative Protein Structure Prediction
MODELLER tutorial

```
$>mod9v14 model.py
```

The current release of Modeller is **9.14**, which was released on July 29th, 2014

Obtaining MODELLER and related information

- MODELLER (9v14) web page
- http://www.salilab.org/modeller/
  - Download Software (Linux/Windows/Mac)
  - HTML Manual
  - **Online tutorial.**
  - **The password is...**

Using MODELLER

- No GUI! ☹
- Controlled by command file ☹☹
- Script is written in PYTHON language ☺
- You may know Python language is simple ☺☺

“MINIMAL” MODELLER

- INPUT:
  - **Target** Sequence (FASTA/PIR format)
  - **Template** Structure (PDB format)
  - **Python** script file
- OUTPUT:
  - Target-Template **Alignment**
  - **Model** in PDB format
  - Additional data (in .log)

Modeling of BLBP
Input

- Target: Brain lipid-binding protein (BLBP)
- BLBP sequence in PIR (MODELLER) format:

blbp.seq

```
>P1;blbp
sequence:blbp:::::::
VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSIDDRNCKSVVRLDGD
KLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA*
```

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Python script for target-template alignment

align.py

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:


log.verbose()
env = environ()


aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v11 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP
STEP 1: Align blbp and 1hms sequences
Python script for target-template alignment

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:


log.verbose()
env = environ()


aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v11 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP
STEP 1: Align blbp and 1hms sequences
Python script for target-template alignment

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:


log.verbose()
env = environ()


aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v11 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Python script for target-template alignment

```
# Example for: alignment.align()

# This will read two sequences, align them, and write the alignment
# to a file:


log.verbose()
env = environ()


aln = alignment(env)
mdl = model(env, file='1hms')
aln.append_model(mdl, align_codes='1hms')
aln.append(file='blbp.seq', align_codes=('blbp'))

# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1hms.ali', alignment_format='PIR')
aln.write(file='blbp-1hms.pap', alignment_format='PAP')
```

Run by typing `mod9v11 align.py` in the directory where you have the python file.
MODELLER will produce a `align.log` file

Modeling of BLBP
STEP 1: Align blbp and 1hms sequences
Output

blbp-1hms.ali

```
>P1;1hms
structureX:1hms:    1 :A: 131 :A:undefined:undefined:-1.00:-1.00
VDAFLGTWKLVDSKNFDDYMKSLGVGFATRQVASMTKPTTIIEKNGDILTLKTHSTFKNTEISFKLGVEFDETTA
DDRKVKSIVTLDGGKLVHLQKWDGQETTLVRELIDGKLILTLTHGTAVCTRTYEKE*
>P1;blbp
sequence:blbp:      : :      : : : : 0.00: 0.00
VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSI
DDRNCKSVVRLDGDKLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA*
```

Modeling of BLBP
STEP 1: Align blbp and 1hms sequences
Output

```
>P1;1hms
structureX:1hms:    1 : : 131 : :undefined:undefined:-1.00:-1.00
VDAFLGTWKLVDSKNFDDYMKSLGVGFATRQVASMTKPTTIIEKNGDILTLKTHSTFKNTEISFKLGVEFDETTA
DDRKVKSIVTLDGGKLVHLQKWDGQETTLVRELIDGKLILTLTHGTAVCTRTYEKE*
>P1;blbp
sequence:blbp:      : :      : : : : 0.00: 0.00
VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSI
DDRNCKSVVRLDGDKLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA*
```

Modeling of BLBP

STEP 1: Align blbp and 1hms sequences

Output

blbp-1hms.pap

```
 _aln.pos          10         20         30         40         50         60
1hms       VDAFLGTWKLVDSKNFDDYMKSLGVGFATRQVASMTKPTTIIEKNGDILTLKTHSTFKNTEISFKLGV
blbp       VDAFCATWKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGE
 _consrvd  ****   **** ** *** *** **********    **** **    *       *  ******* * **


 _aln.p    70         80         90         100        110        120        130
1hms       EFDETTADDRKVKSIVTLDGGKLVHLQKWDGQETTLVRELIDGKLILTLTHGTAVCTRTYEKE
blbp       EFEETSIDDRNCKSVVRLDGDKLIHVQKWDGKETNCTREIKDGKMVVTLTFGDIVAVRCYEKA
 _consrvd  ** **   ***   ** * *** ** * ***** **    **   ***    *** *   *   * ***
```

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

```
# Homology modelling by the automodel class
from modeller.automodel import *       # Load the automodel class
log.verbose()                          # request verbose output
env = environ()                        # create a new MODELLER environment

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'


a = automodel(env,
              alnfile  = 'blbp-1hms.ali',     # alignment filename
              knowns   = '1hms',              # codes of the templates
              sequence = 'blbp')              # code of the target


a.starting_model= 1                 # index of the first model
a.ending_model  = 1                 # index of the last model
                                    # (determines how many models to calculate)
a.make()                            # do the actual homology modelling
```

Run by typing `mod9v11 model.py` in the directory where you have the python file.
MODELLER will produce a `model.log` file

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

```
# Homology modelling by the automodel class
from modeller.automodel import *    # Load the automodel class
log.verbose()                       # request verbose output
env = environ()                     # create a new MODELLER environment

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'


a = automodel(env,
              alnfile  = 'blbp-1hms.ali',     # alignment filename
              knowns   = '1hms',              # codes of the templates
              sequence = 'blbp')              # code of the target


a.starting_model= 1                 # index of the first model
a.ending_model  = 1                 # index of the last model
                                    # (determines how many models to calculate)
a.make()                            # do the actual homology modelling
```

Run by typing `mod9v11 model.py` in the directory where you have the python file.
MODELLER will produce a `model.log` file

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

```
# Homology modelling by the automodel class
from modeller.automodel import *    # Load the automodel class
log.verbose()                       # request verbose output
env = environ()                     # create a new MODELLER environment

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'


a = automodel(env,
              alnfile  = 'blbp-1hms.ali',     # alignment filename
              knowns   = '1hms',              # codes of the templates
              sequence = 'blbp')              # code of the target


a.starting_model= 1                 # index of the first model
a.ending_model  = 1                 # index of the last model
                                    # (determines how many models to calculate)
a.make()                            # do the actual homology modelling
```

Run by typing `mod9v11 model.py` in the directory where you have the python file.
MODELLER will produce a `model.log` file

Modeling of BLBP

STEP 2: Model the blbp structure using the alignment from step 1.

Python script for model building

PDB file

Can be viewed with Chimera

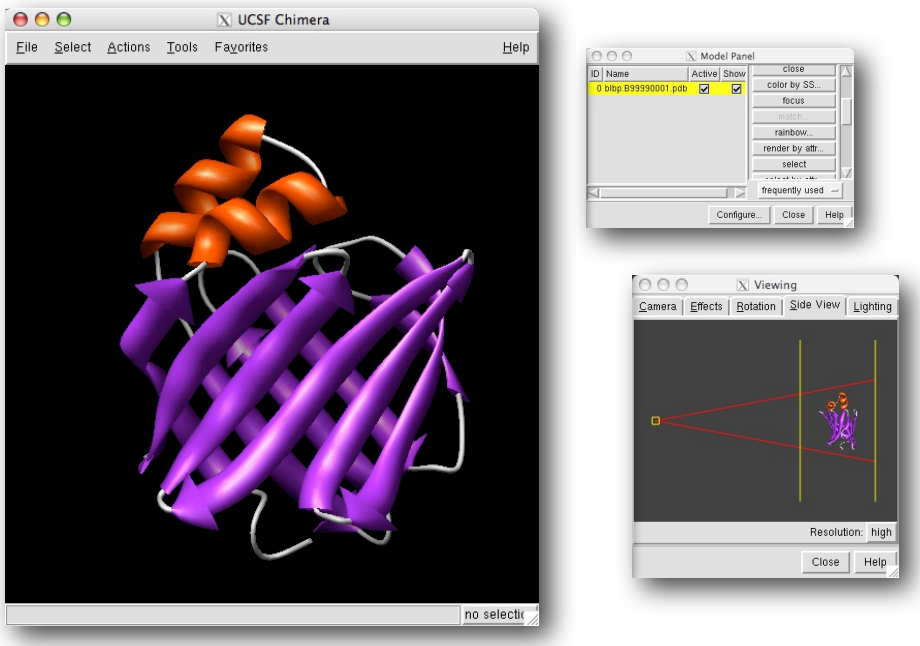
http://www.cgl.ucsf.edu/chimera/

PyMol

http://www.pymol.org

Rasmol

http://www.openrasmol.org

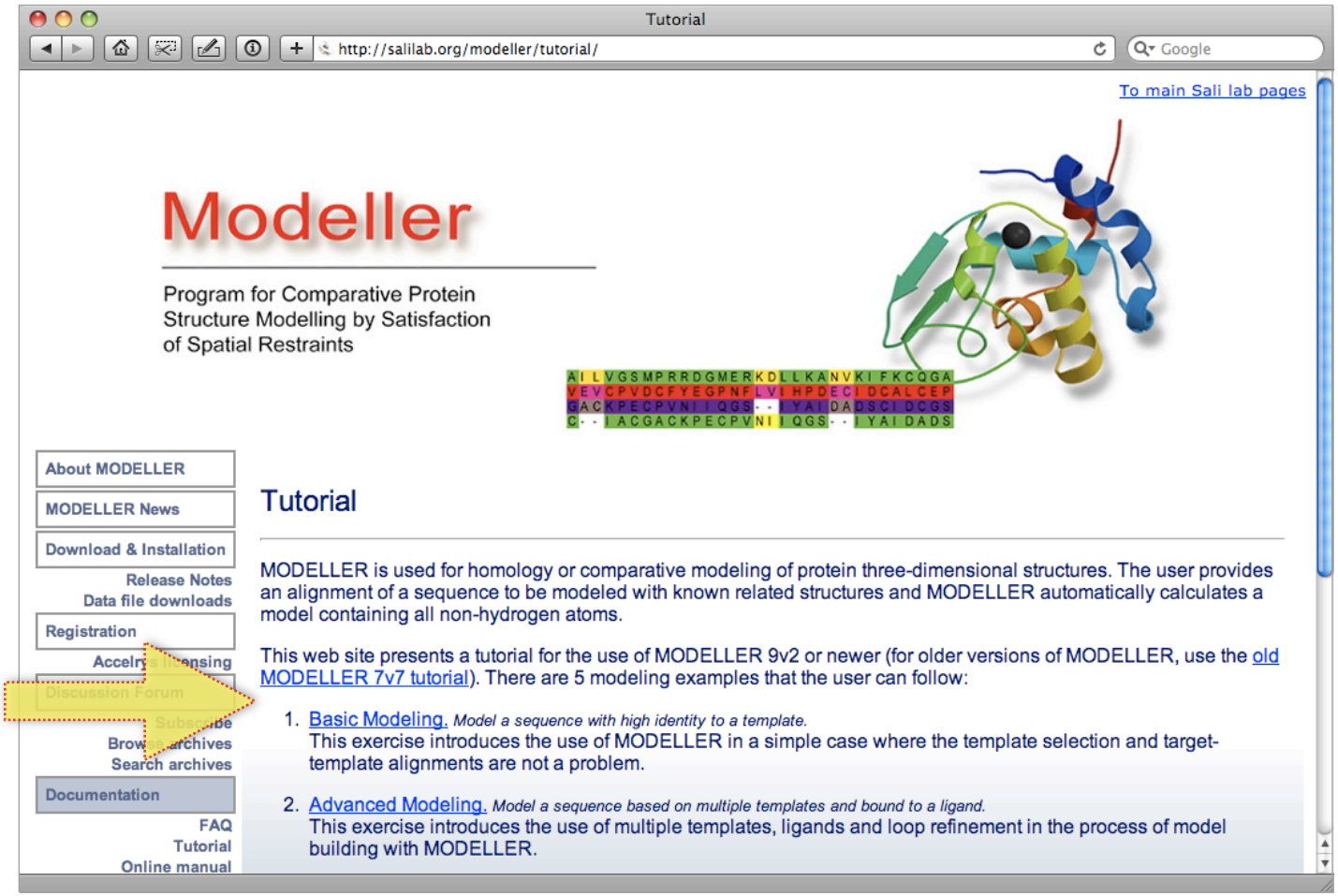


- Model file →

  `blbp.B99990001.pdb`

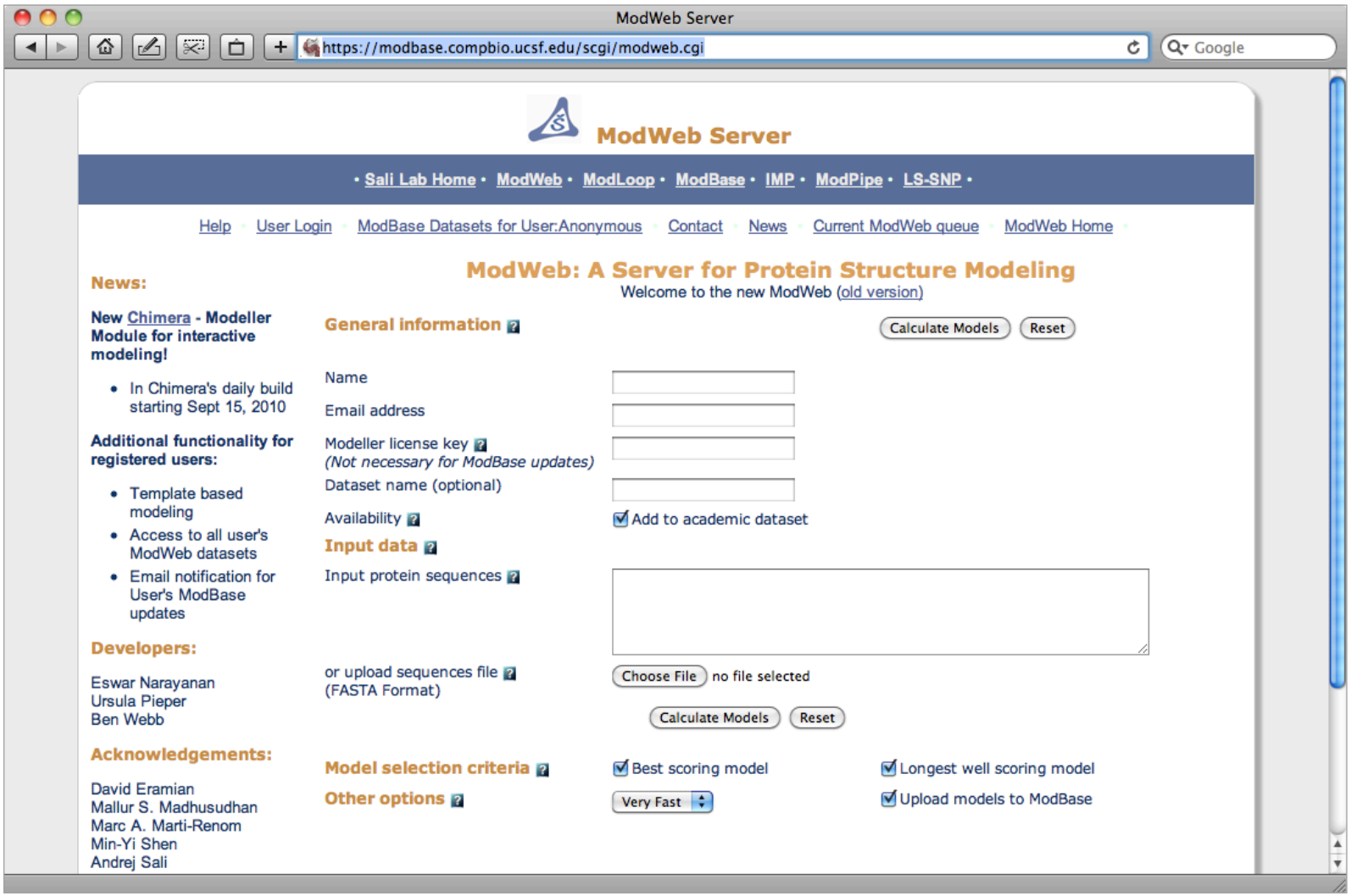
Get ready for 11th May!

http://www.salilab.org/modeller/tutorial/



FULLY AUTOMATIC MODWEB

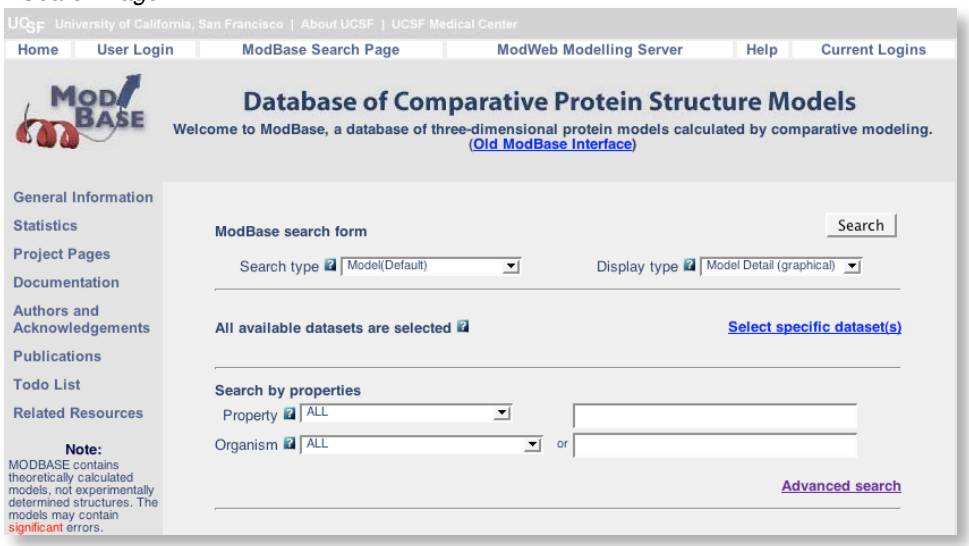
http://salilab.org/modweb



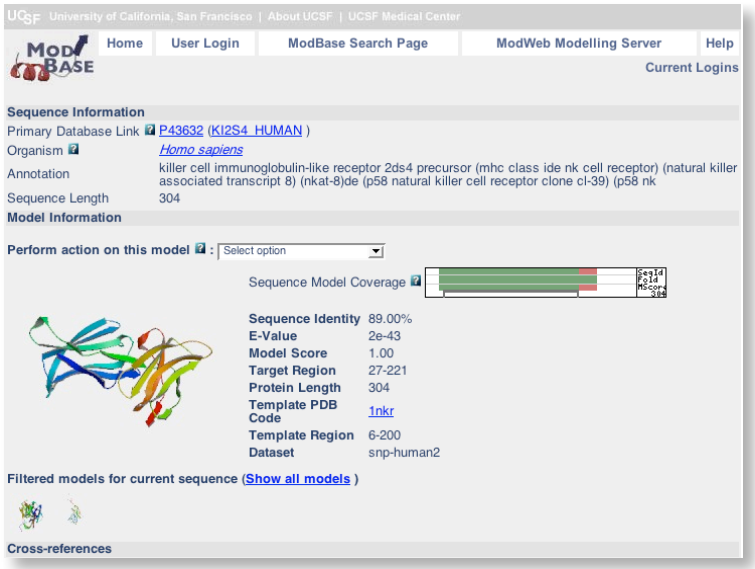
MODBASE

http://salilab.org/modbase

Search Page



Model Details



Sequence Overview

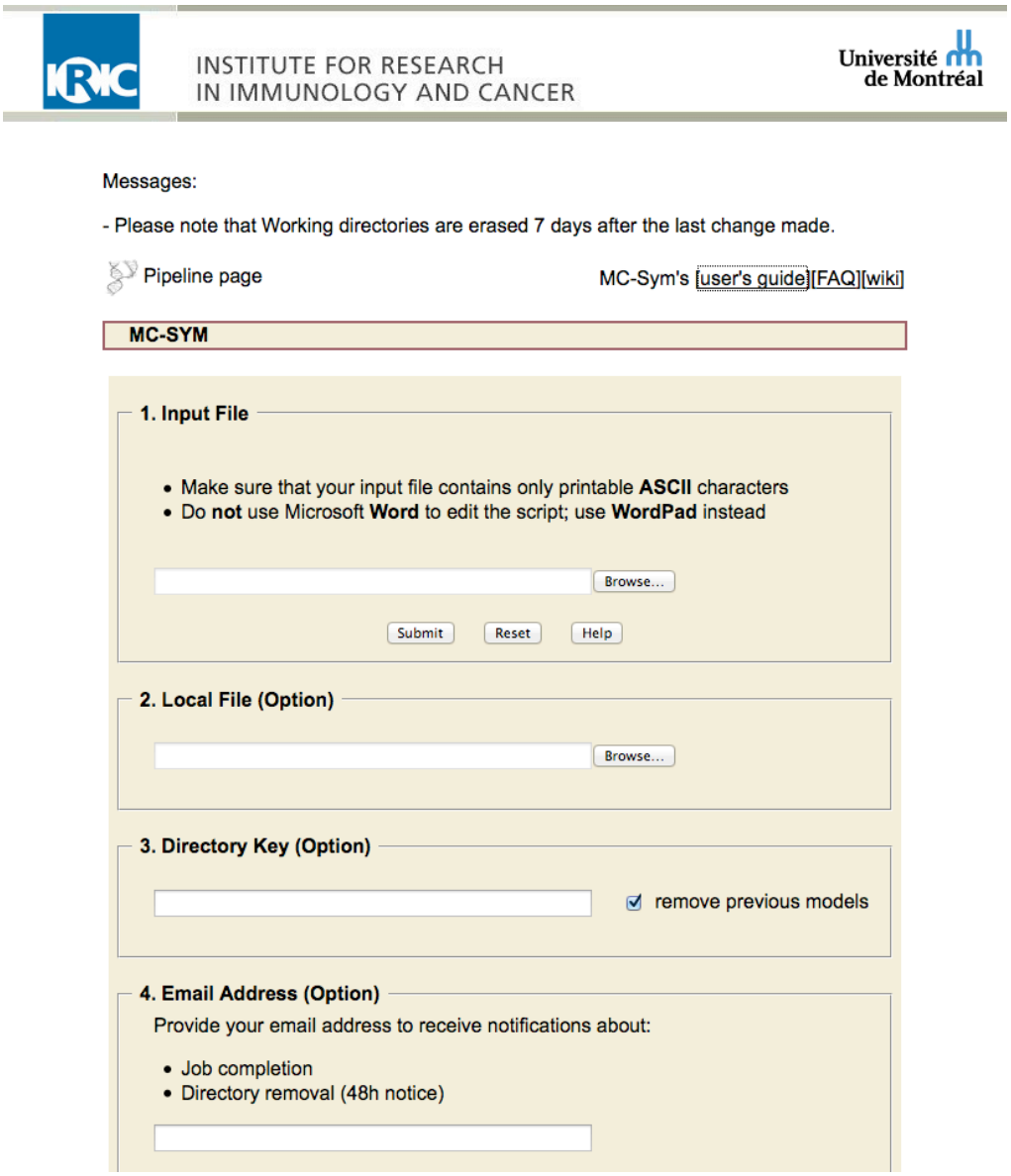
Model Overview

- *Pieper et al. (2004) Nucleic Acids Research 32, D217-D222*

Modeling RNA

MC-FOLD / MC-SYM

http://www.major.iric.ca/MajorLabEn/MC-Tools.html
http://www.major.iric.ca/MC-Sym/



- Coarse-grained model
- Fragment-based approach
- Fully automated

- Slow
- Sampling bottleneck
- Relative poor accuracy

ROSIE

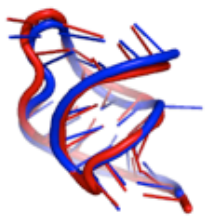
http://rosie.rosettacommons.org/rna_denovo

Welcome to ROSIE

Rosetta Online Server that Includes Everyone

Welcome | Queue | Documentation | Support | Login | Create an account

Rosetta RNA_Denovo Protocol



[Submit RNA_Denovo task]

[RNA Denovo Queue]

[RNA Denovo Documentation]

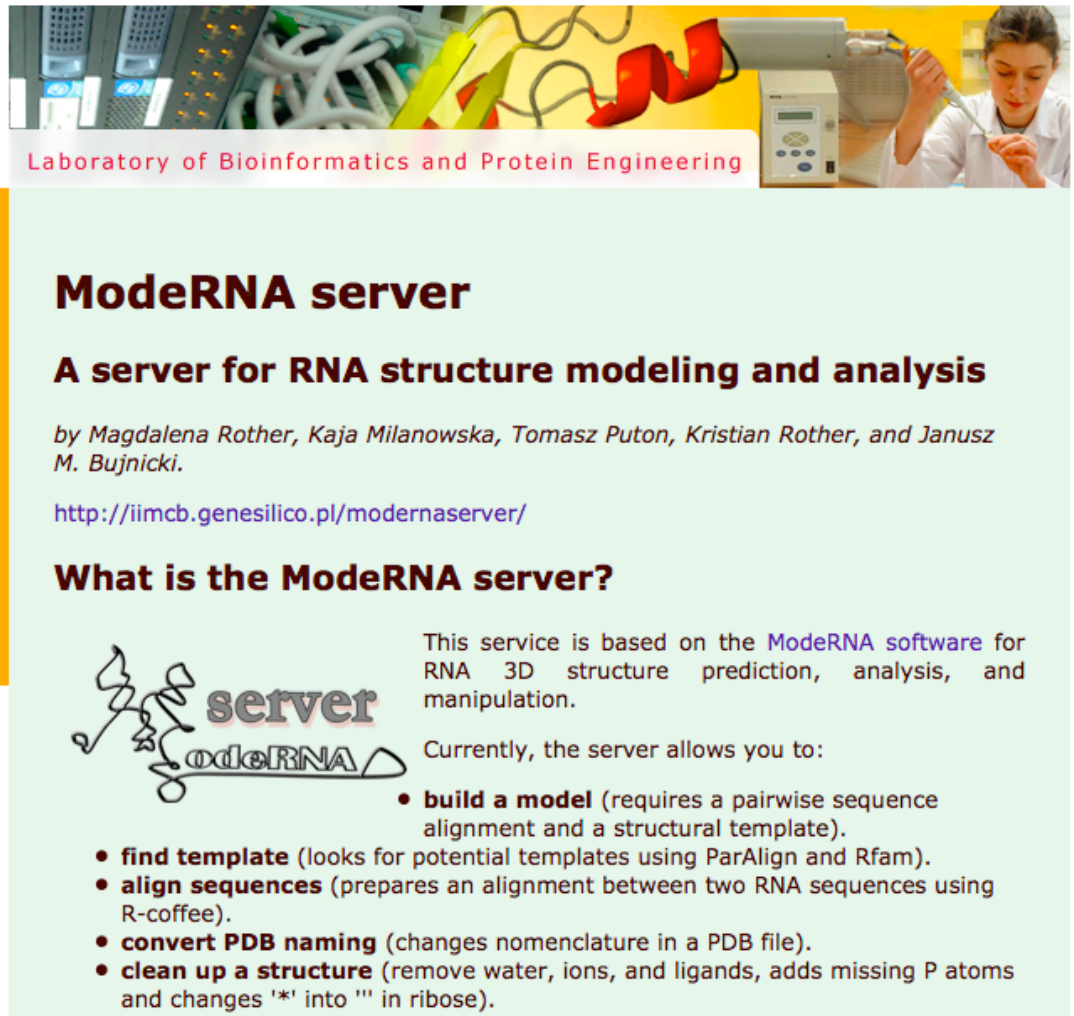
Please cite the following article when referring to results from our RNA_Denovo server:

- Das, R., Karanicolas, J., Baker., D. "Atomic accuracy in predicting and designing noncanonical RNA structure". *Nature Methods* 7:291–294. (2010). Online | PDF

rosie is web front-end for Rosetta software suite. Developed by Sergey Lyskov, GrayLab at JHU. Copyright © 2012 Rosetta Commons Member Institutions.

MODERNA

http://iimcb.genesilico.pl/modernaserver/

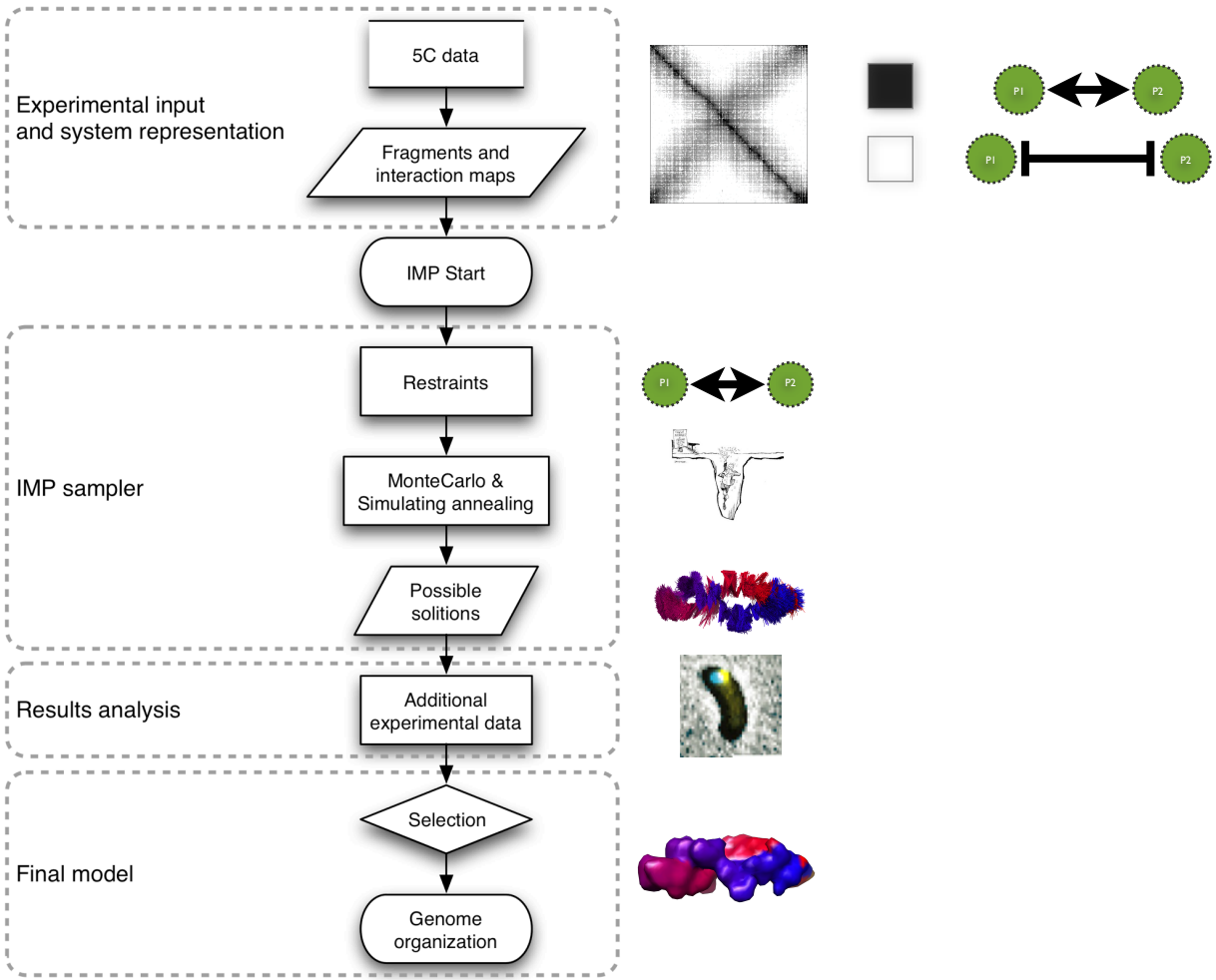


- Comparative RNA modeling
- Able to handle 115 nucleotide modifications

Modeling Chromatin

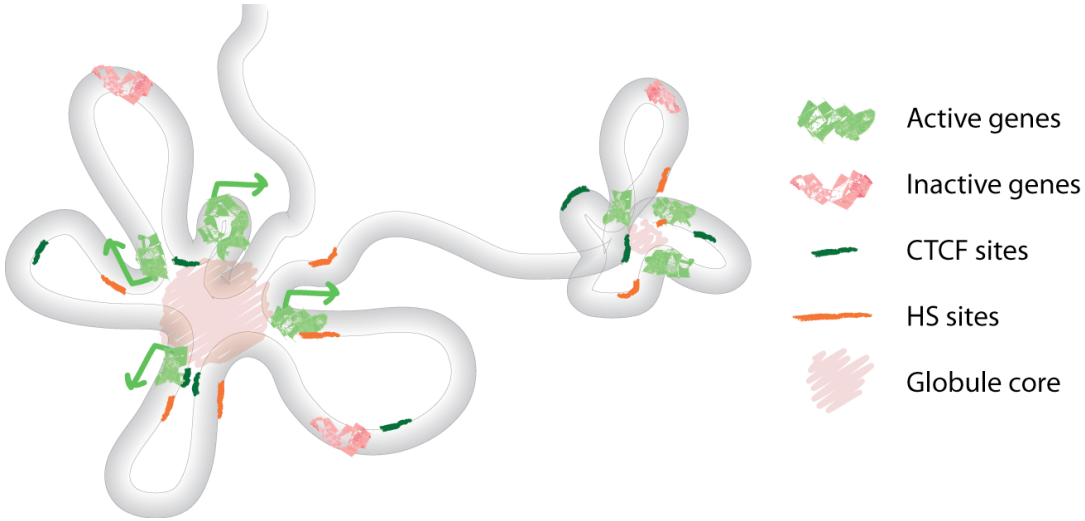
Integrative Modeling with RNA and chromatin

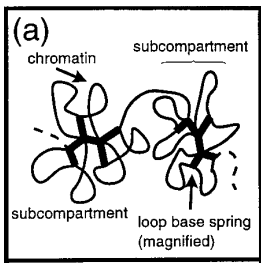
http://www.integrativemodeling.org



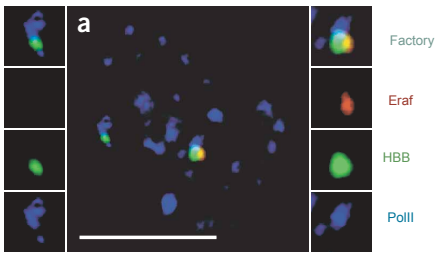
The "Chromatin Globule" model

D. Baù et al. Nat Struct Mol Biol (2011) 18:107-14
A. Sanyal et al. Current Opinion in Cell Biology (2011) 23:325–33.

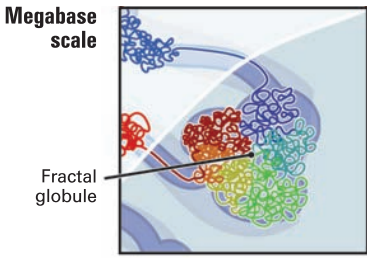




Münkel et al. JMB (1999)



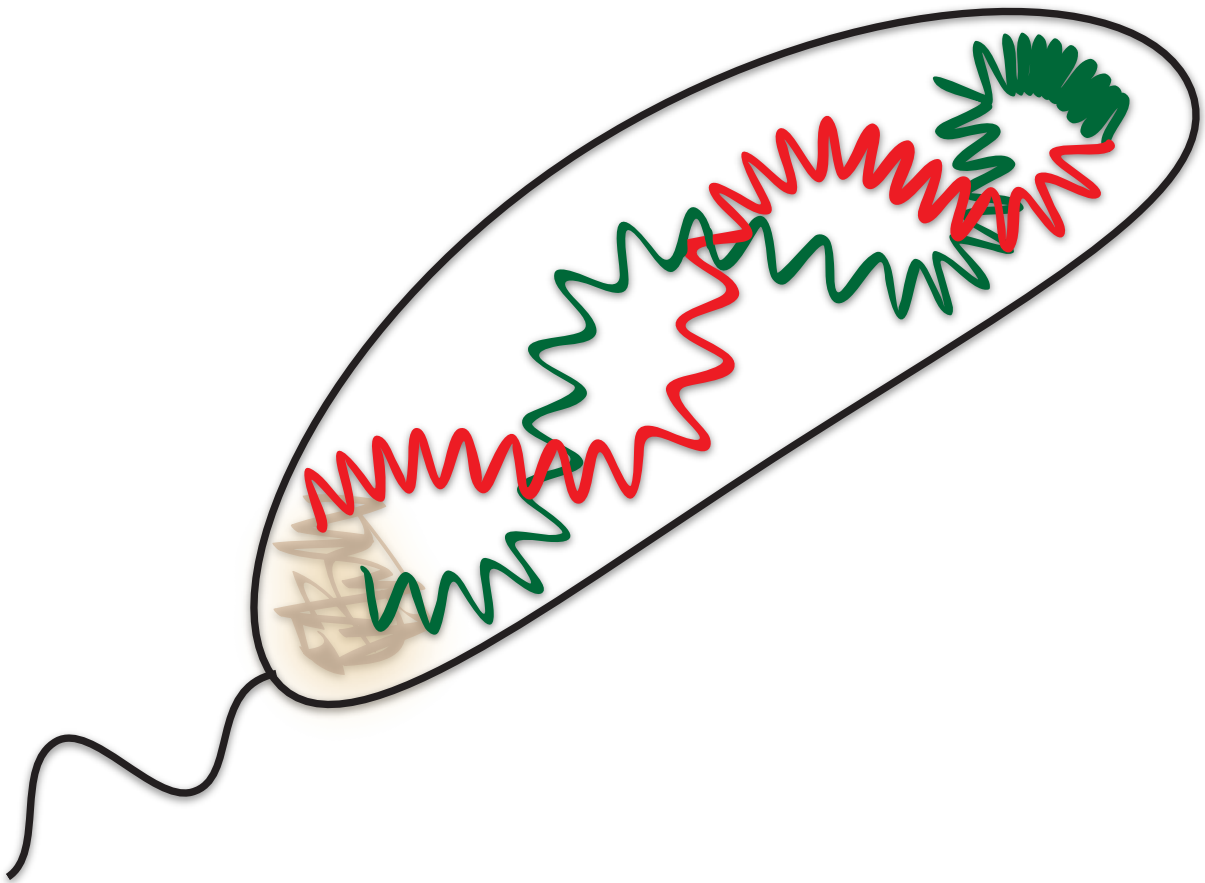
Osborne et al. Nat Genet (2004)



Lieberman-Aiden et al. Science (2009)

Caulobacter crescentus 3D genome

M.A. Umbarger, et al. Molecular Cell (2011) 44:252–264



cnag CRG Centre for Genomic Regulation

Take-home messages

- Homology != Identity
- Comparative modeling needs prior knowledge and a template.
- Our model relies on the **alignment** step...
- Evaluation is the key step, statistical potential.
- You should play with modeller, tutorial 11th may!

General References

Protein Structure Prediction:

Marti-Renom et al. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
Baker & Sali. Science 294, 93-96, 2001.Cited by 1338
Pavlopoulou, A., & Pavlopoulou, A. (2011). State-of-the-art bioinformatics protein structure prediction tools (Review). International Journal of Molecular Medicine, 28, 295-310. http://dx.doi.org/10.3892/ijmm.2011.705

Comparative Modeling:

Madhusudhan et al. The Proteomics Protocols Handbook. Ed. Walker. Humana Press Inc., Totowa, NJ. 831-860, 2005.
Sánchez, Roberto, and Andrej Šali. "Comparative protein structure modeling: introduction and practical examples with modeller." Protein Structure Prediction. Humana Press, 2000. 97-129.
Zhang, Yang. "I-TASSER server for protein 3D structure prediction." BMC bioinformatics 9.1 (2008): 40.
Kim, David E., Dylan Chivian, and David Baker. "Protein structure prediction and analysis using the Robetta server." Nucleic acids research 32.suppl 2 (2004): W526-W531.

MODELLER:

Sali & Blundell. J. Mol. Biol. 234, 779-815, 1993.
Eswar, M. A. et al. Comparative Protein Structure Modeling With MODELLER.
Current Protocols in Bioinformatics, John Wiley & Sons, Inc.,Supplement 15, 5.6.1-5.6.30, 2006.

Structural Genomics:

Sali. Nat. Struct. Biol. 5, 1029, 1998.
Burley et al. Nat. Genet. 23, 151, 1999.
Sali & Kuriyan. TIBS 22, M20, 1999.
Sanchez et al. Nat. Str. Biol. 7, 986, 2000.
Baker & Sali. Science 294, 93-96, 2001.
Terwilliger et al. Annu. Rev. Biophys. 38, 371-383, 2009.