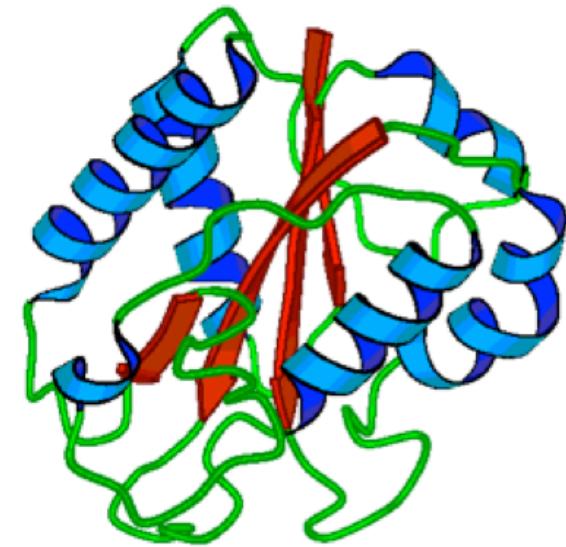


Bioinformática Estructural

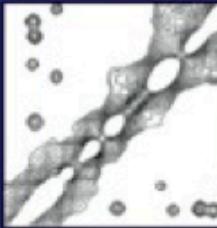
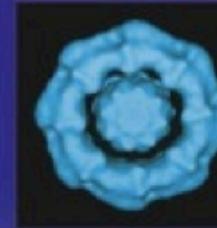
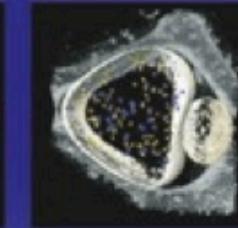
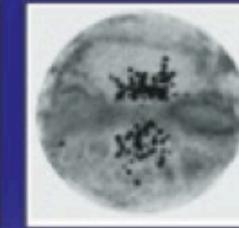
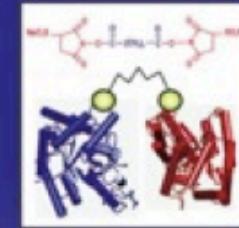
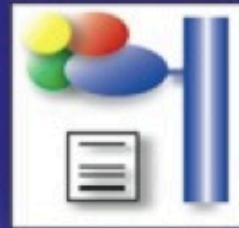
Bio 393
Bioquímica e Ing. Biotecnología

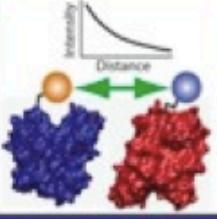
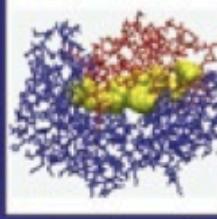
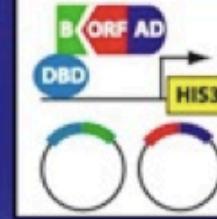
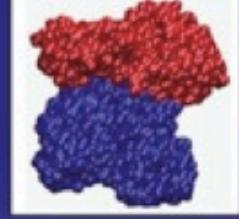
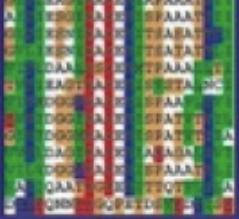
Protein structure prediction

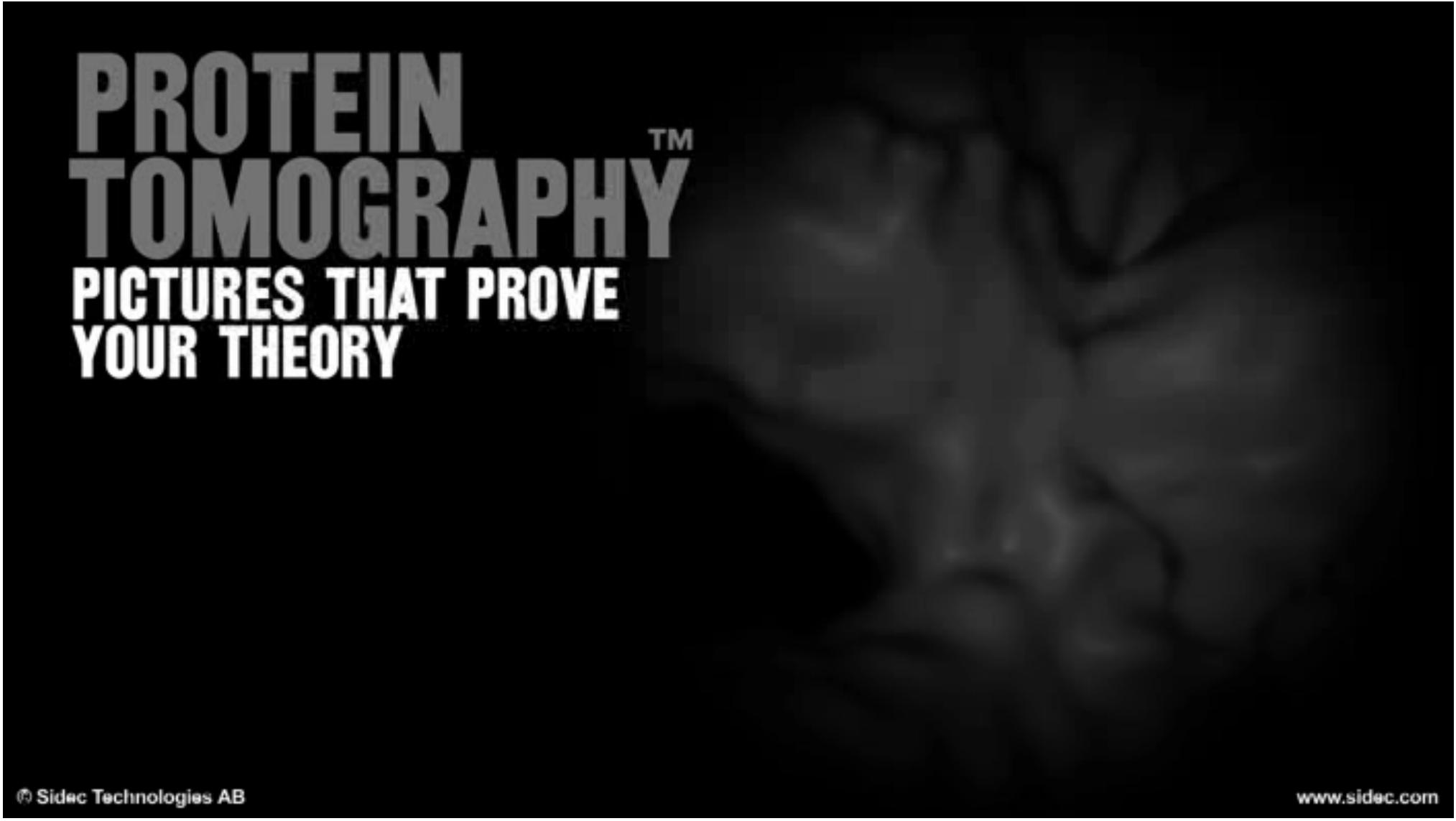
...SDVIFTEDGILICNRK...



Todo es un modelo

						
X-ray crystallography	NMR spectroscopy	2D & single particle electron microscopy	electron tomography	immuno-electron microscopy	chemical cross-linking	affinity purification mass spectroscopy
subunit structure	subunit structure	subunit shape	subunit shape		subunit structure	
subunit shape	subunit shape	subunit-subunit contact	subunit-subunit contact		subunit-subunit contact	subunit-subunit contact
subunit-subunit contact	subunit-subunit contact	subunit proximity	subunit proximity		subunit proximity	subunit proximity
subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity
subunit stoichiometry	subunit stoichiometry					
assembly symmetry	assembly symmetry	assembly symmetry	assembly symmetry	assembly symmetry		
assembly shape	assembly shape	assembly shape	assembly shape			
assembly structure	assembly structure					

						
FRET	site-directed mutagenesis	yeast two-hybrid system	gene/protein arrays	PDB	computational docking	bioinformatics
subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit structure subunit shape	subunit-subunit contact	subunit-subunit contact
subunit proximity		subunit proximity	subunit proximity			



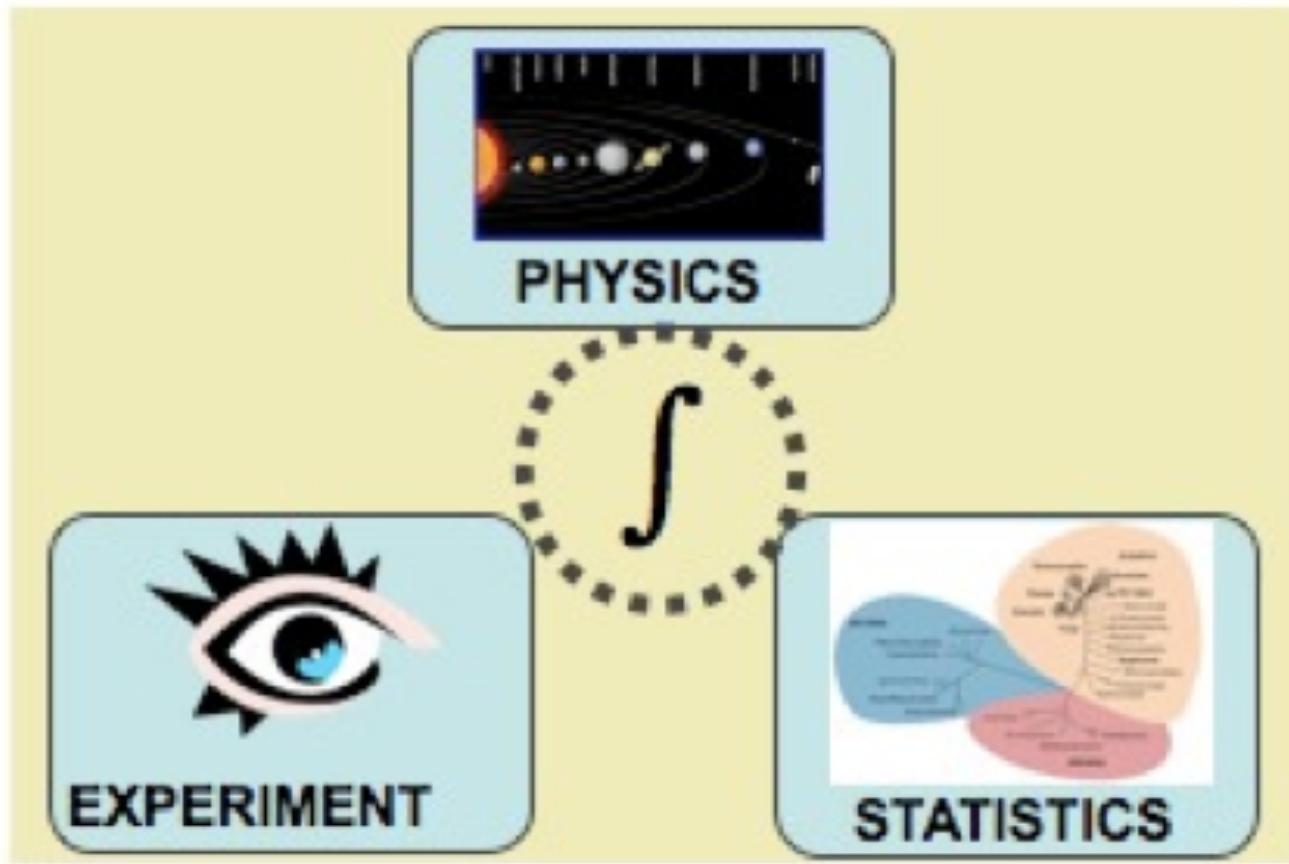
PROTEIN TOMOGRAPHY™

PICTURES THAT PROVE YOUR THEORY

© Sidec Technologies AB

www.sidec.com

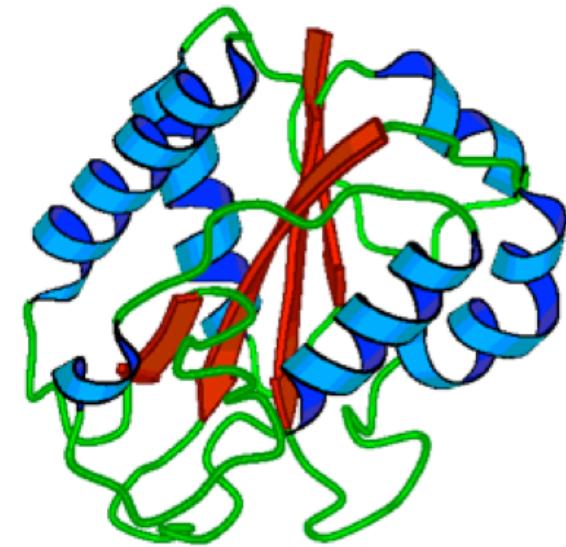
Todo es un modelo



Sali, Earnest, Glaeser, Baumeister. From words to literature in structural proteomics. *Nature* 422, 216-225, 2003.

Protein structure prediction

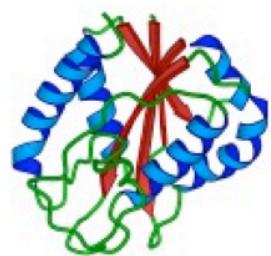
...SDVIFTEDGILICNRK...



- **LIMDPSERTVKMCQYGPSHGLISVMRYDVTNMSI**
F
- **Secondary Structure** • Helix, Strand, Coil
- **Solvent Accessibility** • Exposed, Buried
- **Transmembrane Helix & Topology** •
Membrane, Loops (inside/out)

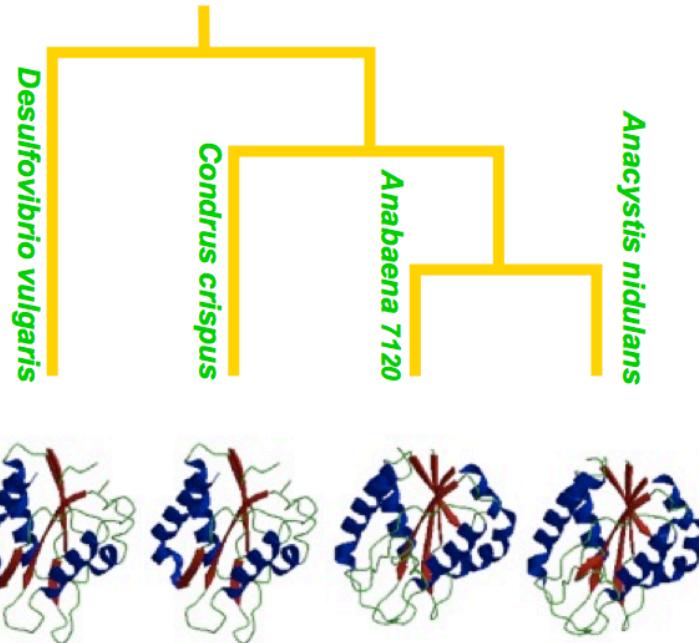
3D

GFCHIKAYTRLIMVG...



Folding
(physics)

Ab initio (de novo) prediction
NEW FOLDS



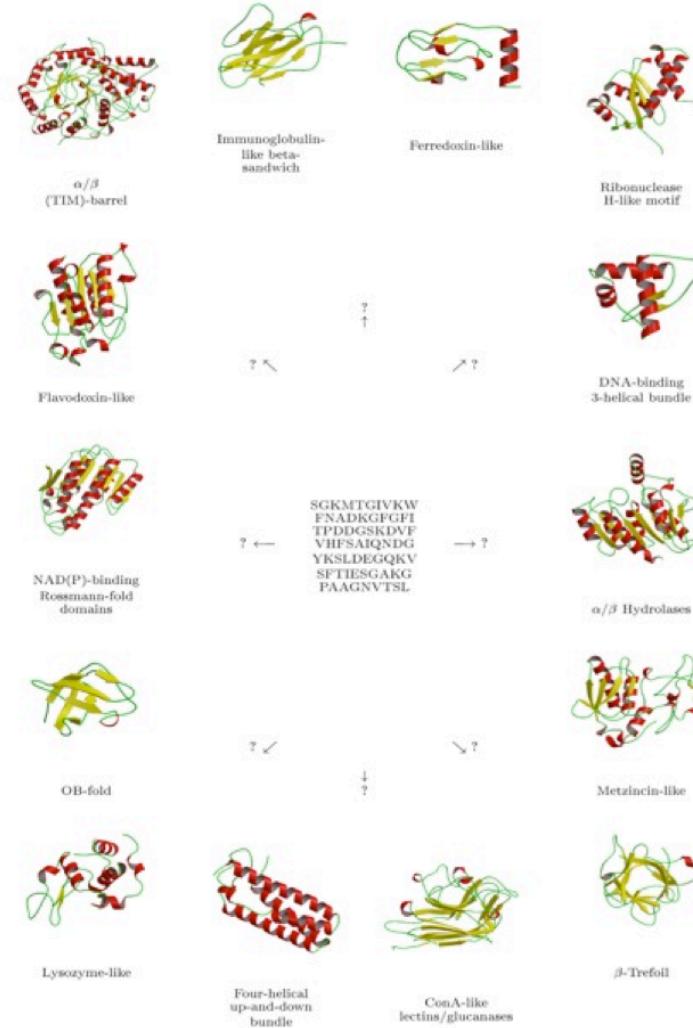
Evolution
(data/statistics)

Fold Recognition
Comparative Modeling
KNOWN FOLDS

Fold Recognition

FR = searching against
the PDB database

- Using Sequence
- Using Structure
- Using Sequence and Structure



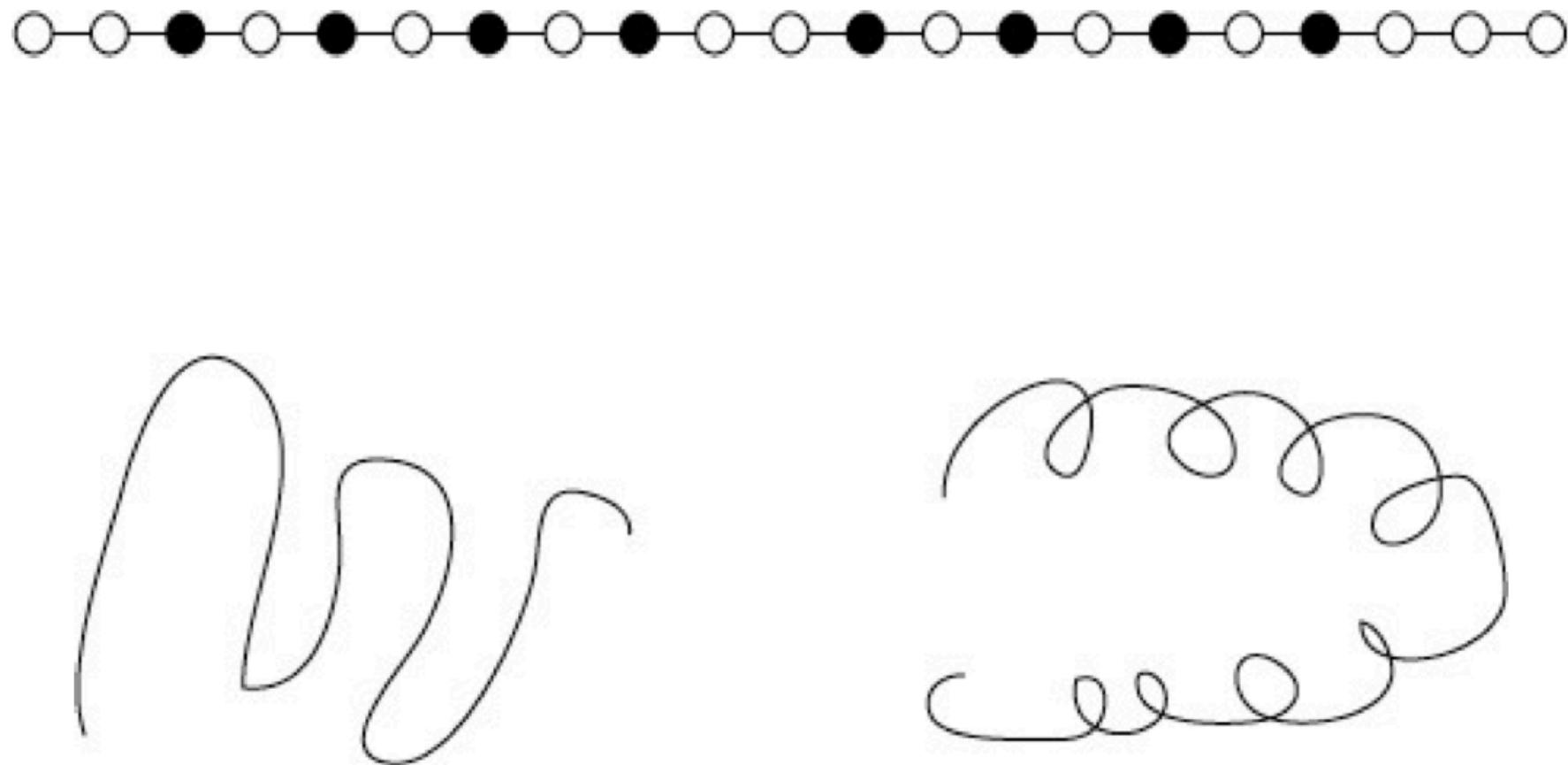
Fold recognition by sequence

- Sequence similarity searches (simple)
 - e.g. BLAST
- Profile Searches
 - PSI-BLAST
 - HMMs

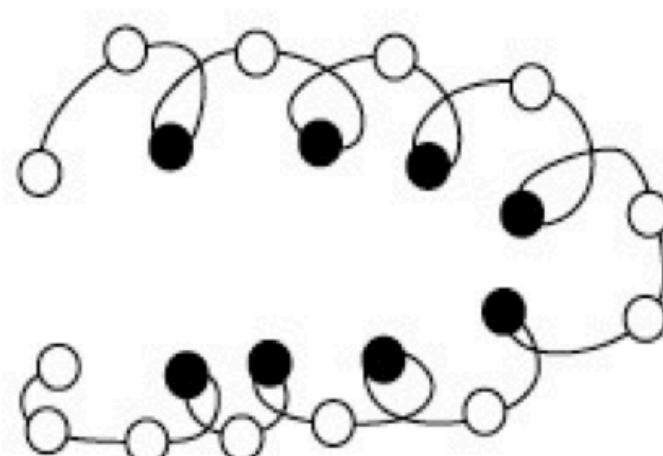
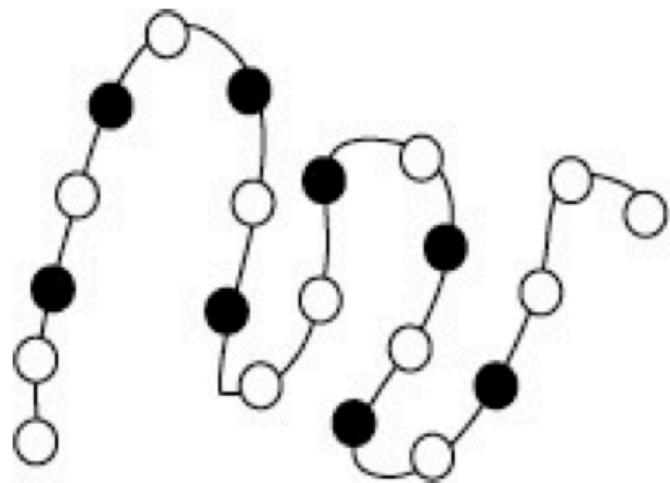
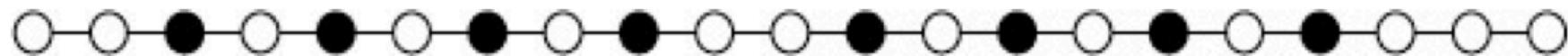
Fold recognition by structure

- **THREADING**
- Use statistical potential to evaluate sequence/structure compatibility.
- Evaluate compatibility of a given sequence with all structures in the PDB database.

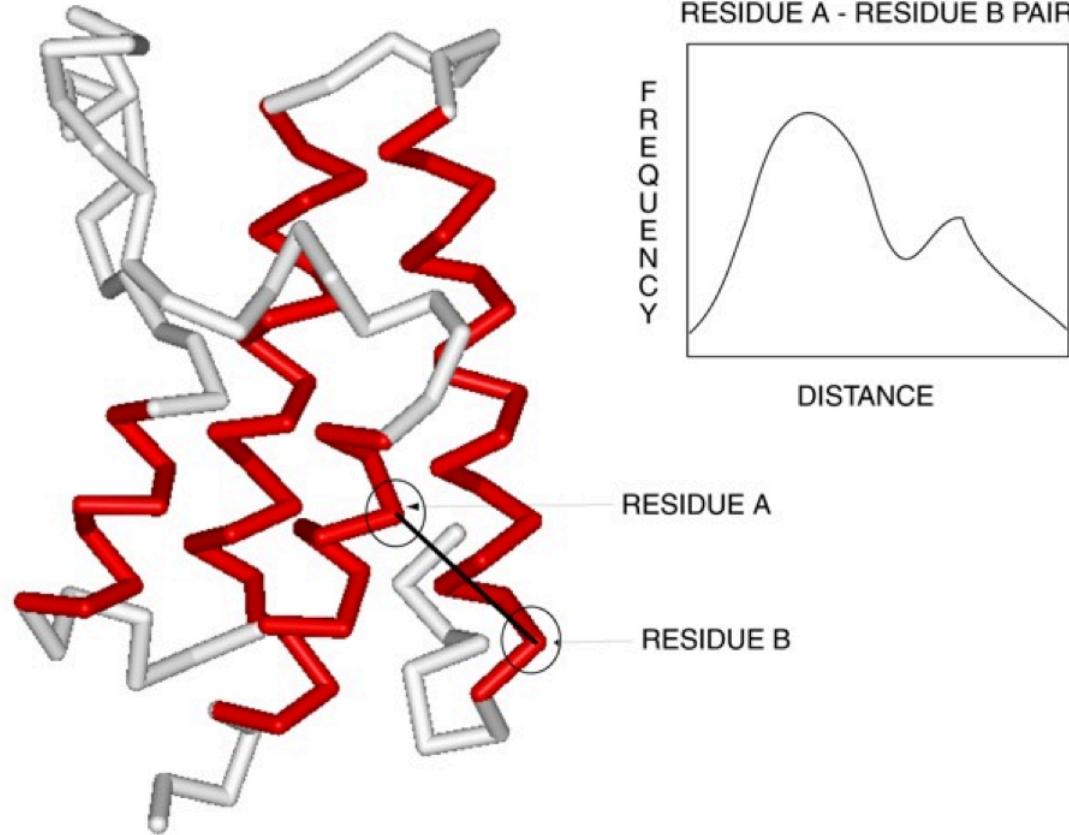
Threading



Threading



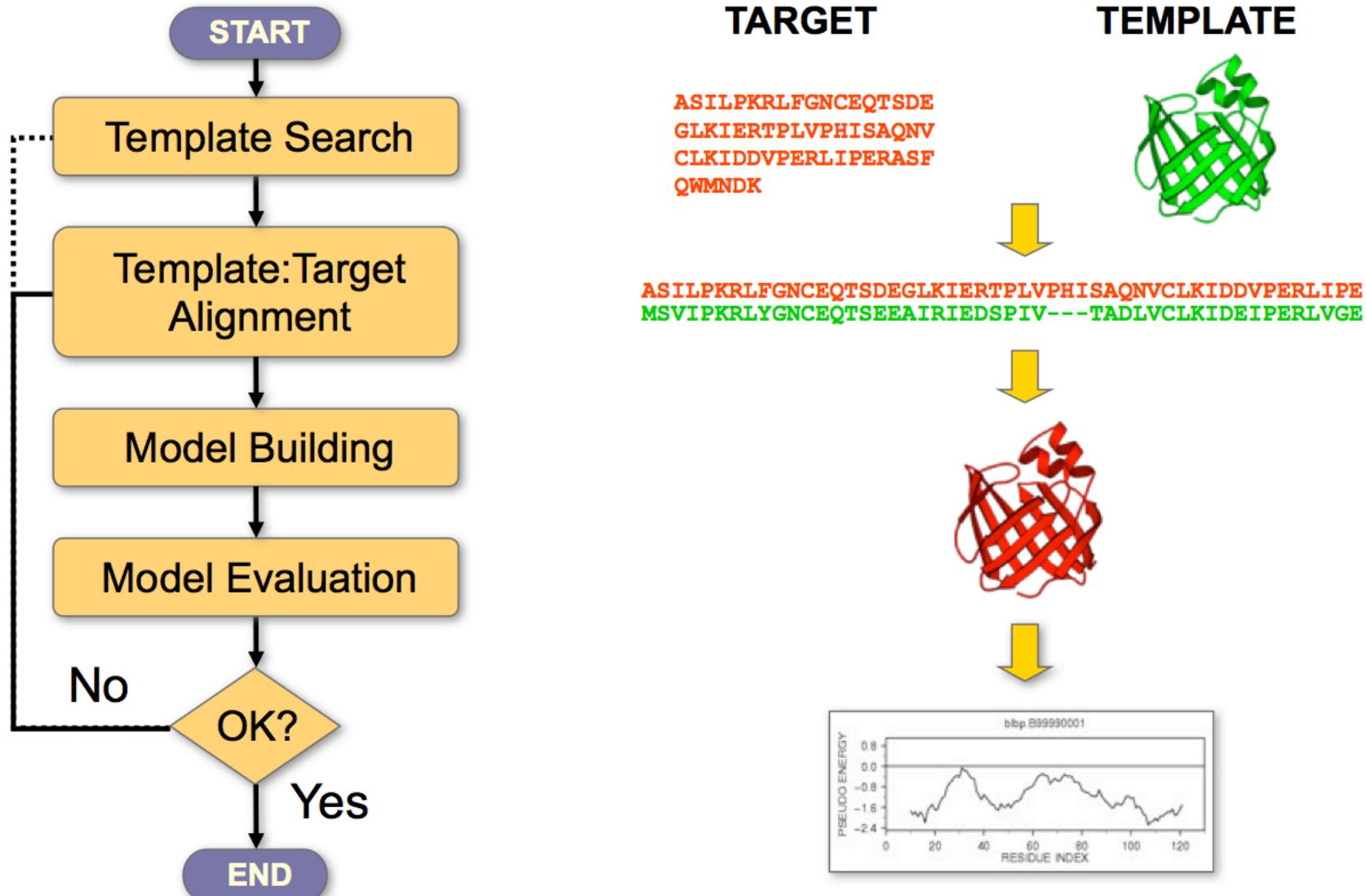
Statistical Potential



SPs represent frequencies of atom-atom distances observed in experimental structures.

**How do we go from fold recognition to
the explicit 3D structure?**

Comparative Modeling

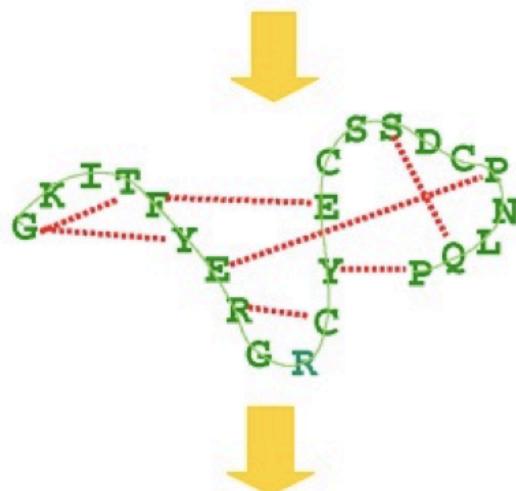


Modeller

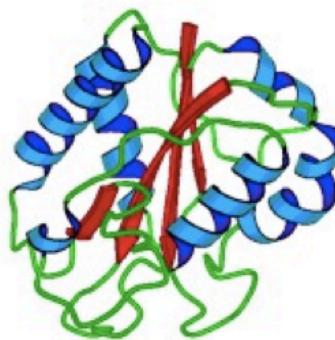
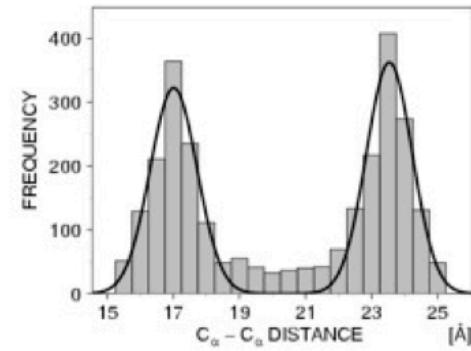
3D GKITFYERGFQGHCYESDC-NLQP...

SEQ GKITFYERG---RCYESDCPNLQP...

1. Extract spatial restraints



2. Satisfy spatial restraints



$$P(\mathbf{R} / \mathbf{I}) = \prod_i p_i (\mathbf{r}_i / \mathbf{l}_i)$$

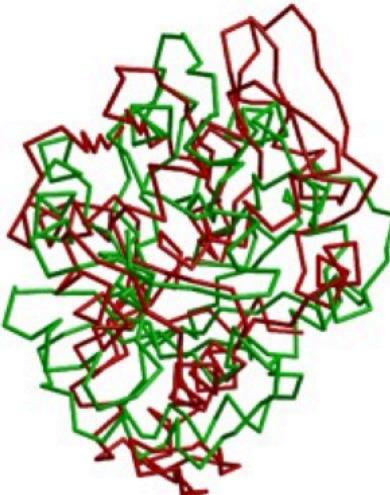
Errors in comparative modeling

MODEL

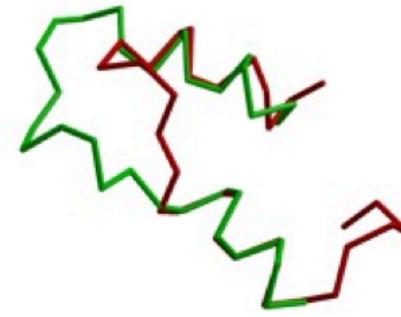
X-RAY

TEMPLATE

Incorrect template

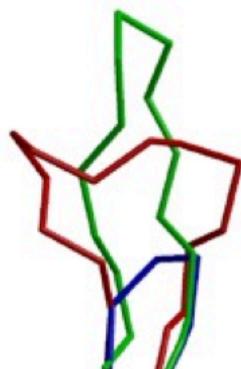


Misalignment

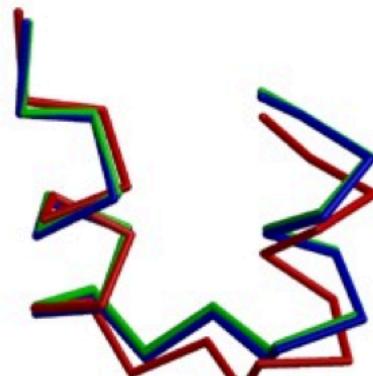


EDB
---EPQFTWAAQPFETQHINHESGQQCTSHANGVHSFQGRCEHQHTPLLTTFAVVNNYCCSSEHNTCPHN
TRSA
ETAAAMFERQKRSQTSIAASNSWTCQHQSSEHLLTSCFPVTPVPHETLADNGVACSGEHH...-HN
.....aaaaaaa
EDB
70 80 90 100 110 120 130
ETEHCNHCNGQVILIHONLTTFPGNISHCIVYAQTPAAMPYIVAVACDSDQGRDFPQYVVVFWLDRII
TRSA
70 80 90 100 110 120 130
GQTCYQSYSTMSTIDCRETSQ...FYPICATEETRQANHIIIVACEGH...-PTVPIVIGTADY
bbbbb bbbbbb aaaaaaaaaaaaaaaaaaaaaaa bbbbbb

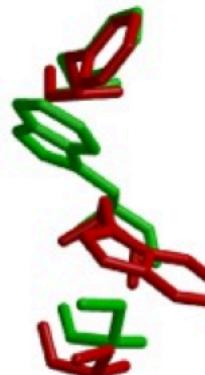
Region without a template



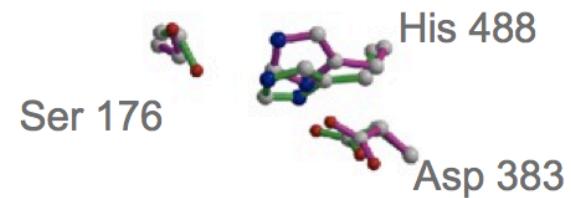
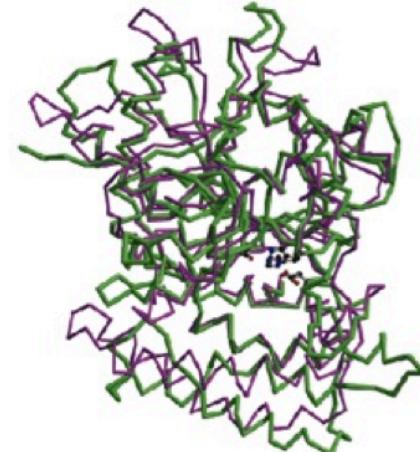
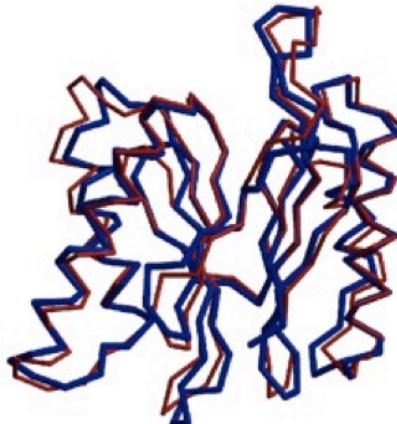
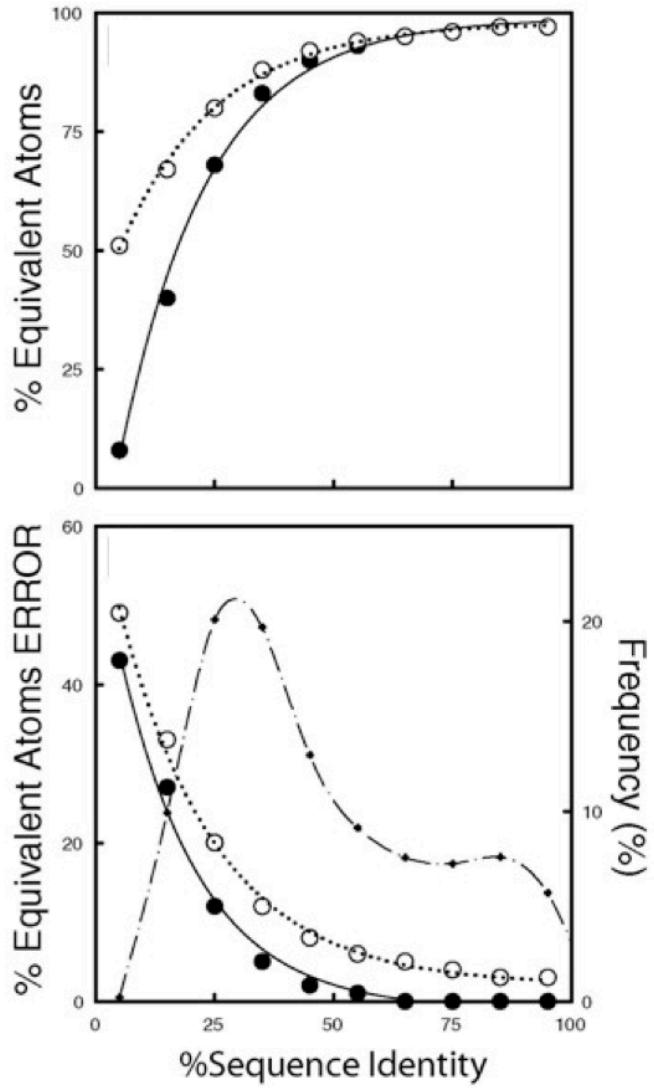
Distortion in correctly aligned regions



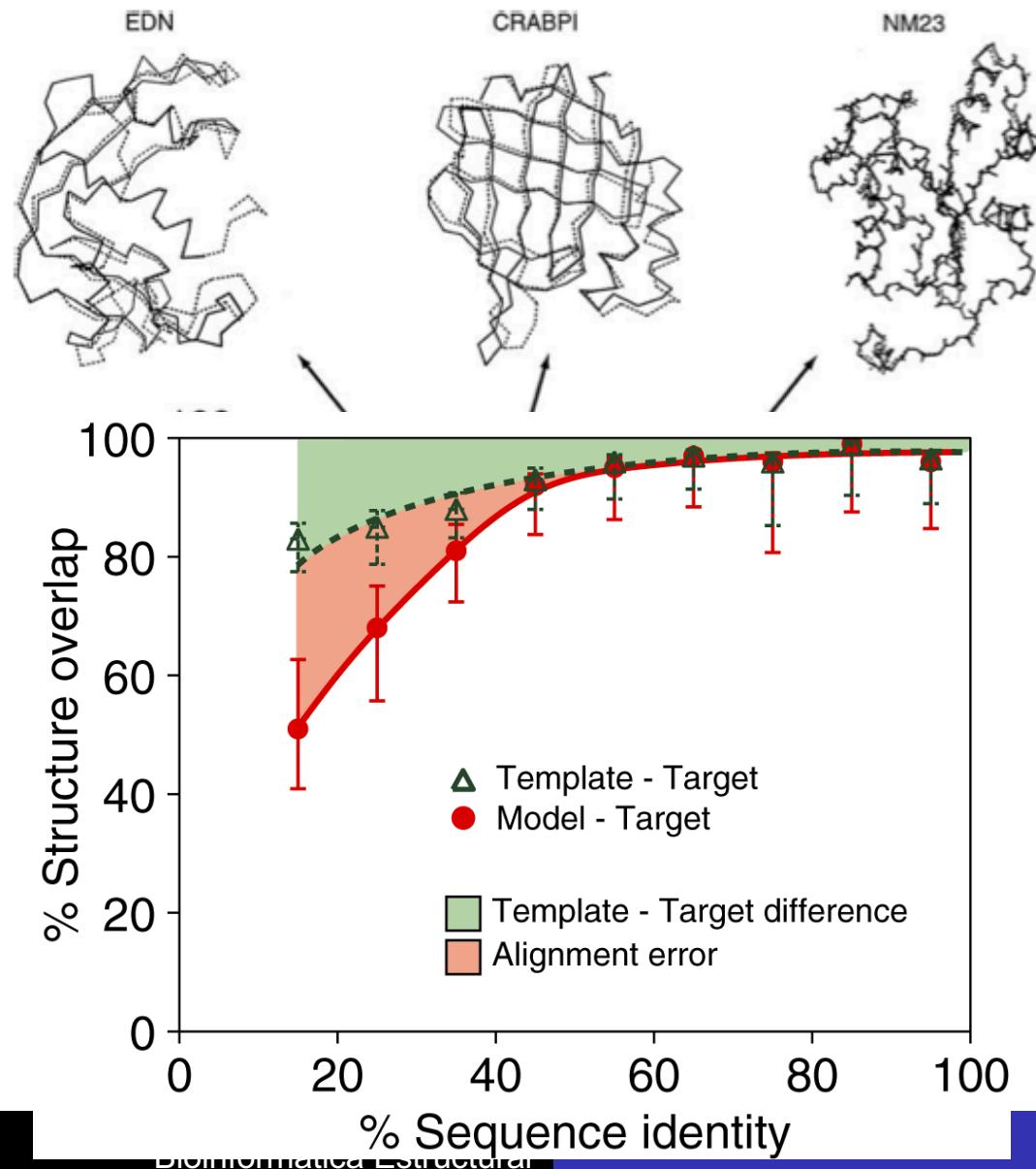
Sidechain packing



Comparative modeling accuracy



- Target-Template Difference
- Alignment Accuracy

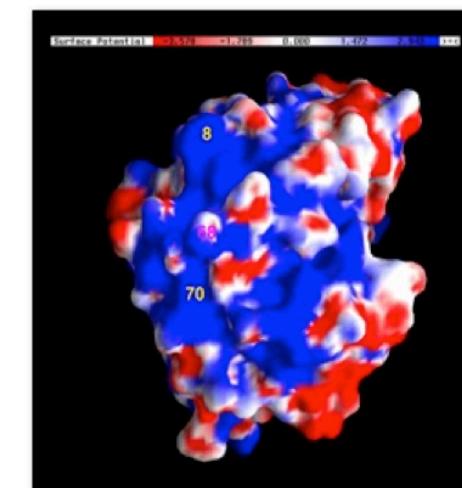
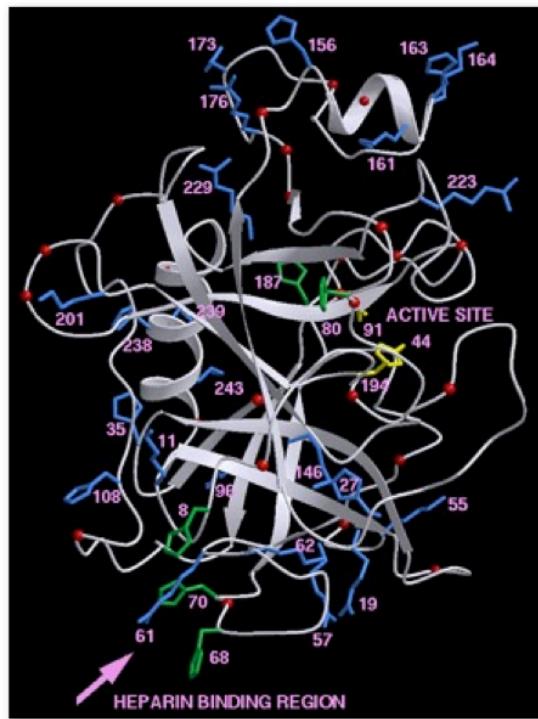
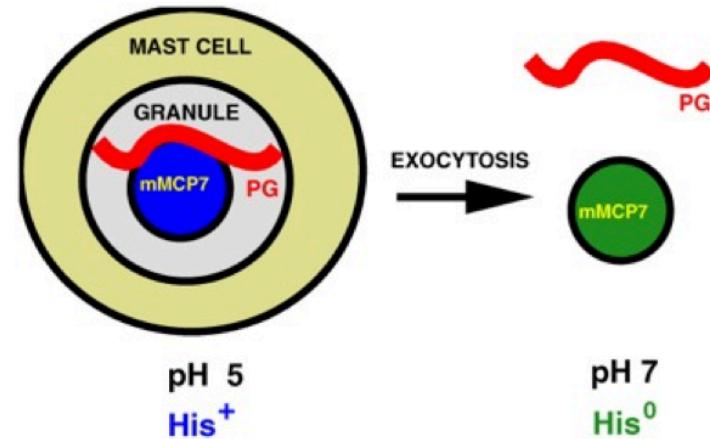


Aren't models just copies of the template? What do we gain?

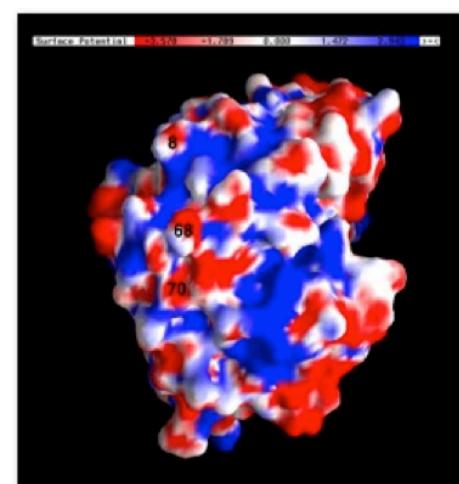
Predicting features of a model that are not present in the template

Do mast cell proteases bind proteoglycans? Where? When?

1. mMCps bind negatively charged proteoglycans.
2. Comparative models used to find clusters of positively charged surface residues.
3. Tested by site-directed mutagenesis.



Native m MCP-7 at pH=5 (His⁺)



Native m MCP-7 at pH=7 (His⁰)

Huang et al. *J. Clin. Immunol.* **18**, 169, 1998.
Matsumoto et al. *J. Biol. Chem.* **270**, 19524, 1995.
Šali et al. *J. Biol. Chem.* **268**, 9023, 1993.

how accurate are comparative
models of protein structure?

Sequence similarity vs structure similarity

The EMBO Journal vol.5 no.4 pp.823–826, 1986

The relation between the divergence of sequence and structure in proteins

Cyrus Chothia¹ and Arthur M.Lesk²

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH,
and ¹Christopher Ingold Laboratory, University College London, 20 Gordon
Street, London WC1H 0AJ, UK

²Permanent address: Fairleigh Dickinson University, Teaneck-Hackensack
Campus, Teaneck, NJ 07666, USA

Communicated by M.F.Perutz

Homologous proteins have regions which retain the same general fold and regions where the folds differ. For pairs of distantly related proteins (residue identity ~ 20%), the regions with the same fold may comprise less than half of each molecule. The regions with the same general fold differ in structure by amounts that increase as the amino acid sequences diverge. The root mean square deviation in the positions of the main chain atoms, Δ , is related to the fraction of mutated residues, H , by the expression: $\Delta(\text{\AA}) = 0.40 e^{1.87H}$.

Key words: evolution/protein homology/model building

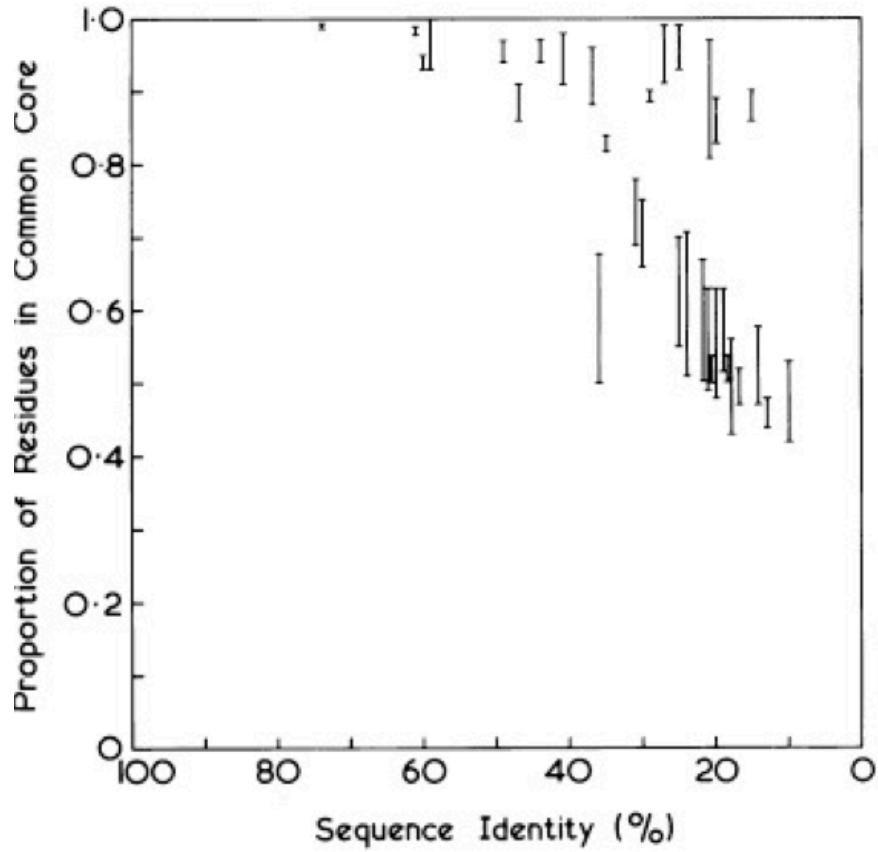


Fig. 1. Size of common cores as a function of protein homology. If two proteins of length n_1 and n_2 have c residues in the common core, the fractions of each sequence in the common core are c/n_1 and c/n_2 . We plot these values, connected by a bar, against the residue identity of the core (see Table II).

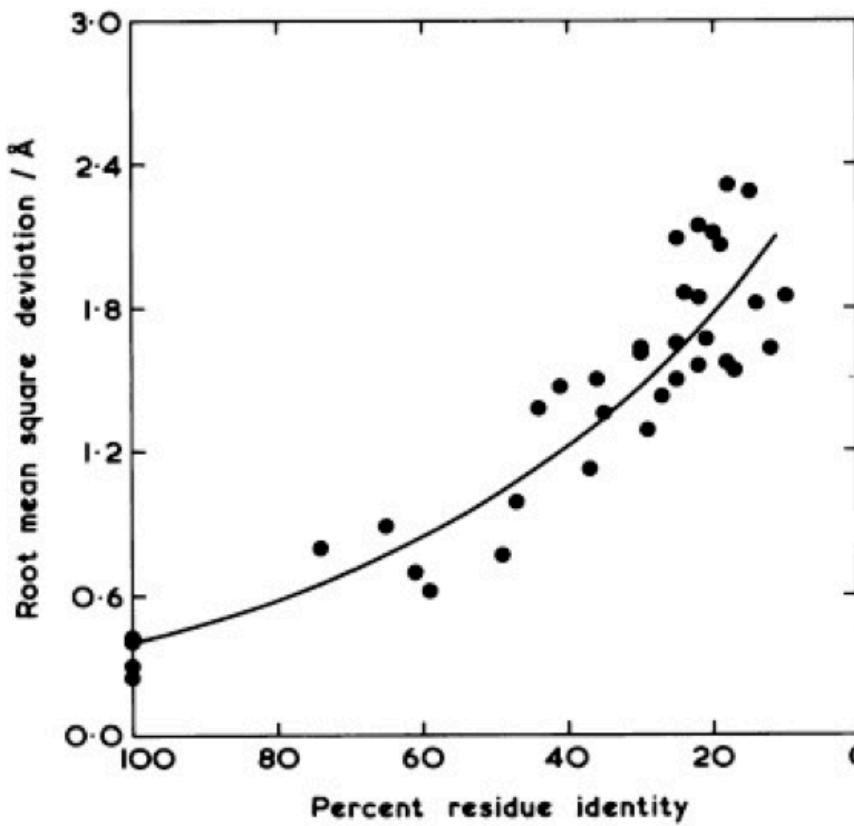
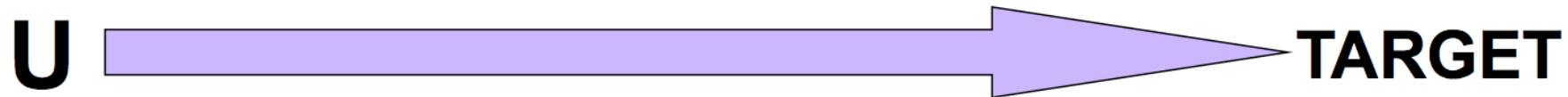
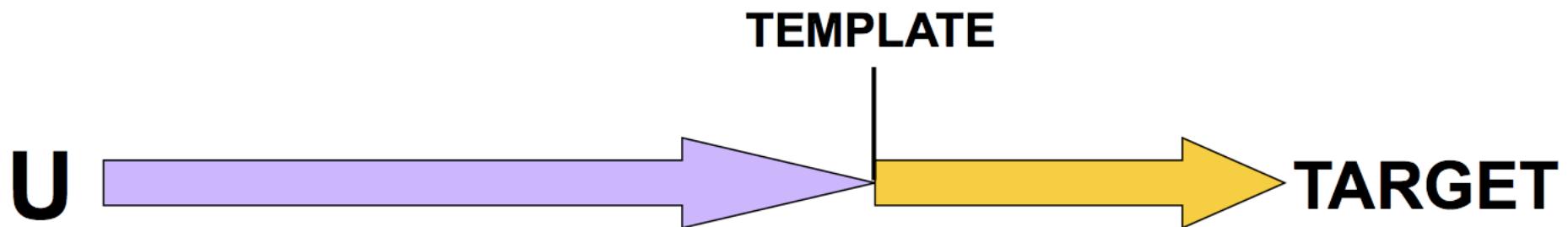


Fig. 2. The relation of residue identity and the r.m.s. deviation of the backbone atoms of the common cores of 32 pairs of homologous proteins (see Table II).

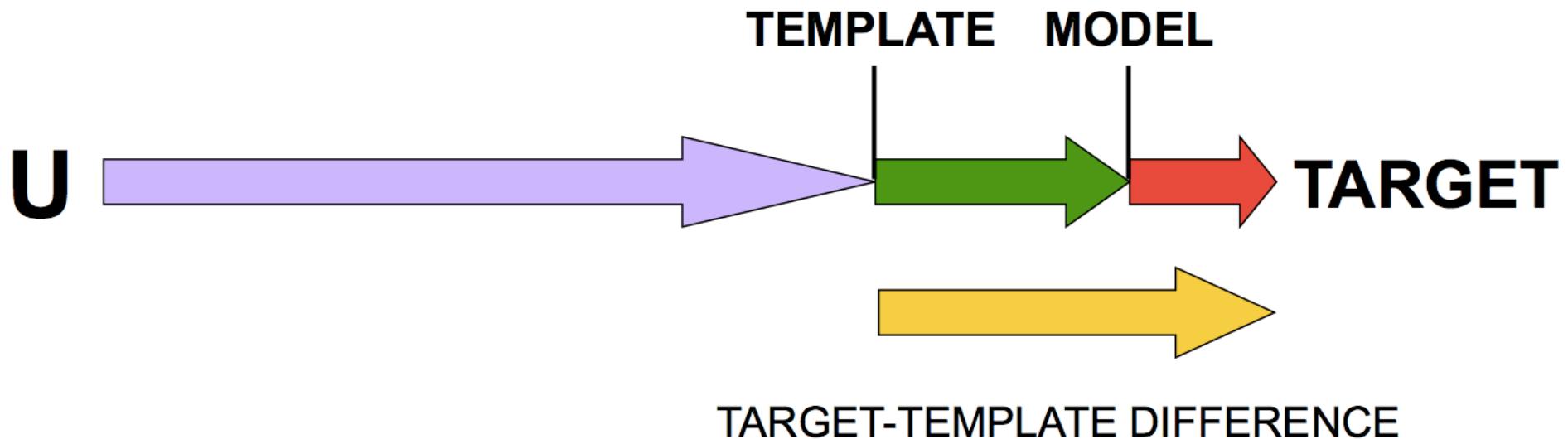
- **Accuracy** : How close is a model to the experimental structure of the target.

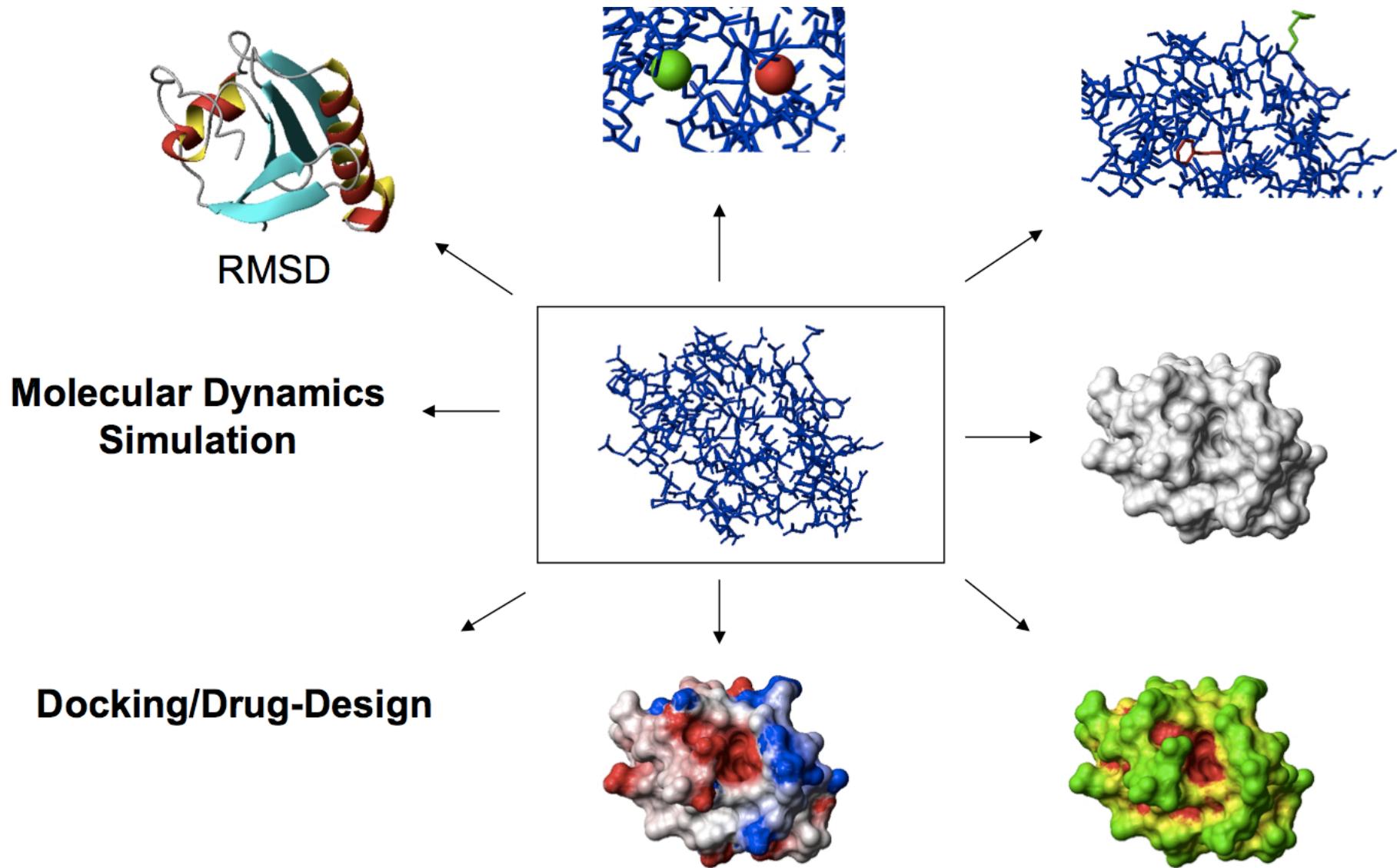


- **Accuracy** : How close is a model to the experimental structure of the target.



- **Accuracy** : How close is a model to the experimental structure of the target.
- **Added-value**: How much closer to the target is the model vs. its template.





Measurements

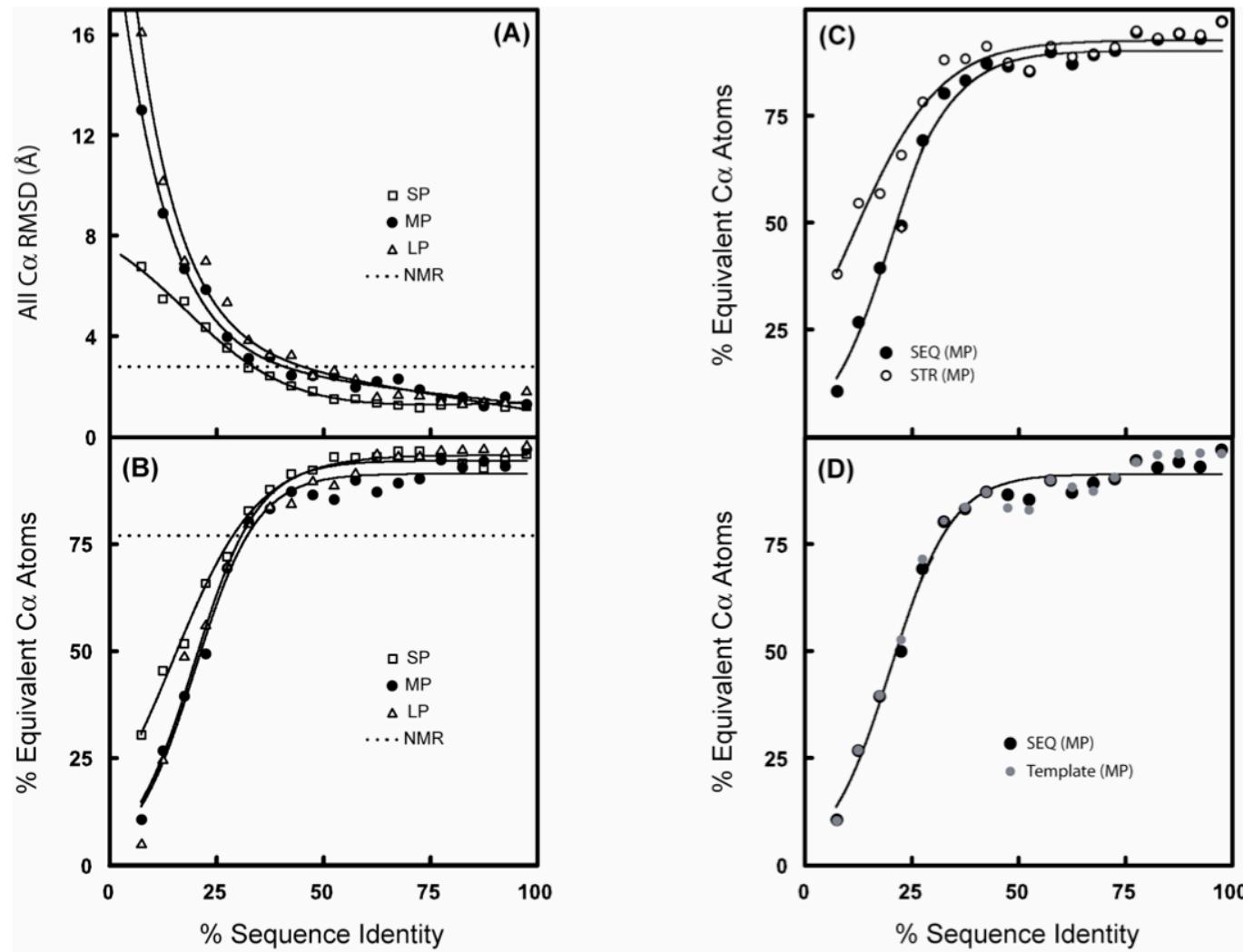
Structure-derived properties

- “Overall accuracy” : RMSD, % Equivalent C α
- Inter-residue distance
- Residue neighborhood
- Residue exposure state
- Accessible surface area
- Surface pockets
- Pocket composition
- Salt-bridges
- Etc ...

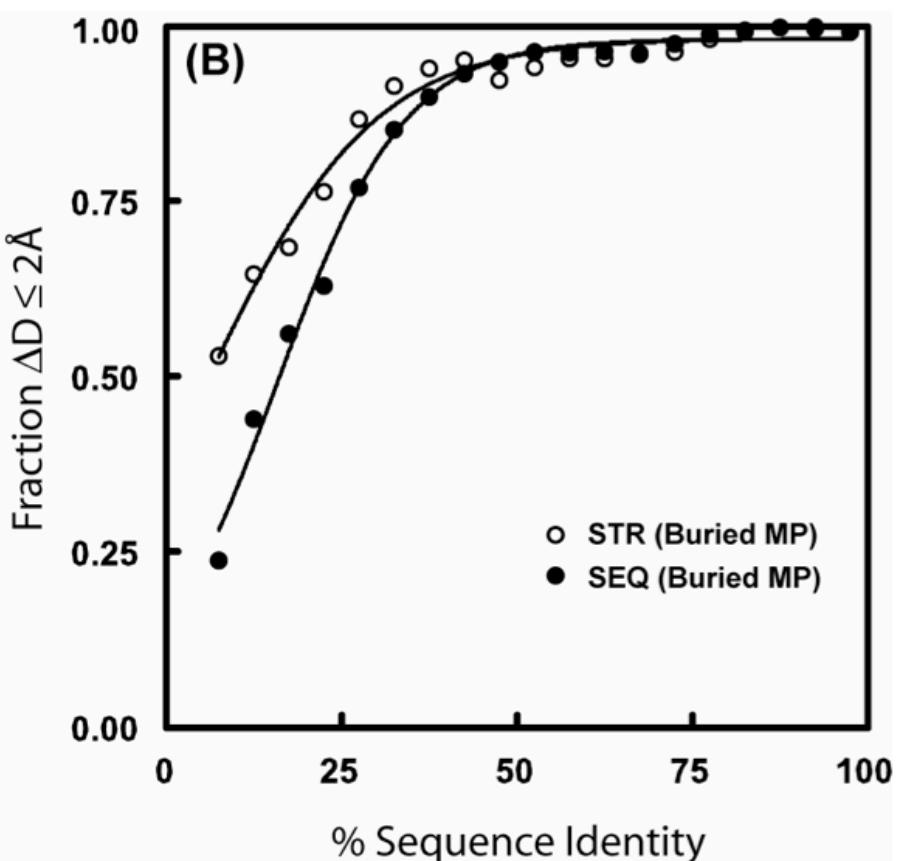
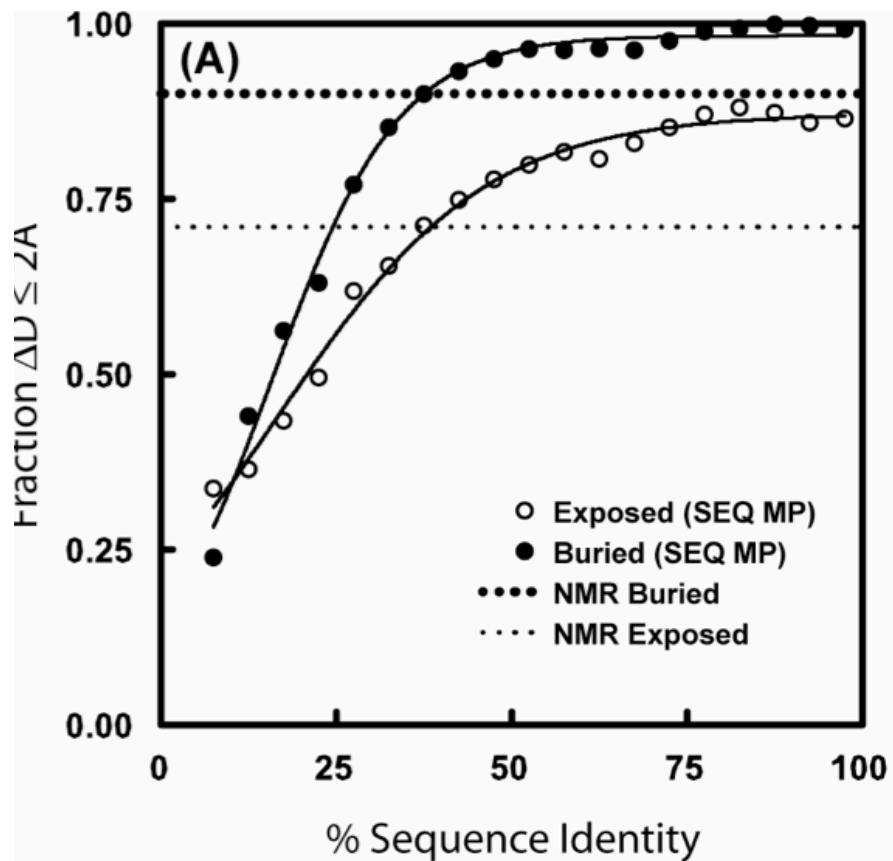
Comparisons

1. MODEL vs. TARGET : model accuracy
2. TEMPLATE vs. TARGET: target-template difference
(use modeling alignment as equivalence guide)
3. 2 – 1 = added-value
4. SEQ-MODEL vs. STR-MODELS: alignment effect
5. NMR vs. X-RAY: experimental variation

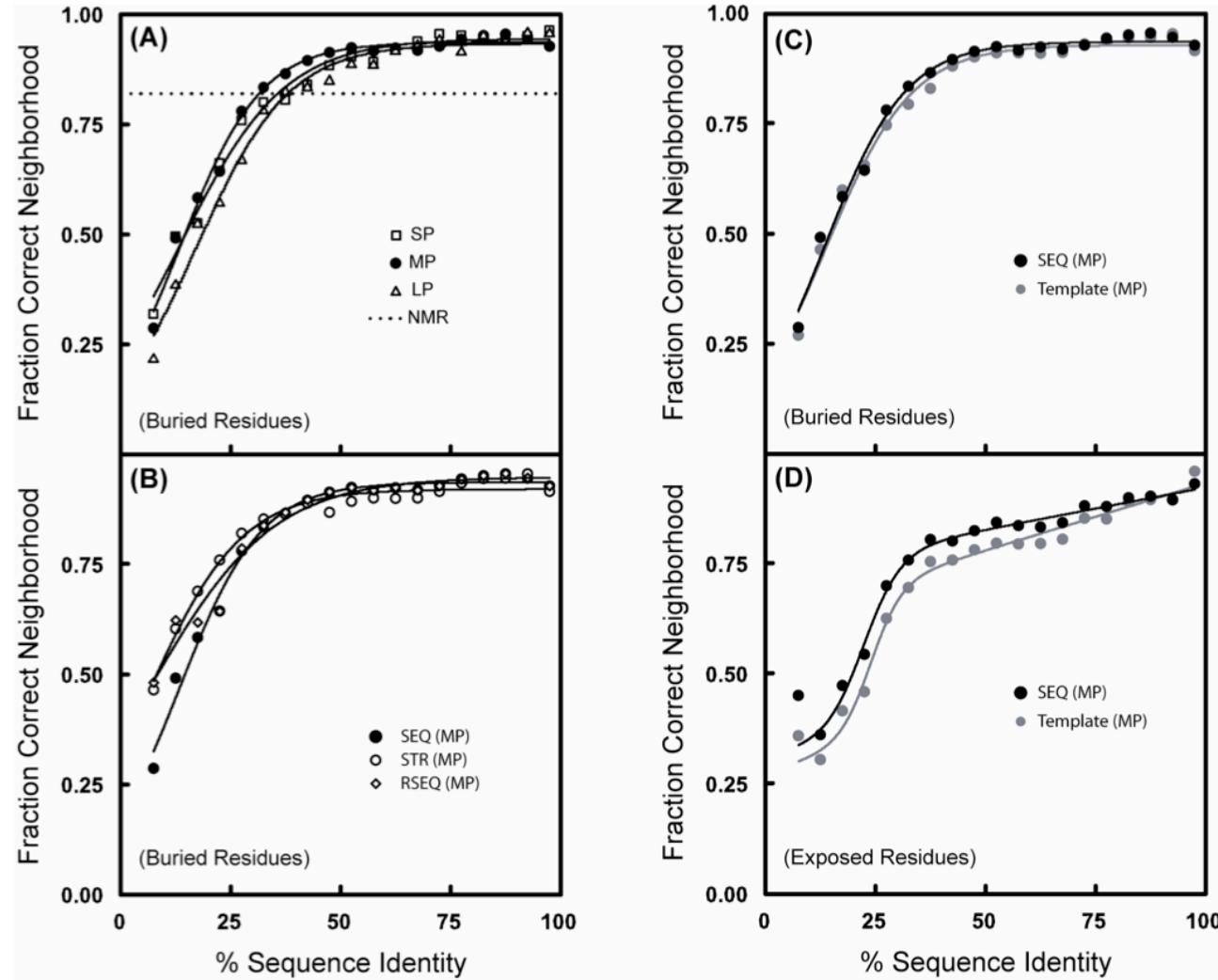
Overall Accuracy



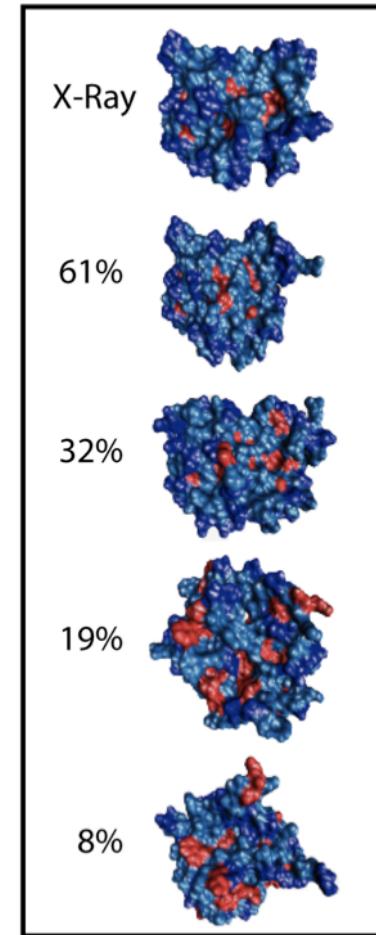
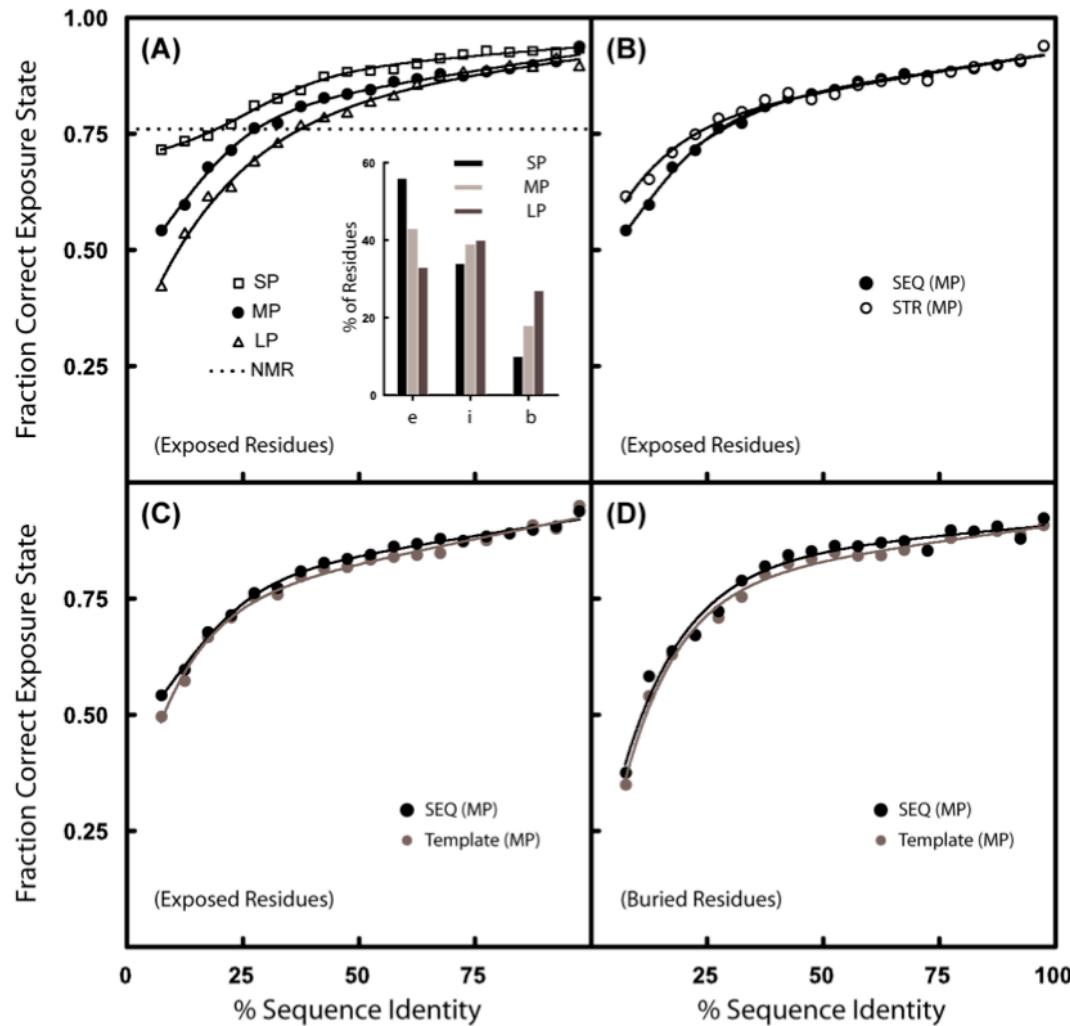
Inter-residue Distance



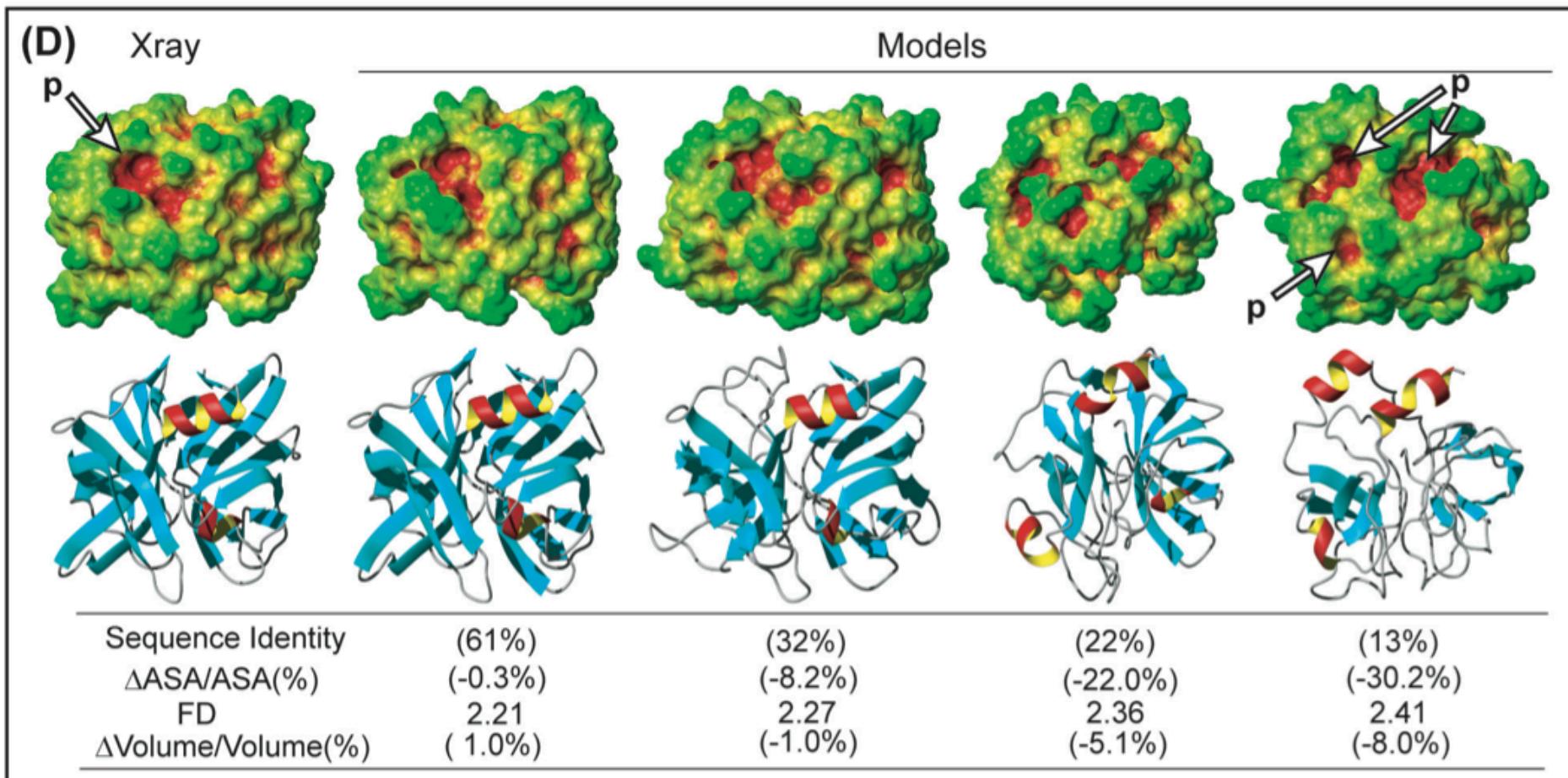
Residue Neighborhood



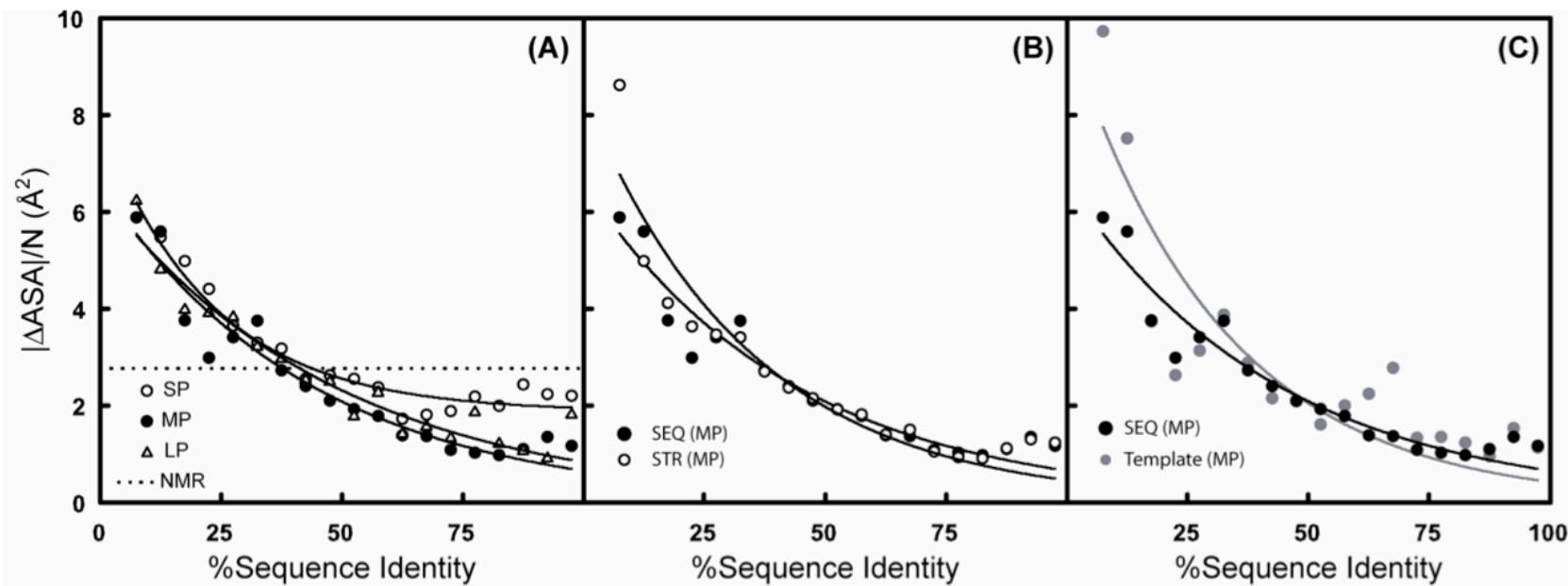
Residue exposure state.



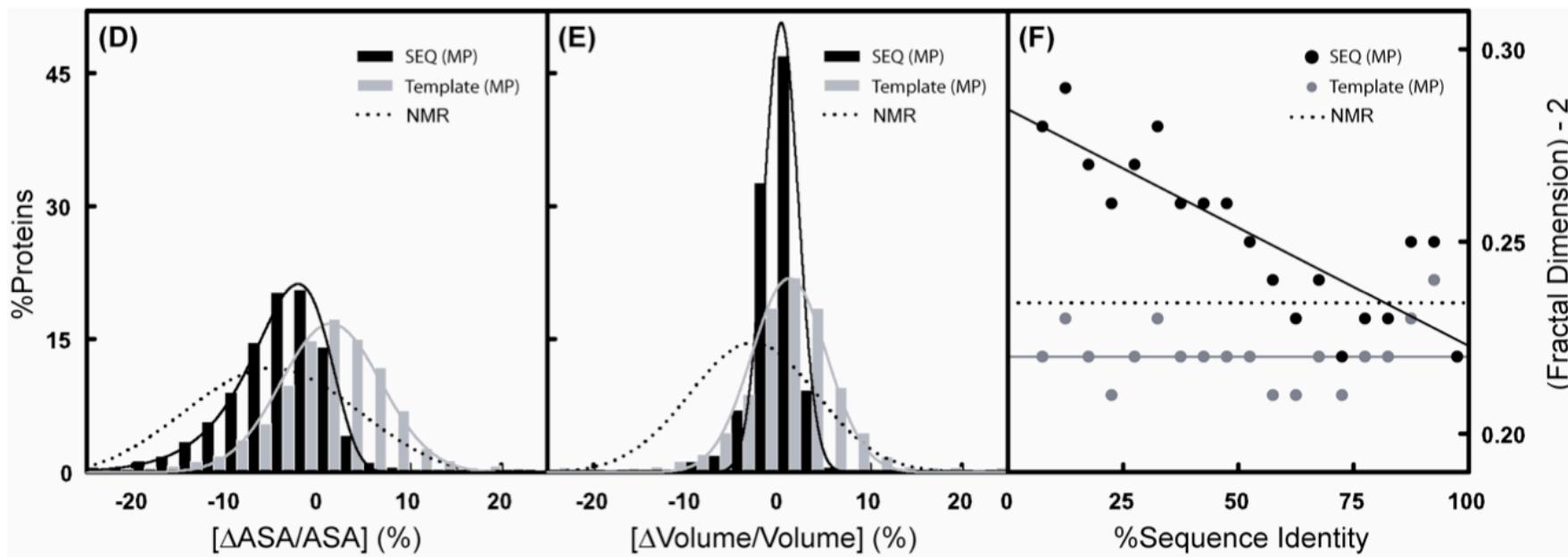
SASA and Volume



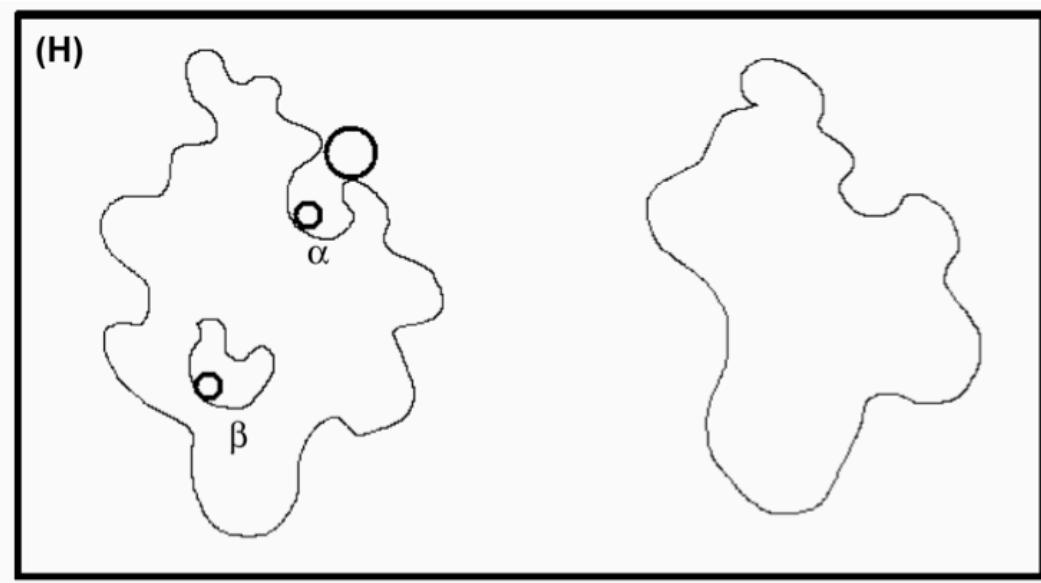
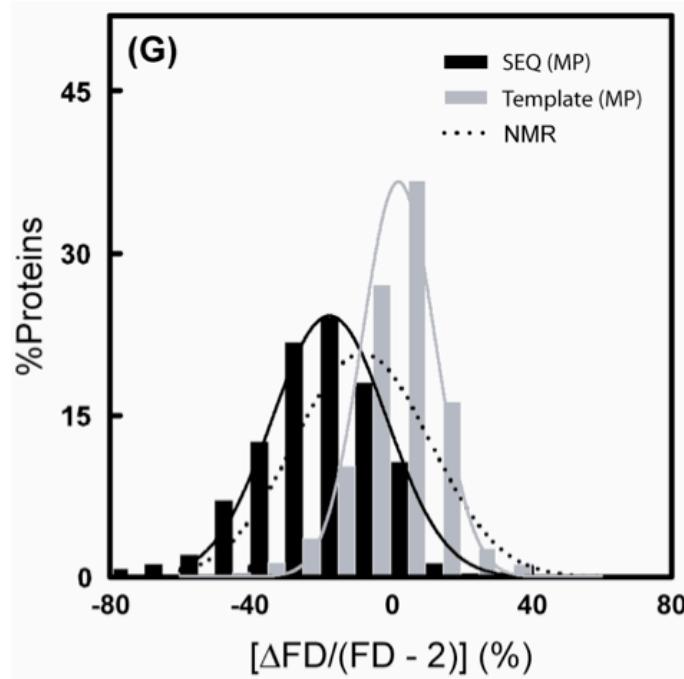
SASA



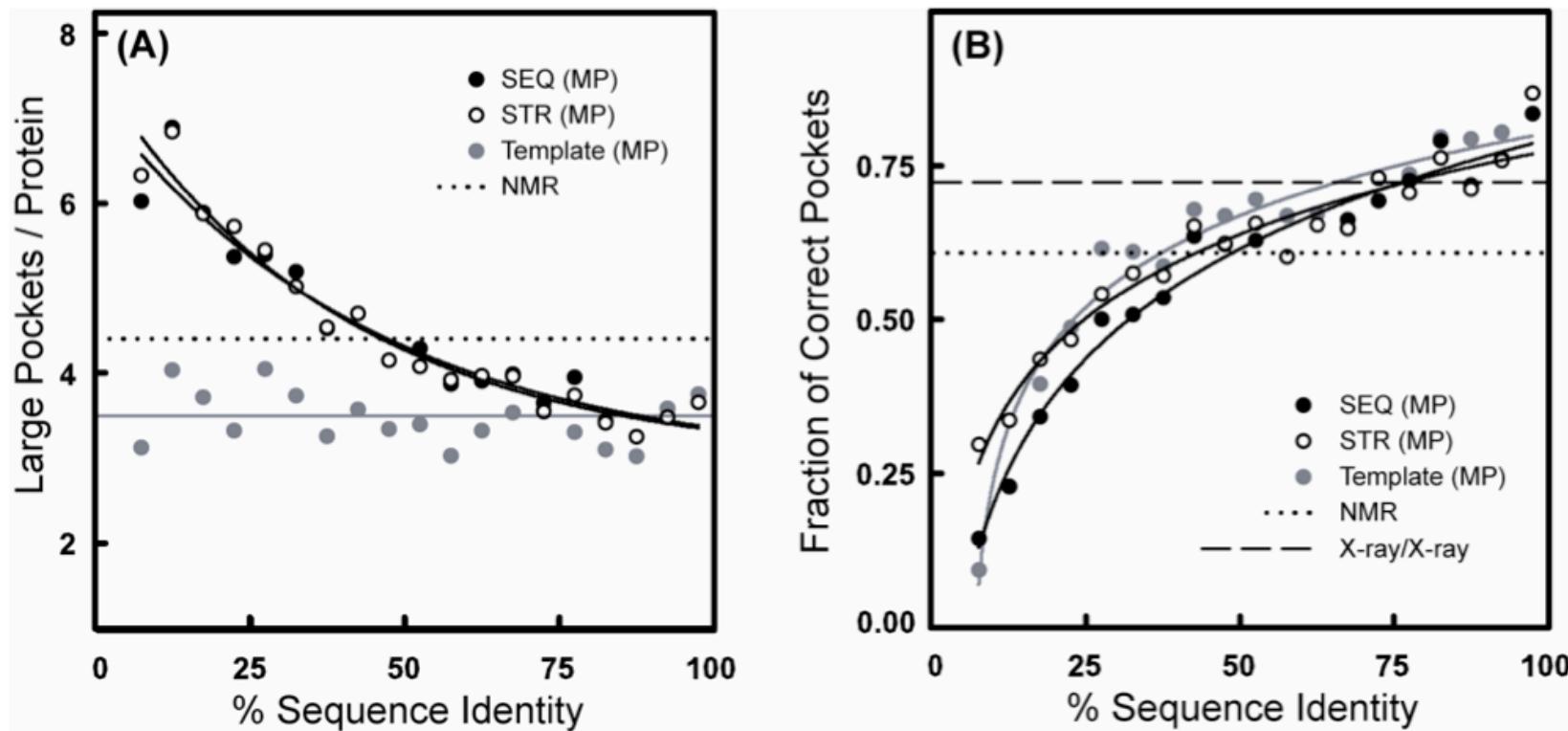
SASA



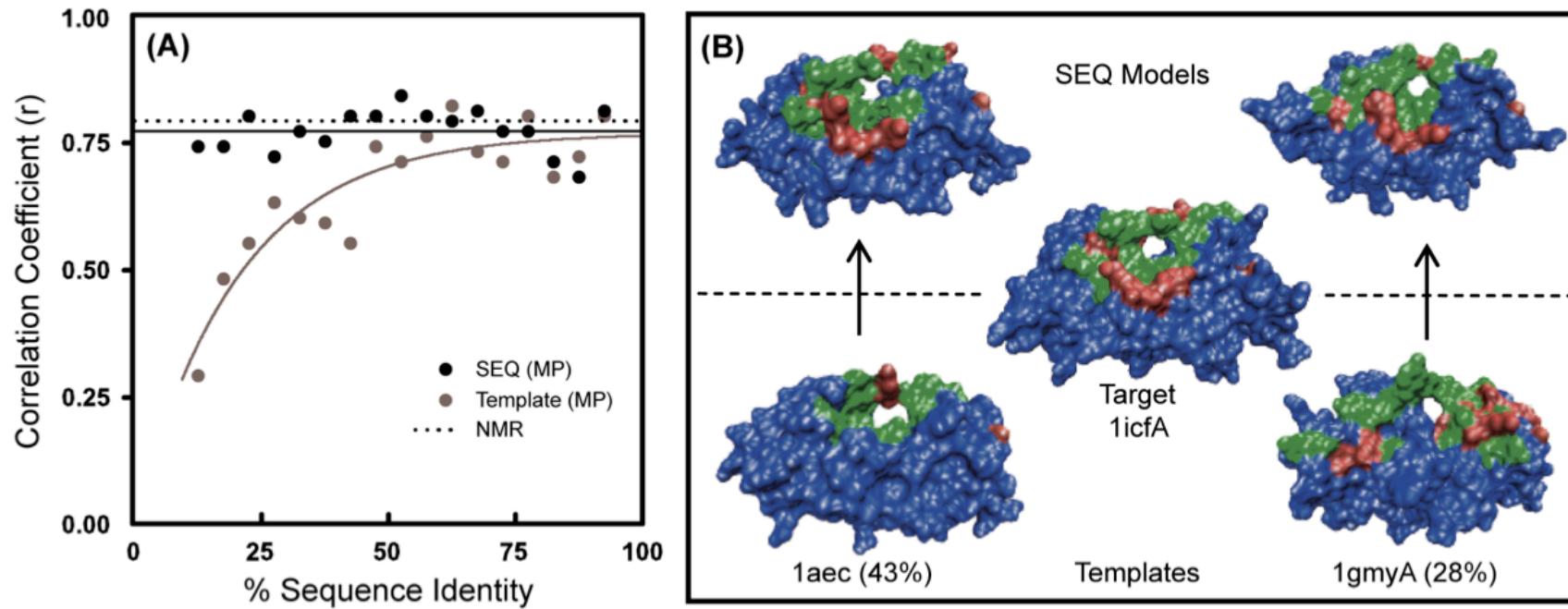
SASA



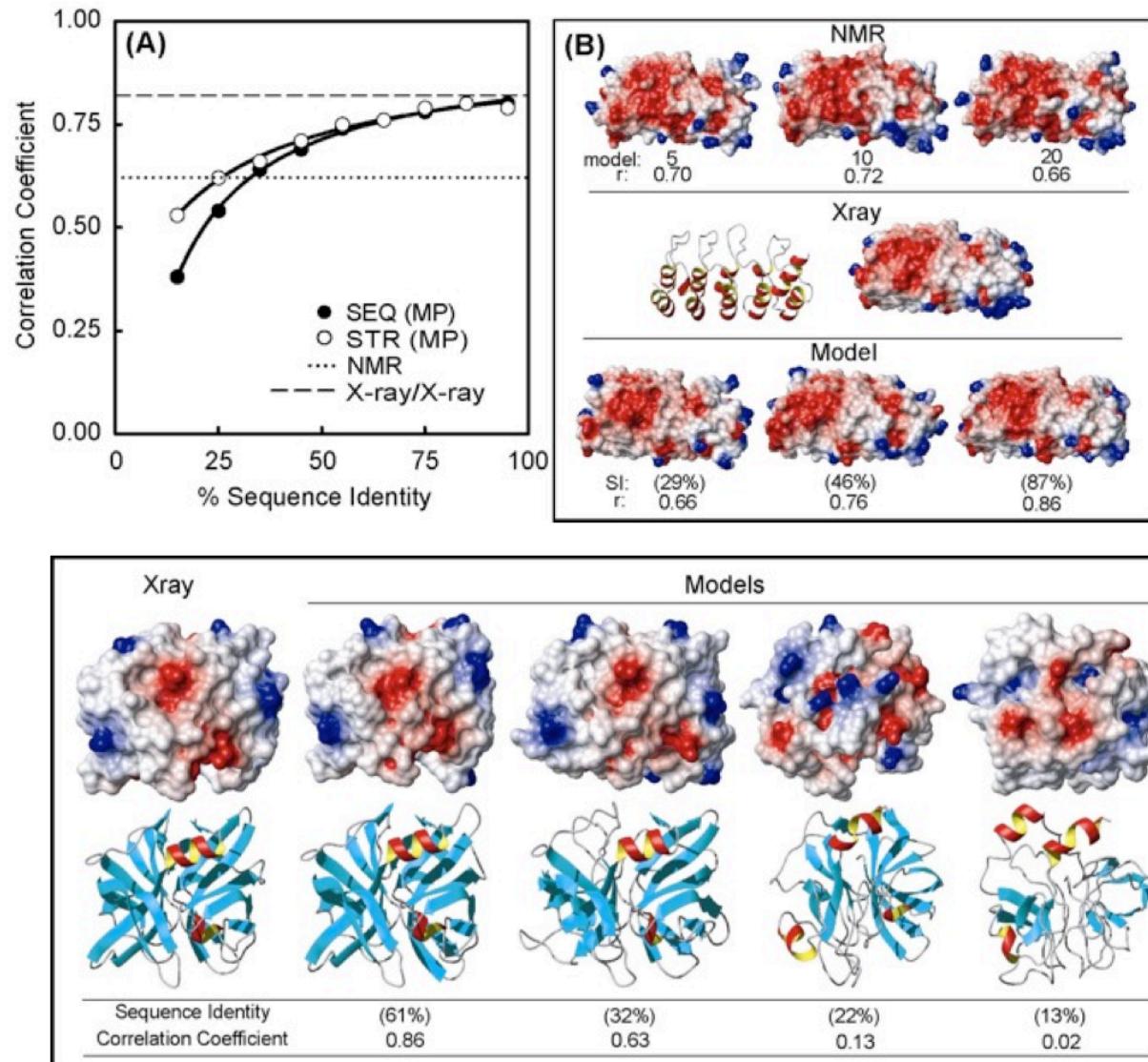
Pockets



Pocket Comparison



Electrostatic potential accuracy



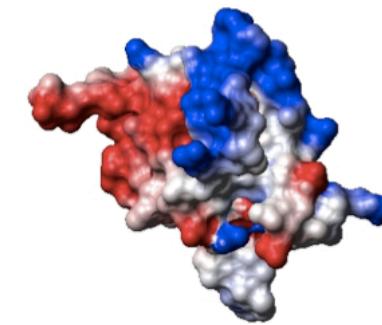
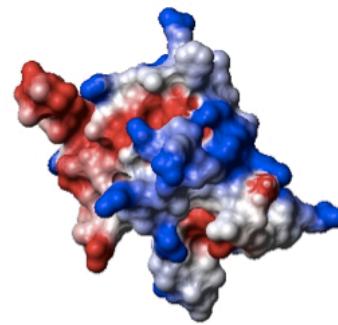
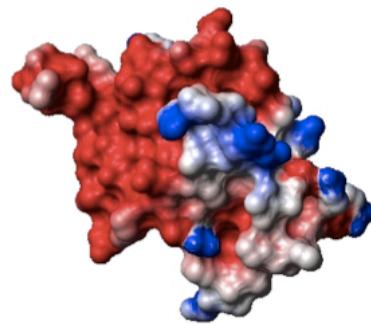
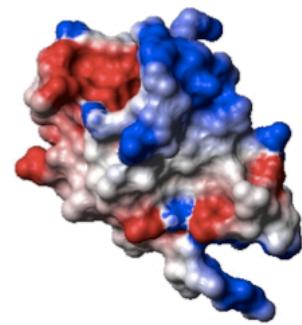
TARGET

TEMPLATE

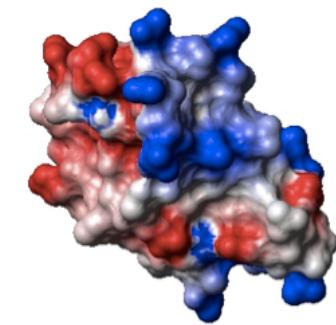
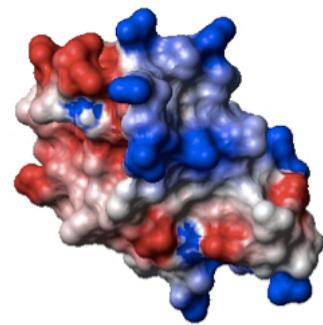
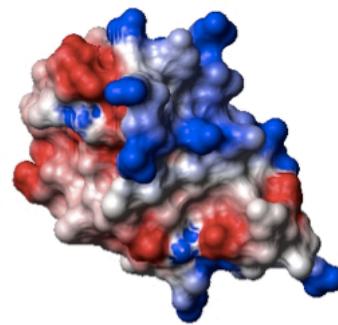
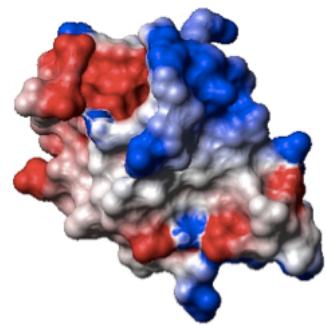
SEQ-MODEL

STR-MODEL

20%



89%



Conclusions

- Not all properties behave the same (e.g. added-value, effect of alignment errors).
- Accuracy increases with target-template sequence similarity.
- Added-value decreases with target-template sequence similarity.
- Surface of models is rugged resulting in increased ASA and artificial pockets.
- Error in models based on 35% target-template sequence identity (or more) is of similar magnitude to NMR:X-RAY differences.

Conclusions

- Properties that depend only on geometry (e.g. exposure state) do not show much added-value.
- Properties that depend on the size and chemistry of residues show more added-value (e.g. neighborhood, composition, EP).
- Added-value comes from combining the right sequence with the right template structure.
- Alignment accuracy does have an impact on added-value for “chemistry dependent” properties.

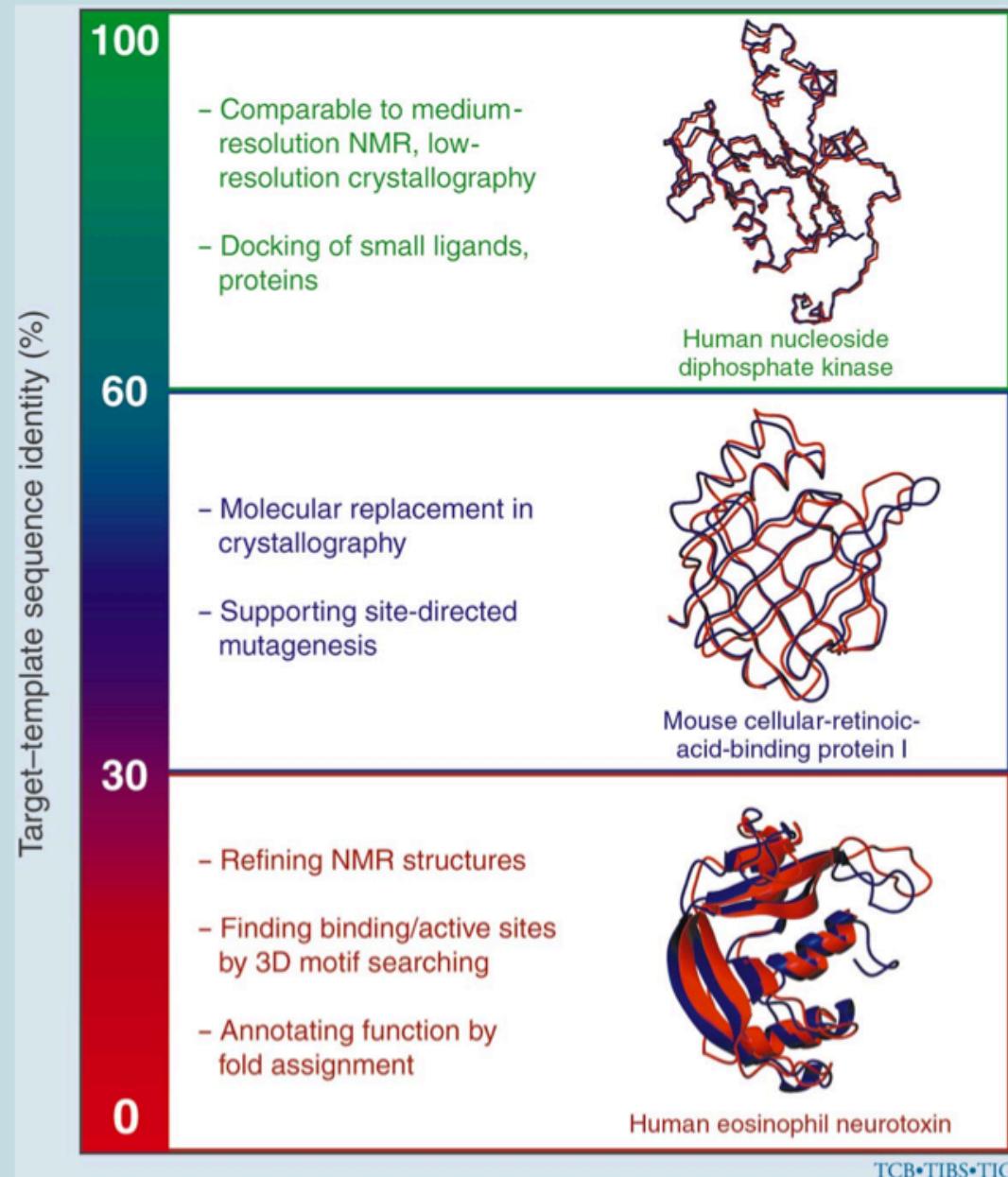
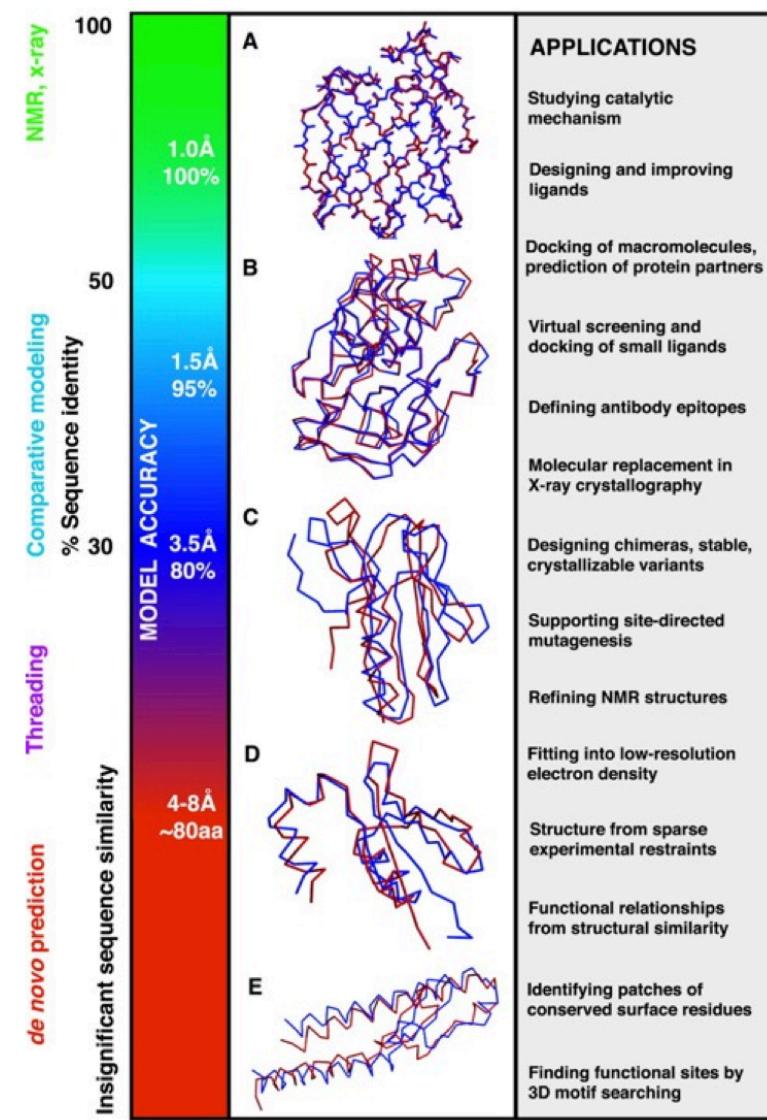
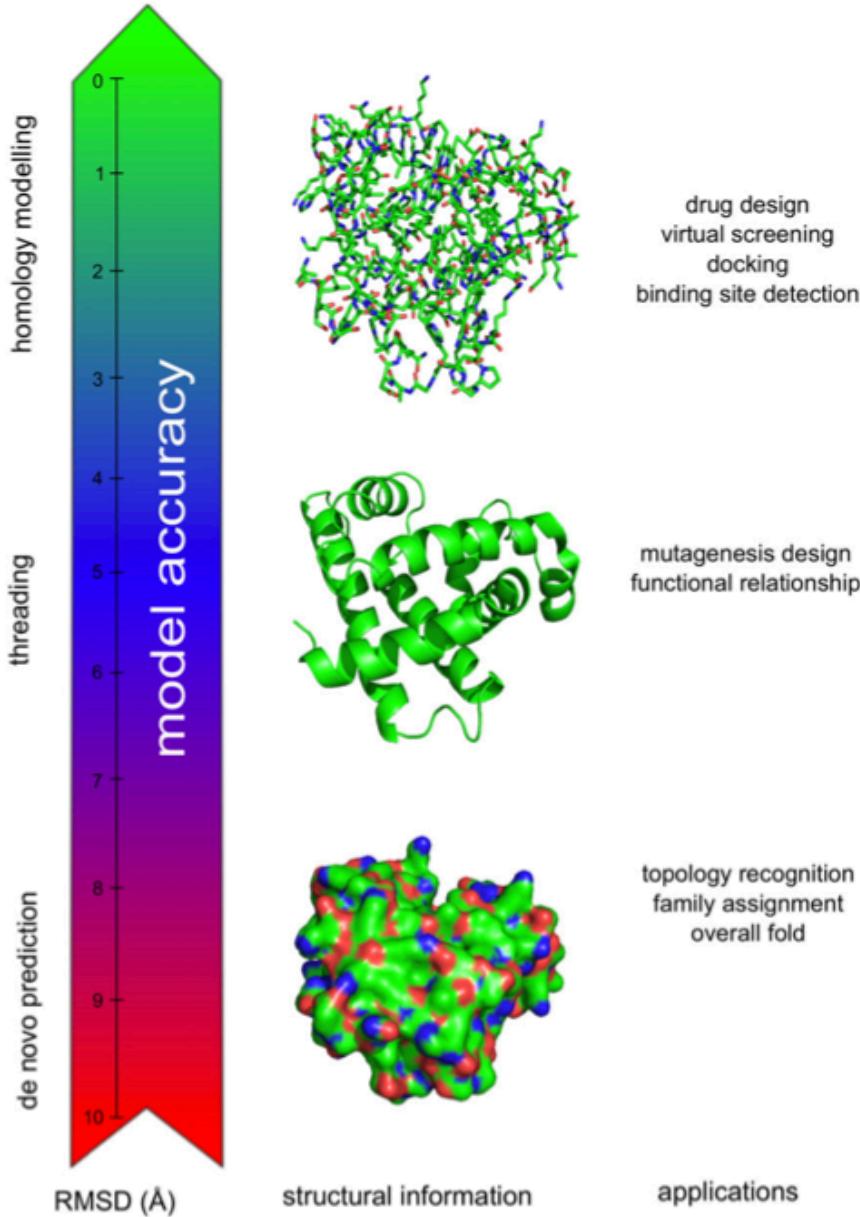


FIGURE 1. Schematic diagram showing the range of accuracy obtained by comparative modelling²³. The potential uses of comparative models depend on their accuracy. This in turn depends significantly on the sequence identity between the sequence modelled and the known structure on which the model was based. Sample models (red) are compared with the actual structures (blue).



Simple Homology Modelling

- We are going to use Modeller
- Free for academic use
- <http://salilab.org/modeller/9v6/modeller9v6.exe>
- Licence key: MODELIRANJE
- Modeller is a very sophisticated tool where you can control almost any aspect of the homology modelling process
- Here we are only going to use the simplest options
- Modeller has no interface. To use it we have to write python scripts

Chain we are going to model

ENLYFQSMINSFYA**F**EVKDAKGRTVSLEKYKGKVSLV
VNVAASDCQLTDRNYLGLKELHKEFGPSHFSVLA~~F~~PCN
QFGESEPRPSKEVESFARKNYGVTFP**I**FHKIKILGSE
GEPAFRFLVDSSK**K**EPRWNFWKYLVNPEGQVVKFWRP
EEPIEVIRPDIAALVRQV**I**IKKKEDL

T0388 LOC493869A, *Homo sapiens*

↑
CASP target ID

This sequence was one of the targets of the CASP8 experiment

1st step: BLAST against PDB

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence [?](#) [Clear](#) **Query subrange** [?](#)

```
ENLYFQSMINSFYAFEVAKDAKGRTVSLEKYKGKVSLVVNVASDCQLTDRNYLGLKELHKE  
FGPSHFSQLAFTPNCNQFGESEPRPSKEVESFARKNYGVTFFPIFHKIKILGSEGEPAFRFLV  
DSSKKEPRWNFWKYLVNPEGQVVKFWRPEEPIEVIRPDIAALVRQVIIKKEDL
```

From
To

Or, upload file [Browse...](#) [?](#)

Job Title
Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database [?](#)

Selecting the template

- The perfect match exists, because right now the structure for this target is already public
- We are going to ignore it, and use chain A of pdb entry 2p31 instead

```
>□pdb|3CYN|A S Chain A, The Structure Of Human Gpx8
  pdb|3CYN|B S Chain B, The Structure Of Human Gpx8
  pdb|3CYN|C S Chain C, The Structure Of Human Gpx8
Length=189

Score = 357 bits (917), Expect = 9e-100, Method: Compositional matrix adjust.
Identities = 174/174 (100%), Positives = 174/174 (100%), Gaps = 0/174 (0%)

Query  1      ENLYFQSMINSFYAFEVKDAKGRTVSLEKYKGKVSLVNVASDCQLTDRNYLGLKELHKE  60
          ENLYFQSMINSFYAFEVKDAKGRTVSLEKYKGKVSLVNVASDCQLTDRNYLGLKELHKE
Sbjct   16     ENLYFQSMINSFYAFEVKDAKGRTVSLEKYKGKVSLVNVASDCQLTDRNYLGLKELHKE  75
          ENLYFQSMINSFYAFEVKDAKGRTVSLEKYKGKVSLVNVASDCQLTDRNYLGLKELHKE

Query  61     FGPISHFSVLAFFPCNQFGESEPRPSKEVESFARKNYGVTFFPIFHKKIKILGSEGEPAFRFLV  120
          FGPISHFSVLAFFPCNQFGESEPRPSKEVESFARKNYGVTFFPIFHKKIKILGSEGEPAFRFLV
Sbjct   76     FGPISHFSVLAFFPCNQFGESEPRPSKEVESFARKNYGVTFFPIFHKKIKILGSEGEPAFRFLV  135
          FGPISHFSVLAFFPCNQFGESEPRPSKEVESFARKNYGVTFFPIFHKKIKILGSEGEPAFRFLV

Query 121     DSSKKEPRWNFWKYLVNPPEGQVVFKWRPEEPIEVIRPDIAALVRQVIKKKEDL  174
          DSSKKEPRWNFWKYLVNPPEGQVVFKWRPEEPIEVIRPDIAALVRQVIKKKEDL
Sbjct  136     DSSKKEPRWNFWKYLVNPPEGQVVFKWRPEEPIEVIRPDIAALVRQVIKKKEDL  189
          DSSKKEPRWNFWKYLVNPPEGQVVFKWRPEEPIEVIRPDIAALVRQVIKKKEDL
```

```
>□pdb|2P31|A S Chain A, Crystal Structure Of Human Glutathione Peroxidase 7
  pdb|2P31|B S Chain B, Crystal Structure Of Human Glutathione Peroxidase 7
Length=181

Score = 210 bits (534), Expect = 3e-55, Method: Compositional matrix adjust.
Identities = 95/166 (57%), Positives = 123/166 (74%), Gaps = 2/166 (1%)

Query  1      ENLYFQSMIN--SFYAFEVKDAKGRTVSLEKYKGKVSLVNVASDCQLTDRNYLGLKELH  58
          ENLYFQSM      FY F+ + G+ VSLEYK+G VSLVVNVAS+C TD++Y L++L
Sbjct   16     ENLYFQSMQQEQDFYDFKAVNIRGKLVSLEKYRGSVSLVNVASECGFTDQHYRALQQLQ  75
          ENLYFQSMQQEQDFYDFKAVNIRGKLVSLEKYRGSVSLVNVASECGFTDQHYRALQQLQ

Query  59     KEFGPSHFSVLAFFPCNQFGESEPRPSKEVESFARKNYGVTFFPIFHKKIKILGSEGEPAFRF  118
          ++ GP HF+VLAFFPCNQFG+ EP +KE+ESFAR+ Y V+FP+F KI + G+ PAF++
Sbjct   76     RDLGPHHFNVLAFFPCNQFGQQEPDSNKEIESFARRTYSVSFPMSKIAVTGTGAHPAFKY  135
          RDLGPHHFNVLAFFPCNQFGQQEPDSNKEIESFARRTYSVSFPMSKIAVTGTGAHPAFKY

Query 119     LVDSSKKEPRWNFWKYLVNPPEGQVVFKWRPEEPIEVIRPDIAALVR  164
          L +S KEP WNFWKYLW P+G+VV W P +E +RP I ALVR
Sbjct  136     LAQTSGKEPTWNFWKYLVAPDGKVVGAWDPTVSVEEVRPQITALVR  181
          LAQTSGKEPTWNFWKYLVAPDGKVVGAWDPTVSVEEVRPQITALVR
```

2nd step: Creating an alignment

- Modeller has a sophisticated alignment tool
 - Uses structural information from the template
 - Dynamic programming instead of the approximate method of blast
- To create the alignment you need to:
 1. Download the [PDB file](#) of the template
 2. Put your sequence in PIR format ([example](#))
 3. Edit the [alignment script](#) to set the template and chain
 4. Call modeller: mod9v6.exe align.py

PIR file

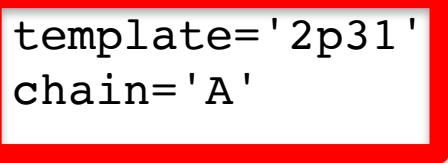
- Just replace the sequence with your own one
- The last line in the sequence needs to end in *
- Do not touch anything else from the file, or the alignment script will not work
- File name: target.ali

```
>P1;target
sequence:target::::::::::0.00: 0.00
ENLYFQSMINSFYAF'EVKDAKGRTVSLEKYKGKVSLVVNVASDCQLTDRNYLGLKELHKE
FGPSHF'SVLA'FPCNQFGESEPRPSKEVESFARKNYGVTFP'IFHKIKILGSEGEPAFRFLV
DSSKKEPRWNFWKYLVNPEGQVVKFWRPEEPIEVIRPDIAALVRQVIKKKEDL*
```

Align.py

```
from modeller import *
from modeller.automodel import *

env = environ()
aln = alignment(env)

template='2p31'
chain='A'  

Just change the value of these 2 lines  
with your template


tc=template+chain

mdl = model(env, file=template,
model_segment=('FIRST:'+chain,'LAST:'+chain))
aln.append_model(mdl, align_codes=tc, atom_files=template+'.pdb')
aln.append(file='target.ali', align_codes='target')
aln.align2d()
aln.write(file='target-'+tc+'.ali', alignment_format='PIR')
aln.write(file='target-'+tc+'.pap', alignment_format='PAP')
```

- Alignment is different from that produced by BLAST
 - Modeller has ignored the residues lacking structural information

Creating the model

```
from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

template='2p31'
chain='A'

tc=template+chain

class MyModel(automodel):
    def get_model_filename(self,sequence,id1,id2,
file_ext):
        return sequence+'_'+id2+file_ext

    def special_restraints(self, aln):
        rsr = self.restraints

a = MyModel(env, alnfile='target-'+tc+'.ali',
            knowns=tc, sequence='target',
            assess_methods=(assess.DOPE, assess.GA341))
a.starting_model = 1
```

- 5 models are created
- Each of them can be slightly different
- Models are going to be assessed using 2 different criteria

Results of the modelling

>> Summary of successfully produced models:

Filename	molpdf	DOPE score	GA341 score

target_1.pdb	1280.53101	-19077.32812	1.00000
target_2.pdb	1570.33606	-18480.83008	1.00000
target_3.pdb	960.32550	-19365.79102	1.00000
target_4.pdb	1415.41724	-18980.71094	1.00000
target_5.pdb	1463.82593	-19077.91016	1.00000

- According to DOPE score, 3 is the best model and 2 the worst
- The lowest the DOPE score, the better
- Let's see how different are the models

```
from modeller import *
from modeller.scripts import complete_pdb
log.verbose()      # request verbose output
env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters
# read model file
mdl = complete_pdb(env, 'hvdfr.B99990001.pdb')
# Assess with DOPE: s = selection(mdl)    # all atom selection
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='hvdfr.profile',
normalize_profile=True, smoothing_window=15)
```

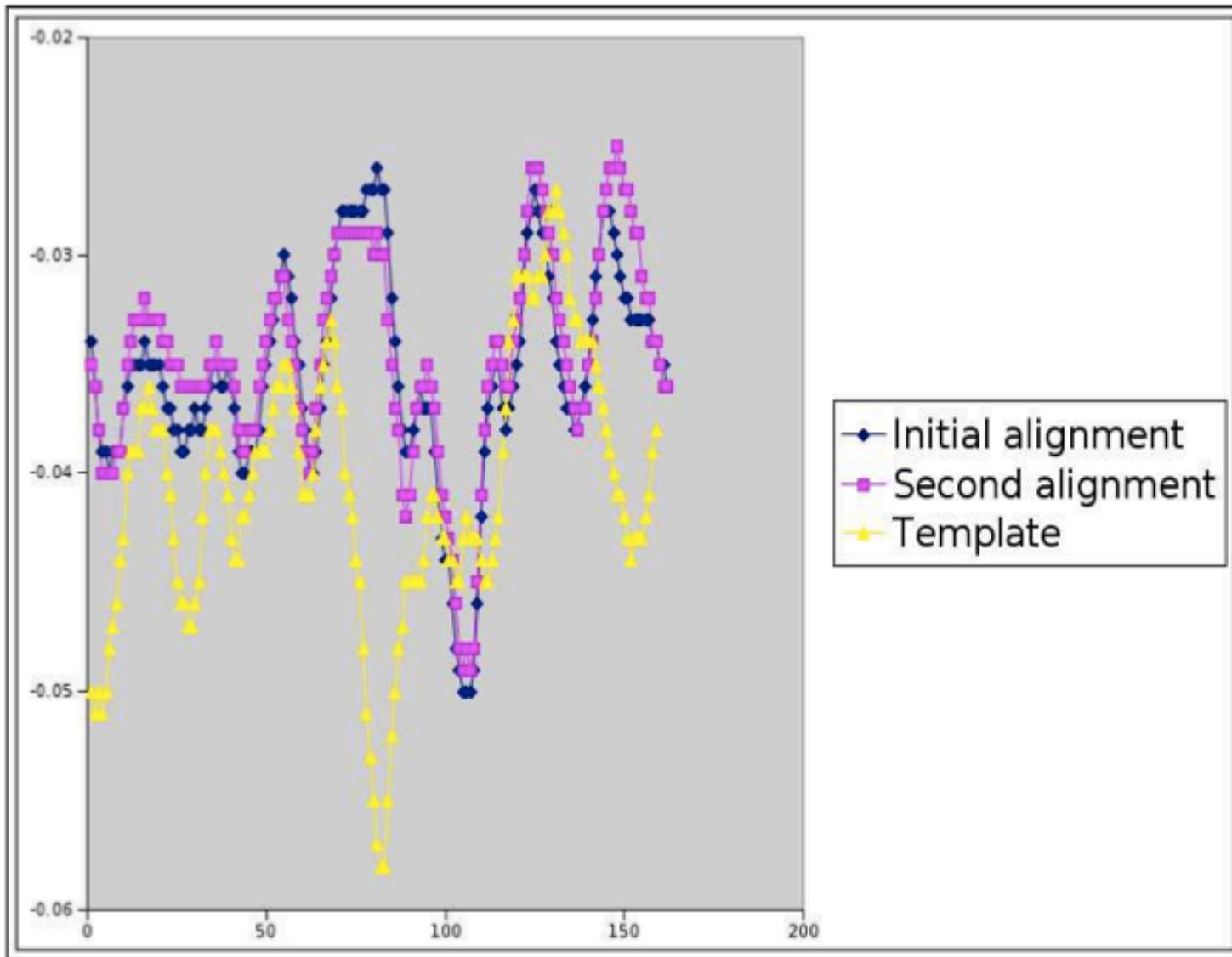
Analysis

The results can be graphed using the gnumeric (or any other) spreadsheet.

- Launch gnumeric from the main menu or entering “gnumeric &” on the command line.
- Go to Data -> Get external data -> Import text file.
- Change the spreadsheets option to “all files” and open “hvdfr.profile” .
- Run through the wizard (clicking forward), selecting the “See two separators as one” option before importing.

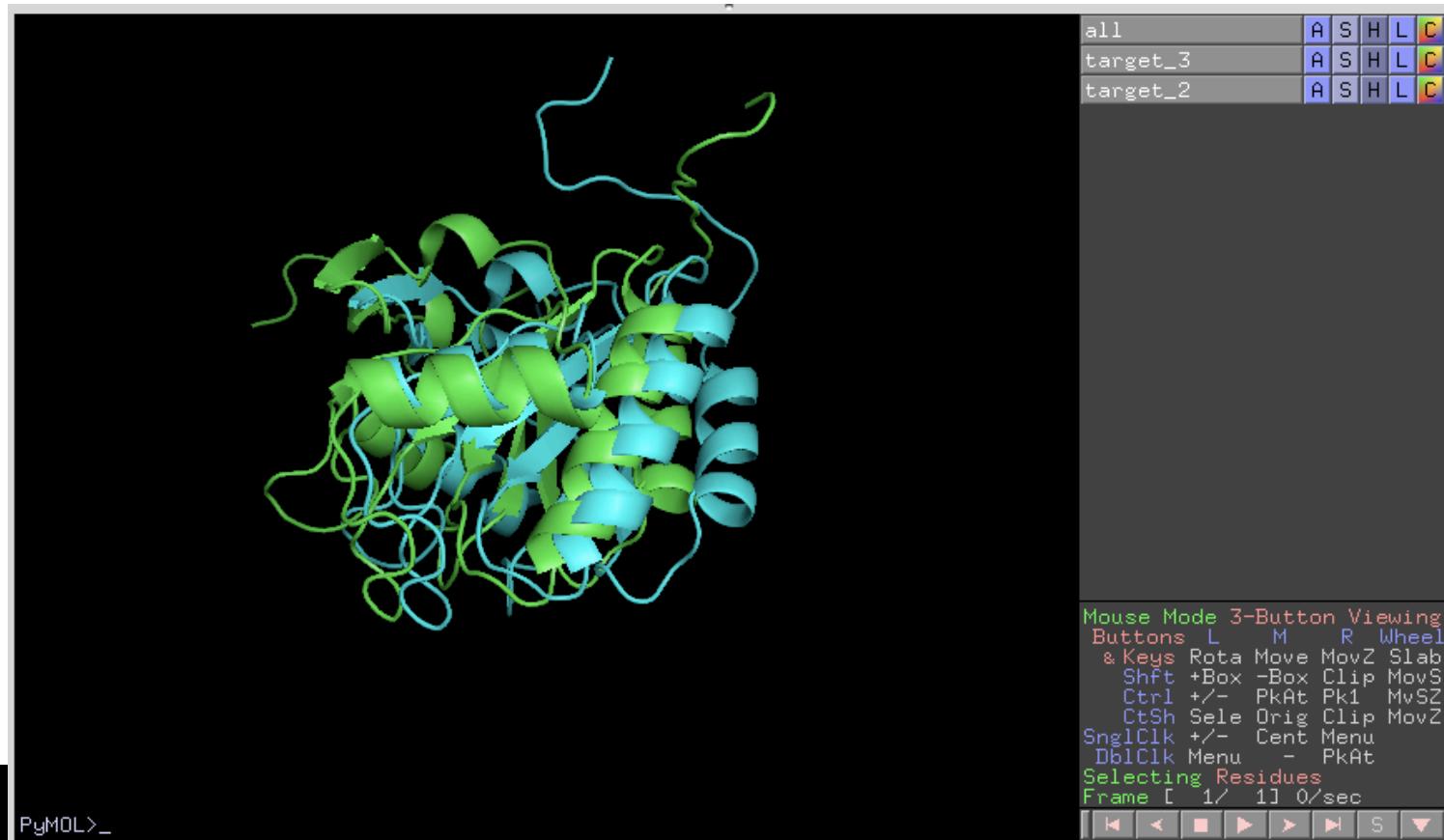
The spreadsheet will now be populated with profile data. Plot the first and last columns of data.

- Select the data (hold down Ctrl and use the arrow keys to select the second column),
- Click the chart button (bar graph icon).
- Drag over the area where you want to place the graph.
- Double click the graph to edit and add new data series
- Add plots for the model/s and template



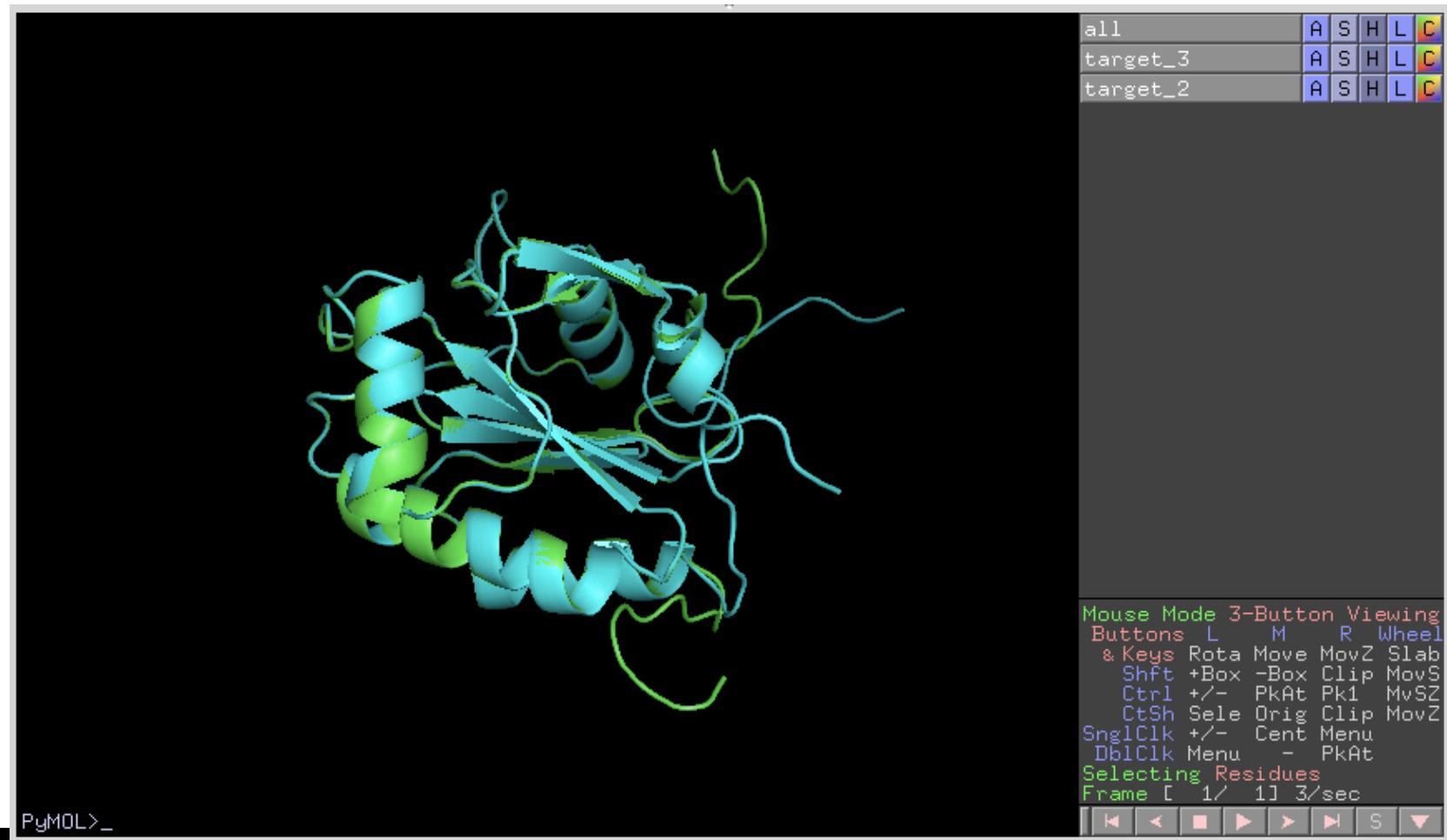
Viewing the two models from pymol

1. Open model 3 as usual
2. But then, instead of double-clicking model 2, open it from inside pymol using File → open
3. The models are not aligned



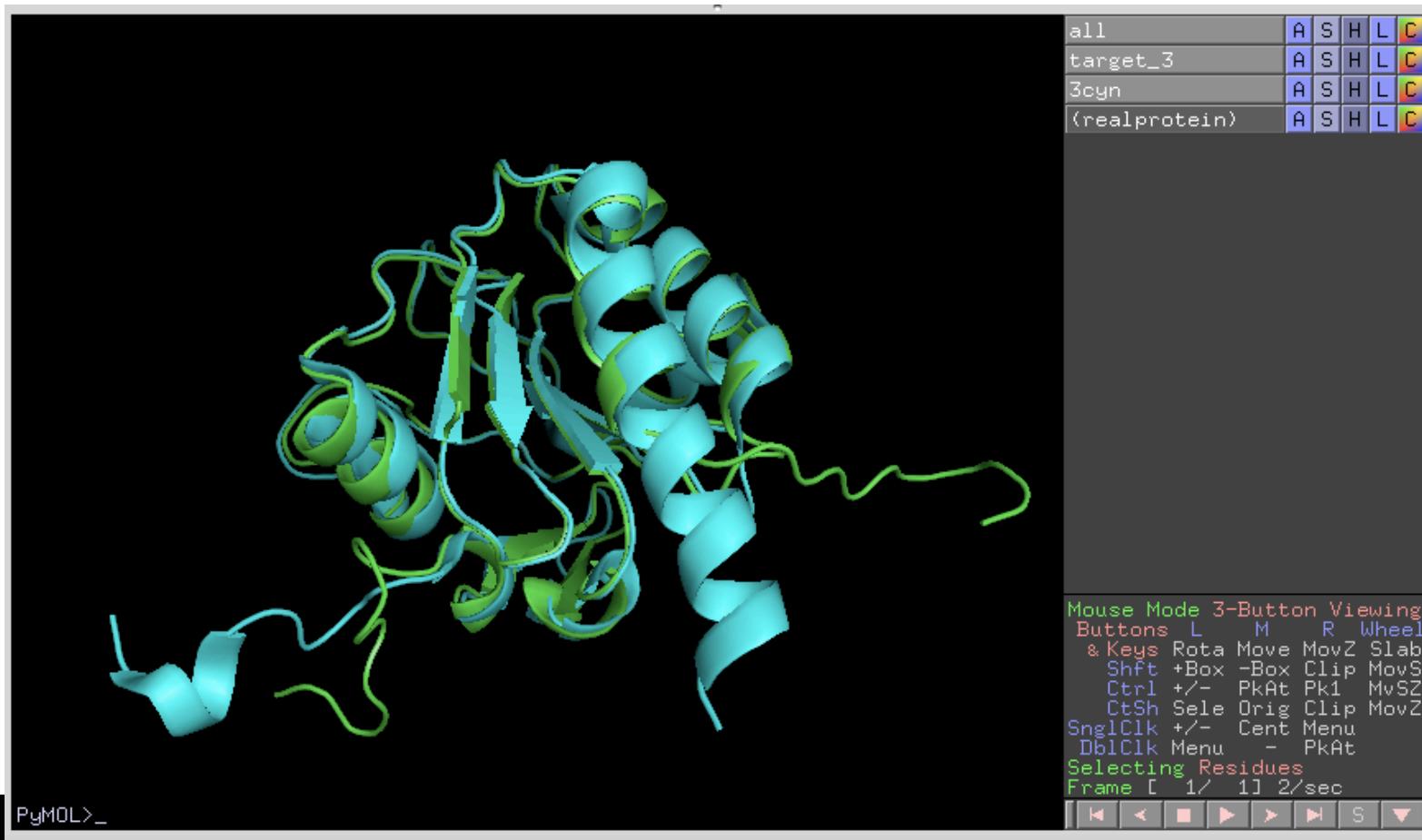
Type: align target_3,target_2

- The only differences are in the two ends of the chain



So how does the model compare to the real protein 3CYN?

- The residues at both ends of the chain are wrong



Can we do any better?

- We can give modeller information about the secondary structure of the target
- We can get these predictions from [PSIPRED](#)

CCCCCCCCCCCCCEEEEEEECCCCCEECHHHHCCCEEE
EEECCCCCCCCCHHHHHHHHHHHHHCCCCEEeeeeeee
CCCCCCCCCCCCCHHHHHHHHHCCCCHHEEEEEECC
CCCCCHHHHHHHHCCCCCCCCCEEeeeeccccee
EECCCCCHHHHHHHHHHHHHHHHHHHHHCCC

- Then, the [modelling script](#) needs to be modified

```

from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

template='2p31'
chain='A'
tc=template+chain

class MyModel(automodel):
    def get_model_filename(self, sequence, id1, id2, file_ext):
        return sequence+'_'+id2+file_ext

    def special_restraints(self, aln):
        rsr = self.restraints
        rsr.add(secondary_structure.strand(self.residue_range('12:', '18:')))
        rsr.add(secondary_structure.strand(self.residue_range('24:', '25:')))
        rsr.add(secondary_structure.alpha(self.residue_range('27:', '30:')))
        rsr.add(secondary_structure.strand(self.residue_range('34:', '39:')))
        rsr.add(secondary_structure.alpha(self.residue_range('48:', '61:')))
        rsr.add(secondary_structure.strand(self.residue_range('66:', '72:')))
        rsr.add(secondary_structure.alpha(self.residue_range('84:', '93:')))
        rsr.add(secondary_structure.alpha(self.residue_range('99:', '100:')))
        rsr.add(secondary_structure.strand(self.residue_range('101:', '106:')))
        rsr.add(secondary_structure.alpha(self.residue_range('114:', '121:')))
        rsr.add(secondary_structure.strand(self.residue_range('132:', '136:')))
        rsr.add(secondary_structure.strand(self.residue_range('142:', '146:')))
        rsr.add(secondary_structure.alpha(self.residue_range('152:', '171:')))

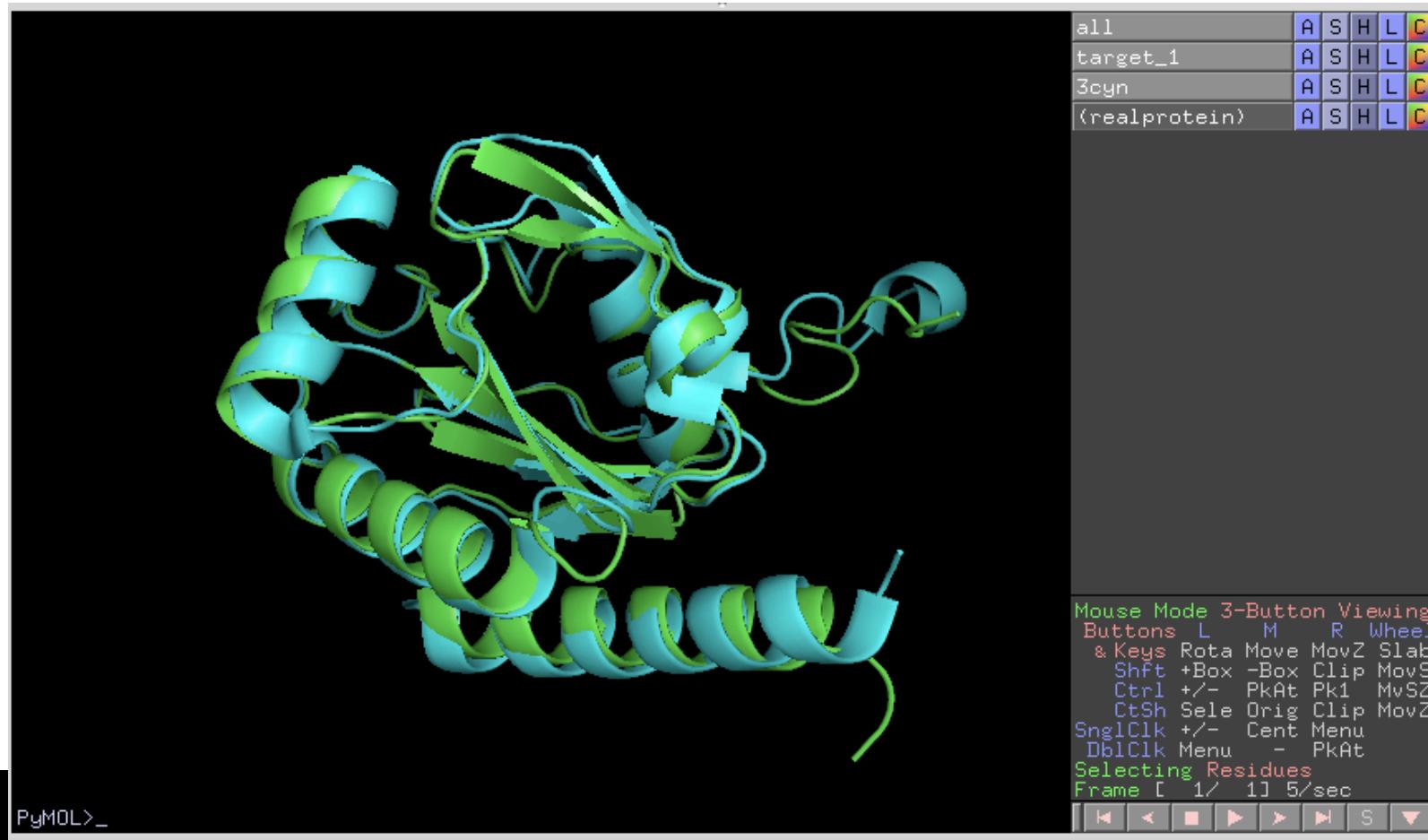
a = MyModel(env, alnfile='target-'+tc+'.ali',
            knowns=tc, sequence='target',
            assess_methods=(assess.DOPE, assess.GA341))
a.starting_model = 1
a.ending_model = 5

```

Pred
SS
info

And here is the new model,

- Now at least we got right one end of the protein



Lets do the same with

- MKEGDSLLRNVAGPLGTPVPMEKKFHKILAIGAYTG
IVEVYPIAKAWQEIGNDVTTLHVTFEPMVILKEELEK
AVTRADEEHIVEPVPLNPNADFLANMKNVSQRLKE
KVRELLESEMAAARDWDLVFMVRPVGDQKQVFEV
VWKEYQVPMKVDLHPIMVDGLE

>[pdb|3LRX|A](#) **S** Chain A, Crystal Structure Of The C-Terminal Domain (Residues 78-226) Of Pf1911 Hydrogenase From Pyrococcus Furiosus, Northeast Structural Genomics Consortium Target Pfr246a

[pdb|3LRX|B](#) **S** Chain B, Crystal Structure Of The C-Terminal Domain (Residues 78-226) Of Pf1911 Hydrogenase From Pyrococcus Furiosus, Northeast Structural Genomics Consortium Target Pfr246a

[pdb|3LRX|C](#) **S** Chain C, Crystal Structure Of The C-Terminal Domain (Residues 78-226) Of Pf1911 Hydrogenase From Pyrococcus Furiosus, Northeast Structural Genomics Consortium Target Pfr246a

[pdb|3LRX|D](#) **S** Chain D, Crystal Structure Of The C-Terminal Domain (Residues 78-226) Of Pf1911 Hydrogenase From Pyrococcus Furiosus, Northeast Structural Genomics Consortium Target Pfr246a

[pdb|3LRX|E](#) **S** Chain E, Crystal Structure Of The C-Terminal Domain (Residues 78-226) Of Pf1911 Hydrogenase From Pyrococcus Furiosus, Northeast Structural Genomics Consortium Target Pfr246a

[pdb|3LRX|F](#) **S** Chain F, Crystal Structure Of The C-Terminal Domain (Residues 78-226) Of Pf1911 Hydrogenase From Pyrococcus Furiosus, Northeast Structural Genomics Consortium Target Pfr246a

Length=158

Score = 254 bits (648), Expect = 2e-68, Method: Compositional matrix adjust.
Identities = 141/163 (86%), Positives = 141/163 (86%), Gaps = 12/163 (7%)

Query 2 KEGDSLLRNVAGPLGTPVPMEKKFH KILAIGAYTGIVEVYPIAKAWQEIGNDVTLHVTF 61
KEGDSLL NVAGPLGTPVP EK F KILAIGAYTGIVEVYPIAKAWQEIGNDVTLHVTF

Sbjct 2 KEGDSLL-NVAGPLGTPVPXEK-FGKILAIGAYTGIVEVYPIAKAWQEIGNDVTLHVTF 59

Query 62 EPMVILKEELEKAVTRADEEHIVEPVPLNPNAFLANMKNVSQRLKEKVRELLESEAAA 121
EP VILKEELEKAVTR HIVEPVPLNPNAFLAN KNVSQRLKEKVRELLESE

Sbjct 60 EPXVILKEELEKAVTR----HIVEPVPLNPNAFLANXKNVSQRLKEKVRELLESE---- 111

Query 122 RDWDLVFMVRPVGDKQKVFEVVWKEYQVPMKVDLHPIVDGLE 164
DWDLVF V PVGDQKQVFEVV KEY VP KVDLHPI VDGLE

Sbjct 112 -DWDLVFXVGPVGDKQKVFEVV-KEYGVPXKVDLHPIXVDGLE 152

3LRX

[Display Files ▾](#)
[Download Files ▾](#)
[Print this Page](#)
[Share this Page](#)

Crystal Structure of the C-terminal domain (residues 78-226) of PF1911 hydrogenase from Pyrococcus furiosus, Northeast Structural Genomics Consortium Target PfR246A

DOI:10.2210/pdb3lrx/pdb

Primary Citation

Northeast Structural Genomics Consortium Target PfR246A

Forouhar, F., Abashidze, M., Seetharaman, J., Mao, M., Xiao, R., Ciccosanti, C., Foote, E.L., Belote, R.L., Everett, J.K., Nair, R., Acton, T.B., Rost, B., Montelione, G.T., Tong, L., Hunt, J.F.

Journal: To be Published

Not in PubMed

↳ Molecular Description

[Hide](#)

Classification: [Oxidoreductase](#)

Structure Weight: 109110.59

Molecule: Putative hydrogenase

Polymer: 1 Type: polypeptide(L)

Chains: A, B, C, D, E, F

EC#: [1.18.1.2](#) EC

Fragment: sequence database residues 78-226

Length: 158

↳ Source

[Hide](#)

Polymer: 1

Scientific Name: [Pyrococcus furiosus](#)



Expression System:

[Escherichia coli](#)

[View in Jmol](#) [SimpleViewer](#)
Other Viewers ▾ [Protein Workshop](#)

Biological assembly 1 assigned by authors
and generated by PISA (software)



Make a Model...

- Automodel

```
from modeller import *
from modeller.automodel import *
env = environ()
a = automodel(env, alnfile='hvdfr-4dfr.ali', knowns='4dfr', sequence='hvdfr')
a.starting_model = 1
a.ending_model = 1
a.make()
```

At the End!!!

- REPRESENT YOUR RESULT IN VMD
- NEXT WEEK HOMEWORK
 - CREATE A MUTANT OF YOUR PROTEIN
 - 1 RESIDUE
 - Active Site
 - BAD TEMPLATE RESULT

At the End!!!

