Protein Structure Prediction and Refinement

Chaok Seok

Department of Chemistry
Seoul National University

Structural and biophysical methods for biological macromolecules in solution

EMBO global exchange lecture course, Sungkyunkwan University, Jun 19-26, 2016

Contents

- Introduction: principles of structure prediction methods
- Computational prediction methods: theory and practice
  - Protein structure prediction (Robetta, I-TASSER, HHpred, SWISS-MODEL)
  - Protein structure refinement, loop modelling, and docking (GALAXY)
- Conclusion: when/how to use computational methods

Introduction: principles of structure prediction methods

Natural law

Principles of physics or physical chemistry

Evolutionary principle

Related sequences, structures, functions

Physical principles for structure prediction

Experimentally observed states (or structures) corresponds to thermodynamically stable states in the given condition (pH, salt concentration, temperature, etc)

Protein molecules often have enough time to reach stable states before they act -> reproducible/reliable function

(in equilibrium)

Physical law for stability?

Stable states have the lowest free energy

$G=H-TS$

Transitions between states corresponds to the minimum-free energy path

Free energy as a function of state (structure/conformation)

G (Free energy at constant pressure) =
H (enthalpy) -T (temperature) S (entropy)

H (enthalpy) = E (energy) - pV (energy needed to maintain pressure)

E = potential energy (interaction energy)
+ kinetic energy (determined by temperature)

Physical principle for potential energy of molecules

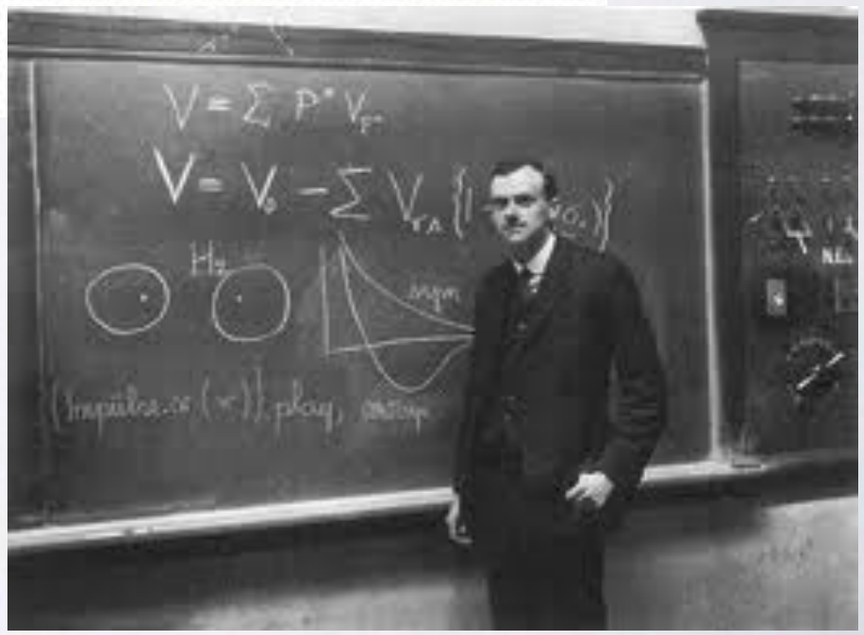
Quantum mechanics

Energy is determined by

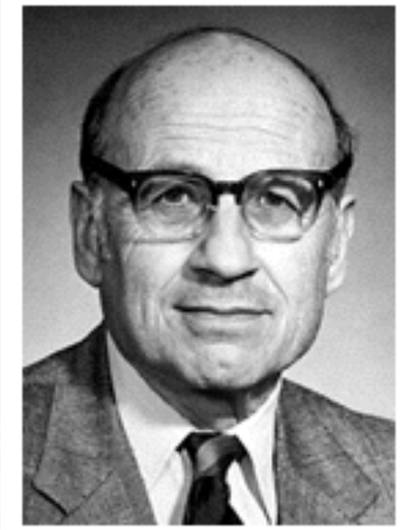
electrostatic interactions between electrons and nuclei

Quantum Mechanics of Many-Electron Systems (Dirac '29)

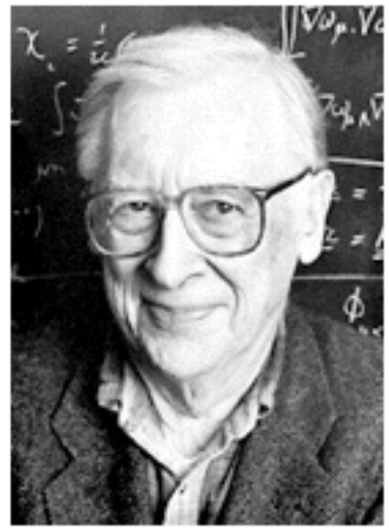
"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations that are much too complicated to be soluble. **It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to explanation of the main features of complex atomic systems without too much computation.**"



Nobel prize in chemistry, 1998



Walter Kohn



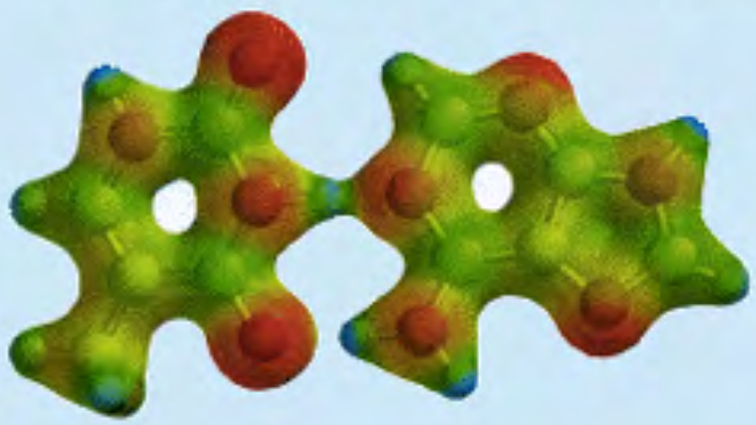
John A. Pople

Chemistry with computers $\hat{H}\Psi = E\Psi$

Chemistry is not only test tubes and chemicals. In quantum chemistry, quantum mechanics is used to *compute* the properties of molecules and their inter… laureates have made it possible to us… quantum mechanics to study molecul… with the help of computers.

H_2
→
H + HF

Charge distributions



Charge distributions in molecules can be computed with quantum-chemical methods. Excess of electrons (red in the illustration) in one molecule is drawn to a more positively charged portion of another molecule (blue). In this way the base pairs bind DNA together.

Computer time required for quantum calculations

Example: water molecules

1 water molecule: ~1 min

10 water molecules: ~1 day (for a single configuration)

10 water molecules, 1 microsecond: >1000 days

10000 water molecules, 1 microsecond: >million years

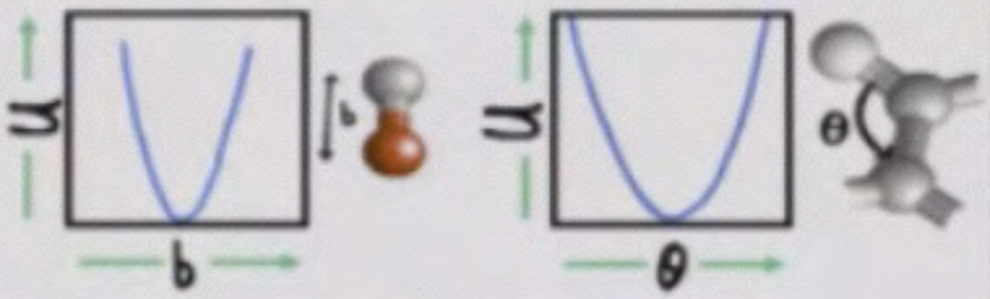
-> approximate energy with that of molecular mechanics (classical mechanics)

Molecular mechanics energy

MOLECULAR POTENTIAL ENERGY

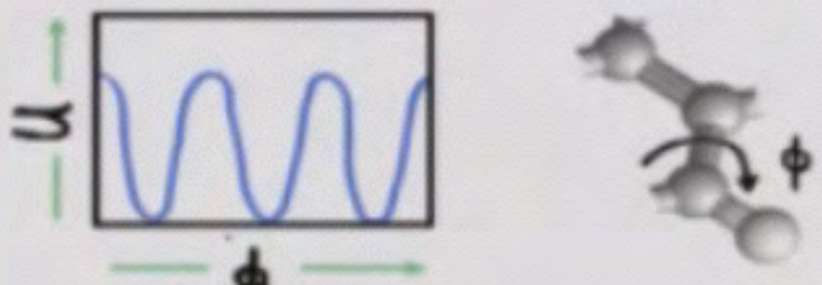
$$U = \sum_{\text{All}} \tfrac{1}{2} K_b (b - b_o)^2 + \sum_{\text{All}} \tfrac{1}{2} K_\theta (\theta - \theta_o)^2$$

Hooke



$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

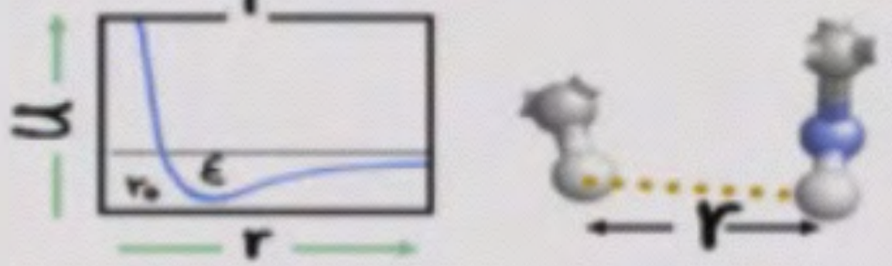
All Torsion



Fourier

$$+ \sum \epsilon \left[\left(\frac{r_o}{r}\right)^{12} - 2\left(\frac{r_o}{r}\right)^{6} \right]$$

All Nonbonded



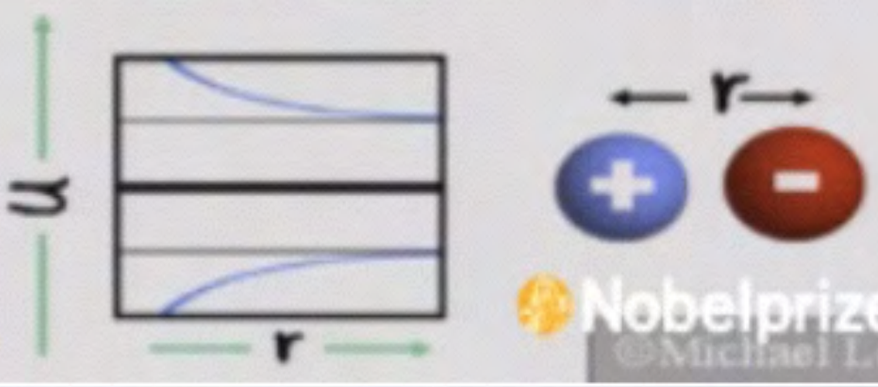
Van der Waals

$$+ \sum 332 q_i q_j / r$$

All partial

Coulomb

Simple sum over many terms



Nobelprize.org

©Michael Levitt '13

The Nobel Prize in Chemistry 2013

Martin Karplus, Michael Levitt, Arieh Warshel

The Nobel Prize in Chemistry 2013

© Nobel Media AB

Martin Karplus

Harvard Univ.

Photo: Keilana via Wikimedia Commons

Michael Levitt

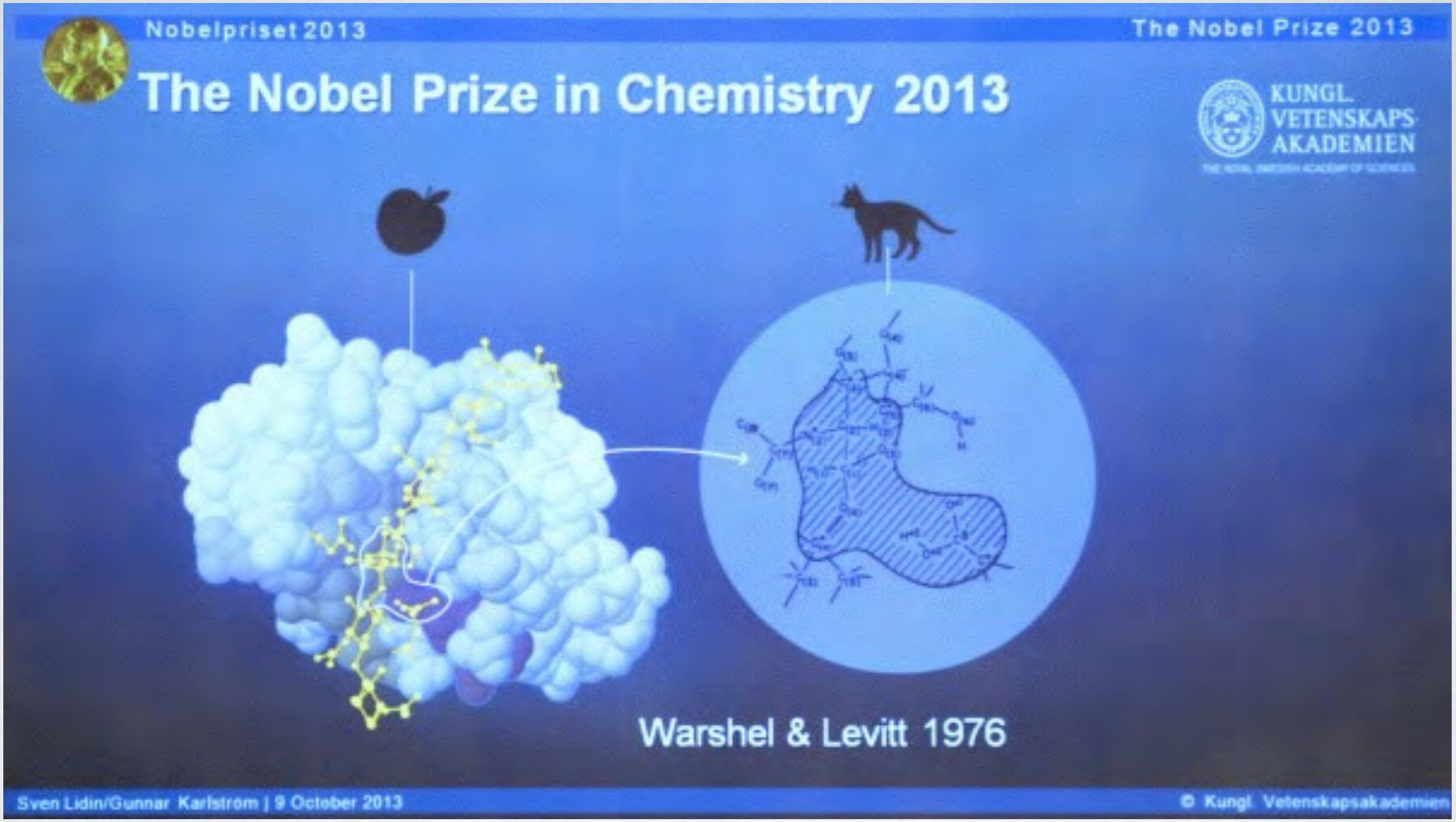
Stanford Univ.

Photo: Wikimedia Commons

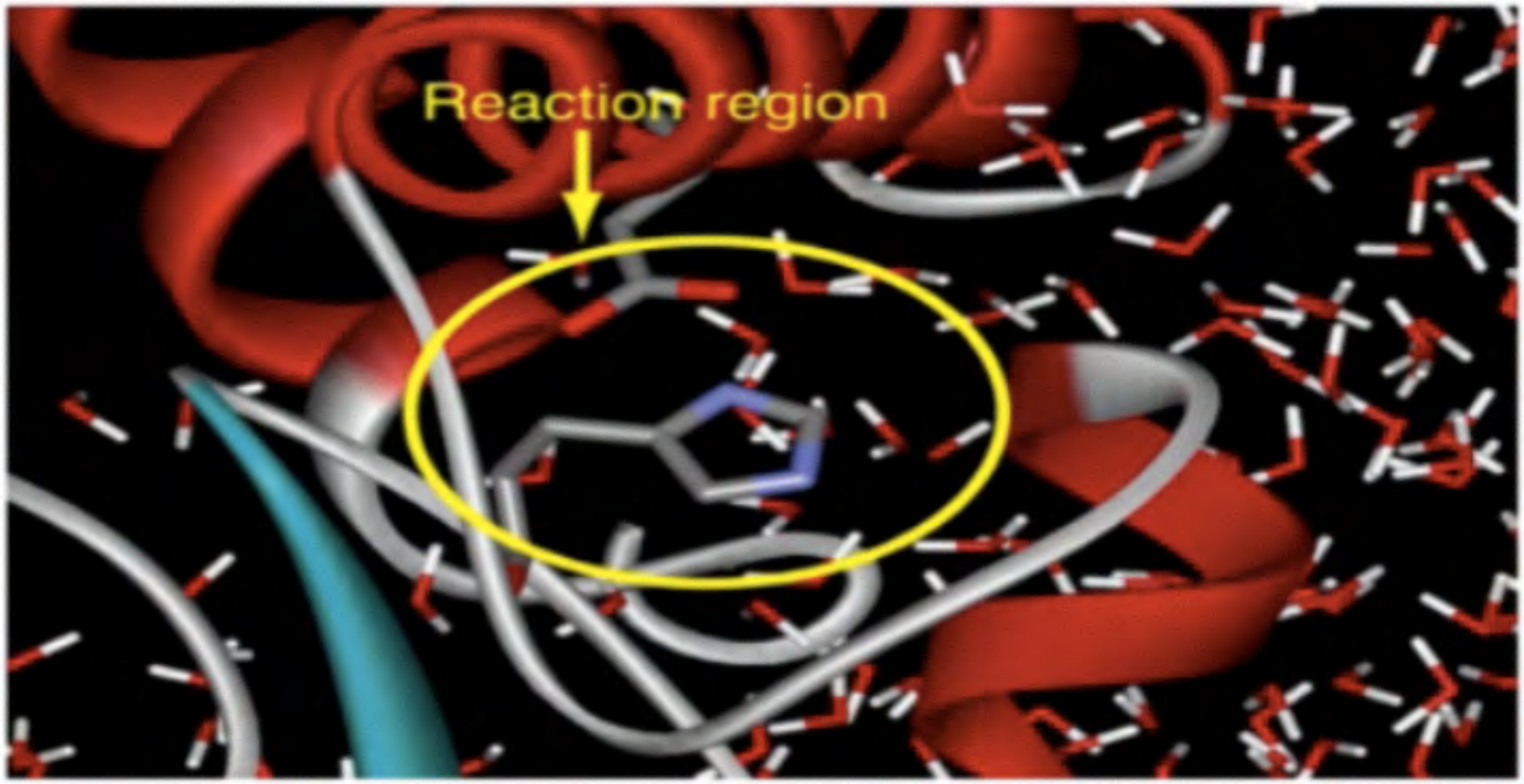
Arieh Warshel

Univ. Southern California

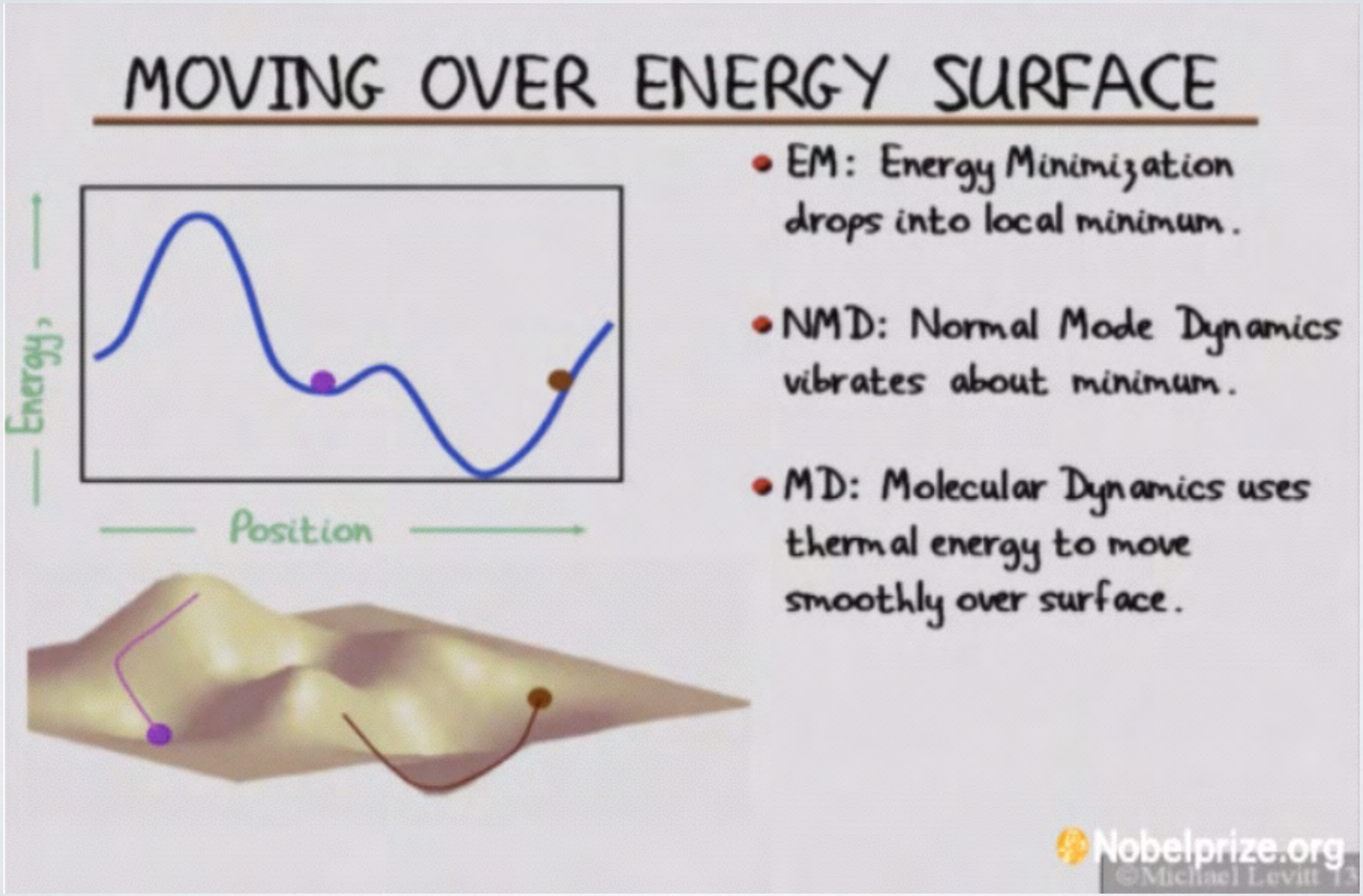
"Development of Multiscale Models for Complex Chemical Systems"


Nobelpriset 2013
The Nobel Prize 2013
The Nobel Prize in Chemistry 2013
KUNGL. VETENSKAPS-AKADEMIEN
Warshel & Levitt 1976
Sven Lidin/Gunnar Karlström | 9 October 2013
© Kungl. Vetenskapsakademien

QM/MM: To study enzymatic reactions, we divide the system in two parts (Warshel & Levitt, JMB 1976)



States on the energy surface



“…everything that living things do can be understood in terms of the jigglings and wigglings of atoms.”

The Feynman Lectures in 1963

States observed by molecular dynamics simulation

Correspond to experimentally observed states in the same condition (pH, salt, temperature, etc) if

the potential energy function is accurate,

and

the simulation is long enough (in the experimental time scale).

Problems of force field:

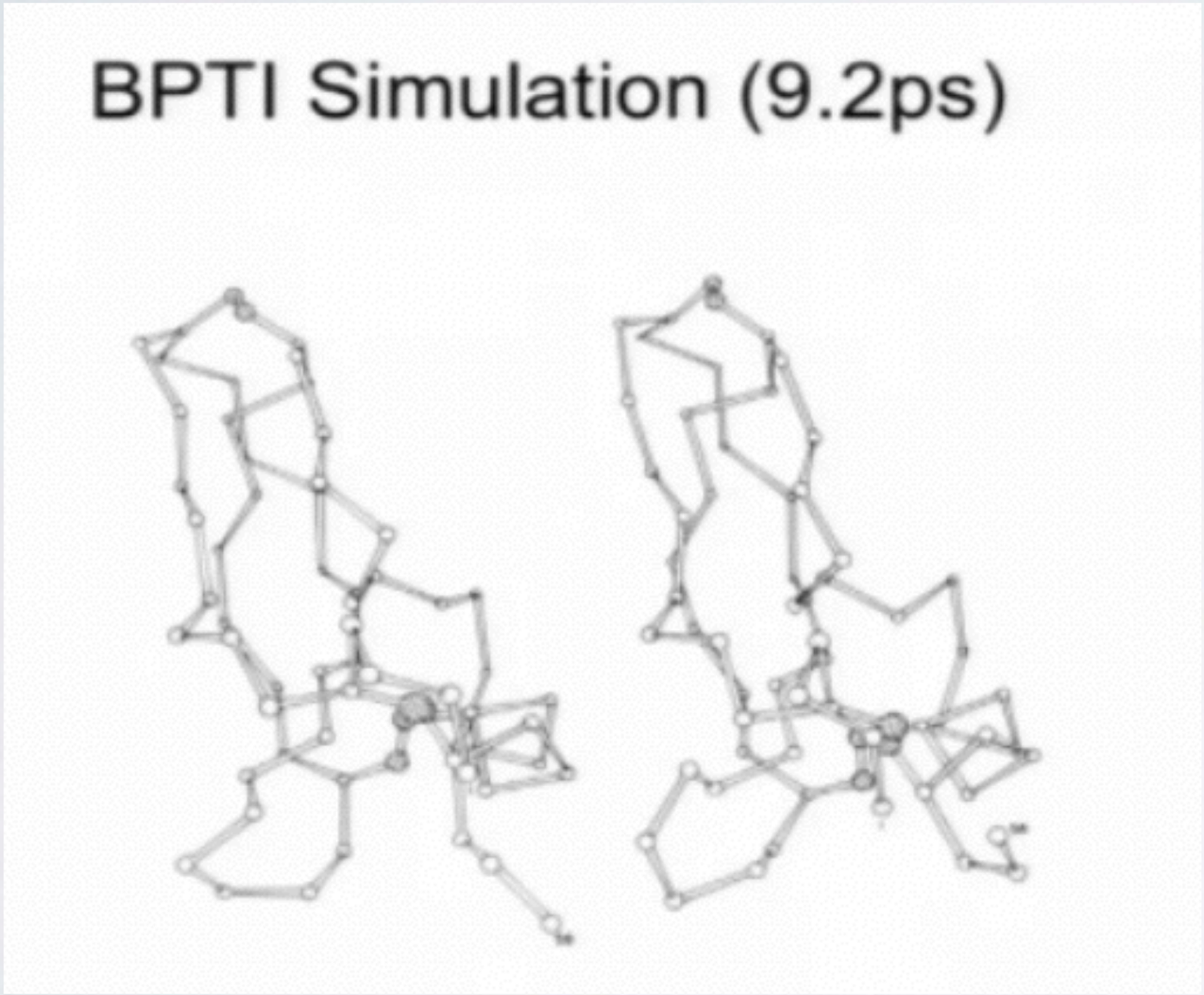
Simple functional form,

Lack of polarization,

Parameter training on small molecules,

+ Heuristic adjustments

First molecular dynamics (MD) simulation of protein



Computer time required for a folding protein by molecular dynamics simulation

A millisecond-simulation of a protein in solution

with a time step of 1 femtosecond requires

10^12 steps!

(each step requires energy and gradient evaluation
for protein and all solvent molecules)

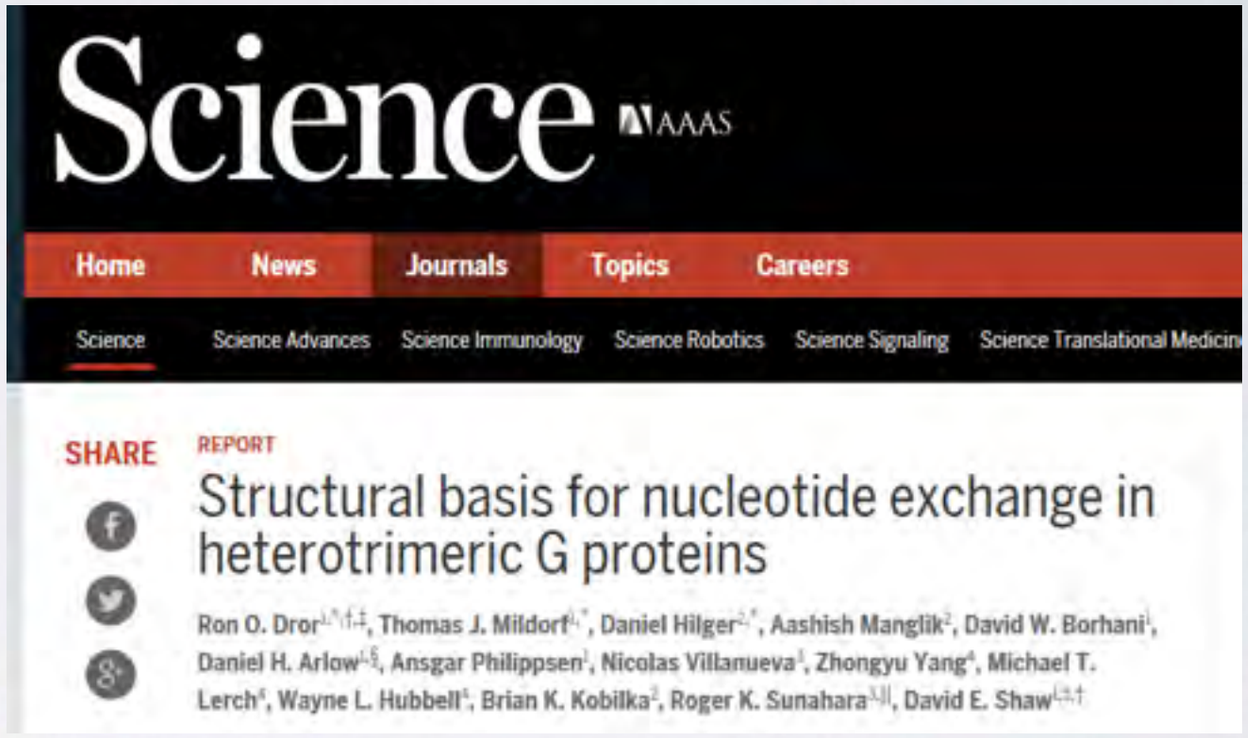
-> new method needed!

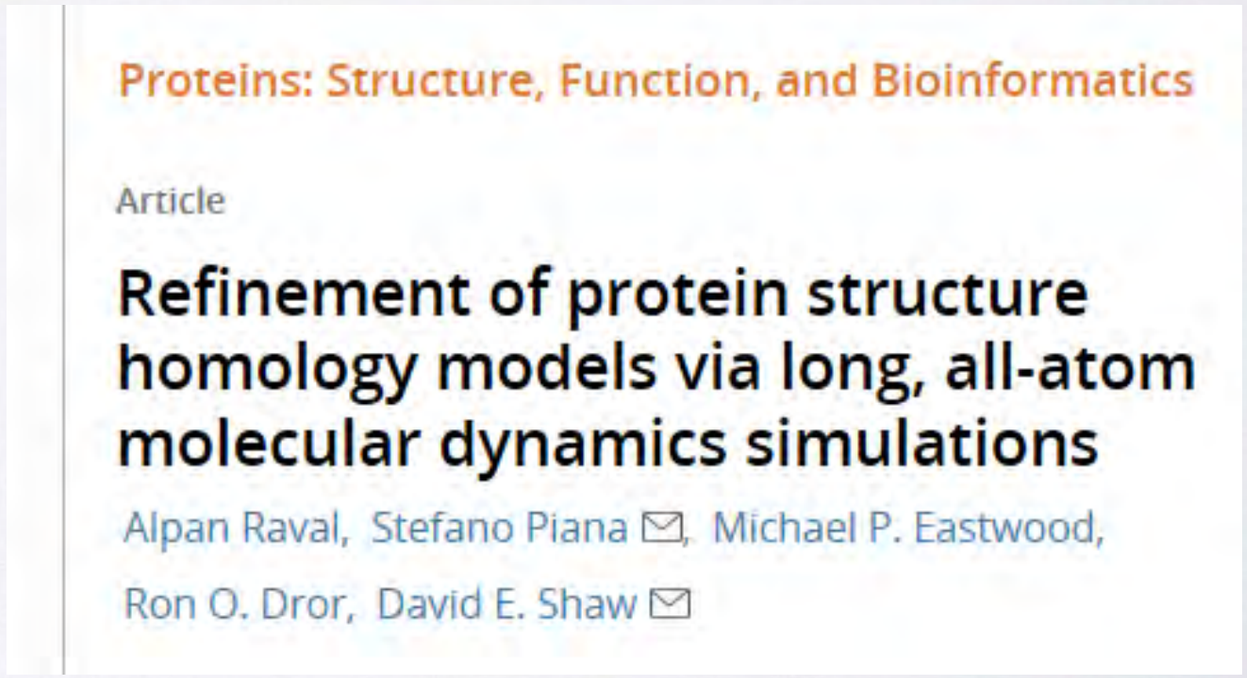
Failure of Force Field in Protein Structure Prediction

David E. Shaw

Anton

Energy problem





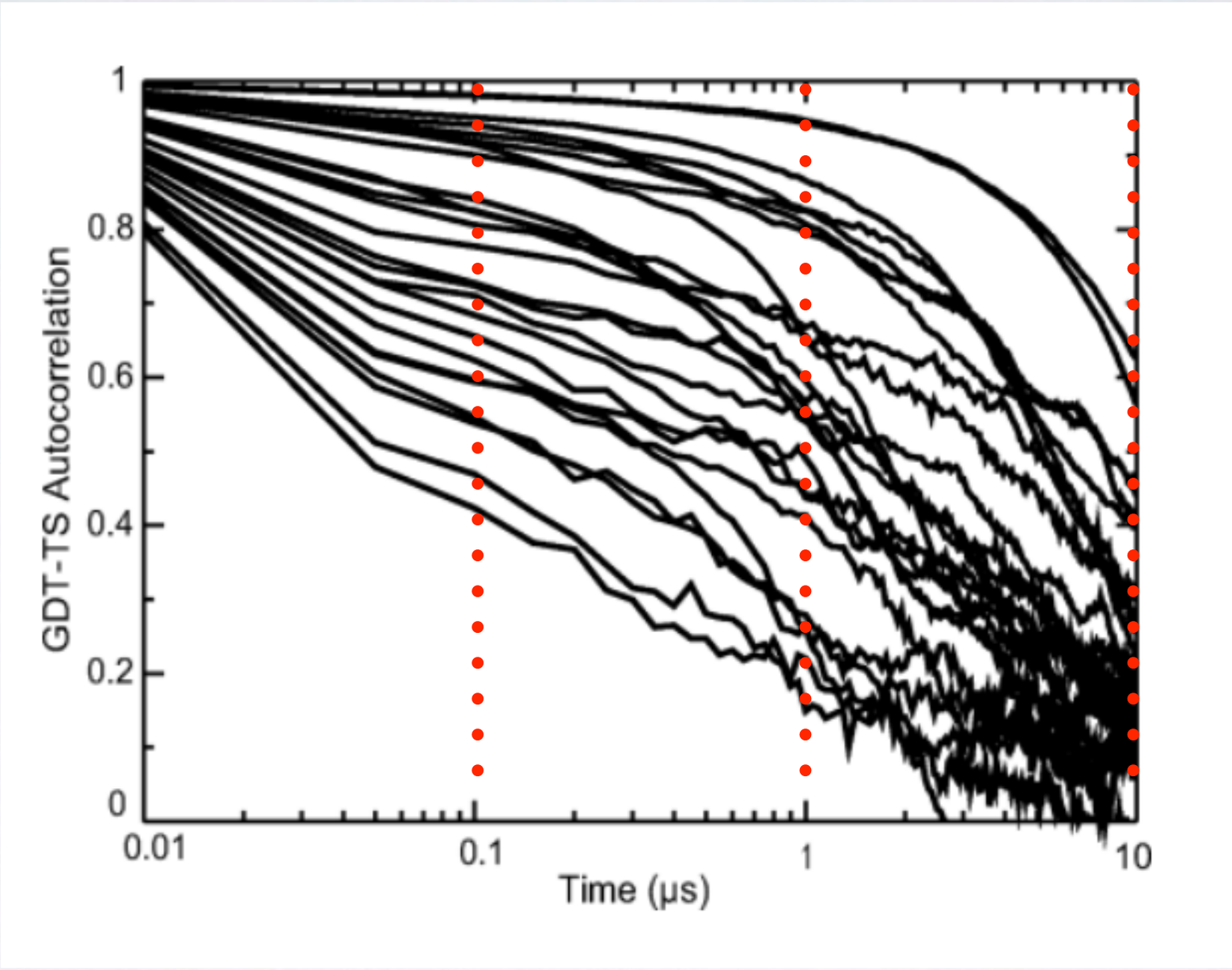
CASP (Critical Assessment of techniques
for protein Structure Prediction)

Community-wide experiment on the comparative evaluation
of protein structure prediction methods

held every two years since 1994

Refinement of protein structure homology models via long, all-atom molecular dynamics simulations

Closeness to the crystal structure from the crystallographic structure



David E. Shaw

Knowledge-based Potential in Protein Structure Prediction

Inverse-Boltzmann law

$$E(\mathbf{r}_i, \mathbf{r}_j) = -kT \log \frac{N(r_{ij})}{N_{\text{ref}}(r_{ij})}$$

N from the structure database

A popular type of free energy function used in structure prediction.

Stable states in the free energy surface

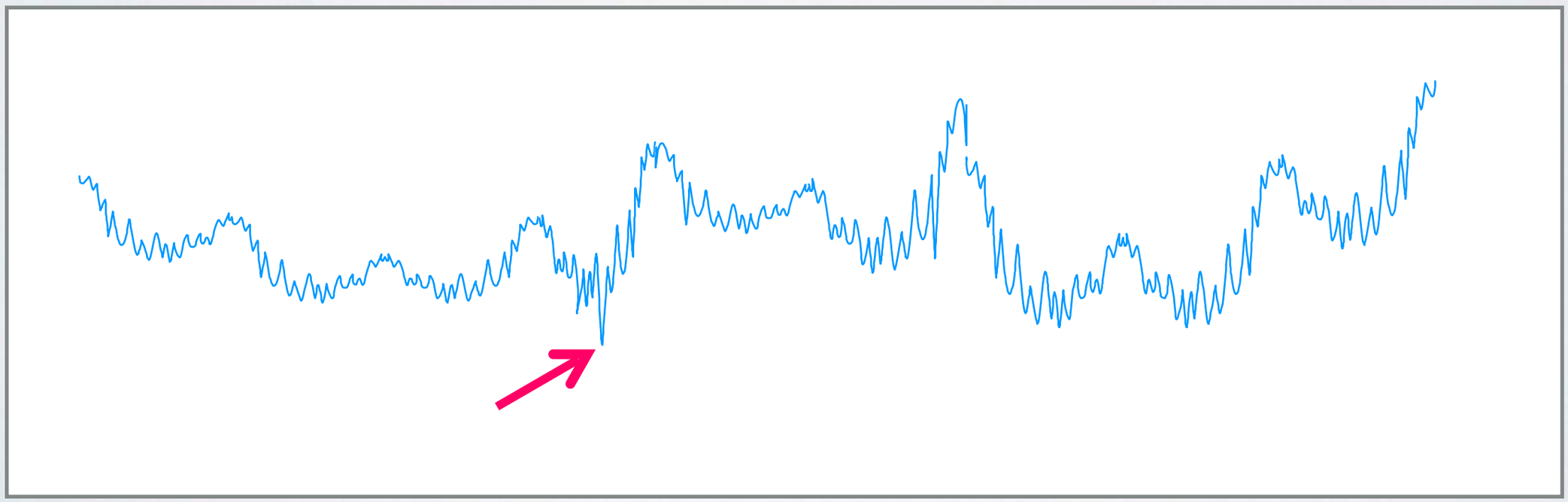
MD simulation

potential energy -> sampling -> distribution & free energy

Global optimisation

global minimum (& sometimes suboptimal minima),

but not (thermodynamic) distribution

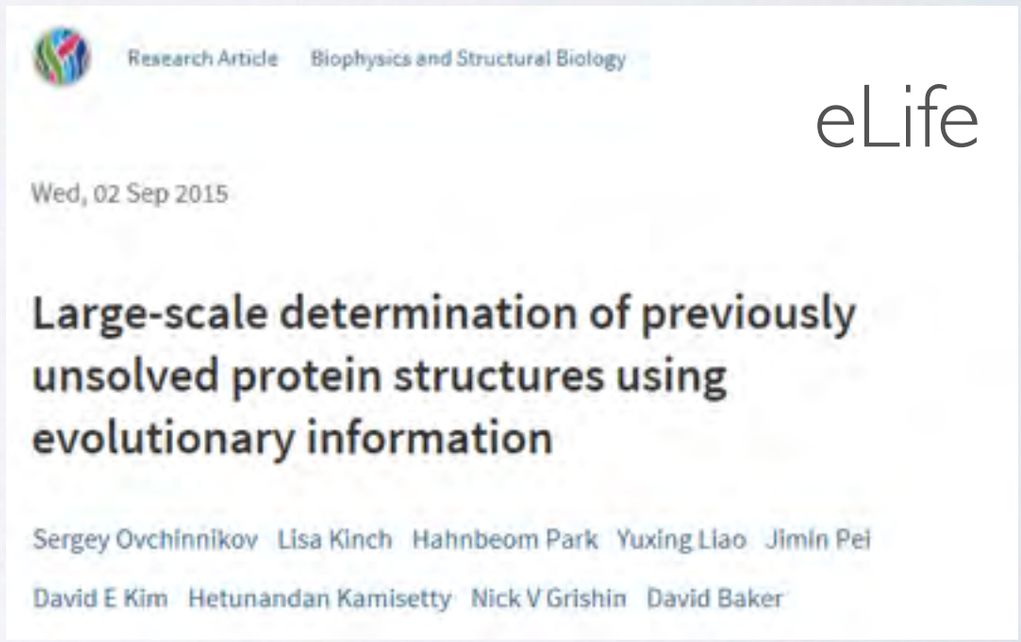


However, ab initio structure prediction is still extremely difficult

Reports on successful predictions for small proteins (~100aa)

Best predictors in CASP: Rosetta (Robetta), I-TASSER

Recent advance (by Rosetta)
extract contact information
from related sequences (GREMLIN)
->
restraint optimisation



Protein structure prediction

Ab initio prediction is usually inaccurate.

Homology modeling
(or Template-based modeling or Comparative modeling)
based on evolutionary principle

can be very accurate if close template(s) can be found.

Homology modeling

Ab initio prediction is usually inaccurate.

Homology modeling
(or Template-based modeling
or Comparative modeling)

based on evolutionary principle

can be very accurate if close template(s) can be found.

Procedure of homology modeling

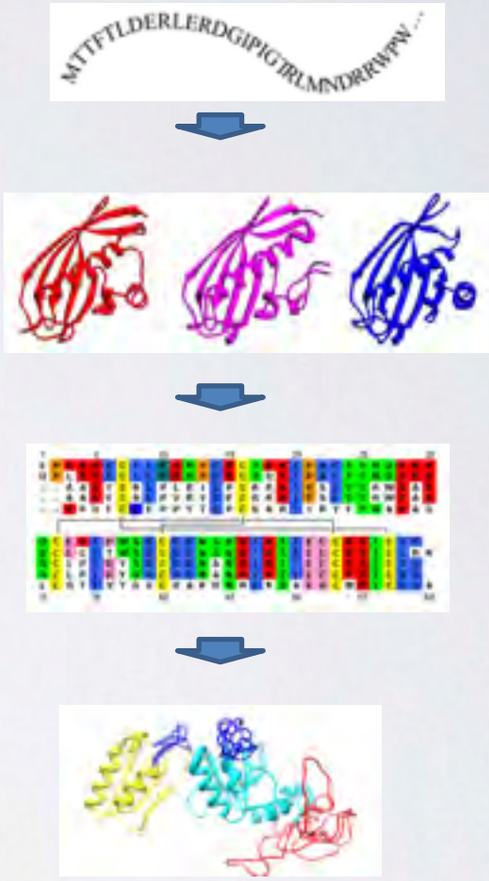
Domain parsing

-> Template selection

-> (Multiple) sequence alignment

-> Model building

-> Loop modeling, Refinement



Web servers for protein structure prediction

- I-TASSER
  - http://zhanglab.ccmb.med.umich.edu/I-TASSER/
- Robetta
  - http://robetta.bakerlab.org/
- HHpred
  - https://toolkit.tuebingen.mpg.de/hhpred
- SWISS-MODEL
  - http://swissmodel.expasy.org/

Example: IL-4

- Interleukine-4 in Sus scrofa (Pig)

- Involved in at least several B-cell activation processes as well as in other cell types.

- Its structure has not been determined yet,
  but structures of homologs in other organisms have been determined.

- http://www.uniprot.org/uniprot/F1CFE1

I-TASSER

- One of the best prediction servers in recent CASPs
- Automatic prediction server without any intermediate steps for the user
- Requires registration
- Predicts:
  - Protein tertiary structure
  - Ligand binding sites
  - Functions (GO, …)

I-TASSER

http://zhanglab.ccmb.med.umich.edu/I-TASSER/

I-TASSER

I-TASSER On-line Server (View an example of I-TASSER output):

Copy and paste your sequence below ([10, 1500] residues in FASTA format). Click here for a sample input:

>tr|F1CFE1|F1CFE1_PIG Interleukin-4 OS=Sus scrofa GN=IL-4 PE=2 SV=1
MGLTSQLIPTLVCLLACTSNFVHGHKCDITLQEIIKTLNILTARKNSCMELPVTDVFAAP
ENTTEKETFCRASTVLRHIYRHHTCMKSLLSGLDRNLSSMANMTCSVHEAKKSTLKDFLE
RLKTIMKEKYSKC

Paste amino acid sequence

Or upload the sequence from your local computer:

파일 선택 선택된 파일 없음

Email: (mandatory, where results will be sent to)

chaok@snu.ac.kr

Password: (mandatory, please click here if you do not have a password)

••••••••

Put your registered information

ID: (optional, your given name of the protein)

IL-4

▶ **Option I:** Assign additional restraints & templates to guide I-TASSER modeling.

▶ **Option II:** Exclude some templates from I-TASSER template library.

▶ **Option III:** Specify secondary structure for specific residues.

☑ Keep my results public (uncheck this box if you want to keep your job private. A key will be assigned for you to access the results)

Run I-TASSER Clear form

(Please submit a new job only after your old job is completed)

I-TASSER

The sequence has been successfully submitted to the I-TASSER server.

Your job id number is S279887. You will be notified by email once the job is completed if an email has been provided. The results of structure and function predictions for your submitted sequence will be available at (You may bookmark this link for your convenience of future visit): http://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S279887/.

Your submitted sequence is of 133 residues:
>IL-4
MGLTSQLIPTLVCLLACTSNFVHGHKCDITLQEIIKTLNILTARKNSCMELPVTDVFAAP
ENTTEKETFCRASTVLRHIYRHHTCMKSLLSGLDRNLSSMANMTCSVHEAKKSTLKDFLE
RLKTIMKEKYSKC

Back to Home
Back to I-TASSER server
Back to I-TASSER Queue

Bookmark this link
or
The server will sent an e-mail when the prediction is done.

It takes ~72 hours for a single prediction.

I-TASSER

- On the "Queue" page, you can track the job status.

ID	Protein Name	Length	C-score	Estimated TM-score	Estimated RMSD(Å)	Submission date	User's email address	User's IP
S279887	IL-4	133	NA	NA	NA	2016-06-24	xxx@snu.ac.kr	147.47.216.xxx

This job is pending in this queue.

ID	Protein Name	Length	C-score	Estimated TM-score	Estimated RMSD(Å)	Submission date	User's email address	User's IP
S279435	V144N	380	NA	NA	NA	2016-06-20	xxx@modares.ac.ir	194.225.168.xxx

This job is running and should be completed in approximately 35hrs.

ID	Protein Name	Length	C-score	Estimated TM-score	Estimated RMSD(Å)	Submission date	User's email address	User's IP
S279431	GLUTAREDOXIN	112	0.13	0.73±0.11	3.9±2.7	2016-06-20	xxx@live.utm.my	161.139.222.xxx

Model1 Model2 Model3 Model4 Model5

I-TASSER

- Secondary Structure / Solvent Accessibility Prediction

Predicted Secondary Structure

```
                    20        40        60        80        100
                    |         |         |         |         |
Sequence    MALAKAKETVASAPVVVYSKSYCPFCVRVKKLFEQLGATFKAIELDGESDGSELQSALAEWTGQRTVPNVFINGKHIGGCDDTLALNNEGKLVPLLTEAGAIASSAKTTITA
Prediction CHHHHHHHHHHCCCSSSSSCCCCHHHHHHHHHHHHHCCCCCSSSSCCCCCCHHHHHHHHHHHCCCCCCSSSSCCSSSCCCHHHHHHHHCCCHHHHHHCCCCCCHHHHHHCC
Conf.Score 958999999755999999899890799999999975999889982899987999999999858998193879999787729999999869839999885886420234314 9
           H:Helix; S:Strand; C:Coil
```

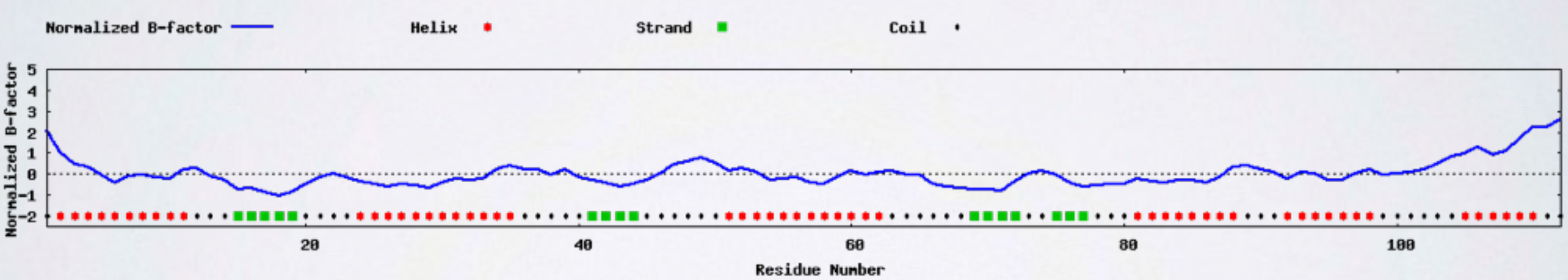
Predicted Solvent Accessibility

```
                    20        40        60        80        100
                    |         |         |         |         |
Sequence    MALAKAKETVASAPVVVYSKSYCPFCVRVKKLFEQLGATFKAIELDGESDGSELQSALAEWTGQRTVPNVFINGKHIGGCDDTLALNNEGKLVPLLTEAGAIASSAKTTITA
Prediction 72373046207734020102344320320241057361623313024474164025104736743210001044420013320330367441351056033324564443 68
           Values range from 0 (buried residue) to 9 (highly exposed residue)
```

I-TASSER

- Predicted normalized B-factor,
  which is related to per-residue model accuracy

Predicted normalized B-factor



I-TASSER

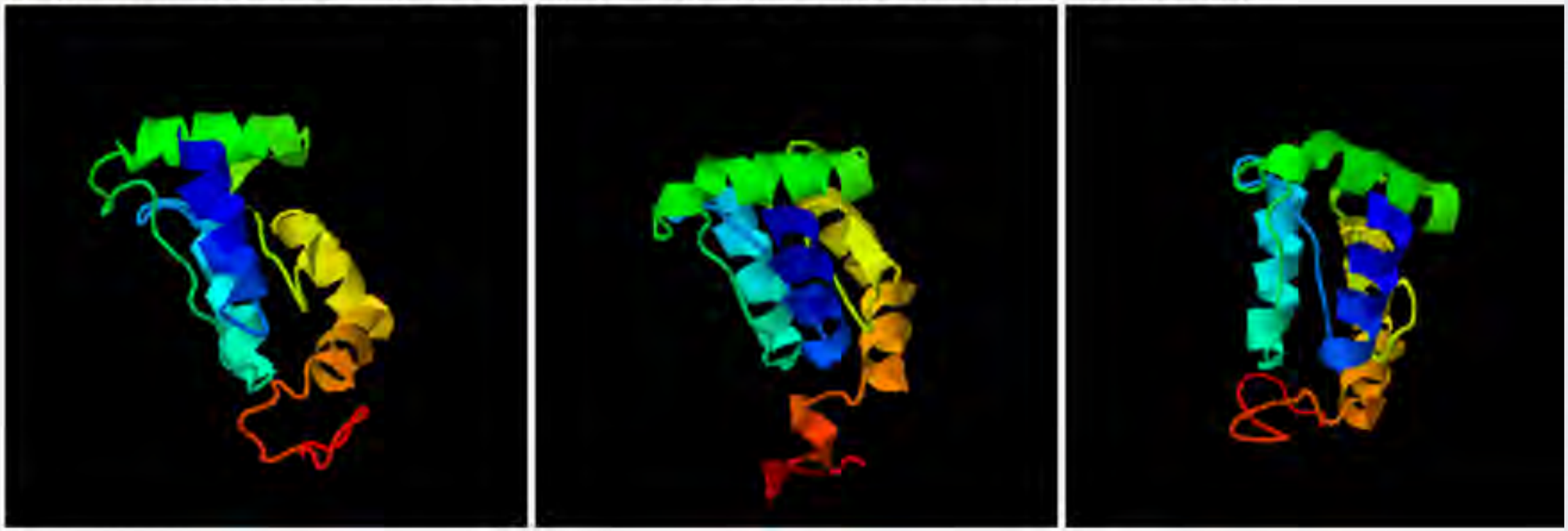
Top 5 final models predicted by I-TASSER

It provides up to 5 models

(For each target, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and reports up to five models which corresponds to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these qualities. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models as seen in our benchmark tests. If the I-TASSER simulations converge, it is possible to have less than 5 clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.)

- More about C-score
- Local structure accuracy profile of the top five models

(By right-click on the images, you can export image file or change the configurations, e.g. modifying the background color or stopping the spin of your models)



- Download Model 1
- C-score=0.13 (Read more about C-score)
- Estimated TM-score = 0.73±0.11
- Estimated RMSD = 3.9±2.7Å

- Download Model 2
- C-score = -1.07

- Download Model 3
- C-score = -1.08

It provides estimated model accuracies

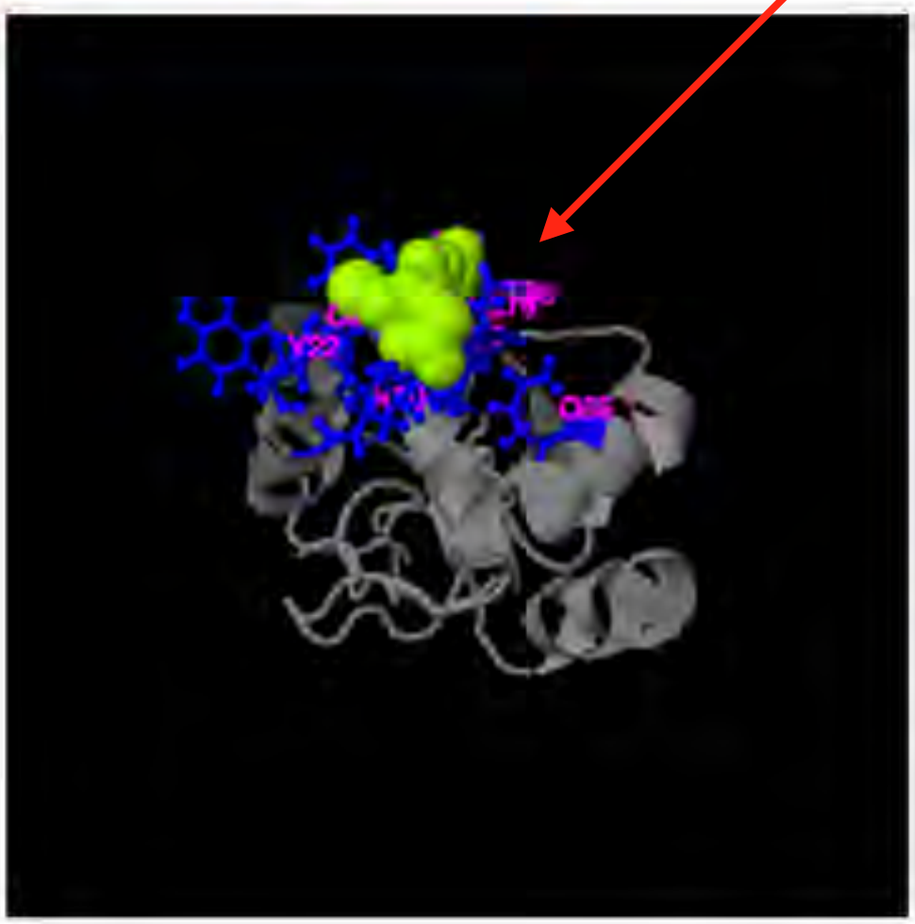
I-TASSER

It predicts ligand binding sites

Predicted function using COACH

(This section reports biological annotations of the target protein by COACH based on the I-TASSER structure prediction. COACH is a meta-server approach that combines multiple function annotation from the COFACTOR, TM-SITE and S-SITE programs.)

Ligand binding sites



Reset to initial orientation Spin On/Off

Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
●	1	0.95	194	2e7pC	GSH	Rep, Mult	20,22,23,25,55,66,67,68,79,80,81
	2	0.05	9	2yanB	GSH	Rep, Mult	20,22,23,25,55
	3	0.04	5	3fzaA	BME	Rep, Mult	25,28,80,81,84
	4	0.02	5	3rhcB	k-mer	N/A	22,23,24,25
	5	0.02	4	3mszA	GSH	Rep, Mult	25,66,68,79,80,81,82

Download the residue-specific ligand binding probability, which is estimated by SVM.
Download the all possible binding ligands and detailed prediction summary.
Download the templates clustering results.

(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a
(b) **Cluster size** is the total number of templates in a cluster.
(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the BioLiP databas
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in t
Mult is the complex structures with all potential binding ligands in the cluster.

I-TASSER

It also predicts GO terms and their reliability

Gene Ontology (GO) terms

Top 10 homologous GO templates in PDB

Rank	Cscore^{GO}	TM-score	RMSD^{a}	IDEN^{a}	Cov	PDB Hit	Associated GO Terms
1	0.41	0.8106	2.22	0.29	0.96	2x8cB	GO:0000166 GO:0016491 GO:0046872 GO:0055114 GO:0004791 GO:0050660 GO:0050661
2	0.37	0.8226	1.07	0.46	0.88	3rhcA	GO:0009055 GO:0015035 GO:0045454
3	0.37	0.8763	0.99	0.49	0.92	3fz9A	GO:0009055 GO:0015035 GO:0045454
4	0.36	0.9294	0.94	0.57	0.97	1z7pA	GO:0051536 GO:0051537 GO:0046872 GO:0009055 GO:0015035
5	0.36	0.8325	1.27	0.40	0.89	3h8qA	GO:0004791 GO:0009055 GO:0015035 GO:0016668 GO:0045454
6	0.36	0.8474	1.09	0.41	0.90	3d5jA	GO:0009055 GO:0015035 GO:0045454
7	0.35	0.8097	1.67	0.36	0.91	1kteA	GO:0005737 GO:0006810 GO:0022900 GO:0009055 GO:0015035
8	0.35	0.8078	1.67	0.31	0.91	1b4qA	GO:0005737 GO:0005829 GO:0022900 GO:0015949 GO:0015038 GO:0015035
9	0.35	0.8444	1.09	0.39	0.90	3c1rA	GO:0004364 GO:0022900 GO:0004602 GO:0005634 GO:0010731 GO:0015035
10	0.34	0.8135	1.53	0.26	0.88	3ipzA	GO:0009055 GO:0015035 GO:0045454

Consensus prediction of GO terms

Molecular Function	GO:0009055	GO:0015035	GO:0050660	GO:0004791	GO:0050661	GO:0046872	GO:0051537
GO-Score	0.91	0.91	0.62	0.62	0.62	0.62	0.36
Biological Process	GO:0045454	GO:0055114					
GO-Score	0.91	0.62					
Cellular Component	GO:0005737						
GO-Score	0.41						

Robetta

- One of the best prediction servers in recent CASPs
- Requires registration
- Automatic prediction server with **a single** pause for the user
  - It predicts domains first.
  - And you have to re-submit a prediction job for each domain.
- You can run one prediction job at a time.

Robetta

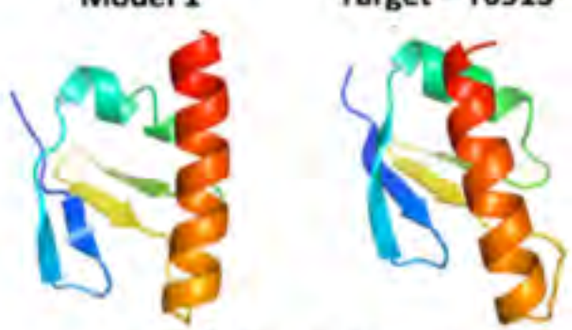
http://robetta.bakerlab.org/

www.bakerlab.org

ROBETTA BETA

Full-chain Protein Structure Prediction Server

Model 1 Target – T0513



2.66 Å over 62 residues

0.84 Å over 39 residues

de novo prediction by Robetta in CASP-8

REGISTRATION

[**Register / Update**] [**Login**]

DOCUMENTATION

[**Docs / FAQs**]

SERVICES

Domain Parsing & 3-D Modeling
[**Queue**] [**Submit**]

Interface Alanine Scanning
[**Queue**] [**Submit**]

Fragment Libraries
[**Queue**] [**Submit**]

DNA Interface Residue Scanning
[**Queue**] [**Submit**]

RELATED SITES

Rosetta Commons
Rosetta Commons ROSIE server *NEW*
RosettaBackrub Server
RosettaDesign Server
FoldIt
Rosetta@home
Human Proteome Folding Project
Rosetta@Cloud

Robetta

Submit a job to the Server

If you submit more than one job using different logins, the jobs will be deleted and the IP may be locked out.

Required

Prediction Type: Ginzu : Domain Prediction
Structure : 3-D Model (available per domain after Ginzu completes from results page)

Registered Username: chaok **or** Registered Email Address:

Target Name:
IL-4

Paste Fasta (AA sequence only!!) TRANSLATE DNA TO AA

```
>tr|F1CFE1|F1CFE1_PIG Interleukin-4 OS=Sus scrofa GN=IL-4 PE=2 SV=1
MGLTSQLIPTLVCLLACTSNFVHGHKCDITLQEIIKTLNILTARKNSCMELPVTDVFAAP
ENTTEKETFCRASTVLRHIYRHHTCMKSLLSGLDRNLSSMANMTCSVHE[illegible]RKSTEKDFLE
RLKTIMKEKYSKC
```

or Upload Fasta: 파일 선택 선택된 파일 없음

Put your registered information

Paste amino acid sequence

Optional

Reply Email:

Do not warn me if my sequence matches one already submitted

Note: please do not submit known PDB sequences or CASP targets intentionally

Submit

It takes ~72 hours for a single prediction.

Robetta

- On the "Queue" page, you can track the job status.

Num	ID	Status	Method	Username	Target
1	66648	Queued	Ginzu	bawer178	TodS
2	66647	Queued	Ginzu	meghau	H_doris_luciferase_4
3	66646	Queued	Ginzu	bo15mtech11003	tdp43-2c
4	66645	Queued	Ginzu	bootpp	ctClpP1
5	66644	Queued	Ginzu	kellyj2	H_melpomene_luciferase_7
6	66643	Queued	Ginzu	a.jahangiri	Omp34-no signal
7	66642	Active	Ginzu	iphan	rickA
8	66641	Queued	Ginzu	amishra2	1NYC
9	66640	Queued	Ginzu	emte1992	hhhh
10	66639	Active	Ginzu	divnand	ORF 66
11	66638	Active	Ginzu	biotcorp	haru1
12	66637	Complete	Ginzu	johnjames	set
13	66636	Complete	Ginzu	jaboyer	ml28a
14	66635	Complete	Ginzu	efischer	kiaa1524
15	66634	Complete	Ginzu	ekaterinakarpova	NEP1.5
16	66633	Complete	Ginzu	yuanjiaqi273	NCX1_Cyto1
17	66629	Complete	Ginzu	anfa	IgE
18	66628	Complete	Ginzu	jaobrien	PA2B3_BUNFA
19	66627	Complete	Ginzu	zhangxin	zx
20	66626	Active	Structure	dkhago	Eueides_isabella_luciferase_1
21	66625	Active	Structure	rwmartin	D_plexippus_DPOGS206945
22	66624	Complete	Ginzu	amakiriaustin	medin
23	66623	Active	Ginzu	renata.tisi@unimib.it	ScRad50
24	66622	Complete	Ginzu	sprajap	LD3
25	66621	Queued	Structure	itoni	hc_full

Status

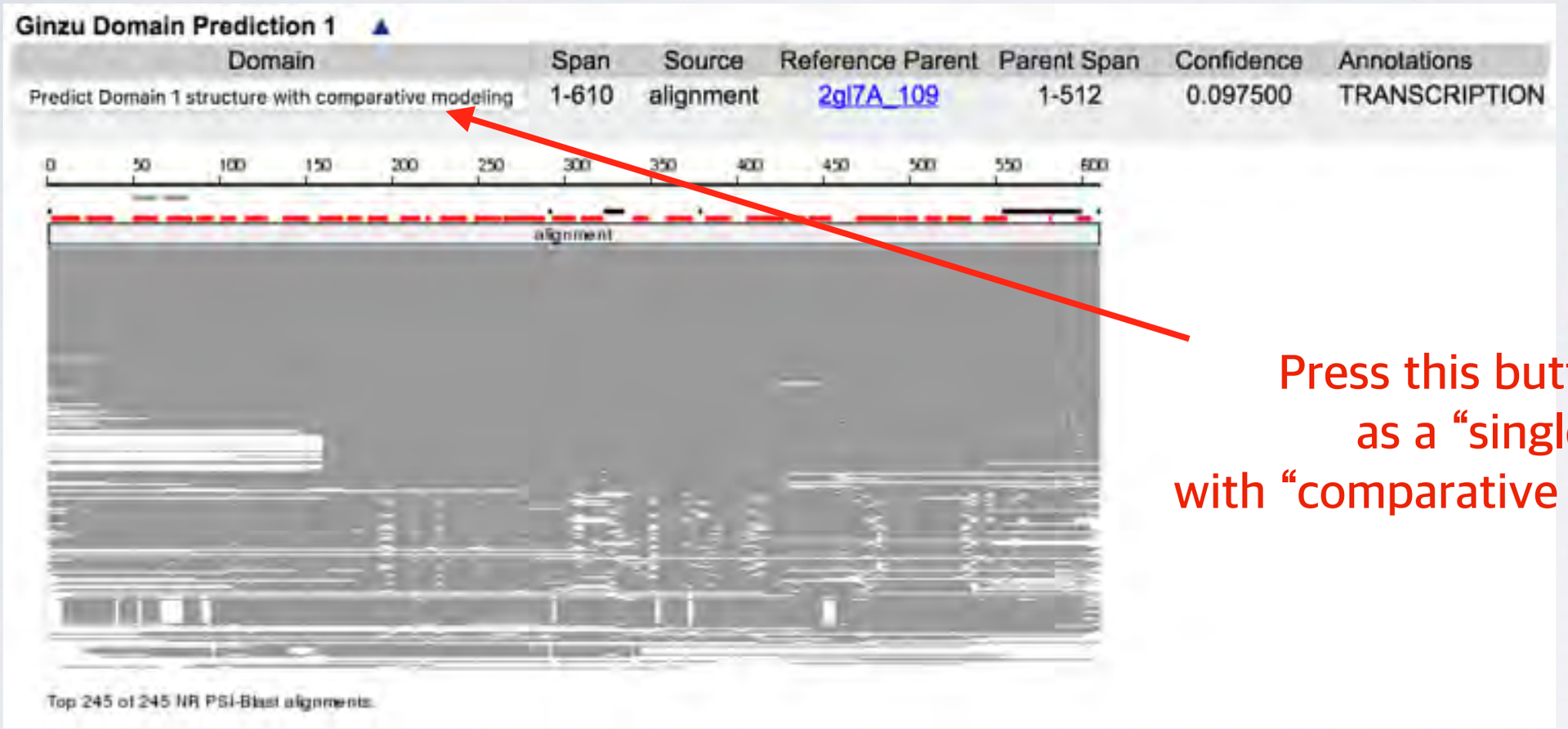
Queued: waiting for running
Active: running now
Complete: finished

Method

Ginzu: domain prediction stage

Structure:
tertiary structure prediction stage

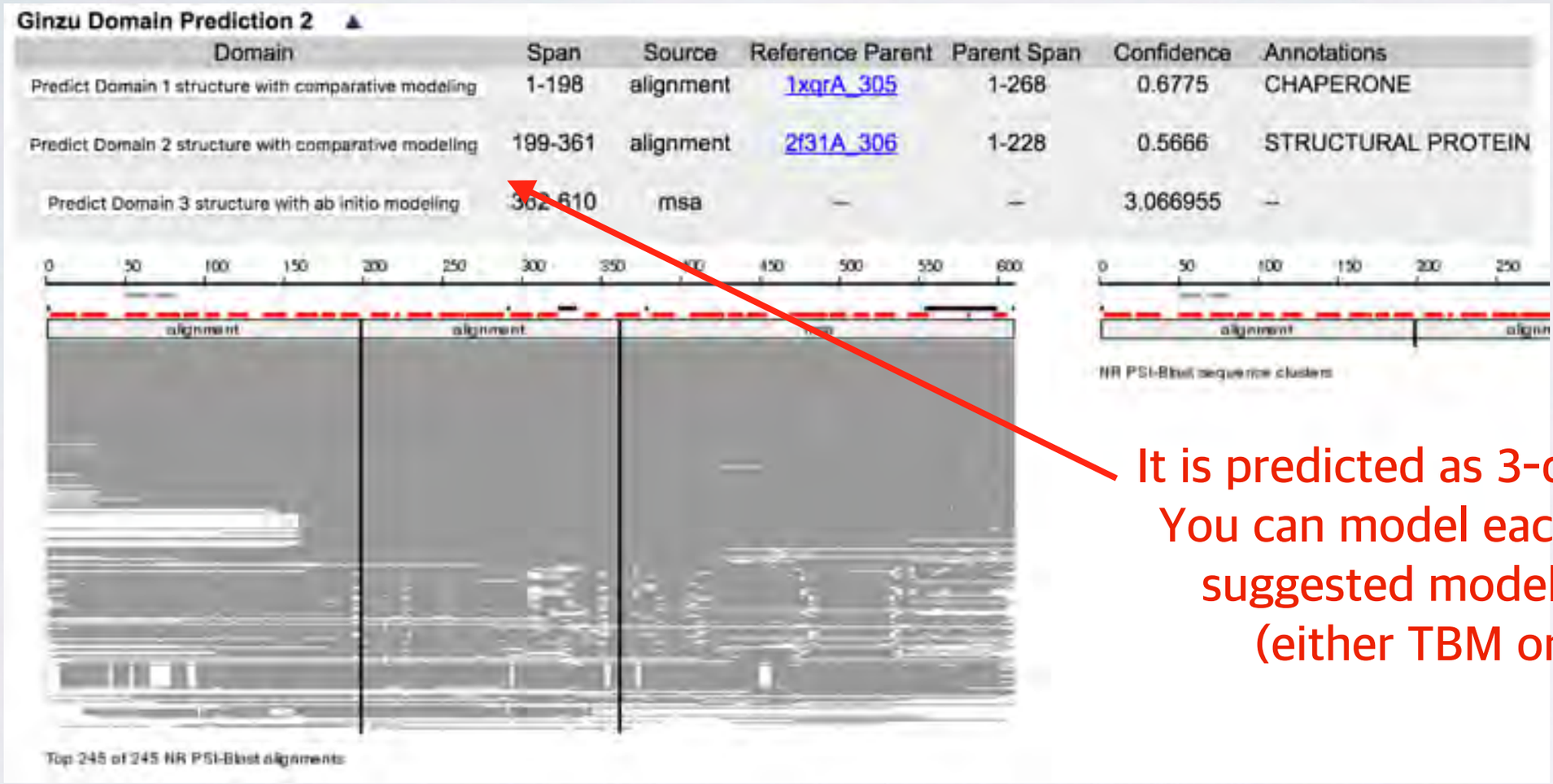
Robetta



Press this button to predict
as a “single” domain
with “comparative modeling” method

Robetta

- Once you got "domain prediction" results,
  you can choose domains to predict 3D structures



Ginzu Domain Prediction 2

Domain	Span	Source	Reference Parent	Parent Span	Confidence	Annotations
Predict Domain 1 structure with comparative modeling	1-198	alignment	1xqrA_305	1-268	0.6775	CHAPERONE
Predict Domain 2 structure with comparative modeling	199-361	alignment	2f31A_306	1-228	0.5666	STRUCTURAL PROTEIN
Predict Domain 3 structure with ab initio modeling	362-610	msa	–	–	3.066955	–

It is predicted as 3-domain protein.
You can model each domain with
suggested modeling method.
(either TBM or ab initio)

Robetta

It provides up to 5 models and related predicted features.

Full Structure Predictions

Click the icons below each image to download the prediction in PDB format (PDB)

Features and Secondary Structure

HHpred

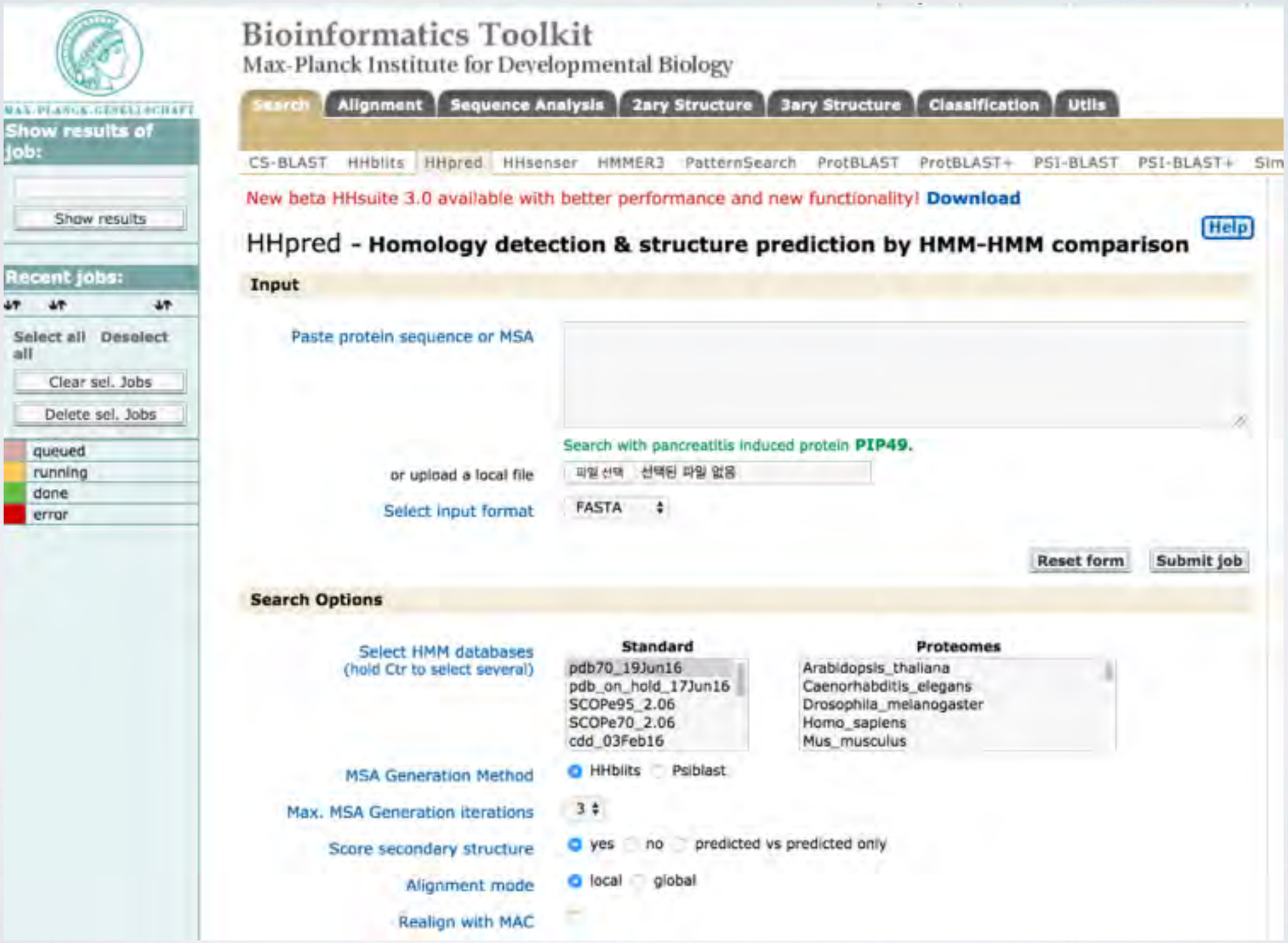
- One of the best prediction servers in recent CASPs that are the fastest.

- It does NOT require registration,
  but requires a MODELLER license key to build 3D model.

- It provides an interactive prediction mode
  and also provides automatic selection options.

HHpred

http://toolkit.tuebingen.mpg.de/hhpred



HHpred

Paste amino acid sequence

HHpred - Homology detection & structure prediction by HMM-HMM comparison

Help

Input

Paste protein sequence or MSA

```
>tr|F1CFE1|F1CFE1_PIG Interleukin-4 OS=Sus scrofa GN=IL-4 PE=2 SV=1
MGLTSQLIPTLVCLLACTSNFVHGHKCDITLQEIIKTLNILTARKNSCMELPVTDVFAAP
ENTTEKETFCRASTVLRHIYRHHTCMKSLLSGLDRNLSSMANMTCSVHEAKKSTLKDFLE
RLKTIMKEKYSKC
```

Search with pancreatitis induced protein **PIP49.**

or upload a local file — 파일 선택 선택된 파일 없음

Select input format — FASTA

Reset form | Submit job

Search Options

Select HMM databases (hold Ctr to select several)

Standard	Proteomes
pdb70_19Jun16	Arabidopsis_thaliana
pdb_on_hold_17Jun16	Caenorhabditis_elegans
SCOPe95_2.06	Drosophila_melanogaster
SCOPe70_2.06	Homo_sapiens
cdd_03Feb16	Mus_musculus

MSA Generation Method — HHblits / Psiblast

Max. MSA Generation iterations — 3

Score secondary structure — yes / no / predicted vs predicted only

Alignment mode — local / global

Realign with MAC — ☑

Check this for the better sequence alignment

HHpred

This bar shows aligned regions for each templates
(red: reliable matches; blue: unreliable matches)

Press this menu for 3D model building

HHpred - Results Job-ID: 5842525 Date: 15:50 on Jun 24 2016

Help

Results | Histogram | Reduced alignment | Representative alignment | Full alignment

Create model | Merge Q/T alignments | Forward to PColls | Forward MSA | Save | Export

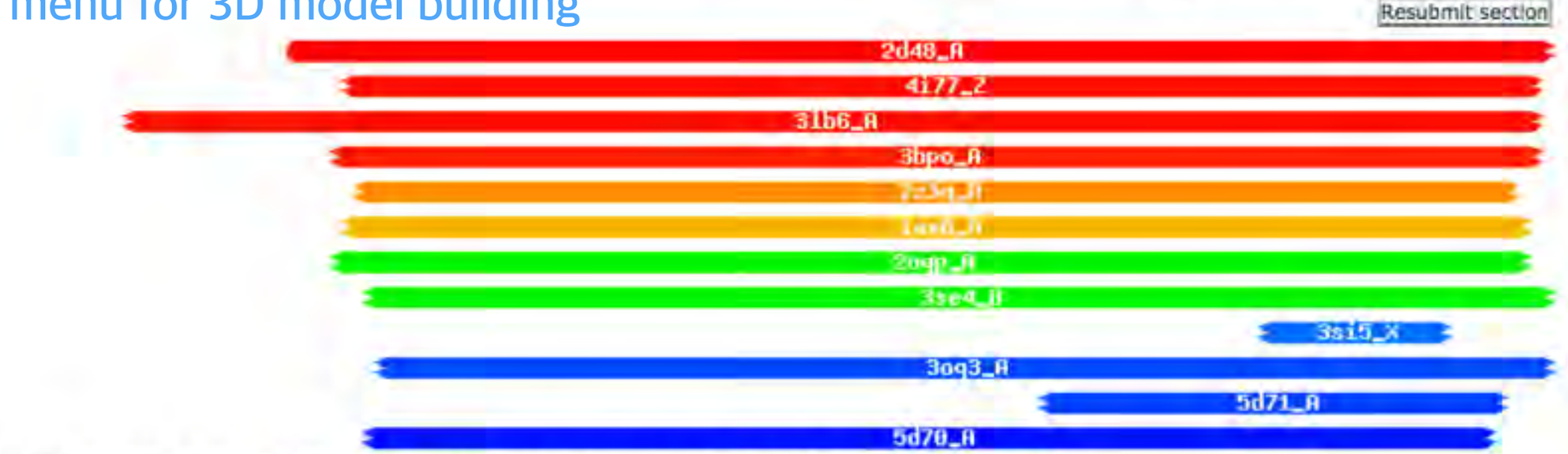
Color alignments

color only SS | color alignments | color alignments

View Alignment

23 … 100 … 133

Resubmit section



Query tr|F1CFE1|F1CFE1_PIG (seq=MGLTSQLIPT...TIMKEKYSKC Len=133 Neff=4.3 Nseqs=85)
Parameters score SS:yes search:local realign with MAP:yes

No	Hit		Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	2d48_A	Interleukin-4; four hel	100.0	3E-54	8E-59	324.9	10.7	109	23-133	1-127 (129)
2	4i77_B	Interleukin-13, IL-13;	98.0	7.2E-05	1.9E-09	55.3	9.2	93	28-132	6-108 (112)
3	3lb6_A	Interleukin-13, IL-13;	97.9	0.00011	3.1E-09	55.4	9.5	112	9-132	5-128 (132)
4	3bpo_A	Interleukin 13; IL4, IL	96.9	0.0027	7.3E-08	47.7	6.3	95	27-132	6-110 (127)
5	2x3q_A	Interleukin-15, IL-15;	88.1	5.9	0.00016	28.9	9.0	90	29-130	5-116 (119)
6	1ax8_A	Obesity protein, leptin	84.4	3.9	0.00011	31.5	6.6	104	28-131	11-138 (146)
7	2oqp_A	Interleukin-21, IL-21,	72.3	4.9	0.00013	30.2	3.8	94	27-131	7-121 (134)
8	3se4_B	Interferon omega-1; typ	71.8	27	0.00074	26.9	8.1	97	30-133	15-144 (177)
9	3si5_X	Protein CASC5; BUBR1-bi	49.5							
10	3oq3_A	IFN-alpha-5, interferon	46.8							

Detailed template detection informations

HHpred

Manual template selection

Automatic template selection

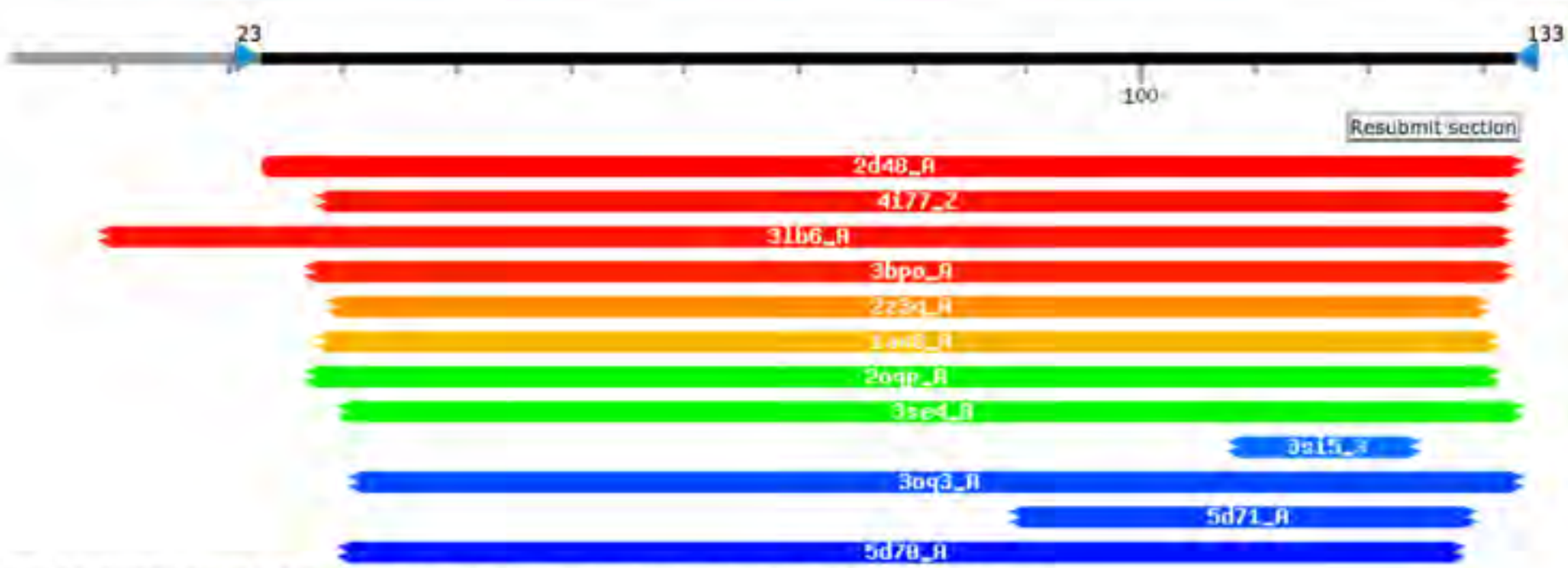
Results | Histogram | Reduced alignment | Representative alignment | Full alignment

Create model | Merge Q/T alignments | Forward to PColis | Forward MSA | Save | Export

1. **Manual template selection: Generate a PIR-alignment of your sequence with the selected template or templates in order to build a 3D model with MODELLER.** [Create model frommanually selected te]

2. **Automatic template selection: Optimize diversities of query and template HMMs, rerank templates and automatically select best set. In further steps a multiple alignment is created from this set, and a 3D model is build with MODELLER using this alignment.** [Automatically select best templates]

Only hits found in PDB or SCOP can be used to create a model (other hits are disabled).



Query tr|F1CFE1|F1CFE1_PIG (seq=MGLTSQLIPT...TIMKEKYSKC Len=133 Neff=4.3 Nseqs=85)
Parameters score SS:yes search:local realign with MAP:yes

	No	Hit		Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
☑	1	2d48_A	Interleukin-4; four hel	100.0	3E-54	8E-59	324.9	10.7	109	23-133	1-127 (129)
	2	4i77_Z	Interleukin-13, IL-13;	98.0	7.2E-05	1.9E-09	55.3	9.2	93	28-132	6-108 (112)
	3	3lb6_A	Interleukin-13, IL-13;	97.9	0.00011	3.1E-09	55.4	9.5	112	9-132	5-128 (132)

HHpred

HHpred makemodel - Results Job-ID: 5842525_1 Date: 15:58 on Jun 24 2016 Help

Results | Histogram | Show query alignment

Select templates ◉ optimal single template ○ optimal multiple templates ○ user-defined

You can build your model using an optimal single template per query sequence domain (HHpred4 in CASP8), optimal multiple templates (HHpred5 in CASP8) or your own selected templates. We recommend not to select more than 5 template matches per query sequence domain.

The hit list below may contain several matches per template with similar alignments. These differ by the diversity (thickness) of the query and template HMMs used to generate the HMM-HMM alignment. A neural network is used to predict the expected TMScore for each alignment (column 6) and template alignments are reranked by this score. (TMScore &isin [0,1]; TMScore &ge 0.4 corresponds to meaningful predictions.)

Press "Generate alignment for MODELLER" to generate an PIR-alignment from the (automatically or manually) selected alignments.

Generate alignment for MODELLER

Press this button to generate sequence alignment between query and selected templates

HHpred

Generated sequence alignment is placed at here automatically

Modeller

Help

Input

Paste multiple sequence alignment [Target sequence + template(s)]

```
>P1;tr|F1CFE1|F1CFE1_PIG
sequence:tr|F1CFE1|F1CFE1_PIG:  1: : 133: :Interleukin-4 OS=Sus scrofa GN=IL-4
MGLTSQLIPTLVCLLACTSNFVHGHKCDITLQEIIKTLNILTARKNSCMELPVTDVFAAPENTTEKETFCRASTVLRHI
-----LLSGLDRNLSSMANM-TCSVHEAKKSTLKDFLERLKTIMKEKYSKC*
>P1;2d48_A
structureX:2d48_A: :A: :A:Interleukin-4; four helix bundle, cytokine; 1.65A {Ho
----------------------HKC--DITLQEIIKDLNSLTEQKTLCTELTVTDIFAASKNTTEKETFCRAATVLRQF
KQLIRFLKRLDRNLWGLAGLNSCPVKEANQSTLENFLERLKTIMREKYSKC*
```

The first sequence must be the target, the other sequences serve as templates. Sequence templates have to be preceeded by a PDB- or SCOP-identifier. **Paste example sequence.**

or upload a local file: 파일 선택 선택된 파일 없음

Select input format: PIR/NBRF

Options

Enter own alignment identifier: (must be the identifier from the alignment above!)

upload PDB files: 파일 선택 선택된 파일 없음

More uploads Add additional upload possibilities

The MODELLER license requires you to enter a MODELLER-key!

Please insert your MODELLER-key

Put your MODELLER license key

Job Options

Job-ID: (leave empty for automatic creation)

Send notification to (optional)

New form | Restore input | Submit job

HHpred

You can see the model structure on web browser

Modeller - Job-ID: 1515044 Date: 16:05 on Jun 24 2016

PDB-file | View 3D structure | Quality by VERIFY3D | Quality by SOLVX | Quality by ANOLEA

Save Export

Back to HHpred results to view query-template 3D-superpositions

```
EXPDTA    THEORETICAL MODEL, MODELLER 9.16 2016/06/24 16:06:06
REMARK    6 MODELLER OBJECTIVE FUNCTION:       983.4588
REMARK    6 MODELLER BEST TEMPLATE % SEQ ID:  55.906
REMARK    6 SEQUENCE: 1515044_temp
REMARK    6 TEMPLATE: 2d48
ATOM      1  N   MET     1      44.056  58.247   2.669  1.00 88.00           N
ATOM      2  CA  MET     1      43.210  58.884   3.702  1.00 88.00           C
ATOM      3  CB  MET     1      41.782  58.316   3.644  1.00 88.00           C
ATOM      4  CG  MET     1      40.784  59.076   4.519  1.00 88.00           C
ATOM      5  SD  MET     1      39.056  58.575   4.269  1.00 88.00           S
ATOM      6  CE  MET     1      38.932  59.312   2.614  1.00 88.00           C
ATOM      7  C   MET     1      43.780  58.633   5.056  1.00 88.00           C
ATOM      8  O   MET     1      44.869  59.105   5.379  1.00 88.00           O
ATOM      9  N   GLY     2      43.053  57.867   5.891  1.00 67.36           N
ATOM     10  CA  GLY     2      43.543  57.600   7.209  1.00 67.36           C
ATOM     11  C   GLY     2      44.739  56.722   7.064  1.00 67.36           C
ATOM     12  O   GLY     2      44.849  55.951   6.112  1.00 67.36           O
ATOM     13  N   LEU     3      45.674  56.821   8.028  1.00191.44           N
```

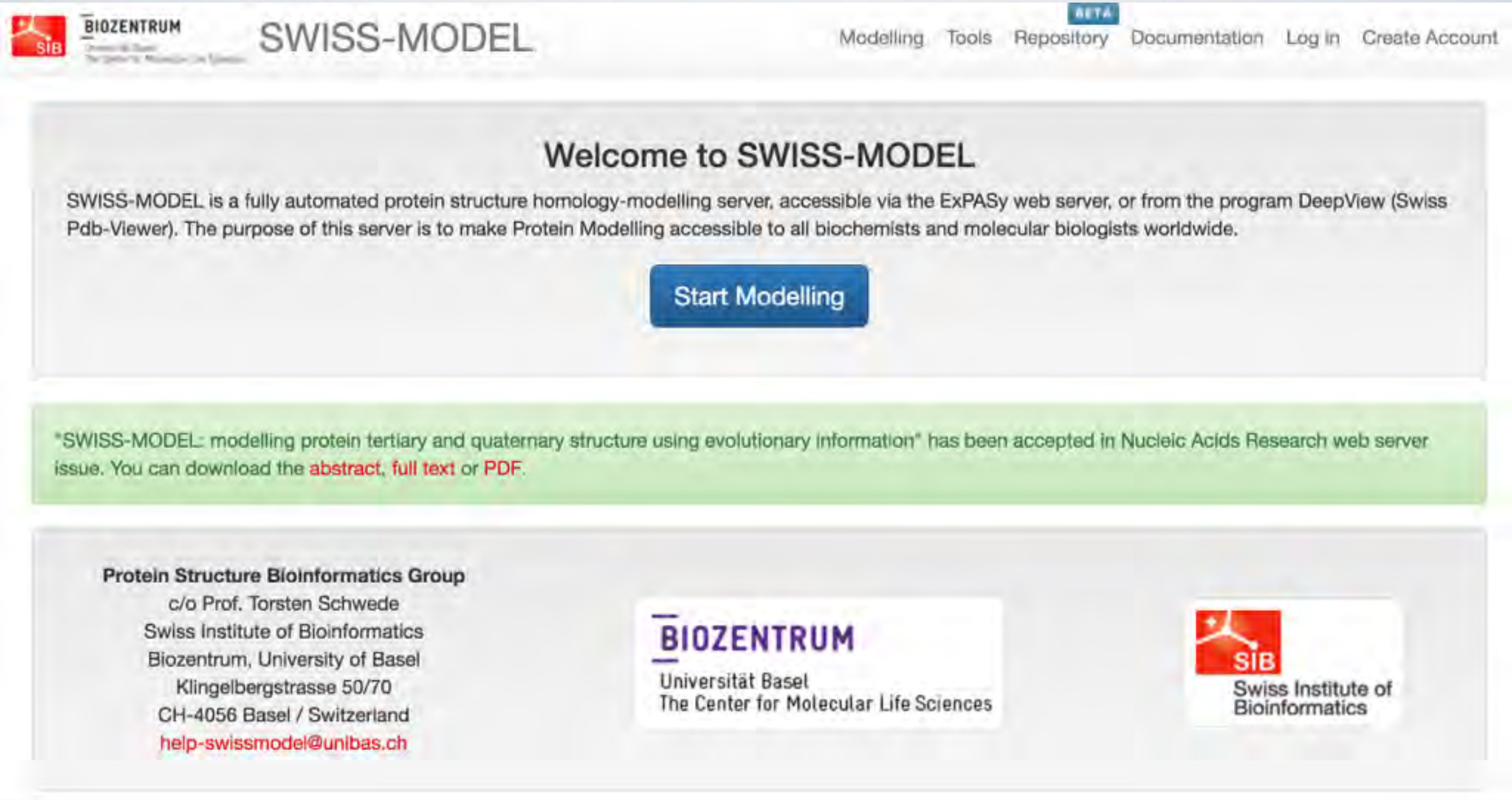
You can predict model accuracies

SWISS-MODEL

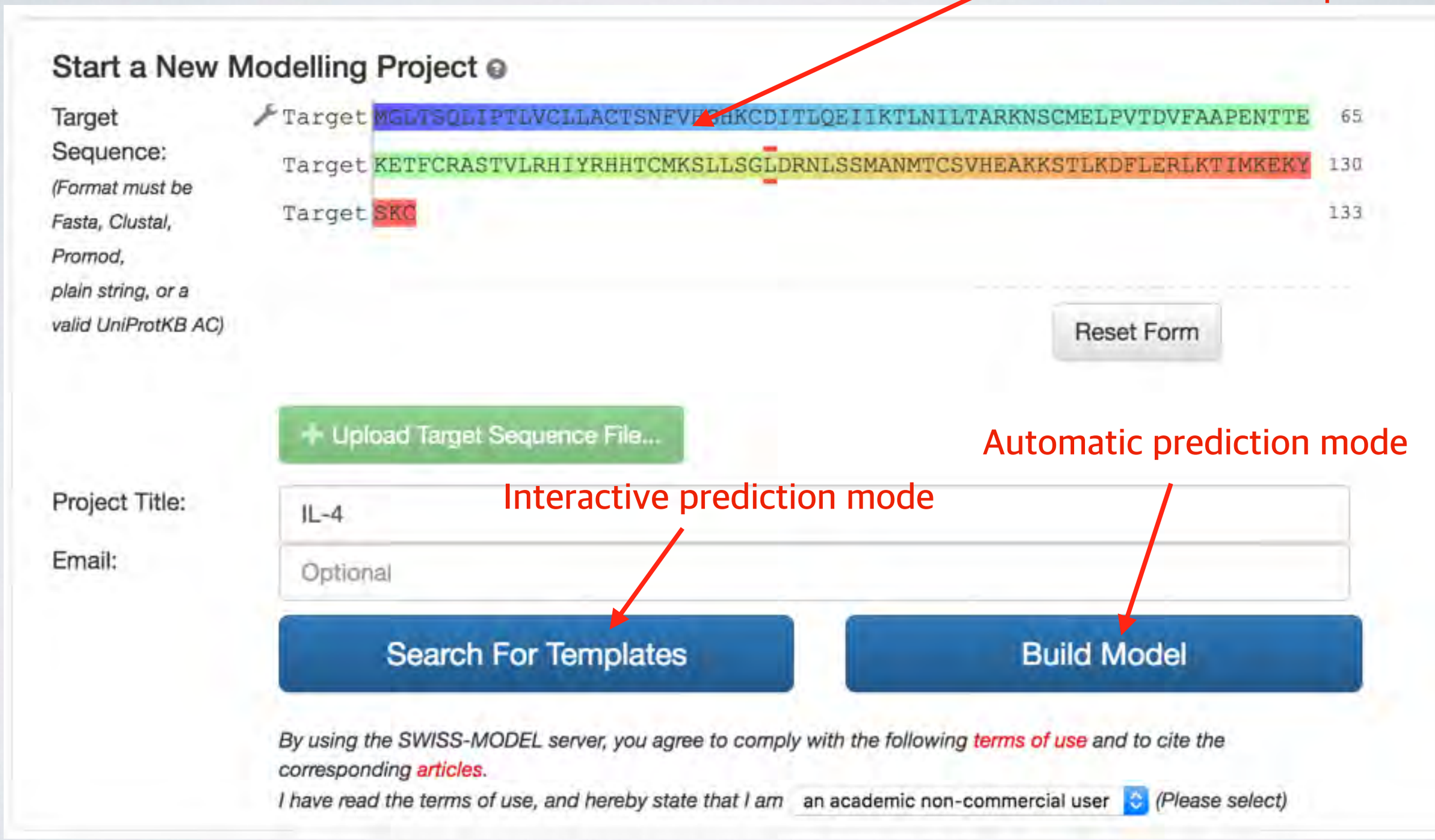
- It does NOT require registration.
- It provides both interactive and automatic predictions.
- It uses only single templates for each prediction.

SWISS-MODEL

http://swissmodel.expasy.org/



SWISS-MODEL



Both modes take few minutes for a single prediction.

SWISS-MODEL

- Interactive prediction mode

You can select templates to build 3D models

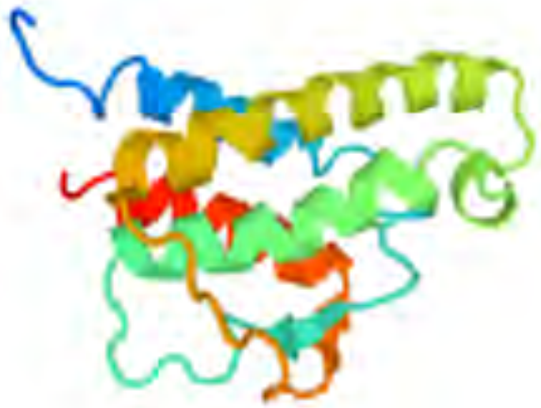
Template Results

Templates | Sequence Similarity | Alignment of Selected Templates | More

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
1bcn.1.A	INTERLEUKIN-4		66.07	NMR	monomer	None
1bbn.1.A	INTERLEUKIN-4		66.07	NMR	monomer	None
1itl.1.A	INTERLEUKIN-4		66.07	NMR	monomer	None
2b8z.1.A	Interleukin-4		68.81	X-ray, 2.5Å	monomer	None
2b91.1.A	Interleukin-4		68.81	X-ray, 2.0Å	monomer	None
2cyk.1.A	INTERLEUKIN-4		68.81	NMR	monomer	None
2b8u.1.A	Interleukin-4		68.81	X-ray, 1.8Å	monomer	None
1cyl.1.A	INTERLEUKIN-4		68.81	NMR	monomer	None
1iar.1.A	PROTEIN (INTERLEUKIN-4)		68.81	X-ray, 2.3Å	hetero-oligomer	None
3bpn.1.A	Interleukin-4		68.81	X-ray, 3.0Å	hetero-oligomer	2 x NAG
2b8x.1.A	Interleukin-4		68.81	X-ray, 1.7Å	monomer	None
2b90.1.A	Interleukin-4		67.89	X-ray, 2.1Å	monomer	None
1hij.1.A	INTERLEUKIN-4		67.89	X-ray, 3.0Å	homo-dimer	None
2d48.1.A	Interleukin-4		67.89	X-ray, 1.6Å	monomer	None

Build Models 1

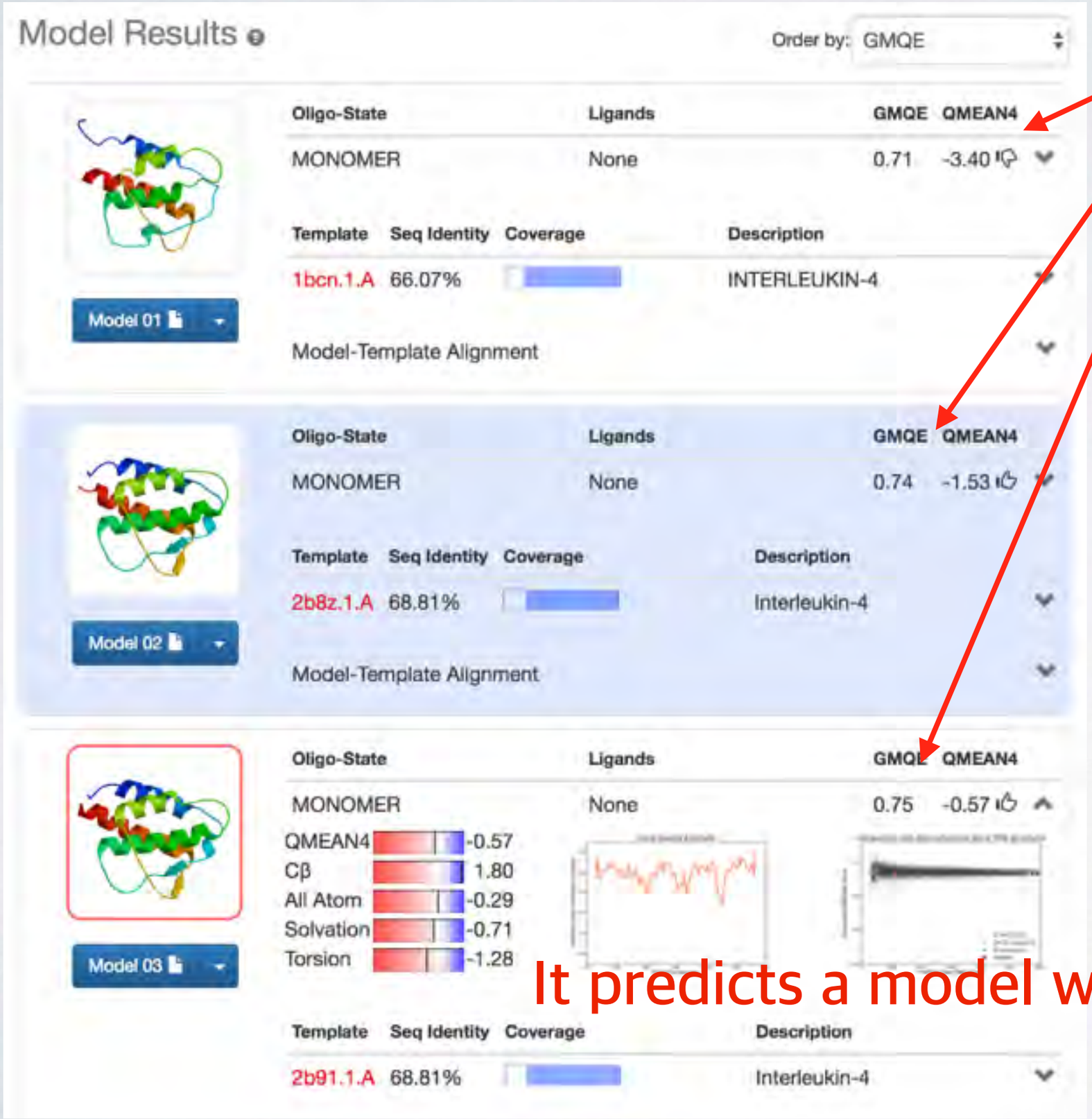
Clear Selection



View | pv

1bcn.1.A

SWISS-MODEL



Here are predicted model accuracies

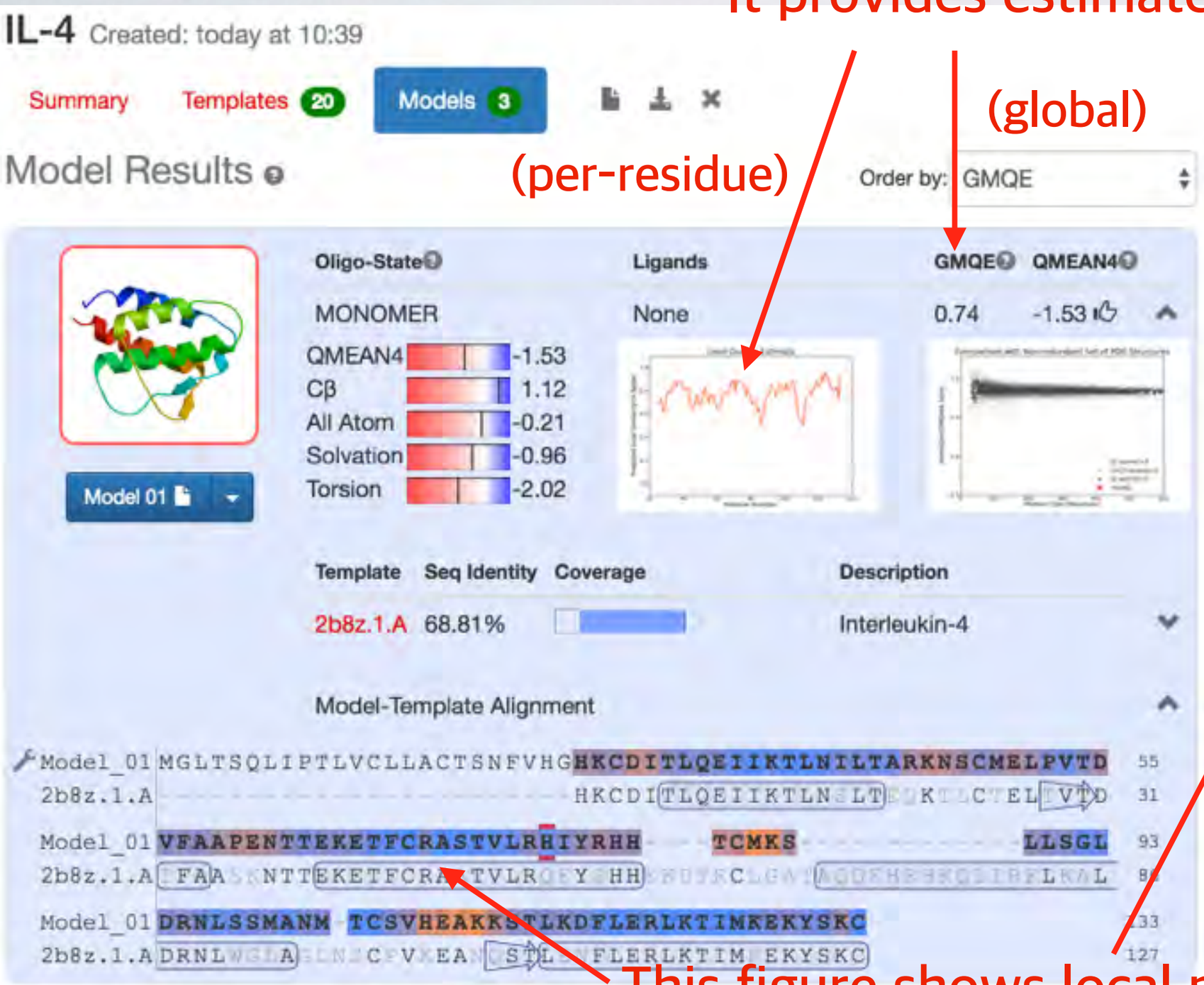
It predicts a model with for each template

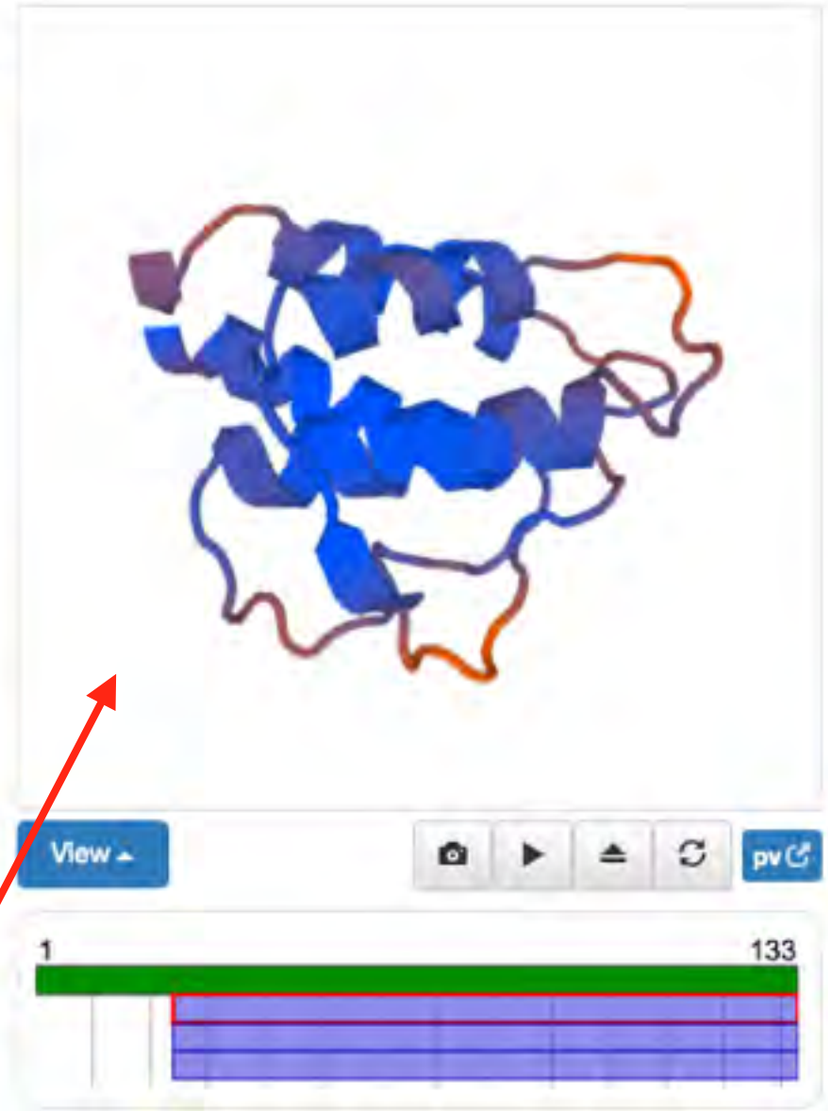
SWISS-MODEL

It provides estimated model accuracies

(per-residue)

(global)





This figure shows local model accuracies
(blue means reliable)

GALAXY programs for protein structure prediction, refinement, and docking

GALAXY

Structure Prediction

- *Protein structure refinement* — Galaxy Refine
- *Template-based modeling* — Galaxy TBM
- *Domain prediction* — Galaxy Dom
- *Loop modeling* — Galaxy Loop

Protein Docking

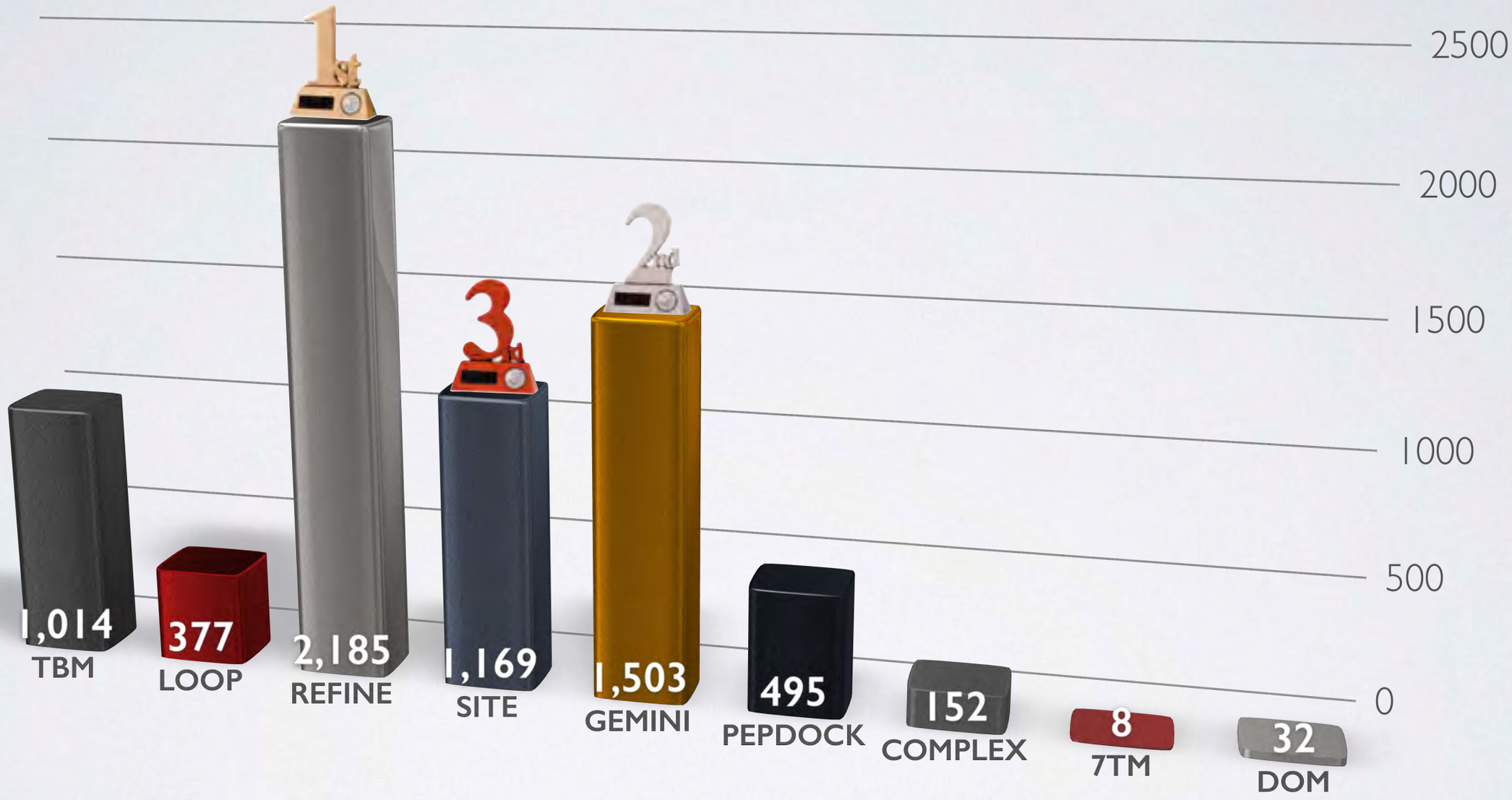
- *Oligomer prediction* — Galaxy Gemini
- *Protein-protein docking* — Galaxy PPDock
- *Protein-peptide docking* — Galaxy PepDock

Ligand Docking

- *Binding site prediction* — GalaxySite
- *Protein-ligand docking* — GalaxyDock
- *GPCR docking* — Galaxy 7TM

GalaxyWEB server
(http://galaxy.seoklab.org)

What are the most used programs?

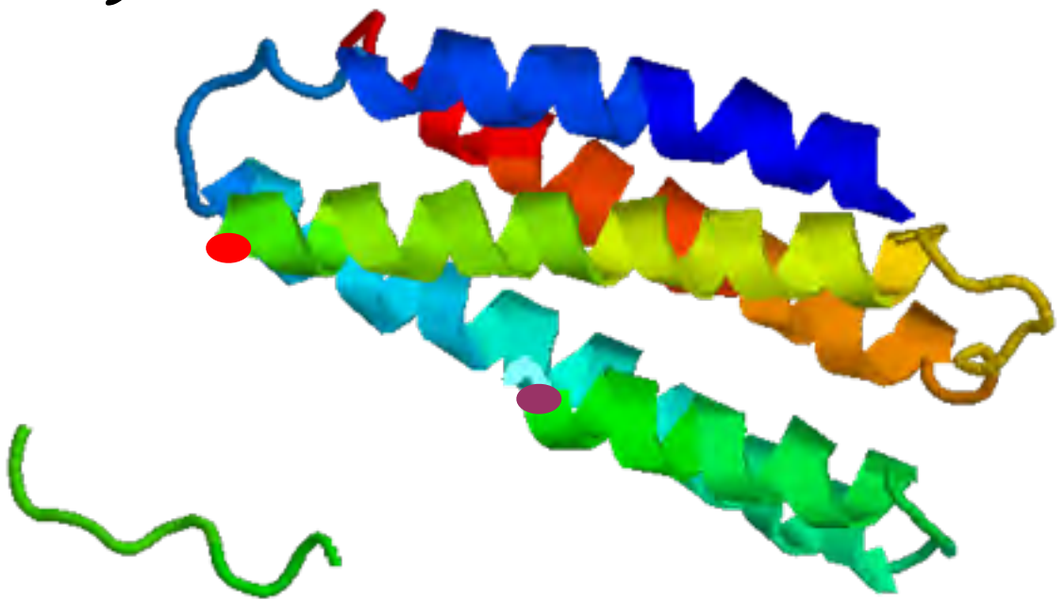


Out of 6,936 runs after excluding the jobs submitted through SNU IPs
(as of Jan 17, 2016)

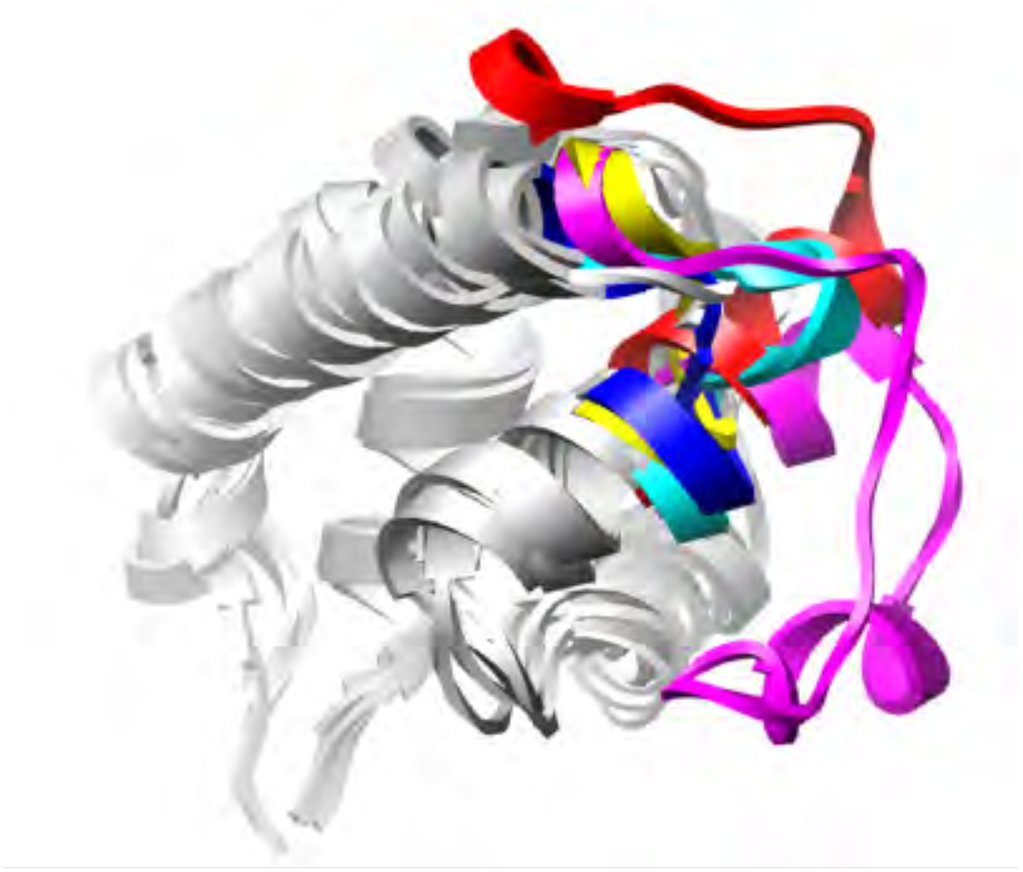
Protein loop modeling

Sequence Alignment

Loop Modeling



Variability of protein loops



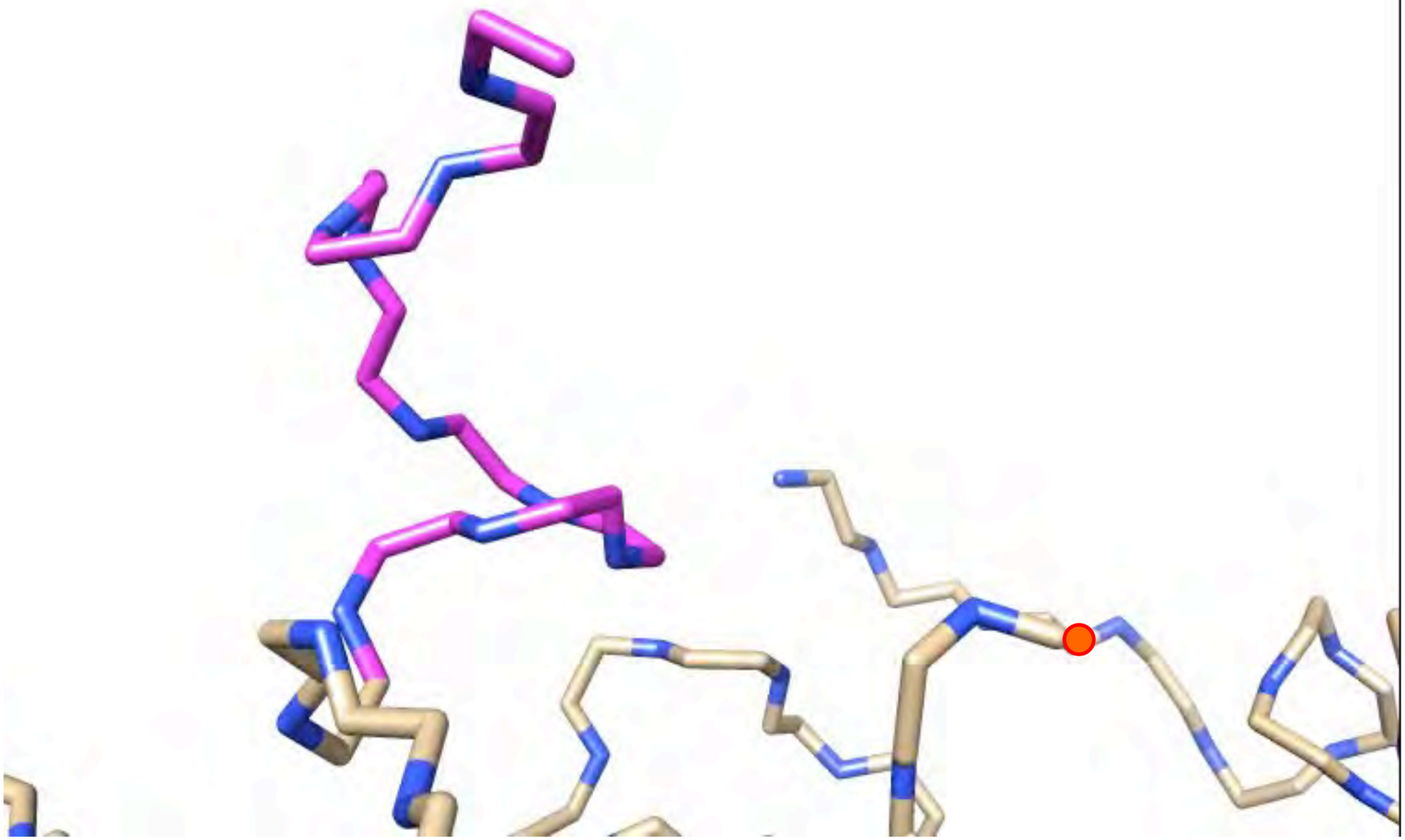
Sequence and structure variability
among homologous proteins

↓

Functional specificity

Loop closure: an important issue in loop modeling

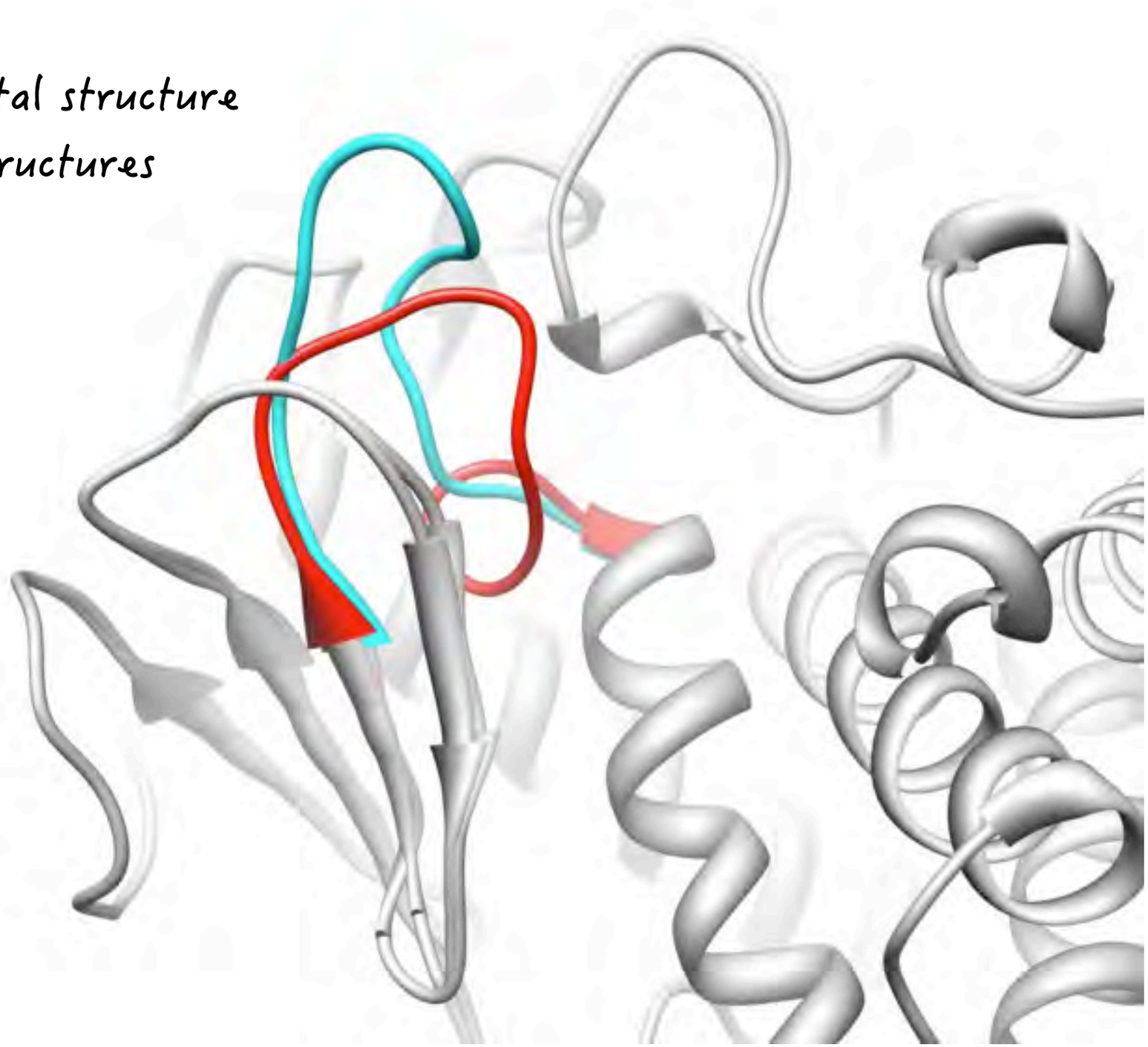
To close the backbone loop correctly between fixed structures.



Example of loop sampling

Experimental structure

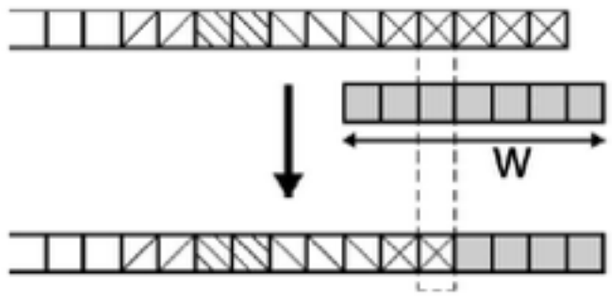
Sampled structures



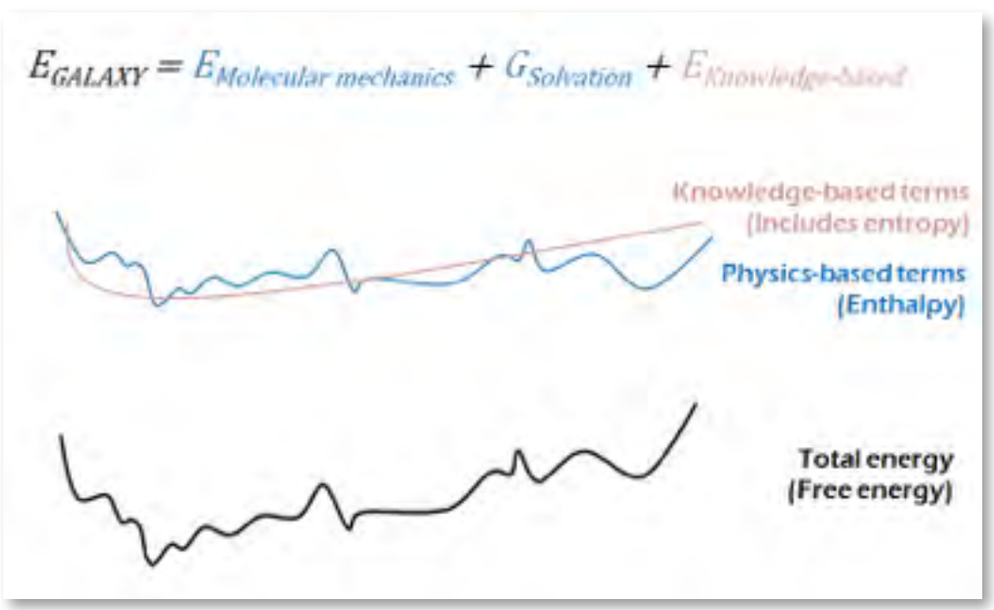
GalaxyLoop protein loop modeling

Loop Closure Algorithm

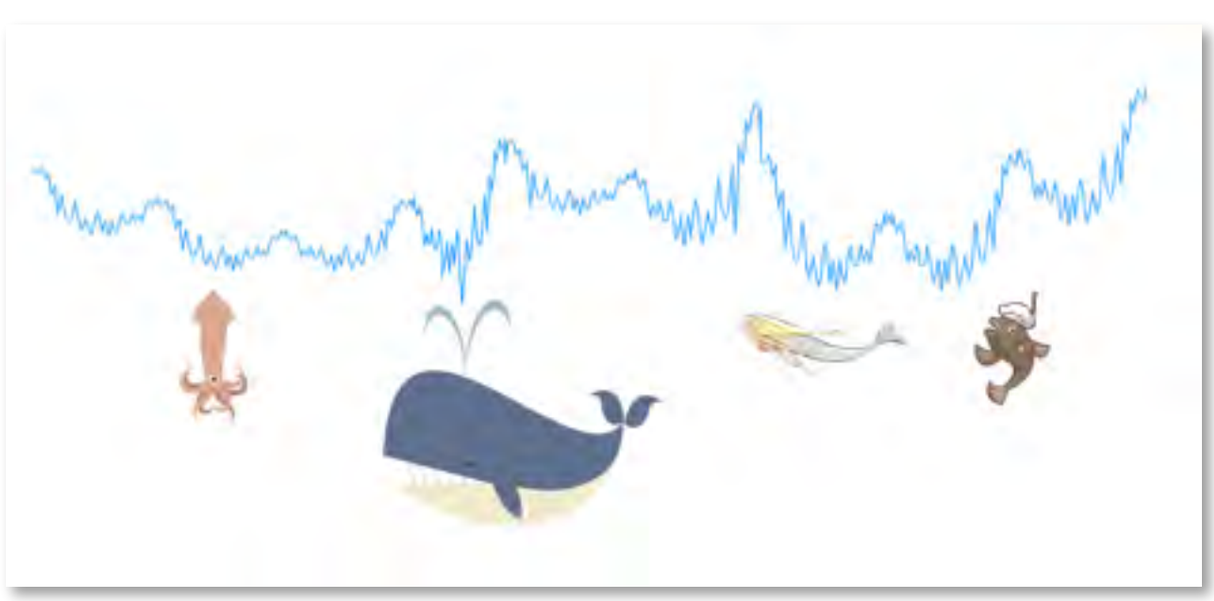
Fragment Assembly



Energy Function

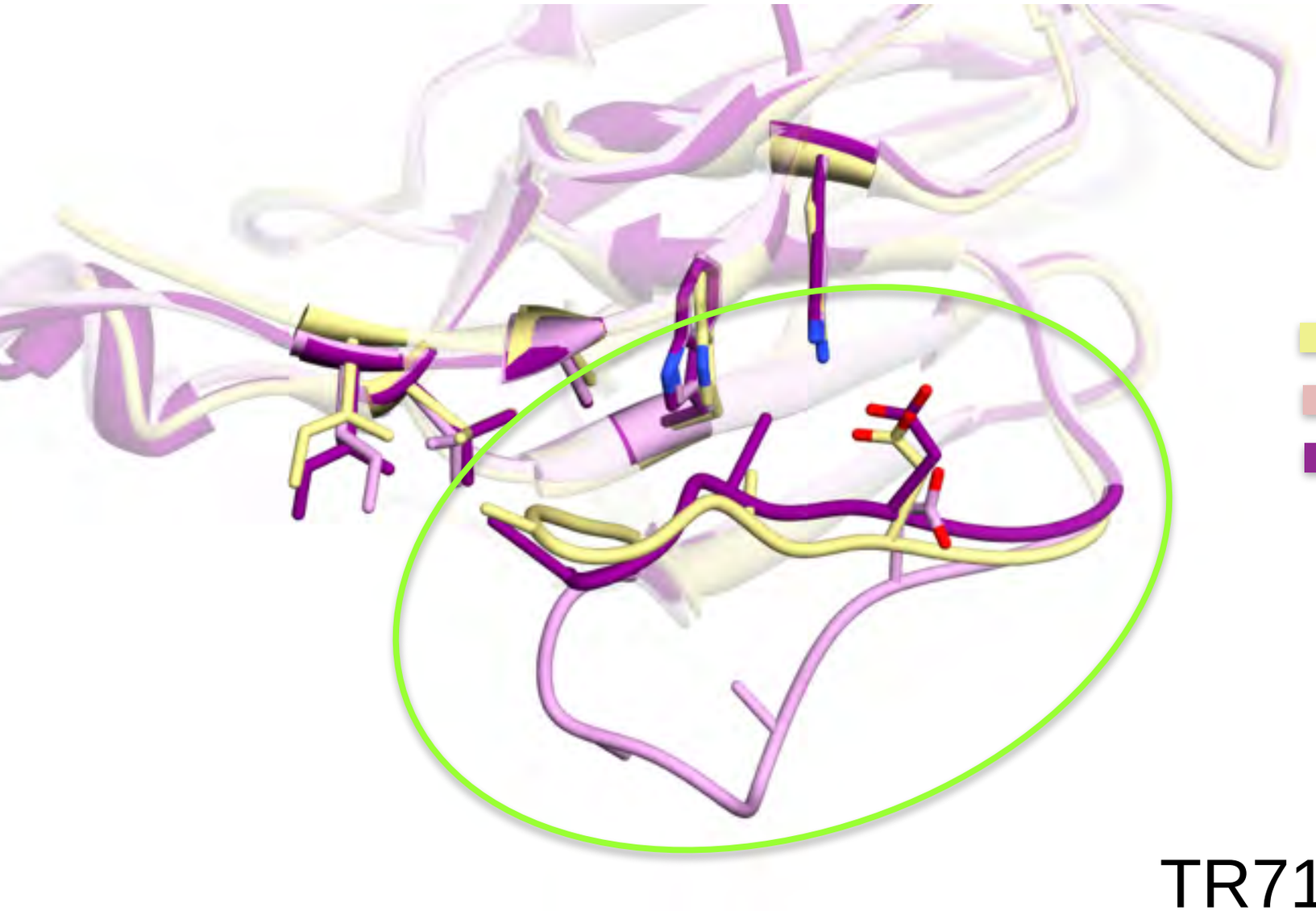


Global Optimization



E. A. Coutsias, C. Seok*, M. P. Jacobson, and K. A. Dill, A Kinematic View of Loop Closure, J. Comput. Chem. 25, 510-528 (2004).
E. A. Coutsias, C. Seok, M. J. Wester, and K. A. Dill, Resultants and Loop Closure, Int. J. Quantum Chem. 106, 176 (2006).
J. Ko, D. Lee, H. Park, E. A. Coutsias, J. Lee*, and C. Seok*, The FALC-Loop web server for protein loop modeling, Nucleic Acids Res. 39, W210-W214 (2011).
J. Lee*, D. Lee, H. Park, E. A. Coutsias, and C. Seok*, Protein loop modeling by using fragment assembly and analytical loop closure, Proteins, 78, 3428-3436 (2010).
H. Park, J. Ko, K. Joo, J. Lee, C. Seok*, and J. Lee*, Refinement of protein termini in template-based modeling using conformational space annealing, Proteins, 79, 2725-2734 (2011).
H. Park and C. Seok*, Refinement of unreliable local regions in template-based protein models, Proteins, 80 (8), 1974-1986 (2012).

GalaxyLoop: blind prediction example on model framework structure



Experimental structure
Initial structure
After refinement

TR712

GDT-HA: 80.2 -> 82.1
HB-bb: 82.3 -> 92.4
MP score: 2.20 -> 1.28 (Polar H only)

GalaxyLoop server: input page

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home | Services | Queue | Help | Softwares | Suppl

GalaxyLoop

Loops or termini of an initial structure may be refined. One may specify up to three loops or termini regions to model.

User Information

Job name

E-mail address (Optional)

Initial Structure

PDB File — 찾아보기...

Protein Structure File (allowed file extensions: pdb, txt)

Modeling options

Scoring method — PS1tbm protocol / PS2 protocol

Loops or Termini to be Refined (5aa<=len<=20aa)

Region 1 — Start , End

Region 2 — Start , End

Region 3 — Start , End

Submit

Help

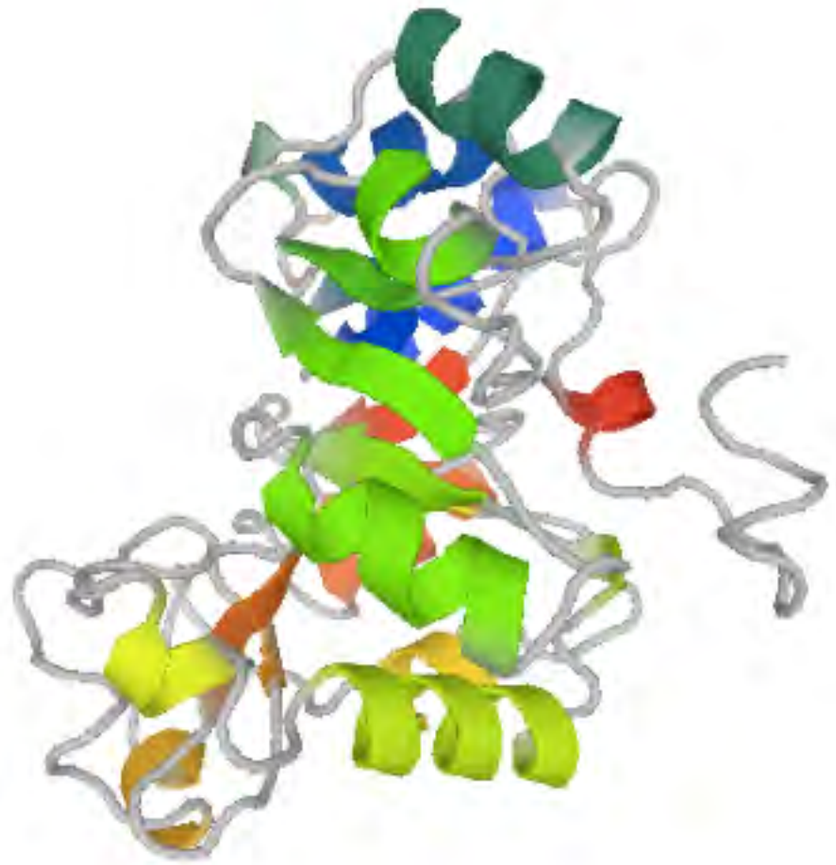
- Information
- PDB File: File Format
- E-mail: Average run time is ~2h. If e-mail address is given, the server sends notifications automatically. If not, the user has to bookmark the report page.

Example

- PDB file: T0570.M1.pdb
- Region 1: Start(3), End(15)
- Region 2: Start(126), End(134)
- Report: [View]

GalaxyLoop server: output page

R0570

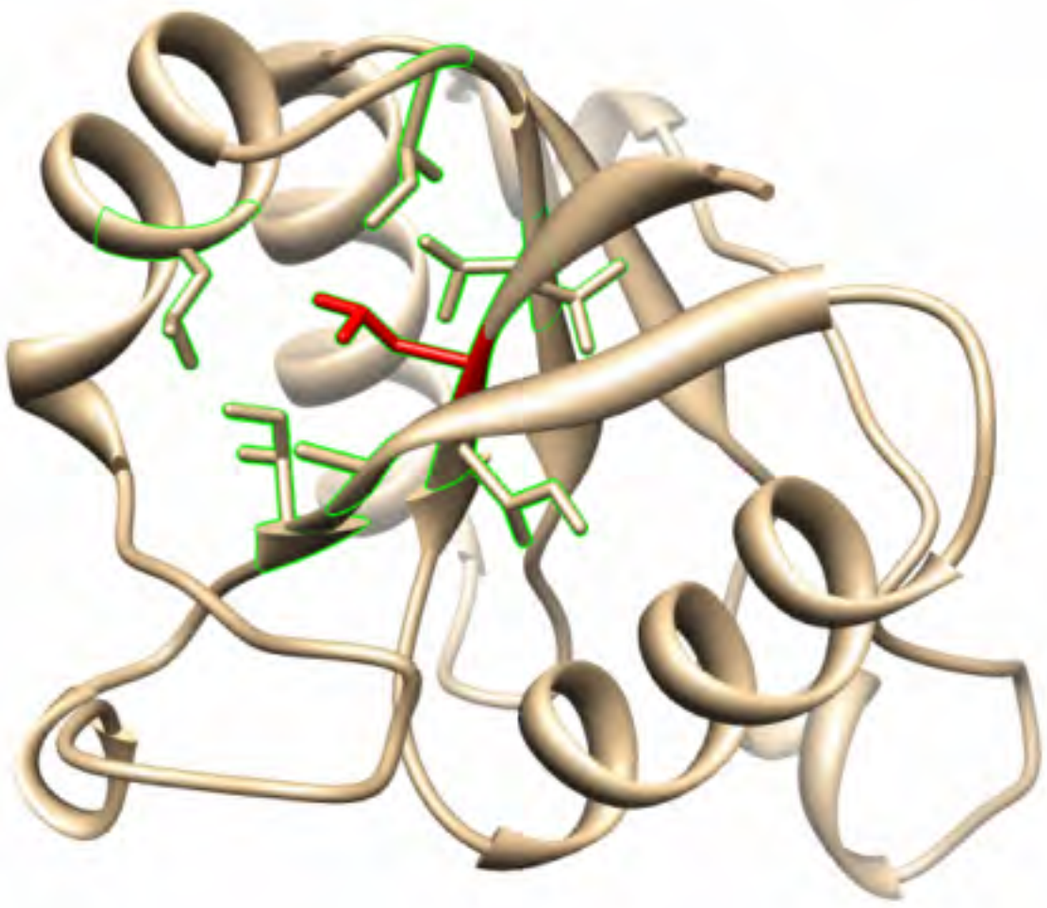


View in PV [Model 1] [Model 2] [Model 3] [Model 4] [Model 5] [All Models]

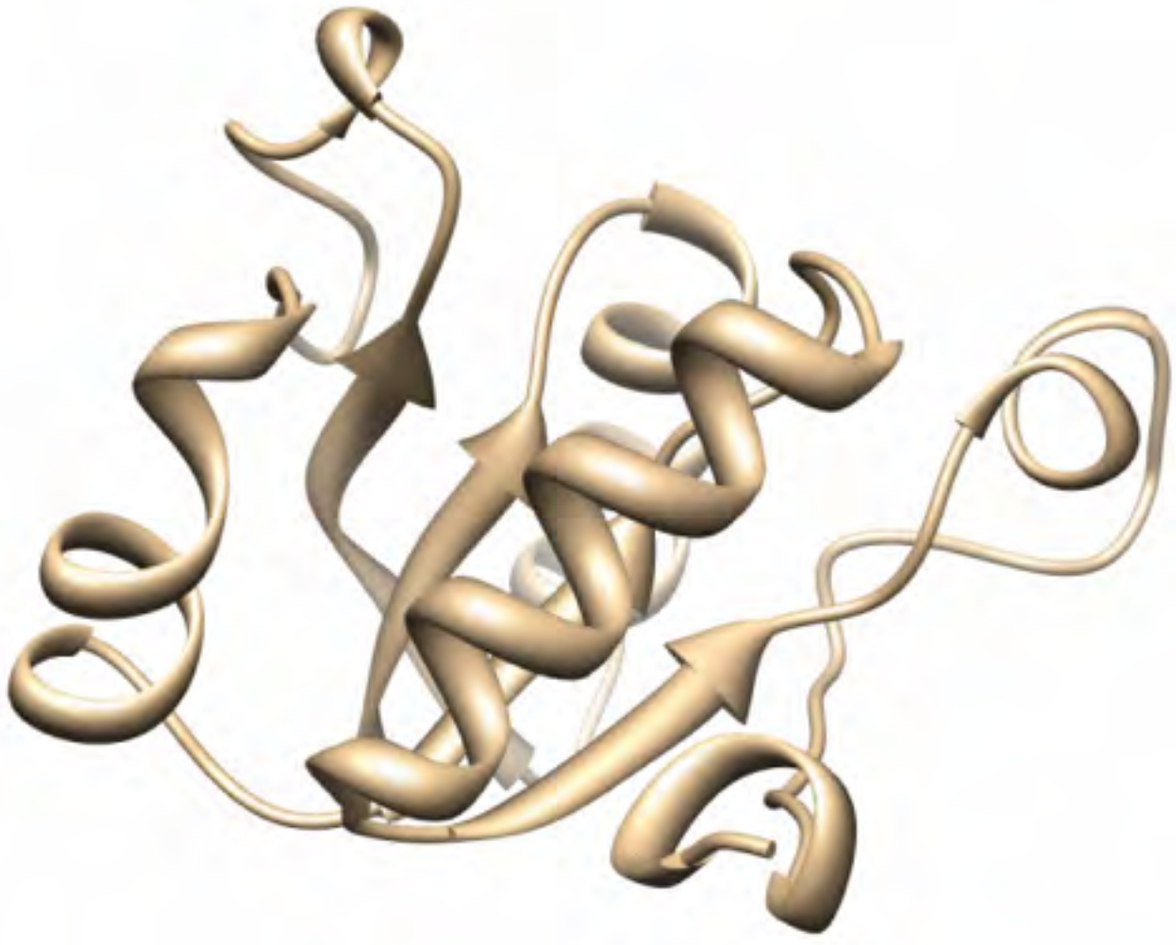
Download [Model 1] [Model 2] [Model 3] [Model 4] [Model 5]

GalaxyRefine: Overall relaxation

Mild relaxation
(w/ stronger initial-structure restraint)



Aggressive relaxation
(w/ weaker restraint)



GalaxyRefine: A successful blind prediction example

Experimental structure
Initial structure
After refinement

TR681

GDT-HA: 56.9 -> 63.9
HB-bb: 82.3 -> 92.4
MP score: 2.63 -> 1.44 (Polar H only)

GalaxyRefine server: input page

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

GalaxyRefine

Model structures generated by protein structure prediction methods can be refined. No gaps are allowed in the middle of initial protein structure.

User Information

Job name

E-mail address (Optional)

Model Structure to be refined

PDB File
(≤1000 AA)

찾아보기...

Protein Structure File (allowed file extensions: pdb, txt)

Submit

submit reset

Help

- Information
- PDB File: File Format
- E-mail: Average run time is 1~2h. If e-mail address is given, the server sends notifications automatically. **If not, the user has to bookmark the report page.**

Example

- PDB file: TR747.pdb
- Report: [View]

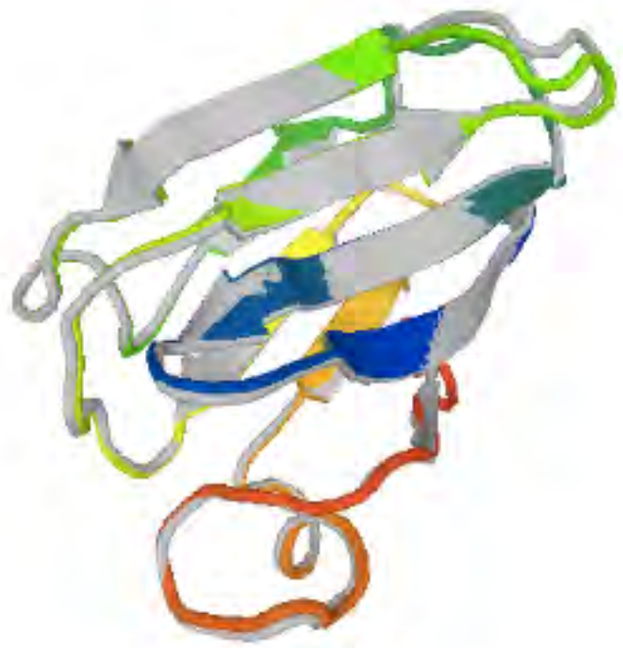
Software

- You can download a standalone version at here.

compbio.galaxy@gmail.com | Lab. of Computational Biology and Biomolecular Engineering

GalaxyRefine server: output page

TR747



View in PV [Model 1] [Model 2] [Model 3] [Model 4] [Model 5] [All Models]
Download [Model 1] [Model 2] [Model 3] [Model 4] [Model 5] [All Models]

Structure Information

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	1.951	14.4	0.0	95.8
MODEL 1	0.9808	0.516	1.543	10.6	0.0	99.0
MODEL 2	0.9411	0.978	1.700	6.8	1.1	95.8
MODEL 3	0.9397	0.978	1.676	5.0	0.0	93.8
MODEL 4	0.9538	0.839	1.924	11.2	0.0	94.8

Community-wide blind prediction experiments

CASP (**C**ritical **A**ssessment of techniques for protein **S**tructure **P**rediction)
Comparative evaluation of protein structure prediction methods

CAPRI (**C**ritical **A**ssessment of **PR**ediction of **I**nteractions)
Comparative evaluation of methods for prediction of protein-protein interactions

CSAR (**C**ommunity **S**tructure-**A**ctivity **R**esource)
Comparative evaluation of methods for prediction of protein-ligand interactions

GPCR Dock
Comparative evaluation of methods for GPCR modeling and docking

CAPRI

CAPRI (Critical Assessment of PRediction of Interactions)

Communitywide experiment on the comparative evaluation of methods for prediction of protein-protein interactions

held irregularly

We participated since Round 20, Jan 2010.

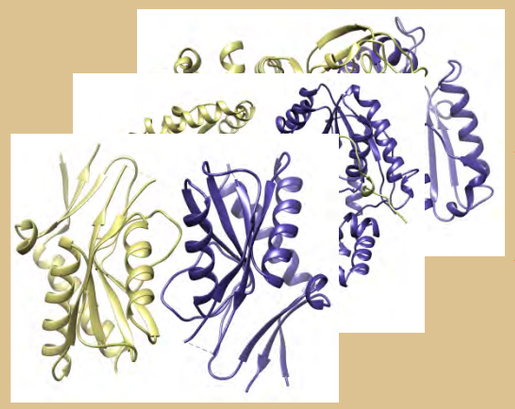
Prediction of Protein-Protein Interactions by GALAXY in CAPRI Round 30

Participant	Participated targets	Performance
CAPRI Predictor Ranking		
Seok	25	**15/14****
Huang	25	**16/13****
Guerois	25	**16/12****
Zou	25	**14/11****
Shen	25	**13/11****
Grudinin	24	**11/10****
Weng	25	**13/9****
Vakser	25	**11/9****
Vajda/Kozakov	24	**15/8****
Fernandez-Recio	25	**11/8****
Lee	20	**10/7****
Tomii	20	**8/6****
Sali	12	**6/4****
Negi	25	**7/3****
Eisenstein	6	**3****
Bates	25	**7/2****
Kihara	23	**7/2****
Zhou	25	**4/2****
Tovchigrechko	12	**3/1****
Ritchie	8	**2/1****
Fernandez-Fuentes	14	**1**
Xiao	11	**1**
Gray	1	**1**
Gong	8	**0**
Del Carpio	3	**0**
Wade	2	**0**
Haliloglu	1	**0**

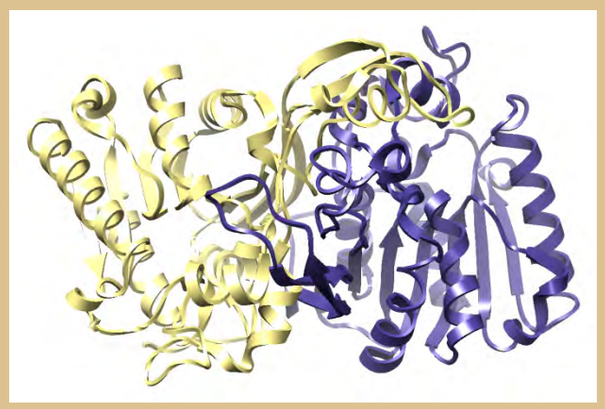
Prediction of homo-oligomer structure by GALAXY

Target Protein

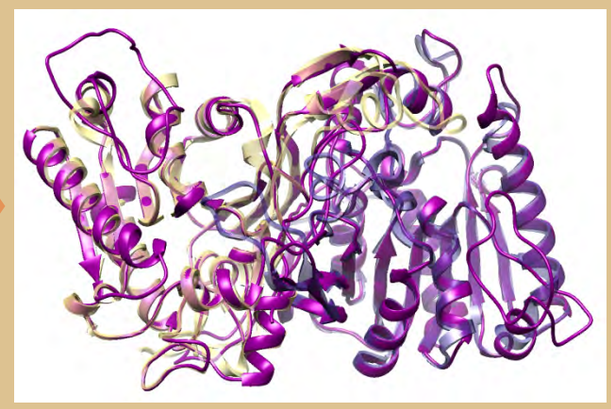
Oligomer Template Search (GalaxyGemini) and Model Building (GalaxyTBM)



HHsearch
on oligomer database

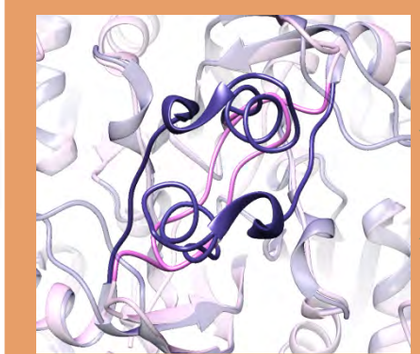


Template selection with
interface weighted rescoring

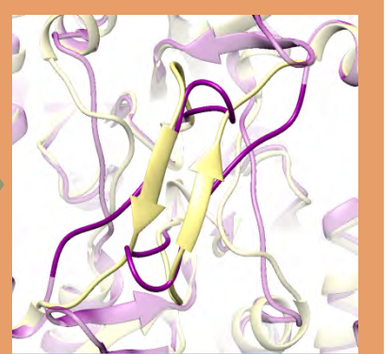


Oligomer model building

Symmetric ULR modeling (GalaxyLoop)

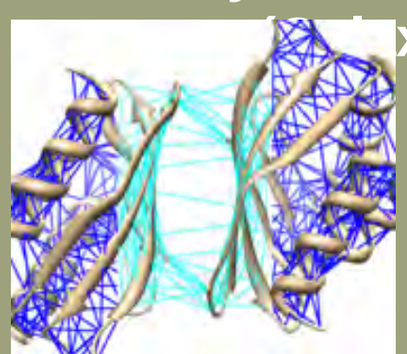


ULR detection

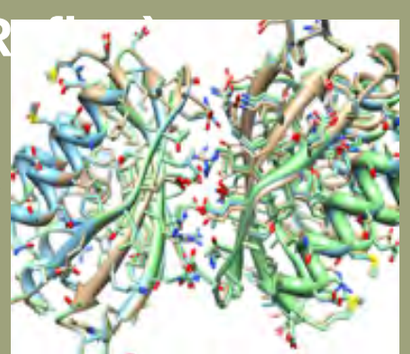


ULR modeling

Global Refinement with Symmetry Restraint

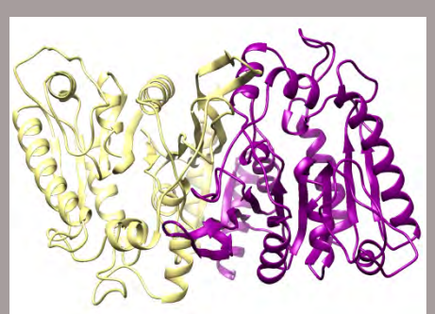


Restraint from
initial structure



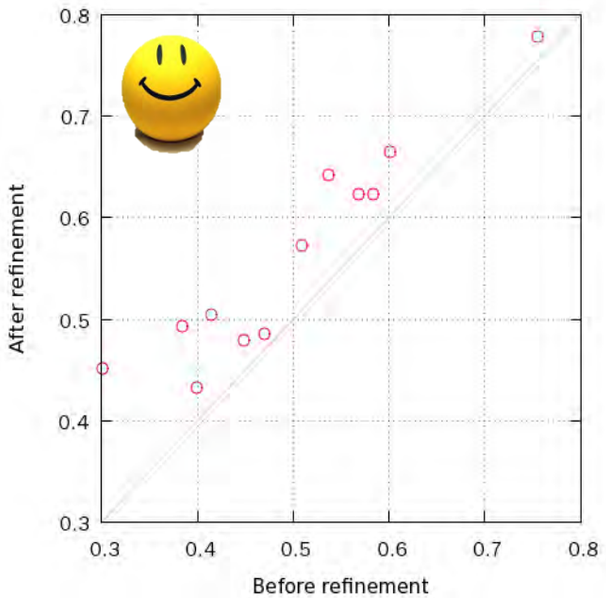
Complex structure
refinement

Final Model

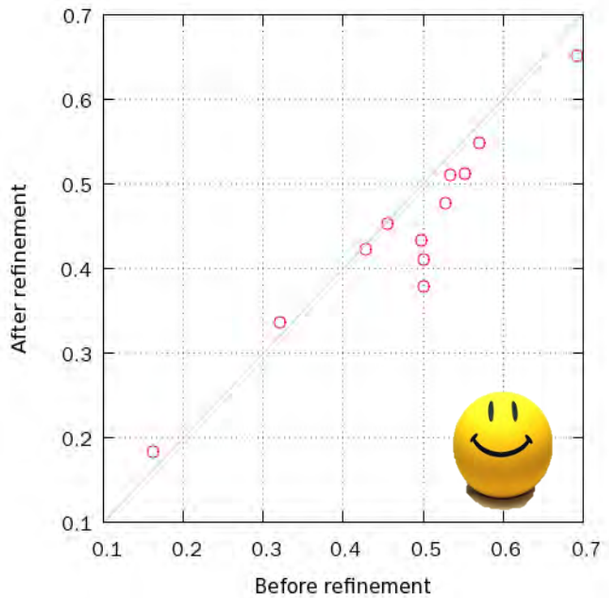


Impact of refinement

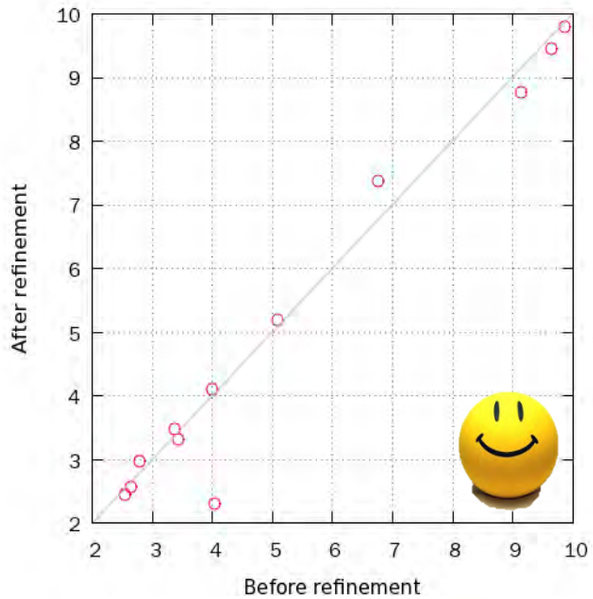
F_{nat}



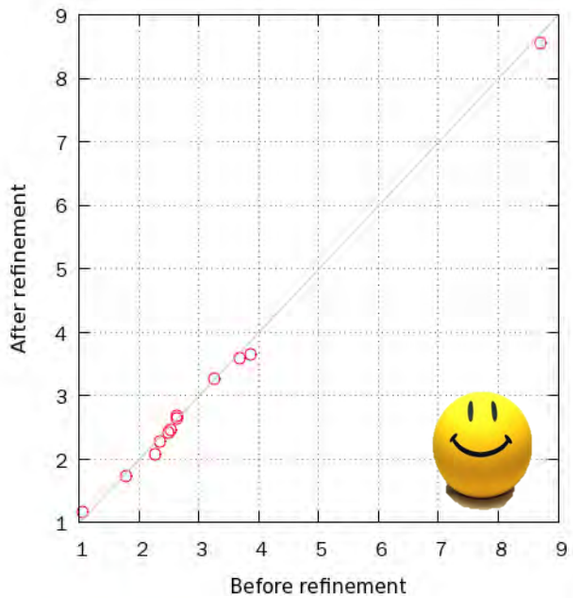
F_{nonnat}

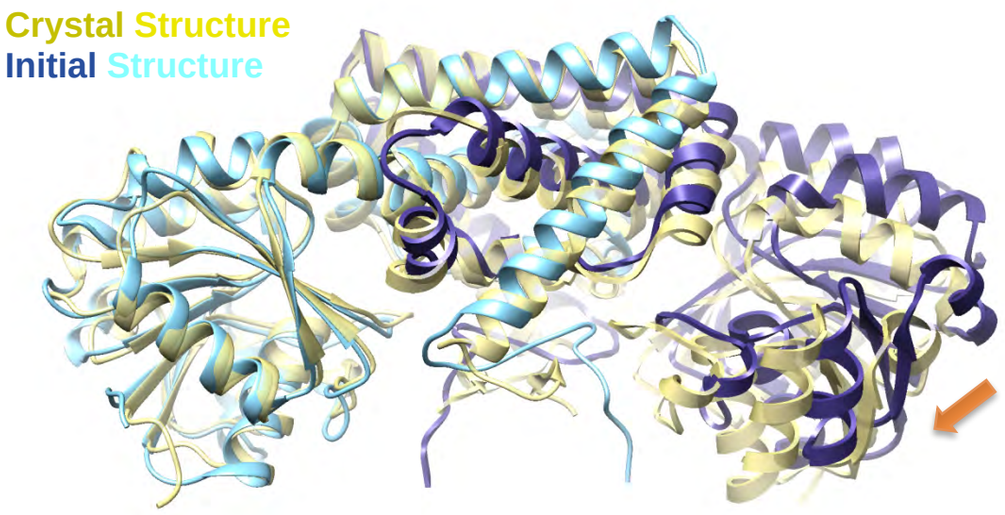


LRMSD

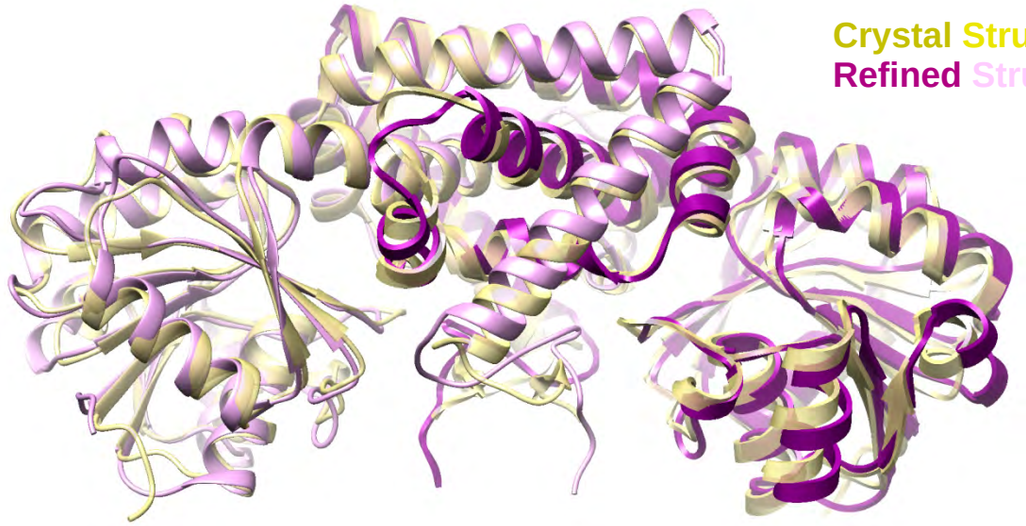


IRMSD





Refinement



T85	F_{nat}	F_{nonnat}	LRMSD	IRMSD
Before	0.60	0.32	4.02	2.27

T85	F_{nat}	F_{nonnat}	LRMSD	IRMSD
After	0.66	0.34	2.31	2.07

GalaxyGemini server: input page

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

GalaxyGemini

Prediction of protein oligomeric structure is performed from protein monomer. The input protein structure may be either an experimental structure or a model structure.

User Information

Job name

E-mail address (Optional)

Protein for oligomeric structure prediction

PDB File 찾아보기...

Protein Structure File (allowed file extensions: pdb, txt)

Energy minimization ○ YES ◉ NO

Submit

submit reset

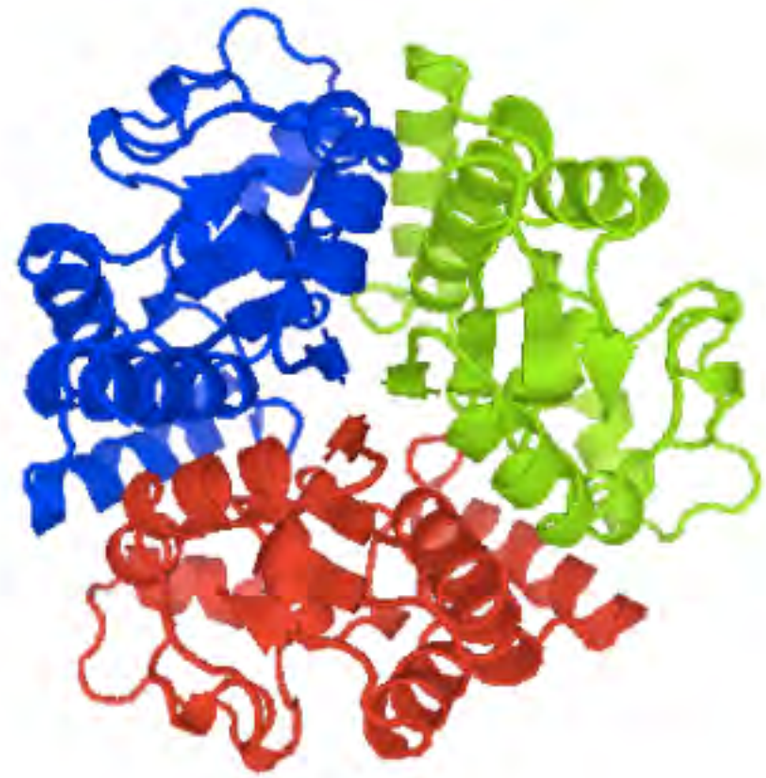
Help

- Information
- PDB File: File Format
- E-mail: Average run time is 30min~1h. If e-mail address is given, the server sends notifications automatically. If not, the user has to bookmark the report page address.
- Energy minimization: Energy minimization is used to remove steric clashes at the oligomer interface.

Example

- PDB file: Gemini.pdb
- Report: [View]

GalaxyGemini server: output page



View in PV [Model 1] [Model 2] [Model 3]

Predicted Oligomeric structures

No	Number of Subunits	Score1	Score2	Oligomer template	Oligomer Structure
1	3-mer	594.146	504.821	2qs7_A	[DOWNLOAD]
2	6-mer	332.601	186.66	1jx7_A	[DOWNLOAD]
3	1-mer	326.535	0.0	2d1p_B	[DOWNLOAD]

GalaxyRefineComplex server: input page

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

GalaxyRefineComplex

Protein-protein complex structures generated by protein-protein docking methods can be refined. Protein chains must have different chain IDs and be separated by "TER" line.
No gaps are allowed in the middle of input protein chains.

User Information

Job name

E-mail address (Optional)

Model Structure to be refined

PDB File
(≤ 800 AA)

찾아보기...
Protein Structure File (allowed file extensions: pdb, txt)

Refinement options

Symmetric refinement for homo-oligomer

Yes No

Submit

submit reset

Help

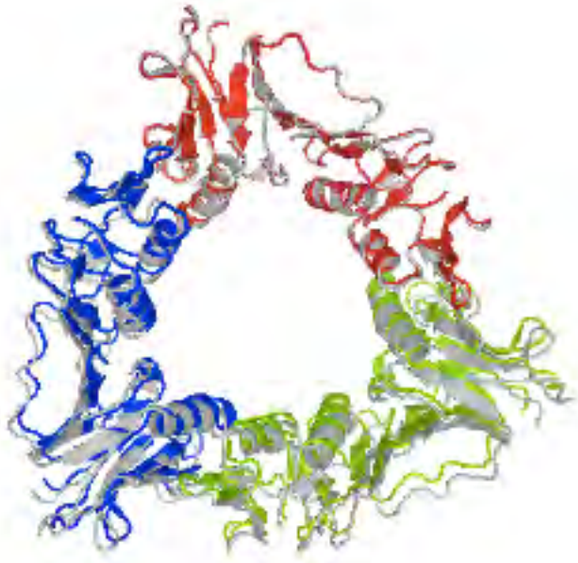
- Information
- PDB File: File Format
- E-mail: Average run time is 1~2h. If e-mail address is given, the server sends notifications automatically. If not, the user has to bookmark the report page.

Example

- Hetero-oligomer example
  - PDB file: hetero_complex.pdb
  - Report: [View]
- Homo-oligomer example
  - PDB file: homo_oligomer.pdb
  - Report: [View]

GalaxyRefineComplex server: output page

Homo.Oligomer



View in PV [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [All Models]
Download [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [All Models]

Download

- 10 refined models [DOWNLOAD]

Global Structure Information

Model	Input	Model 1	Model 2	Model 3	Model 4	Model 5
Interface area	3097.6	3217.2	3227.9	3219.5	3211.0	3221.3
MolProbity	2.609	1.097	1.153	1.287	1.260	1.260
Clash score	37.8	3.0	3.6	3.6	3.3	3.3
Rotamer outlier	3.0	1.0	1.0	1.5	1.5	1.5
Rama favored	96.9	98.2	98.2	98.2	98.2	98.2

Model	Input	Model 6	Model 7	Model 8	Model 9	Model 10

GalaxySite server: input page

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

GalaxySite

Prediction of ligand binding site of a query protein is performed. Up to three ligands that are likely to bind to the protein and their predicted binding poses are provided.

User Information

Job name

E-mail address (Optional)

Query Protein Information

SEQUENCE
(≤500 AA)

or PDB file
(≤500 AA)

찾아보기...

Protein Structure File (allowed file extensions: pdb, txt)

Submit

submit reset

Help

- Information
- PDB File: file format
- E-mail: Average run time ranges from 2h (for a structure input) to 4h (for a sequence input). If e-mail address is provided, the server sends notifications automatically. **If not, the user has to bookmark the report page.**

Examples

- **Prediction from a protein sequence**
  - SEQ: Uniprot Q8CNK1
  - Report: [View]
- **Prediction from a protein structure**
  - PDB file: SITE.pdb
  - Report: [View]

GalaxySite server: output page

FMN_binding

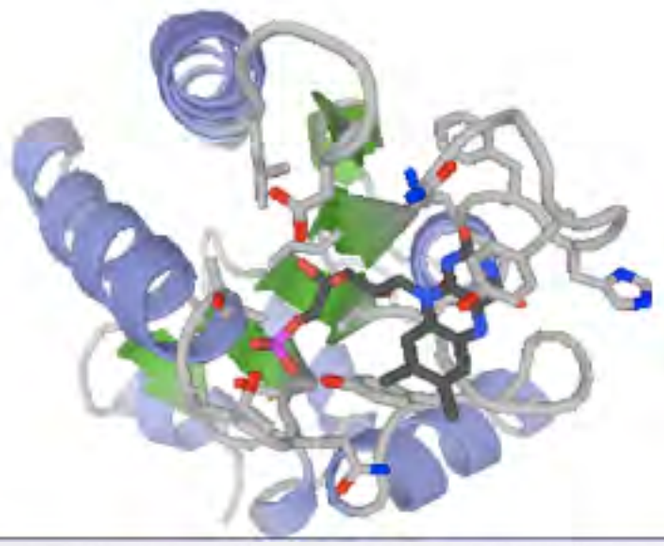
Ligands predicted to bind

No	Ligand Name	Ligand Structure	Templates for protein-ligand complex
1	FMN		3qe2_A, 1ttl_A, 2bpo_A, 3hr4_A, 1bwy_F, 1ykg_A, 3f6r_A, 1f4p_A, 2avc1_A, 2fcr_A, 1yob_A, 1ddo_A, 1czn_A, 1ag9_A, 5nul_A, 2q9u_A, 1e5d_A, 1ycg_A, 2ohh_A, 2xod_A, 3n3a_C, 1rlj_A
2	FAD		2bpo_A
3	BEN		2bmv_A

Predicted ligand-binding residues

No	Ligand Name	Binding Residues	Interaction Analysis
1	FMN	26S 27Q 28T 30T 31A 78A 79T 80Y 81G 113L 114G 115N 118Y 120H 121F 122N 148D 152L	LINK
2	FAD	28T 30T 80Y 115N 118Y 148D	LINK
3	BEN	39S 40K 42A 43H 48R 49G	LINK

Predicted binding poses



Help

- Information
- The data will be stored in the server only for 30 days.
- **Ligands predicted to bind:** GalaxyWEB provides up to 3 predictions for each query.
- **Predicted ligand-binding residues:** Ligand-binding residues depends on the definition of residue-ligand contact. If the distance between an amino acid residue and a ligand atom is less than the sum of van der waals radii of the two atoms + 0.5 A, the residue is considered as a binding site residue.
- **Binding site interaction analysis:** Interactions at the predicted ligand binding site are analyzed using LIGPLOT.
- **Predicted binding poses:** Predicted binding poses can be downloaded in the PDB format or viewed in PV.

GalaxyDom server: input page

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

GalaxyDom

A method to assign modeling units for protein structure prediction.

User Information

Job name

E-mail address (Optional)

Sequence Information

SEQUENCE
(≤1000 AA and >40 AA)

Or SEQ file 찾아보기...
(allowed file extensions: fa, fasta, txt, seq)

Submit

submit reset

Help

- Information
- SEQ file: a SEQ file in the FASTA format is required.
- E-mail: If e-mail address is given, the server sends notifications automatically. If not, the user has to bookmark the report page.

Example

- SEQ file: example.fa
- Report: [View]

compbio.galaxy@gmail.com | Lab. of Computational Biology and Biomolecular Engineering

GalaxyDom server: output page

None

Modeling Unit Information

No	Regions (Length)	Type	Sequence
1	1-26 (26)	SEQADV	QLYTRASQPELAPEDPEDLEHHHHHH
2	27-50 (24)	signaling peptide	MFQKKTYAVFLILLLMMFTAACSG
3	51-380 (330)	modeling unit	SKTSAEKKESETEKSSDIAQVKIKDVSYTLPSKYDKSTSDDQLVLKVNVA VKNTGKDPLNVDSMDFTLYQGDTKMSDTDPEDYSEKLQGSTINADKSVEG NLFFVVDKGKQYELNYTPESYGDKKPKSVTFKIDGKDKKILATADKLQDS AKALSAYVDVLLFGKDNADFEKITGANKNEIVNDFNESAKDGYLSASGLS STYADSKALDNIVNGIKEGLSKNSSIQAKTTSISKDEAIVEATVKPVDAS SLSDRIEDKVKDYYSKNSSASYEEAVKYALQVYPEEFKKLGPASSEKTVE VKMKKNDIDQWQLDMDDYRAAELVEAFIKE

GalaxyTBM server: input page

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

GalaxyTBM

Structure prediction and refinement of up to three loops or termini is automatically performed.

User Information

Job name

E-mail address (Optional)

Sequence Information

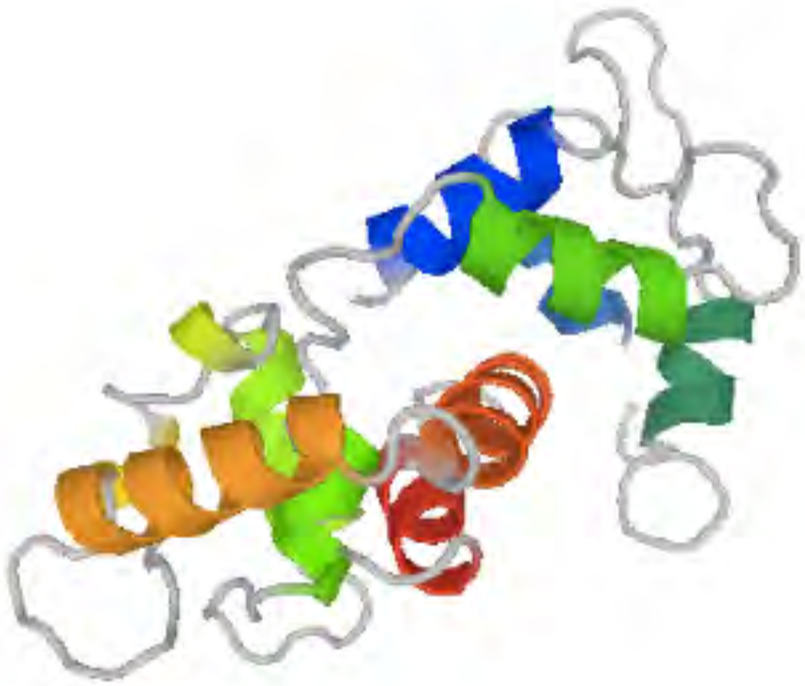
SEQUENCE
(≤500 AA)

Or SEQ file 찾아보기...
(allowed file extensions: fa. fasta, txt, seq)

Submit

submit reset

GalaxyTBM server: output page



View in PV [Model 1] [Model 2] [Model 3] [Model 4] [Model 5] [All Models]

Download [Model 1] [Model 2] [Model 3] [Model 4] [Model 5]

GalaxyPepDock server: input page

Nucleic Acids Research Advance Access published May 12, 2015

Nucleic Acids Research, 2015 1
doi: 10.1093/nar/gkv495

GalaxyPepDock: a protein–peptide docking tool based on interaction similarity

Hasup Lee[1], Lim Heo[1]

[1]Department of Chemistry, S
Sciences, College of Medici

Received February 11, 2015; Revi

GalaxyWEB

A web server for protein structure prediction, refinement, and related methods

Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

GalaxyPepDock

Protein-peptide docking based on interaction similarity

User Information

Job name

E-mail address (Optional)

Protein-peptide Docking

PROTEIN structure (≤ 900 AA) — 찾아보기... (allowed file extensions: pdb, txt)

PEPTIDE sequence (≤ 30 AA) — 찾아보기... (allowed file extensions: fa, fasta, txt, seq)

Submit

submit reset

Help

- **Protein structure**: PDB Format
- **Peptide sequence**: FASTA format
- E-mail: Average run time is ~12h. If e-mail address is given, the server sends notifications automatically. If **not, the user has to bookmark the report page address.**
- **More Information**

Example

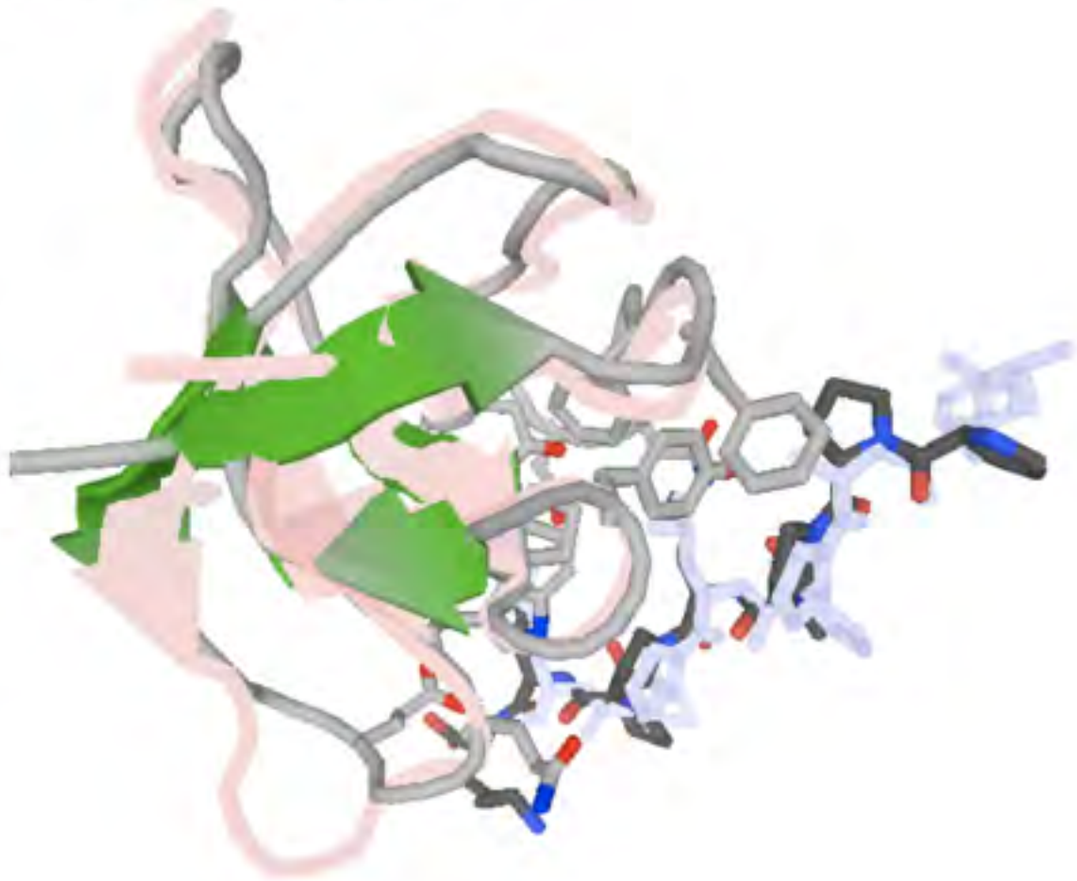
- **Protein structure**: protein.pdb
- **Peptide sequence**: peptide.fa
- Report: [View]

compbio.galaxy@gmail.com | Lab. of Computational Biology and Biomolecular Engineering

GalaxyPepDock server: output page

Predicted protein-peptide complex structures

(Template structures are shown in light colors)



View in PV [1] [2] [3] [4] [5] [6] [7] [8] [9] [10]

Download [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [All]

Galaxy7TM server: input page

OXFORD JOURNALS
Nucleic Acids Research

Galaxy7TM: flexible GPCR–ligand docking by structure refinement

Gyu Rie Lee and Chaok Seok*
+ Author Affiliations
* To whom correspondence shoul
1568; Email: chaok@snu.ac.kr

GalaxyWEB
A web server for protein structure prediction, refinement, and related methods
Computational Biology Lab, Department of Chemistry, Seoul National University

Home Services Queue Help Softwares Suppl

Galaxy7TM

Given a GPCR structure and a ligand structure, optimized complex structures are generated by docking and refinement. Input GPCR structure without gaps in the middle is recommended. Up to five gaps in the input GPCR structure can be filled if its full sequence is submitted together.

User Information

Job name

E-mail address (Optional)

Input GPCR and ligand structures

PDB File (≤1000 AA) — 찾아보기... — Protein Structure File (allowed file extensions: pdb, txt)

Sequence File (Optional) (≤1000 AA) — 찾아보기... — Protein Sequence File (allowed file extensions: fa, fasta, seq)

Ligand File (≤150 atoms) — 찾아보기... — Ligand Structure File (allowed file extensions: mol2, pdb, xyz)

Help

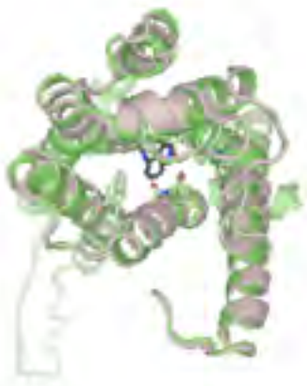
- Information
- PDB File: File Format
- Mol2 File: File Format
- E-mail: Average run time is ~2h. If e-mail address is given, the server sends notifications automatically. If not, the user has to bookmark the report page.

Example

- PDB file: P07700.pdb
- Ligand file: XF5.mol2
- Binding pocket: 89,90,93,178,179,246,250,269,273
- Report: [View]

Galaxy7TM server: output page

Galaxy7TM_P07700_XF5



View in PV [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [All Models]

Download [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [All Models]

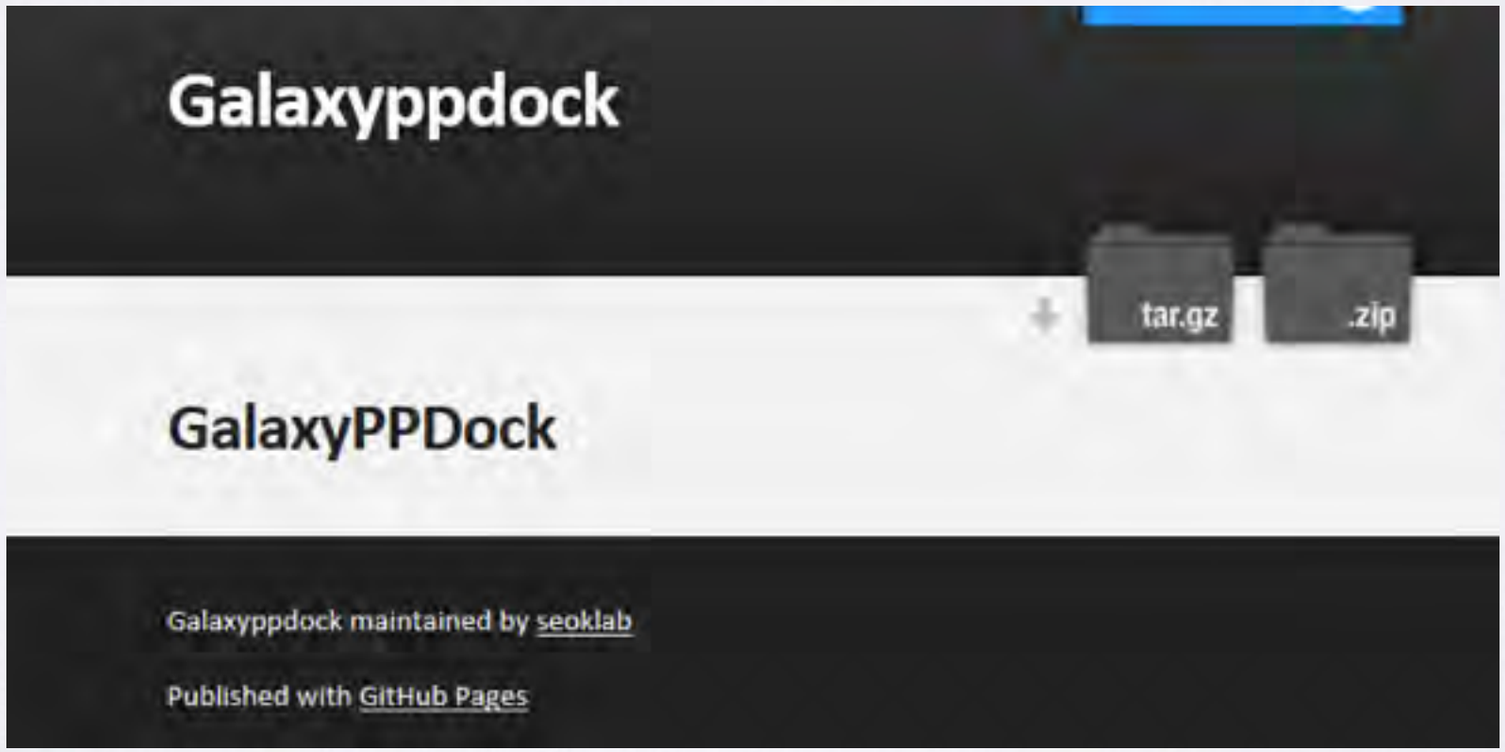
Model Information

Model	Refinement energy	Docking energy	Ligand RMSD from Model 1	Binding site interactions	Residues in contact
MODEL 1	-17247.468	-7.439	0.000	LINK	LINK
MODEL 2	-17216.997	-7.665	5.091	LINK	LINK
MODEL 3	-17237.360	-7.202	5.066	LINK	LINK
MODEL 4	-17281.641	-6.912	1.748	LINK	LINK
MODEL 5	-17247.782	-6.852	4.851	LINK	LINK

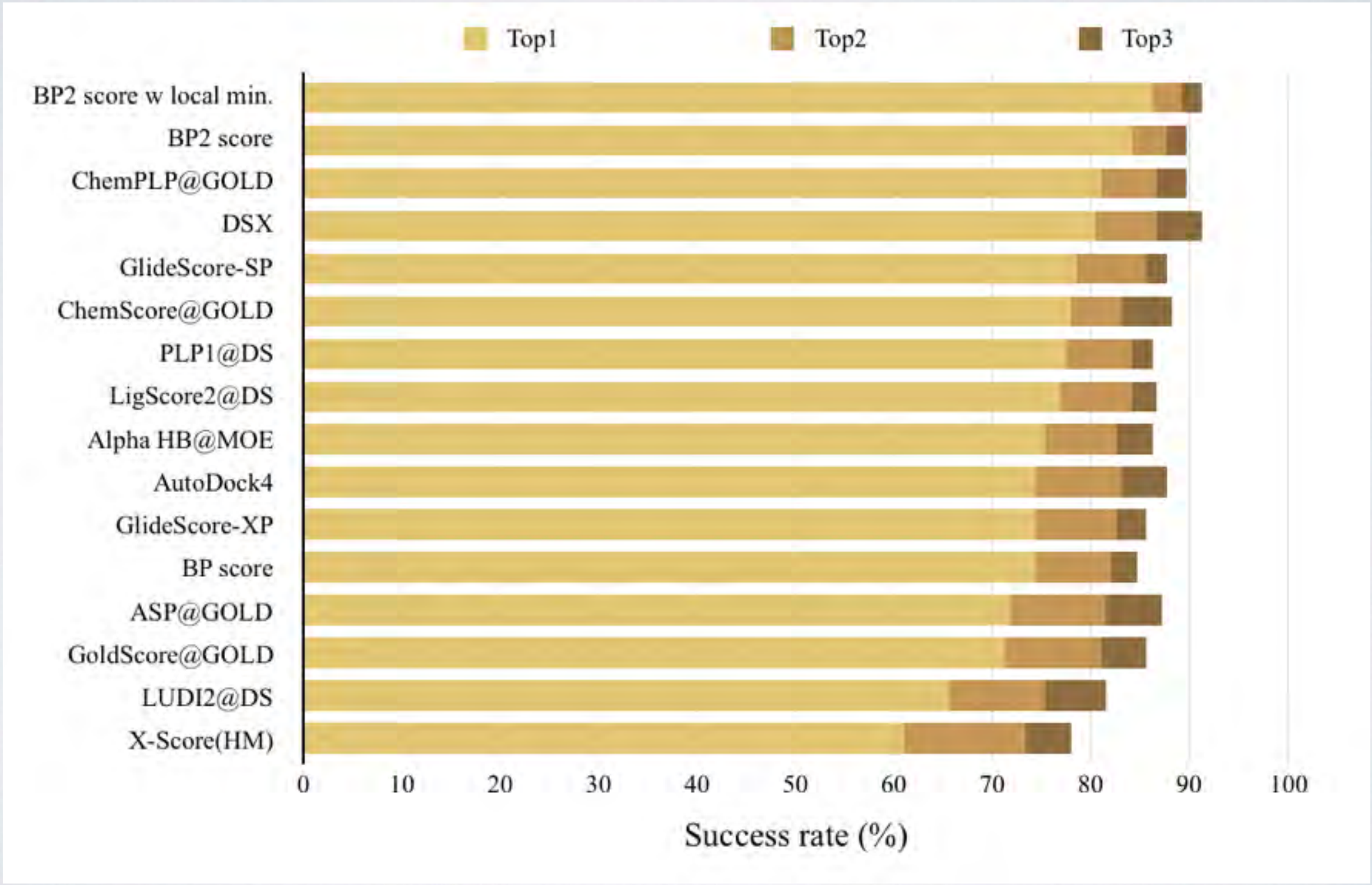
Downloadable GALAXY softwares (http://galaxy.seoklab.org/softwares)

GALAXY Programs

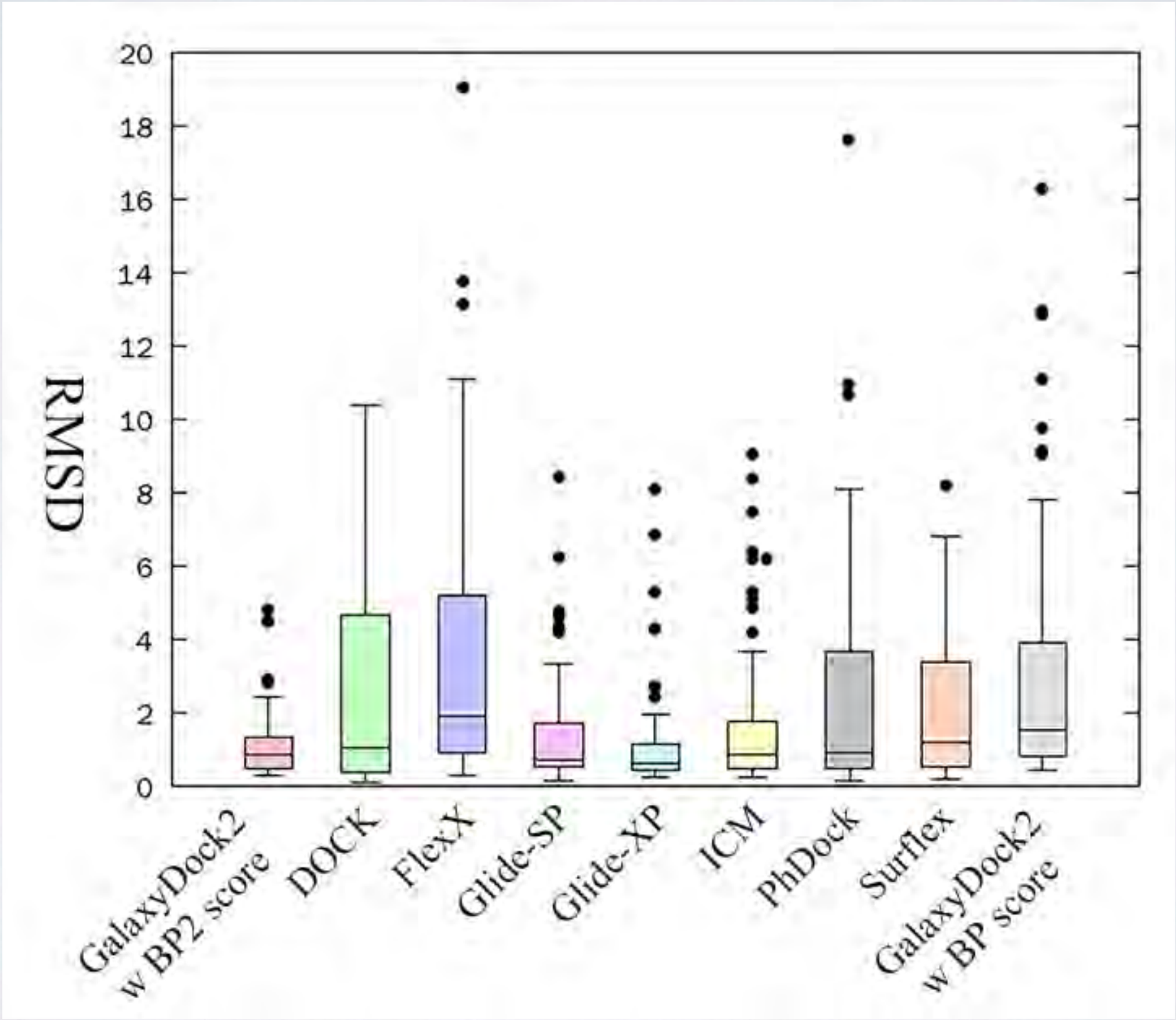
- GalaxyRefine
- GalaxyPPDock
- GalaxyFill
- GalaxyDock
- FALC



A new energy function (BP2 score) for protein-ligand docking



GalaxyDock with BP2 score

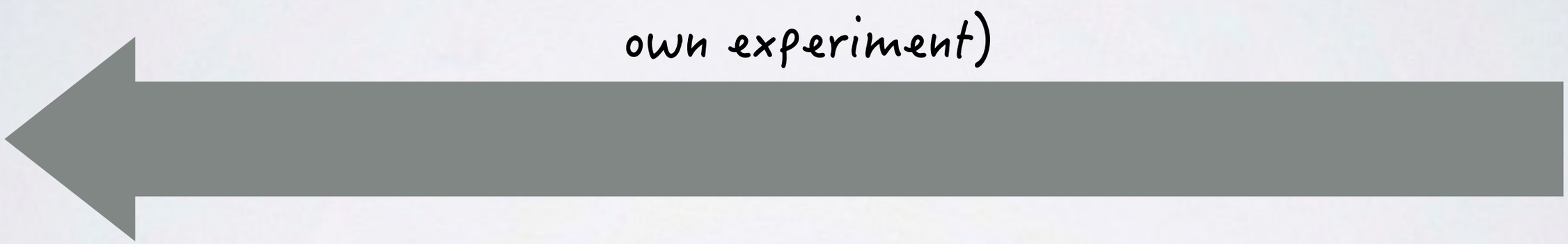


Conclusion: when/how to use computational methods?

Validate/Compare with experimental results

Decide on how much information you would like to use in your computation.

The amount of (experimental) information you use (from literature or your own experiment)



Accuracy of your results

New information you obtain

Compare different computational methods

Consensus among the results from different methods

Reliability of your results

Different hypotheses you may test by independent means

My own examples

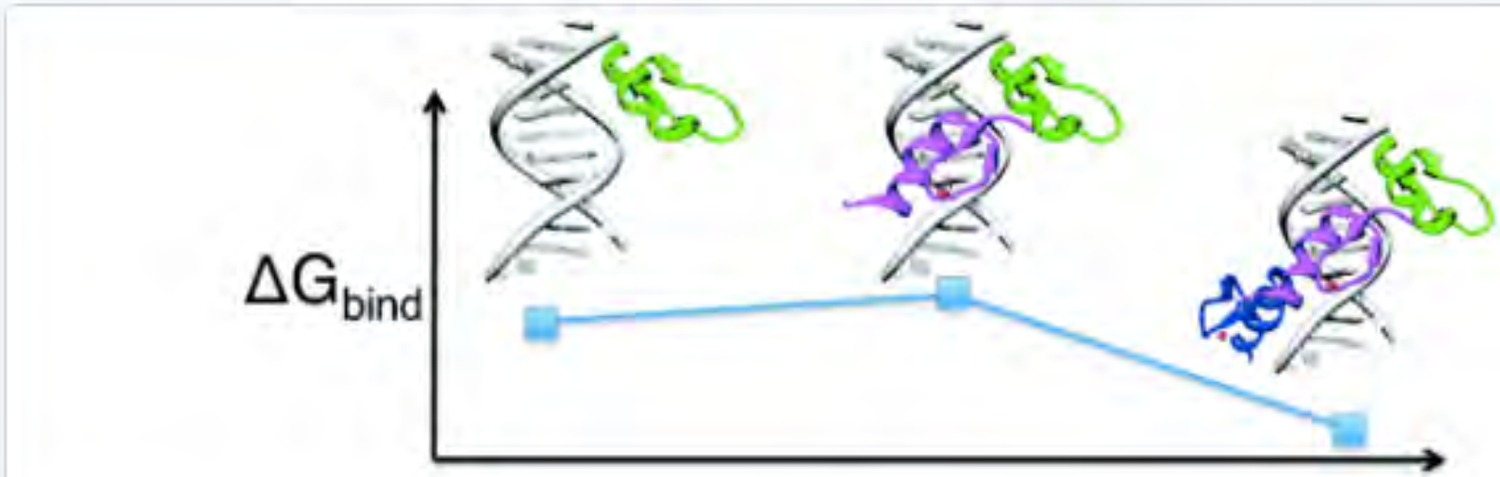
Cooperativity and Specificity of Cys_2His_2 Zinc Finger Protein−DNA Interactions: A Molecular Dynamics Simulation Study

Juyong Lee, Jin-Soo Kim and Chaok Seok*
Department of Chemistry, Seoul National University, Seoul, Republic of Korea

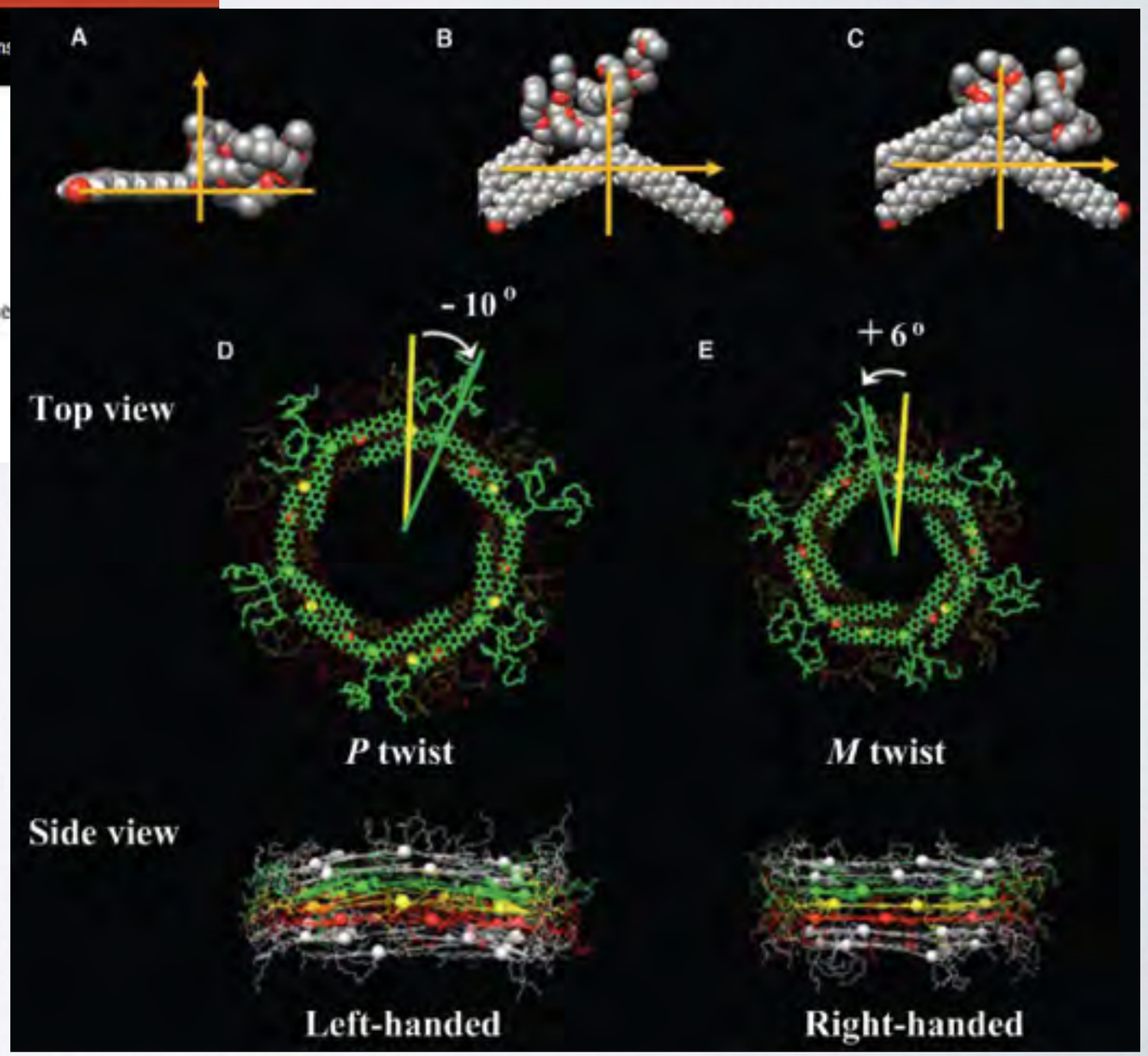
J. Phys. Chem. B, **2010**, *114* (22), pp 7662–7671
DOI: 10.1021/jp1017289
Publication Date (Web): May 14, 2010

* To whom correspondence should be addressed. E-mail: chaok@snu.ac.kr. Phone: 82-2-880-9197. Fax: 82-2-871-8119.

Abstract



My own examples



My own examples

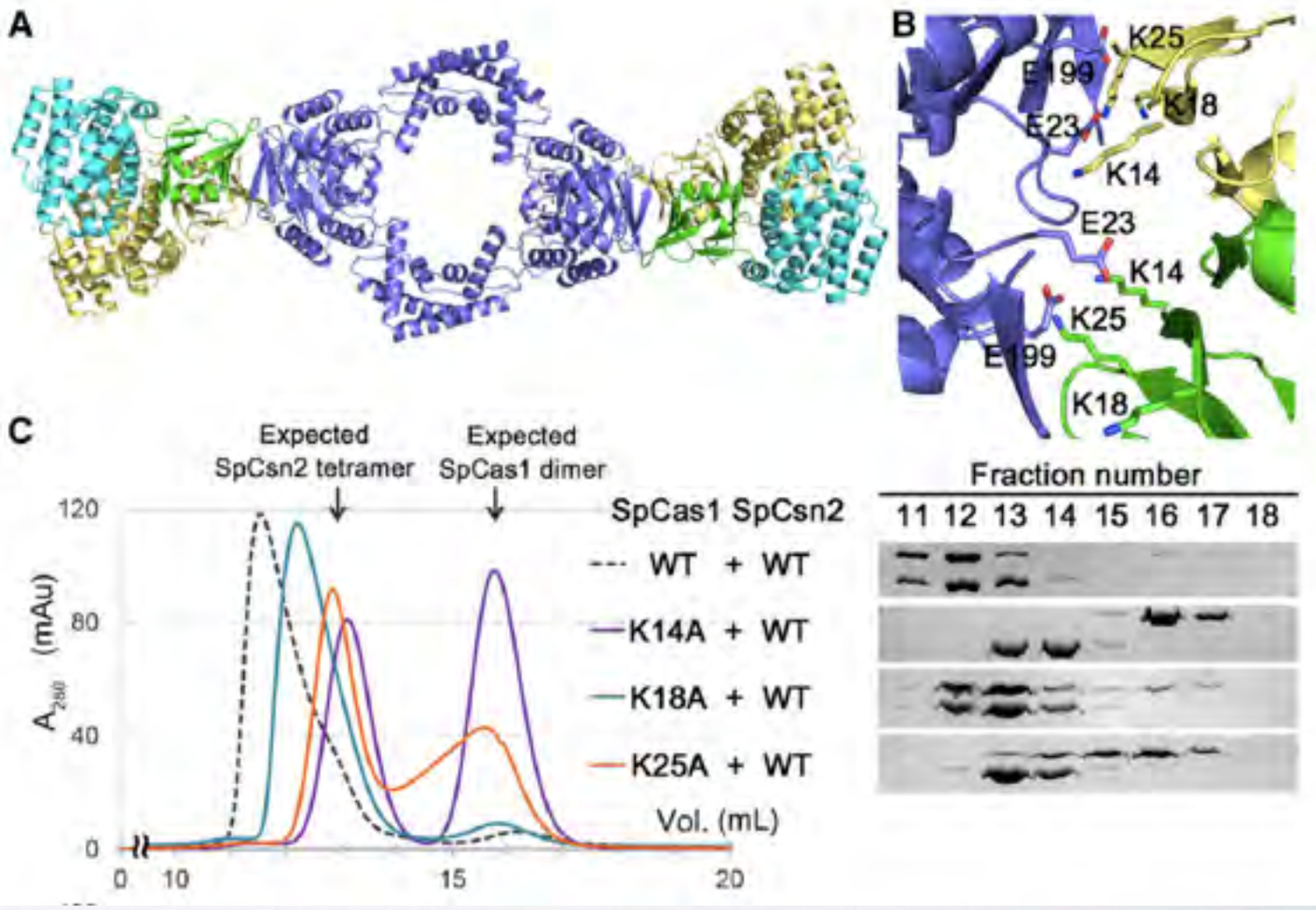
Structure

Volume 24, Issue 1, 5 January 2016, Pages 70–79

Article

Crystal Structure of *Streptococcus pyog*
Interaction with Csn2 in the Type II CRIS

Donghyun Ka[1], Hasup Lee[2], Yi-Deun Jung[3], Kyunggon Kim[4,5], Chac
Bae[1,6,7]



Thank You!