



Demo: I-TASSER Gateway for Protein Structure Prediction and Structure-Based Function Annotation

Chengxin Zhang
Yang Zhang lab, University of Michigan
Gateways 2017
10/25/2017

Outline

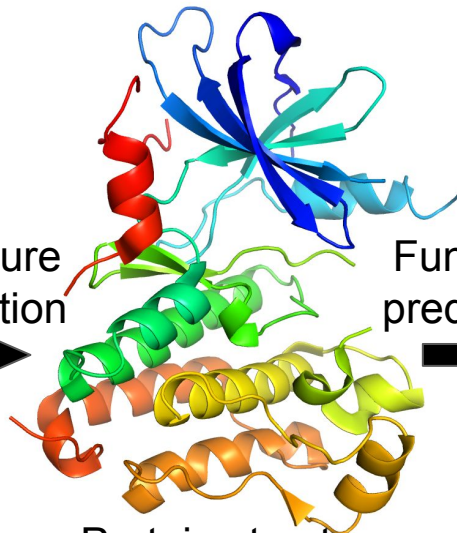
- **How I-TASSER generates protein structure and function predictions using XSEDE Comet?**
- How to use the I-TASSER Gateway and interpret the output results?

Protein structure and function prediction

ALLRILKETELRKVKVLGSGAFGT
VYKGIWIPDGENVKIPVAIKVLRP
KANKEILDEAYVMAGVGSPYVSRL
LGICLTSTVQLVTQLMPYGCLLDH
VRENRGRLGSQDLLNWCMQIAKGM
SYLEDVRLVHRDLAARNVLVKSPN
HVKITDFGLVPIKWMMALESILRRR
FTHQSDVWSYGVTVWELMTFGAKP
YDGIPAREIPDLLEKGERLPQPPI
CTIDVYMIMVKCWMIDSECRPRFR
ELVSEFSRMARDPQRFVVIQNLD
TFYRSILLEDDDLVDAEE

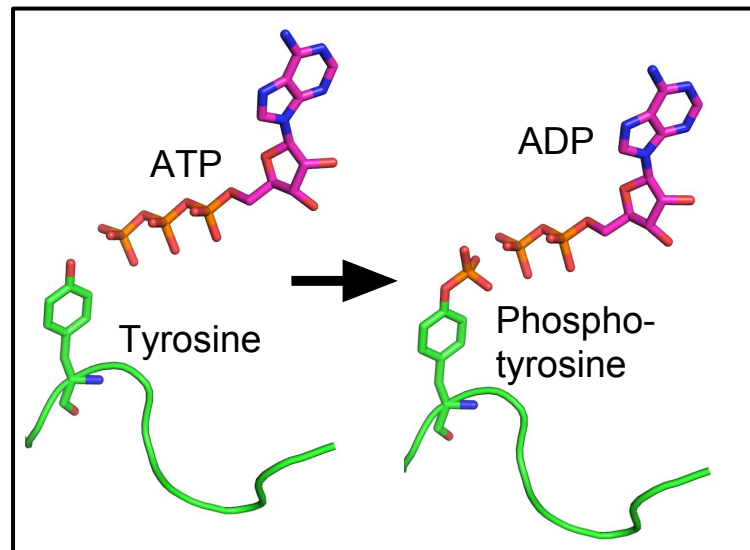
Protein sequence

Structure
prediction

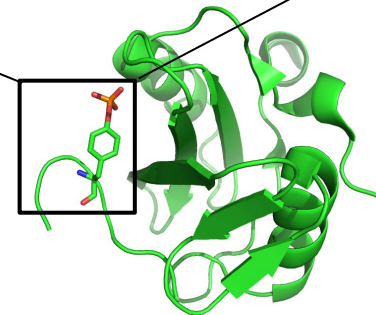


Protein structure

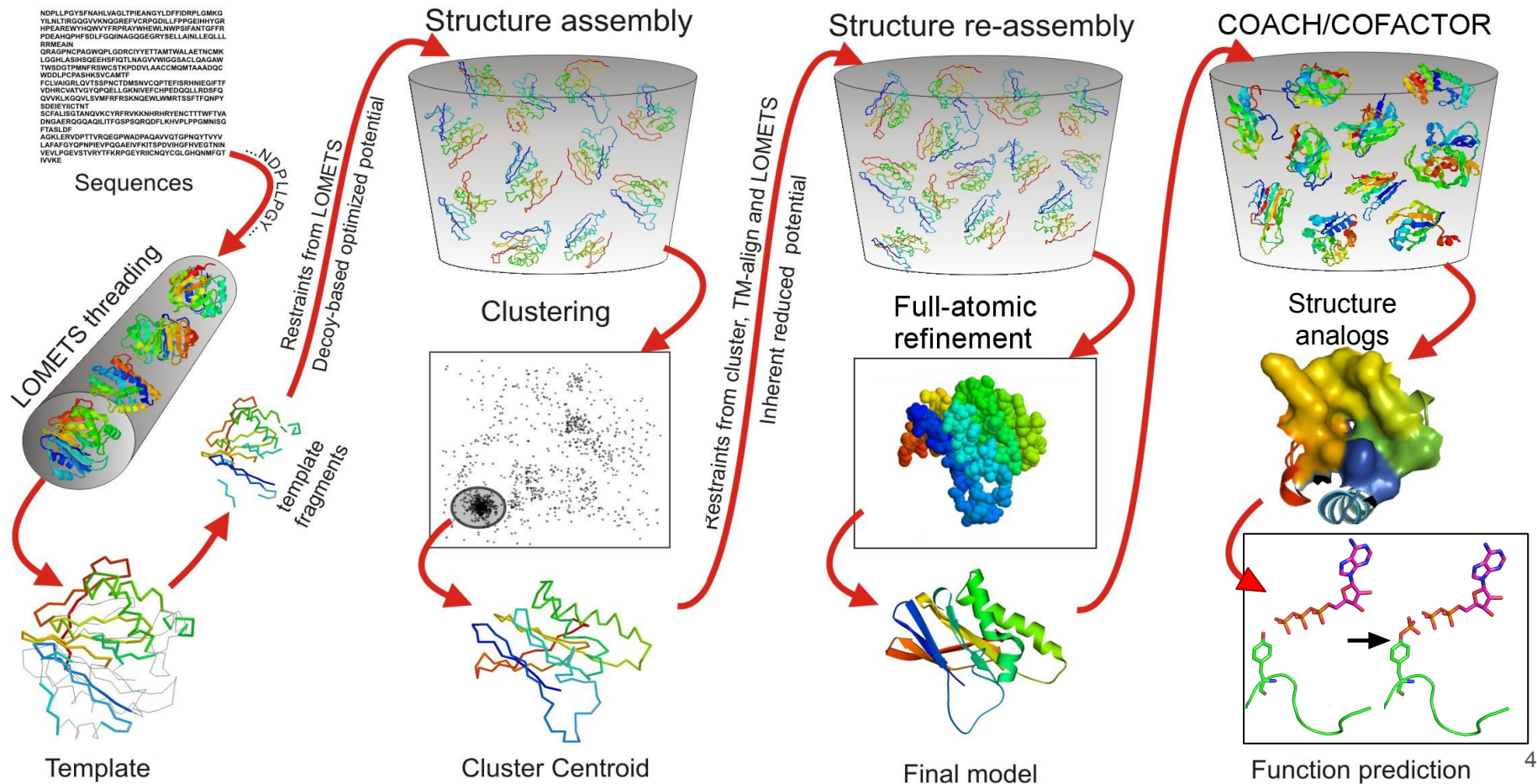
Function
prediction



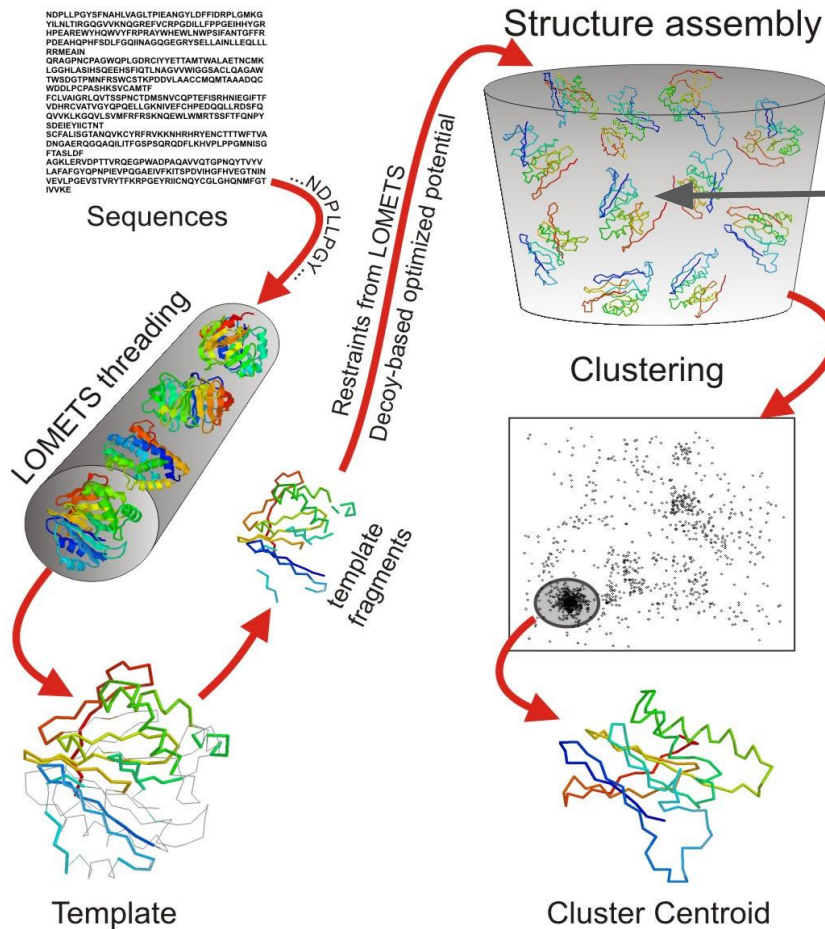
Protein function



I-TASSER Gateway: structure and function prediction



I-TASSER Gateway: energy function for structure prediction

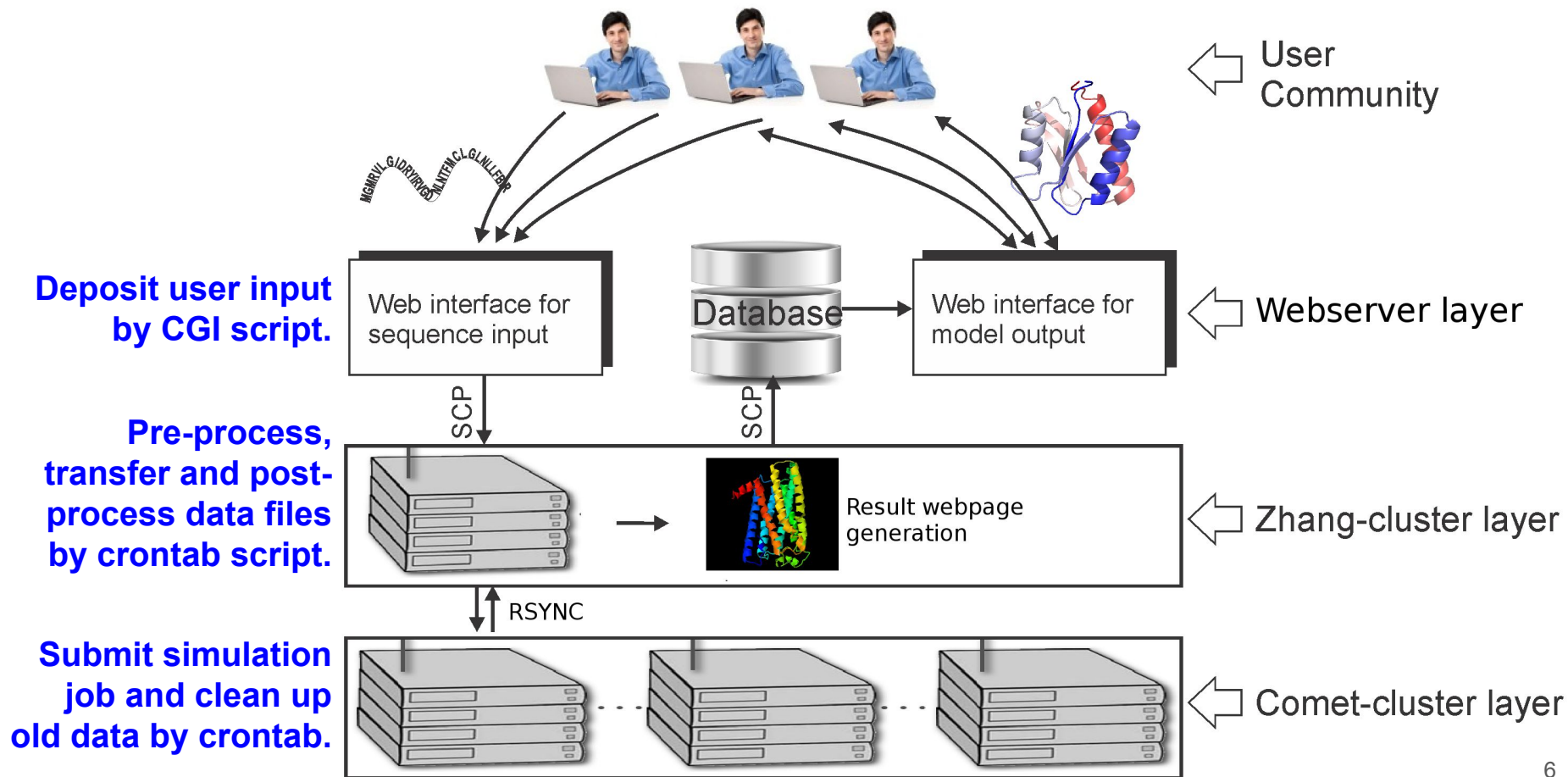


I-TASSER energy function:

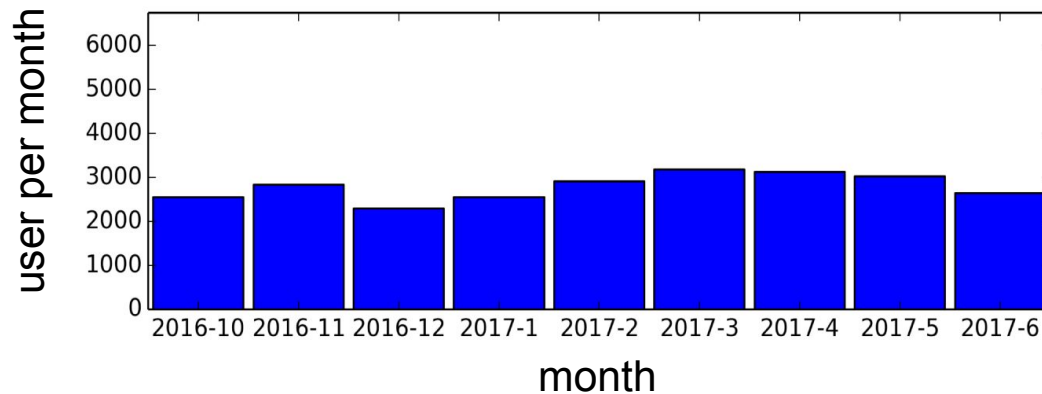
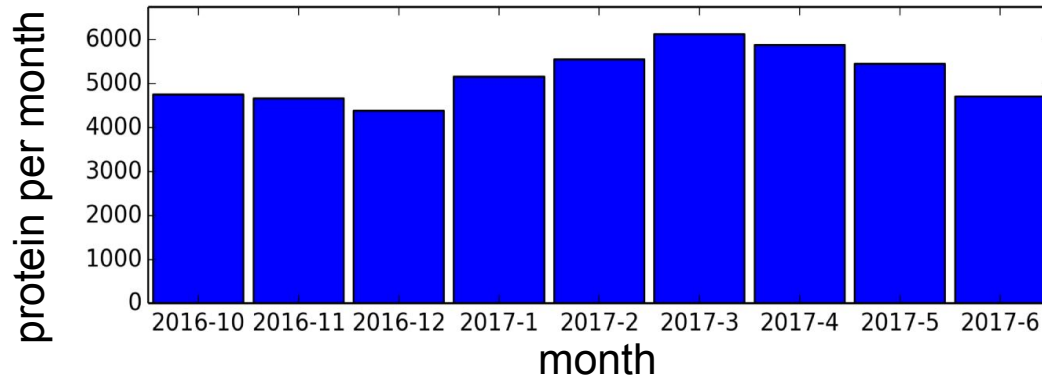
$$E_{\text{I-TASSER}} = E_{\text{knowledge-based}} + E_{\text{threading-template-derived}} + E_{\text{user-specified}}$$

<https://zhanglab.ccmb.med.umich.edu/potential/>

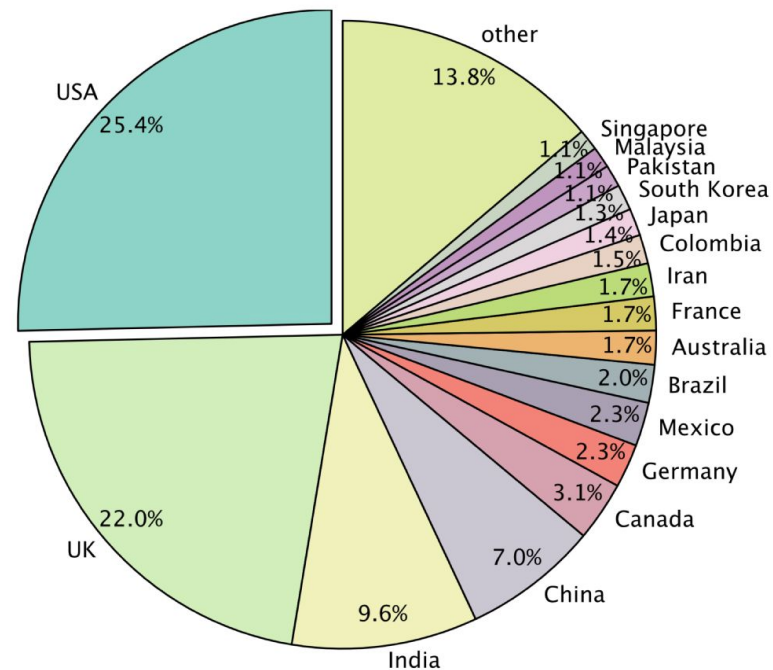
I-TASSER Gateway: integration with XSEDE Comet



I-TASSER Gateway: popularity



Distribution of 87,223 gateway users among 134 countries.



Outline

- How I-TASSER generates protein structure and function predictions using XSEDE Comet?
- **How to use the I-TASSER Gateway and interpret the output results?**
<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>

Acknowledgement

- Jianyi Yang, and Yang Zhang. *Nucleic acids research* 43.W1 (2015): W174-W181.
- Ambrish Roy, Alper Kucukural, and Yang Zhang. *Nature protocols* 5, no. 4 (2010): 725-738.
- Yang Zhang. *BMC bioinformatics* 9, no. 1 (2008): 40.



National Institutes of Health
Turning Discovery Into Health

XSEDE

Extreme Science and Engineering
Discovery Environment

SDSC: Nancy Wilkins-Diehr, Mahidhar Tatineni
UMICH: Katherine Lawrence



PI: Yang Zhang



Ambrish Roy

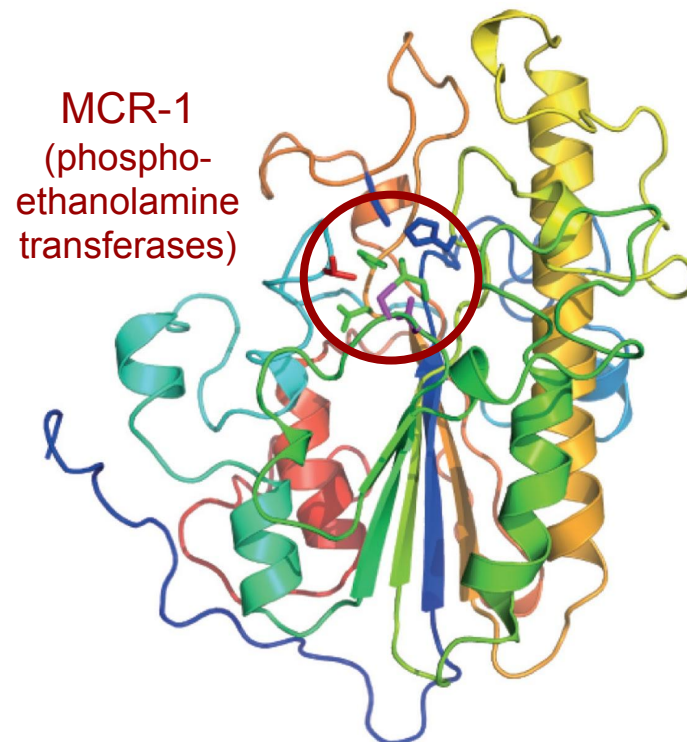
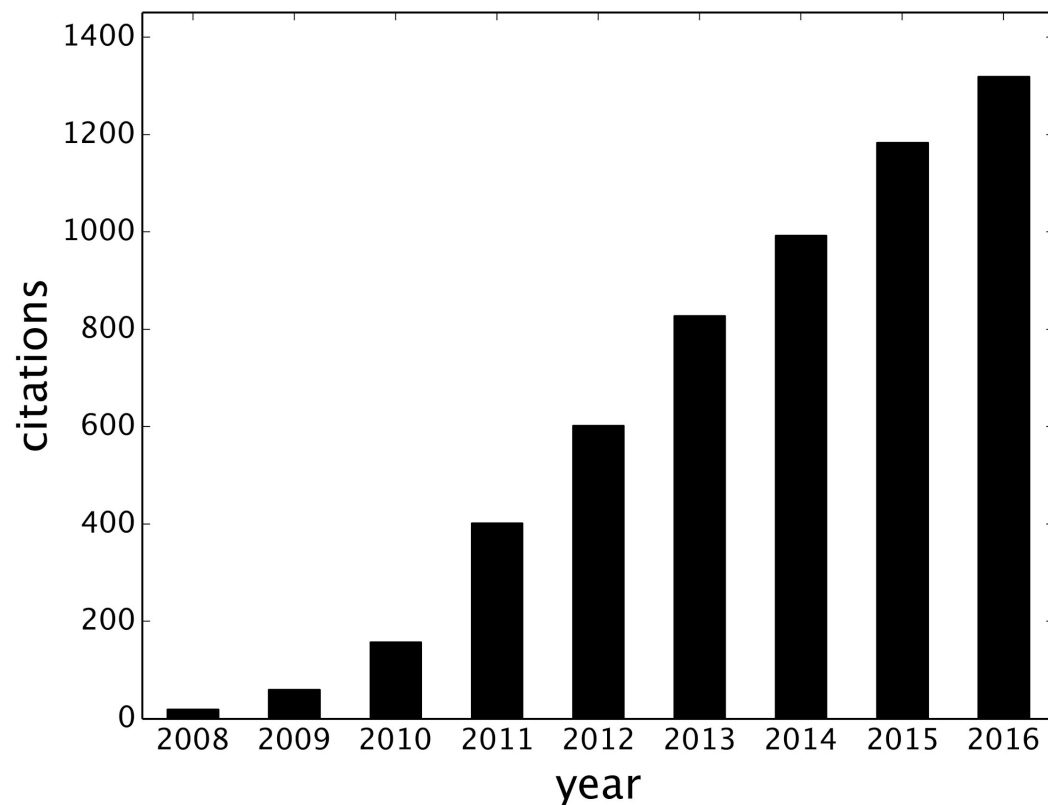


Jianyi Yang



S.M. Mortuza

I-TASSER Gateway: citations



Liu, Yi-Yun, et al.
The Lancet infectious diseases
16.2 (2016): 161-168.

How to submit I-TASSER jobs?

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



I-TASSER

Protein Structure & Function Predictions

(The server completed predictions for 351289 proteins submitted by 85624 users from 134 countries)

(The template library was updated on 2017/09/07)

I-TASSER (Iterative Threading ASSEMBly Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach [LOMETS](#), with full-length atomic models constructed by iterative template fragment assembly simulations. Function insights of the target are then derived by threading the 3D models through protein function database [BioLiP](#). I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide [CASP7](#), [CASP8](#), [CASP9](#), [CASP10](#), [CASP11](#), and [CASP12](#) experiments. It was also ranked as the best for function prediction in [CASP9](#). The server is in active development with the goal to provide the most accurate structural and function predictions using state-of-the-art algorithms. Please report problems and questions at [I-TASSER message board](#) and our members will study and answer the questions asap. ([>> More about the server ...](#))

[\[Queue\]](#) [\[Forum\]](#) [\[Download\]](#) [\[Search\]](#) [\[Registration\]](#) [\[Statistics\]](#) [\[Remove\]](#) [\[Potential\]](#) [\[Decoys\]](#) [\[News\]](#) [\[Annotation\]](#) [\[About\]](#) [\[FAQ\]](#)

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](#):

Or upload the sequence from your local computer:

No file selected.

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

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

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

► **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)

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

Or upload the sequence from your local computer:

No file selected.

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

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

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

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

[\(Read more explanation on how to add restraints\)](#)

☐ Assign contact/distance restraints No file selected.

☐ Specify template without alignment

☐ Specify template without alignment No file selected.

☐ Specify template with alignment No file selected.

How to introduce external restraints?

```
DIST 124 CA 84 CA 17.4  
CONTACT 37 345
```

Atom-atom restraints [Explanation](#)

Type a PDB ID [Explanation](#)

Upload a structure [Explanation](#)

Template & alignment [Explanation](#)

[\[Home\]](#) [\[Server\]](#) [\[Queue\]](#) [\[About\]](#) [\[Remove\]](#) [\[Statistics\]](#)

[Click on [S205619_results.tar.bz2](#) to download the tarball file including all modeling results listed on this page]

(Click on [Annotation of I-TASSER Output](#) to read the instructions for how to interpret the results on this page)

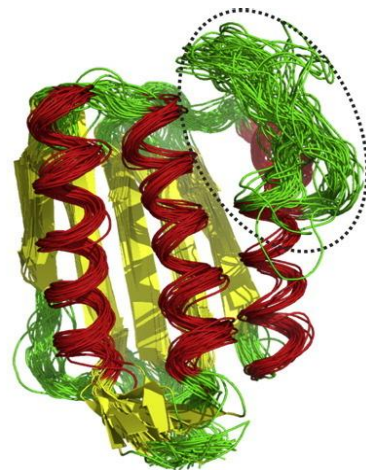
>protein
 IFCLLGFSDDHPRLEAVLFVFLFFYLTLVGNFTIIISYLDPLPHTMYFFLSNLSLLD
 ICFTTSLAPQTLVNLQRPKKTIYGGCVAQLYISLALGSTECILLADMALDRYIAVCKPL
 HYVVMNPRLCQQLASISWGLASSLIHATFTLQLPLCGNHRLDHFICEVPALLKACV
 DTDVNELVLFVVSFLVYPPALISISYGFITQAVRLIKSVLEARKAFKSTCSSHLTVII
 FYGTIITYVYVPSVFAADQGGKISLFYMTVPTLNPITLYLRNDKMKAEALRLKLSGK

[illegible]

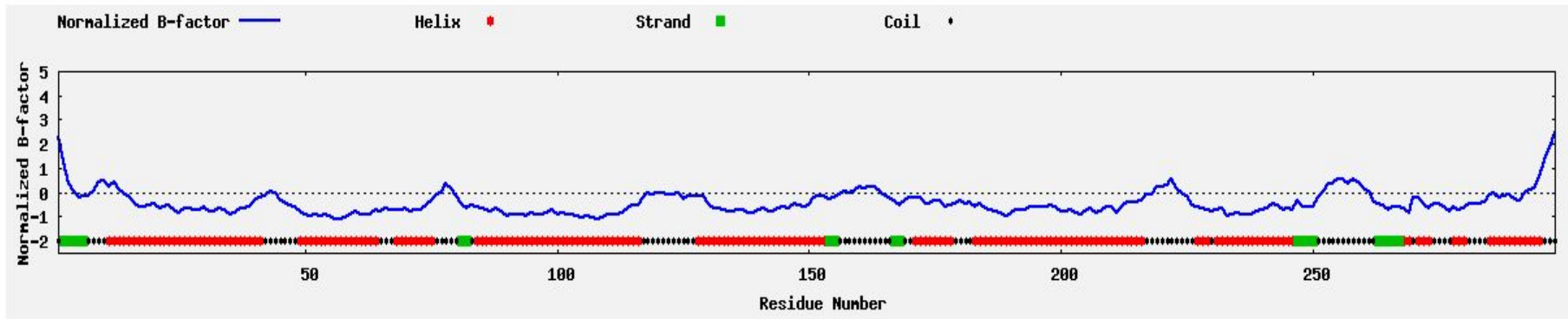
	20	40	60	80	100	120	140
Sequence	DFILLGFSHPRLAEVLFVFLFYLLTLVGNTIIISYLDPLLHTPMYFFLSNLSLLDICFTTSLAPQTLVNLQRPKKITYGGCVAQLYISALGSTECILLADMALDRYIAVCKPLHYVVMNPRLCQQLASISWLS						
Prediction	70000000432400110023033133133332002002001300000010200000010100200300000116623010100220122013002100200010030000000210200000033000000110011						

Values range from 0 (buried residue) to 9 (highly exposed residue)

How to interpret I-TASSER result?

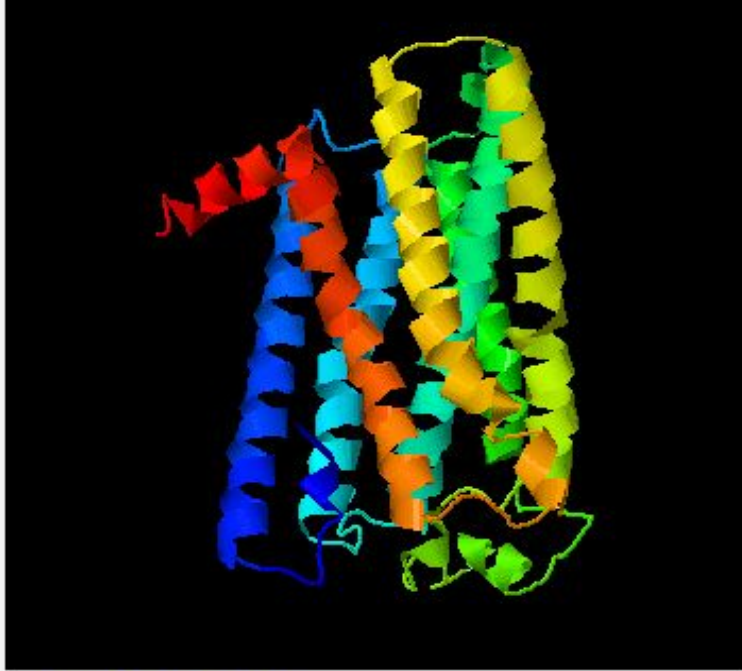


Predicted normalized B-factor

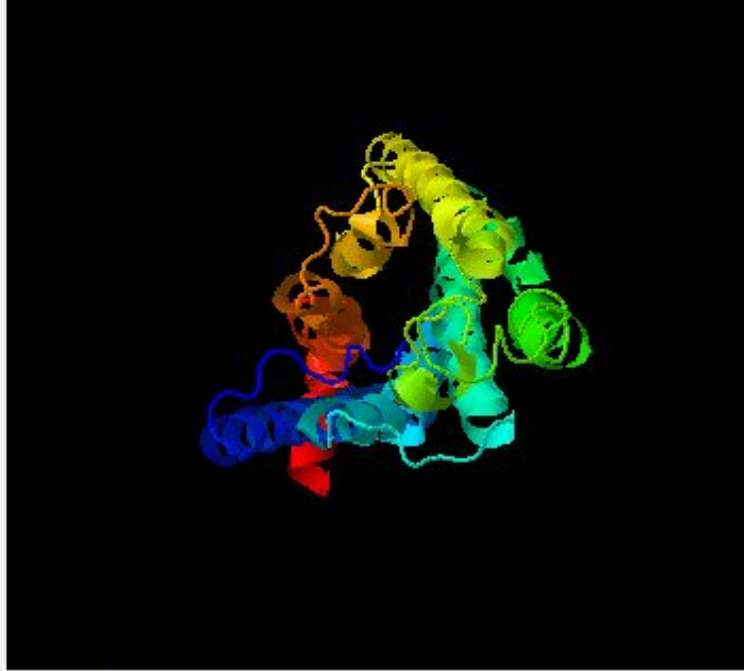


How to interpret I-TASSER results?

Top 5 final models predicted by I-TASSER <http://zhanglab.ccmb.med.umich.edu/I-TASSER/example/>



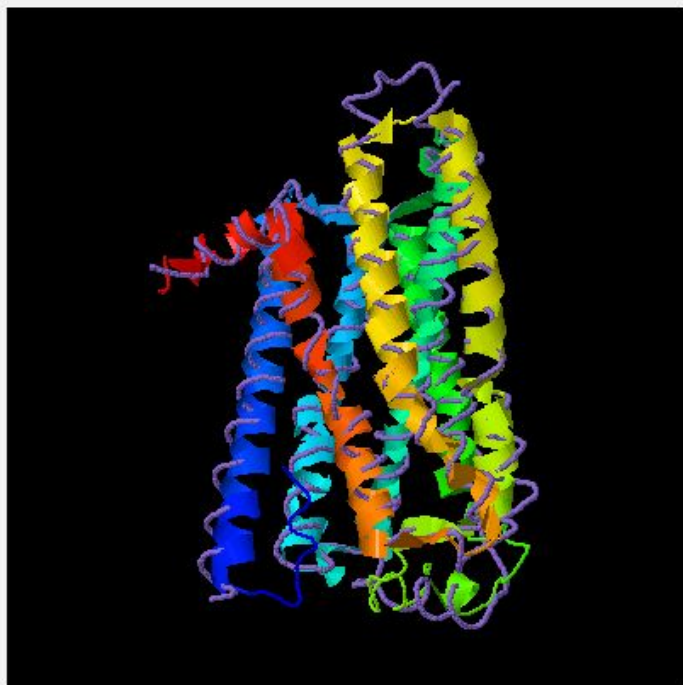
- [Download Model 1](#)
- C-score=-0.08
- Estimated TM-score = 0.70 ± 0.12
- Estimated RMSD = $6.3 \pm 3.9 \text{ \AA}$



- [Download Model 2](#)
- C-score=-0.78

How to interpret I-TASSER results?

Proteins structurally close to the target in the PDB (as identified by [TM-align](#))

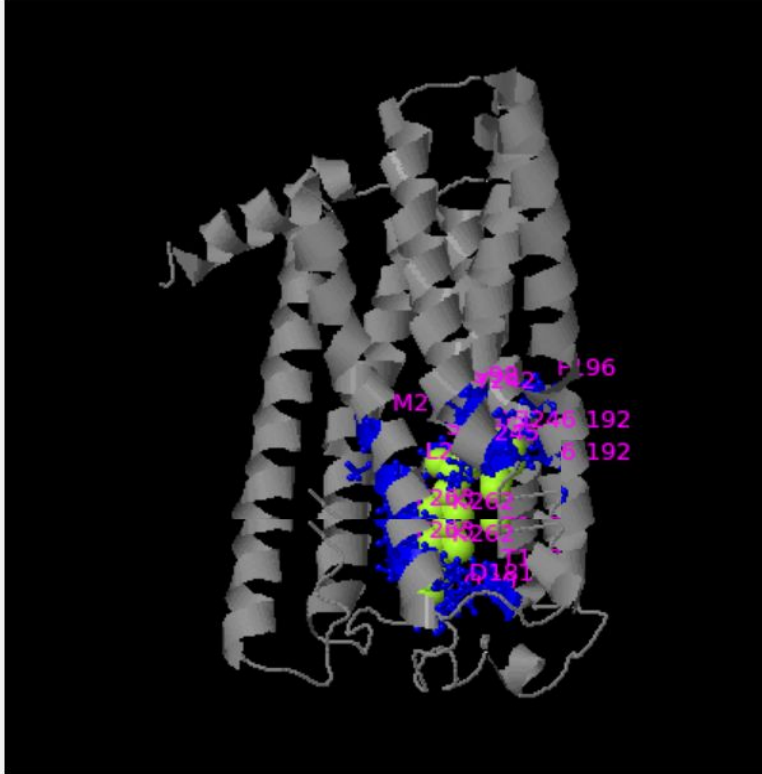


Top 10 Identified structural analogs in PDB






Click to view	Rank	PDB Hit	TM-score	RMSD ^a	IDEN ^a	Cov	Alignment
<input type="radio"/>	1	4bvnA	0.905	1.31	0.165	0.936	Download
<input type="radio"/>	2	2rh1A	0.895	1.63	0.168	0.936	Download
<input type="radio"/>	3	3uonA	0.848	2.14	0.176	0.916	Download
<input type="radio"/>	4	2vt4C	0.847	1.26	0.172	0.876	Download
<input type="radio"/>	5	3emlA	0.846	2.35	0.173	0.923	Download
<input type="radio"/>	6	2ksbA	0.844	2.62	0.143	0.923	Download
<input type="radio"/>	7	4iaqA	0.843	2.00	0.196	0.903	Download
<input type="radio"/>	8	3pblA	0.843	1.43	0.176	0.879	Download
<input type="radio"/>	9	3p0gA	0.842	2.47	0.157	0.926	Download
<input type="radio"/>	10	3rzeA	0.827	1.92	0.171	0.883	Download

How to interpret I-TASSER results?

Ligand binding sites



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

Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
	1	0.29	53	4amiA	G90	Rep , Mult	90,91,94,95,98,180,181,182,192
	2	0.16	36	2y04A	Y01	Rep , Mult	51,55,58,96,128,132,135,139
	3	0.08	24	4ea3A	0NN	Rep , Mult	71,91,94,95,98,99,192,196,242,2
	4	0.05	12	3dqba	PEPTIDE	Rep , Mult	48,112,115,116,118,211,215,218,
	5	0.04	18	2ycwA	2CV	Rep , Mult	208,211,212,225,228,229,235

[Download](#) the residue-specific ligand binding probability, which is estimated by SVM.

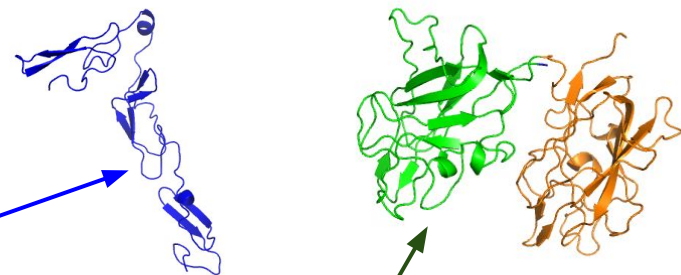
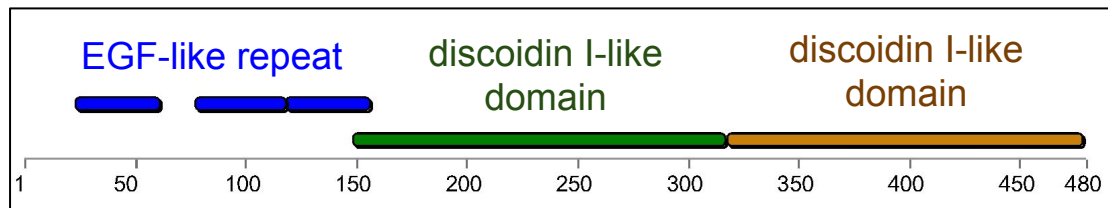
[Download](#) the all possible binding ligands and detailed prediction summary.

Download the templates clustering results.

Outline

- How I-TASSER generates protein structure and function predictions using XSEDE Comet?
- How to use the I-TASSER Gateway and interpret the output results?
- How to improve the quality of the I-TASSER model?
 - Multi-domain protein structure prediction. (Domain partition, domain structure prediction, domain assembly)

How to predict multi-domain protein structure?



Top 10 threading templates used by I-TASSER

Rank	PDB Hit	I den1	I den2	Cov	Norm. Zscore	Download Align.
1	1sddb	0.40	0.27	0.67	3.23	FF3
2	2qqm	0.30	0.26	0.87	2.97	SPX
3	4bxs	0.29	0.28	0.97	2.20	MUS
4	2qqj	0.37	0.24	0.64	2.66	HHP
5	1sdd	0.44	0.29	0.65	7.04	RAP2
6	2qqj	0.37	0.23	0.63	1.81	IIIj
7	1sdd	0.31	0.31	0.98	4.38	IIIb
8	4qz9	0.30	0.28	0.91	2.75	IIIe
9	2r7e	0.31	0.29	0.94	3.65	pge
10	2qqo	0.27	0.24	0.86	1.00	RRR6

Sec.Str Seq	80	100	120	140	160	180	200	220	2
1	CCCCCCCCCCCCCCCC	CCCCCCCCCCCC	CCCCCCCCCCCC	CCCCCCCCCCCC	CCCCCCCCCCCC	CCCCCCCCCCCC	CCCCCCCCCCCC	CCCCCCCCCCCC	CCCCCCCCCCCC
2	DEEPTSGAGCTPNP	CHNGGTCEISEAYR	GDTFIGYVCKCP	RGFNGIHCQHNINE	CEVEPCNGGICT	LDLVANYSCECP	GEFMRNCQYK	CSGPLGIEGGI	ISNQITASSTH
3	DGGMF-----	CRYDRLEIWD	---GFPDVGPH	---IGRYCGKTP	GRISSSGILSM	VFYTDSIAIAK	FSANYSVLQSS	VSFKMEALGME	SIEHSDQITASS
4	EESWYFPCSLHTF	PAINGVHWHLLN	MGGPKDIHVVNF	HGQTFTEEGRE	DNLGVLPPLPG	TFAISKMPKSK	IGTWLLETVR	GMOALFTVIDK	DCKLPMGLAS
5	-----	-----	-----	-----	-----	-----	-----	-----	-----
6	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	-----	-----	-----	-----	-----	-----	-----	-----	-----

ThreaDom ON-LINE

Predicting protein domain boundary by multiple threading alignments

<http://zhanglab.ccmb.med.umich.edu/ThreaDom/>

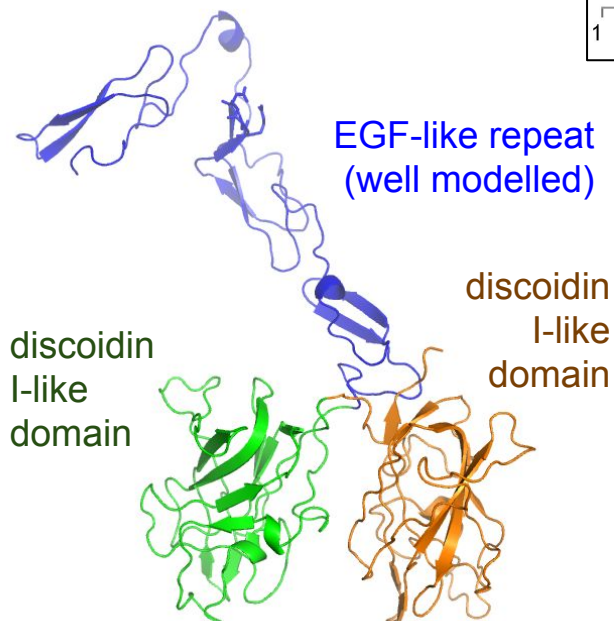
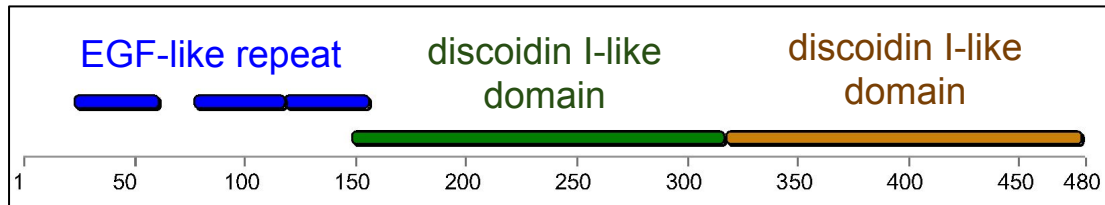
ThreaDomEx

Predicting Protein Domain Boundary and Detecting Discontinuous Domain

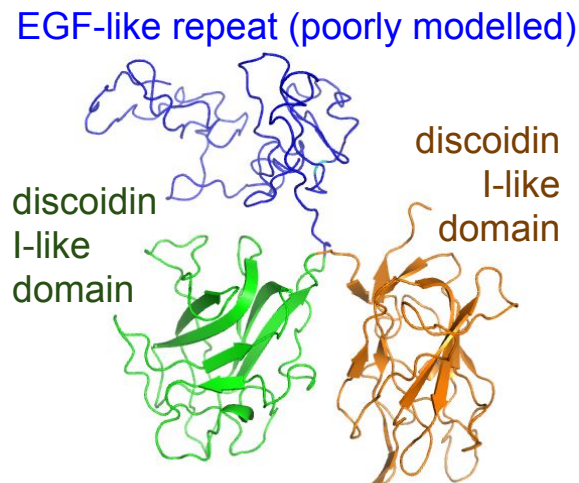
<http://zhanglab.ccmb.med.umich.edu/ThreaDomEx/>²⁰

How to predict multi-domain protein structure?

DEL1 (UniProt EDIL3_HUMAN)

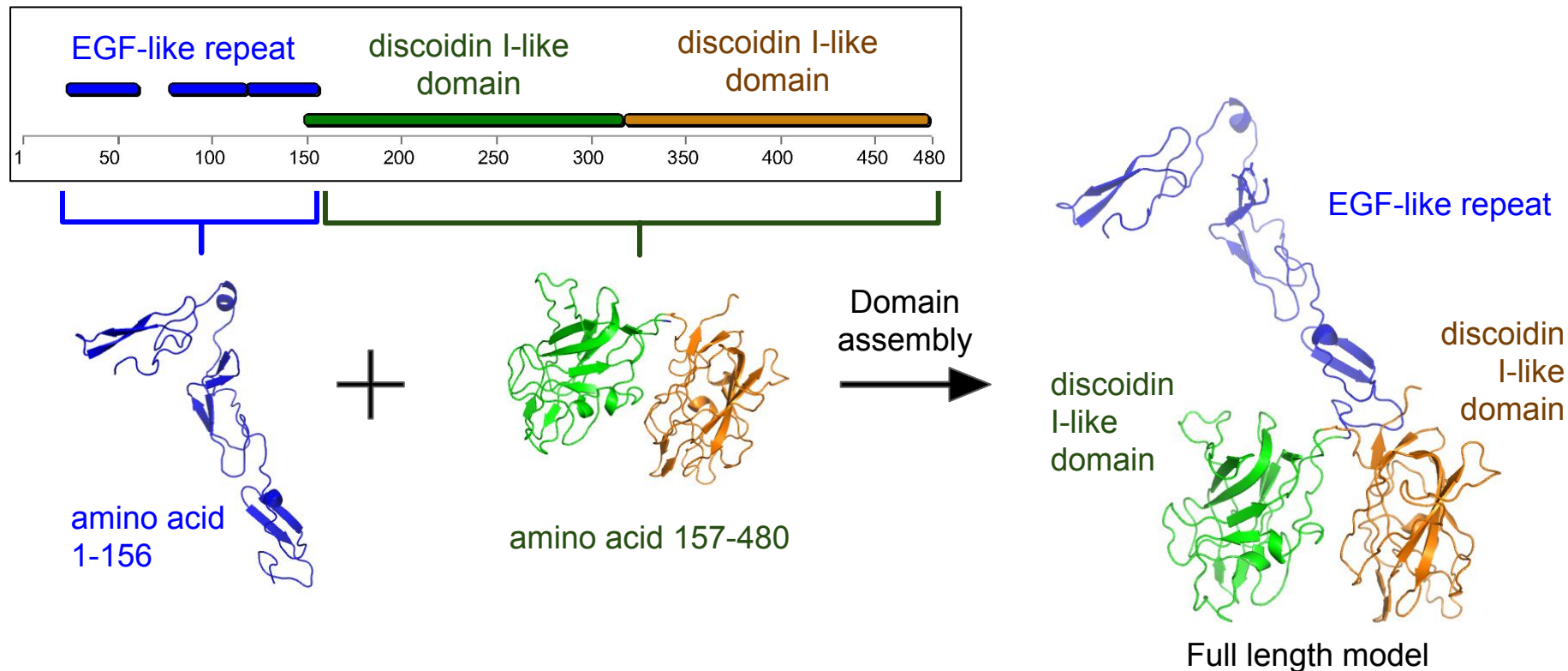


Full length model assembled from individual domains

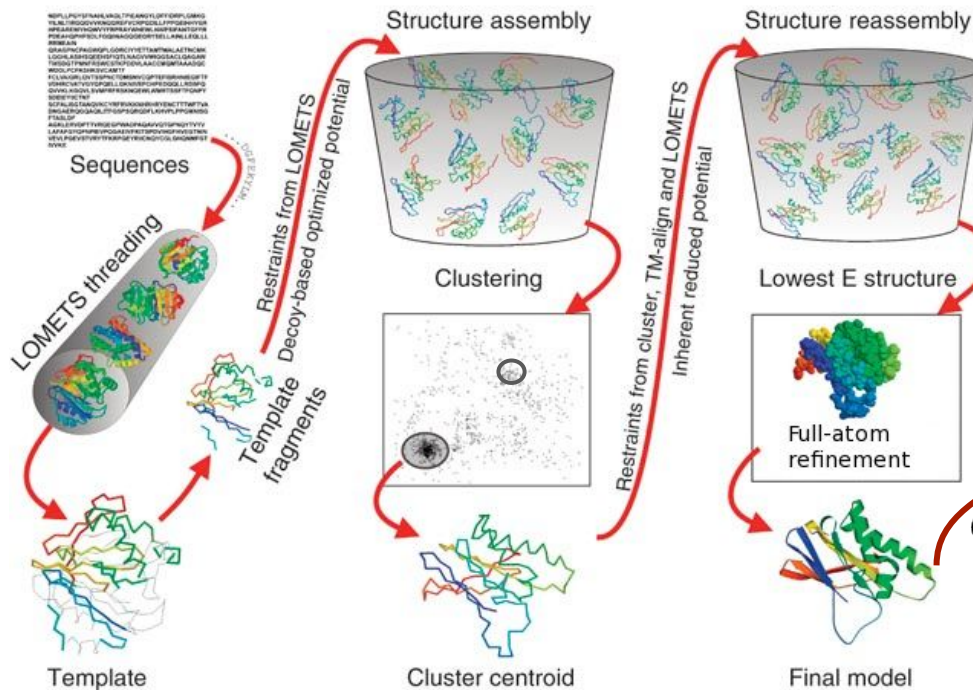


Full length model directly predicted by I-TASSER

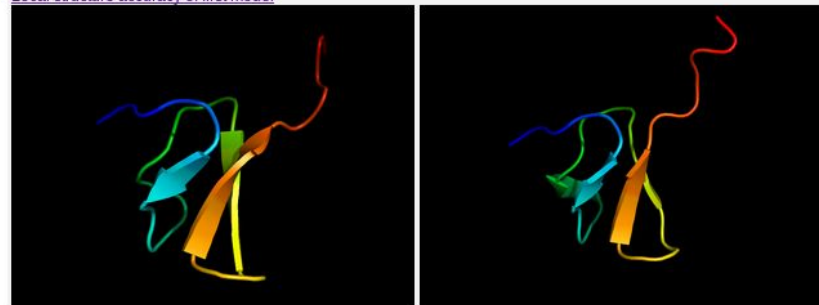
How to predict multi-domain protein structure?



What is C-score?



- [More about C-score](#)
- [Local structure accuracy of first model](#)



- [Download Model 1](#)
- C-score=1.45
 - Estimated TM-score = 0.92 ± 0.06
 - Estimated RMSD = $0.5 \pm 0.5 \text{ \AA}$

- [Download Model 2](#)
- C-score=-0.65

$$Cscore = \ln \left(\frac{M}{M_{tot}} \cdot \frac{1}{\langle RMSD \rangle} \cdot \prod_i \frac{Z_i(i)}{Z_0(i)} \right)$$

$$TMscore = \frac{1}{L} \sum_{i=1}^{L_{ali}} \frac{1}{1 + \left(\frac{d_i}{d_0} \right)}$$

Why does I-TASSER generates < 5 models?

