I-TASSER results for job id S229343

[Click on S229343_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)

Submitted Sequence in FASTA format

```
>seq
MSMAVETFGFFMATVGLLMLGVTLPNSYWRVSTVHGNVITTNTIFENLWFSCATDSLGVY
NCWEFPSMLALSGYIQACRALMITAILLGFLGLLLGIAGLRCTNIGGLELSRKAKLAATA
GALHILAGICGMVAISWYAFNITRDFFDPLYPGTKYELGPALYLGWSASLISILGGLCLC
SACCCGSDEDPAASARRPYQAPVSVMPVATSDQEGDSSFGKYGRNAYV
```

Predicted Secondary Structure

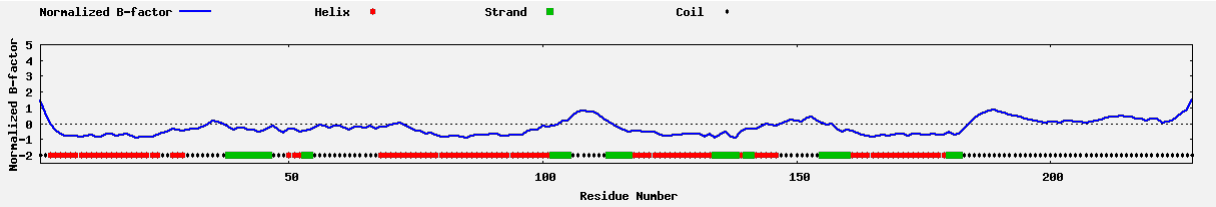
```
                  20        40        60        80        100       120       140       160       180       200       220
                   |         |         |         |         |         |         |         |         |         |         |
Sequence      MSMAVETFGFFMATVGLLMLGVTLPNSYWRVSTVHGNVITTNTIFENLWFSCATDSLGVYNCWEFPSMLALSGYIQACRALMITAILLGFLGLLLGIAGLRCTNIGGLELSRKAKLAATAGALHILAGICGMVAISWYAFNITRDFFDPLYPGTKYELGPALYLGWSASLISILGGLCLCSACCCGSDEDPAASARRPYQAPVSVMPVATSDQEGDSSFGKYGRNAYV
Prediction    CCHHHHHHHHHHHHHHHHHHHHHCCHHHCCCCCCCCSSSSSSSSCCCHHHSSCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHSSSSCCCCCCCSSSSSHHHHHHHHHHHHHHHSSSSSHSSHHHHHCCCCCCCSSSSSSHHHHHHHHHHHHHHHHHHSSSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Conf.Score    9389999999999999999999978874426146898832202553764988478877724557770420005799999999999999999999876244514777340233307989999999997671134320102331488886534560002999999999999999862544467767877667777889998888777788888888888975769
              H:Helix; S:Strand; C:Coil
```

Predicted Solvent Accessibility

```
                  20        40        60        80        100       120       140       160       180       200       220
                   |         |         |         |         |         |         |         |         |         |         |
Sequence      MSMAVETFGFFMATVGLLMLGVTLPNSYWRVSTVHGNVITTNTIFENLWFSCATDSLGVYNCWEFPSMLALSGYIQACRALMITAILLGFLGLLLGIAGLRCTNIGGLELSRKAKLAATAGALHILAGICGMVAISWYAFNITRDFFDPLYPGTKYELGPALYLGWSASLISILGGLCLCSACCCGSDEDPAASARRPYQAPVSVMPVATSDQEGDSSFGKYGRNAYV
Prediction    4210001001311331221000000123121114233313331213303322033231333033121213333313021000000113113022002201210211444542211001100110121132002000000120233233432453323211001101201331331121000001344575444444333443444344444474644456243641 4
              Values range from 0 (buried residue) to 9 (highly exposed residue)
```

Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by B=(B'-u)/s, where B' is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. Click here to read more about predicted normalized B-factor)



Top 10 threading templates used by I-TASSER

(I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.
Sec.Str						CCHHHHHHHHHHHHHHHHHHHHHCCHHHCCCCCCCCSSSSSSSSSCCCHHHSSCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHSSSSCCCCCCCSSSSSHHHHHHHHHHHHHHHSSSSSHSSHHHHHCCCCCCCSSSSSSHHHHHHHHHHHHHHHHHHSSSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Seq						MSMAVETFGFFMATVGLLMLGVTLPNSYWRVSTVHGNVITTNTIFENLWFSCATDSLGVYNCWEFPSMLALSGYIQACRALMITAILLGFLGLLLGIAGLRCTNIGGLELSRKAKLAATAGALHILAGICGMVAISWYAFNITRDFFDPLYPGTKYELGPALYLGWSASLISILGGLCLCSACCCGSDEDPAASARRPYQAPVSVMPVATSDQEGDSSFGKYGRNAYV
1	4or2A	0.10	0.15	0.94	1.11	Download
2	3a0oA	0.09	0.16	0.95	1.11	Download
3	4l6rA	0.10	0.20	0.98	1.04	Download
4	4rdqA	0.11	0.20	1.00	1.10	Download
5	4k5yA	0.11	0.15	0.89	1.01	Download
6	4kzsA	0.09	0.18	0.95	1.10	Download
7	4um2A	0.07	0.19	0.96	1.08	Download
8	4l8tA	0.06	0.18	0.92	1.06	Download
9	4k5yA	0.13	0.15	0.96	0.92	Download
10	4xnwA	0.10	0.15	0.82	0.68	Download

(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. (more about the colors used)

(b) Rank of templates represents the top ten threading templates used by I-TASSER.

(c) Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.

(d) Iden2 is the percentage sequence identity of the whole template chains with query sequence.

(e) Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.

(f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

(g) Download Align. provides the 3D structure of the aligned regions of the threading templates.

(h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:

1: Neff-PPAS 2: PROSPECT2 3: Neff-PPAS 4: PROSPECT2 5: Neff-PPAS 6: PROSPECT2 7: PROSPECT2 8: PROSPECT2 9: MUSTER 10: FFAS-3D

Top 5 final models predicted by I-TASSER

(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-clicking on the images, you can change the configurations, e.g. modifying the background color or stopping the spin of your models)

- Download Model 1
- C-score=-2.48 (Read more about C-score)
- Estimated TM-score = 0.43±0.14
- Estimated RMSD = 11.3±4.5Å

- Download Model 2
- C-score = -4.53

- Download Model 3
- C-score = -3.91

- Download Model 4
- C-score = -5

- Download Model 5
- C-score = -5

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

(After the structure assembly simulation, I-TASSER uses the TM-align structural alignment program to match the first I-TASSER model to all structures in the PDB library. This section reports the top 10 proteins from the PDB that have the closest structural similarity, i.e. the highest TM-score, to the predicted I-TASSER model. Due to the structural similarity, these proteins often have similar function to the target. However, users are encouraged to use the data in the next section 'Predicted function using COACH' to infer the function of the target protein, since COACH has been extensively trained to derive biological functions from multi-source of sequence and structure features which has on average a higher accuracy than the function annotations derived only from the global structure comparison.)

Top 10 Identified stuctural analogs in PDB

Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
	1	4l6rA	0.773	3.11	0.079	0.939	Download
	2	4k5yA	0.749	3.24	0.065	0.934	Download
	3	3oduA	0.738	3.31	0.060	0.925	Download
	4	4xnwA	0.736	3.62	0.092	0.956	Download
	5	4jkvA	0.735	3.51	0.078	0.934	Download
	6	4ib4A	0.735	3.50	0.070	0.943	Download
	7	4mbsA	0.730	3.67	0.065	0.947	Download
	8	4or2A2	0.727	3.36	0.082	0.908	Download
	9	3qakA	0.727	3.68	0.060	0.952	Download
	10	2ziyA	0.727	3.74	0.070	0.934	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.
(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

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 results from the COFACTOR, TM-SITE and S-SITE programs.)

Ligand binding sites

Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
	1	0.03	9	2y04B	2CV	Rep, Mult	12,13
	2	0.03	9	3zpgA	2CV	Rep, Mult	51,88,91,95,98
	3	0.02	5	3A0BC	3A0BC14	Rep, Mult	157,166,216,224
	4	0.02	6	4amiA	2CV	Rep, Mult	78,82,85
	5	0.02	4	3nyaA	CLR	Rep, Mult	62,81,85,88

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 more reliable prediction.
(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 database.
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column.
Mult is the complex structures with all potential binding ligands in the cluster.

Enzyme Commission (EC) numbers and active sites

Click to view	Rank	Cscore[EC]	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	EC Number	Active Site Residues
	1	0.154	3d4sA	0.711	3.85	0.041	0.947	3.2.1.17	NA
	2	0.123	1m56A	0.491	5.21	0.050	0.789	1.9.3.1	NA
	3	0.116	1qleA	0.562	4.49	0.051	0.789	1.9.3.1	NA
	4	0.114	2occN	0.559	4.58	0.063	0.803	1.9.3.1	11

5	0.114	1occA	0.558	4.58	0.063	0.803	1.9.3.1	NA	

Click on the radio buttons to visualize predicted active site residues.

(a) $Cscore^{EC}$ is the confidence score for the EC number prediction. $Cscore^{EC}$ values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) $RMSD^{a}$ is the RMSD between residues that are structurally aligned by TM-align.

(d) $IDEN^{a}$ is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

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	0.16	0.7189	3.80	0.08	0.96	3emlA	GO:0003796 GO:0007186 GO:0009253 GO:0016021 GO:0016998 GO:0016798 GO:0008152 GO:0042742 GO:0019835 GO:0016787 GO:0003824
2	0.16	0.7065	3.89	0.07	0.93	3rzeA	GO:0005737 GO:0004872 GO:0007186 GO:0016021 GO:0010894 GO:0042742 GO:0008152 GO:0016798 GO:0003796 GO:0019835 GO:0003824 GO:0009253 GO:0016787 GO:0016998 GO:0004930 GO:0009629 GO:0045907 GO:0005887 GO:0045429 GO:0006954 GO:0005730 GO:0004969 GO:0051381 GO:0007165 GO:0007200 GO:0004871 GO:0005634 GO:0016020 GO:0005624 GO:0007268 GO:0005886
3	0.16	0.7384	3.31	0.06	0.93	3oduA	GO:0019835 GO:0016998 GO:0003796 GO:0016798 GO:0008152 GO:0042742 GO:0016787 GO:0003824 GO:0009253 GO:0007186 GO:0016021
4	0.16	0.7105	3.89	0.08	0.95	2ks9A	GO:0004995 GO:0005886 GO:0007186 GO:0016021
5	0.15	0.7265	3.74	0.07	0.93	2zjyA	GO:0004871 GO:0016021 GO:0007601 GO:0018298 GO:0007186 GO:0007602 GO:0004872 GO:0009881 GO:0007165 GO:0004930 GO:0050896 GO:0016020
6	0.10	0.3908	5.85	0.09	0.66	3atlA	GO:0016829
7	0.09	0.3622	5.46	0.04	0.59	2hp3A	GO:0047547 GO:0019543
8	0.07	0.2933	5.65	0.08	0.50	2qb0B	GO:0003796 GO:0005515 GO:0005634 GO:0009253 GO:0016998 GO:0043565
9	0.06	0.1623	1.55	0.00	0.17	2x7rC	GO:0005198 GO:0019031
10	0.06	0.1526	1.13	0.06	0.16	1tiiC	GO:0003824 GO:0005576 GO:0009405

Consensus prediction of GO terms

Molecular Function	GO:0003796	GO:0008227	GO:0043176	GO:0008188						
GO-Score	0.41	0.33	0.33	0.31						
Biological Process	GO:0007186	GO:0019835	GO:0016998	GO:0009253	GO:0042742	GO:0045428	GO:0051240	GO:0019229	GO:0007267	GO:0019226
GO-Score	0.58	0.41	0.41	0.41	0.41	0.33	0.33	0.33	0.33	0.33
Cellular Component	GO:0016021	GO:0005626	GO:0031981	GO:0043232	GO:0031226					
GO-Score	0.58	0.33	0.33	0.33	0.33					

(a) $Cscore^{GO}$ is a combined measure for evaluating global and local similarity between query and template protein. It's range is [0-1] and higher values indicate more confident predictions.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) $RMSD^{a}$ is the RMSD between residues that are structurally aligned by TM-align.

(d) $IDEN^{a}$ is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

(f) The second table shows a consensus GO terms amongst the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on $Cscore^{GO}$ of the template.

[Click on S229343_results.tar.bz2 to download the tarball file including all modeling results listed on this page]

Please cite the following articles when you use the I-TASSER server:

1. J Yang, R Yan, A Roy, D Xu, J Poisson, Y Zhang. The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12:7-8, 2015.
2. A Roy, A Kucukural, Y Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5:725-738, 2010.
3. Y Zhang. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9:40, 2008.