# I-TASSER results for job id S613477

**[Click on S613477_results.tar.bz2 to download the tarball file including all modeling results listed on this page]**

## Submitted Sequence in FASTA format

```
>seq
MANSGLQLLGFSMALLGWVGLVACTAIPQWQMSSYAGDNIITAQAMYKGLWMDCVTQSTG
MMSCKMYDSVLALSAALQATRALMVVSLVLGFLAMFVATMGMKCTRCGGDDKVKKARIAM
GGGIIFIVAGLAALVACSWYGHQIVTDFYNPLIPTNIKYEFGPAIFIGWAGSALVILGGA
LLSCSCPGNESKAGYRVPRSYPKSNSSKEYV
```

## Predicted Secondary Structure

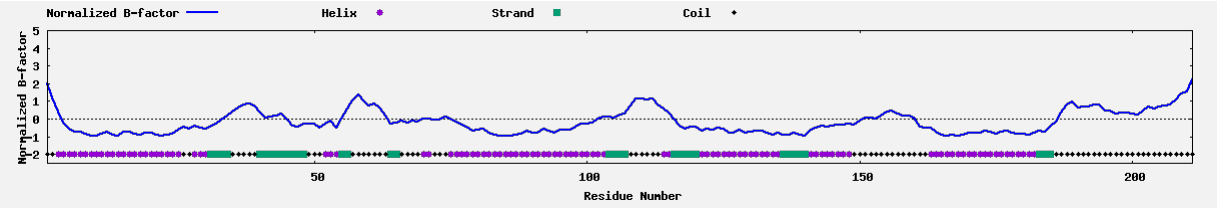
```
                  20        40        60        80        100       120       140       160       180       200
                   |         |         |         |         |         |         |         |         |         |
Sequence      MANSGLQLLGFSMALLGWVGLVACTAIPQWQMSSYAGDNIITAQAMYKGLWMDCVTQSTGMMSCKMYDSVLALSAALQATRALMVVSLVLGFLAMFVATMGMKCTRCGGDDKVKKARIAMGGGIIFIVAGLAALVACSWYGHQIVTDFYNPLIPTNIKYEFGPAIFIGWAGSALVILGGALLSCSCPGNESKAGYRVPRSYPKSNSSKEYV
Prediction    CCHHHHHHHHHHHHHHHHHHHHHHCCHHHSSSSCCCCCSSSSSSSSSCCCHHHSSCCCCCCCSSCCCCHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHSSSSCCCCCCHHSSSSSHHHHHHHHHHHHHHSSSSSHHHHHHHHCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHSSSCCCCCCCCCCCCCCCCCCCCCCCCCC
Conf.Score    952788999999999999999996874445820689850011248877778675788998423687502013168889999999999999999999875204414767451125510436999999987533134556655887518777765201041059999999999999873565789877778988888999998899
              H:Helix; S:Strand; C:Coil
```

## Predicted Solvent Accessibility

```
                  20        40        60        80        100       120       140       160       180       200
                   |         |         |         |         |         |         |         |         |         |
Sequence      MANSGLQLLGFSMALLGWVGLVACTAIPQWQMSSYAGDNIITAQAMYKGLWMDCVTQSTGMMSCKMYDSVLALSAALQATRALMVVSLVLGFLAMFVATMGMKCTRCGGDDKVKKARIAMGGGIIFIVAGLAALVACSWYGHQIVTDFYNPLIPTNIKYEFGPAIFIGWAGSALVILGGALLSCSCPGNESKAGYRVPRSYPKSNSSKEYV
Prediction    732100010013113311200000001221102223443222321312134032210332313330332222232343230220000001311330320022012103114546423120011001111212311100000101302332223323443333211000001203313311210000023466664343444445674466 27
              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 CCHHHHHHHHHHHHHHHHHHHHHHCCHHHSSSSCCCCCSSSSSSSSSCCCHHHSSCCCCCCCSSCCCCHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHSSSSCCCCCCHHSSSSSHHHHHHHHHHHHHHSSSSSHHHHHHHHCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHSSSCCCCCCCCCCCCCCCCCCCCCCCCCC |
| | | | | | | Seq MANSGLQLLGFSMALLGWVGLVACTAIPQWQMSSYAGDNIITAQAMYKGLWMDCVTQSTGMMSCKMYDSVLALSAALQATRALMVVSLVLGFLAMFVATMGMKCTRCGGDDKVKKARIAMGGGIIFIVAGLAALVACSWYGHQIVTDFYNPLIPTNIKYEFGPAIFIGWAGSALVILGGALLSCSCPGNESKAGYRVPRSYPKSNSSKEYV |
| 1 | 6akfA | 0.49 | 0.42 | 0.87 | 2.59 | Download |
| 2 | 6akfA | 0.49 | 0.42 | 0.86 | 2.80 | Download |
| 3 | 6ov2A | 0.47 | 0.41 | 0.86 | 2.82 | Download |
| 4 | 6akf | 0.48 | 0.42 | 0.87 | 4.08 | Download |
| 5 | 6ov2 | 0.48 | 0.41 | 0.86 | 3.03 | Download |
| 6 | 6akfA | 0.49 | 0.42 | 0.86 | 3.33 | Download |
| 7 | 6ov2 | 0.48 | 0.41 | 0.86 | 4.55 | Download |
| 8 | 4p79A | 0.37 | 0.32 | 0.85 | 2.92 | Download |
| 9 | 6akfA | 0.49 | 0.42 | 0.87 | 2.87 | Download |
| 10 | 6ov2A | 0.48 | 0.41 | 0.86 | 2.59 | 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) Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.

(d) Ident2 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.

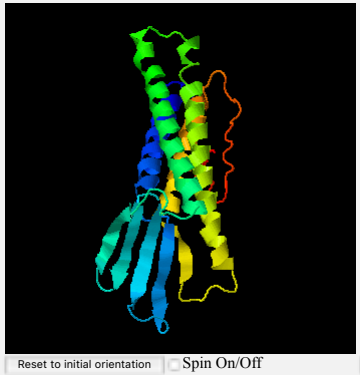
(h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:
1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: HHSEARCH I 6: Neff-PPAS 7: HHSEARCH 8: pGenTHREADER 9: wdPPAS 10: PROSPECT2

## 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.)

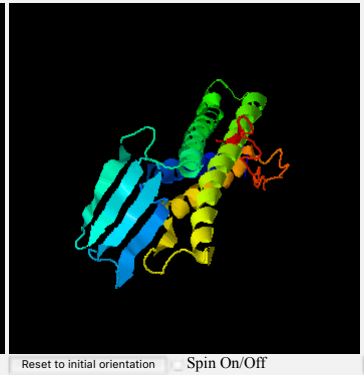
- More about C-score
- Local structure accuracy profile of the top five models

(By right-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)



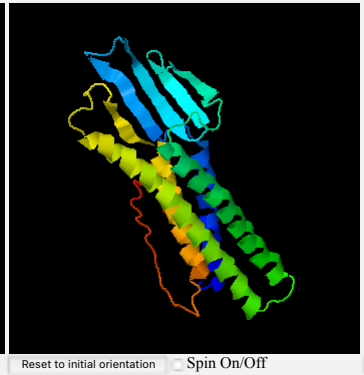
Reset to initial orientation Spin On/Off

- Download Model 1
- C-score=-0.46 (Read more about C-score)
- Estimated TM-score = 0.65±0.13
- Estimated RMSD = 6.4±3.9Å



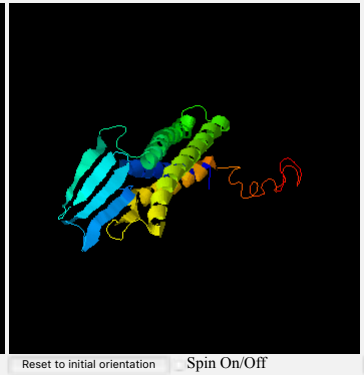
Reset to initial orientation Spin On/Off

- Download Model 2
- C-score = -0.89



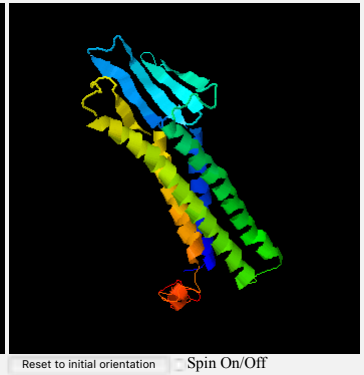
Reset to initial orientation Spin On/Off

- Download Model 3
- C-score = -0.71



Reset to initial orientation Spin On/Off

- Download Model 4
- C-score = -4.26

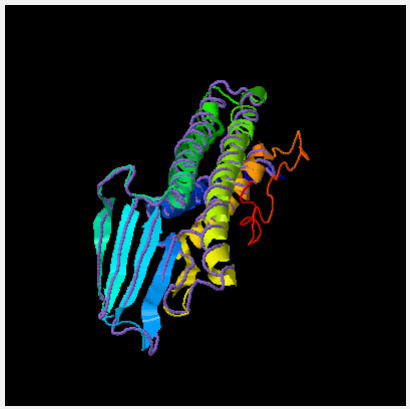


Reset to initial orientation Spin On/Off

- Download Model 5
- C-score = -4.33

## 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.)



Reset to initial orientation Spin On/Off

**Top 10 Identified stuctural analogs in PDB**

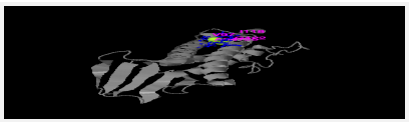
| Click to view | Rank | PDB Hit | TM-score | RMSD[a] | IDEN[a] | Cov | Alignment |
|---|---|---|---|---|---|---|---|
| | 1 | 6akfA | 0.843 | 1.07 | 0.484 | 0.872 | Download |
| | 2 | 6ov2A | 0.813 | 1.57 | 0.473 | 0.863 | Download |
| | 3 | 5b2gA | 0.774 | 2.02 | 0.438 | 0.834 | Download |
| | 4 | 3x29A | 0.749 | 1.32 | 0.536 | 0.787 | Download |
| | 5 | 4p79A | 0.735 | 2.28 | 0.357 | 0.810 | Download |
| | 6 | 5vhxE | 0.648 | 3.28 | 0.109 | 0.801 | Download |
| | 7 | 6qkcl | 0.627 | 2.88 | 0.184 | 0.739 | Download |
| | 8 | 6c14B | 0.622 | 2.59 | 0.115 | 0.739 | Download |
| | 9 | 6wxrB | 0.599 | 3.55 | 0.150 | 0.801 | Download |
| | 10 | 3jbrE | 0.588 | 1.95 | 0.110 | 0.644 | 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 COFACTOR and COACH

(This section reports biological annotations of the target protein by COFACTOR and COACH based on the I-TASSER structure prediction. While COFACTOR deduces protein functions (ligand-binding sites, EC and GO) using structure comparison and protein-protein networks, COACH is a meta-server approach that combines multiple function annotation results (on ligand-binding sites) 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 | 0.11 | 5 | 2y02A | 2CV | Rep, Mult | 94,97,118,122,125 |
| | 2 | 0.07 | 3 | 2q8hA | TF4 | Rep, Mult | 21,84,129,168,171 |
| | 3 | 0.07 | 3 | 5d51L | KR | Rep, Mult | 79,80,83,136,137,140 |

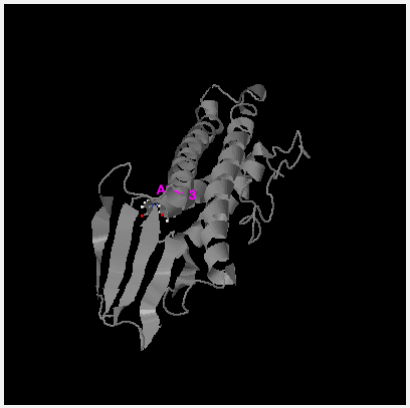
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 0.03 | 1 | 2w6dA CPL | Rep, Mult | 88,89,90,91,126 |
| | 5 | 0.02 | 1 | 5d51L KR | Rep, Mult | 6,9,10,181 |

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.

Reset to initial orientation Spin On/Off

## 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 | 0.211 | 2pfdB | 0.555 | 3.88 | 0.040 | 0.716 | 2.1.2.5 4.3.1.4 | 23,26 |
| | 2 | 0.210 | 2o6yA | 0.516 | 5.18 | 0.077 | 0.782 | 4.3.1.- | NA |
| | 3 | 0.206 | 1i0aA | 0.512 | 4.64 | 0.082 | 0.725 | 4.3.2.1 | NA |
| | 4 | 0.203 | 1k62B | 0.512 | 4.69 | 0.064 | 0.725 | 4.3.2.1 | NA |
| | 5 | 0.201 | 1k7wD | 0.513 | 4.97 | 0.051 | 0.749 | 4.3.2.1 | 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.

Reset to initial orientation Spin On/Off

## 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 | 0.23 | 0.5551 | 4.06 | 0.05 | 0.73 | 1tt9A | GO:0007010 GO:0019215 GO:0005737 GO:0005542 GO:0005814 GO:0005794 GO:0003824 GO:0044237 GO:0016740 GO:0030412 GO:0016829 GO:0005856 GO:0030409 GO:0006547 GO:0008152 |
| 2 | 0.22 | 0.5026 | 4.11 | 0.07 | 0.70 | 3gf9A | GO:0005089 GO:0005515 GO:0005622 GO:0035023 |
| 3 | 0.21 | 0.5479 | 3.87 | 0.05 | 0.74 | 1qovA | GO:0005576 GO:0042597 GO:0044179 GO:0016020 GO:0020002 GO:0033644 GO:0051715 GO:0016021 GO:0019835 |
| 4 | 0.21 | 0.5062 | 5.22 | 0.09 | 0.77 | 2ohyB | GO:0016740 GO:0016829 GO:0006547 GO:0003824 GO:0009058 GO:0016841 |
| 5 | 0.21 | 0.5381 | 4.66 | 0.09 | 0.78 | 1dbhA | GO:0005089 GO:0005515 GO:0005622 GO:0035023 |
| 6 | 0.20 | 0.5282 | 3.60 | 0.06 | 0.67 | 1o5hA | GO:0003824 GO:0044237 |
| 7 | 0.20 | 0.5422 | 3.82 | 0.06 | 0.73 | 2nrjA | GO:0009405 GO:0044179 GO:0020002 GO:0016020 GO:0005576 GO:0033644 GO:0016021 GO:0019835 |
| 8 | 0.20 | 0.5444 | 4.65 | 0.07 | 0.78 | 1xdvB | GO:0005085 GO:0005089 GO:0005515 GO:0005622 GO:0007264 GO:0035023 GO:0051056 |
| 9 | 0.20 | 0.5236 | 4.20 | 0.05 | 0.73 | 3ky9B | GO:0005085 GO:0005089 GO:0005515 GO:0005622 GO:0035023 GO:0035556 |
| 10 | 0.20 | 0.5170 | 4.34 | 0.07 | 0.73 | 3jv3A | GO:0005085 GO:0005089 GO:0005515 GO:0005622 GO:0035023 GO:0035556 |

### Consensus prediction of GO terms

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Molecular Function** | GO:0043176 | GO:0032403 | GO:0019842 | GO:0016841 | GO:0030407 | GO:0031406 | GO:0019239 | GO:0005089 |
| **GO-Score** | 0.46 | 0.46 | 0.46 | 0.46 | 0.46 | 0.46 | 0.46 | 0.38 |
| **Biological Process** | GO:0006996 | GO:0008219 | GO:0051715 | GO:0006547 | GO:0035023 | | | |
| **GO-Score** | 0.46 | 0.42 | 0.42 | 0.39 | 0.38 | | | |
| **Cellular Component** | GO:0005813 | GO:0044450 | GO:0043231 | GO:0044218 | GO:0031224 | GO:0033643 | | |
| **GO-Score** | 0.46 | 0.46 | 0.46 | 0.42 | 0.42 | 0.42 | | |

(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 S613477_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, Y Zhang. I-TASSER server: new development for protein structure and function predictions, Nucleic Acids Research, 43: W174-W181, 2015.
2. C Zhang, PL Freddolino, Y Zhang. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. Nucleic Acids Research, 45: W291-W299, 2017.