Kishan Sarpangala

Modelling Track Assignment 4

### [Secondary Structure Prediction & Homology Modeling](https://canopy.uc.edu/webapps/assignment/uploadAssignment?content_id=_27584020_1&course_id=_347988_1&group_id=&mode=view)

Using SwissModel, build homology models for 3 representatives target-template pairs in each of the 3 families, e.g.,

1. Neuroglobin – Myoglobin
2. Myoglobin - Hemoglobin a
3. Hemoglobin a - Hemoglobin b

Representing different level of structural (and sequence) similarity.

Assess the quality of resulting models, including by superimposing the models obtained by using SwissModel with experimentally resolved structures (using Superpose, Pymol or another suitable tool for structure to structure alignment/superposition).

**Summary**:

Secondary Structure

For structures sharing remote homologies, sequence alignments can be poor, whereas the structural similarity may very well be preserved.

For alignments with low sequence identity, SuperPose will attempt a secondary structural alignment, and use this information to guide the superposition.

The *Secondary Structure Alignment* option allows you to determine the percent-identity cutoff for sequence-vs-secondary-structure alignments.

Detecting secondary structure is done by using a set of techniques in bioinformatics that aims to predict the secondary structures of protiens and nucleic acud sequences based only on knowledge of their primary structure.

For proteins, this means predicting the formation of protein structures such as alpha helices and beta strands, while for nucleic acids it means predicting the formation of nucleic acid structures like helixes and stem-loop structures through base pairing and base stacking interactions.

Secondary structure prediction can refer to:

* Protein structure prediction
* Nucleic Acid structure prediction.

SuperPose can handle superpositions with:

1. Identical sequences but slightly different structures
2. Identical sequences but profoundly different structures (e.g. open and closed forms of calmodulin)
3. Modestly dissimilar sequences, lengths and structures.
4. Different sequence lengths but similar structures or sequences
5. Largely different sequences but largely similar structures.

* Superpose can calculate both pairwise and multiple structure superpositions.
* It can also generate average and pairwise RMSD values for alpha carbons, backbone atoms, heavy atoms and all atoms.
* The “SuperPose” web server uses “ RMSD” or Root-Mean-Suare-Deviation” as a difference measure to find the optimal pairwise or multiple protein structure superposition.

There are lot of Structural and sequence similarity when you look at

[**MolScript Superposition Image**](http://wishart.biology.ualberta.ca/cgi-bin/SuperPose/superpose_output.cgi?time=1524555831&action=image&chain1=PDBB_model_default_chain_A&chain2=PDBA_model_default_chain_B&method=structure) and **Sequence Alignment** of

1. Neuroglobin - Myoglobin
2. Myoglobin - Hemoglobin a
3. Hemoglobin a - Hemoglobin b
   1. 2) 3) are Two molecules of identical sequence and structure

**The documentation is available below.**

The resulting **GMQE score** is expressed as a number **between 0 and 1**, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target.

**QMEAN** is a composite scoring function which is able to derive both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates on the basis of one single model. There are two global score values, QMEAN4 and QMEAN6. QMEAN4 is a linear combination of four statistical potential terms. QMEAN6 additionally uses two agreement terms evaluating the consistency of structural features with sequence based predictions. Both global scores are originally in a range [0,1] with one being good

Per default **they are transformed into Z-scores** to relate them with what we would expect from high resolution X-ray structures. That's also what is displayed below.

The SuperPose web server uses a combination of sequence alignment, difference distance matrix comparison and quaternion eigenvalue superposition to allow users to superimpose 3D structures in 5 different ways:

1) Two molecules of identical sequence and structure

2) Two molecules of identical sequence but profoundly different structure (i.e. open and closed forms of calmodulin).

3) Two molecules of modestly dissimilar sequence, length and structure (i.e. hemoglobin A and B chains).

4) Two molecules that are profoundly different in sequence but similar in structure (i.e. ubiquitin and elongin).

5) Two or more molecules of identical sequence but slightly different structure (i.e. crystal subunit isoforms, NMR structure ensembles).

**Note**:

1. **Myoglobin** and **Hemoglobin** each bind oxygen, we can assume that there should be some similarities between their structures. However, there should also be some dissimilarities because of their different roles as oxygen binding proteins. One obvious difference between the two proteins is that myoglobin is a single polypeptide chain while vertebrate hemoglobins are tetrameric
2. **Neuroglobin** are structurally similar to **myoglobin**, but they contain distinct features like extraordinarily high temperature resistances and unusual cavities inside the molecules.

|  |  |
| --- | --- |
| **3 representatives target-template** | **Structural and Sequence**  **(SuperPose Results)** |
| 1. Neuroglobin – Myoglobin | Structurally same & Sequence alignment is same about 98.7% |
| 1. Myoglobin - Hemoglobin a | Structurally same but sequence alignment is different about 39.4% |
| 1. Hemoglobin a - Hemoglobin b | Structurally same & Sequence alignment is almost same about 96.5% |

1. The affinity of hemoglobin for oxygen is less than its structural analog myoglobin. Interestingly enough, however, this does not affect hemoglobin's usefulness for the body; on the contrary, it allows hemoglobin to be a more efficient oxygen carrier than myoglobin. This is so because hemoglobin can release oxygen more easily than can myoglobin

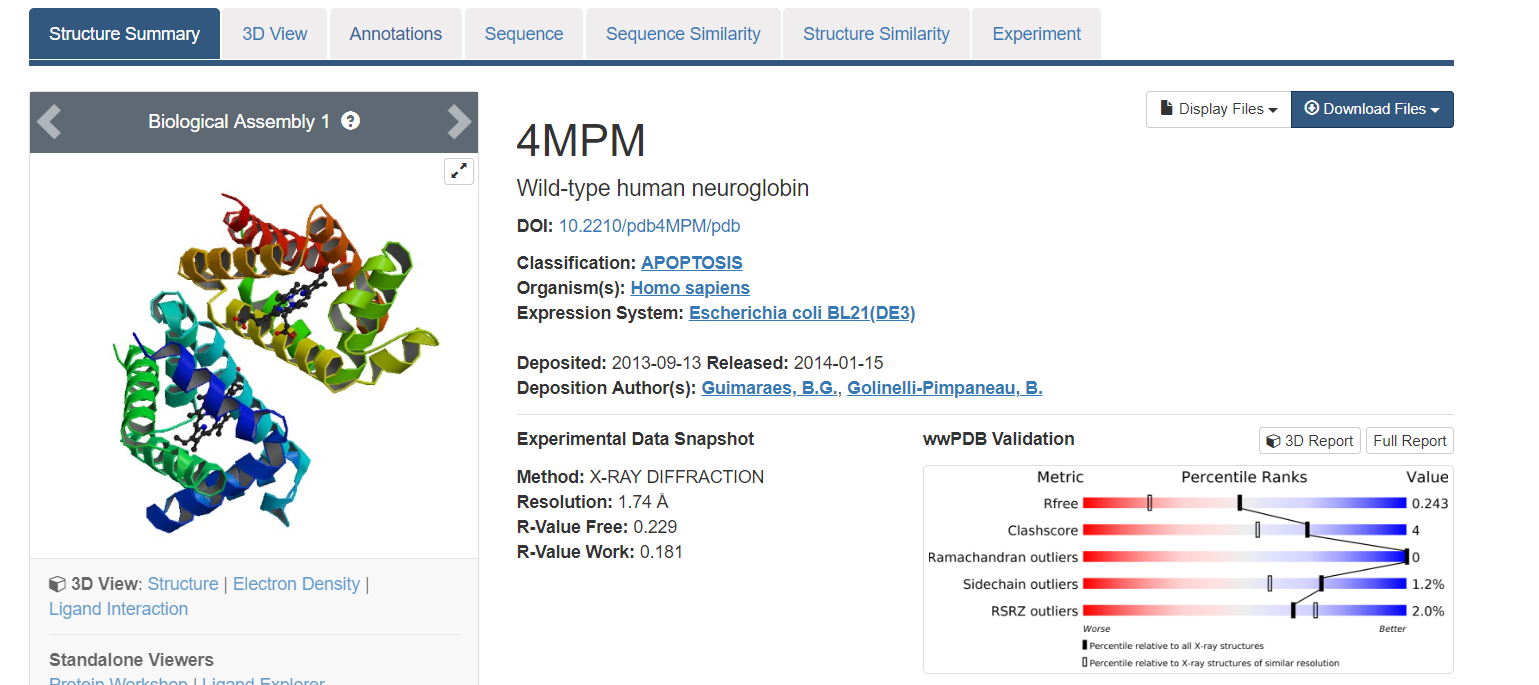
Hemoglobin exists in two distinct states: the T-state and the R-state. The T-state of hemoglobin is the more "Tense" of the two; this is the deoxy form of hemoglobin (meaning that it lacks an oxygen species) and is also known as "deoxyhemoglobin". The R-state of hemoglobin is more "Relaxed" and is the fully oxygenated form; it is also known as "oxyhemoglobin."

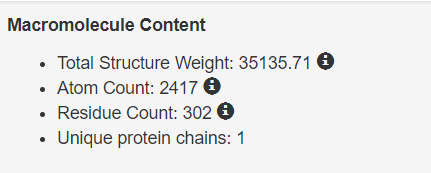
Hemoglobin is quite flexible in terms of how much oxygen it can bind and therefore how much it can release. This is one of the reasons that the body prefers to use hemoglobin, as opposed to myoglobin, for oxygen transport: hemoglobin can modify its own affinity for oxygen to suit the situation at hand, making it capable of handling a wider variety of chemical environments and organ systems while still being able to distribute oxygen effectively.

1. Hemoglobin a - Hemoglobin b are structurally same & sequence alignment is almost same about 96.5%. The difference distance matrix of Hemoglobin a - Hemoglobin b is quite interesting. They are so similar; the dissimilarity is very low.
2. Neuroglobin - Myoglobin

**Neuroglobin**: **4MPM**

Took Fasta file from PDB website







Fasta File of Neuroglobin

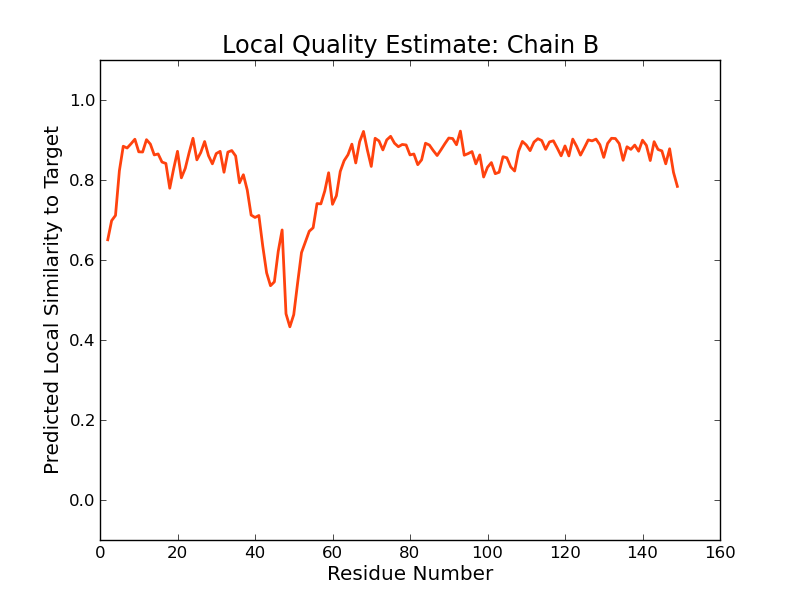
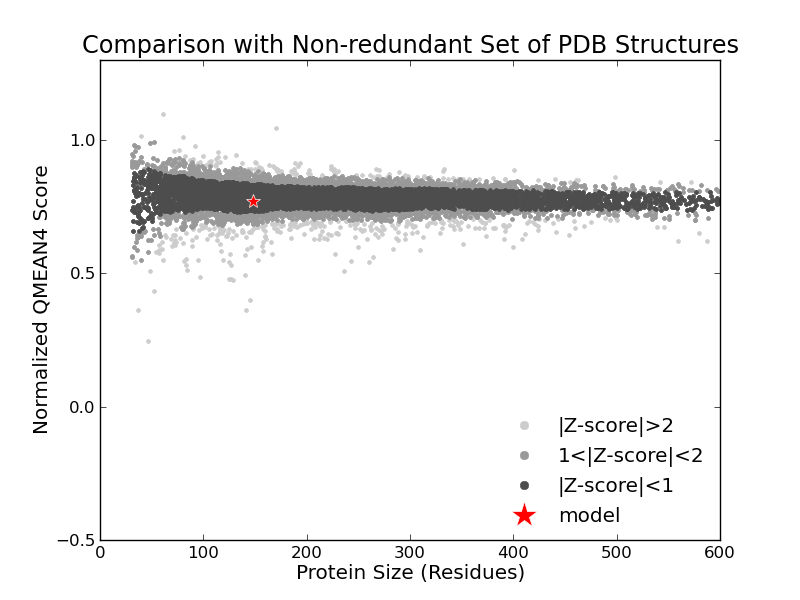
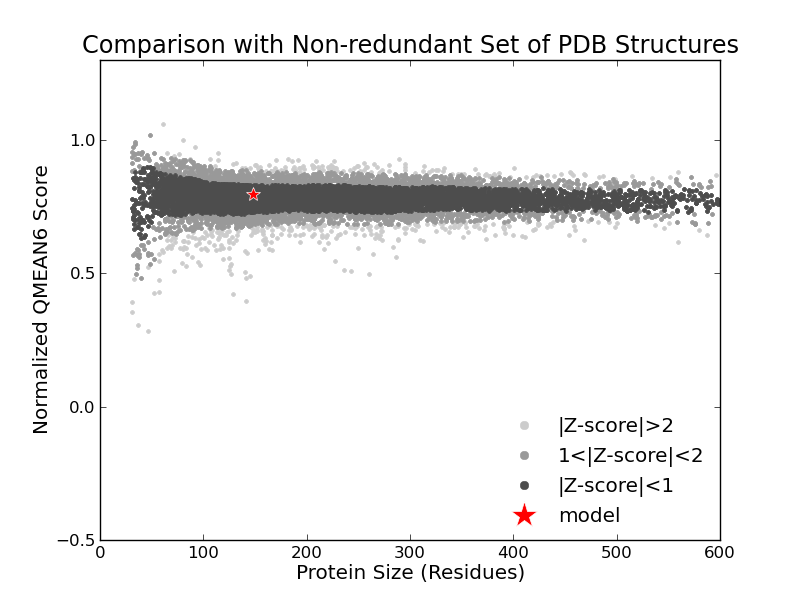
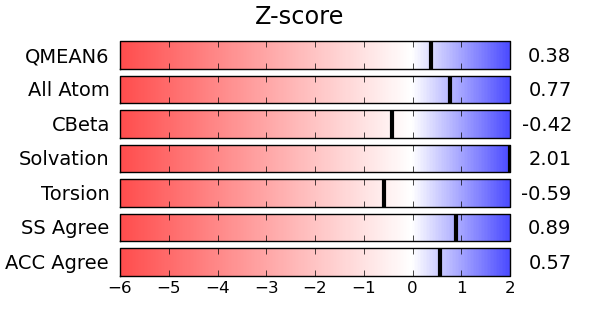
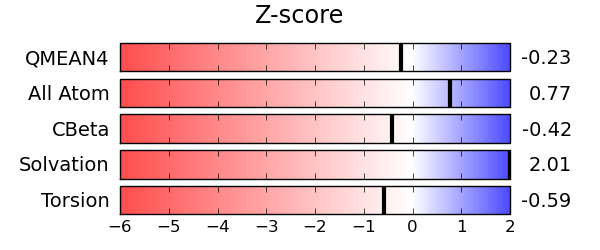
Neuroglobin: pdb\_id: 4mpm

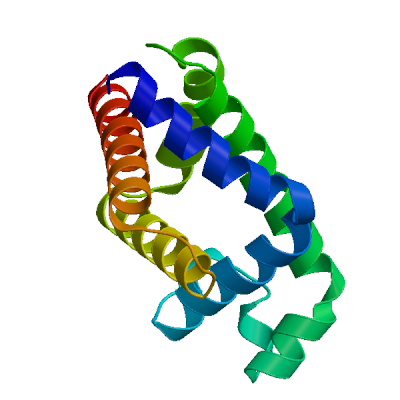


Swiss Model Neuroglobin Neuro-globin Template PDB file This zip file contains all the Template report (will be input to Superpose) files obtained from Swiss Website

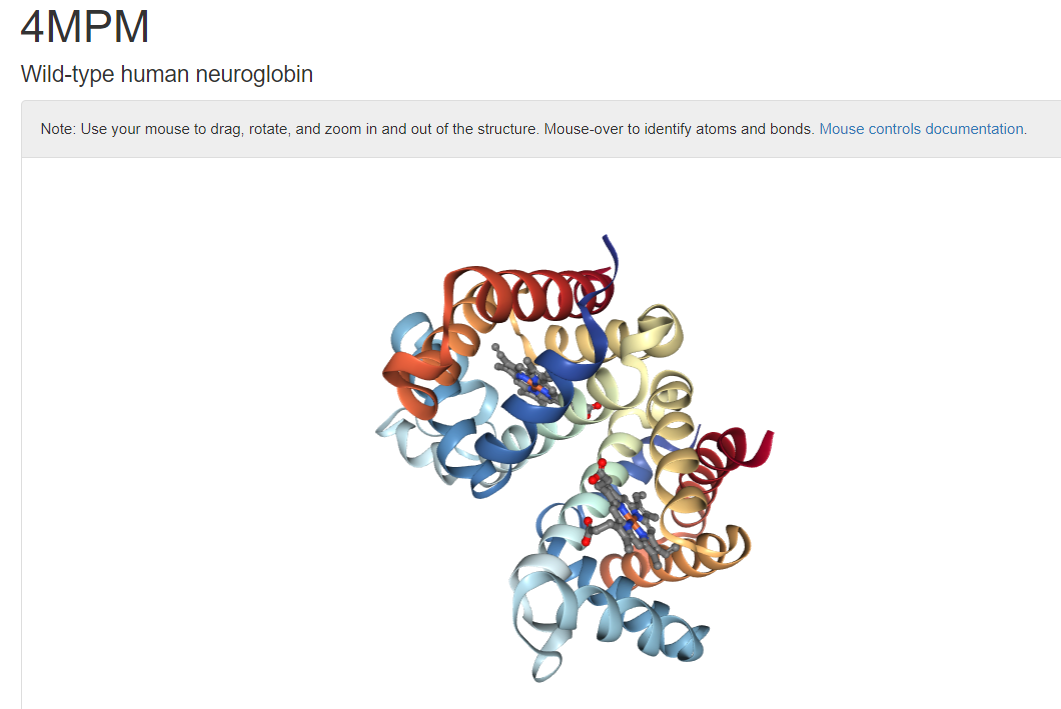
for Neuroglobin

Important Charts with respect to Neuroglobin  



Neuroglobin structure (obtained from Swiss site)

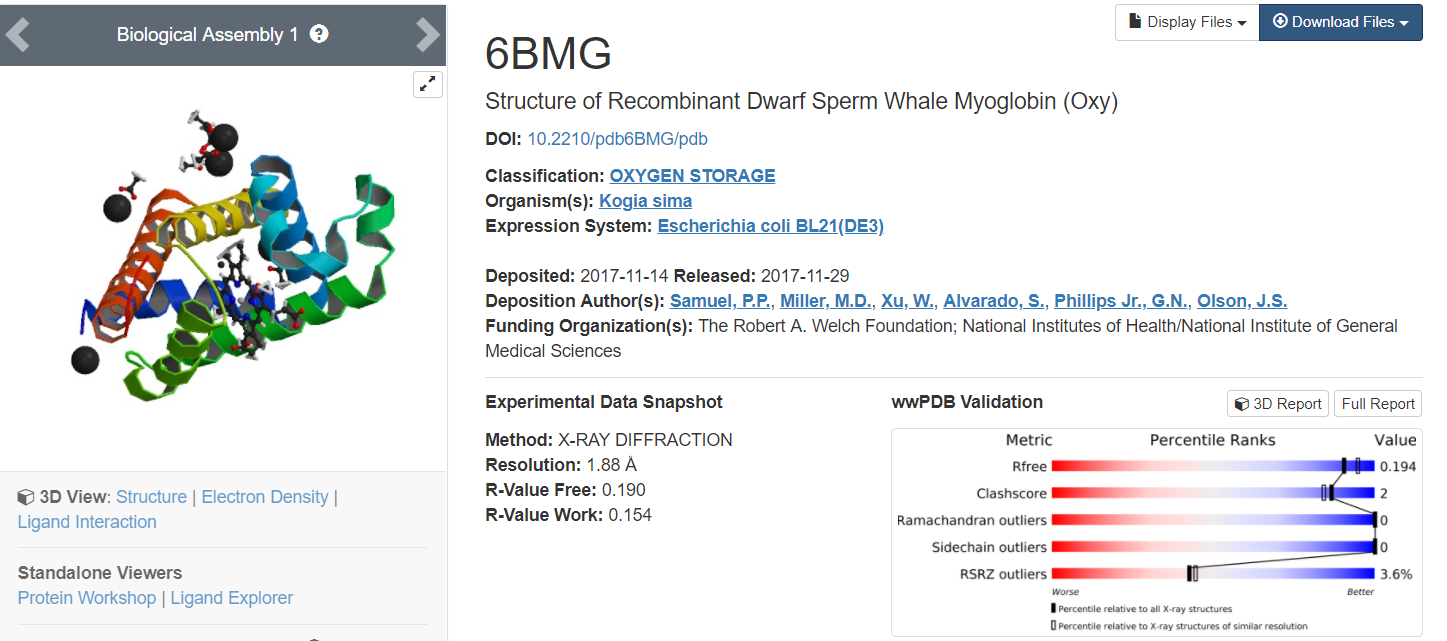
Below Neuroglobin Structure Image is obtained from RCSB PDB protein Data Bank

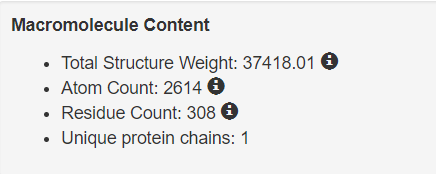


This files below contains local & global scores calculated by the QMEAN scoring function…

Myoglobin pdb\_id: 6bmg







Fasta File of Myoglobin

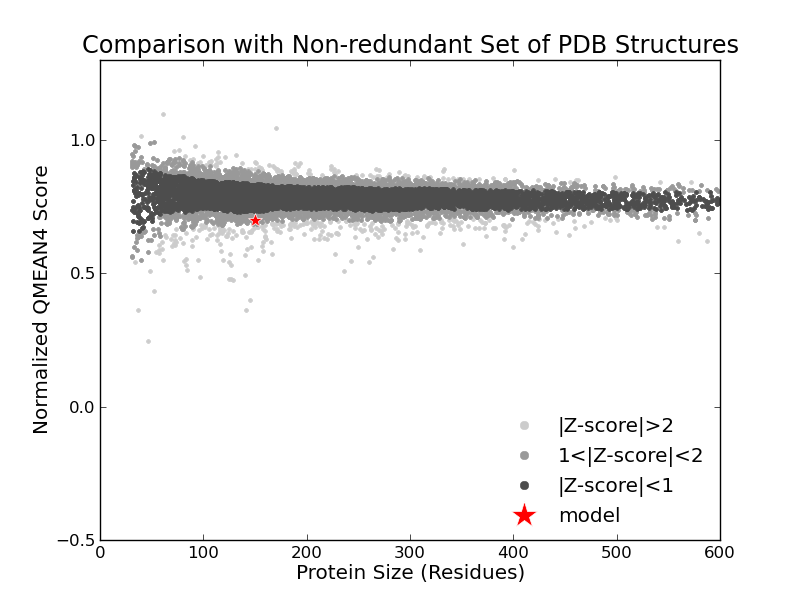
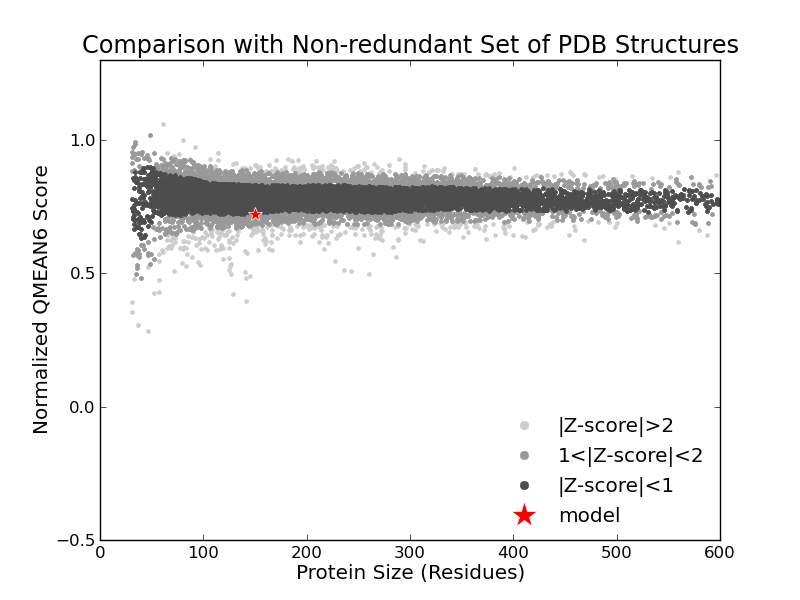
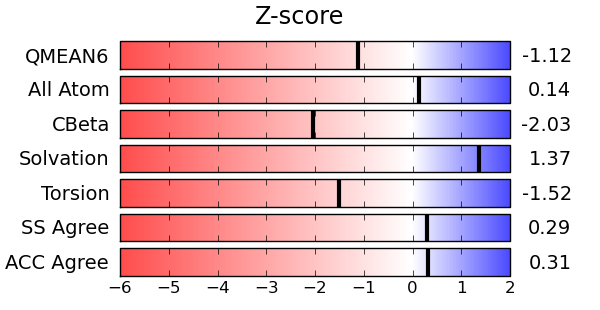
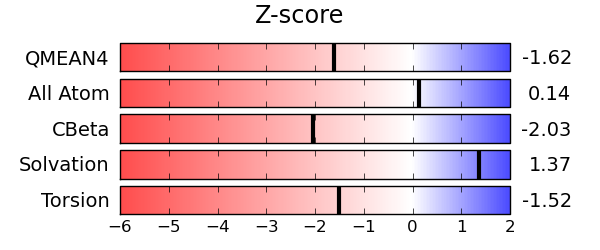
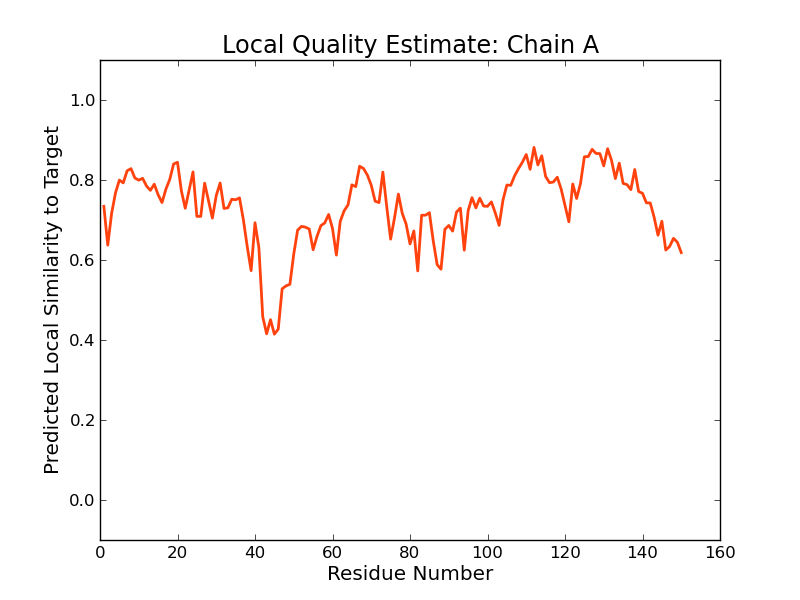
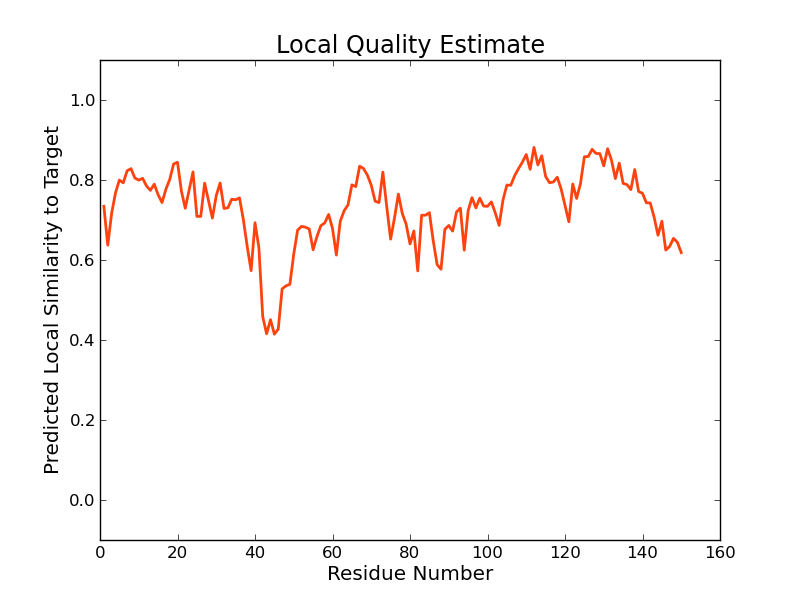


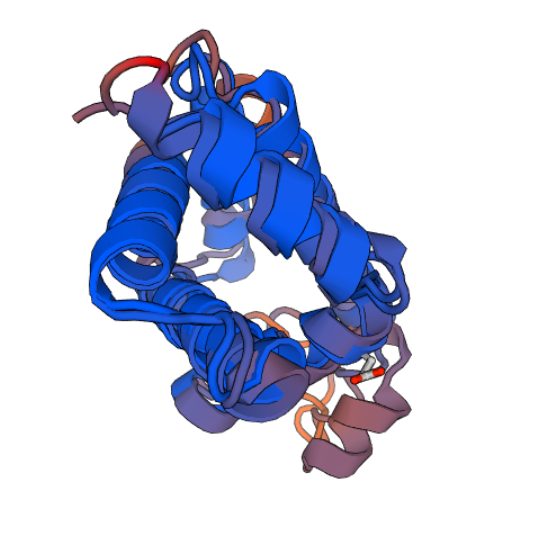
  

Myoglobin Template PDB file This zip file contains all the template files obtained from Swiss Swiss Model Myoglobin Report

(will be input to Superpose) website for Myoglobin

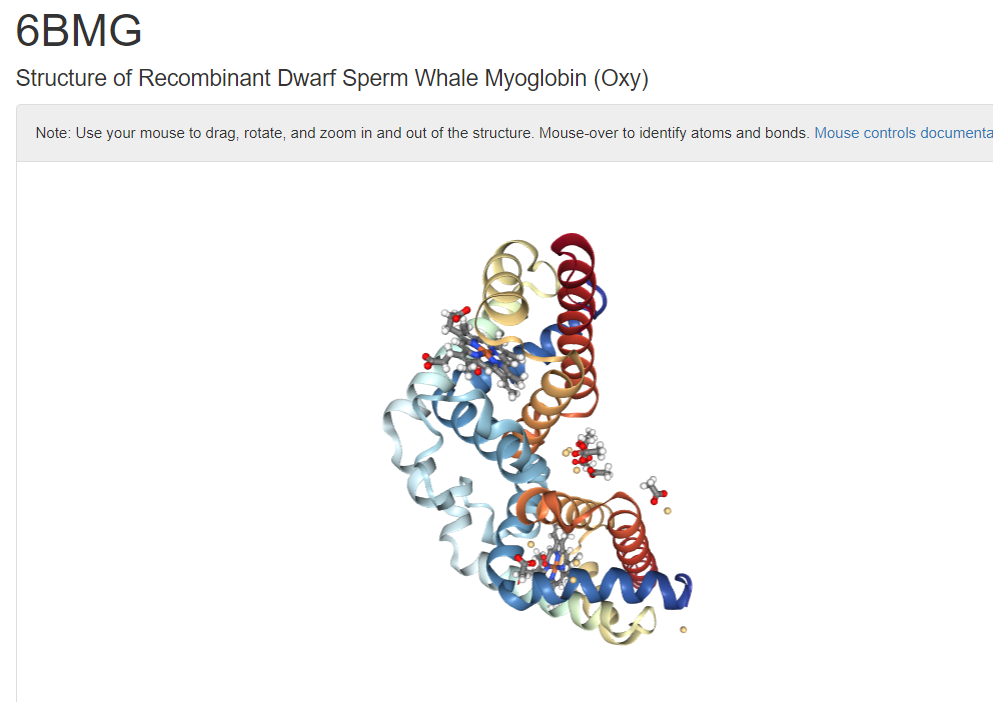
Important Charts with respect to Myoglobin





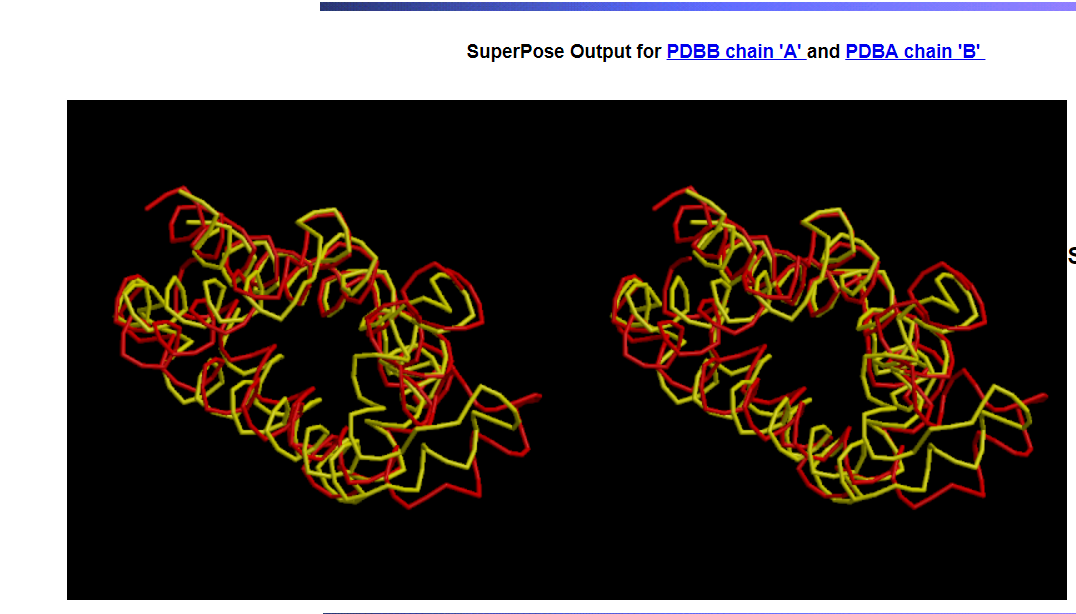
Myoglobin structure (obtained from Swiss site)

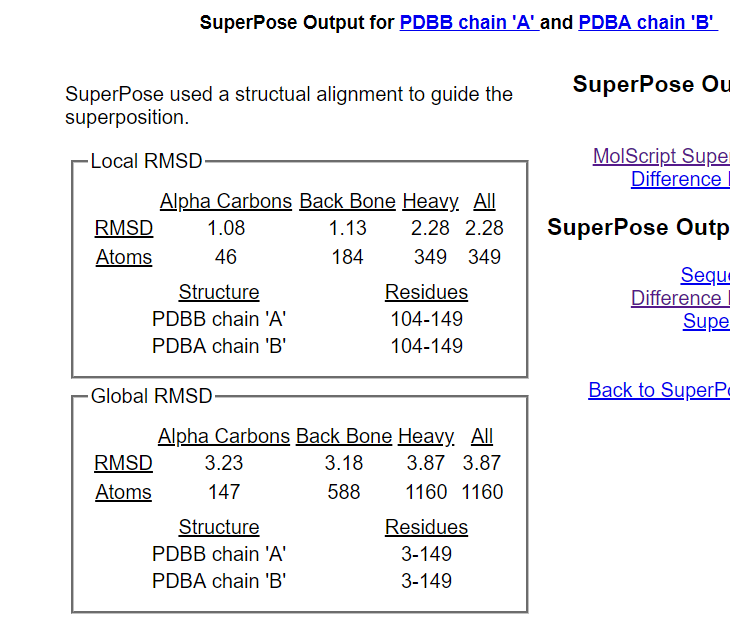
Below Myoglobin Structure Image is obtained from RCSB PDB protein Data Bank



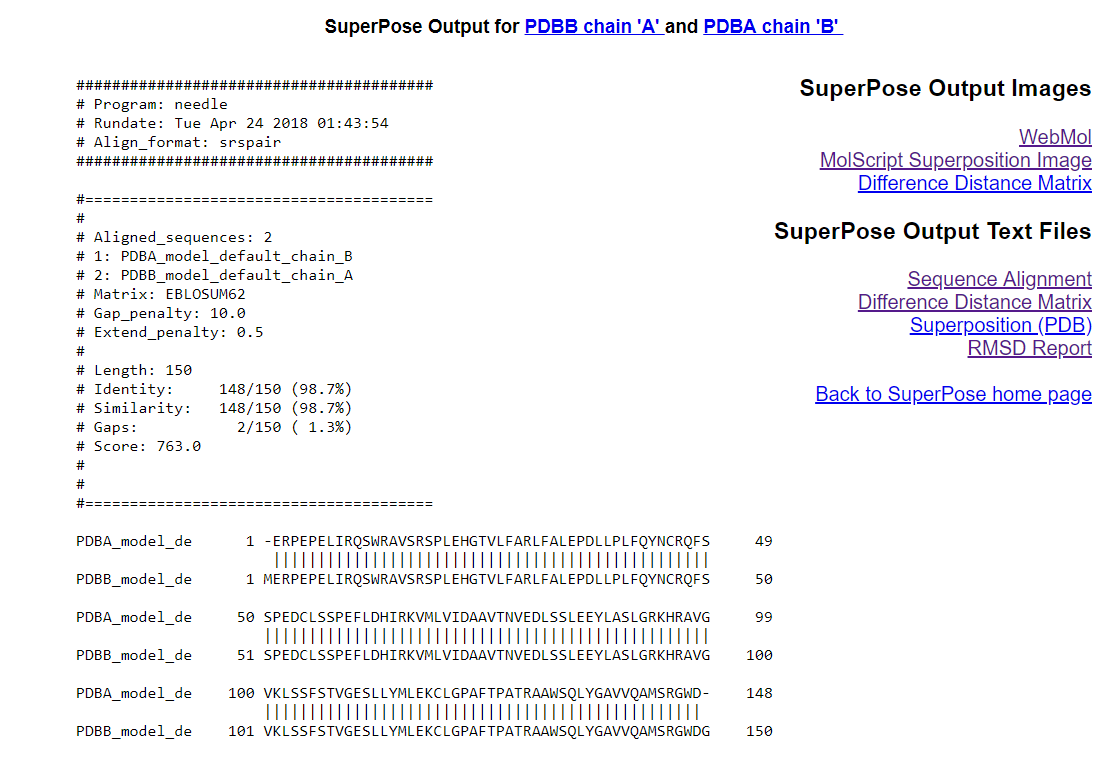
**SuperPose Results:** BetweenNeuroglobin - Myoglobin

[**MolScript Superposition Image**](http://wishart.biology.ualberta.ca/cgi-bin/SuperPose/superpose_output.cgi?time=1524555831&action=image&chain1=PDBB_model_default_chain_A&chain2=PDBA_model_default_chain_B&method=structure)

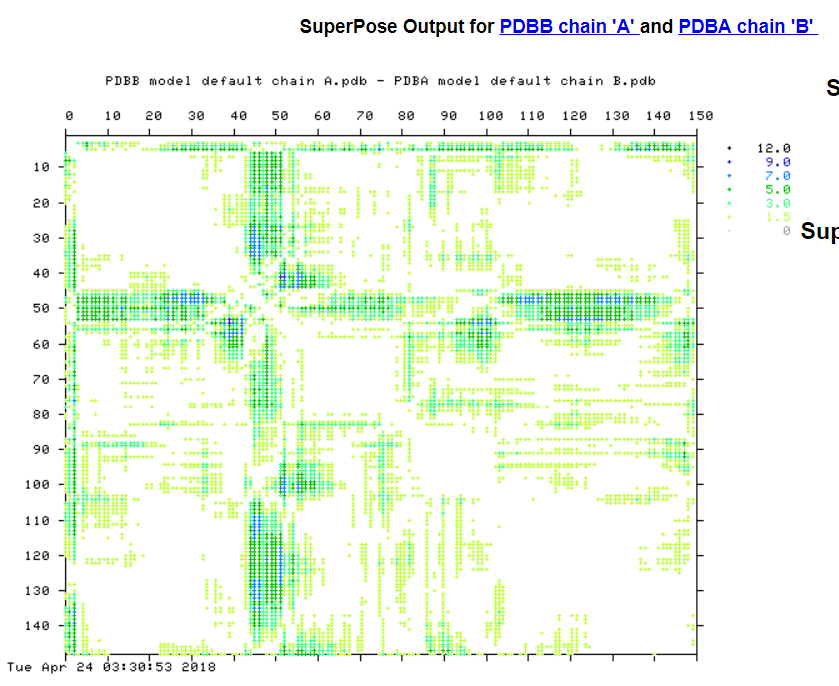




**Sequence Alignment**



Difference Distance Matrix



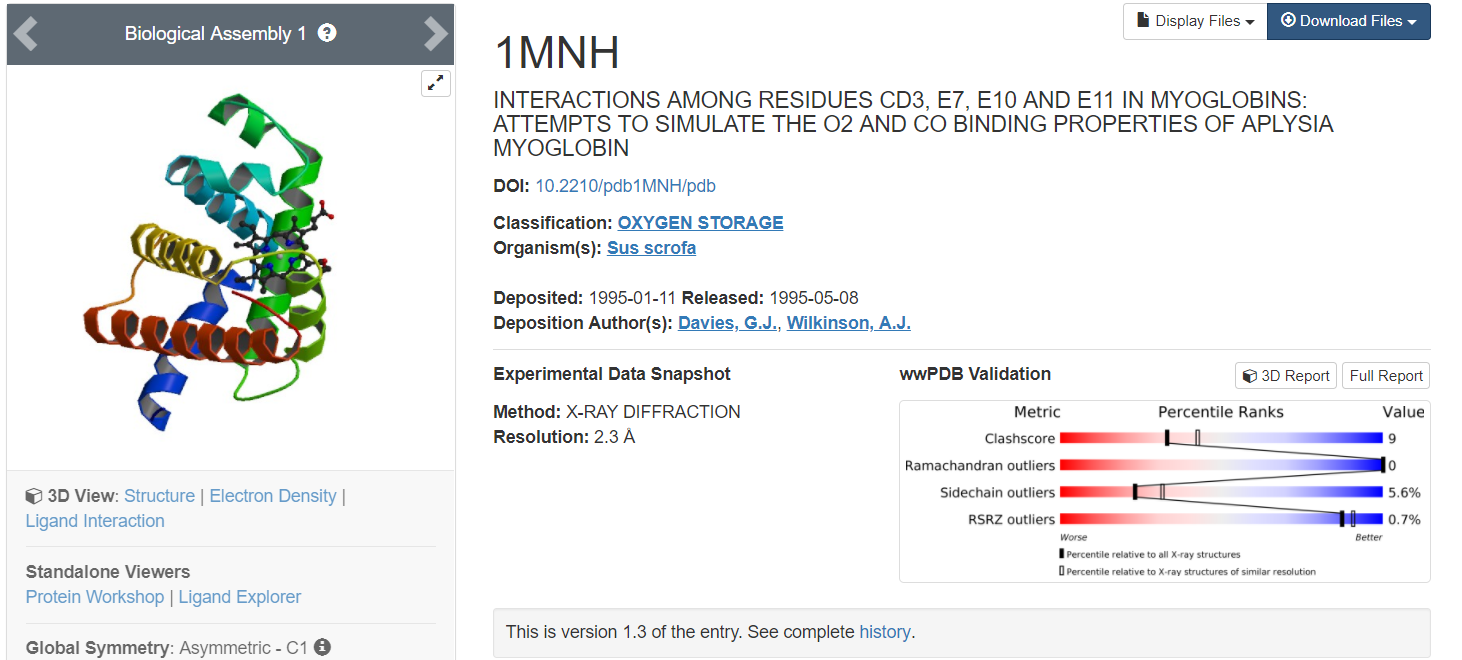
SuperPose Results

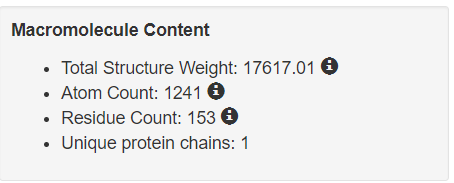
1. Difference Distance Matrix
2. Final SuperPose Result

1. Myoglobin - Hemoglobin a

Myoglobin pdb\_id : 1MNH







Myoglobin Fasta File

The above fasta file was used as input to Swiss website.

<https://swissmodel.expasy.org>

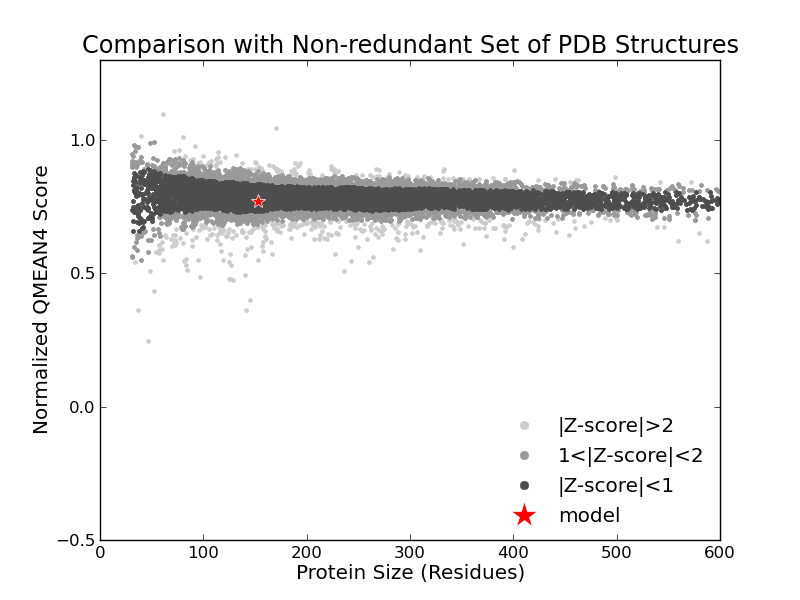
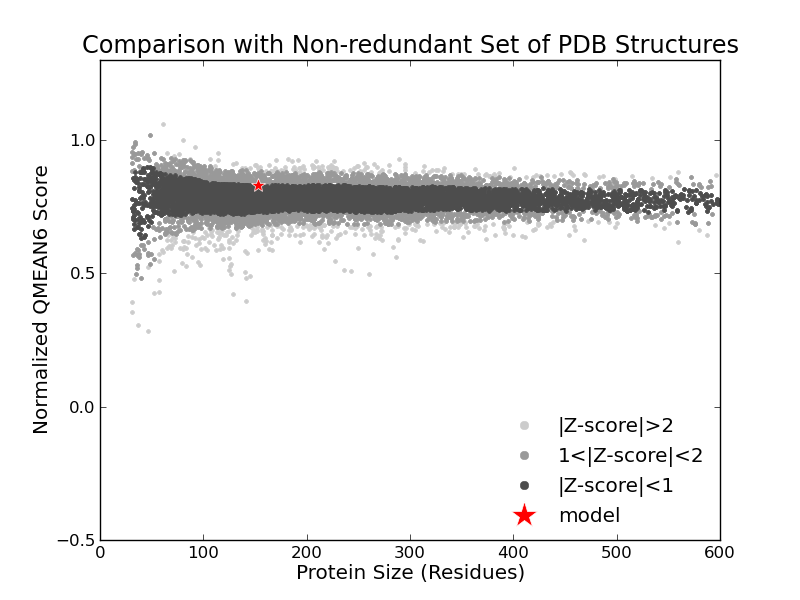
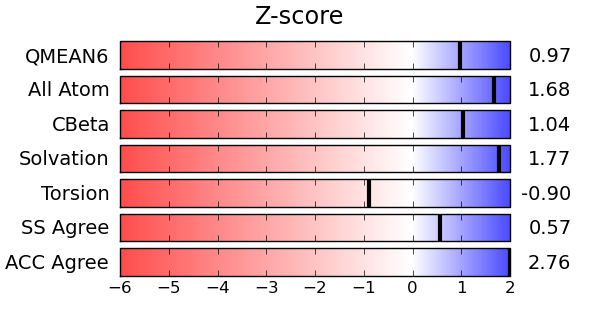
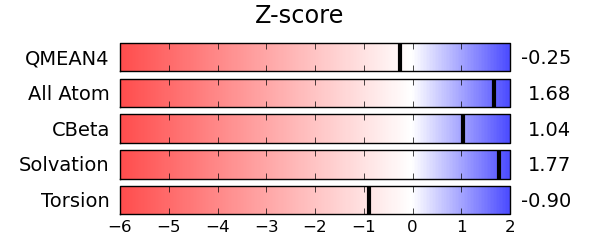
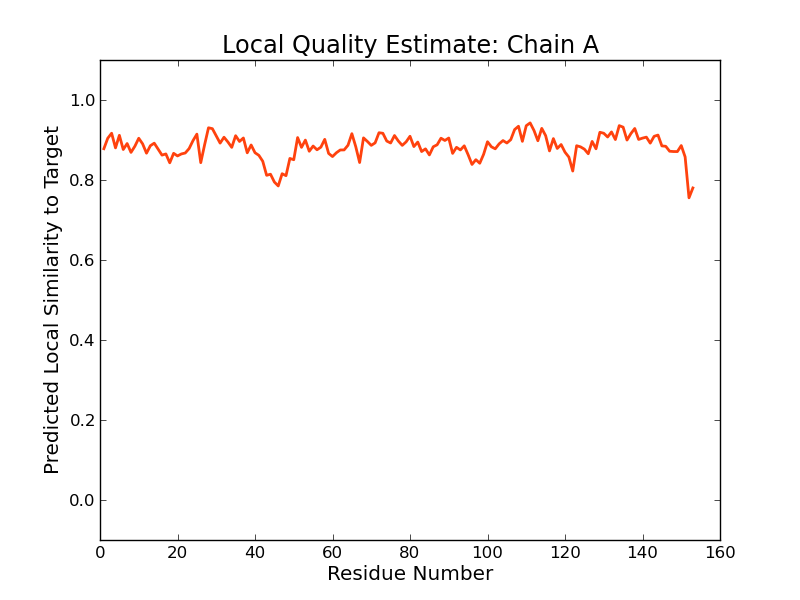
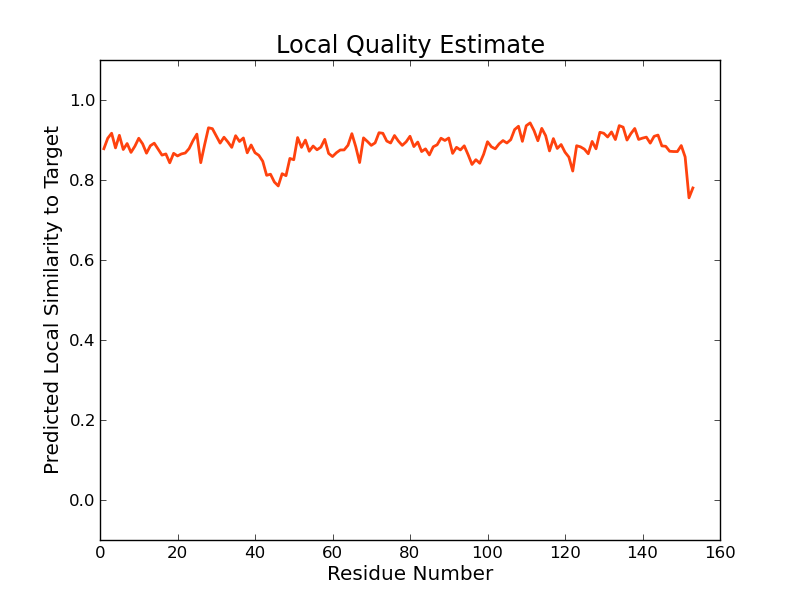


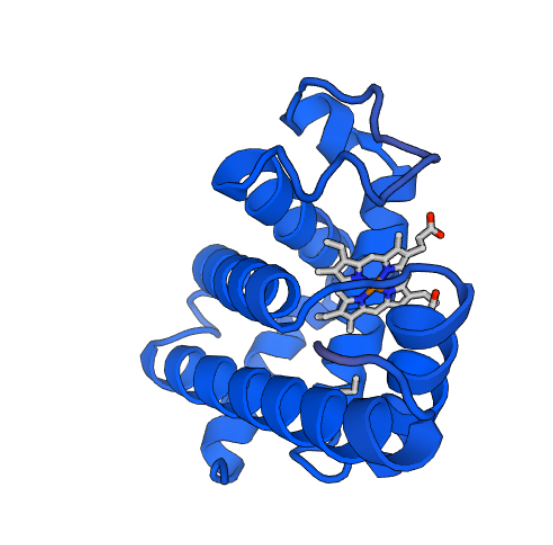
  

Myoglobin Template PDB file This zip file contains all the template files obtained from Swiss Swiss Model Myoglobin Report

(will be input to Superpose) website for Myoglobin

Important Charts with respect to Myoglobin





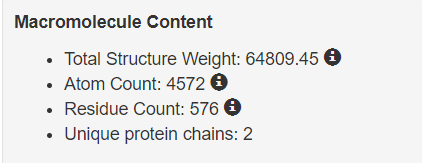
Myoglobin structure (obtained from Swiss site)

Below Myoglobin Structure Image is obtained from RCSB PDB protein Data Bank



Hemoglobin a PDB\_ID: 1BZ1



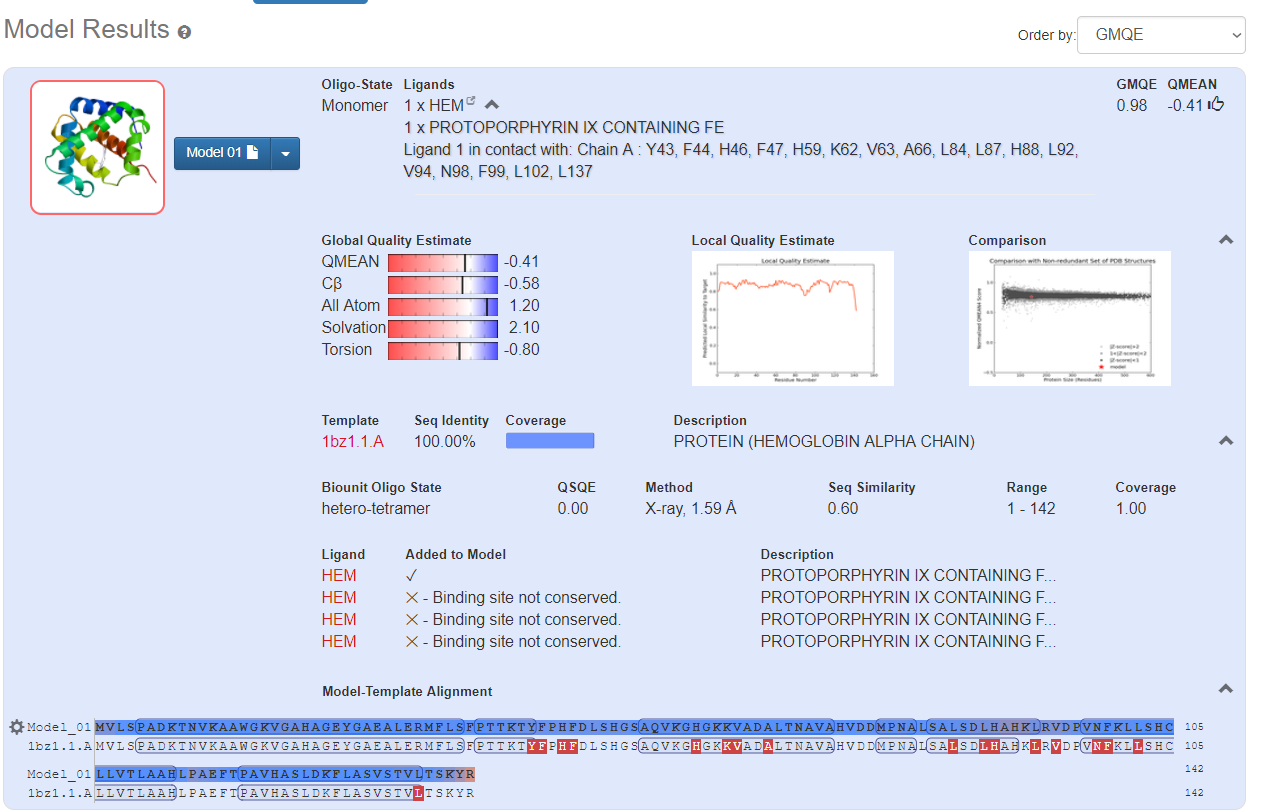




Hemoglobin\_a Fasta File

The above fasta file was used as input to Swiss website.

<https://swissmodel.expasy.org>

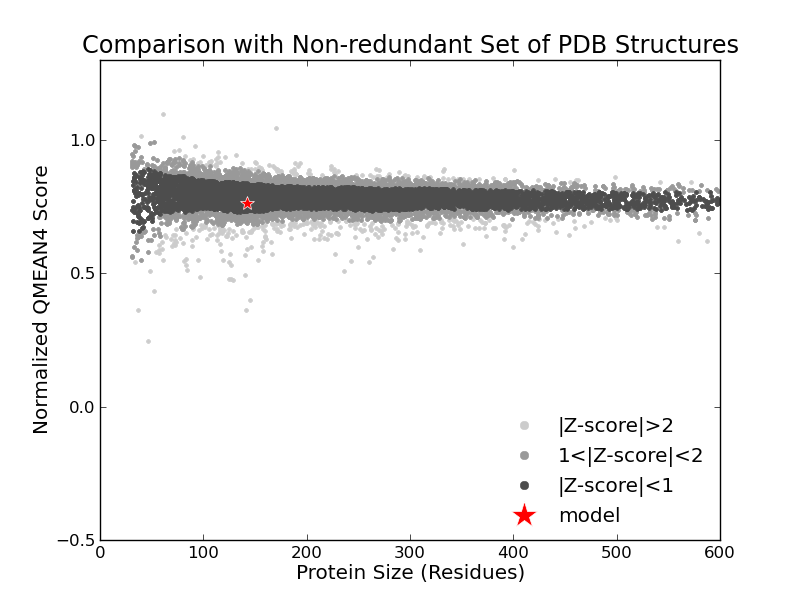
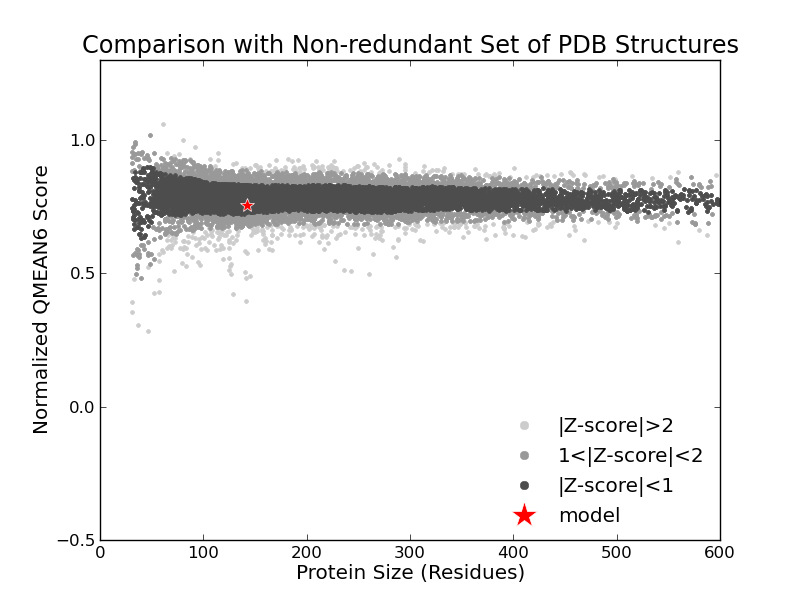
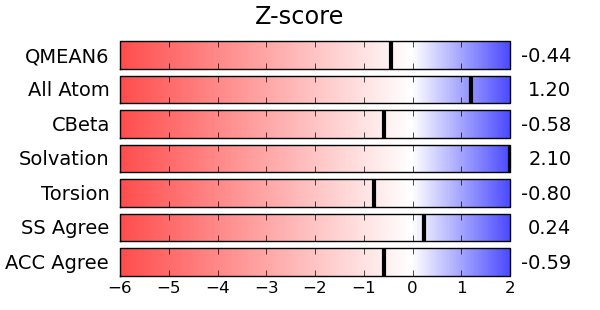
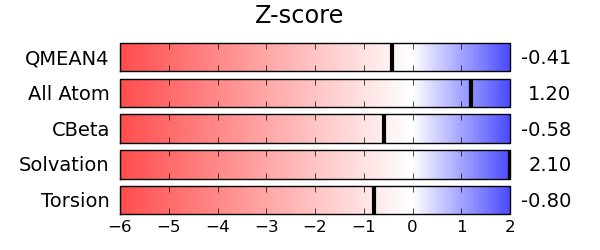
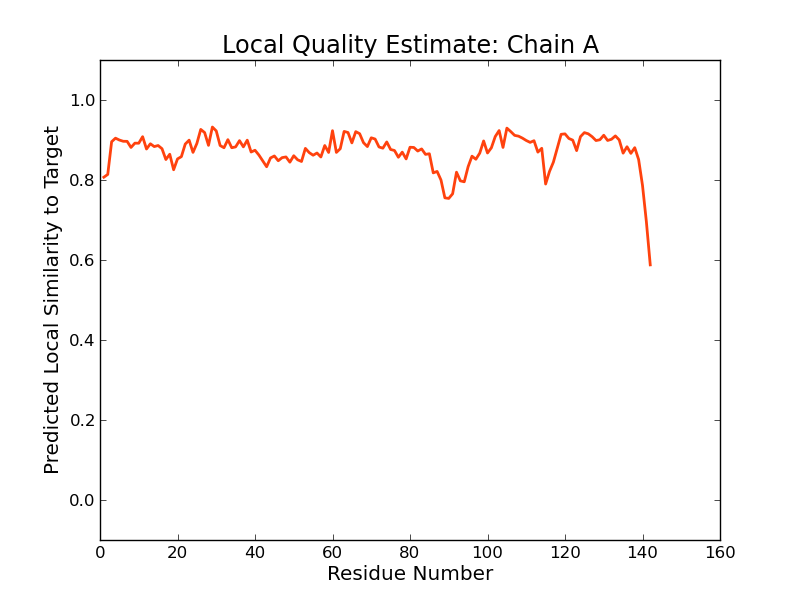
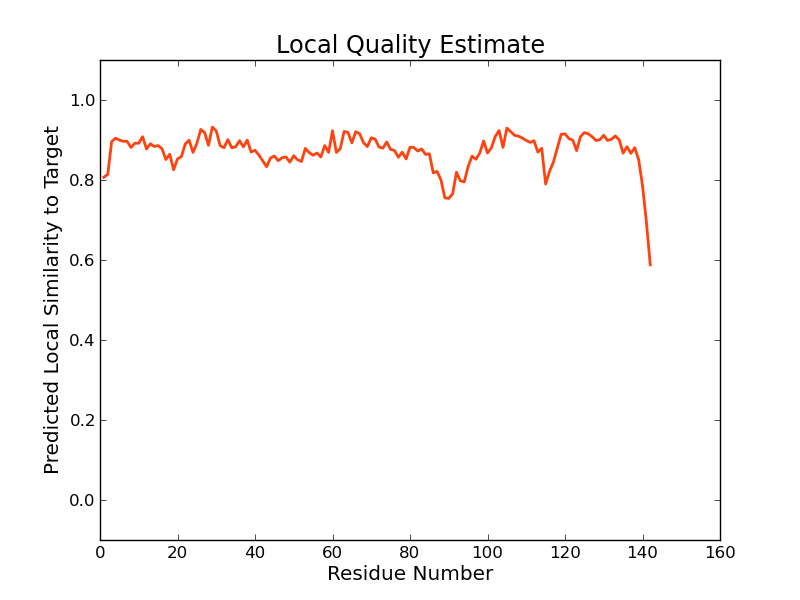


Myoglobin Template PDB file This zip file contains all the template files obtained from Swiss Swiss Model Myoglobin Report

(will be input to Superpose) website for Myoglobin

Important Charts with respect to Myoglobin





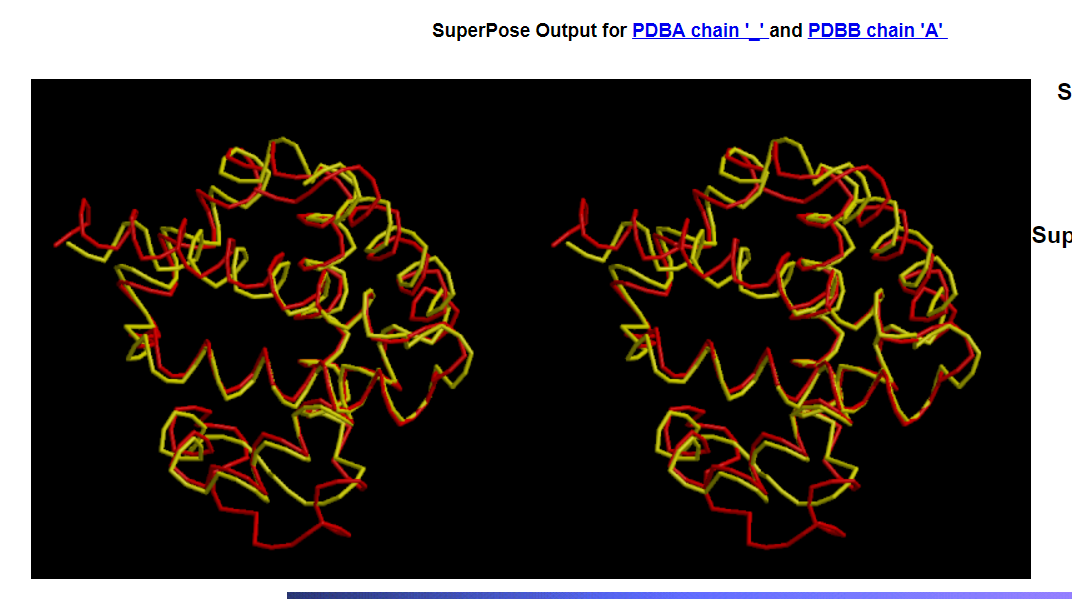
Hemoglobin\_a structure (obtained from Swiss site)

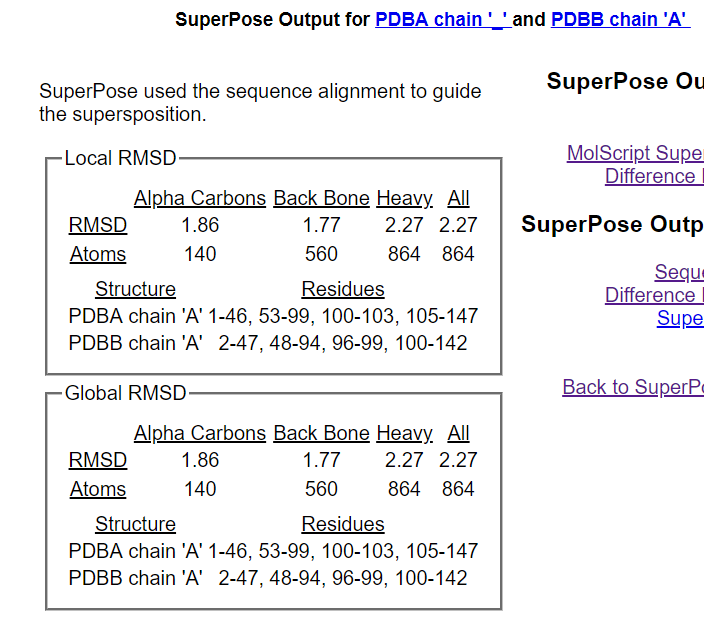
Below Hemoglobin\_a Structure Image is obtained from RCSB PDB protein Data Bank



SuperPose Results: BetweenMyoglobin - Hemoglobin a

[**MolScript Superposition Image**](http://wishart.biology.ualberta.ca/cgi-bin/SuperPose/superpose_output.cgi?time=1524555831&action=image&chain1=PDBB_model_default_chain_A&chain2=PDBA_model_default_chain_B&method=structure)

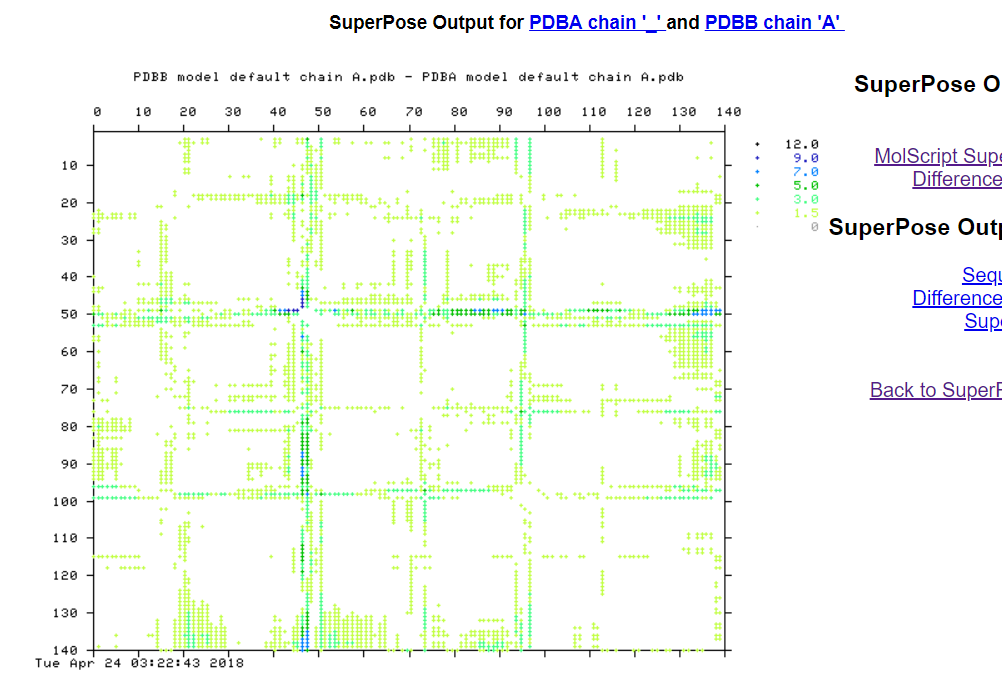




**Sequence Alignment**



**Difference Distance Matrix**



SuperPose Results

1. Difference Distance Matrix (dd\_matrix)
2. Final SuperPose Result

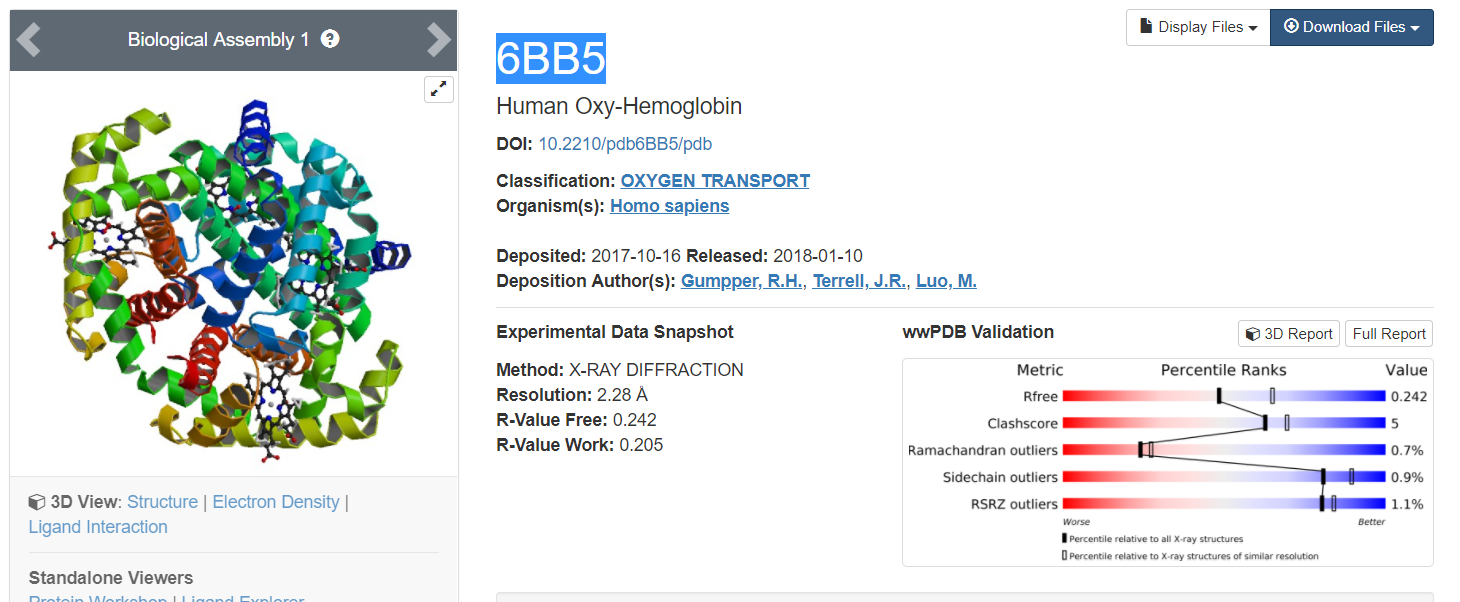
  

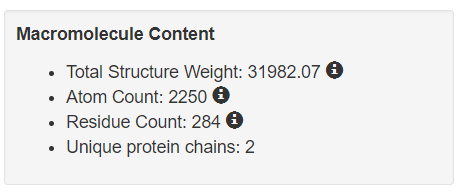
1. Hemoglobin a - Hemoglobin b

Hemoglobin b

* Beta-globin is a component (subunit) of a larger protein called hemoglobin, which is located inside red blood cells.
* In adults, hemoglobin normally consists of four protein subunits: two subunits of beta-globin and two subunits of another protein called alpha-globin, which is produced from another gene called HBA.
* Total or partial absence of HBB causes a genetic disease called beta thalassemia
* HBB interacts with Hemoglobin, alpha 1 (HBA1) to form haemoglobin A, the major haemoglobin in adult humans.

# **Hemoglobin b** PDB\_ID: 6BB5



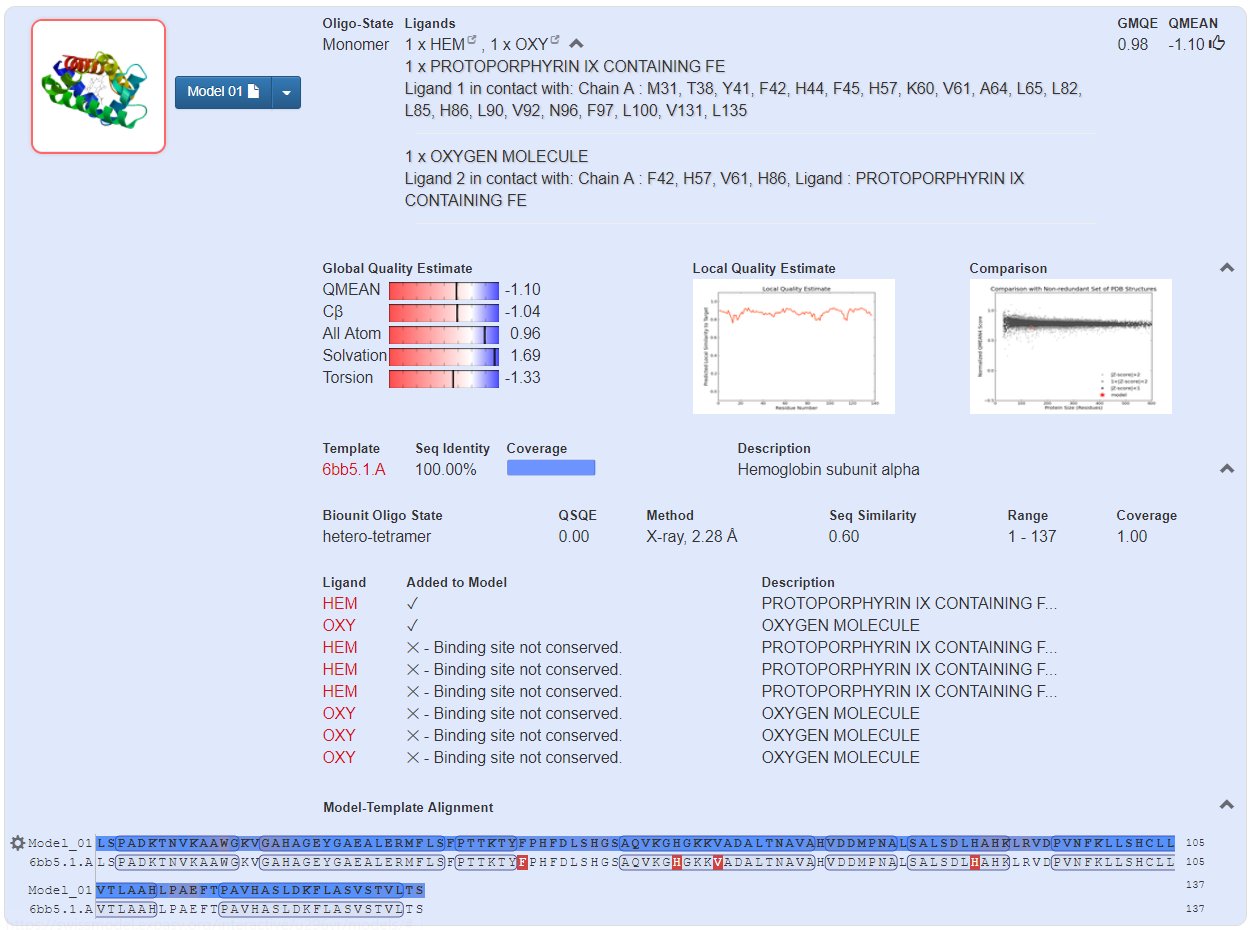




Hemoglobin\_b Fasta File

The above fasta file was used as input to Swiss website.

<https://swissmodel.expasy.org>

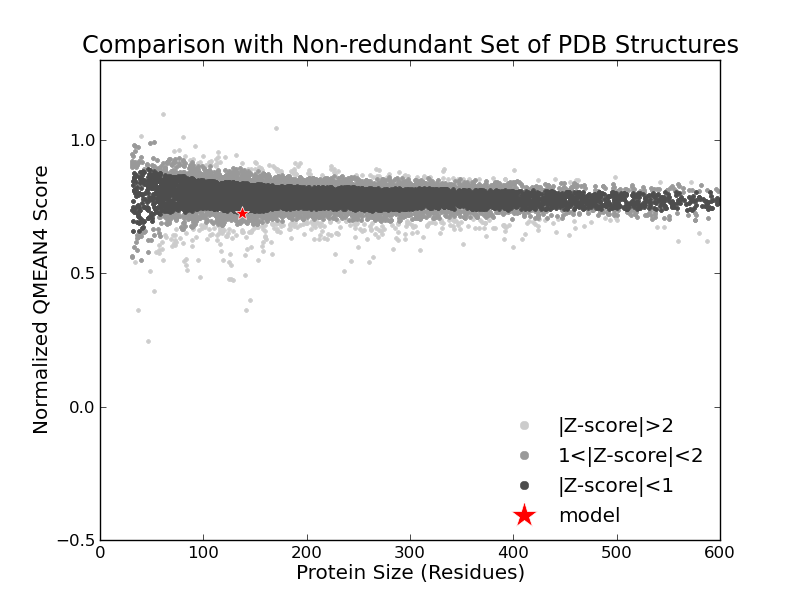
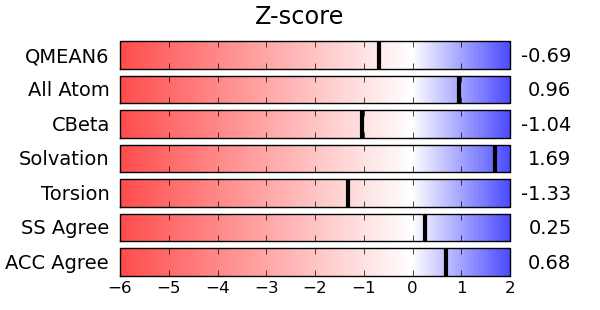
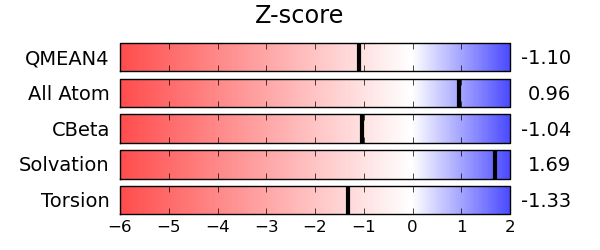
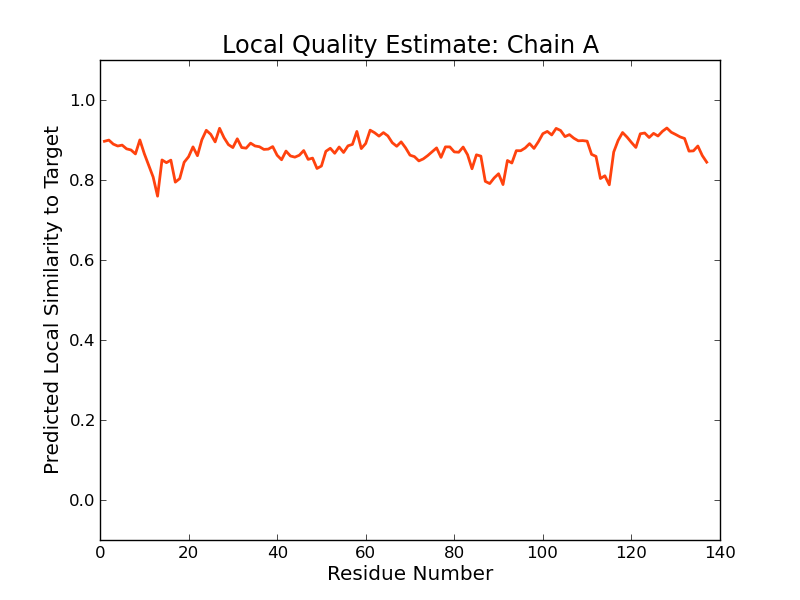
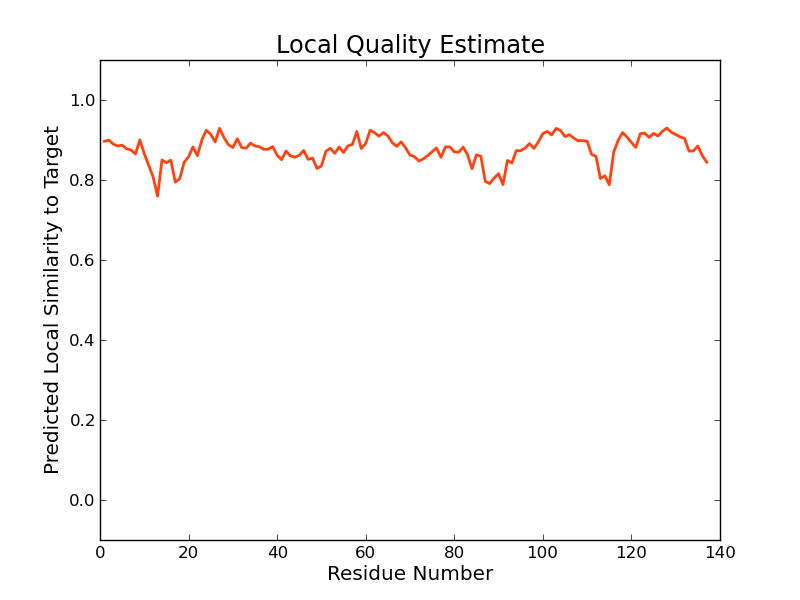


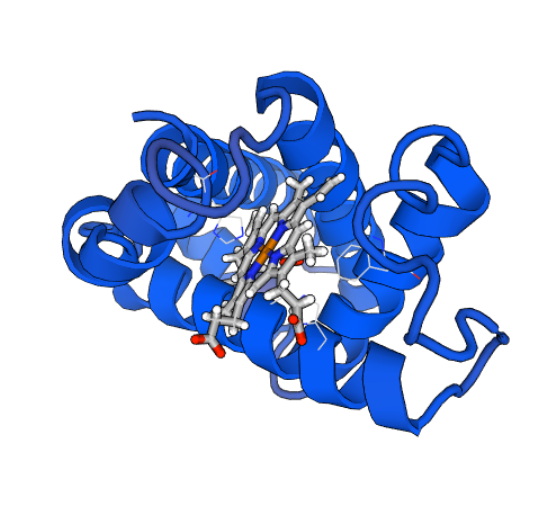
  

Myoglobin Template PDB file This zip file contains all the template files obtained from Swiss Swiss Model Myoglobin Report

(will be input to Superpose) website for Myoglobin

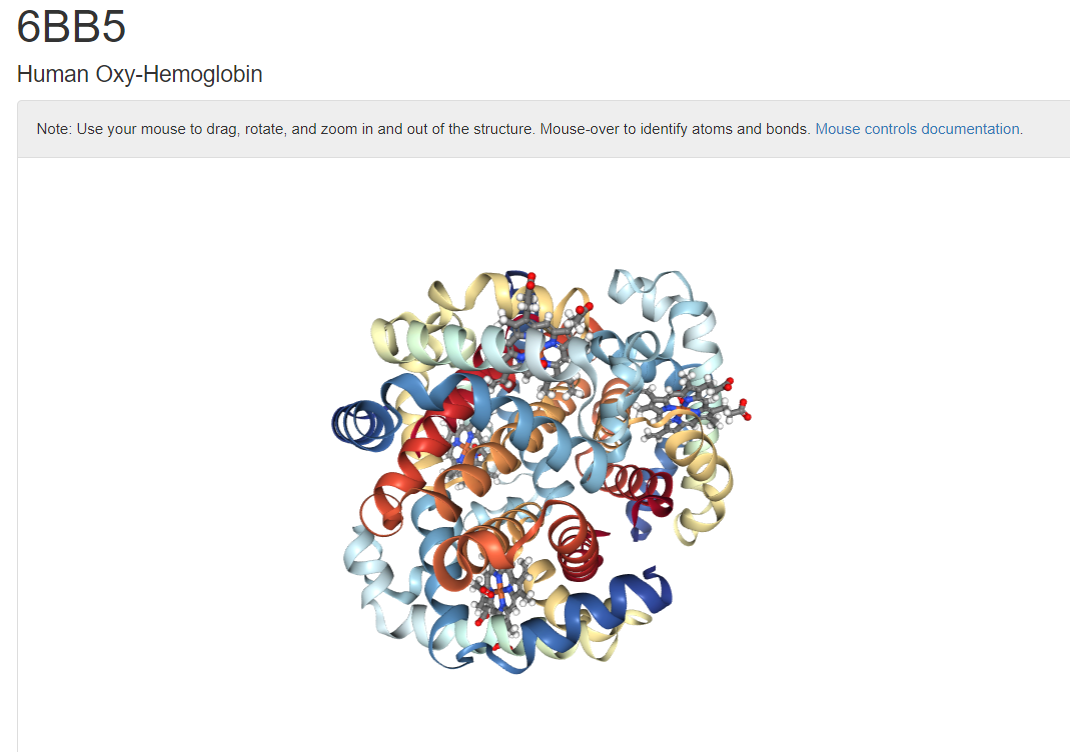
Important Charts with respect to Hemoglobin b



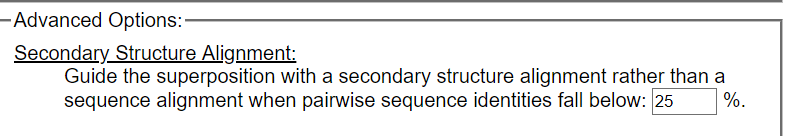


Hemoglobin\_b structure (obtained from Swiss site)

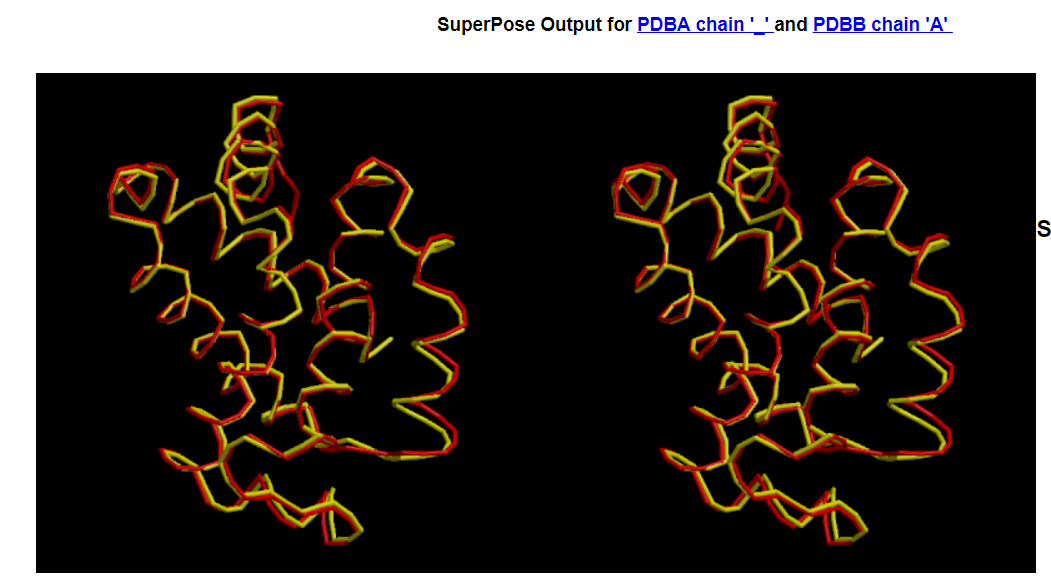
(Below) Hemoglobin\_b Structure Image is obtained from RCSB PDB protein Data Bank

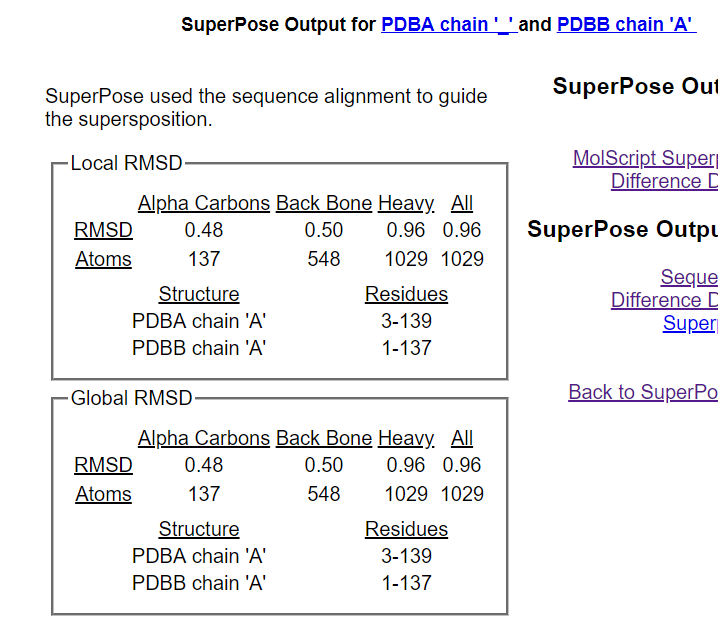


SuperPose Results: Hemoglobin a - Hemoglobin b **(Normal Run)**

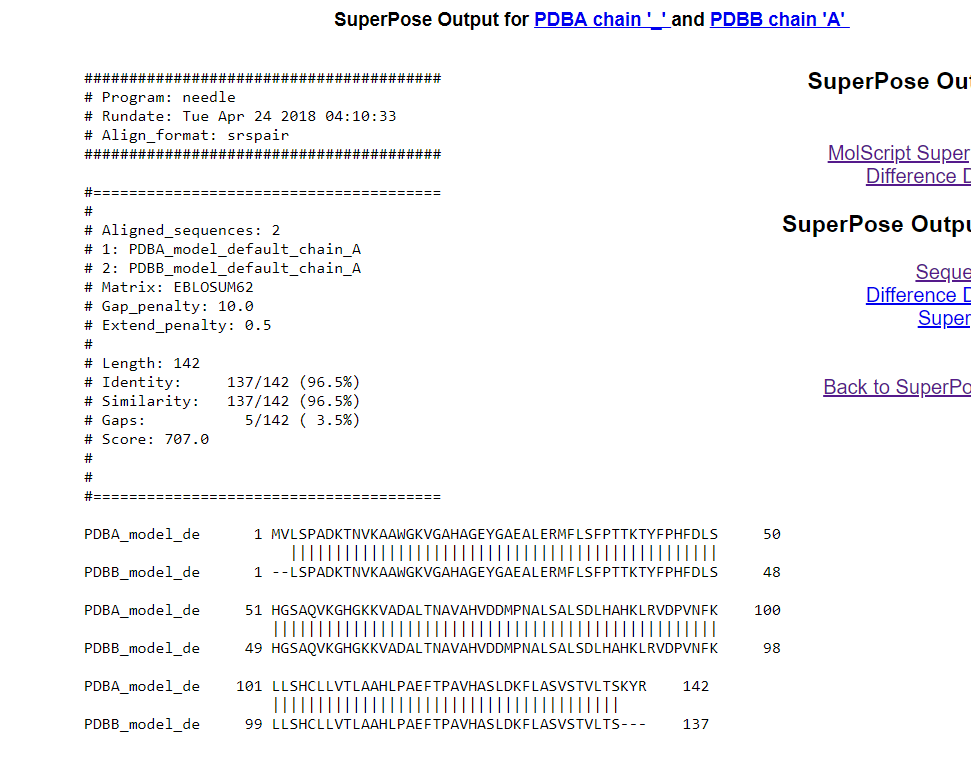
(Without selection of advanced tab or Secondary structure alignment) 

[**MolScript Superposition Image**](http://wishart.biology.ualberta.ca/cgi-bin/SuperPose/superpose_output.cgi?time=1524555831&action=image&chain1=PDBB_model_default_chain_A&chain2=PDBA_model_default_chain_B&method=structure)





**Sequence Alignment**



**Difference Distance Matrix**



**SuperPose Results**

1. Difference Distance Matrix (dd\_matrix)
2. Final SuperPose Result

**Similarly I repeated the exercise for rest of the target protein including cytochrome and also Zinc fingers.**

**Homology Modelling results:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Protein** | **Template Protein** | **Sequence Identity** | **GMQE** | **Q-**  **mean** | **Post Superpose RMSD** | | | |
| **Alpha Carbons** | **Back Bone** | **Heavy** | **All** |
| Wild-type human neuroglobin <http://www.rcsb.org/structure/4MPM> | Horse heart ferric myoglobin <https://swissmodel.expasy.org/templates/3rj6.1> | 14.8 | 0.62 | -1.75 | 1.683 | 1.647 | 2.127 | 2.127 |
| Bar-headed goose haemoglobin <https://swissmodel.expasy.org/templates/1a4f.1> | 27.97 | 0.66 | -2.01 | 1.680 | 1.647 | 2.173 | 2.173 |
| murine neuroglobin mutant V140W <https://swissmodel.expasy.org/templates/4o2g.1> | 92.05 | 0.95 | -0.59 | 1.437 | 1.413 | 1.84 | 1.84 |
| PR domain zinc finger protein9 <https://www.rcsb.org/structure/5v3g> | Human PRDM9 allele-A ZnF Domain with Associated Recombination Hotspot <https://swissmodel.expasy.org/templates/5eh2.1> | 95.07 | 0.66 | 0.25 | 0.425 | 0.490 | 0.828 | 0.828 |
| Zinc finger and BTB domain-containing protein 17 <https://swissmodel.expasy.org/templates/2n25.1> | 43.93 | 0.09 | -0.85 | 0.61 | 0.648 | 1.005 | 1.005 |
| MouseZFP568-ZnF1-11  <https://swissmodel.expasy.org/templates/5v3m.1> | 58.27 | 0.71 | -1.34 | 0.553 | 0.570 | 0.998 | 0.998 |
| Human Cytochrome P450 1A1 <http://www.rcsb.org/structure/4I8V> | Cytochrome P450 1A2 <https://swissmodel.expasy.org/templates/2hi4.1> | 74 | 0.86 | -0.72 | 0.494 | 0.516 | 0.878 | 0.878 |
| Mammalian cytochrome P450 2B4 <https://swissmodel.expasy.org/templates/1suo.1> | 32.04 | 0.64 | -2.19 | 0.772 | 0.792 | 1.142 | 1.142 |
| Cytochrome P450 2R1 <https://swissmodel.expasy.org/templates/3czh.1> | 27.53 | 0.68 | -2.40 | 0.694 | 0.706 | 1.068 | 1.068 |

**Greater the sequence identity lower the RMSD value from superpose. Proteins that are less homologus have higher RMSD value on superpose.**

**Just globin family:**

|  |  |
| --- | --- |
| **3 representatives target-template** | **Structural and Sequence**  **(SuperPose Results)** |
| 1. Neuroglobin – Myoglobin | Structurally same & Sequence alignment is same about 98.7% |
| 1. Myoglobin - Hemoglobin a | Structurally same but sequence alignment is different about 39.4% |
| 1. Hemoglobin a - Hemoglobin b | Structurally same & Sequence alignment is almost same about 96.5% |

**Reference**

[1] <http://wishart.biology.ualberta.ca/SuperPose/help.html>