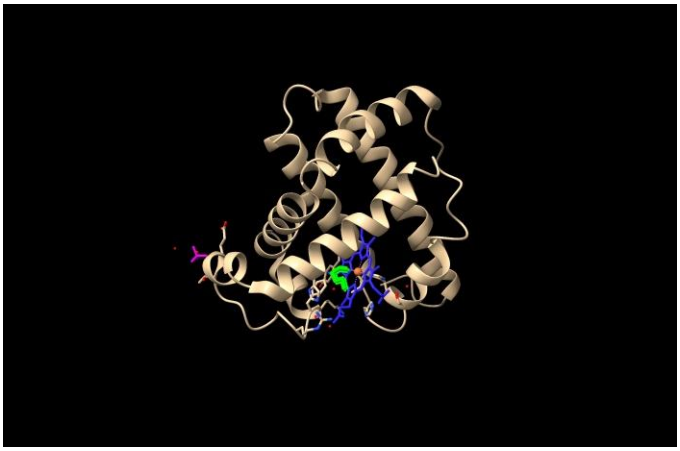
Designing a Protein

Michael Pacheco

CS491: Computational Structural Biology



*Myoglobin 101 3, Ligands Colored*

**1) Scaffold protein**

Myoglobin 101

Scaffold protein origin: Myoglobin can be used for allergy research, and this protein model was originally sourced from sperm whales.

**2) Selecting epitope(s)**

Epitope origin: *Phleum pratense (Timothy Grass)*, is the source of the epitope. I chose the option for allergic diseases in humans, I chose my MHC Restriction from *iedb.org* at random which ended up being class 1 and chose the epitope to be linear and all other options were left at default for *iedb.org*.

LELQFRRVKCKY

**3) Grafting altered epitopes to scaffold protein**

*N1 model sequence:*

LELQFRRVKCKYMVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRVKHLKTEAEMKASEDLKKHGV
TVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQGAMNKALELFRKDIAAKYKEL
GYQG

*C2 model sequence:*

MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRVKHLKTEAEMKASEDLKKHGVTVLTALGAILKK
KGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQGAMNKALELFRKDIAAKYKELGYQGLELQFR
RVKCKYGGGGGGLELQFRRVKCKY

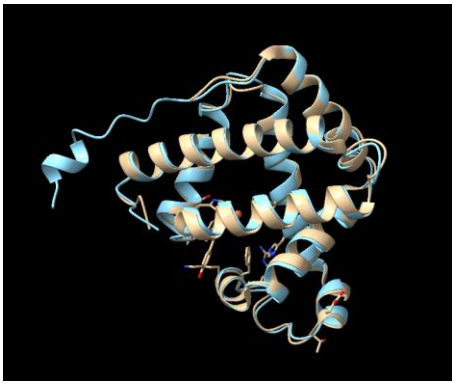
*L3 model sequence:*
MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRVKHLKTEAEMKASEDLKKHGVTVLTALGAILKK
KGHLELQFRRVKCKYHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQGAMNKALELFRKDIAAKYKE
LGYQG

**4) AlphaFold and Robetta model comparisons to original scaffold protein Myoglobin 101**

Needleman-Wunsch/BlOSUM62 Aligned Robetta models of tagged epitope changes to original scaffold protein.

<u>The original scaffold protein, Myoglobin 101, is colored tan below.</u>

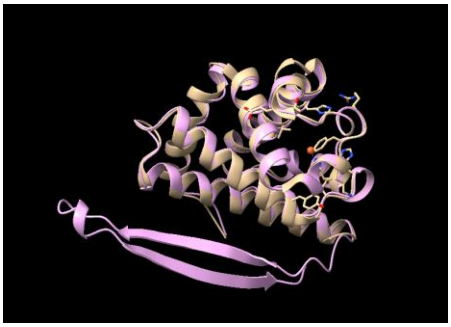
*N1 - Robetta Model Image*



| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker 101m.pdb, chain A (#1) with N1_model1.pdb, chain A (#5), sequence alignment score = 795.5
RMSD between 151 pruned atom pairs is 0.545 angstroms; (across all 154 pairs: 0.644)

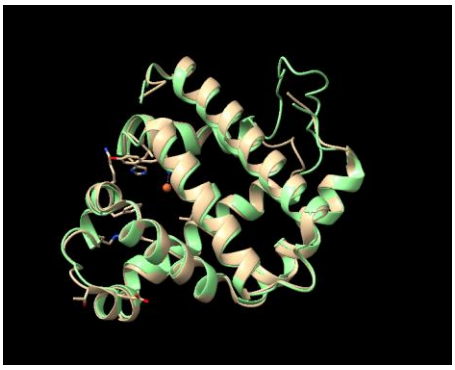
*C2 - Robetta Model Image*



| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker 101m.pdb, chain A (#1) with C2_model1.pdb, chain A (#6), sequence alignment score = 795.5
RMSD between 152 pruned atom pairs is 0.559 angstroms; (across all 154 pairs: 0.731)

*L3 - Robetta Model Image*



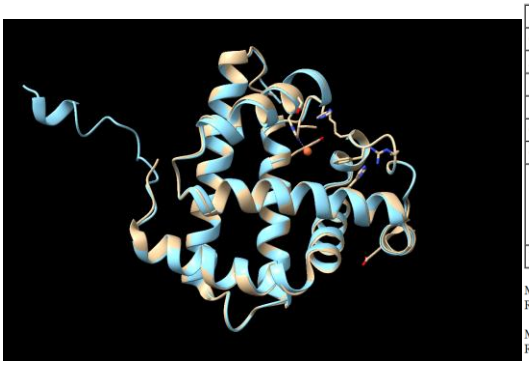
| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker 101m.pdb, chain A (#1) with L3_model1.pdb, chain A (#7), sequence alignment score = 765.5
RMSD between 149 pruned atom pairs is 0.644 angstroms; (across all 154 pairs: 0.877)

<u>Observation</u> – Yes, the scaffold protein changed conformation approaching the epitopes, but only minor changes along the rest of the two models.

Needleman-Wunsch/BlOSUM62 Aligned AlphaFold models of tagged epitope changes to original scaffold protein.

*N1 - AlphaFold Model Image*

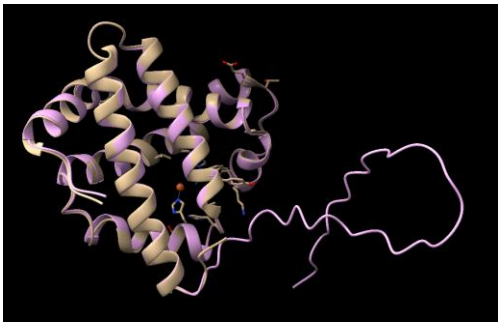


| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker 101m.pdb, chain A (#1) with ranked_0.pdb, chain A (#2), sequence alignment score = 802.7
RMSD between 153 pruned atom pairs is 0.369 angstroms; (across all 154 pairs: 0.437)

Matchmaker 101m.pdb, chain A (#1) with ranked_0.pdb, chain A (#3), sequence alignment score = 768.5
RMSD between 153 pruned atom pairs is 0.293 angstroms; (across all 154 pairs: 0.380)

*C2 - AlphaFold Model Image*

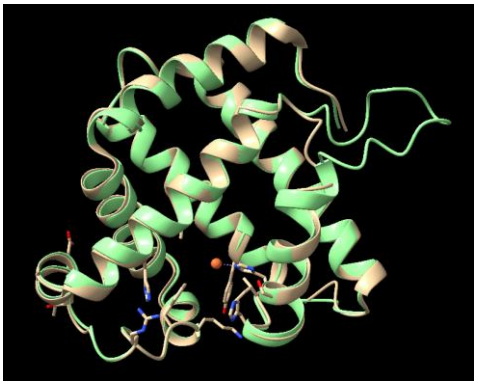


| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker 101m.pdb, chain A (#1) with ranked_0.pdb, chain A (#2), sequence alignment score = 802.7
RMSD between 153 pruned atom pairs is 0.369 angstroms; (across all 154 pairs: 0.437)

Matchmaker 101m.pdb, chain A (#1) with ranked_0.pdb, chain A (#3), sequence alignment score = 768.5
RMSD between 153 pruned atom pairs is 0.293 angstroms; (across all 154 pairs: 0.380)

*L3 - AlphaFold Model Image*



| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker 101m.pdb, chain A (#1) with ranked_0.pdb, chain A (#2), sequence alignment score = 802.7
RMSD between 153 pruned atom pairs is 0.369 angstroms; (across all 154 pairs: 0.437)

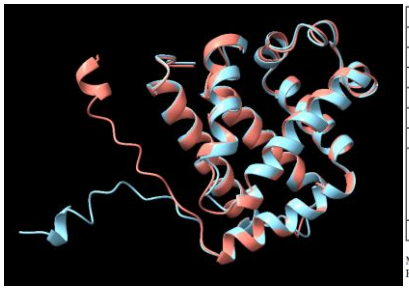
Matchmaker 101m.pdb, chain A (#1) with ranked_0.pdb, chain A (#3), sequence alignment score = 768.5
RMSD between 153 pruned atom pairs is 0.293 angstroms; (across all 154 pairs: 0.380)

Matchmaker 101m.pdb, chain A (#1) with ranked_0.pdb, chain A (#4), sequence alignment score = 744.5
RMSD between 149 pruned atom pairs is 0.403 angstroms; (across all 154 pairs: 0.672)

<u>Observation</u> – Similar to the Robetta models, the most significant conformational changes were around the atoms approaching the tagged epitopes.

**5) Model-to-model comparions between AlphaFold and Robetta tagged epitopes.**

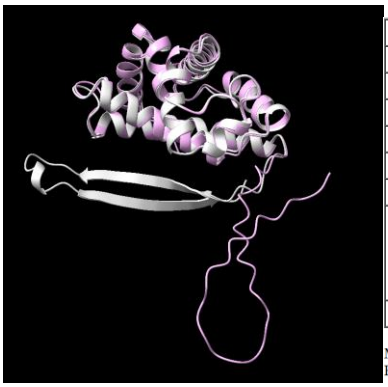
*N1 - Robetta: red, AlphaFold: blue*



| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker ranked_0.pdb, chain A (#2) with N1_model1.pdb, chain A (#5), sequence alignment score = 866.7
RMSD between 145 pruned atom pairs is 0.515 angstroms; (across all 166 pairs: 5.258)

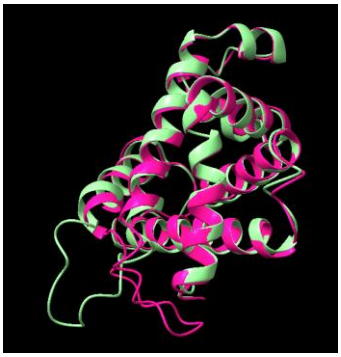
*C2 - Robetta: silver, AlphaFold: pink*



| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker ranked_0.pdb, chain A (#3) with C2_model1.pdb, chain A (#6), sequence alignment score = 892.3
RMSD between 146 pruned atom pairs is 0.533 angstroms; (across all 184 pairs: 14.060)

*L3 - Robetta: pink, AlphaFold: green*



| Parameters | |
|---|---|
| Chain pairing | bb |
| Alignment algorithm | Needleman-Wunsch |
| Similarity matrix | BLOSUM-62 |
| SS fraction | 0.3 |
| Gap open (HH/SS/other) | 18/18/6 |
| Gap extend | 1 |
| SS matrix | H S O<br>H 6 -9 -6<br>S 6 -6<br>O 4 |
| Iteration cutoff | 2 |

Matchmaker ranked_0.pdb, chain A (#4) with L3_model1.pdb, chain A (#7), sequence alignment score = 852.3
RMSD between 146 pruned atom pairs is 0.648 angstroms; (across all 166 pairs: 3.465)

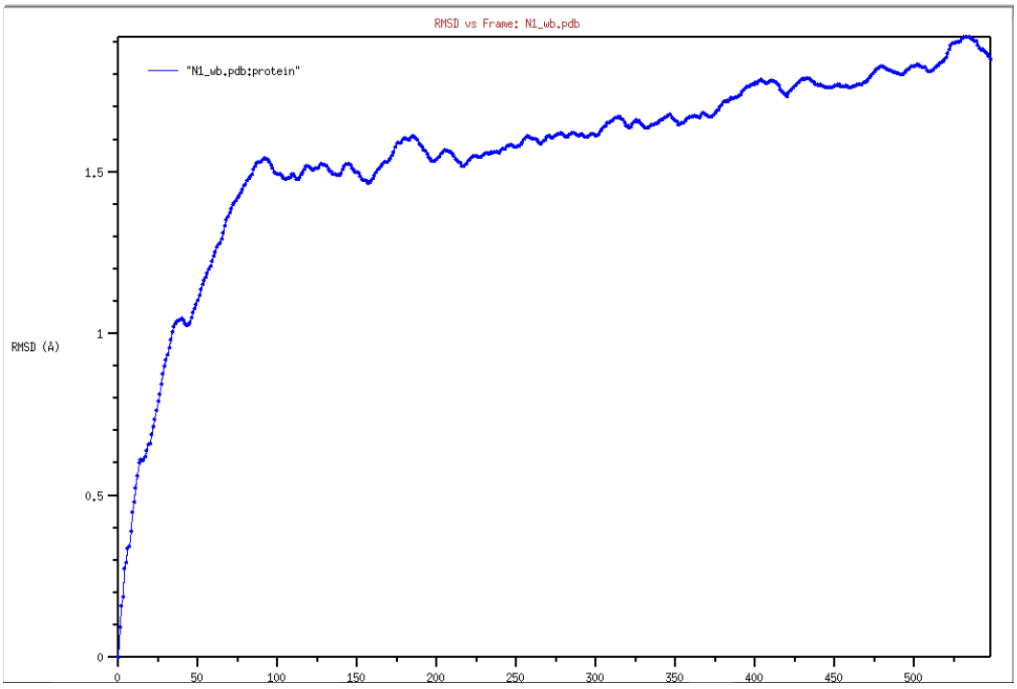
Observation - The structural prediction differences between both AlphaFold and Robetta are most significant for the tagged epitopes. Because Myoglobin's structure has been experimentally determined, both models are able to reference it, while the tagged epitope's structure may not be experimentally determined for reference or disordered, forcing the two methods to have reached different conclusions on the structure. You can observe that all the structures for the tagged epitopes from the AlphaFold method tend to be more 'spread out' from the Myoglobin molecule while the Robetta models tend to 'stay closer' to it.

**6) Molecular Dynamics Simulation**

Plot of energy over time and rmsd over time?

*Energy vs Timestep/Time*


*RMSD vs Frame/Time*




Observation: I would say that the molecule has equilibrated. We can try to keep looking for lower energies but the simulation had stopped at 1 million steps with a fairly stable RMSD and a more stable energy value for their times.

**References:**

PDB source: https://www.rcsb.org/structure/101M

Epitope source: https://iedb.org/result_v3.php?cookie_id=1a446b