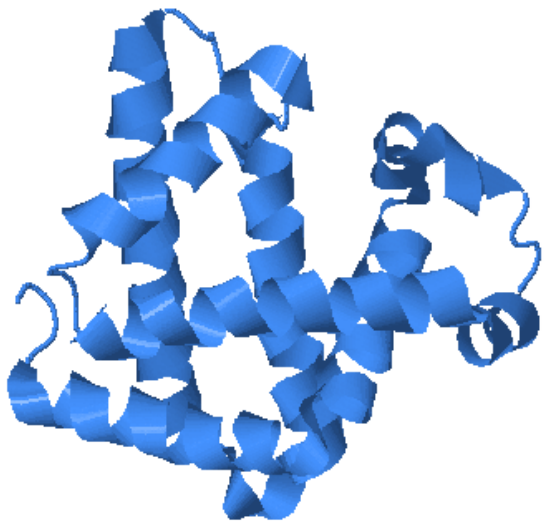
Protein folding!

Team 1:
Saketh Marrapu and Matthew Russell

December 2, 2025

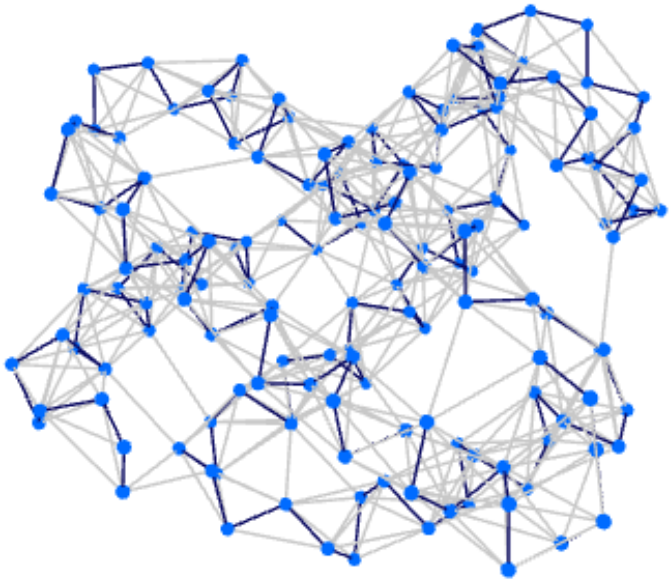
Proteins and networks

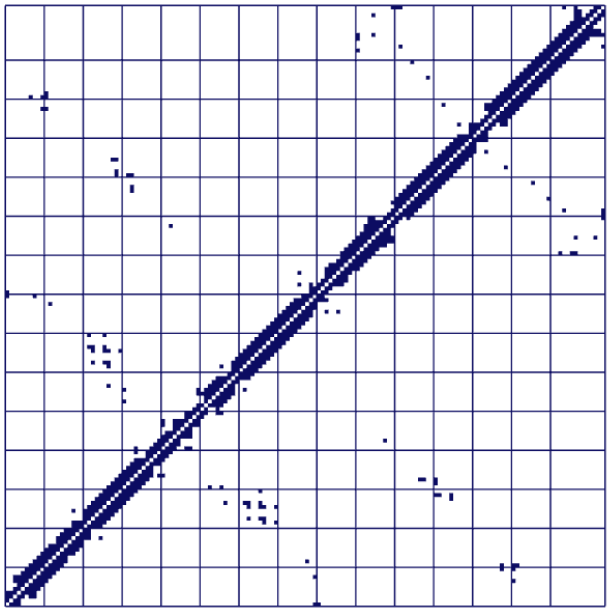
- Proteins: biomolecules comprised of amino acids.
- Amino acids: alanine, cysteine, ...
- Each protein contains one or more chains of amino acids (*residues*).
- Residues are sequentially ordered along the chain: residue 1 is adjacent to residue 2, which is adjacent to residue 3, and so on.
- However, in reality, proteins end up as a three-dimensional jumble.
- Some residues may end up close to each other, even though they are sequentially adjacent.



How to construct the network

- Pick some cutoff value (we used 7 Å).
- If the "backbone atoms" of two residues are closer than the cutoff, then say that the corresponding nodes are adjacent.





Which amino acids are most central?

- Pick a particular measure of centrality — in our case, we chose eigenvector centrality.
- Choose some collection of proteins (structures available from the Protein Data Bank).
- Are some amino acids more likely to be central in the corresponding networks than others (once we account for the relative frequency of each amino acid)?
- First attempt: look only at myoglobin proteins (from different species, different mutations, etc.).
- These were too similar to each other.

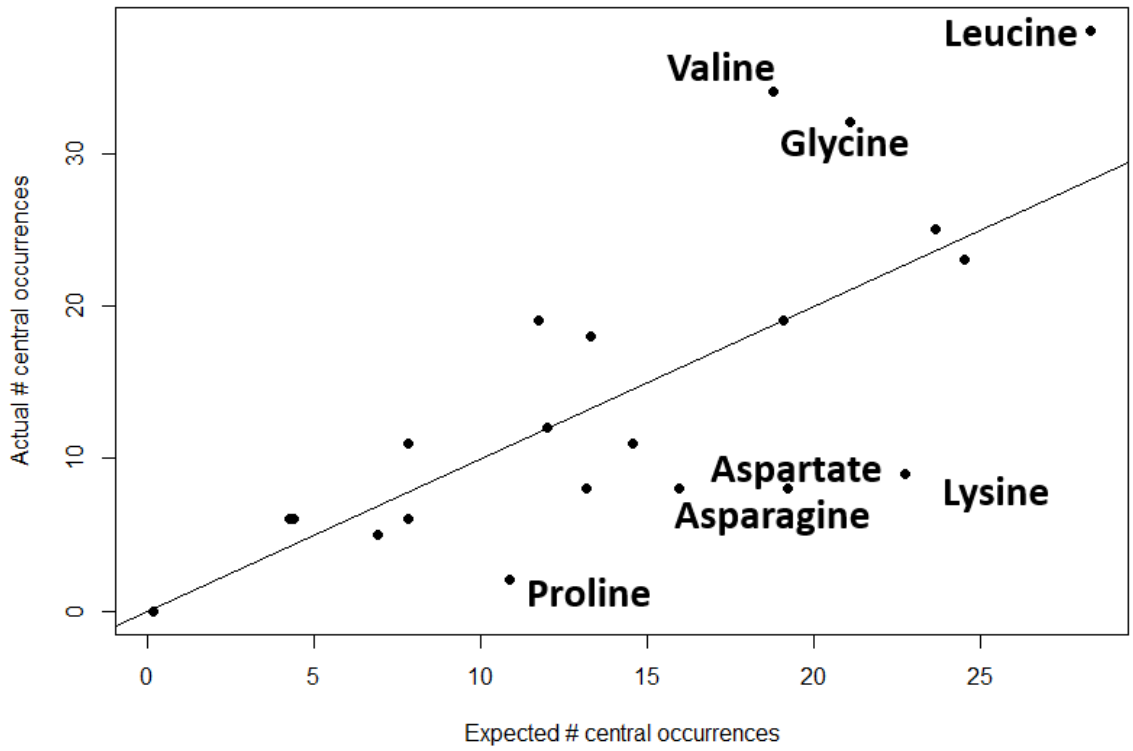
Which amino acids are most central?

- Second attempt: consider six families of proteins (myoglobin, lysozyme, calmodulin, carbonic anhydrase, PGK, AHSP).
- Choose five examples of each from the Protein Data Bank.
- Compute the 10 most central residues (using eigenvector centrality).
- Examine which amino acids are most likely to be central.

Results

- Most likely to be central: glycine, leucine, valine
- Least likely to be central: asparagine, aspartate, lysine, proline

- Biochemistry connection: asparagine, aspartate, lysine are all electrically charged and/or polar.
- On the other hand, glycine, leucine, and valine are all both neutral in charge and nonpolar.



Discussion and future work

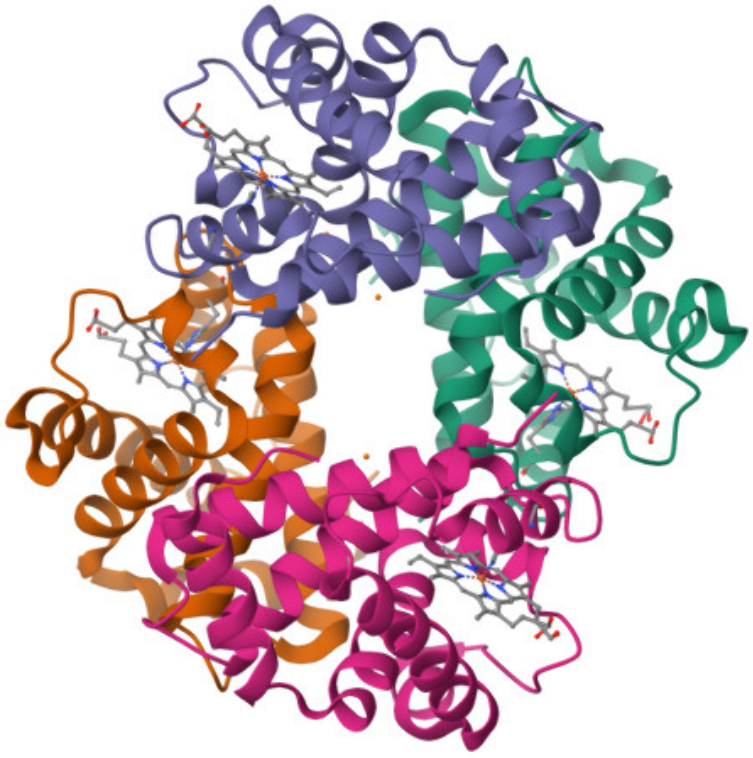
Ideas come from Broto Chakrabarty and Nita Parekh, "NAPS: Network analysis of protein structures". *Nucleic Acids Research*, 44(W1):W375–W382, 05 2016.

Other possible ways to construct networks:

- Change the cutoff value of 7 Å
- Replace "distance between backbone atoms" with "distance between any two atoms"
- Weighted versions of these: for example, use weight equal to the reciprocal of the distance between the residues, or the number of pairs of atoms inside the cutoff

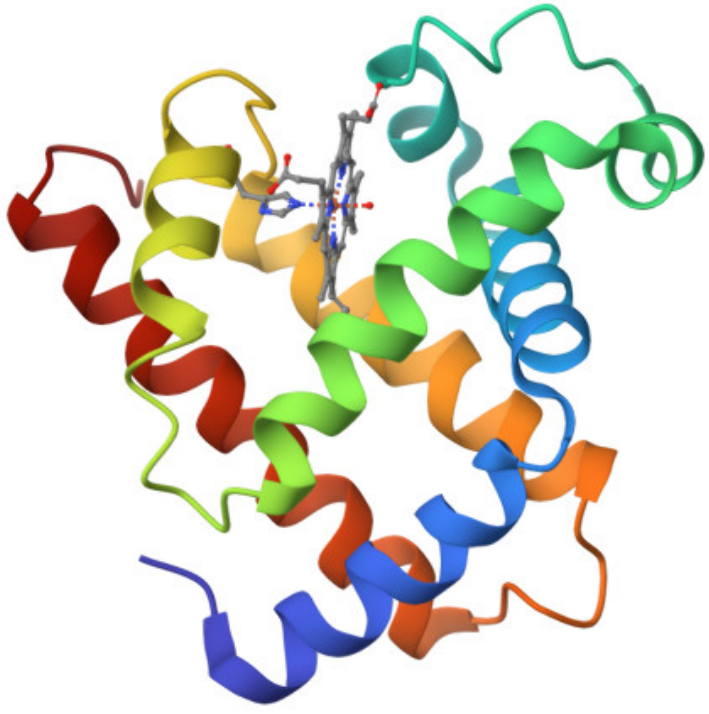
Repeat eigenvector centrality analysis in these new cases.

Background Knowledge



Hemoglobin

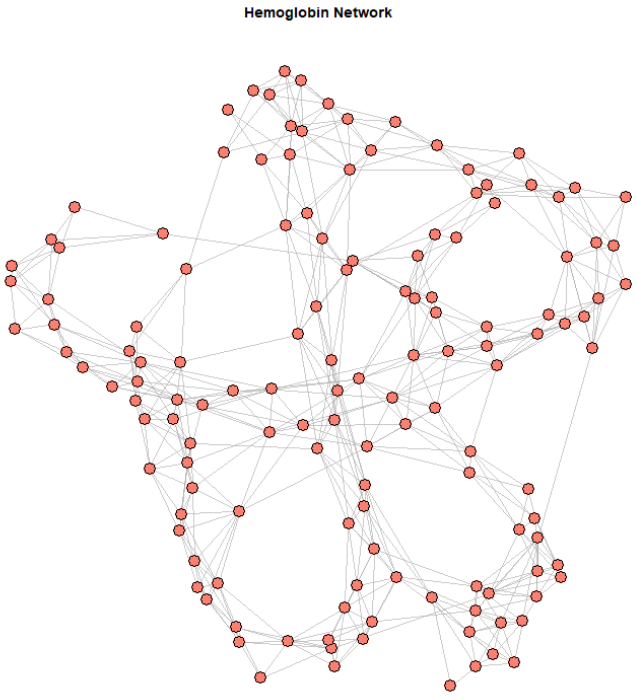
- Globin Fold
- Tetramer so look at α_1!
- Transports Oxygen

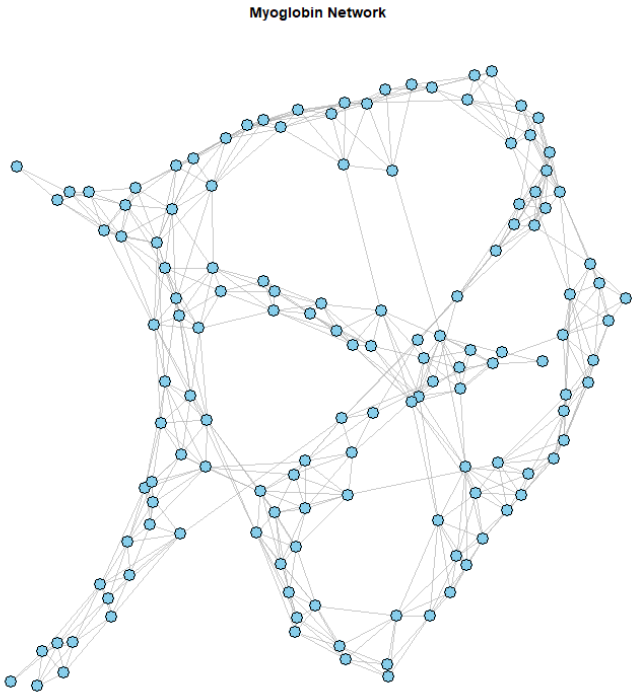


Myoglobin

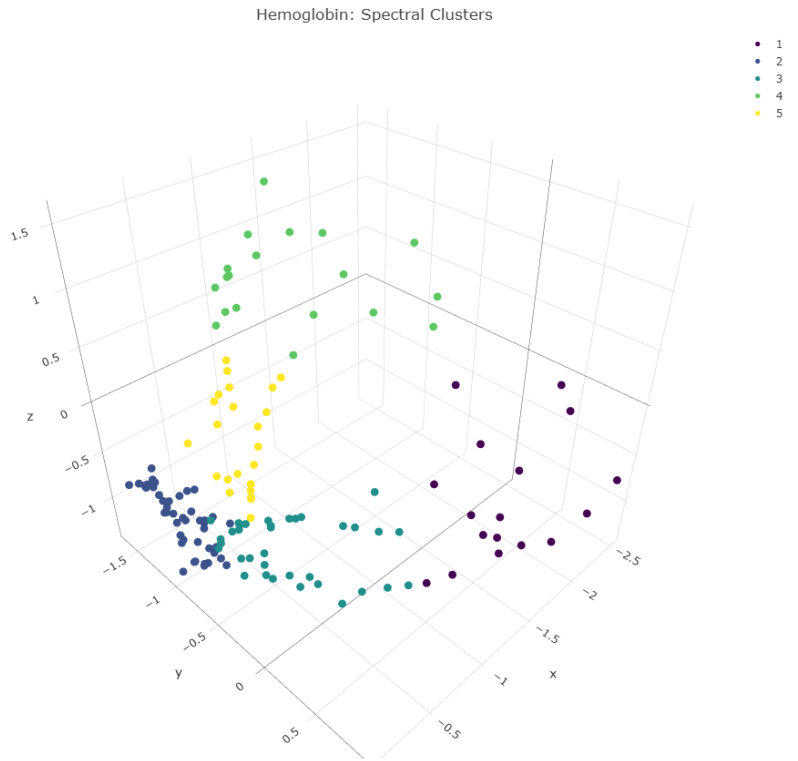
- Globin Fold
- Monomer
- Only holds Oxygen

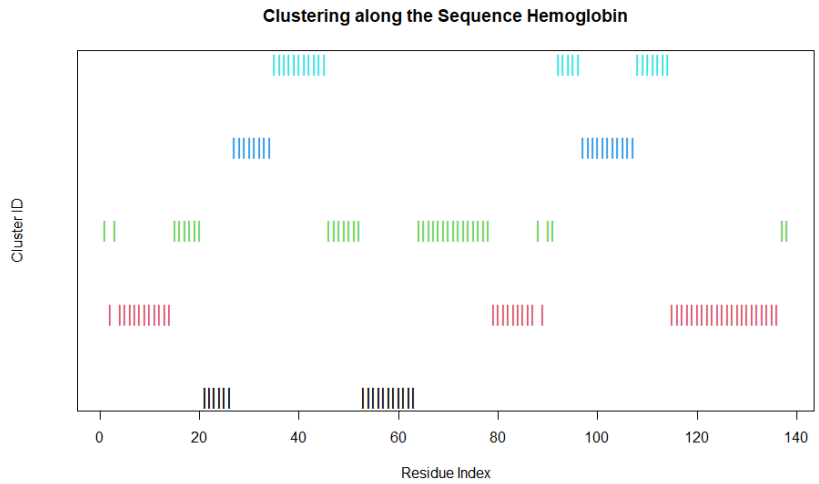
Graph Representations





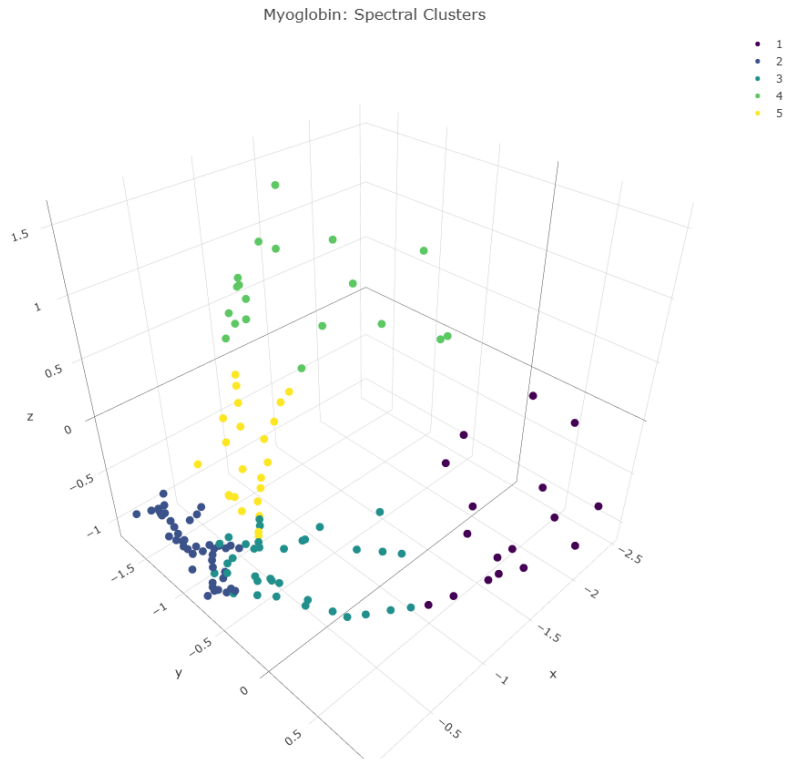
Spectral Clustering for Hemoglobin





1. $\alpha_1\beta_1$ **interface** - holds protein together
2. **Heme-Pocket** - globin fold that allows it to hold oxygen
3. **Allosteric Switch** - switches from Tense to Relaxed
4. **Scaffold** - provides structural stability
5. **Bridge** - connects the scaffold to the active site

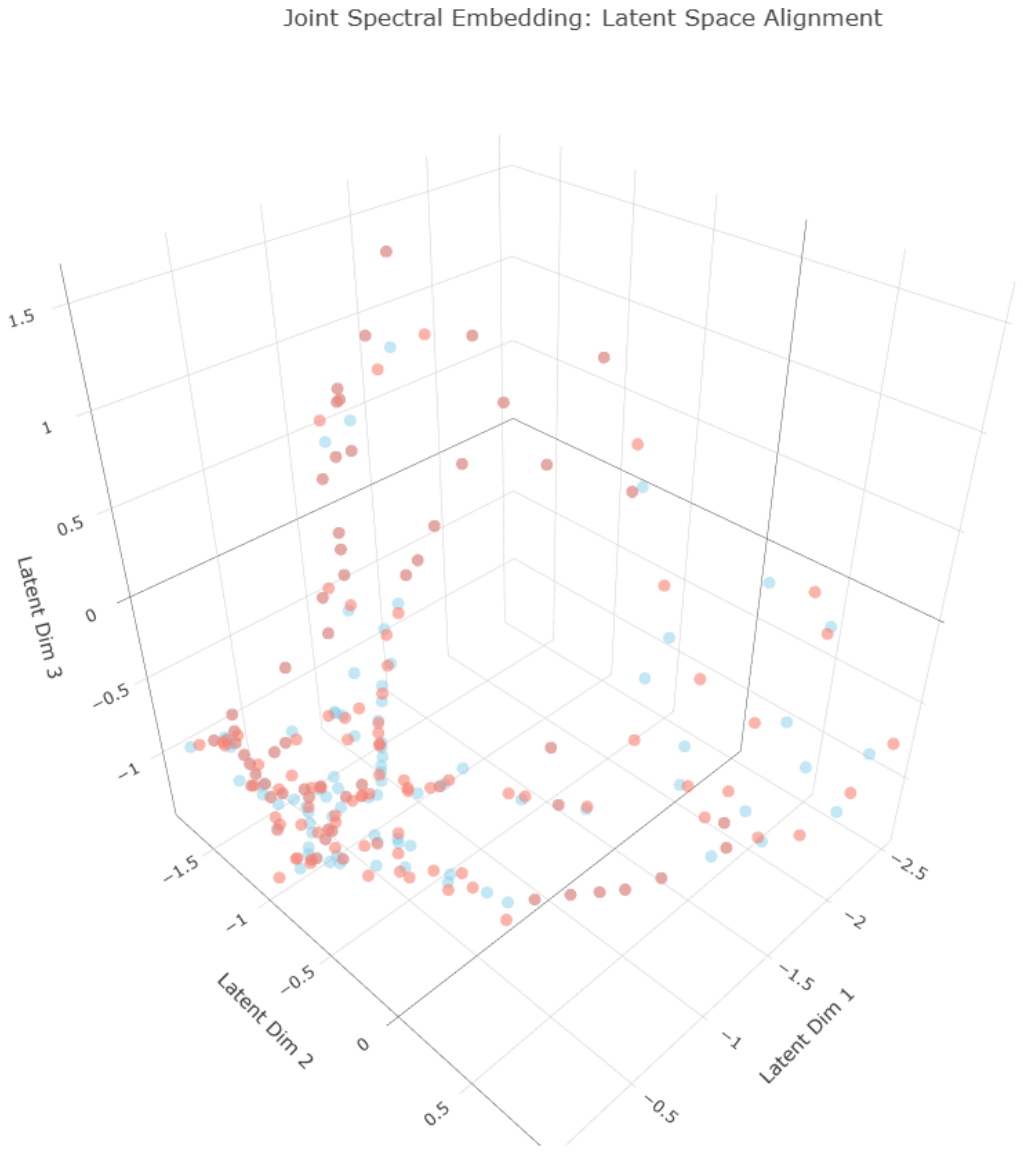
Spectral Clustering for Myoglobin



1. **E-F-A Corner** - entrance to heme pocket
2. **Scaffold** - provides structural stability
3. **Heme-Pocket** - rigid cluster that wraps around oxygen
4. **Hinge** - floppy loops that allow for movement
5. **Exposed Surface** - interacts with solvent

,

Spectral Mapping



- Hb and Mb share a very similar "protein core"
- Embed them into the same 3d space
- Look at the distance between them to determine "difference in function" of each residue in the protein

Distance Difference

- **Residue 137** - *Tyrosine-140 (Hb) vs Tyrosine-146 (Mb)* - This residue is useless in Myoglobin, but in Hemoglobin it is particularly significant because it enables important function in Allostery by triggering the T state as it forms a hydrogen bond
- **Residue 54** - *Lysine-56 (Hb) vs Lysine-62 (Mb)* - In myoglobin this residue faces the water and does little. In Hemoglobin this is part of the E-helix that forms the $\alpha_1\beta_1$ interface, essentially working as the glue to make it a Tetramer.
- **Residue 82** - *Serine-84 (Hb) vs Alanine-90 (Mb)* - These correspond to the residue that pulls the F-helix when Oxygen is bound. In Myoglobin this movement is small and local, but in Hemoglobin this movement is transferred to other chains and thus has to be more connected in order to transmit this signal to other chains.