Detection of Cysteine-Tyrosine Crosslinks Using Quantitative Metrics

Terryce Nederhood
Benson Lab
Summer 2016 Research Summary

Posttranslational modifications of tyrosine molecules are of interest because they extend biological functions while remaining attached to the protein backbone. These modifications give the protein the ability to do reversible one-electron redox chemistry. Similar to metal ion cofactors. The cysteine-tyrosine crosslink that is formed by bonding the epsilon carbon of the tyrosine molecule with the sulfur gamma of the cysteine. In 2010 the Benson lab searched the protein data bank in order to identify any proteins with crosslink known and they stumbled upon BF4112, an orphan protein, with a pair of cysteine and tyrosine side chains with correct geometry for crosslink formation located near the copper metal center. Our goal this summer was to use computational analysis to develop a quantitative way to identify the presence of previously formed cysteine-tyrosine crosslinks already in the protein data bank

We began the project by downloading experimental(structure factor) and coordinate (PDB) files from the protein databank to generate electron density maps for visual analysis of how the coordinates represent diffraction data. Phasing (Phenix Phaser) converted the experimental reflections (structure factor files) into an electron density map using the molecular replacement method (coordinate files included as well). Following phasing, the modelled coordinates were refined (Phenix Refinement) to best fit the electron density map.The refined coordinates and electron density maps were visually examined using the Coot program. Each refinement was ran twice; with no adjustments to the protein structure or with. A parameter file that added a bond between cysteine sulfur gamman and tyrosine carbon epsilon (crosslink). In order to easily identify new crosslinks a metric was needed that could serve as a filter for potential structures.

The clashschore metric was the only metric that changed in our refinement comparisons. Clashscore is the measure of the number of clashes the protein per one thousand atoms. We expected that if the parameter file was added we would see an increase in the clashscore because bringing two molecules closer together would create a larger clash, but after our results showed the opposite. We talked with Dave and Jane Richardson (authors) and we learned that once two residues get pushed too close together instead of being classified as a clash, they are classified as a bond, decreasing the clashscore. The MolProbity score from the Phenix packages was suggested as an alternative metric which combines the clashscore, rotamer and Ramachandran data into a single score. For data and structures known to contain a cysteine-tyrosine crosslink there was a decrease in the MolProbity score as expected.

Unfortunately clashscores and MolProbity scores were positively, This correlation suggests that changes in the MolProbity score is dominated by the clashscore when adding cysteine-tyrosine bond constraints. This suggested that MolProbity score is not the metric we are searching for.

We have now decided to focus on metrics for the individual atoms/residues rather than global metrics for cystallographic fit. The two metrics currently investigated are the carbon-sulfur bond length ($r_{c\text{-}s}$) from refinement with the bonded constraint and the sharpness of the electron density around each atom (individual B factor) These metrics were chosen because the addition of one bond does not change global parameters (clashscore and MolProbity score), and need individual parameters to describe goodness-of-fit. After comparing the bond length to the final atomic B factor of structures known to contain a cysteine-tyrosine crosslink, it was noted that a B factor under 70 and a bond length less than 2.0 angstroms was a good filter for candidate structures. When these constraints were applied to PDB files previously identified from C-S contacts less than 3.2 angstroms there were two PDB entries from the ubiquitin-ligase UBE2G2 (3H8K and 4LAD), with electron densities consistent with a cysteine-tyrosine crosslink. PDB entries from this search with carbon-sulfur bond lengths between 2.0 and 2.6 angstroms will also be examined as these structures might represent partial cysteine-tyrosine crosslink occupancies. One PDB entry of interest is the B chain of alaine racemase (1EPV). I plan on wrapping my summer up by taking a deeper look at these and other potential cysteine-tyrosine crosslink containing structures.

I was really excited to start this research because protein structure and crystallization are things that I have always found interesting and wanted to learn more about. The incredible thing about research is that you learn so much more than you ever expected to. This summer not only did I learn more about protein structure, but I was able to learn how to navigate a Linux command line and how structures are refined in order to create an accurate model. Doing research gives you the chance to apply what you have learned in the classroom to a real life problem which is something incredibly amazing about science. This summer I was able to develop my problem solving skills in order to help think of alternate solutions to a problem. This summer has been a great learning experience and I look forward to continuing this research and hopefully finding a solution to the problem we set out to solve.