


Spring 2015

# Structural studies of the Ig58 domain of the giant muscle protein obscurin

Matthew C. Oehler  
*James Madison University*

Follow this and additional works at: <https://commons.lib.jmu.edu/honors201019>

 Part of the [Medical Biophysics Commons](#), and the [Musculoskeletal Diseases Commons](#)

---

## Recommended Citation

Oehler, Matthew C., "Structural studies of the Ig58 domain of the giant muscle protein obscurin" (2015). *Senior Honors Projects, 2010-current*. 49.  
<https://commons.lib.jmu.edu/honors201019/49>

This Thesis is brought to you for free and open access by the Honors College at JMU Scholarly Commons. It has been accepted for inclusion in Senior Honors Projects, 2010-current by an authorized administrator of JMU Scholarly Commons. For more information, please contact [dc\\_admin@jmu.edu](mailto:dc_admin@jmu.edu).

**Structural studies of the Ig58 domain of  
the giant muscle protein obscurin**

---

An Honors Program Project Presented to  
the Faculty of the Undergraduate  
College of Science and Mathematics  
James Madison University

---

by Matthew Curtis Oehler

May 2015

---

---

Accepted by the faculty of the Department of Chemistry and Biochemistry, James Madison University, in partial fulfillment of the requirements for the Honors Program.

FACULTY COMMITTEE:

---

Project Advisor: Nathan Wright, Ph. D.,  
Assistant Professor, Biochemistry

---

Reader: Gina Macdonald, Ph. D.,  
Professor, Biochemistry

---

Reader: Christopher Berndsen, Ph. D.,  
Assistant Professor, Biochemistry

---

---

HONORS PROGRAM APPROVAL:

---

Philip Frana, Ph.D.,  
Interim Director, Honors Program

---

---

PUBLIC PRESENTATION

This work is accepted for presentation, in part or in full, at the Department of Chemistry and Biochemistry Spring Symposium on 03/26/2015.

## Table of Contents

List of Figures	3
Acknowledgements	4
Abstract	5
Introduction	6
Experimental	17
Results	20
Discussion	28
Conclusions	30
References	31
Appendix	35

## List of Figures

### Figures

1	Schematic of the sarcomere with reference terms	7
2	Diagram of actin, myosin, troponin, and tropomyosin	8
3	Schematic of the sarcomere with giant muscle proteins highlighted	10
4	Schematic of titin and it's spring-like capabilities	11
5	Obscurin schematic with highlighted HCM pathology	13
6	Schematic of nucleus in applied magnetic field	16
7	X-ray structure of Ig58	22
8	20 best NMR structures and NMR/X-ray overlay	25
9	HSQC of Arg→Gln mutation projected onto the high-resolution NMR structure	26
10	Electron density fit of charge surface of Ig58	27
11	Molecular dynamics simulation of charge surface	28

### Tables

1	Statistics of the X-ray structure of Ig58, at 2.41 Å resolution	23
2	NMR-derived restraints and statistics of 20 NMR structures	24

## **Acknowledgments**

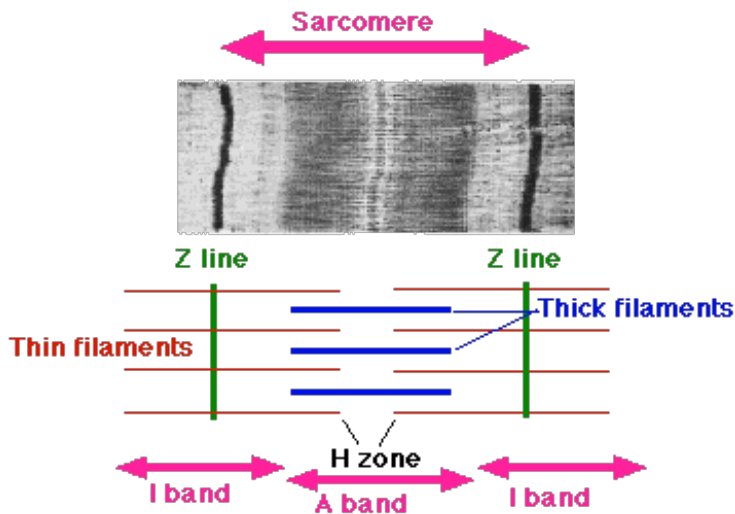
I would like to thank Dr. Wright and Dr. Berndsen for assistance in data collection and analysis for these experiments, as well as the JMU Dept. of Chemistry and Biochemistry for use of instrumentation. I would also like to thank the Virginia Tech crystallography lab and Dr. Carla Slebodnick for use of their instrumentation and assistance in data collection of the low-resolution data set of Ig58. Finally I would like to thank Logan Meyer for her initial work on Ig58 that provided a platform for my investigations and interpretations and Kelly DuPont for her assistance in data analysis and structural refinement of the crystal structure of Ig58, as well as the creation of the appendix herein.

## **Abstract**

Obscurin (720-900 kD) is a giant sarcomeric signaling protein that is the only known link between the cytoskeleton and the surrounding membrane structure. Mutations to obscurin and to obscurin binding partners have been linked to human muscle diseases such as hypertrophic cardiomyopathies and muscular dystrophy. These diseases likely occur due to the abrogation of specific molecular interactions necessary for suitable function. To more fully understand how specific mutations lead to disease, here we solve the high-resolution structure of obscurin Ig58. The literature shows that an Arg8Gln mutation to the Ig58 domain of obscurin is associated with hypertrophic cardiomyopathy (HCM). Chemical shift changes of this mutation and MD simulations suggest that this mutation disrupts a large charge-charge surface of Ig58, perturbing the titin-binding interface.

## Introduction

Muscular motion, at its most basic level, results from the motions of multiple connected proteins, each with one or many unique functions, within the cell. In order to understand these molecular motions, we can study each of these unique proteins and investigate their role in force generation, stretch resistance, and stretch signaling in muscle. This will provide insight into why certain mutations lead to a breakdown of muscle cells and a pathological loss of the components that underlie controlled muscle motion. Muscle proteins are exquisitely arranged within the myofibril, and make up the macromolecular structure called the sarcomere. Sarcomeres are the most basic unit of contraction in striated muscle<sup>1</sup>. Figure 1 shows a histological slide of striated muscle under a microscope with the two main components of the sarcomere, Actin and Myosin, appearing as light and dark horizontal lines, respectively (Figure 1). These two long filamentous fibers that slide past each other during muscular contraction and relaxation<sup>2,3</sup>. In order to understand where different protein components of the sarcomere localize, it is necessary to recognize the specific terminology used for the arrangement of the sarcomere: I-band, A-band, Z-lines, and M-line. The isotropic band, or I-band, is the area of actin polymers that have no overlap with myosin<sup>3</sup>. The length of the I-band is subject to change as the sarcomere shortens, resulting from the actin and myosin filaments sliding past each other upon



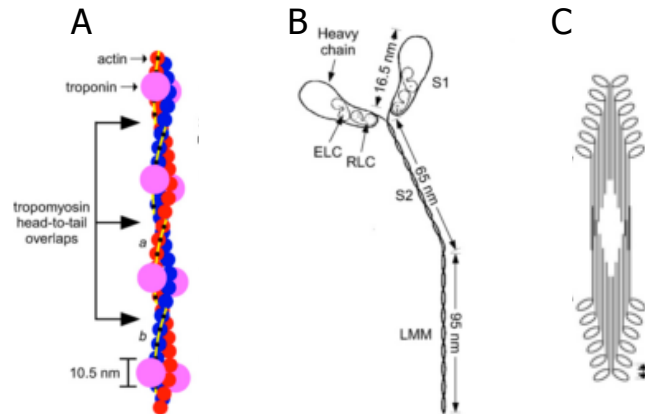
**Figure 1.** Schematic of the sarcomere directly aligned with a histological slide of striated muscle<sup>4</sup>.

contraction, thus producing a shortened area of actin filaments without myosin overlap. The darker anisotropic band, or A-band, refers to the entire length of myosin filaments and does not change in length upon contraction<sup>3</sup>. The length of a single sarcomere is defined as the distance between two Z-lines, which correspond to the dark vertical bands of the histological slide in Figure 1. The Z-line of striated muscle is also the point where the pointed termini of actin polymers are anchored<sup>5</sup>. The M-line is defined as the middle of the sarcomere and refers to the cross connecting of the two bipolar chains of a single thick filament<sup>6</sup>.

F-actin, referred to as the thin filament, is a double helix of actin monomers and is roughly 40 Å in diameter but is as much as several microns in length. Two important thin filament associated proteins in striated muscle are the globular protein troponin and the filamentous protein tropomyosin shown in pink and blue of Figure 2A, respectively. For every helical turn of actin, there are two troponin complexes lying in direct contact with



both actin and tropomyosin. Tropomyosin, twists around the actin double helix and sterically blocks the myosin binding sites in the absence of  $\text{Ca}^{2+}$  ion<sup>2,1</sup> (Figure 2A).



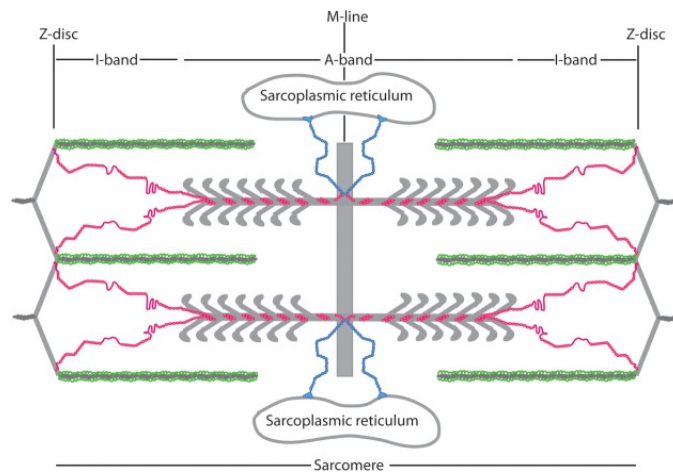
**Figure 2.** Schematic of **A)** actin filament wrapped by tropomyosin with troponin **B)** single myosin filament including light and heavy chain regions **C)** myosin filament bundle. Figure adapted from Hooper *et al*<sup>2</sup>.

Myosin, often referred to as the thick filament of about 110 Angstroms in diameter, is composed of three pairs of molecules<sup>1 2 3</sup>. The first pair is known as the heavy chain, which wind together to form a coiled-coil tail, both shown in Figure 2B as two globular heads and thin long tails, respectively. The other two are known as the essential light chain and the regulatory chain, which combined with the other end of the heavy chain form one of the combined molecule's two globular heads<sup>2</sup>. All of the components of the thin and thick filament are necessary for contraction to occur.

Muscle contraction takes place when calcium ions are released from the sarcoplasmic reticulum and are able to move in to the sarcomere and bind to troponin<sup>1,7</sup>. This calcium binding alters the structure of troponin molecules and causes them to move

the helical tropomyosin chain they are attached to, revealing previously covered myosin-binding sites on the actin surface<sup>1,7</sup>. The myosin head binds and hydrolyzes ATP at its globular head region. This hydrolysis then initiates a conformational change where the myosin head is then 'cocked'<sup>1,7</sup>. This complex then binds to the actin strand. After binding to actin, the myosin head produces the power stroke, where the head repositions to its original orientation and the filaments slide past each other<sup>1,7</sup>. ADP then leaves the Myosin head and ATP is free to enter. This event results in myosin detaching from actin, thus allowing the contractile apparatus and muscle cell to relax. The cycle can then repeat as long as tropomyosin does not occlude the myosin binding site on actin. Large-scale motions that result from muscular contractions are therefore, at their most basic level, the direct result of microscopic molecular movements.

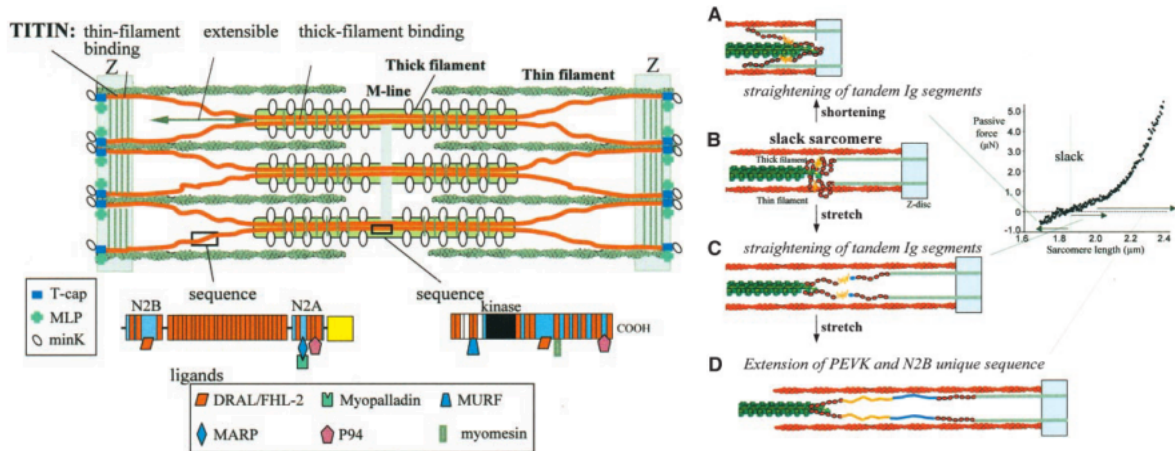
The sliding filament hypothesis<sup>7</sup> is almost always presented in textbooks exclusively for its properties of the generation of force involved in the contraction of muscle and presents several important protein structures that work in tandem to make this possible. A complete understanding of the sarcomere however, is much more involved. The main aim of this research is to investigate the equally important ability of muscle to not only contract, but to expand. However, there is a general lack of understanding about the properties of striated muscle for regulated expansion, and resistance to expansion. While there are likely many proteins involved in this process, there is a limited understanding of what players are involved. Probably the most well understood component of the regulation of muscle expansion is the giant filamentous protein titin and secondly obscurin. Both obscurin and titin are represented below in Figure 3, in their likely locations based on known binding targets.



**Figure 3.** Schematic of the sarcomere with giant muscle proteins highlighted. Obscurin, in blue, links titin (red) to the surrounding SR membrane and T-tubule structure in striated muscle. Figure from Meyer and Wright<sup>8</sup>.

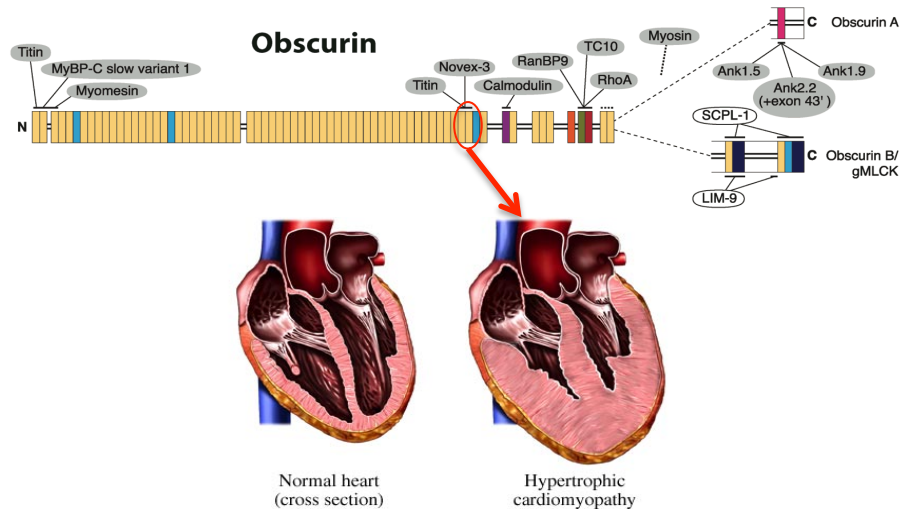
Full length titin, is about 2993 kD with a pI of 6.35<sup>9</sup>, is expressed in multiple isoforms, which vary in both length and strength, and is necessary for proper myocyte structure and function. Titin is made up of over 300 individual Ig-like folds, arranged in tandem and interspersed with occasional enzymatically active regions or other signaling motifs<sup>10,11,12</sup> (Figure 4A). The protein in total resembles a long rope or chain and has been referred to as the “molecular ruler” of the sarcomere. Titin is a stretch sensor that regulates the length of the sarcomere<sup>10,11</sup>. While each individual domain is relatively immobile, the linkers between the domains are dynamic. Thus, titin can form many different conformations, much like a section of chain links can form many different kinds of shapes<sup>7</sup>. Titin, with its N and C terminal binding to the Z-band and M-band respectively, becomes elongated when the muscle stretches<sup>7</sup> (Figure 4B). Titin therefore acts like an enthalpic-entropic spring<sup>7</sup>. As it is gradually elongated, individual titin Ig-like

domains straighten out, relative to each other. This action breaks inter-Ig noncovalent forces, and this produces an enthalpic component to stretching. Additionally, as titin stretches out it can conform to fewer and fewer conformations, and thus has a lower amount of entropy in the stretched form. These two basic forces allow titin to gradually resist more and more force the further it is stretched, thus giving titin a bungee-cord like property to resist physical manipulations. Titin is extensively cross-linked at both the Z-disk and M-band. The interactions with proteins in these areas such as  $\alpha$ -actinin, MyBP-C, and obscurin not only generate the cytoskeletal architecture that the sarcomere relies on, but also provides malleable and dynamic protein mesh that can both resist and sense physical stretching forces.



**Figure 4.** **A)** Schematic of titin in the sarcomere with domains and binding partners highlighted for regions near the M-band and Z-line. **B)** Sequence of stretching showing titin’s spring-like properties. Figure adapted from Granzier *et al*<sup>7</sup>.

Similar in structure to titin is the protein obscurin. The structural similarity between the two suggests that they may also share a functional similarity. Obscurin is the most recently discovered giant protein involved in muscle contraction and was first identified as a ligand of the Z-disk titin domains Z9 and Z10<sup>12</sup>. Full length obscurin is roughly 720 kD and is known to play an important role in skeletal and cardiac muscle assembly<sup>12</sup>. The human isoform-a of obscurin has a pI of 5.54<sup>9</sup>. While titin runs longitudinally through muscle cells, obscurin localization is more complicated. Obscurin intercalates through the sarcomere in an apparently random pattern, and links the sarcomere to the surrounding membrane structures in muscle cells<sup>13,14</sup>. In recent findings, obscurin has also been found in a diffuse pattern in non-muscle cell types<sup>15</sup> however, it is currently unclear why obscurin expresses in these other cells, what it might bind to, or what is its true subcellular distribution. Obscurin knockout and knockdown studies in mice show myocytes with poorly organized M and A bands of the sarcomere, poor sarcomere organization, altered muscle development, and generalized muscle weakening<sup>13,16</sup>. Obscurin is also a highly modular giant muscle protein and contains more than 60 Ig-like and FnIII-like domains, kinase signaling regions, calmodulin binding sites, and a RhoGEF signalling domain, among others. These domains and binding targets are highlighted in Figure 5.



**Figure 5.** Representation of the domains and target binding proteins of obscurin. Ig58/59 circled in red above, is the titin binding site on Ig58<sup>2</sup>. A mutation in Ig58 (Arg4344Gln, here referred to as Arg8Gln) is the only known mutation in obscurin that is directly related to human disease<sup>3</sup>.

Most obscurin targets bind to discrete domains of obscurin, and are either other proteins involved in the structural organization of the sarcomere or are signaling proteins, which are freely diffuse<sup>13,17,18</sup>. The protein ankyrin, for example, aids in the functionality of the sarcomere through its interaction with obscurin. This interaction, along with titin, is thought to connect the contractile apparatus to the sarcoplasmic reticulum<sup>13,17,19</sup>.

The Ig58/59 tandem domains of obscurin bind to the ZIg9/10 tandem domains of titin, most likely at the Z-line<sup>12,20</sup>. Young, *et al.* proved that both obscurin and both titin domains must be present in order for binding to occur. This was demonstrated through Y2H and binding assay experiments and also suggested through isothermal calorimetric (ITC) experiments, although this group does not present the ITC data<sup>12</sup>. A population genetics study showed that a patient populace with hypertrophic cardiomyopathy (HCM)

also had a genetic variation that coded for a mutant variant of obscurin Ig58. This single amino acid mutation of Arg4333Gln is thought to be enough to disrupt obscurin-titin binding of these four domains completely, suggesting that this mutation may be responsible for the disease at which point we can only speculate over the reasons why<sup>10</sup>. HCM is characterized by a swelling or hypertrophy of the left ventricle due to enlarged myofibrils. This swelling obstructs the appropriate flow of blood in that the left ventricle does not achieve optimal preload volume and thus fails to deliver an adequate stroke volume of oxygenated blood to the rest of the body<sup>21</sup>. This disease manifests itself in a variety of symptoms including, but not limited to, arrhythmias and premature death<sup>20</sup>. HCM is also the most common cause of sudden cardiac death in adolescents<sup>22</sup>. Therefore, in order better understand the specific binding interactions between Ig58/59 and ZIg9/10 and the role of this connection in the function of the cytoskeleton, we used both high and low resolution techniques to probe the structure of these domains, particularly Ig58. The high and low-resolution techniques used are X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and small angle neutron scattering (SAXS). The information gathered can afford us information that has implications to pathologies like HCM.

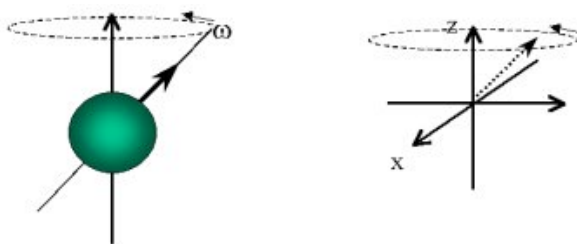
The two common biophysical techniques used to solve the high-resolution structure of a protein or large molecule are nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography.

Like many other types of molecules, such as salts, proteins can be prompted to form crystals. The most commonly used method of protein crystallization is the hanging drop technique. In this technique, protein crystals form as vapor is displaced in a hanging

droplet of solution. Individual crystals can then be shot by an X-ray source with diffracted rays collected in a dot pattern, rotating the mounted crystal as the diffraction is collected. The position and intensity of the dots, within this pattern, contains information about the space group, and with the aid of mathematical interpretations, the electron density within the space group. In this type of data collection, phase information is lost. The loss of the phase is a fundamental limitation, which is ultimately related to the nature of the measurement and is due to a quantum mechanical phenomenon related to the wave particle duality of matter<sup>23,24</sup>. Techniques like molecular replacement, where the phases of a known structure are applied to the experimental data, can in an iterative technique, be used to obtain the phases of all the reflections. Once the phases are known Fourier transformation, sometimes called “the frequency domain representation of the original signal,” can be used to back-calculate the electron density of the protein within the crystal<sup>23,24</sup>. After obtaining three-dimensional data the protein molecule within the crystalline lattice can be fitted into the resulting electron density.

While obtaining a high-resolution structure through X-ray crystallography requires a solid protein crystal, NMR spectroscopy is performed on protein in solution. Nuclear magnetic resonance spectroscopy relies on the magnetic properties of isotopic atoms, which can absorb and re-emit electromagnetic radiation applied by a strong magnetic field<sup>25,26</sup>. The nuclei of the isotopic atoms are sent into discrete spin states by the applied magnetic field (Figure 6).





**Figure 6.** Green sphere, representing magnetic nucleus can be sent into a discrete spin state with angular momentum  $\omega$  with the application of a strong magnetic field in the direction of the Z-axis<sup>27</sup>.

These nuclear spin states have discrete angular momentums, which are manifested as specific peaks in an NMR spectrum. Protein NMR experiments can assign individual peaks in an NMR spectrum to specific atoms within the protein, thus allowing sequence-specific assignments of the atoms within the protein<sup>25,26</sup>. Through-space NMR experiments can then be conducted to measure how close the nuclei are to each other. The compilation of all of the data collected from these experiments gives rise to an atomic-scale, high-resolution solution model. Similar to NMR spectroscopy, low-resolution SAXS is performed on a protein in solution.

Low resolution SAXS data is collected from a special instrument that elastically scatters X-rays into the sample at low angles, a technique called small angle X-ray scattering (SAXS). The low angle range gives information about the shape and size of macromolecules, characteristic distances of partially order materials, and other important information<sup>28</sup>. Data can then analyzed using the program SASSIE, which analyzes data collected from fundamentally disordered biological molecules and can predict basic molecular parameters using experimental scattering restraints.

## Experimental

### *Purification*

The GenBank protein database resource was used to identify the DNA sequence encoding human obscurin-a, accession number CAC44768.1. This sequence was then used to identify residues 4336-4428 of human obscurin-a as the predicted domain sequence of Ig2 via homology modeling with the program Phyre. This sequence was in accordance with the sequence determined by Kontrogianni-Konstantopoulos, et al. The codon optimized synthetic synthesis of the gene was ordered and obtained, with five additional residues added to the ends of the sequence in an attempt to ensure that the complete domain would be obtained. Extra residues were added due to a small amount of uncertainty in the exact size of the domain. This product was subsequently digested and ligated into the NdeI and XhoI restriction sites of the pET24a vector (Novagen). BL21(DE3) *E. coli* containing the Ig2 vector were grown in 5 L of  $^{15}\text{N}$  or  $^{15}\text{N}$ ,  $^{13}\text{C}$  MOPS minimal media containing 30 mg/L kanamycin<sup>29</sup>.  $^{15}\text{NH}_4$  and  $^{13}\text{C}$  glucose was obtained from Cambridge Isotopes, Andover, MA. Cells were grown to an OD<sub>600</sub> of 0.6-0.8 and expression was induced with 1 mM IPTG for 4 h. Cells were then pelleted by centrifugation and stored at  $-80^\circ\text{C}$ . For purification, cells were thawed and resuspended in 25 mL of 50 mM Sodium Phosphate (pH 8.0) with 300 mM NaCl and 10 mM imidazole. Cells were lysed by sonication and cellular debris was removed by centrifugation at 15,000 rpm. Purification of Ig58 was performed using a Ni-NTA nickel resin packed gravity column equilibrated with 50 mL of 50 mM Sodium Phosphate (pH 8.0) with 300 mM NaCl and 100 mM imidazole. Using a 10 mM to 500 mM gradient of imidazole, Ig58 eluted into one fraction after the 500 mM imidazole elution step.

Fractions were analyzed by 10% Tris-Tricine SDS-PAGE gels stained with Coomassie Blue. The elution fraction was then concentrated down to 500  $\mu$ l using an 5,000 MWCO Spin-X-UF Corning concentrator, and run over a G75 resin packed size-exclusion, gravity column, equilibrated with two column volumes of 20 mM Tris pH 7.5 and 50 mM NaCl, over a period of six hours. Protein eluted at fraction 30, consistent with other Ig domains. The protein was then concentrated to >1.5 mM as quantified by absorbance ( $\epsilon_{280} = 11,620 \text{ M}^{-1} \text{ cm}^{-1}$ ) using a 5,000 MWCO Spin-X-UF Corning concentrator.

**X-ray:**

Ig58 protein was concentrated to 10.5-11.3 mg/mL and crystal screening was performed using solutions from Hampton Research. Significant crystal growth was noticed for all of the following screens: PEG/Ion (HR2-126), PEG/Ion2 (HR2-098), PEGRx 1 (HR2-082), PEGRx 2(HR2-084), Index 1-48(HR2-144), and Index 49-96(HR2-144). The most prominent crystal growth appeared in solutions containing between 19-23% PEG 3350 and various salts. After crystal growth optimization, Ig58 crystals were grown in 0.3-0.5 M NaCl, 19-23% PEG 3350, for eight days via hanging-drop vapor diffusion. Crystals were harvested using MicroMeshes (MiTeGen, Ithica, NY) and cryo-protected in 50% glycerol (v/v) prior to freezing in LN<sub>2</sub>. An initial 2.7 Å data set was collected using an Oxford Diffraction Gemini diffractometer and a rotating copper anode. Indexing and processing of these data were performed with iMosflm with a sigma cutoff of 2<sup>30,31</sup>. Data was subsequently scaled and merged with the program Aimless<sup>32</sup> within the CCP4 program suite<sup>33,34</sup>. Detailed instructions on how to operate these programs are included in the appendix. Phasing of the Ig58 reflections data was solved in PHENIX by molecular replacement using PHASER-MR crystallographic software with PDB # 2YZ8

as the search model<sup>35,36</sup>. Autobuild was used to generate a starting structure that then underwent 76 iterative rounds of manual model building and refinement, performed with COOT and PHENIX Refine<sup>35,38</sup>. A subsequent higher resolution data set was collected at the Structural Biology Center beamline 19-ID-D at the Advanced Photon Source, Argonne National Laboratory to 2.411 Å. The final data set used to solve the structure was collected from two different crystals. Indexing and processing of high-resolution data were performed with HKL2000 by integration of single crystal diffraction data from area detector with a sigma cutoff of 2<sup>30</sup>. Data was subsequently scaled, merged, and phased as described above, with the previously determined lower resolution structure used for molecular replacement, with the lower resolution structure undergoing its own, completely different set of refinement. The structure then underwent 160 iterative rounds of manual model building, as described above. Atomic coordinates and reflections of the X-ray structure have been deposited in the RCSB Data Bank under accession number 4RSV. Statistics are presented in Table 1.

#### **NMR:**

NMR experiments, completed by Logan Meyer, were completed in a manner similar to those in Rudloff and Wright 2015<sup>39</sup>. The protein was purified using a modified version of the purification scheme described previously. The final 20 structures were selected (from 200) based on lowest energy and were of high quality based on the statistical criteria listed in Table 2. Chemical shifts for Ig58 have been deposited in BMRB under accession number 25308 and the 20 best structures in the PDB under accession number 2MWC.

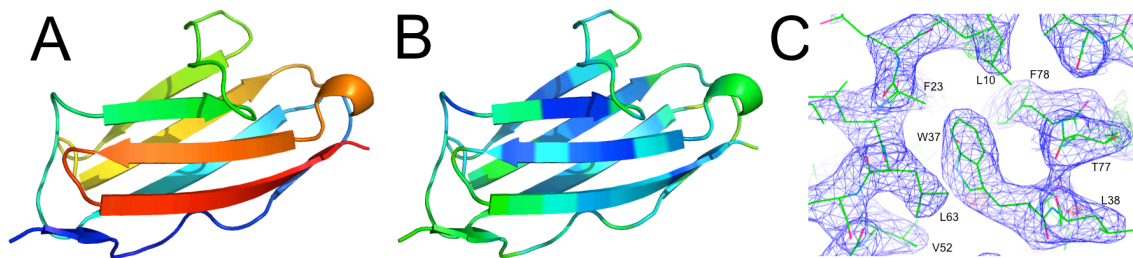
**Molecular Dynamics:**

All MD simulations were performed using the YASARA 12.4.1 software package, the AMBER03 force field, and explicit solvent (with 150 mM NaCl) in a size 65 x 65 x 65 Å<sup>3</sup> box at 25 °C<sup>40</sup>.

## Results

This project was originally aimed at studying data from two domains of obscurin: Ig58/59 and Ig1/2. Most of this manuscript however, deals with data on Ig58/59. While initial experiments of Ig1/2 were promising, another lab solved the crystal structure of titin M10 in complex with Ig1 of obscurin<sup>41</sup>, and low resolution SAXS data were not interpretable due to polydispersity within the sample. For SAXS analysis, it is critical that the experimental sample is completely monodisperse, meaning that the protein in solution is entirely pure and is in no way aggregated. The lack of monodispersity of our sample is most likely due to the fact that the purified solution sat for too long before analysis, giving the protein time to break down and aggregate.

To begin probing the molecular mechanism of why the Ig58 and Ig59 domains of obscurin are important in myocyte organization and how the mutant phenotype of Arg→Gln at the eighth residue of Ig58 is implicated in the disease HCM, we solved the high-resolution structure of Ig58 using both X-ray crystallography and NMR spectroscopy. The work here deals specifically with the high-resolution X-ray structure, solved in conjunction with Kelly DuPont, and its correlation to the NMR structure previously solved by Logan Meyer. The X-ray structure was solved by molecular replacement to 2.41 Å and is shown below in Figure 7A.



**Figure 7.** **A)** X-ray structure of Ig58 solved to 2.41 Å resolution with an  $R_{\text{free}}$  value of 0.28. **B)** X-ray structure with ‘hot’ colors representing areas of high B-factors and ‘cool’ colors areas with low B-factors. **C)** Manual refinement of Ig58 at 1.5 sigma in COOT. Final structure determined from merging two data sets from two crystals.

X-ray data showed that Ig58 folds into a typical Ig-like fold (Figure 7A). This fold consists of six core beta strands and one peripheral beta strand arranged in a beta-sandwich like fold, characterized by two opposing antiparallel beta sheets<sup>8</sup>. In crystallography, the degree of relative vibrational motion of specific regions of a protein are represented by values termed, B-factors, measured in  $\text{Å}^2$ . Atoms that are well ordered within the structure and experience little vibrational motion have low B-factors, and atoms belonging to flexible regions have high B-factors<sup>23,24</sup> (Figure 7B). Upon structural refinement of the X-ray structure of Ig58, areas with high B-factors were often the most difficult to fit to the data. Due to their lack of non-covalent interactions, the loops display the most elevated B-factors and appear in warmer colors in Figure 7B. The total B-factor value, giving a readout of the overall flexibility of the protein, as well as many other statistical parameters, are listed below in Table 1. The use of both NMR spectroscopy and X-ray crystallography as structural techniques can provide complementary structural

information, as the true structure of the protein *in vivo* may be slightly misrepresented by one of these two techniques alone. This misrepresentation can occur through crystal packing artifacts in X-ray structures, which can affect the positions of the disordered regions such as loops tremendously. Problems may also arise due to areas of high motion in NMR structures making it hard to resolve the positions of atoms that are close together. It is therefore complementary to have both of these high-resolution structures, for the purpose of structural validation. Recognizing these facts, here we present both the NMR and the X-ray structure of the Ig58 domain of obscurin.

**Table 1:** Statistics of the X-ray structure of Ig58, at 2.41 Å resolution.

Resolution range (Å)	37.64 - 2.411 (2.496 - 2.411)
Space group	P 61 2 2
Unit cell	44.482 44.482 177 90 90 120
Total reflections	11201 (1170)
Unique reflections	4329 (347)
Multiplicity (°)	3.2 (3.4)
Completeness (%)	95.84 (80.14)
Mean I/sigma(I)	29.47 (3.11)
Wilson B-factor (Å <sup>2</sup> )	66.54
R-merge	0.03344 (0.3173)
R-meas	0.04104
CC(1/2)	0.995 (0.931)
R-work	0.2463 (0.3529)
R-free	0.2649 (0.3225)
Protein residues	90
RMS(bonds)	0.020
RMS(angles)	2.08
Ramachandran favored (%)	84.6
Ramachandran allowed (%)	13.4
Ramachandran outliers (%)	0.0
Clashscore <sup>molprobity</sup>	37.75
Average B-factor (Å <sup>2</sup> )	81.40

\*Values in brackets refer to highest resolution shell (2.496 - 2.411)<sup>34</sup>

The Ig58 NMR ensemble generated by Meyer, L., includes greater than 14 restraints per residue, with no distance constraint violations greater than 0.4 Å and a Q-



factor of 0.27 (Table 2). The NMR-based structures agree well with each other, with a backbone RMSD of 0.471 Å for the 20 best structures (Figure 8A).

**Table 2:** NMR-derived restraints and statistics of 20 NMR structures.

	<20>	best
<b>rmsd from distance constraints (Å)<sup>1</sup></b>		
total (1270)	0.027 ± 0.002	0.026
intraresidue (276)	0.005 ± 0.004	0.005
sequential ( i - j  = 1) (345)	0.015 ± 0.004	0.012
medium range (1 <  i - j  ≤ 1) (109)	0.048 ± 0.007	0.045
long range ( i - j  = 1) (472)	0.029 ± 0.003	0.029
hydrogen bonds (68)	0.049 ± 0.008	0.053
<b>rmsd from exptl dihedral constraints (°)</b>		
Φ, Ψ (169)	0.463 ± 0.142	0.690
<b>rmsd from dipolar coupling restraints (Hz)</b>		
D <sub>NH</sub> (82)	0.74 ± 0.03	0.74
<b>rmsd from exptl <sup>13</sup>C chemical shifts</b>		
<sup>13</sup> Ca (ppm)	1.34 ± 0.23	1.53
<sup>13</sup> Cb (ppm)	1.78 ± 0.19	1.73
<b>rmsd from idealized geometry</b>		
bonds (Å)	0.004 ± 0.001	0.004
angles (°)	0.691 ± 0.020	0.668
impropers (°)	0.618 ± 0.026	0.627
<b>Lennard-Jones potential energy (kcal/mol)<sup>2</sup></b>	-350 ± 12	-340
<b>Q-value<sup>3</sup></b>	0.27 ± 0.04	0.24
<b>% most favorable region in the Ramachandran plot<sup>4</sup></b>	70.6 ± 2.9	72.2
<b>all backbone atoms (3-92)</b>	0.471 ± 0.075	0.340
<b>all heavy atoms (3-92)</b>	1.088 ± 0.081	1.070

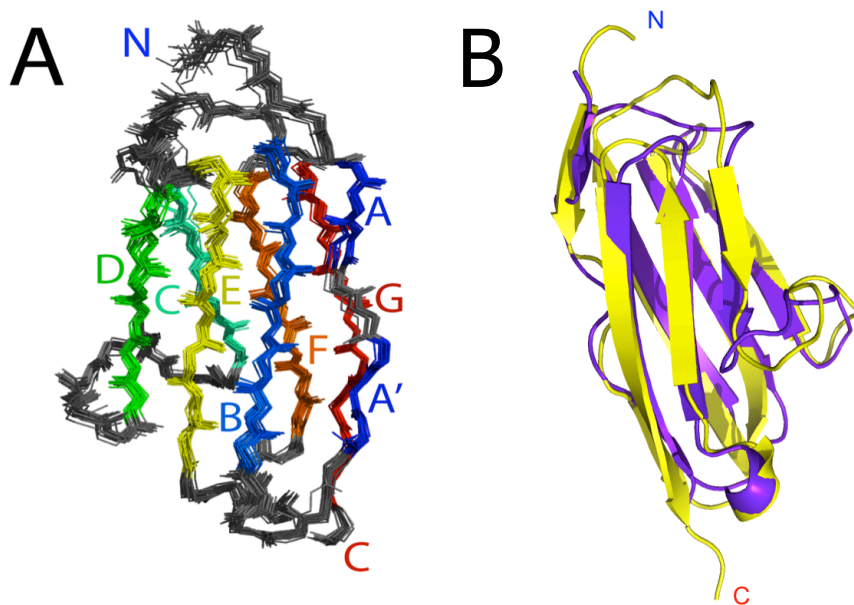
<sup>1</sup> None of the 20 structures has a distance violation > 0.4 Å or a dihedral angle violation of > 5°. The

<sup>2</sup> Lennard-Jones van der Waals energies were calculated using CHARMM parameters and were not used in any stage of the structure determination

<sup>3</sup> Q-values were determined by randomly removing 10% of all RDC values. To ensure accuracy, an ensemble of structures with a second randomly removed subset of RDCs was also run. The Q-value of this second set was similar to the first.

<sup>4</sup> PROCHECK was utilized to generate the Ramachandran plot

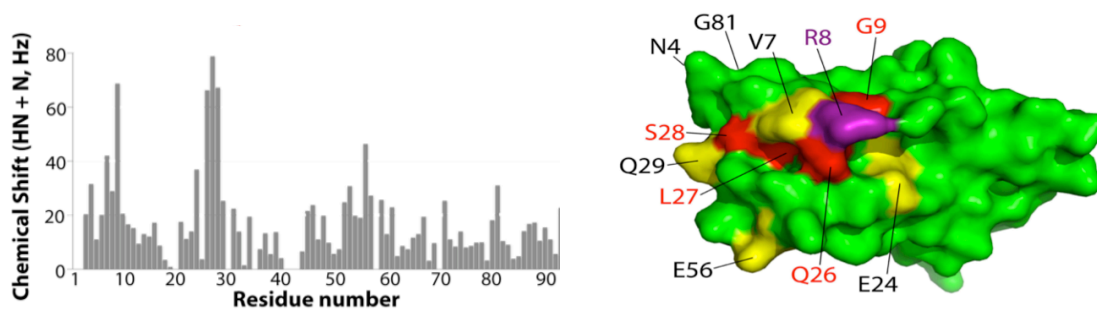
The NMR and X-ray structures are similar (heavy atom RMSD = 1.4 Å), with regions of the greatest difference in the B-C loop and the D-E loop (Figure 8B). These are long loops, and display slightly elevated B-factors in the X-ray structure and fewer-than-normal distance restraints in the solution ensemble. Thus, these slight structural differences are likely a function of local protein dynamics. The analysis of both of these high-resolution structures makes it possible to understand the molecular mechanism of how mutation to Ig58 leads to disease.



**Figure 8.** **A)** 20 best NMR structures from the fully assigned Ig58 construct at 20 mM Tris, 50 mM NaCl, 350 mM azide, 10% D<sub>2</sub>O, and pH 7.5 at 25° C. **B)** Overlay of X-Ray (purple) and NMR (gold) structures.

The 8<sup>th</sup> residue of Ig58 corresponds to Arg4344 of full-length human obscurin-a, and an Arg→Gln mutation at this site has been linked to hypertrophic cardiomyopathy<sup>10</sup>, as previously stated. An Ig58 construct containing the Arg→Gln mutation resulted in

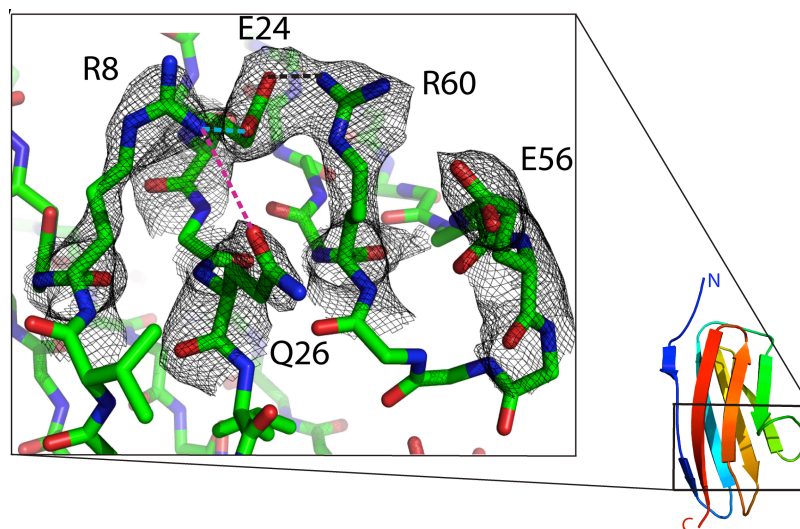
only minor changes in the HSQC spectrum, indicating that this mutant version remains similarly folded (Figure 9). The chemical shift changes between the wild type and mutant Ig58 were plotted based on standard deviation and are unique for each residue, represented in Figure 8. Projecting the resulting chemical shift changes onto the high-resolution NMR structure shows that adjacent residues to the mutation display significant chemical shift perturbations (Figure 9).



**Figure 9.** Ig58 construct containing the Arg→Gln mutation resulting in minor changes in the HSQC spectrum. Projecting the resulting chemical shift changes onto the high-resolution structure shows that adjacent residues to the mutation (purple) display significant (>2x st. dev. for yellow and >3x st. dev. for red) chemical shift perturbations.

Several residues on neighboring  $\beta$ -strands are also in different molecular environments (Figure 9). These include residue Glu24, Gln26, Leu27, Ser28, Gln29, Glu56, and Gly81 (Ig58 numbering). Close inspection of the Ig58 electron density map reveals a series of charge-charge interactions between residues Gln26-Arg8, Arg8-Glu24, Glu24-Arg60, and conceivably between Arg60-Glu56 (Figure 10). The electron density, represented by

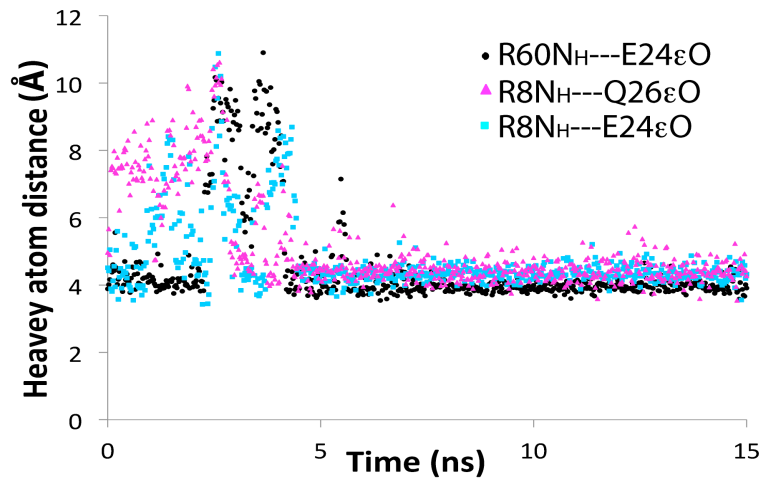
the black mesh in Figure 10, is actual crystallographic data. The location of the mesh indicates that, within the crystal, the residues at these positions are in close proximity. At many points the mesh is continuous between residues, and is evidence of the strength of the interaction between charge-charge pairs.



**Figure 10.** Ig58 electron density map revealing a series of charge-charge interactions between residues Gln26-Arg8 (pink), Arg8-Glu24 (blue), Glu24-Arg60 (black).

Such charge-charge pairs could explain the long-range influence of the Arg→Gln mutation on the chemical shifts. Molecular dynamics simulations of the Ig58 structure, run by Berndsen, C., reveal that Arg8-Glu24, Arg8-Gln26, and Glu24-Arg60 all can exist in stable (>10 ns) interactions (Figure 11), and that these electrostatic bonds stabilize and shape a significant part of the Ig58 surface. Each trace of Figure 11 corresponds to one of the charge-charge interactions between two residues of Figure 10. The large, irregular peaks at the beginning, from 0 to 5 nanoseconds, are indicative of an equilibration period.

The flatness of the lines after that point indicates that the distances between these residues does not change and that these interactions are stable.



**Figure 11.** Simulation of Ig58 in explicit solvent at 310 K and 100 mM NaCl, run by Berndsen, C. Each trace represents the distances between two adjacent residues involved in a charge-charge interaction with colors corresponding to the distance measurements between residues of Figure 9.

These three non-covalent interactions display long-lived charge-charge conformations, suggesting that all three contribute to the surface topology of this Ig58 face (Figure 10). Thus, while the Arg8Gln mutation may affect target binding through direct ablation of a protein-target coulombic interaction, it is equally feasible that this mutation disrupts target binding more generally by altering the overall target binding site topology or dynamics.

## Discussion

The significance of solving the structure of Ig58 lies in the fact that this particular domain is the only region of obscurin that has both a known binding partner<sup>12</sup> and a pathological mutation<sup>10</sup>. Therefore, the structural effects of the mutation adopt a more significant context when linked with the domain's function, in this case, target protein binding. Analysis of the arrangement of residues provides strong insights into the effects of the mutation on the local and global structure, and consequently function, since structure dictates function. After obtaining information about the domain's structure and function, we can begin to understand how the mutation in Ig58 effects myocyte function and how this is related to disease.

While more work must be done in order to make any claims about how this mutation is linked to HCM, it does seem possible that this mutation is responsible for the disease due to interference between obscurin-titin binding. Without obscurin and titin linked together as they would be in a wild-type cell, there could be moderate to severe cytoskeletal dysfunction<sup>42</sup>. Abrogation of this interaction seems to weaken the myofibril cytoskeleton. The Arg8Gln mutation to Ig58 does not appear to be a major weakening event, as the cytoskeleton remains in tact and only mildly dysfunctional. But the model is that like a hole in a wicker basket, over time it causes the rest of the cytoskeletal web to have to overcompensate, leading to thicker cardiomyocytes, and ultimately to HCM.

Mutations to obscurin binding partners have also been linked to disease such as multiple mutations to the M10 domain of titin, which normally binds to Ig1 of obscurin<sup>39</sup>. These mutations have been linked to limb-girdle and tibial muscular dystrophy<sup>41</sup>, although the effects of the mutations on the protein structure and function have not been

fully characterized. It may be the case that a perturbed binding interface, similar to that of the Arg→Gln obscurin Ig58 mutant, is also what is responsible for the diseases linked to other cytoskeletal proteins.

It is perhaps surprising that a one-residue mutation on the surface of a single Ig-like domain could lead to disease, especially since the overall conformation of the domain is largely maintained in spite of the mutation. When the Arg8Gln mutation is introduced into Ig58, the domain maintains its fold, and in fact is only slightly distorted from the original structure based on chemical shift changes. Structural analysis suggests that this mutation disrupts a moderately large network of electrostatic interactions on the Ig58 surface. If charge-charge network is the titin binding site, as is postulated, then this mutation not only perturbs a potential inter-protein charge-charge interaction, but also changes the side-chain orientation of many nearby residues. We also note that while Ig58 was relatively easy to crystalize, the Arg8Gln mutation has still not been crystallized, despite repeated efforts. The Arg8Gln mutation is not predicted to directly perturb any crystal packing interactions in the original Ig58 space group, and thus it is tempting to speculate that the Arg8Gln construct is more difficult to crystallize due to an increase in surface entropy<sup>43</sup>. This parameter could then adversely affect target binding, leading to disease.

## **Conclusions**

Significant structural information complementing the NMR work of Meyer, L. was obtained through X-ray crystallography and molecular dynamics. There is strong evidence that the Arg8Gln mutation disrupts a tight, multi-residue charge interface on the surface of Ig58, and also has some non-local effects on structure. The mutation of Arg8Gln at residue number 8 of this structure may be causing cytoskeletal dysfunction and may ultimately be responsible for HCM due to ablated binding between obscurin and titin however, more work must be done in order to definitively prove that there is a direct correlation between the mutation and the disease.



## References

1. Squire, J. M. Architecture and function in the muscle sarcomere. *Curr. Opin. Struct. Biol.* **7**, 247–257 (1997).
2. Hooper, S. L., Hobbs, K. H. & Thuma, J. B. Invertebrate muscles: Thin and thick filament structure; molecular basis of contraction and its regulation, catch and asynchronous muscle. *Prog. Neurobiol.* **86**, 72–127 (2008).
3. Huxley, A. F. and Niedergerke, R. Structural changes in muscle during contraction. *Nature Publishing Group.* **173**, 971-973 (1954).
4. University, A. Biology 5600 > Mendonca > Flashcards > Sarcomere | StudyBlueStudyBlue (2015). At <<https://www.studyblue.com/notes/n/sarcomere/deck/5439488>>
5. Young, P., Ferguson, C., Bañuelos, S. & Gautel, M. Molecular structure of the sarcomeric Z-disk: Two types of titin interactions lead to an asymmetrical sorting of  $\alpha$ -actinin. *EMBO J.* **17**, 1614–1624 (1998).
6. Granzier, H. L. & Labeit, S. The Giant Protein Titin: A Major Player in Myocardial Mechanics, Signaling, and Disease. *Circ. Res.* **94**, 284–295 (2004).
7. Galler, S. Forgotten research from 19th century: science should not follow fashion. *J. Muscle Res. Cell Motil.* **36**, 5–9 (2014).
8. Meyer, L. C. & Wright, N. T. Structure of giant muscle proteins. *Front. Physiol.* **4**, 1–12 (2013).
9. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis *Nucleic Acids Res.* 31:3784-3788(2003).
10. Arimura, T. *et al.* Structural analysis of obscurin gene in hypertrophic cardiomyopathy. *Biochem. Biophys. Res. Commun.* **362**, 281–287 (2007).
11. Wang, K., McClure, J. & Tu, a. Titin: major myofibrillar components of striated muscle. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3698–3702 (1979).
12. Young, P., Ehler, E. & Gautel, M. Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly. *J. Cell Biol.* **154**, 123–136 (2001).

13. Kontrogianni-Konstantopoulos, A. *et al.* Obscurin modulates the assembly and organization of sarcomeres and the sarcoplasmic reticulum. *FASEB J.* **20**, 2102–2111 (2006).
14. Kontrogianni-Konstantopoulos, A., Ackermann, M. a, Bowman, A. L., Yap, S. V & Bloch, R. J. Muscle giants: molecular scaffolds in sarcomerogenesis. *Physiol. Rev.* **89**, 1217–1267 (2009).
15. Ackermann, M. a., Shriver, M., Perry, N. a., Hu, L. Y. R. & Kontrogianni-Konstantopoulos, A. Obscurins: Goliaths and Davids take over non-muscle tissues. *PLoS One* **9**, (2014).
16. Borisov, A. B. *et al.* Essential role of obscurin in cardiac myofibrillogenesis and hypertrophic response: Evidence from small interfering RNA-mediated gene silencing. *Histochem. Cell Biol.* **125**, 227–238 (2006).
17. Yap, S. V & Bloch, R. J. NIH Public Access. *Mol. Biol.* **89**, 1217–1267 (2011).
18. Otey, C. a., Dixon, R., Stack, C. & Goicoechea, S. M. Cytoplasmic Ig-domain proteins: Cytoskeletal regulators with a role in human disease. *Cell Motil. Cytoskeleton* **66**, 618–634 (2009).
19. Kimata, Y. *et al.* Genetic Evidence for a Role of BiP / Kar2 That Regulates Ire1 in Response to Accumulation of Unfolded Proteins. *Mol. Biol. Cell* **14**, 2559–2569 (2003).
20. Konno, T., Chang, S., Seidman, J. G. & Seidman, C. E. Genetics of hypertrophic cardiomyopathy. *Curr. Opin. Cardiol.* **25**, 205–209 (2010).
21. Ball, W. *et al.* Long-term survival in patients with resting obstructive hypertrophic cardiomyopathy: Comparison of conservative versus invasive treatment. *J. Am. Coll. Cardiol.* **58**, 2313–2321 (2011).
22. Maron, B. J. *et al.* Contemporary definitions and classification of the cardiomyopathies: An American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functio. *Circulation* **113**, 1807–1816 (2006).
23. Kleywegt, G. J. Validation of protein crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 249–265 (2000).
24. Laskowski, R. a., MacArthur, M. W. & Thornton, J. M. Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.* **8**, 631–639 (1998).

25. Bax, A. & Grishaev, A. Weak alignment NMR: A hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.* **15**, 563–570 (2005).
26. Bax, A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* **12**, 1–16 (2003).
27. Ray Fort, K. NMR Lectures. Chemistry.umeche.maine.edu (2015). at <http://chemistry.umeche.maine.edu/CHY431/NMR/NMR-3.html>
28. Shih, P.-M. *et al.* Random Coil Behavior of Chemically Denatured Topologically Knotted Proteins Revealed by Small Angle X-Ray Scattering. *J. Phys. Chem. B* (2015).
29. Wright, N. T. *et al.* The three-dimensional solution structure of Ca<sup>2+</sup>-bound S100A1 as determined by NMR spectroscopy. *J. Mol. Biol.* **353**, 410–426 (2005).
30. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 307–326 (1997).
31. Leslie, A.G.W. and Powell, H.R. Evolving Methods for Macromolecular Crystallography, **245**, 41-51(2007).
32. Evans, P.R. Scaling and assessment of data quality. *Acta Cryst.* **D62**, 72-82 (2006).
33. Winn, M. D. *et al.* *Acta. Cryst.* **D67**, 235-242 (2011)
34. Peters, G. M. *et al.* truncate Z. Dauter, *Acta Cryst.* **D62**, 867 (2006)
35. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J, Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **D66**, 213-221 (2010).
36. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L.C., and Read, R.J. Phaser crystallographic software. *Appl. Cryst.* **40**, 658-674 (2007).
37. Iwasaki, L. R. *et al.* A pilot study of ambulatory masticatory muscle activities in temporomandibular joint disorders diagnostic groups. *Orthod. Craniofac. Res.* **18**, 146–155 (2015).
38. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. Features and Development of Coot. *Acta Crystallographica Section D - Biological Crystallography.* **66**, 486-501 (2010).

39. Rudloff, M. and Wright, N. Biophysical characterization of naturally occurring titin M10 mutations. *Protein Sci.* (2015)
40. Nieh, M.-P., Glinka, C. J., Krueger, S., Prosser, R. S. & Katsaras, J. SANS study on the effect of lanthanide ions and charged lipids on the morphology of phospholipid mixtures. Small-angle neutron scattering. *Biophys. J.* **82**, 2487–2498 (2002).
41. Pernigo, S. *et al.* The Crystal Structure of the Human Titin:Obscurin Complex Reveals a Conserved yet Specific Muscle M-Band Zipper Module. *J. Mol. Biol.* **427**, 718–736 (2015).
42. Pernigo, S. *et al.* Structural insight into M-band assembly and mechanics from the titin-obscurin-like-1 complex. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2908–2913 (2010).
43. Loll, P. J., Xu, P., Schmidt, J. T. & Melideo, S. L. Enhancing ubiquitin crystallization through surface-entropy reduction. *Acta Crystallogr. Sect. F* **70**, 1434–1442 (2014).

## Appendix

### ***Solving a Crystal Structure***

#### ***HKL Tutorial (Indexing)***

*Written with Kelly DuPont*

### ***HKL2000, Phenix, and Coot Tutorial***

#### **HKL Steps:**

##### **Step 0: Open HKL2000**

Open a terminal window, go all the way up in directories. Then go through 'usr' to 'local' to 'bin.'

##### **Step 1: Load Data**

Once HKL2000 has opened fully, look under the space titled 'directory tree' and click on 'berndsce,' 'Desktop,' 'APS4102014,' then on the data folder you will be working on (eg. 'JMU113' or 'JMU211').

Designate your input folder, which is the data folder you just chose. This is where HKL will scan for data files. Create or assign where the output files will go and click 'load data sets.' To create a new folder, click 'Create Directory' and type the name.

Click 'Load Data Sets.' You will choose the file with 'data' in the name, and NOT with 'scan' in the name. Once this has opened, record your wavelength and frame width. Then, click on the tab entitled 'Index.'

##### **Step 2: Indexing**

Click on 'fit basic' and make sure that there are exactly 6 red boxes. Set resolution maximum to 'corner' if you think there is data that was not included and to 'edge' if you only want to go to the user set boundaries of the original data. For Synchrotron data, 'corner' is more appropriate. The minimum resolution should be 50.00. Click on 'Peak Sear,' then, in the new window that appears, click on 'Peak Search,' and minimize the random peaks. This will usually involve lessening your peaks until they fall within the range 300-800. Record the number of peaks. Once this is done, click 'OK' then 'Index.'

##### **Step 3: Choosing a Space Group**

When the window appears for you to choose a space group, choose the group that has the highest symmetry with the lowest percent (usually it is green and closest to the top). Then click 'Apply and Close.'

##### **Step 4: Refining**

Your  $\chi^2$  values should be less than 2. Click 'Refine' and continue to refine until your  $\chi^2$  values level out (are constant). At this time, the 'Integrate' button should be green (ideally). If it is, press it.

If not, click 'Fit All' then 'Refine' again until the  $\chi^2$  values level out. You may also need to change your space group, the values of which will have change due to your refining. Click 'Bravais Lattice' to open up your space group options. Again, choose the one with the most symmetry. If you do change your space group, you will need to refine again. Record the space group at the top of this screen. 'Integrate' should be green now, so press it.

### **Step 5: Integrate**

Your mosaicity value should be less than 1. Record both it and your resolution range. Ideally, the graph on the left should look like one downward slope while the graph on the right should be fairly constant. Record your positional, partiality, X-  $\chi^2$ , Y-  $\chi^2$ , and  $\chi^2$  values.

### **Step 6: Scale**

Click on the 'Scale' tab at the top and make sure your space group matches with the one you previously chose. Make sure that the box next to 'Small Slippage, Imperfect Goniostat' is red. Click 'Fix B' and make sure that 'Write Rejection File' is red also. Then click 'Scale Sets,' which should be green.

Pay close attention to the graphs titled Completeness vs. Resolution and I/Sigma and CC1/2 vs. Resolution. These will aid you in determining the minimum value of your resolution. On the Completeness vs. Resolution graph, move your cursor to where the blue line begins to drop and remember this number. Scroll down to the CC1/2 vs. Resolution graph and move your cursor along the grey limit to where the grey meets the blue. This number should match up fairly decently with the number from before. However, to be more accurate, click on the 'Show Log File' button at the bottom and scroll down on the window that opens to the bottom. Compare the values in the Average I column to those in the Average error column and record the Shell lower limit for the line in which the Average I is approximately 2 times that of the Average error. This will be the minimum value of your resolution. Hit 'Close' on this window when done, set your minimum resolution, and the maximum for 50.

Click 'Delete Rejection File' then 'Scale Sets.'

Record a, b, c and alpha, beta, gamma.

Your output file will have the name output.sca and will be wherever you specified it to go in the 'Data' tab. This is what you will use in CCP4.

### **CCP4 Steps:**

#### **Step 0: Opening CCP4**

From the Finder, go to 'Applications,' then 'ccp4-6.4.0' and 'ccp4.' (If it is not already down on the tabs, which it should be)

#### **Step 1: Running CCP4**

On the right side of the screen there is a program list. Under 'Data Reduction and Analysis,' click 'Symmetry, Scale, and Merge (Aimless).'

In the window that pops up, make sure that 'Option to skip scaling & just merge' is black. Also, make sure the FreeR column has 0.05 in the blank and that 'Copy FreeR from another file' is grey. Change the Input reflection file type to 'Scalepack file' and choose your input/output destinations. SCA #1 is your input. You must input the 'output.sca' file, NOT the 'scalepack.sca' file.

If there are 2 input files, click 'Add File' and then browse to find that second file. This will merge both and output a single .mtz file for you to use.

HKLOUT is where you type/choose the output destination. When choosing your output destination, you will have to name the file. To do this, choose the folder destination, click 'OK' and then add to the blank a /\_\_\_\_.mtz where the \_\_\_\_ is the desired file name. Then click 'Run' and 'Run Now.'

If the process is finished quickly, chances are that you did something wrong. It should at least take a few seconds.

### **Phenix Steps:**

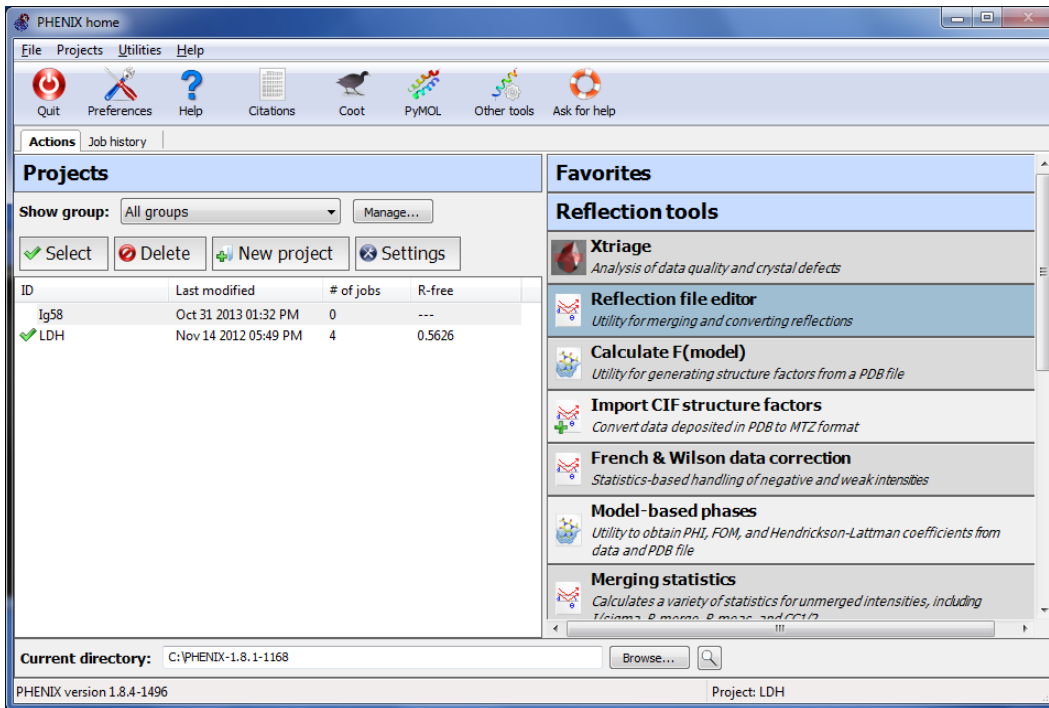
#### **Step 0: Creating a Folder for Data Collect in Phenix**

Create a folder on the phenix home desktop and name it something related to your project. This new folder will be where all of the output files will be directed to from Phenix and Coot.

You can create subfolders for the preliminary data that you will need throughout the process of solving a crystal structure. This may include the sequence file for your protein as well as the .mtz file created from Indexing.

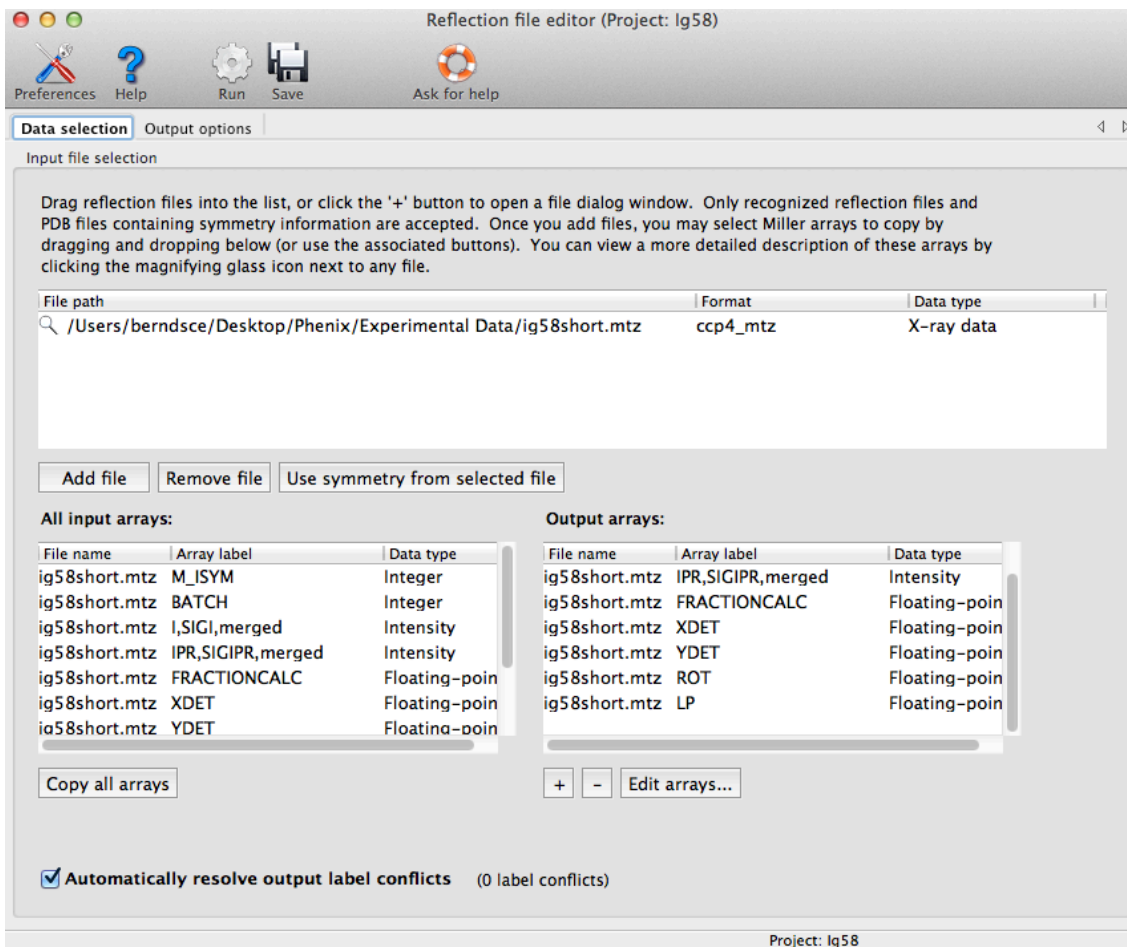
#### **Step 1: Creating a Reflection File**

To create a reflections file, open Phenix and create a new project ID. The following window will open.



Open the 'Reflection Tools' tab on the right hand side of the window. Then select the 'Reflection file editor' and the following window will open.

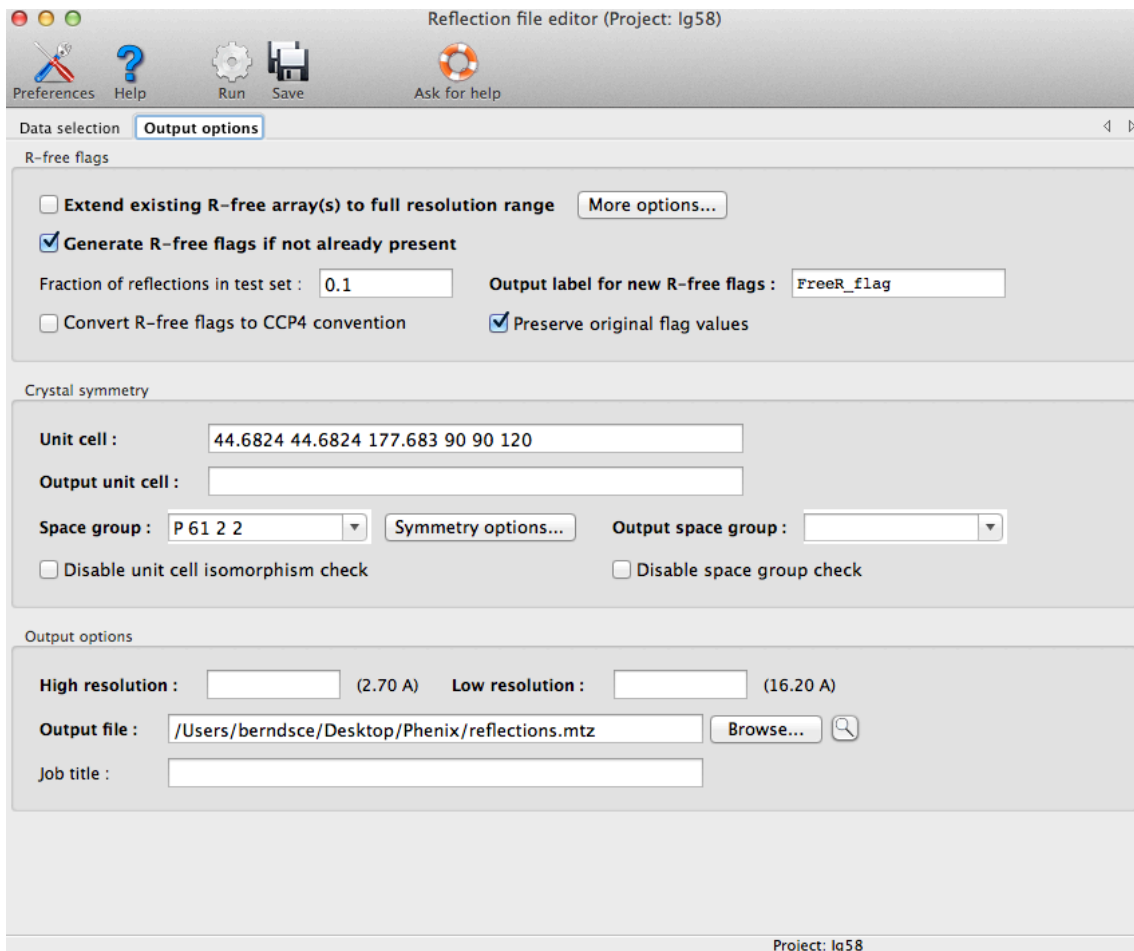




Insert your short.mtz file created from HKL. This .mtz file will be used to create reflections of your protein structure. When you have added the .mtz file, a list of input arrays will pop-up in the section labeled **All input arrays**.

Hit the copy all arrays button below the selection. This command will copy all file to the output arrays. Delete the files with the data type of integer.

Then open the output options tab. The window should look as follows.



Make sure that **Generate R-free flags if not already present** is checked.

Also check that the **Fraction of Reflections in test set** is at 0.1.

Check to make sure the space group is correct.

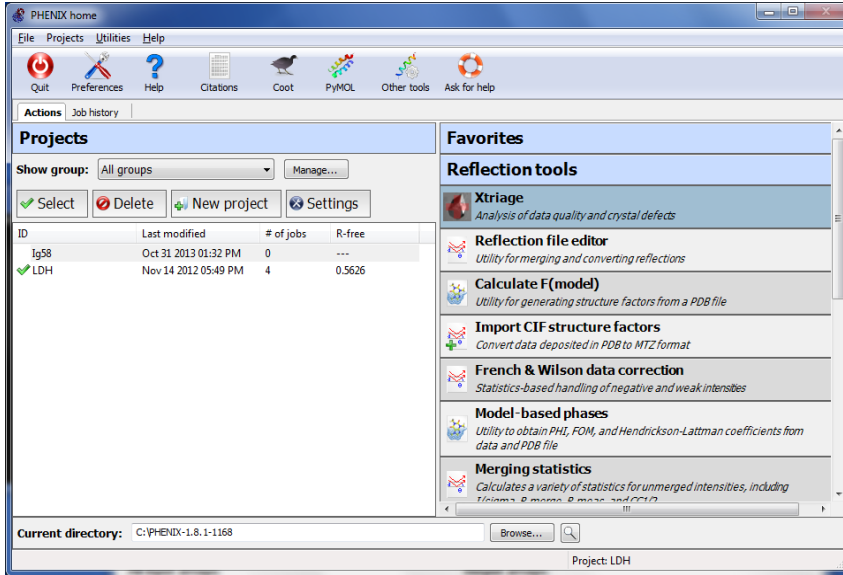
The rest of the output options will fill in automatically

Make sure the output file is directed to the correct location

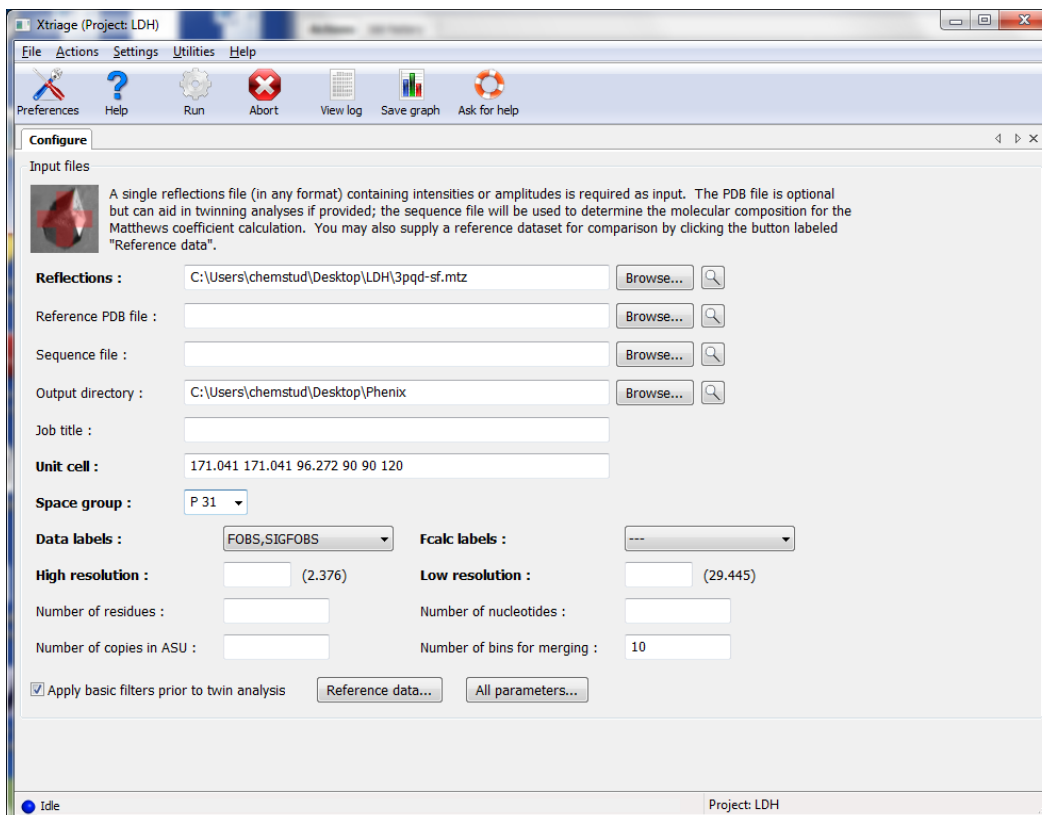
Run the Reflection file editor. After the program is done, a reflections.mtz file will be added to the output location. Confirm the reflections.mtz file is in the correct location before opening the Xtriate tab on the home page of phenix.

## Step 2: Xtrriage 2

Select the Xtrriage option under Reflection tools on the Phenix home screen



Once the xtrriage program is open, find the reflections.mtz that was just generated from the reflection file editor.



Make sure the output directory is going to the correct folder. The purpose of performing the xtriage program is to check the quality of data via checking the space groups, unit cell, and resolution.

The output file is created in a folder called Xtriage\_2. This will be found in the phenix folder that you created on the desktop.

### **Step 3: Models and Templates for Molecular Replacement**

To obtain templates for generating an electron density map use Basic Local Alignment Search Tool (BLAST). Go to the website: [blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi). Once on this site, go to the protein blast found under the heading Basic BLAST.

The protein blast web page would look as follows.

**BLAST®** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastp suite **Standard Protein BLAST**

blastn **blastp** blastx tblastn tblastx

Enter Query Sequence BLASTP programs search protein databases using a protein query. [more...](#)

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

TAKNTVVRGLENVEALEGGGALFECQLSQPEVAHAHTWLLDDEPVRTSENAEVVFFENGLRHLLLLKNL  
RPQDSCRVTFLAGDMVTS AFLTVRGGLEHHHHHH

Query subrange From   
To

Or, upload file  No file chosen

Job Title   
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Organism Optional   Exclude   
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional  Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query Optional

Program Selection

Algorithm  blastp (protein-protein BLAST)  
 PSI-BLAST (Position-Specific Iterated BLAST)  
 PHI-BLAST (Pattern Hit Initiated BLAST)  
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)  
Choose a BLAST algorithm

**BLAST** Search database Protein Data Bank proteins(pdb) using Blastp (protein-protein BLAST)  
 Show results in a new window

[Algorithm parameters](#) Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

On this screen, insert the sequence of the protein of interest into the large empty box at the top of the screen. Next select Protein Data Bank under the **Database** dropdown menu. Then click the BLAST button at the bottom of the screen.

On this preceding screen there will be some statistics at the top that show how well your protein sequence matches those in the .pdb

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastp suite/Formatting Results - 8UNEUM1Z01R

Edit and Resubmit Save Search Strategies Formatting options Download You Tube How to read this page Blast report description

**Protein Sequence (102 letters)**

RID 8UNEUM1Z01R (Expires on 11-22 07:37 am)

Query ID lcj289253 Database Name pdb  
 Description None Description PDB protein database  
 Molecule type amino acid Program BLASTP 2.2.26+ Citation  
 Query Length 102

Other reports: Search Summary Taxonomy reports Distance tree of results Related Structures Multiple alignment

**Graphic Summary**

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. Superfamilies Multi-domains

Distribution of 29 Blast Hits on the Query Sequence

Color key for alignment scores

Score Range	Color
<40	Black
40-50	Blue
50-80	Green
80-200	Red
>=200	Dark Red

Query 1 20 40 60 80 100

**Descriptions**

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Chain A, Solution Structure Of The Ig Domain (2998-3100) Of Human Obscurin	50.4	50.4	84%	5e-09	35%	<a href="#">2CR6_A</a>
<input type="checkbox"/> Chain A, Crystal Structure Of The 32th Ig-Like Domain Of Human Obscurin (Kiaa1556)	48.9	48.9	80%	2e-08	40%	<a href="#">2YZ8_A</a>
<input type="checkbox"/> Chain A, Solution Structure Of The 6th Ig-Like Domain From Human Kiaa1556	48.5	48.5	80%	2e-08	40%	<a href="#">2ETB_A</a>
<input type="checkbox"/> Chain A, Solution Nmr Structure Of Titin I27 Domain Mutant	47.0	47.0	90%	7e-08	35%	<a href="#">2RQ8_A</a>
<input type="checkbox"/> Chain A, Ig?? Protein Domain: sash1HMA1B Chain B, Ig?? Protein Domain: sash1HMA1C Chain C, Ig?? Protein Domain: sash1HMA1D Chain D, Ig?? Protein Domain:	47.0	47.0	86%	7e-08	34%	<a href="#">1W6A_A</a>

At the bottom there will be a long list of solved structures of similar proteins. Each of the solved structures will have a percent sequence identity listed next to the four-character name for that particular structure. You want to have the highest percent sequence identity for your sequence meaning it could be 35% or greater. Once you choose a solved structure (or a few solved structures), you can preform molecular replacement. The higher the percent identity, the better your structure will be in the end because you will not have to build as much into the structure, which may lead to worse  $R_{work}$  and  $R_{free}$  values. If there are no solved structures that are similar to your protein sequence, then it may be useful to create poly-alanine models or truncated loop models. These can be used in molecular replacement to help...

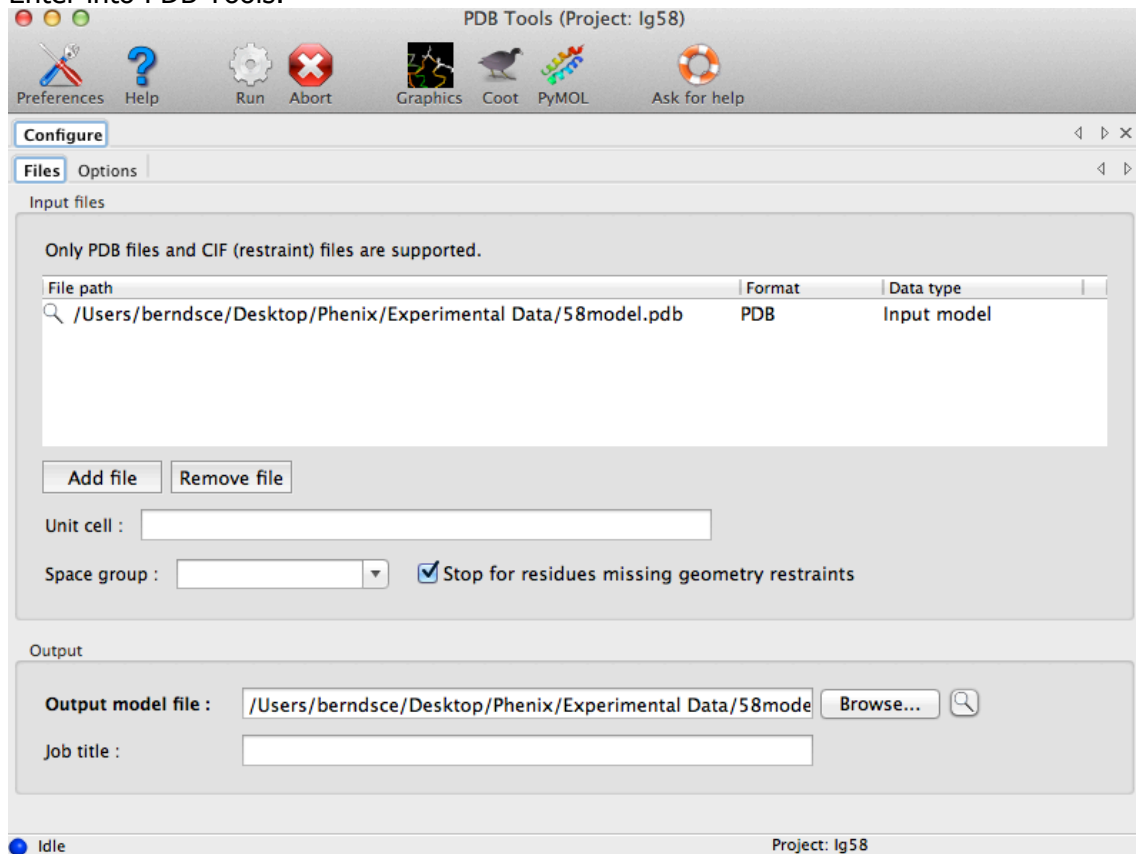
Once you have chosen the PDB structure that you want to use, go to rcsb.org to obtain the PDB file.

Here is where you select to download as a .pdb file

Right next to the ascension number

Under the tab for Model tools on the phenix home page, you will find a sub heading for PDB tools.

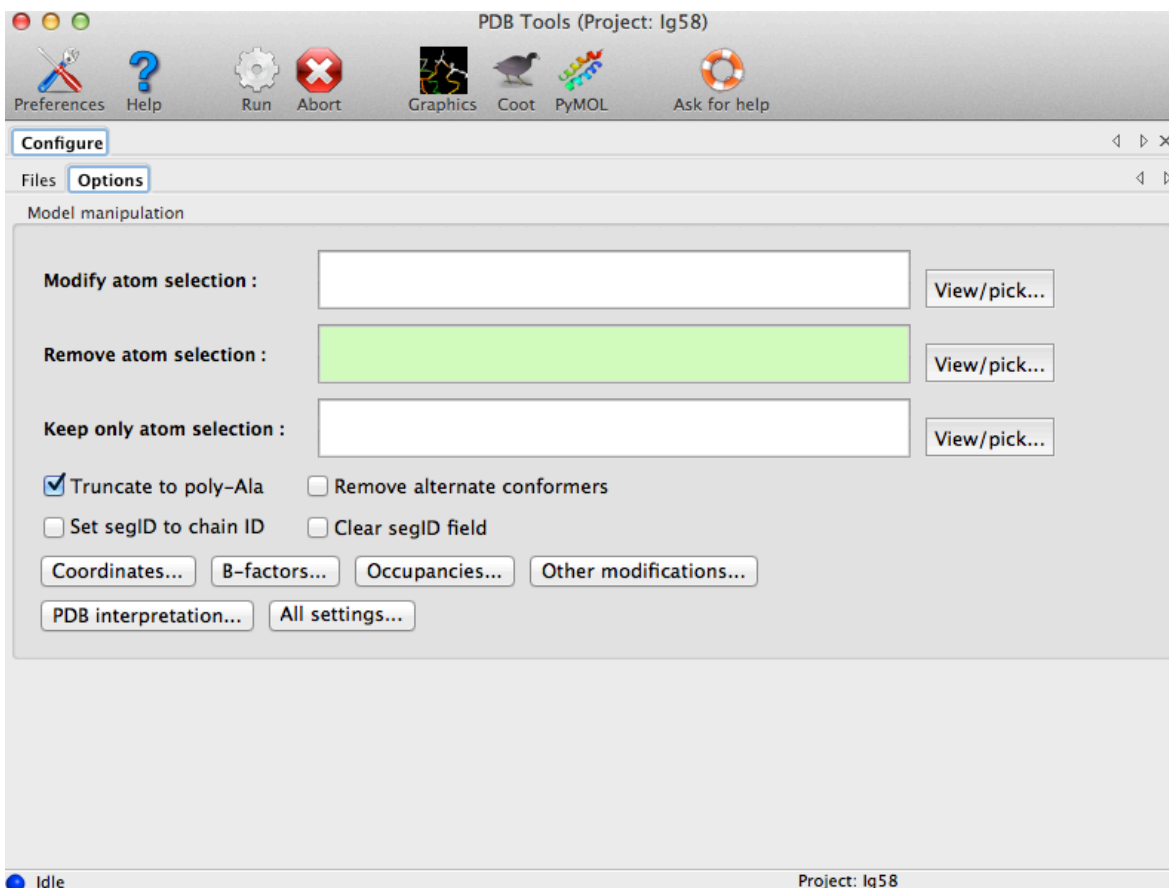
## Enter into PDB Tools.



Add the original .pdb for you protein. In this case it was 58model.pdb

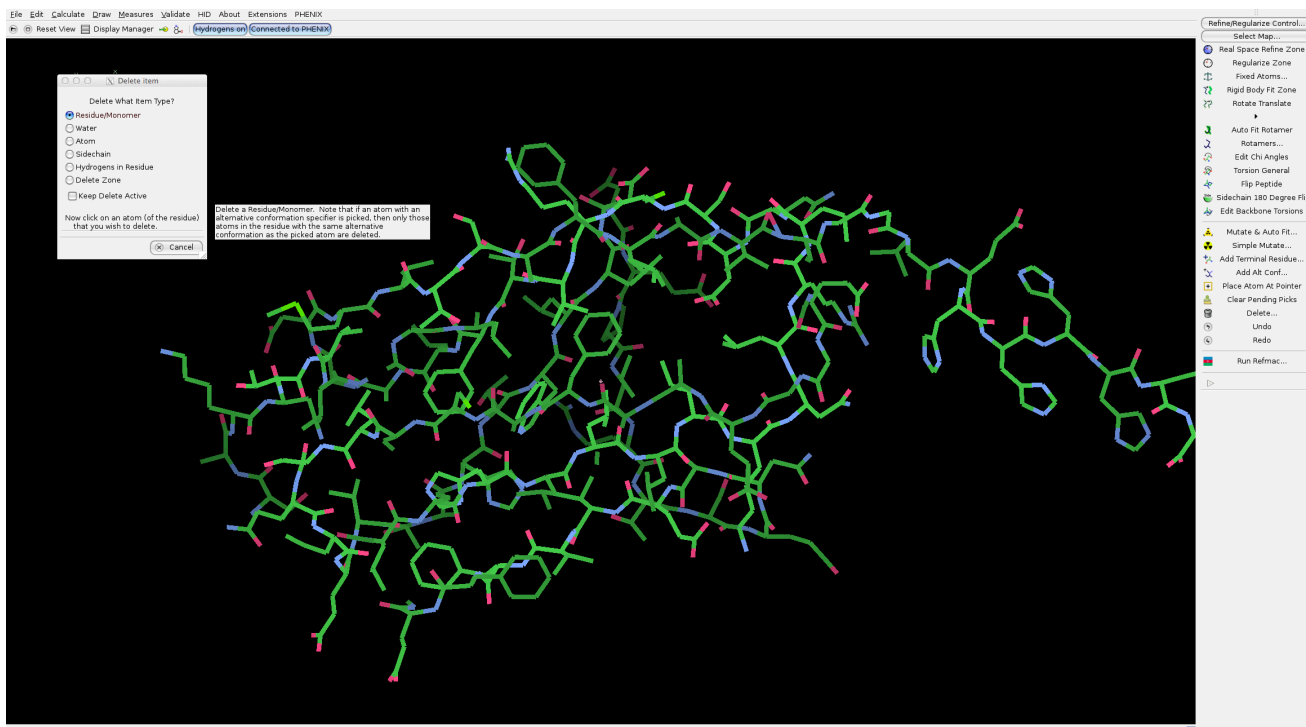
The output file will be directed to the folder that you created on the desktop.



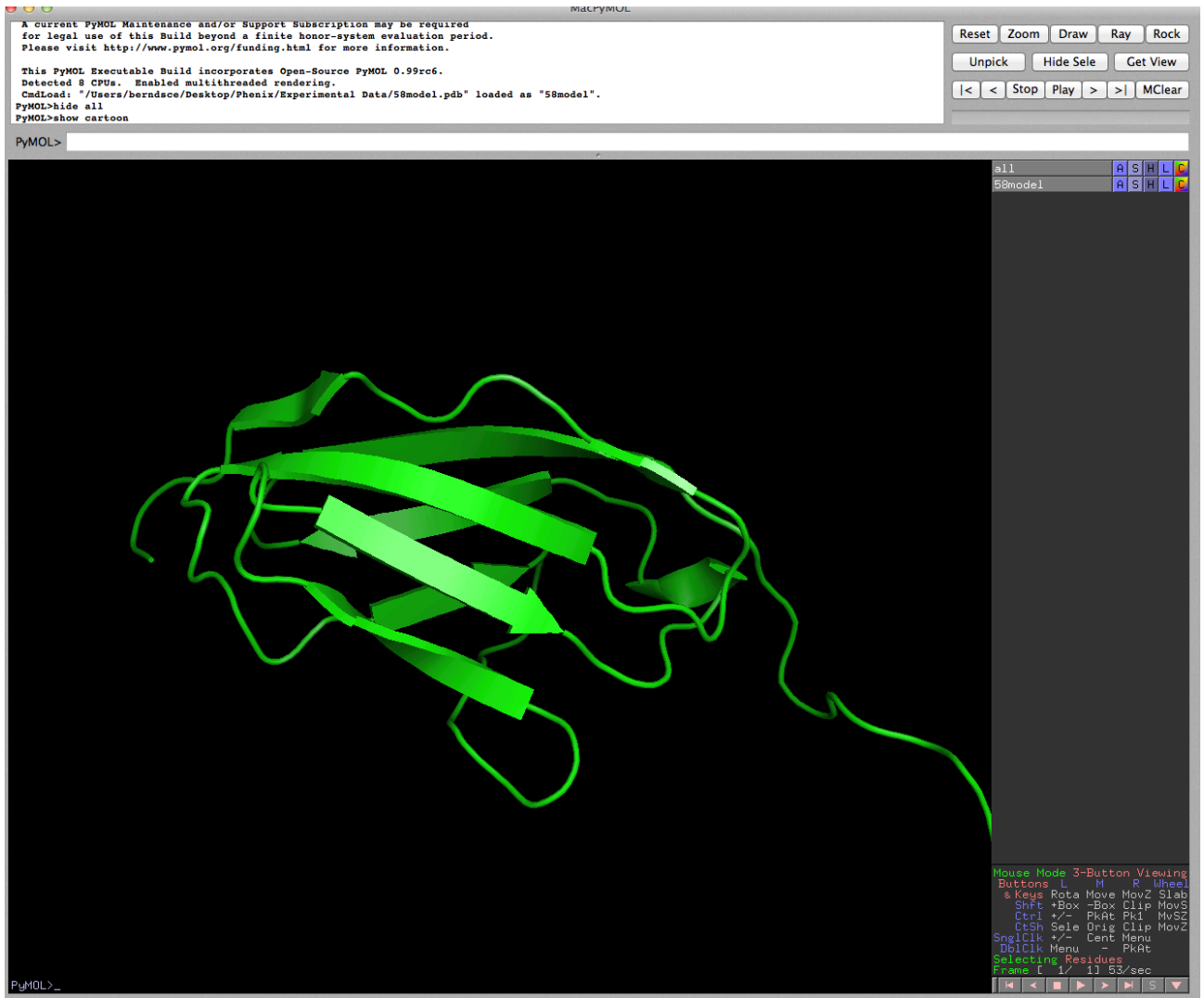


Under the options tab in PDB Tools, select **truncate to poly-Ala**  
Then press run on the top of the tool bar

To create the deleted loops model, open the original .pdb file in the program COOT. The delete icon on the right side on the figure below will pull up the choices in the secondary window for deleting an item in COOT. Choose the residue/monomer option. Then use the cursor to click on the carbonyl carbon of the backbone. This will delete the backbone and the side chain of the amino acid in the loop.

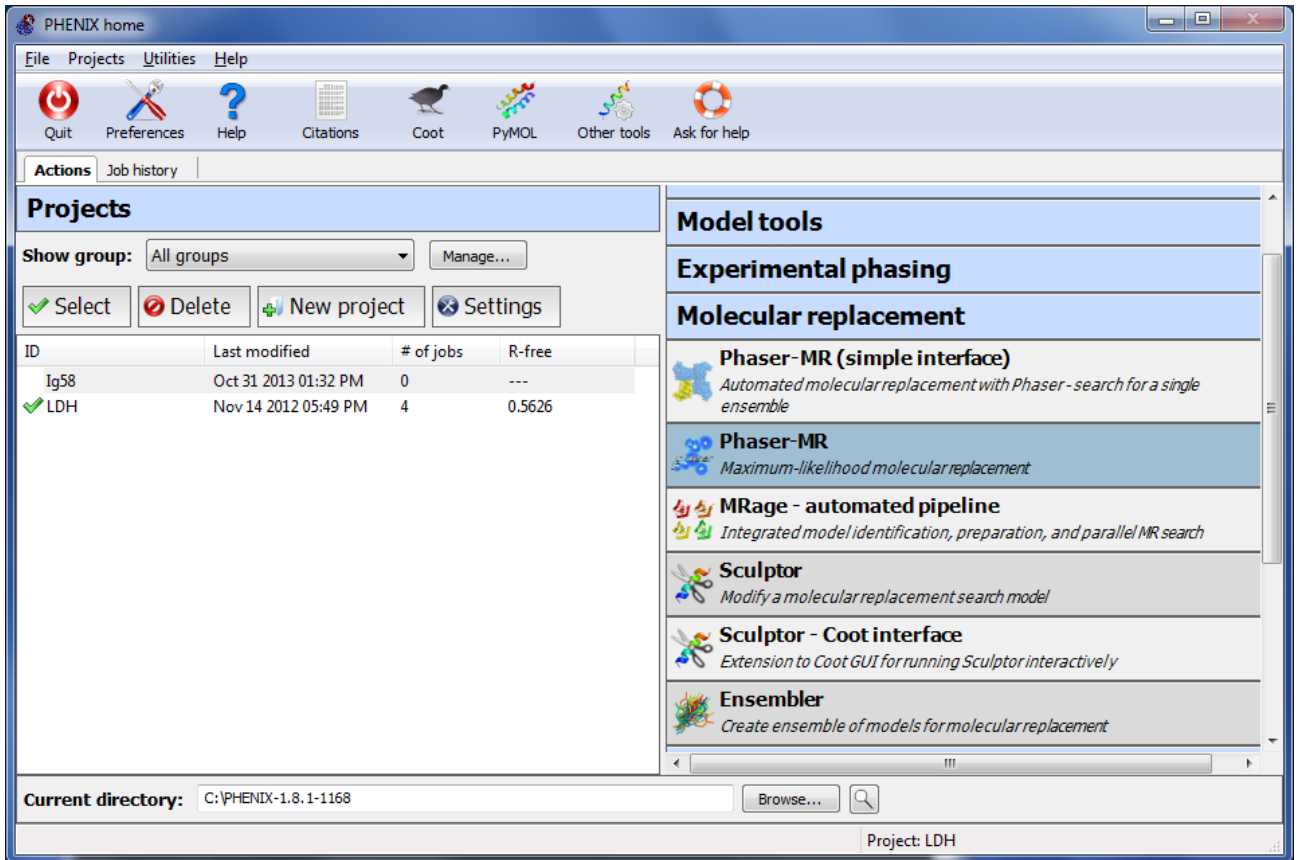


It is difficult to see what secondary structure the amino acids are in within COOT, so use PyMOL to view the cartoon

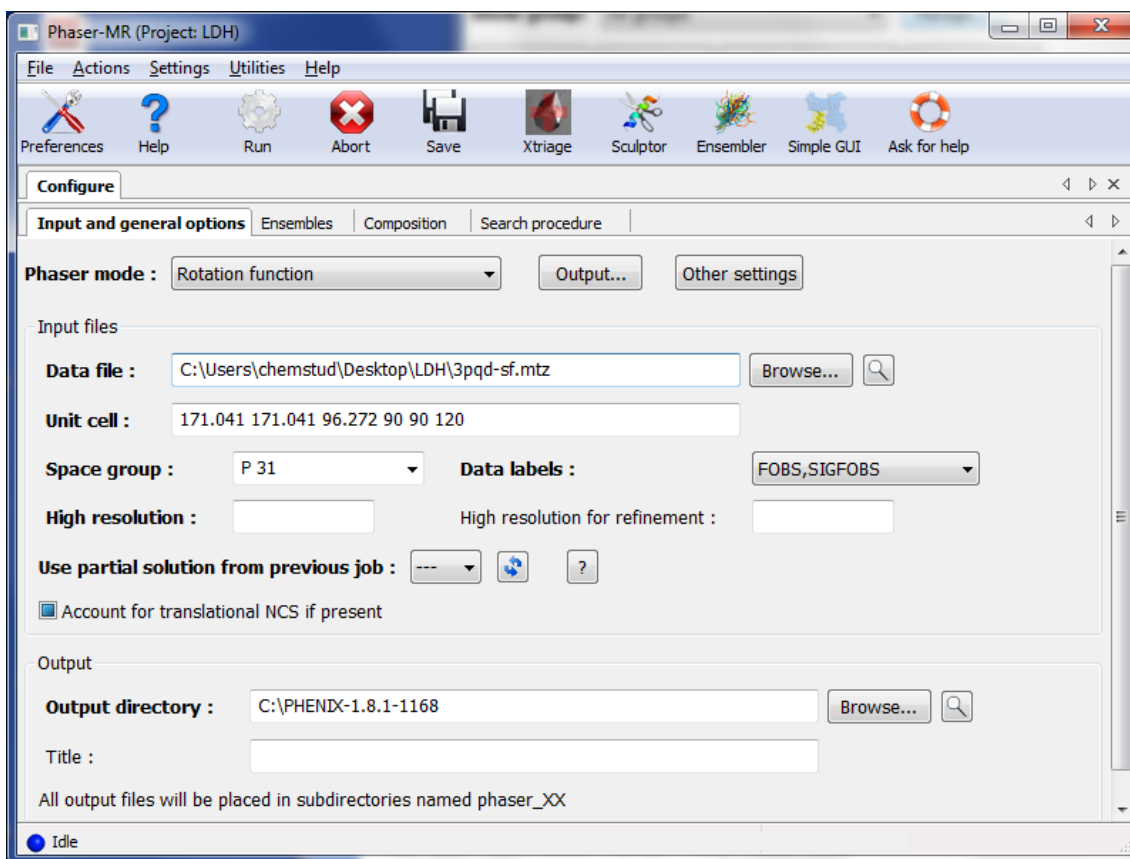


Show the protein as a cartoon so that it is easy to see what is a loop and what is not. A loop is a turn between a secondary structure or in a random coil.

## Step 4: Phaser-MR



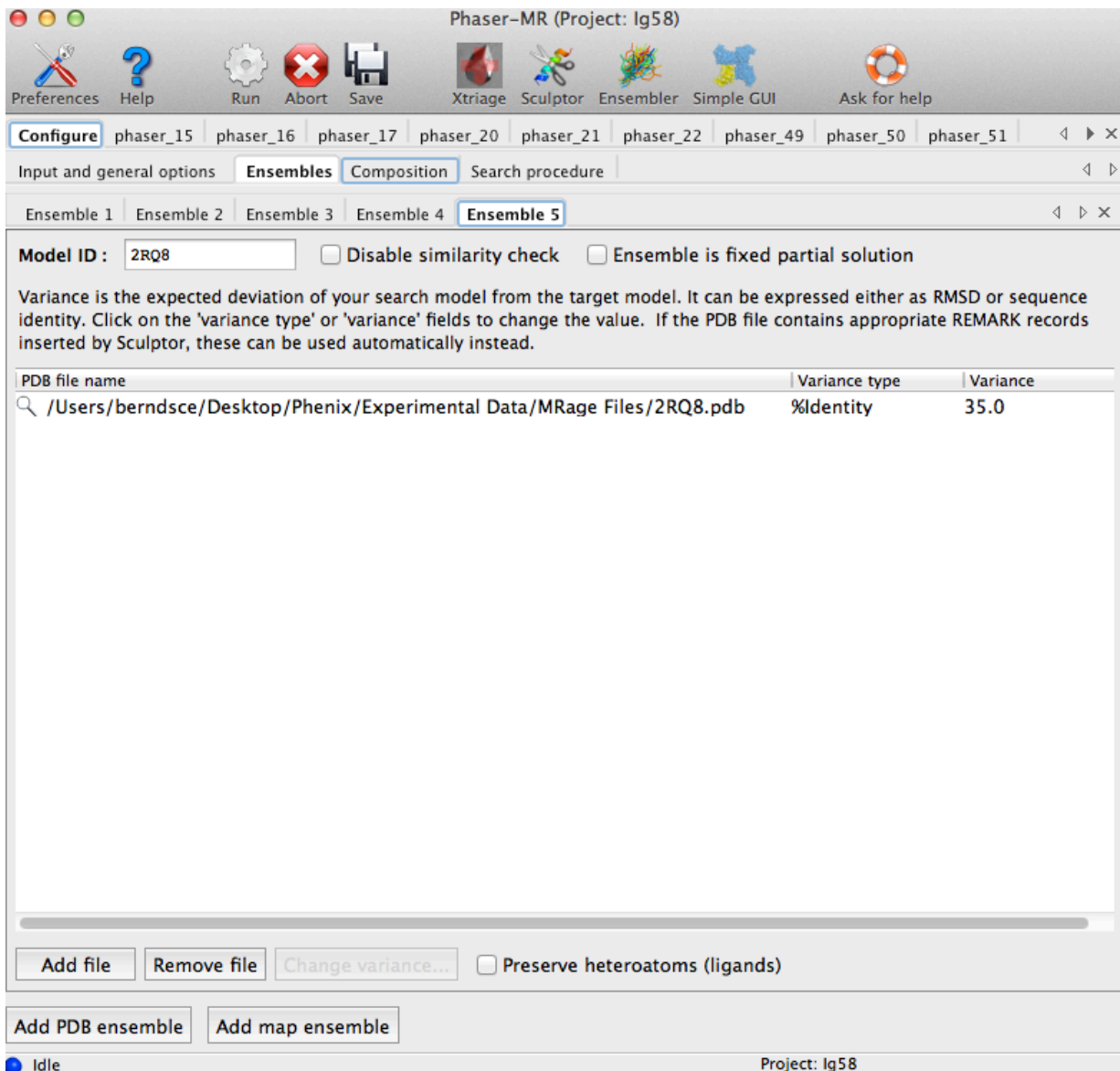
There are multiple programs within the molecular replacement tab on the home screen of Phenix. We advise that you begin using Phaser-MR.



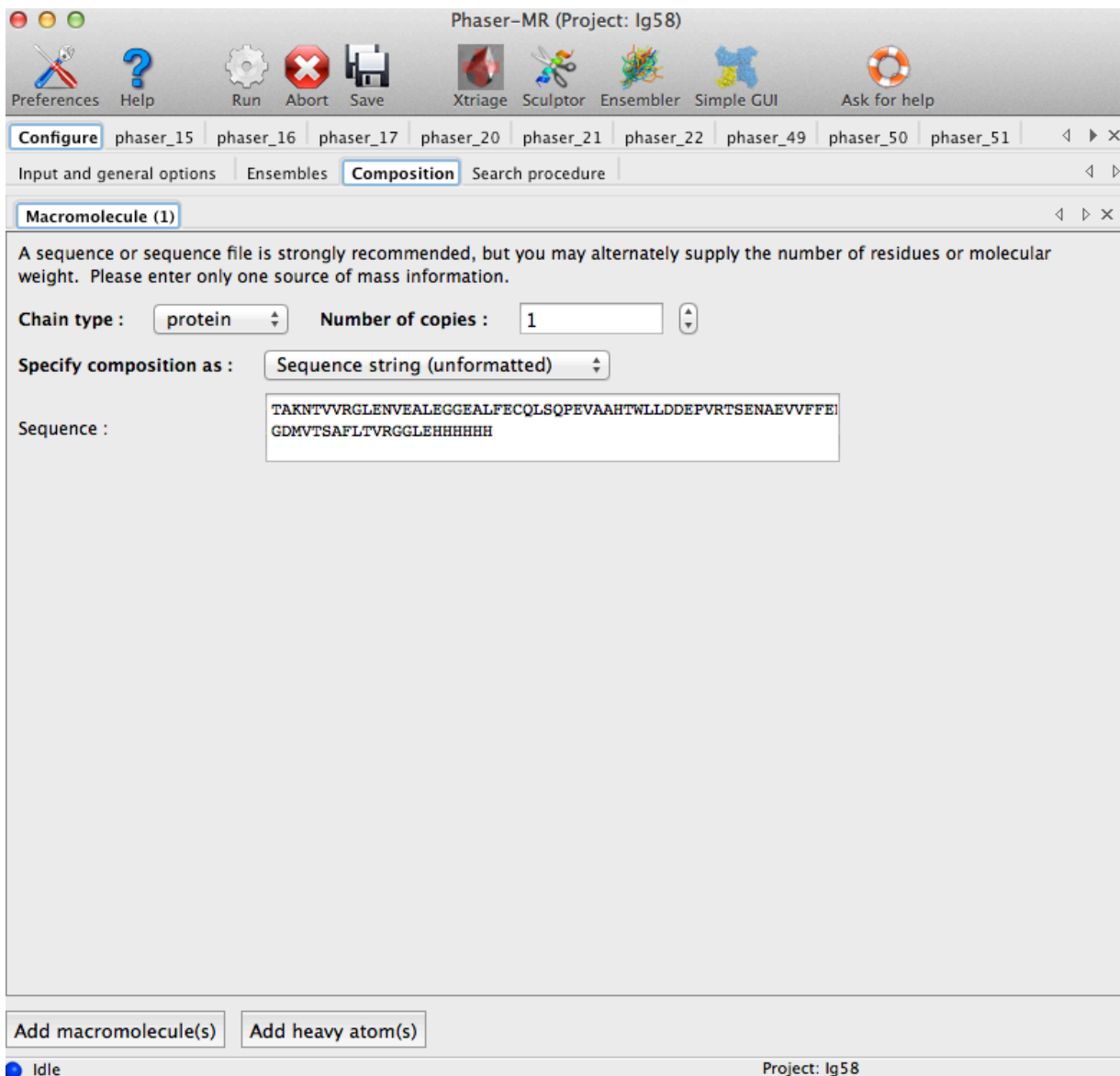
After selecting Phaser-MR insert your reflections file into the **Data file** input bar under the input and general options tab. Also, select the correct output directory [your phenix (your name) folder] and make sure that the space group is correct.

Next click the ensembles tab and add the file of one of your models or templates. Give them a title in the space labeled model ID. You can add multiple ensembles for molecular replacement to use by clicking the add pdb ensembles button at the bottom of the window.

After you have added an ensemble click change variance and depending on whether it was a template or a model, enter the percent sequence identity or an RMSD of 2, respectively.



After you have added the desired amount of ensembles, click the composition tab and enter your sequence after choosing sequence string in the drop down menu of the **Specify composition as:** option.



Click the Search Procedure tab and under search method, select full. Also check the boxes of each ensemble you wish to use in the **Model IDs:** box.

For each new run of molecular replacement, we recommend deleting ensembles entirely rather than simply unchecking them in the **Model IDs:** box if you wish to not use that particular ensemble.

Now click the run icon. This should take a while to run. When the run is complete you will see a window similar to this.

Phaser-MR (Project: Ig58)

Preferences Help Run Abort Save Xtrriage Sculptor Ensembler Simple GUI Ask for help

Configure phaser\_15 phaser\_16 phaser\_17 **phaser\_20** phaser\_21 phaser\_22 phaser\_49 phaser\_50 phaser\_51

Run status Graphs **Results**


Summary


**Phaser has found 167 MR solution(s).**

**Top LLG:** 29.171

**Top TFZ:** 5.0

**Spacegroup:** P 61 2 2

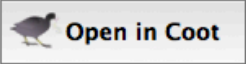

 **Run phenix.refine**

 **Run AutoBuild**

Output files

**Directory:** /Users/berndsce/Desktop/Phenix/phaser\_20

File name	Contents
lg58_phaser.1.mtz	Phases and maps
lg58_phaser.1.pdb	Output model
lg58_phaser_data.mtz	Processed input data

 **Open in Coot**  **Open in PyMOL**

Idle Project: Ig58

Here you are looking for an LLG score that is as high as possible. My cohort and I were able to obtain values between 25 and 60, but the higher the better. You also want a TFZ score of 5 or very close to it.



## Step 5: Autobuild

The screenshot shows the PHENIX software interface. At the top, there is a menu bar with icons for Quit, Preferences, Help, Citations, Coot, PyMOL, KING, Other tools, and Ask for help. Below the menu bar, there are two tabs: 'Actions' and 'Job history'. The main window is divided into two panes. The left pane is titled 'Projects' and contains a 'Show group:' dropdown menu set to 'All groups', a 'Manage...' button, and buttons for 'Select', 'Delete', 'New project', and 'Settings'. Below these buttons is a table with the following data:

ID	Last modified	# of jobs	R-free
✓ Ig58	Nov 22 2013 04:3...	61	0.3719

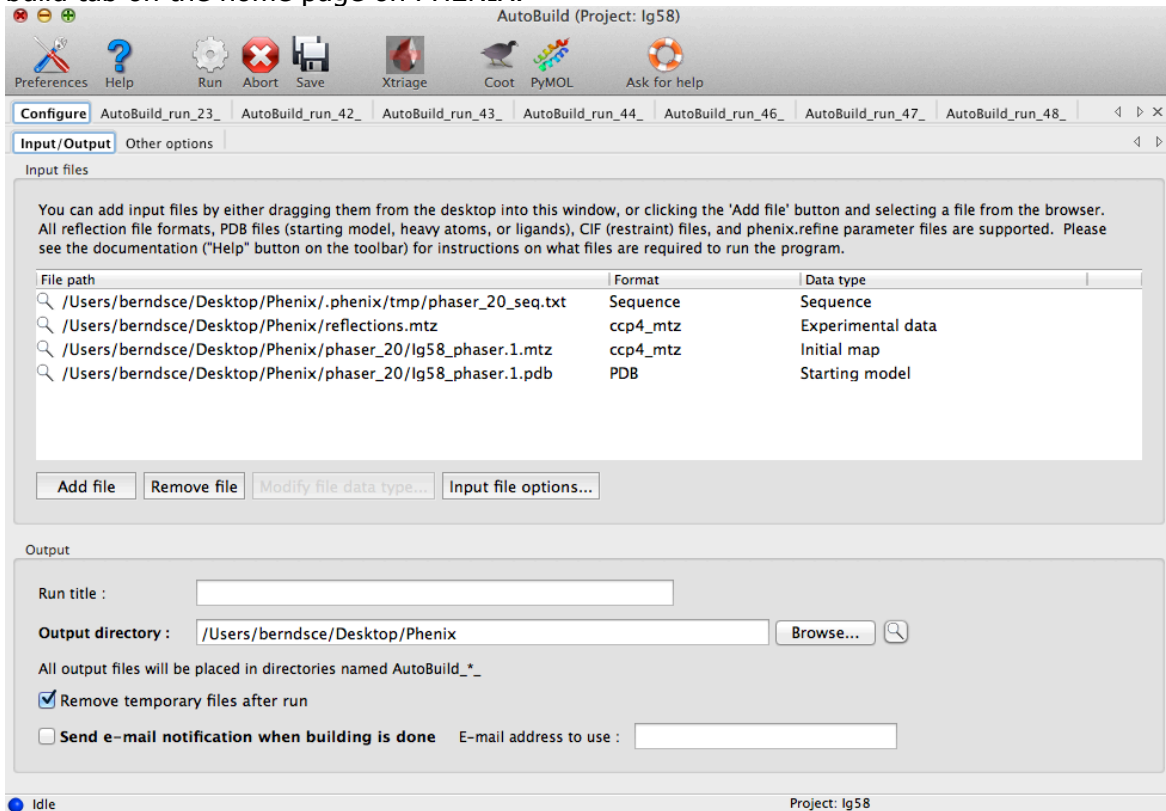
The right pane is titled 'Favorites' and contains a list of tool categories: Reflection tools, Model tools, Experimental phasing, Molecular replacement, Model building, Refinement, Maps, Ligands, Validation, and Utilities. The 'Model building' category is expanded, showing the following options:

- AutoBuild**: Automated model-building and refinement
- Phase and build**: Faster auto-building combined with density modification
- Find Helices and Strands**: Fast chain tracing
- Fit Loops**: Fast placement of missing loops in electron density
- Morph model**: Model improvement for poor molecular replacement solutions

At the bottom of the interface, there is a 'Current directory:' field containing '/Users/berndsce/Desktop/Phenix' and a 'Browse...' button. The status bar at the very bottom shows 'PHENIX version 1.8.4-1496' on the left and 'Project: Ig58' on the right.

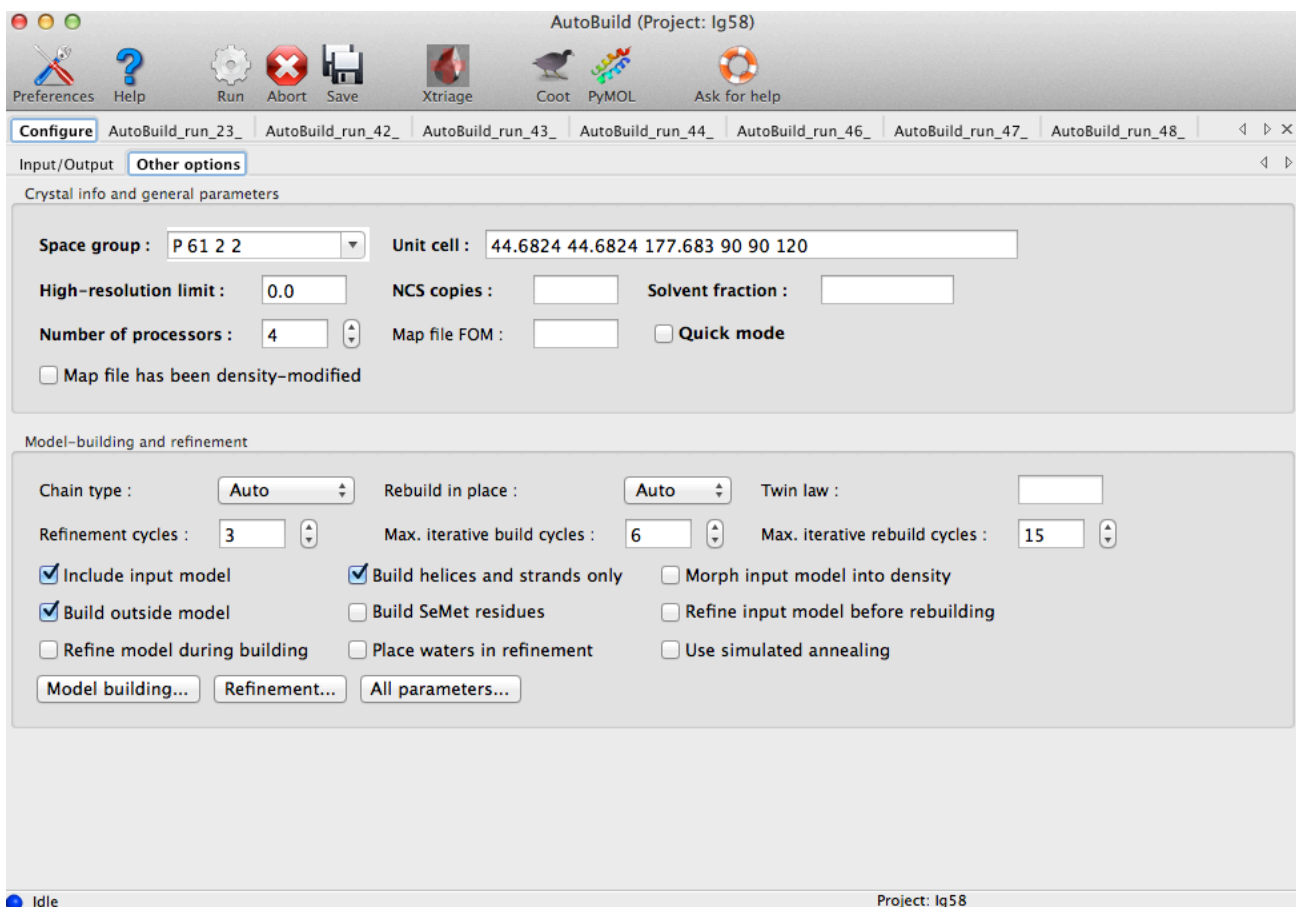
The next step in the process is to do autobuild on the results from Phaser-MR. If you are not already in Phaser-MR then you can find the autobuild program under the model

build tab on the home page on PHENIX.



To run autobuild you need to have the following files:  
the sequence of your protein. (it should be a .txt file)  
the original reflections file for your protein (.mtz)  
the .mtz from the phaser run (the initial map in .mtz file)  
the .pdb file from phaser (starting model)

Before you move to the other options make sure the output directory is the correct place



Under the Other Options tab at the top of the screen, you need to fill the necessary refinement options. Most of the time you will want to include input model, build outside the model, refine model during building, refine input model before rebuilding. Depending on how many cycles you want to perform for the autobuild, the number of refinement cycles will change. This goes for max. iterative build cycles and max. iterative rebuild cycles as well. Usually the refinement cycles is 3, the build cycles is 6 and the rebuild cycles is 15.

AutoBuild (Project: Ig58)

Configure **AutoBuild\_run\_23** AutoBuild\_run\_42\_ AutoBuild\_run\_43\_ AutoBuild\_run\_44\_ AutoBuild\_run\_46\_ AutoBuild\_run\_47\_ AutoBuild\_run\_48\_

Status Summary Model-building Structure status

Wizard status

**FINISHED**

Overall best on cycle 4 R= 0.29 Rfree= 0.44  
Residues built: 82 Side chains: 56

Model-map correlation: 0.72

Last cycle=5 R= 0.31 Rfree= 0.41  
Residues built: 81 Side chains: 77

Log output

```
Terwilliger TC. (2004) SOLVE and RESOLVE: Automated structure solution, density
modification, and model building. J. Synchrotron Rad. 11:49-52.

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung
L-W, Read RJ, Adams PD. (2008) Iterative model building, structure refinement
and density modification with the PHENIX AutoBuild wizard. Acta Cryst.
D64:61-69.

Zwart PH, Grosse-Kunstleve RW, Adams PD. (2005) Xtrriage and Pest: automatic
assessment of X-ray data and substructure structure factor estimation. CCP4
Newsletter Winter:Contribution 7.

Cleaning up AutoBuild_run_23_/TEMP0 now...
Done cleaning up ...
```

Pause Abort View log

Idle Project: Ig58

after the autobuild run is complete, you will see this screen under the status tab at the top. The goal here is to have a low R ( $\sim 0.2$ ) and a low R-free ( $\sim 0.2$ ). The R-free will always be above the R value, but you want to get the R-value and the R-free value within 0.05 of each other. The closer the model-map correction is to 1.0 the better. This indicates that the model produced from autobuild is close to the density map from the diffractions. The closer you are to 1.0 the less you have to manual build in COOT.

AutoBuild (Project: Ig58)

Configure **AutoBuild\_run\_23\_** AutoBuild\_run\_42\_ AutoBuild\_run\_43\_ AutoBuild\_run\_44\_ AutoBuild\_run\_46\_ AutoBuild\_run\_47\_ AutoBuild\_run\_48\_

Status **Summary** Model-building Structure status

Output files

Directory: /Users/berndsce/Desktop/Phenix/AutoBuild\_run\_23\_

File name	Contents
AutoBuild_run_23_1.log	AutoBuild log
overall_best.pdb	Best model from RESOLVE/phenix....
overall_best_denmod_map_coeffs.mtz	Best density-modified map
overall_best_placed.pdb	Best model (docked sequence only)
overall_best_refine_data.mtz	Data and R-free flags
overall_best_refine_map_coeffs.mtz	Best refined maps (2Fo-Fc, Fo-Fc)
AutoBuild_summary.dat	Summary of run results

Data analysis

AutoBuild has analyzed your X-ray data with Xtrriage. This will indicate whether you have any pathologies such as twinning or pseudosymmetry, as well as providing information on data quality and anomalous signal.

Xtrriage log file Results and graphs

Final model

R-work:	0.2935	R-free:	0.4534	Residues:	82
Fragments:	4	Waters:	9		

Run phenix.refine Run validation

Idle Project: Ig58

The summary tab shows you the output files that came from the autobuild run. From this page you can Run phenix.refine on the results that can out of the autobuild. Refinement will try to get the R-work and R-free closer together and closer to 0.2.

The screenshot shows the AutoBuild software interface for Project: Ig58. The main window displays the 'Model-building' tab, which provides a summary of the current best model and a table of model-building cycles.

**Current best model**

<b>R-work:</b>	0.2935	<b>R-free:</b>	0.4534	<b>CC:</b>	---
----------------	--------	----------------	--------	------------	-----

**Model-building cycles**

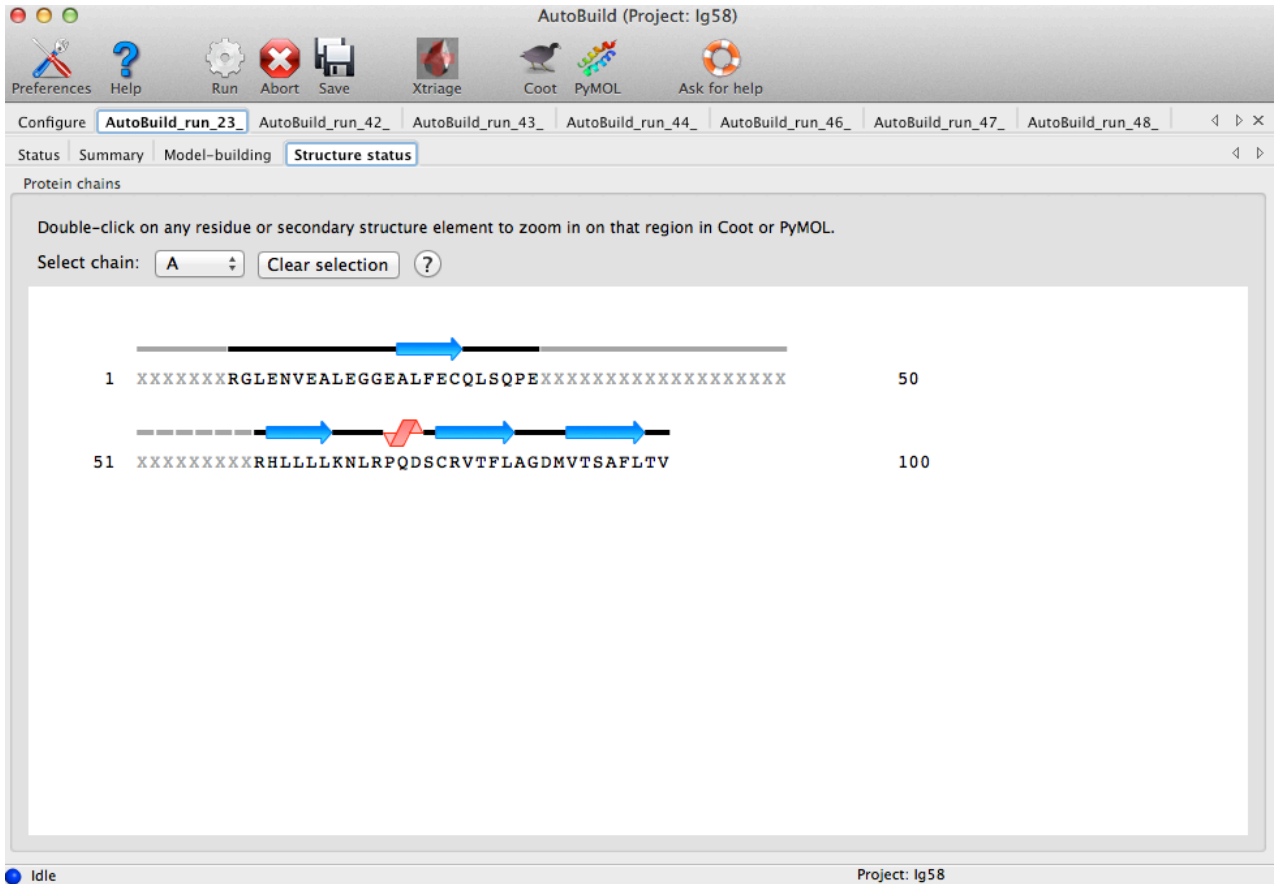
AutoBuild has set up parallel model-building runs. The best model from these will be automatically selected.

Run	Status	Residues	R	R-free	CC
1	Finished	82	0.2879	0.4433	n/a
2	Finished	82	0.2879	0.4433	n/a
3	Running	81	0.3158	0.4328	n/a
4	Finished	79	0.3189	0.4023	n/a

Below the table is a button labeled 'Open in Coot' with a Coot icon.

The status bar at the bottom shows 'Idle' on the left and 'Project: Ig58' on the right.

The model-building tab shows you the best R-work and R-free values that came from the autobuild.



The structure status tab will show how much of the structure has been identified for your protein. In this case there is a large portion missing from the middle of the protein sequence. This means that you will either have to go forward with this structure and run the phenix.refine and then manually build in that portion of the protein sequence in COOT or you can keep performing autobuilds until the structure gets better.

## Step 6: phenix.refine

phenix.refine (Project: Ig58)

Configure Refine\_24 Refine\_25 Refine\_26 Refine\_27 Refine\_28 Refine\_29 Refine\_30 Refine\_31 Refine\_32 Refine\_33 Refine\_34 Refine\_35

Input data Refinement settings Output

Input files

File path	Format	Data type
/Users/berndsce/Desktop/Phenix/reflections.mtz	ccp4_mtz	X-ray data, X-ray data
/Users/berndsce/Desktop/Phenix/Ig58_refine_37-coot-1.pdb	PDB	Input model

Add file Remove file Modify file data type... Use symmetry from selected file

Space group: P 61 2 2 Unit cell: 44.6824 44.6824 177.683 90 90 120

X-ray data and experimental phases

Data labels: I,SIGI R-free label: FreeR\_flag Test flag value: 1

High resolution: (2.701) Low resolution: (16.197) Phase labels: ---

Options...

Neutron data

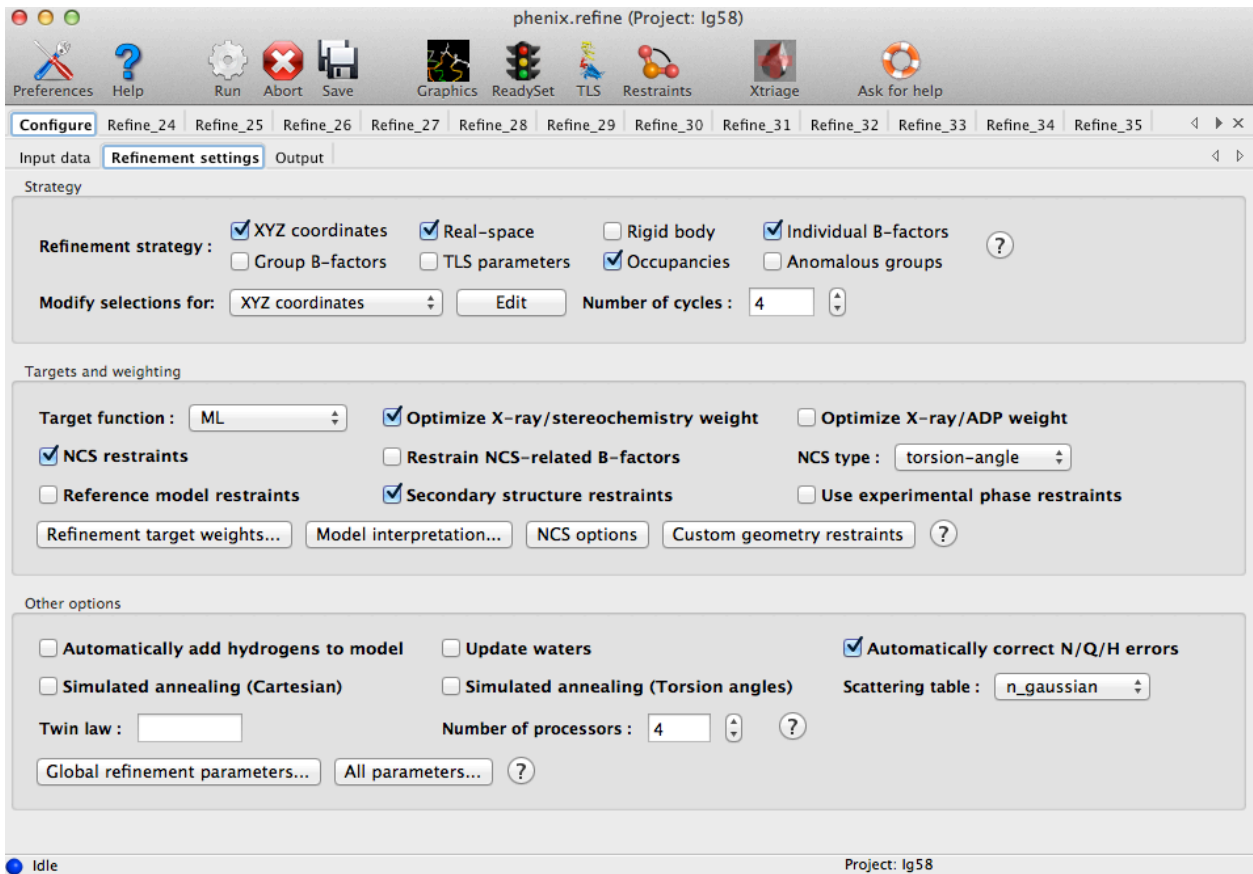
Data labels: --- R-free label: --- Test flag value:

High resolution: Low resolution: Options...

Idle Project: Ig58

For phenix.refine you need two files in the file path. One needs to be the original reflections.mtz file and the other needs to be the .pdb file from the autobuild or the .pdb that you would like to refine.





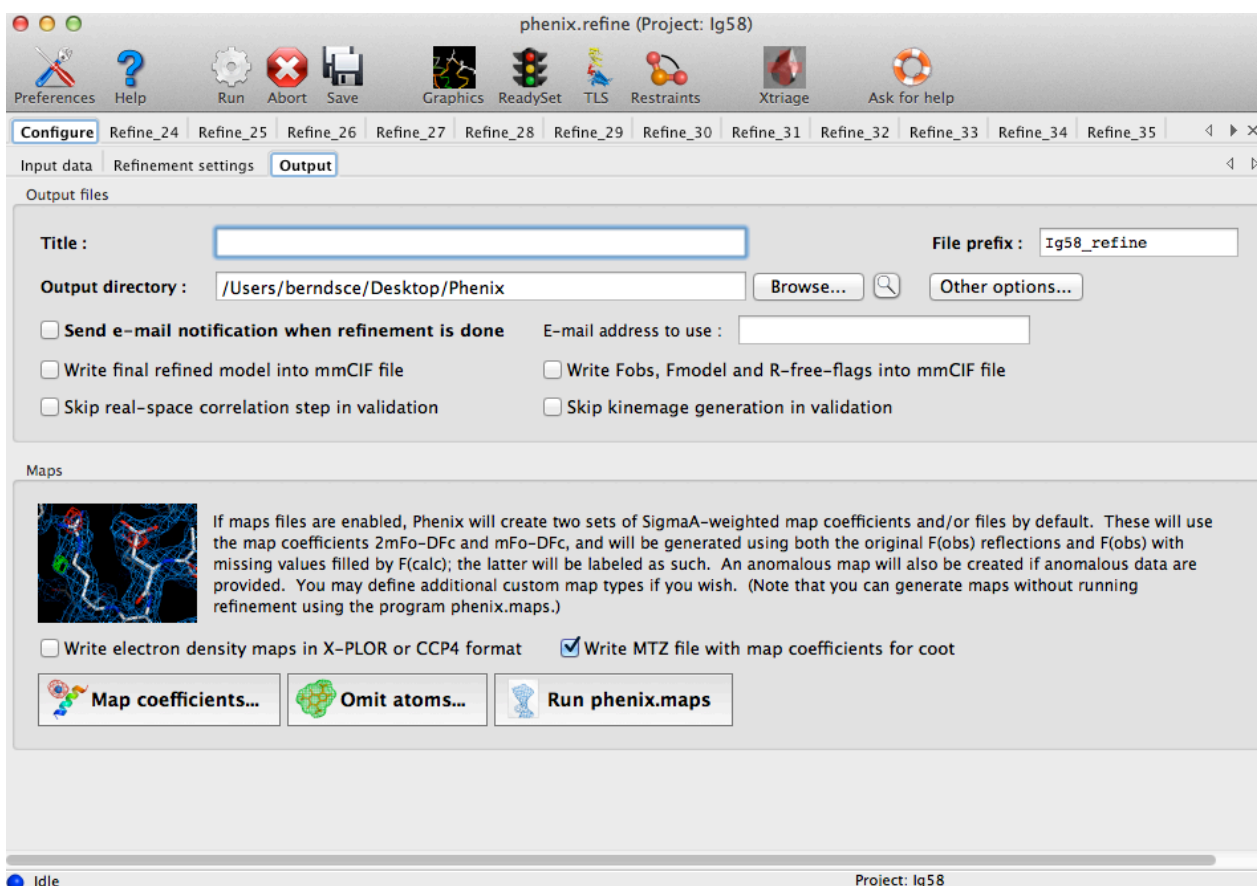
Usually when you run phenix.refine the refinement setting are as follows:

The strategy usually has the XYZ coordinates checked off, as well as real-space, rigid body, and individual B-factors

The targets and weighing usually has NCS restraints, checked off as well as secondary structure restraints

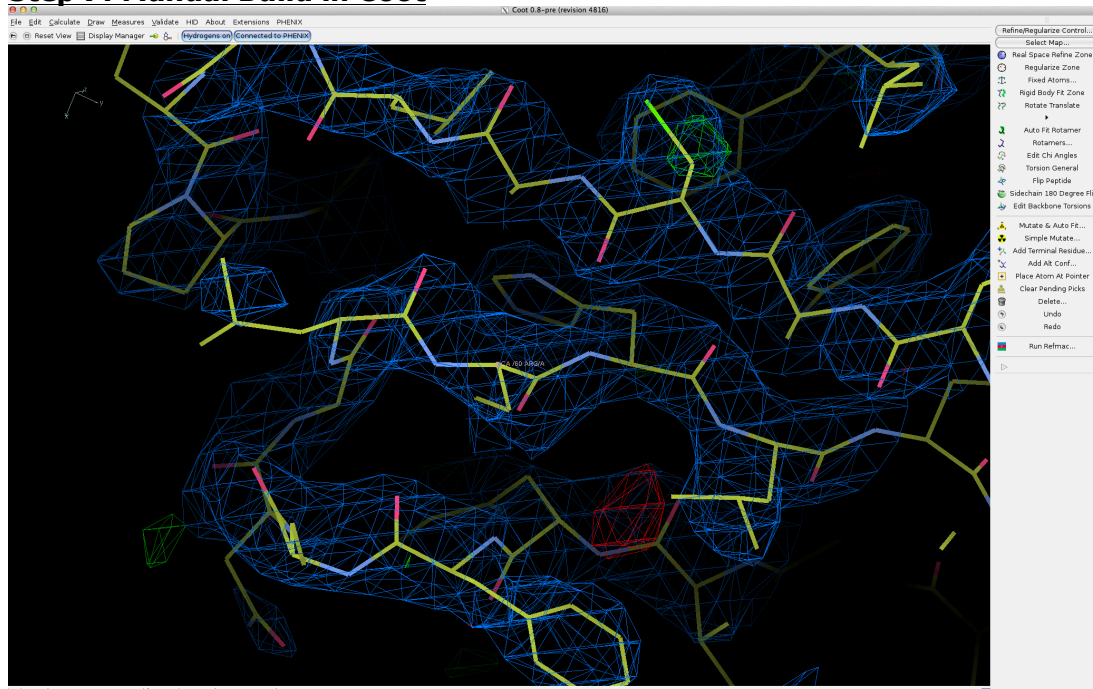
The other options usually has automatically correct N/Q/H errors

HOWEVER, these do not always need to be checked off. For more information on what needs to be checked off, click on the question mark to read more about each of the options on this tab. You will learn what each parameter will do during the refinement. This will allow you to pick and choose what restraints or parameters to use when performing a refinement on your structure.



Before performing the refinement on your protein, make sure the results are being placed in the correct location.

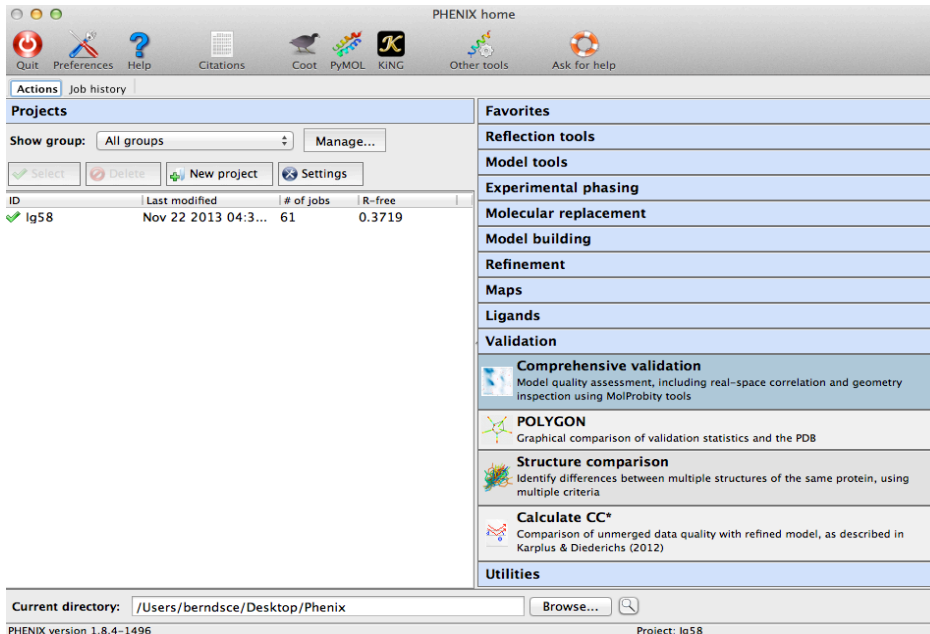
## Step 7: Manual Build in Coot



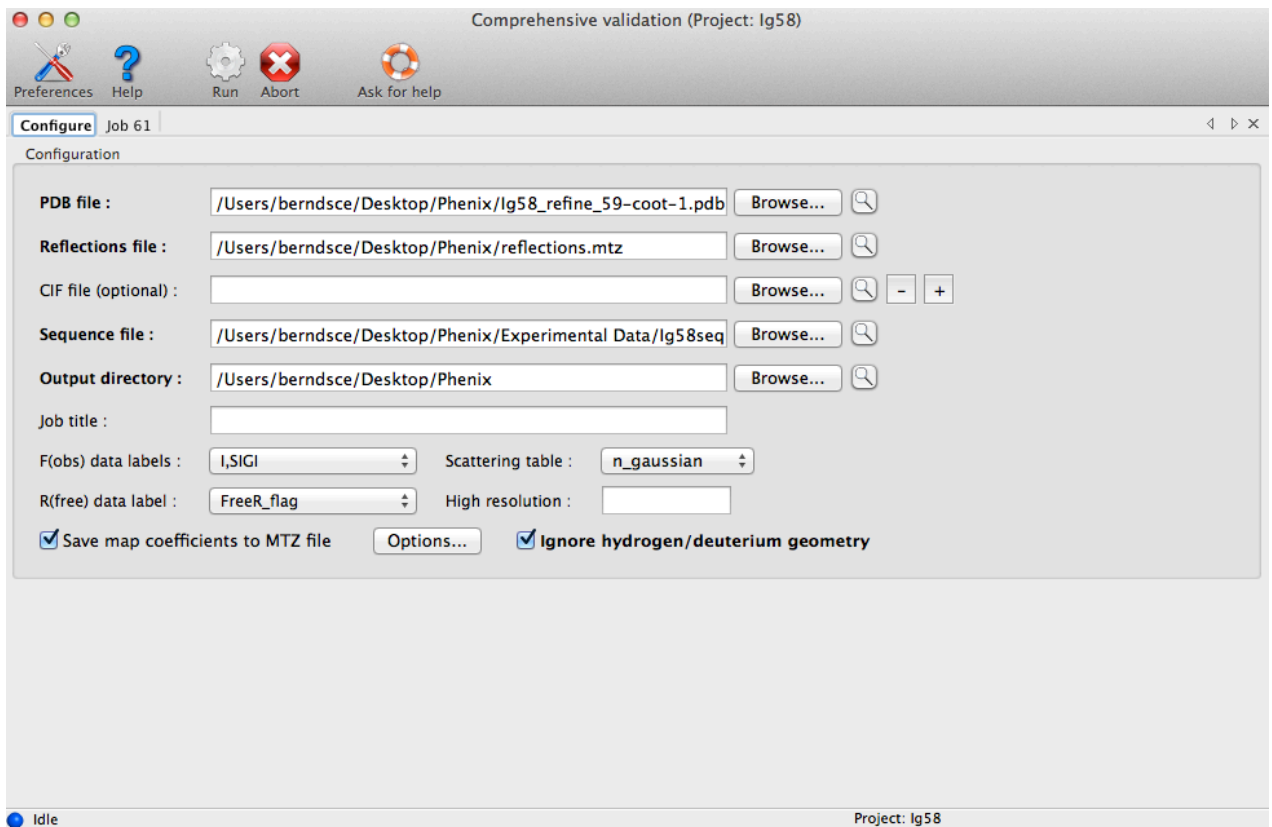
For more information on how to manually build in COOT, please refer to the COOT SECTION.

## Step 8: Comprehensive Validation

Once you have a good looking structure from performing the manual build in COOT, you need to confirm that the structure looks OKAY. This means the R-work and R-free are within 0.05 of each other and they are near 0.2. To see what the R-work and R-free values are you need to run a comprehensive validation.



The comprehensive validation program is under the validation tab on the right hand side on the home page.



To run the validation program you need the following files:

the .pdb file that you want to validate

the original reflections.mtz file

the starting sequence file (.dat or .fasta)

Also make sure the output files are going to the correct location/

Once you have all the requirements for running the comprehensive validation program click RUN on the top bar.

The screenshot shows the 'Comprehensive validation (Project: Ig58)' window. The top bar contains icons for Preferences, Help, Run, Abort, and Ask for help. Below the top bar, the 'Configure' section shows 'Job 61' and tabs for 'Model and data', 'MolProbity', 'Real-space correlation', 'Atomic properties', and 'Sequence check'. The 'Basic statistics' section displays the following information:

PDB file: /Users/berndsoe/Desktop/Phenix/Ig58\_refine\_59-coot-1.pdb  
Reflection file: /Users/berndsoe/Desktop/Phenix/reflections.mtz  
Data labels: I,SIGI      Resolution: 16.20 - 2.70  
R-work: 0.3038      R-free: 0.3978  
RMS(bonds): 0.0097      RMS(angles): 2.417  
Avg. B-factor: 43.40

By default, if you have not defined other map types, Coot and PyMOL will display the 2mFo-DFc and mFo-DFc maps that were generated for real-space correlation calculations; these are essentially the same as the maps created by phenix.refine, but without missing Fobs filled in. (Note: you must turn on XPLOr map output for the maps to appear in PyMOL.)

Buttons: Open in Coot, Open in PyMOL, Compare statistics, Sequence viewer

The 'Detailed statistics' section shows 'X-ray statistics by resolution bin:' and a 'Plot statistics by resolution' button. Below is a table:

	R-work	R-free	%complete	FOM	Phase error	Scale factor	#work	#test
16.1976 - 3.3974	0.2823	0.3830	98.8%	0.71	32.41	1.01	1523	168
3.3974 - 2.7007	0.3667	0.4426	100.0%	0.56	48.54	0.96	1429	159

Project: Ig58

The results from the comprehensive validation will several important tabs to look at. The model and data tab will tell you about the R-work, R-free, RMS (bonds), RMS (angles) and the average B-factor. These are important to note especially the R-work and R-free because once they are close to 0.2 and within 0.05 of each other then you can run MolProbity on your structure.

Comprehensive validation (Project: Ig58)


Preferences Help Run Abort Ask for help

Configure Job 61

Run status Model and data MolProbity Real-space correlation Atomic properties Sequence check

Summary Basic geometry Protein Clashes

Validation summary



These statistics are computed using the same underlying distributions as the MolProbity web server. The overall score represents the experimental resolution expected for a model of this quality; ideally the score should be lower than the actual resolution.

Basic statistics for Ig58\_refine\_59-coot-1.pdb:

Ramachandran outliers:	0.0%	(Goal : < 0.2%)	Ramachandran favored:	95.4%	(Goal : > 98%)
Rotamer outliers:	1.3%	(Goal : 1%)	C-beta outliers:	4	(Goal : 0)
Clashscore:	9.41		Overall score:	1.91	

Show validation in KiNG

Missing atoms

No missing non-hydrogen atoms detected.

Idle Project: Ig58

The comprehensive validation will give you a summary from MolProbity but this is not good enough. You need to go on to the MolProbity website and run your structure through those programs.

## Step 9: MolProbity <http://molprobity.biochem.duke.edu/>

ONLY do MolProbity when the R-work and R-free are close to 0.2 and within 0.05 of each other. If they are not within this range, then repeat the refinement (change the parameters if needed), then the manual build, and the comprehensive validation.

**Main page**

FILE UPLOAD/RETRIEVAL (MORE OPTIONS)

PDB/NDB code:  type: PDB coords

No file chosen type: PDB coords

We have updated Reduce to add hydrogens at a length more consistent with electron-cloud positions, and accordingly adjusted the Van der Waals radii in Probe to compensate for the change. This will affect comparison of results calculated with older versions of MolProbity, but generally results in lower clashscores. For analyses using nuclear-position hydrogens, you have the option of selecting nuclear x-H positions when adding hydrogens. Read more about this change [here](#).

Ramachandran scoring has also been updated to use new six-category distributions, derived from a larger **Top8000** dataset of high quality PDB files.

Please don't hesitate to report any **bugs** you may encounter.

If you prefer to use the previous MolProbity3, please go to <http://helix.research.duhs.duke.edu>.

**Walk-thrus & tutorials:**

**Evaluate X-ray structure:** Typical steps for a published X-ray crystal structure or one still undergoing refinement.

**Evaluate NMR structure:** Typical steps for a published NMR ensemble or one still undergoing refinement.

**Fix up structure:** Rebuild the model to remove outliers as part of the refinement cycle.

**Work with kinemages:** Create and view interactive 3-D graphics from your web browser.

**What's new in 4.02b:**

- Updated Reduce-added hydrogen lengths to be more consistent with other crystallography software.
- New Top8000 Ramachandran validation information.
- Allows use of both electron cloud and nuclear x-H bond-lengths, monitored in real-time.
- Full support for 2-character CHAINIDs added. 4-character SEGIDs also supported in place of CHAINIDs.
- Occupancy cutoff for clashscore now set to 0.1. Clashscore calculation includes all alternates.

**Common questions:**

**Cite MolProbity:** Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66:12-21.

and/or

Davis et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Research 35:W375-W383.

**Cite KING:** Chen et al. (2009) KING (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. Protein Science 18:2403-2409.

**About hydrogens:** Why have the hydrogen bondlengths changed?

**Installing Java:** how to make kinemage graphics work in your browser.

**Download MolProbity:** how can I run a private MolProbity server, or run from the command line?

*NB: the back button doesn't work inside MolProbity*

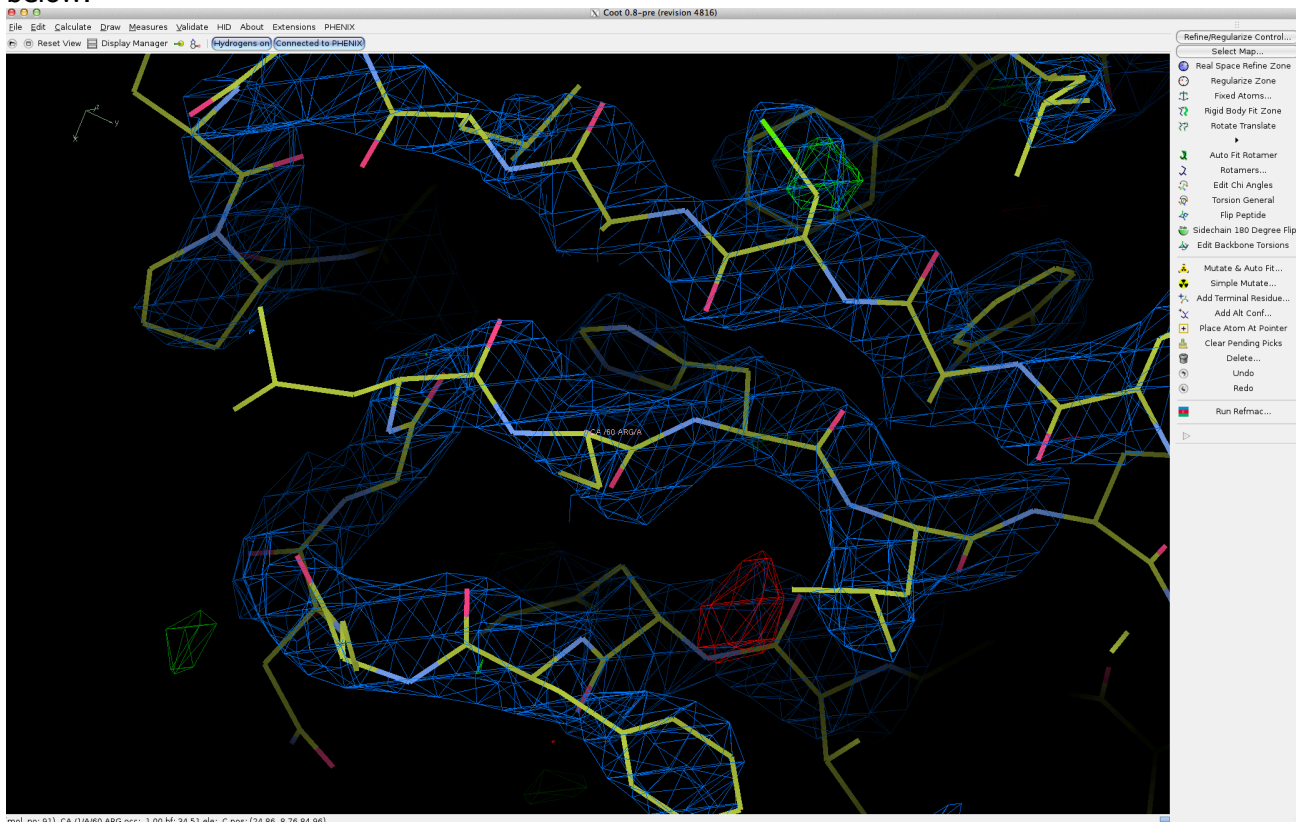
About MolProbity | Website for the Richardson Lab | Using ecloud x-H | Internal reference 4.02b-509

For the MolProbity use the Evaluate X-ray structure under the Walk-thrus & tutorials. FOLLOW the instructions.

If MolProbity says the structure is not good, then repeat the refinement step, along with the manual build in COOT and the comprehensive validation.

## **COOT SECTION:**

The figure below shows the electron density map for the protein. You need to make the sequence fit into the blue or green parts of the map. To do this follow the steps provide below.



### **Step 0: Opening a Structure in Coot**

Click on 'Open Coords...' and select the .pdb file you received in the above steps. It will probably either be in an Xtriage folder or a Refine folder. You can also do this by clicking on 'File' then 'Open Coordinates' and selecting your .pdb file. This will open the chain of amino acids.

You will then need to open up the electron density map. To do this, click on 'File,' then 'Auto Open .mtz' and select the .mtz file that does NOT contain data in the name.

Go to 'Draw,' then 'Cell and Symmetry,' 'Yes,' 'Show,' and 'Apply' to show the unit cells.

### **Step 1: Basic Coot**

Right clicking and scrolling will allow you to zoom in and out of your structure.

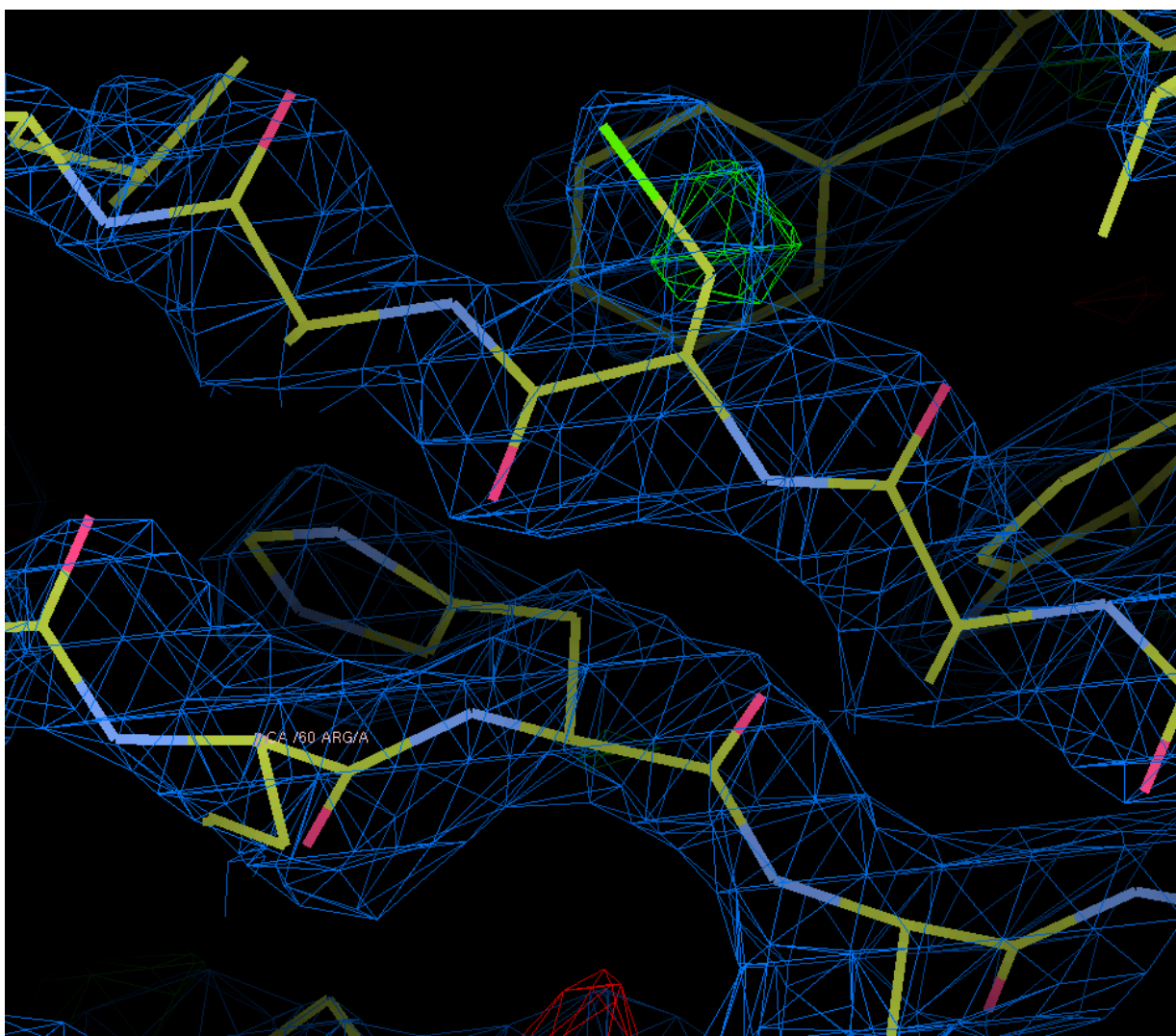


Left scrolling will allow you to change how much electron density it shown. The numerical value will be displayed at the top of the structure picture. You should operate within the 1.5-2 range of these values.

Left clicking on an amino acid will label it. To remove labels, go to 'Measures,' 'Distances and Angles' and click 'Clear All Atom Labels.'

Left clicking and dragging will allow you to rotate the x-, y-, and z-axes around the center of your screen, which is represented by a pink square in the center of Coot.

Holding down the control button on your keyboard while left clicking and dragging will allow you to move the center of your screen.

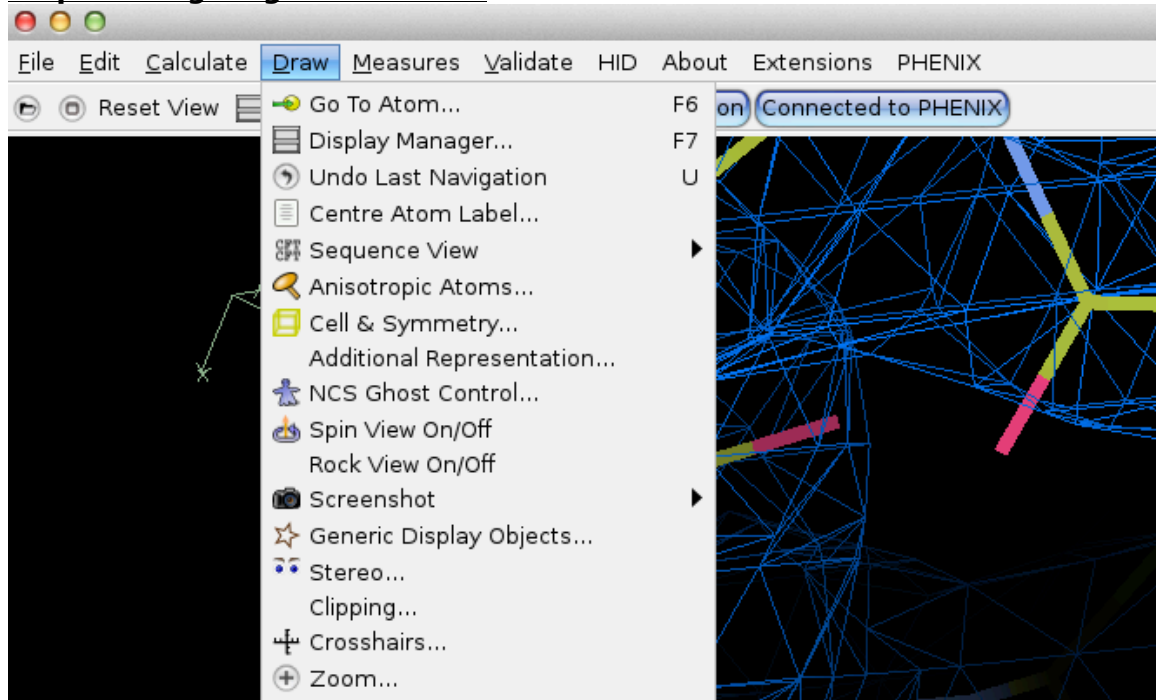


When you are manually fitting the sequence into the density map, you want to get the sequence into the blue or green density as shown above. When there is red density, this

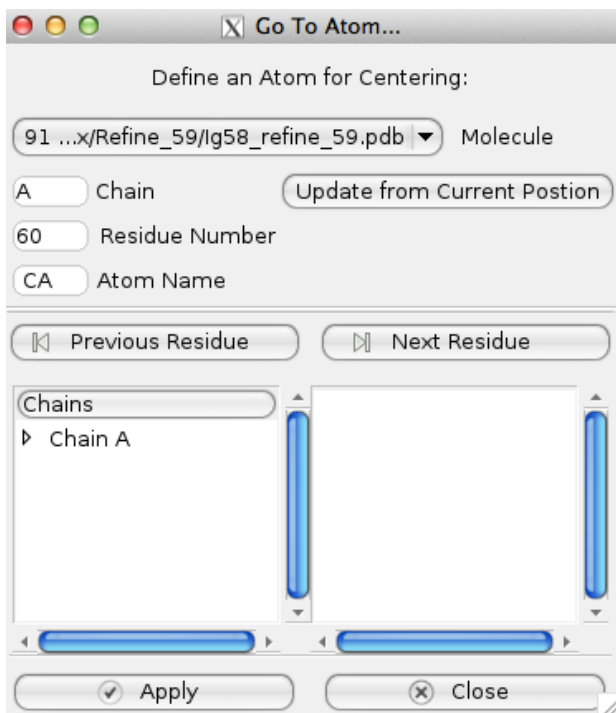
means that the sequence is not favorable to be in that area. You want to AVOID the red density\*. The blue density is the occupied area that is acceptable.

\*When there is red density within your blue density, to use Dr. Berndsen's words, "cry." There is not much you can do here except use the 'Real Space Refine Zone' feature to move it out of the way as best as possible. This feature will auto fit your amino acid to the electron density map.

## Step 2: Navigating the Structure



In able to move from amino acid to amino acid, select the 'Go To Atom' under the 'Draw' pulldown menu. This will allow you to move throughout the sequence one amino acid to the next. If there is more than one chain, it may be necessary to designate which one you want to navigate through.



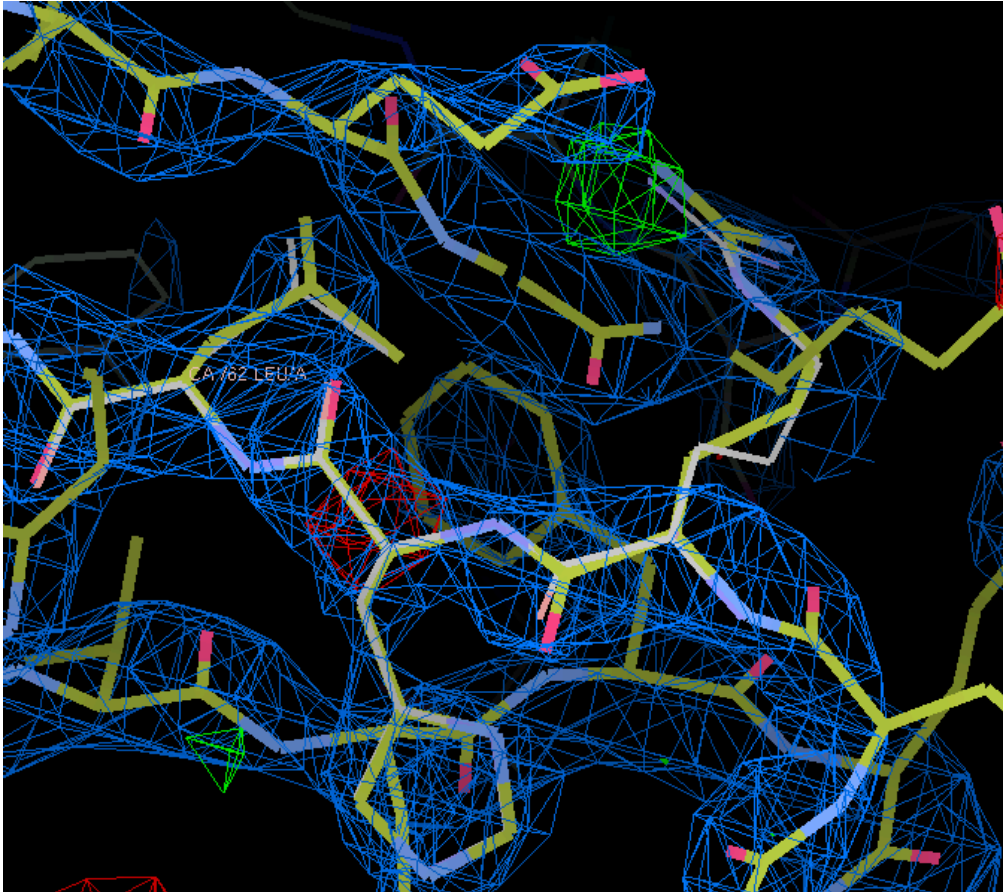
The 'Go To Atom' screen looks like the figure above. You can move forward in the chain by clicking "Next Residue". To move backward click on "Previous Residue."

An easier way to do this is to go to 'Draw' then 'Sequence View' and your .pdb file. This will display the entire chain in a window. Moving this window to the side will allow a quick jump from amino acid to amino acid. Clicking on the desired amino acid number will place you at that spot.



### **Step 3: Real Space Refine Zone**

The 'Real Space Refine Zone' may be the most helpful tool that Coot has to fit your structure. This will fit your amino acid to the electron density and allow you to position it the way you want.

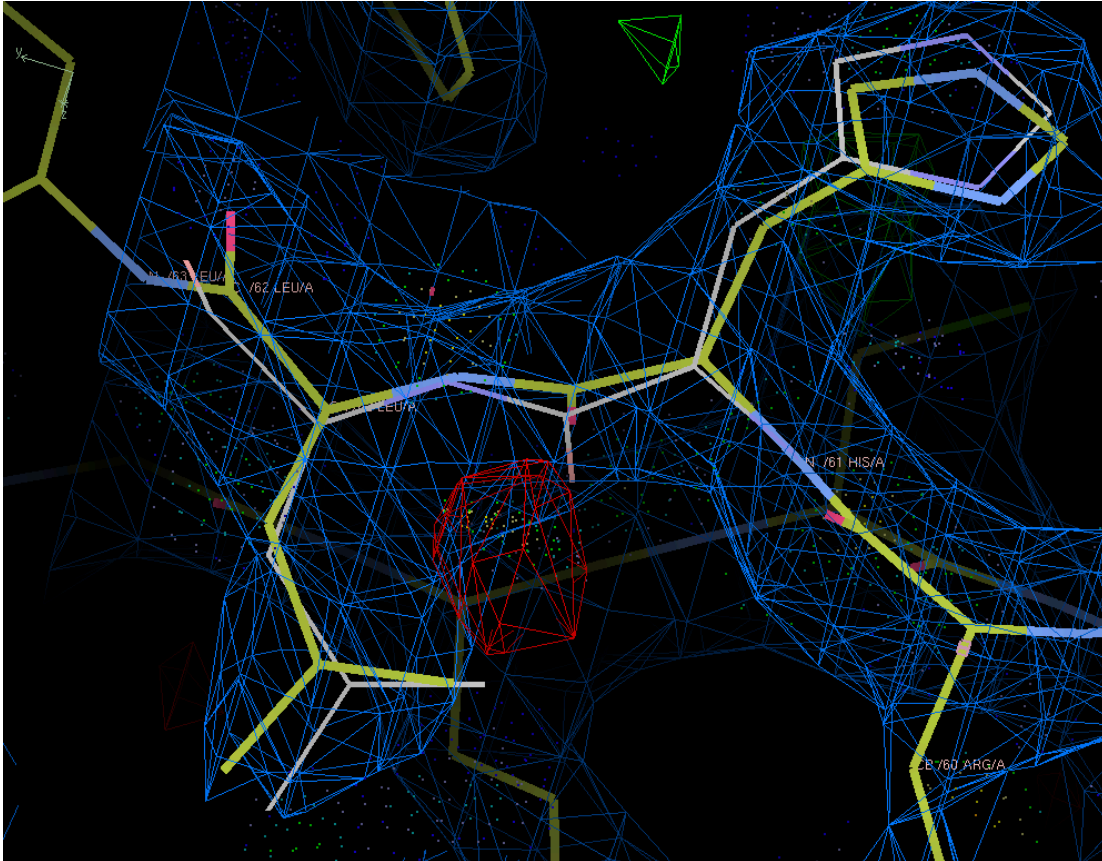


The 'Real Space Refine Zone' will auto fit an area that you select. This will get the angles and bonds in optimal orientations for the sequence. To select one amino acid to refine, double click on some part of the amino acid. To select a larger region, single click on each side of the region. . Dragging an amino acid or region will allow you to finagle the region where you want it. It is easiest to grab the region by clicking and dragging from a "corner" or "bend" of the region. You will then have to accept or reject the refinement. Ideally, all of the categories below will be green.



Once the bonds and angles are in the optimal orientation press the accept button to keep the changes you made on the structure. Use discretion to determine whether or not an orientation is suitable. It is more important to match an amino acid to the given electron density than it is to have the orientation be optimal (have everything be green).

There will be many green and blue regions surrounding the atom where waters, calcium, etc. will eventually go. Do not panic if you see these seemingly random regions go unfilled.



'Regularize Zone' is similar to 'Real Space Refine Zone' but may make more drastic changes. If you are having a lot of difficulty with a region, this may be good place to go as more of a last resort and may point you in the right direction.

#### **Step 4: Mutate and Auto Fit**

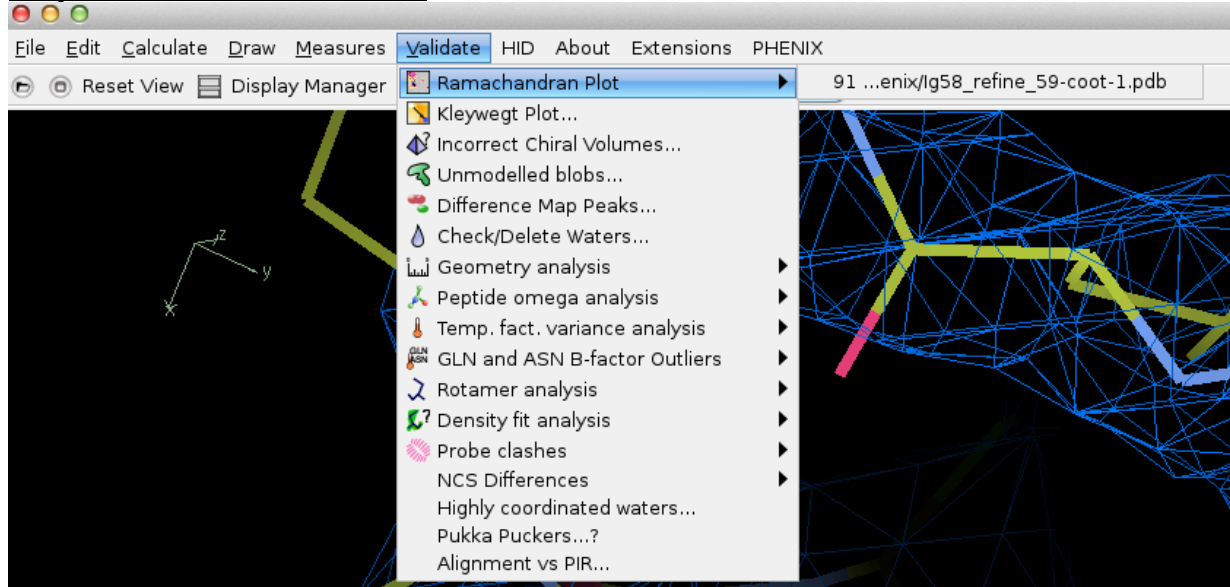
If you want to change an amino acid, use the 'Mutate & Auto Fit' button from the side menu that is shown above then select the amino acid you want to change it to. This can also be done with the 'Simple Mutate...' feature, but this will not fit the amino acid to your electron density.

#### **Step 5: Add Terminal Residue**

If there is an amino acid missing from the sequence you can add in an amino acid by using the 'Add Terminal Residue' button. After clicking it, click on the terminal carbon of your structure and then on 'Accept' in the window that pops up. Make sure that the added amino acid (it will be faded grey) is inside the electron density on your screen (you can move it by clicking on the grey amino acid and dragging it) before hitting 'Accept.'

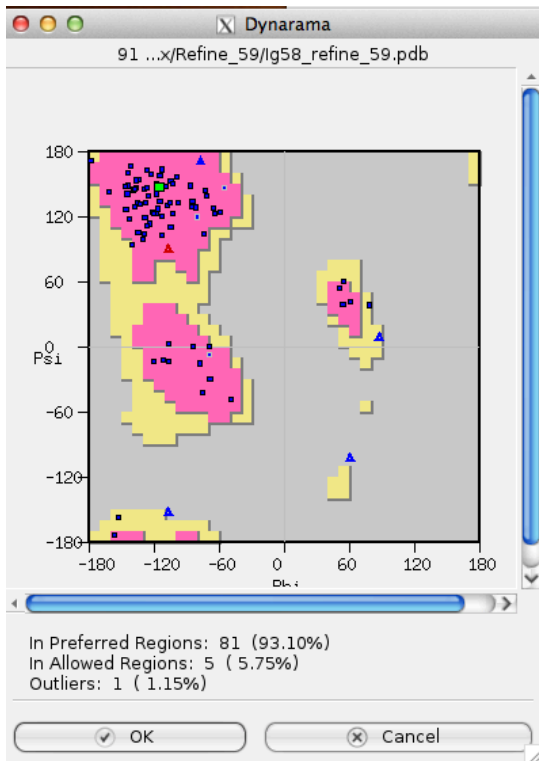
If the amino acid appears and is not connected to the rest of your chain, simply click 'Real Space Refine Zone' and select an area containing the terminal residue and the newly added amino acid. The amino acid should then connect.

### Step 6: Ramachandran Plots



To view the Ramachandran Plot, go to Validate, then Ramachandran Plot, then click on the .pdb file that you would like to view in the Ramachandran plot. You need to check the Ramachandran plot before you run the comprehensive validation because to have a good structure meaning that all the amino acids are in the pink regions on the Ramachandran plot.





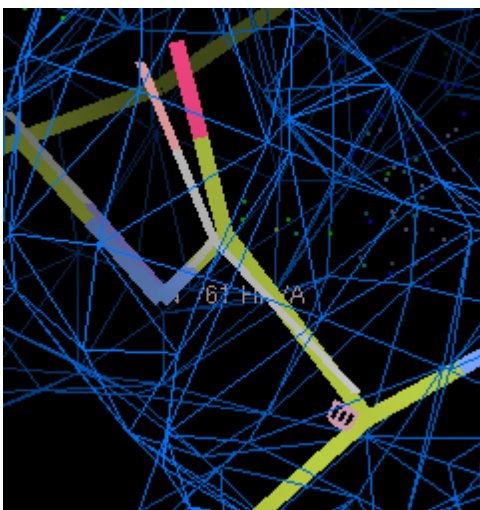
As shown in the Ramachandran plot, all the residues need to be in the pink regions (ideally). If they are not you need to move the residue into the pink region, or at least the yellow region. When hovering your mouse over a square or triangle, it will show which amino acid that shape relates to. Those amino acids with squares have different Ramachandran favored regions than those with triangles.

For a structure to be considered good, you need to have zero "Ramachandran outliers" and minimal "in allowed regions" (maximum in "favored" regions). Once you have checked this plot than you can go forward with validating your structure.

A red shape means that the amino acid is a Ramachandran outlier, and should be fixed. A blue shape means that the amino acid is not an outlier, although it may or may not be in the favored region. A green shape represents the amino acid you currently have your screen's focus on. Note that fixing an amino acid may involve moving the amino acids immediately before and after it. To fix or move an amino acid's configuration, you can use the 'Real Space Refine Zone,' 'Regularize Zone,' or 'Edit Backbone Torsions.' All of these can be found on the right side of the Coot window, and 'Edit Backbone Torsions' may be the most helpful.

### **Step 7: Edit Backbone Torsions**

Once in the 'Edit Backbone Torsions' window, you can either rotate from the peptide or the carbonyl. Note that this will change the Ramachandran plot for both the amino acid you are rotating and the one(s) around it. Try to get both shapes in the desired pink region, but make sure that they stay within the electron density.



### **Step 8: Check/Delete Waters**

When you have refined a structure using the 'Update Waters' feature in the refinement settings, the program will place waters in spots they do not actually belong. One way to check whether or not these are likely water spaces, click on 'Validate,' then on 'Check/Delete Waters' and 'OK.' This will bring up a small window in which questionable waters are displayed. Check to see if the waters in these spots are too close or have too small of an electron density to constitute true waters. Note that there may be more incorrect waters than those in the questionable water window and that not all of the questionable waters may be incorrect. Use your best abilities to discern whether or not there should be waters in each place.

