Application of (3+1)D based motion restraints to improve the refinement of a superspace approximation of an incommensurately modulated protein crystal

Jeffrey J. Lovelace[1], Garib Murshudov[2], Václav Petříček[3], Gloria E. O. Borgstahl[1]

[1]*Eppley Institute for Research in Cancer, Omaha, NE, USA*

[2]*MRC Laboratory of Molecular Biology, Cambridge, UK*

[3]*Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic*

gborgstahl@unmc.edu

Regular protein structure refinement is a large problem. Incommensurately modulated protein structure solution is orders of magnitude more challenging. Crystals grown from a complex of profilin and actin (PA) (~4000 total atoms in the complex) can be incommensurately modulated [1]. Work has been ongoing to solve this structure. We have successfully collected and processed three datasets [2] and assigned the superspace group [3]. Unfortunately, there is no direct approach for modulated protein crystals available at the moment. One option is to treat the reflections and structure solution and refinement as a supercell approximation. As long as the supercell approximation is close to the q-vector then the resulting refined structure will be close to the actual one. The number of smaller cells that make up the supercell represents the number of samples that will be observed along the atomic modulation function in higher dimensional space in this case (3+1)D. As long as the sampling rate (from the supercell) is higher than the Nyquist frequency of the expected atomic modulation function it should be possible to reasonably fit the atomic modulation function using the points sampled by the supercell. For PA a supercell of 7x is used resulting in a q vector of 0.286 which is within 6% of the measured q-vector of 0.270.

Although you cannot make predictions about the actual shape of the modulation function it is possible, under some situations, to restrict the shape of the modulation function. In the case of PA, only first order satellites are observed which indicates that the modulation functions should be smooth. As a first guess at the modulation function a simple sine wave can be used.

How can this be implemented?

One of the most commonly used refinement software tools for protein crystallography is Refmac [4]. Refmac is fast and provides a variety of options for refining protein structures but it doesn't, at the moment, have any options for higher dimensional approaches. As a proof of concept a small R-script was written to restrain the atomic motions in (3+1)D space to sinusoids. The average structure was found by using the program Phaser [5] followed by several cycles of Refmac to fit the mains. In this case, the intensity for the main reflections for the average structure were calculated by adding the intensity of the main and the intensity of its satellites. The average structure was expanded by 7x in the x_2 direction and served as the starting position of the refinement. This was found to produce a slightly more favorable starting position for the supercell than simply using the intensity of the mains based on the R and R_{free} values from the resulting supercell refinement. After the script restrained the atoms in the supercell to sinusoidal motions, Refmac was run for several cycles of refinement against the restrained structure. Multiple cycles of R followed by Refmac were executed where (other than the initial run) the output from the previous Refmac cycle was used as the input to the R script. This approach is similar to a manager giving an employee a task and then periodically checking on the state of that task and making corrections to it before it goes off in an unwanted direction.

How to approximate the atomic modulation functions?

Atom positions from the supercell were transformed into their corresponding x_4 coordinates and these coordinates are translated into the x_4 range of 0 to 1. For each atom the positions were represented as displacements from the average position along each of the principle axes (x_1, x_2 and x_3). Once these matrices of data were assembled the DFT was applied and the first harmonic (amplitude and phase) was extracted representing a simple sinusoidal motion. The first harmonics were used to calculate new displacement values for each atom in all three axes. The positions were then transformed back into normal 3D space and the atomic position data was updated in the structure file which was given to Refmac.

Results

The script was fast (1.2 seconds) relative to the five cycles of Refmac (several minutes) for each global refinement adding around a minute of extra time to the 50 global cycles that were done. Total refinement time took a couple of hours on a marginally fast machine (i5-4570 3.20GHz, 8 GB, 64 bit Linux).

Refinements are represented as rainbow plots (colored by the t value of the center of mass of the protein chain), with only some of the carbon backbone plotted, Fig. 1 shows an unrestrained (in the (3+1)D space sense) refinement and Fig. 2 shows the restrained results. The restrained results have a nice smooth motion (black circular arrow in Fig. 2). Although the restrained refinements have a higher R value they have a much better R to R_{free} agreement than the unrestrained refinement. We are only showing an x_2-x_3 projection because the majority of the motion is contained in these two directions as shown by the center of mass plots as a function of x_4 in Fig 3.

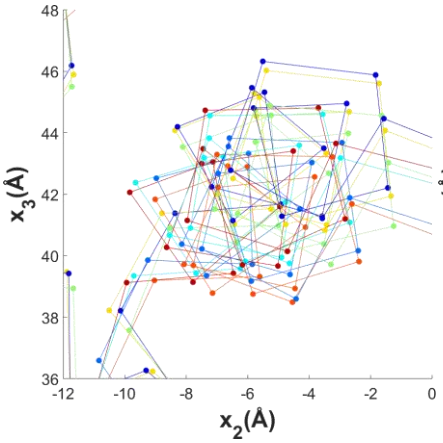

Figure 1. Normal.

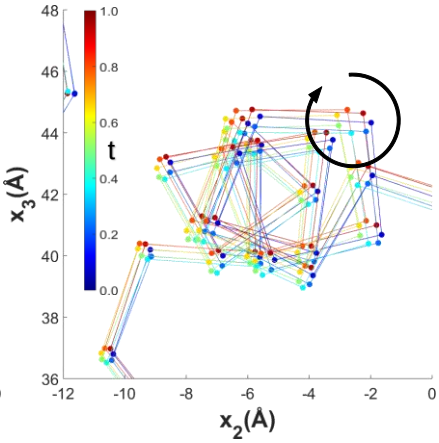

Figure 2. (3+1)D Restraints.

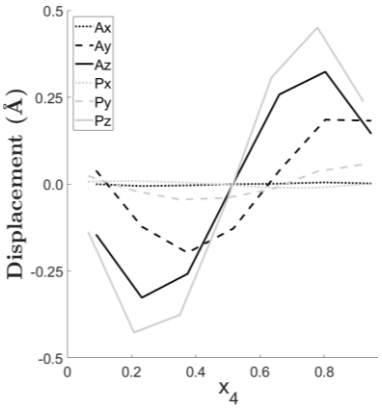

Figure 3. Center of Mass Motion for Actin (A) and Profilin (P).

1. J.J. Lovelace, C.R. Murphy, K. Narayan, C.E. Schutt, U. Lindberg, C. Svensson, M. Wynn, and G.E.O. Borgstahl, *J. Appl. Cryst.*, **41**, (2008), 600.

2. J.C. Porta, J.J. Lovelace, A.M.M. Schreurs, L.M.J. Kroon-Batenburg and G.E.O. Borgstahl, *Acta Cryst.*, **D67**, (2011) 628.

3. J.C. Porta, J.J. Lovelace, G.E.O. Borgstahl, *J. Appl. Cryst.*, **50**, (2017), 1200.

4. G.N. Murshudov, A.A. Vagin, E.J. Dodson, *Acta Cryst.*, **D53**, (1997), 240.

5. A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni and R.J. Read *J. Appl. Cryst.*, **40**, (2007), 658.

This work was supported by NSF Grant MCB-1518145 and the Fred & Pamela Buffett Cancer Center Support Grant (P30CA036727).