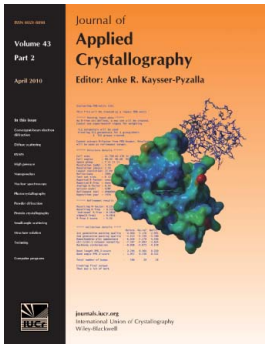
Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Anke R. Pyzalla**

Simulation of modulated reflections

Jeffrey J. Lovelace, Martyn D. Winn and Gloria E. O. Borgstahl

J. Appl. Cryst. (2010). **43**, 285–292



Journal of
Applied Crystallography
ISSN 0021-8898

Received 9 September 2009
Accepted 14 January 2010

Simulation of modulated reflections

Jeffrey J. Lovelace,[a] Martyn D. Winn[b] and Gloria E. O. Borgstahl[a]*

[a]The Eppley Institute for Research in Cancer and Allied Diseases, 987696 Nebraska Medical Center, Omaha, NE 68198-7696, USA, and [b]Computational Science and Engineering Department, STFC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, UK. Correspondence e-mail: gborgstahl@unmc.edu

In a modulated crystal, the repeating unit is not periodic but contains a disorder of the molecules that can sometimes be described with a mathematical function. In the diffraction pattern from such a modulated crystal the standard periodic main reflections are flanked by satellite reflections. Generally, incommensurately modulated crystal diffraction cannot be simply described using integers along the reciprocal-lattice directions. However, in the special commensurate case where the satellite spacing is rational relative to the main reflections, a supercell can be used to describe the modulation. Using a supercell allows structural processing to proceed in a 'normal' fashion but with the downside of dealing with many more atoms. Not much is known about the relationship between a highly modulated macromolecular crystal and the resulting satellite intensities so in this study a modulated protein crystal was simulated using a supercell approach. The protein superstructure was modulated to varying degrees and the resultant diffraction patterns and electron-density maps were studied to understand better how a modulation may manifest itself in real protein data. In the case that was evaluated, relatively small structural modulation resulted in significant satellite intensities. Interesting cases were observed where extinguished main reflections had strong satellites.

1. Introduction

Crystallographic methods for solving periodic crystal structures are well developed. In an ideal periodic crystal (Fig. 1*a*) the unit translations and symmetry operators of the space group replicate the asymmetric unit. If satellite reflections are observed around the main Bragg reflections the crystal is aperiodic. These satellite reflections can be as intense as, and even more intense than, the main spots and result from a structural modulation within the crystals (Fig. 1*a*). These modulated structures can be described with a function that expresses the deviation from translational symmetry and can be thought of as a form of systematic or smoothly varying order to the position or occupancy of the structure (van Smaalen, 2007). Structural analysis of modulated crystals is well developed (van Smaalen, 2004; Petricek & Dusek, 2004) and modulated crystals of small molecules are frequently observed and can be solved. A popular software package for modulated small-molecule crystals is *JANA2006* (Petricek & Dusek, 2006). For macromolecular crystals, satellite reflections from modulated crystals have been reported (Rupp, 2009; Aschaffenburg *et al.*, 1972; Lovelace *et al.*, 2008; Schutt *et al.*, 1989; Pickersgill, 1987) and discussed amongst crystallographic circles, and methods are being developed for solving incommensurately modulated protein structures (Lovelace *et al.*, 2008).

The mathematical approach used to describe multi-dimensional modulations for a diffraction data set is very detailed and complex with excellent descriptions available elsewhere (van Smaalen, 2007; Janssen *et al.*, 1999). Here, the much simpler four-dimensional case is described. In this case, only one modulation is considered where main reflections are flanked on either side by satellites along a common direction (Fig. 1*b*). The normal h, k and l indices are used to describe main reflection positions and a single $\mathbf{q}$ vector is added to describe the satellite positions [equation (1)]. The m term in equation (1) is an integer that describes the order of the satellite reflection. The satellites that are closest to the main reflection are first order ($m = \pm 1$) and the next closest are second order ($m = \pm 2$) *etc.*, as shown in (Fig. 1*b*).

$$\mathbf{H} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* + m\mathbf{q}, \quad (1)$$

$$\mathbf{q} = q_1\mathbf{a}^* + q_2\mathbf{b}^* + q_3\mathbf{c}^*. \quad (2)$$

The $\mathbf{q}$ vector consists of fractional indices describing the spacing of the first-order satellite reflections relative to the main reflection [equation (2)]. Modulations come in two varieties – commensurate or incommensurate. The type of modulation within the crystal can be distinguished by the spacing of the satellite reflection from the main reflection. For commensurate crystals, all components of the $\mathbf{q}$ vector are rational, while for incommensurate crystals, at least one component is irrational (Fig. 1*a*). The direction of the modulation is given by the direction of the satellite reflections. For the commensurate case, the modulation is a special type of

superlattice; the distortion is smoothly varying and can be described with an integer multiple relationship to the main lattice (Fig. 1*a*). Commensurate crystals can be indexed normally by three integer indices and then solved as a supercell of the original unit cell. The supercell can be many times larger than the main unit cell. Currently there is no software to deal with incommensurately modulated protein crystals. The primary obstacles for using *JANA2006* for proteins are that the phasing methods, display of electron density and stereochemical refinement of structures are not appropriate for protein data; however, we are collaborating with the software authors to adapt *JANA2006* for our modulated protein crystals. Therefore, in this study we used the supercell concept to explore the effect of commensurate modulations of protein structure on X-ray diffraction using existing protein crystallography software.

1.1. Selection of a model system

Our current research objective is to solve the structure of incommensurately modulated profilin:actin (PA) crystals. By solving this one structure, a collection of software tools would be developed so that other modulated protein structures could be solved, instead of their data being left as an interesting observation in the back of a thesis (Chik, 1996). So, in order to understand better how a modulation in the PA crystal will manifest in satellite reflections modulated supercells were simulated. An example of incommensurately modulated diffraction from PA crystals is shown in Fig. 1(*c*). Normally, the satellites are weaker than the main reflections but sometimes the satellites are much stronger than the main (Fig. 1*c*). In this study, the size of structural modulation needed to observe satellites as well as the effects on the satellite-to-main intensity ratio were tested. Incommensurately modulated PA has a single **q** vector of $\mathbf{q}_1 = 0.29\ldots\mathbf{b}^*$ (Lovelace *et al.*, 2008). The nearest rational number that could be used to approximate PA's **q** vector with the smallest denominator is 2/7 (0.285). Here, over seven unit cells in the **b** direction there are two repeats of the modulation. So, to simulate PA's modulation a sevenfold supercell in the **b** direction would be required. With 514 residues in the asymmetric unit this supercell would contain 14 392 residues and 56 chain IDs. Thus, to reduce the size of the problem and perform the simulation with reasonable processing speed, a similar crystal with fewer atoms and chains was selected. The dendrotoxin from the venomous green mamba snake, commonly known as ToxD, was used (Skarzynski, 1992). ToxD conveniently has an identical space group to PA but only 59 residues in one chain for the asymmetric unit.

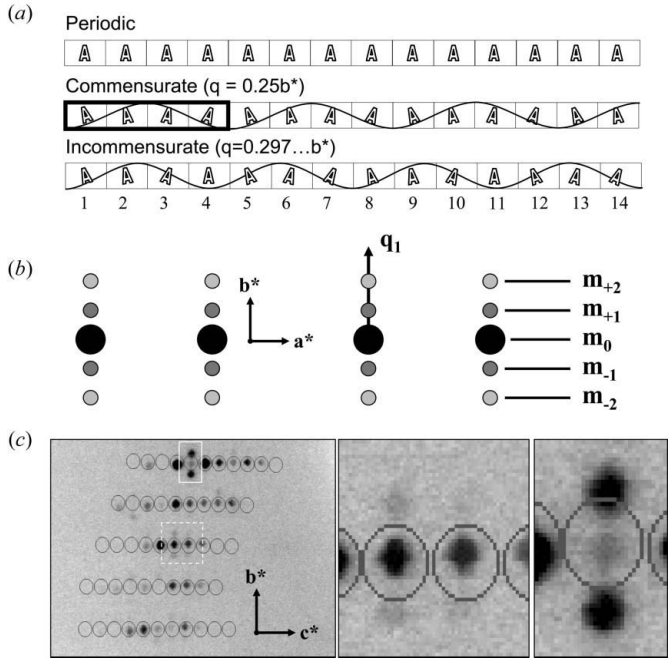


Figure 1
Introduction to modulated diffraction. (*a*) Three types of crystal lattices (periodic, commensurate, incommensurate); (*b*) incommensurate 3 + 1 diffraction model with second-order satellites along the **b*** direction; (*c*) actual incommensurately modulated profilin:actin (PA) data showing **b*** satellites (white dashed box) with some satellites having a higher intensity than the main reflection (white box).

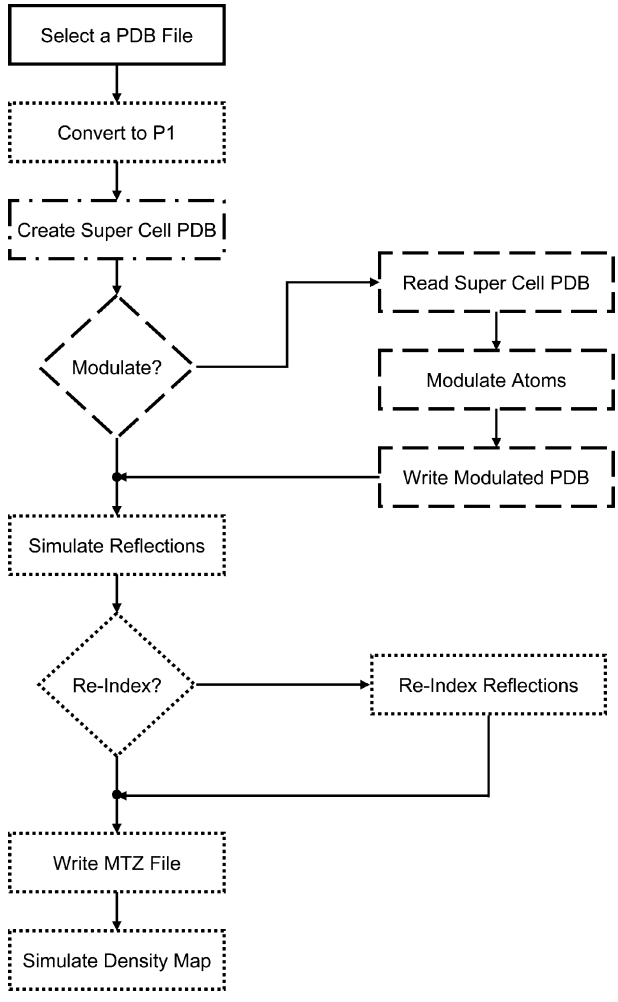


Figure 2
Process flow chart for the simulation of a modulated protein crystal. Line type indicates the processing software involved (solid = none, dotted = CCP4, long dash = *MATLAB*, dotted long dash = both).

2. Method of simulation

The simulation was performed using CCP4 software (Collaborative Computational Project, Number 4, 1994) and custom *MATLAB* scripts (The Mathworks, 2007) using the process outlined in Fig. 2. The following CCP4 programs were used: *PDBCUR*, *SFALL*, *REINDEX* and *HKLVIEW*. The first step was to select a PDB (Protein Data Bank; Berman *et al.*, 2000) file of atomic coordinates (in this case ToxD, PDB code 1dtx; Skarzynski, 1992). The next step was to convert to a *P*1 space group using *PDBCUR*. Next, *PDBCUR* was used to extend the unit cell by seven times in the **b** direction (Fig. 3). The atoms in the PDB file are stored in such a format that the atoms forming the proteins are listed in a continuous fashion from the N- to the C-terminus. The atoms are spanning unit cells while never completely filling a single unit cell. Before the modulation was carried out, the atoms were repositioned so that they were all inside the unit-cell boundaries before the modulation function was applied. Additionally, all atoms were converted to nitrogen, which is roughly representative of the average scattering power of the protein, and stripped of chain and amino acid information to simplify reading, writing and modifying the PDB files. This also makes it possible to get around some of the file limitations in the PDB format when you have a large number of chain IDs. All of the steps used to extend a PDB file to a supercell representation are summarized in Fig. 3. The modulation was carried out using *MATLAB*. The appropriate atoms (based on the modulation function) were modulated and a new modulated PDB file was written. *SFALL* was used to calculate structure factors from the modulated supercell PDB file. Optionally, *MATLAB* and *REINDEX* were used to change point groups or extract and reindex the main reflections, in order to evaluate what effect using the average structure approach would have on the electron density in regions containing the modulation. Results were visualized with a combination of *MATLAB* (atoms), *Coot* (atoms and electron density; Emsley & Cowtan, 2004) and *HKLVIEW* (reflections).

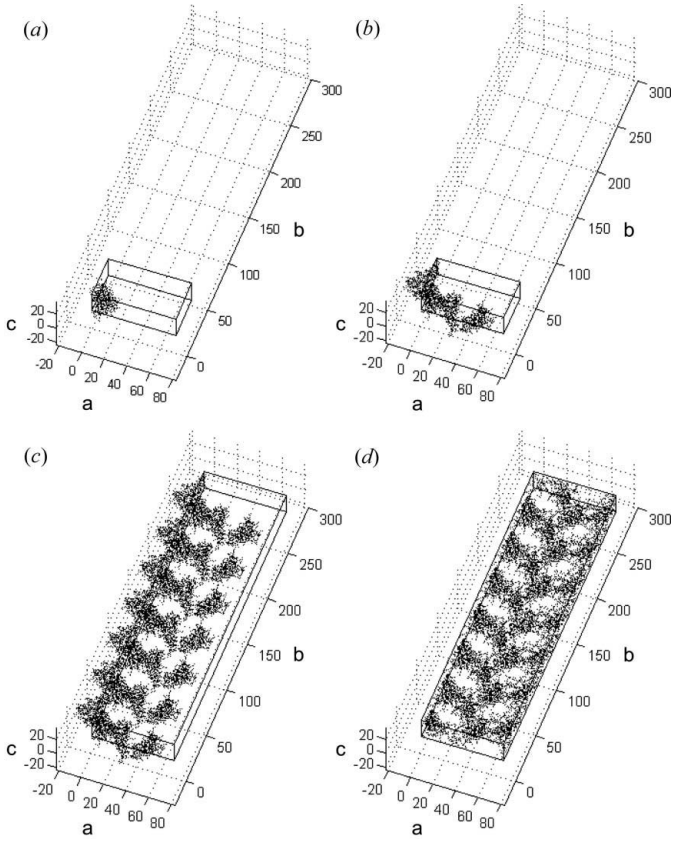


Figure 3
Generation of the supercell for the simulation. The (*a*) original ToxD cell in $P2_12_12_1$ (a = 73.5, b = 39.0, c = 23.1 Å) is expanded to the ToxD supercell along the **b** direction in *P*1 (A = 73.5, B = 273.0, C = 23.1 Å). The three steps include (*b*) the generation of symmetry-related atoms; (*c*) expansion along the **b** direction by seven unit cells; and (*d*) all atoms are placed in the supercell.

2.1. Modulation function

Existing protein supercell structures typically include translational symmetry that involves a subunit taking on multiple conformations. Instead of manipulating a subunit by hand to create the modulation in the supercell, a mathematical modulation function was used. Modulation functions are assigned to individual atoms to solve incommensurately modulated small-molecule crystals. The difference for this case is that a global function was used to cause an overall modulation of the superstructure. It is also advantageous to use a global function to describe the modulation as it can easily be scaled to any supercell based on any PDB file. For PA it was hypothesized that the superstructure is twisted along the **b** direction (Schutt *et al.*, 1997). This twist could be a function of distance along axis *b* with the atoms moving in the *ac* plane only. This hypothesis was considered in the selection of the global modulation function.

The modulation function was required to be consistent at the boundaries of the unit cell and provide a modulation that was orthogonal to the direction of the modulation. A twisting function was designed that rocks the atoms back and forth, similar to the agitator motion of a top-loading washing machine (Fig. 4; see also supplementary material[1]). This function carves out a cylinder in the unit cell. Along the length of the cylinder are an integer number of sine waves with a minimum of one. The value of the sine function indicates the strength and direction of the rotation of atoms in the plane as a function of **b**. The function also uses a half sine wave extending from the center to the edge of the cylinder along the radius. This forces the motion to a maximum at half the radius of the cylinder. Rotation is relative to a user-defined maximum rotation. The rotation for any atom in the modulation cylinder is given by the following equation:

$$\theta_{\text{modulation}}(r, y_0) = \frac{\theta_{\max}}{2} \sin\left(\frac{r}{R_{\max}}\pi\right) \sin\left(\frac{y_0}{B} 2\pi n\right) \qquad (3)$$

[1] A movie of the agitator motion discussed in this paper is included as part of the supplementary material available from the IUCr electronic archives (Reference: HE5459). Services for accessing this material are described at the back of the journal.

for all r such that $r \leq R_{max}$, where $r = [(x_0 - A/2)^2 + (z_0 - C/2)^2]^{1/2}$. θ_{max} is the maximum amount of angular oscillation; n is the number of modulation waves per supercell; x_0, y_0, z_0 are the position of an atom within the supercell; a, b, c the unit-cell parameters; A, B, C the supercell parameters; R_{max} the radius of the modulation function; r the perpendicular distance to the modulation axis; and $\theta_{modulation}$ the amount of angular motion to modulate an atom.

The new coordinates (x, y, z) of the atoms (x_0, y_0, z_0) that are modulated by a rotation about the b axis are given by the following equations:

$$x = \cos(\theta_{modulation})(x_0 - A/2) + \sin(\theta_{modulation})(z_0 - C/2) + A/2, \quad (4)$$

$$y = y_0, \quad (5)$$

$$z = -\sin(\theta_{modulation})(x_0 - A/2) + \cos(\theta_{modulation})(z_0 - C/2) + C/2. \quad (6)$$

There is no change in the y position because the motion is perpendicular to the axis of modulation in this case.

A routine, *PDBRead.m* (Nikumbh, 2002), was initially used to manipulate PDB files in *MATLAB*. This implementation had very poor performance on large PDB files (several minutes to read and write a PDB), so more efficient *MATLAB* routines were created resulting in PDB reading and writing performance of less than 1 s. The atoms were modulated through additional *MATLAB* scripts which implemented the washing machine modulation. All of the *MATLAB* routines are available as supplementary material.

3. Results and discussion

3.1. Effect of rotation range

The effect of rotation range on diffraction was initially analyzed by visual inspection of simulated intensities for the appearance of satellite reflections. In this analysis the modulations are reported in terms of θ_{max} or the theoretical maximum for angular atom movement. On average, the actual motion of any atom is much lower. Simulated intensities were created for a range of modulation angles, spaced logarithmically (0.0, 1.1, 1.7, 2.5, 3.8, 5.8, 8.7, 13.2, 19.8, 29.9 and 45.0°);

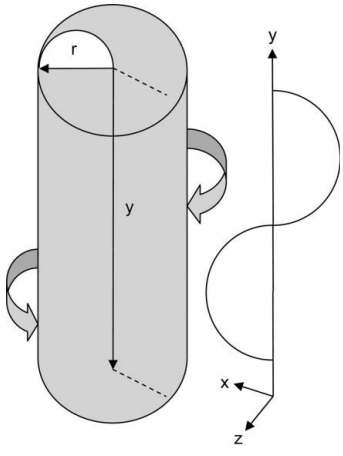


Figure 4
Washing machine function used to simulate a modulated protein structure. The position of an atom (x, y, z) determines the direction and amount of modulation (shift in atomic position that is applied) as described in equations (3)–(6). This function rocks the atoms back and forth in a manner similar to a top-loading washing machine.

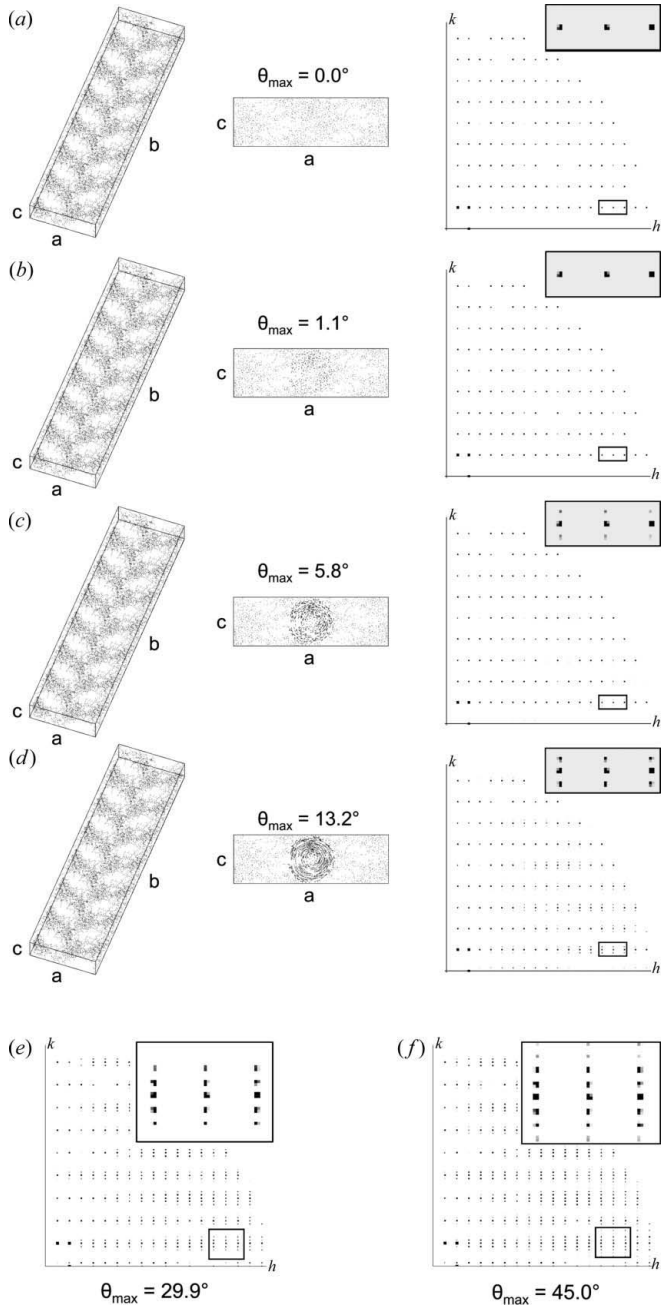


Figure 5
Modulated supercell and corresponding simulated diffraction pattern for the *HK*0 plane. Examples of the simulated data calculated over a range of modulation angle are shown as follows: (*a*) normal supercell with θ_{max} of 0°; (*b*) θ_{max} of 1.2°; (*c*) θ_{max} of 5.8°; (*d*) θ_{max} of 13.2°; (*e*) θ_{max} of 29.9° (note, second-order satellites visible); (*f*) θ_{max} of 45.0° (note, third-order satellites visible). The middle image shows the *ac* plane of the supercell. Here all the atoms along the entire depth of the **b** direction of the supercell are shown so it is easier to see the effect of the modulation function on the structure, in this direction, even for small values of θ_{max}.

each had one oscillation cycle over the entire supercell ($n = 1$). Fig. 5 shows the modulated supercell and simulated intensity pattern for several cases. The effect of the modulation on the structure is not visible in the top view (Fig. 5, left image) but looking down the b axis of the sevenfold supercell (middle image) makes it very obvious where the modulation is occurring even for very small values of the modulation functions. The first panel (Fig. 5*a*) has no modulation and therefore no satellite reflections. The modulation is visible when looking down the b axis with θ_{max} of 1.1° (circular outline of the modulation function, Fig. 5*b*) but there are no visible satellites in the simulated intensity pattern. It is not until θ_{max} is set to 5.8° that the satellite reflections are visible (Fig. 5*c*). The satellite reflections fill in from high resolution to low resolution as the modulation is increased (Fig. 5*d*). Second- and third-order satellites are visible at θ_{max} of 29.9° (Fig. 5*e*) and 45.0° (Fig. 5*f*), respectively.

3.2. Effect of number of modulations per supercell on diffraction

From the **q** vector determined for PA it was expected that there would be about two modulation cycles ($n = 2$) for every seven unit cells. To verify the effect the number of modulation waves would have on reflection spacing, the number of modulation waves per unit cell was increased from one to two. Fig. 6(*a*) shows a single modulation over the supercell at a modulation of 13.2° and Fig. 6(*b*) shows a double modulation over the supercell. Similar reflection spacing to that observed in the actual PA crystal is observed in the double modulation case. Effectively, the first-order reflections are extinguished and only the second-order reflections are visible. As expected, the satellite-to-main spacing goes from 1/7 to 2/7 relative to the next main reflection along the b direction.

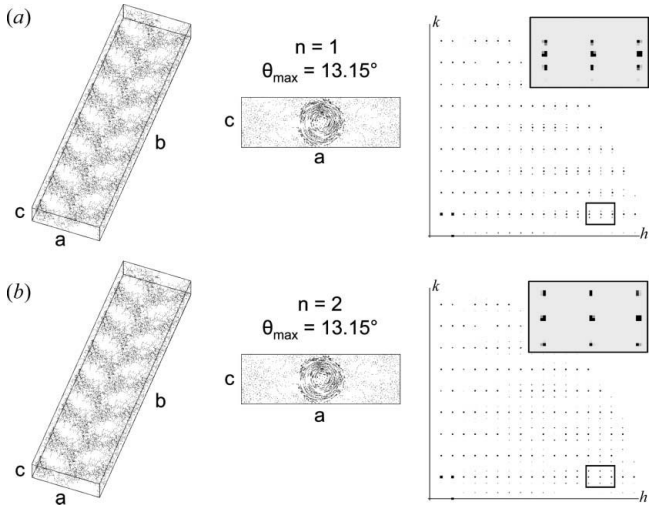


Figure 6
The effect of the number of modulation waves per unit cell (n) on diffraction. In part (*a*) $n = 1$, and in (*b*) $n = 2$. The structural modulation was for θ_{max} of 13.15° for both parts. The primary difference between the two parts is that the reflection spacing goes from 1/7 to 2/7 as n goes from 1 to 2, as expected.

3.3. Angular rotation *versus* satellite-to-main intensity ratio

Is there a relationship between the amount of modulation and the ratio of satellite-to-main intensity? This is a common question that has arisen multiple times during discussions. In order to answer this question, it is first necessary to reindex the supercell reflections using the mtz (CCP4 reflection) file as a starting point. Main reflections were selected as the reflections whose k index was divisible by 7. This was determined using the modulo operator. Satellite reflections were selected by searching for cases where the same operator was applied with a 1 added to all **b** indices and a 1 subtracted from all **b** indices. Since the reflections were generated based on a resolution cut-off (2.3 Å), the pool of reflections was further trimmed to include only mains that contained both satellites. Intensity ratios were calculated by averaging the satellite intensities together and then dividing this intensity by the main reflection. The average ratio was plotted against θ_{max}. Points were measured every 0.1° from 0 to 45° (Fig. 7). The relationship appears to follow a monotonically increasing function, possibly logarithmic in overall shape. This monotonic relationship is disrupted by peaks at 25, 38, 35 and 44°. The appearance of what seem to be resonant peaks was unexpected. If it was not for these resonant peaks, then there would be a simple correlation between θ_{max} and satellite-to-main ratio.

To help isolate the cause of the unexpected peaks the data were graphed in a different manner. For each modulation the ratios where grouped into a histogram. Several ranges of binning were used. Initially all ratios were binned for values falling between 0 and 1 with the initial assumption that there would be no values greater than 1 (the case where the satellites were stronger than the main reflections). Bin frequencies were assigned colors. The colors were assigned as follows: blue for a frequency of 1 to red for the maximum binned frequency over all modulations. Bins with no count were assigned white.

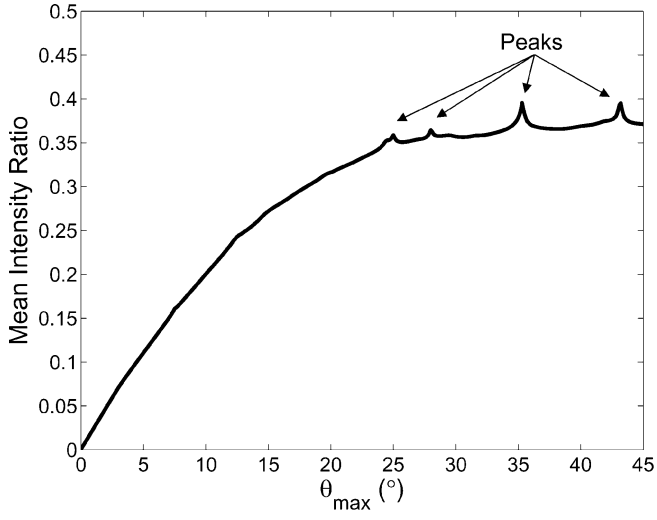


Figure 7
Analysis of the satellite-to-main reflection ratio (mean ratio) *versus* the amount of structural modulation (θ_{max}). The arrows point to peaks that appear superimposed on an otherwise positive monotonic function.

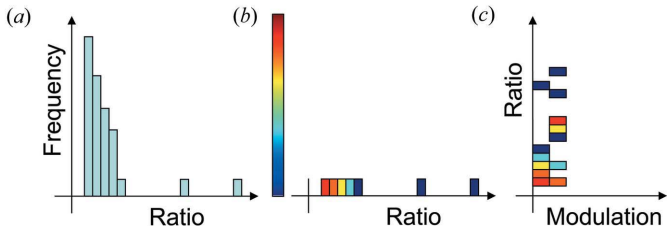


Figure 8
Method used to transform the satellite-to-main reflection ratio data to a histogram representation of the binned ratio frequency data *versus* the amount of structural modulation (θ_{max}). The first step (*a*) is to bin the ratio data by frequency. This frequency binning was done for each of the simulated θ_{max} values. Next (*b*) the frequency values are replaced by a color code (white, no counts; blue, 1 count; to red, maximum counts). Finally (*c*) the ratio histograms for all θ_{max} angles are combined on a single plot where location and color show the number of reflections with a given ratio value at a θ_{max} angle.

The histograms were then plotted vertically for each modulation, resulting in a three-dimensional picture of bins to modulation with the third-dimension, frequency, assigned a color. The process is shown in Fig. 8. The results using a binning range of 0 to 1 are shown in Fig. 9(*a*). Fig. 9(*a*) shows that as θ_{max} increases both the average ratio increases as well as the deviation in ratios. It was still not clear from this data why peaks were observed in Fig. 7. One clue is visible in the upper left of Fig. 9(*a*) as there seems to be some structure to the data in the low-frequency bins extending beyond a ratio of 1. In order to explore this possibility the histogram data were recalculated by extending the binning range from 0 to 50 instead of the original 0 to 1. The results are shown in Fig. 9(*b*). At this scale there are a lot of interesting patterns for low-frequency bins in the histogram data that were simply not visible in the 0–1 range. Initially, we thought these peaks were

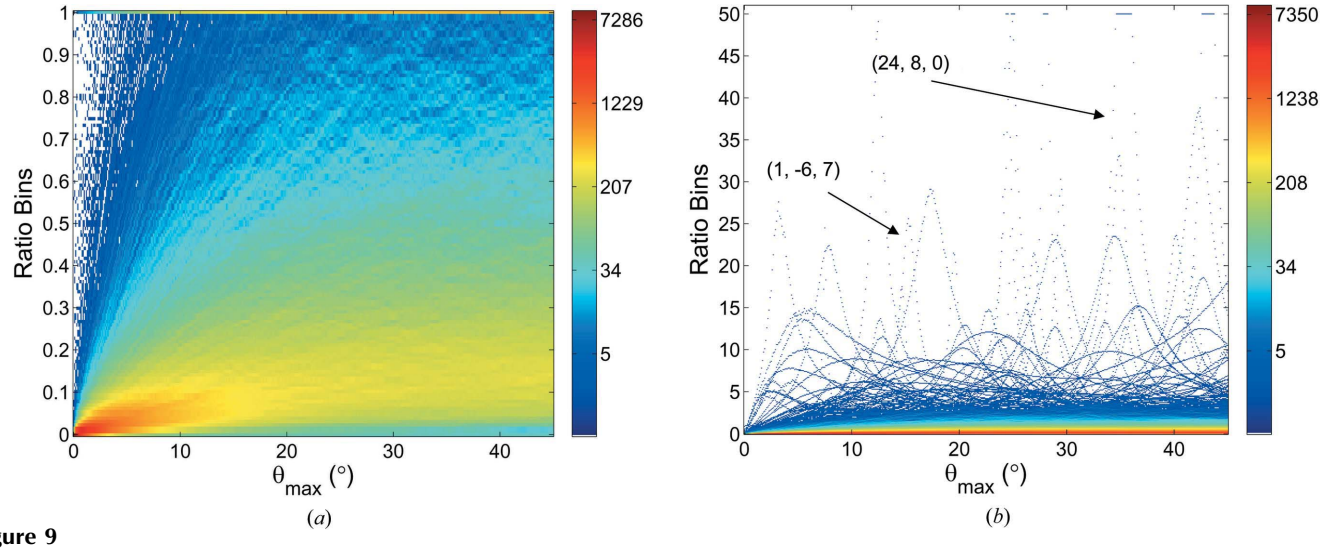


Figure 9
Analysis of the satellite-to-main reflection ratio *versus* the amount of structural modulation (θ_{max}) using histograms. Part (*a*) shows ratio data binned from 0 to 1; part (*b*) shows ratio data binned from 0 to 50. The arrows point to two reflections selected for detailed intensity plots (Fig. 10).

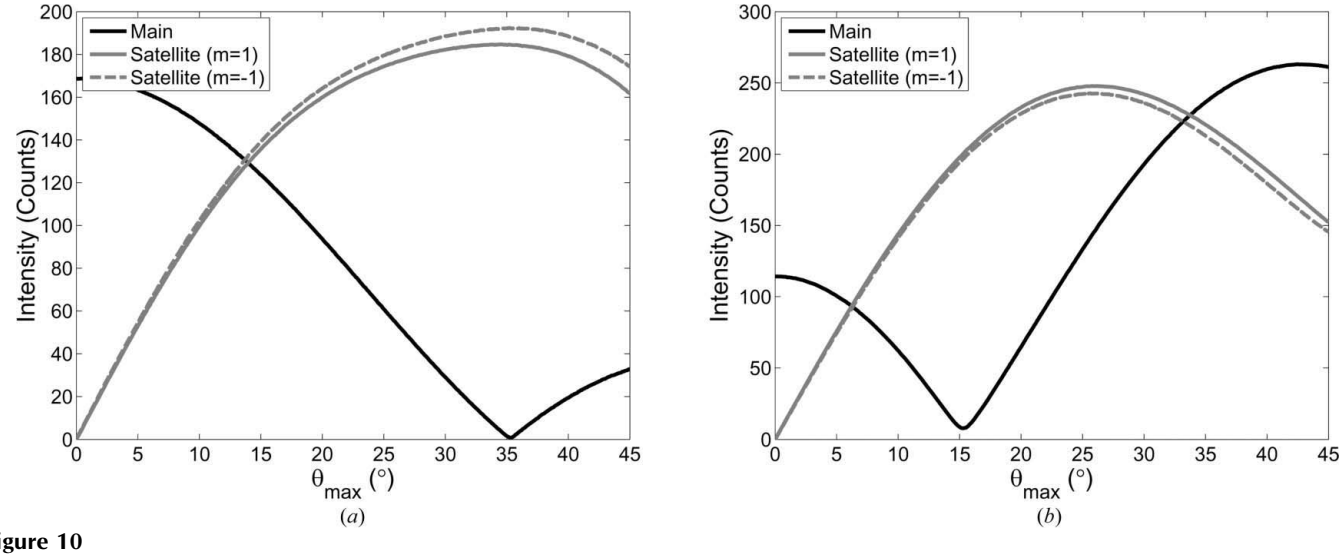


Figure 10
Intensity profiles *versus* θ_{max} for two reflections with large (>1) satellite-to-main ratios. In part (*a*) the reflection (24 8 0) shows an example where the satellite reflection intensities peak when the main reflection is extinguished. In part (*b*) the reflection ($1\bar{6}7$) shows an example of a lower-intensity main reflection where the peak of the satellite intensities does not coincide with the minimum of the main reflection.

individual reflections (dark-blue color) that had very strong satellite intensities. They were not caused by systematic absences because those reflections were filtered out.

Two peaks were selected (arrows in Fig. 9*b*) from the histogram data for further study to understand better what occurred. Searching the ratio data using the histogram to limit the search proved that each peak was the result of intensity modulations for only a single reflection. Plots of the intensities for two examples are shown in Fig. 10. Reflection (24 8 0) in Fig. 10(*a*) had a maximum ratio of 265.3 at θ_{max} of 35.5°. For this case, the primary contributor for the large and sharp increase in ratio is that the main reflection is nearly extinguished. Additionally, the peak in the satellite intensity coincides with the extinguished main. Reflection ($1\bar{6}7$) in Fig. 10(*b*) had a maximum ratio of 26.2 at θ_{max} of 15.5°. The primary contributor is a reduction in main intensity but it is not extinguished. Also, for this case the peak in satellite intensity for this reflection does not coincide with the reduction of the main reflection. Thus, studying the satellite-to-main intensity ratio is a good way to find cases where the satellites become more intense than the main reflection.

Another way to estimate the amount of modulation is described by Lam *et al.* (1992). In this paper, a method of estimating the modulation is found by fitting a three-parameter function to the reflection data (grouped in resolution shells) as a function of *d* spacing. One of the fit parameters is an estimation of the modulation in the structure. Since this method uses a large number of reflections, averages them and then fits a function, it avoids the problems of dealing with cases that have a very high ratio.

3.4. Average structure analysis

When working with small-molecule modulated data the first step in determining the solution is to solve the average structure. An average structure is calculated in the normal unit cell by using only the main reflections. By analyzing thermal ellipsoids, it is normally possible to get an understanding of what atoms are involved and overall how they are moving in the structure (Koutentis *et al.*, 2001). In a protein structure, how will the modulated areas manifest themselves in the electron-density map? To investigate this process, the main reflections were selected from the modulated supercell reflections and reindexed into a smaller unit cell by compressing the **b** dimension back to its original size. The main reflections were selected by using only reflections where **b** was divisible by 7. The new reflections were reindexed in **b** by dividing the original index by 7.

Electron-density maps were created from both normal data and the mains from the structure simulation that was modulated at 13.2°. Fig. 10 shows several different views of the electron-density maps. Blue/black density is shown for the normal data and purple/gray density for the modulated data. The maps were made at a σ level of 2.0 using *Coot*. Looking down the *c* axis as shown in Fig. 11(*a*) there is no noticeable difference between the two densities (about half and half). Along the *a* axis (Fig. 11*b*) the region of modulation is clearly

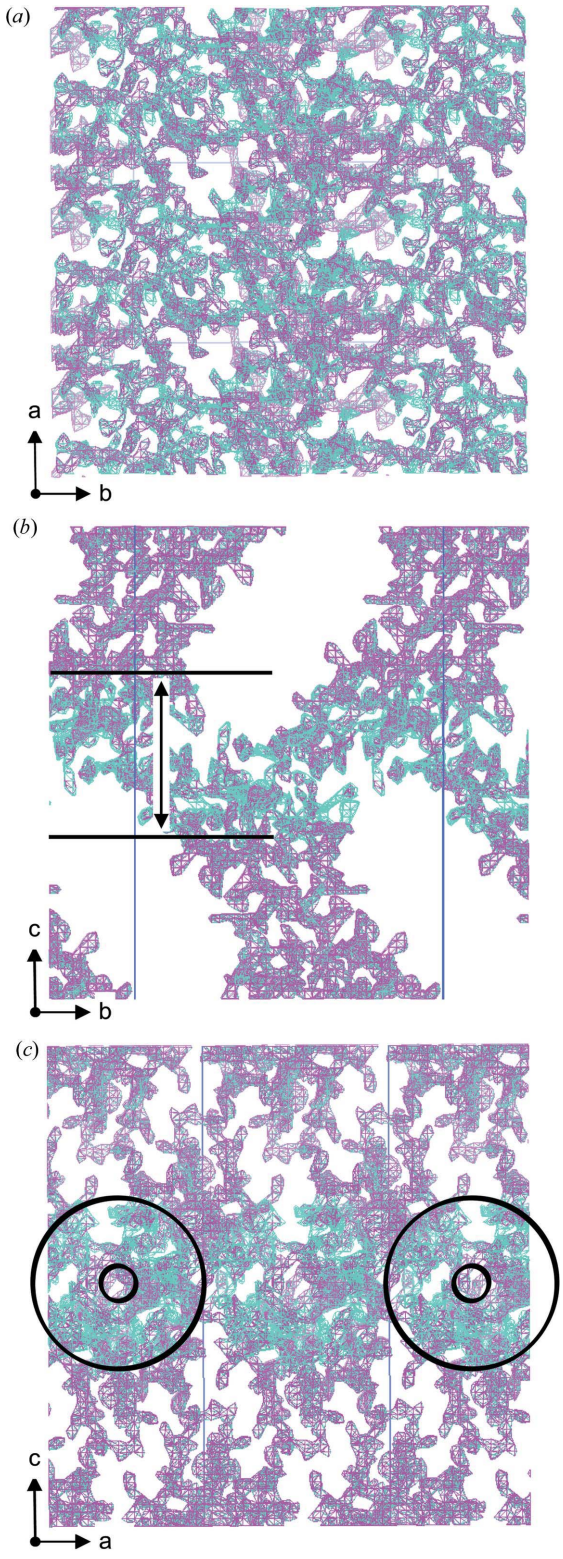


Figure 11
Electron-density comparison of normal (blue/black) with averaged modulated data (purple/gray). In part (*a*) the map is displayed looking down the *c* axis and the effect of the modulation is not visible. In part (*b*) the map is displayed looking down the *a* axis where a double arrow indicates the region where the modulation is along the entire **b** direction but only spans about a third of the **c** direction. In part (*c*) the map is displayed looking down the *b* axis showing the donut shape of the modulation function.

visible with the normal density dominating over the modulated density where the modulation function was applied (highlighted middle section in Fig. 11*b*). Along the *b* axis (Fig. 11*c*) the donut shape of the modulation function is clearly visible. From these results, the conclusion is that the modulated areas will have a lower overall electron density in the average structure when compared with the normal electron density. This lower density is due to the fact that we are considering a positional average of the modulated structures. Therefore, much as small-molecule crystallographers use displacement ellipsoid size and shape to help identify the modulated atoms in the average structure, it should be possible to use electron-density variations to help visualize the modulated regions of protein structures and gain some insight into the appropriate modulation function. Also, analysis of the average electron density may help the X-ray data-to-parameter ratio for structural refinement. For example, *JANA2006* (Petricek *et al.*, 2000), the incommensurately modulated structure-solving program, uses up to three modulation functions per atom to accurately refine a modulated, small-molecule structure. For the refinement of a modulated protein structure, the number of parameters could be reduced by identifying which atoms are and are not significantly modulated and applying the modulation function(s) accordingly.

4. Conclusions

Current protein crystallography tools were used to evaluate commensurately modulated data sets using a global modulation function applied to a supercell. Satellites become visible with only a very small amount of modulation (5.79°, Fig. 5*c*). One indicator of a potentially large amount of modulation would be large numbers of reflections whose satellite-to-main ratio is much greater than one, as shown in Fig. 9(*b*). Since this case was observed in actual modulated PA diffraction data multiple times, this would indicate that the modulation will be large. In difficult cases, where a supercell does not completely predict the diffraction image or the resultant electron-density maps are uninterpretable, it may be useful to perform a **q**-vector analysis of the reflections. **q**-Vector analysis will help assign the proper supercell or can determine if the structure is actually incommensurately modulated. Finally, in much the same way that displacement ellipsoids refined for average structures are used to determine the location of the modulation for small molecules, electron-density strength may be used to determine areas that are significantly modulated in a protein crystal. Here, the modulation relationship is inversely proportional to the density. Additionally, this information should help in the determination of incommensurate protein structures by allowing identification of the atoms that are strongly modulated and application of the modulation function to only those regions, thereby increasing the X-ray data-to-parameter ratio.

This work was supported and funded by NSF grant No. 0718661 and the Nebraska Research Initiative. We would like to thank Clarence E. Schutt (Princeton University) and Uno Lindberg (Stockholm University) for useful discussion. The PA incommensurate diffraction image is courtesy of Jason C. Porta.

References

Aschaffenburg, R., Fenna, R. E. & Phillips, D. C. (1972). *J. Mol. Biol.* **67**, 529–531.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalow, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Chik, J. K. (1996). PhD thesis, Princeton University, USA.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Janssen, T., Janner, A., Looijenga-Vos, A. & Wolff, P. M. D. (1999). *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson & E. Prince, pp. 899–947. Dordrecht: Kluwer Academic Publishers.
Koutentis, P. A., Haddon, R. C., Oakley, R. T., Cordes, A. W. & Brock, C. P. (2001). *Acta Cryst.* B**57**, 680–691.
Lam, E. J. W., Beurskens, P. T. & Smaalen, S. V. (1992). *Solid State Commun.* **82**, 345–349.
Lovelace, J. J., Murphy, C. R., Daniels, L., Narayan, K., Schutt, C. E., Lindberg, U., Svensson, C. & Borgstahl, G. E. O. (2008). *J. Appl. Cryst.* **41**, 600–605.
Nikumbh, S. (2002). *PDBRead*. http://www.mathworks.com/matlabcentral/fileexchange/1288.
Petricek, V. & Dusek, M. (2004). *Z. Kristallogr.* **219**, 692–700.
Petricek, V. & Dusek, M. (2006). *JANA2006*. Institute of Physics, Czech Academy of Sciences, Prague, Czech Republic.
Petricek, V., Dusek, M. & Palatinus, L. (2000). *JANA2000.* Institute of Physics, Czech Academy of Sciences, Prague, Czech Republic.
Pickersgill, R. W. (1987). *Acta Cryst.* A**43**, 502–506.
Rupp, B. (2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. New York: Garland Science.
Schutt, C. E., Kreatsoulas, C., Page, R. & Lindberg, U. (1997). *Nat. Struct. Biol.* **4**, 169–172.
Schutt, C. E., Lindberg, U., Myslik, J. & Strauss, N. (1989). *J. Mol. Biol.* **209**, 735–746.
Skarzynski, T. (1992). *J. Mol. Biol.* **224**, 671–683.
Smaalen, S. van (2004). *Z. Kristallogr.* **219**, 681–691.
Smaalen, S. van (2007). *Incommensurate Crystallography*. New York: Oxford University Press.
The Mathworks (2007). *MATLAB*. Version 2007b. http://www.mathworks.co.uk/.