research papers

Are the St John's wort Hyp-1 superstructures different?

Jeffrey J. Lovelace and Gloria E. O. Borgstahl*

Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, 986805 Nebraska Medical Center, Omaha, NE 68198-6805, USA. *Correspondence e-mail: gborgstahl@unmc.edu

Two commensurately modulated structures (PDB entries 4n3e and 6sjj) were solved using translational noncrystallographic symmetry (tNCS). The data required the use of large supercells, sevenfold and ninefold, respectively, to properly index the reflections. Commensurately modulated structures can be challenging to solve. Molecular-replacement software such as *Phaser* can detect tNCS and either handle it automatically or, for more challenging situations, allow the user to enter a tNCS vector, which the software then uses to place the components. Although this approach has been successful in solving these types of challenging structures, it does not make it easy to understand the underlying modulation in the structure or how these two structures are related. An alternate view of this problem is that the atoms and associated parameters are following periodic atomic modulation functions (AMFs) in higher dimensional space, and what is being observed in these supercells are the points where these higher dimensional AMFs intersect physical 3D space. In this case, the two 3D structures, with a sevenfold and a ninefold superstructure, seem to be quite different. However, describing those structures within the higher dimensional superspace approach makes a strong case that they are closely related, as they show very similar AMFs and can be described with one unique (3+1)D structure, *i.e.* they are two different 3D intersections of the same (3+1)D structure.

1. Introduction

Commensurately modulated structures are a unique subset of entries in the Protein Data Bank (PDB). The diffraction patterns from these types of structures are distinct in that they consist of strong main reflections with, on average, less intense satellite reflections around the main reflection. Indexing software can usually index the main reflections, but in some cases can have difficulty indexing all of the reflections. For commensurate cases, the indexing software will lock into a solution that is some multiple of the basic unit cell, resulting in a supercell. This supercell will be made of an integer multiple of basic cells along one or more of the unit-cell directions. An excellent approach to solving these structures has successfully been implemented by using the theory behind translational noncrystallographic symmetry (tNCS) to enable molecular-replacement programs (McCoy *et al.*, 2007; Read & McCoy, 2016) to arrive at solutions where they may have been unable to do so in the past.

Our laboratory has been working on the incommensurately modulated structure of the profilin–actin complex (Lovelace *et al.*, 2008; Porta *et al.*, 2011, 2017) and has developed a set of superspace-analysis tools to aid us in the analysis of our refinements. Our incommensurate case has proven to be more challenging because there is not a supercell that will predict
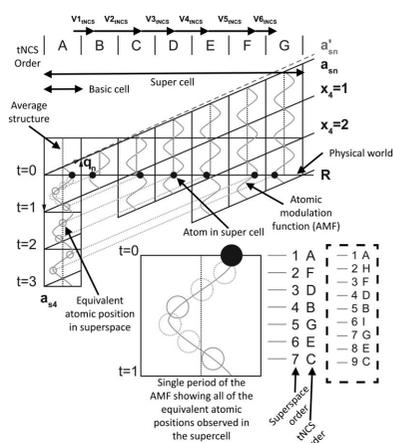
the satellite reflections, so we have had to use an approximation. Additionally, we have a systematic absence in our reflections when the data are represented as a supercell approximation, which is a characteristic of the underlying modulation that has led to some interesting issues that have required some more innovative solutions during refinement (Lovelace *et al.*, 2018). The modulated Hyp-1–ANS complex (PDB entry 4n3e, with a sevenfold supercell) and its cousin (PDB entry 6sjj, with a ninefold supercell) provided us with two distinct large commensurately modulated structures (Sliwiak *et al.*, 2015; Smietanska *et al.*, 2020) to evaluate our superspace-analysis tools. Dusek *et al.* (2003) demonstrated for a small molecule that the correctly chosen superspace group can describe the array of crystalline phases (including commensurate and incommensurately modulated phases) observed for sodium carbonate. In superspace, they were able to observe how the atomic modulation functions (AMFs) were similar and determine which atom was responsible for the transition to the incommensurate phase. Using this paper as a template, we suspected that by observing the commensurately modulated superstructures in superspace it should be possible to visualize how these modulated structures are related. The authors of the Hyp-1 structures speculated that there must be some kind of deeper connection between the structures, but pointed out that it is difficult to detect one by comparing the 3D structures. The primary difference in the crystallization conditions was the addition of melatonin for the PDB entry

6jss structure, which leads to a supercell requiring two more basic cells.

1.1. tNCS

tNCS occurs when the asymmetric unit is made up of a 3D array of identical chains. The chains in this array have slight occupancy, displacement, rotational and/or conformational differences that disrupt the basic cell to basic cell periodic order. If this disruption results in a longer range periodic order, then additional diffraction spots appear, which lead to the need to use a supercell to describe the observed diffraction pattern. This type of case is diagnosed by studying the Patterson map (Patterson, 1935; Fig. 1). For PDB entry 4n3e the Patterson map has non-origin strong peaks about every 1/7 of the unit cell (Fig. 1a), and the Patterson map for PDB entry 6sjj has non-origin strong peaks about every 1/9 of the unit cell (Fig. 1b); these spacings are consistent with an offset of a whole protein molecule when compared with the unit cell along c. This type of protein-sized spacing in the Patterson maps is a strong indication of tNCS. This challenging molecular-replacement problem was solved using tNCS in *Phaser* (McCoy *et al.*, 2007). Details of the tNCS implementation in *Phaser* and of how it was adapted for PDB entry 4n3e can be found in Sliwiak *et al.* (2014). Both structures required an innovative multistep approach to arrive at the final solution. Although the structures were solved, it was difficult to use the
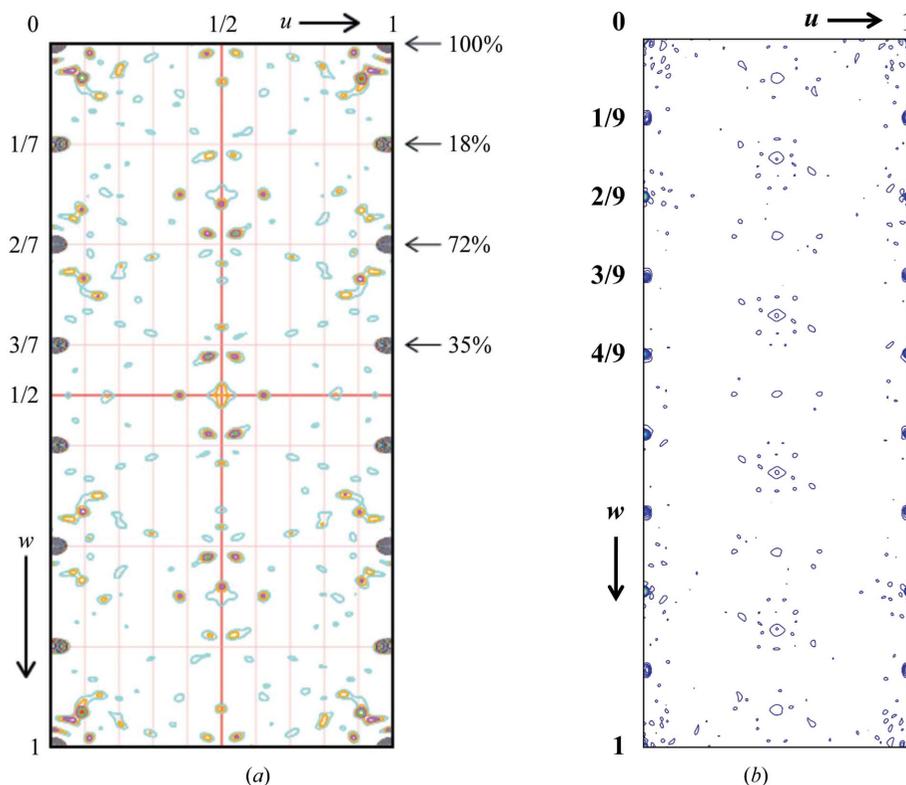
Figure 1
Patterson maps of (*a*) the PDB entry 4n3e data showing the tNCS with a strong peak every 1/7 along the *c* unit-cell direction (*w* Patterson direction; Sliwiak *et al.*, 2015) and (*b*) the PDB entry 6sjj data showing strong peaks every 1/9 along the *c* unit-cell direction. These strong peaks indicate tNCS in the structures. (*a*) is reproduced from Sliwiak *et al.* (2014) and (*b*) was generated from the deposited structure factors using Python with *numpy*, *matplotlib* and *gemmi*.

tNCS data to arrive at an understanding of the underlying modulation. This analysis was complicated by the PDB entry 6sjj structure, which should be related to the PDB entry 4n3e structure but appeared to be unique, requiring an additional two basic cells in the supercell. We decided to investigate whether things may become more evident by looking at the results of refinement from the superspace perspective. Visualization of supercells in superspace is not a new idea and has been successfully used to help to solve incommensurately modulated small-molecule crystals (Schönleber & Chapuis, 2004). These approaches are no longer seen as often in this field due to the ability of *Superflip* (Palatinus & Chapuis, 2007) to now directly solve these types of structures.

1.2. Superspace notation

When the underlying periodic structure of a crystal is broken by a periodic distortion of the structure, the resulting superposition of periodic structure and periodic distortion will produce a diffraction pattern that may not be periodic but will still have long-range order and produce diffraction spots. This case is known as a modulated structure. These structures can be described by superspace (higher than three dimensions). For this paper, we are using a variation of the crystallographic notation that can easily be extended into higher dimensions (van Smaalen, 2005, 2007). Superspace modulations are identified by the number of extra dimensions needed to describe the diffraction pattern: $(3+d)$D, where d is the number of extra dimensions. The simplest of these would be $(3+1)$D. The unit cell is defined by $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ instead of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. The reciprocal-space vectors are $\mathbf{a}_1^*, \mathbf{a}_2^*, \ldots, \mathbf{a}_n^*$ instead of $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$. Distances are measured as the values $\mathbf{x}_1$, $\mathbf{x}_2, \ldots, \mathbf{x}_n$ as opposed to $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Reflections are labelled h_1, $h_2, \ldots, h_n$ instead of h, k, l, m.

1.3. Superspace

Superspace, as developed by Janner and Janssen (Janner & Janssen, 1977, 1980a,b), describes how periodic atomic functions in higher dimensional space can be used to explain the diffraction patterns of modulated structures observed in physical space. This theory provides a very powerful approach that even allows the indexing of diffraction patterns of incommensurately modulated structures and interpretation of the resulting structures. These types of structures have a very distinct diffraction pattern of main reflections with associated satellites. For diffraction, the superspace approach uses the concept of $\mathbf{q}$ vectors to describe the location of satellite reflections relative to their main reflection. The number of $\mathbf{q}$ vectors also indicates the number of extra dimensions needed to describe the modulation. The easiest modulation case is described using $(3+1)$D superspace, where only one $\mathbf{q}$ vector is needed to describe the satellite reflections. The $\mathbf{q}$ vector (or the spacing between the main reflection and the first-order satellite) that describes the diffraction pattern of the commensurately modulated structure observed for the PDB entry 4n3e data can be determined from an analysis of the average reflection intensities arranged as a function of l

(Sliwiak *et al.*, 2015; Fig. 2a). In the same manner, the $\mathbf{q}$ vector for PDB entry 6sjj (Smietanska *et al.*, 2020) can be determined by a similar analysis (Fig. 2b). The intensity relationship for main reflections and satellites is that on average first-order satellites will be less intense than the main reflection, second-order satellites will be less intense than first-order satellites, and so on. Locally, this trend may not hold due to the actual shape of the AMFs. Averaged over many reflections, the trend in reflection intensities will hold. Fig. 2(a) shows that the first-order satellites are spaced three reflections away from each main reflection, and other related satellites and the main reflections are at positions divisible by seven as a function of l. Taken together, main reflections every seven reflections and first-order satellites three reflections away from the main reflections lead to a $\mathbf{q}$-vector definition of $3/7\mathbf{a}_3^*$. Using the $3/7\mathbf{a}_3^*$ $\mathbf{q}$ vector, which can be interpreted as having three periods of the periodic modulation function every seven basic cells, allows the construction of a superspace diagram (Fig. 3) to demonstrate how a periodic AMF in this higher dimensional space (grey wavy periodic lines in Fig. 3) can account for the observed atomic positions (black circles in Fig. 3) in the
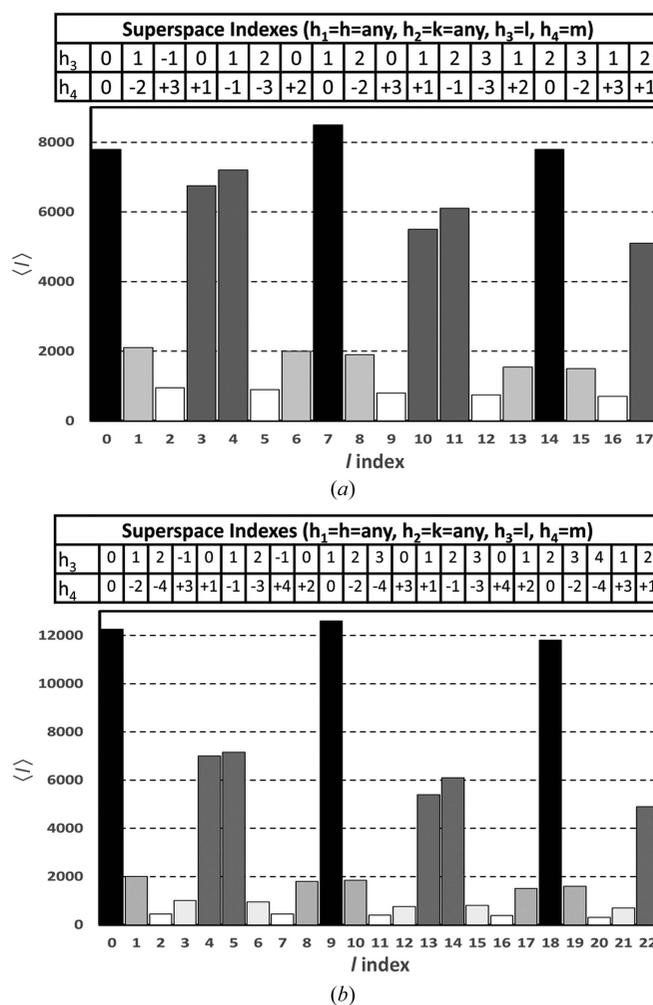
Figure 2
Average reflection intensities as a function of the l index. The superspace indexing of the same average reflection intensities (h_3 and h_4) is shown for (a) PDB entry 4n3e and (b) PDB entry 6sjj. These charts were adapted from Sliwiak *et al.* (2015) and Smietanska *et al.* (2020).

supercell where the AMFs intersect with physical space (**R** in Fig. 3). Due to the periodicity of the AMF, all of the observed atomic positions in the supercell can be translated to a single period of the AMF in t (Fig. 3, enlarged area), where t has a value from 0 to 1 over a single period. In superspace, data are usually shown as a function of t (lines in superspace but parallel to **R**) or x_4 (lines parallel to $\mathbf{a}_{sn}$). The choice depends on the information that is trying to be conveyed. Additionally, it is possible to see that the states encountered in the supercell by traversing the basic cells in physical space (A–G) are not in the same order as encountered in superspace (1–7) (enlarged area in Fig. 3). For (3+1)D superspace, the AMFs are line functions, where the x_4 value of the position is determined by the dot product of the average position of the atom in the basic cell in fractional coordinates with the **q** vector plus a phase offset t,

$$x_4 = t + \mathbf{q} \cdot \bar{\mathbf{x}}, \tag{1}$$

where

$$\mathbf{q} = \sigma_1 \mathbf{a}_1^* + \sigma_2 \mathbf{a}_2^* + \sigma_3 \mathbf{a}_3^*, \tag{2}$$

$$\bar{\mathbf{x}} = \bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2 + \bar{\mathbf{x}}_3, \tag{3}$$

$$\mathbf{q} \cdot \bar{\mathbf{x}} = \sigma_1 \bar{\mathbf{x}}_1 + \sigma_2 \bar{\mathbf{x}}_2 + \sigma_3 \bar{\mathbf{x}}_3. \tag{4}$$

$\bar{\mathbf{x}}$ is the average position of an atom in a basic cell in the crystal and σ_n is the **q**-vector coefficient. All atomic parameters in superspace are governed by AMFs.

In the case of the reflection data for PDB entry 6sjj, the main reflections (highest intensity) appear for values of l that are divisible by 9. The next highest (first-order satellites) appear four reflections away. This leads to a **q**-vector of $4/9\mathbf{a}_3^*$. In this case, there are four periods of the modulation wave for
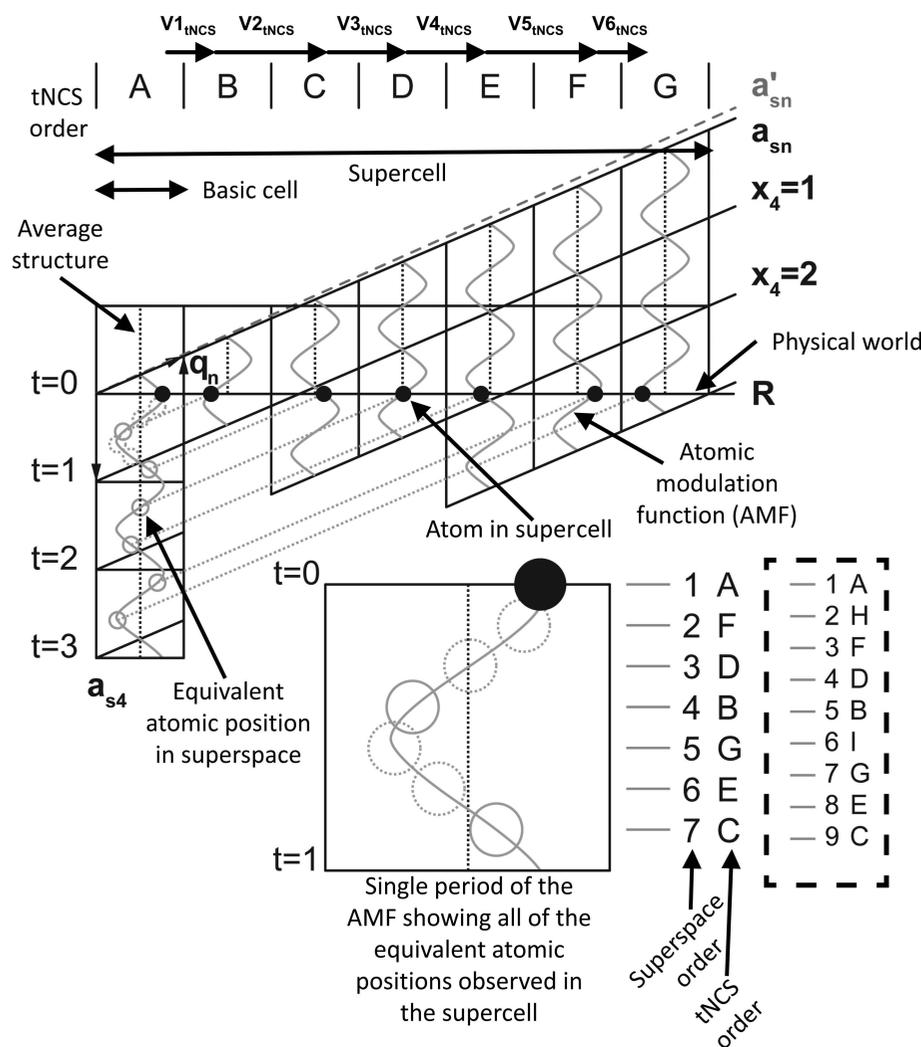
Figure 3
Superspace diagram for (3+1)D superspace with three periods of the modulation function every seven basic cells. A periodic function in superspace is used to describe the seemingly random appearance of tNCS-related atoms in a supercell (along **R**, where **R** represents 3D space in the physical world). These displacements are extremely exaggerated as the modulations are typically small compared with the unit cell. An average structure position is given by the vertical dotted line. The $\mathbf{q}_n$ vector determines the angle between x_4 and physical space ($\mathbf{a}_{sn}$). The enlarged region for $t = 0$ to $t = 1$ shows the reordering of basic cells that occurs in superspace. In the dashed box to the right of the 3/7 reordering is the reordering for the 4/9 supercell. The grey $\mathbf{a}_{sn}'$ line shows the vector that would need to be used for a 4/9 modulation; although not shown, this line would interect at four periods after nine basic cells, leading to a ninefold supercell.

PDB entry 4n3e
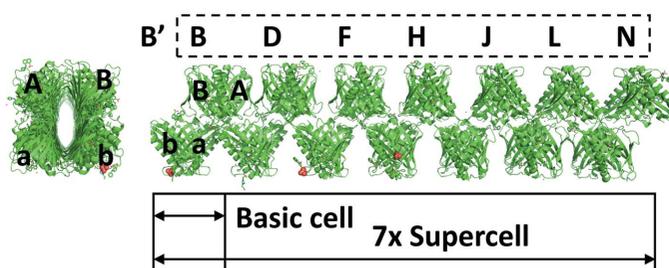
Figure 4
The superstructure of PDB entry 4n3e with views looking down and along the structure. The supercell is a sevenfold expansion in the **c** direction. Chains in the first basic cell (a, b, A and B) use a notation with a prime to denote all related chains in subsequent basic cells; for example, B′ refers to chains B, D, F, H, J, L and N. The PDB entry 6sjj supercell looks similar but has two additional basic cells and uses a slightly different naming scheme for the chains.

every nine basic cells. The reordering of the basic cells in superspace for this 4/9 case is shown in the dashed box in Fig. 3.

2. Methods

Details of the processing will focus on the PDB entry 4n3e structure as an example. The processing is the same for the PDB entry 6sjj structure, with the primary difference being that the supercell is a ninefold expansion and the **q** vector is $4/9\mathbf{a}_3^*$ as opposed to $3/7\mathbf{a}_3^*$. The PDB entry 4n3e structure consists of a sevenfold basic cell expansion into a supercell with four protein chains and some small molecules in the asymmetric unit (ASU) of the basic cell where the modulation was along the $\mathbf{x}_3$ dimension (Fig. 4). The resulting supercell has 28 chains in the ASU. The chains composing the first basic cell are A, B, a and b. All of the chains related by the modulation in other individual basic cells will be annotated with a prime. For example, B′ represents the chains in basic cells 1–7 (B, D, F, H, J, L and N), which are all related by tNCS to the B chain in the first basic cell (Fig. 4). The PDB entry 6sjj structure consists of a ninefold basic cell expansion into a supercell (not shown), and like the PDB entry 4n3e structure has four protein chains in the ASU of the basic cell. In this case, chains j′, B′, i′ and A′ of PDB entry 6sjj correspond to chains b′, B′, a′ and A′ of PDB entry 4n3e. Superspace figures for chain B′ of PDB entry 4n3e have previously been published (Lovelace & Borgstahl, 2020) but the processing details of how they were generated were not discussed.

Matlab (MathWorks) scripts were used to manipulate the PDB files to display the data in superspace. These scripts are available in the supporting information and will generate many of the figures shown in this paper. The process of breaking down a supercell (Fig. 5a) and transforming it into superspace starts with dividing it into basic cells (Fig. 5b). The number of basic cells is the product of the denominators of the rationalized **q** vector. The number of basic cells along any dimension will be equal to the denominator of the **q** vector in that dimension. For PDB entry 4n3e the number of total basic cells is seven because $1 \times 1 \times 7 = 7$ and they are along $\mathbf{x}_3$
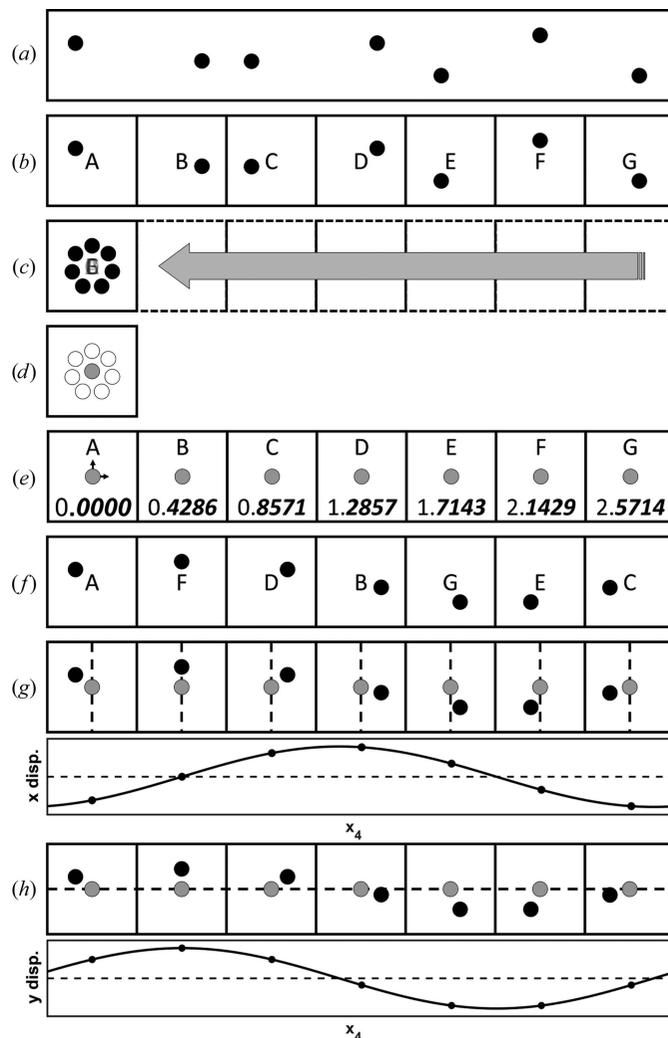
Figure 5
The process of converting a supercell into its subspace representation: (*a*) the supercell, (*b*) partitioning of the supercell into the basic cell, (*c*) translation of all of the basic cells to overlap with the first basic cell, (*d*) calculation of the average positions from the translated positions of related elements, (*e*) using the average position in each basic cell to determine $\mathbf{x}_4$, (*f*) sorting the basic cells by fractional $\mathbf{x}_4$, (*g*) determining horizontal displacements from the average as a function along $\mathbf{x}_4$ and (*h*) determining vertical displacements from the average as a function of $\mathbf{x}_4$.

because that is the only dimension with a denominator ($\sigma_3 = 3/7$) greater than 1. The function *superorder.m* (provided in the supporting information) will return a rationalized **q** vector to within a tolerance given the **q** vector in decimal form. The basic cells are then all translated to the coordinates of a single basic cell (Fig. 5c). An average atomic position (grey dot in Fig. 5d) is calculated averaging the related basic cell atoms (grey circles in Fig. 5d). The average position is used to calculate an $\mathbf{x}_4$ position for that atom in all of the basic cells (Fig. 5e). Sorting can then be performed on the fractional component of $\mathbf{x}_4$ to determine the frame order to animate the basic cell (Fig. 5f). Additionally, plots of the approximated AMFs can be created by measuring the displacement from the average position of an atom (black circle) for a single dimension (dashed line) ($\mathbf{x}_1$ in Fig. 5g or $\mathbf{x}_2$ in Fig. 5h) and then plotting this information as a function of $\mathbf{x}_4$.

The values in $\mathbf{x}_4$ were calculated (Fig. 6) using the *Matlab* function *superorder.m* (available in the supporting information). Firstly, *superorder.m* assigns a numerical index to the basic cells and then uses the most straightforward starting position (0, 0, 0) and determines equivalent positions in all of the basic cells. Next, it calculates $\mathbf{x}_4$ and then extracts the fractional part $\mathbf{x}_4'$ by subtracting the lowest integer from the value using the floor function. The floor function is used so that negative values are correctly calculated. The basic cells are sorted by $\mathbf{x}_4'$ and the basic cell sort order is returned as a list of the numerical indices as well as the associated values of $\mathbf{x}_4'$.

The *superorder.m* script will work with full 3D supercells for (3+1)D superspace (a single $\mathbf{q}$ vector with all nonzero components). The supporting information provides a 2D supercell (along $\mathbf{x}_1$ and $\mathbf{x}_3$) example found in Wagner & Schönleber (2009). In that paper there are two plots, one of the supercell (Fig. 15) and one showing the basic cells of the supercell rearranged in superspace order (Fig. 23). Although they do not discuss how they performed this rearrangement, the *superorder.m* function can determine the necessary reordering of the 35 basic cells and reproduce the published results.

	$q = 3/7a_3^*$	x_{avg}	x_4 ($q \cdot x_{avg}$)	$x_4' = x_4 - fl(x_4)$	Sort x_4'	Superspace order
Supercell 7c	1(A)	0,0,0	0/7	0/7	0/7	1(A)
	2(B)	0,0,1	3/7	3/7	1/7	6(F)
	3(C)	0,0,2	6/7	6/7	2/7	4(D)
	4(D)	0,0,3	9/7	2/7	3/7	2(B)
	5(E)	0,0,4	12/7	5/7	4/7	7(G)
	6(F)	0,0,5	15/7	1/7	5/7	5(E)
	7(G)	0,0,6	18/7	4/7	6/7	3(C)

Figure 6
Sample calculations were performed by the *superorder.m Matlab* script, where the basic cells are assigned a numerical index (1, 2, . . . , n) and an average position that is the same in each basic cell (x_{avg}). The value of x_4 is calculated by the dot product of $\mathbf{q}$ with x_{avg}. The x_4 values are translated to the range 0 to 1, resulting in x_4', using the floor function, which returns the lowest integer value. x_4' is sorted and the resulting reordered indices are returned.
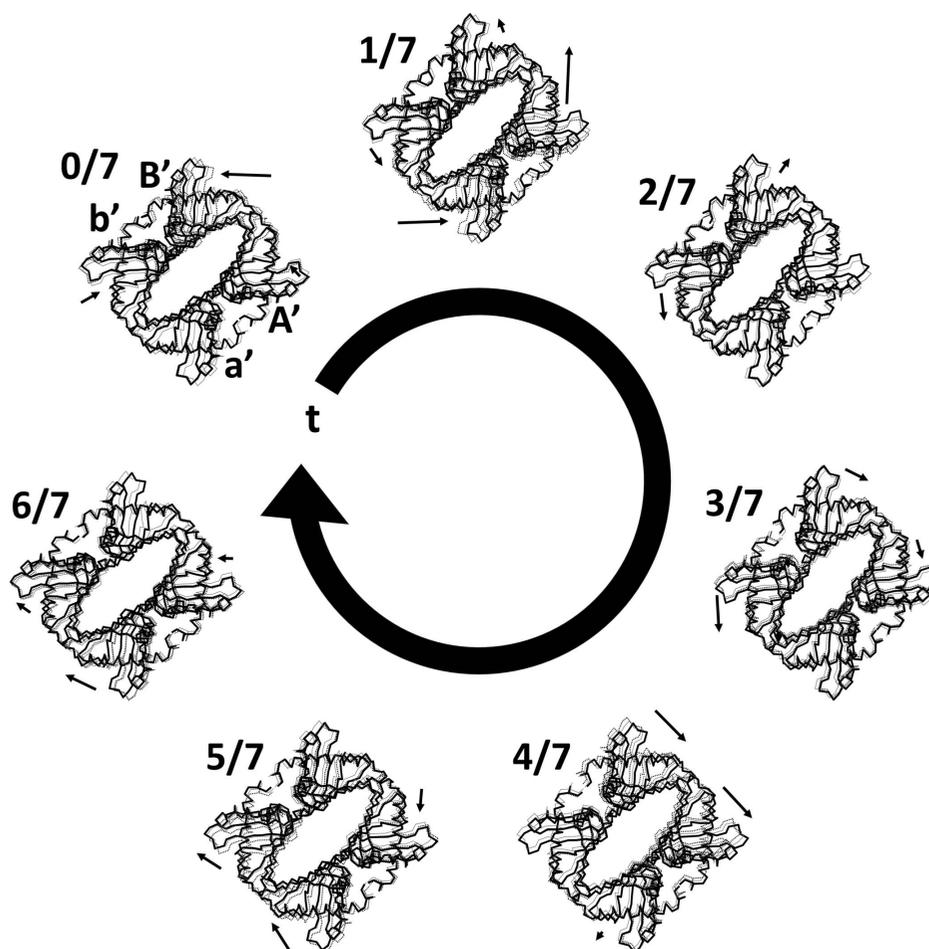
Figure 7
PDB entry 4n3e basic cells arranged in periodic superspace order ($t = 0$ to $t = 1$). Asymmetric unit basic cells were plotted to show the current data (black line; C^α trace of the four chains a′, A′, b′ and B′) as well as the previous basic cell (black dashed lines) and average position (grey line). The arrow indicates the relative displacement from the previous basic cell. Arrows are on the same relative scale for all basic cells.
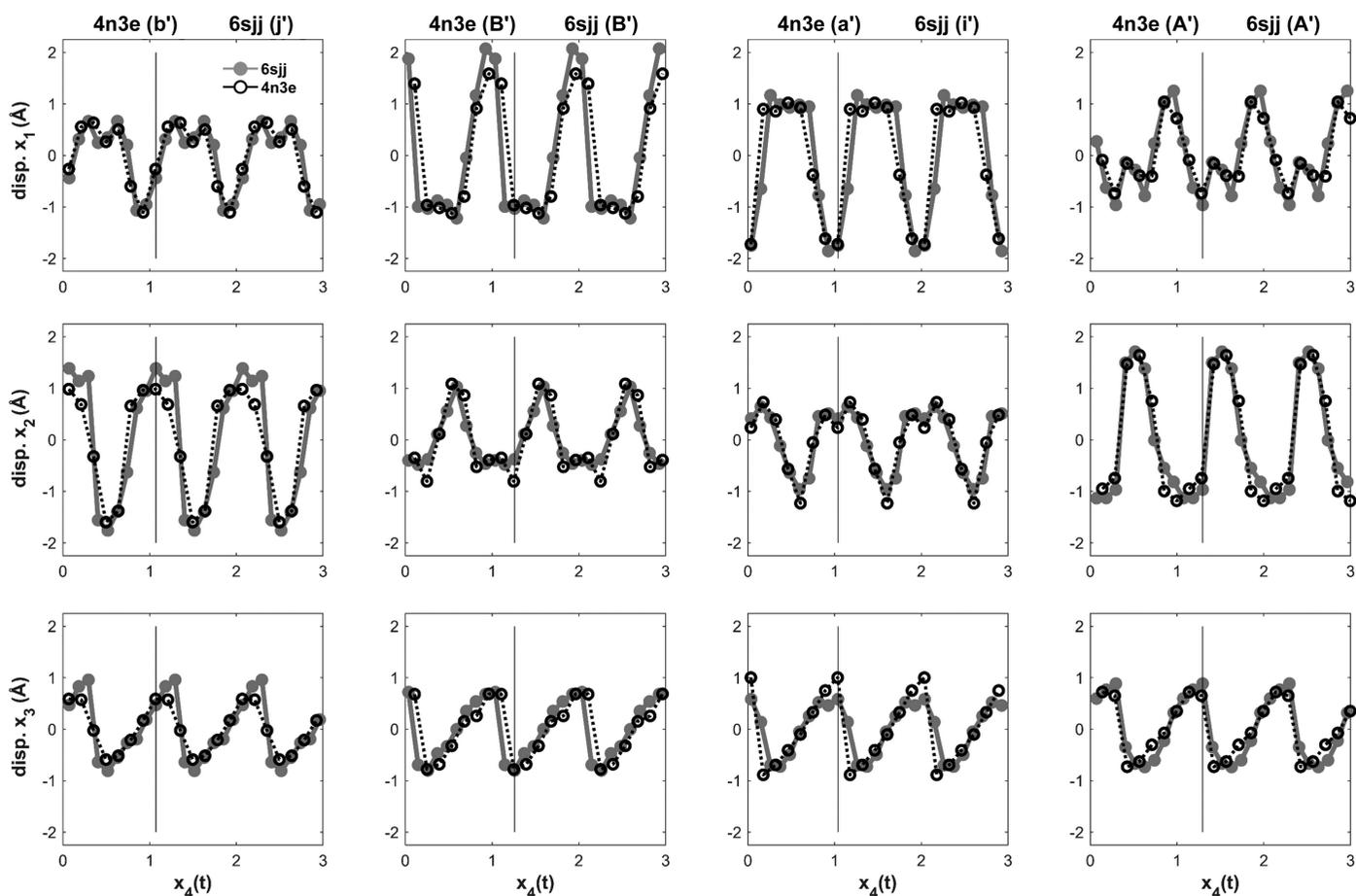
Figure 8
AMF approximations using the average centre of mass for each of the four chains in the ASUs for PDB entry 4n3e ($\mathbf{q} = 3/7\mathbf{a}_3^*$) and PDB entry 6sjj ($\mathbf{q} = 4/9\mathbf{a}_3^*$) in the $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ directions as a function of $\mathbf{x}_4$.

Table 1
Space-group and unit-cell data for PDB entries 4n3e and 6sjj.

PDB code	4n3e	4n3e, (3+1)D	6sjj	6sjj, (3+1)D
Space group	$C2$	$C222_1(00\gamma)$	$C2$	$C222_1(00\gamma)$
a, b, c (Å)	146.29, 146.29, 298.56	146.29, 146.29, 42.65	145.85, 145.85, 385.4	145.85, 145.85, 42.83
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
$\mathbf{q}_1$		0, 0, 3/7		0, 0, 4/9

3. Results

The results of transforming the supercell of PDB entry 4n3e into superspace (Fig. 7) demonstrates the power of this method to make the underlying periodic modulation accessible for analysis. Fig. 7 shows the basic cell states in a circle arranged by the value of t. The value of t was calculated using the corner of the basic cell as the origin. It is shown as a circle to reinforce the idea that AMFs are periodic along the $\mathbf{a}_{s4}$ direction. The rearrangement of the basic cells transforms what previously appeared to be random motions between neighbouring basic cells in the supercell into a cleaner sequence of consecutive transitions. It is more visually appealing and easier to visualize this by watching the movie (supporting information file `4n3e_ans-3-7-mas.gif`),

which shows the basic cells as frames animated in superspace order. These seven snapshots of the continuous AMFs in (3+1)D superspace show that most of the modulation (as viewed down $\mathbf{x}_3$) is related to an opening and closing of the space between chains b′ and B′ (Fig. 7). The modulation is not smooth in terms of the amount of displacement between frames being similar. Some consecutive frames show a large amount of displacement relative to the previous frame and other frames show very little relative displacement. AMFs that were not smooth were expected because of the higher order observed satellites, which are indicative of more sharp displacements. Additionally, all of the chains modulate relative to the average structure. Black arrows show the relative displacement of the four chains in each basic cell relative to the previous position. The black dashed line is the previous state and the grey line is the average position. When the same analysis is applied to PDB entry 6sjj it shows a very similar modulation (supporting information file `6sjj-ans-4-9-mas.gif`). With an understanding of how to transform the data into superspace, an interesting question would be: is there a relationship between the two 3D superstructures in (3+1)D superspace?

PDB entries 4n3e and 6sjj were compared by generating AMF approximations for the centre-of-mass displacements

with respect to the three crystal directions x_1, x_2 and x_3 of the four chains in the basic cell ASU. Fig. 8 shows plots of the displacements (in each of the three primary axes: x_1, x_2, x_3) as a function of x_4 (Fig. 8) for both PDB entry 4n3e (black circles and dashed black lines) and PDB entry 6sjj (grey dots and grey lines). Data from both superstructures overlay with each other in (3+1)D superspace. These superspace-plot data support the idea that both observed superstructures are 3D intersections in physical space from the same (3+1)D structure. In their paper, Sliwiak and coworkers observed that the change between the two structures was quite large given that the only difference in the crystallization conditions was the addition of melatonin (Smietanska *et al.*, 2020). They believed that this addition would result in at most a small change to the structure. They had also hoped to see melatonin binding to the complex. Unfortunately, they did not observe melatonin binding in the electron-density maps. In (3+1)D superspace, melatonin does appear to make a small change to the **q** vector (less than 1° rotation) and this small change results in changes to the intersection of the higher dimensional structure with physical space, resulting in the observed increase in supercell dimensions as well as the changes to the arrangement of the chains within the supercell compared with the PDB entry 4n3e structure.

Investigation of the AMFs reveals some interesting properties that apply to both the PDB entry 4n3e and 6sjj structures. The AMFs for x_1 and x_2 seem to be mirrored representations, relative to x_4, between b′/j′ and A′ as well as between B′ and a′/i′. In x_3, all four chains follow a sawtooth function. The sawtooth function increases linearly up to some maximum followed by a discontinuous or nearly discontinuous reset back to a minimum, where it then begins the linear increase again. The major difference between the four chains is where the reset of the sawtooth occurs in x_4. Chains a′/i′ and A′ experience the reset simultaneously in the same basic cell (vertical black lines in Fig. 8), whereas chain b′/j′ is lagging and has not reached the peak of the sawtooth and chain B′ has already experienced the sawtooth reset.

The PDB entry 4n3e structure has another interesting superspace feature where there are some occupancy modulations for some of the bound small molecules. An example of an AMF for a small molecule near **a′** is shown in the supporting information. Superspace represents occupancy modulations by having no value for the AMF at these positions by using a function with a discontinuity.

Another aspect that can be explored is the (3+1)D space group. In this case we are going from the information provided in the original papers to arrive at the correct (3+1)D space group. Both structures were in space group $C2$. Using *Superspace Group Finder* (https://it.iucr.org/resources/finder/; Orlov *et al.*, 2008) and inputting $C2$ in the 3D group set results in 21 potential parent (3+1)D space groups. These space groups can be further reduced using the directions provided in Porta *et al.* (2017) by first limiting the list to those that are possible for protein crystals and then reducing the remaining space groups to those which match the observed parameters of the **q** vector. In this case, there is only one remaining (3+1)D space group:

$C222_1(00\gamma)$. Table 1 summarizes the space-group data. The main difference between the two (3+1)D unit cells is that the **q** vectors are slightly different (0.015). The superspace group indicates that there should be a screw axis down a_3. The previously observed relationships in the AMFs between b′/j′ and A′ as well as between B′ and a′/i′ seem to support a screw axis as expected for the (3+1)D space group.

4. Conclusions

Looking at modulated atomic structures in superspace has provided an insightful understanding of how two modulated structures that are related in (3+1)D superspace appear very different in physical space. In this case, the difference can be attributed to slightly different 3D intersections of the same (3+1)D structure caused by a small change in the **q** vector. Additionally, these structures provided validation of the approaches that we are using to evaluate our PA refinements. They also provide independent validation of the tNCS refinement approach that was originally used to solve these structures. By providing the associated scripts as supporting information, we hope that other research groups will be able to more easily perform this analysis in the future on other modulated structures.

References

Dusek, M., Chapuis, G., Meyer, M. & Petricek, V. (2003). *Acta Cryst.* B**59**, 337–352.

Janner, A. & Janssen, T. (1977). *Phys. Rev. B*, **15**, 643–658.

Janner, A. & Janssen, T. (1980a). *Acta Cryst.* A**36**, 399–408.

Janner, A. & Janssen, T. (1980b). *Acta Cryst.* A**36**, 408–415.

Lovelace, J. J. & Borgstahl, G. E. O. (2020). *Crystallogr. Rev.* **26**, 3–50.

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.

Lovelace, J. J., Murshudov, G., Petříček, V. & Borgstahl, G. E. O. (2018). *Acta Cryst.* A**74**, a229.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Orlov, I., Palatinus, L. & Chapuis, G. (2008). *J. Appl. Cryst.* **41**, 1182–1186.

Palatinus, L. & Chapuis, G. (2007). *J. Appl. Cryst.* **40**, 786–790.

Patterson, A. L. (1935). *Z. Kristallogr.* **90**, 517–542.

research papers

Porta, J., Lovelace, J. & Borgstahl, G. E. O. (2017). *J. Appl. Cryst.* **50**, 1200–1207.

Porta, J., Lovelace, J. J., Schreurs, A. M. M., Kroon-Batenburg, L. M. J. & Borgstahl, G. E. O. (2011). *Acta Cryst.* D**67**, 628–638.

Read, R. J. & McCoy, A. J. (2016). *Acta Cryst.* D**72**, 375–387.

Schönleber, A. & Chapuis, G. (2004). *Acta Cryst.* B**60**, 108–120.

Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). *Acta Cryst.* D**71**, 829–843.

Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). *Acta Cryst.* D**70**, 471–480.

Smaalen, S. van (2005). *Z. Kristallogr.* **219**, 681–691.

Smaalen, S. van (2007). *Incommensurate Crystallography*. Oxford University Press.

Smietanska, J., Sliwiak, J., Gilski, M., Dauter, Z., Strzalka, R., Wolny, J. & Jaskolski, M. (2020). *Acta Cryst.* D**76**, 653–667.

Wagner, T. & Schönleber, A. (2009). *Acta Cryst.* B**65**, 249–268.