

A non-mathematical introduction to the superspace description of modulated structures

Trixie Wagner^{a*} and Andreas Schönleber^b

^aNovartis Institutes for BioMedical Research, 4002 Basel, Switzerland, and ^bLaboratory of Crystallography, University of Bayreuth, 95440 Bayreuth, Germany

Correspondence e-mail:
trixie.wagner@novartis.com

Received 13 February 2009

Accepted 27 April 2009

The X-ray analysis of (6*R*,7*aS*)-6-(*tert*-butyl-dimethylsilyloxy)-1-hydroxy-2-phenyl-5,6,7,7*a*-tetrahydropyrrolizin-3-one, C₁₉H₂₇NO₃Si, revealed a diffraction pattern which is typical for modulated structures: strong Bragg peaks surrounded by weaker reflections which cannot be indexed with the same three reciprocal lattice vectors that are used to describe the strong peaks. For this class of crystal structures the concept of superspace has been developed which, however, for many crystallographers still constitutes a Gordian Knot. As a possible tool to cut this knot the crystal structure of the above-mentioned tetrahydropyrrolizinone derivative is presented as an illustrative example for handling and describing the modulated structure of a typical pharmaceutical (*i.e.* molecular) compound. Having established a working knowledge of the concepts and terminology of the superspace approach a concise and detailed description of the complete process of peak indexing, data processing, structure solution and structure interpretation is presented for the incommensurately modulated crystal structure of the above-mentioned compound. The superspace symmetry applied is $P2_1(\alpha 0 \gamma)0$; the (incommensurate) \mathbf{q} vector components at 100 K are $\alpha = 0.1422$ (2) and $\gamma = 0.3839$ (8).

1. Introduction

Together with quasicrystals and composite crystals incommensurately modulated structures constitute the category of aperiodic crystals. Since modulated structures will be discussed in detail over the course of this paper the interested reader is referred to 'An elementary introduction to superspace crystallography' by van Smaalen (2004) for a comprehensible description of quasicrystals and composite crystals. The diffraction patterns of modulated structures are characterized by the existence of additional Bragg reflections so that an integer indexing with three indices hkl is not possible; instead, $3 + d$ indices are required ($d = 1, 2$ or 3). The necessity for using four or more indices must be understood as a loss of periodicity in three dimensions. To restore the periodicity, the concept of higher-dimensional superspace was developed (de Wolff, 1974, 1977). Atoms are no longer points in space, but are envisioned as d -dimensional atomic domains (Janner & Janssen, 1977).

The theoretical foundation for the superspace approach is now well established (van Smaalen, 2007). Furthermore, considerable effort has been put into making JANA2006, a crystallographic computing system that can conveniently handle the data and structures of modulated compounds (Petricek *et al.*, 2006), available and user-friendly.

However, for the majority of the crystallographic community many of the concepts and terms used in this context are

still not easily accessible because superspace descriptions tend to be rather mathematical and based on inorganic examples. In this manuscript a more practical approach to this topic is presented, based on an illustrative example from a pharmaceutical crystallographic service lab. The manuscript is divided into three main sections: in the first section a working knowledge of modulated structures will be established and key terms such as modulation vector, atomic modulation function, atomic domain *etc.* will be introduced and explained. After a brief excursion into the historical and recent developments in the field of aperiodic structures this working knowledge will be illustrated in the third section where the complete process of peak indexing, data processing, structure solution and structure interpretation for an incommensurately modulated structure is presented for the crystal structure of (6*R*,7*aS*)-6-(*tert*-butyl-dimethyl-silanyloxy)-1-hydroxy-2-phenyl-5,6,7,7*a*-tetrahydropyrrolizin-3-one, C₁₉H₂₇NO₃Si.

1.1. Modulated structures – a first definition

Most crystal structures are periodic in three dimensions and show a diffraction pattern that can be indexed with three integer numbers (Fig. 1). Modulated structures can be derived from those structures: here atoms, groups of atoms or even whole molecules are shifted or rotated with respect to their neighbours such that the three-dimensional translational symmetry, often considered as *the* characteristic feature of a crystal structure, is destroyed. These shifts and rotations in modulated structures, however, are not arbitrary; they follow distinct rules and within these distortions (or better: modulations) there is additional periodicity which can mathematically be described by so-called **atomic modulation functions** (AMFs). AMFs can be harmonic (continuous) and therefore be expressed by a sine/cosine combination or they may be

discontinuous, in which case crenel or sawtooth functions are needed for their description (Fig. 2; Petricek *et al.*, 1995). Based on the periodicity of the modulation wave a distinction can be made between **commensurately** modulated structures (periodicity matches an integral number of lattice translations of the basic cell) and **incommensurately** modulated structures (periodicity does not match an integral number of lattice

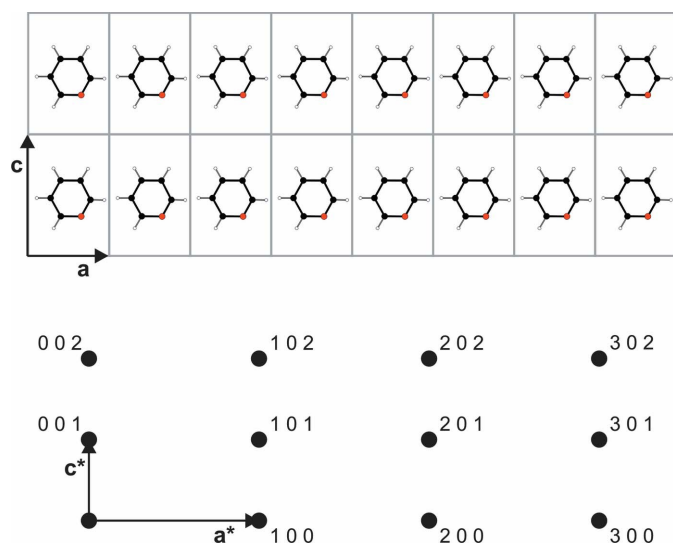


Figure 1
Two-dimensional schematic representation of a hypothetical crystal structure that is periodic in three dimensions and its schematic diffraction pattern, indexed with three integer numbers based on the reciprocal lattice vectors **a***, **b*** (not shown) and **c***. Lattice/unit cells in light grey.

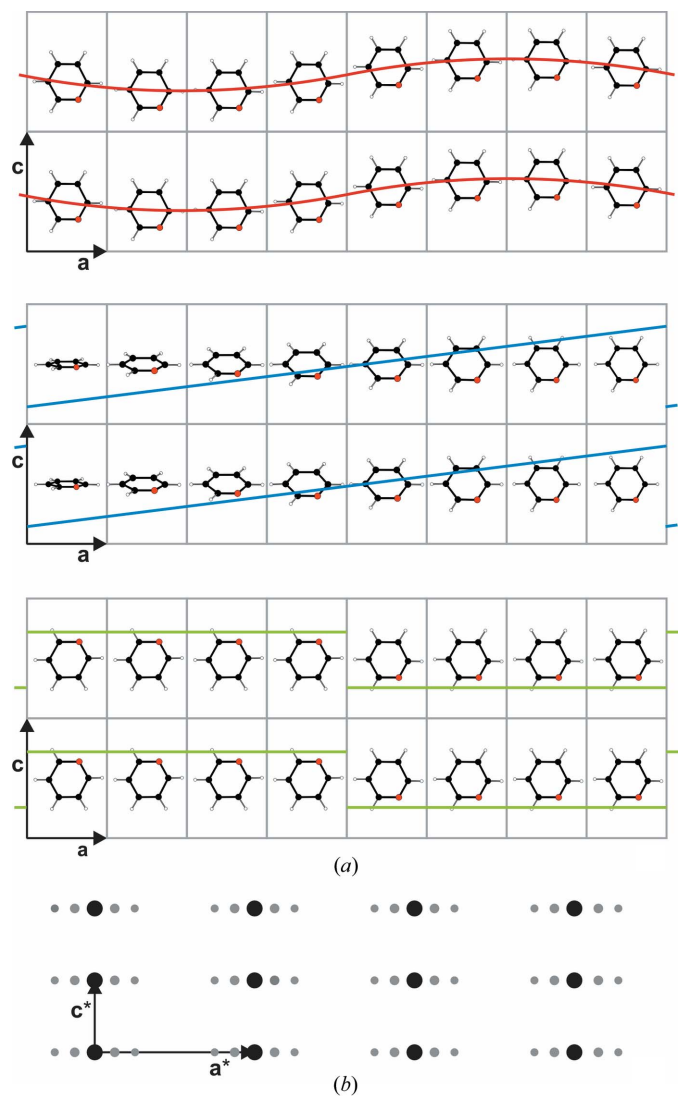


Figure 2
Schematic representation of three manifestations of a modulated structure with lost translational symmetry along the **a** axis. All three drawings in (a) are derived from the periodic structure shown in Fig. 1 by shifting or rotating the molecules. The atomic modulation functions suitable for the description of the modulated atomic positions are shown as an overlay: in the top drawing the molecules are shifted up and down parallel to **c** in a continuous harmonic (sinusoidal) way (red curve); in the middle drawing the molecules are rotated around an axis parallel to **a**, the rotation angle can be described using a sawtooth function (blue) with a discontinuity between molecules 8 and 1; in the bottom drawing the molecule adopts two different orientations which can be described by a step-like crenel function (green). Note that the modulation proceeds only along **a**, the **c** direction is not affected. (b) In the associated schematic diffraction pattern additional reflections (grey circles) appear along **a***. The number of satellite reflections and their intensity distribution depend on the strength and nature of the modulation. For simplicity, only one diffraction scheme was drawn.

translations of the basic cell and is therefore incommensurate with the periodic basic structure).

Due to the periodic character of the modulation, additional sharp peaks appear in the diffraction pattern, just as the Bragg reflections produced by a three-dimensional crystal are a result of the periodic terms in the structure factor equation. These additional peaks are referred to as **satellite reflections** and usually they are weaker than the **main** reflections. The satellite reflections can lie parallel to one reciprocal axis, but they do not have to. It should be emphasized that modulated structures are not disordered but have long-range order which is reflected in those discrete additional peaks rather than in diffuse streaks. For **displacive modulations** as described above (*i.e.* the atomic positions are affected), the number of satellite reflections that can be observed depends on the amplitude of the AMFs or, in other words, on the degree of distortion/modulation present in the three-dimensional crystal structure. The stronger the modulation, the stronger the satellite reflections and also the higher the order of the satellite reflections that have measurable intensity (see §1.2.3).

Even though the distribution of satellite reflections in reciprocal space is not arbitrary, they cannot be described with the same three reciprocal lattice vectors that allow an integer indexing of the main reflections. Any attempt to do so results in non-integer values for h and/or k and/or l (Fig. 3). For an incommensurately modulated structure there is not *any* set of

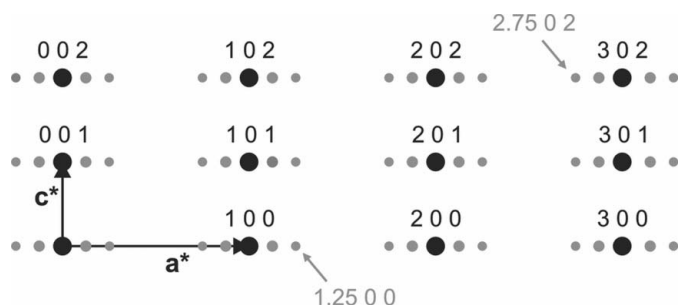


Figure 3

Indexed schematic diffraction pattern of the modulated structure of Fig. 2, main reflections black, satellite reflections grey. Indices for main reflections are based on the lattice vectors \mathbf{a}^* , \mathbf{b}^* (not shown) and \mathbf{c}^* . Note that an integer indexing for the satellite reflections is not possible using those vectors. The deviations from the indices of the closest main reflection are $(-0.25, 0, 0)$ for the marked satellite reflection at the top and $(+0.25, 0, 0)$ for that at the bottom.

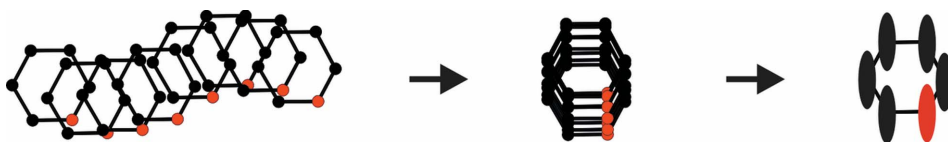


Figure 4

Neglecting satellite reflections and solving the structure in the *basic cell* corresponds to averaging the contents of the eight unit cells shown in either of the top rows of Fig. 2(a) by shifting the molecules parallel to \mathbf{a} into one cell (H atoms omitted for clarity; the molecules from Fig. 2 have been moved towards each other already). Consequently, the resulting *average structure* is characterized by large displacement ellipsoids and, in the case of a strong modulation, also by unrealistic geometric parameters.

three vectors that allows an integer indexing of all (main and satellite) reflections.

1.2. Handling the diffraction pattern of modulated structures

The diffraction pattern of a modulated structure (strong main reflections surrounded by weaker satellite reflections) can be approached in three different ways, two of which are based on a description in the traditional three-dimensional space, and one that is not. Since the structure, however, is not periodic in three dimensions (except for commensurate cases) any three-dimensional approach can only be regarded as an approximation. The three-dimensional approaches shall be mentioned nonetheless since the results may contain valuable information that can be used at a later stage (space-group determination in superspace, identification of groups with strongest modulation, type of AMF to describe the modulation *etc.*).

1.2.1. Basic cell and average structure. The first option is to ignore the weak satellite reflections altogether and concentrate on the main reflections. The indexing procedure then delivers a so-called **basic cell** (sometimes also referred to as a **subcell**) with a volume that can accommodate a realistic number of the molecule under study (1, 2, 3 *etc.*). In the three-dimensional crystal structure using the main reflections only corresponds to averaging the contents of several unit cells and looking at a structure which is commonly referred to as the **average structure**. It is usually characterized by large anisotropic atomic displacement parameters (ADPs) as well as unrealistic bond lengths and bond angles. In the hypothetical structure of Fig. 2, for example, the averaging would include eight unit cells (Fig. 4). The inherent disadvantage of this approach is that a certain amount of information provided by the diffraction pattern, *i.e.* satellite intensity, is neglected.

1.2.2. Supercell and superstructure. The second way of handling the diffraction pattern is to drop the distinction between main and satellite reflections and to use all reflections equivalently for indexing (Fig. 5). This results in a smaller reciprocal unit cell and a reciprocal lattice where, depending on the maximum order of the satellite reflections that are observed, many reciprocal lattice points may be empty, *i.e.* have no observable intensity. In direct space the unit cell is accordingly larger and is commonly referred to as a **supercell**. The resulting crystal structure is called a **superstructure** and is characterized by a large number of independent molecules in

the asymmetric unit (Z' ; $Z' = 8$ in the example of Fig. 5). In incommensurate cases where the positions of the satellite reflections do not perfectly fit the grid of the supercell lattice a deformation of reciprocal space has to be accepted which usually manifests itself in poor agreement factors, large standard deviations of refined parameters, split atoms, large ADPs *etc.* In

such cases the term superstructure approximation would be more appropriate (see also §1.4).

For commensurate cases the description as a (three-dimensional) superstructure constitutes a valid approach (e.g. Hao *et al.*, 2005; Siegler *et al.*, 2008) and can even be used in combination with a (higher-dimensional) superspace treatment (Schmid & Wagner, 2005).

1.2.3. Moving into superspace. In the third approach, too, all reflections are used, but the distinction between main and satellite reflections is maintained. In a first step the reciprocal unit cell is established using the main reflections (*cf* §1.2.1). The second step makes use of the fact that the distribution of satellite reflections in reciprocal space is not arbitrary; they can be divided into groups in which they are equidistant not only from each other but also from the main reflection to

which they belong. This systematic distribution allows the definition of a fourth vector, the so-called **modulation wavevector \mathbf{q}** , which describes the satellite reflections with respect to their main reflection. Now every satellite peak can be uniquely identified as being $n \cdot \mathbf{q}$ ($n = \pm 1, \pm 2, \dots$) away from its main reflection, and one speaks of the n th-order satellite reflection of this main reflection. This manuscript will be restricted to four-dimensional cases, *i.e.* one modulation wavevector \mathbf{q} . The theory for five- or six-dimensional cases, where a second or even a third set of satellite reflections requires the use of one or two additional vectors, can be derived accordingly. Since five- and six-dimensional cases, however, occur mainly with higher symmetry (hexagonal, cubic) and a standardized treatment is still in the process of being developed, they are beyond the scope of this introductory guide.

Within the framework of the three reciprocal lattice vectors \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* , which describe the basic cell, the modulation wavevector \mathbf{q} can be expressed only with the help of fractional components: $\mathbf{q} = \alpha \mathbf{a}^* + \beta \mathbf{b}^* + \gamma \mathbf{c}^*$. In other words, the components of \mathbf{q} are given with respect to the basis vectors of the reciprocal lattice of the basic cell or average structure.¹ Please note that the number of non-zero components of \mathbf{q} is not related to the dimensionality of the modulation, which is defined by the number of \mathbf{q} vectors necessary for the description of reciprocal space. Depending on the rationality of α , β and γ there is now a second way (*cf* §1.1) to classify modulated structures as **commensurate** (all components rational) or **incommensurate** (at least one of the components irrational). In direct space commensurate cases are characterized by a supercell in which all lattice parameters are integer multiples (1, 2, ..., n) of the lattice parameters of the basic cell. In practice, however, it is often not easy to distinguish between commensurate and incommensurate cases, especially when the multipliers involve larger integers and the differences between a commensurate and an incommensurate approach disappear.

Working with one modulation wavevector \mathbf{q} implies a transition into four-dimensional space (*cf* §1.2.4) and every reflection of the diffraction pattern can now be described in a unique way by four integer indices h , k , l and m . This higher-dimensional concept affects all subsequent steps of structure analysis and therefore will be discussed in detail in connection with the sample structure in §3. A schematic illustration of the relation between superspace and three-dimensional space is given in Fig. S1 of the supplementary material.²

1.2.4. The definition of reciprocal superspace. It is difficult to envision the fourth dimension introduced in the preceding section by looking at the \mathbf{q} vector alone because \mathbf{q} itself is only part of the definition of reciprocal superspace. To illustrate the

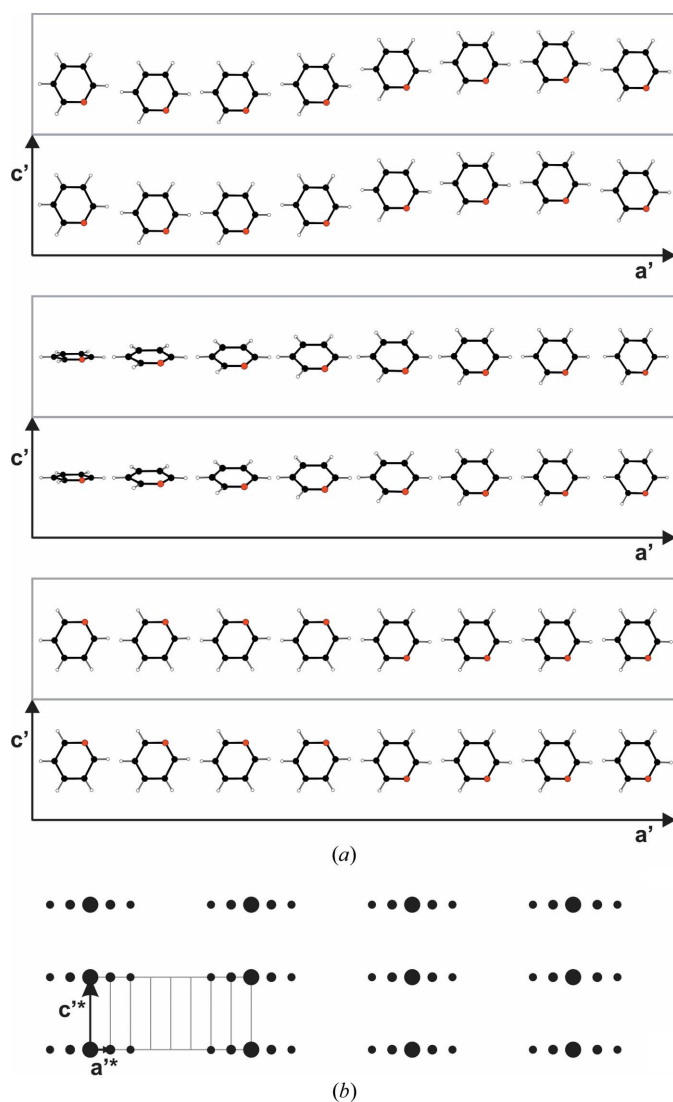


Figure 5 Using main and satellite reflections equivalently results in a large unit cell (light grey) with $\mathbf{a}' = 8 \cdot \mathbf{a}$ or, in reciprocal space, $\mathbf{a}'^* = 1/8 \cdot \mathbf{a}^*$. Note that unless satellite reflections up to higher orders are observed, numerous reciprocal lattice points are empty (b). This holds especially if the satellite reflections are not along one reciprocal axis but are located in a plane (e.g. the $\mathbf{a}^* \mathbf{c}^*$ plane, *cf* §3).

¹ The \mathbf{q} vector components α , β and γ have to be distinguished from the unit-cell parameters α , β and γ . Alternative notations such as σ_1 , σ_2 , σ_3 can be found in the literature (van Smaalen, 2007). In this manuscript, however, the nomenclature used in Vol. C of the *International Tables of Crystallography* (α , β and γ) will be followed (Janssen *et al.*, 2006).

² Supplementary data for this paper are available from the IUCr electronic archives (Reference: BK5084). Services for accessing these data are described at the back of the journal.

relationship between three-dimensional reciprocal space and reciprocal superspace it is more convenient to focus on one line of the diffraction pattern shown in Fig. 6(a). In this figure \mathbf{q} is chosen to be parallel to \mathbf{a}^* (*i.e.* only the α component of \mathbf{q} is non-zero). The additional fourth dimension is defined by the unity vector \mathbf{e}_4^* (Fig. 6b) which is perpendicular to \mathbf{a}^* in the plane of the paper (\mathbf{e}_4^* is also perpendicular to \mathbf{b}^* and \mathbf{c}^* , which are not shown in Fig. 6). The linear combination of \mathbf{e}_4^* and \mathbf{q} defines the fourth reciprocal lattice vector, \mathbf{a}_{s4}^* (*cf.* §1.3). Extending this vector \mathbf{a}_{s4}^* illustrates that every satellite reflection along the (three-dimensional) reciprocal lattice vector \mathbf{a}^* can now be understood as a projection of a reflection along the reciprocal superspace vector \mathbf{a}_{s4}^* (dotted lines), or, in other words: \mathbf{q} is the result of the projection of the additional dimension \mathbf{a}_{s4}^* into three-dimensional reciprocal space \mathbf{R}^* .

1.3. Superspace and direct space – where are the atoms?

Up to this point the terminology of superspace has been used in conjunction with reciprocal space only. Now, how can the structure of an aperiodic crystal in direct space be imagined? Where are the atoms? And what does it mean that ‘... the aperiodic structure in real space is interpreted as a cut through the $(3 + d)$ -dimensional superspace description’ (Janner & Janssen, 1977)?

Before answering these questions, the following convention shall be introduced: the vectors \mathbf{a} , \mathbf{b} and \mathbf{c} and their respective coordinates x , y and z used in classical three-dimensional crystallography will now be denoted \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 , and the coordinates as x_1 , x_2 and x_3 . That way the notation can easily be extended to an arbitrary number of dimensions. Axes and coordinates in an n -dimensional superspace will then be referred to as \mathbf{a}_{s1} , \mathbf{a}_{s2} , \mathbf{a}_{s3} , ..., \mathbf{a}_{sn} and $x_1, x_2, x_3, \dots, x_n$ (just as the reciprocal superspace direction in §1.2.4 was denoted \mathbf{a}_{s4}^*).

Since in the superspace approach the additional dimensions are defined to be perpendicular to all three dimensions of physical space, it is impossible to illustrate all four dimensions on paper. One has to use projections (in reciprocal space) or sections (in direct space) of the higher-dimensional space, just as one would draw a two-dimensional projection of a three-dimensional crystal structure on graph paper. This way the complexity of four-dimensional space can be reduced by drawing two-dimensional sections consisting of one of the three principal axes (\mathbf{a}_{s1} , \mathbf{a}_{s2} or \mathbf{a}_{s3}) and the axis that runs parallel to the fourth dimension.

Using this technique, a hypothetical four-dimensional (or better: $(3 + 1)$ -dimensional) modulated structure with one atom in the unit cell shall now be considered (Fig. 7). In the $(D + d)$ nomenclature the first number D refers to the dimensions in physical space and the second number d to the dimensions in additional space. As mentioned above, the illustration is simplified by drawing four unit cells in a $(1 + 1)$ -dimensional superspace section.

To facilitate recognizing the three-dimensional crystal structure in the superspace section that will be described in Fig. 7(c) a periodic one atom structure is shown first (Fig. 7a). In the aperiodic variant below (Fig. 7b) the atoms have been

shifted from their original positions (small black circles) and the translational symmetry from Fig. 7(a) is lost: the distance between P and P', for example, is smaller than the distance between P' and P''.

The periodic superspace structure (Fig. 7c) is now defined by two basis vectors \mathbf{a}_{s1} (or \mathbf{a}_{s2} or \mathbf{a}_{s3}) and \mathbf{a}_{s4} . This vector \mathbf{a}_{s4} is perpendicular to the corresponding three-dimensional axis \mathbf{a}_1 (if \mathbf{a}_{s1} is chosen) and the angle between \mathbf{a}_{s1} (superspace) and \mathbf{a}_1 (three-dimensional space) is defined by the \mathbf{q} -vector component α : $\tan(\mathbf{a}_1, \mathbf{a}_{s1}) = \frac{\alpha}{|\mathbf{a}_1|}$. In $(3 + 1)$ -dimensional superspace atoms can no longer be envisioned as points in space. Instead they have to be understood as one-dimensional objects, the so-called **atomic domains**, along the fourth dimension and they are represented by a curve (mathematically: the atomic modulation function, AMF) which is periodic along \mathbf{a}_{s4} . The shape of the AMF has to be determined experimentally during

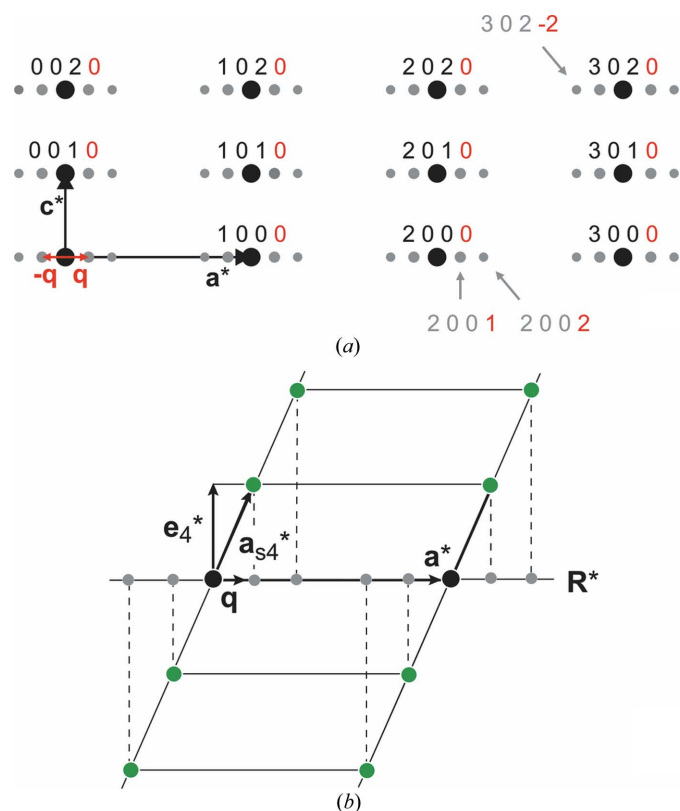


Figure 6

Four-dimensionally indexed diffraction pattern of a hypothetical modulated structure (a). The vectors \mathbf{a}^* and \mathbf{c}^* which define the *basic cell* are derived from the positions of the main reflections. The additional vector \mathbf{q} , reaching from the main reflection to the satellite, can be expressed in fractions of those basic vectors: in the present example $\mathbf{q} = 0.125 \cdot \mathbf{a}^* + 0 \cdot \mathbf{b}^* + 0 \cdot \mathbf{c}^*$. Note that by introducing only one additional vector it is possible to uniquely index all reflections with four integer numbers. The fourth index also reveals the satellite order: the marked satellite reflection at the top is the minus second-order satellite reflection of the main reflection $3\ 0\ 2$. (b) Diffraction pattern generated by a one-dimensionally modulated crystal. The reflections (green circles) along \mathbf{a}_{s4}^* (defined by the linear combination of \mathbf{e}_4^* and \mathbf{q} in the plane of the paper, see §1.2.4) are projected as satellite reflections (grey circles) onto \mathbf{R}^* , the reciprocal three-dimensional space in which the reciprocal lattice vectors \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* (not shown) are defined (see also supplementary material Fig. S1).

structure solution and refinement. The same holds for the AMFs in the sections (\mathbf{a}_{s2} , \mathbf{a}_{s4}) and (\mathbf{a}_{s3} , \mathbf{a}_{s4}). Note that in the two dimensions shown in Fig. 7(c) the AMF does exhibit translational symmetry, or, in more general words: within the superspace approach the additional dimensions are introduced to recover translational symmetry and the structure that is aperiodic in three dimensions becomes periodic in higher-dimensional space.

The aperiodic atomic structure in three-dimensional space (cf. Fig. 7b) can now be deduced by a one-dimensional (horizontal) cut \mathbf{R} (parallel to \mathbf{a}_1) through this two-dimensional representation of superspace. The intersections of \mathbf{R} with the AMFs of the superspace atomic domains give the positions of the atoms in three-dimensional space. Note again that along the three-dimensional space line \mathbf{R} there is no translational symmetry (periodicity).

If all intersections (P' , P'' , P''' etc.) are shifted by translational symmetry into the left unit cell (i.e. parallel to \mathbf{a}_{s1} , dotted lines) the resulting curve is the atomic modulation function for the depicted atom. Or, in more general words, the shape of the AMF for any atom can be reconstructed *via* the positions of this atom in three-dimensional space and *vice versa*.

The type of representation depicted in Fig. 7(c) occurs very frequently in the description of modulated structures and will reappear in §3, for example in conjunction with two-dimensional sections out of four-dimensional electron-density maps.

The atomic modulation functions are wavefunctions with a period of 1, therefore a **phase** t of the modulation can be defined which runs parallel to \mathbf{a}_{s4} and ranges from 0 (origin of cell) to 1 (origin of next cell). The coordinate of the fourth dimension, x_4 (which also runs parallel to \mathbf{a}_{s4}) has to be distinguished from this new parameter t : in Fig. 7(c) the atoms P , P' , P'' and P''' have the same phase ($t = 0$), but different coordinates x_4 . The value of t is found from a projection onto \mathbf{a}_{s4} along \mathbf{R} while the value for x_4 is found by a projection onto \mathbf{a}_{s4} along \mathbf{a}_{s1} . A different value of t will lead to a different set of points $P \dots P'''$ in three-dimensional space which, however, is related to the one obtained for $t = 0$ by a simple origin shift. For a visual demonstration of the impact of t onto the description of the crystal structure the interested reader is referred to the 'superspace playground', a website maintained by the Laboratory of Crystallography at the École Polytechnique Fédérale de Lausanne (Chapuis & Orlov, 2005).

As the whole crystal structure in three-dimensional space is defined as a three-dimensional cut \mathbf{R} of the $(3 + 1)$ -dimensional superspace, all atoms in the crystal (i.e. in the aperiodic three-dimensional structure) have the same phase t , but, as mentioned before, not the same coordinate x_4 . Therefore, for the crystal-chemical analysis only atoms with the same phase of the modulation, i.e. with the same value of t , can be considered. Shifting these atoms from the three-dimensional space line \mathbf{R} into one unit cell (*via* the restored translational symmetry of the superspace construction) provides a very elegant way for this analysis, realised, for example, in the so-called **t -plots** shown in §3: varying t for a certain geometric parameter from $t = 0$ to $t = 1$ provides all values for this

parameter occurring anywhere in the three-dimensional crystal structure along \mathbf{R} . This is visualized in Fig. 7(d): the interatomic distance at the phase of the modulation $t = t_0$ is equivalent to the distance between atoms P and P' in three-dimensional space, the distance for $t = t_1$ represents the three-dimensional space distance between P' and P'' , the one for $t =$

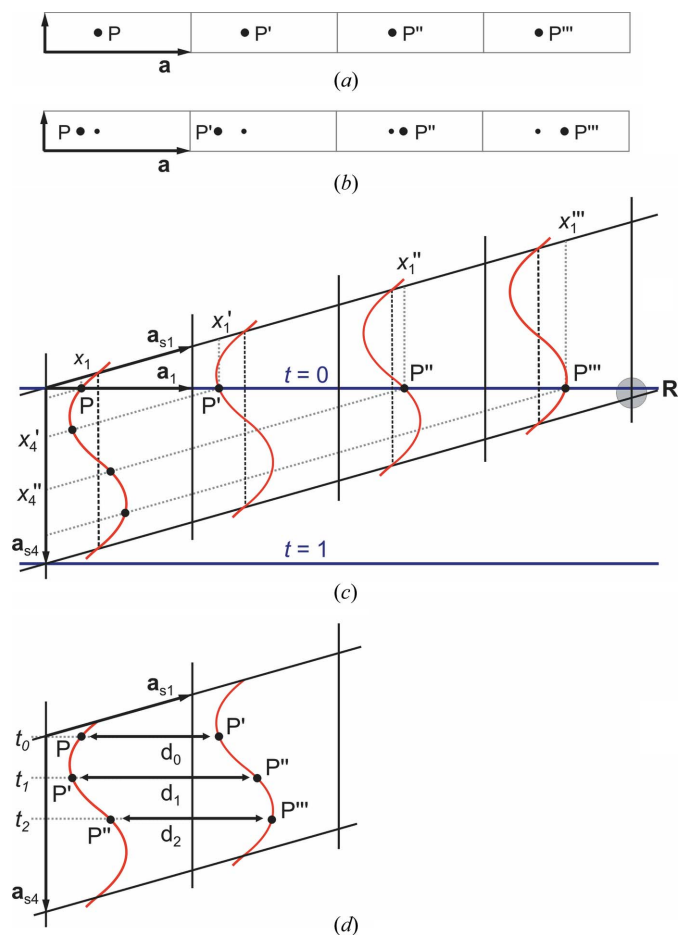


Figure 7

The superspace approach for aperiodic structures. At the top a periodic one-atom structure is shown with four equivalent positions P , P' , P'' and P''' (a). In the aperiodic structure the four atoms have been shifted out of their original positions (small black circles) by apparently arbitrary amounts (b). $(1 + 1)$ -Dimensional sketch (c) of the superspace approach around the atomic positions of (b). A section defined by the superspace vectors \mathbf{a}_{s1} and \mathbf{a}_{s4} is shown, with \mathbf{a}_{s4} perpendicular to \mathbf{a}_1 and the angle between \mathbf{a}_1 and \mathbf{a}_{s4} defined by the \mathbf{a}_{s4} component of \mathbf{q} (see text). Average positions (cf a) depicted as dashed vertical lines, 'modulated positions' as solid curves (red). Please note that these red curves correspond to the red lines shown in the top row of Fig. 2(a). The positions of the atoms in three-dimensional space (P , P' , P'' , P''') can be derived from the intersections of the three-dimensional space line \mathbf{R} (parallel to \mathbf{a}_1) and the atomic modulation functions. Shifting P , P' , P'' , P''' into the unit cell on the left by lattice translations (dotted lines) gives the shape of the AMF of this atom, or, in other words: the AMF representing atom P in superspace is the mathematical expression of all the positions of P in the three-dimensional crystal. \mathbf{R} can be drawn for different values of t , the phase of the modulation (here: $t = 0$). A more detailed definition of t is given in §1.3. (d) Superspace section with two unit cells to illustrate the significance of t . To calculate all interatomic distances on \mathbf{R} it is sufficient to calculate all distances for the first unit cell from $t = 0$ to $t = 1$ (see text). For example, the distance d_2 at $t = t_2$ in (d) is equivalent by translational symmetry to the distance between P'' and P''' on the three-dimensional line \mathbf{R} in (c) where $t = 0$.

t_2 the three-dimensional space distance between P'' and P''' , and so on (note that the two atoms defining one distance still have the same t value). Of course, other structural features such as coordinates, bond lengths, torsion angles, *etc.* also change with t , and it is exactly this variation in geometry with t that is the core of the discussion of molecular geometry for modulated structures (*cf.* §3).

1.4. Commensurate approximation or superspace description?

Based on this superspace description one of the fundamental questions related to modulated structures can be addressed: under what circumstances is a superstructure approximation of an incommensurately modulated structure valid enough and when is it necessary, or at least more appropriate, to use a higher-dimensional approach? Unfortunately, unlike in the case of the *Ultimate Question of Life, the Universe, and Everything* from Douglas Adams' famous book *The Hitchhiker's Guide to the Galaxy* there is no simple answer to this question. In general, one can say that a commensurate/superstructure approach is valid as long as it properly reflects the three-dimensional structure and does not lead to too many problems such as unrealistic geometric features, poor data-to-parameter ratio, or strange displacement ellipsoids which have to be tamed with the help of (usually undesired) restraints or constraints.

In incommensurate cases such a commensurate approach – as already mentioned – always has to be regarded as an approximation because the atomic positions used in the superstructure are not the positions of the atoms in three-

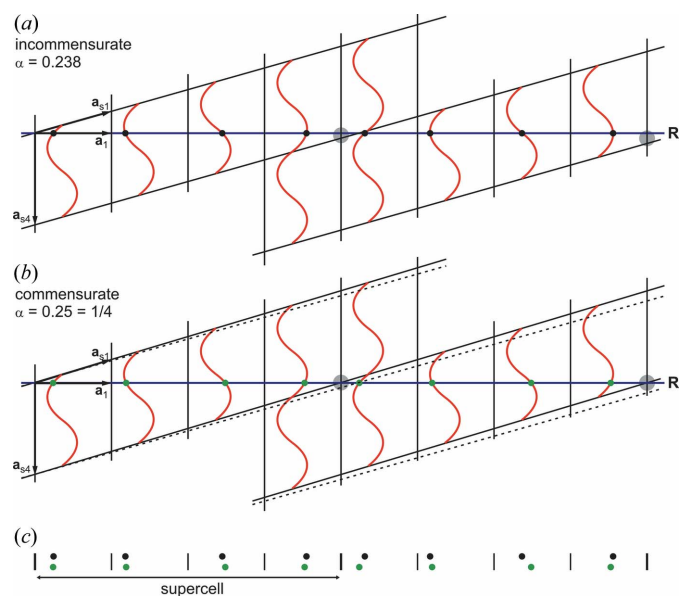


Figure 8

Extended version of the sketch shown in Fig. 7. An incommensurate case with a \mathbf{q} vector of $0.238 \cdot \mathbf{a}_1^*$ is assumed (a). The closest simple rational number here is $0.25 = \frac{1}{4}$, which results in a fourfold superstructure (b). Note the skew of the lattice compared with the original geometry (dashed lines). The resulting (different!) atomic positions from the incommensurate representation (black circles) and the superstructure (green circles) are depicted at the bottom (c).

dimensional space. This is illustrated in Fig. 8. To display the atomic positions of an aperiodic structure by a superstructure approximation, the superspace lattice has to be deformed in such a way that one of the lattice points coincides with \mathbf{R} . In Fig. 8(a) this is the lattice point of the fourth unit cell (grey circle). Therefore, the corresponding superlattice is fourfold, and the respective coordinates x_1 of the commensurate approximation can be calculated from the fractional coordinates x_1, x_1', x_1'' and x_1''' (Fig. 7) of the incommensurate structure.

Owing to the lattice deformation which is necessary to approximate a \mathbf{q} vector of (*e.g.*) 0.238 by the rational number $0.25 = \frac{1}{4}$ (Fig. 8b) the resulting positions in the superstructure (green circles) are visibly different from those which are derived from the superspace representation (Fig. 8a) and which are the real positions of the atoms in physical space (black circles). This effect becomes the more pronounced the further one proceeds along the three-dimensional space line \mathbf{R} and usually manifests itself in problems during the refinement.

In cases where the satellite reflections occur along one of the reciprocal axes and the multiplier n for the n -fold superstructure is high, a commensurate three-dimensional refinement can still behave properly (and thus be pursued) but more often the mentioned warning signs (well known from classical three-dimensional crystallography) appear already at an early stage of the refinement and should be taken as an indication that a transition into a superspace description might be more promising.

At first sight it may seem cumbersome and puzzling to introduce an additional dimension with no physical meaning. However, there is one inherent advantage: the superspace approach is not limited to aperiodic structures but can also be applied to commensurate superstructures and non-modulated structures. Therefore, it can provide, *e.g.* in the case of phase transitions, one structural model for all different phases where the three-dimensional space group of the non-modulated high temperature structure may be derived as a subgroup of the superspace group of the modulated low-temperature structure (Chapuis, 1996; Schönleber *et al.*, 2003). In the case of an n -fold superstructure, the superspace description might improve the data-to-parameter ratio, which is especially important for superstructures with large n .

2. Development and present situation

A glance into the crystallographic literature suggests that aperiodic structures have received much more attention in the past one or two decades than previously. However, the recognition of aperiodic structures is by no means a recent phenomenon in the long history of crystallography. The following section will summarize the developments in this field from early findings to the current status of research.

2.1. Historical perspective

Shortly after the *law of constancy of angles* by Nicolaus Steno heralded the birth of modern crystallography, Jean-

Baptiste Romé-de-Isle proposed in 1772 that crystals are made of microscopic building blocks (the unit cell). René-Just Haüy renewed this thought by proposing that these building blocks are small parallelepipeds and he enunciated the *law of rational indices*, according to which each crystal face is uniquely characterized by three small integers. This law suggested that crystals are built up of a periodic array of those basic structural units in all three dimensions. This three-dimensional periodicity was then believed for almost the next 200 years to be one of the most fundamental principles underlying crystalline architecture.

This concept of periodicity associated with space-group symmetry represents a powerful tool for structural studies and is the basis for successful investigations of crystals by optics and diffraction. Since the discovery of X-ray diffraction by von Laue and co-workers in 1912 (Friedrich *et al.*, 1912, 1913), it was believed that the discrete distribution of diffracted intensities was a direct consequence of the periodic arrangement of the atoms in the three dimensions of ordinary space. However, indications questioning the paradigm of three-dimensional periodicity of crystals appeared all during the history of modern crystallography. Many crystals have been found which do not fulfil the periodicity criterion, but nevertheless give diffraction patterns with perfectly discrete peaks. The discovery of these new types of structures is mainly a result of the problem of trying to index the diffraction peaks with three small integers, but also of applying the law of rational indices to the crystal faces during studies of crystal morphology.

The law of Haüy on rational indices was questioned during optical studies on the mineral calaverite AuTe_2 (Smith, 1903). Later Dehlinger explained the broadening of Debye diffraction lines in metals with a *periodic deformation* of the lattice (Dehlinger, 1927). As it was already known that periodic perturbations in optical gratings generate additional diffraction spots, Dehlinger thus concluded that in analogy a deformation of the crystal lattice should generate secondary lines next to the main lines in the X-ray diffraction pattern, which he called *Gittergeister* (lattice ghosts). While re-investigating calaverite, $\text{Au}_{1-p}\text{Ag}_p\text{Te}_2$ with $p < 0.15$, Goldschmidt and co-workers concluded that the law of rational indices is not a general law and cannot be applied to this mineral (Goldschmidt *et al.*, 1931). While studying the diffraction pattern of the aluminium–copper alloy *Duralumin*, Preston observed additional reflections next to the diffraction peaks of the lattice (Preston, 1938). He called these additional diffraction spots *satellites*. Daniel and Lipson also found additional reflections while studying the alloy Cu_4FeNi_3 (Daniel & Lipson, 1943). They identified a regular deformation of the original cubic lattice as the origin of those. Tanisaki explained additional sharp satellite reflections in the diffraction images of NaNO_2 by a *micro-domain structure* (Tanisaki, 1961, 1963).

de Wolff and his co-workers investigated $\gamma\text{-Na}_2\text{CO}_3$ and found that the additional reflections could be indexed with small integers only if a fourth index was included (Brouns *et al.*, 1964) and that the fourth direction (giving the fourth index) does not coincide with any set of lattice planes of the

main lattice (defined by the first three directions). The structure of $\gamma\text{-Na}_2\text{CO}_3$ was then refined as a *modulated structure* in an harmonic approximation (de Wolff, 1974; van Aalst *et al.*, 1976), *i.e.* the displacive modulation was described with a continuous function by one harmonic wave. In a re-investigation (Dusek *et al.*, 2003) satellite reflections up to fifth order and additional harmonic waves were used to better model the anharmonic features of the structure.

At the same time Nowotny and co-workers identified the so-called chimney ladder compounds $\text{Cr}_{11}\text{Ge}_{19}$, $\text{Mo}_{13}\text{Ge}_{23}$ and $\text{V}_{17}\text{Ge}_{31}$ as *incommensurate intergrowth compounds* (Völlenknecht *et al.*, 1967). Some years later Johnson and Watson determined the atomic displacements, *i.e.* the atomic modulation functions of the molecular composite crystal hepta-(tetrathiafulvalene) pentaoidide, which they called a *dual sublattice system*, via applying a superstructure approximation (Johnson & Watson, 1976).

The final step to recognizing that translational symmetry is not a *conditio sine qua non* for the crystalline state was the discovery of a (non-crystallographic) icosahedral point group in the diffraction pattern of aluminium–manganese alloy (Al with 14 at.% Mn; Shechtman *et al.*, 1984). This alloy was then called a *quasicrystal*. The sharp diffraction spots are explained by a long-range orientational order. This discovery of quasicrystals forced the crystallographic community ultimately ‘... to reconsider the concept of crystals’ (Yamamoto, 1996).

Following this new idea, calaverite was re-investigated as an *incommensurately modulated crystal* and all faces could be indexed by four small integers. Now it was concluded, that ‘(...) the classical law of rational indices still holds for incommensurate crystals, provided the correct number of indices is used (four in the case of calaverite)’ (Dam *et al.*, 1985).

The two common features of all these examples are that the crystal exhibits *flat faces* and that they are characterized by *sharp and well separated peaks*, as is the case for classical three-dimensional periodic crystals. But in contrast to ‘classical’ crystals the examples discussed above lack three-dimensional translational symmetry and consequently are classified in the new category of so-called *aperiodic crystals*. More detailed explanations can be found in reviews like those of Janssen & Janner (1987), Janssen (1988), Bertaut (1990), van Smaalen (1995), Chapuis (1996) and Yamamoto (1996) or in the *International Tables for Crystallography*, Vol. C in ch. 9.8 about *Incommensurate and Commensurate Modulated Structures* by Janssen *et al.* (2006).

2.2. Current status

The increasing number of publications in the past decades on modulated structures described within the superspace approach demonstrates that this concept is well established now and is becoming more and more known, accepted and familiar to the scientific community. This development is also reflected in the two textbooks *Aperiodic Crystals* by Janssen *et al.* (2007) and *Incommensurate Crystallography* by van Smaalen (2007) within the IUCr Monographs on Crystal-

lography series. The interested reader is referred to these two volumes.

The recent technical developments (two-dimensional detectors, powerful computers with large memory and fast processors *etc.*) allow the treatment of large numbers of reflections and complicated structures on an acceptable timescale. The implementation of higher dimensions, *i.e.* more than three indices, to allow handling of satellite reflections is available in almost all commercial and non-commercial diffractometer software packages. The release of *JANA2006* (Petricek *et al.*, 2006) enables not only ‘specialists’ but ‘everybody’ to apply the superspace approach in their research.

Scanning through the reviews and databases on modulated structures, the dominance of inorganic compounds is striking. A search in the ‘*Incommensurate Structures Database*’ at the ‘*Bilbao Crystallographic Server*’ (Aroyo *et al.*, 2006) for example reveals only 15% organic or metal-organic structures. However, despite this dominance of inorganic compounds, there are also some impressive examples of modulated organic materials.

For modulated molecular compounds there is a general agreement in the literature that one possible origin of modulation can be found in the interaction between crystal packing and molecular conformation, *i.e.* in the interplay between intramolecular and intermolecular forces (Dzyabchenko & Scheraga, 2004; Herbststein, 2005). Another reason for the loss of three-dimensional periodicity might be conflicting packing tendencies, even if the molecular conformation is fixed.

To understand the structural features and the chemical behaviour of modulated organic compounds an exact crystal-chemical analysis is necessary, based on an accurate description of the crystal structure. As discussed above, this can be performed in a very elegant and precise way by applying the superspace approach. To illustrate the different implications of the modulation on molecular crystals the following structures are briefly presented:

The classical example of the influence of the crystal packing on the molecular conformation is biphenyl, $C_{12}H_{10}$ (Baudour & Sanquer, 1983; Dzyabchenko & Scheraga, 2004). In the gas phase the torsion angle φ between the two phenyl rings is $42(2)^\circ$ (Fig. 9), which can be presumed to be the optimal conformation of the biphenyl molecule. In the crystal structure at room temperature the molecule is planar ($\varphi = 0^\circ$), at least on average. This change of the molecular conformation is

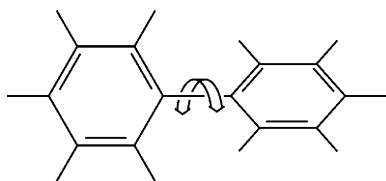


Figure 9

Sketch of the intramolecular torsion in biphenyl $C_{12}H_{10}$, re-drawn after Dzyabchenko & Scheraga (2004).

caused by intermolecular interactions in the crystal. During cooling to lower temperatures ($T < 38$ K) two other crystalline phases appear which are characterized by molecules with $\langle \varphi \rangle \neq 0^\circ$ (Baudour & Sanquer, 1983). In particular, for the incommensurately modulated low-temperature structure at $T = 20$ K the average torsion angle varies between $\varphi = -11^\circ$ and $\varphi = +11^\circ$. The modulation of the structure can be explained as the result of a competition between intra- and intermolecular contributions to the lattice energy (Dzyabchenko & Scheraga, 2004).

A similar behaviour with respect to intramolecular torsion can be found in 4,4'-dichlorobiphenylsulfone, $(ClC_6H_4)_2SO_2$ (Zúñiga & Criado, 1995) and in 2-phenylbenzimidazole, $C_{13}H_{10}N_2$ (Zúñiga *et al.*, 2006). In these structures, too, the modulation is interpreted as a consequence of the interaction between intramolecular forces and the crystal packing. For 2-phenylbenzimidazole, for example, a variation in the central torsion angle of $\pm 5^\circ$ is found.

In contrast, the incommensurately modulated structure of 3,4-diphenyl-2a,5a,6,7,8,8a,8b-heptahydro-furo[4,3,2-de]-chromen-2-one, $C_{22}H_{20}O_3$, cannot be explained by a torsion of the molecular conformation (Guiblin *et al.*, 2006). Here the molecule can be divided into three rigid bodies. Two of these are phenyl rings, but only one of them forms an intermolecular contact with a neighbouring molecule *via* a $C-H \cdots O$ interaction. Due to the modulation the complete molecule is shifted and the contact acts as a kind of mechanical spring. As a consequence, the shifts of the two phenyl rings with respect to the rest of the molecule are slightly different, and the origin of the modulation lies in the mutual frustration of molecular conformation and hydrogen-bond formation (Guiblin *et al.*, 2006).

The incommensurately modulated structure of the organic salt quinium (*R*)-mandelate, $C_{20}H_{25}N_2O_2^+ \cdot C_8H_7O_3^-$, which exhibits a strongly pronounced displacive modulation (Schönleber, 2002; Schönleber & Chapuis, 2004), is slightly more complicated – different parts of the quinium cation follow different modulation functions: the quinium segment, for example, shows a rather undulating movement, while the vinyl group exhibits two preferred orientations, which are occupied as a function of the phase t of the modulation. As a consequence not only does the conformation of the molecules vary, but, due to the different atomic environments, the anisotropic ADPs also vary. Furthermore, the (intermolecular) hydrogen-bond scheme is modulated, which means that the hydrogen bonds are formed and broken as a function of t (see §1.3).

Other recent examples of modulated organic structures comprise the channel inclusion compound of nonadecane/urea, $CH_3(CH_2)_{17}CH_3/(NH_2)_2CO$, which was refined as an incommensurate host-guest system at low temperature (Toudic *et al.*, 2008) with an alternating intermodulation between host and guest substructures from channel to channel, and the modulated structure of 1-phenyl-4-cyclohexylbenzene, $C_{14}H_{20}$ (Evain *et al.*, 2009), in which the cyclohexyl group is rotated with respect to the benzene ring.

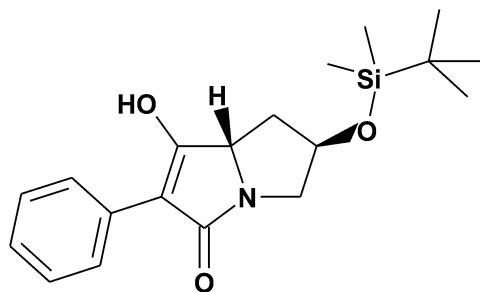


Figure 10
Structural formula of (6*R*,7*aS*)-6-(*tert*-butyldimethylsilyloxy)-1-hydroxy-2-phenyl-5,6,7,7*a*-tetrahydropyrrolizin-3-one, $C_{19}H_{27}NO_3Si$ (VAR205).

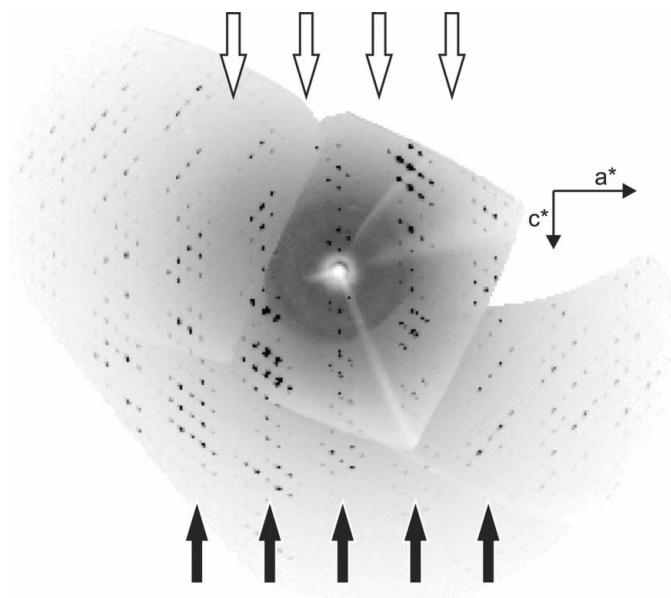


Figure 11
Reconstruction of the $h0l$ layer of the diffraction pattern of VAR205. Note the alternating vertical bands of spots with high intensity (black arrows) and no intensity (white arrows). The reconstruction was calculated using the *APEX2* suite of programs (Bruker AXS, 2007*d*).

3. Modulated structures – a case study

In the following section, the modulated crystal structure of (6*R*,7*aS*)-6-(*tert*-butyl-dimethylsilyloxy)-1-hydroxy-2-phenyl-5,6,7,7*a*-tetrahydropyrrolizin-3-one (from now on referred to by its Novartis code VAR205), an organic compound (Fig. 10) with pharmaceutical impact, will be described and discussed. A detailed description of the complete procedure of structure analysis starting with the determination of the \mathbf{q} vector from the diffraction pattern will be given. Intensity data were collected at $T = 100$ K on a Bruker AXS three-circle goniometer with a SMART6000 CCD detector, using Cu $K\alpha$ radiation from a fine focus rotating anode generator equipped with Montel multilayer mirrors (Table 1). A standard data collection protocol was followed with 12 180° ω scans (scan width 0.6°) at different φ positions and two different detector positions (crystal-to-detector distance 40 mm). The software utilized comprises the standard Bruker AXS suite *SMART*,

Table 1

Experimental data.

Crystal data	
Chemical formula	$C_{19}H_{27}NO_3Si$
Chemical formula weight	345.51
Temperature (K)	100
Crystal system, superspace group	Monoclinic, $P2_1(\alpha 0 \gamma)0$
a, b, c (Å)	6.425 (2), 6.522 (2), 22.716 (6)
β ($^\circ$)	90.739 (11)
Modulation wavevector \mathbf{q}	[0.1422 (2), 0, 0.3839 (8)]
V (Å ³)	951.8 (5)
Z	2
D_x (Mg m ⁻³)	1.2052 (6)
$F(000)$	372
Data collection	
Diffractometer	Bruker AXS three-circle goniometer with SMART6000 CCD detector
Radiation type	Cu $K\alpha$
Absorption correction	Semi-empirical from equivalents
μ (mm ⁻¹)	1.215
Data collection method	ω scans
Range of h, k, l, m	$-8 \leq h \leq 8$ $-6 \leq k \leq 7$ $-28 \leq l \leq 28$ $-4 \leq m \leq 4$
No. of measured, independent and observed reflections	52 466, 28 265, 24 813
Criterion for observed reflections	$I > 3\sigma(I)$
$R_{int,all}, R_{int,obs}$	0.023, 0.020
$R_{int,all}, R_{int,obs}$ (main reflections)	0.017, 0.017
$R_{int,all}, R_{int,obs}$ (first-order satellite reflections)	0.023, 0.023
$R_{int,all}, R_{int,obs}$ (second-order satellite reflections)	0.020, 0.020
$R_{int,all}, R_{int,obs}$ (third-order satellite reflections)	0.028, 0.027
$R_{int,all}, R_{int,obs}$ (fourth-order satellite reflections)	0.048, 0.041

RLATT, *SAINT* (Bruker AXS, 2007*a,b,c*) and *SADABS* (Bruker AXS, 2001),³ the *SHELXTL* program package (Sheldrick, 2008), and the crystallographic computing system *JANA2006* (Petricek *et al.*, 2006). Graphical representations of molecules were generated with *PLATON* (Spek, 2003) and *DIAMOND* (Brandenburg, 1999).

VAR205 is a key intermediate in the synthesis of potential lymphocyte function-associated antigen 1 inhibitors (Zecchinon *et al.*, 2006) bearing a pyrrolizinone scaffold (Baumann, 2007). Structural analysis was carried out to establish the relative stereochemistry at the fused ring core and the degree of pyramidalization at the central nitrogen atom.

3.1. The diffraction pattern

The diffraction pattern of VAR205 shows irregularities – visible either with a suitable program that allows a schematic representation of reflections in reciprocal space (in the present study *RLATT*) or in a reconstruction of reciprocal space (Fig. 11): indexing using all observed spots (*i.e.* a supercell approach) results in a huge monoclinic unit cell with a volume

³ Please note that many other commercial and non-commercial software packages are also capable of handling diffraction patterns of modulated structures with satellite reflections.

of $30\,500\text{ \AA}^3$, which would correspond to approximately 70 molecules per unit cell. Such large Z values usually indicate either a wrong unit cell, a twinned crystal or a modulated structure. A closer look at various orientations of the diffraction pattern reveals a distinction between reflections based on their intensities: layers of very strong spots are surrounded in a non-arbitrary way by weaker reflections. This intensity pattern strongly suggests that the possibility of having a twinned crystal can be eliminated and that the crystal structure of VAR205 is modulated.

As discussed in §1 there are three ways of handling the diffraction pattern of a modulated structure: indexing the basic cell with main reflections only (ignoring the satellite reflections and therefore as a consequence neglecting a large fraction of diffracted intensities), indexing the supercell using all reflections equivalently (which is in the case of incommensurateness only an approximation) and indexing the basic cell followed by the determination of the modulation wavevector \mathbf{q} . These options along with their subsequent steps are summarized in the flowchart in Fig. 12. In many cases, especially when the modulation is not too strong, the fastest way to obtain structural information and an idea of the extent of the modulation is to start with the basic cell and try to determine the average structure.

3.1.1. Basic cell and average structure. The main reflections in the diffraction pattern of VAR205 yield a monoclinic basic cell with a volume of 950 \AA^3 , corresponding to two molecules per unit cell (Fig. 13). The satellite peaks were eliminated ‘by hand’ with *RLATT* from the peak list, the remaining main

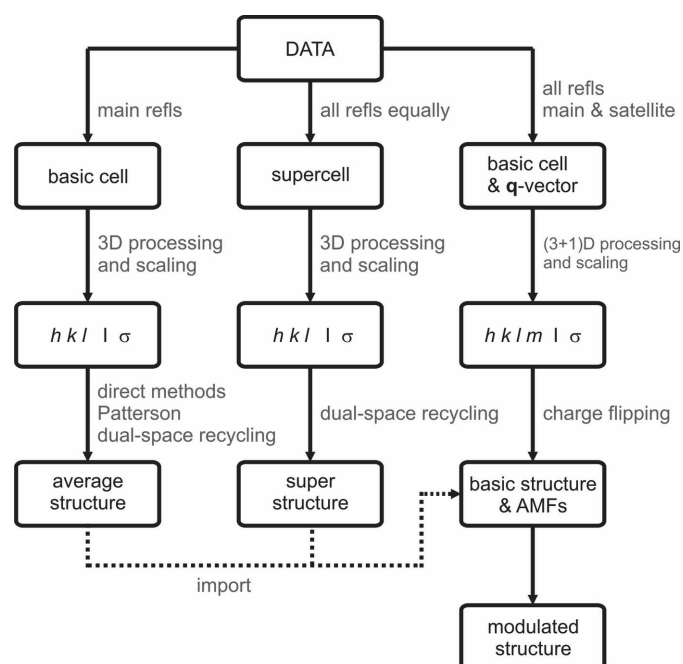


Figure 12 Flowchart illustrating different possible ways of handling the diffraction data of one-dimensionally modulated structures. Coordinates obtained from the average structure or the superstructure (after transformation) can be used as a starting model for the refinement in $(3+d)$ -dimensional superspace.

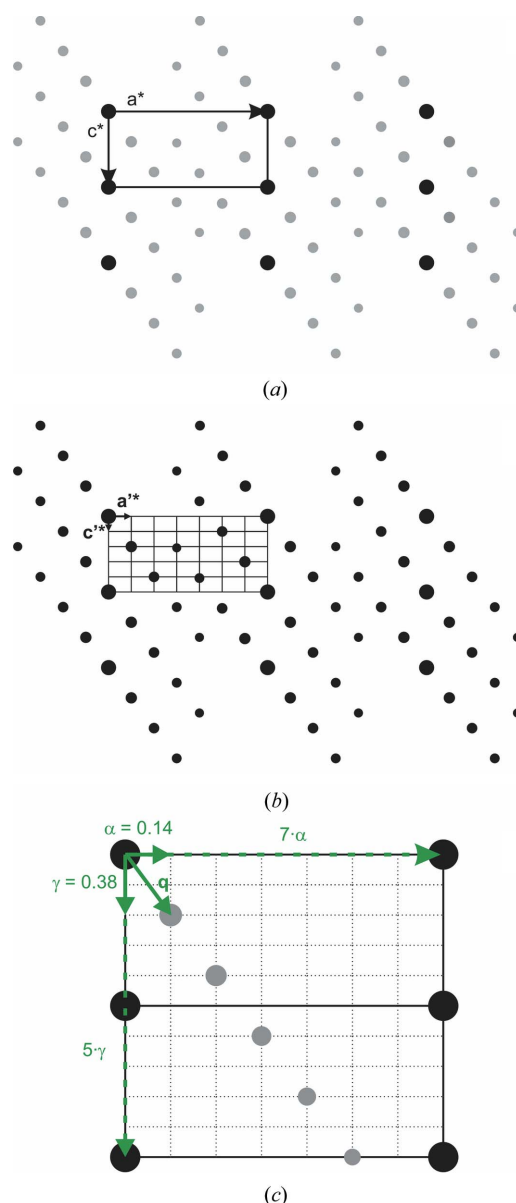


Figure 13 Schematic depiction of the diffraction pattern of VAR205 with main and satellite reflections and modulation wavevector $\mathbf{q} = 0.14\mathbf{a}^* + 0.38\mathbf{c}^*$. (a) Neglecting satellite reflections (light grey) results in a three-dimensional unit cell of $\mathbf{a} = 6.4$, $\mathbf{b} = 6.5$, $\mathbf{c} = 22.7\text{ \AA}$ (average structure). (b) Using all reflections equivalently results in a three-dimensional unit cell of $\mathbf{a}' = 7\mathbf{a} = 44.8\text{ \AA}$, $\mathbf{b}' = \mathbf{b} = 6.5\text{ \AA}$ and $\mathbf{c}' = 5\mathbf{c} = 113.5\text{ \AA}$ (supercell). The volume of this superstructure unit cell is 35 times that of the unit cell of the average structure. Once the \mathbf{q} vector is known the unit cell can also be derived by looking for the smallest integer numbers with which the components of the \mathbf{q} vector (approximated by rational numbers, here: $2/5$ and $1/7$) have to be multiplied to reach a lattice point of the main reflections (c). Those integers have to be applied to the cell parameters of the small unit cell from (a) to give the big unit cell from (b).

reflections were re-imported into the indexing routine. Data processing and scaling results in classical three-dimensional data, which can then be inspected for systematic absences to determine the space-group symmetry of the main reflections. The reflection condition $0k0$, $k = 2n$ and the enantiopurity of

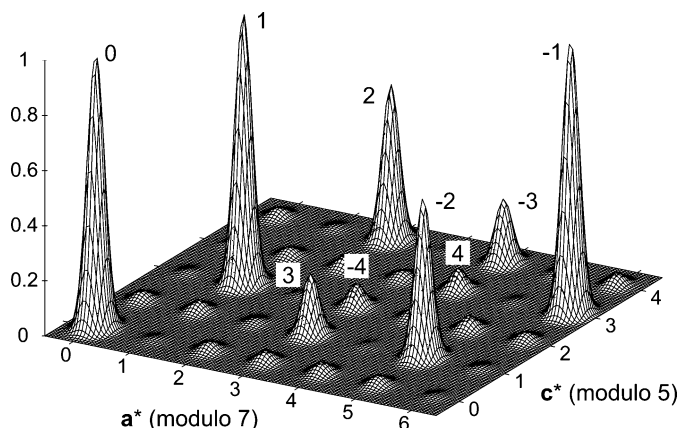


Figure 14

Intensity distribution in the reciprocal supercell of VAR205. All measured intensities were summed up and collapsed into one unit cell (compare with Figs. 11 and 13b). Corresponding orders of the reflections in the superspace approach are given. Note that the intensity of the first-order satellite reflections is of the same order of magnitude as the intensity of the main reflections, indicating a very strong modulation, and that many of the reciprocal lattice points have no measurable intensity.

the compound led to the monoclinic space group $P2_1$ with $Z = 2$. All attempts to solve the structure with classical direct (*SHELXS*) or dual-space recycling methods (*SHELXM*) failed, which was the first indication of a strong modulation.

If an average structure can be obtained this way, the coordinates from the structural (isotropic) refinement can later be used as a starting model for the refinement of the modulated structure (to find the proper shapes of the AMFs). Useful information can also be obtained from any disorder sites. Equally important for a later (3 + 1)-dimensional treatment of the problem is the knowledge of the space group. If $P2_1$ is the space group of the average structure (even if the structure cannot be determined this way) $P2_1$ will very likely also be the basis of the superspace group symbol (see §3.2.1).

If the modulation is weak, it is often not only possible to solve the average structure but also to refine it to very reasonable agreement factors, usually, however, at the cost of large displacement parameters or extensive disorder modeling.

3.1.2. Supercell and superstructure. For the supercell processing the orientation matrix can either be obtained directly by indexing all reflections or, if this is not possible, by combining the information from the basic cell parameters and \mathbf{q} -vector components (Fig. 13). In the case of VAR205 the three \mathbf{q} -vector components are $\alpha = 0.143$ (2), $\beta = 0$ and $\gamma = 0.384$ (8) (cf. §3.1.3). To obtain an appropriate supercell, the non-zero components will be approximated by $\alpha = 1/7$ (≈ 0.1429) and $\gamma = 2/5$ ($= 0.4000$), then the respective lattice parameters will be multiplied by the denominators, i.e. \mathbf{a} by 7 and \mathbf{c} by 5. As a result a supercell is obtained with a volume which is $7 \times 5 = 35$ times that of the basic cell. Please note

⁴ The chemical synthesis of VAR205 started from naturally occurring hydroxyproline, the chiral carbon C09 remained unchanged in all synthetic steps. This was verified in one of the final refinement cycles [Flack $x = 0.04$ (1)].

again that in this incommensurate case the supercell represents only an approximation: the value for $\gamma = 0.384$ (8) was approximated with $\gamma = 2/5$; another possible approximation could be $\gamma = 3/8$ ($= 0.375$), resulting in a $7 \times 8 = 56$ -fold supercell with a somewhat different description of the atomic arrangement for the same crystal structure!

Data processing and scaling can be pursued in the same way as for a classical three-dimensional periodic structure. However, one has to be aware of the fact that even if satellite reflections to high order are observed, most of the reciprocal lattice points are 'empty', i.e. no intensity will be detectable (Fig. 14). Accordingly, profile fitting will be poor, average intensities during data processing will be low and a large percentage of the reflections will be classified as weak or unobserved.

Nonetheless, there can be reasons to use a supercell integration and then later transform the resulting three-dimensional hkl file into higher-dimensional data: conventional three-dimensional software can be used and all data are available at any time so that e.g. the decision up to which order satellite reflections shall be included can be postponed.

In the present example of VAR205 structure solution was possible for the commensurate 7×5 -fold approximation in the space group $P2_1$ using the dual-space recycling algorithm employed in *SHELXM*. Subsequent cycles of refinement and difference-Fourier map calculations allowed completion of the model (35 independent molecules = 840 non-H atoms). Owing to the large number of unobserved reflections the data-to-parameter ratio was very poor and anisotropic refinement (7560 parameters!) in *SHELXL* was only possible using block-diagonal or conjugate-gradient least-squares instead of full-matrix least-squares. Nonetheless, it was learned from the superstructure approximation that the molecules show a strong modulation along \mathbf{a} (Fig. 15).

Even if the refinement of a superstructure could be carried out with a good data-to-parameter ratio to acceptable agreement factors, it is important to emphasize once again that the superstructure solution in the present case can only be regarded as an approximation. Since γ is equal to 0.384 (8), treating the data in a 7×8 -fold supercell with $\gamma = 3/8$ is at least as valid as in the 7×5 -fold supercell with $\gamma = 2/5$

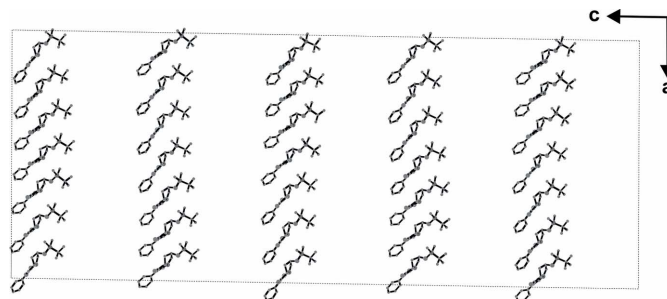


Figure 15

Superstructure of VAR205 as obtained from structure solution and refinement in the 7×5 -fold supercell approximation (H atoms omitted for clarity). View along \mathbf{b} , 35 molecules in asymmetric unit shown. Note the sinusoidal modulation along \mathbf{a} .

discussed above. This problem occurs whenever the \mathbf{q} -vector components are between two rational numbers, *e.g.* 0.186, which is as close to 1/5 as it is to 1/6, so that neither a fivefold nor a sixfold supercell will mirror the true structure (see, for example, the discussion in Schönleber & Chapuis, 2004). Usually this problem will reveal itself in strange geometric and/or displacement parameters in one or more of the molecules in the asymmetric unit.

In principle, one of the molecules in the asymmetric unit could be used as a starting model for the refinement of the structure in superspace (in analogy to the average structure of §3.1.1) by transforming the coordinates into the basic cell (in the present case by taking, for example, the molecule closest to the origin and multiplying the fractional coordinates x by 7 and z by 5). If the modulation is not too strong, *i.e.* if the amplitudes of the AMFs are not too large, the probability is high of finding appropriate starting values for the AMFs in the first cycles of refinement. In general, a starting model for (3 + 1)-dimensional refinement derived from the superstructure can be assumed to be better than a starting model from the three-dimensional refinement of the average structure in the basic cell. As in the present case the modulation is rather strong it was, however, not possible to establish a good structural model with proper AMFs for all atoms starting from

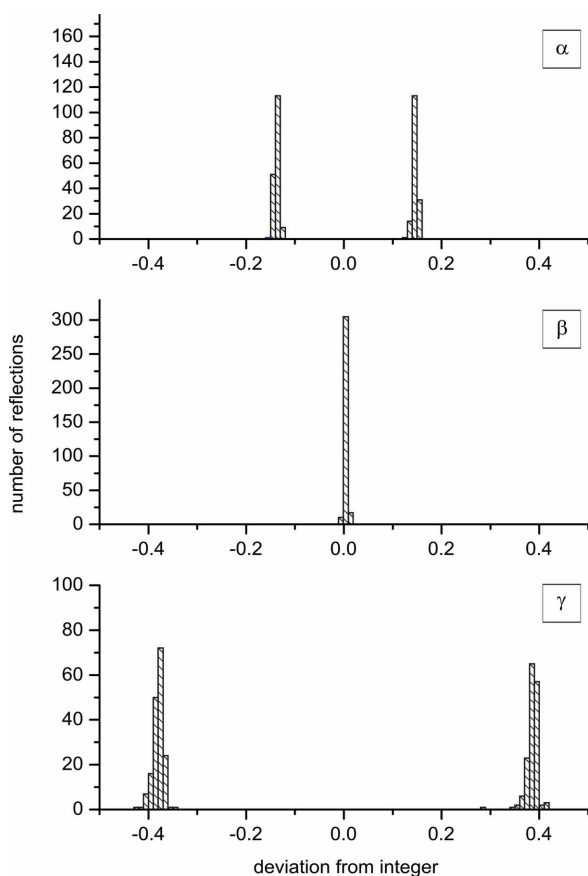


Figure 16
Histogram of the deviations from integer numbers of the first-order satellite reflections of the compound under study as obtained during the indexing routine. It clearly shows a modulation wavevector ($\alpha\gamma$) with the components α and γ having irrational values.

the fractional coordinates of one molecule of the superstructure approximation.

3.1.3. Indexing in superspace. The crucial step for all subsequent action in superspace is the determination of a suitable \mathbf{q} vector from the original raw data. For this purpose it is important to separate main reflections from satellite reflections and obtain fractional indices (*i.e.* hkl values of the closest main reflection and the deviations from these integers) for the latter based on the orientation matrix of the basic cell defined by the main reflections. If this procedure is carried out with first-order satellites only (the higher-order satellite peaks were eliminated from the reflection list with *RLATT*), a very clean and clear distribution of fractional indices (Fig. 16) was obtained (*via* the indexing routine in *SMART*) resulting in an initial \mathbf{q} vector of (0.14, 0, 0.38).

Once the starting values for the \mathbf{q} -vector components have been determined this way, the data can be processed using the orientation matrix of the basic cell along with these \mathbf{q} -vector components and a maximum satellite order n_{\max} , up to which satellite peaks should be processed.

Depending on the integration software, the output format of the resulting hkl file can vary. Using *SAINT*, the resulting reflection file is always (3 + 3)-dimensional (indices h, k, l, m, n, p). In the present (3 + 1)-dimensional case it contains main reflections ($m = n = p = 0$) and satellite reflections ($m \neq 0, n = p = 0$). Overlapping satellite reflections (which could occur with one of the \mathbf{q} -vector components close to 0.5) are treated the same way as overlapping reflections from a twinned crystal, *i.e.* a 'twin component' batch number is given at the very end of the lines in the reflection file.

The raw reflection files can then be scaled and corrected for absorption effects, *e.g.* with *SADABS* (unless the file contains overlapping reflections, this option is not yet implemented), and the resulting [again: (3 + 3)-dimensional] hkl file can be directly imported into *JANA2006*. The guided input routine of *JANA2006* (file > structure > new) leads through data reduction, analysis of systematic absences and space-group determination. It closes with the option to attempt a structure solution in higher-dimensional space *via* the external program *SUPERFLIP* (Palatinus & Chapuis, 2007).

3.2. Structure solution and completion in superspace

Just as for three-dimensional periodic crystals the first step in the structure solution process for modulated structures in superspace is the determination of the correct superspace group. This can be done manually or, as mentioned above, automatically during the data importing routine within *JANA2006* by checking the systematic absences for both main and satellite reflections. The superspace groups for (3 + 1) dimensions are tabulated along with their special reflection conditions in the *International Tables for Crystallography*, Vol. C in ch. 9.8, Incommensurate and Commensurate Modulated Structures, by Janssen *et al.* (2006).

3.2.1. Superspace-group determination. A first glance at Table 9.8.3.5 '(3 + 1)-dimensional superspace groups' in the *International Tables for Crystallography*, Vol. C, suggests that

the superspace-group symbol consists of the basic space-group symbol (here: $P2_1$) extended by an expression in parentheses which denotes the general form of the \mathbf{q} vector followed by additional letters.

However, this is only partially true. The Bravais symbol refers now to a $(3 + 1)$ -dimensional lattice. Looking at superspace group No. 53, $B2(00\gamma)s$, for example, the letter B does not mean an additional centring of $(\frac{1}{2}, 0, \frac{1}{2})$ as in three-dimensional space groups, but of $(\frac{1}{2}, 0, \frac{1}{2}, 0)$ since the centring can also occur along the fourth dimension. If this is the case, the Bravais symbol B' (or X) is used to describe the additional centring of $(\frac{1}{2}, 0, \frac{1}{2}, \frac{1}{2})$. The letters following the \mathbf{q} vector indicate additional translational components of the symmetry operators along the fourth dimension (screw and glide components: if there is none, a '0' is written, an 's' stands for a translation by $\frac{1}{2}$). For a more sophisticated discussion the interested reader is referred to ch. 3 of the IUCr textbook *Incommensurate crystallography* by van Smaalen (2007).

In the case of VAR205 the choice of the superspace group is straightforward: the only special reflection conditions are $0k00$ with $k = 2n$, indicating the twofold screw axis along \mathbf{b} (which is already known from the basic cell). Therefore, the lattice in superspace is primitive, the resulting superspace group is $P2_1(\alpha 0\gamma)0$.

Now imagine, for example, a monoclinic \mathbf{b} unique lattice with a \mathbf{q} vector $(\alpha 0\gamma)$ and systematic absences for the reflections $0k0m$ with $m \neq 2n$. This indicates a screw component of the twofold axis along the fourth dimension, the resulting superspace group would be $P2(\alpha 0\gamma)s$. The systematic absences might also be $0k00$ with $k \neq 2n$ and $h0lm$ with $m = 2n$. Here the first condition indicates as above a twofold screw axis along \mathbf{b} , the second one a glide plane perpendicular to \mathbf{b} with a glide component along the fourth dimension. The resulting superspace group would be $P2_1/m(\alpha 0\gamma)0s$.

3.2.2. Structure solution. Having established the superspace group, the input file for *SUPERFLIP* can be generated and the program can be called directly from within *JANA2006*. *SUPERFLIP* extends the charge-flipping algorithm introduced by Oszlányi & Sütö (2004, 2008) to an arbitrary number of dimensions and can therefore solve modulated structures *ab initio* in higher-dimensional space. This capability is especially important if no three-dimensional structure (average structure or superstructure) is available from which a starting model for the basic structure can be deduced. *SUPERFLIP* generates a $(3 + 1)$ -dimensional electron-density map which is then searched for atoms by *JANA2006*. Since atoms in $(3 + 1)$ -dimensional superspace have to be imagined as one-dimensional waves, the interpretation of the electron-density maps already requires the use of at least one modulation wave (automatically employed by *JANA2006*), *i.e.* the program puts out starting values for the fractional coordinates of the atoms and starting values for the parameters describing the AMFs.

From the *SUPERFLIP* solution of the present example a peak list was extracted from which the first six peaks were selected based on peak height and interatomic distances (see discussion below) as silicon (peak1), oxygen (peaks 2–3) and

carbon (peaks 4–6) atoms. Starting from this set of six atoms the model was completed in subsequent cycles of refinement, Fourier-map calculation, and peak search followed by manual identification of possible new atomic positions.

Analyzing the surroundings of atoms in modulated structures differs considerably from doing so in non-modulated structures: for every distance of interest there would be an infinitely long table of values, owing to the missing translational symmetry. Here the superspace concept with the modulation parameter t (indicating the phase of the modulation, see §1.3 and Fig. 7) provides a very convenient way to visualize and analyse structural properties such as interatomic distances. A listing of a certain distance as a function of t , *i.e.* varying t from 0 to 1, provides all values for this distance occurring anywhere in the structure (van Smaalen, 2004). In this context it is important to keep in mind that the extent of variation which is found in the geometric parameters of modulated structures resembles the findings from packing analyses of non-modulated structures: bond lengths are expected to show less variation than bond angles, which again vary less than torsion angles and intermolecular distances.

There are two ways of checking the variation of a structural parameter with t in *JANA2006*: the subroutine *Dist* produces a listing of the structural parameter in user-defined intervals, the subroutine *Graph* provides a plot of this relation. When looking for missing atoms in a structure based on a Fourier-map peak list it is usually faster to tabulate the distances at intervals of 0.1 or 0.2 in t than to make graphs. Fig. 17 shows the results of such a search in the vicinity of *e.g.* atom Si01. Atom C03 is one of the carbon atoms bonded to Si01 and at this early stage of refinement shows a rather large variation in the bonding distance to the silicon atom (*cf.* the difference of 0.111 Å between minimum and maximum value). Peak max8 from the Fourier peak list to Si01 varies from 1.2 to 2.8 Å and therefore will probably not be a suitable candidate for one of the missing C atoms. Peak max7, on the other hand, shows a relatively even distribution of distances and therefore will be selected as carbon atom C04. Another cycle of refinement, Fourier-map calculation and peak search would then follow.

In three-dimensional space refining a structure means finding values for the coordinates x, y, z and the six anisotropic atomic displacement parameters U^{ij} which result in the best agreement of F_{obs} and F_{calc} . In superspace this concept has to be extended: the refined coordinates x, y, z (or better: x_1, x_2, x_3) are the positions of the atoms in the basic structure and, in addition, for each atom the deviation from this average position has to be determined and refined. This deviation can be expressed by a variety of AMFs (*cf.* Fig. 2), which, in the case of a superposition of harmonic waves up to second order for the x_1 coordinate looks like

$$f(x_4) = x_1 + A_1 \cdot \cos(2\pi \cdot 1 \cdot x_4) + B_1 \cdot \sin(2\pi \cdot 1 \cdot x_4) + A_2 \cdot \cos(2\pi \cdot 2 \cdot x_4) + B_2 \cdot \sin(2\pi \cdot 2 \cdot x_4). \quad (1)$$

The coefficients A_1, B_1 and A_2, B_2 are the amplitudes of the harmonic waves of first and second order. These coefficients are determined by *JANA2006* on the basis of the diffraction

data and occur as parameters in the atomic parameter file m40. Therefore, for each of the positional parameters (fractional coordinates x_1 , x_2 and x_3) two more parameters per harmonic wave will occur to describe the AMF which characterizes the deviation from the basic position along the fourth dimension.

During the structure completion cycle it is important to verify for the atoms in the structural model that their position has been determined correctly and that their AMFs follow the observed electron density. This procedure resembles very much the $2F_{\text{obs}} - F_{\text{calc}}$ map calculation and inspection that structural biologists use to check if their protein model is placed correctly and fits the observed electron density. For this purpose an F_{obs} map is calculated in superspace around the atom of interest as a function of x_4 and an overlay of the map and the AMF is generated with the subroutine *Contour* (Fig. 18). Here again the two-dimensional projection technique introduced in Fig. 7 is used: the vertical axis always represents the superspace coordinate x_4 (along \mathbf{a}_{s4}), the second axis is one of the three remaining axes \mathbf{a}_{s1} , \mathbf{a}_{s2} or \mathbf{a}_{s3} with their fractional coordinates x_1 , x_2 or x_3 (as the angle between the two axes is defined *via* the respective \mathbf{q} -vector component, the images differ). The three sections in Fig. 18 show clearly that there is density around the position of Si01 and that the AMF follows the observed electron density. It

atom Si01					

C03.....	1.601(5)	max7.....	1.8371(9)	max8.....	1.6299(13)
ave	1.607(9)	ave	1.900(3)	ave	1.838(3)
min	1.546(9)	min	1.723(3)	min	1.200(3)
max	1.657(9)	max	2.095(3)	max	2.800(3)
t	distance	t	distance	t	distance
0.000	1.622(9)	0.000	1.928(3)	0.000	1.637(3)
0.200	1.656(9)	0.200	2.094(3)	0.200	1.253(3)
0.400	1.635(9)	0.400	1.886(3)	0.400	1.300(3)
0.600	1.565(9)	0.600	1.733(3)	0.600	2.265(3)
0.800	1.557(9)	0.800	1.858(3)	0.800	2.747(3)
1.000	1.622(9)	1.000	1.928(3)	1.000	1.637(3)

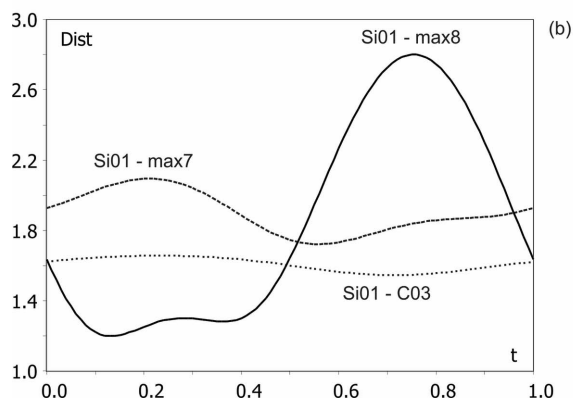


Figure 17
Distances of atom C03 and Fourier peaks max7 and max8 from Si01 along the phase of the modulation t as obtained from the *JANA2006* subroutine *Dist* (a) and *Graph* (b). The large variation between minimum and maximum distance for max8 indicates that this peak might not be a suitable new atomic position.

also becomes obvious, however, that, at a later stage, higher harmonic functions will have to be added to model the details of the electron density better. Please note again the one-dimensional character of the atomic domain along \mathbf{a}_{s4} , represented by its electron density and its AMF (red line).

When the structural model has been completed (*i.e.* all atoms have been found and assigned) according to the above-mentioned procedure the refinement is completed by adding harmonic functions of higher order, introducing anisotropic ADPs and calculating H-atom positions. For VAR205 R_{obs} for all (main and satellite) reflections at this stage had already dropped to 0.146, compared with 0.400 for the first model which included only six atoms.

3.2.3. Completion of refinement. In the case of a strong modulation, it is generally advantageous to use not only harmonic waves of first and second order, but also of the third order for the AMFs of the positional parameters before switching to anisotropic refinement. By summing up the higher-order waves, it is possible to better model the details of the electron density. In Fig. 19 an example is shown where a hypothetical 'electron-density distribution' is fitted quite satisfactorily using a superposition of three harmonic waves of higher order, while the first-order wave alone is a rather rough approximation. Introducing anisotropic ADPs too early can hinder the proper determination of the positional AMFs.

In the refinement process for VAR205 three harmonic waves for the positional AMFs were added ($R_{\text{obs}} = 0.083$) before switching to anisotropic ADPs ($R_{\text{obs}} = 0.070$). Just as in the refinement of a three-dimensional periodic structure it is important to make sure at all times that the introduction of additional harmonic waves (= more parameters) is justified by a statistically significant drop in the crystallographic agreement factors.

Up to this point only AMFs for the positional parameters were used, *i.e.* only the so-called **displacive modulation**, which

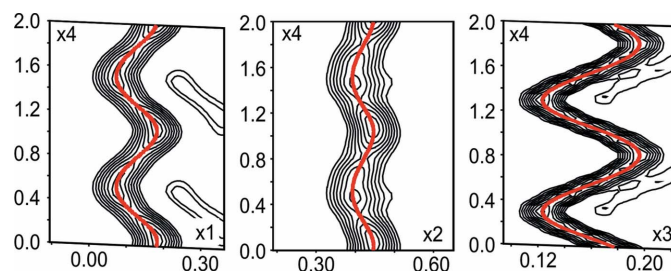


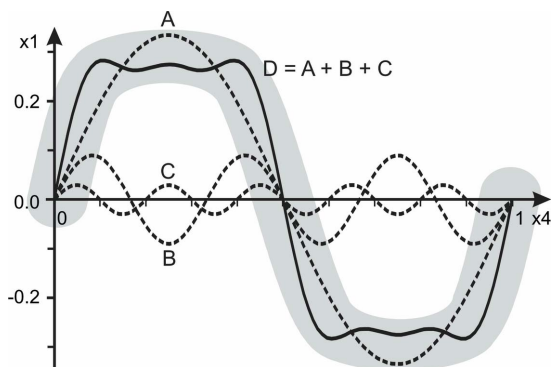
Figure 18
Fit of AMFs of Si01 (red lines) to the electron density for all three superspace sections. Each of the three diagrams is a variation of the superspace representation shown in Fig. 7, to which the observed electron density has been added. For a better illustration two periods are drawn along x_4 . Note that not all the details of the electron density can be accurately described with only one harmonic function. The horizontal width corresponds to 3 \AA in all three pictures, showing the different amplitudes along the three dimensions of physical space and the large amplitude of the deviation from the average position along x_3 . Hypothetical examples of electron-density distributions requiring the use of sawtooth and crenel type AMFs are shown in Fig. S2 of the supplementary material.

Table 2

Refinement data.

Refinement on	F
$R_{\text{all}}, wR_{\text{all}}$	0.042, 0.042
$R_{\text{obs}}, wR_{\text{obs}}$	0.036, 0.041
$R_{\text{obs}}, R_{\text{all}}, wR_{\text{all}}$ (main reflections)	0.033, 0.034, 0.043
$R_{\text{obs}}, R_{\text{all}}, wR_{\text{all}}$ (first-order satellite reflections)	0.031, 0.033, 0.039
$R_{\text{obs}}, R_{\text{all}}, wR_{\text{all}}$ (second-order satellite reflections)	0.032, 0.035, 0.038
$R_{\text{obs}}, R_{\text{all}}, wR_{\text{all}}$ (third-order satellite reflections)	0.040, 0.049, 0.041
$R_{\text{obs}}, R_{\text{all}}, wR_{\text{all}}$ (fourth-order satellite reflections)	0.063, 0.093, 0.064
$GoF_{\text{obs}}, GoF_{\text{all}}$	2.52, 2.38
No. of restraints	2
No. of parameters	1396
Weighting scheme	$w = [\sigma^2(F) + (0.01F)^2]^{-1}$
$(\Delta/\sigma)_{\text{max}}$	-0.0082
$\Delta\rho_{\text{max}}, \Delta\rho_{\text{min}}$ ($e\text{\AA}^{-1}$)	0.27, -0.39
Extinction correction	Gaussian isotropic type I

affects the positions of the atoms, was considered. However, just as the coordinates x_1 , x_2 and x_3 can be modulated, other atomic parameters such as the occupation factors or the ADPs can also vary periodically throughout the structure, owing to the varying environments of the atoms. Again, these modulations can be described by harmonic, sawtooth or crenel functions (*cf.* Fig. 2). **Occupational modulation** is frequently encountered in alloys but can also occur when, for example, solvent positions in organic modulated structures are not always fully occupied and the amount to which they are occupied varies periodically. ADPs, on the other hand, tend to be modulated whenever a strong displacive modulation is present, because a change in interatomic distances and the variation of the atomic surroundings often go hand in hand with a change in the vibrational amplitudes. Accordingly AMFs can also be added for the atomic displacement parameters U^{ij} (**thermal modulation**). Since six values describe the shape of the anisotropic displacement ellipsoid, every additional harmonic wave for the AMFs will add 12 more para-

**Figure 19**

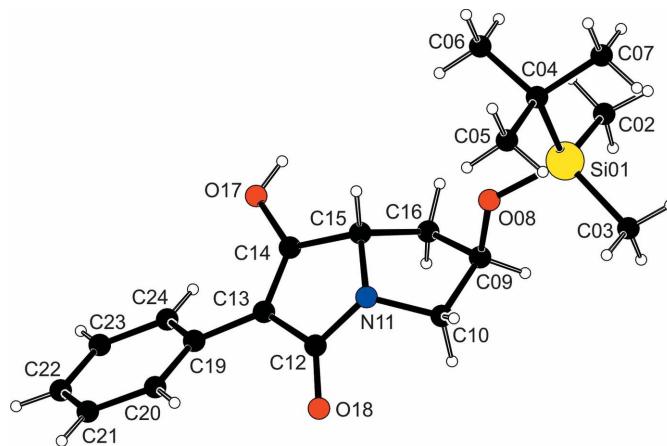
Summing up harmonic waves of higher order can help to model the path of the electron density more accurately with the AMF. In the depicted example the (hypothetical) electron density (grey) is better fit with a superposition (D) of the three waves A, B and C. Using only the harmonic wave of first order (A) does not allow the description of the anharmonic details.

meters to the refinement [one sine and one cosine component, see (1)]. Again, it is important to constantly monitor improvements in the fit of the electron density and the agreements factors to check whether the introduction of new higher harmonic waves is reasonable.

In the case of VAR205, in the final stages of refinement four harmonic waves for the displacive and two harmonic waves for the ADP modulation were used and along with the addition of H atoms (riding model or free refinement possible) the R factor for all (main and satellite) reflections dropped to $R_{\text{obs}} = 0.036$.

3.3. Interpretation and presentation of results

While presenting the refinement results for a modulated structure, it is important to keep in mind that there is not just 'one bond length' or 'one crystal environment' for a given group of atoms. Often, the first step following the summary of the refinement details is to divide the molecule into so-called rigid units (if not already done during the refinement process, see §3.3.2) and describe their movements relative to each other. During the discussion of individual geometric parameters of interest it is necessary to pay attention to the fact that all these values vary with the phase of the modulation t around a mean value which usually corresponds to the values expected for a non-modulated structure. Sections of electron density can be used to show that the determined AMFs adequately describe the experimental electron density (resembling the $2F_{\text{obs}} - F_{\text{calc}}$ maps in structural biology papers) and to provide a graphical representation of the extent of modulation present in the structure. If there is interest in the crystal packing, the atomic coordinates of the modulated structure can be exported as a commensurate superstructure. Here, however, the same approximations have to be accepted which have already been discussed in the context of the superstructure approach (*cf.* §3.1.2).

**Figure 20**

Atomic labelling scheme of (6*R*,7*aS*)-6-(*tert*-butyl dimethylsilyloxy)-1-hydroxy-2-phenyl-5,6,7,7*a*-tetrahydropyrrolizin-3-one. All atomic radii are arbitrary. Figure drawn with *PLATON* (Spek, 2003).

3.3.1. Refinement parameters. The final structural model of VAR205 is based on continuous atomic modulation functions for displacive (x_1, x_2, x_3) and thermal (U^{ij}) modulation. Higher harmonics up to fourth order were applied for the displacive modulation of all atoms and higher harmonics up to second order for modulation of the ADPs of the 24 non-H atoms. In total 1396 parameters were refined. The atomic labels are introduced in Fig. 20. Except atom H17, all hydrogen atoms are coupled *via* a riding model to their respective carbon atoms. The position of atom H17 (of the OH group) was refined isotropically applying bond-length [O–H: 0.84 (10) Å] and bond-angle [C–O–H: 109.47 (50)°] restraints.

The refinement was based on all main and satellite reflections up to fourth order, yielding a total of 28 265 reflections, of which 24 813 are observed with $I > 3\sigma(I)$. The ratio between the number of refined parameters and the number of reflections used is > 20 . In the last refinement cycle no correlation coefficient exceeded 0.7 (with three correlation coefficients larger than 0.6, all affecting the positional parameters of atom H17). The final residual parameters of the refinement are given in Table 2.

3.3.2. Rigid units. Following the above-mentioned concept of dividing the molecule into suitable rigid units the VAR205 molecule can be described as follows (Fig. 20): the silicon atom with its *tert*-butyl substituent and two methyl groups is connected *via* an oxygen atom to a non-planar tetrahydropyrrolizinone moiety, to which a flat phenyl ring is attached. With respect to the modulation, these individual molecular units act like rigid bodies, the modulation mainly affects the torsion angles between the three segments, changing the conformation of the molecule as a whole, but not of the molecular units. This variation will be discussed in §3.3.3.

This concept of dividing the molecule into suitable rigid units can also be employed during the refinement procedure, especially in cases with poorer data quality or fewer observed satellite reflections. In the present study the high quality of the

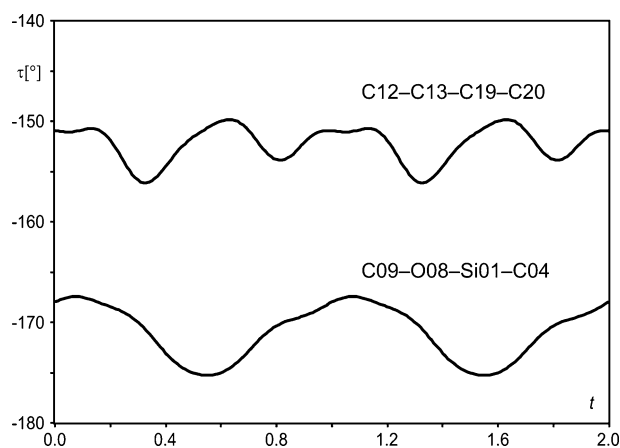


Figure 21
Torsion angles C12–C13–C19–C20 and C09–O08–Si01–C04 as a function of the phase of the modulation t . Curve drawn with the JANA2006 subroutine *Graph*.

Table 3
Selected bond lengths (Å) and bond angles (°).

	Mean	Minimum	Maximum
C04–C07	1.538 (8)	1.528 (8)	1.542 (8)
O08–C09	1.428 (5)	1.424 (5)	1.432 (5)
C20–C21	1.384 (8)	1.377 (8)	1.395 (8)
C21–C22	1.383 (8)	1.367 (8)	1.393 (8)
C20–C21–C22	120.4 (4)	119.6 (4)	120.9 (4)
Si01–O08	1.653 (7)	1.649 (7)	1.658 (7)
O08–C09	1.428 (5)	1.424 (5)	1.432 (5)
Si01–O08–C09	125.7 (2)	124.3 (2)	127.6 (2)

dataset with satellite reflections observable up to fourth order allowed all atoms to be treated individually, *i.e.* each with its own set of AMFs, while still retaining a very high data-to-parameter ratio (see Table 2). If this is not the case a so-called rigid-body refinement can help to reduce the number of parameters and/or avoid refinement problems such as unreasonable geometries. For VAR205 this option was tested using the above-mentioned rigid groups with Si01, C15 and C19 as reference atoms. For all non-H atoms in those rigid units the fractional coordinates and anisotropic ADPs were refined individually, while positional modulation waves up to fourth order and TLS (translation, libration, screw formalism) modulation waves up to second order were refined for each group as a whole (Schomaker & Trueblood, 1968). The H atoms were attached *via* riding models. The atoms O08, O17

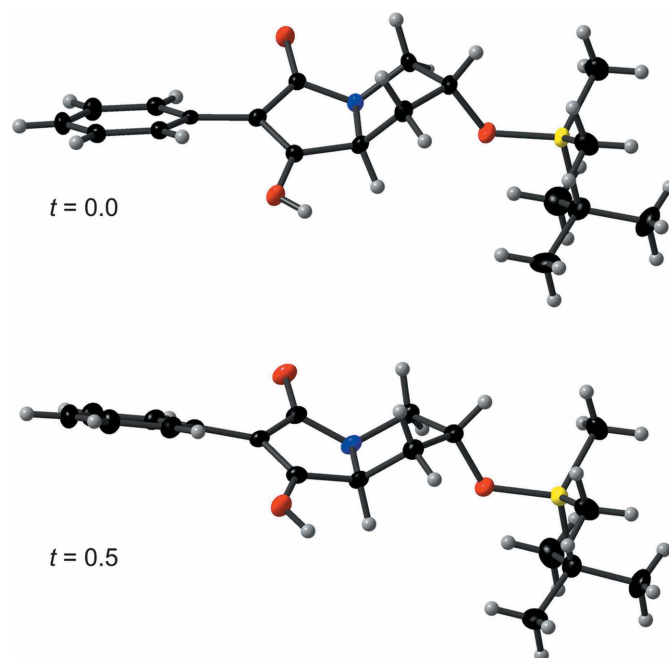


Figure 22
Side view of the molecule in the crystal for two different values of the phase of the modulation t . The anisotropic displacement ellipsoids are drawn at the 50% probability level, the radii of H atoms are arbitrary. The movie attached to this figure in the online version of the manuscript shows the variation of the molecule (in conformation and position) over a complete period of t . Please note that t is a spatial and not a temporal variable and has to be distinguished from the variable t that is used in physics for time. The movie corresponds to a 'walk' along the molecules in Fig. 23.

Table 4
Selected torsion angles (°).

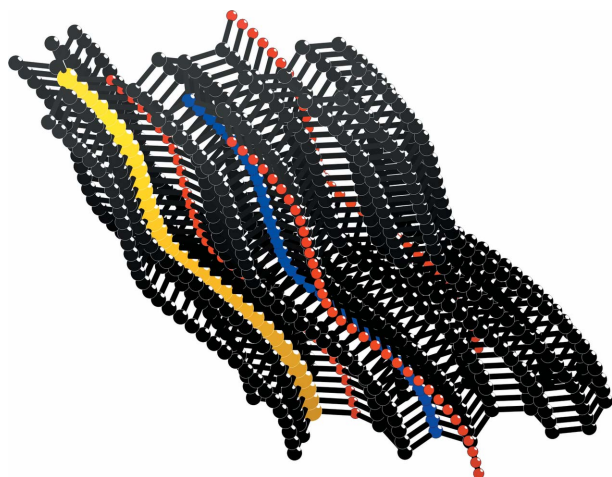
	Mean	Minimum	Maximum
N11–C10–C09–C16	–26.27 (17)	–28.47 (17)	–24.28 (18)
C10–N11–C15–C16	23.32 (18)	22.48 (18)	24.54 (18)

with H17, and O18 were refined as individual atoms as described in §3.2.3.

This strategy with common modulation parameters for all atoms in the respective rigid units reduced the number of refined parameters from 1396 to 775 and resulted in good agreement factors ($R_{\text{obs}} = 0.040$, $R_{\text{all}} = 0.046$ and $wR_{\text{all}} = 0.047$ for all reflections, $\Delta\rho_{\text{max}} = 0.31 \text{ e \AA}^{-1}$ and $\Delta\rho_{\text{min}} = -0.42 \text{ e \AA}^{-1}$, compared with values in Table 2). However, the refinement suffered heavily from correlations and did not converge properly. Therefore, this strategy of rigid-body refinement was not pursued any further. However, as mentioned above, in cases with limited numbers of reflections, such a refinement strategy is an option to consider.

3.3.3. Selected geometrical parameters. Mean bond lengths and angles in VAR205 are in very good agreement with the expected values for non-modulated structures, as published in the *International Tables for Crystallography*, Vol. C in ch. 9.5 (Allen *et al.*, 2006). The distances and angles remain constant within the standard uncertainties along the fourth dimension, *i.e.* along the phase of the modulation t (see Table 3 for some selected values).

Consider, *e.g.* the bond lengths and the bond angle within the Si01–O08–C09 group, connecting the *tert*-butyldimethylsilyl substituent with the tetrahydropyrrolizinone. They are rather constant as a function of t (Table 3), the torsion angles around these bonds (*e.g.* C09–O08–Si01–C04), however, vary by $\sim 10^\circ$. The same holds for the C13–

**Figure 23**

Alternative way of presenting the superstructure of VAR205. The 35 molecules of the asymmetric unit are shown along the direction of the modulation. Note the sinusoidal spatial wave of the atoms and, as a consequence, the missing translational symmetry: in a non-modulated structure all molecules would lie perfectly behind each other on a straight line.

C19 bond linking the tetrahydropyrrolizinone moiety to the phenyl ring. Here the torsion angles vary by $\sim 5^\circ$ (see Fig. 21). While the relative movement of the *tert*-butyldimethylsilyloxy and the tetrahydropyrrolizinone moieties with respect to each other has a quite undulating character, the relative movement of the tetrahydropyrrolizinone moiety and the phenyl ring is less smooth (see animated gif attached to Fig. 22 in the electronic version of the paper).

The degree of pyramidalization of the central N atom is not affected by the modulation, N11 is on average $0.331(1) \text{ \AA}$ out of the plane defined by the three carbon atoms C10, C12 and C15. This distance varies only slightly with t : the minimum and maximum distances are 0.320 and 0.343 \AA . For related compounds with a fully sp^3 -hybridized nitrogen (four substituents) distances around 0.53 \AA are found (Wagner, 2008).

The constancy with t also holds for the envelope structure C15–N11–C10–C09–C16, as can be seen from the torsion angles in Table 4. C16 is on average $0.625(1) \text{ \AA}$ out of the plane defined by the four atoms C09, C10, N11 and C15 (minimum: -0.651 \AA , maximum: 0.594 \AA).

The relative stereochemistry of H15 and the *tert*-butyldimethylsilyloxy substituent at the pyrrolidine five-ring can be unambiguously defined as *cis*.

The two oxygen atoms O17 and O18 are clearly out of the best plane defined by the five atoms N11, C12, C13, C14 and C15. The average distance for O17 is $0.0934(9) \text{ \AA}$ (minimum: 0.067 \AA , maximum: 0.145 \AA) and for O18 it is $-0.2639(9) \text{ \AA}$ (minimum: -0.253 \AA , maximum: -0.283 \AA).

As already mentioned above, the hydrogen atom H17 of the OH group was refined with restraints. As it is involved in the hydrogen bond O17–H17 \cdots O18ⁱ along **b** [with symmetry code (i) $x, y + 1, z$], the corresponding bond lengths and angles will be mentioned, but not further discussed. The average distance O17–O18ⁱ is $2.564(5) \text{ \AA}$ [minimum: $2.538(6) \text{ \AA}$, maximum: $2.588(5) \text{ \AA}$] and for O17–H17 it is $0.84(5) \text{ \AA}$ [minimum: $0.78(5) \text{ \AA}$, maximum: $0.88(5) \text{ \AA}$], the angle O17–H17–O18 is almost linear with an average value of $176(5)^\circ$ [minimum: $172(5)^\circ$, maximum: $179(5)^\circ$].

3.3.4. Superstructure. The atomic coordinates from the modulated structure can be converted into a commensurate superstructure, with, however, the same approximations that have to be accepted during a superstructure solution (*cf.* §3.1.2). Looking at the unit cell of the superstructure alone may not provide any insights regarding the direction of the modulation, *i.e.* may not provide any answer to the question: how does the variation in geometry and/or environment proceed throughout the crystal? It can help to align the individual molecules of the asymmetric unit according to the modulation. For the crystal structure of VAR205 this arrangement is found by selecting molecules perpendicular to the direction of the **q** vector (for details, see supplementary material Fig. S3). The resulting line of molecules is shown in Fig. 23. While arranging the molecules in such a way reveals the lack of translational symmetry in the crystal very clearly, it is not the most convenient way to visualize the change of the intramolecular geometry with t . For this purpose it is much more illustrative to export the coordinates of the individual

molecule in suitable steps of t and create an animated gif of these ‘snapshots along the fourth dimension’ (cf. Fig. 22).

4. Summary

The crystal structure of (6*R*,7*aS*)-6-(*tert*-butyldimethylsilylanyloxy)-1-hydroxy-2-phenyl-5,6,7,7*a*-tetrahydropyrrolizin-3-one, a typical organic compound, has been successfully refined using the superspace approach. This approach was introduced in a non-mathematical way. All steps from data handling to structure solution, refinement and presentation have been discussed in detail.

The description of the diffraction pattern in (3 + 1)-dimensional superspace delivered a unique indexing of the peaks, the final model during structure refinement was based on 1396 parameters with a data-to-parameter ratio above 20. Looking at the structure, the three individual molecular fragments act as rigid units, while the overall shape, or conformation, of the molecule varies significantly with the phase of the modulation t . This variation is reflected in the torsion and dihedral angles between these rigid units and in the intermolecular distances. Nevertheless, all parameters associated with intra- and intermolecular geometry correspond to values and distributions found for similar molecules in the Cambridge Structural Database (Allen, 2002).

In contrast, treating the structure as a superstructure (approximation) resulted in a series of problems: while indexing the diffraction pattern, it was not clear whether a 7×5 -fold or a 7×8 -fold supercell should be used. In the structure refinement, because of the large number of molecules in the asymmetric unit (35 and 56, respectively) and because of the large number of unobserved reflections, the data-to-parameter ratio was very poor. For anisotropic refinement 7560 parameters have to be considered for the 7×5 -fold and 12 096 for 7×8 -fold superstructure. Reasonable geometric parameters and anisotropic displacement parameters can only be obtained by employing large numbers of restraints.

The authors thank Karl Baumann for providing the compound under study and Vaclav Petricek, Michal Dusek and Lukas Palatinus for very valuable discussions and assistance with the programs *JANA2006* and *SUPERFLIP*. Ina Dix is acknowledged for the reconstruction of reciprocal space (Fig. 11) and Ekkehard Görlach for the calculations behind Figs. 14 and 16. Many thanks also to Carol Brock for her help and encouragement.

References

Aalst, W. van, den Hollander, J., Peterse, W. J. A. M. & de Wolff, P. M. (1976). *Acta Cryst.* **B32**, 47–58.
 Allen, F. H. (2002). *Acta Cryst.* **B58**, 380–388.
 Allen, F. H., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (2006). *International Tables for Crystallography*, Vol. C, pp. 790–811. Berlin: Springer.

Aroyo, M. I., Perez-Mato, J. M., Capillas, C., Kroumova, E., Ivantchev, S., Madariaga, G., Kirov, A. & Wondratschek, H. (2006). *Z. Kristallogr.* **221**, 15–27.
 Baudour, J. L. & Sanquer, M. (1983). *Acta Cryst.* **B39**, 75–84.
 Baumann, K. (2007). *PCT Int. Appl.* pp. 1–67.
 Bertaut, E. (1990). *Crystallogr. Rev.* **2**, 107–132.
 Brandenburg, K. (1999). *DIAMOND*. Crystal Impact GbR, Bonn, Germany.
 Brouns, E., Visser, J. W. & de Wolff, P. M. (1964). *Acta Cryst.* **17**, 614.
 Bruker AXS (2001). *SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
 Bruker AXS (2007*a*). *SMART*. Bruker AXS Inc., Madison, Wisconsin, USA.
 Bruker AXS (2007*b*). *RLATT*. Bruker AXS Inc., Madison, Wisconsin, USA.
 Bruker AXS (2007*c*). *SAINTE*. Bruker AXS Inc., Madison, Wisconsin, USA.
 Bruker AXS (2007*d*). *APEX2*. Bruker AXS Inc., Madison, Wisconsin, USA.
 Chapuis, G. (1996). *Crystallogr. Rev.* **5**, 109–131.
 Chapuis, G. & Orlov, I. (2005). Superspace Playground, <http://superspace.epfl.ch>.
 Dam, B., Janner, A. & Donnay, J. D. H. (1985). *Phys. Rev. Lett.* **55**, 2301–2304.
 Daniel, V. & Lipson, H. (1943). *Proc. R. Soc. London Ser. A*, **181**, 368–378.
 Dehlinger, U. (1927). *Z. Kristallogr.* **65**, 615–631.
 Dusek, M., Chapuis, G., Meyer, M. & Petricek, V. (2003). *Acta Cryst.* **B59**, 337–352.
 Dzyabchenko, A. & Scheraga, H. A. (2004). *Acta Cryst.* **B60**, 228–237.
 Evain, M., Péres, L. O., Mevellec, J.-Y., Froyer, G. & Petricek, V. (2009). *Z. Kristallogr.* Submitted.
 Friedrich, W., Knipping, P. & von Laue, M. (1912). *Sitzungsber. Kgl. Akad. Wiss.*, pp. 303–322.
 Friedrich, W., Knipping, P. & von Laue, M. (1913). *Ann. Phys.* **41**, 971–988.
 Goldschmidt, V., Palache, C. & Peacock, M. (1931). *Neues Jahrb. Mineral. A*, **63**, 1–58.
 Guiblin, N., Fuhrer, C. A., Häner, R., Stoeckli-Evans, H., Schenk, K. & Chapuis, G. (2006). *Acta Cryst.* **B62**, 506–512.
 Hao, X., Parkin, S. & Brock, C. P. (2005). *Acta Cryst.* **B61**, 675–688.
 Herbststein, F. H. (2005). *Crystalline Molecular Complexes and Compounds: Structure and Principles*. IUCr Monographs on Crystallography 18. Oxford University Press.
 Janner, A. & Janssen, T. (1977). *Phys. Rev. B*, **15**, 643–658.
 Janssen, T. (1988). *Phys. Rep.* **168**, 55–113.
 Janssen, T., Chapuis, G. & de Boissieu, M. (2007). *Aperiodic Crystals*. IUCr Monographs on Crystallography 20. Oxford University Press.
 Janssen, T. & Janner, A. (1987). *Adv. Phys.* **36**, 519–624.
 Janssen, T., Janner, A., Looijenga-Vos, A. & de Wolff, P. M. (2006). *International Tables for Crystallography*, Vol. C, pp. 907–955. Berlin: Springer.
 Johnson, C. K. & Watson, C. R. (1976). *J. Chem. Phys.* **64**, 2271–2286.
 Oszlányi, G. & Sütö, A. (2004). *Acta Cryst.* **A60**, 134–141.
 Oszlányi, G. & Sütö, A. (2008). *Acta Cryst.* **A64**, 123–134.
 Palatinus, L. & Chapuis, G. (2007). *J. Appl. Cryst.* **40**, 786–790.
 Petricek, V., Dusek, M. & Palatinus, L. (2006). *JANA2006*. Institute of Physics, Praha, Czech Republic.
 Petricek, V., van der Lee, A. & Evain, M. (1995). *Acta Cryst.* **A51**, 529–535.
 Preston, G. D. (1938). *Philos. Mag. A*, **26**, 855–871.
 Schmid, S. & Wagner, T. (2005). *Acta Cryst.* **B61**, 361–366.
 Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* **B24**, 63–76.
 Schönleber, A. (2002). Doctoral thesis, University of Lausanne, Switzerland.
 Schönleber, A. & Chapuis, G. (2004). *Acta Cryst.* **B60**, 108–120.
 Schönleber, A., Pattison, P. & Chapuis, G. (2003). *Z. Kristallogr.* **218**, 507–513.

- Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. (1984). *Phys. Rev. Lett.* **53**, 1951–1953.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Siegler, M. A., Hao, X., Parkin, S. & Brock, C. P. (2008). *Acta Cryst.* **B64**, 738–749.
- Smaalen, S. van (1995). *Crystallogr. Rev.* **4**, 79–202.
- Smaalen, S. van (2004). *Z. Kristallogr.* **219**, 681–691.
- Smaalen, S. van (2007). *Incommensurate Crystallography*. IUCr Monographs on Crystallography 21, Oxford University Press.
- Smith, G. F. H. (1903). *Z. Kristallogr.* **37**, 209–234.
- Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.
- Tanisaki, S. (1961). *J. Phys. Soc. Jpn*, **16**, 579.
- Tanisaki, S. (1963). *J. Phys. Soc. Jpn*, **18**, 1181–1191.
- Toudic, B., Garcia, P., Odin, C., Rabiller, P., Ecolivet, C., Collet, E., Bourges, P., McIntyre, G. J., Hollingsworth, M. D. & Breczewski, T. (2008). *Science*, **319**, 69–71.
- Völlenkne, H., Preisinger, A., Nowotny, H. & Wittmann, A. (1967). *Z. Kristallogr.* **124**, 9–25.
- Wagner, T. (2008). Unpublished results.
- Wolff, P. M. de (1974). *Acta Cryst.* **A30**, 777–785.
- Wolff, P. M. de (1977). *Acta Cryst.* **A33**, 493–497.
- Yamamoto, A. (1996). *Acta Cryst.* **A52**, 509–560.
- Zecchinon, L., Fett, T., Van den Bergh, P. & Desmecht, D. (2006). *Clin. Appl. Immunol. Rev.* **6**, 173–189.
- Zúñiga, F. J. & Criado, A. (1995). *Acta Cryst.* **B51**, 880–888.
- Zúñiga, F. J., Palatinus, L., Cabillo, P., Claramunt, R. M. & Elguero, J. (2006). *Z. Kristallogr.* **221**, 281–287.