Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Pseudo-symmetry characterization and refinement of a trigonal crystal form of naphthalene 1,2-dioxygenase

Enrique Carredano, Björn Kauppi, Devapriya Choudhury and S. Ramaswamy*†

Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, Biomedical Center, S-751 24 Uppsala, Sweden

† Present address: Department of Biochemistry, University of Iowa, Iowa City, IA 52242-1109, USA.

Correspondence e-mail: rams@xray.bmc.uu.se

Received 17 August 1999
Accepted 20 December 1999

Two trigonal crystal structures of naphthalene 1,2-dioxygenase from *Pseudomonas sp*. NCIB 9816-4 have been refined at 2.6 Å resolution. The space group is *R*3, with four heterodimers in the asymmetric unit. The crystallographic threefold axis coincides with the symmetry axis of the active molecule, a mushroom-shaped $\alpha_3\beta_3$ hexamer. The crystal is formed by symmetrical contacts between the hexamers on three different interaction surfaces, one on the β-subunit and the other two on the α-subunits. Nickel ions mediate one of the α-subunit interactions. The two other types of packing contacts sustain two interlaced and almost independent crystal patterns with significantly different temperature factors. The space group of the individual crystal patterns is *R*32, with the corresponding twofold axes parallel to each other. The interactions between the crystal patterns separate the two parallel twofolds, eliminating the twofold symmetry for the whole crystal. The differences in temperature factors among the molecules in the asymmetric unit have been refined and are different for the two refined structures. An analysis of the structure factors of the pseudo-equivalent reflections showed that their differences lie in their phases and not in their amplitudes, suggesting that $R_{\rm merge}$ is not an appropriate indicator for revealing the correct point group.

1. Introduction

Naphthalene 1,2-dioxygenase from *Pseudomonas sp*. NCIB 9816-4 (NDO) oxidizes naphthalene to (+)-*cis*-(1*R*,2*S*)-dihydroxy-1,2-dihydronaphthalene and belongs to a family of aromatic ring-hydroxylating dioxygenases that oxidize aromatic hydrocarbons and related compounds to *cis*-arene diols (Resnick, 1996). The enzyme is the terminal component of a microbial system that makes use of a reductase (an iron–sulfur flavoprotein) to transport electrons from NAD(P)H to a Rieske-type [2Fe–2S] ferredoxin. The latter protein then reduces the terminal oxygenase through a Rieske-type [2Fe–2S] centre. A mononuclear non-heme iron centre contributes thereafter to the addition of dioxygen to the substrate using a supply of two electrons from the [2Fe–2S] centre (Haigler & Gibson, 1990*a*,*b*).

The structure of NDO was determined recently using a combination of the multiple isomorphous replacement (MIR) and the multiple-wavelength anomalous diffraction (MAD) methods (Kauppi *et al.*, 1998). The active molecule is an $\alpha_3\beta_3$ hexamer, the shape of which resembles that of a mushroom, with the three β-subunits forming the foot and the three α-subunits forming the cap.

The α-subunit has a β-sheet domain, which contains the Rieske [2Fe–2S] centre, and a catalytic domain with a fold

dominated by an antiparallel nine-stranded pleated β-sheet with helices packed against it. The Rieske domain (residues 38–158) consists of two β-sheets and two β-hairpins. The sheets form a β-sandwich with a β-sheet from the catalytic domain; residues from the loops in the hairpins bind the Rieske centre. A nine-stranded antiparallel β-sheet constitutes the core of the catalytic domain.

The structure of NDO was solved in a trigonal crystal form. On the basis of merging statistics of the diffraction data the space group *R*32 was assumed, with two molecules in the asymmetric unit (Lee *et al.*, 1997; Kauppi *et al.*, 1998). A pseudo-symmetry between the two molecules was indicated by a large peak in the self-Patterson at about $(0, 0, \frac{1}{2})$. It was soon realised that one of the molecules 'behaved badly' in three respects: poor electron density, short contacts with crystallographic neighbours and, after refinement, extremely high *B* factors. With this model, the structure was refined to an *R* factor of 0.31 and a free *R* factor of 0.34 for all data to a resolution of 2.5 Å.

A new orthorhombic crystal form with space group *I*222 and three heterodimers in the asymmetric unit diffracting to 2.25 Å resolution was found. Molecular replacement with the best refined model of NDO in the trigonal system as a search model resulted in a clear solution. This structure was refined to an *R* factor of 0.20 and a free *R* factor of 0.24.

All details relating to the NDO structure reported by Kauppi *et al.* (1998) concerned the *I*222 crystal form, although the structure was originally solved in the trigonal crystal form. The objective of this study is to return to the trigonal rhombohedral crystal form and find a more accurate model to explain the observed data.

2. Materials and methods

Details concerning the purification of the protein, crystallization and data collection are given elsewhere (Lee *et al.*, 1997; Kauppi *et al.*, 1998).

2.1. Determination of the correct space group

The data used in this part of the work was the same as that used in the refinement of the *R*32 model in Kauppi *et al.* (1998). The data extend to 2.6 Å. The crystal system is rhombohedral, with unit-cell dimensions $a = b = 178.9$, $c = 323.6$ Å in the hexagonal setting. The starting point was the *R*32 model (Kauppi *et al.*, 1998) with one 'good' and one 'bad' molecule.

The model of the 'bad' molecule had short crystal contacts in the region between residues 280–290 in the α-subunit and no crystal contacts at all in the region between residues 523 and 529 in the β-subunit. Using the refined model of the *I*222 crystal structure (Kauppi *et al.*, 1998) as a search model in *AMoRe* (Navaza, 1994), only one of the two heterodimers in the *R*32 asymmetric unit was found by molecular replacement. Using the program *BRUTE* (Fujinaga & Read, 1987) and by translating the missing molecule along the *c* axis of the hexagonal cell and keeping the 'good' molecule fixed, two minima in the *R* factor which corresponded to two maxima in the correlation coefficient were found. The correlation coefficient and the *R* factor in the calculation are a result of the contribution of the translation plus the contribution of the 'good' molecule. The two positions were symmetrically placed on either side of the centre of the unit-cell dimension and at a distance of about 4.0 Å from each other (Fig. 1).

Combining the data with the best experimental phases from Kauppi *et al.* (1998) and a model containing the 'good' molecule and a half-occupancy version of the 'bad' molecule, an $mF_o - DF_c$ difference map was calculated with the programs *REFMAC* (Murshudov *et al.*, 1997) and *FFT* (Collaborative Computational Project, Number 4, 1994). This map was then skeletonized with *MAPMAN* (Kleywegt & Jones, 1996) and read into *O* (Jones *et al.*, 1991) for display. The skeleton revealed the trace of another molecule translated a distance of 4.0 Å along *c* from the half-occupancy 'bad' molecule. Its position agreed with the second peak from the *BRUTE* calculation.

Acoordingly, a model was built where the missing molecule had two conformations, each with an occupancy of 0.5. Using alternative conformations in *REFMAC*, it was possible to refine this model in *R*32 at a resolution of 2.6 Å. During the *REFMAC* cycles, the molecules were refined without NCS restraints and the *B* factors were refined individually. After each cycle, however, the transformation matrices from the full-occupancy molecule to the two half-occupancy molecules were calculated and the latter were generated from the coordinates of the former. It was necessary to apply this kind of pseudo-strict non-crystallographic symmetry in order to increase the data-to-parameter ratio.

A program was written based on the Newton–Raphson algorithm to optimize additional overall temperature factors

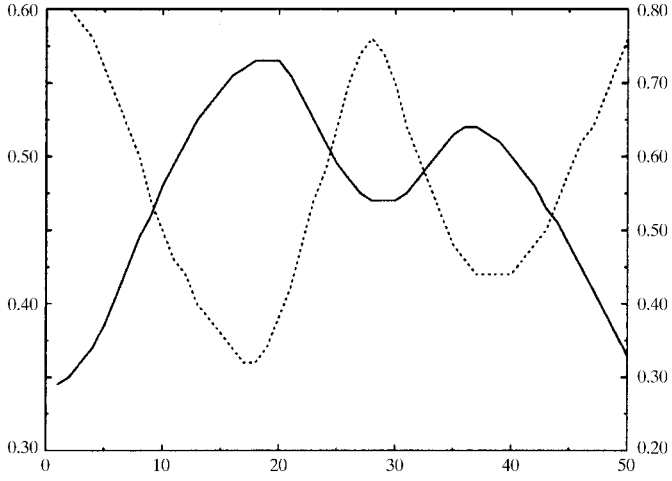

Figure 1
Correlation coefficient and *R* factor as functions of molecular translation. The *x* axis indicates the number of 0.2 Å steps from half the cell dimension along the *z* axis minus 5.0 Å. Two minima in the *R* factor (dashed line, left-hand scale) correspond to two maxima in the correlation coefficient (full line, right-hand scale) when translating the 'badly behaved' molecule in the asymmetric unit of the *R*32 model along the *c* axis.

related to each of the two half-occupancy molecules. These parameters were then added to the *B* factors inherited from the master molecule. There was no manual intervention during the refinement of the *R*32 model. Some 100 water molecules could be added automatically to the model using *ARP* (Lamzin & Wilson, 1993). The final *R* factor was 0.225, with a free *R* factor of 0.278, for all data.

According to *PROCHECK* (Laskowski *et al.*, 1993), 85.3% of non-glycine residues are in the most favoured regions, 14.0% are in additionally allowed regions, 0.2% are in generously allowed regions and 0.5% are in the disallowed regions of the Ramachandran plot.

The quality of the $2mF_o - DF_c$ map was good for the full-occupancy molecule and ambiguous for the half-occupancy ones. Nevertheless, it was now possible to actually see the density of each of the half-occupancy molecules in the $mF_o - DF_c$ map when its model was left out for the map calculation (Fig. 2).

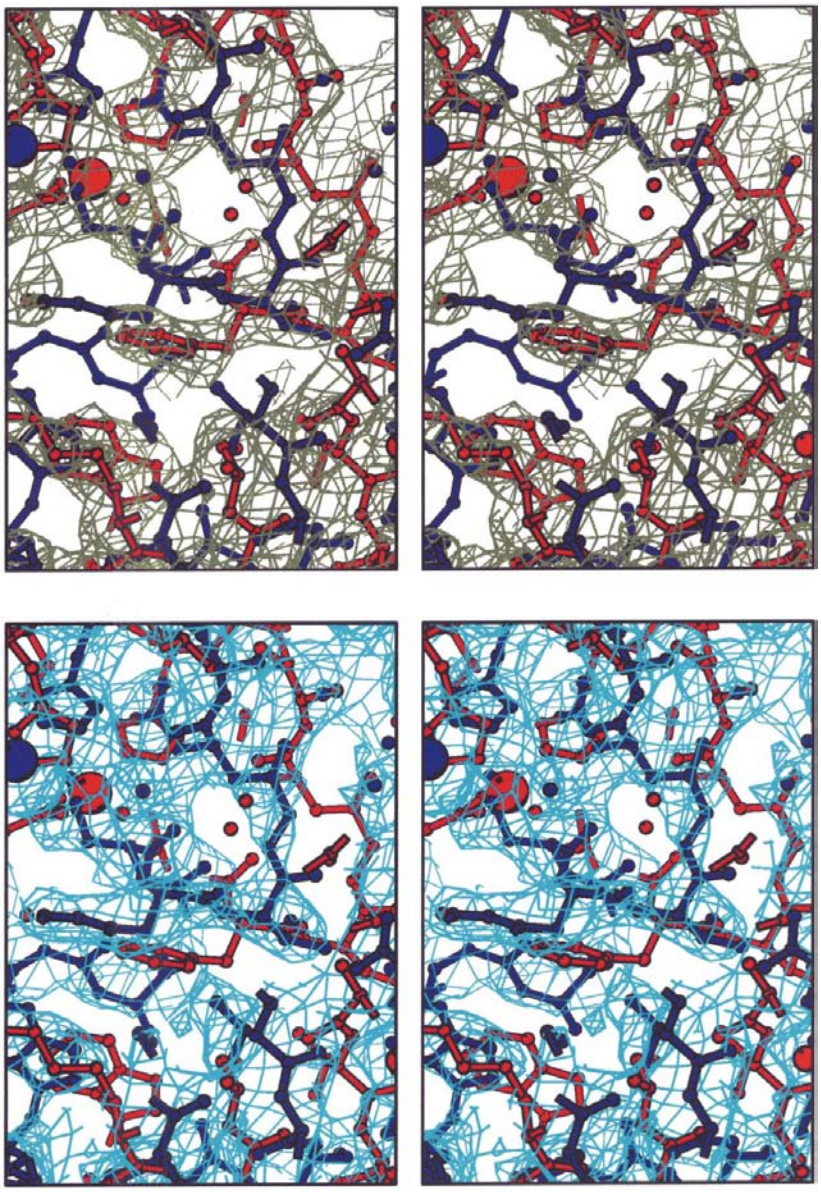

Figure 2
Space-group error modelled as static disorder. Since the twofold symmetry is valid for half of the molecules in the crystal, the crystal structure may be approximately modelled by giving half occupancy to the molecules not obeying the symmetry. These $mF_o - DF_c$ maps were computed after omitting one of the half-occupancy molecules. To remove model bias, a random shift was applied to the coordinates. Four cycles of refinement were carried out before map calculation. The maps are contoured at 1 × r.m.s. around the active site. Density for the catalytic irons and ligands appears which belongs to the missing half-occupancy molecule. A different half-occupancy molecule has been omitted in each one of the figures. This stereoview was produced with *BOBSCRIPT* (Esnouf, 1997; Kraulis, 1991).

The assumption of static disorder allows two possible conformations of the disordered molecules. In a specific location of the physical crystal, however, the disordered molecule must be found in only one of the alternative conformations. Neighbouring molecules, on the other hand, must be in the state of the other conformation since otherwise short contacts would occur. By induction, it follows that the two conformations alternate along the crystal and, as the orientation of the molecules also alternates by 180°, the conformation chosen is always on the same side, sacrificing the twofold symmetry and giving rise to *R*3 packing (Fig. 3).

A model was built with *R*3 symmetry and consequently all the problems (bad contacts with crystallographic neighbours, molecules with half occupancy and poor electron density) were eliminated. In *R*3, there are four molecules in the asymmetric unit, each one related to two of the others by a twofold rotation about the face diagonal $a + b$ followed by a translation along the *c* axis of the hexagonal cell and related to the third molecule by a pure translation along the *c* axis.

2.2. Solution and refinement of two *R*3 structures

2.2.1. Data processing. Two data sets (15.0–2.6 Å) collected from soaked crystals at beamline 19 (BM14) at the ESRF (Grenoble) were used. The first data set (IR) consisted of 72° of data collected from a crystal soaked in iridium (III) chloride. The second data set (NDOSHG) was collected from a crystal soaked in thimerosal (EMTS) and consisted of 91° of data. The data were processed and scaled with *DENZO* and *SCALEPACK* from the *HKL* suite of programs (Otwinowski & Minor, 1997). The merging statistics in both space groups *R*3 and *R*32 are presented in Table 1.

For reasons that will be discussed in §3 and despite the fact that the space group of the crystals is *R*3, the pseudo-equivalent reflections have the same amplitude within experimental error. This is also proved by the data, where it is clear that $R_{\rm merge}$ does not discriminate between

Table 1
Merging statistics.

Values are given for overall resolution (15.0–2.6 Å); values in parentheses correspond to the highest resolution shell (2.64–2.60 Å).

	IR *R*3	IR *R*32	NDOSHG *R*3	NDOSHG *R*32
χ^2†	1.34 (1.81)	1.17 (1.42)	1.37 (1.60)	1.19 (1.44)
Linear $R_{\rm merge}$‡	5.5 (25.9)	6.0 (27.1)	8.0 (31.7)	8.5 (31.9)
Square $R_{\rm merge}$§	4.3 (21.1)	4.7 (22.6)	6.8 (25.1)	7.2 (25.5)
Completeness (%)	86.5 (89.0)	95.5 (96.0)	87.5 (76.8)	96.7 (93.1)
Reflections with at least three observations (%)	30.3 (18.2)	73.5 (70.4)	39.5 (23.4)	79.7 (63.6)
No. of unique reflections	99867 (5129)	56936 (2827)	97219 (4259)	55520 (2639)
No. of measured reflections	222574 (9742)	209863 (9307)	239991 (8687)	231621 (8917)
Multiplicity	2.2 (1.9)	3.7 (3.3)	2.5 (2.0)	4.2 (3.4)
Reflections with $I/\sigma(I) \geq 3$.	74.2 (23.8)	74.9 (41.4)	62.1 (23.4)	59.3 (25.6)

† $\chi^2 = \sum_{hkl}\sum_i \big(I_{hkl,i} - \langle I_{hkl}\rangle\big)^2 \Big/ \big[\sigma^2(I) N_{hkl}/(N_{hkl}-1)\big]$. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i \big|I_{hkl,i} - \langle I_{hkl}\rangle\big| \Big/ \sum_{hkl} N_{hkl}\langle I_{hkl}\rangle$. § $R_{\rm merge} = \sum_{hkl}\sum_i \big(I_{hkl,i} - \langle I_{hkl}\rangle\big)^2 / \sum_{hkl} N_{hkl}\langle I_{hkl}\rangle^2$.

the assumption of the two different space groups. Therefore, the data were scaled as if the space group of the crystal had been *R*32 and then expanded to *R*3 using *SFTOOLS* (Collaborative Computational Project, Number 4, 1994). Care was taken to choose the test data for cross validation (Brünger, 1993) before the expansion so that both related reflections belonged to the same set (either the test or the working set).

The unit-cell dimensions (in Å) are 177.3, 177.3 and 321.4 for IR, and 175.2, 175.2 and 316.8 for NDOSHG. The data are not isomorphous to each other (overall $R_{\rm merge}$ is 0.29).

2.2.2. Structure solution. One heterodimer from the refined *I*222 structure from Kauppi *et al.* (1998) was used as a search model in *AMoRe* in space group *R*3 to find four heterodimers in the asymmetric unit. The highest solutions were correct and were obtained using reflections in the resolution range 15.0–3.0 Å. The values obtained for the correlation coefficient and the *R* factor were 0.78 and 0.34, respectively, for IR and 0.72 and 0.39, respectively, for NDOSHG.

2.2.3. Structure refinement. The structures were refined by alternate rounds of automatic minimization using the program *REFMAC* and model building using *O* assisted by *OOPS* (Kleywegt & Jones, 1997, 1998). Geometrical restraints for the protein chains were set up using *PROTIN* (Collaborative Computational Project, Number 4, 1994). Bulk-solvent correction was applied as implemented in *REFMAC* and the temperature factors of the atoms were refined individually and isotropically. $2mF_o - DF_c$ and $mF_o - DF_c$ maps were calculated from the *REFMAC* output.

Pseudo-strict non-crystallographic symmetry was applied as described previously for the *R*32 case. The dimer with the lowest *B* factors was chosen as the master molecule.

Water molecules were added with *ARP* around the master heterodimer and the NCS operators were applied to their coordinates to obtain the waters in the entire asymmetric unit using *PDBSET* (Collaborative Computational Project, Number 4, 1994).

The models were validated during the refinement with the programs *OOPS* and *PROCHECK*.

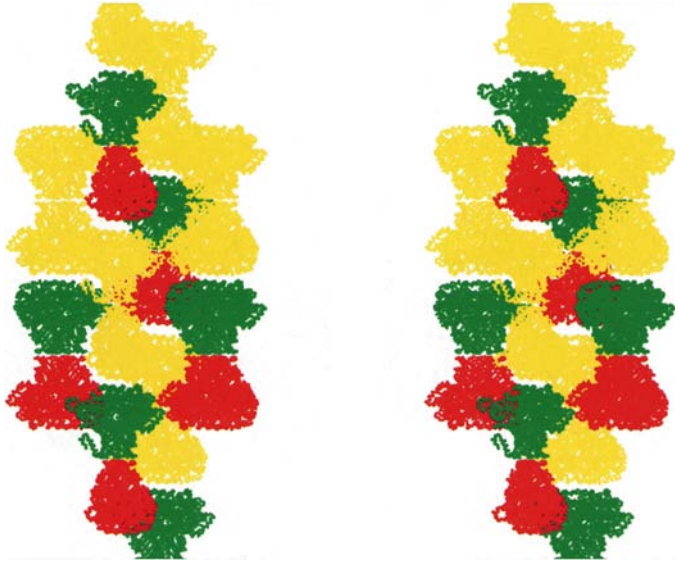

Figure 3
Packing in *R*3. Using inductive reasoning it can be shown (see text) from the results of the refinement of the *R*32 model that the actual packing in the crystal is *R*3. The twofold symmetry may only be obeyed by half of the molecules (yellow), which form a crystal pattern with space-group type *R*32. The green and red molecules also form such a crystal pattern, although a difference in *B* factor makes them more distinct from one another. Ignoring *B* factors, the space group of the crystal pattern made of the green and red molecules is also *R*32. However, the twofold axis relating the yellow molecules is shifted by about 2.0 Å from the twofold axis relating the green and the red molecules. The two axes are parallel to each other and normal to the plane of the paper and the shift is along the *c* axis (vertical). As a result of this shift, the twofold symmetry of the total crystal structure is lost. The nickel ions (see below) are bound at the interface between the yellow and the green molecules. This stereoview was produced with *MOLSCRIPT* (Kraulis, 1991).

3. Results and discussion

3.1. The refined *R*3 structures

Two crystal structures of a trigonal crystal form of naphthalene 1,2-dioxygenase from *Pseudomonas sp*. NCIB 9816-4 have been refined to conventional and free *R* factors of 0.228 and 0.284, respectively, for IR, and 0.253 and 0.308, respectively, for NDOSHG. Other refinement statistics are presented in Table 2. The results from the refinement of the partial *B* factors for the different NCS-related mates are also shown in this table in terms of average *B* factors. The four $\alpha\beta$-dimers in the asymmetric unit are otherwise identical after superimposition.

Table 2
Model and refinement statistics.

	IR	NDOSHG
No. of residues in asymmetric unit	2560	2560
No. of non-H atoms in asymmetric unit		
Protein	20352	20352
Rieske centres	16	16
Catalytic Fe	4	4
Water molecules	748	880
Ni	0	2
SO_4^{2-}	20	40
All	21252	21122
Average *B* factors (Å^2)		
NCS molecule 1	26.5	19.4
NCS molecule 2	26.8	23.4
NCS molecule 3	61.4	30.7
NCS molecule 4	62.9	39.2
All protein	44.4	28.2
Heteroatoms	49.4	42.4
All	44.6	28.7
R.m.s. deviations from ideal values		
Bond lengths (Å)	0.023	0.014
Bond angles (°)	2.3	2.6
Peptide-planarity angles		
% outliers (cutoff 5.8°)	24.2	18.2
RSC values		
% outliers (cutoff 1.0 Å)	13.8	14.1
Pepflip values		
% outliers (cutoff 2.5 Å)	1.9	1.9
RS-fit values		
% outliers (cutoff 0.64)	4.3	3.8
Ramachandran statistics (*PROCHECK*)		
Percentage of residues (non-Gly) in		
most favoured regions	87.6	86.9
additional allowed regions	12.1	12.4
generously allowed regions	0.0	0.4
disallowed regions	0.4	0.4

The final electron-density $2mF_o - DF_c$ map, contoured at 1σ, is well defined for this resolution with the exception of two short stretches, 233–236 in the α-subunit and 523–526 in the β-subunit, the side chains of which are not very well defined. The models have no distorted bonds and angles according to *PROCHECK*.

The structure of the heterodimer is essentially the same as that of the *I*222 crystal form, with the exception of minor differences in loop regions and a slight difference in the relative orientation of the subunits. The latter is correlated to differences in the packing and is reflected in a larger r.m.s. deviation from the overall r.m.s. compared with each of the individual subunits. Superimposing the C^{α} atoms of the heterodimer onto the dimer of the *I*222 crystal structure (PDB code 1ndo) gives r.m.s. deviations of 0.4 and 0.6 Å for IR and NDOSHG, respectively. Superimposing the subunits individually gives 0.2 and 0.4 Å r.m.s. deviation, respectively, for the α-subunits (447 atoms) and 0.4 and 0.4 Å r.m.s. deviation, respectively, for the β-subunits (192 atoms).

3.2. Crystal packing

The axis of symmetry of the hexamer coincides with a crystallographic threefold rotation axis and is parallel to the *c* axis in the hexagonal cell. Along the threefold, the hexamers adopt one of two orientations, corresponding to 0 and 180° rotation, in an alternate fashion so that α-subunits face α-subunits and β-subunits face β-subunits. Using the mushroom analogy, caps face caps and feet face feet along the threefold axis (see Fig. 3).

The asymmetric unit is formed by four heterodimers. Two of these have lower overall *B* factors than the other two (Table 2). When the threefold is applied to the asymmetric unit, two pairs of mushroom-shaped hexamers are formed.

3.2.1. Two interlaced crystal patterns. The partition of the molecules in the crystal into low and high *B*-factor molecules can be related to the packing interactions. With one exception (see below), all of the crystal contacts involve residues belonging to the same type of molecule. High *B*-factor molecules interact almost exclusively with high *B*-factor molecules and low *B*-factor molecules with low *B*-factor molecules. The crystal structure may be described as two almost independent crystal patterns which are interlaced. Each of these are built up by crystal contacts in which each α-subunit interacts with another α-subunit and each β-subunit with another β-subunit. The interactions are symmetrical and the regions of interaction are those between residues 280 and 290 for both the α-subunits and between 523 and 529 for both the β-subunits.

Without taking into account the *B*-factor distribution, the space group of the individual crystal patterns is *R*32. The corresponding twofold axes are perpendicular to *c* and parallel to each other.

3.2.2. Nickel-mediated interactions bring the two together. If the two interlaced crystal patterns were placed symmetrically and equidistant to one another then the contact distance between them would be too great. Such a theoretical crystal would probably be too unstable. The crystal patterns of the low and the high *B*-factor molecules are joined to one another through an interaction mediated by Ni atoms. In the NDOSHG structure, this was modelled as a full occupancy nickel ion. For IR, the occupancy for the ion is partial as suggested by the density at the binding site in the $2mF_o - DF_c$ and $mF_o - DF_c$ maps when contoured at 1σ and 3σ, respectively. The closest ligand to the nickel ion is His434 from one of the interacting molecules. Glu11 from the other molecule is at a hydrogen-bonding distance and Glu442 is at a slightly longer distance (Fig. 4).

All these side chains come from the α-subunits and only one (not both) of the hexamers from the pairs described above are involved. Since the interaction is symmetrical, there are two nickel-mediated interactions per asymmetric unit. These interactions ensure a separation distance between the parallel twofold symmetry axis of the individual crystal patterns. This distance is 2.8 Å for IR and 1.9 Å for NDOSHG. Therefore, the twofold rotational symmetry of the individual crystal patterns is not carried forward to the total crystal structure and the space group of the latter is *R*3 and not *R*32. Apparently, the nickel-mediated interactions hold the crystal structure together by linking the two otherwise independent crystal patterns. This is remarkable, as there are as many as 2560 residues in the asymmetric unit. Nickel thus plays a key role in the formation and stability of the crystal; the crystals do not grow in the absence of nickel in the crystallization buffer (Lee

et al., 1997). Each of the high *B*-factor pairs of mushroom-shaped hexamers described above is brought together to a low *B*-factor pair by six nickel-mediated interactions. Of the high *B*-factor molecules, those in contact with the low *B*-factor molecules through the nickel-mediated interactions have lower *B* factors than those which are not. Additionally, as a consequence of the lower occupancy of the nickel-binding site of IR, the unit-cell dimension *c* is larger and the *B* factors are higher.

3.3. Analysis of the structure-factor equation

In the previous section, it was shown that the *R*3 crystal structure is made up of two crystal patterns of space group *R*32. The twofolds of the individual patterns are parallel but are separated by a significant shift along the *c* axis. After choosing the origin so that a twofold of the low *B*-factor pattern passes through it, the effect of the separation Δz = 2.0 Å on the structure factors, their amplitudes and phases are investigated below. For the sake of clarity, temperature factors are not included.

The structure factors $F(h, k, l)$ may be written as a sum of the contributions from the crystal patterns $F_1(h, k, l)$ and $F_2(h, k, l)$,

$$F(h, k, l) = F_1(h, k, l) + F_2(h, k, l), \tag{1}$$

where

$$F_1(h, k, l) = \sum_{m=1}^{M} f_m \exp[2\pi i(hx_m + ky_m + lz_m)], \tag{2}$$

$$F_2(h, k, l) = \sum_{n=1}^{N} f_n \exp[2\pi i(hx_n + ky_n + lz_n)] \tag{3}$$

and where the sums over *m* and *n* comprise atoms belonging to the low and the high *B*-factor patterns, respectively, and where

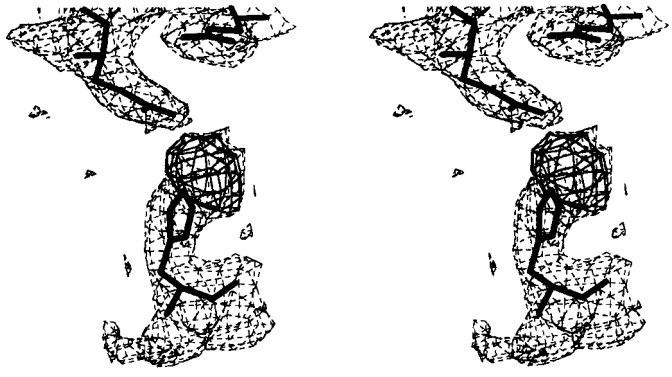

Figure 4
Stereoview of nickel-binding site. Nickel is essential for the formation of the crystals. The ion binds to His434 and interacts electrostatically with Glu11 from a symmetry-related molecule. The side chain of Glu442 from the same molecule is at a slightly longer distance. The lighter contour corresponds to the $2mF_o - DF_c$ map and the darker one to the $mF_o - DF_c$ map at 1σ and 3σ levels, respectively, when the ion has been modelled as an O atom. Inserting a nickel ion gives a *B* factor of about 40 Å^2 after refinement. These maps corresponds to data from NDOSHG. The IR data gives considerably weaker density, suggesting partial occupancy for the ion.

$M = N = \mathcal{N}/2$, with $\mathcal{N}$ being the total number of atoms in the unit cell.

The relation between the coordinates of the corresponding low and high *B*-factor atoms is given by the pseudo-symmetry, which is the NCS relation

$$x_n = x_m, \tag{4}$$

$$y_n = y_m, \tag{5}$$

$$z_n = z_m + \tfrac{1}{2} + \Delta z, \tag{6}$$

where Δz is the distance between the parallel twofolds defined above.

We choose our origin such that the twofold applies to low *B*-factor molecules. This means that for every atom at (x_m, y_m, z_m), there is an atom at $(y_m, x_m, -z_m)$ (*R*32 symmetry). Therefore, F_1 in (2) may be rewritten as

$$\begin{aligned} F_1(h, k, l) &= \sum_{m=1}^{M/2} f_m\{\exp[2\pi i(hx_m + ky_m + lz_m)] \\ &\quad + \exp[2\pi i(kx_m + hy_m - lz_m)]\} \\ &= F_1(k, h, -l), \end{aligned} \tag{7}$$

which is a symmetry relation of $R\bar{3}m$.

The NCS high *B*-factor mate of position $(y_m, x_m, -z_m)$ is located at position $(y_m, x_m, -z_m + \frac{1}{2} + \Delta z)$, in agreement with (4) to (6). Therefore, it is possible to write the contribution (3) to the structure factor as a sum over half the number of atoms (M) belonging to the low *B*-factor pattern in the unit cell,

$$\begin{aligned} F_2(h, k, l) &= \sum_{j=1}^{M/2} f_j\big(\exp\{2\pi i[hx_m + ky_m + l(z_m + \tfrac{1}{2} + \Delta z)]\} \\ &\quad + \exp\{2\pi i[kx_m + hy_m + l(-z_m + \tfrac{1}{2} + \Delta z)]\}\big) \\ &= \exp(\pi il)\exp(2\pi i\Delta z l)F_1(h, k, l). \end{aligned} \tag{8}$$

Theoretically, letting $\Delta z \rightarrow 0$ would bring the twofolds together and make the space group of the crystal structure become *R*32. Setting $\Delta z = 0$ into (8) gives

$$^{R32}F_2(h, k, l) = \exp(\pi il)F_1(h, k, l) = \overline{F}_2(h, k, l). \tag{9}$$

Using this notation, we may re-write (8) as

$$F_2(h, k, l) = \exp(2\pi il\Delta z)\overline{F}_2(h, k, l). \tag{10}$$

Thus, the effect of a translation by Δz on the partial contribution $F_2(h, k, l)$ is to increase its phase by $2\pi l\Delta z$ radians without affecting its amplitude. Accordingly, the effect on the pseudo-equivalent contribution $F_2(k, h, -l)$ is

$$F_2(k, h, -l) = \exp(-2\pi il\Delta z)\overline{F}_2(k, h, -l). \tag{11}$$

The phase shift of the pseudo-equivalent reflection thus has the same magnitude but is opposite in direction. The total effect is a splitting of the *R*32 reflection into two pseudo-equivalent reflections in *R*3. The two reflections have the same first component $F_1(h, k, l) = F_1(k, h, -l)$, but whereas the amplitude of the second components are also equal, the phases are not; the magnitude of the difference is $4\pi l\Delta z$.

The change in phase and amplitude of the total structure factor (1) can be calculated using (9). To repeat,

$$\overline{F}_2(h, k, l) = \exp(\pi i l) F_1(h, k, l). \quad (12)$$

This means that, neglecting the temperature-factor distribution, the two contributions $F_1(h, k, l)$ and $\overline{F}_2(h, k, l)$ do not only have the same amplitude but are exactly in phase or exactly out of phase depending upon the parity of l. The reason for this is that the phase difference is πl and therefore even values of l give a phase difference which is a multiple of 2π (in phase), whereas odd values of l give a phase difference equivalent to π (out of phase).

Including the B factors in the equations has the effect that the magnitudes of $F_1(h, k, l)$ and $\overline{F}_2(h, k, l)$ become different. This has no consequence on the amplitudes of the total reflections being equal for the pseudo-equivalent reflections, but has a minor effect on the magnitude of the phase shift.

The corresponding Argand diagrams are shown in Fig. 5. There, the resultant structure factors of the pseudo-equivalent reflections are represented as dashed arrows and it is obvious from the figure that both have the same amplitude. Nevertheless, they definitely do not have the same phase. Using Fig. 5, it can be shown from simple geometrical considerations that the phase error resulting from assuming the wrong space group ($R32$ instead of the correct $R3$) is dependent upon the parity of l according to

$$\Delta\Phi = \pi/2 - \pi l \Delta z \quad (13)$$

for reflections with odd l and

$$\Delta\Phi = \pi l \Delta z \quad (14)$$

for reflections with even l. Additionally, the maximum phase error in both cases is $\pi/2$.

Applying the law of cosines to Fig. 5 and neglecting the B-factor distribution, it is also possible to obtain an approximate formula for the intensities in terms of $\big||F_1(h, k, l)|\big|$. The formula is different depending again on the parity of l,

$$\big||F|\big|^2 = 2\big||F_1|\big|^2[1 - \cos(\pi - 2\pi l \Delta z)] \quad (15)$$

for reflections with even l and

$$\big||F|\big|^2 = 2\big||F_1|\big|^2[1 - \cos(2\pi l \Delta z)] \quad (16)$$

for reflections with odd l.

The multiplying factor modulating $\big||F_1|\big|^2$ in (15) and (16) has been plotted in Fig. 6, together with a plot of the theoretical phase error in (13) and (14). This has been performed as a function of l and for even and odd values. Note that the phase errors and the modulating factor for the amplitudes are anticorrelated in the sense that the minima of one correspond to the maxima of the other.

A phase splitting of π or 180°, *i.e.* when

$$4\pi l \Delta z = \pi, \quad (17)$$

represents a special situation in which the partial contributions from $F_2(h, k, l)$ and $F_2(h, k, -l)$ from the pseudo-equivalent reflections are out of phase so that their vector average vanishes. This is independent of the parity of l and has consequences for both the phase errors and the amplitudes of the reflections: at this point on the reciprocal l axis, the phase error is the same for odd and even reflections and is equal to $\pi/4$ or 45°. Also at this point, the multiplying factor modulating the amplitudes is the same for both odd and even reflections. From the crystal structure, we know that Δz = 2.0/323.6 = 0.0062 and therefore the 'crossover point' obtained by solving for l in (17) occurs at about $l \simeq 40$.

3.4. $R_{\rm merge}$ does not always reveal the correct point group

Fig. 5 shows why it is impossible to determine the correct point group on the basis of $R_{\rm merge}$. Independent of the parity of l, both pseudo-equivalent reflections have the same amplitude. This is therefore a case of an $R3$ structure where the point

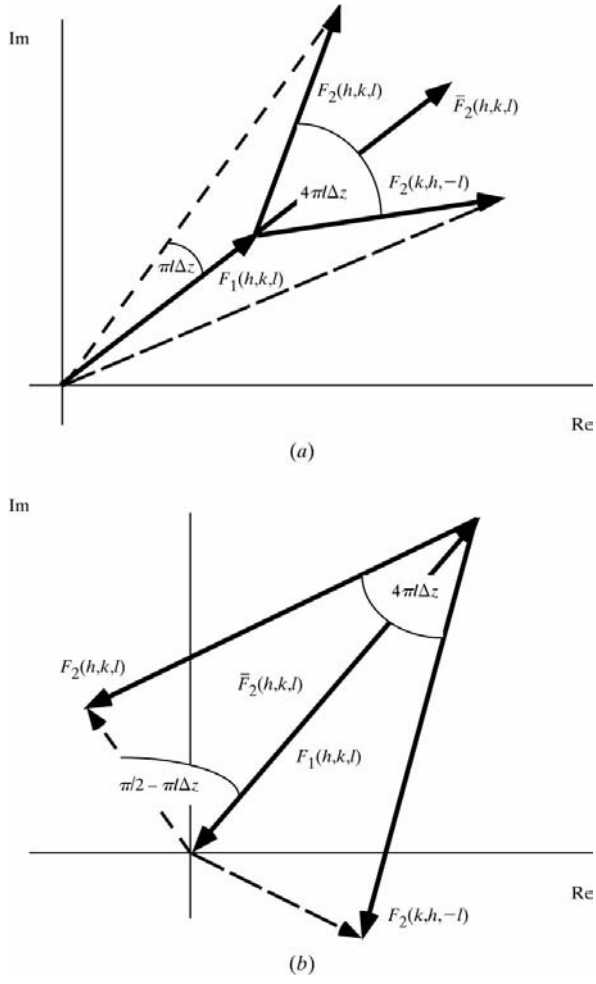

Figure 5
Argand diagrams with the contributions from the two crystal patterns to the total structure factors. The resultant structure factors (dashed lines) are the vector sum of $F_1(h, k, l)$ and $\overline{F}_2(h, k, l)$. The geometry is simplified by the fact that $\Delta z = 0$ means a translation vector between the crystal patterns *exactly* equal to half the cell dimension. The two contributions $F_1(h, k, l)$ and $F_2(h, k, l)$ do not only have the same amplitude but they are exactly in or out of phase depending upon the parity of l. (*a*) Even l gives a phase difference which is a multiple of 2π (in phase, the two arrows in the same direction). The phase error that results from the space-group error as obtained from simple geometrical considerations is $\pi l \Delta z$. (*b*) Odd l gives a phase difference between the two contributions equivalent to π (out of phase, the two arrows having opposite direction) and the phase error arising from the space-group error is $\pi/2 - \pi l \Delta z$.

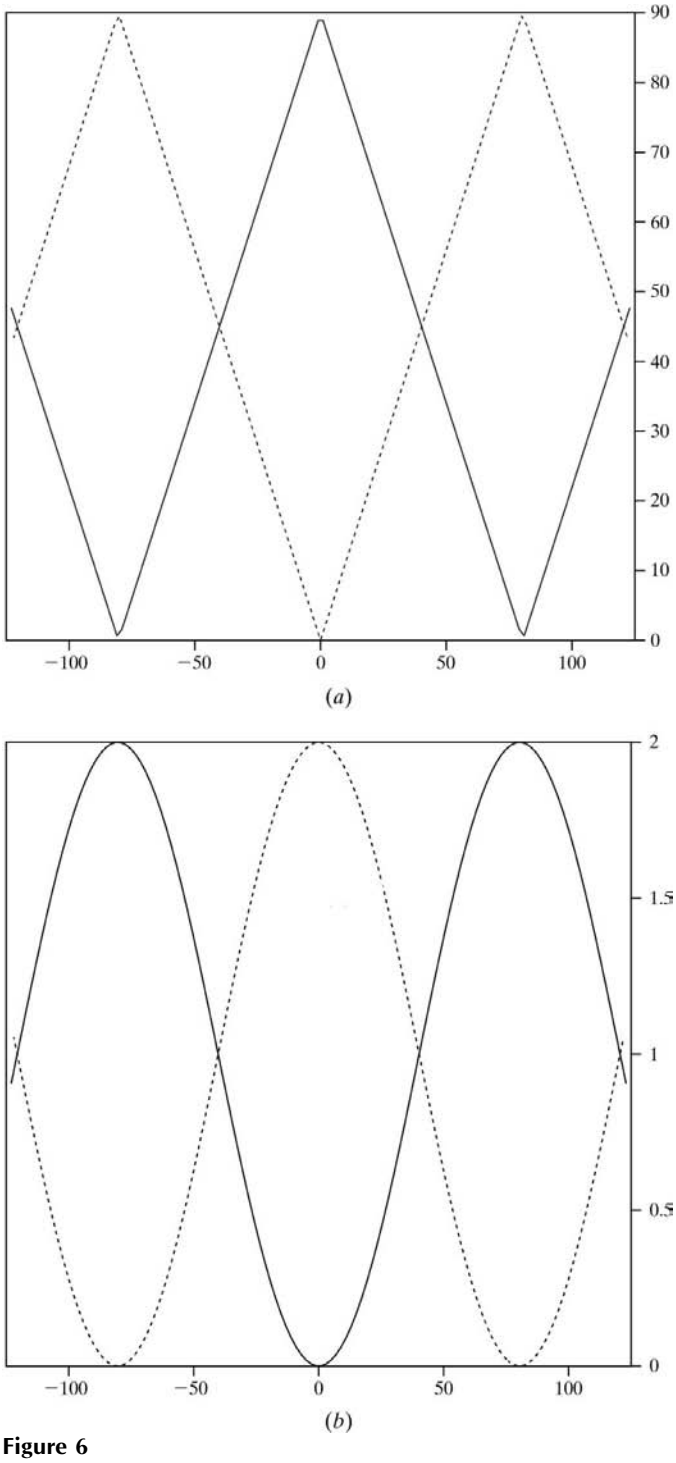

Figure 6
Anti-correlated properties of phase-errors and amplitudes. (*a*) In this figure, continuous lines refer to odd values of *l* and broken lines to even values of *l*. The theoretical phase error obtained from simple geometrical considerations differs depending upon the parity of *l*. The phase splitting can be quite substantial. Low-resolution reflections with odd *l*s are affected more, since in this case the reflections with smaller *l*s have considerable phase errors. Reflections with even *l*s, on the other hand, are affected more with increasing resolution. At about $l \simeq 40$, the phase error arising from space-group error is $\pi/4$ or 45° and is the same for both parities. (*b*) The sinusoidal modulating factor affecting the amplitudes of the resultant reflections. Low-resolution reflections with odd *l* are suppressed, while even-*l* reflections with low values are spared. At higher resolution the roles are reversed. The 'crossover' point occurs again at about $l \simeq 40$.

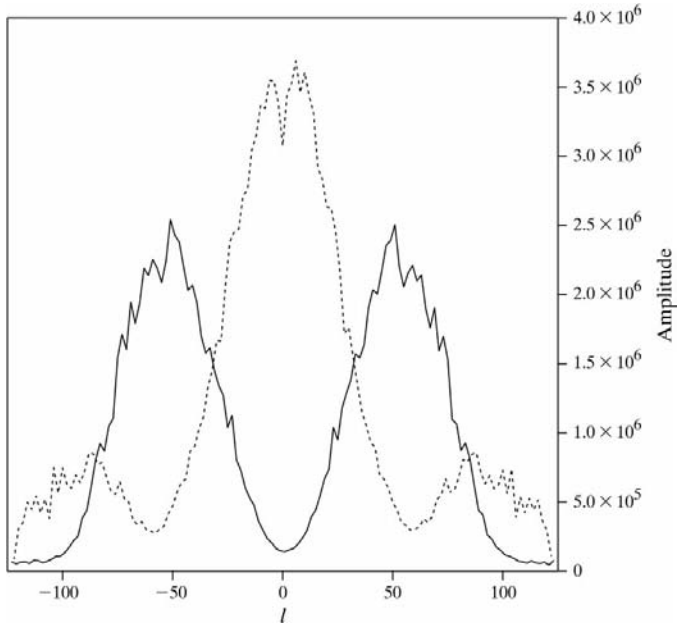

Figure 7
The amplitude of the reflections as a function of *l*. There is a clear-cut distinction between the distributions followed by those reflections with odd values of *l* (full line) and those with even values of *l* (dashed line).

group of the data is $R\bar{3}m$, illustrating that such a condition is necessary but not sufficient for an *R*32 structure.

Accumulating all reflections with the same value of *l* in one bin, calculating the average amplitude and plotting this average as a function of *l* results in the plot shown in Fig. 7. This plot has two main characteristics. First of all, it is highly symmetrical. Since Friedel pairs are not included, this reflects the equality of the amplitudes of the pseudo-equivalent reflections. The most striking feature is, however, the clear-cut partition of the reflections, with those with odd values of *l* following one distribution and those with even values of *l* following another. This partition is a consequence of the modulating sinusoidal factor (equations 15 and 16 or right side of Fig. 6) discriminating the reflections by the parity of *l*.

A plot of the average amplitudes as function of a reciprocal value which shows the characteristics of the plot in Fig. 7 may thus be considered to be a strong sign of a translational pseudo-symmetry of the type described for NDO.

We would like to thank Professor Hans Eklund for support during the work and Drs Phil Evans, Ignacio Fita and Eleanor Dodson for useful discussions. In addition, we thank the referees for valuable comments. The project was funded by a grant from the Swedish Agriculture and Forestry Research Fund to SR.

References

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.

Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
Haigler, B. E. & Gibson, D. T. (1990*a*). *J. Bacteriol.* **172**, 457–464.
Haigler, B. E. & Gibson, D. T. (1990*b*). *J. Bacteriol.* **172**, 465–468.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kauppi, B., Lee, K., Carredano, E., Parales, R. E., Gibson, D. T., Eklund, H. & Ramaswamy, S. (1998). *Structure*, **6**, 571–586.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 108–230.
Kleywegt, G. J. & Jones, T. A. (1998). *Acta Cryst.* D**54**, 1119–1131.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Lamzin, V. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, K., Kauppi, B., Parales, R. E., Gibson, D. T. & Ramaswamy, S. (1997). *Biochem. Biophys. Res. Commun.* **241**, 553–557.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Resnick, S. M. (1996). *J. Ind. Microbiol. Biotechnol.* **17**, 438–457.