Low resolution *ab initio* phasing of *Sarcocystis muris* lectin SML-2

Jürgen J. Müller,^a Natalia L. Lunina,^b Alexandre Urhumtsev,^c
Edgar Weckert^d, Udo Heinemann^{a,e} and Vladimir Y. Lunin^b

^a*Macromolecular Structure and Interactions, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Str. 10, D-13092 Berlin, Germany*

^b*Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia*

^c*Département de Physique, Université H.Poincaré Nancy 1, 54506 Vandoeuvre-lès-Nancy, France*

^d*Hasylab at DESY, Notkestr.85, D-22603 Hamburg, Germany*

^e*Institut für Chemie und Biochemie, Freie Universität, Takustr. 6, D-14195 Berlin, Germany*

Running title: Low resolution *Sarcocystis muris* lectin SML-2 structure.

Correspondence e-mail: jjm@mdc-berlin.de

Synopsis

The connectivity-based *ab initio* phasing method was applied to obtain the first low resolution crystallographic images from crystals of the lectin SML-2.

Abstract

The structure analysis of the lectin SML-2 faced difficulties when applying standard crystallographic phasing methods. The connectivity-based *ab initio* phasing method allowed to compute a 16-Å resolution Fourier synthesis and derive primary structural information. It was found that SML-2 crystals have three dimers in the asymmetric part of the unit cell linked by a non-crystallographic symmetry close to the translation by $(0,0,\frac{1}{3})$. A clear identification of the non-crystallographic two-fold axis explains a space group transformation from the primitive $P2_12_12_1$ to C -centered $C222_1$ observed during annealing procedures within an N_2 cryostream for co-crystals of SML-2 and galactose. Related packing considerations predict a possible arrangement of SML-2 molecules in a tetragonal unit cell. Multiple non-crystallographic symmetries and crystal forms provide a basis for further image improvements.

1. Introduction

Sarcocystis muris is an intracellular cyst-forming parasite that propagates in mice as the intermediate and cats as the final host (Müller *et al.*, 2001; Klein *et al.*, 2003). The parasite lectin SML-2 belongs to a family of three highly conserved galactose specific lectins (Müller *et al.*, 2001). Entzeroth *et al.* (1992) demonstrated the secretion of these proteins during an early phase of host cell invasion, and localized the lectins at the moving junction where they contribute to parasite motility. The biologically active unit is a homodimer (Montag *et al.*, 1997), and each monomer consists of 138 amino acids corresponding to a molecular mass of 15,066 Da.

Traditional ways to solve the structure were unsuccessful. Because of low sequence match to any other protein of known structure, molecular replacement failed. Furthermore, no heavy atom phasing was successful so far, and selenomethionine derivatives were not available because the protein has been prepared directly from the parasites (Müller *et al.*, 2001). At the same time, a complete very-low resolution data set collected for the native

crystal let us to apply *ab initio* phasing methods (Lunin *et al.*, 2002) and obtain the first low resolution structure images.

2. Experimental part and programs

Protein preparation and crystallization has been described previously (Müller *et al.*, 2001). The crystal with dimensions of $0.2 \times 0.2 \times 0.2 \text{ mm}^3$ belongs to the primitive orthorhombic space group $P2_12_12_1$ and diffracts to about 5-6 Å resolution at 100 K at the HASYLAB PETRA I undulator beamline at the PETRA storage ring, DESY, Hamburg. We used a 6-circle goniometer (Huber, Germany) in combination with a 165 mm MARCCD detector at a 425 mm crystal to detector distance, and with a non-focused parallel beam. A dataset was collected with an overall R_{sym} of 6.6% up to 6-Å resolution with 100% completeness up to 10 Å, and 98.9% up to 6 Å (Table 1). The more detailed statistics as determined by the data evaluation program *XDS* (Kabsch, 1993) are given in Table 2.

Co-crystallization of the lectin SML-2 with a modified galactose (Müller *et al.*, 2001) resulted in a second crystal form, $C222_1$ with a unit cell of $a=74.70 \text{ Å}$, $b=81.97 \text{ Å}$, and $c=131.00 \text{ Å}$ (Table 1). In an N_2 cryostream, after annealing this crystal switched to space group $P2_12_12_1$ yielding a crystal isomorphous to that of the native apo-protein.

Only one of multiple efforts to obtain heavy atoms derivatives was successful. A *Pt* derivative data set (formal resolution from 3.8 to 19.9 Å; practically no data at the resolution lower than 16.6 Å; details of statistics not shown) has been collected at rotating anode at home. Program *OASIS* (Hao *et al.*, 2000), with six heavy atoms per asymmetric part of the unit cell, followed by *DM* (Cowtan, 1994) from the program suite *CCP4* (1994) provided us phase values for data in the resolution range 5.0-15.0 Å. The phasing power was very low and the corresponding Fourier syntheses were completely noisy and useless.

Further work with the native data set was done using program *MOLREP* (Vagin & Teplyakov, 1997), *ab initio* phasing program *GENMEM* (Lunina *et al.*, 2000), program *IMP* from the program suite *RAVE* (Kleywegt *et al.*, 2001) and graphic programs *FFT-CAN* (Vernoslova & Lunin, 1989) and *PyMOL* (DeLano, 2002), as described below.

3. Preliminary analysis

For the Matthews coefficient to be in the range of 2.3-3.0 Å³/Da, the asymmetric part of the unit cell of the $P2_12_12_1$ space group crystals could contain 6-8 monomers. The protein content of the unit cell is then 41.2-54.2%. The unit cell in space group $C222_1$, being roughly 1.5 times smaller in volume, provides space for 4-6 monomers in the asymmetric unit.

The self-rotation function calculated by *MOLREP* with the available native data strongly indicated a non-crystallographic two-fold axis orthogonal to the b axis and forming an angle with the a axis of approximately 45° (Fig. 1a). The observed height of 6.96 σ of the corresponding self-rotation peak is comparable with the height of 7.60 of the origin peak and exceeds more than 4-fold the height of all other peaks. It is worth noting that, together with the crystallographic 2₁ axis parallel to a , this non-crystallographic symmetry generates four-fold non-crystallographic axis parallel to b producing an equally strong peak in the self-rotation function in the $\kappa=90^\circ$ section (not shown) of the self-rotation function.

4. Connectivity-based *ab initio* phasing

An application of the connectivity-based *ab initio* phasing method to the lectin data was feasible due to exceptional completeness of this data set at low resolution. The details of this method have been described previously (Lunin *et al.*, 2000; Lunina *et al.*, 2003; Urzhumtseva *et al.*, 2004) and here we recall only that the main steps of the procedure are:

- Large numbers of random phase sets are generated; at the initial run the phases are generated uniformly, at the following iterations they are generated taking into account values obtained previously and their reliability.
- For each phase set the corresponding Fourier map is calculated with experimental magnitudes; in each map the region of a relatively high density is analyzed, the connected components in this region are identified, and their volumes are estimated.

- The maps (and corresponding phase sets) that satisfy prescribed conditions (for example, a given number of connected components, absence of small ‘noisy drops’, etc.) are accepted for further analysis.
- Optimal alignment is performed followed by averaging of selected images; the averaging produces approximate phase values and also their figures of merits; this information is used in next iterations.

Giving initially phases for a few tens of reflections, the procedure improves and extends this phase set.

5. *Ab initio* phasing of the native SML-2 data

To start the procedure of connectivity-based phasing we should have some ideas how the high-density region looks like, in terms of connectivity analysis. It is reasonable to expect that at low resolution molecules appear as isolated ‘blobs’ of approximately equal volume, if an appropriate cut-off level is used (our experience suggests to choose the level resulting in a molecular mask with a volume of 25-30 Å³/residue). Nevertheless, if two molecules form, for example, a dimer with a close interface, then such dimer may be seen as a single ‘blob’ of a larger volume. In the study of SML-2 the precise number of monomers (or dimers) in the unit cell was not known, so that we tried the phasing using different start hypotheses, namely 2 or 3 or 4 dimers in the asymmetric part of a molecular mask covering 15% of the unit cell volume, or 4 or 6 or 8 monomers in a molecular mask covering 7% of the unit cell.

The search for a phase set that gives a 21-Å resolution Fourier synthesis with 3 connected isolated components (supposed to correspond to three dimers) per asymmetric unit resulted in a number of images with 3 blobs of approximately the same volume. The search for two blobs resulted in a number of images with two blobs, one of them being twice as large than the other. This may be interpreted as three components of equal volume, two of them being merged. The search for 4 dimers resulted in a number of images with 3 blobs of nearly the same volume and an additional very small, marginal ‘drop’. This may be interpreted just again as three equal blobs plus noise. The attempts to search for isolated monomers resulted in images with drops of very different volumes. These tests led us to

accept, accordingly to Urzhumtseva *et al.* (2004), the hypothesis of three dimers in the asymmetric unit seen as 3 blobs of connected density as a selection rule for initial steps of phasing.

At the first three steps of phasing the selection rule was simply ‘3 blobs per asymmetric unit’ for Fourier syntheses in several resolution zones. The particular resolutions were (27, 25, 22, 20 Å), (27, 25, 22, 20, 18 Å), and (27, 18, 17 Å), respectively for these steps. At every step 100 selected phase sets were aligned and averaged producing approximate phases and their individual figures of merit as the output. The alignment was carried out at resolutions of 21, 19 and 18 Å, respectively.

At the next three steps the additional requirement was introduced into the phasing, that the three blobs have roughly the same volume. In addition, the resolution zone was extended to 15 Å (228 unique reflections). The results obtained are discussed in the next section. The variation of figures of merit from step to step is shown in Table 3.

6. Results

6.1. Molecular packing

The Fourier maps obtained after first three steps of *ab initio* phasing showed very clearly the molecular packing (Fig. 2). These maps reveal three high-density regions in the asymmetric part of the unit cell, similar in their shapes and volumes. It is worth noting that neither equivalent volumes nor similar shapes were demanded during these steps of phasing. An internal symmetry of the obtained high-density regions (this symmetry is discussed in more detail below) made us conclude that each of these regions corresponds to a lectin dimer. This was confirmed by the size of these regions when an appropriate density cut-off level was chosen so that the corresponding envelopes touch each other. Therefore, the agreement of the number, shape and size of obtained regions with independent experimental information can be considered as an experimental evidence of the results in the situation when an atomic model is still unknown.

The three dimer regions demonstrate quasi regular arrangement in the unit cell close to the translation by a $(0,0,\frac{1}{3})$ vector. This is taken to indicate that at a low resolution this molecular packing may be described by a crystal with the same parameters a and b and three times smaller parameter c . This ‘hidden’ additional periodicity in the c -direction implies that the reflections with indices $h,k,3n+1$ and $h,k,3n+2$ should be negligible. Fig. 3 indicates a weak form of such a phenomenon in the 16-Å resolution zone and its absence at a higher resolution. As a result, two non-crystallographic symmetries close to $(0,0,\frac{1}{3})$ and $(0,0,\frac{2}{3})$ are indeed approximate and non-crystallographic (with a possible rotation component) rather than additional exact periodicity. The presence of quasi periodicity in the crystal is confirmed by Patterson syntheses (Fig. 4) that demonstrate prominent peaks at $(0,0,\frac{1}{3})$ and $(0,0,\frac{2}{3})$. In a 16-Å resolution map these peaks have the height of 10 σ (to be compared with 14 σ for the origin peak) and are about five times stronger than the other peaks. In a 6-Å resolution map these peaks are of 14 σ height and still dominate all other peaks with the exception of a 83 σ origin peak. However now they are split demonstrating a component along a -axis, a small shift additional to $(0,0,\frac{1}{3})$ and $(0,0,\frac{2}{3})$. We used the *ab initio* phased syntheses and the program *IMP* (Kleywegt *et al.*, 2001) to refine non-crystallographic symmetry transformations. This analysis revealed indeed a presence of about 5 degrees rotation component.

The same program *IMP* estimated the similarity of these regions of Fourier syntheses numerically (Table 4). Corresponding density correlation is high especially when the weighted synthesis was analysed. A lower correlation in the non-weighted synthesis and lower values of figures of merits for higher resolution reflections indicate a need in a further refinement of corresponding phase values.

6.2. Non-crystallographic dyad

A visual analysis of the shape of three dimer regions linked by non-crystallographic symmetry show that each of them has an approximate two-fold symmetry (Figs. 5, 6). It is present both at low and high density cut-off levels, while some break in the symmetry caused by phase errors are observable at high levels. The directions of dyads for three

crystallographically independent dimers are close to each other and coincide with the direction of the rotation axes indicated by the self-rotation function for $\kappa=180^\circ$ (Fig. 1). It is worth to note that no condition relevant to this symmetry was included into the phasing procedure. We concluded that the appearance of this symmetry in the images obtained may serve as one more indicator of the correctness of the phase values obtained by our procedure.

This internal symmetry detected by a visual analysis of maps has been confirmed numerically (Table 4) especially when a proper weighting is used.

6.3. Switch between crystal forms

To simplify the presentation, in this section we use notation $^{[a]}2_1$, $^{[b]}2_1$, $^{[c]}2_1$ to distinguish 2_1 axes parallel to a , b and c coordinate axes, respectively. a' , b' , c' stand for the coordinate axes in new coordinate systems and x' , y' , z' for corresponding coordinates.

The observed molecular packing suggests an explanation of possible switches between the crystal forms for lectin SML-2. First, pseudo-translation symmetry suggests the existence of crystals with a 3-fold shortened c -axis, *i.e.* unit cell dimensions of around $53 \times 130 \times 53$ Å. These crystals possess $P2_12_12_1$ space group symmetry. The asymmetric part of the unit cell contains two monomers that form a dimer through a non-crystallographic dyad.

The self-rotation function and analysis of *ab initio* phased Fourier maps define that the dimer's dyad is in the plane parallel to ac -plane and constituting the angle about 45° with the a -axis. Furthermore, the map analysis places this dyad into the plane $y = \frac{1}{8}$. Together with the $P2_12_12_1$ symmetry, this results in $P4_32_12$ space group symmetry (Fig. 7) with the same unit cell about $53 \times 53 \times 130$ Å, but now with a single monomer per asymmetric unit. This means that modification of crystallization conditions may lead to a small shift and rotation of dimers resulting in crystals with $P4_32_12$ symmetry. Taking into account that at this stage of phasing the enantiomer choice is free, the $P4_12_12$ space group must be considered as a possible alternative to $P4_32_12$.

Small rotations and shifts of dimers may result in one more different crystal form. Let us suppose that, as a result of a small modification of molecule packing, the non-crystallographic translations and dyad become true crystallographic symmetry. This dyad, that

initially passes very close to the crystallographic $^{[b]}2_1$, after such a minor rearrangement may intersect this axis exactly. Let us suppose, too, that this rearrangement slightly distorts two other two-fold crystallographic axes $^{[a]}2_1$ and $^{[c]}2_1$, making them non-crystallographic. In this case the structure adopts the $C222_1$ space group symmetry (Fig. 8) with the unit cell parameters of about $75 \times 75 \times 130$ Å. Now an asymmetric part of this unit cell contains four monomers linked by non-crystallographic symmetry operators that are close to the former crystallographic axes $^{[a]}2_1$ and $^{[c]}2_1$. These non-crystallographic axes must be nearly parallel to the new $a'b'$ -plane and constitute the angles close to 45° with the new a' and b' axes. It is interesting that crystals of the lectin co-crystallised with galactose belong to this space group and have quite similar unit cell dimensions (see Tab. 1). A low resolution (*e.g.*, at 6 Å) self-rotation function (Fig. 1b) reveals the peaks corresponding to two-fold rotation axes in accordance with our analysis. When compared with Fig. 1a, these peaks are weaker (4.3σ in comparison with 10.1σ for the peak at the origin and 1.7σ for the next peaks). When the self-rotation function is calculated at a higher resolution (*e.g.*, at 3 Å) the peaks decrease to the level of noise and may be missed.

The position of non-crystallographic axes in all these crystal forms is in agreement with the observation by Wang & Janin (1993) on the closeness of non-crystallographic symmetry axes to the crystallographic ones.

6.4. Heavy atom analysis

One more independent source of information was *Pt* derivative. A direct comparison of phases obtained by the two approaches was impossible because for no reflection both phase values were available simultaneously. The same reason prevented us to use difference Fourier synthesis with *ab initio* phases to check the heavy atom positions. At this level of studies, the only way to compare results of these approaches was to analyze the positions of *Pt* atoms.

When comparing two phase sets, the possibility of a different choice of the unit cell origin and enantiomer must be taken into account (for a general case analysis see Lunin & Lunina, 1995). The same alignment procedure should be applied when comparing heavy atom positions with *ab initio* phased images. In a $P2_12_12_1$ crystal there are 8 possibilities of

permitted origin choice and both enantiomers are compatible with the space group. All possible 16 transformations were applied to the set of heavy atoms. For all but two transformations heavy atoms were obtained either between the envelopes or inside them, and the corresponding transformations were eliminated. The transformation $(x,y,z) \rightarrow (-x+1/2, -y, -z+1/2)$ positioned the heavy atoms more or less at the surface of the molecular envelope when the envelope volume is equal to 20% of the total unit cell volume (envelope volume is defined by a density cut-off level applied). This volume is unrealistically small, and therefore the heavy atoms seem to be too close to the centers of dimers. The transformation $(x,y,z) \rightarrow (-x+1/2, -y+1/2, -z+1/2)$ positioned the heavy atoms nicely at the surface of the molecular envelope when the corresponding density cut-off level is significantly lower and the envelope volume is equal to 50% of the total unit cell volume (Fig. 9). It is worth noting that the heavy atom positions are in agreement with the non-crystallographic dyads of dimers, as defined from *ab initio* phasing.

This result shows that the found *ab initio* phases do not contradict to preliminary information obtained with Pt-derivative. Unfortunately for the moment low quality of SIR phases together with the lost central zone reflections make a more detailed comparison impossible.

7. Conclusions

This work illustrates possibilities of *ab initio* phasing in solving problems in protein crystallography that cannot be approached otherwise. Low-resolution images give an explanation of the existence of several crystal forms and predict the presence of new crystals, in particular a tetragonal form. In the crystals with multiple copies (6 independent monomers in $P2_12_12_1$ crystals), these phases and envelopes are a good starting point for further phase extension through non-crystallographic symmetry averaging. Based on the *ab initio* obtained images, a link between $P2_12_12_1$ and $C222_1$ crystals has been established. This information permits a direct transfer, without any molecular replacement searches, of the (improved) lectin envelopes into the $C222_1$ crystals for which higher resolution data are available. These will be the obvious next steps in the crystal structure analysis of the lectin SML-2.

The currently available low-resolution set of experimental magnitudes and *ab initio* determined phases has been deposited as a supplementary file to this article.

Acknowledgement

We are grateful to H. Klein, Th. Montag, N. Zyto and B. Löschner from the Paul-Ehrlich-Institut, Bundesanstalt für Seren und Impfstoffe, FG Parasitologie/Diagnostika, 63225 Langen, Germany, who provided the original lectin SML-2. The modified galactose was a gift from U. Pfüller, University of Witten-Herdecke, Germany. This work was supported by the Fonds der Chemischen Industrie. NLL and VYL were supported by RFBR grant 03-04-48155 and joint RFBR-CNRS grant 05-01-22002-CNRS_a. AU was supported by the CNRS-RAS collaboration, GdR 2417 CNRS and by Pole “Intelligence Logicielle”, Nancy.

References

- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.*, **D50**, 760-763.
- Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34-38.
- DeLano, W.L. (2002). *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>
- Entzeroth, R., Kerkhoff, H. & Koenig, A. (1992). *Eur. J. Cell. Biol.*, **59**, 405-413.
- Hao, O., Gu, Y.X., Zheng, C.D. & Fan, H.F. (2000). *J. Appl. Cryst.* **33**, 980-981.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795-800.
- Klein, H., Mueller, S., Loeschner, B., Toenjes, R. R., Braun, G., Müller, E.-Ch., Otto, A. & Montag, T. (2003). *Parasitol. Res.*, **90**, 84-86.
- Kleywegt, G.J., Zou, J.Y., Kjeldgaard, M. & Jones, T.A. (2001). In: *International Tables for Crystallography*, Vol.F, (Rossmann, M.G. & Arnold, E., Editors). Chapter 17.1, pp. 353-356. Dordrecht: Kluwer Academic Publishers, The Netherlands
- Lunin, V.Y. & Lunina, N.L. (1996). *Acta Cryst.*, **A52**, 365-368.

- Lunin, V.Y., Lunina, N.L. & Urzhumtsev, A. (2000). *Acta Cryst.*, **A56**, 375-382.
- Lunin, V.Y., Lunina, N., Podjarny, A., Bockmayr, A. & Urzhumtsev, A. (2002). *Z. Kristallogr.*, **217**, 668-685.
- Lunina, N., Lunin, V.Y. & Urzhumtsev, A. (2003). *Acta Cryst.* **D59**, 1702-1715.
- Montag, T., Bornhak, M., Loeschner, B., Klein, H., Otto, A., Zyto, N. & Entzeroth, R. (1997). *Eur. J. Cell Biol.* **74**, Suppl. 46, 21.
- Müller, J. J., Müller, E.-Ch., Montag, Th., Zyto, N., Löschner, B., Klein, H., Heinemann, U. & Otto, A. (2001). *Acta Cryst.* **D57**, 1042-1045.
- Urzhumtseva, L., Lunina, N., Fokine, A., Samama, J.-P., Lunin, V.Y. & Urzhumtsev, A. (2004). *Acta Cryst.*, **D60**, 1519-1526.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022-1025.
- Vernoslova, E.A. & Lunin, V.Y. (1993). *J. Appl. Cryst.*, **26**, 291-294.
- Wang, X. & Janin, J. (1993). *Acta Cryst.*, **D49**, 505-512.

Table 1. Statistics for native *Sarcocystis muris* lectin SML-2 crystals and SML-2/galactose co- crystals.

Crystal	Native SML2	Co-crystal SML-2/galactose	Co-crystal SML-2/galactose
X-ray source	HASYLAB PETRA I MARCCD-165	Home RU H2B MAR300	BESSYII BL1 MARCCD-165
Space group	$P2_12_12_1$	$C222_1$	$P2_12_12_1$
Unit cell axes a,b,c [Å]	53.30, 129.48, 157.63	74.70, 81.97, 131.00	53.49, 130.71, 159.86
Mosaicity [°]	0.69	0.37	0.37
Resolution [Å]: overall	110.-6.0	20.-2.45	50.-2.56
outer shell	20.0-6.0	2.54-2.45	2.70-2.56
Observ./unique reflect.	45,384 / 2,974	73,314 / 15,121	485,661 / 36,856
R_{sym} [%], overall/outer	6.6 / 6.9	4.9 / 15.7	6.9 / 18.7
Completeness [%], overall/outer	100. / 98.8	99.8 / 99.0	99.6 / 98.3
Average redundancy	see Table 2	4.85	13
$\langle I \rangle / \langle \sigma_I \rangle$, overall/outer	32.7 / 30.3	6.9 / 18.7	28.7 / 14.3

Table 2: Very-low resolution data statistics calculated with XDS (Kabsch, 1993)

Resolution limit [Å]	No. of reflections			Completeness	R-factor	$\langle I \rangle / \langle \sigma_I \rangle$	R_{meas}
	observed	unique	possible	[%]	[%]		[%]
100	4	1	1	100	5.4	27.9	6.2
70	3	1	1	100	43.6	4.5	53.3
50	27	5	5	100	39.7	15.0	43.7
30	227	26	26	100	3.9	31.8	4.3
20	964	70	70	100	9.2	32.9	9.6
10	10,073	589	597	98.7	5.9	44.2	6.0
6	34,086	2,282	2,308	98.9	6.9	30.3	7.1

Table 3. **Mean figure of merit during *ab initio* phasing of lectin**

Resolution zone (Å)	∞ -28.0	∞ -25.0	∞ -20.0	∞ -16.0
# refl.	38	56	103	189
Step of phasing				
1	.26	.21	.17	.13
2	.31	.26	.23	.18
3	.34	.30	.27	.22
4	.36	.32	.31	.25
5	.39	.36	.34	.28
6	.41	.38	.36	.30

Table 4. Internal correlation of the Fourier syntheses calculated with *ab initio* phases at the resolution of 16 Å. The table shows correlation of Fourier syntheses values in three regions presumably corresponding for the dimers A, B, C and also the internal two-fold symmetry for each of these regions.

<i>Unweighted</i>	A	B	C	<i>Weighted</i>	A	B	C
A	1.00	0.65	0.76	A	1.00	0.95	0.97
B	0.65	1.00	0.74	B	0.95	1.00	0.95
C	0.76	0.74	1.00	C	0.97	0.95	1.00
Internal	0.78	0.31	0.64	Internal	0.84	0.83	0.84

FIGURE LEGENDS

Fig. 1. **Self-rotation functions (a) for native SML-2 ($P2_12_12_1$), and (b) for native SML-2 co-crystallized with galactose ($C222_1$) at 6 Å resolution.** The sections corresponding to rotation by $\kappa=180^\circ$ are shown.

Fig. 2. **Fourier synthesis at 18 Å resolution calculated with the phases obtained after 3 cycles of connectivity-based *ab initio* phasing.** Each ‘blob’ corresponds to a lectin dimer. The slice $0 \leq y \leq \frac{1}{2}$ is shown in projection along b -axes. The unit cell border is shown in red, and $P2_12_12_1$ space group symmetry in blue. The dimers A, B and C are linked by non-crystallographic symmetry close to the translation by $(0,0,\frac{1}{3})$ vector. The dimers A and A', B and B', C and C' are linked by crystallographic symmetry. The used cut-off level 1.0σ isolates 15% of the unit cell volume that corresponds to the specific volume of 50.0 Å^3 per residue.

Fig. 3. **Observed structure factor amplitudes *versus* l -index value.** (a) $d \geq 16.0 \text{ Å}$ resolution zone; (b) $16.0 \text{ Å} > d \geq 8.0 \text{ Å}$ resolution zone. Low resolution data demonstrate relatively large values for $h,k,3n$ reflections. This is not the case for medium resolution data.

Fig. 4. **Section $y=0$ of Patterson maps calculated with SML-2 data.** (a) 16 Å resolution, contours starting from 2σ with the step 2σ are shown; (b) 6 Å resolution, contours starting from 2σ with the step 4σ are shown. Strong peaks in the 16 Å map, comparable in height with the peak in the origin, indicate pseudo-translation symmetry. These peaks are split in 6 Å resolution map indicating the non-crystallographic nature of the translation close to $(0,0,\frac{1}{3})$.

Fig. 5. **3D image of three dimers.** Fourier synthesis (18 Å resolution) calculated with phases obtained after 3 cycles of connectivity-based *ab initio* phasing viewed following *b*-axis. Three dimers linked by non-crystallographic symmetry are shown. The blue contour corresponds to a 1.0 σ cut-off level, the green contour corresponds to 2.0 σ . The image demonstrates an approximate symmetry of the dimers. The high level region is broken in two parts for two of three dimers due to phase errors present. Broken arrows indicate the direction of the non-crystallographic dyad as defined by the self-rotation function.

Fig. 6. **3D image showing a lectin dimer.** Fourier synthesis (18 Å resolution) calculated with phases obtained after 3 cycles of connectivity-based *ab initio* phasing looking down the direction of the two-fold non-crystallographic symmetry axis as defined from the self-rotation function. The blue contour corresponds to a 1.0 σ cut-off level, the green contour corresponds to 2.0 σ .

Fig. 7. **Fourier synthesis map demonstrates an approximate P4₃2₁2 space group symmetry.** One of possible choices of the unit cell is indicated. Two subsequent slices (a) $-\frac{1}{8} \leq z' \leq \frac{1}{8}$ and (b) $\frac{1}{8} \leq z' \leq \frac{3}{8}$ are shown. Symmetry elements of the initial P2₁2₁2₁ space group are in red; symmetry elements induced by non-crystallographic translations are in green; symmetries induced by a non-crystallographic dyad are in blue, (x', y', z') are coordinates in the P4₃2₁2 coordinate system. The map contours are shown for 16 Å resolution. The Fourier synthesis was obtained after 6 steps of *ab initio* phasing. The cut-off level corresponds to 2.0 σ (13 Å³ per residue).

Fig. 8. **Fourier synthesis map demonstrates an approximate $C222_1$ space group symmetry.** Two subsequent slices (a) $-\frac{1}{8} \leq z' \leq \frac{1}{8}$ and (b) $\frac{1}{8} \leq z' \leq \frac{3}{8}$ are shown. Symmetry elements of the initial $P2_12_12_1$ space group are shown in red; symmetry elements induced by non-crystallographic translations are in green; symmetries induced by non-crystallographic dyad are in blue, (x', y', z') are coordinates in the $C222_1$ coordinate system. Dimers B and B' are linked now by non-crystallographic symmetry close to the former $^{[a]}2_1$ crystallographic axes (shown by dashed red line). The map contours are shown for 18 Å resolution Fourier synthesis as obtained after 3 steps of *ab initio* phasing. The cut-off level corresponds to 1.5σ (36 Å^3 per residue).

Fig. 9. **The positions of heavy atoms in a Pt derivative superposed with *ab initio* defined envelopes.** Three dimers linked by non-crystallographic pseudo-translation symmetry are shown. The purple spheres indicate positions of Pt atoms (see Section 6.4 for detail). The arrows indicate the direction of the non-crystallographic dyad, as defined by the self-rotation function. The envelope shown occupies 40% of the unit cell volume (see Section 3 for volume estimation). The green surface corresponds to 2.0σ cut-off.

1.5.1.

Fig. 2.

Fig. 3.

Fig. 4.

Fig. 5.

Fig. 6.

Fig. 7.

Fig. 8.

Fig. 9.