Acta Cryst. (1996). D**52**, 1146–1152

X-ray Structure Solution of Amaryllis Lectin by Molecular Replacement With Only 4% of the Total Diffracting Matter

LAURENT CHANTALAT,[a] STEPHEN D. WOOD,[a] PIERRE RIZKALLAH[b] AND COLIN D. REYNOLDS[a]

[a]Biophysics Department, School of Biomolecular Sciences, Liverpool John Moores University, Liverpool L3 3AF, England, and [b]CCLRC, Daresbury Laboratory, Warrington WA4 4AD England. E-mail: bmsswood@vax.livjm.ac.uk

(*Received 10 February 1996; accepted 29 April 1996*)

Abstract

It is often the case that analogous proteins from different species crystallize in a different form. These structures can usually be easily solved by the molecular-replacement (MR) technique, as the protein folding is very often conserved. However, the results from MR become more uncertain as the proportion of diffracting matter decreases as a result of multimericity and/or absence of some of the atoms in the model. In this paper results are presented on the structure solution of amaryllis lectin (109 residues per monomer) containing two protein molecules in the asymmetric unit. The structure was solved by MR using the Cα coordinates of one monomer from snowdrop lectin which has 85% amino-acid sequence identity to amaryllis lectin. This represents only 6% of the non-H atoms of the protein molecule to be used for structure determination and it is a major improvement on previous reports. Further calculations were carried out in order to establish the minimum number of atoms which could be included in the model before a clear solution to the MR problem was revealed. This study showed that the structure of amaryllis lectin could still have been solved easily with 3.85% of the model, which even in the most favourable cases, will probably constitute a minimum for molecular-replacement structure solution.

1. Introduction

The use of the molecular-replacement method (Crowther, 1972) has become more and more popular as the number of protein structures determined by X-ray crystallography has increased dramatically in the last few years. The technique is primarily used to solve protein structures which are expected to show high three-dimensional homology with one of the known structures. These could be either the same protein in a different crystal form, or a structurally related molecule which will share some of the important features of the structure under investigation. The problem is simplest when the native protein belongs to a different crystal form and gets more complicated as the sequence homology decreases. For our study, the difficulty would have been low because of the high sequence homology between the two molecules involved (17 sequence changes per monomer). As at first only the Cα atoms of one monomer were available as a search model, the molecular-replacement problem could have been much harder to solve owing to the lack of diffracting matter in this fragment.

Lectins are oligomeric proteins composed of subunits, usually with one carbohydrate binding site per subunit, and can be grouped into families of homologous proteins. The best characterized group of lectins are those from the seeds of leguminous plants (Van Driessche, 1988). In the last ten years a number of new lectins have been isolated from the bulbs, corms, rhizomes and tubers of perennial plants and these storage organs have been shown to be a rich source of lectins of biologically interesting properties (Van Damme, Smeets & Peumans, 1995). Research on non-seed lectins has been focused on the monocotyledonous families Amaryllidaceae, Alliaceae, Orchidaceae, Araceae and Liliaceae. This research has shown that all the lectins characterized to date from Amaryllidaceae, Alliaceae and Orchidaceae species belong to a super-family of mannose-binding proteins (Van Damme *et al.*, 1995) and are potent inhibitors of retroviruses (Balzarini *et al.*, 1991). The crystal structures of the Amaryllidaceae lectin from snowdrop bulbs was determined very recently (Hester, Kaku, Goldstein & Schubert-Wright, 1995). This structure contains a new class of protein fold which consists of three subdomains (I, II and III) each made up of an antiparallel four-stranded β-sheet. The three sub domains are arranged as a 12-stranded β-barrel around an internal pseudo threefold axis.

We have crystallized the tetrameric amaryllis lectin (*Hippeastrum* hybrid agglutinin, HHA) (Wood *et al.*, 1994) in space group $C2$ with cell parameters $a = 73.4\,(1)$, $b = 100.3\,(1)$, $c = 62.2$ Å, $\alpha = \beta = 90$ and $\gamma = 137.2\,(2)^\circ$. Data were collected at the SRS Daresbury Laboratory using the MAR imaging-plate scanner, yielding a data set which was 97% complete to 2.25 Å resolution, with 69% of the data greater than 3σ. The unit-cell volume indicated that there were two molecules per asymmetric unit yielding a packing density of 2.86 Å^3 Da^{-1} (Matthews, 1968). A self-rotation search clearly indicated the presence of non-crystallographic twofold symmetry. The dimeric structure could be determined using only the Cα coordinates of a single

monomer of the snowdrop lectin (*Galanthus nivalis* agglutinin, GNA) structure (Hester *et al.*, 1995) as a search model. This study reports the first example of a multimeric structure containing two molecules in the asymmetric unit solved by molecular replacement with only the Cα coordinates of a monomer. This corresponds to a probe with only 6% of the diffracting matter of the total dimer. The technique was further applied to a smaller probe where Cα coordinates at both ends of the polypeptide chain were removed until the molecular-replacement problem could not be solved any more. This occurred when just 65 Cα atoms or less than 4% out of the 1688 total protein atoms present in the atomic coordinate file. A further experiment showed that when using main-chain and side-chain atoms, the structure could be solved with a truncated fragment (*i.e.* 46 residues or one subdomain and a half) which represented much more of the diffracting matter (21%).

To our knowledge this constitutes the first time that the full molecular-replacement problem [rotation (RF) and translation function (TF)] could be solved with such a small fraction of the total structure in the asymmetric unit. Recently, a hexameric structure in the asymmetric unit was solved using a partial model from one monomer (Oh, 1995). This probe with 16–17% of the total diffracting matter contained slightly more than three times the number of atoms that we used in this work. A few examples of small probe experiments in molecular replacement can be found in the literature. A similar investigation to ours was made with a Fab subdomain, where it was shown that a model made of Cα atoms of the C domain (approximately 6%) could be used to solve the rotation function unequivocally (Cygler & Anderson, 1988*a*). These workers also showed that a correctly orientated Fab subdomain containing only 24% of the asymmetric subunit could be used to solve the translation function (Cygler & Anderson, 1988*b*). Schierbeek, Renestseder, Dijkstra & Hol (1985) have also shown that when an actinidin model, containing only the Cα atoms (13% of the total number of atoms), was exactly in the correct orientation, the highest peak in the TF occurred at the correct position. Thus, in order to obtain phase information, it is essential to have a complete enough model which can be used to solve both the rotation and translation function (except in space group $P1$ where only the orientation is required). In our study we have shown that the use of as little as 4% of the asymmetric unit-cell contents for phasing is enough to determine the correct molecular-replacement solution.

2. Results

The three-dimensional structure of amaryllis lectin was assumed to be similar to that of snowdrop lectin because of the high sequence homology. Both lectins belong to the Amaryllidaceae plant family and exhibit strict specificity for α-D-mannose. They were both characterized to be tetramers of identical subunits (each of about 11 700 Da). Snowdrop lectin was therefore regarded as a good candidate for structure determination by molecular replacement.

AMoRe (Navaza, 1994), an integrated suite of molecular-replacement programs, and *POLARRFN* (W. Kabsch, Collaborative Computational Project, Number 4, 1994) were used to determine the structure of amaryllis lectin. *AMoRe* had previously been used with success in the group to solve a particularly difficult structure (Chantalat, Jones, Körber, Navaza & Pavlovsky, 1995), and therefore we decided to use it for this study.

2.1. *Self-rotation*

The self-rotation function was calculated with *POLARRFN*. As there is only a dimer in the asymmetric unit and the molecule is naturally found as a tetramer, we expected to find the dimer perpendicular to the crystallographic dyad axis along y. Fig. 1 shows the 180° section of the self-rotation function looking along the z axis. Apart from the crystallographic dyad at $\psi = 90$, $\varphi = 90°$, a strong non-crystallographic twofold rotation axis emerges at $\psi = 90$, $\varphi = 0°$, and at the symmetry-related position ($\psi = 0$, $\varphi = 0°$) when using reflections between 15 and 6 Å resolution and an integration radius of 20 Å. The height of the local dyads are 65.2% of that obtained for the crystallographic dyad. The local twofold axes are found in the a/c plane, with one parallel to the a axis, with the other one 90° away from it, therefore

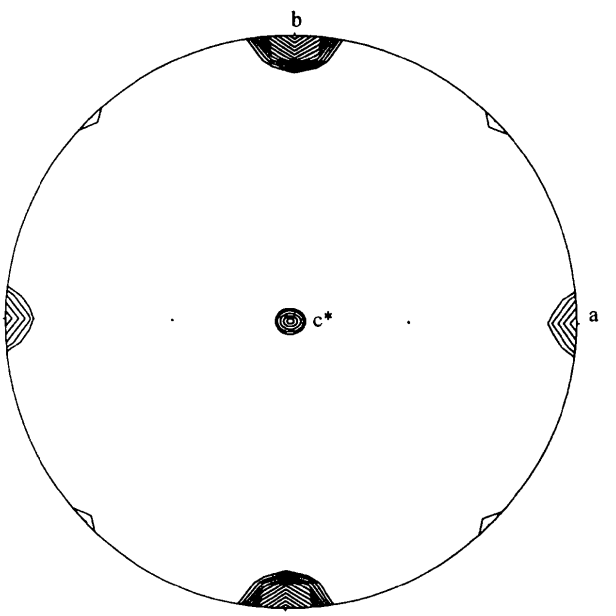

Fig. 1. Section $\kappa = 180°$ of the self-rotation function. Polar angles: ψ, inclination against c^* (z); φ, azimuth measured relative to a (x). Contour levels from 40% in steps of 5%. Peak height for crystallographic twofold along b (y): 100% and peak height for local twofold at (ψ/φ) = (90, 0) and (0, 0): 65.2%.

forming a 222 set with the crystallographic dyad as it was expected. The third highest peak had a height of 45.6%.

2.2. *Rotation function*

The model of GNA which has been refined to an R value of 17% to 2.3 Å resolution (Hester *et al.*, 1995), was first used with Cα atoms only. Later, when the full model was made available from the Protein Data Bank (entry 1MSA) at Brookhaven National Laboratory (Bernstein *et al.*, 1977), it was used to check the correctness of the molecular-replacement solution. The Cα search model contained 109 atoms out of 1688 of the full dimer. Several parameters used for the cross-rotation function, are shown in Table 1, the $P1$ cell size was chosen approximately equal to the smallest box containing the model, plus the integration radius, plus the required resolution (Navaza, 1994). Although we used data between 15 and 4 Å for this work, our calculation also showed that the structure could also have been solved with data from 15–10 to 5–3 Å. The top 50 rotation-function values for the two models are shown in Fig. 2. At first glance, it was clear that in both cases, the top two peaks were the correct solutions, as their α angles, which is the last rotation to be performed, differed by 180° and their β and γ angles were approximately identical. These two solutions were 1.2 and 1.1σ above the highest noise peak (Fig. 2).

2.3. *The translation function and rigid-body refinement*

The one-body translation function was carried out for the 50 RF solution peaks using all data between 15 and 4 Å, this resolution range was also used for the two-body TF and rigid-body refinement. The main parameter to be considered in finding a solution is the correlation coefficient (Navaza & Vernoslova, 1995). Of the TF peaks found, 15 of these were sorted in descending order of their correlation coefficient. It is important to select a large enough number of TF peaks since the highest correlation coefficient might not correspond to the highest TF peak.

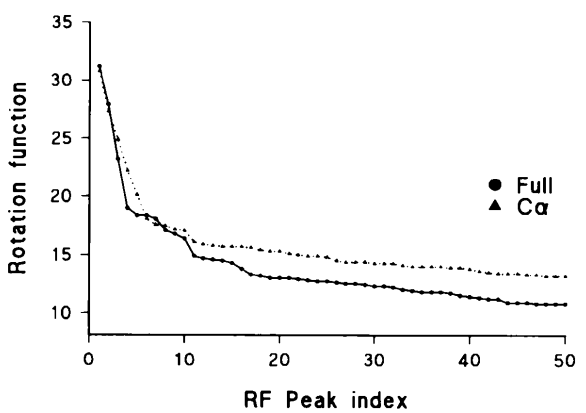

Fig. 2. Rotation search using Cα (▲) and all coordinates (●) of a monomer of snowdrop lectin as search model. The top two peaks correspond to the rotations shown in Table 2.

Table 1. *Parameters used for the cross-rotation function*

P1 cell for the model (Å)	78 × 63 × 58
Resolution range (Å)	15–4
Patterson vector length (Å)	20
Sampling interval for β* ( )	2.5
Assumed overall temperature factor (Å^2)	15
Sharpening factor (Å^2)	−10

* For definition, see Rossmann & Blow (1962).

Fig. 3(*a*) shows that the translation function gives consistent results with the rotation function; the first two peaks give the highest correlation coefficients. Rigid-body (RB) refinement was then applied to each top translation solution (Fig. 3*b*). As there was more than one molecule in the asymmetric unit, a series of two-body TF's and rigid-body refinement were performed for the top one-body TF solutions with the highest ranking one being fixed (Figs. 3*c* and 3*d*). In the two-body situation, the structure factors of the supposedly correct solution, usually the highest overall after RB refinement, contribute to the TF and RB refinement calculations. If the fixed position for one of the molecules has been assigned correctly, the improvement of the calculated phases will probably indicate clearly the position of the second molecule by a much higher correlation coefficient than for the other solutions. The second peak of the rotation function when combined with the top peak gave the best result for the two-body TF and RB refinement (Figs. 3*c* and 3*d*). After the final RB refinement the correct solution given by the Cα model had a correlation coefficient 10% higher than the first noise peak (Table 2) and it also had the lowest R factor of 42.4% (4.5% lower than the highest noise peak).

2.4. *Verification of the solution*

The initial verification was to inspect the model on the graphics system. For a good solution very few clashes should be observed. Also from the self-rotation function and biochemical studies two dimers were expected to form a tetramer around the crystallographic twofold axis. The rotation function appeared to be correct since the first two peaks were related by nearly 180° about the last rotation axis. Inspection of packing within the cell showed the dimers were packed sensibly (Fig. 4). The electron-density map made from this Cα model also showed excellent continuity and some side chains could already be distinguished.

The molecular-replacement solution was confirmed when the full coordinate set was released. The same procedure described above using *AMoRe* was employed with all the atoms of a snowdrop lectin monomer (Table 1) with atomic temperature factors set to 15 Å^2. As there were only 17 residues which were different

Table 2. *Molecular-replacement parameters and S/N ratio for the final dimer*

Model*	α ( )	β ( )	γ ( )	x (Å)†	y (Å)†	z (Å)†	CC (%)	Rf (%)	*S/N* ratio‡
All atoms	353.47	103.83	112.59	0.692	0.627	0.121	73.6	32.2	1.71
(844)	173.10	104.44	111.72	0.006	0.658	0.151	73.6	32.2	
All Cα atoms	352.99	103.82	111.98	0.690	0.628	0.132	52.8	42.4	1.25
(109)	172.32	104.52	112.12	0.014	0.644	0.145	52.8	42.0	
Less Cα atoms	351.90	102.48	111.90	0.707	0.654	0.092	34.5	48.8	1.06
(65)	171.10	104.54	112.44	0.016	0.639	0.137	34.5	48.8	

* Atoms selected for a single monomer. † Fractional coordinates. ‡ *S/N* ratio as (CC for solution peak)/(CC for highest noise peak).

between the GNA and HHA molecules, no changes were made to the search model. Figs. 2 and 3, also show the RF, TF and RB refinement for the full model. After the two-body refinement both models gave similar molecular-replacement parameters differing only by a few angstroms and degrees (Table 2). Although the final CC for the full model is very high (73.6%, Fig. 3*d*), interestingly the *S/N* ratio is 1.71 for all atoms compared to a lower but still significant value of 1.25 for the Cα atoms only (Table 2).

After replacing the 17 residues that differed between the two structures, the complete dimer was refined by simulated annealing with *X-PLOR* (Brünger, Kuriyan & Karplus, 1987). The current model yields a crystallographic *R* factor of 20% and a free *R* factor (Brünger, 1992) of 28%, including all data between 8.0 and 2.3 Å.

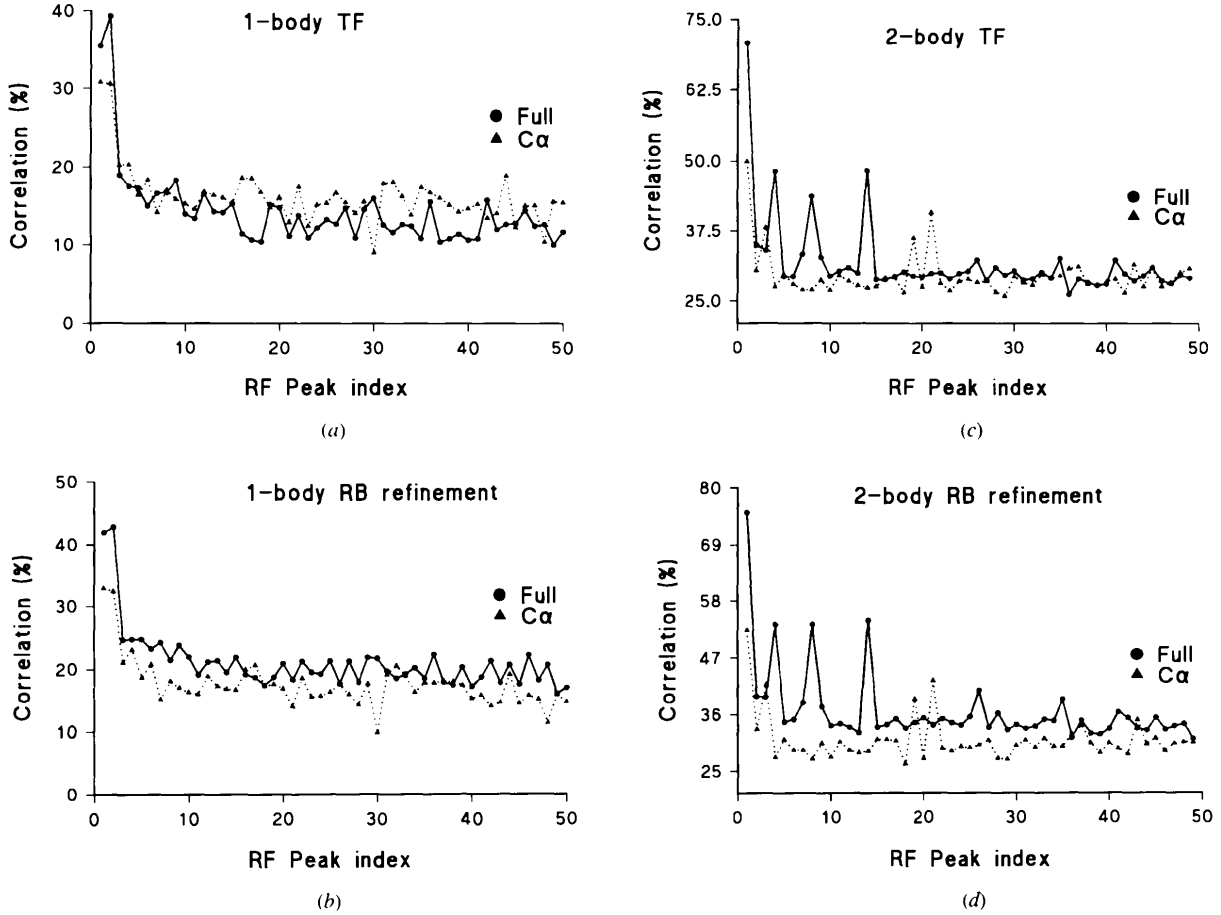

Fig. 3. Translation search and rigid-body refinement using the Cα atoms only (▲) and the full search model (●) of snowdrop lectin against the amaryllis diffraction data. The TF and RB refinement were first applied to each orientation of the search model [one-body TF (*a*) and one-body RB refinement (*b*)]. Then the top solution, in first or second position was fixed, the TF was once more run to find the translation parameters of the second molecule with respect to the fixed one [two-body TF (*c*)]. The orientation and translation parameters of the two molecules was refined in pairs by RB [two-body RB refinement (*d*)].

The electron-density map is of good quality as shown in Fig. 5.

2.5. *Experiments with smaller models*

To find out exactly how incomplete a model for MR could be, we tested more extreme models where further Cα atoms were removed from the calculations. The smallest fragment that could be used to solve the HHA structure easily (*i.e.* the correct solution coming first after the TF or RB refinement) consisted of 65 Cα atoms or 3.85% of the total structure; 22 Cα atoms on both ends of the chain were removed from the initial 109 atoms of one monomer. The correct rotation solutions were still in first and second positions. If one more atom was omitted the solution was drawn down to the tenth place after the translation function. This would not have been too critical since the rotation-function solution was clear, and the incorrect solutions quickly excluded because of bad packing using, for example, the program *CONTACT* (Collaborative Computational Project, Number 4, 1994). All of the remaining solutions could then be checked on the graphics system and further refined if they seemed sensible.

Although our work shows that only a small fragment is required to solve the molecular-replacement problem, it may not be a very practical case, since adding the main-chain atoms to a Cα trace can easily be performed and we have shown that the structure could be solved just as well (unpublished results). In another experiment, we wanted to determine in terms of the number of residues

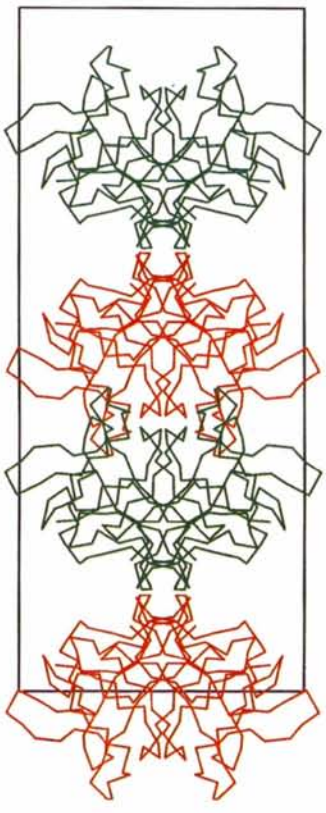

Fig. 4. Packing of four dimers (Cα atoms only) in the unit cell of amaryllis lectin after rigid-body refinement of the search model with *AMoRe*. The view is down the *a* axis with the *b* axis along the length of the paper.

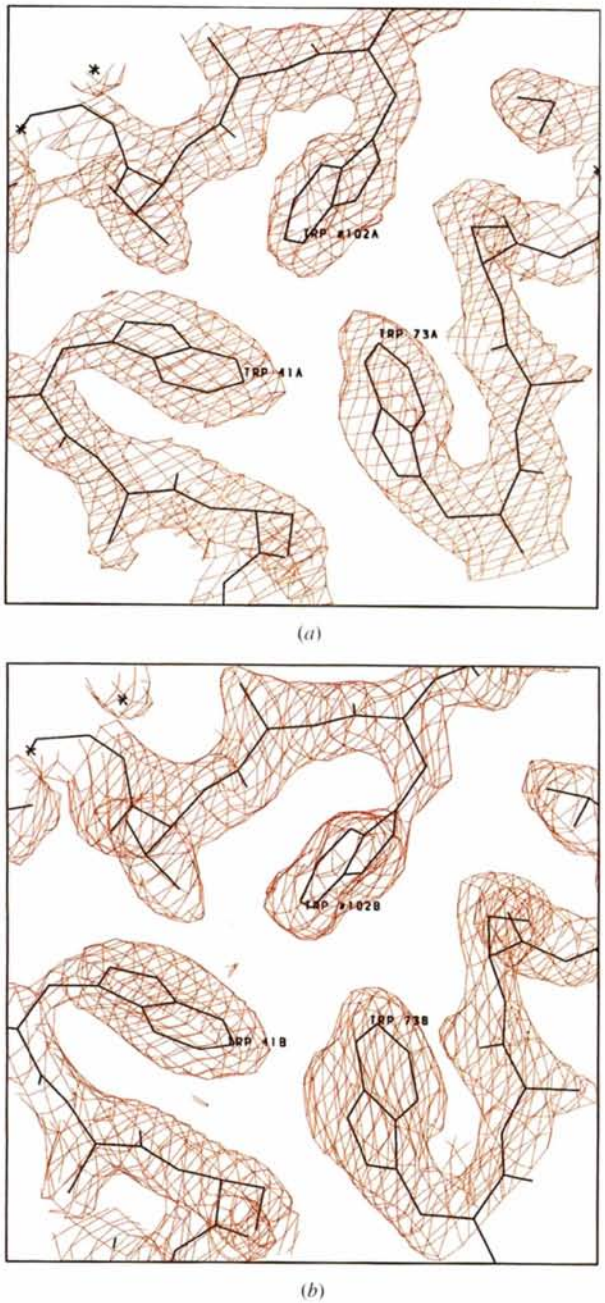

Fig. 5. Equivalent views of the $2F_o - F_c$ electron density around a tryptophan cluster (Trp41, Trp73 and Trp102) for the two independent molecules (*a*) molecule *A* and (*b*) molecule *B*. The map is contoured at 0.6 e Å^{-3}.

employed, the smallest fragment that could be used to solve the HHA structure. For this test all the atoms from each residue were used and the shortest segment which could be used successfully included residues 33–79 (21% of the protein atoms). This corresponds roughly to one subdomain (four β-strands) and a half.

3. Concluding remarks

In the HHA protein, the polypeptide fold consists of three sequential β-sheet subdomains (I, II and III) and each subdomain forms a flat four-stranded antiparallel β-sheet. The β-sheets are inter-related by pseudo three-fold symmetry to form a 12-stranded β-barrel (Fig. 6). Subdomain I is a hybrid β-sheet. It contains both the N-terminal region and C-terminal region (but only up to residue 101). The outer strand of subdomain I is donated from the dimer-related subunit. The centre of the β-barrel is filled with conserved hydrophobic side chains. The high sequence homology (85%), the globular shape of the protein and the essentially identical fold of the polypeptide backbone of the target and probe models were probably important factors in the structure solution by molecular replacement.

Our experience indicates that with only the Cα atoms of one monomer (6% of the non-H atoms) the rotation and translation functions can clearly be solved (Figs. 2 and 3). Up until now all experiments involving so few atoms had always failed to give the correct translation parameters (*e.g.* Cygler & Anderson, 1988*b*). In our study, we also found out that just 3.9% of the atoms were sufficient to give the correct solution, that is 65 atoms out of approximately 1690. We also showed that the HHA structure could be determined by MR using a Cα search model with data in the range 15–10 to 5–3 Å.

We further showed that the HHA structure could be solved using a search model containing one and a half subdomains (*i.e.* approximately one half of a monomer) with only 21% of the protein atoms of the asymmetric unit. Experiments using a probe consisting of only a four-stranded antiparallel β-sheet (one subdomain) failed, possibly as a result of the pseudo threefold symmetry which relates subdomains I, II and III (Fig. 6).

We expect that the polypeptide fold described above to be a common feature of all monocot bulb lectins, which we plan to establish by determining the structures of representative lectins from the Alliaceae, Orchidaceae and Liliaceae families by the approach described above.

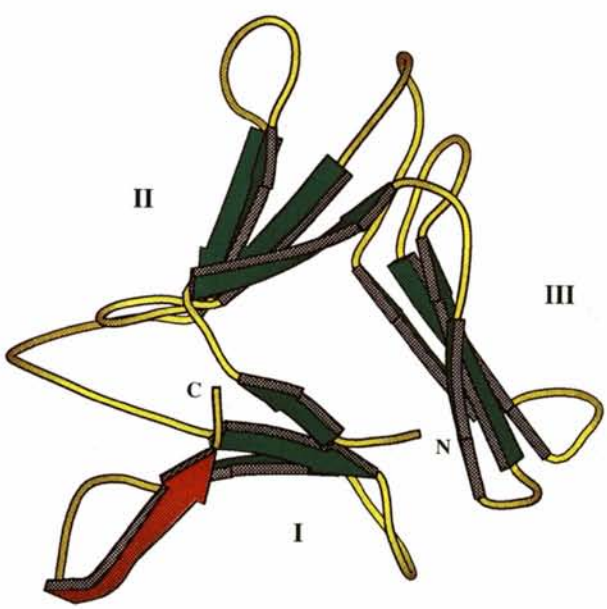

Fig. 6. The folding of the HHA monomer showing the pseudo threefold axis relating subdomains I, II and III. The β-strands are shown as arrows in green for strands 1 to 11 and the C-terminal strand, 12, is shown in red. Subdomain I is a hybrid β-sheet, strand 12 is donated from the dimer-related subunit. The figure was prepared using *MOLSCRIPT* (Kraulis, 1991).

This work was supported by a grant from the Mizutani Foundation for Glycoscience (to CDR). Synchrotron radiation facilities and support were kindly provided by CCLRC, Daresbury Laboratory. Liverpool John Moores University are thanked for general support. We are indebted to Professor C. S. Wright (Virginia Commonwealth University, USA) for supplying us with the Cα coordinates of GNA prior to publication. We thank Dr A. K. Allen (Charing Cross and Westminster Medical School) for his generous help and advice, and Professor E. H. Evans and Professor P. F. Lindley for their encouragement. Dr S. J. Lambert is thanked for his excellent technical assistance, as well as Dr E. Dodson for stimulating discussions whilst writing this paper.

References

Balzarini, J., Schols, D., Neyts, J., Van Damme, E. J. M., Peumans, W. & De Clerq, E. (1991). *Antimicrob. Agents Chemother.* **35**, 410–416.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **236**, 458–460.

Chantalat, L., Jones, N. D., Körber, F., Navaza, J. & Pavlovsky, A. G. (1995). *Protein Pept. Lett.* **2**, 333–340.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–183. New York: Gordon and Breach.

Cygler, M. & Anderson, W. F. (1988*a*). *Acta Cryst.* A**44**, 38–45.

Cygler, M. & Anderson, W. F. (1988*b*). *Acta Cryst.* A**44**, 300–308.
Hester, G., Kaku, H. Goldstein, I. J. & Schubert-Wright, C. (1995). *Nature Struct. Biol.* **2**, 472–479.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. & Vernoslova, E. (1995). *Acta Cryst.* A**51**, 445–449.
Oh, B.-H. (1995). *Acta Cryst.* D**51**, 140–144.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Schierbeek, A. J., Renestseder, R., Dijkstra, B. W. & Hol, W. G. J. (1985). In *Molecular Replacement, Proceedings of the Daresbury Study Weekend, 15–16 February 1985*, edited by P. A. Machin, pp. 16–21. Warrington: Daresbury Laboratory
Van Damme, E. J. M., Smeets, K. & Peumans, W. J. (1995). *Lectin: Biomedical Perspectives*, edited by A. Pusztai and S. Bardocz, pp. 59–80. Basingstoke: Taylor and Francis
Van Driessche, E. (1988). *Adv. Lectin Res.* **1**, 73–134.
Wood, S. D., Reynolds, C. D., Lambert, S. & McMichael, P. A. D. Allen, A. K. & Rizkallah, P. J. (1994). *Acta Cryst.* D**50**, 110–111.