Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structure of the native (unligated) mannose-specific bulb lectin from *Scilla campanulata* (bluebell) at 1.7 Å resolution

**Stephen D. Wood,[a] Lisa M. Wright,[a]† Colin D. Reynolds,[a]* Pierre J. Rizkallah,[b] Anthony K. Allen,[c] Willy J. Peumans[d] and Els J. M. Van Damme[d]**

[a]School of Biomolecular Sciences, Max Perutz Building, Liverpool John Moores University, Liverpool L3 3AF, England, [b]CCLRC Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, England, [c]Molecular Pathology Section, Division of Biomedical Sciences, Imperial College, School of Medicine, London SW7 2AZ, England, and [d]Laboratory for Phytopathology and Plant Protection, Willem de Croylaan 42, Catholic University of Leuven, 3001 Leuven, Belgium

† Present address: York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO1 5DD, England.

Correspondence e-mail: bmscreyn@livjm.ac.uk

Received 13 November 1998
Accepted 23 April 1999

**PDB Reference:** mannose-specific bulb lectin, 1b2p.

The X-ray crystal structure of native *Scilla campanulata* agglutinin, a mannose-specific lectin from bluebell bulbs and a member of the Liliaceae family, has been determined by molecular replacement and refined to an $R$ value of 0.186 at 1.7 Å resolution. The lectin crystallizes in space group $P2_12_12$ with unit-cell parameters $a = 70.42$, $b = 92.95$, $c = 46.64$ Å. The unit cell contains eight protein molecules of $M_r = 13143$ Da (119 amino-acid residues). The asymmetric unit comprises two chemically identical molecules, $A$ and $B$, related by a non-crystallographic twofold axis perpendicular to $c$. This dimer further associates by crystallographic twofold symmetry to form a tetramer. The fold of the polypeptide backbone closely resembles that found in the lectins from *Galanthus nivalis* (snowdrop) and *Hippeastrum* (amaryllis) and contains a threefold symmetric $\beta$-prism made up of three antiparallel four-stranded $\beta$-sheets. Each of the four-stranded $\beta$-sheets (I, II and III) possesses a potential saccharide-binding site containing conserved residues; however, site II has two mutations relative to sites I and III which may prevent ligation at this site. Our study provides the first accurate and detailed description of a native (unligated) structure from this superfamily of mannose-specific bulb lectins and will allow comparisons with a number of lectin–saccharide complexes which have already been determined or are currently under investigation.

## 1. Introduction

Lectins are a structurally diverse group of proteins of non-immune origin which exhibit antibody-like carbohydrate-binding specificity (Sharon & Lis, 1989*a*; Rini, 1995). These proteins are ubiquitous in nature and are found in all types of living organisms. Lectins have the capability to serve as recognition molecules within cells, between cells or between organisms by virtue of their binding specificity (Drickamer & Taylor, 1993) and thus find extensive applications in biological and medical research (Sharon & Lis, 1989*b*).

Plant lectins have been most widely studied. Owing to the important progress in the molecular biology and structural analysis of these proteins, plant lectins have been redefined recently on a functional basis as proteins possessing at least one non-catalytic domain which binds reversibly to a specific mono- or oligosaccharide (Peumans & Van Damme, 1995). The increasing availability of detailed sequence information has enabled the apparently highly heterogeneous group of plant lectins to be subdivided into a limited number of evolutionary and structurally related protein families (Van Damme *et al.*, 1999). Most lectins can be classified into four main families of homologous proteins: (*a*) legume lectins, (*b*) chitin-binding lectins containing hevein domains, (*c*) type 2

ribosome-inactivating proteins (RIPs) and (*d*) monocot mannose-binding lectins. In addition to these four large families, three small lectin families are distinguished, namely the jacalin-related lectins and amaranthins, which clearly exhibit new folds (Wright, 1997), and the so-called Cucurbitaceae phloem lectins, the structure of which has not yet been resolved.

The largest and best characterized family is that from the seeds of leguminous plants (Van Driessche, 1988; Sharon & Lis, 1990; Loris *et al.*, 1998). Among some of the best known lectins of this group are concanavalin A (Sumner, 1919), lentil lectin (Landsteiner & Raubitshek, 1907), favin (Tomita *et al.*, 1972), pea lectin (Trowbridge, 1974) and *Lathyrus ochrus* lectin (Rouge & Sousa-Cavada, 1984). Concanavalin A is the most notable of these and was the first phytohaemagglutinin to be crystallized (Sumner & Howell, 1936) and have its structure determined by X-ray crystallography (Edelman *et al.*, 1972; Hardman & Ainsworth, 1972). Subsequently, ConA has been refined at high resolution (Emmerich *et al.*, 1994; Parkin *et al.*, 1996) and several carbohydrate complexes have been reported (Derewenda *et al.*, 1989; Naismith *et al.*, 1994; Naismith & Field, 1996; Moothoo & Naismith, 1998). The ConA monomer contains a six- and a seven-stranded antiparallel $\beta$-sheet and this fold is common to all other members of the legume lectin family.

Chitin-binding lectins containing hevein domains (Raikhal *et al.*, 1993) are very different from the legume lectins and include wheatgerm agglutinin (WGA; Allen *et al.*, 1973), rice (Tsuda, 1979), rye and barley lectins (Peumans *et al.*, 1982). The X-ray crystal structures of two isoforms of WGA, WGA1 and WGA2, and several carbohydrate complexes have been reported (Wright, 1987, 1989, 1990; Wright & Jaeger, 1993). WGA is dimeric, with two identical 17 kDa monomers containing four isostructural domains with a high cysteine content and relatively little secondary structure.

Ribosome-inactivating proteins (RIPs) occur as two types (Stirpe & Barbieri, 1986). Type 1 proteins are single-chained and include trichosanthin, which is reported as being homologous to the ricin A chain (Zheng & Wang, 1986), $\alpha$-momorcharin (Ren *et al.*, 1994) and $\beta$-momorcharin (Xiong *et al.*, 1994). Type 2 proteins consist of an *A* chain and *B* chain, where the *A* chain is the toxin and the *B* chain (a lectin) is responsible for the binding of the whole protein molecule to the target-cell surface and also for helping the *A* chain to cross the cell membrane. The ricin *B* chain is composed of two domains: a linker domain and a second domain extending over three homologous subdomains related by a pseudo-threefold axis (Rutenber & Robertus, 1991).

Since 1986, when the first members of the monocot mannose-binding lectin superfamily were isolated, several proteins have been characterized from the families Amaryllidaceae, Alliaceae, Orchidaceae, Araceae and Liliaceae (Van Damme *et al.*, 1995) and shown to belong to a group of lectins with strict specificity for $\alpha$-D-mannose. These monocot lectins exhibit a strong affinity for $\alpha$(1,3)- and $\alpha$-(1,6)-linked mannosyl residues (Shibuya *et al.*, 1988; Kaku *et al.*, 1991), although the fine specificity for oligosaccharides may vary even within the same family (Kaku & Goldstein, 1992). Some of these monocot mannose-binding lectins display potent inhibitory activity on retroviruses including human immunodeficiency virus (HIV), recognizing either the core regions or the outer chains of N-linked oligosaccharides of the envelope glycoprotein gp120 and thus interfering with infectivity or virus-induced cell fusion (Feizi *et al.*, 1990; Balzarini *et al.*, 1991). Preliminary results (Balzarini, private communication) show that the Liliaceae lectin (SCA) from bluebell bulbs exhibits anti-retroviral activity, but that it is an order of magnitude lower than the inhibitory effects reported for the Amaryllidaceae and Orchidaceae lectins (Balzarini *et al.*, 1991). *Galanthus nivalis* agglutinin (GNA) from snowdrop bulbs complexed with methyl $\alpha$-D-mannose was the first member of the monocot lectins to have its crystal structure reported (Hester *et al.*, 1995) at 2.3 Å resolution. Subsequently, the structure of *Hippeastrum hybrid* agglutinin (HHA) from amaryllis bulbs in complex with $\alpha$-D-mannose was determined by molecular replacement (Wood, 1995; Chantalat *et al.*, 1996) at 2.3 Å resolution. Both GNA and HHA belong to the Amaryllidaceae family. In the current study, we report the structure solution and refinement at 1.7 Å resolution of native *Scilla campanulata* agglutinin (SCA), the first Liliaceae lectin structure to be determined by X-ray crystallography.

## 2. Experimental

### 2.1. Protein purification, sequence and crystallization

*Scilla campanulata* agglutinin (SCA) was prepared from bluebell bulbs by affinity chromatography according to the method previously described (Wright *et al.*, 1996). The cDNA sequence has been determined (Van Damme & Peumans, unpublished results) and the corresponding amino-acid sequence comprising 119 residues and with an $M_r$ of 13143 Da is shown in Fig. 1(*a*). Bipyramidal crystals of maximum dimensions 0.7 × 0.7 × 1.0 mm were grown by the hanging-drop method at pH 4.7, using 20 µl droplets of 10 mg ml$^{-1}$ protein in 600 m*M* phosphate-buffered saline equilibrated against 70% saturated ammonium sulfate as precipitant (Wright *et al.*, 1996).

### 2.2. Data collection and processing

Data to 1.7 Å resolution were collected using a MAR image plate at a wavelength of 0.80 Å on station 9.5 at the CCLRC Daresbury Laboratory. Data were collected at 277 K over 93° in steps of 1.0°. Diffraction data were processed using *MOSFLM* (Leslie, 1992) and, after scaling and averaging with *ROTAVATA* and *AGROVATA* (Collaborative Computational Project, Number 4, 1994), yielded a unique data set of 33837 reflections with a merging *R* factor of 0.039. The unit-cell dimensions were determined to be $a$ = 70.42, $b$ = 92.95, $c$ = 46.64 Å. Subsequent analysis of the measured data confirmed the crystals as belonging to the orthorhombic space group $P2_12_12$. The statistics of the data collection and processing are summarized in Table 1.

### 2.3. Molecular replacement

The structure of the native SCA crystals was solved with molecular replacement using the program *AMoRe* (Navaza, 1994). A self-rotation function based on data in the resolution range 8–4 Å indicated twofold symmetry perpendicular to the crystallographic twofold axis (Wright *et al.*, 1996). Biochemical studies have also shown that the SCA molecule associates into tetramers. The refined monomer coordinates obtained for the HHA lectin structure (Chantalat *et al.*, 1996) were used as a

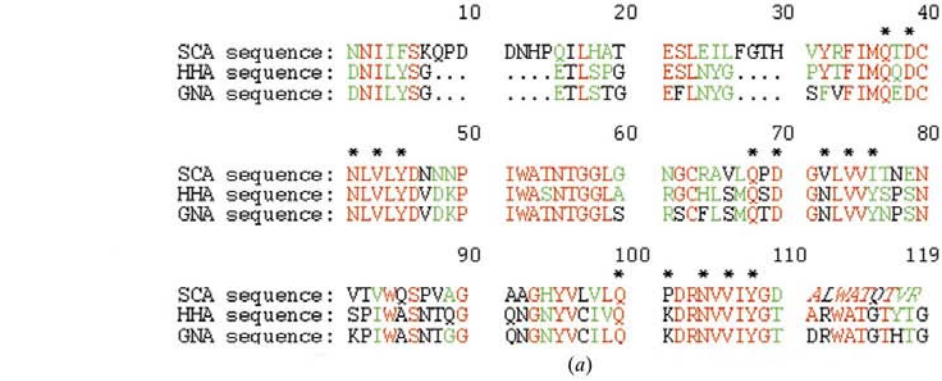


**Figure 1**

(*a*) Sequence alignment between bluebell lectin (SCA), amaryllis lectin (HHA) and snowdrop lectin (GNA). Identical residues are shown in red and similar residues are shown in green; others are in black. The percentage identity between HHA and SCA is 54% and that between GNA and SCA is 52%. The similarity of HHA and SCA is 81% and that of GNA and SCA is 75%. The position of the extra residues for SCA are shown as dots in the other sequences and the C-terminal exchange residues are shown in italics. The residues involved in carbohydrate binding at each of the sub-domains are indicated with an asterisk. (*b*) Schematic drawing depicting sub-domains I, II and III of monomer *A* and the C-terminal exchange of monomer *B*. (*c*) Stereo plot of the monomer. (*d*) Ribbon diagram of the monomer indicating sub-domains I, II and III.

**Table 1**
Data collection and processing statistics.

| | |
|---|---|
| Data collected at | CCLRC Daresbury Laboratory |
| Wavelength (Å) | 0.8 |
| Temperature (K) | 277 |
| Range collected (°) | 93 |
| Oscillation range (°) | 1.0 |
| Number of crystals used | 1 |
| Space group | $P2_12_12$ (No. 18) |
| Number molecules in asymmetric unit | 2 |
| $a$ (Å) | 70.42 |
| $b$ (Å) | 92.95 |
| $c$ (Å) | 46.64 |
| Volume (Å$^3$) | 305283.94 |
| $M_r$ (Da) | 13143 |
| $V_M$ (Å$^3$ Da$^{-1}$) | 2.90 |
| Solvent content (%) | 57.6 |
| Number of reflections measured | 161054 |
| Number of independent merged reflections | 33837 |
| Overall $R_{\rm merge}$ | 0.039 |
| Completeness (%) | 98.5 |
| Multiplicity | 3.8 |
| Resolution limit (Å) | 1.7 |
| Overall $I/\sigma(I)$ (in last shell) | 8.3 (3.2) |

search model in cross-rotation-function searches. The percentage identity and similarity of HHA (109 residues) compared with SCA (119 residues) are 54 and 81%, respectively (Fig. 1*a*). Two clear solutions to both the rotation and translation functions were obtained and, after rigid-body refinement the $R$ factor was 45% and the correlation coefficient was 64%. The packing function indicated very few close contacts.

### 2.4. Refinement

After replacing the various residues which differed between the two structures and including the two insertions and the one deletion (Fig. 1*a*), the complete dimer was refined by simulated annealing with *X-PLOR* ($T_{\rm init}$ = 3000 K, timestep = 0.001 s, $T_{\rm final}$ = 300 K; Brünger *et al.*, 1987). This reduced the conventional $R$ factor to 30% ($R_{\rm free}$ = 37%; Brünger, 1992). Further improvements in the agreement between the model and diffraction data were obtained using least-squares refinement with *X-PLOR* (Brünger, 1990). The model was periodically rebuilt on an interactive graphics system and solvent molecules were gradually incorporated. For the final round of refinement, *REFMAC* (Murshudov *et al.*, 1997) was used. The various stages of refinement are summarized in Table 2. In each stage, 20 cycles of least-squares refinement were carried out and the progress of refinement was assessed using $R_{\rm free}$ (Brünger, 1992). Electron-density maps were calculated, using both $F_o - F_c$ and $2F_o - F_c$ as coefficients, and the model was examined on a Silicon Graphics workstation using the program *O* (Jones & Kjeldgaard, 1993).

## 3. Results and discussion

### 3.1. Least-squares refinement

The final model contains 1852 protein atoms and 246 solvent molecules with a final $R$ factor of 18.6% (and

**Table 2**
Various stages of refinement.

| Stage | Number of cycles | Resolution (Å) | Final $R$ (%) | $R_{\rm free}$ (%) | Comments |
|---|---|---|---|---|---|
| 0 | 0 | 10–1.7 | 45 | — | Starting model |
| 1 | 20 | 10–1.7 | 30.2 | 37.3 | Simulated annealing |
| 2 | 20 | 10–1.7 | 28.6 | 34.9 | Replaced differing residues and included inserts |
| 3 | 20 | 10–1.7 | 27.4 | 32.8 | Least-squares refinement with $B_{\rm iso}$ = 20 Å$^3$ |
| 4 | 20 | 10–1.7 | 25.5 | 29.6 | Individual $B$ factors refined |
| 5 | 20 | 10–1.7 | 24.3 | 28.4 | 50 solvent sites added |
| 6 | 20 | 10–1.7 | 23.5 | 27.5 | 50 solvent sites added |
| 7 | 20 | 10–1.7 | 22.6 | 26.5 | 50 solvent sites added |
| 8 | 20 | 10–1.7 | 21.7 | 25.4 | 50 solvent sites added |
| 9 | 20 | 10–1.7 | 20.8 | 24.6 | 46 solvent sites added |
| 10 | 20 | 20–1.7 | 19.5 | 23.3 | Refinement and side-chain adjustment |
| 11 | 20 | 20–1.7 | 18.8 | 23.1 | Structure has 246 solvent sites and 1852 protein atoms, and was refined using *X-PLOR* |
| 12 | 20 | 20–1.7 | 18.6 | 20.8 | Final model was refined using *REFMAC* (Murshudov *et al.*, 1997) |

$R_{\rm free}$ = 20.8%) for the resolution range 20.0–1.7 Å. The model was checked for stereochemical correctness using the program *PROCHECK* (Laskowski *et al.*, 1993) and had an average standard deviation from ideal stereochemistry of 0.011 Å for main-chain bond lengths and 1.91° for main-chain bond angles. Ramachandran plots (Ramachandran & Sassiekharan, 1968) of the main-chain torsion angles ($\psi$,$\varphi$) for each subunit show that no residues lie in disallowed regions (Fig. 2). Residues Asn47 from both subunits (*A* and *B*) lie in the generously allowed regions of the Ramachandran plot. These slightly non-standard $\varphi$,$\psi$ values arise as a result of dimer–

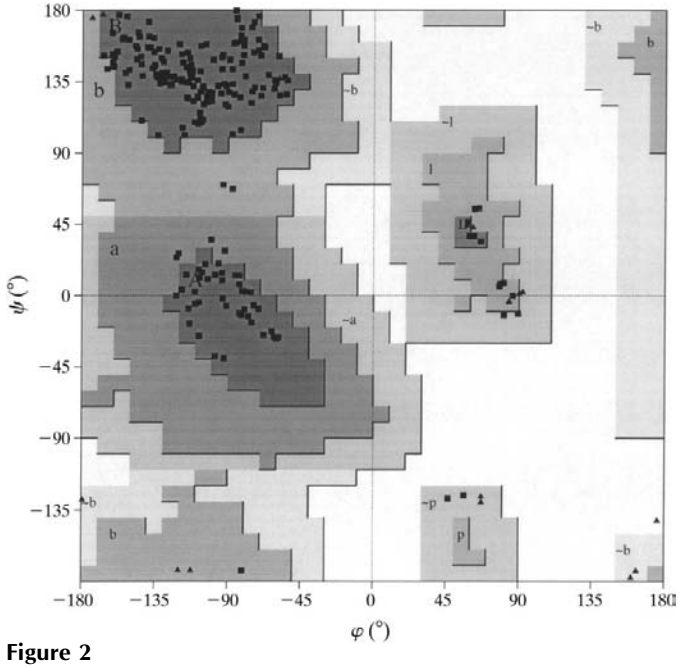


**Figure 2**
Ramachandran plots of the *A* and *B* chains showing 90.2 and 91.2% of residues in the most favoured regions for *A* and *B*, respectively. No residues reside in disallowed regions for either chain.

**Table 3**
Refinement statistics.

| | |
|---|---|
| Resolution range (Å) | 20.0–1.7 |
| Number of reflections used (no cutoff applied) | 33837 |
| Crystallographic $R$ factor† for all data [$F_{obs} > 2\sigma(F_{obs})$] (%) | 18.6 |
| $R_{free}$‡ for all data [$F_{obs} > 2\sigma(F_{obs})$] (%) | 20.8 |
| Number of protein atoms | 1852 |
| Number of water molecules | 246 |
| Average $B$ values (Å$^2$) | |
| Subunit $A$ main-chain atoms | 18.1 |
| Subunit $A$ side-chain atoms | 23.1 |
| Subunit $B$ main-chain atoms | 17.8 |
| Subunit $B$ side-chain atoms | 23.0 |
| Average temperature factor for solvent atoms (Å$^2$) | 39.8 |
| Estimated coordinate error§ (Å) | 0.221/0.225 |
| *X-PLOR* stereochemical parameters¶ | |
| R.m.s deviations from ideal | |
| Bond lengths (Å) | 0.008 |
| Bond angles (°) | 1.709 |
| Dihedral angles (°) | 29.278 |
| Improper angles (°) | 0.769 |

† Crystallographic $R$ factor = $\sum(||k|F_{obs}| - |F_{calc}||)/\sum k|F_{obs}|$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure-factor amplitudes, respectively. ‡ $R_{free}$ is the crystallographic $R$ factor calculated for a subset of randomly selected reflections (5%) not used in the phasing process (Brünger, 1992). § Estimated according to Luzzati (1952) and Read (1986). ¶ The parameters described by Engh & Huber (1991) were used.

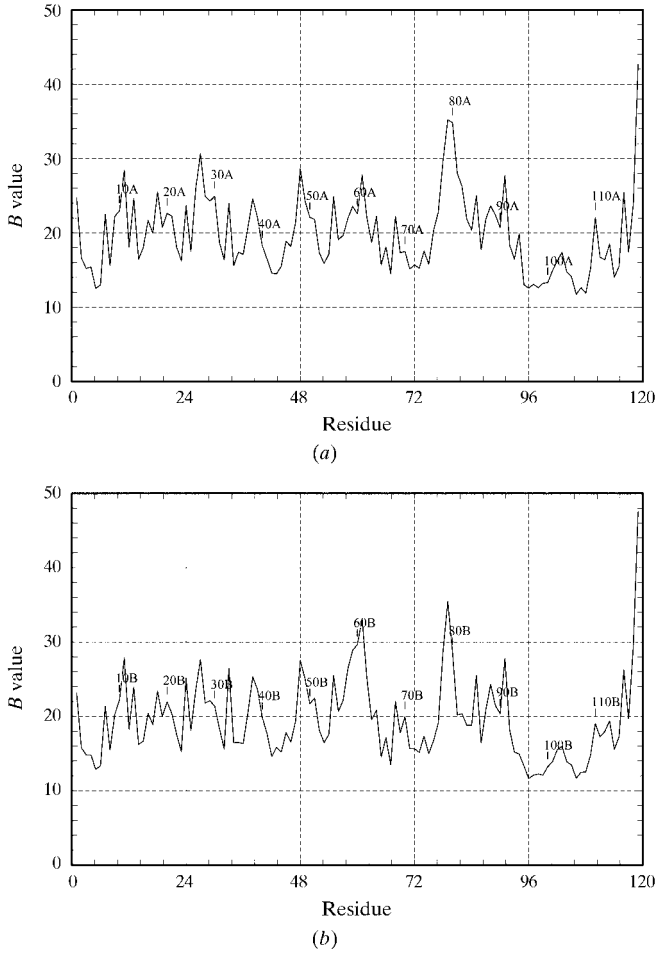


**Figure 3**
The distribution of average main-chain and side-chain temperature factors for (*a*) the *A* chain and (*b*) the *B* chain.

**Table 4**
Subunit structure comparison of native SCA with GNA, HHA and complex SCA, calculated using *LSQKAB* (Kabsch, 1976).

Subunit superpositions.

| | R.m.s Δ (Å) |
|---|---|
| SCA subunit *A* *versus* SCA subunit *B* | 0.24 |
| SCA subunit *A* *versus* GNA subunit *A* | 0.78 |
| SCA subunit *B* *versus* GNA subunit *A* | 0.80 |
| SCA subunit *A* *versus* HHA subunit *A* | 0.74 |
| SCA subunit *B* *versus* HHA subunit *A* | 0.77 |

Superposition of native SCA binding sites.

| | R.m.s Δ (Å) |
|---|---|
| SCA subunit *A* site I *versus* SCA subunit *B* site I | 0.09 |
| SCA subunit *A* site II *versus* SCA subunit *B* site II | 0.39 |
| SCA subunit *A* site III *versus* SCA subunit *B* site III | 0.58 |

Superposition of native SCA binding sites with complex SCA binding sites.†

| | R.m.s Δ (Å) |
|---|---|
| SCA subunit *A* site I *versus* complex SCA subunit *A* site I | 0.08 |
| SCA subunit *A* site II *versus* complex SCA subunit *A* site II | 0.09 |
| SCA subunit *A* site III *versus* complex SCA subunit *A* site III | 0.08 |
| SCA subunit *B* site I *versus* complex SCA subunit *B* site I | 0.07 |
| SCA subunit *B* site II *versus* complex SCA subunit *B* site II | 0.08 |
| SCA subunit *B* site III *versus* complex SCA subunit *B* site III | 0.08 |

Superposition of native SCA binding sites with complex GNA binding sites.

| | R.m.s Δ (Å) |
|---|---|
| SCA subunit *A* site I *versus* complex GNA subunit *A* site I | 0.59 |
| SCA subunit *A* site II *versus* complex GNA subunit *A* site II | 0.77 |
| SCA subunit *A* site III *versus* complex GNA subunit *A* site III | 0.45 |
| SCA subunit *B* site I *versus* complex GNA subunit *B* site I | 0.59 |
| SCA subunit *B* site II *versus* complex GNA subunit *B* site II | 0.86 |
| SCA subunit *B* site III *versus* complex GNA subunit *B* site III | 0.46 |

Superposition of native SCA binding sites with complex HHA binding sites.

| | R.m.s Δ (Å) |
|---|---|
| SCA subunit *A* site I *versus* complex HHA subunit *A* site I | 0.66 |
| SCA subunit *A* site II *versus* complex HHA subunit *A* site II | 0.82 |
| SCA subunit *A* site III *versus* complex HHA subunit *A* site III | 0.62 |
| SCA subunit *B* site I *versus* complex HHA subunit *B* site I | 0.67 |
| SCA subunit *B* site II *versus* complex HHA subunit *B* site II | 0.90 |
| SCA subunit *B* site III *versus* complex HHA subunit *B* site III | 0.40 |

† Wright, Wood, Reynolds & Rizkallah, unpublished work; Wright (1998).

dimer interactions within the SCA tetramer. The distribution of the average main-chain and side-chain temperature factors is given in Fig. 3. A statistical summary of the final model is presented in Table 3.

### 3.2. Structure of the monomer

Ribbon diagrams, drawn with *MOLSCRIPT* (Kraulis, 1991), of the SCA monomer (subunit *A*) and tetramer

**Table 5**
List of contacts involved in the SCA tetramer interfaces (<4 Å).

Calculated using *X-PLOR*.

| Interface | Hydrogen-bonding contacts (Å) | | | Total number of hydrogen contacts | Close contacts (Å) | | | Total number of close contacts |
|---|---|---|---|---|---|---|---|---|
| | Min | Max | Avg | | Min | Max | Avg | |
| *AB* | 2.71 | 2.94 | 2.83 | 13 | 3.04 | 3.57 | 3.29 | 29 |
| *AB′* | — | — | — | — | 3.45 | 3.92 | 3.70 | 3 |
| *A′B* | — | — | — | — | 3.37 | 4.00 | 3.72 | 3 |

(showing the two independent monomers *A* and *B* in the asymmetric unit related by non-crystallographic twofold symmetry and the two dimers associated through the crystallographic twofold axis) are illustrated in Figs. 1(*d*) and 6, respectively.

Previous work on the monocot mannose-binding lectins has established the structures of snowdrop (*Galanthus nivalis*) lectin (Hester *et al.*, 1995; Hester & Wright, 1996) and amaryllis (*Hippeastrum*) lectin (Wood, 1995; Chantalat *et al.*, 1996). In both of these lectins, and also in SCA, the polypeptide fold consists of three $\beta$-sheet subdomains (I, II, and III) and each subdomain forms a flat four-stranded antiparallel $\beta$-sheet. The $\beta$-sheets are interrelated by pseudo-threefold symmetry to form a 12-stranded $\beta$-barrel-like arrangement (Figs. 1*c* and 6), referred to as a $\beta$-prism. The quasi-threefold axis coincides with the barrel axis. Subdomain I is a hybrid $\beta$-sheet which contains both the N-terminal and C-terminal regions. The outer strand of subdomain I is donated from the dimer-related subunit (Figs. 1*b* and 6). The centre of the $\beta$-barrel is highly hydrophobic and is filled with conserved non-polar side chains (Figs. 1*a* and 4). The consensus carbohydrate-recognition sequence Q*x*D*x*N*x*V*x*Y (found, for example, in all of the Amaryllidaceae lectins; Wright & Hester, 1996) occurs in the putative carbohydrate-recognition domains CRDI and CRDIII of SCA. CRDII, however, contains two mutations (N and Y replaced by V and I, respectively) which probably prevents binding of $\alpha$-D-mannose. The two independent subunits *A* and *B* of the native SCA were found to possess very similar folds to each other and to those found in GNA and HHA (Fig. 5). A least-squares superposition of main-chain atoms of the two SCA subunits gives an r.m.s. difference of 0.24 Å (Table 4). Comparisons of SCA with the GNA and HHA subunits yielded r.m.s. discrepancies in the ranges 0.78–0.80 and 0.74–0.77 Å, respectively (Table 4). An example of these comparisons is shown in Fig. 5.

Each SCA subunit contains 119 amino-acid residues (Fig. 1*a*) and is organized into three antiparallel $\beta$-sheets approximately 35 amino-acid residues in length, as illustrated in Figs. 1(*b*), 1(*c*) and 6. Pairwise comparisons using main-chain atoms only and excluding the residues in the loop regions revealed essentially identical topology for the three $\beta$-sheets in each subunit, as listed in Table 4.

Superpositions of the conserved residues in the two subunits of SCA for the putative mannose-binding sites 1, 2 and 3, located in subdomains I, II and III, respectively, compare well, with r.m.s. discrepancies in the range 0.09–0.58 Å. There is also close agreement between these conserved saccharide-binding residues for the native SCA structure and the positions found

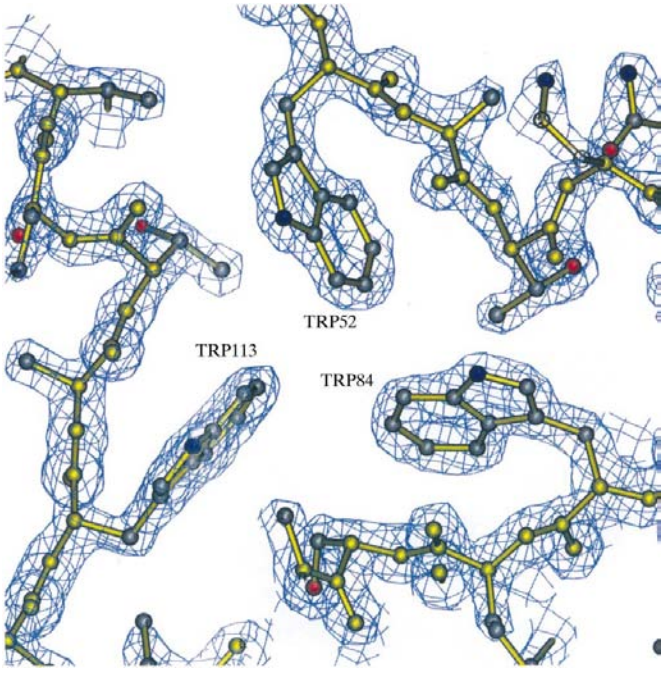


**Figure 4**
Part of the hydrophobic centre of the $\beta$-barrel, showing some of the conserved non-polar side chains and their associated electron density.

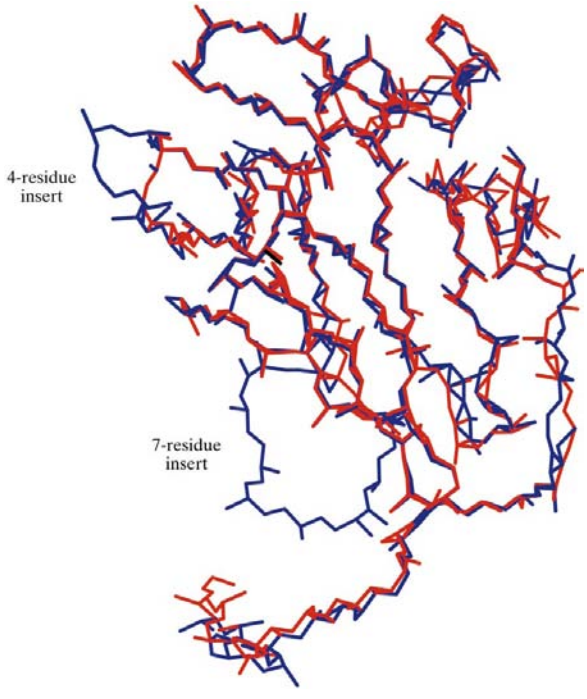


**Figure 5**
Comparison of the *A*-chain backbone for SCA (shown in blue) with the *A*-chain backbone of GNA (shown in red).

in GNA, HHA and SCA complexes (Wright, Wood, Reynolds & Rizkallah, unpublished work), as shown in Table 4. These results confirm that formation of mannose complexes gives rise to relatively small conformational changes in the residues involved in binding.

### 3.3. Quaternary structure, subunit interactions and crystal packing

Biochemical evidence has demonstrated that the functional SCA molecule is organized as a tetramer of four identical subunits (Wood *et al.*, 1996). In the crystal, the crystallographic twofold axis parallel to the *c* axis relates the *A*–*B* dimer of the asymmetric unit to generate an SCA tetramer (*A*–*B*, *A*′–*B*′) with approximate 222 point symmetry (Fig. 6). The SCA tetramer contains two different types of interfaces between the subunits. Analysis of these interfaces in terms of the non-covalent interactions (protein–protein hydrogen bonds, protein–water hydrogen bonds and van der Waals contacts) and the buried surface area show that the *A*–*B* pair (identical to *A*′–*B*′) form one type of dimer and the *A*–*B*′ pair (identical to *A*′–*B*) form a quite different dimer. Table 5 lists the contacts involved in tetramer formation. The *A*–*B* dimer association involves C-terminal strand exchange (Fig. 6) and as a result gives rise to extensive contacts and a buried surface area of much greater magnitude (approximately 5:1) than for the *A*–*B*′ dimer (Table 6). The packing of the SCA dimers in the

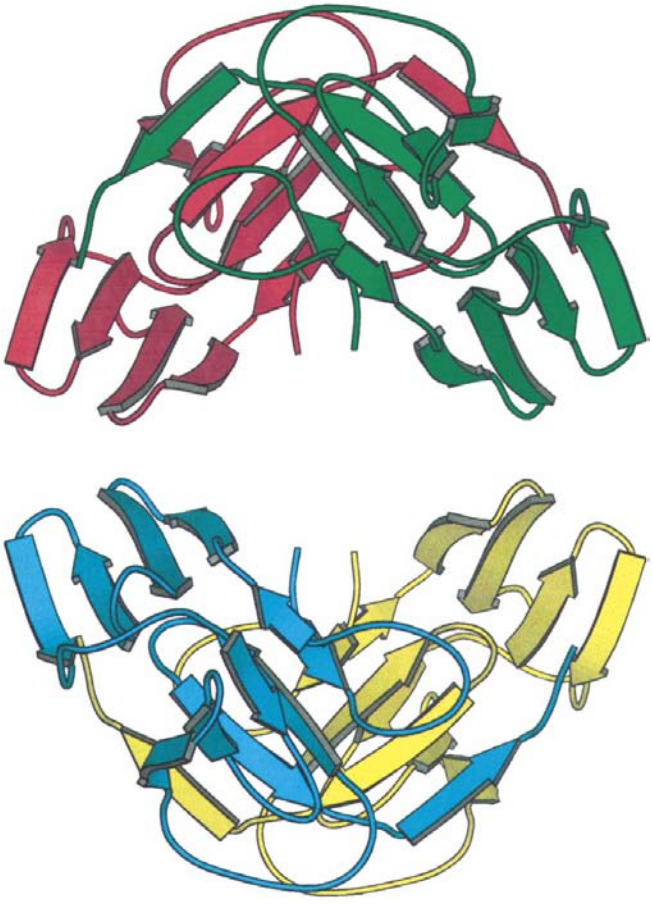

**Figure 6**
Tetrameric arrangement of SCA molecule. The monomers *A* and *B* are coloured red and green, respectively; the monomers *A*′ and *B*′ of the twofold-related dimer are coloured yellow and blue, respectively.

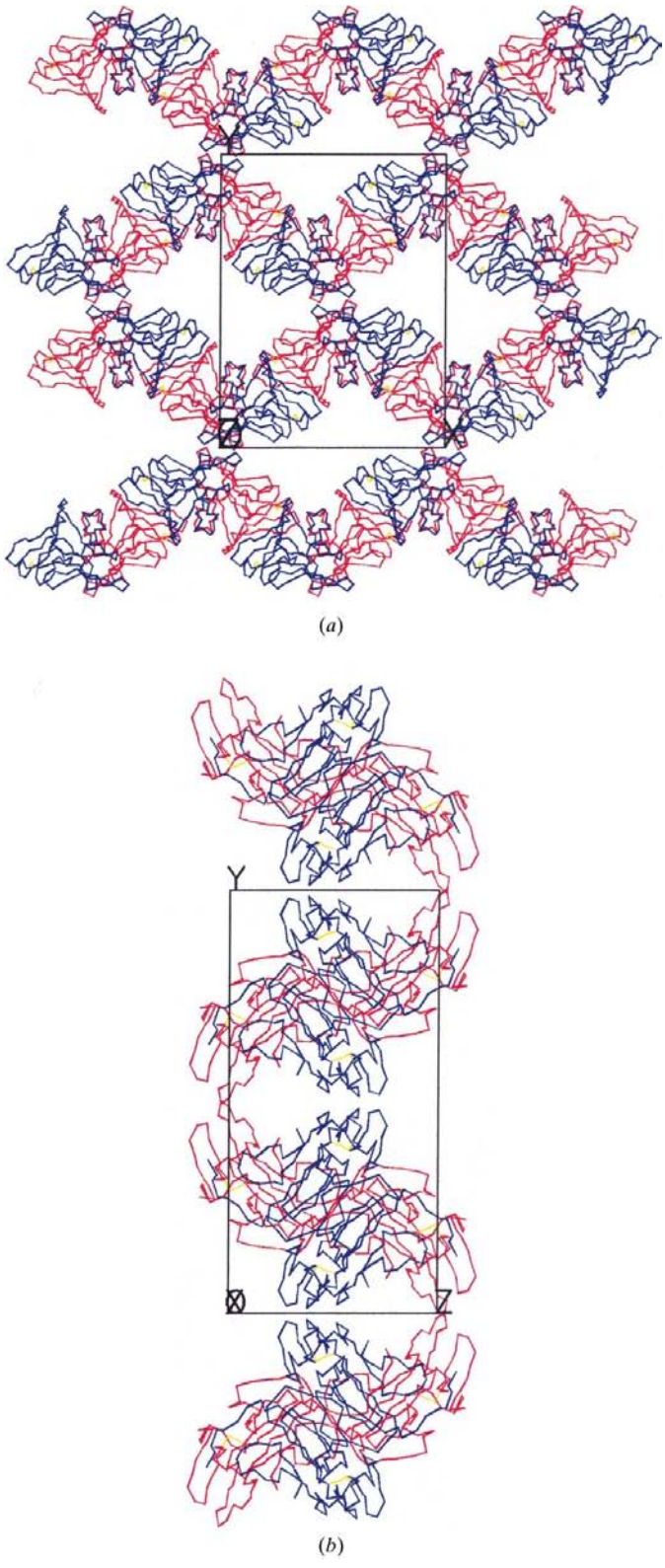


**Figure 7**
Packing diagram of SCA molecules in the orthorhombic unit cell viewed (*a*) down the *c* axis and (*b*) down the *a* axis.

**Table 6**
Analysis of SCA subunit interfaces and comparison with GNA interfaces.

| SCA interface | Buried surface area (Å$^2$)† | GNA interface | Buried surface area (Å$^2$)‡ | Difference (Å$^2$) |
|---|---|---|---|---|
| *AB* | 1854 | AD | 1728 | +126 |
| *A′B′* | 1854 | BC | 1748 | +106 |
| *AB′* | 369 | AB | 478 | −109 |
| *A′B* | 369 | CD | 475 | −106 |

† The buried surface areas were derived from the solvent-accessible surface values calculated with the program *X-PLOR* (Brünger *et al.*, 1987), version 3.851.f, using a probe radius of 1.6 Å. ‡ Hester & Wright (1996).

unit cell of the crystal, viewed down the *c* axis, is shown in Fig. 7(*a*). The SCA molecules in the crystal lattice make contact with each other *via* the weaker of the subunit–subunit interfaces (*A* with *B′* and *A′* with *B*). There are few direct protein–protein interactions, most contacts being mediated by solvent molecules.

## 4. Conclusions

Our studies of the native (unligated) structure of *S. campanulata* agglutinin have shown that the overall fold of the monomer is essentially similar to that found in the snowdrop and amaryllis lectin complexes recently determined by X-ray crystallography (Hester *et al.*, 1995; Hester & Wright, 1996; Wright & Hester, 1996; Chantalat *et al.*, 1996). Two of the putative carbohydrate-binding domains (I and III) in the SCA monomer contain key residues which are identical to those found in the three saccharide-binding regions of snowdrop and amaryllis lectins (Wright & Hester, 1996) and have similar geometry. Interestingly, at site II in SCA, two out of the five residues normally involved in sugar binding have been mutated, which may well prevent mannose binding at this site. We have recently crystallized complexes of SCA with a monosaccharide (Wright *et al.*, 1997), a disaccharide (Wright, Wood *et al.*, 1998) and a trisaccharide (Wright, Rizkallah *et al.*, 1998) and analyses of the carbohydrate-binding modes within these complexes is in progress. The coordinates and structure factors of the SCA structure have been deposited in the Protein Data Bank.

We thank The Leverhulme Trust (F/754/A) and the Mizutani Foundation for Glycoscience for financial support for this research. Synchrotron radiation beam time and facilities were kindly provided by the CCLRC Daresbury Laboratory and general support by Liverpool John Moores University. Professor E. H. Evans and M. J. Donovan are thanked for their help and encouragement. WJP is a Research Director and EJMVD is a Postdoctoral Fellow of the Fund for Scientific Research, Flanders. The authors are grateful to the referees for their helpful comments on the manuscript.

## References

Allen, A. K., Neuberger, A. & Sharon, N. (1973). *Biochem. J.* **131**, 155–162.
Balzarini, J., Schols, D., Neyts, J., Van Damme, E. J. M., Peumans, W. & DeClercq, E. (1991). *Antimicrob. Agents Chemother.* **35**, 410–416.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **236**, 458–460.
Chantalat, L., Wood, S. D., Rizkallah, P. J. & Reynolds, C. D. (1996). *Acta Cryst.* D**52**, 1146–1152.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z., Wan, T. & Campbell, J. (1989). *EMBO J.* **8**, 2189–2193.
Drickamer, K. & Taylor, M. E. (1993). *Annu. Rev. Cell Biol.* **9**, 237–264.
Edelman, G. M., Cunningham, B. A., Reeke, G. N., Becker, J. W., Waxdal, M. J. & Wang, J. L. (1972). *Proc. Natl Acad. Sci. USA*, **69**(9), 2580–2584.
Emmerich, C., Helliwell, J. R., Redshaw, M., Naismith, J. H., Harrop, S. J., Raftery, J., Kalb (Gilboa), A. J., Yariv, J., Dauter, Z. & Wilson, K. S. (1994). *Acta Cryst.* D**50**, 749–756.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Feizi, T., Larkin, M., Childs, R. A., Matthews, T., Thiels, S., Mizuochi, T. & Lawson, A. M. (1990). *AIDS Res. Hum. Retrovir.* **6**, 99–100.
Hardman, K. D. & Ainsworth, C. F. (1972). *Biochemistry*, **11**, 4910–4919.
Hester, G., Kaku, H., Goldstein, I. J. & Wright, C. S. (1995). *Nature Struct. Biol.* **2**, 472–479.
Hester, G. & Wright, C. S. (1996). *J. Mol. Biol.* **262**, 516–531.
Jones, T. A. & Kjeldgaard, M. (1993). *O Version* 5.9, *The Manual.* Uppsala University, Uppsala, Sweden.
Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.
Kaku, H. & Goldstein, I. J. (1992). *Carbohydr. Res.* **229**(2), 337–346.
Kaku, H., Goldstein, I. J. & Oscarson, S. (1991). *Carbohydr. Res.* **213**, 109–116.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Landsteiner, K. & Raubitshek, H. (1907). *Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg.* **45**, 660–667.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newslett. Protein Crystallogr.* **26**. Warrington: Daresbury Laboratory.
Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). *Biochim. Biophys. Acta*, **1383**, 9–36.
Luzzati, P. V. (1952). *Acta Cryst.* **5**, 802–810.
Moothoo, D. N. & Naismith, J. H. (1998). *Glycobiology*, **8**, 173–181.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Raftery, J., Kalb (Gilboa), A. J. & Yariv, J. (1994). *Acta Cryst.* D**50**, 847–858.
Naismith, J. H. & Field, R. A. (1996). *J. Biol. Chem.* **271**, 972–976.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–166.
Parkin, S., Rupp, B. & Hope, M. (1996). *Acta Cryst.* D**52**, 1161–1168.
Peumans, W. J., Stinissen, H. M. & Carlier, A. R. (1982). *Planta*, **154**, 568–572.
Peumans, W. J. & Van Damme, E. J. M. (1995). *Plant Physiol.* **109**, 347–352.
Raikhal, N. V., Lee, H. I. & Broekaert, W. F. (1993). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 591–615.
Ramachandran, G. N. & Sassiekharan, V. (1968). *Adv. Protein Chem.* **28**, 283–437.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Ren, J. S., Wang, Y. P., Dong, Y. C. & Stuart, D. I. (1994). *Structure*, **2**, 7–16.
Rini, J. M. (1995). *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.
Rouge, P. & Sousa-Cavada, B. (1984). *Plant Sci.* **37**, 21–27.
Rutenber, E. & Robertus, J. (1991). *Proteins*, **10**, 260–269.
Sharon, N. & Lis, H. (1989*a*). *Lectins.* London: Chapman & Hall.

Sharon, N. & Lis, H. (1989*b*). *Science*, **246**, 227–234.
Sharon, N. & Lis, H. (1990). *FASEB J.* **4**, 3198–3208.
Shibuya, N., Goldstein, I. J., Van Damme, E. J. M. & Peumans, W. J. (1988). *J. Biol. Chem.* **263**, 728–734.
Stirpe, F. & Barbieri, L. (1986). *FEBS Lett.* **195**, 1–8.
Sumner, J. B. (1919). *J. Biol. Chem.* **37**, 137–142.
Sumner, J. B. & Howell, S. F. (1936). *J. Bacteriol.* **32**, 227–237.
Tomita, M., Kurokawa, T., Onozaki, K., Osawa, T., Sakurai, Y. & Ukita, T. (1972). *Int. J. Cancer*, **10**, 602–606.
Trowbridge, I. S. (1974). *J. Biol. Chem.* **249**, 6004–6012.
Tsuda, M. (1979). *J. Biochem.* **86**, 1451–1461.
Van Damme, E. J. M., Peumans, W. J., Barre, A. & Rougé, P. (1999). In the press.
Van Damme, E. J. M., Smeets, K. & Peumans, W. J. (1995). *Lectins: Biomedical Perspectives*, edited by A. Pusztai & S. Bardocz, pp. 59–80. London: Taylor & Francis.
Van Driessche, E. (1988). In *Advances in Lectin Research*, edited by H. Franz. Berlin: VEB Verlay Volk & Gesundheit.
Wood, S. D. (1995). PhD thesis, Liverpool John Moores University, England.
Wood, S. D., Allen, A. K., Wright, L. M. & Reynolds, C. D. (1996). In *Lectins: Biology, Biochemistry and Clinical Biochemistry*, Vol. 11, edited by E. Van Driessche, S. Beeckmans & T. C. Bøg-Hanson, pp. 86–90. Hellerup, Denmark: Textop.
Wright, C. S. (1987). *J. Mol. Biol.* **194**(3), 501–529.
Wright, C. S. (1989). *J. Mol. Biol.* **209**, 475–487.
Wright. C. S. (1990). *J. Mol. Biol.* **215**, 635–651.
Wright, C. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 631–636.
Wright, C. S. & Hester, G. (1996). *Structure*, **4**, 1339–1352.
Wright, C. S. & Jaeger, J. (1993). *J. Mol. Biol.* **232**, 620–638.
Wright, L. M. (1998). PhD thesis, Liverpool John Moores University, England.
Wright, L. M., Rizkallah, P. J., Wood, S. D. & Reynolds, C. D. (1998). *Acta Cryst.* D**54**, 665–667.
Wright, L. M., Wood, S. D., Reynolds, C. D., Rizkallah, P. J. & Allen, A. K. (1997). *Protein Pept. Lett.* **4**, 343–348.
Wright, L. M., Wood, S. D., Reynolds, C. D., Rizkallah, P. J. & Allen, A. K. (1998). *Acta Cryst.* D**54**, 90–92.
Wright, L. M., Wood, S. D., Reynolds, C. D., Rizkallah, P. J., Peumans, W. J., Van Damme, E. J. M. & Allen, A. K. (1996). *Acta Cryst.* D**52**, 1021–1023.
Xiong, J. P., Xia, Z. X., Zhang, L., Ye, G. J., Jin, S. W. & Wang, Y. (1994). *J. Mol. Biol.* **238**, 284–285.
Zheng, X. J. & Wang, J. H. (1986). *Nature (London)*, **321**, 477–478.