Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Concanavalin A in a dimeric crystal form: revisiting structural accuracy and molecular flexibility

Katherine A. Kantardjieff,[a] Peter Höchtl,[b] Brent W. Segelke,[b] Fu-Ming Tao[a] and Bernhard Rupp[b]*

[a]Department of Chemistry and Biochemistry and W. M. Keck Foundation Center for Molecular Structure, California State University Fullerton, Fullerton, CA 92834, USA, and [b]Biology and Biotechnology Research Program, L-448, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

Correspondence e-mail: br@llnl.gov

Received 26 September 2001
Accepted 15 November 2001

PDB Reference: dimeric concanavalin A, 1gkb, r1gkbsf.

A structure of native concanavalin A (ConA), a hardy perennial of structural biology, has been determined in a dimeric crystal form at a resolution of 1.56 Å (space group $C222_1$; unit-cell parameters $a = 118.70$, $b = 101.38$, $c = 111.97$ Å; two molecules in the asymmetric unit). The structure has been refined to an R_{free} of 0.206 (R = 0.178) after iterative model building and phase-bias removal using *Shake&wARP*. Correspondence between calculated water–tyrosine interactions and experimentally observed structures near the saccharide-binding site suggests that the observed interactions between Tyr12 and water in various crystal forms are to be expected and are not unique to the presence of an active site. The present structure differs from previously reported atomic resolution structures of ConA in several regions and extends insight into the conformational flexibility of this molecule. Furthermore, this third, low-temperature, structure of ConA in a different crystal form, independently refined using powerful model-bias removal techniques, affords the opportunity to revisit assessment of accuracy and precision in high- or atomic resolution protein structures. It is illustrated that several precise structures of the same molecule can differ substantially in local detail and users of crystallographic models are reminded to consider the potential impact when interpreting structures. Suggestions on how to effectively represent ensembles of crystallographic models of a given molecule are provided.

1. Introduction

Concanavalin A (ConA) is a saccharide-binding protein from the jack bean (*Canavalia ensiformis*) which was first isolated in crystalline form by Sumner (1919). It is a member of a larger family of plant lectins (Sharon, 1993). Although lectins were initially discovered in plants, they are widely distributed in nature (Sharon & Lis, 1989). The ability of plant lectins to bind to mammalian cells and discriminate between different carbohydrate structures has led to a substantial body of research (Lis & Sharon, 1973, 1986; Sharon & Lis, 1975; Bittiger & Schnebli, 1976; Goldstein & Poretz, 1986). Structurally homologous animal lectins are believed to be important for cell adhesion and cross-linking by viruses (Weis *et al.*, 1998; Nagar *et al.*, 1998; White *et al.*, 2000). Despite their abundance, the true physiological functions of plant lectins such as ConA are not yet clearly understood (Chrispeels & Raikhel, 1991; Peumans & Van Damme, 1995).

Native ConA has two metal-binding sites and one saccharide-binding site near the distal end of the molecule (Hardman & Ainsworth, 1976). The first metal site, S1, binds transition metals Ni, Co, Zn, Mn and Cd, while the second metal site, S2, binds only Ca and Cd. Native ConA contains a

© 2002 International Union of Crystallography
Printed in Denmark – all rights reserved

mixture of transition metals at site S1, while site S2 is occupied by Ca ions. Ions can be removed in acidic media at pH 1.2 and the protein can be reconstituted in the presence of Ca as a single transition-metal species at pH 5.2, which has been shown to improve the diffraction quality of ConA crystals (Kalb *et al.*, 1988). Binding of saccharide requires the presence of divalent cations (Sumner & Howell, 1936) and both metal-binding sites must be occupied for binding to occur (Kalb & Levitzki, 1968). Atomic details of saccharide binding to ConA have been described previously (Bradbrook *et al.*, 1998) and solvent interactions, to which some significance has been attributed, have been discussed (Deacon *et al.*, 1997). ConA has typically been crystallized in space group *I*222 by the dialysis method (Derewenda *et al.*, 1989) and 16 *I*222 structures to varying degrees of resolution may be found in the PDB. Parkin *et al.* (1996) reported the first atomic resolution (1.2 Å) structure of ConA. Subsequently, an ultrahigh resolution (0.94 Å) structure of ConA was published and the authors assert small-molecule precision and accuracy in this structure (Deacon *et al.*, 1997). Atomic resolution data have been defined as extending to at least 1.2 Å, with 50% or more of the theoretically measurable reflections in the outer range having $I > 2\sigma(I)$ (Dauter *et al.*, 1995, 1997; Sheldrick, 1990).

In this paper, we discuss the structure of native ConA in a different crystal form, space group $C222_1$ (unit-cell parameters a = 118.70, b = 101.38, c = 111.97 Å; two molecules in the asymmetric unit; resolution 1.56 Å). The new $C222_1$ structure differs in several loops which form crystal packing contacts or are exposed to solvent. Moreover, portions of these loops which had very weak density in the atomic resolution structures could be rebuilt into clear electron density. The $C222_1$ structure also shows solvent rearrangement around the saccharide-binding site and within the dimer interface. In an earlier publication reporting the 1.2 Å structure of ConA in space group *I*222 (Parkin *et al.*, 1996), difficulties in properly assessing accuracy and precision in the absence of both independent low-temperature refinement and of coordinate standard uncertainties for each atom (as obtained *via* matrix inversion) were discussed. This third, low-temperature, structure of ConA in a different crystal form, independently refined using powerful model-bias removal techniques, affords the opportunity to revisit the assessment of accuracy and precision in high- or atomic resolution protein structures. Furthermore, although two previously reported structures of ConA have been determined from higher resolution data (Parkin *et al.*, 1996; Deacon *et al.*, 1997), the structure in this work extends insights into the conformational flexibility for this molecule and suggests a more practical definition of accuracy in macromolecular structures compared with small-molecule accuracy and precision.

2. Materials and methods

2.1. Crystallization

Jack bean ConA was purchased from Sigma–Aldrich (St Louis, MO, USA), type 5V, C7275. Crystals grew in hanging drops containing 10 µl of protein solution (30 mg ml^{-1} in H_2O) and 10 µl precipitant solution [100 m*M* Tris–HCl pH 8.0 and 25%(*v*/*v*) ethanol] equilibrated against 1 ml of precipitant solution.

2.2. Data collection

A single crystal (0.4 × 0.4 × 0.4 mm) was cryoprotected by swiping it once through buffered 30% 2-methyl-2,4-pentanediol (MPD) using a loop and was immediately mounted and flash-cooled in the nitrogen-gas stream from a Siemens LT-2 cryocooler mounted on a Huber four-circle goniostat. Data were collected using Cu $K\alpha$ radiation (73 179 unique reflections) to 1.56 Å resolution in two sets of high- and low-resolution ω sweeps on dual multiwire area detectors from Area Detector Systems Corporation (ADSC, Poway, CA, USA). Data were integrated using the ADSC Multiwire software, space group $C222_1$ (No. 20), with unit-cell parameters at 120 K of a = 118.70, b = 101.38, c = 111.97 Å. Calculation of the Matthews coefficient (Matthews, 1968) and solvent density indicated two molecules in the asymmetric unit.

2.3. Molecular replacement

The structure was determined by molecular replacement using PDB code 1jbc (Parkin *et al.*, 1996) as a search model. *EPMR* (Kissinger *et al.*, 1999) was used in default automated partial structure mode, searching for two molecules. The search for the first molecule in the asymmetric unit converged at a correlation coefficient (CC) of 0.32. With partial structure factors for the first solution, the search for the second molecule reached convergence within three iterations (CC = 0.63). After rigid-body refinement of the two individual molecules the R value was 0.37.

2.4. Refinement

To ensure effective phase-bias removal, the model was iteratively built into maps generated by a modified *wARP* procedure, *Shake&wARP* (Segelke *et al.*, 2000). Six different starting models were created by randomly deleting 10% of all atoms and randomly perturbing the atomic coordinates by 0–0.5 Å (r.m.s.d. = 0.25 Å). Dummy atoms were automatically built into each model using *ARP* (Lamzin & Wilson, 1993) followed by unrestrained maximum-likelihood refinement with *REFMAC* (Collaborative Computational Project, Number 4, 1994; Murshudov *et al.*, 1997). After 30 cycles, the resulting six individual maps were averaged and weights (w) calculated as described in *wARP* (Perrakis *et al.*, 1997). The $C222_1$ model was iteratively rebuilt into the resulting wF_o/φ_c maps using the program *XFIT* from the *XTALVIEW* package (McRee, 1992). Residues or regions with significantly different conformations in the two molecules were excluded from the NCS restraints. After repeated cycles of water building and real-space refinement, followed by restrained *REFMAC5* maximum-likelihood refinement, the final structure (PDB code 1gkb) refined to R = 0.178 and $R_{\rm free}$ = 0.206. For each of the models discussed in this paper, real-space correlation

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution bin (1.60–1.56 Å).

Data collection	
Space group	$C222_1$
Wavelength (Å)	1.54178
Unit-cell parameters (Å)	
a	118.70
b	101.38
c	111.97
Observed reflections	234173 (7300)
Unique reflections	73179 (2920)
Multiplicity	3.2 (2.5)
R_{sym}	0.035 (0.249)
$\langle I/\sigma(I)\rangle$	16.8 (2.1)
V_M (Matthews coefficient)	3.4
Solvent (%)	63.4
Resolution (Å)	38.6–1.56
Refinement	
Free *R* value, random, 10%	0.206 (0.273)
R value	0.178 (0.275)
No. of protein atoms	3677
No. of water molecules	688
No. of metal atoms	6
R.m.s.d. bond length† (Å)	0.02
R.m.s.d. bond angle† (°)	1.85
Overall coordinate error‡ (Å)	0.094
RSCC (*Shake&wARP*)§	0.93
RSCC (*REFMAC*5)¶	0.97

† Deviations from restraint targets (Engh & Huber, 1991). ‡ Estimated standard uncertainty, diffraction precision index (DPI) based on R_{free} (Cruickshank, 1999). § Real-space correlation coefficient, F_c map against averaged and weighted *Shake&wARP* map. ¶ Real-space correlation coefficient, F_o map against F_c map, as reported by *REFMAC*5.

coefficient plots have been calculated from *Shake&wARP* maps using *OVERLAPMAP* from the *CCP*4 suite and in-house programs (B. Rupp, unpublished work). Details of the refinement and data-collection statistics are tabulated in the header file of PDB code 1gkb and summarized in Table 1.

2.5. Computational analysis

Calculations have been performed on phenol–water clusters to analyse the tyrosine–water interactions around Tyr12, to which some significance has been attributed (Deacon *et al.*, 1997). Interactions of water molecules with a protein are determined by specific local protein–water substructures and interaction energies diminish quickly with distance (Scheiner, 1997; Jeffrey, 1997; Fersht, 1999*a*,*b*). Therefore, density functional theory (DFT) and *ab initio* molecular orbital theory were used to calculate the equilibrium geometries and relative energies of C_6H_5OH–$(H_2O)_n$ (n = 1, 2) clusters in several different configurations. The DFT method chosen for this study is the well established Becke's three-parameter function (Becke, 1992*a*,*b*, 1993) with the non-local correlation (B3LYP) provided by Lee *et al.* (1988). The *ab initio* method is frozen-core second-order Møller–Plesset perturbation approximation (MP2) (Møller & Plesset, 1934; Brinkley & Pople, 1975). Both methods are known to be reliable, particularly for calculations of closed-shell stable molecules and hydrogen-bonded complexes (Kim & Jordan, 1994; Rablen *et al.*, 1998). A reasonably large basis set, 6-311++G**, (Krishnan *et al.*, 1980; Frisch *et al.*, 1984; Clark *et al.*, 1983) was used in all calculations. B3LYP and MP2 calculations were carried out with the *GAUSSIAN*98 program package (Frisch *et al.*, 1998).

Molecular-cluster models were constructed beginning from the three most stable water–phenol clusters, which were then permuted as starting models for the energy minimizations including two water molecules. Initial geometry optimization led to three distinct stable C_6H_5OH–H_2O binary clusters: water structure 1 (WS 1), where H_2O accepts a hydrogen bond from the OH group of C_6H_5OH, WS 2, where H_2O donates a hydrogen bond to the lone electron pair of the OH group of C_6H_5OH, and WS 3, where H_2O donates a hydrogen bond to the π-system of C_6H_5OH. To extend the hydrogen-bonded solvent 'network', these primary single water clusters were combined to produce dual water–phenol clusters with water positions (1, 2), (1, 3) and (2, 3) (WS 4, 5 and 6, respectively) and relative energies for the equilibrium geometries of these clusters were calculated.

3. Results and discussion

3.1. Overall features

The $C222_1$ crystal structure contains two molecules of ConA in the asymmetric unit related by a non-crystallographic twofold axis tilted 9.3° to the *c* axis. Structural features have been described in detail by Naismith *et al.* (1993), Parkin *et al.* (1996) and Deacon *et al.* (1997) and the general characteristics of the present ConA model are largely unchanged. Noteworthy differences between the new $C222_1$ structure refined at 1.56 Å and the models refined at higher resolution (1.2 Å and effectively 1.0 Å) arise from changes in crystal packing contacts and rearrangements in the solvent network. Several loops which had very weak electron density in the atomic resolution structures were rebuilt into clear electron density in the present model. Rearrangements in the water structure arising from different crystal packing are significant. In all crystal forms, the hydrogen-bonding network around Tyr12 is consistent with DFT and MP2 calculations of favorable phenol–H_2O geometries. A fourth, solvent-accessible, metal site (S4) is present. Structural features representing significant changes or additional information compared with the previous models are presented. The two atomic resolution structures and the present $C222_1$ structure are compared based on electron-density maps calculated using the same phase-bias removal procedure (described in §2.5) for each set of deposited coordinates and structure factors.

3.2. Loop regions and crystal contacts

Two loops in the present structure are considerably different from the previously determined atomic resolution structures (Fig. 1). Although the other two structures were refined at higher resolution, the electron density for these particular loops is weak, whereas the density in the $C222_1$ structure is well defined. Loop Lys114–Leu126 in the $C222_1$ crystal form is stabilized by crystal contacts that do not exist in the *I*222 crystal form. Weak density visible at 0.8σ suggests that Lys114 may be partially succinylated, but this feature has

not been modelled and refined. Loop His180–Ser190 is also stabilized by different crystal contacts to those seen in crystal form *I*222. Furthermore, because these loops are adjacent to one another in the $C222_1$ structure, these crystal contacts help to stabilize both loops through side-chain and backbone interactions. The densities for Tyr120 and His121, which are poorly defined in the *I*222 structures, are distinct in the $C222_1$ structure.

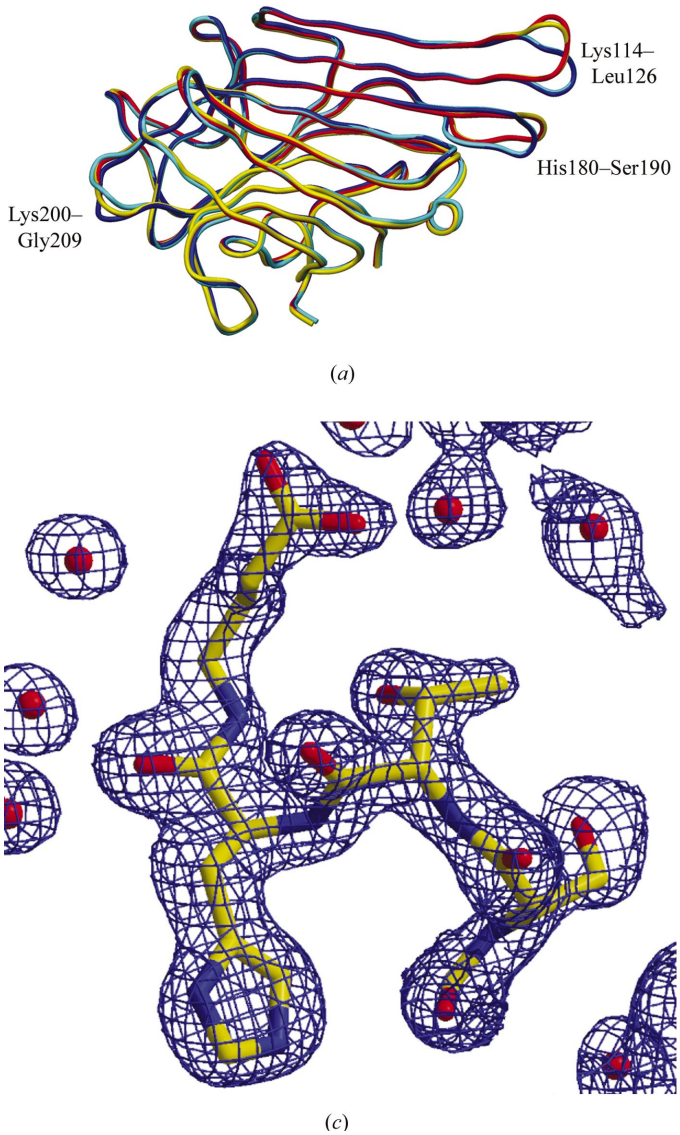
Loop Lys200–Gly209, which is very near to the saccharide-binding site, is also in a different conformation in the $C222_1$ structure to that found in the *I*222 crystal forms. Furthermore, this loop exhibits a different conformation in each of the two independent chains of the dimer. Loop *A* makes different crystal packing contacts to loop *B*, which affects the local solvent structure in the saccharide-binding site (§3.3).

A solvent-exposed loop Ser161–Ser164, exhibiting disorder in the atomic resolution structures, has continuous backbone density in chain *A* of the present structure and side-chain density is clearly visible (Fig. 1). Although there is a break in main chain *B* at 1σ, the backbone structure is clear and side-chain density is apparent. We note that surprisingly the electron density for this solvent-exposed loop is most poorly defined in the highest resolution structure (Deacon *et al.*, 1997) and clearest in the 1.56 Å structure. As the data for 1nls were merged from four different crystals, it is possible that individual multi-conformer arrangements do not create sufficiently localized electron density to discern any defined loop.

3.3. Solvent structure

The bound solvent structure in the $C222_1$ model, as in the *I*222 structures (Parkin *et al.*, 1996; Deacon *et al.*, 1997), reveals several regions where water molecules are surrounded by many potential hydrogen-bond acceptors and the solvent has been described as 'dynamic' (Deacon *et al.*, 1997). The dimer interface in the $C222_1$ structure also contains an additional hydrogen-bonded network of trapped waters near Tyr176.

Three highly conserved water molecules located in the saccharide-binding site near Asp208 have been observed at



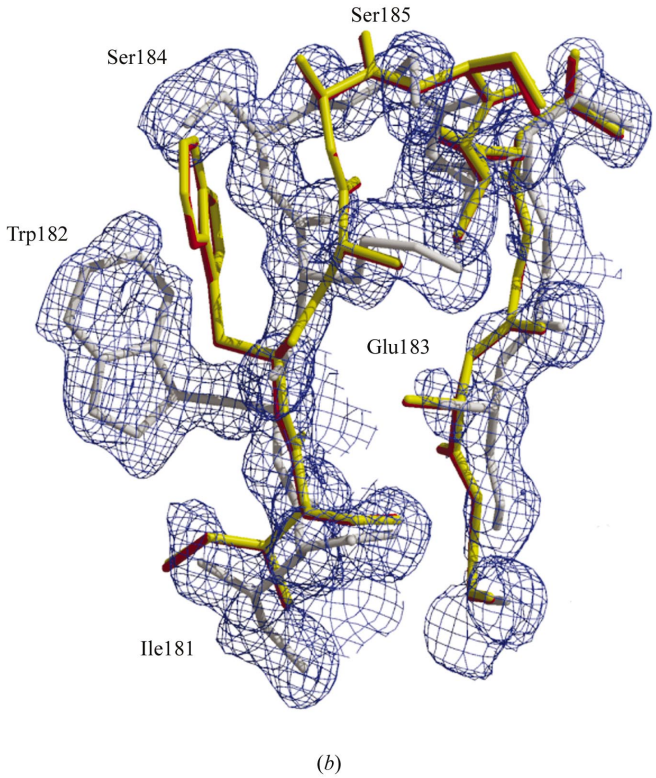


Figure 1
(*a*) A superposition of four high-resolution ConA C^{α} backbone traces: cyan, $C222_1$ structure 1gkb (*A*); blue, 1gkb (*B*); red, *I*222 structure 1jbc; yellow, *I*222 structure 1nls. R.m.s. deviation of backbone atoms: 1gkb *A*/*B*, 0.33 Å; 1gkb (*A*)/1jbc, 0.71 Å; 1gkb (*B*)/1jbc, 0.70 Å; 1gkb (*A* or *B*)/1nls, 0.73 Å; 1jbc/1nls, 0.37 Å. Loops that exhibit considerably different backbone conformations among the structures are labelled. (*b*) Electron density (*Shake&wARP* map, 1σ level) for loop 180–190 with tube models of the polypeptide for the three structures: white, 1gkb (*A*); red, 1jbc; yellow, 1nls. (*c*) *Shake&wARP* electron density (1σ) and polypeptide for Tyr120 and His121 of 1gkb (*A*) within loop 114–126.

room temperature and in the previous atomic resolution structures (Naismith *et al.*, 1994; Parkin *et al.*, 1996; Deacon *et al.*, 1997). These waters are again present in the $C222_1$ structure. A fourth water molecule has been observed directly over the aromatic ring of Tyr12 in both of the two previously determined low-temperature atomic resolution *I*222 structures (Deacon *et al.*, 1997; Parkin *et al.*, 1996). This fourth water is also present in our low-temperature $C222_1$ structure, but it is seen in the proximity of the aromatic ring of Tyr12*B* only, not Tyr12*A*.

In all three structures, waters in the region of Tyr12 are part of a well defined extended hydrogen-bonding network. Whereas the solvent environment around Tyr12*B* more closely resembles that of the *I*222 structures, water around Tyr12*A* is hydrogen bonded to the phenolic oxygen (O^{η}). Hydrogen bonding in this region of the $C222_1$ model is also satisfied by crystal contacts. In the absence of these contacts in *I*222, the solvent network locally rearranges to satisfy hydrogen-bonding requirements (details discussed in §3.5).

Based on density apparent in the *Shake&wARP* map of the ultrahigh-resolution *I*222 structure, a well ordered MPD molecule is probably present in 1nls (Fig. 2). MPD is a common cryoprotectant and was in fact used in the case of 1nls (soaks in 25% MPD; Deacon *et al.*, 1997; Deacon, personal communication). An additional feature of 1nls not previously reported appears to be O-linked carbohydrate to Ser134 (not shown), which might be further refined and analysed.

3.4. Fourth metal-binding site

A new metal-binding site, S4, located near one of the solvent-exposed loops is evident in the $C222_1$ structure. S4 is distinct from the S3 site reported by Naismith *et al.* (1993). The ion in the present structure has been refined as Mg^{2+}, sixfold coordinated, with slightly distorted octahedral geometry (Fig. 3). We have also observed the S4 site in *Shake&wARP* maps of both the *I*222 structures 1jbc and 1nls. However, the S4 site is less well defined in these structures and the ions appear to be pentacoordinate, with slightly distorted square-pyramidal geometry. There is no biological relevance evident for metal site S4.

3.5. Computational analysis of the hydrogen-bonded solvent network in the saccharide-binding site

A conserved water molecule has been observed directly over the aromatic ring of Tyr12 in both of the two low-temperature *I*222 structures determined previously (Parkin *et al.*, 1996; Deacon *et al.*, 1997) and it is also seen in our low-temperature $C222_1$ structure in proximity to the aromatic ring of Tyr12*B*. Some significance has been attributed to this particular water (Deacon *et al.*, 1997), because its position corresponds to that of the exocyclic C6 of bound monosaccharide (Bradbrook *et al.*, 1998). However, because this water is seen in the low-temperature $C222_1$ structure above Tyr12*B* only and not above Tyr12*A*, we performed theoretical calculations on phenol–water clusters to better understand the interactions of tyrosine with water. Our calculations show that binary water structure (WS) 1, where the OH group of phenol donates a traditional hydrogen bond to the O atom of water, is the most stable geometry for a C_6H_5OH–H_2O dimer, with a

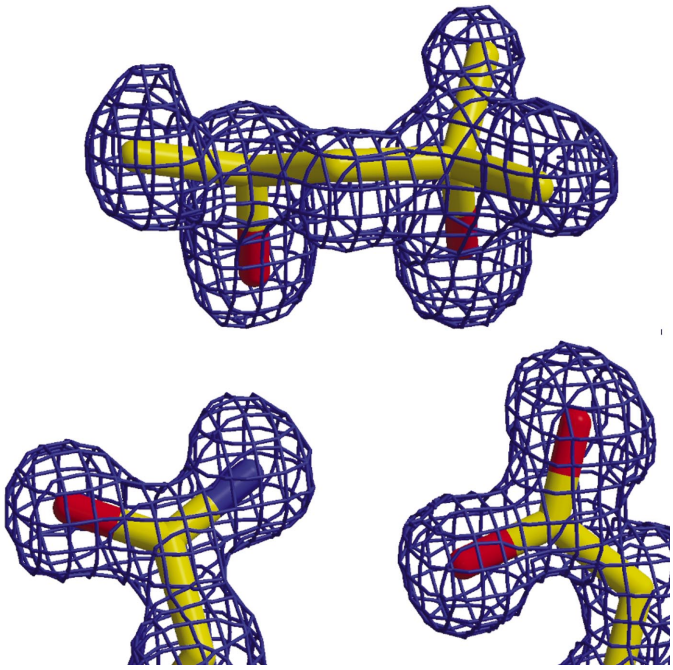


Figure 2
Putative MPD molecule found in *Shake&wARP* map (1σ contour level) of 1nls; shown is an unrefined stick model of MPD placed in density. The MPD molecule appears to be hydrogen bonded to Gln137 NE2 and Asp139 OD2 of the same molecule. Gln137 NE2 and OE1 could be flipped to best satisfy the hydrogen-bond network. Figure created with *XTALVIEW* (McRee, 1992) and *Raster3D* (Merritt & Bacon, 1997)

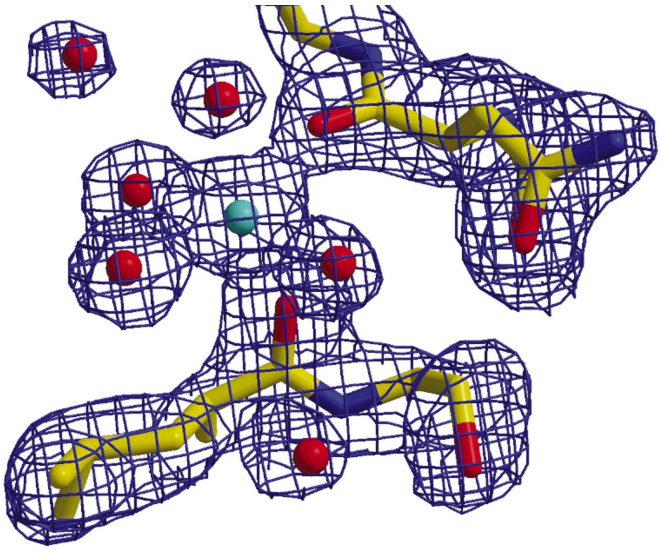


Figure 3
The S4 ion-binding site Mg803A. The geometry of this sixfold coordinated site is distorted octahedral. Two residues, Leu107 and Asn131 *via* main-chain carbonyl groups, and four waters bind to this ion. The average bond length between the magnesium ion and each ligand is 2.4 Å. Bond distances, angles and connectivity are listed in the LINK and SITE records of PDB entry 1gkb.

binding energy of 35 kJ mol^{-1} at the MP2 level (25.5 kJ mol^{-1} at B3LYP). WS 2, where H_2O donates a hydrogen bond to the OH group of phenol, is less stable by 12.6 kJ mol^{-1} than WS 1. WS 3, where H_2O donates a hydrogen bond to the π-system of phenol, is less stable by 14.6 kJ mol^{-1} than WS 1 and is nearly as stable as WS 2. For the ternary clusters C_6H_5OH–$(H_2O)_2$, WS 4 and 5 are more stable than WS 6, with binding energies of 78.3, 79.5 and 59.4 kJ mol^{-1}, respectively.

In our $C222_1$ structure, we find different hydrogen-bond patterns in the solvent networks around Tyr12*A* and Tyr12*B*, respectively, resulting from varying environments arising from the crystal packing arrangements. Equilibrium geometries of the dual water–phenol clusters WS 5 and WS 6 from our DFT and MP2 calculations are consistent with those configurations observed in all three high-resolution ConA structures in the two crystal forms. The combined multiple hydrogen-bonding dual-water configuration of WS 5, which is consistent with the configuration around Tyr12*A*, is favored by almost 21 kJ mol^{-1}. Equilibrium geometries for WS 5 and WS 6 (water directly above the tyrosine π-system) are shown below their corresponding Tyr12*A* and Tyr12*B* electron densities in Fig. 4. The correspondence between the experimentally observed structures and our calculations suggest that the observed interactions between Tyr12 and water in the different crystal forms, although perhaps interesting, are to be expected and are not unique to the presence of an active site. They arise from reorganization of the hydrogen-bonded water network owing to the presence of protein. This largely entropy-driven propensity has been discussed by Fersht (1999*a*,*b*). In the absence of conventional hydrogen-bond acceptors, as for example in the Tyr12*A* pocket of ConA,

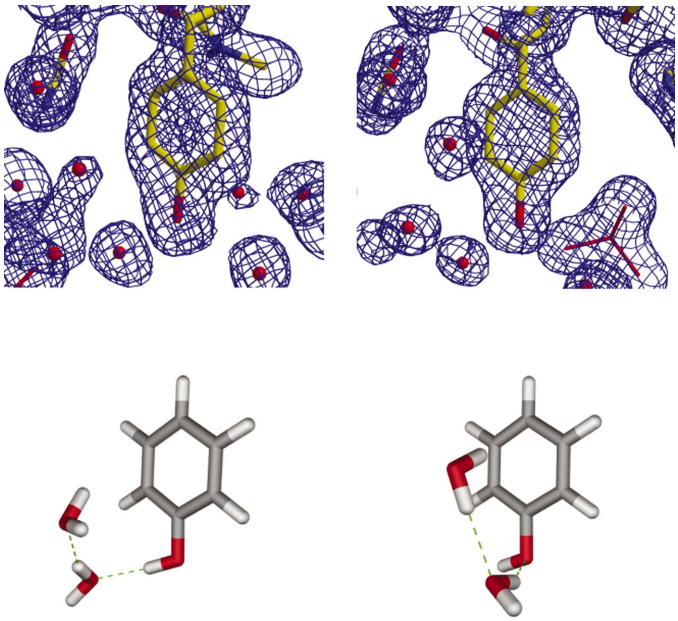


Figure 4
Electron densities (*Shake&wARP*, 1σ level) for Tyr12*A* (left) and Tyr12*B* (right) from 1gkb compared with calculated equilibrium geometries for ternary water–phenol conformations WS 5 (left) and WS 6 (right). The red stick triangle (right panel, right bottom) is view-clipped symmetry-related residue Leu99*A*.

X—H$\cdots\pi$(Ph) bonds satisfy hydrogen-bond potentials (Steiner, 1998).

Hydrogen bonds between donors X—H and the π-electron cloud of an aromatic moiety (aromatic or π-facial) were discovered by Wulf *et al.* (1936) and interactions of this type are well documented in structural organic chemistry (Malone *et al.*, 1997). These and other 'non-conventional' hydrogen bonds have been documented in the structural biology literature and have been shown to have increasingly important roles in protein structure and function. For a general survey of this field, see Desiraju & Steiner (1999). More recently, analysing high-resolution structures ($\leq$ 1.6 Å), Steiner & Koellner (2001) have discussed the frequency of hydrogen bonds with π-acceptors and their roles in stabilizing local three-dimensional structures in proteins.

The metrics of water hydrogen bonding to proteins in crystals have been analyzed previously and reviewed by, for example, Burley & Petsko (1988) and Jeffrey & Saenger (1994). Interactions of water with aromatic ring systems have been described by others, including Suzuki *et al.* (1992), Rodham *et al.* (1993) and Dougherty (1996). Energy values range from 2.1 to 125 kJ mol^{-1} and distances between X and the aromatic π-center cluster around 3.2–3.8 Å. The geometry of X—H$\cdots\pi$(Ph) hydrogen bonds tends to be 'softer' than that of conventional hydrogen bonds, allowing for large lateral displacements of the donor and varying hydrogen-bond angles without much change in energy (Suzuki *et al.*, 1992).

3.6. Quality assessment

PROCHECK (Laskowski *et al.*, 1993) and *WHAT_CHECK* (Hooft *et al.*, 1996) reports were created upon coordinate deposition and are available from the PDB entry 1gkb. The model has 87% of the residues in most favored regions and none in disallowed regions. Real-space fit correlation between the structure model and the electron-density maps is high over most parts of the structure. The average real-space correlation coefficient (RSCC) of F_o *versus* F_c reported by *REFMAC*5 is 0.97. Fig. 5 shows the real-space correlation coefficient obtained from bias-minimized, averaged and weighted *Shake&wARP* maps (average RSCC = 0.93) and the corresponding *B* factor as a function of residue.

3.7. Flexibility and the accuracy of macromolecular models

The meaning of precision and accuracy in macromolecular structure determination is somewhat ambiguous. While in small-molecule structures great coordinate precision nearly always implies high structural accuracy, in macromolecular crystallography, even with high-resolution data, refinement practice and phase bias affect the accuracy of models (Fig. 2; Kleywegt & Jones, 1997). Moreover, the flexibility of macromolecules frequently results in differences between independently determined structures and thus suggests a more subtle definition of structural accuracy in a biological context. Based on the following discussion, we present three different presentations of multiple structure models, all of which allow pinpointing of local variability.

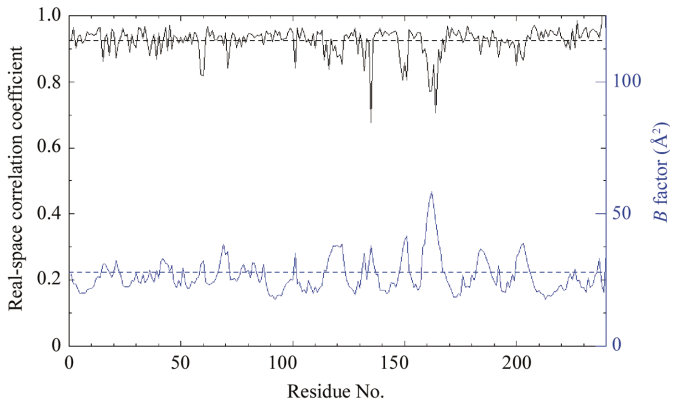


Figure 5
Real-space correlation (black) and B factors (blue) as a function of residue for chain A of 1gkb. The plot was created automatically by a *Shake&wARP* script (Rupp, unpublished). Dashed lines represent corresponding mean values.

The precision with which coordinate positions of an atom are determined in crystallographic refinement depends on the 'sharpness' of the electron-density peak, which directly relates to the temperature factor B (Cruickshank, 1999) and correlates with the mean displacement $\langle u \rangle$ of atoms: $B = 8\pi^2\langle u^2 \rangle$. A similar correlation can be used to analyse ensembles of independently determined structure models. For an ensemble of structures, the average per-residue B factor and its correlation with the mean-square coordinate differences between superimposed models can be used to pinpoint local variation between the individual models.[1] Significant deviations from such a correlation are indicative of local 'hot spots' in a structure and defining accuracy – as far as the overall structure is concerned – becomes ambiguous. When large mean positional displacements between models are observed while the mean B factors remains small, the corresponding regions are well defined in each model, but they differ significantly from each other. Model differences may arise from multiple loop conformations caused by varying crystal contacts or they could be indications of uncertain tracing in one of the models. Large mean per-residue B factors combined with large mean displacements are indicative of regions of generic disorder across the models. In either case, care must be taken when assigning significance (or accuracy) based on the features of one structure alone. The correlations between mean per-residue B and the mean-square displacement between models can be represented in a variety of ways. In Figs. 6, 7 and 8 we show three different representations for ConA, using either a simple line graph, a correlation plot and, most satisfying in our view, a structural model presentation using *MOLMOL* (Koradi *et al.*, 1996).

For our ensemble of ConA structures, 1jbc, 1nls and two molecules of 1gkb, we demonstrate flexibility not evident from separate crystal structures, however precise. In Fig. 6 the average per-residue B factors (black) and mean coordinate variations between models (red, expressed as B-factor equivalents) are plotted. The B equivalent is again calculated from the relation between B factor and mean-square displacement $B = 8\pi^2\langle u^2 \rangle$. Immediate attention is attracted to areas where large coordinate deviations exist that do not correlate with large mean B factors. The same residues are easily identified from a correlation plot in Fig. 7 of average per-residue B factors (black) and mean coordinate variations. The *MOLMOL* representation of this information (Fig. 8) clearly indicates areas where packing contacts lead to large coordinate variation with low mean B factors and where loops with genuine flexibility have both large B factors and large coordinate variations. Fig. 8 thus provides a very accurate picture of the entire ConA structure.

[1] To allow a meaningful analysis of large multidomain structures, independent superpositions for each domain may be necessary.

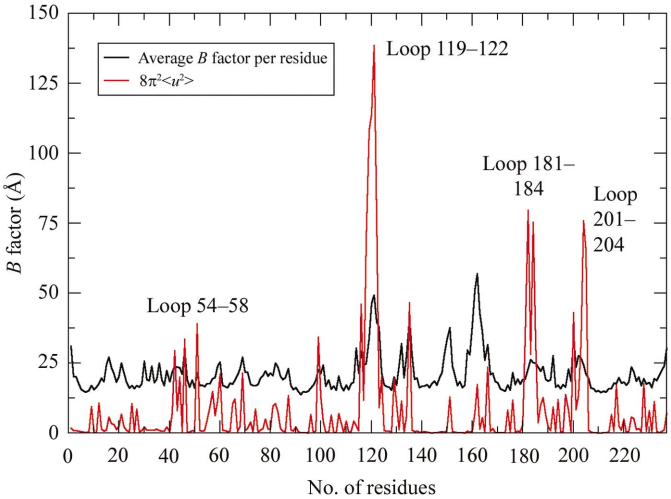


Figure 6
Average residual B factors (black) and mean coordinate variations (red) for the ensemble of ConA structures 1jbc, 1nls and the two 1gkb models. Mean coordinate variations are expressed as B-factor equivalents $8\pi^2\langle u^2 \rangle$. Immediate attention is drawn to areas where large coordinate deviations do not correlate with large average B factors.

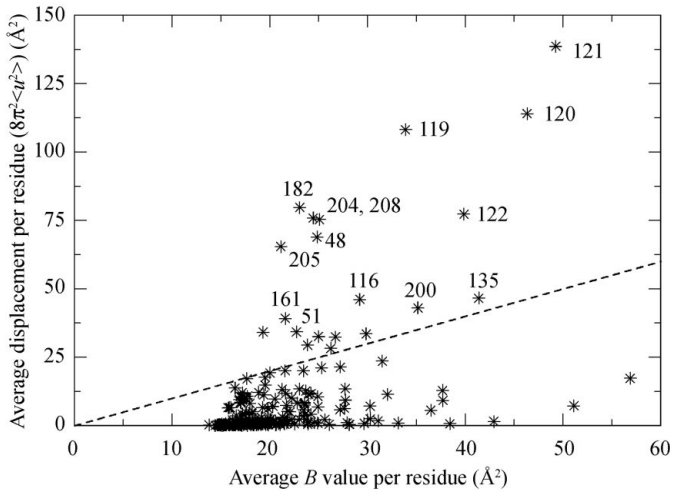


Figure 7
Correlation plot of the average per-residue B factors *versus* mean coordinate deviation. The same outliers as indicated in Fig. 6 are readily identified as loop regions in Fig. 8. The dashed line is a guide to the eye drawn at $x = y$.

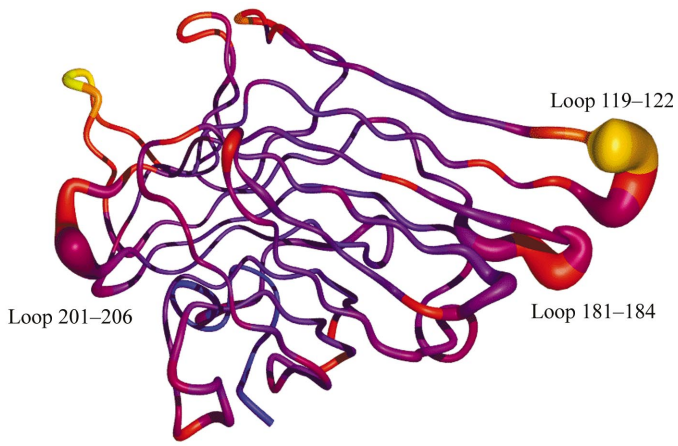


Figure 8
A *MOLMOL* representation of the average per-residue *B* factors (color coded from low to high: blue, purple, red, orange, yellow, respectively) mapped onto mean coordinate deviation (represented by the width of model) per residue. A value of 0.1 Å has been added to the mean coordinate deviations to create a minimal ribbon width suitable for display. Program *MOLMOL* (Koradi *et al.*, 1996), rendered with *POV-Ray* (Persistance of Vision Raytracer V3.1g).

4. Concluding remarks

Although high model coordinate precision (and crystallographic accuracy) is achievable based on properly refined atomic resolution data, the accuracy of a particular structural model, as it relates to the molecule in biological context, remains difficult to define. Structures modelled in more than one crystal form can provide useful insights into flexibility of molecules and may allow discrimination between genuine flexibility and variation resulting from specific crystal packing. Disordered loops may be stabilized in another crystal form and fluid water structures may exhibit rearrangements purely to preserve or optimize the hydrogen-bonded solvent network. The demonstration that several precise structures of the same molecule can differ substantially in local detail should remind users of crystallographic models to be aware of interactions with symmetry-related molecules and to consider their potential impact when interpreting structural models.

This work was conducted under the auspices of the US Department of Energy at Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48. Portions of this work were supported by Faculty Fellowships awarded to KAK by Associated Western Universities. F-MT was supported by the Research Corporation (Grant No. CC 4713) and the Camille and Henry Dreyfus Foundation (Grant No. TH-00-028). We thank Dr George Craig, LLNL, for participation in the crystal-growth experiments and Dr Sean Parkin, University of Kentucky, for the data collection on the ADSC Multi-Wire system. A crystallization and data-collection paper is planned (Parkin & Craig, in preparation).

References

Becke, A. D. (1992*a*). *J. Chem. Phys.* **96**, 2155–2160.
Becke, A. D. (1992*b*). *J. Chem. Phys.* **97**, 9173–9177.
Becke, A. D. (1993). *J. Chem. Phys.* **98**, 5648–5652.
Bittiger, H. & Schnebli, H. P. (1976). Editors. *Concanavalin A as a Tool.* New York: John Wiley & Sons.
Bradbrook, G. M., Gleichman, T., Harrop, S. J., Habash, J., Raftery, J., Kalb (Gilboa), J., Yariv, J., Hillier, I. H. & Helliwell, J. R. (1998). *J. Chem. Soc. Faraday Trans.* **94**, 1603–1611.
Brinkley, J. S. & Pople, J. A. (1975). *Int. J. Quantum Chem.* **9**, 229.
Burley, S. K. & Petsko, G. A. (1988). *Adv. Protein Chem.* **39**, 125–189.
Chrispeels, M. J. & Raikhel, N. V. (1991). *Plant Cell*, **3**, 1–9.
Clark, T., Chandrasekhar, J., Spitznagel, G. W. & Schleyer, P. von R. (1983). *J. Comput. Chem.* **4**, 294–301.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). *Curr. Opin. Struct. Biol.* **5**, 784–790.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
Deacon, A., Gleichman, T., Kalb (Gilboa), A. J., Price, H., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **93**, 4305–4312.
Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb (Gilboa), A. J., Dodson, E. J., Papiz, M. Z., Wan, T. & Campbell, J. (1989). *EMBO J.* **8**, 2189–2193.
Desiraju, G. R. & Steiner, T. (1999). *The Weak Hydrogen Bond in Structural Chemistry and Biology.* Oxford University Press.
Dougherty, D. A. (1996). *Science*, **271**, 163–168.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Fersht, A. (1999*a*). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, ch. 11, p. 332. New York: Freeman & Company.
Fersht, A. (1999*b*). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, ch. 17, p. 510. New York: Freeman & Company.
Frisch, M. J., Pople, J. A. & Brinkley, J. S. (1984). *J. Chem. Phys.* **80**, 3265–3269.
Frisch, M. J. *et al.* (1998). *GAUSSIAN*98, Revision A.2. Gaussian Inc., Pittsburgh, PA, USA.
Goldstein, I. J. & Poretz, R. D. (1986). *The Lectins. Properties, Function and Application in Biology and Medicine*, edited by E. E. Liener, N. Sharon & I. J. Goldstein, pp. 35–247. London/Orlando: Academic Press.
Hardman, K. D. & Ainsworth, C. F. (1976). *Biochemistry*, **15**, 1120–1128.
Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272–272.
Jeffrey, G. A. (1997). *An Introduction to Hydrogen Bonding.* Oxford University Press.
Jeffrey, G. A. & Saenger, W. (1994). *Hydrogen Bonding in Biological Structures*, ch. 23. New York: Springer–Verlag.
Kalb, A. J. & Levitzki, A. (1968). *Biochem. J.* **109**, 669–672.
Kalb, A. J., Yariv, J., Helliwell, J. R. & Papiz, M. Z. (1988). *J. Cryst. Growth*, **88**, 537–540.
Kim, K. & Jordan, K. D. (1994). *J. Phys. Chem.* **98**, 10089–10094.
Kissinger, C. R., Gelhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 208–230.
Koradi, R., Billeter, M. & Wüthrich, K. (1996). *J. Mol. Graph.* **14**, 51–55.
Krishnan, R., Brinkley, J. S., Seeger R. & Pople, J. A. (1980). *J. Chem. Phys.* **72**, 650–654.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785–789.
Lis, H. & Sharon, N. (1973). *Annu. Rev. Biochem.* **42**, 541–574.

Lis, H. & Sharon, N. (1986). *Annu. Rev. Biochem.* **53**, 35–67.
McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.
Malone, J. F., Murray, C. M., Charlton, M. H., Docherty, R. & Lavery, A. J. (1997). *Chem. Rev.* **97**, 1303–1324.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497 .
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Møller, C. & Plesset, M. (1934). *J. Phys. Rev.* **26**, 618–622.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. & Rini, J. M. (1998). *Science*, **280**, 1277–1281.
Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb (Gilboa), A. J. & Yariv, J. (1994). *Acta Cryst.* D**50**, 847–858.
Naismith, J. H., Habash, J., Harrop, S., Helliwell, J. R., Hunter, W. N., Wan, T. C. M., Weisgerber, S., Kalb (Gilboa), A. J. & Yariv, J. (1993). *Acta Cryst.* D**49**, 561–571.
Parkin, S., Rupp, B. & Hope, H. (1996). *Acta Cryst.* D**52**, 1161–1168.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Peumans, W. J. & Van Damme, E. J. M. (1995). *Histochem. J.* **27**, 253–271.
Rablen, P. R., Lockman, J. W. & Jorgensen, W. L. (1998). *J. Phys. Chem. A*, **102**, 3782–3797.
Rodham, D. A., Suzuki, S., Suenram, R. D., Lovas, F. J., Dasgupta, S., Goddard, W. A. III & Blake, G. A. (1993). *Nature (London)*, **362**, 735–737.
Scheiner, S. (1997). *Hydrogen Bonding, a Theoretical Perspective.* Oxford University Press.
Segelke, B. W., Forstner, M., Knapp, M., Trakhanov, S. D., Parkin, S., Newhouse, Y. M., Bellamy, H. D., Weisgraber, K. H. & Rupp, B. (2000). *Protein Sci.* **9**, 886–897.
Sharon, N. (1993). *Trends Biol. Sci.* **18**, 221.
Sharon, N. & Lis, H. (1975). *Methods Membr. Biol.* **3**, 147–200.
Sharon, N. & Lis, H. (1989). *Science*, **246**, 227–234.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Steiner, T. (1998). *Acta Cryst.* D**54**, 584–588.
Steiner, T. & Koellner, G. (2001). *J. Mol. Biol.* **305**, 535–557.
Sumner, J. B. (1919). *J. Biol. Chem.* **37**, 137–142.
Sumner, J. B. & Howell, S. F. (1936). *J. Biol. Chem.* **113**, 607.
Suzuki, S., Green, P. G., Bumgarner, R. E., Dasgupta, S., Goddard, W. A. III & Blake, G. A. (1992). *Science*, **257**, 942–945.
Weis, W. I., Taylor, M. E. & Drickamer, K. (1998). *Immunol. Rev.* **163**, 19–34.
White, M. R., Crouch, E., Chang, D., Sastry, K., Guo, N., Engelich, G., Takahashi, K., Exekowitz, R. A. & Hartshorn, K. L. (2000). *J. Immunol.* **165.** 2108–2115.
Wulf, O. R., Liddel, U. & Hendricks, S. B. (1936). *J. Am. Chem. Soc.* **58**, 2287–2293.