Structural characterisation of the native fetuin-binding protein *Scilla campanulata* agglutinin: a novel two-domain lectin

Lisa M. Wright[a,1], Colin D. Reynolds[a,*], Pierre J. Rizkallah[b], Anthony K. Allen[c], Els J.M. Van Damme[d], Michael J. Donovan[a], Willy J. Peumans[d]

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

Received 8 November 1999; received in revised form 16 December 1999

Edited by Richard Cogdell

Abstract The three-dimensional structure of a 244-residue, multivalent, fetuin-binding lectin, SCAfet, isolated from bluebell (*Scilla campanulata*) bulbs, has been solved at 3.3 Å resolution by molecular replacement using the coordinates of the 119-residue, mannose-binding lectin, SCAman, also from bluebell bulbs. Unlike most monocot mannose-binding lectins, such as *Galanthus nivalis* agglutinin from snowdrop bulbs, which fold into a single domain, SCAfet contains two domains with approximately 55% sequence identity, joined by a linker peptide. Both domains are made up of a 12-stranded β-prism II fold, with three putative carbohydrate-binding sites, one on each subdomain. SCAfet binds to the complex saccharides of various animal glycoproteins but not to simple sugars.

Key words: Fetuin-binding lectin; Monocot mannose-binding lectin; Two-domain lectin; β-Prism II fold; X-ray crystallography; *Scilla campanulata* agglutinin

1. Introduction

Two lectins have recently been isolated from bluebell (*Scilla campanulata*) bulbs: a mannose-binding lectin, SCAman, and a fetuin-binding lectin, SCAfet [1]. SCAman is a tetramer made up of four identical unglycosylated subunits of 13 143 Da, containing 119 amino acid residues and binds specifically to α(1,3)- and α(1,6)-linked mannopyranosides. SCAfet which is also not glycosylated forms tetramers of identical 26 224 Da protomers, with 244 amino acid residues in a single polypeptide chain, but unlike SCAman, SCAfet does not bind any simple sugars [1]. Both SCAman and SCAfet are members of the superfamily of monocot mannose-binding lectins but SCAfet, unlike the majority of proteins that have been characterised in this group [2], contains a polypeptide chain that is folded into two distinct domains, SCAfet-DOM1 and SCAfet-DOM2, respectively [1].

To date, there have been only a few reports in the literature of monocot mannose-binding lectins which have been shown to contain either intact or cleaved two-domain protomers. Native T×LCI from tulip bulbs is a tetramer of four, two-domain, 28 kDa subunits, which undergoes partial proteolysis, yielding both cleaved and uncleaved protomers [3]. The N-terminal and C-terminal domains of T×LCI have only 20% sequence identify and show specificities for the monosaccharides, mannose and GalNAc, respectively. *Arum maculatum* agglutinin (AMA) is a dimeric lectin made up of two completely cleaved protomers which share approximately 40% sequence identity and probably exhibit a different specificity towards monosaccharides [4]. *Allium sativum* agglutinin I (ASAI) from garlic bulbs is completely cleaved into two mannose-binding domains which are highly homologous (85% sequence identity) [5]. The structure of the garlic lectin heterodimer, containing two subunits of approximately 11.5 and 12.5 kDa, has recently been determined [6]. Here we present the first report of the 3D structure of the novel intact two-domain lectin, SCAfet, at 3.3 Å resolution.

2. Materials and methods

SCAfet has been isolated from bluebell bulbs using a fetuin-Sepharose affinity column, cloned from a cDNA library and sequenced, as described earlier [1].

SCAfet crystals were grown from 70% saturated ammonium sulphate, pH 4.8, by the hanging drop vapour diffusion method, as previously described [7]. X-ray intensity data were measured at

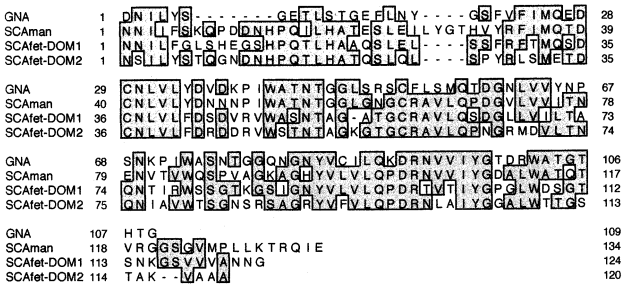

Fig. 1. Alignment of the amino acid sequence of *Galanthus nivalis* agglutinin (GNA) with those of the *Scilla* mannose-binding (SCAman) and fetuin-binding (SCAfet-DOM1 and SCAfet-DOM2) lectins. Dashes represent deletions and identical residues are boxed.

*Corresponding author.
E-mail: bmscreyn@livjm.ac.uk

[1] Present address: York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO1 5DD, UK.

Abbreviations: AMA, *Arum maculatum* agglutinin; ASAI, *Allium sativum* agglutinin; DOM, domain; SCAman, mannose-binding *Scilla campanulata* agglutinin; SCAfet, fetuin-binding *Scilla campanulata* agglutinin; T×LCI, tulip lectin with complex specificity I

PII: S0014-5793(00)01109-1

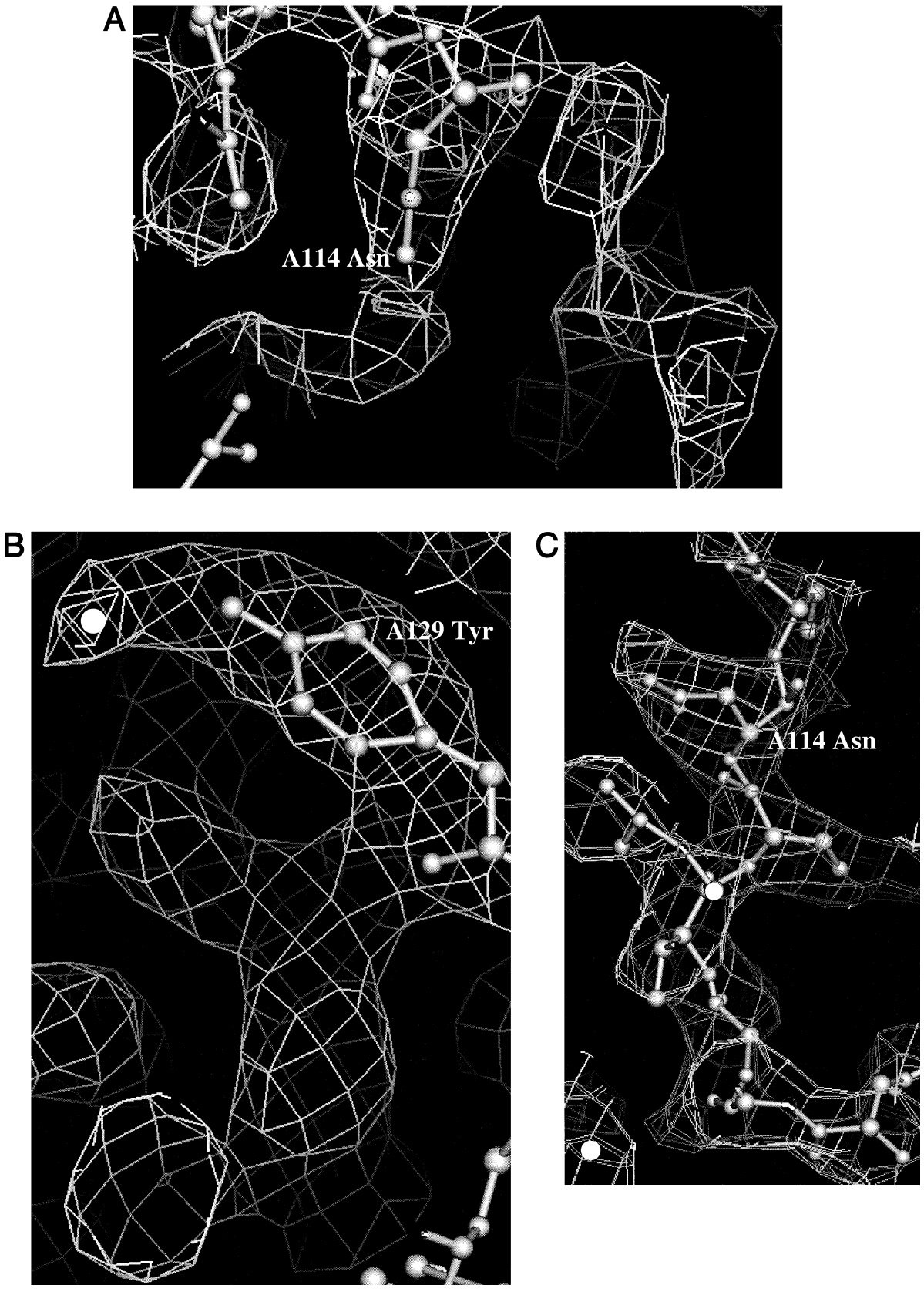

Fig. 2. Vacant Fo-Fc electron density in the region of (A) 114 Asn and (B) 129 Tyr of subunit A. The linker peptide 114 Asn-125 Asn, as well as the residues 126 Ser, 127 Ile, 128 Leu were excluded from the model at this stage of the analysis. (C) Modelling of part of the 11-residue linker peptide into the difference density. The figures were produced with O [15].

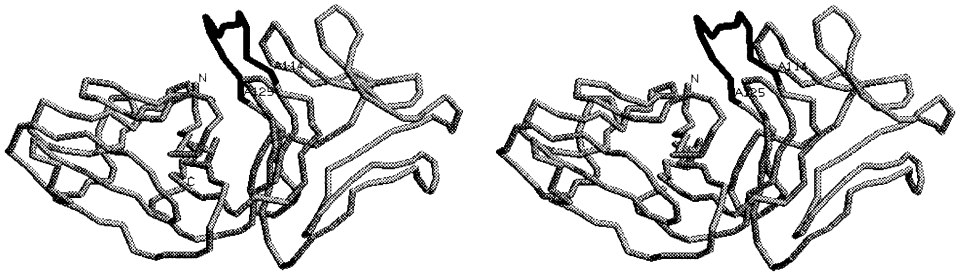

Fig. 3. Stereoview of the three-dimensional backbone fold of the SCAfet monomer showing the two domains, DOM1 and DOM2. The linker peptide is shown in black. The figure was produced with MOLSCRIPT [19].

293K, with $\lambda = 0.92$ Å, on station 9.5 at the Synchroton Radiation Source, Daresbury Laboratory, UK, using a large MAR Research image plate detector system. A total of 180° of X-ray data was collected from two crystals to a resolution of 3.3 Å. These data were processed with MOSFLM [8] and scaled using the CCP4 suite of programmes [9] yielding a 95.3% complete data set. The monoclinic unit cell parameters refined to $a = 277.94$ Å, $b = 164.10$ Å, $c = 53.60$ Å and $\beta = 95.38°$, in space group C2, with six SCAfet monomers per asymmetric unit, including 68.2% solvent. The crystal structure of SCAfet was determined by the molecular replacement method, using the programme package AMoRe [10] and the SCAman dimer [11] as a search model, with data in the resolution range 8–4 Å. The sequence of the SCAman monomer shows 48% and 53% sequence identity to the first and second domains of SCAfet, respectively [1]. Six clear solutions to both the rotation and translation function were obtained, consistent with the V_m value of 3.87 Å^3 Da^{-1} [12]. Following optimisation with the rigid-body refinement protocol of AMoRe, all six solutions had *R*-factors of 44.0% and correlation coefficients of 52.7%. The packing function indicated very few close packing contacts. Refinement of the six monomers within the asymmetrical unit was undertaken against data in the resolution range 20–3.3 Å, using the program REFMAC [13]. After a round of refinement (20 cycles), corrections to the amino acid sequence of each of the six SCAfet monomers were made, according to the sequence given in Fig. 1. At this stage each SCAfet subunit was comprised of two separate polypeptide chains corresponding to domains 1 and 2 of the molecule. In order to form the complete SCAfet monomer, an insert of 11 residues was modelled into the difference electron density maps to join the two chains together. Density modified maps were calculated using DM [14], which improved the electron density for the linker peptides and enabled the inserts to be built (Fig. 2) for the SCAfet monomers. NCS restraints were used throughout the refinement except for the final 50 cycles. Electron density maps ($|F_o|-|F_c|$) and ($2|F_o|-|F_c|$) were computed within the resolution range 3.3–20.0 Å and displayed using the modelling programme O [15] at various stages during refinement.

3. Results and discussion

The final model of the SCAfet structure contains 10 890 protein atoms with a crystallographic *R*-factor of 19.0% ($R_{free} = 29.3\%$, the *R*-factor calculated for 5% of the data omitted from the refinement) for 34 139 reflections in the resolution range 20.0–3.3 Å.

Our structural studies show that the polypeptide chain of each SCAfet monomer is 244 amino acid residues long and is folded into two distinct domains, 1 and 2, joined by a linker peptide, as illustrated in Fig. 3. Both domains of SCAfet are composed of three subdomains (I, II and III) related by pseudo-three-fold symmetry with each subdomain containing a 4-stranded antiparallel β-sheet. Thus, each domain consists of a 12-stranded β-prism II fold, similar to that reported in the other single-domain monocot mannose-binding lectins [6,11,16–18]. The root mean square deviations for the 244 α atoms of monomer 1 and the other five independent monomers were in the range 0.6 to 0.7 Å.

The 11-residue linker peptide forms a β-turn between the C-terminal strand of domain 1 and the N-terminal strand of domain 2 (Fig. 3). As a result of this extra loop, subdomain I of domain 2 contains an extended β-sheet compared to the other β-sheets found in the SCAfet monomer.

Superposition of the SCAman dimer (made up of subunits A and B) onto the SCAfet monomer shows that in subdomain I of both domains 1 and 2 of SCAfet, there is a shorter loop, with four amino acid residues deleted, compared with the same regions of the SCAman monomers. It is clear from an examination of the 3D structure of SCAfet that the reason for these shorter loops is to allow for the presence of the 11-residue peptide link which connects domains 1 and 2 (Fig. 4).

SCAfet contains six potential carbohydrate recognition domains per monomer, it associates to form tetramers in solution so that the functional SCAfet molecule has a total of 24 binding sites. Although SCAfet readily agglutinates trypsin-treated rabbit erythrocytes, it is inactive towards human erythrocytes from any of the blood groups [1]. Hapten inhibition assays with SCAfet have shown that none of the mono-

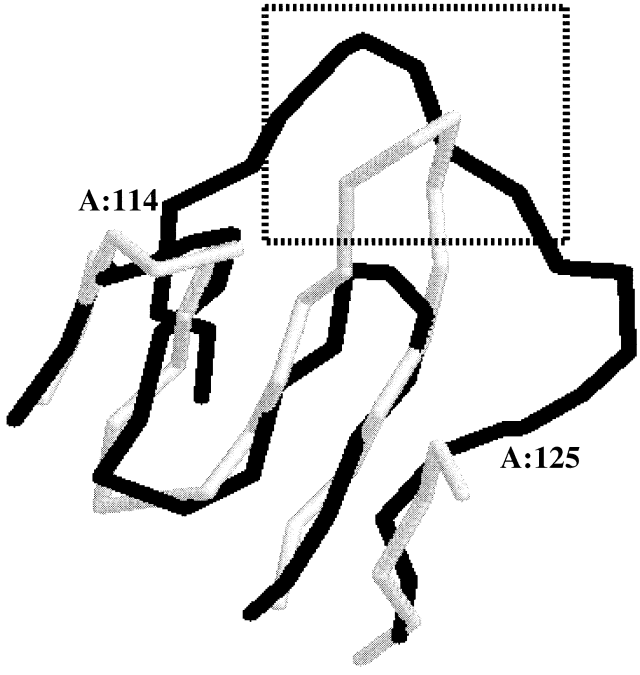

Fig. 4. Superposition of SCAman (light gray) onto SCAfet (black) in the region of subdomain I showing the shorter loop with the four-residue deletion for SCAfet compared to the corresponding loop in SCAman and its position relative to the linker peptide. The figure was produced with MOLSCRIPT [19].

saccharides or oligosaccharides tested had an inhibitory effect on the agglutination of rabbit erythrocytes [1]. However, assays with selected animal glycoproteins revealed that the agglutination activity of SCAfet can be inhibited by asialofetuin, thyroglobulin, ovomucoid and fetuin suggesting that SCAfet binds complex *N*-linked oligosaccharides. The lack of interaction of SCAfet with simple sugars appears to arise as a result of the replacement of key amino acid residues (Asn, Asp, Gln, Tyr) within the monosaccharide-binding pocket by hydrophobic residues. Molecular docking studies with SCAfet have indicated that only four out of the 24 putative monosaccharide-binding sites are able to accommodate mannose [1]. Analyses to determine the precise specificity of SCAfet are in progress and these studies will enable future structural work on carbohydrate complexes of this novel two-domain lectin with an unusually high number of potential saccharide-binding sites.

The coordinates and structure factors of SCAfet have been deposited in the PDB with accession code 1DLP.

Acknowledgements: We thank The Leverhulme Trust (F/754/A) and the Mizutani Foundation for Glycoscience for financial support for this research. X-ray data were collected on the Synchrotron Radiation Source, Daresbury Laboratory, U.K. Liverpool John Moores University provided general support. Dr. S.D. Wood is thanked for his help and for many useful discussions and Professor E.H. Evans for her encouragement. W.J.P. is a Research Director and E.J.M.V.D. is a Postdoctoral Fellow of the Fund for Scientific Research, Flanders.

References

[1] Wright, L.M., Van Damme, E.J.M., Barre, A., Allen, A.K., Van Leuven, F., Reynolds, C.D., Rouge, P. and Peumans, W.J. (1999) Biochem. J. 340, 299–308.
[2] Van Damme, E.J.M., Peumans, W.J., Pusztai, A. and Bardocz, S. (1998) Handbook of Plant Lectins: Properties and Biomedical Applications, Wiley, Chichester.
[3] Van Damme, E.J.M., Briké, F., Winter, H.C., Van Leuven, F., Goldstein, I.J. and Peumans, W.J. (1996) Eur. J. Biochem. 326, 419–427.
[4] Van Damme, E.J.M., Goossens, K., Smeets, K., Van Leuven, F., Verhaert, P. and Peumans, W.J. (1995) Plant Physiol. 107, 1147–1158.
[5] Van Damme, E.J.M., Smeets, K., Torrekens, S., Van Leuven, F., Goldstein, I.J. and Peumans, W.J. (1992) Eur. J. Biochem. 206, 413–420.
[6] Chandra, N.R., Ramachandraiah, G., Bachhowat, K., Dam, T.K., Surolia, A. and Vijayan, M. (1999) J. Mol. Biol. 285, 1157–1168.
[7] Wright, L.M., Reynolds, C.D., Rizkallah, P.J., Allen, A.K., Peumans, W.J., Van Damme, E.J.H. and Donovan, M.J. (1999) Protein Pept. Lett. 6, 253–258.
[8] Leslie, A.G.W. (1992) in: Int. CCP4 ESF-EACBM Newslett. Protein Crystallogr. Vol. 26, Daresbury Laboratory, Warrington.
[9] Collaborative Computational Project, Number 4 (1994) Acta Cryst. D50, 670–693.
[10] Navaza, J. (1994) Acta Cryst. A50, 157–166.
[11] Wood, S.D., Wright, L.M., Reynolds, C.D., Rizkallah, P.J., Allen, A.K., Peumans, W.J. and Van Damme, E.J.M. (1999) Acta Cryst. D55, 1264–1272.
[12] Matthews, B.W. (1968) J. Mol. Biol. 33, 491–497.
[13] Marshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Acta Cryst. D53, 240–255.
[14] Cowtan, K. (1994) in: Int. CCP4 ESF-EACBM Newslett. Protein Crystallogr. Vol. 31, Daresbury Laboratory, Warrington.
[15] Jones, J.A. and Kjeldgaard, M. (1993) O Version 6.1.2, The Manual, Uppsala University, Uppsala, Sweden.
[16] Hester, G., Kaku, H., Goldstein, I.J. and Wright, C.S. (1995) Nat. Struct. Biol. 2, 472–479.
[17] Chantalat, L., Wood, S.D., Rizkallah, P.J. and Reynolds, C.D. (1996) Acta Cryst. D52, 1146–1152.
[18] Sauerborn, M.K., Wright, L.M., Reynolds, C.D., Grossmann, G.J. and Rizkallah, P.J. (1999) J. Mol. Biol. 290, 185–199.
[19] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.