X-ray Crystal Structure of a Galactose-Specific C-Type Lectin Possessing a Novel Decameric Quaternary Structure[†,‡]

John R. Walker,[§,||] Bhushan Nagar,[§,⊥] N. Martin Young,[@] Tomoko Hirama,[@] and James M. Rini*[,§]

Departments of Molecular and Medical Genetics and Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada, and Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

Received October 18, 2003; Revised Manuscript Received January 27, 2004

ABSTRACT: Rattlesnake venom lectin (RSL) from the western diamondback rattlesnake (*Crotalus atrox*) is an oligomeric galactose-specific C-type lectin. The X-ray crystal structure of RSL, in complex with lactose and thiodigalactoside, at 2.2 and 2.3 Å resolution, respectively, reveals a decameric protein composed of two 5-fold symmetric pentamers arranged in a staggered, back-to-back orientation. Each monomer corresponds to a single canonical C-type lectin carbohydrate recognition domain devoid of accessory domains and is disulfide-bonded to a monomer in the other pentamer. The structure is the first example of that of a carbohydrate complex of a vertebrate galactose-specific C-type lectin. The 10 carbohydrate-binding sites, located on the rim of the decamer, suggest a role for multivalent interactions and a mechanism for RSL's ability to promote receptor cross-linking and cell aggregation.

C-Type lectins play roles in a number of biologically important processes, including, among others, lymphocyte homing, innate immunity, serum glycoprotein clearance, and organization of the extracellular matrix (*1*−*4*). Although often mosaic in nature, they are characterized by the presence of at least one canonical C-type lectin carbohydrate recognition domain (CRD).[1] The CRD itself is an ∼110-amino acid module that binds carbohydrate in a calcium-dependent fashion (*5*), and among members of the family, both mannose and galactose monosaccharide specificities have been observed. The ability to bind saccharide distinguishes the CRD from the evolutionarily related noncarbohydrate binding C-type lectin-like domains (CTLDs) (*6*).

Structural studies have shown that direct calcium ion−saccharide interactions are a hallmark of C-type lectin complexes (*7*). Moreover, in experiments aimed at probing the determinants of carbohydrate binding specificity, a series of mutant MBP-A proteins, engineered to bind galactose, have been generated (*8*−*10*). These studies defined the position and orientation of the bound mannose and galactose moieties in these complexes and led to insight into the role played by the calcium and carbohydrate-binding residues as determinants of specificity. In addition, the X-ray crystal structure of the galactose complex of the tunicate lectin from *Polyandrocarpa misakiensis* has been determined (*11*). This lectin binds galactose in an orientation that differs by 180° from that seen in the MBP-A's engineered to bind galactose, a result that led to a rationalization (*11*) for the observed correlation between hydrogen bond donor and acceptor distribution and carbohydrate binding specificity shown by members of the C-type lectin family (*8*).

Many lectins are oligomers containing more than one CRD, and in addition to their monosaccharide binding specificities, the way in which these CRDs are spatially arranged is an important determinant of specificity (*2*, *7*, *12*). The need for multivalent interactions stems directly from the fact that the affinities of the CRDs for small saccharides are weak (K_d's in the low millimolar range) and that in most cases they recognize commonly occurring monosaccharide residues such as mannose and galactose. The organization of CRDs into clusters or oligomers can be achieved via contacts between the CRD subunits themselves and/or by interactions of accessory domains such as the coiled-coil domains found in the C-type lectins, human tetranectin, MBP-A, hSP-D, and the recently described DC-SIGN and DC-SIGNR, cell surface receptors implicated in dendritic cell−T-cell interactions and the HIV infection process (*13*−*17*). Other lectins, such as the macrophage mannose receptor (also a C-type lectin), are inherently multivalent as they contain multiple copies of the CRD in a single polypeptide chain, a property shared by members of the galectin family, P-type lectins, and a number of plant lectins (*18*).

Rattlesnake venom lectin (RSL) is an oligomer of disulfide cross-linked dimers (*19*, *20*) isolated from the venom of *Crotalus atrox*. It is a galactose-specific, soluble C-type lectin

[†] This work was supported by the Medical Research Council of Canada (Grant MT-12458 and a studentship to J.R.W.).

[‡] The coordinates and structure factors for the RSL−lactose (1JZN and r1jznsf.ent, respectively) and RSL−TDG (1MUQ and r1muqsf.ent, respectively) complexes have been deposited in the Protein Data Bank.

* To whom correspondence should be addressed. E-mail: james.rini@utoronto.ca.

[§] Departments of Molecular and Medical Genetics and Biochemistry, University of Toronto.

[||] Present address: C. H. Best Institute, Toronto, Ontario M5G 1L6, Canada.

[⊥] Present address: Department of Molecular and Cell Biology, 401 Barker MC 3202, University of California, Berkeley, CA 94720-3202.

[@] National Research Council of Canada.

[1] Abbreviations: APL, *A. piscivorus* lectin; BFL-1 and -2, *B. fasciatus* lectins 1 and 2, respectively; BML, *B. multicinctus* lectin; CRD, carbohydrate recognition domain; CTLD, C-type lectin-like; HHL, Himehabu lectin; HL-1, hepatic lectin 1; LmsL, *L. muta stenophyrs* lectin; MBP-A, mannose binding protein A; PAL, puff adder lectin; RSL, rattlesnake venom lectin; TDG, thiodigalactoside; TSL, *T. stejnegeri* lectin.

10.1021/bi035871a CCC: $27.50
Published on Web 03/11/2004

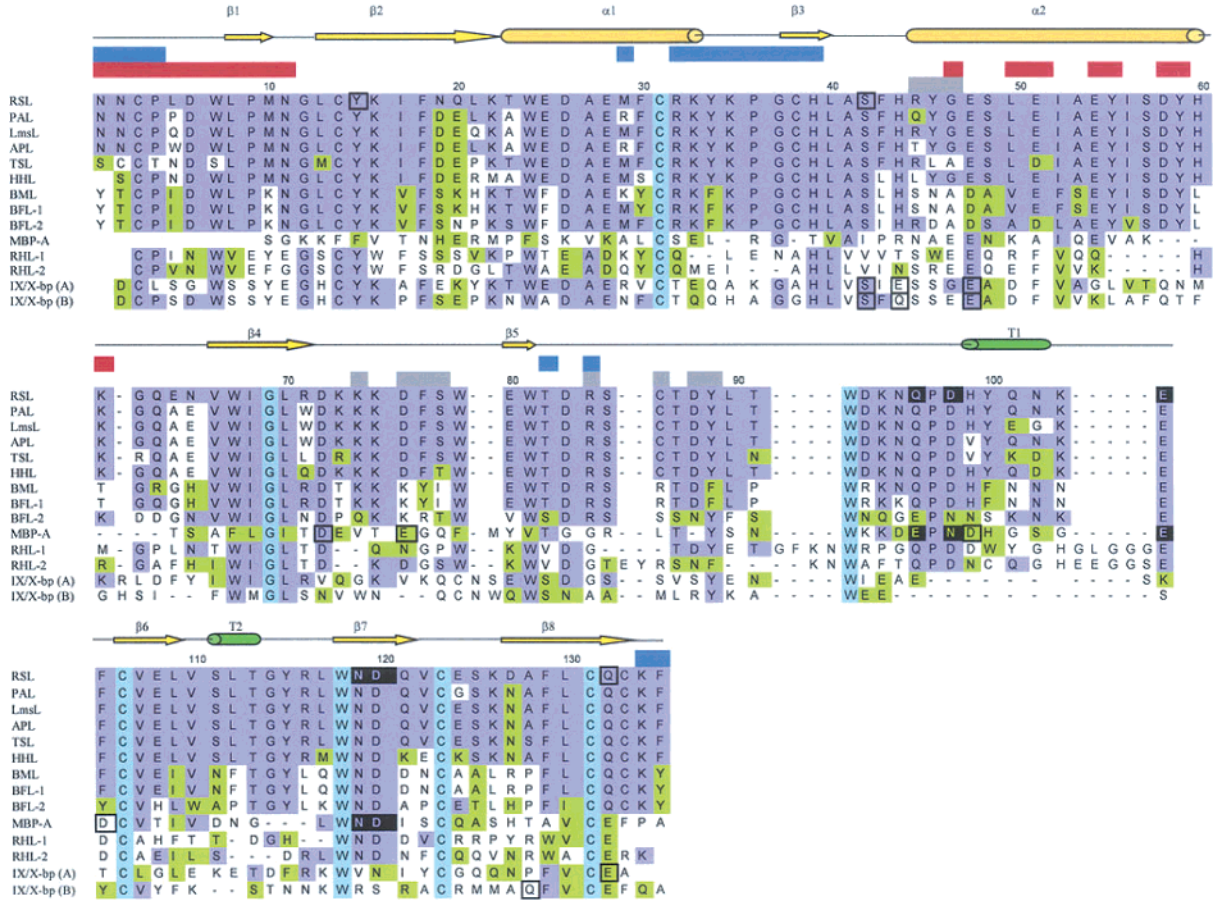

FIGURE 1: Sequence alignment of selected CTLD/C-type lectins with RSL. Secondary structural elements are represented with yellow arrows for β-strands, yellow cylinders for α-helices, and green cylinders for 3_{10}-helices. Blue shading above the sequence indicates the residues in RSL that are part of the basic face of the intrapentamer interface, red shading residues making up the acidic intrapentamer interface, and gray shading residues involved in interpentameric contacts. Residues identical among all the sequences that are listed have a blue background, while residues identical to that of RSL have a purple background. Residues similar to those of RSL have a green background. Residues coordinating the calcium at the carbohydrate-binding site have a black background, while residues involved in coordinating a sodium atom in the RSL structure, and additional calcium atoms in MBP-A and IX/X-bp are boxed. Abbreviations: RSL, rattlesnake venom lectin; PAL, puff adder lectin; LmsL, *L. muta stenophyrs* lectin; APL, *A. piscivorus* lectin; TSL, *T. stenjnegeri* lectin; HHL, Himehabu lectin; BML, *Bu. multicinctus* lectin; BFL-1, *Bu. fasciatus* lectin 1; BFL-2, *Bu. fasciatus* lectin 2; MBP-A, mannose binding protein A; RHL-1 and -2, rat hepatic lectin 1 and 2, respectively; IX/X-bp A and B, factor IX/X binding protein from *T. flavoviridis*.

devoid of accessory domains, and it is one of a family of C-type lectins isolated from viperid and elapid venoms that includes thrombolectin (from *Bothrops atrox*), cottonmouth lectin (from *Ancistrodon piscivorus*), copperhead lectin (from *Agistrodon contortrix contortrix*), *B. jaracara* lectin (from *Bothrops jaracara*), *L. muta stenophyrs* lectin (LmsL, from *Lachesis muta stenophyrs*), Himehabu lectin (HHL, from *Trimeresurus okinavensis*), *T. stejnegeri* lectin (TSL, from *Trimeresurus stejnegeri*), puff adder lectin (PAL, from *Bitis arietans*), *A. piscivorus* lectin (APL, from *Agistrodon piscivorus piscivorus*), *B. fasciatus* lectins 1 and 2 (BFL-1 and BFL-2, respectively, from *Bungarus fasciatus*), and *B. multicinctus* lectin (BML, from *Bungarus multicinctus*) (*20*−*27*). Of these lectins, eight have been completely sequenced and shown to be very similar to RSL (68−92% identical; see Figure 1). These venoms also contain CTLD proteins, examples of which include Botrocetin (from *B. jaracara*) and Habu IX/X bp (from *T. okinavensis*), proteins found to bind coagulation factors (*28*). Although their physiological roles have not been fully characterized, both the C-type lectins and the CTLD proteins found in venom are thought to play roles in disrupting the blood coagulation pathway (*28*). RSL, for example, is a potent inducer of both erythrocyte agglutination (*19*) and platelet aggregation (*29*).

To further our understanding of the structural basis for monosaccharide binding specificity and the role played by multivalency among members of the C-type lectin family, we have determined the X-ray crystal structure of RSL in complex with lactose, the first example of a carbohydrate complex of a galactose-specific vertebrate C-type lectin. Although the galactose moiety of the lactose makes the characteristic calcium ion interactions typical of a C-type lectin, both its orientation and the interactions it makes with the protein provide new insights into the basis for galactose recognition by members of the family. Moreover, the structure reveals a novel decameric oligomer composed of two 5-fold symmetric pentamers, a structural arrangement suggesting the possibility of making multivalent interactions with opposing cell surfaces.

MATERIALS AND METHODS

RSL Purification. RSL was purified from western diamondback rattlesnake (*C. atrox*) venom (Sigma) essentially as described previously (*20*). Up to 5 g of lyophilized venom was dissolved on ice with 30 mL of Tyrode buffer (137 mM

Table 1: Data Collection, Phasing, and Refinement Statistics

Data Collection							
data set	unit cell (Å)[a]			resolution range (Å)	no. of unique reflections	completeness (%) (total/last shell)	R_{sym} (%)[b] (total/last shell)
RSL−TDG1	117.2	153.1	95.8	40.0−2.8	21147	98.0/79.9	9.8/30.2
RSL−TDG2	110.4	151.2	95.0	40.0−2.2	33477	94.8/82.9	10.9/40.2
RSL−Ho	119.6	156.3	98.0	30.0−2.8	23196	95.2/95.9	12.9/33.1
RSL−Gd	117.1	152.7	96.6	18.0−4.0	7612	99.1/99.7	17.8/29.7
RSL−Lac	63.3	178.5	137.8	40.0−2.2	38575	96.2/79.8	6.3/15.6

Phasing Statistics for the RSL−TDG1 Complex	holmium	gadolinium
no. of sites	5	5
resolution (Å)[c]	3.0/3.0	5.0/4.0
R_{merge}[c,d] (%)	11.7/4.0	14.2/7.6
figure of merit[c]	0.19/0.28	0.20/0.26
phasing power[c,e]	1.94/2.23	1.29/1.70

Refinement Statistics	RSL−TDG2	RSL−Lac
resolution range (Å)	35.0−2.3	40.0−2.2
no. of reflections in working set	30154	37405
no. of reflections in test set	3323	1170
R_{cryst}[f]	0.21	0.20
R_{free}[g]	0.26	0.25
average *B* values	48.6	35.8
rmsd bond lengths (Å)	0.007	0.010
rmsd bond angles (°)	1.4	1.5
rmsd dihedral angles (°)	23.5	24.0
rmsd improper torsion (°)	0.76	0.84

[a] $R_{sym} = \sum_h\sum_i|I_i(h) - \langle I(h)\rangle|/\sum_h\sum_i I_i(h)$, where $I_i(h)$ is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$. [b] Both crystal forms in space group $C222_1$. [c] The isomorphous and anomalous values are given as isomorphous/anomalous. [d] $R_{merge} = \sum_r(\sum_i|I_{i,r} - \langle I_r\rangle|)/\sum_r[n(r)\langle I_r\rangle]$, where n is the number of unique reflections, $I_{i,r}$ is the ith measurement of the rth reflection, and $n(r)$ is the number of measurements of the rth reflection in the data set. [e] Phasing power $= [\sum(F_H)^2/\sum E^2]^{1/2}$, where $\sum E^2 = \sum(|F_{PH}|_{obs} - |F_{PH}|_{calc})^2$. [f] $R_{cryst} = \sum_h||F(h)_{obs}| - |F(h)_{calc}||/\sum_h|F(h)_{obs}|$ for reflections in the working set. [g] $R_{free} = \sum_h||F(h)_{obs}| - |F(h)_{calc}||/\sum_h|F(h)_{obs}|$ for reflections in the test set.

NaCl, 12 mM $NaHCO_3$, 2.7 mM KCl, 0.3 mM $NaH_2PO_4\cdot H_2O$, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, and 3 mM NaN_3), centrifuged, and then loaded onto a Sepharose 4B (Pharmacia) column (100 mL). Following washing with Tyrode buffer, RSL was eluted with Tyrode buffer containing 100 mM lactose and dialyzed against a solution containing 10 mM Tris (pH 8.0) and 1 mM $CaCl_2$. The purified yield was typically 7 mg of RSL/g of venom.

Surface Plasmon Resonance Analysis. The relative inhibitory properties of sugar ligands were determined by inhibition analyses on a BIACORE 1000 surface plasmon resonance instrument (Biacore AB), at 25 °C. The surface of a CM5 research-grade sensor chip was modified with approximately 500 RU of lactose-substituted bovine serum albumin, according to the manufacturer's recommended amine coupling procedure. Samples (150 μL) of the lectin [6.25 nM in 10 mM HEPES (pH 7.4), 150 mM NaCl, and 0.005% P20] mixed with various concentrations of ligand were injected at a flow rate of 20 μL/min, and followed by elution with the same buffer. The inhibition plots were constructed from the bound amounts determined immediately after the buffer wash was begun. The sensorgrams showed linear binding, but dissociation was very slow because of the multivalency of the lectin. The surface was regenerated by washing for 30 s with 10 mM lactose in the running buffer.

Crystallization and Data Collection. RSL−thiodigalactoside (RSL−TDG1) crystals were grown at 4 °C using the hanging drop vapor diffusion method by mixing 2 μL of RSL (5 mg/mL), in buffer containing 10 mM Tris (pH 8.0), 1 mM $CaCl_2$, and 100 mM thiodigalactoside, with 0.5 μL of Cymal-3 at 10 times the CMC (Hampton Research), and 2.5 μL of well solution containing 5 M sodium formate, 100 mM HEPES (pH 7.0), and 10% glycerol. The crystals were allowed to equilibrate to room temperature before being flash-frozen in the N_2 gas stream. RSL−TDG2 crystals were prepared the same way as RSL−TDG1 crystals except that over the 1 mL well solution was placed 500 μL of a 50:50 mix of paraffin and silicon oil. After 1 month, the crystallization plate was transferred to room temperature and the crystals were allowed to equilibrate to this temperature. Five days later, the coverslip containing the crystals was placed over a well solution containing 6 M sodium formate and 20% glycerol, and the crystals were allowed to equilibrate over this solution for 1 day before being flash-frozen in the N_2 gas stream. RSL−Ho crystals were grown at room temperature by mixing RSL (8 mg/mL) in buffer containing 10 mM Tris (pH 8.0), 1 mM $CaCl_2$, 100 mM thiodigalactoside, and 25 mM $HoCl_3$ with an equal volume of well solution containing 4 M sodium formate, 10 mM $HoCl_3$, and 100 mM HEPES (pH 7.0). These crystals were cryoprotected in well solution to which 30% glycerol and 50 mM thiodigalactoside had been added. RSL−Gd crystals were prepared by soaking RSL−TDG1 crystals in a droplet containing $^9/_{10}$ of their well solution and $^1/_{10}$ of 75 mM $GdCl_3$. RSL−lactose crystals were grown at room temperature by equilibrating a hanging drop composed of equal volumes of RSL (5−7.5 mg/mL) in buffer containing 10 mM Tris (pH 8.0), 1 mM $CaCl_2$, and 100 mM lactose with well solution that was made up of 1.3−1.5 M NaCl, sodium acetate (pH 4.6), and 40 mM $CaCl_2$. RSL−lactose crystals were cryoprotected in a solution containing 25% ethylene glycol, 1.25 M NaCl, 20 mM $CaCl_2$, 50 mM sodium acetate (pH 5.0),

and 50 mM lactose. All crystals were frozen in loops in the N_2 cooling stream at 100 K.

All diffraction data except those of RSL−Gd crystals were collected at the Cornell High Energy Synchrotron Source (CHESS). RSL−Ho data were collected at beamline A1 using the Princeton 1K CCD. RSL−lactose and RSL−TDG data were collected using the Quantum-4 CCD at beamline F2. RSL−Gd data were collected using a 345 mm MAR Research image plate detector at the Cu Kα wavelength using a Rigaku FR-C generator. DENZO and Scalepack were used to process the data (*30*).

Structure Determination and Refinement. The crystal structure of RSL was determined in a complex with thiodigalactoside and lactose in two different $C222_1$ crystal forms (Table 1). Holmium and gadolinium derivatives were used to determine phases for the RSL−TDG1 complex by the MIRAS method. Heavy atom sites were determined using the program SOLVE (*31*). Solvent flattening, heavy atom positional refinement, and 5-fold noncrystallographic symmetry averaging were performed using PHASES (*32*). A further improvement of the phases was achieved with SHARP (*33*) and DM (*34*). An initial chain tracing, made using O (*35*), allowed for placement of the lithostathine model (*36*), which is 33% identical to that of RSL, into the electron density. The sequence of lithostathine was altered to that of RSL and the model adjusted to fit the density. Knowledge of the 5-fold symmetry operators allowed the other four copies of the monomer in the asymmetric unit to be positioned. This preliminary model of the pentamer was used to determine the structure of the RSL−lactose complex by molecular replacement using CNS (*37*). The resolution of the RSL−TDG structure was extended to 2.3 Å by using higher-resolution data from the TDG2 crystals. Both the RSL−TDG2 (2.3 Å) and RSL−lactose (2.2 Å) structures were refined independently using a combination of manual rebuilding using O and the simulated annealing and energy minimization protocols found in CNS. NCS restraints were used only in the initial stages of refinement.

RESULTS AND DISCUSSION

Overall Structure. The lactose and thiodigalactoside complexes, refined at 2.2 and 2.3 Å resolution, respectively, are structurally very similar, and unless indicated otherwise, reference to the structure pertains to that of the lactose complex. All 135 residues have been modeled in each monomer of the lactose complex, while residues 2−135 are observed in each monomer of the thiodigalactoside complex. As shown in Figure 2, RSL was found to be a decamer formed by two pseudo-5-fold symmetric pentamers. The pentamers themselves are related by a 2-fold rotation axis that in both the lactose and thiodigalactoside complex crystals is coincident with the crystallographic 2-fold rotation axis running along the *b*-axis. The 2-fold rotation axis is approximately perpendicular to the pseudo-5-fold rotation axis, and it is positioned such that the two pentamers of the decamer are oriented in a staggered, back-to-back fashion. Since each monomer of the pentamer is disulfide-bonded to a monomer in the 2-fold related pentamer, the pentamers are linked to each other by five disulfide bonds. The decamer has an overall thickness of 64 Å and a diameter of 104 Å. Consistent with the decamer seen in these two crystal forms,

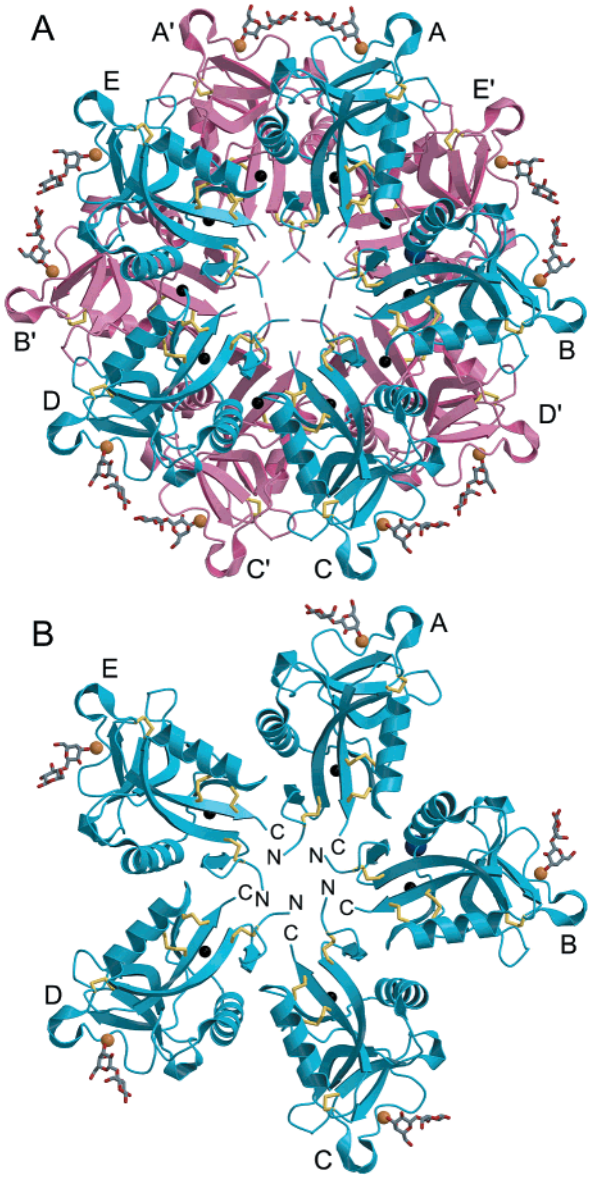

FIGURE 2: Ribbon diagram of the decamer and pentamer. (A) Ribbon diagram of the decamer viewed down the pseudo-5-fold rotation axis. The two pentamers are colored pink and blue. The lactose molecules are represented by stick figures with carbon atoms colored gray and oxygen atoms red. Calcium, sodium, and chloride ions are represented with orange, small black, and large black spheres, respectively. Disulfide bonds are shown in yellow. (B) Ribbon diagram of a single pentamer in the same orientation as in panel A. Molecular images shown here and in subsequent figures were generated using SPOCK (*46*), Raster3D (*47*), and Molscript (*48*).

RSL was found to have a molecular mass of ~150 kDa as determined by gel filtration in the absence of Ca^{2+} (*19*).

The five monomers of both pentamers are oriented such that their N- and C-termini are near the 5-fold rotation axis and the carbohydrate-binding sites are toward the rim of the disk (Figure 2). Each monomer possesses the canonical C-type lectin fold and as shown in Figure 3 is comprised of two four-stranded β-sheets flanked by two α-helices, two short 3_{10}-helices, and a considerable amount of coil structure. The monomers show an average rmsd of 1.5 Å on Cα positions compared with that of MBP-A (PDB entry 2MSB). The three intrachain disulfide bonds (C3−C14, C31−C131, and C106−C123) are found in the corresponding positions

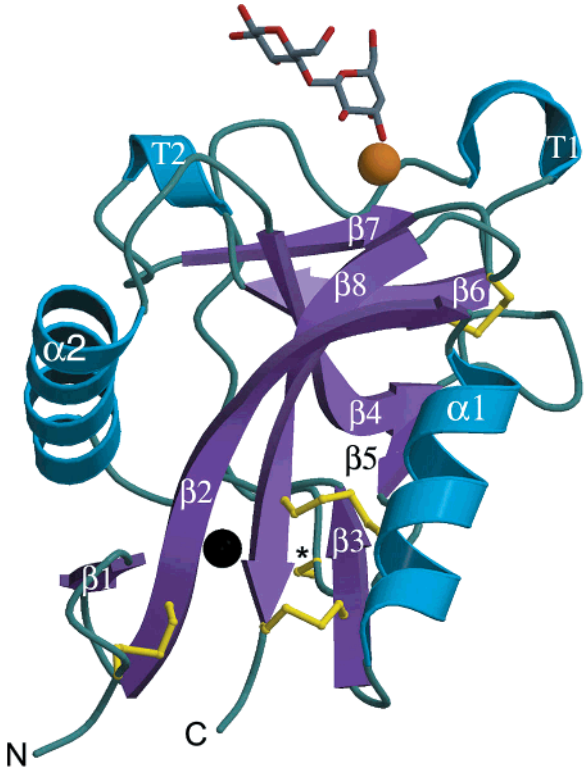

FIGURE 3: Ribbon diagram of the RSL monomer. Helices are colored blue and β-strands magenta, and cysteine side chains are shown in stick representation in yellow. The asterisk labels C86, the residue involved in formation of the disulfide-linked dimers. Calcium and sodium ions are represented as orange and black spheres, respectively, and lactose is shown in stick representation with carbon and oxygen atoms colored gray and red, respectively.

in the structures of lithostathine, IX/X-bp, CD94, and the hepatic lectin (HL-1), but neither the intrachain disulfide bond (C38−C133) nor the disulfide bond between monomers (C86−C86) has been previously observed. The observed disulfide bonding pattern is in complete agreement with that determined by peptide analysis (*19*).

RSL is a so-called long form CRD/CTLDs, molecules characterized by an approximately 10-residue N-terminal extension that typically forms a disulfide-cross-linked (C3−C14 in RSL) hairpin (*6*). These extensions are involved in subunit interactions in a number of CTLDs, and in RSL, these residues make important contacts between monomers of the pentamer.

RSL contains a roughly spherical cavity, with a diameter of ~40 Å, between the two pentamers. The cavity has a surface area of ~10000 Å^2 (the exterior surface of the decamer has an area of ~50000 Å^2) and is accessible to the bulk solvent through the central opening located along the 5-fold axis, as well as through channels between the lobes of the pentamers. The central opening, formed by the two N-terminal residues of the five monomers, has a diameter of ~8 Å, while the channels are approximately 6 Å in diameter. The higher *B*-factors of the N-terminal residues suggest that the diameter of the central opening may fluctuate, and indeed, in the thiodigalactoside complex, the lack of interpretable electron density prevented us from modeling the first residue in each chain. Fourteen residues from each of the monomers, positioned throughout the RSL sequence, contribute to the solvent accessible surface of the internal cavity.

The X-ray crystal structures of a number of CRD/CTLDs have revealed calcium ion-binding sites, in addition to the one involved in carbohydrate binding (discussed below), thought to be important in stabilizing these proteins. Although RSL does not possess additional calcium ion-binding sites, it does bind a sodium ion (see Figures 2 and 3) where both IX/X bp and HL-1 bind calcium ions. This site is exposed to the internal cavity of the RSL decamer, and the ion is coordinated in a distorted octahedral geometry by residues Y15, S42, and Q132 and a water molecule. The designation of this ion as a sodium ion is based on several factors, including its octahedral coordination, an average ligand−ion bond length of 2.5 Å, site valency, electron density, and the presence of only oxygen atoms within a 3.4 Å sphere (*38*). In MBP-A, calcium ion-binding site 1 (site 2 is the carbohydrate-binding site) is thought to play a role in structuring residues adjacent to the carbohydrate-binding site (*39*). In RSL, H99 and F105 replace two of the key calcium ion-coordinating residues found in site 1 of MBP-A. These residues in conjunction with D72, which makes a relatively buried salt bridge with H99, are found in a number of the venom lectins, and structurally, they appear to represent an alternative means by which the loop between β4 and β5, the carbohydrate binding site, and β7 can be stabilized.

Electrostatic surface calculations (Figure 4) show that the decamer has an asymmetric charge distribution. The flat outer surface is predominantly neutral to positive in charge, while the outer rim and the interior of the decamer are predominantly negative, even with the calcium ions included in the calculation.

Intrapentameric Interface. RSL is the first example of a C-type lectin exhibiting pseudo-5-fold rotational symmetry. It is devoid of accessory domains, and therefore, all interactions stabilizing the decamer involve residues of the CRD. Within the pentamer, each monomer has a buried surface area of ~1250 Å^2 formed by contacts with its two adjacent monomers. The buried surface is V-shaped and extends from one side of the monomer to the other. Residues 1−10 (i.e., the N-terminal extension, including β1), 47, 50, 51, 54, 55, 58, 59, and 61 (most of α2), which include one basic and four acidic amino acids, constitute the more acidic face of the V-shaped surface. The more basic face, which includes one acidic and six basic amino acids, is made up of residues 1−4, 29, 32, 33, 35−37, and 39 from the carboxyl end of α1, an intervening type II β-turn, a portion of β3 (residues 82 and 84), and the two C-terminal residues.

Interactions between the acidic face of one monomer and the basic face of an adjacent one constitute the interface between monomers within the pentamer. As shown in Figure 5, four salt bridges (E51−K134, E54−H39, E54−R84, and D58−R32) are made across this interface. Several direct and water-mediated hydrogen bonds also bridge the interface. In addition, important apolar contacts (>50 Å^2 per residue) involving residues N1, L5, P9, E54, D58, and Y59 on the acidic face and K33, K35, P36, R84, K134, and F135 on the basic face are made. Shape complementarity can be seen in the wedgelike nature of the interface, as well as by the protuberance of P36 into a pocket between β1 and α2, where it interacts with L8, Y59, and Y55.

Interpentamer Interface. As discussed above, each monomer of the pentamer is disulfide-bonded (C86−C86) to a monomer of the opposing pentamer (Figure 4C,D). For

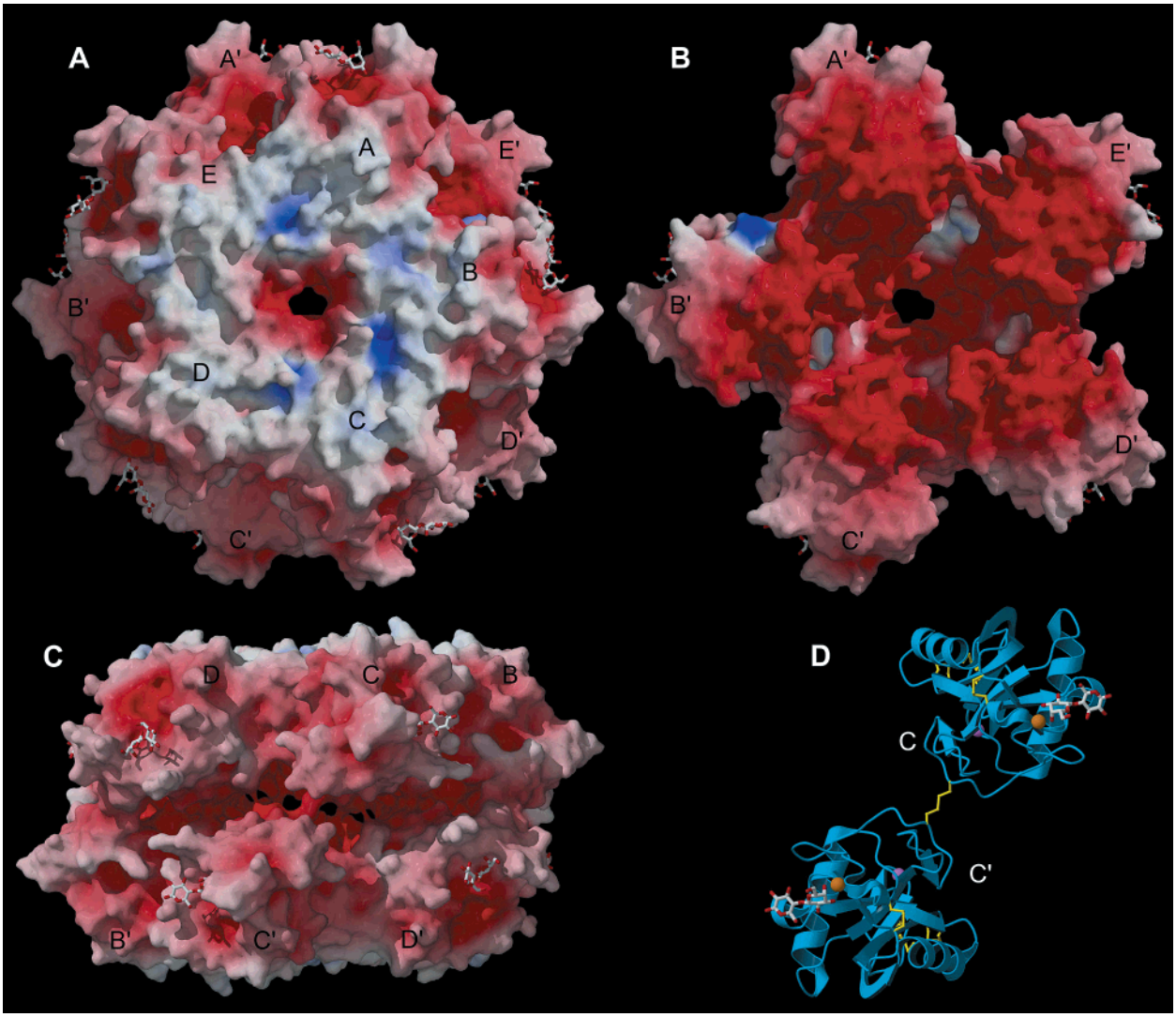

FIGURE 4: Electrostatic surface potential images of RSL. (A) The decamer viewed down the pseudo-5-fold rotation axis. (B) A single pentamer viewed down the pseudo-5-fold axis from the interior of the molecule. (C) The decamer viewed down the crystallographic 2-fold rotation axis. (D) A ribbon diagram of the C−C′ dimer in the same orientation as in panel C (slightly enlarged). Electrostatic surface potentials range from −20 (red) to 20 kT (blue).

discussion purposes, the monomers in one pentamer have been labeled from A to E and their crystallographically related mates labeled from A′ to E′ (see Figure 2). The specific pairing of monomers (A−E′, B−D′, C−C′, D−B′, and E−A′) stems from the position of the 2-fold rotation axis relating the pentamers which intersects the C−C′ disulfide bond. Besides this covalent bond, the monomers disulfide cross-linked to each other also interact across the pentamer−pentamer interface, most notably through hydrogen bonds (involving residues K74, D76, F77, and S78) and a salt bridge (residues K74 and D76). Interactions across the pentamer−pentamer interface are also made between monomers that are not disulfide cross-linked to each other (A−A′, B−E′, C−D′, D−C′, and E−B′; residues 45−47, 88, and 89). In addition, a chloride ion bridges the main chain nitrogen atom of Y46 of monomers B and E′ (see Figure 2) in the lactose complex. In total, the surface area buried on a given monomer, by interactions with monomers in the opposing pentamer, is relatively small (330 Å^2/monomer, averaged over 10 monomers). Were it not for the covalent bonds between the pentamers, the interface would have to be described as weak. Nevertheless, the disulfide bonds can be readily reduced by 10 mM dithiothreitol at room temperature, and this does not disrupt the decameric structure as judged by size-exclusion chromatography in EDTA (data not shown).

Carbohydrate- and Calcium-Binding Site. All 10 carbohydrate-binding sites of the RSL decamer are occupied by lactose [β-D-galactopyranosyl-(1,4)-D-glucopyranose]. There is good electron density for both the galactose and glucose moieties, although the average *B*-factor for glucose (55.3 Å^2) is higher than that of galactose (43.3 Å^2). The galactose moiety is found to interact directly with a bound calcium ion, a hallmark of the C-type lectin−carbohydrate interaction. The calcium ion itself is coordinated by eight ligands (Figure 6A) with the distorted pentagonal bipyramidal geometry seen in other C-type lectin−carbohydrate complexes. In addition to the galactose 3- and 4-OH groups, protein residues Q96, D98, E104, N119, and D120 are found to coordinate the calcium ion.

In addition to their calcium coordinating roles, Q96, D98, N119, and E104 also make direct hydrogen bonds to the galactose 3- and 4-OH groups (Figure 6A and Table 2). In subunit D, N119 and Q121 make additional hydrogen bonds

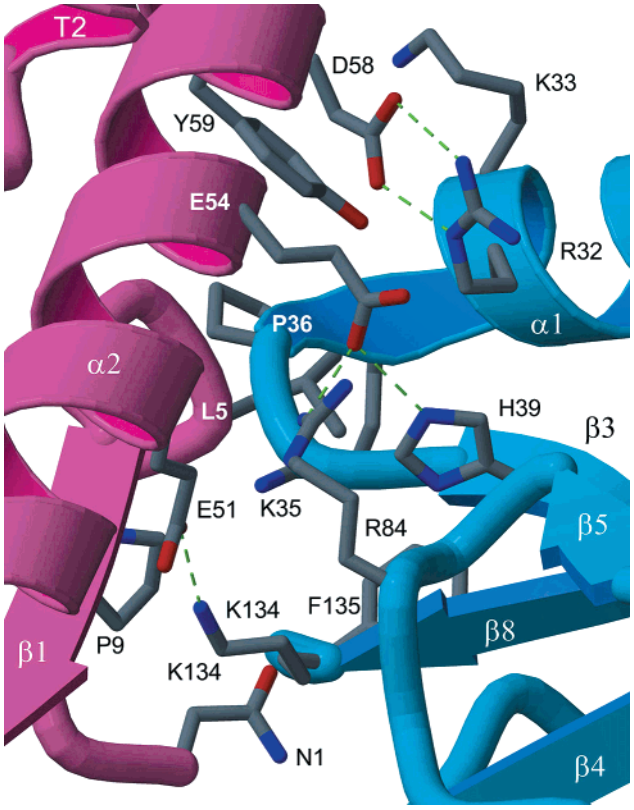

FIGURE 5: Salt bridges and apolar contacts in the intrapentamer interface. Amino acid side chains involved in ion pairs and/or those which contribute to more than 50 Å^2 of the buried surface area between monomer A (pink ribbon) and monomer E (blue ribbon) are shown in stick representation. Carbon, oxygen, and nitrogen atoms are colored gray, red, and blue, respectively. Hydrogen bonds are represented with green dashed lines.

with the galactose 2-OH group, while in subunits A, B, and E, a water molecule mediates the interaction between Q121 and the 2-OH group. Water-mediated hydrogen bonds from the protein and the bound galactose help to define the position of the glucose moiety. No direct contacts are observed between the glucose moiety and the monomer to which it is bound, an observation consistent with the ability of RSL to bind galactose alone. Crystallographic contacts involving glucose at carbohydrate-binding sites D and E result in a change in the observed glycosidic bond angles, but in both cases, the disaccharide remains in the low-energy *syn* conformation (*40*).

Determinants of Carbohydrate Binding Specificity. The X-ray crystal structures of a number of C-type lectin−carbohydrate complexes have shown that there are two general orientations for the monosaccharide bound to the primary calcium ion. In orientation I, typified by that seen in the MBP-A complexes, the monosaccharide ring is oriented such that the 3-OH group of the bound mannose moiety coordinates the calcium ion in the position that allows it to simultaneously hydrogen bond to Glu185 and Asn187 (MBP-A numbering). In orientation II, seen here in this RSL−lactose complex, the 4-OH group of the galactose moiety coordinates the calcium ion at the equivalent position (Figure 6A). In effect, the sugar rings in the two orientations differ by an ~180° rotation, a result of "swapping" the relative position of the 3- and 4-OH groups of the monosaccharide coordinating the calcium ion. Recent analysis has shown that the orientation observed correlates with the

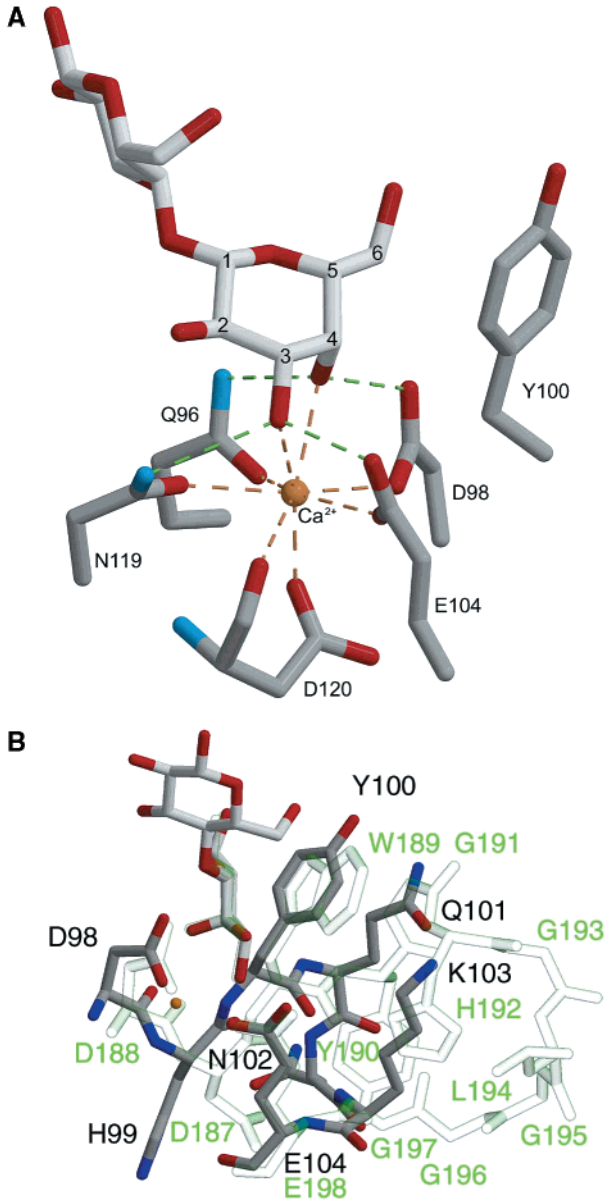

FIGURE 6: Carbohydrate-binding site. (A) Stick figure representation of the carbohydrate-binding site with lactose, calcium, and critical protein residues or side chains depicted. Bonds contributing to the coordination of calcium (yellow sphere) from protein residues and the O3 and O4 hydroxyl groups of galactose are shown as yellow dashed lines. Hydrogen bonds between galactose and Q96, D98, E104, and N119 are depicted as green dashed lines. The side chain of Y100 which makes packing contacts with galactose is also depicted. Carbon, oxygen, and nitrogen atoms are colored gray, red, and blue, respectively. Numerals (1−6) label the carbon atoms of the galactose moiety of the bound lactose. For clarity, water-mediated hydrogen bonds have been omitted, including the interaction involving Q121. (B) Comparison between the carbohydrate-binding site of chain A of the RSL structure and that of the galactose-binding mutant of MBP-A (PDB entry 1AFA). The two carbohydrate-binding sites have been aligned by superimposing the galactose moieties of each structure. The calcium atom of RSL is depicted as a yellow sphere. The stick figure of residues D98−E104 of RSL is colored as in panel A, while residues D187−E198 (including the glycine-rich loop) are colored a semitransparent green. Residues of RSL are labeled in black, while residues of 1AFA are labeled in green.

arrangement of hydrogen bond donors and acceptors that interact with the calcium-coordinating hydroxyl groups of the bound monosaccharide (*11*). MBP-A possesses an

Table 2: Galactose−RSL Hydrogen Bond Distances (Å) in the Lactose Complex[a]

	A			B			C			D			E		
Gal O2	water53	O	2.5	N119	ND2	3.0				N119	ND2	2.9	water216	O	2.8
	water204	O	3.0	water349	O	2.9				water396	O	2.6	water301	O	2.8
Gal O3	N119	ND2	3.2	N119	ND2	3.0	N119	ND2	3.0	N119	ND2	3.2	N119	ND2	3.0
	E104	OE1	2.5	E104	OE1	2.5	E104	OE1	2.8	E104	OE1	2.6	E104	OE1	2.5
Gal O4	Q96	NE2	3.0	Q96	NE2	2.9	Q96	NE2	3.1	Q96	NE2	2.7	Q96	NE2	2.9
	D98	OD2	2.6	D98	OD2	2.6	D98	OD2	2.7	D98	OD2	2.8	D98	OD2	2.6
Gal O5	water320	O	3.0												
Gal O6	water320	O	2.8	water513	O	3.1				water488	O	2.7			
	water471	O	2.7												

[a] Atoms that make hydrogen bonds with the galactose moiety in the binding sites of each of the five monomers (labeled A−E) in the asymmetric unit of the RSL−lactose complex.

acceptor−donor pair at positions 185 and 187 (Glu185 and Asn187, respectively) and binds mannose in orientation I, while RSL and human HL-1 possess donor−acceptor pairs (Gln96 and Asp98 in RSL and Gln239 and Asp241 in HL-1, respectively) at the equivalent positions and bind galactose in orientation II.

In addition to probing the role played by the hydrogen bond geometry, work aimed at engineering a mutant MBP-A to mimic the galactose binding specificity of HL-1 has defined the importance of Trp189 (HL-1 numbering) and a glycine-rich loop that helps to position it (*10*). Tyr100 in RSL corresponds to Trp189 in HL-1, and as shown in panels A and B of Figure 6, it makes a stacking interaction with the apolar face of the galactose moiety. These stacking interactions are typical of galactose-binding lectins of many different types, and they undoubtedly play an important role in defining specificity (*12*). Although RSL lacks the glycine-rich loop (Gly191−Gly197, HL-1 numbering) found to be important in defining the orientation of Trp189 in HL-1, the 3_{10}-helix from H99−N102 (RSL numbering), which packs against Y100 in RSL, may play a similar role (Figure 6B). The importance of both the arrangement of the hydrogen bonding residues and the galactose ring stacking interaction is further supported by the X-ray crystal structure of the tunicate lectin−galactose complex (*11*). Notably, this lectin binds galactose in orientation I, an observation that presumably reflects the fact that the positions of both the hydrogen bonding residues and the stacking aromatic residue are "reversed" relative to that seen in RSL or the galactose-binding MBP-A mutant.

As shown in Figure 1, all of the related snake venom lectins except for BFL-2 possess the Gln96 and Asp98 pair as well as a tyrosine or phenylalanine at the equivalent of position 100. As such, they would all be expected to show galactose binding specificity. In fact, BFL-2 is the only example of a venom C-type lectin that is likely to bind mannose (*27*).

Various cell surface and soluble glycoconjugates contain 2-acetamido-2-deoxygalactose (GalNAc), and like HHL, RSL binds both galactose and GalNAc. Consistent with this specificity is their ability to agglutinate red blood cells of the A and B blood group types (*23*), cells whose blood group substances contain terminal GalNAc and galactose moieties, respectively. Surface plasmon resonance analysis (Figure 7) shows that the K_I for GalNAc inhibition of RSL binding to a lactose−BSA-coated gold chip is 100 μM, while the K_I for lactose is 80 μM. As shown in this RSL−lactose complex, the 2-OH group of the galactose moiety makes a

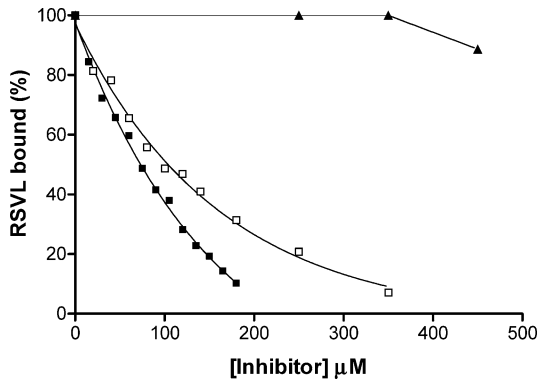

FIGURE 7: Sugar specificity of RSVL determined by BIACORE analysis. Inhibition of binding of the lectin to a lactose−BSA surface by mannose (▲), GalNAc (□), and lactose (■).

water-mediated hydrogen bond with Q121, a residue conserved in five of the six viperid venom lectins (see Figure 1). The more divergent elapid venoms (BML, BFL-1, and BFL-2) have either an aspartate or an alanine in this position. Consistent with the ability of RSL to bind GalNAc, modeling (data not shown) shows that the *N*-acetyl group can be accommodated sterically while at the same time making a hydrogen bond with Q121.

Multivalency. The clustering of CRDs to give multivalent assemblies is a means by which many lectin types confer affinity and specificity on their interactions with multivalent ligands (*7*, *12*). The decameric structure shown by RSL strongly suggests that multivalency plays an important role in how it mediates function. Unlike other C-type lectins such as MBP-A, tetranectin, SP-D, and HL-1, in which accessory domains are responsible for promoting oligomerization, the RSL structure now shows that the C-type lectin CRD itself is capable of mediating higher-order structures. In fact, direct interactions involving the CRD alone are reminiscent of that seen in the AB_5 bacterial toxin B subunits (*41*) and the pentraxins (*42*), pentameric carbohydrate-binding proteins whose subunit folds are unrelated to the C-type lectins. In both cases, multivalent interactions are thought to be important for function. In a model for the cholera toxin−host cell interaction (*43*), for example, the five carbohydrate-binding sites, located in the face of the B subunit pentamer, are proposed to simultaneously interact with its G_{M1} ganglioside receptor on the host cell surface.

As shown in Figures 2 and 4, the carbohydrate-binding sites in RSL are located on the rims of the two pentameric disks. It follows that if all five binding sites of the RSL pentamer were to simultaneously interact with a single cell surface, the glycoconjugates with which it binds would have

to extend significantly from the plane of the membrane. As such, RSL would be expected to interact with cell surface glycoproteins or proteoglycans, rather than the relatively "short" gangliosides found to mediate cholera toxin interactions. Although edge-on modes of interaction might also be possible, Figure 4 shows that the edge of the pentamer or decamer is very negatively charged and therefore not well suited for interactions with the negatively charged cell surface. The ability to mediate multivalent interactions presumably accounts for the fact that RSL is a potent mediator of platelet aggregation, a process involving receptor cross-linking and downstream signaling (*29*). It is also interesting to note that the induction of edema by *B. godmani* (*44*) venom has been attributed to a calcium-dependent galactose-binding lectin, in this case through a process involving mast cell receptor cross-linking. Given that RSL is composed of two back-to-back pentamers, an orientation placing the face of the pentamer parallel to the cell surface also leads naturally to a means of cross-linking opposing cell surfaces during the process of agglutination.

The spacing between carbohydrate-binding sites has also provided insight into the nature of the interactions mediated by members of the C-type lectin family (*2*). In RSL, the carbohydrate-binding sites on each pentamer are located an average of 52 and 83 Å apart, distances consistent with a role in binding the relatively widely spaced determinants found on cell surfaces (*2*). In addition, however, certain pairs of binding sites from the opposing pentamers (e.g., D and C′ in Figure 4C) come to within 27 Å of each other (measured from C4 to C4 of the galactose residues). These more closely spaced pairs of binding sites might be spanned by a single N-linked oligosaccharide whose terminal galactose residues are modeled to be between 15 and 31 Å apart (*45*).

ACKNOWLEDGMENT

We thank Ulu Ünligil for his assistance with several of the computer programs used in this work and Xuekun Xing and David C. Watson for technical assistance. This work is based upon research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the National Science Foundation via Grant DMR-9311772, using the Macromolecular Diffraction facility at CHESS (MacCHESS), which is supported by Grant RR-01646 from the National Institutes of Health.

REFERENCES

1. Drickamer, K., and Taylor, M. E. (1993) Biology of animal lectins, *Annu. Rev. Cell Biol. 9*, 237−264.
2. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) The C-type lectin superfamily in the immune system, *Immunol. Rev. 163*, 19−34.
3. Brissett, N. C., and Perkins, S. J. (1998) Conserved basic residues in the C-type lectin and short complement repeat domains of the G3 region of proteoglycans, *Biochem. J. 329*, 415−424.
4. Vestweber, D., and Blanks, J. E. (1999) Mechanisms that regulate the function of the selectins and their ligands, *Physiol. Rev. 79*, 181−213.
5. Drickamer, K. (1988) Two distinct classes of carbohydrate-recognition domains in animal lectins, *J. Biol. Chem. 263*, 9557−9560.
6. Drickamer, K. (1999) C-Type lectin-like domains, *Curr. Opin. Struct. Biol. 9*, 585−590.
7. Weis, W. I., and Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition, *Annu. Rev. Biochem. 65*, 441−473.
8. Drickamer, K. (1992) Engineering galactose-binding activity into a C-type mannose-binding protein, *Nature 360*, 183−186.
9. Iobst, S. T., and Drickamer, K. (1994) Binding of sugar ligands to Ca^{2+}-dependent animal lectins. II. Generation of high-affinity galactose binding by site-directed mutagenesis, *J. Biol. Chem. 269*, 15512−15519.
10. Kolatkar, A. R., and Weis, W. I. (1996) Structural basis of galactose recognition by C-type animal lectins, *J. Biol. Chem. 271*, 6679−6685.
11. Poget, S. F., Legge, G. B., Proctor, M. R., Butler, P. J., Bycroft, M., and Williams, R. L. (1999) The structure of a tunicate C-type lectin from *Polyandrocarpa misakiensis* complexed with D-galactose, *J. Mol. Biol. 290*, 867−879.
12. Rini, J. M. (1995) Lectin structure, *Annu. Rev. Biophys. Biomol. Struct. 24*, 551−577.
13. Nielsen, B. B., Kastrup, J. S., Rasmussen, H., Holtet, T. L., Graversen, J. H., Etzerodt, M., Thogersen, H. C., and Larsen, I. K. (1997) Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an α-helical coiled coil, *FEBS Lett. 412*, 388−396.
14. Weis, W. I., and Drickamer, K. (1994) Trimeric structure of a C-type mannose-binding protein, *Structure 2*, 1227−1240.
15. Hakansson, K., Lim, N. K., Hoppe, H. J., and Reid, K. B. (1999) Crystal structure of the trimeric α-helical coiled-coil and the three lectin domains of human lung surfactant protein D, *Structure 7*, 255−264.
16. Geijtenbeek, T. B., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C., Adema, G. J., van Kooyk, Y., and Figdor, C. G. (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses, *Cell 100*, 575−585.
17. Geijtenbeek, T. B., Kwon, D. S., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C., Middel, J., Cornelissen, I. L., Nottet, H. S., KewalRamani, V. N., Littman, D. R., Figdor, C. G., and van Kooyk, Y. (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells, *Cell 100*, 587−597.
18. Lis, H., and Sharon, N. (1998) Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition, *Chem. Rev. 98*, 637−674.
19. Hirabayashi, J., Kusunoki, T., and Kasai, K. (1991) Complete primary structure of a galactose-specific lectin from the venom of the rattlesnake *Crotalus atrox*, *J. Biol. Chem. 266*, 2320−2326.
20. Gartner, T. K., and Ogilvie, M. L. (1984) Isolation and characterization of three Ca^{2+} dependent β-galactoside-specific lectins from snake venoms, *Biochem. J. 224*, 301−307.
21. Ozeki, Y., Matsui, T., Hamako, J., Suzuki, M., Fujimura, Y., Yoshida, E., Nishida, S., and Titani, K. (1994) C-Type galactoside-binding lectin from *Bothrops jararaca* venom: comparison of its structure and function with those of botrocetin, *Arch. Biochem. Biophys. 308*, 306−310.
22. Ohkura, M., Miyashita, Y., Nikai, T., Suzuki, J., Komori, Y., Sugihara, H., and Ohizumi, Y. (1996) Properties of Ca^{2+} release induced by puff adder lectin, a novel lectin from the snake *Bitis arietans*, in sarcoplasmic reticulum, *J. Pharmacol. Exp. Ther. 277*, 1043−1048.
23. Nikai, T., Kato, S., Komori, Y., and Sugihara, H. (2000) Amino acid sequence and biological properties of the lectin from the venom of *Trimeresurus okinavensis* (Himehabu), *Toxicon 38*, 707−711.
24. Xu, Q., Wu, X. F., Xia, Q. C., and Wang, K. Y. (1999) Cloning of a galactose-binding lectin from the venom of *Trimeresurus stejnegeri*, *Biochem. J. 341*, 733−737.
25. Aragon-Ortiz, F., Mentele, R., and Auerswald, E. A. (1996) Amino acid sequence of a lectin-like protein from *Lachesis muta stenophyrs* venom, *Toxicon 34*, 763−769.
26. Komori, Y., Nikai, T., Tohkai, T., and Sugihara, H. (1998) Primary structure and biological activity of snake venom lectin (APL) from *Agistrodon p. piscivorus* (Eastern cottonmouth), *Toxicon 37*, 1053−1064.
27. Zha, H. G., Lee, W. H., and Zhang, Y. (2001) Cloning of cDNAs encoding C-type lectins from Elapidae snakes *Bungarus fasciatus* and *Bungarus multicinctus*, *Toxicon 39*, 1887−1892.
28. Markland, F. S. (1998) Snake venoms and the hemostatic system, *Toxicon 36*, 1749−1800.
29. Ogilvie, M. L., Byl, J. W., and Gartner, T. K. (1989) Platelet-aggregation is stimulated by lactose-inhibitable snake venom lectins, *Thromb. Haemostasis 62*, 704−707.

30. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
31. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr. D55*, 849−861.
32. Furey, W., and Swaminathan, S. (1997) PHASES-95: A program package for the processing and analysis of diffraction data from macromolecules, *Methods Enzymol. 277*, 590−620.
33. De la Fortelle, E., and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods, *Methods Enzymol. 276*, 472−494.
34. Cowtan, K., and Main, P. D. (1998) Miscellaneous alogorithms for density modification, *Acta Crystallogr. D54*, 487−493.
35. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A47*, 110−119.
36. Bertrand, J. A., Pignol, D., Bernard, J. P., Verdier, J. M., Dagorn, J. C., and Fontecilla-Camps, J. C. (1996) Crystal structure of human lithostathine, the pancreatic inhibitor of stone formation, *EMBO J. 15*, 2678−2684.
37. Brunger, A. T., and Rice, L. M. (1997) Crystallographic refinement by simulated annealing: methods and applications, *Methods Enzymol. 277*, 243−269.
38. Nayal, M., and Di Cera, E. (1996) Valence screening of water in protein crystals revealed as potential Na^+ binding sites, *J. Mol. Biol. 256*, 228−234.
39. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991) Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing, *Science 254*, 1608−1615.
40. Imberty, A. (1997) Oligosaccharide structures: theory versus experiment, *Curr. Biol. 7*, 617−623.
41. Merritt, E. A., and Hol, W. G. (1995) AB5 toxins, *Curr. Opin. Struct. Biol. 5*, 165−171.
42. Srinivasan, N., White, H. E., Emsley, J., Wood, S. P., Pepys, M. B., and Blundell, T. L. (1994) Comparative analyses of pentraxins: implications for protomer assembly and ligand binding, *Structure 2*, 1017−1027.
43. Merritt, E. A., Sarfaty, S., van den Akker, F., L'Hoir, C., Martial, J. A., and Hol, W. G. (1994) Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide, *Protein Sci. 3*, 166−175.
44. Lomonte, B., Rojas, G., Gutierrez, J. M., and Ramirez, G. (1990) Isolation of a galactose-binding lectin from the venom of the snake *Bothrops godmani* (Godmann's Pit Viper), *Toxicon 28*, 75−81.
45. Lee, Y. C., Lee, R. T., Rice, K., Ichikawa, Y., and Wong, T.-C. (1991) Topography of binding sites of animal lectins: ligands' view, *Pure Appl. Chem. 63*, 499−506.
46. Christopher, J. A. (1997) *SPOCK: The structural properties observation and calculation kit*, The Center for Macromolecular Design, Texas A&M University, College Station, TX.
47. Merritt, E. A., and Murphy, M. E. P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics, *Acta Crystallogr. D50*, 869−873.
48. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.

BI035871A