Glycobiology, 2018, vol. 28, no. 12, 968–977
doi: 10.1093/glycob/cwy072
Advance Access Publication Date: 1 September 2018
Original Article

Structural Biology

Ligand binding and retention in snake gourd seed lectin (SGSL). A crystallographic, thermodynamic and molecular dynamics study

Thyageshwar Chandran, Nukathoti Sivaji, Avadhesha Surolia, and Mamannamana Vijayan[1]

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

[1]To whom correspondence should be addressed: Tel: +91-80-23600765, 22932590; Fax: +91-80-23600535; e-mail: mv@iisc.ac.in (M. V.)

Received 23 August 2017; Revised 3 August 2018; Editorial decision 3 August 2018; Accepted 6 August 2018

Abstract

Snake gourd seed lectin (SGSL) is a non-toxic homolog of type II ribosome-inactivating proteins (RIPs) which contain a catalytic domain and a lectin domain. Isothermal titration calorimetry (ITC) measurements of the interactions of the protein with LacNAc, Lac, Gal, Me-α-Gal were carried out and the crystal structures of the native protein and its complex with Lac were determined. The crystal structure of the Me-α-Gal complex has already been determined. While the crystal structure showed the presence of two-sugar-binding sites, one on each of the two domains of the lectin chain, ITC measurements indicated the presence of only one binding site. In order to resolve this anomaly, molecular dynamics (MD) simulations were carried out on the native protein and on its complexes with Me-α-Gal and Lac. Simulations were also performed on the protein after reducing the inter-chain disulfide bridge between the two chains. The crystal structures and the simulations confirmed the robustness of the protein structure, irrespective of the presence or absence of the disulfide bridge. The simulations indicated that although two sites can bind sugar, only the ligand at one site is retained in a dynamic situation. The studies thus bring out the subtle relationship between binding and retention of the ligand.

Key words: absence of toxicity, binding and retention, ligand binding sites, β-trefoil lectin, Type II RIPs

Introduction

Type II ribosome-inactivating proteins (RIPs), which are believed to be a part of the defense mechanism of plants against pathogens, are in general made up of a catalytic chain and a lectin chain. The catalytic chain removes adenine at a specific location in rRNA (α-sarcin/ricin loop in 28 S rRNA), while the lectin chain is involved in binding to the cell surface receptors and facilitating endocytosis (Olsnes et al. 1974; Endo et al. 1987; Robertus 1991). The two β-trefoil domains in the lectin chain carry one sugar-binding site each. The two chains are connected by means of a disulfide bridge. The first type II RIP whose structure was elucidated was ricin (Montfort et al. 1987), a potent toxin purified from the seeds of castor bean. Subsequently, structures of other type II RIPs such as abrin (Tahirov et al. 1995), *Ricinus communis* agglutinin (RCA) (Sweeney et al. 1997), European mistletoe lectin (Eu-ML) (Sweeney et al. 1998; Jimenez et al. 2005), Himalayan mistletoe lectin (Hm-RIP) (Mishra et al. 2004), *Abrus precatorius* agglutinin (Bagaria et al. 2006) and cinnamomin (Azzi et al. 2009), have been reported.

Most of the type II RIPs reported so far are toxic in nature. Non-toxic or less toxic variants of these RIPs also have been found (Pascal et al. 2001; Sharma et al. 2013; Chandran et al. 2015) and had recently received attention partly on account of the possibility of using them for the construction of non-toxic or mildly toxic immunotoxins and conjugates against specific antigens of tumor

cells (Ferreras et al. 2011; Zeng et al. 2015; Poiroux et al. 2017). However, the physiological role of such proteins is yet to be elucidated. Furthermore, the molecular basis of their non-toxicity is only beginning to be understood. Ebulin is among the first non-toxic type II RIPs to be reported. The absence of toxicity of ebulin has been attributed to a defective oligosaccharide-binding site (Pascal et al. 2001). The structure of another non-toxic analog of type II RIPs, that from *Trichosanthes kirilowii* (TKL-1), determined at 2.7 Å resolution, was reported (Li et al. 2001). However, due to the non-availability of the protein sequence, information provided by the structure is somewhat limited. It has been suggested that the absence of toxicity of this protein is due to a defect in the catalytic chain. Snake gourd seed lectin (SGSL) and bitter gourd seed lectin (BGSL) are two other non-toxic type II RIPs which have been characterized (Mazumder et al. 1981; Padma et al. 1998; Komath et al. 2001; Sultan and Swamy 2005; Kavitha et al. 2009; Swamy et al. 2015). SGSL is a two-chain protein while BGSL is made up of four chains involving two two-chain modules interconnected by a disulfide bridge. The structure of the complex of SGSL with methyl-α-galactoside (Me-α-Gal) and several sugar complexes of BGSL have been determined in this laboratory (Sharma et al. 2013; Chandran et al. 2015) as a part of the long-range program on structural biology of lectins (Chandra and Vijayan 1999; Natchiar et al. 2007; Chandran et al. 2013; Abhinav and Vijayan. 2014). The adenine-binding site in the catalytic chain is defective in both the lectins. The lectin chain of SGSL binds to two sugar molecules while that of BGSL binds only one. A thorough examination of the sequence and structure of the two lectins and their interactions with sugar molecules indicated that the non-toxicity of both of them resulted from a combination of changes in the catalytic and carbohydrate-binding sites.

While the structure of the SGSL–Me-α-Gal complex provided the basic framework of the structure and interactions of SGSL, further exploration of the problem appeared worthwhile. SGSL–carbohydrate interactions in solution have been explored earlier using fluorescence spectroscopy (Komath et al. 2001). It was felt that the investigation using the more rigorous isothermal titration calorimetry (ITC), presented in this paper, might provide additional information. It was also desirable to further crystallographically explore the nuances, if any, of the structure and interactions of the lectin. In the event, crystals of the native protein and its complex with lactose (Lac) could be prepared and analyzed. Comparison between the results of ITC studies and crystallographic investigation brought to light, among other things, an anomaly in relation to the number of sugar-binding sites in the lectin. Molecular dynamics (MD) simulations were carried out on the native protein as well as on the complexes of the lectin with Me-α-Gal and Lac, in order to address this anomaly. MD simulations were also carried out on the protein without the disulfide bridge between the two chains, to explore the source of the structural integrity of the two-chain protein. The results indicated that non-covalent interactions between the two chains rather than the disulfide bridge lead to the robustness of the structure. The MD simulations also brought to light nuances involving the ability to bind a ligand and the ability to retain it.

Results and discussion

Glycan array

Detailed carbohydrate-binding analyses of SGSL were performed by screening the lectin on glycan array (www.functionalglycomics.org). The strongest binding glycan involved Galβ1-4GlcNAc (lactosamine) at the non-reducing end (Supplementary data, Figure S1). The lectin also showed affinity to sialylated glycans. The glycan motifs preferentially bound by SGSL are known to play an important role in cell–cell recognition, blood group determination and adhesion process. Recently, it has been reported that this specific set of glycans is upregulated in neoplastic cells (Varki et al. 2009; Stowell et al. 2015).

Thermodynamic parameters

Results of typical ITC experiments carried out at 298 K, involving the addition of 8-μL aliquots each of Gal, Me-α-Gal, Lac and LacNAc solutions at concentrations mentioned in Materials and methods, to 150 μM of SGSL solution are shown in Figure 1. Observation of a monotonic decrease in the heat evolved with the addition of ligand until saturation is achieved indicates that SGSL has only one type of binding site. Plots of incremental heat evolved as a function of the molar ratio of the ligand to the protein (Figure 1) could be readily fitted to a model involving a single type of sites. The results unambiguously and consistently indicated the

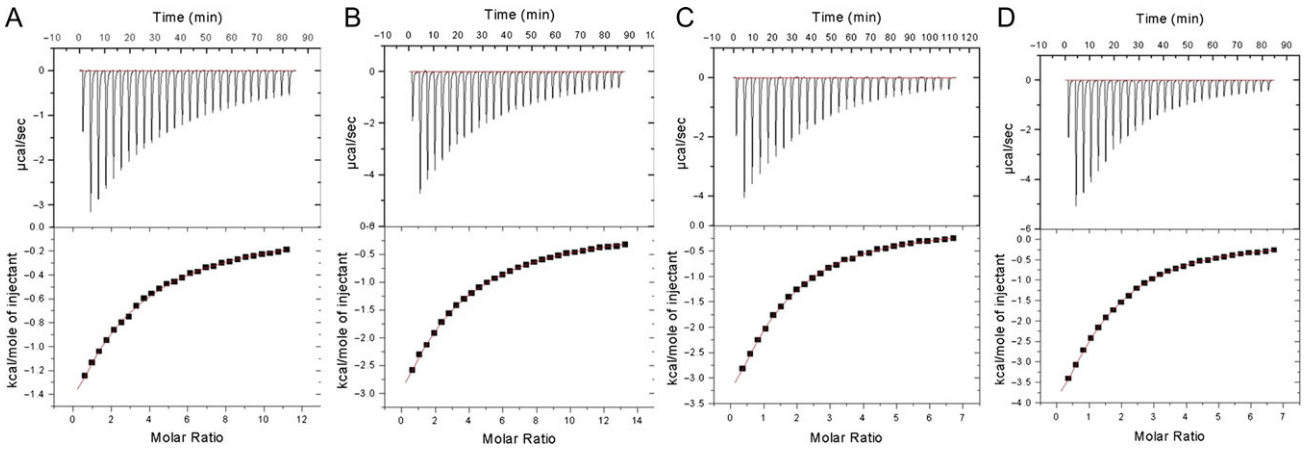

Fig. 1. ITC results on the interaction of SGSL (298 K) with (**A**) Me-α-Gal (**B**) Gal (**C**) Lac and (**D**) LacNAc. The top panel of the figures represents raw data and the lower panel contains plots of the heat released as a function of ligand concentration for the titration shown in the upper panel. Solid lines represent the best least square fits for the obtained data.

number of binding sites as close to one. Attempts to fit the data to a model with two independent sites led to inconsistent and unacceptable results involving abnormal standard deviations for occupancies and all the thermodynamic parameters. This would mean that one of the two possible sites is vacant or binding at that site is too weak to be observed. The low temperature data on LacNAc also could be fitted only to a model involving a single type of sites with n close to one. This is perhaps the most significant result that emerged from ITC studies. It would have been desirable to carryout ITC studies at much higher protein concentrations to explore the weak or no binding of sugar at the second site. This could not, however, be done as the protein precipitated at high concentrations. In any case, it is clear that detectable binding at normal concentrations occurred only at one site. Values of n, K_b, $-\Delta H_b$, $-\Delta G_b$, and $-T\Delta S_b$ for the binding of the sugars at 298 K and those for measurements at low temperatures on LacNAc, are listed in Table I. At the concentrations used, n and K_b values are determined with high accuracy, whereas ΔH values are prone to higher error (Wiseman et al. 1989). The order of binding affinities of the four ligands deduced from ITC experiments is the same as that reported earlier by fluorescence studies (Komath et al. 2001). An examination of thermodynamic data indicates that binding of SGSL to sugars are substantially enthalpically driven in agreement with studies on other plant lectin–sugar interactions (Williams et al. 1992; Schwarz et al. 1993; Surolia et al. 1996; Sharma et al. 1998).

Among the sugars examined, the binding of LacNAc is the strongest. ITC measurements of the binding of this ligand to SGSL were carried out at lower temperatures as well (Table I). An examination of the temperature dependence of the changes in the enthalpies for the binding of LacNAc yielded a moderate change in heat capacity ($d(\Delta H_b/dT) = -200\ kJ\ mol^{-1}\ K^{-1}$) (Supplementary data, Figure S2). This result is in consonance with earlier finding on a number of plant lectin–sugar interactions (Schwarz et al. 1993; Tonne 1994; Surolia et al. 1996). In consonance with the result of fluorescence studies reported earlier (Komath et al. 2001), the present results also indicate a correlation between negative or favorable enthalpy (ΔH) and negative or unfavorable entropy (ΔS) (Supplementary data, Figure S2). Such enthalpy–entropy compensation has been reported in the binding of sugars to other lectins as well (Sharma et al. 1998; Srinivas et al. 1999; Jimenez et al. 2005; Sultan and Swamy 2005; Moulaei et al. 2015).

Overall structure and protein–sugar interactions

The native structure contains two molecules in the asymmetric unit while that of the lactose complex contains one. The two molecules in the native structure have the same geometry with root-mean-square deviation (RMSD) of 0.17 Å in C^α positions when the two are superposed. Four crystallographically independent copies of the SGSL molecule are available, including those in the crystal structures presented here and that in the crystals of the Me-α-Gal complex reported earlier. In spite of differences in crystal packing and bound ligand, the four have nearly the same structure (Figure 2), with RMSD in C^α positions on pairwise superposition ranging between 0.3 Å and 0.5 Å. Thus, SGSL has a robust molecule. The structure of this molecule has been described adequately earlier (Sharma et al. 2013). Contrary to the indication provided by ITC experiments, interpretable electron density for lactose was observed at 1α and 2γ sites (Figure 3). Interactions of the Lac molecule with the lectin at the two sites are illustrated in Figure 4. Interactions of the protein with the galactose moiety are similar to those observed in the complex involving Me-α-Gal (Figure 4 and 5) at both the sites. Additionally, O3 of the glucose residue forms a hydrogen bond with Gly 24 O at 1α. No additional interactions are observed at 2γ. Although complexes with Gal and LacNAc could not be crystallized, possible additional interactions involving them could be discerned through simple modeling. The anomeric state of Gal could be α or β, or both the states could coexist (Abhinav et al. 2015). Modeling indicates the possibility of a hydrogen bond of α-O1 of Gal with Tyr 36 O at 1α and that β-O1 with Arg 239 N at 2γ. The contribution of these possible additional hydrogen bonds in relation to the other common interactions at the binding sites is likely to be very small. Interestingly, the glucose moiety has only one interaction, that too only at 1α site, with the lectin. This part of the ligand molecule substantially points to the solution. Presumably this leads to interference with the water structures in the vicinity of the protein. The resolution at which the structure has been solved does not permit a discussion involving the location of water molecules. The acetyl group of LacNAc, modeled on the basis of similar known structures (PDB ID: 2BSC, 1Y2X), is likely to further interfere with the water structure as it protrudes further into solution.

The discrepancy between the number of binding sites deduced from ITC measurements and observed in crystal structures remains a mystery. Careful examination confirmed that the two binding sites are not directly affected by molecular packing. Comparison with homologous structures indicates that both of them are legitimate sugar-binding sites (Montfort et al. 1987; Tahirov et al. 1995; Sweeney et al. 1997; 1998; Mishra et al. 2004; Jimenez et al. 2005; Bagaria et al. 2006; Azzi et al. 2009). Indeed, the strongest evidence for their being such sites is the unambiguous presence of sugar ligands at these sites in crystal structures with a comparable number of hydrogen bonds with the protein and surface area buried on complexation. It might be instructive to compare the situation in the lectin component of SGSL with that in other similar RIPs. The ideal candidate for this purpose is the lectin chain of RCA. Earlier solution studies on the lectin appeared to give contradictory results on

Table I. Thermodynamic parameters for the binding of Gal and its derivative to SGSL

	N	K_b (× 1000 M^{-1})	$-\Delta G_b$	$-\Delta H_b$	$-T\Delta S$
Me-α-Gal	1.06 ± 0.10	1.12 ± 0.01	17.43 ± 0.55	38.3 ± 3.0	20.9 ± 3.1
Gal	1.01 ± 0.10	3.77 ± 0.09	20.43 ± 0.22	36.0 ± 1.5	15.6 ± 1.5
Lac	1.14 ± 0.05	4.00 ± 0.10	20.59 ± 0.25	32.6 ± 1.5	12.0 ± 1.8
LacNAc	1.06 ± 0.04	4.19 ± 0.01	20.66 ± 0.26	26.3 ± 1.3	5.6 ± 1.3
LacNAc 278 K	1.10 ± 0.04	6.86 ± 0.01	21.06 ± 0.95	22.2 ± 0.1	1.1 ± 1.0
LacNAc 283 K	1.10 ± 0.10	5.42 ± 0.03	20.17 ± 0.80	21.4 ± 1.3	1.7 ± 1.6
LacNAc 288 K	1.13 ± 0.05	4.56 ± 0.01	20.80 ± 0.47	25.7 ± 1.4	4.8 ± 1.0

$-\Delta H$, $-\Delta G$ and $-T\Delta S$ are in kJ mol^{-1}. Unless otherwise stated, the measurements were made at 298 K (see text for details).

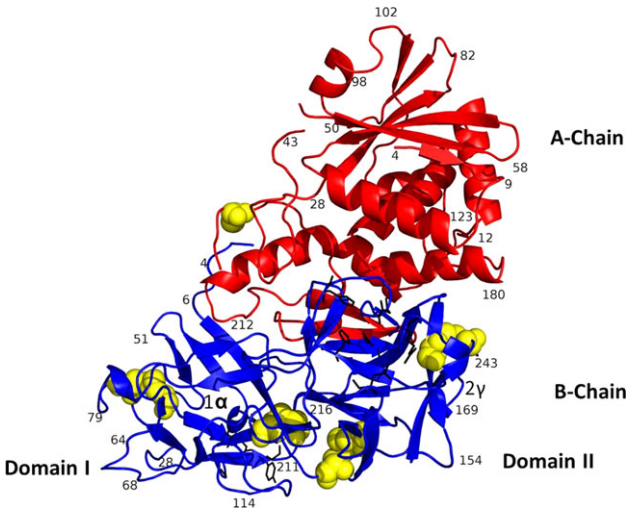

Fig. 2. The 3D structure prepared using the coordinates of an uncomplexed molecule of SGSL. The catalytic (A-chain) chain is shown in red and the lectin (B-chain) chain in blue. The disulfide bonds are shown as spheres. The binding site 1α in domain I of the lectin chain and 2γ in domain II are indicated.

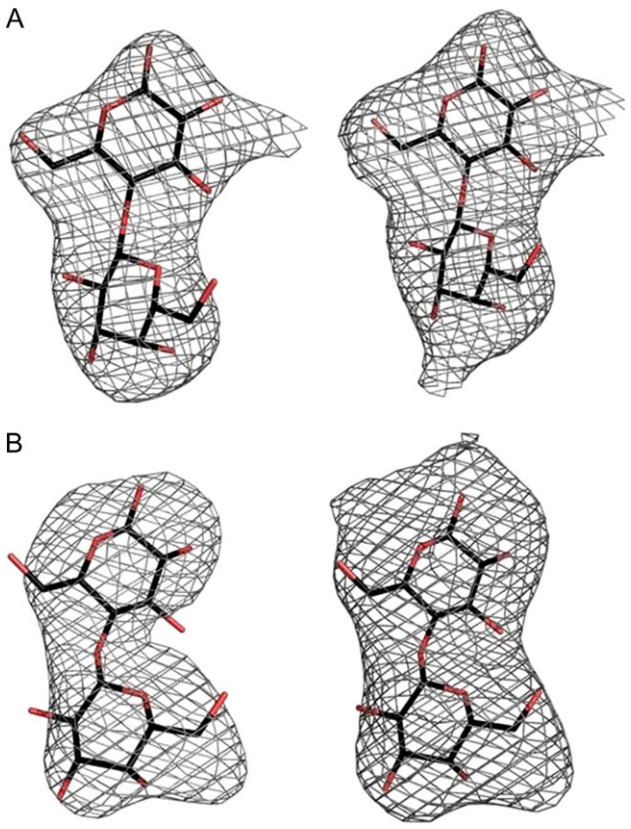

Fig. 3. Fo-Fc simulated annealing omit maps for Lac (**A**) 1α and (**B**) 2γ. Maps at the left and the right are contoured at 3σ and 2σ levels, respectively.

the number of binding sites on the lectin chain of the protein. Eventually the crystal structure indicated only one sugar-binding site on the chain in consonance with the results of the thermodynamic measurements of Podder et al. (1974) and Sharma et al. (1998).

The crystal structure showed that Gal binds at 1α in RCA; 2γ is vacant although this site is substantially, though not entirely, the same as 2γ in SGSL. The lectin–sugar (Me-α-Gal in SGSL and Gal in RCA) interactions in the two proteins are shown in Figure 5. The side chain of a tyrosyl residue stacks against the sugar ring in SGSL at 1α. The corresponding residue in RCA is Trp. Interactions involving an aspartyl, a glycyl and an asparginyl residue are common to both. An interaction involving Gln 41 in SGSL is replaced by that involving Lys 40 in RCA. However, two additional residues, namely, Glu 26 and Gln 35, are involved in interactions with the sugar at 1α in RCA. A histidyl side chain stacks against the sugar ring in SGSL at 2γ. This residue is present at the appropriate location in RCA. The aspartyl, arginyl, histidyl and asparaginyl residues, which are involved in hydrogen bonds with the sugar in SGSL, are present in RCA also at the corresponding locations. The critical difference at 2γ between the two lectins is in the presence of Asp 199 in SGSL which interacts with the sugar. Asp 199 occurs in the kink produced by two insertions in SGSL. This residue does not occur in RCA. Thus, interactions with sugar are weaker at 1α in SGSL than in RCA. On the contrary, the potential for such interactions is stronger at 2γ in SGSL than in RCA. The interactions observed at 1α and the potential for interactions suggested by the disposition of the concerned residue at 2γ, clearly indicate the preference for 1α for sugar binding in RCA. The observed interactions do not indicate any clear preference between the two sites in SGSL. It is, however, clear that 2γ is a better sugar-binding site in SGSL than in RCA while the reverse is true in the case of 1α.

Implications of dynamics to protein stability and ligand binding

An attempt was made to further explore the structure and ligand binding by SGSL using MD simulations. It was hoped that the simulations would help resolve the inconsistency in the results obtained from ITC measurements and crystallographic studies. Simulations were carried out on the native protein and its complexes with Me-α-Gal and Lac. Simulations were also carried out on the protein in which the inter-chain disulfide bridge was reduced in order to explore the source of the integrity of the two-chain protein. Appropriate crystal structures were used to construct the starting models. The simulations on the energy minimized structures were run for 300 ns. The time evolution of the RMSD of C^{α} positions from those in the initial models in the four simulations are shown in (Supplementary data, Figure S3). The RMSD are low throughout the simulations and are stable in the 200–300-ns region. The population distributions with respect to RMSD did not show any appreciable scatter in any of the simulations (Figure 6). The protein structure corresponding to the peak in the distribution in each case was similar to that seen in the respective crystal structure, thus indicating the robust nature of the SGSL molecule.

Interestingly, the population distribution in the simulations involving the native protein and that in which the inter-chain disulfide bridge is reduced are remarkably similar. The structures corresponding to the peaks in the two distributions differ mainly at the C-terminus of the A-chain and the N-terminus of B-chain and, in a less pronounced manner, in some loop regions. It may be recalled that these two termini are involved in the formation of the inter-chain disulfide bond. Naturally, the breakage of the bond results in the movement of the termini. The rest of the molecule remains essentially the same in the structures corresponding to the two peaks, with RMSD in C^{α} positions of 1.14 Å. Thus, the role of the disulfide

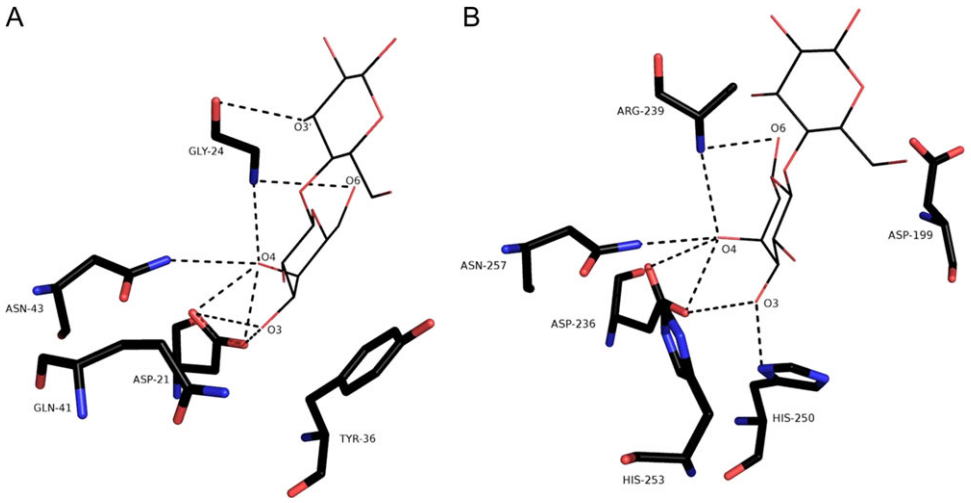

Fig. 4. SGSL–Lac interactions at (**A**) site 1α and (**B**) site 2γ.

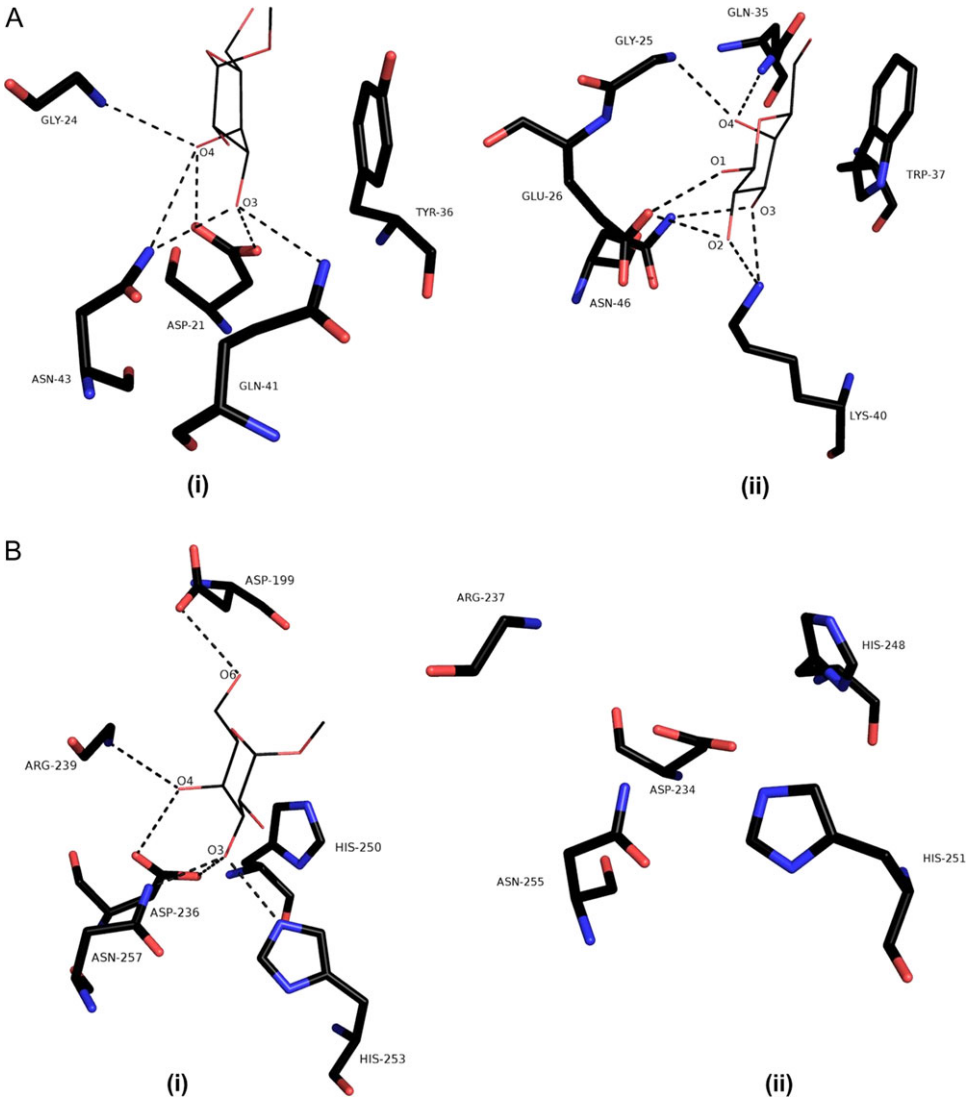

Fig. 5. (**A**) Protein–sugar interactions at 1α in (i) SGSL and (ii) RCA. (**B**) (i) Protein–sugar interactions at 2γ in SGSL and (ii) 2γ site in RCA.

bond in preserving the structural integrity of the protein appears to be marginal. Possibly the disulfide bridge plays a role similar to that of the linker between the two chains, present when the protein is originally synthesized as a single-chain precursor (Di Cola et al. 2001; Frigerio et al. 2001). The linker is probably necessary to ensure the correct folding of the protein. Once the protein is folded, the linker is cleaved off and the disulfide bridge holds the two chains together. It is reduced at the time of the entry of the catalytic chain into the cytoplasm. The two chains in the molecule are primarily held together by non-covalent interactions before their separation. For instance, the surface area buried in the two chains in the molecule is as large as 3500 Å^2.

The focus of the MD simulations was on lectin–sugar interactions. The calculations revealed remarkable differences in the behavior of ligands at 1α and 2γ. The difference between the two sites was apparent even during the equilibration prior to the production MD simulation, in the case of the SGSL–Lac complex. For the initial model at the start of the minimization obtained from the crystal structure, the binding energies of lactose at 1α and 2γ had comparable values at −20.3 and −19.4 kJ mol^{-1}, respectively. During NvT equilibration, the binding at 1α progressively loosened with concurrent movement of the ligand away from the lectin and progressive decrease in the binding energy (Supplementary data, Figure S4). It is the movement of Gly 24 and Gln 41 that substantially contributed to the loosening of the binding site. Lactose at 2γ, however, remains intact during the equilibration phase with a binding energy of −24.9 kJ mol^{-1}, although the second residue (Gln) exhibits some variability in its location.

Ligands at 1α and 2γ remain intact during the equilibration and stabilization steps of the SGSL–Me-α-Gal complex. The initial values of binding energies at the two sites were −16.2 and −18.1 kJ mol^{-1}, respectively. The corresponding final values were −17.5 and −23.2 kJ mol^{-1}. However, the ligands at the two sites behave very differently during simulation. The ligand at 2γ remained intact throughout the simulation. That at 1α detached itself from the binding site by about 97 ns during the course of the simulation (Figure 7), with a progressive decrease in binding energy (Supplementary data, Figure S5).

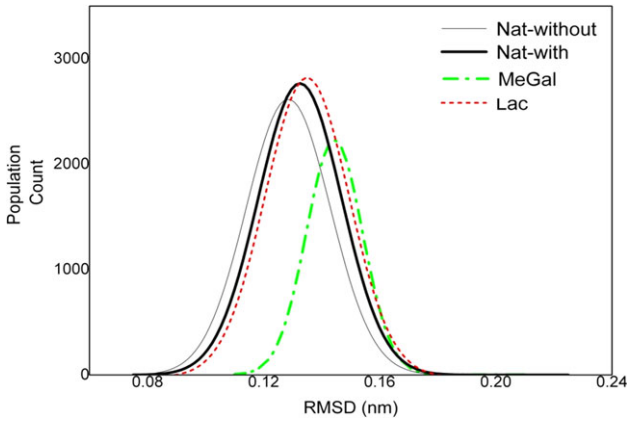

Fig. 6. Population distribution of the four MD simulations. In each case, RMSD is calculated with respect to the starting model of that structure. Nat-without—native structure without inter-chain disulfide bridges. Nat-with—native structure with inter-chain disulfide bridges. MeGal—complex with Me-α-Gal. Lac—complex with Lac.

Conclusion

Comparison of the crystal structures of SGSL and its complex with Lac presented here and that of the complex with Me-α-Gal reported earlier shows that the lectin molecule is substantially rigid with preformed binding sites. Crystal structures indicate the presence of two sugar-binding sites, 1α and 2γ, one each on the two β-trefoil lectin domains. Intriguingly, detailed isothermal titration calorimetric studies on the interaction of several sugars with SGSL show the presence of only one binding site on the lectin. MD simulations of free SGSL with and without the inter-chain disulfide bond and its complexes with Me-α-Gal and Lac confirm the robustness of the protein molecule. The simulations demonstrate that the disulfide bridge is incidental for the structural integrity of SGSL, and perhaps that of other Type II RIPs. They also indicate that the sugar binding at site 1α of SGSL is unstable on account of the flexibility of a couple of residues constituting it, thus providing a plausible explanation for the discrepancy between X-ray and calorimetric results. It would appear that both the sites can bind sugars. In a comparatively rigid framework provided by the crystal, both the ligands are retained, but in a dynamical situation, presumably only one site (2γ) is able to retain the ligand. The mobility of the second site (1α) enables the ligand to detach itself from the protein. The results presented here thus highlight the subtle relation between binding and retention.

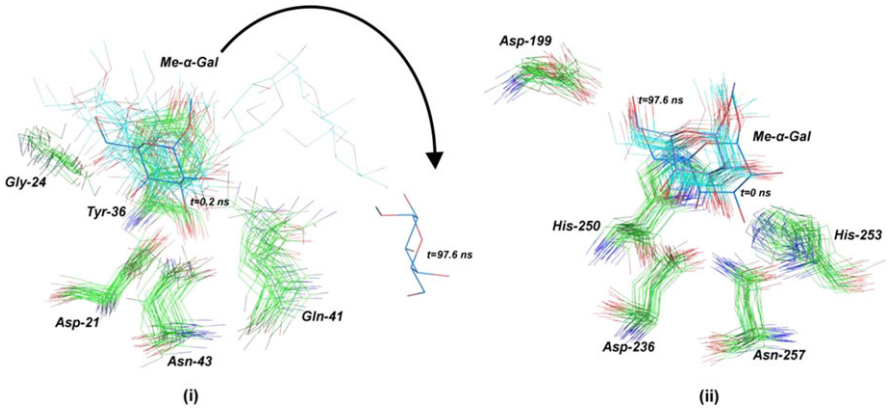

Fig. 7. Scatter of residues and the ligand at (i) 1α and (ii) 2γ during the initial 98 ns of MD simulation of the Me-α-Gal complex. The frames were downloaded at 3-ns intervals. The sugars at the start and end of the simulations are given in thick lines. Arg 239 of 2γ-binding pocket overlaps with ligand in this projection and is therefore omitted from the figure for clarity.

Materials and methods

Materials

Snake gourd seeds were procured from local seed vendors in Bengaluru, India. Sepharose-6B, galactose, Me-α-Gal, β-lactose, β-mercaptoethanol, PEG 400 and ammonium sulfate were purchased from Sigma (St Louis, Missouri, USA). Epichlorohydrin, sodium hydroxide and sodium phosphate (both monobasic and dibasic) were obtained from Merck (Mumbai, India). *N*-acetyl-lactosamine was purchased from Dextra reagents, UK. All solvents used were of HPLC-grade purity unless otherwise mentioned.

Protein purification

SGSL was purified using affinity chromatography as described previously (Komath et al. 1996) with slight modifications when required to increase the yield of the protein. In brief, SGSL seeds were crushed and soaked overnight in phosphate buffer saline (PBS), pH 7.3. The material was centrifuged, and the supernatant passed through a sepharose-6B column cross-linked with galactose. The bound protein was eluted from the affinity column using 200 mM galactose in PBS, pH 7.3. The eluted protein was dialysed extensively against PBS for 48 h and purity examined on a 12% SDS-PAGE gel. Finally, the protein was centrifuged at 12000 rpm, for 50 min. The supernatant was used for further downstream experiments.

Glycan array

The glycan specificity was analyzed on printed glycan array slides at the Consortium for Functional Glycomics (www.functionalglycomics.org). Glycan array version 5.2 containing 609 glycans in replicates of six were used to investigate the protein. The purified protein was biotin labeled by incubating the protein for 6 h at 298 K with 10 mM biotin, which was then applied to the microarray slides (Heimburg-Molinaro et al. 2011).

Isothermal titration calorimetry

Titrations were carried out using a Microcal VP-ITC calorimeter at 277, 283, 288 and 298 K. Purified SGSL was dialysed against Tris–HCl buffer (20 mM Tris/HCl, pH 7.5, 300 mM NaCl) for 12 h with three changes before each titration. The protein concentration was estimated by measuring absorbance at 280 nm. A protein concentration of 150 μM was used in all the experiments. A higher concentration of the protein often led to aggregation. Concentration of ligands used in titrations were 8, 10, 6 and 6 mM for Me-α-Gal, Gal, Lac and LacNAC, respectively. The ligand solutions were prepared from the same dialysate as used for the dialysis of protein, to prevent any artefacts from the heat of mixing. 1.36 ml of the degassed protein solution was carefully added to the sample cell, ensuring that no air bubbles were trapped, and was then equilibrated to the appropriate preset temperature. The ligand solution loaded in the injection syringe was added to the sample cell as a series of injections (aliquots of 7–10 μL) separated by 3-min interval with constant stirring at 307 rev min^{-1}. Control titrations where the ligand is titrated into buffer in the sample cell were carried out to confirm that the heat of dilution for the ligand was negligible. Data were fitted with MicroCal Origin 7 software according to standard procedures. The fitted data yielded the stoichiometry (n), binding affinity (K_b), and enthalpy of binding (ΔH). Other thermodynamic parameters, namely, changes in free energy (ΔG) and entropy (ΔS), were calculated from the equation $\Delta G = \Delta H - T\Delta S = -RT\ln K_b$, in which T is the absolute temperature and $R = 8.314\ J\ mol^{-1}\ K^{-1}$. Three independent titrations were performed for each ligand for measurements at 298 K and average values obtained from these three measurements were calculated. Only two independent measurements were made for low temperature studies on LacNAc.

Crystallization, structure determination and refinement

Prior to crystallization, the protein solution was dialysed extensively in Tris buffer (10 mM Tris pH 7.8, 100 mM NaCl, 5 mM β-mercaptoethanol). Crystals of SGSL in its native form were obtained by the hanging-drop method by equilibrating a 4-μL drop containing 8 mg mL^{-1} of protein, with an equal volume of reservoir solution made up of 0.2 M lithium sulfate monohydrate, 0.1 M Tris–HCl, pH 8.5 and 30% PEG 4000 as a precipitant. After nearly 2 weeks, hexagonal crystals started appearing, which were used for data collection. Crystals of SGSL complexed with Lac were also obtained by the hanging-drop method. The reservoir solution contained 4.0 M sodium formate. The protein solution was incubated with 5 mM lactose for 10 h at 277 K prior to setting up for crystallization. The crystals grew to a maximum dimension of 0.4 × 0.2 × 1.3 mm. Attempts to prepare crystals of complexes with Gal and LAcNac did not succeed.

Intensity data were initially collected at home source using Mar Research MAR345dtb imaging plate mounted on a Rigaku ULTRAX-18 X-ray generator. Subsequently, better data were collected at 100 K on the BM14 beamline at a wavelength of 1.0 Å at the European synchrotron radiation facility (ESRF), Grenoble, France using an MAR Research MAR345 imaging plate. The data were processed and merged using imosflm (Battye et al. 2011) and SCALA of the CCP4 program suite (Winn et al. 2011). TRUNCATE (French and Wilson 1978) from *CCP*4 was used to convert intensities into structure factor amplitudes. Solvent content was estimated using Matthew's method (Matthews 1968). The structures were solved by the molecular replacement method using PHASER (McCoy et al. 2007) with the protein component of the SGSL–Me-α-Gal complex (PDB ID:4HR6) as the search model. Ligands were located from difference Fourier maps when R and R_{free} were 0.276 and 0.327, respectively. The locations were subsequently confirmed using simulated annealing omit maps. Water O atoms were added based on peaks >1σ in 2Fo-Fc maps and 3σ in Fo-Fc maps. The models were validated using PROCHECK (Laskowski et al. 1993) and using MOLPROBITY (Chen et al. 2010). Data collection and refinement statistics are summarized in Table II.

Analysis of structures

Structural alignments were carried out using ALIGN (Cohen 1997) and structure-based multiple sequence alignments were performed using MUSTANG (Konagurthu et al. 2006). The accessible surface area was calculated using the program NACCESS with a probe radius of 1.4 Å (Hubbard and Thornton 1993). CONTACT from the CCP4 suite of programs was used to calculate inter-atomic distances. Possible hydrogen bonds were identified using HBPLUS (McDonald and Thornton 1994).

Molecular dynamics simulations

Simulations were carried out using the Groningen Machine for Chemical Simulations (Gromacs) v5.1.2 package (Berendsen et al.

Table II. Data collection statistics and refinement parameter

	Native	Lactose
Space group	$P6_4$	$P6_122$
Unit cell dimensions		
$a = b$ (Å)	171.2	109.4
c (Å)	75.9	232.7
$\alpha = \beta$ (°)	90	90
γ (°)	120	120
No of molecules/asymmetric unit	2	1
Resolution (Å)	56.8–2.9 (3.0–2.9)	94.8–3.05 (3.2–3.05)
No. of observations	334,967 (46,789)	311,166 (45,533)
No. of unique reflections	28,365 (4115)	16,493 (2340)
Completeness (%)	100 (100)	100 (100)
$I/(\sigma I)$	12.6 (2.1)	13.3 (2.1)
CC(1/2) (%)	99.6 (64.7)	99.9 (80.8)
R_{merge}[a] (%)	19.5 (128.2)	14.6 (153.4)
Multiplicity	11.8 (11.4)	18.9 (19.5)
R-factor[b] (%)	19.0	23.7
R_{free}[b] (%)	25.8	25.4
RMS deviations from ideal values		
Bond length (Å)	0.014	0.017
Bond angle (°)	1.7	2.0
Residues (%) in Ramachandran plot		
Favored (%)	97.1	95.5
Allowed region (%)	2.9	4.5
Disallowed region (%)	0	0

[a]$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all i observations of reflection hkl.

[b]5% of the reflections were used for the R_{free} calculations.

Values within parentheses refer to the last resolution shell.

1995; Van Der Spoel et al. 2005; Abraham et al. 2016) with AMBER99SB force field (Lindorff-Larsen et al. 2010) and TIP3P water model (Jorgensen et al. 1983).

Initial coordinates used for the simulation were obtained from their respective crystal structures. Ligand topologies suitable for AMBER force field were generated using the ACPYPE server (Sousa da Silva and Vranken 2012) and checked manually. Partial charges on the ligands were obtained using the RED server (Vanquelef et al. 2011). Missing side chains were added using COOT. Native protein with and without the inter-chain disulfide bridge and the complexes were positioned inside a cubic box with a distance of 1.2 nm from the walls of the box. Standard protonation states of residues were used and sodium ions were added to neutralize the overall system charge wherever necessary. Particle Mesh Ewald was used for treating electrostatics (Darden et al. 1993), with a Coulomb cut-off of 1.4 nm. The van der Waals interactions were treated using the switch potential applied from 0.9 to 1.0 nm. Bond lengths were restrained using LINCS algorithm (Hess et al. 1997). The energy of the system was minimized using the steepest descent method until a change in the last cycle was less than 1 kJ mol^{-1} nm^{-1}. Later, the system was equilibrated to a required temperature of 300 K using Nosé–Hoover (Nosé 1984; Hoover 1985) thermostat, and to 1 atm using Parrinello–Rahman barostat (Parrinello and Rahman 1981), with position restraints on both protein and ligands. Simulations were carried out for 300 ns. Binding free energies of bound ligand were calculated using Autodock Tools 4 (Morris et al. 2008; El-Hachem et al. 2017). Visual molecular dynamics (Humphrey et al. 1996) molecular visualization software was used for structure visualization and analysis of MD trajectories. Graphs were generated using Xmgrace (Paul J. Turner Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology Beaverton, Oregon).

PDB references

SGSL and its complex, 5y42 (SGSL-Native) and 5y97 (SGSL-Lac).

Supplementary data

Supplementary data are available at *Glycobiology* online.

Funding

The initial X-ray data sets were collected at X-ray facility for Protein Crystal Structure Determination and Protein Design at this Institute, supported by the Science and Engineering Research Board (SERB) of Department of Science and Technology (DST), which were latter, improved at ESRF, Grenoble, facilitated by an arrangement funded by the Department of Biotechnology (DBT). Molecular Dynamics simulations were performed at the Supercomputer Education and Research Centre of the institute. The work is supported by the SERB-DST. AS is an SERB Distinguished Fellow. MV has been successively Einstein Professor of the Indian National Science Academy (INSA) and Platinum Jubilee senior scientist of the National Academy of sciences, India (NASI).

Acknowledgements

The authors acknowledge the help of Aditya Kumar, Asmita Gupta, Ansuman Biswas, Anju Paul, Kishore Babu, Aparna Ashok (Indian Institute of Science, Bangalore, India) at different stages of the work.

Conflict of interest statement

None declared.

Abbreviations

BGSL, bitter gourd seed lectin; Gal, galactose; ITC, isothermal titration calorimetry; Lac, lactose; LAcNac, *N*-acetyl-lactosamine; MD, molecular dynamics; Me-α-Gal, methyl-alpha-galactose; PBS, phosphate-buffered saline; RCA, *Ricinus communis* agglutinin; RFU, relative fluorescence unit; RIP, ribosome-inactivating protein; RMSD, root-mean-square deviation; SGSL, snake gourd seed lectin.

References

Abhinav KV, Sharma K, Swaminathan CP, Surolia A, Vijayan M. 2015. Jacalin–carbohydrate interactions: Distortion of the ligand molecule as a determinant of affinity. *Acta Crystallogr D Biol Crystallogr*. 71:324–331.

Abhinav KV, Vijayan M. 2014. Structural diversity and ligand specificity of lectins. The Bangalore effort. *Pure Appl Chem*. 86:1335–1355.

Abraham MJ, van der Spoel D, Lindahl E, Hess B, team. tGd. 2016. GROMACS User Manual version 5.1.2. wwwgromacsorg.

Azzi A, Wang T, Zhu DW, Zou YS, Liu WY, Lin SX. 2009. Crystal structure of native cinnamomin isoform III and its comparison with other ribosome inactivating proteins. *Proteins*. 74:250–255.

Bagaria A, Surendranath K, Ramagopal UA, Ramakumar S, Karande AA. 2006. Structure-function analysis and insights into the reduced toxicity of *Abrus precatorius* agglutinin I in relation to abrin. *J Biol Chem*. 281: 34465–34474.

Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. 2011. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 67:271–281.

Berendsen HJC, Van Der Spoel D, Van Drunen R. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp Phys Comm*. 91:43–56.

Chandra N, Vijayan M. 1999. Lectins. *Curr Opin Struct Biol*. 9:707–714.

Chandran T, Sharma A, Vijayan M. 2013. Generation of ligand specificity and modes of oligomerization in beta-prism I fold lectins. *Adv Protein Chem Struct Biol*. 92:135–178.

Chandran T, Sharma A, Vijayan M. 2015. Structural studies on a non-toxic homolog of type II RIPs from bitter gourd: Molecular basis of non-toxicity, conformational selection and glycan structure. *J Biosci*. 40: 929–941.

Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 66:12–21.

Cohen GH. 1997. ALIGN: A program to superimpose protein coordinates, accounting for insertions and deletions. *J Appl Cryst*. 30:1160–1161.

Darden T, York D, Pederesen L. 1993. Particle mesh Ewald: An N.log(N) method for sums in large systems. *J Chem Phys*. 98:10089–10092.

Di Cola A, Frigerio L, Lord JM, Ceriotti A, Roberts LM. 2001. Ricin A chain without its partner B chain is degraded after retrotranslocation from the endoplasmic reticulum to the cytosol in plant cells. *Proc Natl Acad Sci U S A*. 98:14726–14731.

El-Hachem N, Haibe-Kains B, Khalil A, Kobeissy FH, Nemer G. 2017. AutoDock and AutoDockTools for Protein-Ligand Docking: Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1(BACE1) as a Case Study. *Methods Mol Biol*. 1598:391–403.

Endo Y, Mitsui K, Motizuki M, Tsurugi K. 1987. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *J Biol Chem*. 262:5908–5912.

Ferreras JM, Citores L, Iglesias R, Jimenez P, Girbes T. 2011. Use of ribosome-inactivating proteins from Sambucus for the construction of immunotoxins and conjugates for cancer therapy. *Toxins*. 3:420–441.

French S, Wilson K. 1978. On the treatment of negative intensity observations. *Acta Crystallogr A*. 34:517–525.

Frigerio L, Jolliffe NA, Di Cola A, Felipe DH, Paris N, Neuhaus JM, Lord JM, Ceriotti A, Roberts LM. 2001. The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting. *Plant Physiol*. 126:167–175.

Heimburg-Molinaro J, Song X, Smith DF, Cummings RD 2011. Preparation and analysis of glycan microarrays. *Curr Protoc Protein Sci*. Chapter 12: Unit 12.10.

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. 1997. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*. 18: 1463–1472.

Hoover WG. 1985. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A*. 31:1695–1697.

Hubbard SJ, Thornton JM. 1993. *'NACCESS'. Computer program, Department of Biochemistry and Molecular Biology*. London: University college.

Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. *J Mol Graph*. 14(33–38):27–38.

Jimenez M, Saiz JL, Andre S, Gabius HJ, Solis D. 2005. Monomer/dimer equilibrium of the AB-type lectin from mistletoe enables combination of toxin/agglutinin activities in one protein: analysis of native and citraconylated proteins by ultracentrifugation/gel filtration and cell biological consequences of dimer destabilization. *Glycobiology*. 15:1386–1395.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 79:926–935.

Kavitha M, Sultan NA, Swamy MJ. 2009. Fluorescence studies on the interaction of hydrophobic ligands with *Momordica charantia* (bitter gourd) seed lectin. *J Photochem Photobiol B*. 94:59–64.

Komath SS, Kenoth R, Swamy MJ. 2001. Thermodynamic analysis of saccharide binding to snake gourd (*Trichosanthes anguina*) seed lectin. Fluorescence and absorption spectroscopic studies. *Eur J Biochem*. 268: 111–119.

Komath SS, Nadimpalli SK, Swamy MJ. 1996. Purification in high yield and characterisation of the galactose-specific lectin from the seeds of snake gourd (*Trichosanthes anguina*). *Biochem Mol Biol Int*. 39:243–252.

Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. 2006. MUSTANG: A multiple structural alignment algorithm. *Proteins*. 64:559–574.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 26:283–291.

Li M, Chai JJ, Wang YP, Wang KY, Bi RC. 2001. Crystal structure of *Trichosanthes Kirilowii* lectin-1 and its relation to the type 2 ribosome inactivating proteins. *Protein Pept Lett*. 8:81–87.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 78:1950–1958.

Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol*. 33: 491–497.

Mazumder T, Gaur N, Surolia A. 1981. The physicochemical properties of the galactose-specific lectin from *Momordica charantia*. *Eur J Biochem*. 113:463–470.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr*. 40: 658–674.

McDonald IK, Thornton JM. 1994. Satisfying hydrogen bonding potential in proteins. *J Mol Biol*. 238:777–793.

Mishra V, Ethayathulla AS, Sharma RS, Yadav S, Krauspenhaar R, Betzel C, Babu CR, Singh TP. 2004. Structure of a novel ribosome-inactivating protein from a hemi-parasitic plant inhabiting the northwestern Himalayas. *Acta Crystallogr Sect D: Biol Crystallogr*. 60:2295–2304.

Montfort W, Villafranca JE, Monzingo AF, Ernst SR, Katzin B, Rutenber E, Xuong NH, Hamlin R, Robertus JD. 1987. The three-dimensional structure of ricin at 2.8 A. *J Biol Chem*. 262:5398–5403.

Morris GM, Huey R, Olson AJ 2008. Using AutoDock for ligand-receptor docking. *Curr Protoc Bioinformatics*. Chapter 8:Unit 8.14.

Moulaei T, Alexandre KB, Shenoy SR, Meyerson JR, Krumpe LR, Constantine B, Wilson J, Buckheit RW Jr., McMahon JB, Subramaniam S et al. 2015. Griffithsin tandemers: flexible and potent lectin inhibitors of the human immunodeficiency virus. *Retrovirology*. 12:6.

Natchiar SK, Suguna K, Surolia A, Vijayan M. 2007. Peanut agglutinin, a lectin with an unusual quaternary structure and interesting ligand binding properties. *Crystallogr Rev*. 13:3–28.

Nosé S. 1984. A unified formulation of the constant temperature molecular-dynamics methods. *J Chem Phys*. 81:511–519.

Olsnes S, Refsnes K, Pihl A. 1974. Mechanism of action of the toxic lectins abrin and ricin. *Nature*. 249:627–631.

Padma P, Komath SS, Swamy MJ. 1998. Fluorescence quenching and time-resolved fluorescence studies on *Momordica charantia* (bitter gourd) seed lectin. *Biochem Mol Biol Int*. 45:911–922.

Parrinello M, Rahman A. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*. 52:7182–7190.

Pascal JM, Day PJ, Monzingo AF, Ernst SR, Robertus JD, Iglesias R, Perez Y, Ferreras JM, Citores L, Girbes T. 2001. 2.8-A crystal structure of a non-toxic type-II ribosome-inactivating protein, ebulin l. *Proteins*. 43:319–326.

Podder SK, Surolia A, Bachhawat BK. 1974. On the specificity of carbohydrate-lectin recognition. The interaction of a lectin from Ricinus communis beans with simple saccharides and concanavalin A. *Eur J Biochem*. 44:151–160.

Poiroux G, Barre A, van Damme EJM, Benoist H, Rouge P. 2017. Plant lectins targeting O-glycans at the cell surface as tools for cancer diagnosis, prognosis and therapy. *Int J Mol Sci*. 18:1232.

Robertus J. 1991. The structure and action of ricin, a cytotoxic N-glycosidase. *Semin Cell Biol*. 2:23–30.

Schwarz FP, Puri K, Bhat RG, Surolia A. 1993. Thermodynamics of monosaccharide binding to concanavalin A, pea (*Pisum sativum*) lectin, and lentil (*Lens culinaris*) lectin. *J Biol Chem*. 268:7668–7677.

Sharma S, Bharadwaj S, Surolia A, Podder SK. 1998. Evaluation of the stoichiometry and energetics of carbohydrate binding to *Ricinus communis* agglutinin: a calorimetric study. *Biochem J*. 333:539–542.

Sharma A, Pohlentz G, Bobbili KB, Jeyaprakash AA, Chandran T, Mormann M, Swamy MJ, Vijayan M. 2013. The sequence and structure of snake gourd (Trichosanthes anguina) seed lectin, a three-chain nontoxic homologue of type II RIPs. *Acta Crystallogr D Biol Crystallogr*. 69:1493–1503.

Sousa da Silva AW, Vranken WF. 2012. ACPYPE—AnteChamber PYthon Parser interfacE. *BMC Res Notes*. 5:367.

Srinivas VR, Bhanuprakash Reddy G, Surolia A. 1999. A predominantly hydrophobic recognition of H-antigenic sugars by winged bean acidic lectin: A thermodynamic study. *FEBS Lett*. 450:181–185.

Stowell SR, Ju T, Cummings RD. 2015. Protein glycosylation in cancer. *Annu Rev Pathol*. 10:473–510.

Sultan NA, Swamy MJ. 2005. Energetics of carbohydrate binding to *Momordica charantia* (bitter gourd) lectin: An isothermal titration calorimetric study. *Arch Biochem Biophys*. 437:115–125.

Surolia A, Sharon N, Schwarz FP. 1996. Thermodynamics of monosaccharide and disaccharide binding to *Erythrina corallodendron* lectin. *J Biol Chem*. 271:17697–17703.

Swamy MJ, Marapakala K, Sultan NA, Kenoth R. 2015. Galactose-specific seed lectins from Cucurbitaceae. *Curr Protein Pept Sci*. 16:17–30.

Sweeney EC, Tonevitsky AG, Palmer RA, Niwa H, Pfueller U, Eck J, Lentzen H, Agapov II, Kirpichnikov MP. 1998. Mistletoe lectin I forms a double trefoil structure. *FEBS Lett*. 431:367–370.

Sweeney EC, Tonevitsky AG, Temiakov DE, Agapov II, Saward S, Palmer RA. 1997. Preliminary crystallographic characterization of ricin agglutinin. *Proteins*. 28:586–589.

Tahirov TH, Lu TH, Liaw YC, Chen YL, Lin JY. 1995. Crystal structure of abrin-a at 2.14 A. *J Mol Biol*. 250:354–367.

Tonne E. 1994. Structure and energetics of protein-carbohydrate complexes. *Curr Opin Strct Biol*. 4:719–728.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. 2005. GROMACS: Fast, flexible, and free. *J Comput Chem*. 26: 1701–1718.

Vanquelef E, Simon S, Marquant G, Garcia E, Klimerak G, Delepine JC, Cieplak P, Dupradeau FY. 2011. R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res*. 39:W511–W517.

Varki A, Kannagi R, Toole BP. 2009. Glycosylation changes in cancer. In: *Essentials of Glycobiology*, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Williams BA, Chervenak MC, Tonne EJ. 1992. Energetics of lectin-carbohydrate binding: A microcalorimetric investigation of concanavalin A-oligomannoside complexation. *J Biol Chem*. 267:22907–22911.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A et al. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 67:235–242.

Wiseman T, Williston S, Brandts JF, Lin LN. 1989. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*. 179:131–137.

Zeng M, Zheng M, Lu D, Wang J, Jiang W, Sha O. 2015. Anti-tumor activities and apoptotic mechanism of ribosome-inactivating proteins. *Chin J Cancer*. 34:325–334.