BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

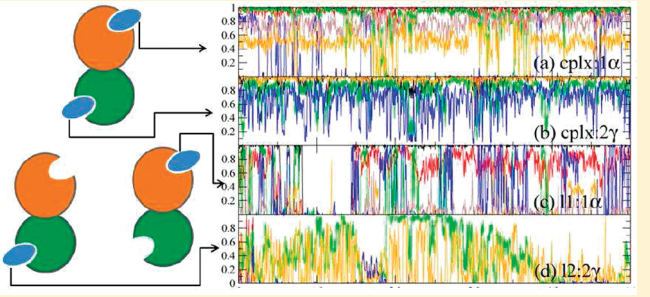
pubs.acs.org/biochemistry

Stability and Sugar Recognition Ability of Ricin-like Carbohydrate Binding Domains

Jianzhuang Yao,[†,‡] Ricky B. Nellas,[†,‡] Mary M. Glover,[‡] and Tongye Shen*[,†,‡]

[†]Department of Biochemistry, Cellular & Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996, United States
[‡]Center for Molecular Biophysics, Oak Ridge National Lab, Oak Ridge, Tennessee 37830, United States

ABSTRACT: Lectins are a class of proteins known for their novel binding to saccharides. Understanding this sugar recognition process can be crucial in creating structure-based designs of proteins with various biological roles. We focus on the sugar binding of a particular lectin, ricin, which has two β-trefoil carbohydrate-binding domains (CRDs) found in several plant protein toxins. The binding ability of possible sites of ricin-like CRD has been puzzling. The apo and various (multiple) ligand-bound forms of the sugar-binding domains of ricin were studied by molecular dynamics simulations. By evaluating structural stability, hydrogen bond dynamics, flexibility, and binding energy, we obtained a detailed picture of the sugar recognition of the ricin-like CRD. Unlike what was previously believed, we found that the binding abilities of the two known sites are not independent of each other. The binding ability of one site is positively affected by the other site. While the mean positions of different binding scenarios are not altered significantly, the flexibility of the binding pockets visibly decreases upon multiple ligand binding. This change in flexibility seems to be the origin of the binding cooperativity. All the hydrogen bonds that are strong in the monoligand state are also strong in the double-ligand complex, although the stability is much higher in the latter form due to cooperativity. These strong hydrogen bonds in a monoligand state are deemed to be the essential hydrogen bonds. Furthermore, by examining the structural correlation matrix, the two domains are structurally one entity. Galactose hydroxyl groups, OH4 and OH3, are the most critical parts in both site 1α and site 2γ recognition.



A class of proteins, generally termed lectins, has the ability to selectively bind to carbohydrates.[1] Their sugar recognition is critical for a diverse set of biological functions, from structural and mechanical roles, such as cell adhesion,[2] material transport such as endocytosis,[3−5] to signaling roles in the immune system.[6] Thus, there is significant interest in studying the protein−carbohydrate interaction and sugar recognition process. It is not trivial to design a protein that recognizes sugar with specificity. Unlike polypeptide recognition, there are less chemically distinct functional groups in sugar that protein can interact with. Hydrogen bonds and hydrophobic packing between the sugar ring and the aromatic residues of proteins seem to be the only two types of important interactions. In contrast to proteomics, glycomics has little direct connection with genomics. The primary structures of sugars are not determined by the genome and are subject to random events.[7] The potential randomness of sugar sequences and the difficulty of detection often are exploited by viruses; i.e., they use seemingly harmless "sugar bushes" to evade antibodies.[8] With a better understanding of nature's design, we may further engineer desired molecular recognitions and signaling pathways involving carbohydrate−protein interactions.

To date, researchers have purified plenty of lectins from bacteria, plants, and animals.[9] Many of their three-dimensional structures have been reported as well. Nevertheless, the actions of many lectins are still poorly understood. Lectins can be classified according to the fold of their carbohydrate-recognition domains (CRDs), such as L-type, M-type, and R-type. One important class of lectins is R-type (ricin-like) β-trefoil lectin. R-type lectin includes the plant toxins such as type II ribosome-inactivating proteins (RIPs). A type II RIP is made up of an A chain and a B chain.[10] Chain B recognizes cell surface sugar and helps chain A penetrate into the cytosol. Once inside, chain A halts protein synthesis.[11,12] Many plant toxins, such as those from barley, have A chain alone.[13] These toxins are called type I RIPs. Compared to type II, type I RIPs are much less toxic because they do not possess the delivery mechanism executed by chain B.

In this work, we focused on the stability and binding properties of R-type lectin's CRDs. By studying two examples of chain B, from *Ricinus communis* and *Trichosanthes kirilowii* lectin-1 (TKL1), we will provide insight into lectin−sugar interactions and R-type CRD binding specificity. Ricin, produced by castor beans, has two chains that are linked by a single disulfide bond. As shown in Figure 1, chain A (RTA), the catalytic domain of the toxin, is a glycohydrolase that cuts an RNA component of ribosomes and causes the shut down of protein synthesis. Chain B (RTB) is the binding domain that recognizes the terminal sugar residues of

Received: December 20, 2010
Revised: March 29, 2011
Published: April 21, 2011

ACS Publications © 2011 American Chemical Society

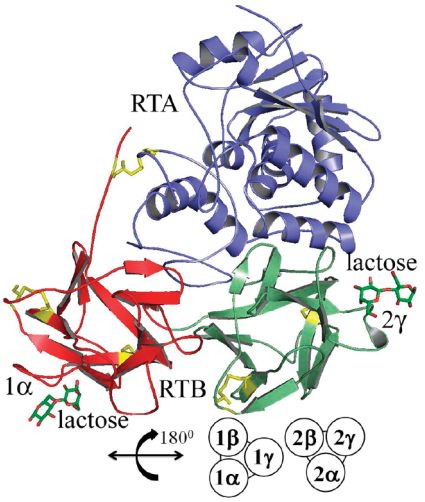


Figure 1. Ricin structure from PDB 2AAI. The two-chain protein (RTA: blue; RTB: red and green) is shown in the ribbon form. Disulfide bonds are shown in yellow. Two ligands are bound at the CRD domains. The bottom view of the subdomain structures of the two β-trefoils is shown in a cartoon representation.

glycoproteins on the cell surface. The last two residues of the nonreducing end of a cell surface polysaccharide are recognized by ricin via direct interaction. In this case, the terminal sugar residue (galactose) plays the central role, while the upstream neighbor residue (glucose) assists in the recognition. Once attached, ricin will accompany these glycoproteins during endocytosis. The structure of chain B contains a tandem R-type CRD.[11,12] The R-type CRD, which folds as a β-trefoil of about 130 residues, is a popular structural motif.[6] This CRD can also be found in other lectins, such as mannose receptors,[14] bacterial hydrolases,[15] and GalNAc-transferases.[16]

Besides the disulfide bond which cross-links chains A and B, there are four other important internal disulfide bonds (two in each of the two CRDs) for stabilizing the structural integrity of chain B.[10] Each trefoil CRD consists of four subdomains: a stem section, called λ, followed by three similar lobes, α, β, and γ. The β-trefoil fold is quite interesting. Each lobe, shown as containing four β-strands which has a very short persistence length (just a few residues), has an important sugar recognition Ω-loop in the middle. It appears that the two disulfide bonds maintain the overall structural stability of each CRD.

Each homologous lobe contains a potential binding site. Altogether, RTB has six potential galactose binding sites. Revealed by crystal structure (PDB: 2AAI) (Figure 1), it turns out that only two (site 1α and site 2γ) of these sites actually retain binding ability.[17] There is a conserved motif Gln-X-Trp in each lope that is critical to the sugar recognition. One of the key conserved protein−ligand interactions is that of Trp residue with the terminal sugar ring.[10,18]

There is a long history of conflicting reports on the binding properties of RTB. In an equilibrium dialysis study by Zentz and co-workers,[19] the results can be fitted into two models: (a) an independent binding model with one site stronger than the other and (b) two binding sites with equal binding strengths but with negative cooperativity between sites; i.e., the early binder inhibits the late binder. There were some very early speculations that cooperativity between the two binding sites do exist.[20]

Researchers believe that if these two sites exhibit cooperative binding, then there should be an accompanying allosteric effect that could be detected with fluorescence polarization or circular dichroism. However, the lack of a significant change in the fluorescence polarization signal upon binding[19] led previous researchers to the conclusion that binding induced conformational changes are quite local at best. Thus, some studies favor model (a) over model (b) and conclude that the binding sites are of different strengths and are independent of each other. In contrast, another experiment involving fluorescence polarization and equilibrium dialysis on ricin−galactosides indicated that the two sites have similar affinities.[21] A quantitative picture of the binding is gradually emerging with recent experiments. Results from mutagenesis investigations have finally pinned down that site 1α has a lower affinity for sugar binding than site 2γ.[7] Our current study will demonstrate that there are some level of positive binding cooperativity between the sites even though there is no major global conformational transitions.

Although the crystal structures have been solved,[22] the binding difference between site 1α and site 2γ is still puzzling, since there are fewer hydrogen bonds between 2γ-lactose (higher binding affinity) than between 1α-lactose (lower binding affinity). In addition, while the experimental study[7] shows that site 2γ is stronger than site 1α, a previous short time free energy perturbation calculation gave an opposite conclusion and strongly favors site 1α.[23] Here, we want to take advantage of long-time molecular dynamics (MD) simulations to provide a dynamic picture of the protein−ligand interaction beyond the crystal structure.

There are many rewards for these types of quantitative studies. To date, there have been some successful vaccines[24,25] that target the relatively well-understood RTA, but vaccines against RTB have been less successful. A further study on the structure and sugar binding of RTB may provide clues on vaccine designs against RTB. Also, fine understanding of the interaction could potentially provide insight into tailor-design lectins for specific cell targets (such as in cancer cells). This may lead to a selective operation of enzyme delivery into targeted cells with distinct surface sugar properties. Available crystal structures of ricin-like CRD extracted from various plants, such as TKL1 (1GGP), abrin-a (1ABR), mistletoe lectin (2MLL), and ricin (2AAI), are valuable to this endeavor.[26] Multidomain lectin seems to be a common feature of many sugar receptors. In fact, the use of cooperative multidomain proteins has many possible advantages, especially in the field of signal transduction.[27,28]

METHODS

In this work, we performed a series of molecular dynamics simulations of ricin−sugar complexes. The systems examined are (1) **apo**: apo form of RTB, (2) **cplx**: RTB complexed with lactoses at both site 1α and site 2γ, (3) **l1**: RTB with one lactose at site 1α, and (4) **l2**: RTB with one lactose at site 2γ. We also investigated another RIP-II, (5) **tkl1**, the chain B of TKL1 for comparison. Crystal structures of ricin chain B complexed with two lactoses(PDB 2AAI[17]) and TKL1 chain B (PDB 1GGP[29]) were used. The initial structures of monoligand (**l1** and **l2**) and apo forms of the ricin were generated by deleting the complementary ligand(s) in the **cplx** structure.

The preparation and molecular dynamics (MD) production runs were performed using Amber10.[30] AMBER99SB force field[31] was used to model the protein and counterions Na^+. GLYCAM 06.f force field[32] and TIP3P[33] were used for the sugar ligands (lactose, a disaccharide β-D-galactopyranosyl-(1−4)-D-glucose) and water molecules, respectively. Hydrogens were added to the

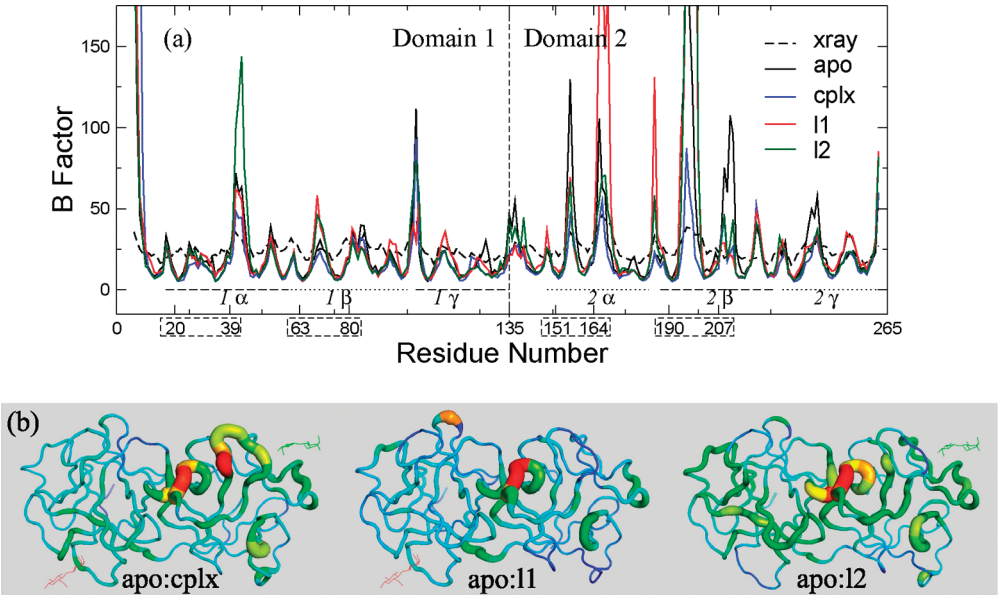


Figure 2. (a) *B*-factors of C_α from X-ray data and computation results for **apo**, **cplx**, **l1**, and **l2**. Residues involved in disulfide bonds are shown in boxes. Homologous subdomains α, β, and γ for both domain 1 and 2 are marked accordingly. (b) C_α *B*-factor ratios were shown in a tube representation by color-coding for **apo**: **cplx**, **apo**: **l1**, and **apo**: **l2**. The values are labeled from smallest to largest by color, from blue to red, and by increasing tube width. The corresponding ligands (red at site 1α and green at site 2γ), if present in the ligand-bound states, are shown.

PDB structures via the LEAP module of AMBER. The protein chain RTB used in the simulation is 257 residues long. The first five residues of the N-terminus of RTB (a very floppy region linked to RTA) were not included in the system. There are 3948, 4038, and 3993 atoms in **apo**, **cplx**, and **l1** or **l2** systems, respectively. Each system was solvated initially with at least 1 nm water layer on each side. That is, 11 281, 11 284, 10 685, and 10 683 water molecules were added to the **apo**, **cplx**, **l1**, and **l2** systems, respectively. There are two histidine residues in RTB. One of them, His251, is important for binding site 2γ. Both are set to be singly protonated at $N_{\varepsilon 2}$. Finally, systems were neutralized by adding 10 Na^+ counterions. Minimization was performed to remove the bad contacts in the initial structure. After the initial minimization and heating to 300 K, a constant pressure (1.01×10^5 Pa) run (NPT) allowed the system to reach equilibrium at 1 atm. The final box dimensions (in Å) are $65.5 \times 70.6 \times 82.0$, $65.5 \times 70.6 \times 82.0$, $64.7 \times 70.3 \times 79.0$, and $64.7 \times 70.3 \times 79.0$ for **apo**, **cplx**, **l1**, and **l2** systems, respectively. In the final stage, we performed a 54 ns constant volume simulation (NVT) at 300 K for each ricin system (**apo**, **cplx**, **l1**, and **l2**). Langevin thermostat[34] was used to regulate the temperature. Bond lengths between hydrogens and heavy atoms were constrained using SHAKE.[35] The particle mesh Ewald method was employed to calculate the long-range interactions. A 2 fs time step was used. Frames were collected at 1 ps interval. The first 4 ns of production runs was not used in the analysis. Each analysis was based on the final 50 ns (50 000 snapshots). Another 50 ns simulation was further performed in the case of **cplx**.

The MM-GB/SA (molecular mechanics, generalized Born model plus solvent-accessible surface area) approach was used to estimate the binding energies.[30] The MM-GB/SA calculation was performed on each snapshot of the complex form. The default surface tension of 0.005 kcal/(mol Å^2) is used for the SA calculation. We calculated the direct potential energy between ligand and protein using Amber force field, the so-called molecular mechanics (MM) part. It is also termed the bare energy as no solvent interaction is directly included. For polar and nonpolar solvation free energy, we separated the complex to its two components and performed GB and SA calculations three times, the difference between the complex and the sum of the components gives the values of the solvation energy. The binding energy, $\Delta G = \overline{G}\,(\text{complex}) - \overline{G}\,(\text{ricin}) - \overline{G}\,(\text{lactose})$, for the ricin−lactose interaction was obtained.

RESULTS AND DISCUSSION

In this work, we performed 50 ns MD simulations on different **apo** or ligand-bound states of ricin (**cplx**, **l1**, and **l2**). We further performed another 50 ns on the double-ligand form, **cplx**, to check the statistics. We examine all three aspects of lectin−sugar interaction in the following order. First, we want to investigate whether there are any important features of the protein conformations before and upon ligand recognition. Second, we study the direct interaction between the sugar and the lectin. The ligands in **cplx** and **l2** stayed in the pocket throughout the corresponding simulations. However, in the course of the simulation, at ~13 ns, the ligand in **l1** moved away from the pocket. This prompted us to impose a ligand−protein distance constraint to finish the remainder of the 50 ns sampling of this single ligand-bound form. The constraining force was turned on when the distance between the galactose-C4 and Asp17-C_α exceeded 10 Å. We present the changes of lactose conformations upon binding at the end.

Protein Stability. One interesting question is how these two binding pockets are altered upon (multi)ligand binding. Does the binding pocket change its flexibility or remain the same? There are various mechanisms that could monitor potential configurational changes upon binding, such as "lock and key", conformational selection, and induced fit models. Is the property of one pocket affected by the other one? To answer these questions, the flexibility changes of ricin induced by ligand(s) were monitored using the *B*-factor (Figure 2a). The Debye−Waller factor, commonly known as *B*-factor, indicates locality of an atom and

dx.doi.org/10.1021/bi102021p |*Biochemistry* 2011, 50, 4097–4104

can be calculated from the mean-square fluctuation using $B = 8\pi^2\langle\Delta r^2\rangle/3$. As shown in Figure 2a, we can easily gauge the qualitative differences among the *B*-factors of C_α for **apo**, **cplx**, **l1**, **l2**, and X-ray. Here, we labeled (boxes) the cysteines forming disulfide bonds. Not surprisingly, the regions near these four bonds have lower *B*-factor values and support the structural integrity of RTB. The peaks and valleys of *B*-factor curves of X-ray and those of the simulations are overall consistent. The difference between the *B*-factor data of X-ray and that of **cplx** could be attributed to the crystal packing present in the X-ray structure.[36] We have used block analysis to ensure the convergence of obtained *B*-factor values. These values reported from different parts of the long simulation are quite consistent except for those at the tip of the loop, which have large values and high variance, e.g., around residues 156, 167, and 198.

Moreover, residues of the unliganded (**apo**) and monoligand forms (**l1** and **l2**) have higher *B*-factor values than those of the double-ligand (**cplx**) form (Figure 2a). This holds true especially at the binding sites of the two lactoses. The relatively higher *B*-factor values in the **apo** form indicate that the corresponding residues are more flexible than those in the **cplx** form. This would suggest that the presence of the two ligands in the **cplx** form has a stabilizing effect. Comparing the flexibility of the site 1α region in the four systems, it is easy to see that their differences are subtle, and the **cplx** is the most stable form. The differences among different ligand-bound states are more significant in the 2γ region. We see a decrease in flexibility upon ligand 2 binding, especially at the regions around residues 237, 246, and 256. Overall, one can also see that there is a cooperative effect of binding on the flexibility change. For example, *B*-factors of site 2γ are further decreased with the presence of the ligand bound to site 1α, and vice versa.

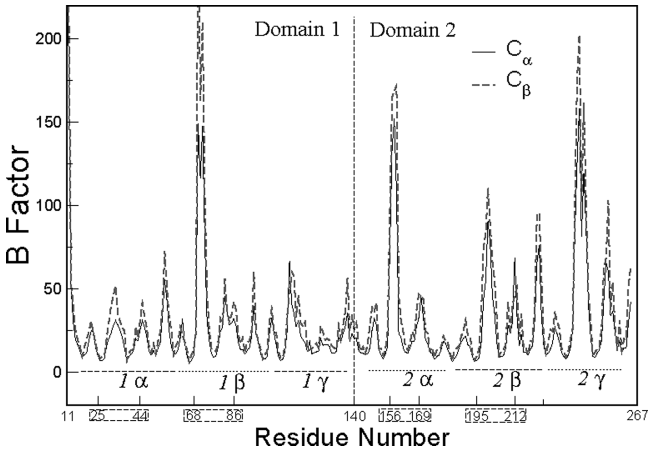


Figure 3. Computed *B*-factors of C_α and C_β for chain B of **TKL1**.

We display the *B*-factor ratios—**apo**: **cplx** (Figure 2b left), **apo**: **l1** (Figure 2b middle), and **apo**: **l2** (Figure 2b right)—of C_α for each residue. Instead of dividing the *B*-factor values of the liganded systems (**cplx**, **l1**, and **l2**), with the *B*-factor values of the apo form, we used the apo form value as the numerator to obtain a more informative picture stressing the regions that are more localized upon binding. Here, the ratio is color-coded on the structure of the protein. The values are expressed from blue (smallest) to red (largest). The portion rendered in red are regions that are stabilized upon ligand binding. As such, it is interesting to note that sugar binding in all the liganded systems studied here have a stabilizing effect on residues 120−126 (in domain 1) and 196−212 (in domain 2). Also, ligand binding at site 2γ alone has a more stabilizing effect on the overall structure of RTB than ligand binding at site 1α alone.

Since TKL1 has a similar structure and function to ricin, its apo form, **tkl1**, was also investigated to see how transferable are our conclusions drawn from ricin dynamics for other type II RIP toxins. Chain B, extracted from the tuber of *Trichosanthes kirilowii*, is 58% identical to chain B of ricin. Their fold and disulfide bond locations are identical as well.[29] As shown in Figure 3, *B* factors of both the side chain and backbone of **tkl1** were larger than those of ricin. It is still puzzling whether all the R-CRDs strongly favor the same location (site 2γ for ricin) as a binding site. Since the simulation indicates a very flexible site 2γ in this case, we speculate that it is unlikely that site 2γ of TKL1 has high binding affinity.

Furthermore, we want to examine whether there are significant ligand-induced conformational changes shown at the mean-structure level. We thus calculated the displacement (Figure 4) between the mean structure of **apo** ensemble and that of **cplx**, **l1**, or **l2**. With respect to **apo**, the monoligand systems, **l1** and **l2**, have a bit higher displacement than the double ligand-bound system, **cplx** in domain 1. This could again be attributed to the stabilizing effect of the bound ligand mentioned above. But overall, it appears that all three ligand-bound complexes retain the apo structure well. In most regions, the differences are less than 1 Å, especially at the binding sites. The observed insignificant mean displacement (less than 1 Å) between **apo** and **cplx** conformation ensembles, with the exception of the loops, is consistent with observed *B*-factor values (Figure 2a). This is also

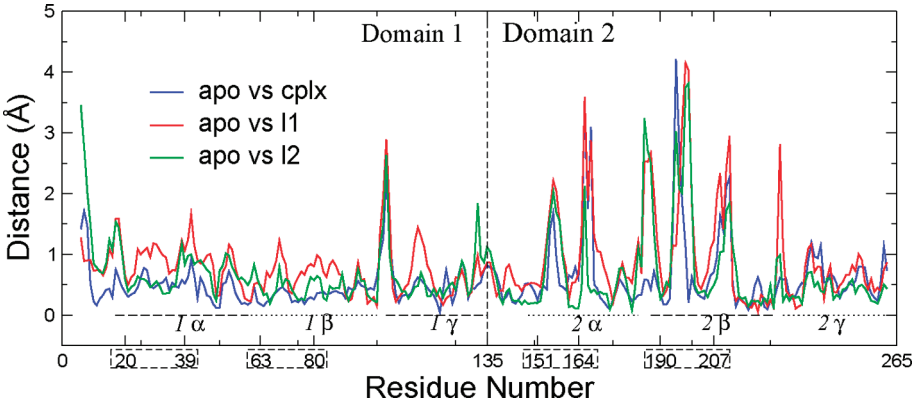


Figure 4. Distance between C_αs for the average structures of **apo** and **cplx**, **apo** and **l1**, and **apo** and **l2**.

dx.doi.org/10.1021/bi102021p |*Biochemistry* 2011, 50, 4097–4104

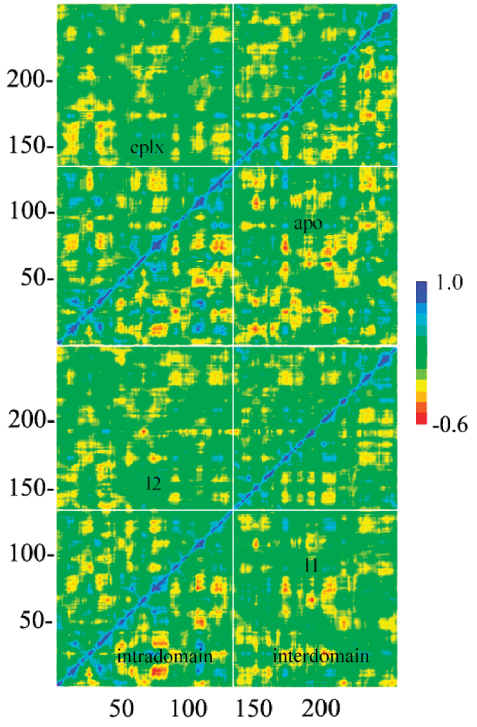


Figure 5. Correlation matrix shown in triangle forms for the **apo** (upper panel, lower triangle), **cplx** (upper, upper), **l1** (lower, lower), and **l2** (lower, upper). Residue numbers are labeled on axes.

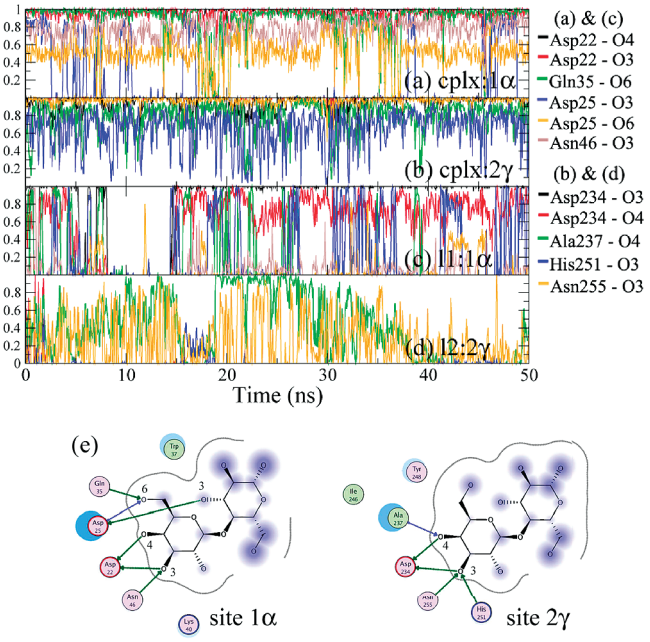


Figure 6. Ratio of hydrogen bond formed as a function of time for (a) **cplx**:1α, (b) **cplx**:2γ, (c) **l1**:1α, and (d) **l2**:2γ, obtained from 50 ns MD simulations. Each data point displayed is from 50 ps local averaging. (e) List of possible hydrogen bond interactions, at site 1α and site 1γ.

in agreement with experimental findings that there is no global conformational change upon ligand binding using fluorescence data.[19] In that study, lacking of fluorescence change was further used to support their conclusion of no allosteric effect. Instead, we show in the current work that cooperativity still exists even there is no change at the mean-structure level.

In terms of overall entropy, mass-weighted covariance matrices of heavy atoms (atoms other than H) were diagonalized to obtain standard phonon-like vibrational entropy terms by quasi-harmonic analysis;[30] total entropies of the ricin chain B (in units of kcal/(mol K)) are 10.9 (**apo**), 10.8 (**cplx**), 11.2 (**l1**), and 11.3 (**l2**). As shown, all four forms have similar entropies with **cplx** being the least flexible one.

To obtain a more detailed understanding of ricin−sugar binding, it is useful to perform a correlation analysis on ricin's conformations. The correlation is defined as the normalized covariance between two different positions, say, of particle i and j, $C_{ij} = \langle \delta\vec{r}_i \cdot \delta\vec{r}_j \rangle / (\langle \delta\vec{r}_i^{\,2} \rangle \langle \delta\vec{r}_j^{\,2} \rangle)^{1/2}$. Here, we have defined the fluctuation around the mean position of particle i as $\delta\vec{r}_i = \vec{r}_i - \langle \vec{r}_i \rangle$. At the residue−residue level covariance, the averaging over the three-dimensional components of the atoms of the residue, i.e., the elements of correlation matrix spanned, renders the correlation of motions to a number between +1 and −1, with +1 indicating a total sync of motions between these two residues. As shown in Figure 5, the correlation matrix plot at the residue level measures how the movement of one residue indexed by x-axis is correlated with the movement of another residue indexed by y-axis. Here, strongly negative (red) regions show that residues are highly anticorrelated with each other, while positive (blue) regions are the opposite and indicate motions are in sync. A near zero value colored by green indicates no correlation. Since the off-diagonal interdomain regions (interaction between domain 1 and domain 2) contain many strongly (anti)correlated regions, it suggests that the two domains are structurally one entity. Thus, one cannot separate two CRDs and study only one-half of it. This is interesting since it is against our initial intuition that the intradomain dynamics should be more correlated than interdomain dynamics. The presence of bound sugar(s), in **cplx** and **l2**, diminishes the correlation between various residues. The system **l1** has almost the same level of correlation as **apo**, especially between residues in domain 1.

Binding Stability. Besides structural analysis of protein's response to different binding states, it is important to directly examine sugar binding properties, hydrogen bonding, and binding free energy. Ligand binding properties of the two sites (site 1α and site 2γ) of **cplx**, **l1**, and **l2** were monitored in terms of hydrogen bond formation as well as their binding free energy (GBSA level). The most important interactions between lectin and sugar are the hydrogen-bonding network and hydrophobic stacking. In the current case of lactose binding, the nonreducing terminal galactose carries the main role, while the upstream glucose assists. To distinguish the atoms from the two residues, we use the symbol prime to label glucose. Specific hydrogen-bonding interactions for both site 1α and site 2γ are labeled in Figure 6e and enumerated in Table 1. In the crystal structure, as shown in Figure 6e, left panel, the protein−ligand complex forms six hydrogen bonds in site 1α. One of them is formed between protein and glucose (Asp25−O3′), and the rest are formed between protein and galactose (Asp22−O4, Asp22−O3, Gln35−O6, Asp25−O6, and Asn46−O3). Whereas in site 2γ (Figure 6e, right), the protein−ligand complex has five hydrogen bonds formed in the crystal structure. All hydrogen bonds are formed between protein and galactose (Asp234−O3, Asp234−O4, Ala237−O4, His251−O3, and Asn255−O3).

When looking at the crystal structures of the pockets of the **cplx**, it is puzzling that site 2γ, with a less number of possible

Table 1. Percentage of Hydrogen Bond Formation

site	interaction	cplx	l1	l2
site 1α				
Asp22−O4	$O_{\delta 1}\cdots HO4$	99.6	82.8	
Asp22−O3	$O_{\delta 2}\cdots HO3$	82.0	59.6	
Gln35−O6	$N_{\epsilon 2}H_{\epsilon 21}\cdots O6$	62.9	12.0	
Asp25−O3′	$O_{\delta 1}\cdots HO3'$	25.3	20.8	
Asp25−O6	NH···O6	29.5	4.2	
Asn46−O3	$N_{\delta 2}H_{\delta 21}\cdots O3$	42.4	4.5	
site 2γ				
Asp234−O3	$O_{\delta 2}\cdots HO3$	72.1		0.3
Asp234−O4	$O_{\delta 1}\cdots HO4$	85.6		1.0
Ala237−O4	NH···O4	77.1		40.7
His251−O3	$N_{\epsilon 2}H_{\epsilon 2}\cdots O3$	54.3		1.5
Asn255−O3	$N_{\delta 2}H_{\delta 21}\cdots O3$	53.7		26.7

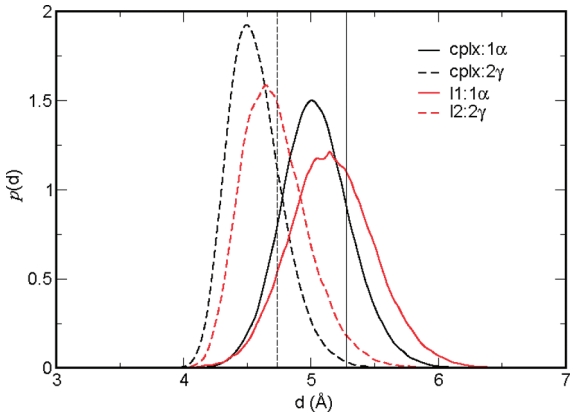


Figure 7. Probability distribution of the ring-to-ring distance, *d*, between the galactose ring and the aromatic ring of Trp37 or Tyr248 in **cplx**:1α, **cplx**:2γ, **l1**, and **l2** obtained from snapshots.

hydrogen bonds than site 1α, has a higher sugar binding affinity than site 1α. As such, we tracked the hydrogen bonds formed throughout our MD simulations. Here, we examined whether each potential hydrogen bond listed above persists throughout the 50 ns MD simulations.

Figure 6a−d presents the hydrogen bonds formed at site 1α and site 2γ for **cplx**, **l1**, and **l2**. Each data point is a local average of 50 ps for an easier display. As shown, hydrogen bonds form and break during the 50 ns MD simulations. Some of those motions are at the time scale of nanosecond. This dynamics plot indicates, especially for hydrogen-bonding properties, that short time nanosecond or less MD simulations may not capture the sugar binding phenomenon as complete. Percentage of hydrogen bond formation is shown in Table 1.

By comparing hydrogen-bonding statistics in the 50 ns MD simulations performed for **cplx**, **l1**, and **l2**, it is clear that the essential hydrogen bonds in site 1α are formed by Asp22−O4 and Asp22−O3, while in site 2γ, these are formed by Ala237−O4 and Asn255−O3 (Table 1). This would suggest that OH4 and OH3 hydroxyl groups are the most critical sugar recognition sites for both 1α and site 2γ. This is in agreement with the experimental results of Niwa and co-workers.[37] If we look at the difference in the hydrogen-bonding dynamics between the monoligand (**l1**, Figure 6c; **l2**, Figure 6d) and two-ligand (**cplx**, Figure 6a,b) systems, it is clear that the possible hydrogen bonds enumerated in the crystal structure (Figure 6e) persist throughout the 50 ns MD simulations in **cplx** system but not in the monoligand systems, **l1** and **l2**. This would suggest that the presence of lactose in both site 1α and site 2γ has an overall positive cooperativity effect. Also, Figure 6a−d supports our notion that Asp22−O4 (black) and Asp22−O3 (red) and Ala237−O4 (green) and Asn255−O3 (orange) are the essential hydrogen bonds in site 1α and site 2γ, respectively. A constraint was added in **l1** system as mentioned earlier; there is a brief area of no bonding in Figure 6c.

Besides hydrogen bonding, we also investigated the hydrophobic packing effect between the sugar ring and the aromatic residue. Here, we define the ring-to-ring distance *d* as the distance from the center of the galactose ring to the center of the six-membered ring part of the indole of Trp37 for 1α or the benzene ring of Tyr248 for 2γ. We show the probability distribution of the distances $p(d)$ in Figure 7. The mean ring-to-ring distances are, in increasing order, 4.57, 4.72, 5.04, and 5.17 Å for **cplx**:2γ, **l2**:2γ, **cplx**:1α, and **l1**:1α. Note that by doing block analysis, we ensure the reported results are statistically robust for this and other properties. For example, in the case of **cplx**, the difference between the mean ring-to-ring distance for the first and second half of the 100 ns simulation is less than 2%. A shorter distance for site 2γ presents a more favorable configuration for ring hydrophobic stacking interaction than that for site 1α. The corresponding distance distributions are wider in the monoligand form than those in the double-ligand form. This also indicates that the interactions in the bound form are stronger in the **cplx** form than those in the monoligand forms and thus suggests the cooperativity of the binding (Figure 7).

A previous thermodynamic perturbation calculation got entirely incorrect results as they favor the weaker binding site 1α over 2γ by more than 60 kcal/mol.[23] In the current study, we did not perform the perturbation calculation which spreads the efforts along "reaction" pathway. Rather we focus on all the sampling at the end-point states, the apo and complex states. Following the molecular mechanics calculation from snapshots of 50 ns trajectories, we found the bare binding energies are the following: −91.1 ± 9.2, −75.8 ± 11.4, −75.7 ± 24.5, and −0.9 ± 10.3 kcal/mol for **cplx**:1α, **cplx**:2γ, **l1**, and **l2**, respectively. With GBSA solvation energy added, we found that for **cplx** site 1α's binding energy is $\Delta G_{mmgbsa} = -34.5 \pm 4.5$ kcal/mol, while site 2γ's binding energy is −26.2 ± 9.6 kcal/mol. We thus conclude that when both ligands are bound, the energies of the two sites are similar. In the cases of single ligands, for system **l1**, the binding energy is −24.3 ± 10.3 kcal/mol. While for system **l2**, 2γ site has −0.2 ± 2.1 kcal/mol. The high fluctuation of binding energy in system **l1** may be caused by the transient escape of the ligand around 14 ns. It is difficult to simply compare the two single bound energies as in the case of **l1**; a constraint was enforced to prevent ligand escape. In contrast, no constraints were added in **l2**. From the fact that **l2** can hold the ligand for the whole run (50 ns) while the ligand escaped in less than 14 ns in **l1**, it is tempting to suggest that site 2γ is stronger than site 1α. Of course, an accurate, statistically meaningful setup involves a much longer time scale simulation involving sugar's association and escape to fully sample the bound and unbound forms of the complex. The time scale required will be at least 2 orders longer than that of this study. However, it is certain that the current results suggest that the binding strengths of monoligand systems are much weaker compared to that of the **cplx** case. This is an indication of cooperative binding, a very important concept for allosteric mechanism of signal transduction.[27,28]

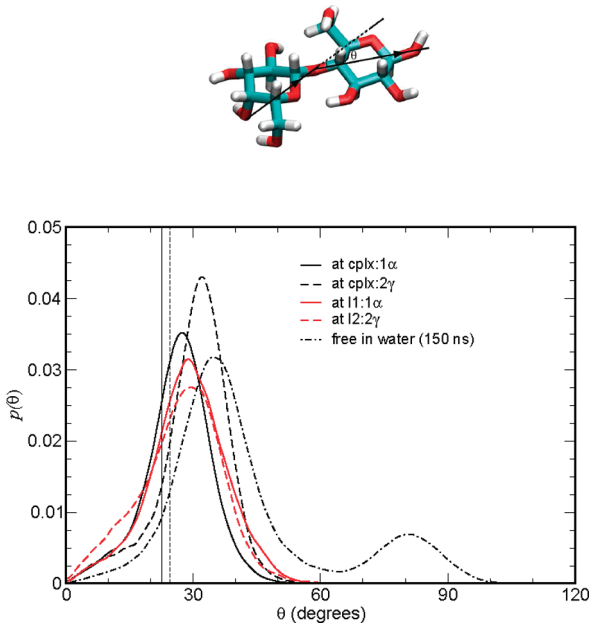


Figure 8. Probability distribution of the super backbone angle θ of lactose in four protein-bound forms and free in solution. The vertical lines indicate the values of θ in the crystal structure 2AAI.

For example, the binding cooperativity is achieved when the interaction strength of molecules A and B is changed as a consequence of molecule A being bound to molecule C. Apparently, the physical reason is that, through interaction, molecule C alters the distribution of molecule A's conformations. The first order effect of a probability distribution comes from its mean value, thus the mean structure, which is the location of the ground state, the minimum of the free energy basin. If the presence of molecule C could significantly change the ground state, then an allosteric effect emerges. Beyond the first order change, the second moment of a distribution of the conformations indicates the flexibility. In this current case, even though the first moment (ground state) did not change much, allosteric effect can come from the change of the second moment of the distribution of the conformations. Our results show that there is little change in the mean structure, and the strengthened interaction comes from the decreased flexibility at site 1α upon site 2γ's binding and vice versa, site 2γ upon site 1α's binding. We believe this type of cooperativity is difficult to detect by fluorescence signal since fluorescence signal is sensitive to the major conformational changes between different free energy basins, not the level of the fluctuation strength of the conformation ensemble.

Lactose Conformations. There are three aspects of protein−sugar interactions. The current work focuses on two aspects: the viewpoint of the protein's respond and direct protein−sugar interaction. We now briefly present the third aspect, sugar's responses.

As mentioned in a previous work, the backbone flexibility of polysaccharides can be measured using the super backbone angle θ, which is defined by the angle of direction changes of the sequential backbone linkage oxygen.[38] Specifically, in this case, we label the position of atom O1′, O4′(=O1), and O4 with $\vec{r}_a$, $\vec{r}_b$, and $\vec{r}_c$, respectively. We thus have the direction vector $\vec{n}_{ij} = (\vec{r}_i - \vec{r}_j)/|\vec{r}_i - \vec{r}_j|$. We can then define $\theta = \arccos(\vec{n}_{ab} \cdot \vec{n}_{bc})$. Thus, a fully extended backbone gives $\theta = 0$. We calculated θ for each frame of the complex simulations and obtained the probability distribution $p(\theta)$. For comparison, we also collect the distribution of θ for free lactose in water.

As shown in Figure 8, we see that ricin keeps the disaccharide's backbone in a more extended form, compared to that of free in water. In the latter case, the simulation of the lactose in water, we observed that lactose starts to explore another basin besides the basin of the bound form. The new basin gives lactose a more curved backbone. Note that due to the infrequent transitions between the two basins in the simulation (only a few transitions observed in 150 ns) of lactose alone, we are certain that the two-basin feature of the landscape of lactose conformation and that ricin suppresses lactose being in the more curved conformations, but are less so on the relative free energy difference between the two basins for the case of free lactose. A much longer and/or specialized simulation can further clarify the protein-free lactose free energy landscape. For the comparison among four different protein-bound forms, again we see that the monoligand form has a more dispersed distribution than the double-ligand complex forms. This confirms that the cooperativity between the ligands strengthens the binding.

We also examined the sugar ring conformations, which can be in various puckering states, from twist boat to chair.[39] Generally, sugar-interacting proteins, such as glycosidases, may strongly influence the sugar ring. However, in this case, all the sugar rings are in the usual chair conformations. No significant distortion of sugar conformations was detected.

CONCLUDING REMARKS

In this work, we performed 50 ns MD simulation on each of the four binding states of RTB system. Statistics and dynamics information on the systems, including the protein's responses, were obtained. We focused on examining the protein in different sugar-bound states and the cooperativity of binding. The origin of the cooperativity was examined in detail. On the stability of ricin, we showed that each sugar is much more stably bound in the presence of the other sugar. Rather than relying on allosteric effects of changing pocket shapes, the cooperativity seems to be originating from the decreased flexibility of one binding pocket when the other pocket has a ligand in it. We think that this flexibility effect can be very important for sugar−lectin interaction. Other similar multidomain sugar binding systems can be investigated.

In addition, the correlation matrix was able to show the structural connection between the two domains and that the protein chain B, despite being made from two homologous domains, is structurally one entity. We also found that the time scale of some hydrogen-bonding fluctuation is at several nanoseconds. As such, current study captures behavior of sugar binding at the time scale of tens of nanoseconds which is much longer than the previous 1 ns MD analysis.

Important hydrogen bonds were identified in this study; H-bonded to galactose's OH3 or OH4 are the most critical ones. It appears that using a two-residue strategy of separately forming the essential H-bonds with OH3 and OH4 of the lactose (used by site 2γ) has a better result than using one residue interacting with OH3 and OH4 at the same time (used by site 1α).

AUTHOR INFORMATION

Corresponding Author

*E-mail: tshen@utk.edu. Phone: (865)974-4088. Fax: (865)-974-6306.

■ ACKNOWLEDGMENT

Computational support was provided in part by the computer cluster at ORNL/UT-Center for Molecular Biophysics. We thank Dr. R. Ynalvez for valuable discussions.

■ ABBREVIATIONS

RIP, ribosome-inactivating protein; CRD, carbohydrate recognition domains; RTB, ricin B-chain; rmsd, root-mean-square-deviation; TKL1, *Trichosanthes kirilowii* lectin-1.

■ REFERENCES

(1) Lis, H., and Sharon, N. (1991) Lectin-carbohydrate interactions. *Curr. Opin. Struct. Biol. 1*, 741–749.

(2) McEver, R. P. (2002) Selectins: lectins that initiate cell adhesion under flow. *Curr. Opin. Cell Biol. 14*, 581–586.

(3) Falnes, P. O., and Sandvig, K. (2000) Penetration of protein toxins into cells. *Curr. Opin. Cell Biol. 12*, 407–413.

(4) Sandvig, K., and van Deurs, B. (2000) Entry of ricin and shiga toxin into cells: molecular mechanisms and medical perspectives. *EMBO J. 19*, 5943–5950.

(5) Sandvig, K., and van Deurs, B. (2005) Delivery into cells: lessons learned from plant and bacterial toxins. *Gene Ther. 12*, 865–872.

(6) Taylor, M. E., and Drickamer, K. (2003) *Introduction to Glycobiology*, Oxford University Press, New York.

(7) Sphyris, N., Lord, J. M., Wales, R., and Roberts, L. M. (1995) Mutational analysis of the Ricinus lectin B-chains galactose-binding ability of the 2γ subdomain of Ricinus communis agglutinin B-chain. *J. Biol. Chem. 270*, 20292–20297.

(8) Kwong, P. D., Doyle, M. L., Casper, D. J., Cicala, C., Leavitt, S. A., Majeed, S., Steen-beke, T. D., Venturi, M., Chaiken, I., Fung, M., Katinger, H., Parren, P. W., Robinson, J., Van Ryk, D., Wang, L., Burton, D. R., Freire, E., Wyatt, R., Sodroski, J., Hen-drickson, W. A., and Arthos, J. (2002) HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature 420*, 623–624.

(9) Rini, J. M. (1995) Lectin Structure. *Annu. Rev. Biophys. Biomol. Struct. 24*, 551–577.

(10) Lord, J. M., Roberts, L. M., and Robertus, J. D. (1994) Ricin: structure, mode of action, and some current applications. *FASEB J. 8*, 201–208.

(11) Hazes, B., and Read, R. J. (1997) Accumulating evidence suggests that several AB-toxins subvert the endoplasmic reticulum-associated protein degradation pathway to enter target cells. *Biochemistry 36*, 11051–11054.

(12) Wales, R., Roberts, L. M., and Lord, J. M. (1993) Addition of an endoplasmic eticulum retrieval sequence to ricin A chain significantly increases its cytotoxicity to mammalian cells. *J. Biol. Chem. 268*, 23986–23990.

(13) Legname, G., Bellosta, P., Gromo, G., Modena, D., Keen, J. N., Roberts, L. M., and Lord, J. M. (1991) Nucleotide sequence of cDNA coding for dianthin 30, a ribosome inactivating protein from Dianthus cariophyllus. *Biochim. Biophys. Acta 1090*, 119–122.

(14) Stahl, P. D. (1992) The mannose receptor and other macrophage lectins. *Curr. Opin. Immunol. 4*, 49–52.

(15) Dodd, R. B., and Drickamer, K. (2001) Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology 11*, 71R–79R.

(16) Wandall, H. H., Irazoqui, F., Tarp, M. A., Bennett, E. P., Mandel, U., Takeuchi, H., Kato, K., Irimura, T., Suryanarayanan, G., Hollingsworth, M. A., and Clausen, H. (2007) The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc:lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation. *Glycobiology 17*, 374–387.

(17) Rutenber, E., Katzin, B. J., Ernst, S., Collins, E. J., Mlsna, D., Ready, M. P., and Robertus, J. D. (1991) Crystallographic refinement of ricin to 2.5 Å. *Proteins: Struct., Funct., Bioinf. 10*, 240–250.

(18) Katzin, B. J., Collins, E. J., and Robertus, J. D. (1991) Structure of ricin A-chain at 2.5 Å. *Proteins: Struct., Funct., Genet. 10*, 251–259.

(19) Zentz, C., Frenoy, J. P., and Bourrillon, R. (1978) Binding of galactose and lactose to ricin. *Biochem. Biophys. Acta 536*, 18–26.

(20) Adair, W. L., and Kornfeld, S. (1974) Isolation of the receptors for wheat germ agglutinin and the Ricinus communis lectins from human erythrocytes using affinity chromatography. *J. Biol. Chem. 249*, 4696–4704.

(21) Houston, L. L., and Dooley, T. P. (1982) Binding of two molecules of 4-methylumbelliferyl galactose or 4-methylumbelliferyl N-acetylgalactosamine to the B chains of ricin and Ricinus communis agglutinin and to purified ricin B chain. *J. Biol. Chem. 257*, 4147–4151.

(22) Rutenber, E., and Robertus, J. D. (1991) Structure of ricin B-chain at 2.5 Å resolution. *Proteins: Struct., Funct., Bioinf. 10*, 260–269.

(23) Ganguly, D., and Mukhopadhyay, C. (2006) Binding diversity of the two binding sites of ricin B lectin. *Biopolymers 83*, 83–94.

(24) Olson, M. A., Carra, J. H., Roxas Duncan, V., Wannemacher, R. W., Smith, L. A., and Millard, C. B. (2004) Finding a new vaccine in the ricin protein fold. *Protein Eng., Des. Sel. 17*, 391–397.

(25) O'Hara, J. M., Neal, L. M., McCarthy, E. A., Kasten Jolly, J. A., Brey, R. N. I., and Mantis, N. J. (2010) Folding domains within the ricin toxin A subunit as targets of protective antibodies. *Vaccine 28*, 7035–7046.

(26) Olnes, S. (2004) The history of ricin, abrin and related toxins. *Toxicon 44*, 361–370.

(27) Changeux, J.-P., and Edelstein, S. J. (2005) Allosteric mechanisms of signal transduction. *Science 308*, 1424–1428.

(28) Hill, T. L. (1985) *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*, Springer-Verlag, New York.

(29) Li, M., Chai, J. J., Wang, Y. P., Wang, K. Y., and Bi, R. C. (2001) Crystal structure of Trichosanthes kirilowii lectin-1 and its relation to the type 2 ribosome inactivating proteins. *Protein Pept. Lett. 8*, 81–87.

(30) Case, D. A., Darden, T. A., Cheatham, T. E. I., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowly, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvry, I., Wong, K. F., Pasani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Kollman, P. A. (2008) AMBER 10, University of California, San Francisco, CA.

(31) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf. 65*, 712–725.

(32) Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem. 29*, 622–655.

(33) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926–935.

(34) Adelman, S. A., and Doll, J. D. (1976) Generalized Langevin equation approach for atom/solid - surface scattering: General formulation for classical scattering off harmonic solids. *J. Chem. Phys. 64*, 2375–2388.

(35) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327–341.

(36) Canino, L. S., Shen, T., and McCammon, J. A. (2002) Changes in flexibility upon binding: Application of the self-consistent pair contact probability method to protein-protein interactions. *J. Chem. Phys. 117*, 9927–9933.

(37) Niwa, H., Tonevitsky, A. G., Agapov, I. I., Saward, S., Pfuller, U., and Palmer, R. A. (2003) Crystal structure at 3 Å of mistletoe lectin 1, a dimeric type-II ribosome-inactivating protein, complexed with galactose. *Eur. J. Biochem. 270*, 2739–2749.

(38) Shen, T., Langan, P., French, A., Johnson, G. P., and Gnanakaran, S. (2009) Conformational flexibility of soluble cellulose oligomers: Chain length and temperature dependence. *J. Am. Chem. Soc. 131*, 14786–94.

(39) Cremer, D., and Pople, J. A. (1975) General definition of ring puckering coordinates. *J. Am. Chem. Soc. 97*, 1354–1358.