JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

An Integrated Computational Analysis of the Structure, Dynamics, and Ligand Binding Interactions of the Human Galectin Network

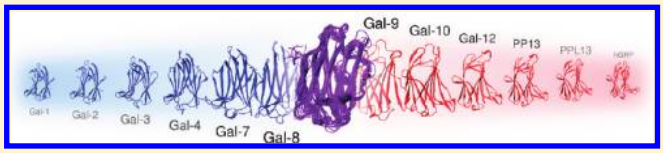
Carlos M. A. Guardia,[†] Diego F. Gauto,[†] Santiago Di Lella,[†] Gabriel A. Rabinovich,[‡,§] Marcelo A. Martí,[*,†,§] and Darío A. Estrin[*,†]

[†]Departamento de Química Inorgánica, Analítica y Química Física, INQUIMAE-CONICET and
[§]Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón II, C1428EHA Ciudad de Buenos Aires, Argentina

[‡]Laboratorio de Inmunopatología, Instituto de Biología y Medicina Experimental (IBYME), CONICET, C1428ADN Ciudad de Buenos Aires, Argentina

S Supporting Information

ABSTRACT: Galectins, a family of evolutionarily conserved animal lectins, have been shown to modulate signaling processes leading to inflammation, apoptosis, immunoregulation, and angiogenesis through their ability to interact with poly-*N*-acetyllactosamine-enriched glycoconjugates. To date 16 human galectin carbohydrate recognition domains have been established by sequence analysis and found to be expressed in several tissues. Given the divergent functions of these lectins, it is of vital importance to understand common and differential features in order to search for specific inhibitors of individual members of the human galectin family. In this work we performed an integrated computational analysis of all individual members of the human galectin family. In the first place, we have built homology-based models for galectin-4 and -12 N-terminus, placental protein 13 (PP13) and PP13-like protein for which no experimental structural information is available. We have then performed classical molecular dynamics simulations of the whole 15 members family in free and ligand-bound states to analyze protein and protein−ligand interaction dynamics. Our results show that all galectins adopt the same fold, and the carbohydrate recognition domains are very similar with structural differences located in specific loops. These differences are reflected in the dynamics characteristics, where mobility differences translate into entropy values which significantly influence their ligand affinity. Thus, ligand selectivity appears to be modulated by subtle differences in the monosaccharide binding sites. Taken together, our results may contribute to the understanding, at a molecular level, of the structural and dynamical determinants that distinguish individual human galectins.



INTRODUCTION

Galectins are a family of multifunctional lectins widely distributed in a variety of tissues and animal species.[1,2] They are commonly defined by their specificity for β-galactoside sugars and their consensus sequence referred to as carbohydrate recognition domain (CRD).[1] Mounting evidence indicates an essential role for these glycan-binding proteins at the interface of physiology and pathology.[3] In fact, galectins can modulate a variety of functions, including immune cell trafficking, T cell fate, dendritic cell biology, angiogenesis, and cell death through fine-tuning cell signaling processes.[4−8] Although all members of the galectin family have a highly homologous CRD, they exhibit considerable variations in their saccharide specificity which might result from differences in the architecture and the dynamics of the ligand-binding groove (LBG).[9] Given the divergent functions of individual galectins, it is of critical importance to dissect the structure−function relationship of these proteins with the ultimate goal of designing selective inhibitors for the treatment of neoplasic and inflammatory disorders.

Up to date, the best studied and characterized human galectins in terms of structure are galectin-1 (Gal-1),[10,11] galectin-3 (Gal-3),[12,13] and galectin-9 (Gal-9).[14−16] Several crystals were obtained in different conditions, i.e., bound states and mutant variants. In most cases, Gal-1 is taken as a reference to compare ligand binding properties.

Recently, the crystal structures of galectin-9 N-terminal CRD (Gal-9N) in the presence of poly-*N*-acetyllactosamine[15] and galectin-9 C-terminal CRD (Gal-9C) in complexes with a biantennary oligosaccharide and sialyllactose[16] were reported, yielding significant insights into the carbohydrate-binding features of galectins. Also, the structure of the C-terminal conserved domain of human GRP (hGRPC), a galectin-related protein (previously known as HSPC159 for hematopoietic stem cell precursor) was obtained in the free form.[17,18] Intriguingly, no apparent lectin activity was detected for this protein. Regarding other members of the human galectin family, the N-terminal CRD of the tandem-repeat lectin galectin-4 from mouse was crystallized in a proteomic initiative,[19] while the structure of human C-terminal CRD was obtained by NMR spectroscopy.[20]

Received: April 20, 2011
Published: June 25, 2011

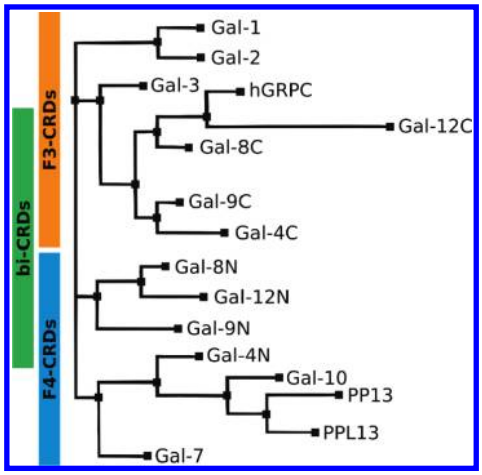


Figure 1. Proposed phylogenetic relationships between galectins based on gene and sequence analysis. Qualitative scheme of a reduced phylogenetic tree involving only human galectins taken from Houzelstein et al. work.[30] The figure is used as a presentation of the human galectin network, their classification, and phylogenetic relationships among the members that have been studied in our work.

A similar situation was observed for galectin-8 (Gal-8), where the N-terminal CRD (Gal-8N) crystal structures were available alone[21] or in the presence of lactose,[22] whereas the structure of C-terminal domain (Gal-8C) was obtained by NMR.[23] Structures for galectin-2 (Gal-2),[24] galectin-7 (Gal-7)[25] and galectin-10 (Gal-10)[26] have been also reported. None of all these structures have been described in detail or compared with other available structures so far. No experimental structures are available for the remaining human members of the galectin family, including Gal-4N, entire Gal-12, placental protein 13 (PP13), and PP13-like protein (PPL13). In fact, attempts to complete the structural study of the newest members of the family were only possible with the use of computer simulation tools when no experimental information was available demonstrating the importance of systematic studies on structure−function relationship. This is the case of PP13[27,28] where homology models are currently used for structural analysis. Yet, to our knowledge there is still no comparative and integrated approach of the structure and the ligand binding specificities of different members of the human galectin network.

Originally proposed by Hirabayashi,[29] galectins are classified into three groups based on their domain organization: (i) prototype (Gal-1, -2, -7, -10, PP13, and PPL13) consisting of one polypeptide chain with one CRD able to dimerize; (ii) tandem repeat-type including one polypeptide chain with two CRDs in tandem (Gal-4, -8, -9, and -12); and (iii) chimera-type consisting of one polypeptide chain bearing a CRD and another domain with no lectin attributes (Gal-3). Although operative, this classification does not take into account phylogeny. This issue was later considered in studies using sequence comparison and intron/exon position analysis,[30] revealing a natural classification of galectins' CRDs into two groups: F3 and F4 CRD types. Interestingly, it was found that all tandem galectins are composed of one CRD of each type: they form the bi-CRD type. The phylogenetic map integrating all human galectins and hGRPC is shown in Figure 1.

As mentioned above, from a total of 16 known human sequences, only 9 galectins have been structurally described.

Table 1. Available Galectin Structures Determined by X-ray Experiments or NMR Spectroscopy

PDB ID	group	description	resolution (Å)[a]
1GZW	F3 mono-CRD	Gal-1	1.65
1HLC	F3 mono-CRD	Gal-2	2.90
1AK3	F3 mono-CRD	Gal-3	2.10
1X50	Bi-CRD	Gal-4C	N.A.
1BKZ	F4 mono-CRD	Gal-7	1.90
2YRO	Bi-CRD	Gal-8C	N.A.
2YV8	Bi-CRD	Gal-8N	1.92
3NV1	Bi-CRD	Gal-9C	1.50
2ZHM	Bi-CRD	Gal-9N	1.84
1LCL	F4 mono-CRD	Gal-10	1.80
3B9C	F3 mono-CRD	hGRPC	1.90

[a] N.A.: Not applicable, NMR experiment.

Furthermore, neither a global structural overlook nor a comparison of the individual monosaccharide binding sites (MBS) across the whole protein family have ever been performed. To bridge this gap, we are presenting here an integrated structural and dynamical comparison of all human members of the galectin subfamily by means of state-of-the-art computer simulation techniques.

Computational techniques applied to model large biological molecules have evolved during the last decades as a fundamental tool to complement experimental information.[31] In silico generated models and the information obtained throughout their study have proved to be extremely useful for analyzing the structural data provided by the experimental methodologies. In particular, computer simulations allow a systematic and economical tool to analyze the dependence of a property of interest on static and dynamical factors and to define biologically relevant conclusions.[32]

Here, starting from the available galectin structures, we have performed classical molecular dynamics (MD) simulations[33] of free and ligand-bound human galectins to analyze protein and protein−ligand interaction dynamics. To complete the analysis of the whole family, we have built homology-based models for those human CRDs lacking an experimentally derived structure. Additionally, we estimated ligand binding interactions in a comparative way by means of computing binding free energies.

EXPERIMENTAL METHODS

Initial Structures. Initial structures of each CRD of all human galectins were retrieved from the Protein Data Bank (http://www.pdb.org) (Table 1) when available. For those CRDs whose structures were not available, we built them by homology modeling using the Modeler software.[34,35] In order to obtain the accurate templates for each target protein, structure similarity searches were done using the ModWeb server for protein structure modeling (https://modbase.compbio.ucsf.edu/scgi/modweb.cgi) (Table 2 and check PDB S1−S4 files of Supporting Information for the structures modeled). Alignments of all galectin CRD sequences were done using the CLUSTAL-W multiple alignment method and program,[36] using BLOSUM62 substitution matrix. Starting from these pairwise target templates, alignments were determined and refined, paying particular attention to the CRD fold. All the alignment trees were constructed with distance matrix method, using as distance

Table 2. Target-Template Correspondence for Missing Human Galectins CRDs Homology Models[a]

target	template (PDB ID)	group
Gal-4N	mouse Gal-4N (2DYC)	Bi-CRD
Gal-12N	Gal-8N (2YV8)	Bi-CRD
PP13	Gal-10 (1G86)	F4 mono-CRD
PPL13	Gal-10 (1G86)	F4 mono-CRD

[a] Since the C-terminal CRD of Gal-12 is less than 20% identical of the other galectin CRDs (Figure 1) and the homology models obtained showed high flexibility in the S5S6 loop, where two tryptophan residues are very close each other, we were not successful in obtaining a stable structure to study with the rest of the CRDs.

between two CRDs, their minimum root-mean-square deviation (RMSD) value.

MD Simulations. Starting from the crystal or the modeled structural model, each CRD structure was subject to the following protocol: Hydrogen atoms were added in order to saturate valences of the heavy atoms. The standard protonation state at physiological pH was assigned to all ionizable residues. Structures were then solvated in an octahedral box of explicit TIP3P model water molecules, localizing the box limits as far as 10 Å from the protein surface. The total number of atoms in each simulated system ranged from 14 500 to 17 700, including solvent molecules. MD simulations were performed at 1 atm and 300 K, maintained with the Berendsen barostat and thermostat,[37,38] using periodic boundary conditions and Ewald sums (grid spacing of 1 Å) for treating long-range electrostatic interactions with a 10 Å cutoff for computing direct interactions. The SHAKE algorithm was applied to all hydrogen-containing bonds, allowing employment of a 2 fs time step for the integration of Newton's equations as are the recommended parameters in the Amber package of programs.[39] The Amber f99SB force field parameters[40] were used for all residues. GLYCAM parameters[41,42] were employed as parameters for carbohydrate ligands. Equilibration protocol consisted of performing an optimization of the initial structure, followed by 500 ps constant volume MD run heating the system slowly to 300 K. Finally, 1 ns MD run at constant pressure was performed to achieve proper density. Different production MD runs (50 ns for the structures experimentally determined and 60 ns for the homology models) were performed. Frames were collected at 1 ps intervals and were subsequently saved on disk in order to perform analyses.

Essential Dynamics. In order to get insight into the dynamical properties of each structure and their influence on the overall structural movements, several essential dynamics (ED) analyses were performed for all production MD runs.[43] The ED for each run is determined by diagonalizing the covariance matrices (cov^T) of the atomic positions along the desired trajectory, obtaining the corresponding eigenvalues and eigenvectors (eq 1):

$$cov^T = \frac{1}{M}\sum_{k=1}^{M}\{[X_i(k) - \langle X_i\rangle][X_j - \langle X_j\rangle]\} \quad (1)$$

where the sum goes over the M configurations or snapshots from the dynamics, $X_i(k)$ corresponds to i^{th} Cartesian coordinate of the system in snapshot number k, and $\langle X_i\rangle$ represents the mean value of this coordinates. Each obtained eigenvector (v_i) corresponds to an essential mode (EM) of the protein. Taken together, all the essential modes describe the motion of the protein along the MD run used to generate the computed matrix. The eigenvalues (λ_i) obtained represent the relative contribution of each EM to the overall dynamics. EMs are ranked according to their eigenvalues and, therefore, their relative weight, being the first EM the one with major contribution or larger eigenvalue. EDs were computed only for the backbone heavy atoms (N, C, Cα, and O).

We performed a comparison between different pairs of root-mean-squared fluctuation (RMSF) profiles using the similarity index (SI), which is a correlation function between both RMSF values at a particular residue corresponding to two selected RMSF protein profiles (A and B) and normalized by dividing by the root-mean-square value of each RMSF (eq 2):

$$SI_{AB} = \frac{\sum RMSF_{A_i}\cdot RMSF_{B_j}}{\sqrt{\sum(RMSF_{A_i})^2\cdot\sum(RMSF_{B_j})^2}} \quad (2)$$

where A_i corresponds to residue number "i" of CRD "A". A value close to one means that both RMSF profiles are almost identical, whereas a value close to zero indicates orthogonal or considerably different fluctuations. Given that total residue number between CRDs is different ($i \neq j$), the dot product is computed only for those RMSF values whose residues have a correspondence between the two CRDs that are being compared, as defined by the sequence alignment done by CLUSTAL-W. The use of the EM and SIs for comparing protein dynamics has already proved a useful method in previous works from our group.[44−46]

Conformational Entropy. Conformational entropy calculations were performed by diagonalization of the mass-weighted Cartesian covariance matrix, by the Schlitter[47] and the Andricioaei and Karplus methods.[48] Since the sampling in a MD simulation depends on the length of the simulation, the calculated entropy also depends on the length of the trajectory used for the calculation. In order to obtain a value independent of the trajectory length, we calculated the entropy for intervals ranging from 4 to 50 ns, then we plotted $S(t)$ versus $1/t$, and the result was fitted using a linear regression. When time approaches to infinite, the intercept corresponds to the S_∞, the conformational entropy parameter independent of the MD simulation. This procedure has been described in a recent publication.[49] For the calculations, we considered only the heavy atoms of all the aminoacid residues of the protein, except three aminoacid residues of both carboxyl and amino terminal regions. This last consideration was taken into account due to the high flexibility of the mentioned residues.

In Silico Construction of Octameric Complex of the Human Gal-9 N-Terminal. Nagae et al. have reported several crystal structures of the human galectin-9 N-terminal CRD in complex with different *N*-acetyllactosamine (LacNAc) oligomers.[15] They found two different crystal structures with LacNAc dimers (PDB IDs 2ZHK and 2ZHL) and trimers (PDB IDs 2ZHM and 2ZHN). In crystal form 2ZHM, they found four protein−ligand complexes (designated complexes A-D) that were divided into two classes by the difference in the interaction modes between the protein and the carbohydrate. The protein molecules in complexes A and B recognize the LacNAc unit of the reducing end, while complexes C and D interact with the middle LacNAc unit. In crystal structure 2ZHN, only one complex was observed, where the second protein recognizes the nonreducing end of the LacNAc trimer. We have aligned the complex A from 2ZHM with the complex from 2ZHN using visual molecular dynamics

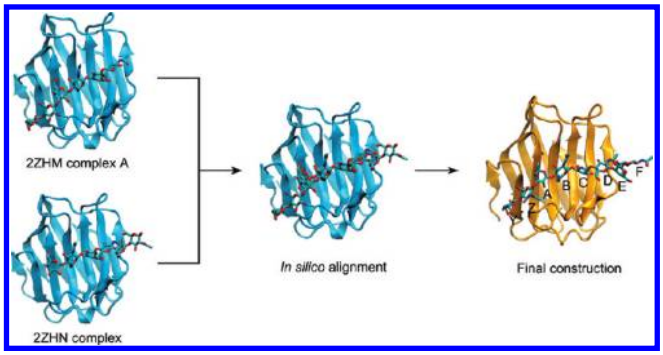


Figure 2. In silico construction of tetra-LacNAc bound human galectin-9 N-terminal. Construction of bound Gal-9N and final structure of the complex showing the total occupation of the LBG and the Y−F MBS. The octameric ligand consists of four repeated subunits of 3-linked *N*-acetyllactosamine disaccharide (LacNAc: β-D-galactopyranosyl-(1→4)-*N*-acetyl-D-glucosamine).

(VMD) tools. The protein structures were superimposable with RMSD value of 0.27 Å, while the second and third LacNAc units from nonreducing extrem of complex A from 2ZHM with the first and second LacNAc units from nonreducing extrem of complex from 2ZHN were superimposed almost perfectly, respectively (Figure 2). From this alignment, we recorded the values of the positions of each atom into a single file of coordinates in order to generate an in silico model of Gal-9N bound to a tetra-LacNAc polysaccharide which occupies the entire groove.

Defining the MBS and Free Energy Calculations. The eight MBS are named Y, Z, A, B, C, D, E, and F with the reducing end usually located in MBS D. Typically LacNAc units bind galectins occupying sites C and D, as observed for LacNAc structures of Gal-1, -3 y -7. Sites E and F are almost completely outside the protein and in the solvent, as can be shown in Figure 2; they were not included in the analysis of structure but in the free energy calculations. To define which residues contribute to shape each MBS, we aligned all CRDs on the Gal-9N octasaccharide complex and defined that a given residue contributes to a MBS when the distance between any heavy atom of the corresponding residue (in the target CRD) was at least less than 5 Å from any heavy atom of the corresponding bound monosaccharide in Gal-9N. Thermodynamic parameters for MD simulations of CRD:ligand complexes were calculated using the single trajectory molecular mechanics/generalized Born surface area (MM/GBSA) approach,[50−52] implemented in the Amber 9 package.[53] It combines molecular mechanical energies, continuum solvent approaches, and solvent accessibility in order to elicit free energies from structural information avoiding the computational intricacy of free energy simulations. The molecular mechanical energies were determined with the *sander* program from Amber and represented the internal energy (bond, angle, and dihedral) and van der Waals and electrostatic interactions.[54] An infinite cutoff for all the interactions was used. The electrostatic contribution to the solvation free energy was calculated with a numerical solver for the generalized Born method.[55,56] Energetic contributions were computed corresponding to the electrostatic energy (ELE) and van der Waals contribution (vdW). Solvation free energy was estimated using the generalized Born approximation (GB_{Solv}), which is based on the use of a cavitation and electrostatic energy components. The total free energy contribution computed by the generalized model is also presented (GB_{Tot}).

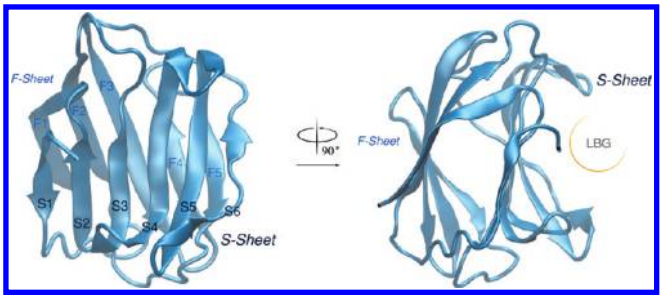


Figure 3. Schematic view of three-dimensional structure of Gal-1. The structure of Gal-1 is presented here as an example of the conserved fold of human galectins. It is also depicted the nomenclature used for each strand and for LBG.

RESULTS

In order to assess structure−function relationships among different members of the galectin family, we first compared the overall structure of all human galectin CRDs. Then, we analyzed the dynamical behavior of the different CRD domains of all human galectins. Finally, we performed a detailed comparison of each CRD ligand binding groove.

Global Comparison of Galectins. The global structure of a galectin CRD consists of about 135−140 residues forming a β-sandwich secondary structure consisting of two slightly bent sheets (Figure 3). The concave side is formed by six strands called S1−S6, and the convex side by five strands called F1−F5, except for some of the members which have an extra F strand at the beginning of the protein sequence (Gal-3 and Gal-8N, for example). Along the protein sequence the strands are ordered S1−F2−S3−S4−S5−S6−F3−F4−F5−S2−F1, with variable long loops connecting them. Loops are named as the two β-strands that the loop connects between. The concave side forms the groove in which carbohydrate binds, called the LBG.

In order to provide a first comparison of all CRD from human galectins subfamily, a multiple sequence alignment is shown in Figure 4. On average, we found 29% of sequence identity and 47% of similarity (conservative substitutions). In this way, human galectins are another example of a protein family with a relatively low sequence identity level and a highly similar 3D structure.[57,58] The crystallographic assessment suggests that several highly conserved amino acids play key roles in carbohydrate recognition and binding. For Gal-1 they are His44, Asn46, Arg48, Val59, Asn61, Trp68, Glu71, and Arg73. In the alignment shown in Figure 4, it is clear that only four residues are highly conserved in the entire family (G25, W80, E83, and R125, indexed as alignment positions), while the rest vary considerably. Interestingly, mutagenesis studies in Gal-1 showed that conserved tryptophan residue in the LBG is not essential and can be substituted by Tyr or Phe.[59,60] As evident, an aromatic ring is fundamental to ensure ligand binding, guaranteeing the correct stacking between the carbohydrate and the amino acid side chain.

Turning to the structural comparison of all CRD from human galectins family, we first performed rmsd-based structural alignments for all experimentally determined structures and also all the present studied human Gal CRDs. The RMSD-based structural alignment was performed with the stamp method as described in the Experimental Methods Section. For this last purpose, average structures of each system were obtained from 50 ns long MD simulation, started either from the X-ray- or homology-based models. The resulting average structure

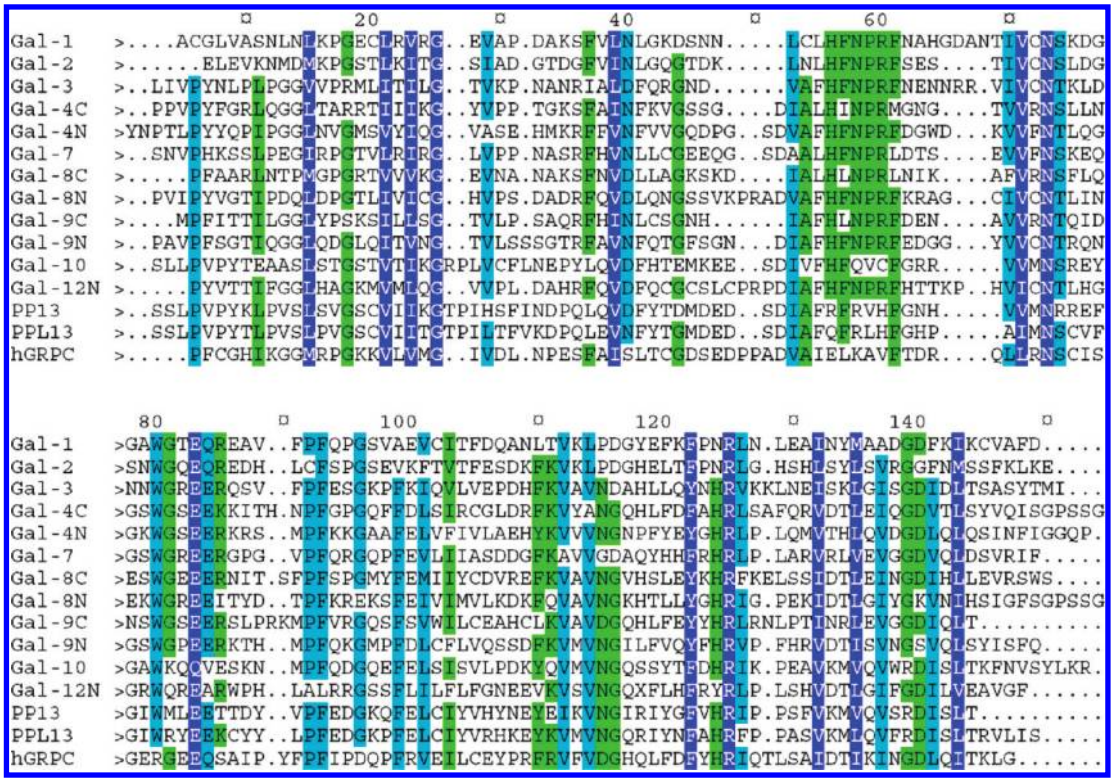


Figure 4. Multiple sequence alignment of the human galectins studied. Gal-4C, -8N, -9N, and -12N sequences were truncated by eight residues to induce the correct alignment on CLUSTALW subroutine implemented on VMD. Strictly conserved residues (blue background) and homologous residues (85% of conservation in cyan and 70% of conservation in green background) are shown. This figure was produced with GeneDoc (http://www.psc.edu/biomed/genedoc).

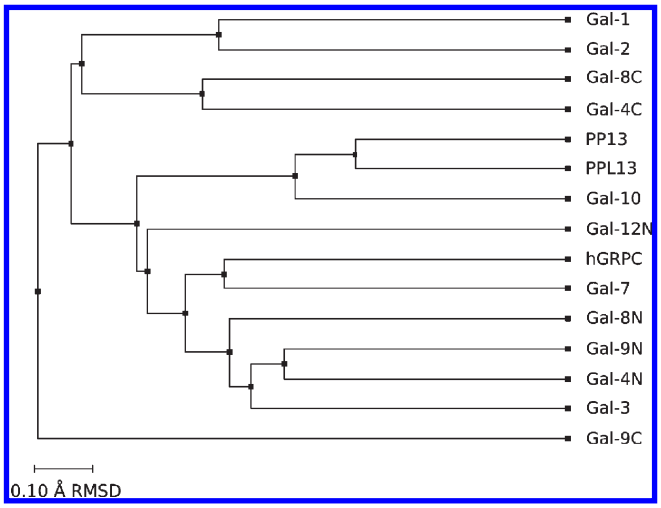


Figure 5. Structural alignment of the average secondary structure obtained from the MD simulation of each galectin. RMSD-based tree representation of the average structures from 50 ns long MD simulations aligned in silico using the CLUSTALW algorithm.

RMSD-based tree is shown in Figure 5 (RMSD-based tree for only experimental structures is shown as Figure S1 and complete alignment results as Table S1 in Supporting Information).

Overall available experimental CRD structures are very similar; the maximum RMSD value is 2.16 Å between Gal-1 and Gal-10. The small differences are maintained after MD relaxation (max. RMSD = 2.18 Å). During the time scale of the simulations, all proteins also remain close to the starting experimental or model structure (max. deviation is 1.87 Å). Interestingly, for those proteins modeled by homology, the possible structural bias produced by the intimate connection between target and template structures was eliminated during the MD simulation, as shown by the fact that the models do not necessarily show higher similarity to their templates than to other structures. Thus, the MD relaxation process seems satisfactory, and the final modeled structures are not affected by the choice of the particular template.

The comparison of the structure based tree (Figure 5) with that based on phylogenetic alignment (Figure 1) shows some similar qualitative trends but notable differences. For example, the F3-CRD Gal-1, -2, -8C, and -4C group is in a different branch from the one observed for the rest of the proteins, mainly F4-CRD type, but some members are clearly missing and in completely different branches (Gal-3, hGRPC). Also interestingly, Gal-9C clearly diverges from the expected behavior based on the sequence information and is allocated in a single branch, separated from the rest of the galectin family. Therefore, the first conclusion of the present analysis is that the global structure of all CRDs is very similar (max. RMSD of only 2.18 Å, which is a value of the same magnitude of that obtained for comparison of a particular protein crystallized in different conditions), and the structural similarity does not seem to follow either the inferred phylogenetic history or the domain organization-based classification.

Although the data shown above demonstrates that galectins CRD structure is highly conserved on average and from the global viewpoint, important differences exist between each CRD,

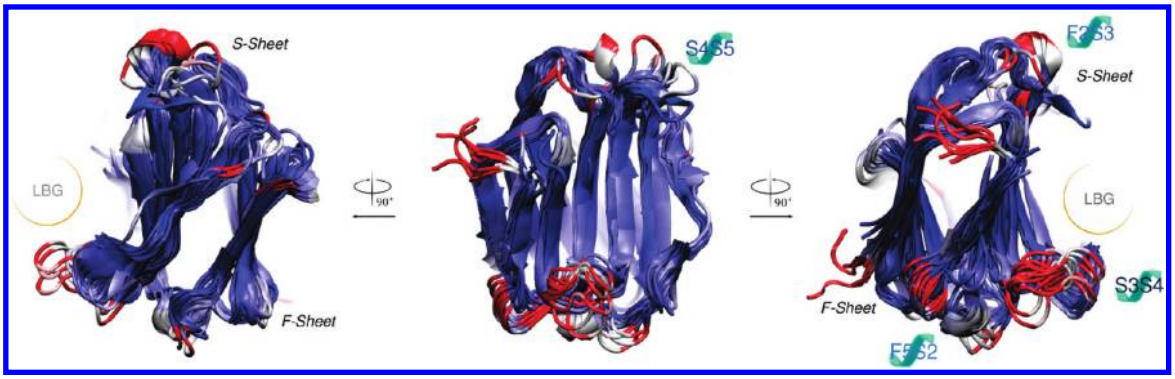


Figure 6. Structural comparison of Gal CRD structures. Blue-colored regions correspond to low RMSD values, while red denotes higher values (Gal-1 fold was used as reference). The most important loops are also indicated, named as the two β-strands that connect, using the nomenclature presented in Figure 3.

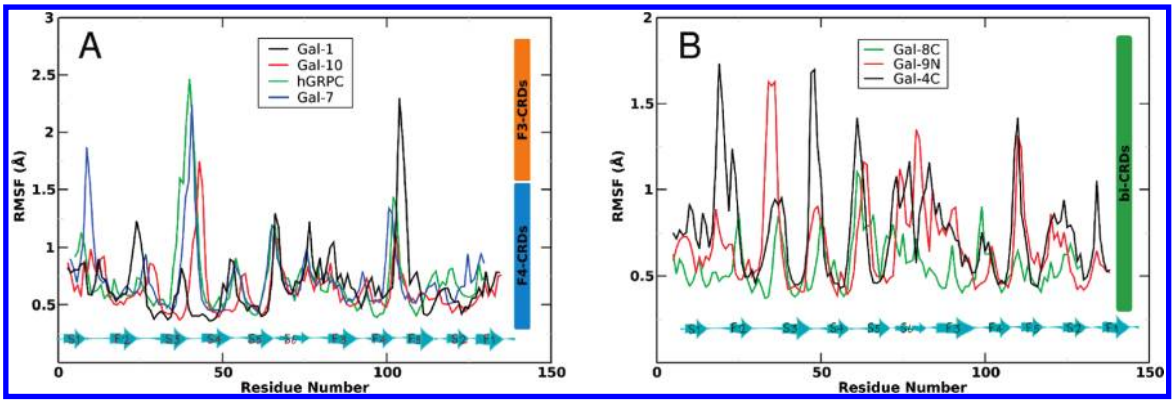


Figure 7. RMSF vs residue plot for selected human galectins. (A) mono-CRD and (B) bi-CRD galectin types.

which are localized in specific structural regions. Figure 6 shows structural alignment of all human galectin CRD average structures colored according to the relative difference using Gal-1 CRD fold as the reference. The figure shows that most of the tertiary structure is almost indistinguishable between them, and significant differences are observed only in the loop regions.

When comparing the CRDs variable loops, it could be observed that among all the loops, those with more conformational variability are F5S2, F2S3, S4S5, and S3S4 sorted in increasing order of variability. The difference in the F5S2 loop is that Gal-1 and -2 are the only ones that deviate significantly from the other CRD conformations. The shift of this pair of galectins (42% sequence identity, 59% similarity) is possibly due to the types of amino acids involved in the loop that do not allow the formation of the small α-helix seen in the majority of the other proteins (see Table S2 in Supporting Information). For the F2S3 loop, the loop structures can be divided into three groups: Gal-9N on the one hand, then Gal-10, PP13 and PPL13 on the other hand, and the rest of CRDs in the middle of both conformations. The grouping of the three F3-CRD type proteins is due to the fact that they exclusively present a α-helix in this portion of amino acids. For the N-terminus of Gal-9 fold, a similar loop is present, although showing an extra residue, which forces the sequence to adopt a different structure. For the S4S5 loop, large differences are found between proteins presenting loops pointing in the direction of the LBG and those with the loop facing entirely the opposite direction. Finally, variability is substantial in the S3S4 loop. Five different groupings can be distinguished: (i) proteins showing the smallest S3S4 loop: Gal-1 and -2, folded in exactly the same way; Gal-3 and -9C, showing slight differences; (ii) Gal-10 and PP13, and a little different PPL13, form a second group; (iii) the loop in Gal-4N and -8C points to the solvent and form a third cluster of S3S4 loop conformation; (iv) the 6 amino acid length loop in Gal-4C, -7, and -9N is mostly hydrophilic, providing a particular conformation; and (v) the longest S3S4 loop is observed for Gal-8N, -12N, and hGRPC, whose conformations also deviate from those adopted by the other proteins.

These data clearly show that human galectin CRD fold is structurally highly conserved. Main structural differences are located in specific loop regions, which allow clustering the different CRDs in common structural groups. However, no correlation with sequence could be inferred from the phylogenetic grouping.

Dynamical Behavior of the Different Human Galectin CRDs. Next, we focused our comparison on the dynamical behavior of all galectins CRDs. In order to get a first insight into domain dynamics, we computed the RMSF of each member. The RMSF value shows the dispersion of the atomic position of each Cα along the simulation and is plotted against residue number, giving an insight into the flexibility/mobility of each segment of the protein. A first look at the data clearly shows that, as expected, main flexible areas of the CRDs are located in the loops. However, the magnitude significantly differs between the loops in the same protein, and the resulting pattern differs within each CRD (Figure 7).

Table 3. Essential Modes (EM) in Human Galectins Mono-CRD Type

CRD type	F3				F4			
protein	Gal-1	Gal-2	Gal-3	hGRPC	Gal-7	Gal-10	PP13	PPL13
EM1	33	24	31	29	27	18	18	19
EM2	12	16	15	13	11	12	9	16
EM3	8	7	10	8	8	10	7	11
no. EM > 50%	3	4	4	5	7	7	6	4

Table 4. EM in Human Galectins Bi-CRD Type

CRD type	F3			F4			
protein	Gal-4C	Gal-8C	Gal-9C	Gal-4N	Gal-8N	Gal-9N	Gal-12N
EM1	25	17	24	26	25	19	37
EM2	11	11	17	10	11	14	13
EM3	9	10	13	7	8	9	9
no. EM > 50%	6	8	6	4	6	6	2

Common features of all CRDs are a main fluctuation in the S3S4 loop (residues ca. 40−50), except for Gal-1 where this main flexibility is shifted to the F4F5 loop and a flexibility of the F4F5 loop (ca. residue 100). Apart from these two cases, the flexibility pattern of the loops seems to distinguish the galectin CRDs. All bi-CRDs show more similar distributed flexibility among all loops below 2.0 Å, while F3 or F4 mono-CRDs have one to three high mobility loops (up to 2.5 Å); others segments are more rigid. Although the results of the numerical comparison of RMSF pairs (check Experimental Methods Section and Table S3 in Supporting Information) are all very high and similar ($SI_{average}$ = 93%), indicating similar patterns of fluctuations, some highlights were found for comparisons between F3- and F4-type galectins, in particular the mono-CRD type. Then, it follows a more heterogeneous trend, dominated by high values between mono- and bi-CRD's of both types (F3 and F4).

In order to analyze the CRD dynamics in more detail, we computed EMs for all cases. The EM is able to detect the main concerted protein/domain motions which are usually related to function as shown in several works.[43,61,62] Tables 3 and 4 show the first three EM contributions to the overall protein dynamics.

As shown in Tables 3 and 4, it is difficult to provide a general trend. The most interesting result is the fact that for all F3-mono CRDs, less EMs are needed to describe the protein dynamics, since more than half of the structural variation is explained with less than 5 EM, while 6−8 EMs are needed in F4 or bi-CRDs. The distribution patterns are similar with the EM1 accounting for ca. 20−40% of the structural variation, the EM2 about 15%, and the EM3 5−10%. Overall analysis of the EM shows that they are mainly located in the loops, as expected from the previous RMSF data and the projection patterns (data not shown). The first EM usually shows movement of one or two loops, while lower amplitude modes involve several loops. Interestingly which and how the loop moves is a particular characteristic of each CRD, and only minor similarities are seen, pointing to completely different ordered dynamical behavior of each CRD. For example, in Gal-1 the first and second EMs are located in the F4F5 loop. The other F3 mono CRDs, Gal-2, -3, and hGRPC all present main movement in the S3S4 loop, accompanied by S5S6 movement and by an open/close movement of the groove (Figure S2, Supporting Information). For the F4-mono CRDs Gal-7, -10, and PPL13 main movements in S3S4 loop were observed. In Gal-7 it was accompanied by S1F2 loop and in PPL13 by F2S3 and F3S6 loops. PP13 has the main modes distributed on all loops. For the bi-CRDs, also no clear pattern is visible. Gal-4C term has EM located mainly in the S1F2 and S3S4 loops, while Gal-4N term presents such EM but located in the F4F5, S4S5, and F5S2 loops. The S3S4 loop also contributes significantly to EM of Gal-9N and -12N and especially in Gal-8N, where the S3S4 loop it is notably larger. The other EM contributions are, however, different for each system. Finally, Gal-8C and -9C are very rigid, and the main motions are located in S5S6 and F5S2 loops, respectively. In summary, each galectin CRD seems to move differently, and the only common movement for most of them seems to be loop S3S4.

Detailed Comparison of All CRDs LBG . To dissect the features of each individual CRDs LBG, we first compared the LBG in the free proteins and then analyzed the effect of structure and dynamics on ligand binding for those cases where a CRD: carbohydrate complex is available.

Global Description of the Ligand Binding Site. As mentioned above, the concave side of the CRD domain is determined by strands S1−S6, forming the LBG, the pocket where carbohydrate binds. In most ligand-bound complexes characterized so far, the carbohydrate (usually a disaccharide) binds to one extreme of the LBG formed by strands S4−S6, where key amino acids are located. However, some CRDs are known to bind oligosaccharides, and therefore the whole LBG may be involved in ligand recognition.[15] To compare in detail each CRD LBG, we have divided the LGB into eight different MBS, denominated Y, Z, A, B, C, D, E, and F, as determined by the in silico combination of two Gal-9N hexasaccharide complexes described in the Methods Section. A structural topology ordering of the data was chosen in order to highlight the relevance of the position of a particular residue in some of the β-strands in LBG. The results listing all residues defining each MBS are shown in Figure 8 (also check Table S5 in Supporting Information).

The results shown in Figure 8A clearly point up that there exist highly conserved topological positions, while others seem to allow the presence of any type of residue. Interestingly, even when only two or three different residues per topological position were found or in cases where a conserved residue is changed, no correlation was observed with defined phylogenetic or domain organization groups. Remarkably, MBS Y, Z, and A are less conserved than MBS B, C, and D (Figure 8B and C). Furthermore, MBS E is almost out of the CRD and could not be well-defined. Most conserved positions are Gln at 5-S5 and hydrophobic residues Ile, Val or Leu at positions 1-S4 and 3-S5. Also highly conserved is Trp at position 2-S6 (except for hGRPC), Glu at position 5-S6 (except for Gal-10 with a Gln), and Pro at -1-S1, found in all CRDs except Gal-1 and 2. Similar situation is observed for His in position 4-S4, Gln in position 6-S4, and Arg in position 8-S4, except in hGRPC, PP13, and PPL13 which have different residues. hGRPC also differs in conserved Arg residue in position 7-S6, together with Gal-8N, Gal-10, and PP13. Gly at position 3-S6 is changed for larger residues in Gal-12N, Gal-10, PP13, and PPL13. Other positions, such as 1-S3, 5-S3, and 7-S5, show also same degree of conservation.

In particular, each CRD has none, one, or at most two changes in the conserved positions, except for hGRPC, Gal-10, PP13 and

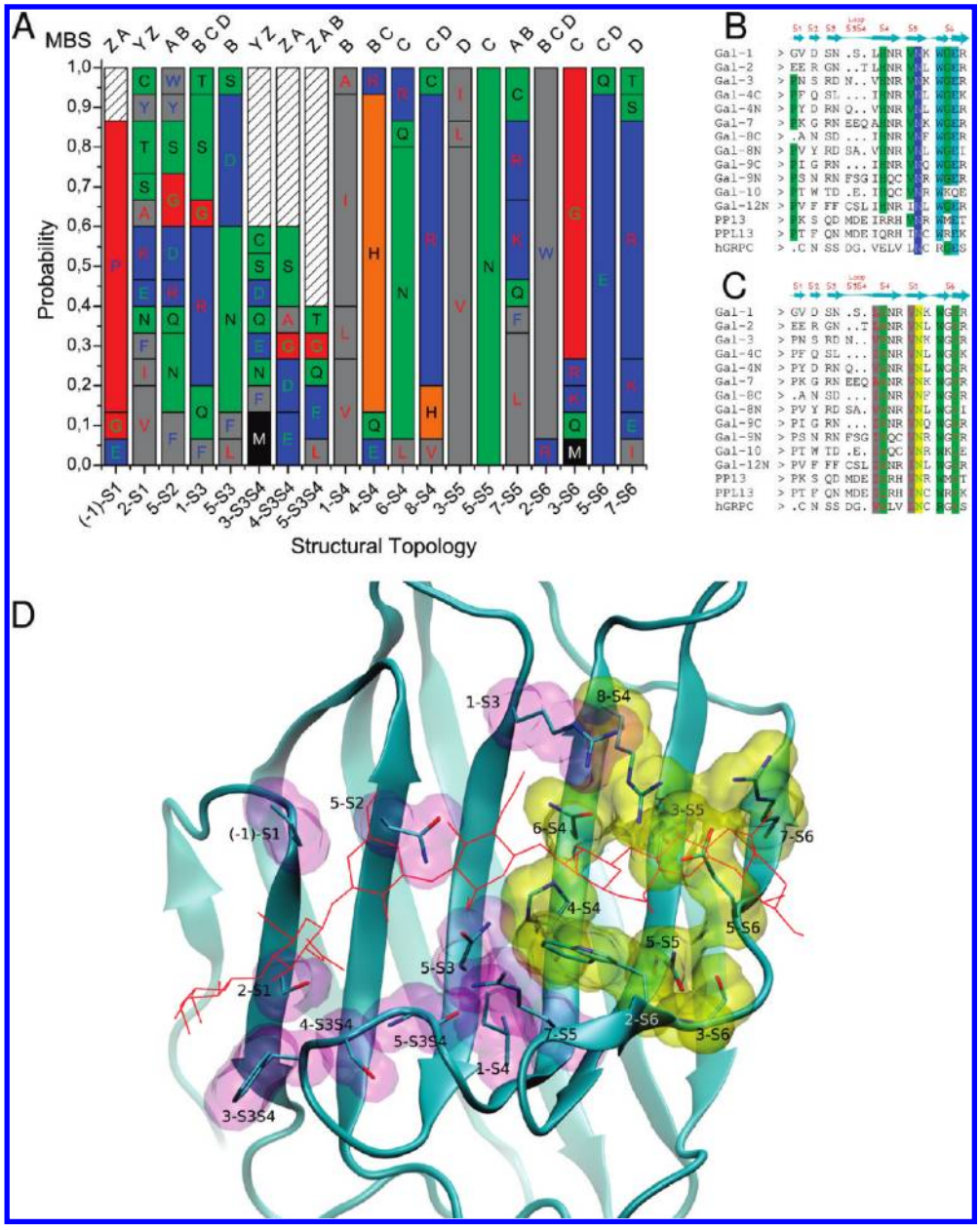


Figure 8. Amino acidic composition of human galectin LBGs based on a topology organization. (A) Bar graph summarizing the results of the topological analysis of all galectins LBGs. Each residue is colored by its physicochemical properties in one-letter code. The structural topology corresponds to the piece of secondary structure under study, preceded by a number which indicates the relative position of some particular residue in that secondary structure element (e.g., 3-S3S4 corresponds to the position no. 3 of the S3S4 loop, counting from the beginning of the S3S4 loop, at the end of S3 β-strand). The probability index was calculated by counting the number of a particular residue that appears in a particular topological position over fifteen, the total of CRDs studied (for more detail, check Supporting Information). (B) Results of the structural alignment based on in silico construction of tetra-LacNAc bound human galectin-9 N-terminal, as explained in the Experimental Methods Section. The topology of the LBG is marked on the top of the figure. The code of colors used is the same as in Figure 3. (C) Structural alignment where residues with similar physicochemical properties are highlighted. (D) 3D representation of the Gal-9N in silico construction with the most relevant aminoacids featured. Based on (A), the most conserved (yellow shadow) and the most variable (purple) amino acids through the global analysis are highlighted.

PPL13 which show more deviations from consensus. Since the analyzed residues are found in the LBG, these differences may explain why Gal-10 lacks affinity for β-galactosides but binds mannose in a unique manner.[63] A similar but unexpected behavior is found for PP13 and PPL13 whose LBG are also different. Finally, the lack of key LBG residues in hGRPC like Trp 2-S6 may explain why this protein does not show affinity for several studied sugars.[17,18] LBG topological residues present some highly conserved positions, with small and unique differences for each CRDs and no direct phylogenetic or domain organization relation. In addition, Gal-10, PP13, PPL13, and specially hGRPC are found to divert more from the LBG conserved residues.

Effect of Ligand Binding on CRD Structure: Structural and Thermodynamic Characterization. As mentioned in the Introduction Section, many CRD structures have been obtained in the

Table 5. Free Energy Calculations for the Occupied MBS in Selected Galectins Using the MM/GBSA Approach and Conformational Entropy Differences Between Apo and Ligand-Bound CRDs

			energy (kJ mol^{-1})[a]					
protein	MBS	ligand	ELE[b]	vdW[c]	GB_{Solv}[d]	GB_{Tot}[e]	OH[f]	ΔS (J K^{-1} mol^{-1})
Gal-1	C	β-D-galactopyranosyl-(1→	−126	−62	137	−51	2	−172.4
	D	4)-*N*-acetyl-D-glucosamine	−97	−35	87	−45	2	
	all	*N*-acetyllactosamine	−223	−97	195	−124	4	
Gal-2	C	β-D-galactopyranosyl-(1→	−122	−53	134	−40	2	−34.3
	D	4)-α-D-glucopyranose	−38	−18	46	−11	1	
	all	α-lactose	−160	−72	164	−68	3	
Gal-3	C	β-D-galactopyranosyl-(1→	−18	−61	33	−46	2	−401.2
	D	4)-*N*-acetyl-D-glucosamine	−195	−27	183	−39	2	
	all	*N*-acetyllactosamine	−213	−88	189	−112	4	
Gal-7	C	β-D-galactopyranosyl-(1→	−44	−58	56	−46	2	−0.4
	D	4)-*N*-acetyl-D-glucosamine	−158	−25	152	−31	2	
	all	*N*-acetyllactosamine	−204	−83	183	−103	4	
Gal-8N	C	β-D-galactopyranosyl-(1→	−26	−62	39	−48	1	−126.4
	D	4)-α-D-glucopyranose	−172	−20	181	−10	2	
	all	α-lactose	−197	−82	203	−76	3	
Gal-9N	Y	β-D-galactopyranosyl-(1→	−10	−21	32	2	0	−22.6
	Z	4)-*N*-acetyl-D-glucosaminyl-(1→	−33	−45	62	−15	1	
	A	3)-β-D-galactopyranosyl-(1→	5	−30	27	3	1	
	B	4)-*N*-acetyl-D-glucosaminyl-(1→	−18	−43	52	−8	1	
	C	3)-β-D-galactopyranosyl-(1→	−80	−64	85	−59	2	
	D	4)-*N*-acetyl-D-glucosaminyl-(1→	−116	−32	129	−19	2	
	all[h]	tetra-*N*-acetyllactosamine	−267	−241	351	−158	7	
Gal-10	C	mannose	−85	−58	96	−47	2	–[g]

[a] For all cases, the errors limits are less than 20 kJ mol^{-1}. [b] ELE: Electrostatic energy. [c] vdW: van der Waals contribution. [d] GB_{Solv}: Solvation free energy. [e] GB_{Tot}: Total free energy contribution. [f] OH: This column corresponds to the number of hydroxyl groups of the ligand which have the ability of hydrogen bonding with some of the protein residues of the LBG. [g] The conformational entropy for Gal-10/mannose complex could not be calculated, since the ligand does not remain in the protein LBG enough simulation time to achieve convergence of entropy. [h] Here we include the energy contributions of entire ligand. For the MBS E and F energy contributions please check Table S6 in Supporting Information.

presence of a carbohydrate ligand. The details of the protein ligand interactions have been exhaustively described in each case[10,13,64] and will not be analyzed in the present work. However, no global comparison of the effect of ligand binding to the CRD structure has been performed so far. To pursue this last issue, we have performed MD simulations of CRD-ligand complexes: Gal-1:lactose (PDB ID: 1GZW), Gal-2:lactose (PDB ID: 1HLC), Gal-3:*N*-acetyllactosamine (PDB ID: 1AK3), Gal-7:*N*-acetyllactosamine (PDB ID: 5GAL), Gal-8N:lactose (PDB ID: 2YXS) and Gal-9N:tetra-*N*-acetyllactosamine (our construction using PDB IDs 2ZHN and 2ZHM; please see Experimental Methods Section).

We analyzed the effect of ligand binding on CRD dynamics by comparing the RMSF of the bound structures with those obtained previously. A general appreciation indicates that ligand binding reduces the mobility of specific loop regions but interestingly not always those close to the LBG. For some cases ligand binding also increases, although at a lesser extent, the mobility of certain residues in other loops.

To illustrate this concept, results shown for Gal-1 CRD (Figure S4A in Supporting Information) demonstrate that no significant changes occurs in the F4F5, S6F3, and S3S4 loops, the last two loops being closer to the LBG than the first one. F2S3 and S6F3 loops, in close contact each other, reduce their mobility. Also, the small S5S6 loop and part of its contiguous region in the S6 strand have a reduced mobility possibly because they are involved in the interaction with the ligand. Interestingly, contact regions on the F-sheet face of the CRD, such as the F3F4 loop and the beginning of the F5S2 loop as well as the beginning of the S2F1 loop, significantly increase their mobility due to ligand binding. These results are consistent with those found by a recent work where the lactose binding to Gal-1 was studied.[65]

For Gal-2, ligand binding also reduces mobility but now mainly in the F5S2 loop (Figure S4B, Supporting Information). For Gal-3, binding of Lac or LacNAc also significantly reduces the mobility of specific regions (Figure S4C, Supporting Information), mainly the S6 strand and F4F5 loop. The main contact responsible for reduced mobility of S6 is the interaction that the strand has with Glu184 (index from 1AK3.pdb), while the effect on F4F5 loop is subtle and cannot be ascribed to a specific interaction. For Gal-8N, also significant reduction in mobility is observed in the small F5F6 and big F6S2 loops. The change in the small loop is due to strengthening of Asn126—Glu111 interaction (index from 2YV8.pdb), which is not properly established in the free protein. The reduced mobility of the F6S2 loop is mainly due to a change in loop structure which adopts a small helical conformation in the ligand bound protein. Reduction in mobility is also observed in the S5S6 loop. For Gal-9N, the S6 strand and the S6F3 loop show a reduced mobility upon ligand binding. The most important effects are evident in

areas that surround the Pro84 and Arg125 (index from chain A of 2HZM.pdb). Both effects could be caused by ligand binding, which induces new conformations on Asn75 and Pro84 residues allowing multiple interactions with the Arg125 NH groups. Arg125 side chain freezes in a conformation closer to the S-sheet than the one in apo Gal-9N, where the Arg125 is much closer to the F-sheet, interacting strongly with Glu86 and Phe123:O but with a conformation where the backbone is much more free to move. Interestingly, ligand binding to galectin-7 does not significantly reduce the mobility of any specific region. In summary despite the general trend which shows that CRD mobility is reduced upon ligand binding, the particular region where the reduction occurs is characteristic of each CRDs.

We computed the corresponding conformational entropy for the free and ligand-bound CRDs, as described in Experimental Methods. The conformational entropy resumes in one value the flexibility (or amplitude of the protein conformational space), where high values mean more flexible and/or wide conformational accessible space. The analysis performed with the 50 ns long MD simulations showed a good convergence of the entropy in most of the structures studied (Figure S3 in Supporting Information). From these values we calculated the change in the CRD conformational entropy due to ligand binding (ΔS). The computed values are reported in Table 5. Consistent with the above observed trend, results are all negative (i.e., upon ligand binding, the CRD reduces its conformational entropy). The values range from almost zero for Gal-7 to almost 420 J K^{-1} mol^{-1}, which would result in about 10 kJ mol^{-1} contribution of entropy loss to complex formation at 300 K. The range of the observed values suggests that conformational entropy is an important contribution to binding free energy and therefore affinity.

Finally, in order to provide a comparative viewpoint of carbohydrate binding thermodynamics for different LBGs, we performed a MM/GBSA analysis for all CRD ligand complexes studied in the present work (Table 5 and Table S6 in Supporting Information). The results show that most disaccharides have overall predicted negative binding free energies between 80 and 125 kJ mol^{-1}, except for mannose-binding Gal-10 and Gal-2 which present weaker binding to the D site. Usually the electrostatic contribution is more than twice the vdW (as expected for this type of ligands), but it is significantly compensated by the solvation penalty. Interestingly, each monomer contribution is significantly different for each CRD, pointing toward high variability and therefore subtle regulation. Looking closer to each MBS, analysis shows substantial variation with values ranging from 0 to −120 kJ mol^{-1}. However, interesting trends can be stated. The vdW contribution of MBS site C is always between −53 and −64 kJ mol^{-1}, possibly due to stacking interaction with conserved Trp. On the other hand, electrostatic interaction energies are more variable. For site D, vdW interaction is significantly smaller (about 26 kJ mol^{-1}); again, the electrostatic contribution shows high variability between different CRDs. Finally, a look at Gal-9N data shows that the MBS that most contribute to binding is MBS C and to a lesser extent MBS D and MBS Z, while other MBSs present negligible or even nonfavorable interactions.

Altogether, the binding thermodynamic results show that apart from conserved vdW interaction of MBS C (due to presence of conserved Trp) each CRD shows significantly different contributions and total binding energies for each MBS and therefore a possibly different affinity. Experimental ligand binding free energy results have been reported for some related systems: bovine spleen Gal-1:LacNAc, −23 kJ mol^{-1} (at 298 K); bovine heart Gal-1:LacNAc, −23.9 kJ mol^{-1} (at 300 K); murine recombinant Gal-3:LacNAc, −24.7 kJ mol^{-1} (at 300 K); human recombinant Gal-7:LacNAc, −18.4 kJ mol^{-1} (at 300 K); and gluthation-*S*-transferase-fused human recombinant Gal-8N: lactose, −23.4 kJ mol^{-1} (at 298 K).[66−68] The overall results are significantly smaller than the computed ones, consistently with the known flaws of this methodology, which tends to overestimate ligand binding free energy.[69] For this reason, the computational results can be taken only in a comparative way. Concerning the hydrogen-bond analysis, it is clear that in each MBS no more than two hydroxyl groups are involved per MBS. This is due to the fact that each monosaccharide binds to the protein exposing some hydroxyl groups, remaining others with one side facing to the solvent. Finally, it is notable that while the binding energy values are all quite similar, the entropy change experienced by each CRD is actually highly variable.

DISCUSSION

In the present study we provide the first comparative analysis of the structure, dynamics, and ligand-binding properties of different members of the human galectin family. The results presented herein can provide rational explanations for different saccharide specificities related to functions of human galectins. The key functional feature that defines each human galectin CRD is linked to its specificity for binding carbohydrates, and its ability to dimerize, cross-link two different binding sites due to tandem architecture, or interact with other proteins (as for Gal-3). Our central hypothesis is that the saccharide-binding specificity and the corresponding affinity are the result of a different architecture and dynamics of each CRD LBG, which in turn might influence biological properties. To elaborate on this idea, we have performed an exhaustive comparative analysis of all human galectin CRD structure and dynamics using state-of-the-art computer simulation techniques. As described in the IntroductionSection, the results are presented in the context of the domain organization classification of human galectins[29] in proto-, tandem- and chimera-type and also in the phylogenetic-based classification[30] of CRDs into two groups: F3 and F4 CRD types, which results in three groups F3-mono, F4-monom and bi-CRD.

The first interesting, but not surprising, result emerging from our analysis is that despite having average percentage identity of only 30%, the structure of all CRDs is very similar with a highly conserved fold, with the only structural differences located in specific loop regions. However, a more subtle result is the high degree of similarity with overall RMSD of less than 2.2 Å, which is similar to values expected for the same protein crystallized under different environments. These differences, although allowing clustering the different CRDs in distinct structural groups, show almost no correlation with sequence or domain organization classification. Therefore, no relation between the same group (either domain or phylogenetic) and function is expected. Also, the differences in the length and characteristics of amino acids in these loops may be crucial for determining the specificity for one or another ligand. In fact, the architecture of a particular carbohydrate can be very complex, and subtle differences in these protein regions, where we observed large changes among individual members of galectin family, maybe sufficient, for example, to explain why some galectins can bind sialylated oligosaccharides, whereas binding of others is restricted by enhanced sialylation.

The second conclusion of the present work is drawn from the dynamical analysis. Again as for the structure, comparative analysis of CRD dynamics shows that global patterns are similar for all CRDs, with the whole mobility mainly located in the loops. Interestingly, which loop and how it moves results a particular characteristic of each CRD, and only minor similarities are seen, pointing to completely different ordered dynamical behavior of each CRD. A possible relevant result of the dynamical analysis relates to the higher and more homogenously distributed motion observed for bi-CRD types compared to mono-CRD type galectins. It is difficult at this point to determine whether this distinct pattern is a basis or a consequence of the domain organization type of galectins. However, given that domain organization bears no association with phylogenetic history, it is reasonable to speculate that particular differential dynamic pattern of mono- and bi-CRD galectins is a consequence of the domain organization. Mono-CRD galectins usually display dimer−monomer equilibrium. Therefore, a higher mobility may interfere with the dimerization process. Another possible functional role of this type of dynamics concerns modulation of ligand affinity by means of CRD conformational entropy change. As shown by our results, the entropy loss contribution to ligand binding free energy may account for up to 120 kJ mol^{-1} (at 300 K), therefore playing a key role in determining CRD affinity. In this context, the less restrained dynamical pattern of bi-CRDs may allows stronger regulation of affinity by these means, while for mono-CRD, the more localized pattern may allow connecting the ligand binding and the dimerization process as observed for Gal-1.[70,71] Finally, a high or low mobility and/or conformational entropy value may explain why some galectins bind linear ligands (such as poly-*N*-acetyllactosamines), while others have high affinity for branched oligosaccharides, given that structural rigidity is less important for the former than for the latter.[3,24,60]

Taken together, the results show that from a global viewpoint all human CRDs are quite similar in both structure and dynamics, with some particular singularities to pay attention to. Consequently, the key remaining question is focused on what are the relevant differences that determine the need for so many human CRDs.

Finally, we provide a detailed comparative study of each LBG relevant residue. The comparative estimated monosaccharide binding energy contributions, which cannot be obtained by experimental methods, show that human galectin CRDs key differences arise from specific details in each MBS. We show that many residues lining the LGB in MBS Z, B, C, and D are highly conserved, such as Trp 2-S6, which determines the high and similar vdW contribution to monosaccharide binding in MBS C. However, for each CRD there are unique substitutions that lack any phylogenetic or domain organization relationship which results in clear differences in the MBS binding energy contributions that are expected to determine different affinities for different carbohydrate ligands, yielding a possibly unique selectivity and therefore biological role. Partial support for this interpretation comes from particular analysis of Gal-10, and specially hGRPC, which is found to have more conserved residues substitutions in the LBG that may explain why Gal-10 binds mannose instead of lactose and why hGRPC does not bind any tested saccharides at all.

CONCLUSIONS

This structural, dynamical, and thermodynamical approach is the first integrated computational analysis of the important family of human galectins. In addition, this study also provides valuable information about those family members still not structurally characterized. We should finally note that further studies are needed in order to address other questions not explored in the present work, such as the role of quaternary structure in galectin function, the oligomerization state relevant for galectin activity, ligand binding specificity, etc. The understanding of the particular and also the global structural behavior of galectins in connection with some of the biological roles identified day by day represents nowadays an exciting area in biomedical research, including cancer, autoimmunity, inflammation, and neurodegeneration.

ASSOCIATED CONTENT

Supporting Information. Figures S1−S4 and Tables S1−S6 and the homology models for Gal-4N, Gal-12N, PP13, and PPL13 as PDB format files (PDB S1−S4, respectively). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: marcelo@qi.fcen.uba.ar, dario@qi.fcen.uba.ar.

ACKNOWLEDGMENT

Authors thank Agencia Nacional de Promoción Científica y Tecnológica (PICT Raices 157 and PICT 2006-603), CONICET, Universidad de Buenos Aires and Fundación Sales for providing financial support for carrying out this work. Calculations were performed in the Centro de Cómputos de Alto Rendimiento (CeCAR) at the Facultad de Ciencias Exactas y Naturales (FCEN) of the Universidad de Buenos Aires and the Hal PME cluster located at University of Córdoba, Argentina.

REFERENCES

(1) Cooper, D. N. W. Galectinomics: finding themes in complexity. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572*, 209–231.

(2) Yang, R. Y.; Rabinovich, G. A.; Liu, F. T. Galectins: structure, function and therapeutic potential. *Expert Rev. Mol. Med.* **2008**, *10*, e17.

(3) Rabinovich, G. A.; Toscano, M. A. Turning 'sweet' on immunity: galectin-glycan interactions in immune tolerance and inflammation. *Nat. Rev. Immunol.* **2009**, *9*, 338–352.

(4) Toscano, M.; Bianco, G. A.; Ilarregui, J. M.; Croci, D. O.; Correale, J.; Hernandez, J. D.; Zwirner, N. W.; Poirier, F.; Riley, E. M.; Baum, L. G.; Rabinovich, G. A. Differential glycosylation of TH1, TH2 and TH-17 effector cells selectively regulates susceptibility to cell death. *Nat. Immunol.* **2007**, *8*, 825–834.

(5) Rabinovich, G. A.; Ilarregui, J. M. Conveying glycan information into T-cell homeostatic programs: a challenging role for galectin-1 in inflammatory and tumor microenvironments. *Immunol. Rev.* **2009**, *230*, 144–159.

(6) Laderach, D. J.; Compagno, D.; Toscano, M. A.; Croci, D. O.; Dergan-Dylon, S.; Salatino, M.; Rabinovich, G. A. Dissecting the signal transduction pathways triggered by galectin−glycan interactions in physiological and pathological settings. *IUBMB Life* **2010**, *62*, 1–13.

(7) Cooper, D.; Ilarregui, J. M.; Pesoa, S. A.; Croci, D. O.; Perretti, M.; Rabinovich, G. A. Multiple Functional Targets of the Immunoregulatory Activity of Galectin-1: Control of Immune Cell Trafficking, Dendritic Cell Physiology, and T-Cell Fate. *Methods Enzymol.* **2010**, *480*, 199–244.

(8) Ilarregui, J. M.; Croci, D. O.; Bianco, G. A.; Toscano, M.; Salatino, M.; Vermeuen, M. E.; Geffner, J. R.; Rabinovich, G. A. Tolerogenic signals delivered by dendritic cells to T cells through a galectin-1 driven immunoregulatory circuit involving interleukin 27 and interleukin 10. *Nat. Immunol.* **2009**, *10*, 981–991.

(9) Dam, T. K.; Brewer, C. F. Lectins as pattern recognition molecules: The effects of epitope density in innate immunity*. *Glycobiology* **2010**, *20*, 270–279.

(10) Lopez-Lucendo, M. F.; Solis, D.; Andre, S.; Hirabayashi, J.; Kasai, K.; Kaltner, H.; Gabius, H. J.; Romero, A. Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. *J. Mol. Biol.* **2004**, *343*, 957–970.

(11) Cho, M.; Cummings, R. D. Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. I. Physical and chemical characterization. *J. Biol. Chem.* **1995**, *270*, 5198–5206.

(12) Liao, D. I.; Kapadia, G.; Ahmed, H.; Vasta, G. R.; Herzberg, O. Structure of S-lectin, a developmentally regulated vertebrate beta-galactoside-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1428–1432.

(13) Seetharaman, J.; Kanigsberg, A.; Slaaby, R.; Leffler, H.; Barondes, S. H.; Rini, J. M. X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1 Å resolution. *J. Biol. Chem.* **1998**, *273*, 13047–13052.

(14) Nagae, M.; Nishi, N.; Nakamura-Tsuruta, S.; Hirabayashi, J.; Wakatsuki, S.; Kato, R. Structural analysis of the human galectin-9 N-terminal carbohydrate recognition domain reveals unexpected properties that differ from the mouse orthologue. *J. Mol. Biol.* **2008**, *375*, 119–135.

(15) Nagae, M.; Nishi, N.; Murata, T.; Usui, T.; Nakamura, T.; Wakatsuki, S.; Kato, R. Structural analysis of the recognition mechanism of poly-N-acetyllactosamine by the human galectin-9 N-terminal carbohydrate recognition domain. *Glycobiology* **2009**, *19*, 112–117.

(16) Yoshida, H.; Teraoka, M.; Nishi, N.; Nakakita, S.; Nakamura, T.; Hirashima, M.; Kamitori, S. X-ray structures of human galectin-9 C-terminal domain in complexes with a biantennary oligosaccharide and sialyllactose. *J. Biol. Chem.* **2010**, *285*, 36969–36976.

(17) Zhou, D.; Ge, H.; Sun, J.; Gao, Y.; Teng, M.; Niu, L. Crystal structure of the C-terminal conserved domain of human GRP, a galectin-related protein, reveals a function mode different from those of galectins. *Proteins* **2008**, *71*, 1582–1588.

(18) Wälti, M. A.; Thore, S.; Aebi, M.; Künzler, M. Crystal structure of the putative carbohydrate recognition domain of human galectin-related protein. *Proteins* **2008**, *72*, 804–808.

(19) Kato-Murayama, M.; Murayama, K.; Terada, T.; Shirouzu, M.; Yokoyama, S. Crystal structure of N-terminal domain of mouse galectin-4. *Riken Structural Genomics/Proteomics Initiative (RSGI)*, RIKEN, Japan, 2006. http://www.rsgi.riken.go.jp/rsgi_e/index.html.

(20) Tomizawa, T.; Kigawa, T.; Saito, K.; Koshiba, S.; Inoue, M.; Yokoyama, S. Solution structure of the C-terminal Gal-bind lectin domain from human galectin-4. *Riken Structural Genomics/Proteomics Initiative (RSGI)*, RIKEN, Japan, 2005. http://www.rsgi.riken.go.jp/rsgi_e/index.html.

(21) Kishishita, S.; Nishino, A.; Murayama, K.; Terada, T.; Shirouzu, M.; Yokoyama, S. Crystal structure of N-terminal domain of human galectin-8. *Riken Structural Genomics/Proteomics Initiative (RSGI)*, RIKEN, Japan, 2007. http://www.rsgi.riken.go.jp/rsgi_e/index.html.

(22) Kishishita, S.; Nishino, A.; Murayama, K.; Terada, T.; Shirouzu, M.; Yokoyama, S. Crystal structure of N-terminal domain of human galectin-8 with D-lactose. *Riken Structural Genomics/Proteomics Initiative (RSGI)*, RIKEN, Japan, 2007. http://www.rsgi.riken.go.jp/rsgi_e/index.html.

(23) Tomizawa, T.; Koshiba, S.; Inoue, M.; Kigawa, T.; Yokoyama, S. Solution structure of the C-terminal Gal-bind lectin protein from human galectin-8. *Riken Structural Genomics/Proteomics Initiative (RSGI)*, RIKEN, Japan, 2007. http://www.rsgi.riken.go.jp/rsgi_e/index.html.

(24) Lobsanov, Y. E. A. X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9 A resolution. *J. Biol. Chem.* **1998**, *268*, 27034–27038.

(25) Leonidas, D. D.; Vatzaki, E. H.; Vorum, H.; Celis, J. E.; Madsen, P.; Acharya, K. R. Structural basis for the recognition of carbohydrates by human galectin-7. *Biochemistry* **1998**, *37*, 13930–13940.

(26) Leonidas, D. D.; Elbert, B. L.; Zhou, Z.; Leffler, H.; Ackerman, S. J.; Acharya, K. R. Crystal structure of human Charcot Leyden crystal protein, an eosinophil lysophospholipase, identifies it as a new member of the carbohydrate-binding family of galectins. *Structure* **1995**, *3*, 1379–1393.

(27) Visegrády, B.; Than, N. G.; Kilár, F.; Sümegi, B.; Than, G. N.; Bohn, H. Homology modelling and molecular dynamics studies of human placental tissue protein 13 (galectin-13). *Protein Eng.* **2001**, *14*, 875–880.

(28) Than, N. G.; Pick, E.; Bellyei, S.; Szigeti, A.; Burger, O.; Berente, Z.; Janaky, T.; Boronkai, A.; Kliman, H.; Meiri, H.; Bohn, H.; Than, G. N.; Sumegi, B. Functional analyses of placental protein 13/galectin-13. *Eur. J. Biochem.* **2004**, *271*, 1065–1078.

(29) Hirabayashi, J.; Kasai, K. The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution. *Glycobiology* **1993**, *3*, 297–304.

(30) Houzelstein, D.; Gonçalves, I. R.; Fadden, A. J.; Sidhu, S. S.; Cooper, D. N. W.; Drickamer, K.; Leffler, H.; Poirier, F. Phylogenetic analysis of the vertebrate galectin family. *Mol. Biol. Evol.* **2004**, *21*, 1177–1187.

(31) Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. Discovery through the computational microscope. *Structure* **2009**, *17*, 1295–1306.

(32) Leach, A. R. Molecular Modelling: *Principles and Applications*; Pearson Education EMA: Harlow, England, 2001, p 744.

(33) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006**, *106*, 1589–1615.

(34) Eswar, N.; Webb, B.; Martí-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative protein structure modelling using MODELLER. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc: New York, 2006; Chapter 5, Unit 5.6, pp 5.6.1−5.6.30.

(35) Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. Comparative protein structure modelling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.

(36) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids. Res.* **1994**, *22*, 4673–4680.

(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(38) van Gunsteren, W. F.; Berendsen, H. J. C. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992–1023.

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

(40) Hornak, V.; Aberl, R.; Okur, A.; Strockbine, B.; Roitberg, A. E.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.

(41) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J. Phys. Chem.* **1995**, *99*, 3832–3846.

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

(43) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425.

(44) Capece, L.; Estrin, D. A.; Marti, M. A. Dynamical characterization of the heme NO oxygen binding (HNOX) domain. Insight into soluble guanylate cyclase allosteric transition. *Biochemistry* **2008**, *47*, 9416–9427.

(45) Martí, M. A.; Estrin, D. A.; Roitberg, A. E. Molecular basis for the pH dependent structural transition of nitrophorin 4. *J. Phys. Chem. B* **2009**, *113*, 2135–2142.

(46) Capece, L.; Marti, M. A.; Bidon-Chanal, A.; Nadra, A.; Luque, F. J.; Estrin, D. A. High pressure reveals structural determinants for globin hexacoordination: Neuroglobin and myoglobin cases. *Proteins* **2009**, *75*, 885–894.

(47) Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993**, *215*, 617–621.

(48) Andricioaei, I.; Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289–6292.

(49) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. Cooperativity in drug-DNA recognition: a molecular dynamics study. *J. Am. Chem. Soc.* **2001**, *123*, 12658–12663.

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

(51) Bashford, D.; Case, D. A. Generalized Born models of macromolecular salvation effects. *Annu. Rev. Phys. Chem.* **2003**, *51*, 129–152.

(52) Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the Generalized-Born model. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.

(53) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A. E.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.

(54) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(55) Constanciel, R.; Contreras, R. Self consistent field theory of solvent effects representation by continuum models: Introduction of desolvation contribution. *Theor. Chim. Acta* **1984**, *65*, 1–11.

(56) Still, W. C.; Tempczyrk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(57) Scheeff, E. D.; Bourne, P. E. Structural evolution of the protein kinase-like superfamily. *PLoS Comput. Biol.* **2005**, *1*, e49.

(58) Moens, L.; Vanfleteren, J.; Van de Peer, Y.; Peeters, K.; Kapp, O.; Czeluzniak, J.; Goodman, M.; Blaxter, M.; Vinogradov, S. Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships. *Mol. Biol. Evol.* **1996**, *13*, 324–333.

(59) Hirabayashi, J.; Kasai, K. Effect of amino acid substitution by sited-directed mutagenesis on the carbohydrate recognition and stability of human 14-kDa beta-galactoside-binding lectin. *J. Biol. Chem.* **1991**, *266*, 23648–23653.

(60) Leffler, H.; Barondes, S. H. Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian beta-galactosides. *J. Biol. Chem.* **1986**, *261*, 10119–10126.

(61) Martin-Galiano, A. J.; Buey, R. M.; Cabezas, M.; Andreu, J. M. Mapping flexibility and the assembly switch of cell division protein FtsZ by computational and mutational approaches. *J. Biol. Chem.* **2010**, *285*, 22554–22565.

(62) Van Aalten, D. M. F.; De Groot, B. L.; Findlay, J. B. C.; Berendsen, H. J. C.; Amadei, A. A comparison of techniques for calculating protein essential dynamics. *J. Comput. Chem.* **1997**, *18*, 169–181.

(63) Swaminathan, G. J.; Leonidas, D. D.; Savage, M. P.; Ackerman, S. J.; Acharya, K. R. Selective recognition of mannose by the human eosinophil Charcot-Leyden crystal protein (Galectin-10): a crystallographic study at 1.8 Å resolution. *Biochemistry* **1999**, *38*, 13837–13843.

(64) Ford, M. G.; Weimar, T.; Köhli, T.; Woods, R. J. Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity. *Proteins* **2003**, *53*, 229–240.

(65) Nesmelova, I. V.; Ermakova, E.; Daragan, V. A.; Pang, M.; Menéndez, M.; Lagartera, L.; Solís, D.; Baum, L. G.; Mayo, K. H. Lactose binding to galectin-1 modulates structural dynamics, increases conformational entropy, and occurs with apparent negative cooperativity. *J. Mol. Biol.* **2010**, *397*, 1209–1230.

(66) Schwarz, F. P.; Ahmed, H.; Bianchet, M. A.; Amzel, L. M.; Vasta, G. R. Thermodynamics of bovine spleen galectin-1 binding to disaccharides: correlation with structure and its effect on oligomerization at the denaturation temperature. *Biochemistry* **1998**, *37*, 5867–5877.

(67) Brewer, C. F. Thermodynamic binding studies of galectin-1, −3 and −7. *Glycoconjugate J.* **2004**, *19*, 459–465.

(68) Ideo, H.; Seko, A.; Ishikuza, I.; Yamashita, K. The *N*-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity. *Glycobiology* **2003**, *13*, 713–723.

(69) Anisimov, V. M.; Cavasotto, C. N. Quantum mechanical binding free energy calculation for phosphopeptide inhibitors of the Lck SH2 domain. *J. Comput. Chem.* **2011**, *32*, 2254–2263.

(70) Stowell, S. R.; Cho, M.; Feasley, C. L.; Arthur, C. M.; Song, X.; Colucci, J. K.; Karmakar, S.; Mehta, P.; Dias-Baruffi, M.; McEver, R. P.; Cummings, R. D. Ligand reduces galectin-1 sensitivity to oxidative inactivation by enhancing dimer formation. *J. Biol. Chem.* **2009**, *284*, 4989–4999.

(71) Di Lella, S.; Martí, M. A.; Croci, D. O.; Guardia, C. M. A.; Díaz-Ricci, J. C.; Rabinovich, G. A.; Caramelo, J. J.; Estrin, D. A. Linking the structure and thermal stability of beta-galactoside-binding protein galectin-1 to ligand binding and dimerization equilibria. *Biochemistry* **2010**, *49*, 7652–7658.