Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 339 (2004) 985–994
Carbohydrate RESEARCH

Toward the understanding of the structure and dynamics of protein–carbohydrate interactions: molecular dynamics studies of the complexes between hevein and oligosaccharidic ligands

Giorgio Colombo,[a,*] Massimiliano Meli,[a] Javier Cañada,[b] Juan Luis Asensio[b] and Jesús Jiménez-Barbero[b,*]

[a]*Istituto di Chimica del Riconoscimento Molecolare, CNR via Mario Bianco, 9, I-20131 Milan, Italy*
[b]*Centre de Investigaciones Biológicas, CSIC, Ramiro de Maetzu 9, E-28040 Madrid, Spain*

Received 3 June 2003; revised 10 September 2003; accepted 15 October 2003

Abstract—Herein we study, through all atom molecular dynamics simulations, the complex between hevein and two N-acetylated chitin oligomers, namely *N*,*N*′-diacetylchitobiose and *N*,*N*′,*N*″-triacetylchitotriose. The results of the simulations for two disaccharide complexes and one trisaccharide complex show that a carbohydrate oligomer is able to move on the surface of the relatively flat binding pocket of hevein, therefore occupying different binding subpockets. Statistical analysis methods were also applied in order to define the principal overall motions in the complexes, showing how the different ligands in the simulations modulate the protein motions. The oligosaccharide binding can be considered as defined by a subtle balance between enthalpic (formation of intermolecular interactions between the ligand and the receptor) and entropic (due mainly to the possibility for the sugar to move on the surface of the protein domain) effects, determining multiple binding conformations. This structural and dynamical view could parallel the results obtained by regularly used restrained MD simulations based on NOE NMR data that provide a well defined structure for both the disaccharide and trisaccharide complexes, and agrees with the observations for longer oligosaccharide chains.

Keywords: Hevein; Lectins; Protein carbohydrate interactions; Molecular dynamics

1. Introduction

Protein–carbohydrate recognition processes play a fundamental role in a range of important biological phenomena such as cell–cell adhesion, cell growth, inflammatory processes, fertilization, and bacterial infection. Therefore, the study of this kind of interactions is currently a topic of major interest because of their potential in pharmaceutical and clinical applications.[1–3] A complete understanding of the molecular mechanisms determining the specific formation of complexes requires the application of experimental and theoretical techniques providing detailed structural information at the atomic level.

From the experimental point of view, X-ray crystallography has provided a wealth of information on the three dimensional structures of protein–carbohydrate complexes.[4,5] In many cases, oligosaccharides, either in the free form in solution or when present as glycoconjugates or complexed to protein receptors such as antibodies, lectins, and enzymes have proved to be difficult to crystallize.[6] In a parallel way, NMR experiments have been frequently used both in the area of oligosaccharide conformational analysis and for studying oligosaccharide–protein interactions. However, and in general, NMR analysis of lectin–saccharide complexes has been hampered by the rather high molecular weight of

* Corresponding authors. Tel.: +39-02-28500031; fax: +39-02-28500-036 (G.C.); tel.: +34-915346623x4370/4371; fax: +34-915531706 (J.J.-B.); e-mail addresses: g.colombo@icrm.cnr.it; jjbarbero@cib.csic.es

0008-6215/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2003.10.030

sugar-binding proteins (lectins). In any case, NMR methods have provided insights into the structural characterization of small lectins and on the thermodynamic driving forces controlling their encounters with carbohydrates.[7,8]

In general, however, experimental studies yield a representation consisting of one single static picture of the system. In order to understand the dynamics of recognition at atomic level, we have little choice but to turn to theoretical methods.

Force field based theoretical methods have been used for instance by Woods and co-workers who,[9] after experimentally determining the X-ray crystal structure of the complex between Gal-α(1-3)Gal epitope (xenograft antigen) and isolectin l-B-4 from *Griffonia simplicifolia* (GS-1-B-4), reported the results of molecular dynamics (MD) simulations on the complex and on the free ligand. The simulations were performed for up to 10 ns in explicit solvent, and the comparison between the results proved that GS-1-B-4 recognizes the low energy conformation of the disaccharide. Moreover, the ensemble of X-ray and MD data provided clear explanations for the specificity of isolectin GS-1-B-4. Lescar and co-workers have used homology and docking approaches to explain the strong preference for α-anomeric sugars in isolectins I-A and I-B from *G. simplicifolia*.[10] Theoretical studies have also been applied to the design of improved mimics of natural protein binders. Colombo and co-workers[11] have successfully applied molecular mechanics and MD methods, to study the interactions between ganglioside GM1 mimics and the heat labile toxin (LT) of *E. coli*. MD simulations of the complexes for the natural ligand and its mimic showed that the chemical nature of the new ligand could trigger a series of rearrangements in the protein, leading to a new H-bonding network, thus resulting in a higher affinity for the rationally designed inhibitor. Indeed, free energy perturbation (FEP) calculations, as well as experimental testing, were applied to verify the observations.

On this basis, this paper is focused on the investigation, at atomic detail, of the features of binding of probe carbohydrates to small sized lectins, by combining extensive MD simulations with previously obtained NMR data. In particular, attention is paid to the complexes between hevein and two chitin oligomers, namely, *N*,*N*′-diacetylchitobiose and *N*,*N*′,*N*″-triacetylchitotriose (Fig. 1).

Hevein is a chitin binding protein, which is present in laticifers of the rubber tree. From the biological viewpoint, this lectin is the major allergen of latex and, in addition, it has been proposed to be involved in the mechanisms of plant defense toward insects and fungi attacks. From a structural point of view, hevein is a small single chain protein of 43 amino acids, especially rich in glycine and cysteine residues. Its 3D structure in solution has been extensively studied by X-ray crystallography and by NMR methods.[12,13] Moreover, its binding to chitin oligosaccharides has been studied through NMR spectroscopy.[7] Since the exoskeleton of insects and fungi is formed by chitin, the first step in the protection provided by hevein could be the molecular recognition of chitin by this small protein domain. Experimental investigations have proposed the existence of an exposed binding site for hevein, comprised by three subsites for *N*-acetylglucosamine (GlcNAc)-containing oligosaccharides: a subsite dubbed +1, defined by the aromatic rings of W23 and Y30, subsite +2 defined by residue W21, and a third subsite, not clearly defined. Based on these observations, some degree of mobility of the disaccharide bound onto the protein surface is considered to be present. A sliding motion of the disaccharide on the binding region of the protein has been proposed as a model to explain the experimental observations on the complexes.[7] In this paper, MD simulations of the complexes between hevein and model

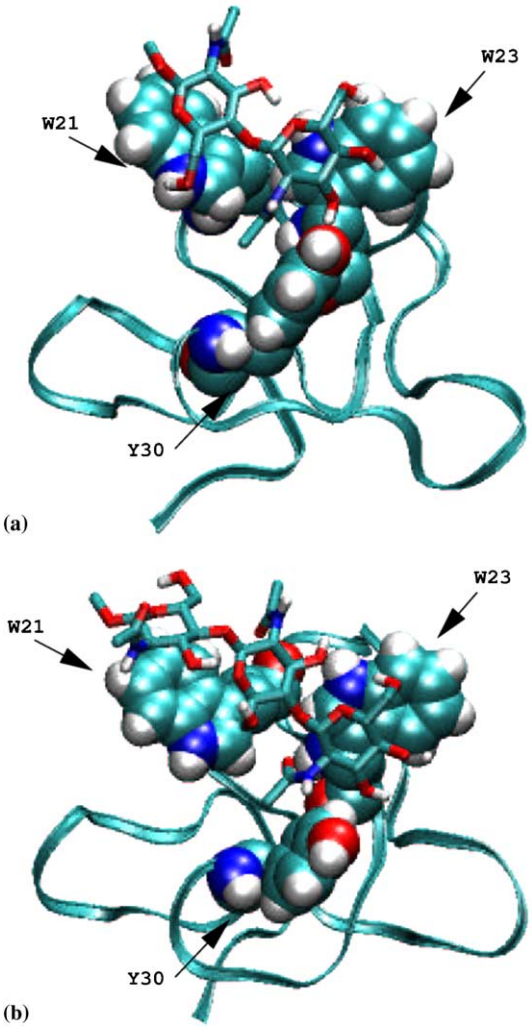

Figure 1. 3D NMR structures of the complexes between hevein and *N*,*N*′-diacetylchitobiose and hevein and *N*,*N*′,*N*″-triacetylchitriose.

chitin oligosaccharides have been used to assess the main determinants of binding and to investigate the possible mobility of the sugars in the time ranges examined.

Two simulations of 40 ns (labeled disac1 and disac2), were run for the complex between hevein and *N*,*N*′-diacetylchitobiose in explicit water solvent, starting from two different models derived from the ensemble of determined from one NMR experiment. The two starting conformations differ in the placing of the disaccharide on the protein surface: the reducing-end sugar moiety (here capped with an OMe group) is stacked in a face to face orientation with the aromatic ring of W21, while, simultaneously, $GlcNAc^{II}$ displays the same kind of carbohydrate–aromatic interaction with W23. This observation is consistent with the hypothesis of the existence of one binding site involving several subsites. Thus, in this case, the sugar binds to subsites +1 (defined by W21) and +2 (defined by W23). In simulation disac2, the starting structure characterized by a longer distance between GlcNAc2 and W23, and the sugar is twisted with respect to the orientation it has in the starting structure of disac1. One 40 ns simulation was run for the hevein–trisaccharide complex (labeled trisac). The structures obtained are compared to the average structures obtained via NMR spectroscopy, and the trajectories are analyzed using statistical approaches such as cluster analysis and the essential dynamics (ED) method:[14] their combination allows to extract the main determinants of the motions in the complexes, shedding further light on the factors influencing the recognition between carbohydrates and proteins in biological processes.

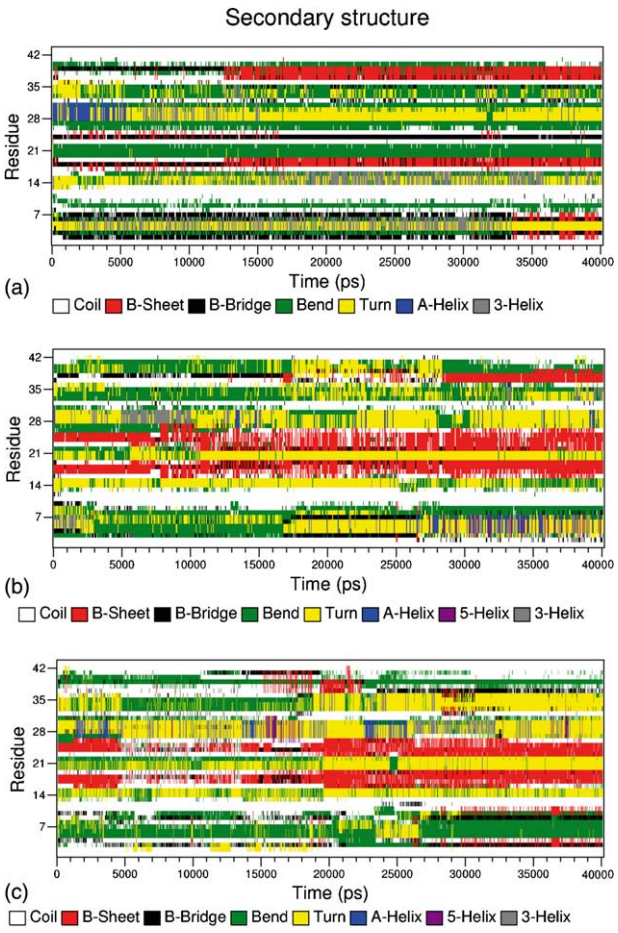

Figure 2. Time dependent secondary structure evolution of hevein in simulations disac1 (a), disac2 (b), and trisac (c).

2. Results

2.1. The protein

The influence of the carbohydrate motions on the three dimensional structure of the protein has been investigated in terms of several structural parameters for all the simulations. The root mean square deviations (RMSD, value calculated with respect to the NMR derived starting structure) show that the global three dimensional arrangement of the protein remains basically unchanged in the simulations, the average root mean square deviation value being around 0.2 nm in all three simulations.

The secondary structure calculations based on the DSSP algorithm,[15] (Fig. 2a–c) present the time variation of the main features of the secondary structure in 40 ns time. It is worth noting a conformational change after 6 ns in disac1 (Fig. 2a) involving the loss of the α-helical geometry spanning residues 28–31 (thus involving Y30) in favor of a more β-turn like geometry. This conformational change is paralleled by an increase in the RMSD value with respect to the NMR starting geometry for the protein. An ordered β-strand is formed after 15 ns of this simulation, and it involves the ordering of residues 37–41 near the C-terminal region of hevein. For simulation disac2 (Fig. 2b), beginning with a different positioning of the carbohydrate on the protein surface, no α-helical secondary structure for residues 28–31 was detected. Only minor transitions to this motif in this region were observed during the 40 ns of the simulation. However, formation of an ordered α-helical structure in the N-terminal region was observed during the last 15 ns of the simulation, together with the formation of a partially ordered β-strand in the C-terminal region. A stable β-sheet is observed in the central region of the protein, present for most of the simulated time, in agreement with the representative structure derived for the protein using NOE restraints and refinement. In the case of trisac, the secondary structure content appears to be intermediate between the two described above: the α-helix involving residues 28–31 appears several times during the simulation, while the C-terminal region remains mostly disordered (Fig. 2c). The central protein

region between residues 16 and 27 was shown by NMR to be particularly structured in the complex: the secondary structure analysis of the three simulations shows that this region can actually be involved in the formation of β-hairpin type of structure. The RMSD values for this region are low, between 0.13 and 0.14 nm. In general, the differences indicate that the description of a complex phenomenon like binding cannot be easily described based only on simulations of only a few ns, or from simple minimization methods.

A more detailed description of the mobility of the protein residues was deduced by calculating the root mean square fluctuation (RMSF) per residue, as shown in Figure 3a–c. RMSF is a good measure of the flexibility of the system during a given time period, and can give valuable information about the deviations of the residues involved in binding and in secondary structure variations with respect to the starting conditions. In disac1, residues W21 and W23, over which the carbohydrate is initially positioned, show the lowest flexibility profile, while residue Y30, with minor interactions with the sugar during this simulation, is characterized by a higher mobility. In disac2, W21 being less involved in the interaction with the ligand is the most flexible of the three putative binding residues, most likely due to the fact that the contacts between the protein and the disaccharide in the starting structure are minimal at this residue. W23 and Y30, over which the ligand is sliding, show a higher flexibility as compared to the previous simulation. The simulation involving the trisaccharide shows a somewhat average situation, with W21 and W23 showing a minimum flexibility and residue Y30, a moderate one. It should be noted that, in all cases, residue Y30 belongs to the protein region undergoing a change in the secondary structure, at the same time being able to establish contacts with the substrate (vide infra). Therefore, depending on the involvement or not of Y30 in sugar binding, different secondary structure conformations of hevein may take place in this region.

These preliminary observations suggest that the secondary structure content and evolution, as well as other global conformational features, have some degree of dependency on motion of the ligand on the receptor. Particularly in the case of the binding to disaccharides, which are hypothesized to have a higher mobility on the protein surface, a mutual interplay of factors should be invoked to modulate the interactions between the binding partners.

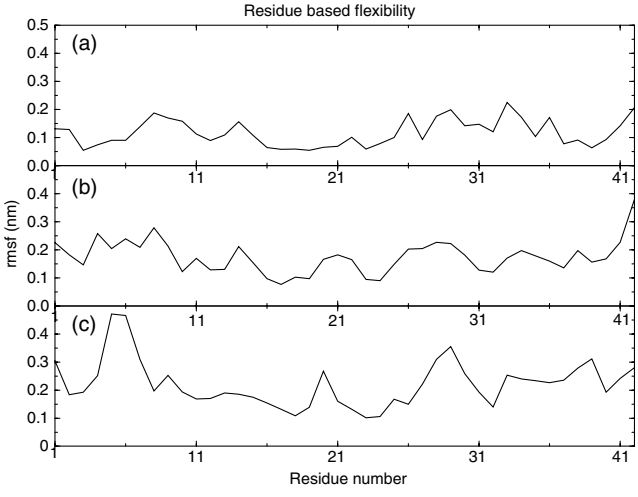

Figure 3. Residue based root mean square fluctuation (flexibility) in hevein: (a) disac1, (b) disac2, (c) trisac.

2.2. The carbohydrates

The conformational mobility of the bound carbohydrates was investigated by analyzing the time variations of the two dihedral angles involving the glycosidic linkages. The torsion angles were defined as $\phi(O'_5–C'_1–O_1–C_4)$ and $\psi(C'_1–O_1–C_4–C_5)$ for the disaccharide complexes, and as $\phi(O'_5–C'_1–O_1–C_4)$, $\psi(C'_1–O_1–C_4–C_5)$, $\phi 2(O''_5–C''_1–O'_1–C'_4)$, and $\psi 2(O''_5–C''_1–O'_1–C'_4)$ for the trisaccharide complex. The time evolution of the angular values is reported in Figure 4 and the average angular values are reported in Table 1. During both

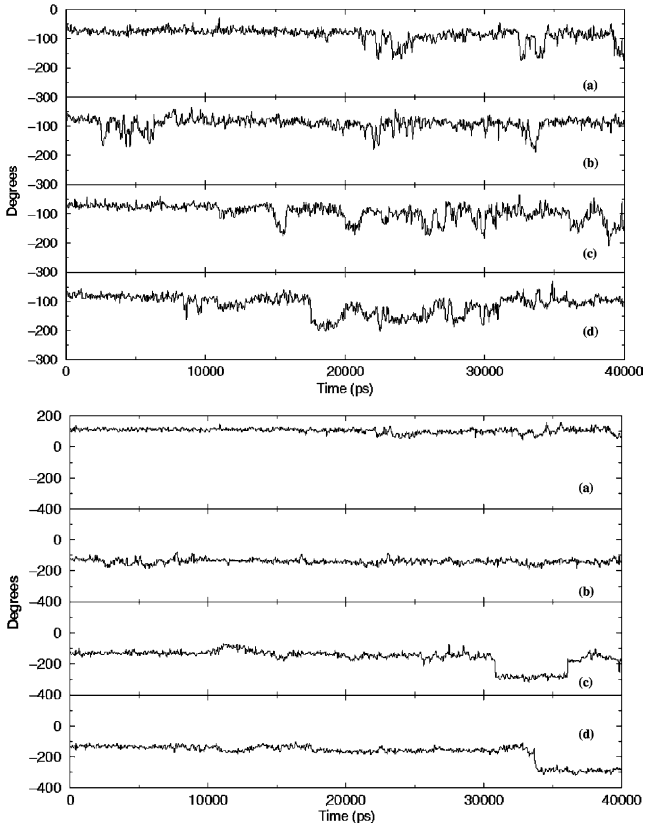

Figure 4. Time evolution of the ϕ (upper panel) and ψ angles. Subfigure labeled (a) refer to simulation disac1, (b) to simulation disac2. Subfigures labeled (c) and (d) refer, respectively, to $\phi 1$ and $\phi 2$ (upper panel) and to $\psi 1$ and $\psi 2$ (lower panel) in the trisaccharide simulation.

Table 1. Summary of the average dihedral values of the simulated carbohydrates in complex with hevein

Simulation	ϕ	ψ	$\phi 2$	$\psi 2$
Disac1	−85.0	−70.0	Abs	Abs
Disac2	−90.7	−130.0	Abs	Abs
Trisac	−80.0	−80.0	−110.7	−169.5

disac1 and disac2 simulations, ϕ (4a and b—upper panel) shows moderate flexibility around the NMR experimental value of −70° (the av value being −85 with a standard deviation of 22 for disac1, and −90 with a standard deviation of 21 for disac2). Several short time transitions to dihedral values of about −170° are observed in the last 20 ns of simulation disac1. These short time range transitions occur much more often during disac2 simulation and characterize the whole 40 ns of the simulation. The values are in good agreement with previously reported NMR and computational studies for β-linked oligosaccharides. The behavior of ψ for the disaccharide complexes (4a and b—lower panel), on the other hand, displays different features between the two simulations. For disac1, a value of −70° remains stable for the whole simulation. In contrast, for disac2 simulation, the value oscillates around −130°. In any case, both values are in agreement with the NMR experimental observations for chitin oligomers.

In the case of the trisaccharide complex (trisac), similar situations are observed, the value of ϕ, involving the reducing-end carbohydrate, is stable around a value of −80° for 17 ns, and shows occasional transitions to −160°, in the second part of the simulation. Dihedral $\phi 2$ value also varies between both conformational states (4c and d—upper panel). The behavior of ψ and $\psi 2$ is similar to that described for the disac simulations, with a certain degree of stability and short time transitions in the last part of the 40 ns simulation (4c and d—lower panel).

The different conformational behavior of the bound carbohydrates as a function of the starting is also evidenced by the different number of transitions involving the rotamer around C-6 as shown in Figure 5. The comparison between the graphs regarding the two dimers prove a definite difference in the frequency of transitions regarding this dihedral angle.

With regards to the behavior of the ring dynamics, no ring-conformational interconversion was observed during the all time length simulated. There may be different causes for that: first of all, the interaction between the protein and the carbohydrates can stabilize the rings in the conformation found in NMR experiments, making a conformational change is not favorable. The second possible cause should be ascribed to insufficient simulation time length. Finally, errors in force field parameterization might artificially 'freeze' the sugar rings.

These preliminary results and analysis are important. Even when bound to the protein, the oligosaccharides

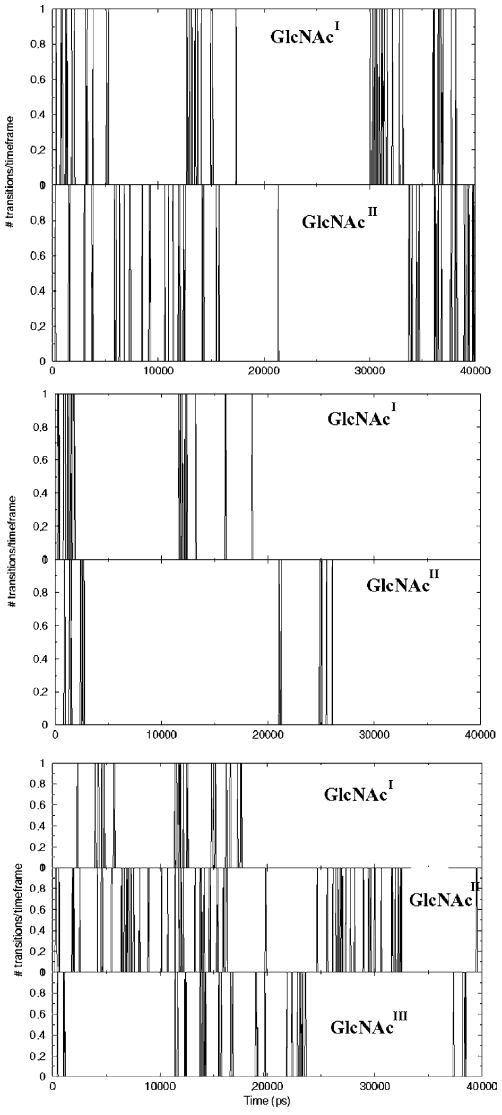

Figure 5. Conformational transitions of the rotamer around C-6 in the three simulations. Panels from top to bottom refer to disac1, disac2, trisac.

under study tend to maintain a certain degree of conformational flexibility, which parallels the flexibility of the protein, and therefore, modulate the possibility of undergoing secondary structure variations. Moreover, and from the MD methodological point of view, it is noteworthy to point out the importance of performing long time-scale solvated simulations. Even under these conditions, it is evident that there is a certain dependence of the results on the starting conditions, and therefore, care should be taken when extracting conclusions from MD simulations alone, since several factors can influence the final results.

2.3. The complexes

As mentioned above, these complexes have been previously studied by NMR methods and their structures

elucidated by conventional methodology using restrained molecular dynamics with the experimental NOE data as distance constraints. The stability of the contacts present in the NMR–NOE starting structures was analyzed as a function of the simulation time. A contact between two groups is considered to be present in the simulation if the minimum distance between atoms of the two groups is lower than 0.6 nm. In disac1 simulation, the contacts between $GlcNAc^{I}$ and W21 are present and stable for the first 25 ns, but they are lost in the remaining part of the simulation. For disac2, the contacts $GlcNAc^{I}$–W21 are present in the first 5 ns, then lost, and recovered between 25 and 28 ns. A second contact, which is initially present is $GlcNAc^{II}$–W23, which displays basically the same intermittent behavior as the previous one. A third putative binding subsite that could be represented by C-24, as hypothesized in the experimental papers is not clear.[7] Our MD simulations show that in both disac1 and disac2 runs, only the GlcNAc-Y30 contact are observed, while C-24 is always far from the sugar. In the trisaccharide simulation, the extended surface of W21 provides contacts with both $GlcNAc^{I}$ and $GlcNAc^{II}$ moieties. These contacts are present for 60% of the simulation time. Additional stabilization of the complex comes from $GlcNAc^{II}$ and $GlcNAc^{III}$, which provide relatively stable contacts with W23 and Y30. Two additional contacts, involving intermolecular hydrogen bonding, have been also characterized by the NMR studies. The first one involves the hydroxyl group of Y30 and the OH-3 group of the nonreducing end ($GlcNAc^{III}$) of the bound oligosaccharide. The second one is featured by the OH group of S19 with the carbonyl moiety of the *N*-acetyl group of the same $GlcNAc^{III}$. The results of the simulations show that the former is present for about 50% of the time in simulation disac1, only 12.5% of the time in disac2, and again for 50% of the time in the simulation with the trisaccharide. Conversely, the second one is only present for about 5% of the time in both disaccharide simulations. For the trisaccharide, this hydrogen bond is more stable and is present for about 50% of the time. All these data show that the complex is a dynamic entity, with the oligosaccharide ligand moving around the exposed binding site of the protein. The Ramachandran plots for the four putative residues involved in binding (Fig. 6a–d) once more underline the dynamic nature of the complexes, showing how the different aminoacids populate different conformational states depending on the motion of the ligand. This feature is also exemplified in Figure 7 showing the evolution of the $GlcNAc^{I}$–Trp21 interaction in the three simulations.

Moreover, the MD simulations show that a higher number of stabilizing interactions are formed by increasing the dimensions of the bound ligand, passing form the di- to the trisaccharide.

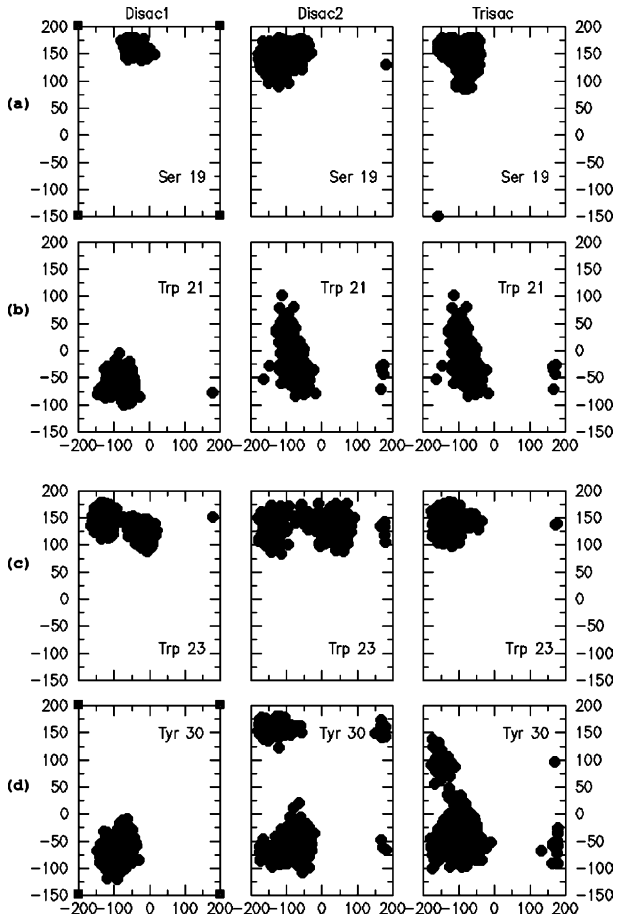

Figure 6. Ramachandran plots for the four putative residues involved in binding. Column one corresponds to disac1 simulation, column two corresponds to disac2, and three to disac3. Row (a) corresponds to Ser19, (b) Trp21, (c) Trp23, and (d) Tyr30.

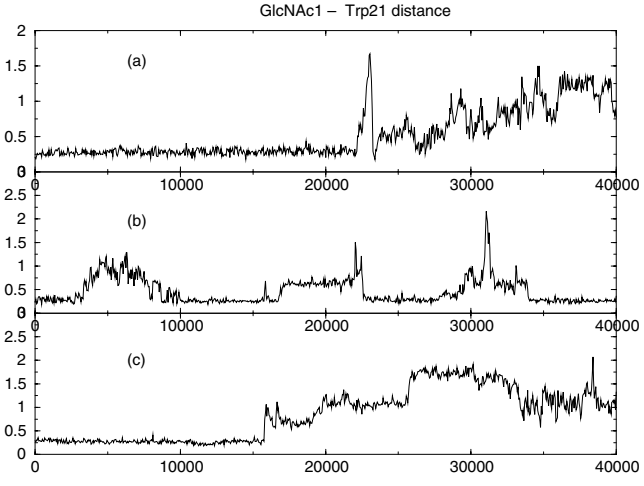

Figure 7. Time evolution of the stacking interaction between $GlcNAc^{I}$–Trp21 in the three simulations, (a) disac1, (b) disac2, (c) trisac.

In a further step, a clustering method was applied to define the most populated structures of the complexes

Table 2. Main distances between carbohydrate subunits and the aromatic protein residues of the binding pocket in disac1, disac2, and trisac complexes

Simulation	1—W21	1—W23	1—Y30	2—W21	2—W23	2—Y30	3—W21	3—W23	3—Y30
Disac1	0.6	0.3	0.6	0.4	0.6	0.9	Abs	Abs	Abs
Disac2	0.6	1.4	0.5	1.0	1.6	0.6	Abs	Abs	Abs
Trisac	0.7	0.6	1.6	0.5	0.5	1.1	1.0	1.4	1.1

Distances in nm. Number 1, 2, and 3 in the first row of the column correspond to the carbohydrate monomers $GlcNAc^{I}$, $GlcNac^{II}$, $GlcNAc^{III}$.

obtained through the long time-scale MD simulations. The main structural features regarding the main clusters are summarized in Table 2.

In principle, this method should be useful to statistically characterize the distinct conformational families, or ensemble of complexes, which are sampled during the simulations. Herein, the applied clustering procedure was based on the method described by Daura et al.[16] A different number of clusters were found for the different simulations. The total number of clusters resulted to be 174 for disac1, and 236 for disac2. This difference indicates that the complexes, in the two different conditions, are exploring different regions of the available conformational space. The higher number of clusters found for disac2 indicates that this simulation has been more efficient in covering a wider portion of the conformational space. In contrast, the hevein–trisaccharide complex simulation is characterized by the presence of only nine different clusters, with the first two of comparable population and representing together almost 90% of the total number of structures. The structures are reported in Figure 8a–d. Thus, the presence of a longer oligosaccharide chain seems to stabilize the complex in different ways.

A detailed analysis was performed for the most representative structures. The most populated cluster for disac1 (Fig. 8a) resembles very closely the starting conformation obtained by the NOE analysis. The reducing-end sugar moiety (here capped with an OMe group) is stacked in a face to face orientation with the aromatic ring of W21, while, simultaneously, the $GlcNAc^{II}$ displays the same kind of carbohydrate–aromatic interaction with W23. This observation is consistent with the hypothesis of the existence of one binding site involving

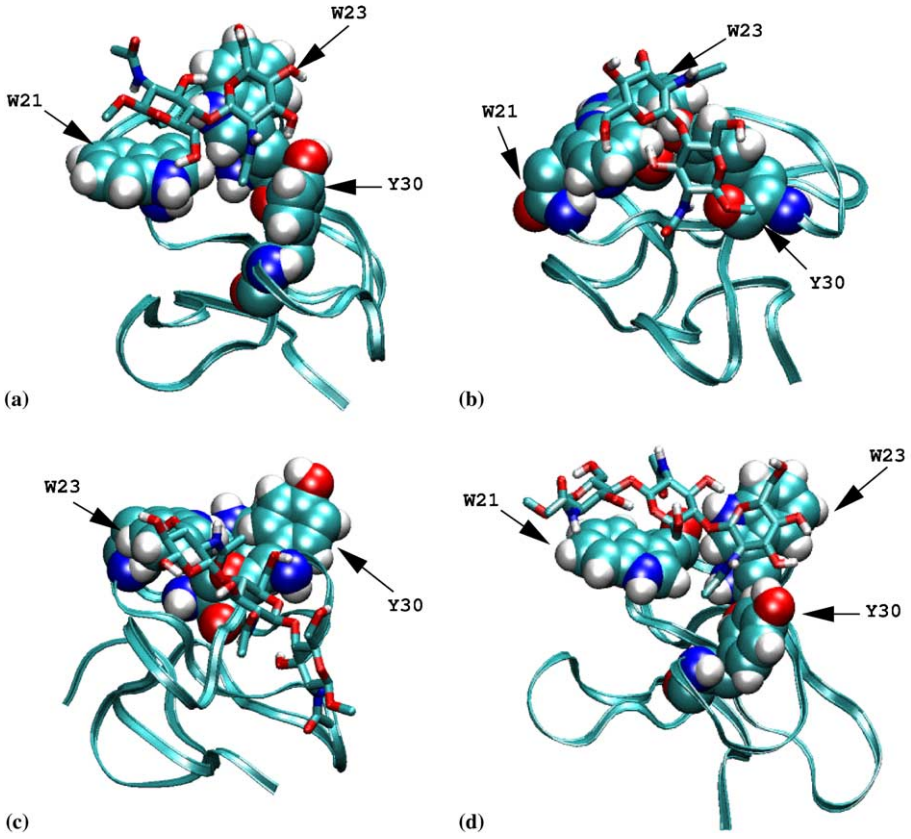

Figure 8. Representative structures of the most populated clusters of conformations from simulations: (a) disac1, (b) disac2, (c) trisac-cluster 1, (d) trisac-cluster2.

several subsites. Thus, in this case, the sugar binds to subsites +1 (defined by W21) and +2 (defined by W23). In disac2 (Fig. 8b), the starting structure is characterized by a longer distance between $GlcNAc^{II}$ and W23, and the sugar is twisted with respect to the orientation it has in the starting structure of disac1. During the MD simulation, the carbohydrate tends to migrate along the exposed binding site at the surface of the protein and, eventually, a new stacking interaction between $GlcNAc^{II}$ and Y30 is established. The contact between W21 and $GlcNAc^{I}$ is present for most of the simulation time and is also present in the representative (central) structure of the most populated cluster. These observations are in good agreement with the previous hypotheses about the existence of a third possible binding site for *N*,*N*′-diacetylchitobiose. In the hevein–trisaccharide complex the situation is somewhat different, but also in agreement with the existence of motion for the sugar along the protein surface. In the two most representative clusters, which are equally populated, and analogously to the observation for the disaccharide simulations, the sugar moieties form favorable stacking-type interactions, with W21 and W23 (Fig. 8d). In the third most populated cluster, the trisaccharide has migrated to a new binding pocket defined by T27, D28, E29, P33, D34, and N35. The terminal sugar moiety of the trimer defines a contact with Y30 via its acetyl group (Fig. 8c).

Finally, the differential behavior of the two hevein–disaccharide complexes was also analyzed using ED analysis,[14] as a function of the starting structures and the different set of initial velocities. This method allows the characterization of relevant collective motions (essential motions) occurring during the MD simulation and their distinction from thermal noise motions. The analysis is performed by building the covariance matrix of the positional fluctuations obtained from the MD simulations. Upon diagonalization of this matrix, a set of eigenvalues and eigenvectors is generated, which define a new set of generalized coordinates. The eigenvectors correspond to directions in a $3N$ dimensional space (where N is the number of atoms) along which collective motions of atoms occur, and yield important insights into the global conformational behavior of the system. The eigenvalues represent the total mean square fluctuation of the system along the corresponding eigenvector.

ED analysis, in this case, provided information about collective motions for the protein under the influence of the di- and trisaccharide. The three trajectories for the proteins were combined to obtain the overall eigenvectors, and subsequently each of the single protein trajectories was projected on the plane defined by the first two eigenvectors associated to first two eigenvalues. From the obtained eigenvalues, it is clear that only the first eight eigenvalues are necessary to describe the 80% of the overall protein fluctuations.

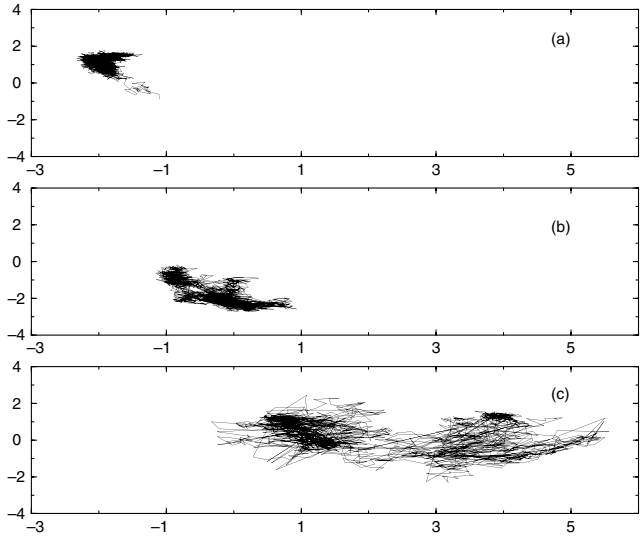

Figure 9. Projections of the protein trajectories along the first two eigenvectors of the covariance fluctuation matrix: (a) disac1, (b) disac2, and (c) trisac.

The projections of the trajectories of the protein in simulations disac1 and disac2 on the planes defined by eigenvectors 1 and 2 (describing the highest nonrandom displacements) show that the two trajectories are sampling a different portion of the conformational space, depending on the starting condition, disac1 simulation is confined to one single region, while disac2 seems to span a wider portion of conformational space (cf. Fig. 9a and b). Again, the trisaccharide–hevein complex samples an extremely different subspace, including part of the subspace visited by simulation disac2 (cf. Fig. 9c): the protein maintains a high flexibility, which could counterbalance for the entropy loss upon binding.

3. Discussion

Carbohydrate–protein recognition processes are involved in the initial steps of a large number of physiological and pathological processes. Interference with these recognition events could be used to modulate or alter signal transmission, or to prevent the onset of diseases. To this aim, a comprehension of the molecular reasons for the binding between carbohydrates and protein is of paramount importance.

Our simulations suggest distinct possible binding mechanisms for different oligosaccharides, even belonging to the same type. These considerations are naturally of qualitative type and aim at providing a possible molecular picture of the events taking place in the recognition phenomena between hevein and its ligands. Present day simulation technology does not give access to the quantitative calculation of the free energy of binding allowing a direct comparison with experi-

mentally determined quantities. However, the use of different conditions, different systems and the statistical analysis of multiple trajectories based on clustering methods should provide a physically based description of the formation of the complex between hevein and the chitin oligomers. Moreover, the use of ED allows the characterization of the relevant (nonrandom) collective motions (called essential motions) and their distinction from thermal noise motions. The first two eigenvectors in general describe the largest motions, which are the most poorly sampled in MD, and can have a large random diffusive component. The use of a combination of statistical based approaches together with structural analysis of the trajectories is thus necessary to extract the most relevant information from the set of configurations obtained.

Despite their extensive nature, disac1 and disac2 simulations do not converge to one single structure: this could be ascribed to either lack of sampling or to the fact that multiple conformations in equilibrium co-participate to complex formation. The results of the simulations for the disaccharide complexes show that a carbohydrate oligomer is able to move on the surface of the relatively flat binding pocket of hevein, thus occupying different binding subpockets. In fact, the combined results of clustering and the ED analysis show how the two dimers in the disaccharide simulations may sample different parts of the conformational space, in this way modulating the protein motions. The two different binding modes, involving either contacts with W21 and W23 or W23 and Y30, involve basically the same type of energetic interactions, mainly aromatic-sugar ring stacking type. The capability for the disaccharide to move (slide) on the surface of the protein domain implies the possibility of multiple binding modes of the disaccharide, thus giving favorable entropic contributions to the complex formation. In the case of the trisaccharide, the longer chain of the ligand allows for establishing a higher number of favorable contacts. The possible lower mobility of the trisaccharide on the surface of the protein may determine a loss of entropy upon binding, which is counterbalanced partly by the higher number of contacts with the protein, and partly by an increase in the overall flexibility (breathing motions) of the protein.

The oligosaccharide binding can thus be considered as defined by a subtle balance between enthalpic (formation of intermolecular interactions between the ligand and the receptor) and entropic (due either to the possibility for the sugar to move on the surface of the protein domain or to an increase in protein 'breathing' motions) effects, giving rise to multiple binding conformations. This is probably the case of phenomena that take place in the biological process. The protection mechanism probably involves the interaction of the long chitin chain with a number of hevein domains, and chitin itself is able to interact with a variable number of protein residues sliding on the accessible protein surface.

As a caveat to the use of all-atom MD simulations in the analysis of such complexes, one should keep in mind that even with very extensive simulations, the sampling of the conformational space for very flexible complexes in not complete. Thus a combination of different conditions and statistical analysis is necessary to improve our understanding of flexible systems, like those studied herein. This should be considered when discussing of very flexible and dynamical complexes, for which a representation based on only one model is often considered as a result of the application of experimental restraints that actually represent an overall averaging over a wide number of possible system configurations. However, the application of suitable analysis techniques (ED and clustering in particular) is able to yield useful information on the description of the dynamics of the bimolecular complexes studied. We feel that the approach outlined herein may also be extended in a general manner to the study of a variety of carbohydrate–protein complexes.

4. Methods

The starting structures for the simulations (Fig. 1) were obtained from the NMR analysis of NOESY spectra recorded for the complexes of hevein to a number of chitin oligomers, as described. The coordinates are available from the authors upon request. The hevein domain was protonated to give a zwitterionic form (with N-terminal NH_3^+ and C-terminal COO^- groups). The oligosaccharides were built using the building blocks obtained from Balaji and co-worker.[17] Three complexes were built according to the NMR structures: two hevein–disaccharide complexes and one hevein–trisaccharide complex.

The complexes were solvated with water in a periodic truncated octahedron, large enough to contain the peptide and 0.9 nm of solvent on all sides. All solvent molecules within 0.15 nm of any peptide atom were removed. The total charge on the complexes was +2. No counterions were added as water is a high dielectric and the inclusion of no-counterions was considered a better approximation to the low salt experimental conditions. The resulting system was composed of the complex and 3016 water molecules in the case of the disaccharide simulations, and of the complex plus 2967 water molecules in the case of the trisaccharide simulation.

The system was subsequently energy minimized with a steepest descent method for 1000 steps. In all simulations, the temperature was maintained close to the intended value of 300 K by weak coupling to an external temperature bath[18] with a coupling constant of 0.1 ps. The protein, sugars, and the rest of the system were

coupled separately to the temperature bath. The GROMOS96 force field[19,20] was used. The simple point charge (SPC)[21] water model was used. The LINGS algorithm[22] was used to constrain all bond lengths. For the water molecules, the SETTLE algorithm[23] was used. A dielectric permittivity, $\varepsilon = 1$, and a time step of 2 fs were used. A twin range cut-off was used for the calculation of the nonbonded interactions. The short range cut-off radius was set to 0.8 nm and the long range cut-off radius to 1.4 nm for both Coulombic and Lennard-Jones interactions. The cut-off values are the same as those used for the GROMOS96 force field parameterization.[19] Interactions within the short range cut-off were updated every time step whereas interactions within the long range cut-off were updated every five time steps together with the pairlist. All atoms were given an initial velocity obtained from a Maxwellian distribution at the desired initial temperature. The density of the system was adjusted performing the first equilibration runs at NPT condition by weak coupling to a bath of constant pressure ($P_0 = 1$ bar, coupling time $\tau_P = 0.5$ ps).[18] All the simulations, starting from the appropriate NMR structure, were equilibrated by 50 ps of MD runs with position restraints on the peptide to allow relaxation of the solvent molecules. These first equilibration runs were followed by other 50 ps runs without position restraints on the peptide. The production runs using NVT conditions, after equilibration, were 40 ns long for all of the three complexes, and labeled, respectively, disac1, disac2, and trisac.

All the MD runs and the analysis of the trajectories were performed using the GROMACS software package.[24] Cluster analysis was performed according to Daura and co-workers.[25]

References

1. Lemieux, R. U. *Chem. Soc. Rev.* **1989**, *18*, 347–374.
2. Lasky, L. A. *Science* **1992**, *258*, 964–969.
3. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.
4. von der Lieth, C. W.; Siebert, H. C.; Kozar, T.; Burchert, M.; Frank, M.; Gilleron, M.; Kaltner, H.; Kayser, G.; Tajkhorshid, E.; Bovin, N. V.; Vliegenthart, J. F. G.; Gabius, H. J. *Acta Anat.* **1998**, *161*, 91–109.
5. Rini, J. M. *Curr. Opin. Struct. Biol.* **1995**, *5*, 617–621.
6. Peters, T.; Pinto, B. M. *Curr. Opin. Struct. Biol.* **1996**, *6*, 710–720.
7. Asensio, J. L.; Siebert, H. C.; von der Lieth, C.; Laynez, J.; Bruix, M.; Soedjanaamadja, U. M.; Beintema, J. J.; Canada, F. J.; Gabius, H. J.; Jimenez-Barbero, J. *Proteins: Struct. Fund. Genet.* **2000**, *40*, 218–236.
8. Asensio, J. L.; Canada, F.; Siebert, H. C.; Laynez, J.; Poveda, A.; Beintema, J.; Gabius, H. J.; Jimenez-Barbero, J. *Chem. Biol.* **2000**, *7*, 529–543.
9. Tempel, W.; Tschampel, S.; Woods, R. *J. Biol. Chem.* **2002**, *277*, 6612–6615.
10. Lescar, J.; Loris, R.; Mitchell, E.; Gautier, C.; Chazolet, V.; Cox, V.; Wyns, L.; Perez, S.; Breton, C.; Imberty, A. *J. Biol. Chem.* **2001**, *277*, 6608–6614.
11. Bernardi, A.; Galgano, M.; Belvisi, L.; Colombo, G. *J. Comp.-Aided Mol. Des.* **2001**, *15*, 117–128.
12. Asensio, J. L.; Canada, F. J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J. *Eur. J. Biochem.* **1995**, *230*, 621–633.
13. Asensio, J. L.; Canada, F. J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J. *Glycobiology* **1998**, *8*, 569–577.
14. Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct. Fund. Genet.* **1993**, *17*, 412–425.
15. Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2576–2637.
16. Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925–932.
17. Vasudevan, S. V.; Balaji, P. V. *J. Phys. Chem. B* **2001**, *105*, 7033–7041.
18. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
19. van Gunsteren, W. F.; Daura, X.; Mark, A. E. *Encyclopedia Comput. Chem.* **1998**, *2*, 1211–1216.
20. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschul, ETH: Zürich, Switzerland, 1996.
21. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
22. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
23. Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
24. van der Spoel, D.; van Drunen, R.; Berendsen, H. J. C. *GRoningen MAchine for Chemical Simulations*, Department of Biophysical Chemistry, BIOSON Research Institute, Nijenborgh 4 NL-9717 AG Groningen, 1994, e-mail to gromacs@chem.rug.nl.
25. Daura, X.; Antes, I.; van Gunsteren, W. F.; Thiel, W.; Mark, A. E. *Proteins: Struct. Fund. Genet.* **1999**, *36*, 542–555.