Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 340 (2005) 1039–1049
Carbohydrate RESEARCH

A dynamic perspective on the molecular recognition of chitooligosaccharide ligands by hevein domains

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

[a]*Istituto di Chimica del Riconoscimento Molecolare, CNR via Mario Bianco, 9, 20131 Milano, Italy*
[b]*Centro de Investigaciones Biologicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain*

Received 13 October 2004; accepted 24 January 2005

Dedicated to Professor David A. Brant

Abstract—The complexes between hevein and different chitin oligomers, from the di- to the penta-saccharide, are studied through all atom molecular-dynamics simulations. The results for the smaller oligosaccharide complexes show that the carbohydrate is able to move on the surface of the relatively flat binding-pocket of hevein, therefore occupying different binding subpockets. The pentasaccharide spans all possible intermolecular interactions with the receptor in a simultaneous manner. Statistical analysis methods were also applied in order to define the principal overall motions in the complexes. The oligosaccharide binding can be considered to be defined by a subtle balance between enthalpic and entropic effects, providing the possibility of the existence of multiple binding conformations. This structural and dynamical view parallels the results based on NOE NMR data for the three disaccharide, trisaccharide, and pentasaccharide complexes.

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

1. Introduction

The clarification of the physicochemical determinants underlying carbohydrate–protein molecular recognition is critical for our understanding of a series of fundamental processes such as cell–cell adhesion, cell growth, inflammatory reactions, fertilization, and bacterial infection. The study of carbohydrate–protein binding is thus a topic of major interest because of its potential in pharmaceutical and clinical applications.[1–3]

A great deal of information has been obtained recently through a variety of studies ranging from X-ray crystallography to NMR spectroscopy, from calorimetry to surface plasmon resonance, and so on.[4] From the experimental point of view, X-ray crystallography has provided a wealth of information on the three-dimensional structures of protein–carbohydrate complexes.[5] In many cases, however, 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. In a parallel way, NMR experiments have been frequently used both in the area of oligosaccharide conformational analysis and for studying oligosaccharide–protein interactions.[6,7] The complexity of the biological problems involved in protein–carbohydrate interactions underlines the need for simplified systems, in which the role of single monomers or chemical groups in the process can be pinpointed at atomic resolution. In order to do this, theoretical investigations on model systems of biological relevance constitute a suitable choice.

In this context, we have undertaken the investigation of the binding features of carbohydrate oligomers of different lengths as probes to a small-sized lectin (namely,

* Corresponding authors. Tel.: +39 02 28500031; fax: +39 02 28500036 (G.C.); tel.: +34 918373112/4370; fax: +34 915360432 (J.J-B.); e-mail addresses: g.colombo@icrm.cnr.it; jjbarbero@cib.csic.es

0008-6215/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.carres.2005.01.044

hevein), by combining long time-scale all-atom molecular dynamics (MD) simulations with previously obtained NMR data.[8]

Hevein is a small single-chain protein of 43 amino acids, especially rich in glycine and cysteine residues, whose biological function is the recognition of chitin oligosaccharides. Hevein is present in laticifers of the rubber tree: from the biological viewpoint, it is the major allergen of latex and, in addition, it has been proposed to be involved in the mechanisms of plant defense against insect and fungal attacks. Its 3D structure in solution has been extensively studied by X-ray crystallography and by NMR methods.[9,10] Its binding to chitin oligosaccharides has been studied by NMR spectroscopy. Since the exoskeleton of insects and fungi is formed of chitin, the first step in the protection provided by hevein could be the molecular recognition of chitin by this small protein domain.[6]

Experimental investigations have proposed the existence of an exposed binding site for hevein, comprised of 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. A third subsite, represented by L16, C24, and S26 was identified based on MD simulations of NMR-selected structures in a previous paper by our group. MD simulations allowed identification of a sliding motion of disaccharide and trisaccharide probes on the binding region of the protein.[8]

In this paper, the requirements for ligand–protein interaction and their connection to recognition mechanisms are investigated through the analysis of four different 40-ns long all-atom MD trajectories in explicit water. Two simulations were run for the complex between hevein and *N*,*N*′-diacetylchitobiose in explicit water solvent (labeled disac1 and disac2), starting from two models derived from NMR structural elucidation. These two starting structures were different in the placing of the carbohydrate dimer on the protein surface. In the first one, $GlcNAc^{I}$ and $GlcNAc^{II}$ are stacked in a face-to-face orientation on the aromatic ring of W21 and W23, respectively. In the second starting structure, the $GlcNAc^{II}$–W23 distance is longer and the disaccharide is twisted with respect to the previous complex (Fig. 1 and Ref. 8). Two more simulations were run for the hevein–trisaccharide (label trisac) and the hevein–pentasaccharide (label penta) complexes, starting from the NMR derived structures (Fig. 1). The trajectories were analyzed by using statistical approaches, such as cluster analysis and the essential dynamics (ED)[11] method: their combination allows extraction of the main features of the dynamics of the ligands and investigation of the conformational space similarities and differences in the carbohydrate-recognition motifs. Understanding minimal

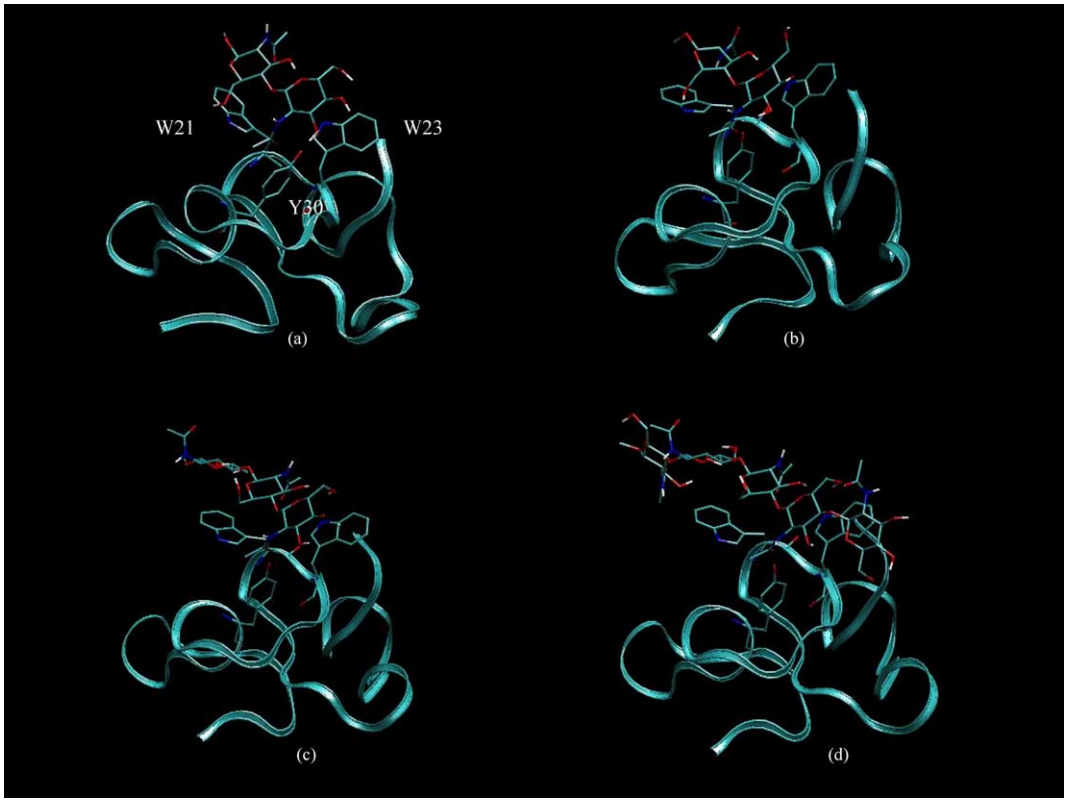

Figure 1. The structures of the starting complexes, as obtained from NMR data. Starting structure for: simulation disac1 (a), disac2 (b), trisac (c), penta (d).

binding requirements is in fact fundamental for the future design of new molecules with improved recognition abilities.

2. Results and discussion

The structure of the receptor protein remained in all cases close to the native conformation, the average RMSD varying between 0.2 and 0.31 nm from the NMR structures (Table 1). The three-dimensional fold of the protein remains globally unchanged, despite the ordering and appearance of β-sheet geometries between 20 and 30 ns in simulations disac2, trisac, and penta, as shown by the DSSP algorithm (see also previous paper).[8] In this paper, however, attention is focused on the behavior of the carbohydrates and on their motions on the surface of the receptor.

Table 1. Structural summary of the simulations

	RMS protein (nm)	Num. of clusters	Ratio (main/second)
Disac1	0.22	174	3.1
Disac2	0.20	236	1.77
Trisac	0.28	9	1.03
Penta	0.3	92	5.8

The first column reports the average protein backbone RMSD from the NMR starting structure. The second column represents the total number of clusters visited in each simulation by the complexes. The third column reports the ratio between the number of structures in the first main structure cluster and the second main structure cluster.

The space sampled by the bound carbohydrates when complexed to the protein was analyzed mainly by the application of statistical analysis methods. In particular, we decided to investigate the similarities between the conformational subspaces spanned by two consecutive carbohydrate units in the bound dimers, in the trimer and in the pentamer. As an example, starting from the non-reducing end (residue 1) and moving towards the reducing end (residue 5) of the pentamer, we could define four different dimeric units composed by residues 1–2, 2–3, 3–4, and 4–5. In the same way, the trimer was divided into two subunits, namely 1–2 and 2–3 (Fig. 2a and b).

The same kind of analysis was also performed comparing the subspaces spanned by three consecutive residues in the pentamer and in the trimer complex. The pentamer was thus divided into three regions (residues 1–3, 2–4, 3–5), and the subspace spanned by these three subregions compared to the subspaces spanned by the trimer bound to the protein (Fig. 2c).

The different conformational behavior of the different subunits identified was first analyzed using the essential dynamics (ED) analysis method.[11] ED 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

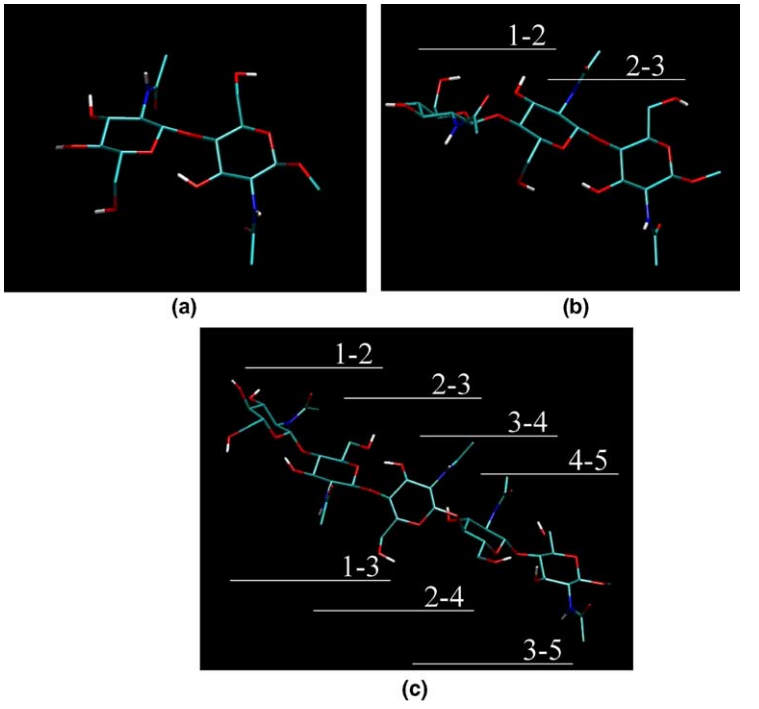

Figure 2. Schematic representation of the 3D structures of the substrates studied, and their subdivision in oligomeric subunits. (a) The dimer, (b) the trimer, and (c) the pentamer.

for each of the afore-defined subunits. The covariance matrix was built using the C atoms defining the pyranose rings of each of the subunits first described (Fig. 2). 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, was used to gain information about collective motions of the carbohydrate subunits as a function of the different oligosaccharide chain-lengths and of the interactions with the receptor protein.

To quantify the similarity of the subspaces spanned, we calculated the subspace overlap (SO) as described in Material and Methods. This number accounts for the (dis)similarity of the essential motions. We have therefore calculated this parameter between the trajectories of the different subunits of the carbohydrates already defined. The results reported in Table 2 show that the subspaces defined for the different subunits partially overlap in some cases, while they might be highly dissimilar.

The trajectories of each of the subunits in the complexes were also projected onto the plane defined by the first two eigenvectors associated to first two eigenvalues.

When the projections for the dimeric sugar subunits are considered, it is immediately apparent that the essential subspaces spanned by the ligand N,N'-diacetylchitobiose, starting from different NMR structures of the complex, are similar (Fig. 3a and b, and Table 2, RMSIP ~ 0.50). A markedly different behavior is seen when the two dimeric subunits in the N,N',N''-triacetylchitotriose are considered. In particular, the subspace spanned by subunit 1–2 seems to be complementary to the one spanned by subunit 2–3 (Fig. 3c and d). Each of the two subspaces partially overlaps with those calculated for the two N,N'-diacetylchitobiose described above (see also Table 2). This observation indicates that the two subunits in the trisaccharide move in a different way than in the previous case on the surface of the protein, trying to establish favorable contacts involving both the reducing and the non-reducing end units. Also based on previously described results, we hypothesized a sliding motion for the trisaccharide in the complex, which is eventually stabilized by a new H-bonding interaction between S19 on the protein and the reducing-end moiety. This sliding motion switches the hydrophobic interactions involving initially the aromatic W21, W23, and Y30 residues of hevein and the 2–3 subunit of the bound carbohydrate, to a situation in which interactions of the sugar with the protein take place mainly with L16, S19, C24, and S26.[7]

In the case of the pentasaccharide–protein complex, the subspaces spanned by each dimeric subunits vary as a function of the position of the subunit in the carbohydrate chain (Fig. 3, subpanels e–h). In particular, the subspace spanned by the 1–2 and 1–3 subunits of the pentamer is more similar to the one spanned by the disaccharide in simulation disac1 (see Table 2.). In contrast, the other three dimeric subunits sample more limited subspaces. This observation, combined with the cluster analysis of the conformations of the complexes visited during the MD, indicates that, in the pentamer, the terminal dimeric subunits defined by 1–2 and 2–3 can partially be reorganized, now defining stable contacts between W21 and subunit 2, and between W23 and subunit 3. This orientation resembles the behavior observed for the N,N'-diacetylchitobiose simulations already described.

In contrast, unlike the previous case, the superposition of the essential subspaces of trimeric units of the pentamer with that of N,N',N''-triacetylchitotriose shows a lower degree of similarity in all cases (Fig. 4, Table 2).

In a further step, the clustering method described by Daura et al.[12] was applied to define the most populated structures of the complexes obtained through the long-timescale MD simulations. To find clusters in a trajectory, the RMSD of atom positions between all pairs of complexes sampled in each simulation was determined. For each structure, the number of other structures for which the RMSD was 0.175 nm or less was calculated.

Table 2. Values of the *normalized overlap* calculated from the covariance matrices of the various trajectories

	Disac2	Trisac(1–2)	Trisac(2–3)	Penta(1–2)	Penta(2–3)	Penta(3–4)	Penta(4–5)
Disac1	0.618	0.22	0.262	0.225	0.2	0.606	0.600
Disac2		0.325	0.234	0.176	0.510	0.536	0.354
Tris(1–2)				0.535	0.506	0.148	0.172
Tris(2–3)				0.282	0.269	0.20	0.270
	Trisac						
Penta(1–3)	0.37						
Penta(2–4)	0.334						
Penta(3–5)	0.218						

The value is 1 for identical matrices and 0 when the sampled subspaces are orthogonal.

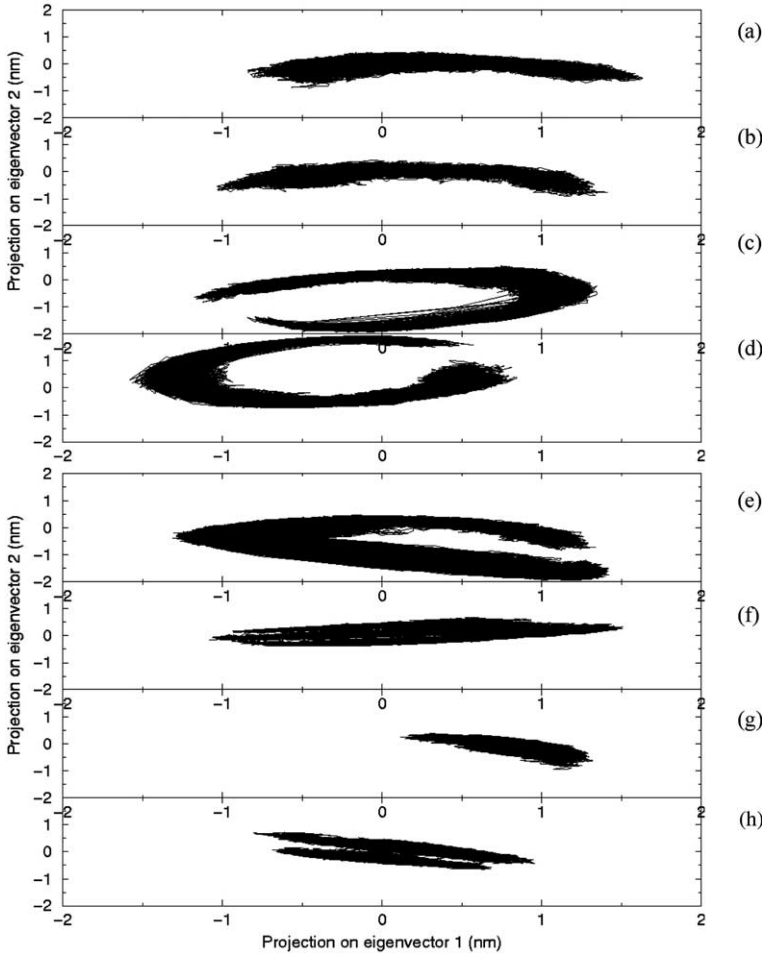

Figure 3. Projections of the carbohydrate dimer trajectories along the two principal components of the covariance fluctuation matrix. (a) Disaccharide 1, (b) disaccharide 2, (c) the 1–2 subunit of the trisaccharide, (d) the 2–3 subunit of the trisaccharide, (e) the 1–2, (f) 2–3, (g) 3–4, (h) 4–5 subunits of the pentasaccharide.

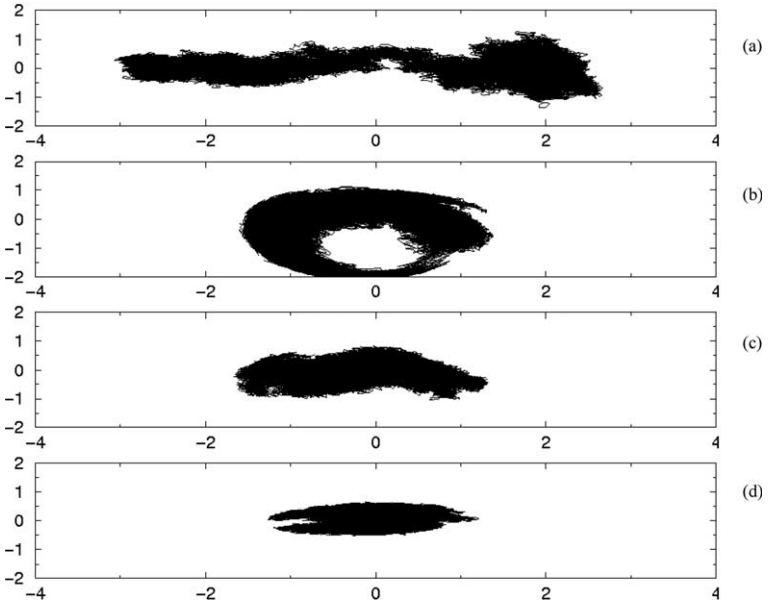

Figure 4. Projections of the carbohydrate dimer trajectories along the two principal components of the covariance fluctuation matrix. (a) Trisaccharide, (b) the 1–3 subunit of the pentasaccharide (c) the 2–4 subunit of the pentasaccharide, (d) the 3–5 subunit of the pentasaccharide.

The structure with the highest number of neighbors was then taken as the center of the cluster, and was taken along with its neighbors to define the first cluster. Then, the structures belonging to this first cluster were removed from the structural ensemble, and the process was repeated with those remaining structures. The process was repeated until the pool of structures was empty. The results are summarized in Table 1.

In principle, this method should be useful to statistically characterize the distinct conformational families, or ensemble of complexes, which are sampled during the simulations. Indeed, different numbers of clusters were found for the different simulations.

In the case of *N*,*N*′-diacetylchitobiose, the total number of clusters turned out to be dependent on the starting conformations. As discussed previously, this difference indicates that the complexes, in the two different conditions, are exploring distinct conformational regions, although the width of the essential subspaces is comparable. These observations suggest that the disaccharide is mainly moving on the surface of the protein around the region determined by W21, W23, and Y30, partially establishing different interactions as a function of the starting structure. Indeed, the *N*,*N*′-diacetylchitobiose–hevein complexes displayed multiple transitions among different clusters, with different profiles depending on the starting conditions. This behavior is shown by the graphs displaying the cluster number populated as a function of time (Fig. 5a and b). However, in both cases, the procedure resulted in the definition of one main structural cluster, whose representative structure is displayed in Figure 6a and b. Interestingly, the relative populations of the first and second main clusters in the two *N*,*N*′-diacetylchitobiose–hevein complex simulations were rather different (see Table 1). This fact could, once more, indicate that the two different starting structures bias the simulations to sample different parts of the phase space, which are accessible and available at the 300 K temperature of the simulations. It is worth pointing out that the transition kinetics among different clusters is also influenced by the starting structures. In the first *N*,*N*′-diacetylchitobiose–hevein (disac1) simulation, a dominant number of transitions to the first (main) cluster could be defined, while for disac2 simulation, a wider distribution of transitions among different clusters was observed (Fig. 7a and b). These observations suggest a dynamic picture of the motion of the ligand on the receptor surface, which should not be limited to the frequently used, but simplified, one-structure for one-complex representation.

In contrast, the hevein–trisaccharide complex simulation is characterized by the presence of nine different clusters, two of them being of comparable population, and representing together almost 90% of the total number of structures. Interestingly, very few transitions among the different clusters were identified (Fig. 7c). The representative structures of these two clusters indicate that the trisaccharide in the hevein complex may sample different subspaces, that include part of those subspaces visited in the two disaccharide simulations, as suggested by ED analysis (Fig. 6c and d).

Finally, the situation is reversed in the case of the pentasaccharide–hevein complex. In this case, the sugar chain is long enough to cover all the possible interaction sites detected by the previous simulations for the shorter oligomers. Cluster analysis yields one major cluster, in

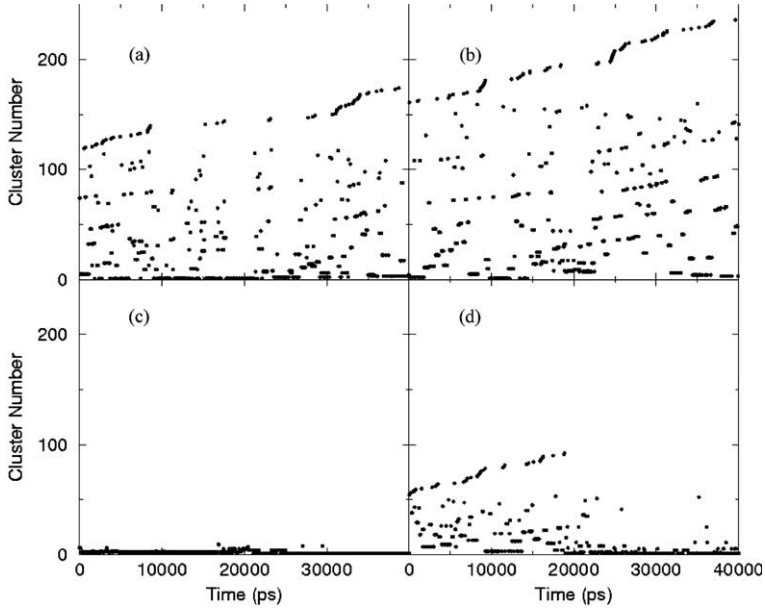

Figure 5. Population of the different clusters as a function of time. Disac1 (a), disac2 (b), trisac (c), penta (d).

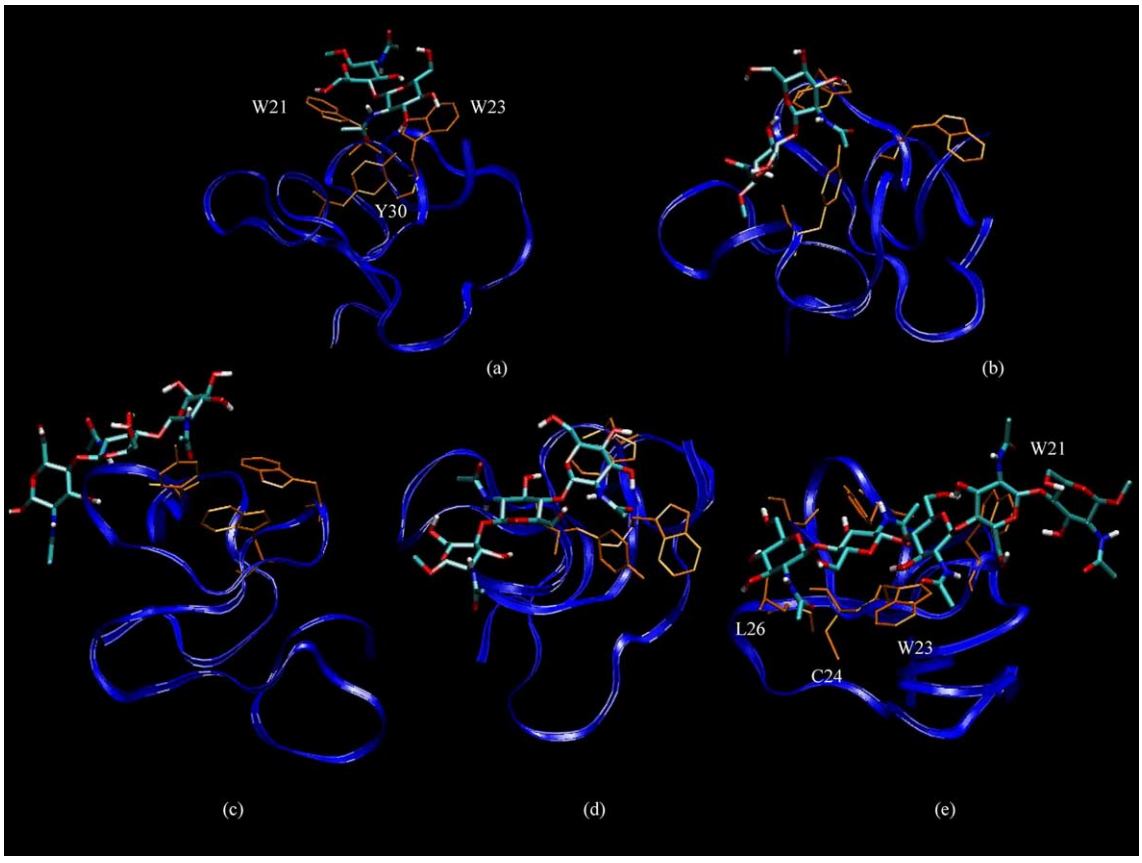

Figure 6. 3D structures of the most representative structure of the most highly populated cluster for the simulations of the disaccharides (a) and (b); of the two most populated clusters (with equal populations) for the trisaccharide simulations (c) and (d); and of the pentasaccharide (e).

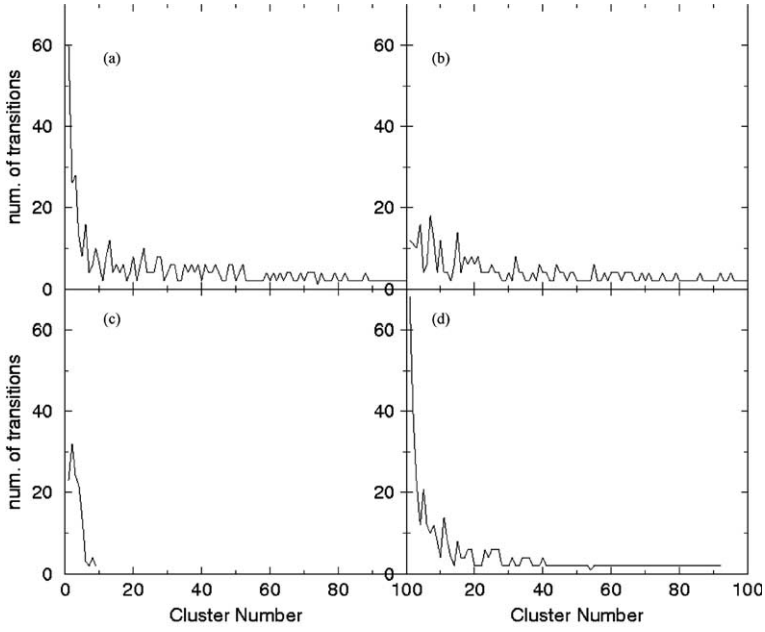

Figure 7. Total number of transitions to each cluster visited during each simulation. Disac1 (a), disac2 (b), trisac (c), penta (d).

which the carbohydrate binds to the W21, W23, Y30 region, as well as to that defined by S26, C24, and L16 (Fig. 6e). Also in this case, the interactions with the protein also modulated the conformational flexibility of the

glycosidic linkages: the analysis of time evolution of the phi and psi torsional angle values showed that the pentasaccharide chain undergoes little conformational changes (Fig. 8), in contrast to the behavior of the trisaccharide that showed larger fluctuations.[8]

These observations were further verified by the calculation of a different geometrical descriptor. In order to characterize the extent of motion of the oligosaccharides on the hevein surface, that is, the time evolution of the minimum distance between the aromatic rings of residues W21, W23, and Y30 and the sugar was calculated. For extracting these values, the central residue of the pentasaccharide, of the trisaccharide, and the GlcNAc^{I} moiety of the disaccharide were considered (Fig. 9). It is immediately apparent that, for the di- and tri-oligomers, the sugar–aromatic distances display a marked variability with the simulation time. In contrast, for the pentasaccharide simulation, the starting configuration of the complex was maintained along the whole simulation time. Nevertheless, this differential behavior could also be due to a slower rearrangement kinetics of the pentasaccharide complex, that might not be caught in the simulation time used in this study. However, this behavior does not hold for the di- and tri-saccharide complexes, where the ligands are not trapped in their initial positions.

We also verified whether the sliding motions of the oligosaccharides were correlated with rotations of the constituting monomers, by looking at the behavior of the vectors that connect the centers of mass of two consecutive subunits. Generally speaking, no correlation was found, and only translational motions of the vectors were shown from this analysis.

Thus, for *N*,*N*′-diacetylchitobiose, the minimal size of the ligand allows it to fit optimally into the binding pocket defined by residues W21, W23, and Y30. For longer oligomers, and first for *N*,*N*′,*N*″-triacetylchitotriose, the additional subunit gives more configurational freedom to the complex, favoring motion on the surface of the protein, and allowing the sugar to establish new interactions with other available sites on the protein, as that defined by S26, C24, and L16. In the case of the pentasaccharide, the length of the carbohydrate chain is sufficient to simultaneously saturate all the possible interactions with the protein, in complete agreement with the NMR observation.[7] The behavior of the pentamer does not parallel that of the trimer, for which sliding of the ligand on the receptor does occur, while the terminal unit can fluctuate in solution, establishing only intermittent contacts with the protein. This situation suggests that the pentasaccharide is indeed ideal for the formation of a stable 1:1 complex with hevein, giving rise to all possible intermolecular interactions. On one hand, it can form a number of stabilizing interactions encompassing all the region from the W21, W23, Y30 patch to the hydrogen-bonding S19 residue. On the other hand, the

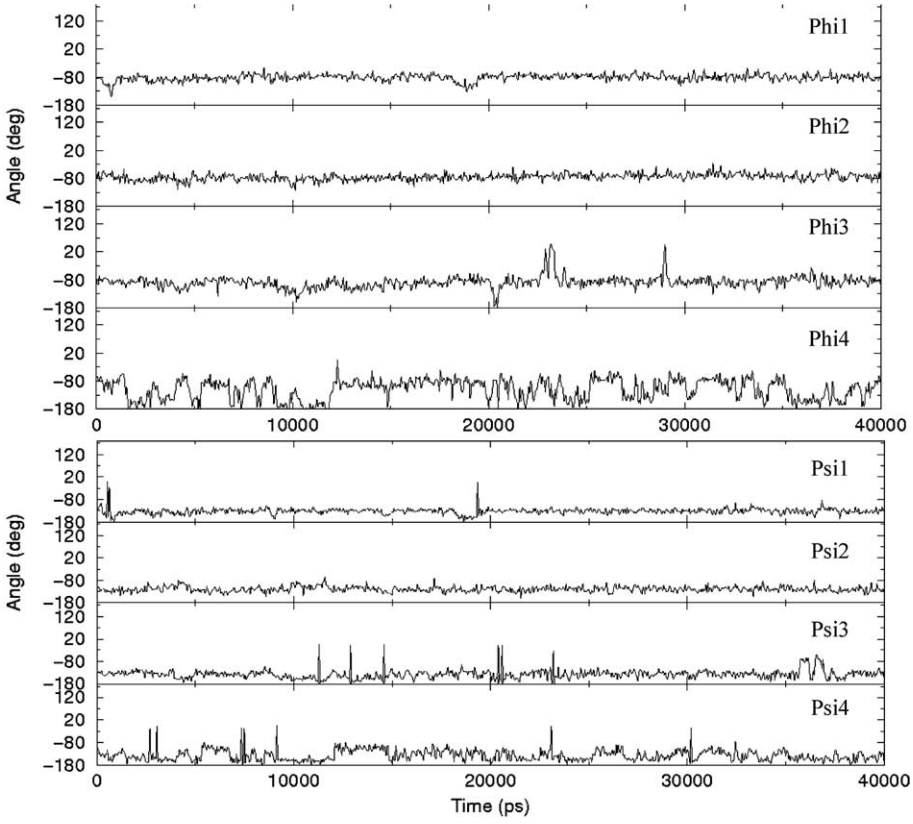

Figure 8. The phi (O'_5–C'_1–O_1–C_4) and psi (C'_1–O_1–C_4–C_5) angle time evolution for the pentasaccharide simulation.

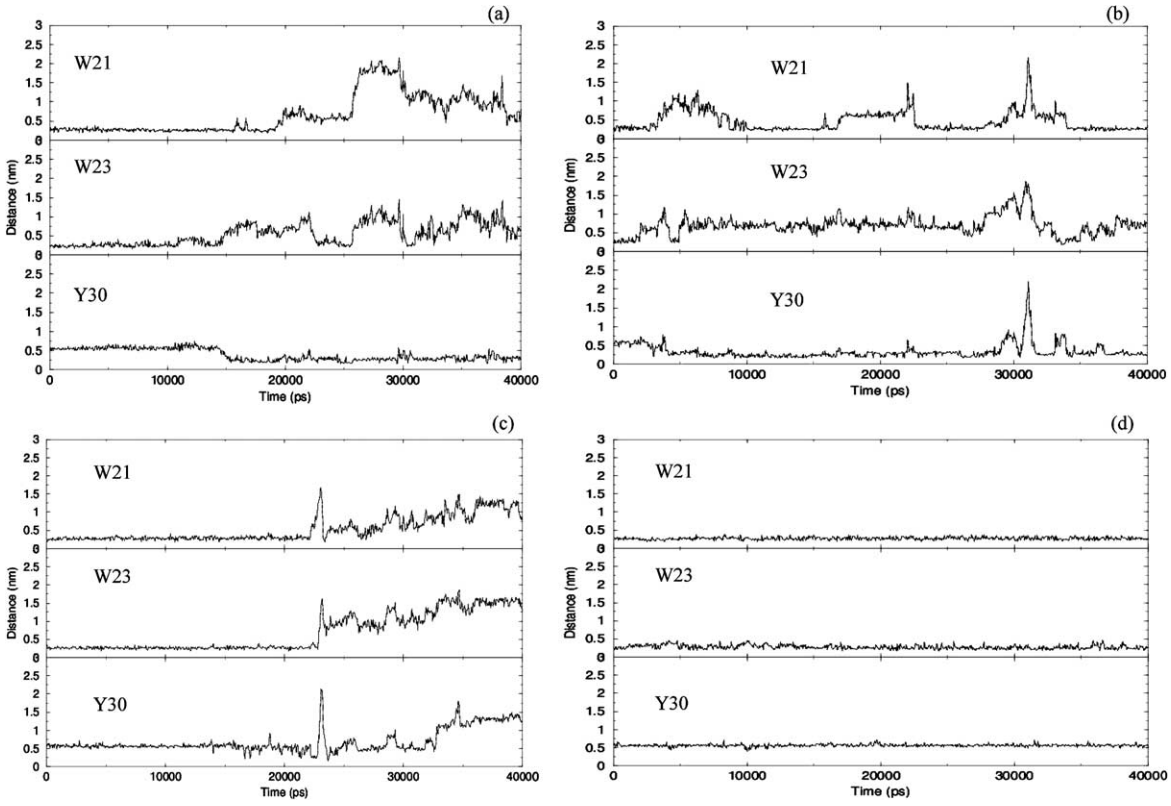

Figure 9. Time evolution of the distance between selected residues on the ligand and W21, W23, or Y30 on the receptor. The time evolution of minimum distance between W21, W23, and Y30 and the $GlcNAc^{I}$ moiety in the disaccharide simulations (a and b), the central residue of the trisaccharide (c) and the central residue of the pentasaccharide (d).

non-reducing end can still be described as having an appreciable degree of conformational freedom and being exposed to solvent, thus providing a favorable entropic and solvation contribution to the complex system.

To summarize, the use of different conditions, different systems, and a statistical analysis of multiple trajectories based on clustering methods provides a physically based description of the recognition between hevein and chitin oligomers, which is consistent with a dynamical picture of the complexes. It is worth noting that even when a stable complex is formed, a high degree of motional freedom is available to the ligand on the rather flat surface of the receptor. Smaller (dimeric or trimeric) subunits of oligosaccharides of different lengths can visit the same conformational subspaces as those taking place in free solution. These flexibility and motional factors should in general be taken into account when analyzing and designing novel molecules with receptor-binding properties, which would be missing by considering a static picture of the complexes.

3. Conclusion

Interference with carbohydrate–protein 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 determinants of the binding between carbohydrates and protein is of paramount importance.

Our simulations suggest distinct possible binding properties for different oligosaccharides, even for those belonging to the same type. These considerations are naturally of semiquantitative 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 full quantitative calculation of the free energy of binding, thus allowing a direct comparison with experimentally determined quantities. However, the use of different conditions, different systems, and the statistical analysis of multiple trajectories based on clustering methods and ED has provided a physically based description of the formation of the complex between hevein and the chitin oligomers.

4. Methods

The starting structures for the simulations were obtained from the NMR analysis of NOESY spectra recorded for the complexes of hevein to a number of chitin oligomers,

as described.[6] 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 Vasudevan and Balaji.[13] Four complexes were built according to the NMR structures: two hevein–disaccharide complexes (the two simulations are labeled disac1 and disac2), one hevein–trisaccharide complex (label trisac) and one hevein–pentasaccharide complex (label penta).

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 counter-ions were added, as water is the highly dielectric and the inclusion of no counter-ions 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, of the complex plus 2967 water molecules in the case of the trisaccharide simulation, and of the complex and 3216 water molecules for the pentasaccharide simulation.

The system was subsequently energy-minimized with the 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[14,15] was used. The simple point charge (SPG)[16] water model was used. The LINCS algorithm[17] was used to constrain all bond lengths. For the water molecules, the SETTLE algorithm[18] was used. A dielectric permittivity, $\varepsilon = 1$, and a time step of 2 fs were used. A twin-range cutoff was used for the calculation of the non-bonded 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. Interactions within the short-range cutoff were updated every time step whereas interactions within the long-range cutoff were updated every 5 time steps together with the pair list. All atoms were given an initial velocity obtained from a Maxwellian distribution at the desired initial temperature. The density of the system was adjusted by performing the first equilibration runs at NPT condition by weak coupling to a bath of constant pressure ($Po = 1$ bar, coupling time $\tau_p = 0.5$ ps).[19] 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 four complexes.

All the MD runs and the analysis of the trajectories were performed using the GROMACS software package.[20] Cluster analysis was performed according to Daura et al.[21]

The subspace overlap (SO) was calculated with the -v, -eig, -v2, and -eig2 options of the g_anaeig of GROMACS. This way, a single number for the overlap between the covariance matrices can be generated. The formulas are:

$$\text{difference} = \text{sqrt}(\text{tr}((\text{sqrt}(M1) - \text{sqrt}(M2))^{\cap 2}))$$

$$\text{normalized overlap} = 1 - \text{difference}/\text{sqrt}(\text{tr}(M1) + \text{tr}(M2))$$

where $M1$ and $M2$ are the two covariance matrices and tr is the trace of a matrix.

The numbers are proportional to the overlap of the square root of the fluctuations. The *normalized overlap* is the most useful number, it is 1 for identical matrices and 0 when the sampled subspaces are orthogonal.

References

1. Lemieux, R. U. *Chem. Soc. Rev.* **1989**, *18*, 347–374.
2. Lasky, L. *Science* **1992**, *258*, 964–969.
3. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.
4. Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rew.* **2002**, *102*, 371–386.
5. Lescar, J.; Loris, R.; Mitchell, E.; Gautier, C.; Cox, V.; Wyns, L.; Perez, S.; Breton, C.; Imberty, A. *J. Biol. Chem.* **2001**, *277*, 6608–6614.
6. 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.; Jimenez-Barbero, J. *Proteins* **2000**, *40*, 218–236.
7. 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.
8. Colombo, G.; Meli, M.; Canada, J.; Asensio, J. L.; Jimenez-Barbero, J. *Carbohydr. Res.* **2004**, *339*, 985–994.
9. Asensio, J. L.; Canada, F. J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J. *Eur. J. Biochem.* **1995**, *230*, 621–633.
10. Asensio, J. L.; Canada, F. J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J. *Glycobiology* **1998**, *8*, 569–577.
11. Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Gen.* **1993**, *17*, 412–425.
12. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem. Int., Ed.* **1999**, *38*, 236–240.
13. Vasudevan, S. V.; Balaji, P. V. *J. Phys. Chem. B* **2001**, *105*, 7033–7041.
14. Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.;

Kruger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
15. van Gunsteren, W. F.; Daura, X.; Mark, A. E. *Encyclopedia Comput. Chem.* **1998**, *2*, 1211–1216.
16. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
17. Hess, B.; Bekker, H.; Fraaije, J. G. E. M.; Berendsen, H. J. C. *J. Comp. Chem.* **1997**, *18*, 1463–1472.
18. Miyamoto, S.; Kollman, P. A. *J. Comp. Chem.* **1992**, *13*, 952–962.
19. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Nola, A. D.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
20. van der Spoel, D.; van Drunen, R.; Berendsen, H. J. C. In *GRoningen MAchine for Chemical Simulations*; van der Spoel, D.; van Drunen, R.; Berendsen, H. J. C., Eds.; Department of Biophysical Chemistry, BIOSON Research Institute: Nijenborgh 4 NL-9717 AG Groningen, 1994.
21. Daura, X.; van Gunsteren, W. F.; Mark, A. E. *Proteins* **1999**, *34*, 269–280.