Carbohydrate Chain of Ganglioside GM₁ as a Ligand: Identification of the Binding Strategies of Three 15 mer Peptides and Their Divergence from the Binding Modes of Growth-Regulatory Galectin-1 and Cholera Toxin

Hans-Christian Siebert,*[a] Karin Born,[b] Sabine André,[a] Martin Frank,[b] Herbert Kaltner,[a] Claus-Wilhelm von der Lieth,[b] Albert J. R. Heck,[c] Jesús Jiménez-Barbero,[d] Jürgen Kopitz,[e] and Hans-Joachim Gabius*[a]

Dedicated to Bernd Aust

Abstract: The branched pentasaccharide chain of ganglioside GM₁ is a prominent cell surface ligand, for example, for cholera toxin or tumor growth-regulatory homodimeric galectins. This activity profile via protein recognition prompted us to examine the binding properties of peptides with this specificity. Our study provides insights into the mechanism of molecular interaction of this thus far unexplored size limit of the protein part. We used three pentadecapeptides in a combined approach of mass spectrometry, NMR spectroscopy and molecular modelling to analyze the ligand binding in solution. Availability of charged and hydrophobic functionalities affected the intramolecular flexibility of the peptides differently. Backfolding led to restrictions in two cases; the flexibility was not reduced significantly by association of the ligand in its energetically privileged conformations. Major contributions to the interaction energy arise from the sialic acid moiety contacting Arg/Lys residues and the N-terminal charge. Considerable involvement of stacking between the monovalent ligand and aromatic rings could not be detected. This carbohydrate binding strategy is similar to how an adenoviral fiber knob targets sialylated glycans. Rational manipulation for an affinity enhancement can now be directed to reduce the flexibility, exploit the potential for stacking and acquire the cross-linking capacity of the natural lectins by peptide attachment to a suitable scaffold.

Keywords: carbohydrates · lectins · NMR spectroscopy · proteins · structure–activity relationship

Introduction

The decoration of the membrane surface with a wide array of biochemical signals is essential for a proper and versatile communication of cells with their environment. Already a brief glance at the nature of suitable signaling systems makes it obvious that hardware for high-density coding is indispensable to produce and present the required large wealth of molecular determinants in the available space. To meet the demands for diversity in code generation with min-

[a] Dr. H.-C. Siebert, Dr. S. André, Dr. H. Kaltner, Prof. H.-J. Gabius
Institut für Physiologische Chemie
Tierärztliche Fakultät, Ludwig-Maximilians-Universität München
Veterinärstrasse 13, 80539 München (Germany)
Fax: (+49) 89-2180-2508
E-mail: hcsiebert@lmu.de
        gabius@lectins.de
        gabius@tiph.vetmed.uni-muenchen.de

[b] Dr. K. Born, Dr. M. Frank, Dr. C.-W. von der Lieth
Zentrale Spektroskopie
Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280
69120 Heidelberg (Germany)

[c] Prof. A. J. R. Heck
Department of Biomolecular Mass Spectrometry
Bijvoet Center for Biomolecular Research and Utrecht Institute
for Pharmaceutical Sciences, Utrecht University
Sorbonnelaan 16, 3584 CA Utrecht (The Netherlands)

[d] Prof. J. Jiménez-Barbero
Centro de Investigaciones Biólogicas, CSIC
Ramiro de Maeztu 9, 28040 Madrid (Spain)

[e] Prof. J. Kopitz
Institut für molekulare Pathologie
Klinikum der Ruprecht-Karls-Universität Heidelberg
Im Neuenheimer Feld 220, 69120 Heidelberg (Germany)

Supporting information for this article is available on the WWW under http://www.chemeurj.org/ or from the author.

imum size, carbohydrates of cellular glycoconjugates are second to no other class of biomolecules.[1] Accordingly, the complex enzymatic machinery for non-random glycan elaboration and the accumulating evidence for ensuing functionality confirm the validity of the concept of the sugar code.[2–5] With endogenous lectins as receptors the sugar-encoded messages are read and translated into biological responses, for example, by glycan cross-linking which then triggers potent signaling.[6–8]

The emerging bioactivity of oligosaccharides has motivated structural studies which defined limited flexibility as distinguishing parameter in contrast to highly flexible oligomers of amino acids and nucleotides.[9–13] The possibility to select preformed, energetically favoured glycan conformations for binding is already well documented for several vertebrate lectins.[13–15] This process effectively reduces the entropic penalty in the thermodynamic equilibrium of a lectin–glycan association compared with the case where a highly flexible ligand needs to be fixed in a single bound-state conformation; this in turn favors binding and the induction of biosignaling on the cellular level mentioned above. It is a fundamental question as to whether the size of a lectin can be reduced while maintaining the target specificity. In order to discern minimal requirements of a peptide to serve as a lectin the examination of the lectin-mimetic peptide interaction with a functionally important glycan is timely and important.

To address this issue, the endogenous lectin galectin-1—with its capacity to act as growth inhibitor on tumor cells—emerges as an attractive model. This lectin's inhibitory effect on the proliferation of neuroblastoma cells, which originate from a childhood tumor of dismal prognosis, is of clinical interest.[16–18] Galectin-1 is a lectin for distinct glycan chains on the cell surface and also has binding sites for nuclear proteins (i.e., Gemin4) and oncogenic H-Ras.[19–22] In the case of the SK-N-MC neuroblastoma cells, the target glycan has been identified.[23] The cellular response is triggered by galectin binding to the pentasaccharide chain of ganglioside GM$_1$ (for glycan structure, see Figure 1a). This interaction accounts for the ensuing switch from tumor cell proliferation to negative growth control.[16–18,23] Thus, a clearcut correlation between a distinct lectin–glycan pair and a medicinally relevant process is given. The ensuing structural analysis of how galectin-1 interacts with this glycan in solution revealed a low-energy conformer with primary contacts to the terminal disaccharide and the branching sialic acid moiety.[24] The amino acids involved in the ligand contact reside at positions widely separated in the sequence, in contrast to receptors which bind ligands through clustering of negative charges. The presence of these negative charges in heparan sulfate and hyaluronic acid had been instrumental for rational peptide design by exploiting the consensus sequences of natural receptors with regular occurrence of basic amino acid residues matching the ligand's charge distribution.[25–28] However, two factors could compensate for the lack of clusters of negative charges in the ligand. In our case: a) allowing the receptor to dock an energetically privi-
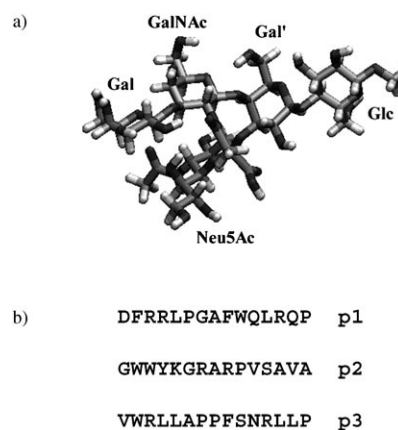

Figure 1. a) Pentasaccharide of ganglioside GM$_1$ and b) amino acid sequences of the three ganglioside GM$_1$-binding 15 mer peptides p1, p2 and p3.

leged conformer and b) use of the readily accessible sialic acid moiety as the docking site. The salt bridge to this moiety and hydrophobic contacts add to the interaction with the terminal galactose residue typical for a galectin. Indeed, in addition to galectin-1, the X-ray structures of complexes of sialoadhesin, cholera toxin and the leukoagglutinin of *Maackia amurensis* with 3′-sialyllactose or pentasaccharide of GM$_1$ illustrate the spectrum how this ligand can be contacted by amino acid side chains. The relative importance of the key interaction site shifts from the sialic acid in sialoadhesin to the two-fingered grip on sialic acid/galactose units in cholera toxin and finally to preferential docking of the galactose unit in the primary site of the plant lectin.[29–32] Thus, not one but several modes of ligand contact are necessary to yield high-affinity binding—a result that has encouraged screening for suitable peptides.

In our context, the screening of a phage library expressing random pentadecamers fused to the pIII minor coat protein on ganglioside GM$_1$ monolayers employing quartz-crystal microbalance technology led to definition of 18 positive clones after five rounds of selection.[33,34] The sequences of the three peptides identified contain both basic and aromatic residues for potential ionic and aromatic stacking interactions (Figure 1b). Interestingly, they share no apparent homology with either galectin-1 or the ganglioside GM$_1$-specific B-chain of the cholera toxin. The critical importance of Arg, Phe and Trp moieties was confirmed by a mutational analysis. It documented loss of sugar binding by substitution of any of these amino acids in the most active peptide p3, whereas a substitution of Pro was tolerated.[34] With these peptides at hand, we addressed the following questions in a stepwise manner, using a previously developed experimental strategy:[35] a) Is the typically high level of intramolecular flexibility of an oligopeptide restrained in these cases with a distinct binding activity? b) Will the ligand exert an influence on the peptide flexibility and/or conformation upon association? c) Where are the key sites of interaction between ligand and receptor peptide? d) Which strategies to gener-

ate affinity are employed by the peptides when comparing our obtained data with the studied cases of lectins?

Results and Discussion

Conformational flexibility of the peptides: The complex structural organization of a lectin protein such as human galectin-1[36] ensures that the key residues for the ligand contact are presented in a rather rigid topology. The more the size of a carbohydrate receptor is reduced, the lower the constraints become to reach this appropriate level of structural organization, until intramolecular flexibility precludes the required tight contacts. We first inspected the sequences of the three peptides depicted in Figure 1b for motifs with a potential for structure building. The presence of charged side chains and terminal positions as well as of aromatic rings and hydrophobic side chains might restrict the flexibility of peptides of this size by intramolecular interactions. As a consequence, a favourable contact site for ganglioside GM_1 might be adopted by the major conformer. Because no discernible consensus sequence is obvious for the three peptides, individual monitoring of each case is required. For example, the hydrophobic cluster of WWY in peptide p2 might impart new properties relative to the other two ligands which were tested. To examine this issue, we performed MD simulations. The results provide predictions which can be used for the comparison to the experimental data. Two sets of force-field parametrization, that is, AMBER and GROMACS, were independently used to separate common from force-field-specific results. In addition, we tested a series of starting conformations (stretched or bent forms) in MD simulations in order to trace commonly obtained structures.

As a clear indication that the particularities of each protocol did not influence the conclusions in these cases, both sets of MD runs yielded comparable results. Changes of intra-protocol parameters such as temperature also did not affect the emerging conclusions. When comparing the behaviour of the three peptides under simulation conditions routinely used, the degree of flexibility of the three peptides appeared to be different. As shown in Figure 2, a bending of the peptide structure is possible for peptide p3 and invariably leads to an energetically favourable structure reached in the last third of the MD runs after starting from a nearly linear chain. To probe its stability, we extended the simulation time 10-fold to 10 ns. Once the bending process took place, no loss of the contact between Arg3/Leu13 was observed within any of the simulations performed. A similar backfolding was detected for peptide p1, whereas peptide p2 maintained a comparatively high degree of flexibility (Figure 3; complete trajectories are available as Supporting Information). Backfolding appeared to involve the orientation of the charged side chains, that is, Arg3 in peptide p3 and Arg4 in peptide p1. They are therefore given in Figure 3; the asterisks denote the potential sites for proton–proton contacts. Of course, special attention was then paid
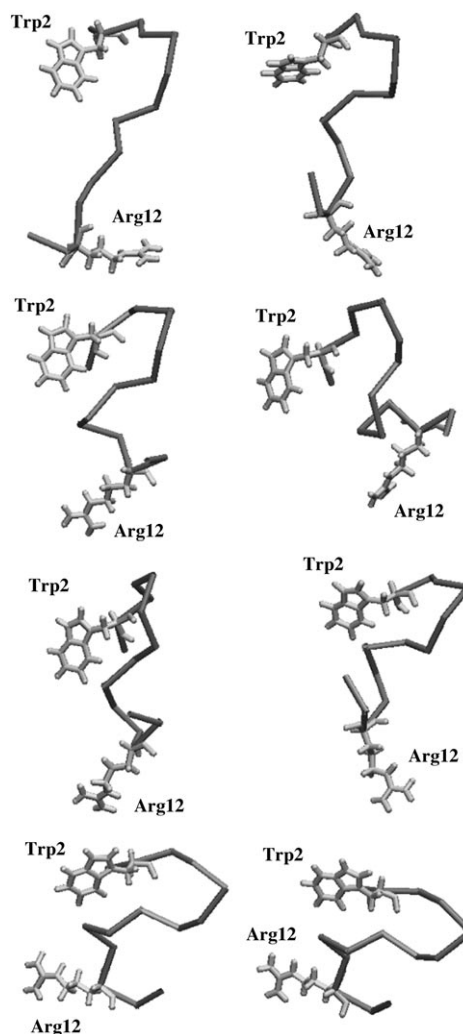

Figure 2. Representative illustration of the flexibility of peptide p3 observed in MD simulations (1000 ps) by using the parametrization of AMBER by presenting the conformation at the start of the simulation (top, left) and snapshots taken in its course. Two side chains in the N- and C-terminal sections (i.e., Trp2, Arg12) are drawn for orientation. Backfolding with establishment of a bend (bottom, right) is energetically privileged.

in the examination of the NOESY spectra in terms of the predicted interresidual proton–proton contacts. Moreover, the calculations did not indicate the presence of intramolecular cluster formation by aromatic stacking interactions in any case. This prediction was also compared with the experimental data (see below).

In order to be able to present the inferred degrees of flexibility as quantitative data, we further processed the results of the MD runs. The structures given in Figures 2 and 3 are based on the values of RMSD compiled in Table 1. In direct comparison between the peptides, it becomes apparent that peptide p2 reaches the comparatively highest value in flexibility despite the occurrence of an aromatic WWY cluster in its sequence. This result was expected due to lack of backfolding, which would stabilize the structure. The calculation of the diffusion coefficient D as a measure of mobility and
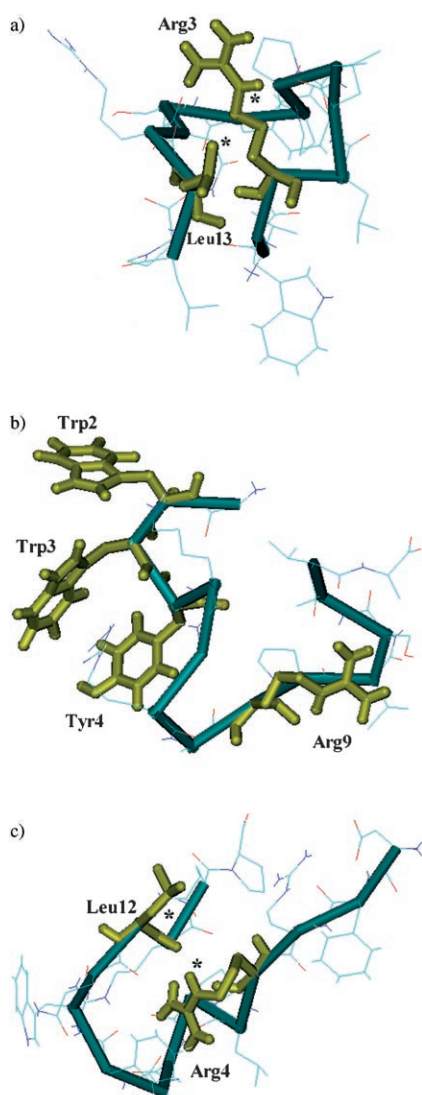
Figure 3. Exemplary snapshots obtained from GROMACS-based MD simulations (1000 ps) of peptides a) p3, b) p2 and c) p1, for which the complete trajectories are available in the Supporting Information. Key proton–proton contacts as revealed by NOESY experiments (see Figure 6a for peptide p3) are highlighted by asterisks in order to visualize the spatial proximity of distinct side chains of peptides a) p3 and c) p1 which are relevant for backfolding (see also Figure 2, bottom), that is Arg3NHε to Leu13CH$_3$ (a) and Arg4NHε to Leu12CH$_3$ (c).

Table 1. Root mean square deviation (RMSD) and diffusion coefficient (D) of the peptides, the pentasaccharide ligand (GM$_1$), and the peptide–ligand complexes

	RMSD [nm]	SD	D [cm^{2}s^{-1}]	SD
p1	0.228	0.0012	4.05×10^{-6}	5.74×10^{-6}
p2	0.415	0.0011	4.35×10^{-6}	2.23×10^{-6}
p3	0.285	0.0006	3.66×10^{-6}	1.72×10^{-6}
GM$_1$	0.238	0.0163	7.61×10^{-6}	0.01×10^{-6}
p1 + GM$_1$	0.234	0.0009	2.17×10^{-6}	1.59×10^{-6}
p2 + GM$_1$	0.266	0.0024	4.04×10^{-6}	2.38×10^{-6}
p3 + GM$_1$ (I)[a]	0.309	0.0017	5.16×10^{-6}	0.79×10^{-6}
p3 + GM$_1$ (II)[a]	0.243	0.0015	4.23×10^{-6}	4.17×10^{-6}

[a] The two binding modes are given in Figure 8.

shape revealed values which agreed with this interpretation. The inherent limits to terms of flexibility of the free oligosaccharide by preferential occupation of few active conformers in the energetic space[24] translated into rather low values of RMSD (Table 1). At this stage, the answer to the first question based on modelling is thus positive. In detail, we could indeed obtain clues which are in favour of restrictions to flexibility generated by the sequence characteristics in two cases. To challenge these conclusions obtained through computational analysis by experimental data we proceeded to perform an NMR analysis of the unbound peptides in solution. In addition to the examination of the spectra for signs of secondary or otherwise ordered structure, decreased flexibility and interresidual contacts, this part of the work was also essential to obtain the signal assignment indispensable for work on peptide–carbohydrate complexes.

The NMR measurements proceeded without problems, because the three peptides were readily soluble in water. Also, no evidence for extensive oligomerization could be concluded from the mostly well-resolved signals, as for examples documented for peptide p3 (Figure 4a). The half-width line broadening of the peaks seen in Figure 4a is in accord with the values predicted by the computer-assisted calculations; its range is also typical for molecules of this size (oligomerization of peptides would have led to an increased line broadening of proton signals). The two-dimen-
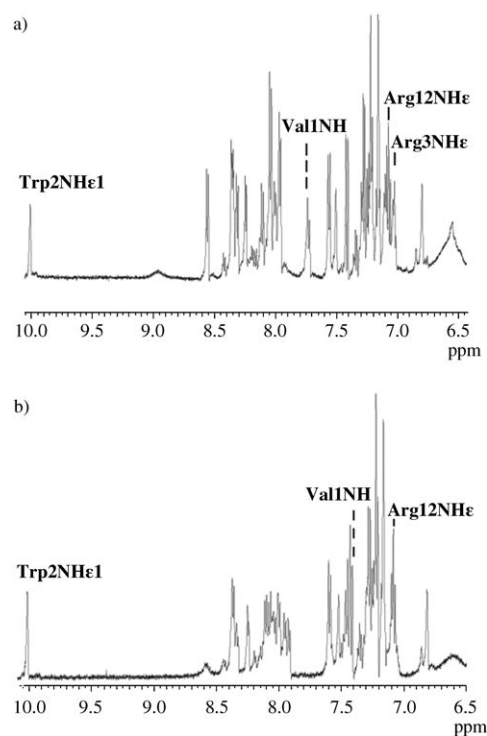

Figure 4. Aromatic/NH part of 1D-NMR spectra of peptide p3 in the absence (a) and presence (b) of ganglioside GM$_1$'s pentasaccharide. Signals were recorded at 750 MHz for experiments performed in 90% H$_2$O/10% D$_2$O at 303 K with the peptide p3 concentration set to 4 mM. Selected cases for signal assignment and ligand-dependent effects are given.
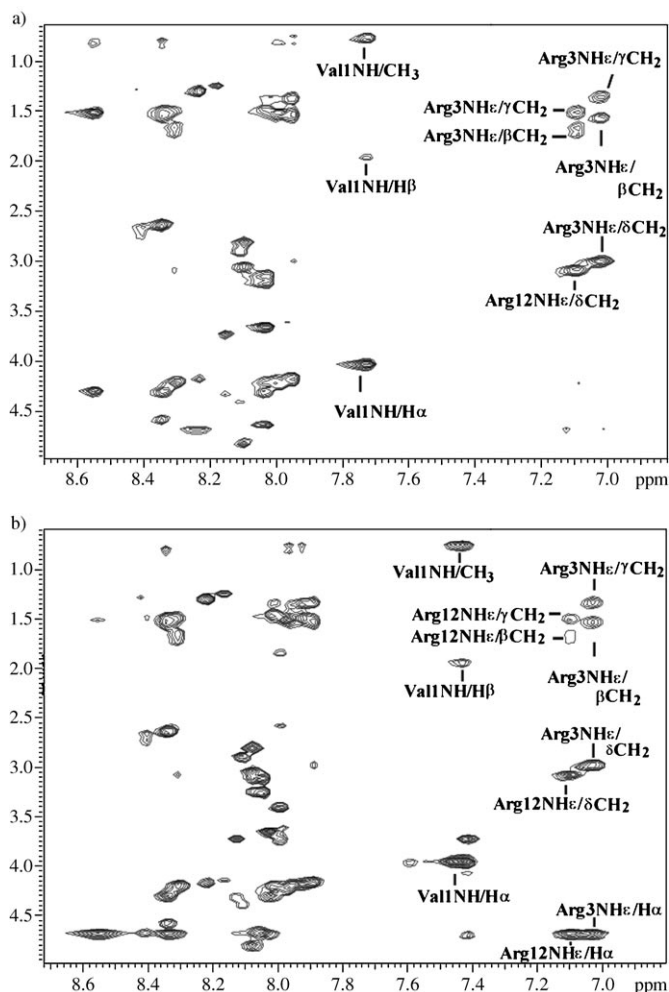
Figure 5. Section of 2D-TOCSY NMR spectra of peptide p3 in the absence (a) and presence (b) of ganglioside GM₁'s pentasaccharide. Signals were recorded at 750 MHz for experiments performed in 90% H₂O/10% D₂O at 303 K with the p3 concentration set to 4 mM. Selected cases for signal assignment and ligand-dependent effects are given.

our NMR-spectroscopical experiments support these findings. A distinct cross-peak was observed for proton Hε of Arg3 and methyl protons of Leu13, which is highlighted in Figure 6a. This experimental result, which indicates spatial vicinity between these two residues at the opposite sites of peptide p3, is also highlighted in Figure 3 by asterisks. Notably, this result is in full agreement with the modelling data. It underlines the physical existence of backfolding. Similarly, NOESY spectra of peptide p1 yielded evidence for a contact between equivalent protons of Arg4 and Leu12, again consistent with the computed bending. On the contrary, no such cross-peaks were present in spectra of peptide p2 (not shown). Thus, bending of the peptide by intramolecular contacts of the amino acids, which are widely separated in the sequence, is experimentally verified in the two cases selected from the modelling studies. The Trp signals of the pep-

sional TOCSY and NOESY spectra support this result (Figures 5a, 6a). The intensities of the cross-peaks of the TOCSY and NOESY spectra of peptide p3 do not show a significant temperature dependence in the range between 293–323 K. Such an effect becomes detectable for peptide oligomers, as their formation is dependent on temperature changes, which was observed previously.[35] Similar data were obtained, when examining peptides p1 and p2 under identical conditions, that is, excluding the fact that the hydrophobic WWY cluster of peptide p2 causes aggregation (see Supporting Information). The quality of the spectra allowed a comparison between the experimental data and the results obtained from the MD runs. At first, we found no evidence for the presence of secondary structure elements. Experimentally, the amide–CαH region of the two-dimensional NOESY spectra is invariably devoid of signals characteristic for secondary structure (see Figure 5a for peptide p3). We next scrutinized the NOESY spectra for interresidual contacts, which were predicted in the MD runs. Indeed,

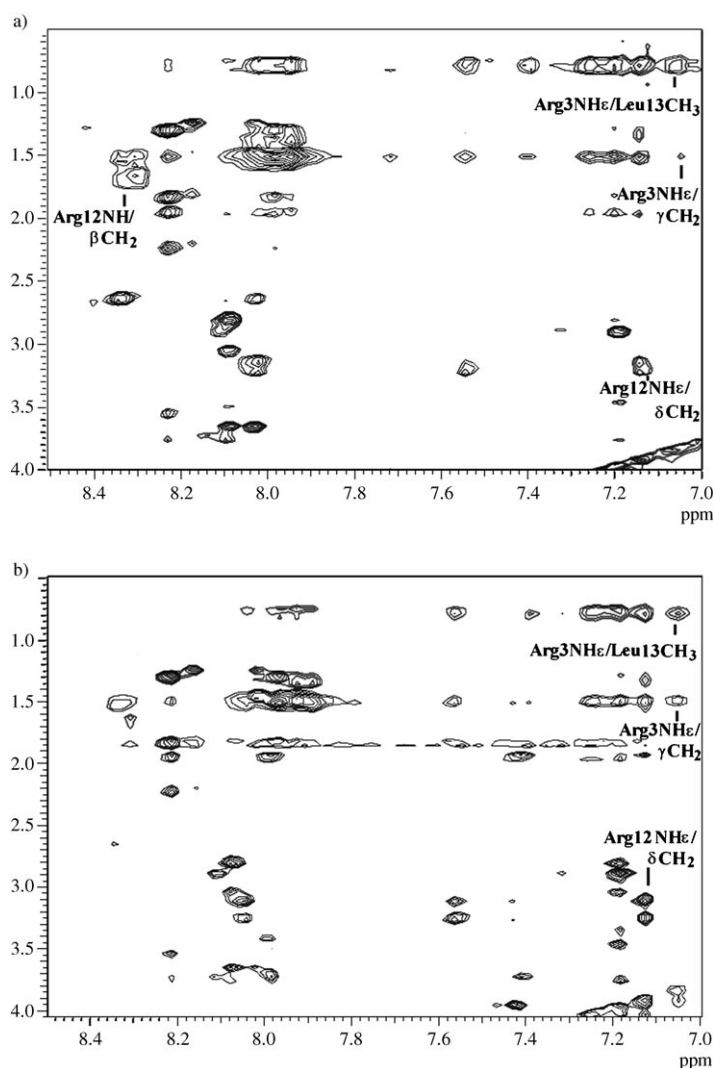

Figure 6. Section of 2D-NOESY NMR spectra of peptide p3 in the absence (a) and presence (b) of ganglioside GM₁'s pentasaccharide. Signals were recorded at 750 MHz for experiments performed in 90% H₂O/10% D₂O at 33°C with the p3 concentration set to 4 mM. Selected cases for signal assignment are given. The interresidual cross-peaks involving Leu13CH₃ might contain a contribution from Leu14CH₃.

tides, especially in the WWY cluster of peptide p2, will deserve special attention to pinpoint any evidence for aromatic stacking, when comparing spectra in the absence and in the presence of the ligand.

Besides the examination of the spectra for indications of interresidual interactions and restrictions to intramolecular flexibility we also completed data analysis by thorough signal assignment, an indispensable prerequisite for pinpointing any ligand-dependent change (see annotations in part a) of Figures 4–6). To explicitly answer the first question when combining results from experiments and computed values: the sequences lead to peptides with a detectable reduction in flexibility in two cases, although no definitive secondary structure is established. Technically, this result underscores the practical validity of the computational data and—due to the internal agreement between the results of the two independent protocols—allowed us to focus on the GROMACS protocol for further calculations during the course of this project. The analysis of the carbohydrate-binding studies show that two of the three peptides studied can acquire a bent low-energy conformation which might be suitable for binding. So far, the binding properties of the free peptides were not determined in direct binding assays but delineated based on inhibition studies with subunit B of cholera toxin and ganglioside GM$_1$ monolayers.[33,34] No phage binding to ganglioside GM$_1$ randomly adsorbed on a polyvinylidene difluoride membrane had been obtained.[33,34] Because the ceramide portion of the clustered ganglioside might therefore be involved in the binding, it was essential for us to document that the peptides interact with the pentasaccharide portion of ganglioside GM$_1$ in the absence of any contact to the sphingolipid's head section. Also, this NMR-spectroscopical analysis would depend on a monovalent interaction without exploitation of the favourable high surface density of a glycolipid monolayer. In order to collect the necessary direct evidence for pentasaccharide binding we performed mass spectrometric measurements, which will be presented in the next paragraph. These experiments are also instrumental to measure the tendency of peptides for self-association independently from the NMR approach.

Complex formation between peptide and pentasaccharide: In order to detect signals of the complex formation with the glycan as proof for the direct binding of the peptides to the pentasaccharide part of the ganglioside only, we enzymatically removed the ceramide portion of the ganglioside. The resulting pentasaccharide is reactive in the molecular recognition, as previously shown for galectin-1.[24] The mass spectra were recorded at the required level of resolution to spot complexes of different stoichiometries. In these experimental series, evidence for a 1:1 stoichiometric complex in each tested case was provided. Figure 7 illustrates occurrence of such a complex in solution at a 5:1 ratio of ligand to peptide for peptide p3. In agreement with the results derived from NMR spectroscopy, no further evidence for peptide aggregation, which would go beyond dimerization, could be detected (Figure 7). Notably, peptide p3 molecule had been reported to bind approximately 2.8 ganglioside GM$_1$ molecules in glycolipid monolayers with their high surface density.[34] Under the given conditions no signals indicative for multivalent interactions were observed with the pentasaccharide as the binding partner. It can be assumed that establishment of this ratio might depend on involvement of secondary interactions in the clustered arrangement extending to the lipid portion. We could also observe no evidence for a distinct behaviour of a certain peptide among the peptides p1–p3. In fact, similar results were obtained for peptides p1 and p2, where peaks at m/z 1440 and 1349 Da, respectively, signify the presence of the (peptide–pentasaccharide)$_1$ complex (not shown). In relation to each other, complex formation was relatively strongest for peptide p3 under these conditions. This evidence implied that the inspection of the NMR spectra of the peptides in the presence of the ligand will allow detection of signal alterations dependent on pentasaccharide binding. Unspecific effects on the signals by a pH alteration after ligand addition were carefully excluded. If such alterations could in fact be recorded, we could immediately take advantage of our detailed signal assignment on the ligand-free peptides. The presence and nature of signal alterations will be discussed in the next three paragraphs.

NMR-Spectroscopical analysis of ligand contact: By comparing the NMR spectra of the peptides in the absence and presence of the sugar compound, we put the given expectation to the experimental test. Indeed, we were able to iden-

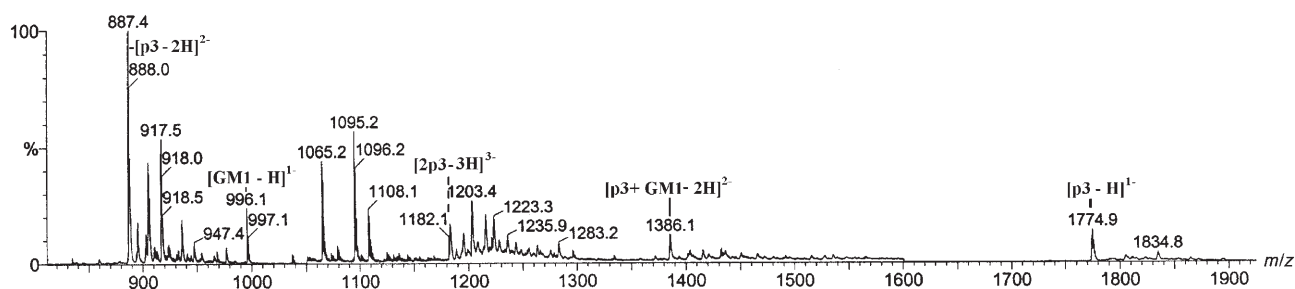Figure 7. Electrospray ionization mass spectrum of peptide p3 in the presence of its ligand, that is, the pentasaccharide chain of ganglioside GM$_1$. Relevant peaks of the free peptide, the ligand as well as of the complex of the peptide with the pentasaccharide are assigned.

tify distinct ligand-dependent signal changes. Globally, no evidence for the acquisition of new long-range connectivities was discernible for peptides p3 and p1. This is in contrast to the behaviour of the heparin-binding peptide (AKKARA)$_6$, where ligand contact converted the conformation from a charged coil to an α-helix, and gains of secondary structure in disordered peptide motifs upon binding for example RNA aptamers.[25,37,38] But the spectra provided clues for sites of ligand contacts to the peptides. Looking at peptide p3, which formed stable complexes under the conditions of mass spectrometry (Figure 7), the signals of the NHε protons from the guanidine groups of Arg3 and Arg12 deserve special attention. They are located at 7.05 and 7.10 ppm, respectively, in the absence of the ligand. Already in the 1D ^{1}H spectrum the Arg3NHε-dependent signal disappeared upon ligand addition (Figure 4a, b). Upon comparison of parts of the two-dimensional TOCSY (Figure 5a, b) and NOESY (Figure 6a, b) spectra, it is obvious that ligand-dependent signal alterations are observable for both Arg residues. Under identical experimental conditions cross-peaks involving the CαH and the NHε protons of both Arg3/12 moieties emerged in the TOCSY spectra, when the ligand is present (Figure 5). Also, the signal of the Val1NH proton is subject to an upfield shift in the presence of the ligand (Figures 4, 5). This shift alteration is likely attributable to interactions of the positively charged section of Val1 at the N-terminal end of the 15 mer peptide to the ligand, an important point for the computational analysis of the complex. In the NOESY spectra the reoccurrence of the cross-peak indicative for backfolding is especially noteworthy, because it excludes a major conformational change after complex formation (Figure 6). Cross-peaks involving the Arg3/12NH backbone and NHε proton signals are highlighted in Figure 6 a, b to emphasize their sensitivity towards ligand binding. This situation for the Arg3/12 residues is different from the TrpNHε1 proton, which reveals no significant indication for a direct ligand contact. In this case, clear preference is given to the basic relative to the aromatic residues.

Similar to the situation for peptide p3, signal alterations also occurred for peptide p1. Comparison of spectra obtained for peptide p1 in the ligand-free and in the ligand-bound states showed signal alterations for the Arg3/13NHε protons (Supporting Information). Notably, an involvement of Arg4NHε protons in ligand contact was not detectable due to its unfavourable position. Two directly neighboring Arg residues tend to adopt orientations, in which their side chains are directed to opposite directions relative to each other, due to repulsion of their positive charges. Therefore, a suitable ligand contact by Arg3 in peptide p1 should be attained at the expense of a less favorable position for Arg4. While using molecular modelling (unrestrained MD runs) together with the NMR signals of the complex, we were able to predict the importance of these two Arg residues. The dominant role of basic amino acids likewise is reflected in the spectrum for peptide p2, with signals of Lys5NH backbone and Arg7/9NHε protons being notably affected (Supporting Information). An immediate question arising in

this context concerns the role of the cluster of aromatic amino acids, fairly common feature of carbohydrate binding by stacking/C-H/π interactions.[10,14,15,39–41] Looking at the half-width line broadening of the Trp2,3NHε1 protons, their slight but significant reduction in the presence of the ligand argues in favour of a diminished inter-Trp interaction (stacking) relative to a likely contact to the ligand (Supporting Information). Compared with peptides p1 and p3, peptide p2 thus shows that a contribution of the aromatic residues to binding is detectable though minor. Based on these results there obviously is a difference in the role of establishing ligand contact between charged and aromatic residues.

For a further characterization of the interaction of the pentasaccharide with the peptides and to assess the affinity of this interaction, we performed systematic titration studies using selected sensor signals to determine the apparent association constant. For peptide p3, $K_A = 116 \pm 26\,\text{M}^{-1}$ was obtained; this value was nearly one magnitude lower for peptide p1 ($15 \pm 3\,\text{M}^{-1}$) and peptide p2 ($11 \pm 3\,\text{M}^{-1}$). The values were obtained by using systematic titrations of 4 mm peptide with an increasing pentasaccharide concentration. For comparison, the IC_{50} values of the inhibition assay were 24, 13, and 1 μm for peptides 1, 2 and 3, respectively.[33,34] The different values closely correspond to the presented data of the analysis of direct binding by NMR titration and of mass spectrometry. This agreement reflects the absence of lipid-dependent interaction. Besides the actual complex formation in solution, and in accord with the mass spectrometric results, our NMR-spectroscopical data on the signal alterations of the peptides have enabled us to take the structural study of the complex formation one step further. Interaction sites have been tracked down on the level of amino acid moieties. With respect to the ligand an increase of half-width line broadening of certain non-overlapping signals of ganglioside GM$_1$'s pentasaccharide, especially the signal of the equatorial proton N3eq of the Neu5Ac residue, could be detected in the presence of peptide p1, p2 or p3 (not shown). This result confirms that sialic acid is a contact partner for the three peptides. The information on the bound-state conformation of the oligosaccharide from the NOESY spectra was, however, rather limited. After all, it enabled the following conclusion: from the set of the three energetically privileged Φ,Ψ angle combinations of the Neu5Acα2–3Gal linkage the conformations 1 and 2, which are in close vicinity in the energy map,[24] are present in the complexes with the peptides p1, p2 and p3 (see next Section). No evidence for a distortion from the low-energy positions of this linkage could be detected.

As key messages from this part of our study the apparent lack of ligand-dependent acquisition of secondary structure and the preferential involvement of Arg (Lys) residues for ligand contact emerged, answering the second and third questions of the introduction. It could thus be argued that positioning of basic amino acids might be the crucial factor irrespective of sequence details. Pertinently, the weak binding properties of phages with the control peptide LGRAGQSYPSFARGL harboring Arg residues at posi-

tions 3 and 13 documented that mere presence of the basic side chains is not sufficient.[33] Also remarkable, a participation of aromatic amino acids in the binding is only measurable for Trp2 and Trp3 in peptide p2. Because mutational substitution of Phe/Trp besides Arg (but not Pro) had been described to abolish the binding properties of peptide p3,[34] the role of these aromatic residues should not generally be assigned to primary ligand contact in monovalent binding. Importantly, charged positions at termini, that is, Val1 in peptide p3, can also matter. At this stage, the binding was experimentally ascertained and contact points were pinpointed by signal alterations. Because we had documented satisfying agreement between computational and experimental data in the first part of our study, we next explored the possibility to visualize ligand contact in greater detail by computational studies. Our experimental data served as a basis for computer-assisted calculations to infer the topology of the complexes in accord with the experimental analysis.

Computational calculations of topology of ligand accommodation: In the first step of these calculations the topology of the complex was modelled in agreement with the NMR-spectroscopical data in order to define the starting conditions for each MD run of the complexes. No restrictions were imposed on the freedom for conformational changes in order to probe the extent of stability of the complexes. Also, the Autodock program (see Experimental Section) was applied to systematically search for other stability-conferring complex configurations. As a final result, the parameters obtained for the complex, which are consistent with our experimental results, implied a remarkable degree of stability.

Peptide p3 showed different results when evaluating the calculations. To match the data of the binding behaviour of peptide p3 with contributions of both Arg3 and Arg12 (see signal shifts in Figure 4a, b) not one but two separate modes of ligand contact are apparently operative (Figure 8). They are readily distinguished by the extent of involvement of the C-terminal peptide part, binding mode II being devoid of respective contacts in this sequence part. The binding modes nevertheless are mutually exclusive, that is, the carbohydrate can only be accommodated at one position at a time. The ensuing formation of the 1:1 stoichiometry has been observed experimentally in mass spectrometry (Figure 7). To match the NMR data, which reveals the role of residue Arg12, the first binding mode (Figure 8a) should thus play a major role. Our input from NMR spectroscopy made this case exceptional among the three peptides, because the experimental data on peptides p1 and p2 fit a single binding mode (Figure 9). Because the analysis of the computational complex formation was in accordance with the experimental results, it further enabled to suggest establishment of inter- and intraresidual hydrogen bonds (Figure 9a–c). As discussed in the previous chapter, the Φ,Ψ angle combination of the Neu5Acα2–3Gal linkage can adopt values of about −180/−40° denoted as conformation 1 and values of about −70/20° denoted as conformation 2.[24] Involvement of certain parts of the Neu5Acα2–3Gal section of the carbohy-
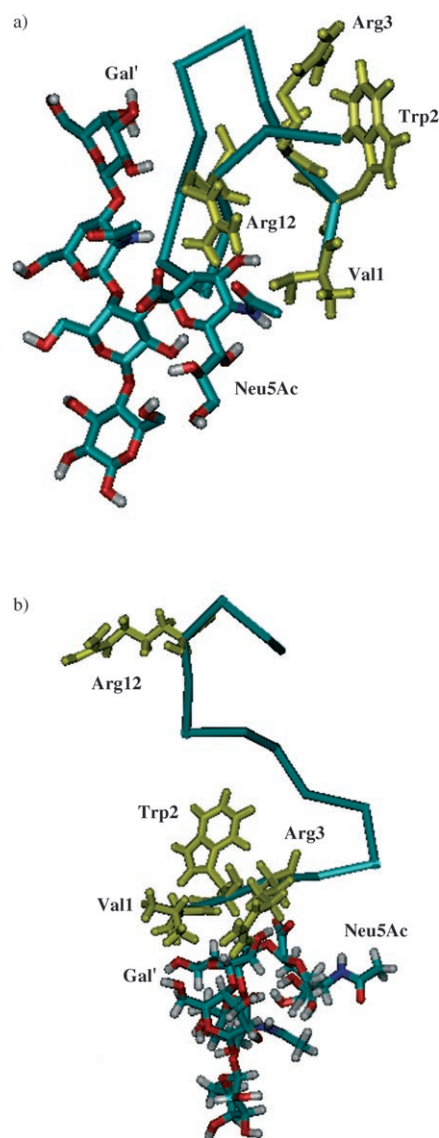

Figure 8. Representative snapshots of peptide-ligand complexes obtained from GROMACS-based MD simulations (1000 ps) of peptide p3 in the presence of ganglioside GM₁'s pentasaccharide. The two configurations illustrate interaction modes of the two binding positions defined experimentally (I, II). Distinct constituents of peptide and ligand are depicted for orientation, for example drawing attention to the exclusive involvement of the N-terminal peptide part for interaction in binding mode II. Complete trajectories are available as Supporting Information.

drate chain allows the occurrence of conformations 1 and 2 of this linkage in the complexes (in rare cases also conformation 3; for the terminology of the conformations, see ref. [24]). After these binding modes were assigned and alternatives by docking analysis with energetically minimized conformations were excluded, we could proceed to further address the issue on intramolecular flexibility in the complex. Respective calculations used the protocols applied for the ligand-free peptides and the pentasaccharide. In full agreement with the experimental data the backfolded peptides p1 and p3 maintained their degree of flexibility after the contact to the ligand (Table 1). Peptide p2 showed a reduced
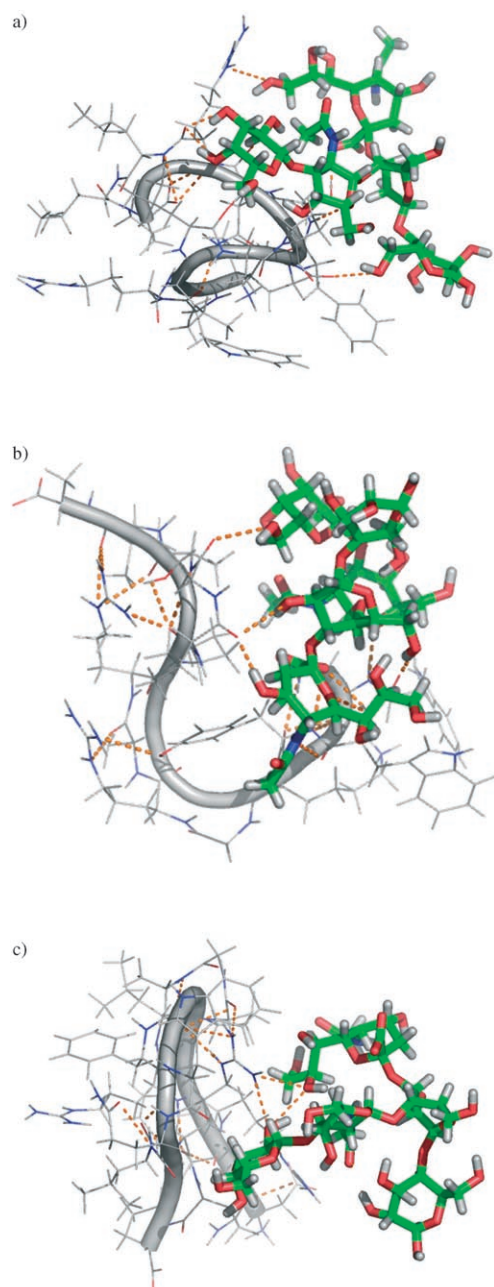
a)

b)

c)

Figure 9. Representative snapshots from a GROMACS-based MD simulation (1000 ps) of peptides p3 in the first binding mode (a), p2 (b) and p1 (c) in the presence of ganglioside GM$_1$'s pentasaccharide taken in the course of the respective MD runs. The peptide backbones with the amino acid side chains are depicted for orientation, and the positions of inter- and intramolecular hydrogen bonds are indicated by dashed lines as follows: Neu5AcC9O to Arg13NHε, Neu5AcC1O1,2 to Ser10OH, GalC3OH and GalC4OH to Arg12O, GalNAcC5O to Ser10NH, Gal′C2OH to Asn11O, GlcC6OH to Pro8O, Asn11NHδ to Trp2O, Arg12NH and Leu13NH to Pro15O1 and O2 (a); Neu5AcC1O1,2 to Gly1NH, Trp3NH and Tyr4NH, Neu5AcNH to Trp3O, Neu5AcC4OH to Val11O, GalC6OH to Ser12C=O, GalNAcC5O to Trp2NH, Gal′C2OH to Val11O, Gal′C6OH to Trp2O, Tyr4OH to Gly1NH, Tyr4NH to Gly1O, Pro10O to Ala13NH, Gly1O to Trp3NH, Ala8O to Arg7NεH (b); Neu5-AcC8O to Arg13NHη1, GalC2O to Arg13NHη1, Arg13NHη2 to Pro6O and Ala8O, Arg13NεH to Ala8O, Arg13NH to Phe9O, Ala8NH to Leu5O, Leu12NH to Phe9O and Gln11Oε, Arg4NH to Arg13O, Gln14Oε to Phe2NH and Arg3NH (c). Water-mediated hydrogen bonds were also delineated (4–6 contacts per peptide) but not included for clarity.

tein, structural alterations beyond the immediate contact site could be imagined. Indeed, already lactose as ligand led to packing of galectin-1, measured as significant decrease in gyration radius by small angle neutron scattering.[42] Conversely, these sequence additions can also modify the ligand specificity, as shown for the siglec CD22 and a 25 mer peptide mimic of its binding site.[43]

With this detailed view on the intermolecular interaction, it was possible to dissect the contributions of the individual monosaccharides of ganglioside GM$_1$'s pentasaccharide and the amino acids of the three peptides. Due to its suggested prime importance the computed interaction energies between Arg12 of peptide p3 and the glycan constituents of the ligand are graphically presented in Figure 10a as a representative example. As can be seen in this figure, the sialic acid of the pentasaccharide is the determinant that interacts most strongly with Arg12, thus leading to this ligand-dependent signal change. After starting the simulation in the energetically most favourable position it is notable that the contact is not lost during the complete MD run. We deliberately show in Figure 10b that the two molecules continue to stay in contact throughout the MD run despite intramolecular fluctuations in both molecules. Reflected in the calculated average diffusion coefficient D and in the complete trajectories open for the reader for inspection, the shape of peptide p3 in binding position I is not markedly altered (Figure 10b). The calculated structures for each pair of interaction between the building blocks of peptide and ligand not only complement the arguments for defining contact points on the level of sequence of the binding partners but also provide a ranking in the energetic driving force for complex formation. As can be seen for peptide p3 in Tables 2 and 3, the residues Val1, Arg3 and Arg12 are the major players in the interaction process between peptide p3 and ganglioside GM$_1$'s pentasaccharide. The prominent role of Val1 had been underscored by respective signal shifts shown in Figures 4 and 5. Also in full agreement with the NMR-derived results no indications for a substantial contribution of the ar-

flexibility. The formation of interresidual bonds in the presence of the ligand can contribute to explain the rigidification of the structure of peptide p2 seen in the RMSD values in Table 1. The changes of the diffusion coefficient D illustrated specific shape alterations, an indication for positional specificity. The diffusion coefficient D is correlated in a reciprocal manner with the radius of a sphere-like particle. Regarding peptide p3, the value of the diffusion coefficient $D = 3.66 \times 10^{-6}$ cm^{2}s^{-1} increased significantly to a value of 5.16×10^{-6} cm^{2}s^{-1} in the first binding mode (Table 1), reflecting structural packing through contact with the ligand. Evidently, this is not the case for peptide p2 (Table 1). When theoretically embedding such a reactive peptide into a pro-
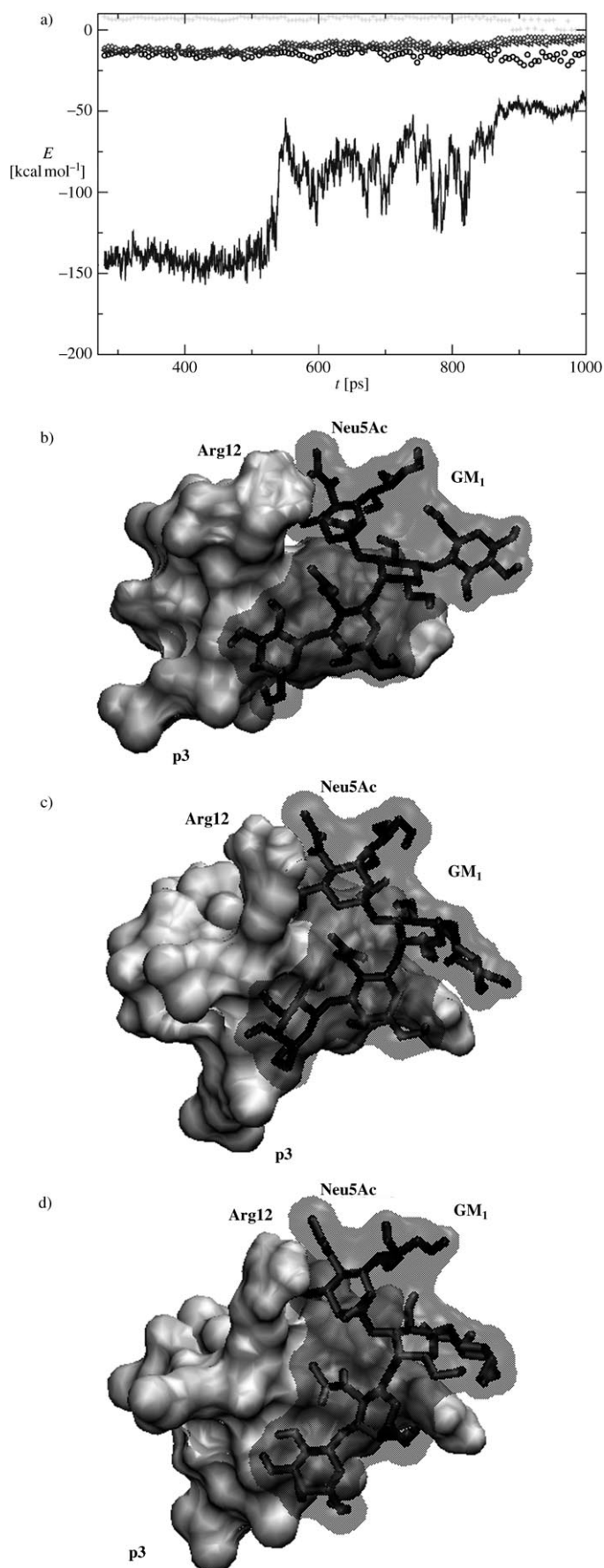
Figure 10. Graphical computation of the contributions of the individual monosaccharide units of ganglioside GM₁'s pentasaccharide to the interaction energy with Arg12 of peptide p3 in binding position I illustrated in Figure 8a based on a GROMACS-directed MD simulation for 1000 ps (a). Three representative snapshots from this simulation of peptide p3 in the presence of ganglioside GM₁'s pentasaccharide taken after equal intervals during the course of the MD run which was started at binding position I, as depicted in Figure 8a (b–d). The two residues involved in the dominant interaction (Arg12, Neu5Ac) are highlighted for orientation. The complete trajectory is available as Supporting Information. ○ Gal, ◇ Gal′, △ GalNAc, + Glc, —— Neu5Ac.

omatic ring of Trp2 of peptide p3 are given (Tables 2, 3). The N-terminal part, as seen in Figure 8, is the contact site for the second binding position, although Val1 plays no dominant role in this complex (Table 3). So far, we have dealt with Coulomb/van der Waals energy terms in these considerations, which exhibit force-field-specific characteristics. To access the ΔG values, we used a different computational approach. The described preference of binding mode I is hereby substantiated, yielding a dramatic difference of about 20 kcal mol⁻¹ in favour of this position. With respect to the polar contributions the interaction energies of Trp2, Trp3 and Tyr4 of peptide p2 with ganglioside GM₁'s pentasaccharide are delivered by their backbone protons only (Figure 9b, Table 4). This result is in line with the observation that the ring NH protons of Trp2 and Trp3 are only slightly altered after ligand addition (see above). The same holds true for Trp10 of peptide p1 (Table 5).

The expected positional effects are clearly present, for example the strong negative effect of the C-terminal residue in peptide p3 (Table 2) or the marked (and expected) difference in energetic contributions for Arg3/Arg4 in peptide p1 (Table 5). The availability of such data for binding of this ligand to galectin-1 made a comparison to the endogenous lectin possible. In that case, binding contributions were mainly distributed between the terminal disaccharide and the sialic acid with Arg 48, His52, Trp68 and Glu71 as key contact points.[24] These side chains are in fact frontrunners in listing natural logarithms of the sugar interface propensity values for the common amino acids.[44] Intriguingly, Arg48 of the lectin interacted more strongly with the GalNAc moiety than with the sialic acid. To underline the requirement for positional effects for the peptides' specificity we added a control experiment by introducing limited structural scrambling in the ligand. Practically, we performed the same set of calculations with a pentasaccharide which we artificially linearized—with the expectation to lower the interaction energy, if the characteristic ligand presentation of ganglioside GM₁ is essential. Squeezing the sialic acid moiety between the GalNAc and Gal′ residues to establish Galβ1–3GalNAcβ1–4Neu5Acα2–3Gal′β1–4Glc indeed caused a drop in binding affinity by about 75% for peptide p3 (position I) to −16.3 kcal mol⁻¹. The same drastic reduction was calculated for peptide p1 with ΔG values of −52.1 kcal mol⁻¹ (GM₁) and −14.7 kcal mol⁻¹ (pseudo GM₁) by using the LIE

Table 2. Contributions of the individual constituents of the pentasaccharide chain of ganglioside GM$_1$ to the interaction energy with peptide p3 and its building blocks.[a]

	Glc	Gal′	GalNAc	Gal	Neu5Ac	Σ
Val1	7.8	−9.2	−8.8	−10.9	−43.9	−65.0
Trp2	−0.2	0.1	0.1	0.3	0.5	0.8
Arg3	5.3	−4.6	−2.8	−15.4	−33.1	−50.6
Leu4	0.0	0.0	−0.2	0.2	0.1	0.1
Leu5	0.0	0.0	−0.1	0.4	0.2	0.5
Ala6	0.0	0.0	0.1	−0.4	−0.2	−0.5
Pro7	−0.1	−0.1	−2.1	−3.1	−0.3	−5.7
Pro8	−0.5	0.2	−1.0	1.0	1.5	1.2
Phe9	−6.5	−0.6	−2.2	0.0	0.1	−9.2
Ser10	−2.7	−4.0	−6.6	−1.6	−5.4	−20.3
Asn11	−0.3	−0.4	−7.2	0.6	−3.4	−10.7
Arg12	6.5	−9.6	−11.1	−14.8	−101.0	−130.0
Leu13	0.1	−0.3	−0.5	−1.2	−2.1	−4.0
Leu14	0.1	−0.1	−0.1	−0.3	−1.4	−1.8
Pro15	−6.6	7.4	5.1	14.6	51.4	71.9
Σ	2.9	−21.2	−37.4	−30.6	−137.0	−223.3

[a] The analysis of the interactions was performed for the binding mode I given in Figure 8; the GROMACS-based Coulomb/van der Waals energy terms are given in kcal mol^{-1}; for the notation of the monosaccharide units in the branched pentasaccharide chain see Figure 1a.

Table 3. Contributions of the individual constituents of the pentasaccharide chain of ganglioside GM$_1$ to the interaction energy with peptide p3 and its building blocks.[a]

	Glc	Gal′	GalNAc	Gal	Neu5Ac	Σ
Val1	−1.3	−5.6	−0.1	0.0	5.3	−1.7
Trp2	−0.3	−6.2	−3.7	−0.7	−8.4	−19.3
Arg3	11.7	−19.1	−13.0	−9.2	−98.5	−128.1
Leu4	0.2	−1.3	−0.7	−0.3	−10.1	−12.2
Leu5	0.1	−0.3	−0.3	−0.2	−5.3	−6.0
Ala6	0.0	0.0	0.0	0.0	0.2	0.2
Pro7	0.1	−0.1	−0.1	0.0	−0.7	−0.8
Pro8	0.0	0.0	0.0	0.0	−0.3	−0.3
Phe9	0.0	0.0	0.0	0.0	−0.2	−0.2
Ser10	0.0	0.0	0.0	0.0	0.0	0.0
Asn11	0.0	0.0	0.0	0.0	0.0	0.0
Arg12	0.0	0.0	0.0	0.0	0.0	0.0
Leu13	0.0	0.0	0.0	0.0	0.0	0.0
Leu14	0.0	0.0	0.0	0.0	0.0	0.0
Pro15	0.0	0.0	0.0	0.0	0.0	0.0
Σ	10.5	−32.6	−17.9	−10.4	−118.0	−168.4

[a] The analysis of the interactions was performed for the binding mode II given in Figure 8; the GROMACS-based Coulomb/van der Waals energy terms are given in kcal mol^{-1}; for the notation of the monosaccharides in the branched pentasaccharide chain, see Figure 1a.

Table 4. Contributions of the individual constituents of the pentasaccharide chain of ganglioside GM$_1$ to the interaction energy with peptide p2 and its building blocks.[a]

	Glc	Gal′	GalNAc	Gal	Neu5Ac	Σ
Gly1	8.4	−12.6	−17.1	−15.3	−13.7	−50.3
Trp2	−0.5	−4.7	−5.7	−0.4	−3.9	−15.2
Trp3	0.2	−1.1	−0.7	−0.2	−11.7	−13.5
Tyr4	0.1	−1.1	−0.8	−0.1	−20.0	−21.9
Lys5	5.1	−4.5	−8.8	−8.7	−50.3	−67.2
Gly6	0.0	0.0	−0.1	0.0	−1.0	−1.1
Arg7	2.9	−3.9	−5.8	−2.7	−40.8	−50.3
Ala8	−0.1	0.1	0.1	0.1	0.9	1.1
Arg9	5.2	−5.9	−5.8	−7.0	−38.4	−51.9
Pro10	0.2	−1.1	−1.6	−0.6	−6.6	−9.7
Val11	−0.6	−1.2	−0.6	−1.0	−5.5	−8.9
Ser12	0.1	−0.7	−0.4	−2.6	−1.0	−4.6
Ala13	0.0	−0.2	−0.9	−1.3	−0.7	−3.1
Val14	0.0	0.0	−0.3	−0.5	0.2	−0.6
Ala15	−2.4	2.9	3.7	5.6	29.4	39.2
Σ	18.6	−34.0	−44.8	−34.7	−163.1	−258.0

[a] The GROMACS-based Coulomb/van der Waals energy terms are given in kcal mol^{-1}; for the notation of the monosaccharide units in the branched pentasaccharide chain, see Figure 1a.

method. In contrast, peptide p2 tolerated the structural change without intramolecular bending.

In conclusion, experimental and computational analysis clearly converge to reveal a dominant role of ionic interactions with additional positional contributions from the vicinity of the sialic acid moiety, that is, the terminal Gal/GalNAc residues. The binding site topology shared at least by the peptides p1 and p3 appears to be a shallow groove on the surface, while at the same time maintaining both the flexibility and the contact area in the ligand's presence (Figure 9a, c, Figure 10b). Peptide p2 can also adopt a groove-like structure in the presence of the pentasaccharide (Figure 9b). We can thus proceed to answer question d) posed in the Introduction.

Comparison of binding modes between the peptides and other lectins: One common feature of the three peptides and pentasaccharide/sialyllactose-specific lectins is the ability to interact with low-energy conformations of the carbohydrate. No deviation from the few, mutually convertible structures readily adopted in solution occurs upon binding. Ultimately, case-specific characteristics of each binding site account for the actual selection of the fitting conformer. In summary, it turned out that all three energetically privileged orientations of the α2–3 linkage in sialyllactose and the pentasaccharide are active in the recognition process. The way each conformer is distinguished from biologically inert structures revealed the intricate strategies of ligand specificity governed by sequence and shape.[24] For example, the sialic acid moiety of the sialyllactose part of the pentasaccharide chain reaches its optimal com-

Table 5. Contributions of the individual constituents of the pentasaccharide chain of ganglioside GM₁ to the interaction energy with peptide p1 and its building blocks.[a]

	Glc	Gal′	GalNAc	Gal	Neu5Ac	Σ
Asp1	−2.5	5.4	2.0	0.9	8.3	14.1
Phe2	−0.3	0.1	0.2	0.0	−0.2	−0.2
Arg3	10.2	−14.7	−10.7	−10.0	−70.8	−96.0
Arg4	4.8	−3.7	−4.6	−9.0	0.0	−12.5
Leu5	−0.1	0.1	0.0	0.3	0.4	0.7
Pro6	−0.1	0.0	−0.2	0.2	−1.9	−2.0
Gly7	0.0	0.0	0.1	0.0	0.0	0.1
Ala8	−0.2	0.1	0.0	0.3	1.5	1.7
Phe9	−0.3	0.2	0.2	0.3	2.5	2.9
Trp10	−0.1	−0.1	−0.3	−4.1	0.2	−4.4
Gln11	−0.3	0.2	−0.3	−3.3	2.4	−1.3
Leu12	−0.5	−0.3	−2.8	−3.3	−1.2	−8.1
Arg13	7.3	−9.0	−7.8	−23.0	−89.2	−121.7
Gln14	−5.0	−0.1	−2.3	−0.4	0.8	−7.0
Pro15	−6.2	6.1	7.4	7.8	2.4	17.5
Σ	6.7	−15.7	−19.1	−43.3	−144.8	−216.2

[a] GROMACS-based Coulomb/van der Waals energy terms are given in kcal mol⁻¹; for notation of the monosaccharide units in the branched pentasaccharide chain, see Figure 1a.

plementarity to the binding site of cholera toxin in a Φ/Ψ angle combination of −172/−26° (conformation 1) by seven direct or solvent-mediated hydrogen bonds especially to Glu11/His13 and also stacking to Tyr12, assisted by three direct or solvent-mediated interresidual hydrogen bonds within the carbohydrate chain.[29,30] Looking at the 15mer peptides, it is intriguing to see that a series of enthalpic gains to compensate the entropic penalty already arising from freezing the fluctuation between conformers can indeed be inferred. Keeping the entropic cost in the thermodynamic balance low is a vital factor for the peptides achieved by backfolding to a suitable structure in the two cases. Their sensitivity to sequence alterations is clearly highlighted by delineating the consequences of artificial linearization of the pentasaccharide. Likewise, a phage clone presenting the pentadecapeptide LGRAGQSYPSFARGL retaining presence of basic and aromatic residues only weakly bound to ganglioside GM₁ monolayer, as already pointed out above as an argument for operative positional specificity.[33]

The presence of aromatic rings ideal for hydrophobic interactions and of basic side chains suitable to make ionic interactions has initially raised the notion for a concerted action to explain how the peptides reach their target specificity. Moreover, the involvement of a main chain amide proton in Asn137 of the influenza virus hemagglutinin instead of a positively charged side chain as contact for the carboxylate added a further option in the cases of peptides p1 and p3.[45] Admittedly, inspection of the binding site topology of this viral protein and of galectin-1, cholera toxin and *Maackia amurensis* agglutinin will rather likely not come up with the properties of a real role model for the peptides, when recalling the inherent involvement of more than one sequence stretch in the architecture of the different binding sites.[24,29,30,32,45] Similar to cholera toxin with its "2-fingered grip",[29] the tetanus toxin Hc fragment, too, accommodates the terminus and branch of its ganglioside ligand in a narrow groove for the galactose section and a shallow pocket for the sialic acid.[46] The notion that peptides might mimic the design of a pocket with a ligand part contacting the peptides from both sides, as seen in wheat germ agglutinin, polyoma virus capsid protein and the N-terminal lectin domain of *Vibrio cholerae* neuraminidase with its total of ten hydrogen bonds to the Neu5Ac moiety,[47–50] can also not be reconciled with our data. Thus, the quest for natural receptor proteins with a selection strategy comparable to that of the peptides is directed to cases with rather limited number of contacts.

With this stipulation set, the following cases could be singled out for further scrutiny in order to answer question d) from the Introduction: the I-type lectin sialoadhesin, the pertussis toxin and the knob domain of the fiber protein of adenovirus serotype 37.[31,51,52] Two of these three cases do not match the prominent role of Arg/Lys residues in the peptides. The salt bridge of the guanidino group of Arg97 with the ligand's carboxylate is complemented by indispensable van der Waals contacts of the sialic acid with Trp2/Trp106 in sialoadhesin,[31] and Ser104 rather than Arg125 is the contact partner for the carboxylate in the pertussis toxin.[51] In contrast, the importance of Lys345 of the viral fiber knob with its K_D value of 5 mM for α2,3-sialyllactose[52] is in fact in agreement with the properties of the peptides. Notably, the hydrophobic contact area of this protein is confined to the sialic acid's N-acetyl group, which offers another option for affinity enhancements in both cases. The apparently enormous potential of a contiguous non-polar sequence stretch for conferring affinity to a short peptide had been deduced with the selectin-based heptamer YYWIGIR, although the binding of the full-length E- and P-selectins to the sialyl Lewisˣ determinant is primarily electrostatic in nature, in line with its binding kinetics required for a molecular braking mechanism.[53,54] The small plant lectins hevein and pseudohevein also illustrate the extent of stacking to generate affinity.[55] To exploit these beneficial effects of stacking for further affinity enhancements in our cases, hydrogen-bond formation to a galactose residue coupled with stacking to the sialic acid by this positioning of Tyr45 in the *Maackia amurensis* agglutinin,[32] for example, can inspire testable concepts.

Conclusion

In order to probe for the mechanism of binding of a natural complex carbohydrate ligand at the lower size limit of the protein we examined the interaction of three ganglioside

GM$_1$-specific peptides with this pentasaccharide. The presence of basic, aromatic and hydrogen-bond-donating/accepting functionalities in the sequences of the peptides requires a theoretical consideration of the various modes of ligand contacts; this is also confirmed by the different experimental docking processes of sialylated ligands to sugar receptors. The concerted strategy by using NMR spectroscopy, mass spectrometry, and molecular modelling revealed binding interactions of the peptides to the pentasaccharide part of the ganglioside. We could detect evidence for restricted flexibility in two cases suitable for an interaction with the pentasaccharide in its low-energy conformations. Our combined approach integrating experimental and computational data enabled us to infer major energetic contributions of the interplay between Arg/Lys residues and the sialic acid part to overall binding. The aromatic rings did not show a significant contribution in the calculations or in the set of signal alterations. But they might be crucial for the affinity enhancement in ganglioside monolayers with a stoichiometry of 2.8:1 for peptide p3.[34] In fact, their presence had been revealed to be essential by mutational analysis in the same experimental setup.[33,34] Because the amino-terminal charge of peptides p3 (position I) and p2 is also a factor to be reckoned with during ligand interaction, the question will next have to be answered whether affinity benefits from restrictions to intramolecular flexibility introduced by peptide cyclization which removes this positive (and also the C-terminal negative) charge. In the calculations the basic amino acids surpass the aromatic rings in their energetic contributions. The potential to use stacking for affinity enhancement thus appears to be rather untouched. This was similarly seen in a natural model case with a comparable preference given to a lysine residue, that is, the knob domain of an adenoviral fiber protein.[52] As a further means to raise the affinity for natural ligand displays such as ganglioside-bearing rafts/glycosynapses mimicking the cross-linking activity of the natural effectors can be considered. The peptides will then have to be attached to suitable scaffolds, as exploited for generating high-affinity ligands for tissue galectins.[56–63] Rational manipulation of receptor structure using the given lead compounds and of receptor-site positioning by clustering will thus be instrumental to gain affinity increases for this class of carbohydrate-binding peptides.

Experimental Section

Materials: Peptides p1, p2 and p3 were chemically synthesized in a fully automated solid-phase synthesizer (MultiSyntech), purified to homogeneity on a Kromasil KR100–10C18 column (250×300 mm) and quality controlled by mass spectrometry.

Preparation of the pentasaccharide chain of ganglioside GM$_1$: Ganglioside GM$_1$ (10 mg; Alexis, Läufelingen, Switzerland) were mixed with 3% (w/v) sodium cholate in chloroform/methanol (2:1, v/v) and dried under a stream of nitrogen. The dried substance was resuspended in 0.1 M sodium acetate buffer (10 mL, pH 5.0) and incubated at 37°C for 48 h in the presence of five units of ceramide glycanase from *Macrobdella decora* (Calbiochem, Bad Soden, Germany). The reaction was stopped by the addition of chloroform/methanol (50 mL, 2:1, v/v) with vigorous mixing. After phase separation, the aqueous phase was obtained, frozen and lyophilized. The product was redissolved in water (2 mL) and applied to prewashed C18 reversed phase SepPak cartridges for removal of free ceramide and residual ganglioside (Waters, Milford, MA). The pentasaccharide was eluted with water (5 mL).[24,64] Quantification was performed by measurement of the amount of bound sialic acid. The yield was approximately 90%.

Mass spectrometry: Aqueous solutions of mixtures of the three peptides p1, p2, and p3 (20 μM) with the pentasaccharide chain of ganglioside GM$_1$ (100 μM) were subjected to analysis by electrospray ionization (ESI) mass spectrometry. Measurements were performed on a quadrupole time-of-flight instrument (Micromass, Manchester, U.K.) operating in negative ion mode, equipped with a "Z-spray" nanoelectrospray source by using in-house pulled and gold-coated needles. Typical conditions were as follows: needle voltage, 1300–1600 V; cone voltage, 25 V; quadrupole pressure, 5×10^{-6} mbar; and TOF analyzer pressure, 3×10^{-7} mbar. Mass spectra were averaged typically over 100 scans. The standard mass scanned was 100–4000 Thomson.[35,65,66]

NMR-Spectroscopical experiments: ^{1}H NMR spectra were recorded in 90% H$_2$O/10% D$_2$O with Bruker AMX 500, AMX 600 and Varian Unity 750 MHz spectrometers. Spectra from the two-dimensional experiments were acquired at 303 K (test measurements in order to define the optimal temperature for the experiments also at 293, 313 and 323 K) with 4 mM solutions of peptides p1, p2, and p3 in the absence or in the presence of an equimolar concentration of the pentasaccharide, an optimal receptor/ligand ratio defined also in another 15 mer peptide case.[35] Total correlation spectroscopy (TOCSY) and nuclear Overhauser and exchange spectroscopy (NOESY) experiments were performed in the phase-sensitive mode using time-proportional phase incrementation method for quadrature detection in F1. Typically, a data matrix of 512×2048 points was chosen to digitize a spectral width of 15 ppm. Eighty scans per increment were used with a relaxation delay of 1 s. Prior to Fourier transformation zero filling was performed in F1 direction to expand the data to 1024×2048 points. Baseline correction was applied in both dimensions. The TOCSY spectra were recorded with mixing times of 10 and 50 ms, respectively, by use of a MLEV-17 isotropic mixing scheme. The NOESY experiments were performed with mixing times of 50, 100, and 200 ms to spot any arising spin diffusion. Titration experiments at a pH value of 5.5 and a temperature of 303 K for estimating K_D values were performed using the signals of Arg3NHε and Arg13NHε for peptide p1, of Arg7NHε and Arg9NHε for peptide p2 and of Val1NH and Arg3NHε for peptide p3 as sensors.[24,67]

Computational calculations based on AMBER: The AMBER (assisted model building with energy refinement) force field as implemented in the program DISCOVER 2.98 (Accelrys Inc., San Diego, CA) was used for molecular dynamics (MD) simulations of peptide p3 in its free form and in complex with ganglioside GM$_1$'s pentasaccharide with explicit inclusion of water molecules. The conformational behaviour of this peptide was calculated using the program DISCOVER 2.98 (Accelrys Inc., San Diego, CA), atomic charge assignment of INSIGHT II and the parametrization of the AMBER 1.6 force field.

Computational calculations based on GROMACS: Initial structures of the three peptides were constructed as a linear chain using the molecular graphics system pymol[68] and subsequently relaxed using the all-atom force field (ffgmx2) implemented in GROMACS (Groningen Machine for Chemical Simulations).[69,70] The individual steps of the applied procedure were as follows: after initial relaxation of the artificially linearized structures by the conjugate gradients algorithm method the peptides were positioned in the center of a box and surrounded by explicit water molecules under periodic boundary conditions. The complete system (peptide and water molecules) was subsequently equilibrated using 1000 steps of the conjugate gradients minimization procedure, followed by a 1000 ps-long MD simulation at 300 K and a pressure of 1 bar. The following cut-off distances for the non-bonded interactions were set: 1.2 nm for Coulomb forces and 1.5 nm for van der Waals interactions. Regarding the carbohydrate ligand its main low-energy conformation in solution was adopted, as described previously.[67] The partial charge for each atom

was calculated using Gaussian03 thereby applying the density functional theory with the Becke3LYP hybrid functional.[71] The topology of ganglioside GM$_1$'s pentasaccharide as required for GROMACS processing was generated using the ProDRG server.[72] Hereby, special care was exercised to avoid ring distortions. Partial charges especially for the sialic acid (Neu5Ac) residue of ganglioside GM$_1$'s pentasaccharide were added manually. Subsequently, the same procedure as used for relaxation of the peptide structures was also applied on ganglioside GM$_1$'s pentasaccharide. Visual inspection of complete trajectories of MD runs, which are available as Supporting Information or from the authors (http://www.dkfz-heidelberg.de/spec/publications/suppl_mat/siebert2005gm1/), necessitates downloading of the freely available CHIME program (for convenience we provide the respective link under the web address given above).

Topology and dynamics of peptide–carbohydrate interaction: The docking procedure implemented in the program Autodock[73] was applied in order to obtain a reasonable starting geometry for the peptide–pentasaccharide complexes. The MD simulations provided the orientation for the selection of suitable structures so that sets of three-dimensional structures of the peptides could be taken from the final period of the MD simulations of the free receptors. Low-energy conformations of ganglioside GM$_1$'s pentasaccharide are described in detail elsewhere.[24] The complex having the lowest energy was chosen after careful examination of the essential parameters. The computational handling of the complexes of the three peptides with ganglioside GM$_1$'s pentasaccharide followed protocols as described for the free peptides, starting with an equilibration period with a length of 200 ps in which all atoms of the peptide–pentasaccharide complexes were restrained to their initial positions, then the actual MD run of a total period of 1000 ps followed.

Calculation of the interaction energy and shape of the complex: The analysis of interaction energy between individual constituents of peptides and ganglioside GM$_1$'s pentasaccharide chain was performed with the g_energy module,[70] which is a part of the GROMACS package. The free energy values of ligand binding were calculated by using the linear interaction energy (LIE) method adapted to process data from molecular dynamics simulations,[74,75] using the g_lie module of GROMACS as essential tool. In order to estimate relative compactness of the complex the diffusion coefficient D was calculated for the free peptides, the ligand and the corresponding complexes using the g_msd module of the GROMACS package according to the Einstein relationship.[76] Relative changes of this parameter are indicative of shape alterations caused by ligand association. In addition, the intramolecular flexibility was calculated as root mean square deviation (RMSD) values by using the g_rms module of the GROMACS package and the peptide backbone as reference.

[1] R. A. Laine, in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman & Hall, Weinheim, London, **1997**, pp. 1–14.

[2] I. Brockhausen, H. Schachter, in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman & Hall, Weinheim, London, **1997**, pp. 79–113.

[3] G. Reuter, H.-J. Gabius, *Cell. Mol. Life Sci.* **1999**, *55*, 368.

[4] H.-J. Gabius, *Anat. Histol. Embryol.* **2001**, *30*, 3.

[5] R. G. Spiro, *Glycobiology* **2002**, *12*, 43R.

[6] H.-J. Gabius, *Eur. J. Biochem.* **1997**, *243*, 543.

[7] H. Kaltner, B. Stierstorfer, *Acta Anat.* **1998**, *161*, 162.

[8] C. F. Brewer, *Biochim. Biophys. Acta* **2002**, *1572*, 255.

[9] B. J. Hardy, *J. Mol. Struct.* **1997**, *395-396*, 187.

[10] H.-J. Gabius, *Pharm. Res.* **1998**, *15*, 23.

[11] C. A. Bush, M. Martin-Pastor, A. Imberty, *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 269.

[12] A. Imberty, S. Pérez, *Chem. Rev.* **2000**, *100*, 4567.

[13] H.-J. Gabius, H.-C. Siebert, S. André, J. Jiménez-Barbero, H. Rüdiger, *ChemBioChem* **2004**, *5*, 740.

[14] H.-C. Siebert, M. Gilleron, H. Kaltner, C.-W. von der Lieth, T. Kozár, N. V. Bovin, E. Y. Korchagina, J. F. G. Vliegenthart, H.-J. Gabius, *Biochem. Biophys. Res. Commun.* **1996**, *219*, 205.

[15] D. Solís, J. Jiménez-Barbero, H. Kaltner, A. Romero, H.-C. Siebert, C.-W. von der Lieth, H.-J. Gabius, *Cells Tissues Organs* **2001**, *168*, 5.

[16] J. Kopitz, C. von Reitzenstein, S. André, H. Kaltner, J. Uhl, V. Ehemann, M. Cantz, H.-J. Gabius, *J. Biol. Chem.* **2001**, *276*, 35917.

[17] J. Kopitz, S. André, C. von Reitzenstein, K. Versluis, H. Kaltner, R. J. Pieters, K. Wasano, I. Kuwabara, F.-T. Liu, M. Cantz, A. J. R. Heck, H.-J. Gabius, *Oncogene* **2003**, *22*, 6277.

[18] S. André, H. Kaltner, M. Lensch, R. Russwurm, H.-C. Siebert, E. Tajkhorshid, A. J. R. Heck, M. von Knebel-Doeberitz, H.-J. Gabius, J. Kopitz, *Int. J. Cancer* **2005**, *114*, 46.

[19] S. André, S. Kojima, N. Yamazaki, C. Fink, H. Kaltner, K. Kayser, H.-J. Gabius, *J. Cancer Res. Clin. Oncol.* **1999**, *125*, 461.

[20] F.-T. Liu, R. J. Patterson, J. L. Wang, *Biochim. Biophys. Acta* **2002**, *1572*, 263.

[21] G. Rappl, H. Abken, J. M. Muche, W. Sterry, W. Tilgen, S. André, H. Kaltner, S. Ugurel, H.-J. Gabius, U. Reinhold, *Leukemia* **2002**, *16*, 840.

[22] B. Rotblat, H. Niv, S. André, H. Kaltner, H.-J. Gabius, Y. Kloog, *Cancer Res.* **2004**, *64*, 3112.

[23] J. Kopitz, C. von Reitzenstein, M. Burchert, M. Cantz, H.-J. Gabius, *J. Biol. Chem.* **1998**, *273*, 11205.

[24] H.-C. Siebert, S. André, S.-Y. Lu, M. Frank, H. Kaltner, J. A. van Kuik, E. Y. Korchagina, N. V. Bovin, E. Tajkhorshid, R. Kaptein, J. F. G. Vliegenthart, C.-W. von der Lieth, J. Jiménez-Barbero, H.-J. Gabius, *Biochemistry* **2003**, *42*, 14762.

[25] A. Verrecchio, M. W. Germann, B. P. Schick, P. Kung, T. Twardowski, J. D. San Antonio, *J. Biol. Chem.* **2000**, *275*, 7701.

[26] B. P. Schick, J. F. Gradowski, J. D. San Antonio, J. Martinez, *Thromb. Haemostasis* **2001**, *85*, 482.

[27] X.-M. Xu, Y. Chen, J. Chen, S. Yang, F. Gao, C. B. Underhill, K. Creswell, L. Zhang, *Cancer Res.* **2003**, *63*, 5685.

[28] N. Liu, X.-M. Xu, J. Chen, L. Wang, S. Yang, C. B. Underhill, L. Zhang, *Int. J. Cancer* **2004**, *109*, 49.

[29] E. A. Merrit, S. Sarfaty, F. van den Akker, C. L'hoir, J. A. Martial, W. G. J. Hol, *Protein Sci.* **1994**, *3*, 166.

[30] E. A. Merrit, S. Sarfaty, M. G. Jobling, T. Chang, R. K. Holmes, T. R. Hirst, W. G. J. Hol, *Protein Sci.* **1997**, *6*, 1516.

[31] A. P. May, R. C. Robinson, M. Vinson, P. R. Crocker, E. Y. Jones, *Mol. Cell* **1998**, *1*, 719.

[32] A. Imberty, C. Gautier, J. Lescar, S. Pérez, L. Wyns, R. Loris, *J. Biol. Chem.* **2000**, *275*, 17541.

[33] T. Matsubara, D. Ishikawa, T. Taki, Y. Okahata, T. Sato, *FEBS Lett.* **1999**, *456*, 253.

[34] T. Matsubara, *Trends Glycosci. Glycotechnol.* **2001**, *13*, 557.

[35] H.-C. Siebert, S.-Y. Lü, M. Frank, J. Kramer, R. Wechselberger, J. Joosten, S. André, K. Rittenhouse-Olson, R. Roy, C.-W. von der Lieth, R. Kaptein, J. F. G. Vliegenthart, A. J. R. Heck, H.-J. Gabius, *Biochemistry* **2002**, *41*, 9707.

[36] M. F. López-Lucendo, D. Solís, S. André, J. Hirabayashi, K.-i. Kasai, H. Kaltner, H.-J. Gabius, A. Romero, *J. Mol. Biol.* **2004**, *343*, 957.

[37] P. E. Wright, H. J. Dyson, *J. Mol. Biol.* **1999**, *293*, 321.

[38] N. Leulliot, G. Varani, *Biochemistry* **2001**, *40*, 7947.

[39] F. A. Quiocho, *Pure Appl. Chem.* **1989**, *61*, 1293.

[40] H.-J. Gabius, *Naturwissenschaften* **2000**, *87*, 108.

[41] A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, *Biochem. J.* **2004**, *382*, 769.

[42] L. He, S. André, H.-C. Siebert, H. Helmholz, B. Niemeyer, H.-J. Gabius, *Biophys. J.* **2003**, *85*, 511.

[43] C. E. von Seggern, R. J. Cotter, *J. Mass Spectrom.* **2004**, *39*, 736.

[44] C. Taroni, S. Jones, J. M. Thornton, *Protein Eng.* **2000**, *13*, 89.

[45] W. Weis, J. H. Brown, S. Cusack, J. C. Paulson, J. J. Skehel, D. C. Wiley, *Nature* **1988**, *333*, 426.

[46] C. Fotinou, P. Emsley, I. Black, H. Ando, H. Ishida, M. Kiso, K. A. Sinha, N. F. Fairweather, N. W. Isaacs, *J. Biol. Chem.* **2001**, *276*, 32 274.

[47] C. S. Wright, *J. Mol. Biol.* **1990**, *215*, 635.

[48] C. S. Wright, J. Jaeger, *J. Mol. Biol.* **1993**, *232*, 620.

[49] T. Stehle, S. C. Harrison, *EMBO J.* **1997**, *16*, 5139.

[50] I. Moustafa, H. Connaris, M. Taylor, V. Zaitsev, J. C. Wilson, M. J. Kiefel, M. von Itzstein, G. Taylor, *J. Biol. Chem.* **2004**, *279*, 40 819.

[51] P. E. Stein, A. Boodhoo, G. D. Armstrong, L. D. Heerze, S. A. Cockle, M. H. Klein, R. J. Read, *Nat. Struct. Biol.* **1994**, *1*, 591.

[52] W. P. Burmeister, D. Guilligay, S. Cusack, G. Wadell, N. Arnberg, *J. Virol.* **2004**, *78*, 7727.

[53] J. B. Briggs, R. A. Larsen, R. B. Harris, K. V. S. Sekar, B. A. Macher, *Glycobiology* **1996**, *6*, 831.

[54] W. S. Somers, J. Tang, G. D. Shaw, R. T. Camphausen, *Cell* **2000**, *103*, 467.

[55] H.-C. Siebert, C.-W. von der Lieth, R. Kaptein, J. J. Beintema, K. Dijkstra, N. Nuland, U. M. S. Soedjanaatmadja, A. Rice, J. F. G. Vliegenthart, C. S. Wright, H.-J. Gabius, *Proteins* **1997**, *28*, 268.

[56] S. André, P. J. Cejas Ortega, M. Alamino Perez, R. Roy, H.-J. Gabius, *Glycobiology* **1999**, *9*, 1253.

[57] S. André, B. Frisch, H. Kaltner, D. L. Desouza, F. Schuber, H.-J. Gabius, *Pharm. Res.* **2000**, *17*, 985.

[58] N. Yamazaki, S. Kojima, N. V. Bovin, S. André, S. Gabius, H.-J. Gabius, *Adv. Drug Delivery Rev.* **2000**, *43*, 225.

[59] S. André, R. J. Pieters, I. Vrasidas, H. Kaltner, I. Kuwabara, F.-T. Liu, R. M. J. Liskamp, H.-J. Gabius, *ChemBioChem* **2001**, *2*, 822.

[60] S. André, B. Liu, H.-J. Gabius, R. Roy, *Org. Biomol. Chem.* **2003**, *1*, 3909.

[61] R. Roy, *Trends Glycosci. Glycotechnol.* **2003**, *15*, 291.

[62] S. André, H. Kaltner, I. Furuike, S.-I. Nishimura, H.-J. Gabius, *Bioconjugate Chem.* **2004**, *15*, 87.

[63] S. André, C. Unverzagt, S. Kojima, M. Frank, J. Seifert, C. Fink, K. Kayser, C.-W. von der Lieth, H.-J. Gabius, *Eur. J. Biochem.* **2004**, *271*, 118.

[64] A. E. Bäcker, J. Holgersson, B. E. Samuelsson, H. Karlsson, *Glycobiology* **1998**, *8*, 533.

[65] T. J. D. Jorgensen, P. Roepstorff, A. J. R. Heck, *Anal. Chem.* **1998**, *70*, 4427.

[66] W. D. van Dongen, A. J. R. Heck, *Analyst* **2000**, *125*, 583.

[67] H.-C. Siebert, S. André, J. L. Asensio, F. J. Cañada, X. Dong, J. F. Espinosa, M. Frank, M. Gilleron, H. Kaltner, T. Kozár, N. V. Bovin, C.-W. von der Lieth, J. F. G. Vliegenthart, J. Jiménez-Barbero, H.-J. Gabius, *ChemBioChem* **2000**, *1*, 181.

[68] W. L. DeLano, The PyMOL Molecular Graphics System (www.pymol.org) **2002**.

[69] H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.* **1995**, *91*, 43.

[70] D. van der Spoel, A. R. van Buuren, E. Apol, P. F. Meulenhoff, D. P. Tielemann, A. L. T. Sijbers, B. Hess, K. A. Feenstra, E. Lindahl, R. van Drunen, H. J. C. Berendsen, Gromacs User Manual version 3.1. (www.gromacs.org), **2001**.

[71] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, L. A. Curtiss, *J. Chem. Phys.* **1989**, *90*, 5622.

[72] A. W. Schuettelkopf, D. van Aalten, *Acta Crystallogr. Sect. D* **2004**, *60*, 1355.

[73] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639.

[74] J. Åqvist, C. Medina, J. E. Samuelsson, *Protein Eng.* **1994**, *7*, 385.

[75] J. Åqvist, V. B. Luzhkov, B. O. Brandsdal, *Acc. Chem. Res.* **2002**, *35*, 358.

[76] M. J. Allen, D. J. Tildesley, Computer Simulations of Liquids. Oxford Science Publications, Oxford, UK, **1987**.