CHAPTER 3

Insights into Carbohydrate Recognition by 3D Structure Determination of Protein–Carbohydrate Complexes Using NMR

MARIO SCHUBERT[a]

[a]Department of Molecular Biology, University of Salzburg, 5020, Salzburg, Austria
*E-mail: mario.schubert@sbg.ac.at

3.1   Introduction

The recognition of carbohydrates by proteins is crucial to many interactions occurring at cell surfaces with roles in recruiting and activating immune cells,[1,2] deciding about cancer progression and metastasis,[3,4] the attachment of pathogens to host cells[5,6] and manipulating host immune pathways.[7] Glycans function as a kind of bar code decorating the cell surface or specific proteins to provide information about the type and status of the cell, which can be read by carbohydrate-recognising proteins and receptors. In addition, glycosylation at asparagines (*N*-glycosylation) is part of a control mechanism that monitors proper protein folding in the endoplasmic reticulum of eukaryotes.[8,9] The *N*-glycans change their composition during folding, and

receptors that recognise these changes decide whether or not the protein is exported to the Golgi, goes to another cycle of folding or is degraded.

In general, a very important function of glycans is to provide a densely packed code—the glycocode—consisting of glycoepitopes that are read by carbohydrate-recognising protein receptors. There are estimates of 3000–7000 functional glycoepitopes in mammals,[10] of which only a fraction have been characterised. To understand the glycocode it is crucial to understand the molecular basis of their recognition by the protein receptors.

Three-dimensional structures of proteins in complex with a carbohydrate catch the proteins in action and provide atomic details of how specificity and affinity is achieved. Structure determination of proteins and protein-complexes is mainly achieved by X-ray crystallography and NMR spectroscopy, whereby crystallography contributes approximately 90% and NMR spectroscopy roughly 10% of all protein structures deposited in the Protein Data Bank (PDB). Determining 3D structures of protein–carbohydrate complexes is challenging for both methods as discussed below.

Carbohydrate-binding protein domains are often extracellular and typically contain disulfide bonds, making it difficult to obtain homogeneous and properly folded proteins in sufficient amounts, however, it is even more challenging to crystallise or isotopically label them. Oligosaccharides are often flexible and can prevent crystallisation and therefore structure determination by X-ray diffraction. NMR spectroscopy is not currently perceived as a method for 3D structure determination of protein–carbohydrate interactions, therefore this review aims to demonstrate the competitive advantages that should be considered to help unravel the enormous challenges of cracking the glycocode and its many crucial functions.

There are a number of excellent reviews on the contribution of NMR spectroscopy to the understanding of protein–carbohydrate interactions that focus on ligand detected methods, mapping the binding site(s) on the protein and the use of docking to build models.[11–15] Here we focus on three-dimensional structures of protein–carbohydrate complexes determined with NMR spectroscopy.

Whereas the number of protein–DNA and protein–RNA complex structures determined with NMR spectroscopy is steadily increasing, only a few protein–carbohydrate complex structures have so far been determined (Figure 3.1). There are many reasons that protein–carbohydrate structures lag behind including technological limitations on the NMR side and obtaining sufficient amounts of active carbohydrate-binding proteins and carbohydrate ligands. Often the glycoepitopes that are recognised are oligosaccharides that are only accessible by sophisticated chemical synthesis.

In the case of protein–DNA and protein–RNA complex structures NMR spectroscopy is actually quite competitive contributing 15% and 35%, respectively, to the complex structures with a size of <40 kDa. These contributions are much higher than the average contribution of NMR to protein structures of ~10% to the PDB. I hypothesise that, if NMR spectroscopy would be applied with the same vigor and persistency to protein–carbohydrate
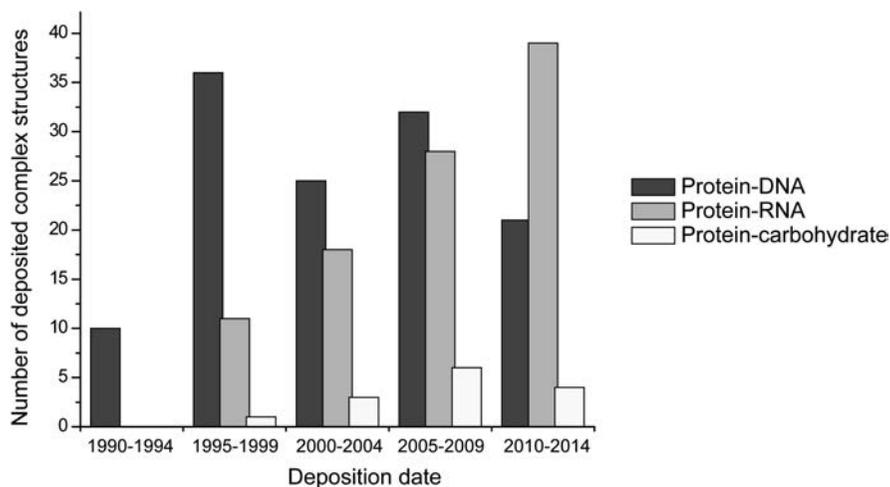
Figure 3.1 Statistics of three-dimensional NMR structures of protein–DNA, protein–RNA and protein–carbohydrate complexes deposited in the Protein Data Bank with a size of <40 kDa correlated with the year of deposition.

interactions, NMR could make a similarly strong contribution. This review serves to illustrate the huge potential of applying NMR spectroscopy to explore the world of protein–carbohydrate recognition. The protein–carbohydrate complex structures that have been determined by NMR spectroscopy can serve as a proof of principle confirming that all the methodology is available for the 3D structure determination of protein–carbohydrate complexes. Advantages, problems and pitfalls of this method are discussed and illustrated by examples.

3.2 Overview of Protein–Carbohydrate Structures Determined by NMR

As already mentioned, 14 biomolecular NMR structures that contain a carbohydrate ligand bound to a protein have been deposited in the Protein Data Bank (PDB). These structures are summarised in Table 3.1. The affinities, expressed as dissociation constants (K_d values), range from nanomolar to millimolar, covering the typical range of affinities generally found for protein–carbohydrate interactions. I chose two examples for illustrating the methodology: the fungal lectin *Coprinopsis cinerea* lectin 2 (CCL2) interacting with the trisaccharide GlcNAcβ1,4[Fucα1,3] GlcNAc with a K_d = 1 μM to represent a tight interacting protein–carbohydrate complex[16] and the complex between rhamnose binding lectin (RBL) domain of Latrophilin-1 and ʟ-rhamnose with a K_d = 1.8 mM,[17] representing a weak interaction. A representative structure of both ensembles is shown in Figure 3.2.

Table 3.1   Protein–carbohydrate complex structures determined by NMR spectroscopy and deposited in the PDB database.

Year of deposition	PDB code	Protein name	K_d	Exchange regime	Number of intermol NOEs	Reference
1997	1ACZ	Glucoamylase[a]	6 μM and 28 μM[o]	Fast	22	55
2001	1IIY	Cyanovirin-N[b]	139 nM	Slow	37	56
2004	1T0W	Hev32[c]	130 μM	Fast	9	57
2004	1WCO	Nisin[d]	50 nM[p]	—[p]	36	58
2005	1ZNT	AcAMP2F18Pff/Y20Pff[e]	2 mM	Fast	13	59
2006	2ERM	FGF-1[f]	n.d.[q,r]	Fast[r]	6	44
2008	2K46	Malectin[g]	26 μM	Intermediate/slow[r]	31	60
2008	2JXA	Latrophilin[h]	1.8 mM	Fast	16	17
2010	2K8R	HFGF[i]	16 μM	Fast	6	61
2010	2KR2	Malectin[j]	50 μM	Intermediate/slow[r]	30	62
2011	2YHH	Microvirin[k]	48 μM	Slow	26	37
2012	2LIQ	CCL2[l]	1 μM	Slow	82	16
2012	2LL4	TgMIC4-A5[m]	110 μM	Intermediate	7	63
2013	2LVZ	ECP[n]	20 μM	Fast	(8)[r,s]	64

[a]Complex with β-cyclodextrin.
[b]Complex with Manα1,2Manα.
[c]Complex with GlcNAcβ1,4 GlcNAc.
[d]Complex with lipid II containing GlcNAcα1,4MurNAc.
[e]Complex with GlcNAcβ1,4 GlcNAcβ1,4 GlcNAc.
[f]Complex with GlcNα1,4IdoAβ1,4GlcNα1,4IdoAβ1,4GlcNα1,4IdoAβ-iPr.
[g]Complex with Glcα1,3Glc.
[h]Complex with Rha.
[i]Complex with inositol hexaphosphate.
[j]Complex with Glcα1,4Glcα.
[k]Complex with Manα1,2Man.
[l]Complex with Galβ1,4[Fucα1,3]GlcNAcβ.
[m]Complex with Galβ1,3GlcNAc.
[n]Complex with O-iPr-GlcNS6Sα1,4IdoA2Sα1,4GlcNS6Sα.
[o]K_d values for the two binding sites from Williamson *et al*.[65]
[p]K_d from Wiedemann *et al*,[66] the solubility of the complex in aqueous solution was so low that the NMR structure was determined based on measurements in DMSO.
[q]n.d.: not determined.
[r]personal communication with the authors.
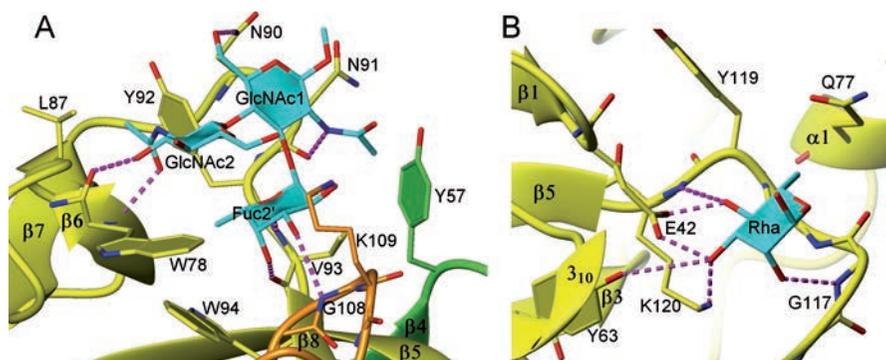[s]not used for the calculations, just for verification.

Figure 3.2   Two examples of protein–carbohydrate structures. (A) The tightly binding Coprinopsis cinerea lectin 2 (CCL2) in complex with the trisaccharide GlcNAcβ1,4[Fucα1,3]GlcNAc and (B) the complex of rhamnose binding lectin (RBL) domain of Latrophilin-1 and ʟ-rhamnose.

3.3   Three-Dimensional Structure Determination by NMR Spectroscopy

3.3.1   The Importance of Intermolecular NOEs

The most reliable NMR method to determine 3D structures of protein–carbohydrate complexes is based on intermolecular nuclear Overhauser effects (NOEs), because an NOE directly establishes the proximity of a protein proton with a ligand proton. If NOEs are observed for example between a methyl group of the protein and several carbohydrate protons, it is a very immediate way to prove such proximity in solution. Such a direct measurement in solution is a huge advantage of NMR spectroscopy. In contrast, deducing such interactions from X-ray crystal structures is more indirect and can be biased, because (a) ligand interactions in the crystal might differ from that in solution, *e.g.* by occluding the binding site due to crystal packing[18] or when the oligosaccharide is contacting and thus bridging several protein molecules within the crystal[19,20] (b) the conditions under which crystals are formed are often quite different than those relevant for the protein–carbohydrate interactions in their natural environment and (c) positioning small saccharides into the electron density, especially for low resolution data, is error prone as illustrated by the high number of inconsistencies or errors in carbohydrate moieties of protein crystal structures.[21] Due to the direct contact information contained in intermolecular NOEs measured in solution, NMR structures determined based on them represent much more reliable structures than those obtained from theoretical calculations without experimental restraints.

The number of intermolecular NOEs critically governs the quality of a protein–carbohydrate complex structure. A handful of NOEs might be sufficient to localise the ligand at the right binding site, but it might be insufficient

to properly orient the ligand. In contrast, a high number of intermolecular NOEs improves the precision and the accuracy of a complex and, thus, reveals much more atomic interactions. In the case of the CCL2–trisaccharide complex[16] 82 intermolecular NOEs could be unambiguously assigned which are schematically illustrated in Figure 3.3(A). This number is surprisingly large for a bound trisaccharide that crucially contributed to a high quality complex structure, in which a large number of individual interactions are observed as shown schematically in Figure 3.3(B). In the case of the low affinity complex between Latrophilin-1 and L-rhamnose[17] the detection of 16 intermolecular NOEs was a sufficiently large number for orienting and defining the monosaccharide (Figure 3.3(C)). These intermolecular NOEs together with 12 docking restraints based on chemical shift changes resulted in a precise structure revealing many atomic contacts (Figure 3.3(D)).
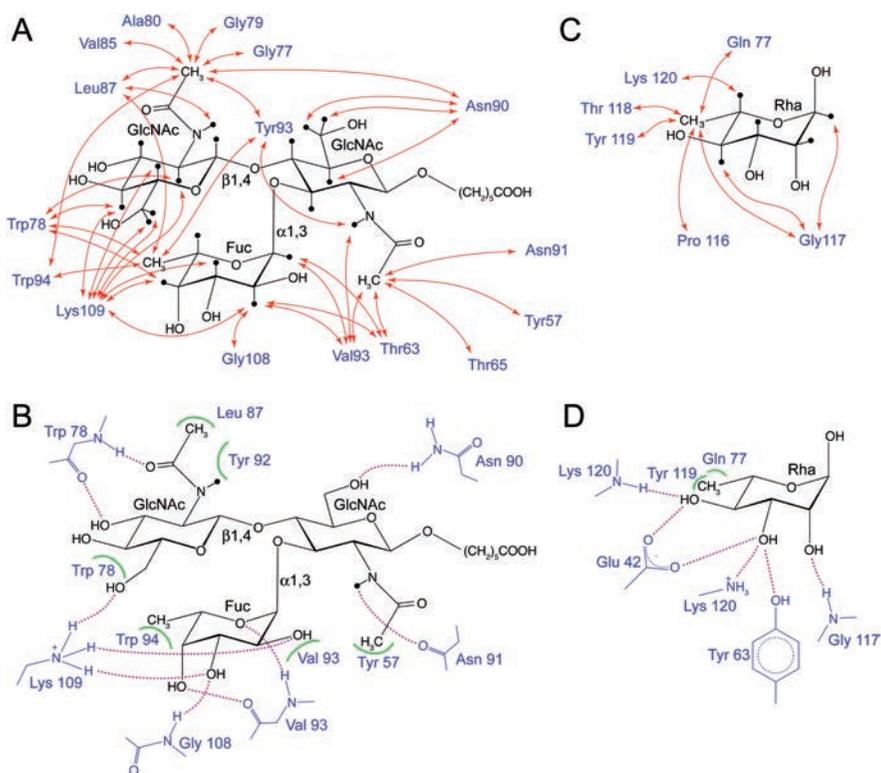

Figure 3.3   Intermolecular NOEs and extracted details of protein–carbohydrate complex structures illustrated with two complexes: (A, B) the complex between the CCL2 and the trisaccharide GlcNAcβ1,4[Fucα1,3]GlcNAc and (C, D) the complex between the Latrophilin-1 and L-rhamnose. In (A) and (C) the observed intermolecular NOEs are shown schematically. Whereas the NOEs are shown for each proton of the ligand, the involved protons on the protein side are combined for each residue. In (B) and (D) the key interactions involved in ligand recognition are displayed schematically.

3.3.2 NMR Pulse Sequences to Detect Intermolecular NOEs

Choosing the most sensitive NMR technique is a critical factor for obtaining many intermolecular NOEs with high quality. To achieve this, the protein needs to be isotope labelled with ^{13}C and ^{15}N close to 100% so that protons of the protein can be distinguished from protons of the ligand by ^{13}C/^{15}N-editing and ^{13}C/^{15}N-filter elements within an NMR pulse sequence. Numerous suitable versions of 2D and 3D ^{13}C-filtered and ^{13}C-edited ^{1}H,^{1}H NOESY experiments are described in excellent reviews.[22–24]

Two-dimensional (2D) ^{1}H–^{1}H NOESY spectra that select protons attached to ^{13}C of labelled protein in one dimension and a ^{13}C-filter to select protons attached to ^{12}C (unlabelled ligand) in the other dimension ensures the exclusive observation of intermolecular NOEs. Either the protein resonances are selected in the indirect dimension which is the case in an F1-edited F2-filtered NOESY or in the direct dimension as in the F1-filtered F2-edited NOESY.[25] The detection of the ligand resonances in the direct dimension has the advantage that the resolution is best for the unlabelled ligand. The resonances of the labelled component can be further resolved by introducing a third dimension to record ^{13}C chemical shift correlations.

About an order of magnitude in sensitivity can be gained if the measurements are carried out in D_2O, although then all NOEs involving exchangeable protons disappear. The most sensitive 3D experiment with the best resolution is an F1 ^{13}C-edited, F3 ^{13}C-filtered HSQC–NOESY described in Dominguez *et al.*[24] It combines the high resolution for the often highly degenerate ^{1}H signals of the unlabelled carbohydrate with the high sensitivity that is obtained for the signals around the water resonance due to the measurement in D_2O. High magnetic fields help to resolve chemical shift degeneracies of the carbohydrate signals together with increasing the sensitivity and the NOE build-up. Although the two indirect dimensions have lower resolution, the combination of ^{1}H with ^{13}C typically allows an unambiguous assignment of an NOE to a protein CH, CH_2 or CH_3 group. Any unnecessary elements of the pulse sequence that reduce signal intensity such as a WATERGATE (WATER-suppression by GrAdient-Tailored Excitation) element[26,27] should be avoided for measurements in D_2O.[24] For detecting intermolecular NOEs of exchangeable protons in aqueous solution, ^{13}C/^{15}N-filter elements either in the direct or indirect dimension can be used. A 3D F1 ^{13}C-edited F3 ^{13}C/^{15}N-filtered HSQC–NOESY with WATERGATE is available as the Bruker standard pulse sequence hsqcgpnoewgx33d. In addition a 3D F1 ^{13}C/^{15}N-filtered F3 ^{15}N-edited NOESY–HSQC with WATERGATE is provided as noesyhsqcf3gpwgx13d.

Critical for all these NOESY experiments is a properly chosen mixing time to avoid spin diffusion. Typical values range in between 80 and 150 ms for complex sizes of <20 kDa. For larger complexes 80–120 ms are appropriate.

3.3.3 Isotope Labelling of Oligosaccharides

In case of chemical shift degeneracies of the oligosaccharide some intermolecular NOEs cannot unambiguously assigned using only one ^{1}H dimension. Uniform isotope labelling of the carbohydrate with ^{13}C and ^{15}N can resolve

such degeneracies by obtaining a better dispersion in two-dimensional ^1H–^{13}C and ^1H–^{15}N correlations or even three-dimensional correlations.[28,29] The use of uniformly $^{13}C/^{15}N$-labelled oligosaccharides in complex with unlabelled protein together with 3D $^{13}C/^{15}N$-filtered and $^{13}C/^{15}N$-edited NOESY experiments as described above allow ambiguities in assigning intermolecular NOEs to be resolved by offering two dimensions (^1H and ^{13}C or ^{15}N) for the identification of the involved ligand protons. Although uniform labelling of oligosaccharides with ^{13}C and ^{15}N isotopes is very difficult, it was successfully achieved either by chemical synthesis,[30] chemoenzymatic synthesis[31] or by bacterial synthesis.[32] The obtained $^{13}C/^{15}N$-isotope labelled carbohydrates include an heparan sulfate analogue,[33] hyaluronan oligomers of different length,[32] α(2,8) polysialic acid,[34] lipopolysaccharide,[35] GM3, Lewis[x] and sialyl Lewis[x].[31] Although, isotope labelled oligosaccharides have to the best of my knowledge not been used so far for detecting intermolecular NOEs, they will be very important for the NMR structure determination of protein interactions with larger glycans in the future.

3.3.4   Line Broadening at the Recognition Interface

Resonances that experience large chemical shift changes upon binding are prone to line broadening, both on the protein and the ligand side, if the exchange rate falls in the same range as resonance frequency deviations upon binding. Especially protons that are involved in proton–π interactions at the protein–carbohydrate interface experience large ring current effects. Line broadening of such resonances is especially unfavourable, because aromatic rings are frequently found at carbohydrate-binding sites and such ring protons could provide intermolecular NOEs that are central to define the 3D structure of the complex. Strategies to avoid this line broadening are to increase or decrease temperature, increasing the fraction of bound ligand, or using an excess of ligand as discussed in the following sections.

The exchange regime of protein resonances can be different compared to ligand resonances, because it depends not only on the kinetics but also on the individual frequency differences between free and bound state (in Hz). Therefore, it is also dependent on the type of experiment. In the case of the CCL2–ligand complex, ^1H and ^{15}N signals of the protein displayed slow chemical exchange behaviour, in a 1:1 complex only signals of the bound protein are seen (Figure 3.4(A)). ^1H signals of the ligand also showed slow chemical exchange behaviour (Figure 3.4(B)), but ^1H–^{13}C cross-peaks of the ligand were never detectable in a natural abundance ^{13}C-HSQC spectrum, likely due to intermediate exchange of the ^{13}C resonances on the NMR time scale.

3.3.5   Binding Equilibria and Their Consequences on Detecting Intermolecular NOEs

To minimise line-broadening effects at the protein–carbohydrate interface, one has to take into account the fraction of bound protein as well as the
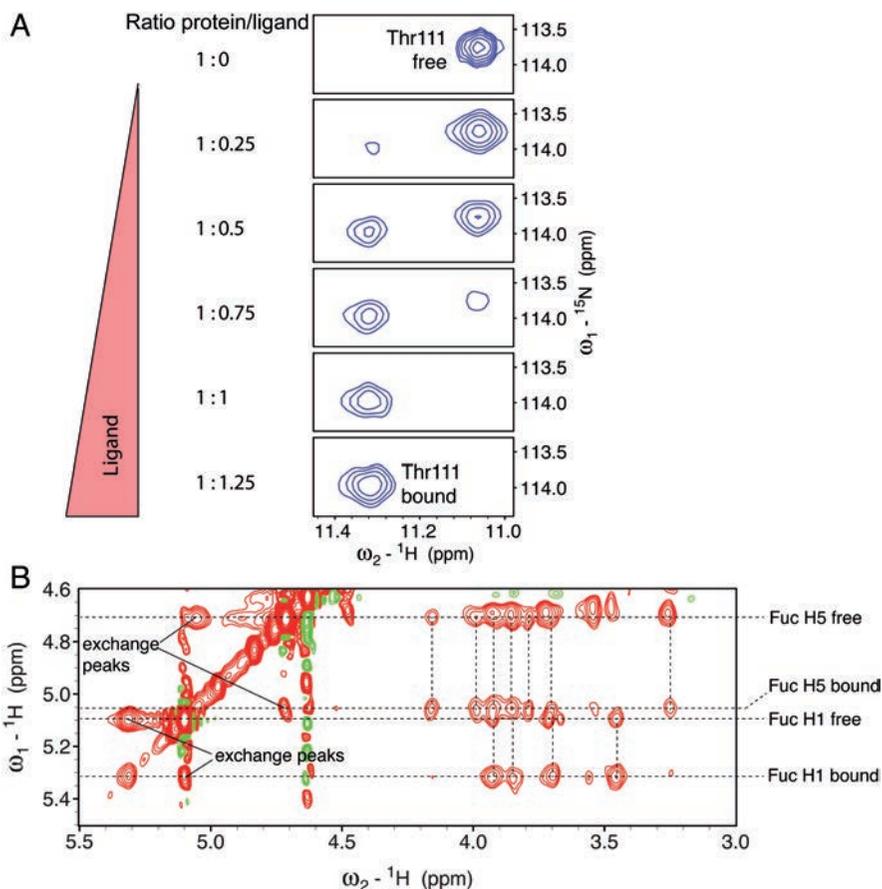
Figure 3.4   Slow exchange behaviour of protein and ligand resonances in the complex of CCL2 and fucosylated chitobiose. (A) Region of a ^{15}N-HSQC spectrum of CCL2 in the presence and absence of the ligand fucosylated chitobiose. The behaviour of the ^{1}H–^{15}N correlation signal of Thr111 with increasing amounts of ligand is shown. (B) Chemical exchange of the ligand fucosylated chitobiose in complex with CCL2 (at a protein–ligand ratio of 1:2.7) illustrated with a 2D F1 ^{13}C-filtered F2 ^{13}C-filtered ^{1}H–^{1}H NOESY recorded at 700 MHz and 310 K with a mixing time of 150 ms. Exchange peaks between free and bound resonances are visible and cross-peak patterns are similar between free and bound resonances.

fraction of bound ligand. For a single binding site interaction the fraction bound can be calculated as follows (see eqn (3.1)).[36]

$$f_b = \frac{\left(P_0 + L_0 + K_d\right) - \sqrt{\left(P_0 + L_0 + K_d\right)^2 - 4P_0L_0}}{2P_0} \tag{3.1}$$

here f_b denotes the fraction bound, P_0 the total protein concentration, L_0 the total ligand concentration and K_d the dissociation constant.
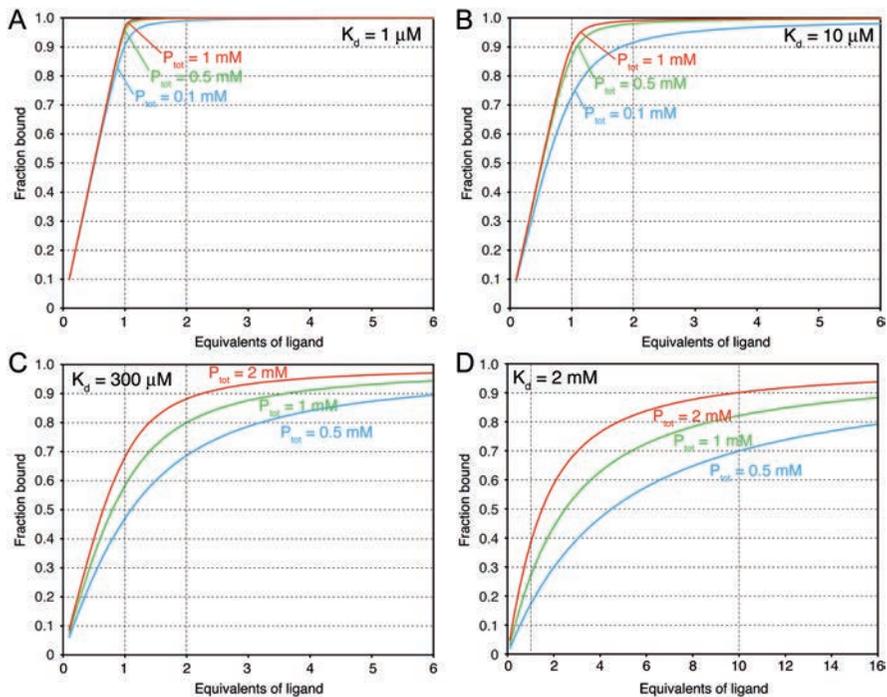
Figure 3.5   Fraction bound for a single binding-site model, calculated with eqn (3.1). Graphs for four dissociation constants are shown with some typical protein concentrations: (A) $K_d = 1\,\mu M$, (B) $K_d = 10\,\mu M$, (C) $K_d = 300\,\mu M$ and (D) $K_d = 2\,mM$.

How the bound fraction depends on the ligand–protein ratio is illustrated for some typical cases in Figure 3.5. In case of a rather high affinity binder with a $K_d = 1\,\mu M$ over 95% saturation is achieved with a 1:1 complex at concentrations >0.5 mM, while only a small excess of ligand pushes the bound fraction to 100% (Figure 3.5(A)). For a ligand that binds ten times weaker ($K_d = 10\,\mu M$) the situation changes slightly: only 90% of the protein is bound in a 1:1 complex at a concentration of 1 mM (Figure 3.5(B)). To achieve saturation an excess of ligand is required. Many protein–carbohydrate interactions display a weak to moderate affinity, which results in a different scenario. An interaction with a $K_d = 300\,\mu M$ is illustrated in Figure 3.5(C). Even at high protein concentrations of 2 mM only 70% of the protein is bound at a 1:1 ratio. To reach 90% saturation an excess of ligand with a ratio of 1:2.2 (for 2 mM protein) or 1:3.6 (for 1 mM protein) is required. Substantial higher ratios are required for even weaker interactions, for example $K_d = 2\,mM$ as shown in Figure 3.5(D). In this case a ratio of 1:10 is necessary to achieve 90% saturation at a protein concentration of 2 mM.

To avoid signal loss due to chemical exchange phenomena the binding sites should be saturated. As a rule of thumb a ligand concentration of $10 \times K_d$ is required to reach the plateau of saturation. However, for the detection of intermolecular NOEs both components are required in sufficient concentrations and exchange broadening on the ligand side has to be prevented as well. There are two approaches to minimise line broadening caused by the exchange in intermolecular NOE experiments: the first approach uses high concentrations of both the protein and the ligand at a ratio of 1 : 1 to push both components close to saturation, the second approach uses the ligand in excess to have the protein fully saturated and the ligand mainly free. For this second approach the ligand has to exchange between free and bound state during the NOESY experiment so that intermolecular NOEs from the bound state can be picked up on the free ligand resonances.

3.3.6 Strategies to Obtain Good Intermolecular NOEs for High Affinity Complexes

The most common approach uses high concentrations of both the protein and the ligand at a 1 : 1 ratio to achieve high saturation for both components. This strategy is illustrated with the fungal lectin *Coprinopsis cinerea* lectin 2 (CCL2) interacting with the trisaccharide GlcNAcβ1,4[Fucα1,3]GlcNAc with a $K_d = 1$ μM. Unusual for this affinity range, the exchange rate is slow ($k_{ex} < 80$ s^{-1}) so that protein amide signals of both the free and the bound protein are visible during the titration with increasing amounts of the ligand (Figure 3.4(A)). In order to observe intermolecular NOEs a concentration of ~1 mM and a 1 : 1 protein–ligand ratio was used, which resulted in a saturation of 97% on the protein side and 97% on the ligand side according to eqn (3.1). Sharp intermolecular signals were observed. In practice the ratio was verified with the disappearance of free protein signals. An excess of ligand pushed the protein resonances towards 100% of bound state leading to the simultaneous observation of free and bound ligand resonances (Figure 3.6(A)). Interestingly, both free and bound state resonances pick up intermolecular NOEs. Although the exchange regime was favourable to observe ^{1}H, ^{13}C and ^{15}N resonances on the protein side, only ^{1}H ligand signals could be observed for the bound form of the ligand. Any attempt to detect ^{1}H–^{13}C correlations at ^{13}C natural abundance did not show any bound state signals, only signals of the free ligand were detected when a slight excess of ligand was used.

The strategy with high concentration at a 1 : 1 ratio also works for fast exchanging resonances. The concentration of the protein–carbohydrate complex should be well above $10 \times K_d$, or even better $100 \times K_d$ to achieve high saturation of both the protein and the ligand. The exact 1 : 1 ratio is critical, a small excess of ligand results in a significant population of free ligand, and an excess of protein reduces the fraction bound.
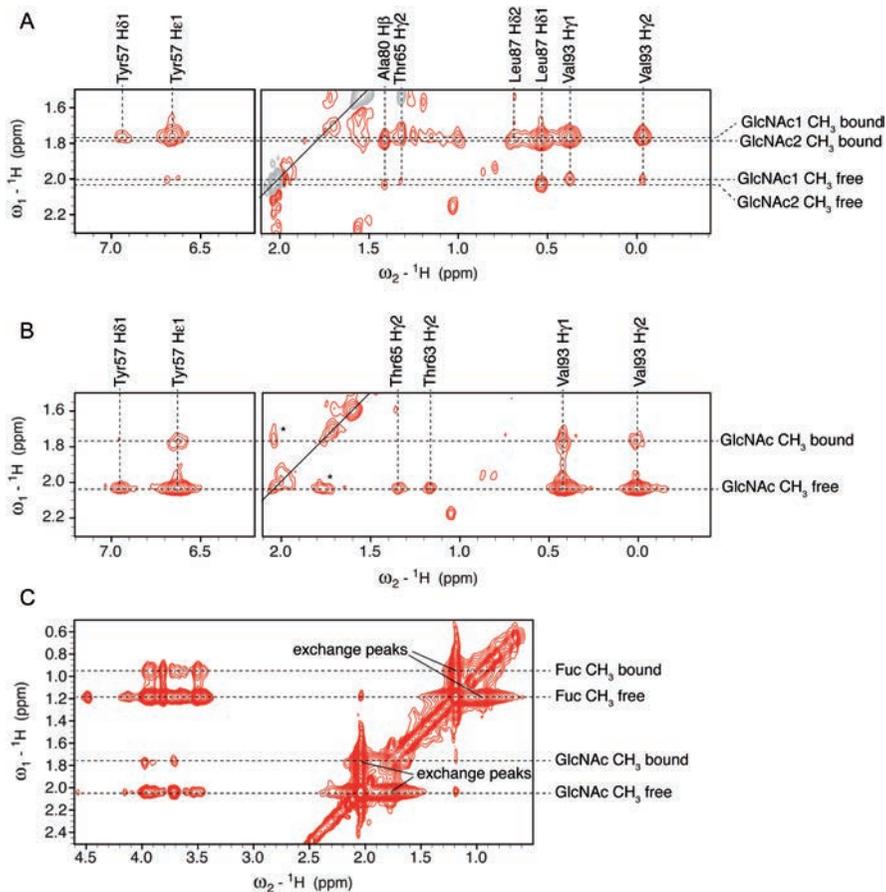
Figure 3.6 Intermolecular NOEs picked up on free and bound ligand resonances in slow exchange. (A) 2D F1 ^{13}C-filtered F2 ^{13}C-edited ^{1}H–^{1}H NOESY measured of the complex between CCL2 (1 mM) and fucosylated chitobiose (2.7 mM) at 500 MHz and 310 K with a mixing time of 150 ms. (B) 2D F1 ^{13}C-filtered F2 ^{13}C-edited ^{1}H–^{1}H NOESY measured of the interaction between CCL2 (1 mM) and Lewis^x (2.7 mM) measured at 500 MHz and 310 K with a mixing time of 150 ms. (C) 2D F1 ^{13}C-filtered F2 ^{13}C-filtered ^{1}H–^{1}H NOESY of the same complex as in panel B recorded at 500 MHz and 310 K with a mixing time of 150 ms.

3.3.7 Low Affinity – A Problem for Protein–Carbohydrate Structure Determination?

Counterintuitively, low to medium affinity protein–carbohydrate interactions (K_d = 10 μM–10 mM) do not prevent NMR structure determination. However, slightly different strategies have to be applied. Basically, a compromise has to be found, to achieve sufficient concentrations of both components and to minimise the line broadening of signals for both the protein and the ligand.

Typically, weak interactions are characterised by fast exchange kinetics, where resonance frequencies and line shapes of both the protein and the ligand depend very much on the grade of saturation. However, slow or intermediate exchange regimes can also occur for low affinity interactions. In any case, resonance assignments of the protein are normally achieved in the fully saturated state with an excess of ligand. Under fast exchange regimes, signal positions depend on the binding equilibrium so that differences in chemical shifts might occur between the assigned protein resonances of the fully saturated state and the observed peak positions.

For moderate affinities (K_d = 10–500 μM) the concentrations required to saturate both components in a 1:1 ratio are already difficult to reach. The stability and solubility of the sample set the limits in that context. Nevertheless, successful examples using high concentrations together with a 1:1 ratio have been reported for protein–carbohydrate and protein–RNA complexes, *e.g.* the microvirin–Manα1,2Man complex[37] with a 1.5 mM complex concentration (K_d = 48 μM) and the SRp20–RNA complex[38] with 1 mM complex concentration (K_d = 20 μM). Let us consider a protein–ligand interaction with a K_d of 300 μM, for which the calculated bound fraction is shown in Figure 3.5(C). A 1:1 mixture results in only 70% saturation even at the high concentration of 2 mM (for both the protein and the ligand). For a fast exchange regime the line widths of both components might be still sufficiently sharp to observe intermolecular NOEs, whereas an intermediate exchange regime would likely prevent this observation. Surprisingly, even for affinities in the micromolar range, slow exchange behaviour can occur, for example in the case of the interaction between CCL2 and Lewis^x (Galβ1,4[Fucα1,3]GlcNAcβ). Lewis^x lacks an acetamido group compared to the preferred ligand fucosylated chitobiose (GlcNAcβ1,4[Fucα1,3]GlcNAcβ; K_d = 1 μM), resulting in an almost 500-fold decrease in affinity (K_d = 456 μM). Despite this low affinity the ligand resonances are in slow exchange as shown in a 2D F1 ^{13}C-filterd F2 ^{13}C-edited NOESY (Figure 3.6(B) and (C)). The intermolecular NOE patterns are very similar to the complex with the preferred ligand fucosylated chitobiose, indicating a very similar ligand recognition and orientation. Interestingly, the intermolecular cross-peaks picked up at the resonances of the free Lewis^x ligand have much higher intensities than the corresponding cross-peaks picked up at the bound state resonances. This finding is contrary to the observation made with the preferred ligand (Figure 3.6(A)) and gives a hint how to study such interactions. The resonances of the free and bound form exchange during the NOE mixing time as indicated by strong exchange peaks in a 2D F1 ^{13}C-filterd F2 ^{13}C-edited ^{1}H–^{1}H NOESY (Figure 3.6(C)), resulting in very similar NOE patterns picked up on the free and the bound state resonances (Figure 3.6(B) and (C)).

Independent of the exchange regime, the best strategy for the structure determination of weak interactions is using an excess of ligand, which results in a complete saturation of the protein. For example, an excess of ligand resulting in <10% bound ligand has the advantage that the chemical shifts of the ligand will be very close to resonances of the free form, so that

an assignment of the ligand in the bound form is not required to interpret intermolecular NOEs. The protein concentration is very important in that context, whereas a ligand excess of $1:10$ is sufficient to reach 90% saturated protein with an $K_d = 2$ mM at a protein concentration of 2 mM, a much higher excess is required to reach the same saturation at a protein concentration of 1 mM (Figure 3.5(D)).

An extreme of this scenario is used for the transfer NOE experiment, in which a small ligand is in large excess, whereas the protein is present in very small amounts and, thus, completely saturated. In this way, the ligand resonances correspond to the free ligand ones and the small bound fraction still dominates the NOE enhancement that reports the conformation of the bound ligand. This scenario is, however, not suited to detect intermolecular NOEs since the protein concentration must be high enough to detect protein signals.

3.3.8 Influence of the Field Strength and Other Parameters on Intermolecular NOEs

The field strength has a marked influence on the exchange behaviour of the NMR resonances and thus on the observation of intermolecular NOEs. Complexes displaying intermediate to slow exchange behaviour benefit from higher field strengths, where the exchange regime is pushed towards slow exchange. With the other advantages of higher resolution, better NOE build-up and sensitivity, significant improvements are expected from 28.2 Tesla magnets (1.2 GHz systems). Such magnets raise the likelihood to observe intermolecular NOEs that could not be observed otherwise. However, complexes displaying intermediate to fast exchange might not profit from higher fields due to increased line broadening. Then a compromise regarding the magnetic field has to be experimentally found.

Beside the field strength, protein and ligand concentrations, there are many parameters that can be changed to improve the quality of intermolecular NOEs: temperature, pH, ionic strength and the type of buffer system. The affinity of charged interactions can be increased by lower salt concentrations due to a larger contribution of Coulomb attractions to the binding. A decrease in ionic strength might help for complexes with sialic acid-containing glycans or sulfated glycans that are negatively charged and for which electrostatic interactions are expected. In that case lower ionic strength might tip the balance to achieve sufficiently sharp lines on both the protein and the ligand side in order to obtain good intermolecular NOEs, similar to observations made with protein–RNA complexes.[38]

3.3.9 Other Complications and Solutions

Line broadening can also have other reasons than the exchange between free and bound states, for example slow ring flipping of Tyr and Phe side chains. Typically these side chains are flipping their aromatic rings very fast leading

to average chemical shifts for both the Hδ and the Hε hydrogens. However, if sterically hindered by their environment, these flipping rates can slow down dramatically so that in extreme cases two resonances for Hδ and two for Hε can be observed.[39] When the flipping rates get close to the chemical shift differences $\Delta\delta = \delta(H\delta1) - \delta(H\delta2)$ or $\Delta\delta = \delta(H\varepsilon1) - \delta(H\varepsilon2)$ this leads to such severe line broadening that the signals disappear. This effect is important since aromatic residues are overrepresented at the binding sites for carbohydrate interactions and the ring flipping of Tyr and Phe side chains is sometimes slowed down due to the ligand. In the case of the CCL2–trisaccharide complex, the Hδ and Hε resonances of Tyr92 are broadened and not observable in most spectra. The aromatic ring of Tyr92 is stacking to the *N*-acetyl group of a GlcNAc that is slowing down the flipping rate. The Hδ resonance of Tyr92 was only observable in the complex in a 3D F1 ^{13}C-edited F3 ^{13}C-filtered HSQC–NOESY albeit with a broad line shape.

Sliding of the ligand within the binding side is another complication that might occur when repetitive units follow each other in a linear oligosaccharide. This behaviour increases the apparent affinity but hampers studying the interaction with NMR spectroscopy due to exchange between different registers of binding. This phenomenon is well-described for example by the protein–RNA interaction of hnRNP C with single stranded poly-U containing RNA oligonucleotides.[40] In that case, the exchange rate between different binding registers was faster than the applied mixing times in the NOESY experiments, so that resonances of one particular uracil picked up NOEs from different register positions within the binding site. Structure calculations with all NOE derived distance restraints resulted in a dilemma, because the restraints could not be fulfilled by one structure with one register. However, after disentangling the contributions of the different registers with the help of an alternate ^{13}C-labelled RNA sample, a 3D structure determination was achieved.[40] The recognition of glucosaminoglycans consisting of repetitive (disaccharide) units displays striking similarities to such a scenario: the binding sites for glucosaminoglycans are probably even more plastic because some seem to accommodate ligands in different orientations, thereby mainly recognising SO_4^- groups, no matter which underlying structure they are attached to, as long as the SO_4^- groups occur on similar sites with similar distances.[41–43] Nevertheless, NMR spectroscopy was successfully used to provide a three-dimensional picture of the human FGF-1 in complex with a hexasaccharide heparin analogue.[44]

3.3.10 Promising Technological Developments Applicable to Protein–Carbohydrate Complexes

Despite the advantages of using NOE-based 3D structure determination, it requires a lot of effort in terms of spectrometer time and data analysis, whereby the data complexity increases exponentially with protein size. Additionally, although NOE-derived NMR structures can reach good local accuracy, the entire structures are usually less accurate than corresponding

crystal structures due to the lack of long-range structural restraints. Long-range restraints obtained by residual dipolar couplings (RDCs) and pseudo contact shifts (PCS) can significantly improve the accuracy and speed of convergence.[45–47] In particular PCS are promising for protein complexes,[48,49] they provide unique inter-molecular long-range restraints that help to orient and position the binding partner. Long-range restraints of up to 40 Å were obtained, providing very valuable restraints.[45] PCS require a metal binding tag on either the protein or the ligand and are easily measured by the chemical shift differences $\Delta\delta^{PCS}$ observed between a sample with a diamagnetic and a paramagnetic metal ion. Using different lanthanoid ions with the same chelator tag results in independent datasets providing a large amount of long-range restraints. The use of PCS for the structure determination of protein complexes evolved rapidly, which is excellently reviewed for protein–protein complexes by Hass and Ubbink.[48] This methodology should also be applicable to protein–carbohydrate complexes and promises exciting molecular details in the near future. Oligosaccharides with lanthanide binding tags have already been reported.[50–52] So far they were mainly used to elucidate the structures of the oligosaccharides, but their application to structure determination of protein–carbohydrate complexes seems feasible. The first PCS- and field-induced RDC-based structural models of a protein–carbohydrate complex were reported for Galectin-3 in complex with lactose,[53,54] although they were not deposited in the PDB.

3.4 Judging the Quality of Protein–Carbohydrate Complex Structures Determined by NMR

How reliable are NMR structures of protein–carbohydrate complexes? Figure 3.7 shows four examples of protein–carbohydrate complex structures obtained by NMR spectroscopy displaying different precision at the binding site. It seems logical that the more restraints are used, the better the precision and accuracy. However, the two examples in Figure 3.6 with most inter-molecular NOEs, namely the CCL2 complex (82 intermol. NOEs) and Malectin complex (31 intermol. NOEs) display not the best precision with RMSDs (all heavy atoms) of 1.17 Å and 1.77 Å, respectively. However, for the complexes of Latrophilin and TgMIC4–A5 with RMDSs of 0.61 Å and 0.54 Å, respectively, a docking algorithm was used that applied chemical shift perturbation-derived ambiguous interaction restraints in addition to intermolecular NOEs. Also the number of restraints has to be considered in relation to the ligand size and the protein size, *e.g.* smaller ligands are easier defined. In addition, an excellent precision might be misleading, because few incorrect restraints can over-restrain the ensemble leading to an artificially high precision but worse accuracy.

Judging the accuracy of a protein–carbohydrate complex structure is more difficult. In general more specific interactions such as hydrogen bonds and H–π interactions are visible in high quality structures. Several observables can be used to verify the correctness of a structure, favourably unbiased
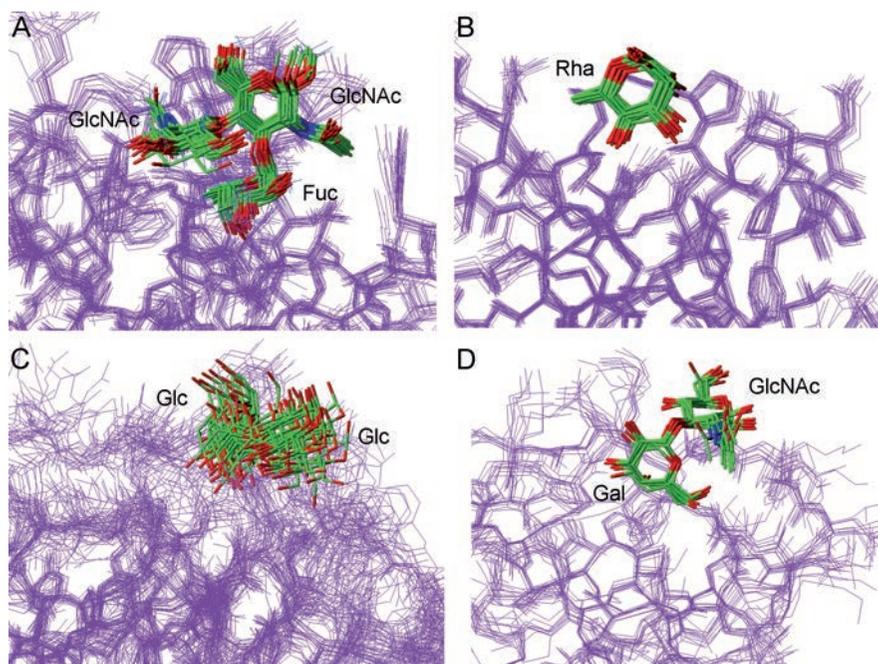
Figure 3.7 Precision of various protein–carbohydrate complexes. (A) Complex between CCL2 and GlcNAcβ1,4[Fucα1,3]GlcNAc (PDB: 2LIQ). (B) Inter-action between rhamnose binding lectin (RBL) domain of Latrophilin-1 and L-rhamnose (PDB: 2JXA). (C) Malectin interacting with Glcα1,3Glc (PDB: 2K46). (D) The Apple-5 domain of *Toxoplasma gondii* microneme protein (TgMIC4 A5) in complex with lacto-*N*-biose (PDB: 2LL4).

observables that were not used for the structure determination. Chemical shift deviation during titration, both on the ligand and the protein side, is one such indicator. The structure of the CCL2–trisaccharide complex revealed three intermolecular N–H···O hydrogen bonds. Interestingly, all three showed large chemical shift deviations during the NMR titration, of which Trp78 and Gly108 represent the largest deviations of all with $\Delta\delta(^1\text{H})$ of 1.66 and 1.21 ppm, and $\Delta\delta(^{15}\text{N})$ of 3.4 and 2.7 ppm, respectively (Figure 3.8). This observation confirms the formation of N–H···O hydrogen bonds upon carbohydrate binding. On the ligand side, the H^N of the acetamido group of GlcNAc1 experiences a downfield shift of 0.29 ppm, confirming its participation in an N–H···O hydrogen bond to the carbonyl of Asn91 as well. Methyl–π interactions also lead to significant chemical shift deviations upon complex formation. The methyl group of GlcNAc1 that is forming a methyl–π interaction with Tyr68 is experiencing an upfield shift of 0.24 ppm. Similar observations were made for the low affinity complex between Latrophilin-1 and L-rhamnose, where the two amides forming direct N–H···O hydrogen bonds to the monosaccharide experience the largest chemical shift devia-tions among all N–H resonances.[17]
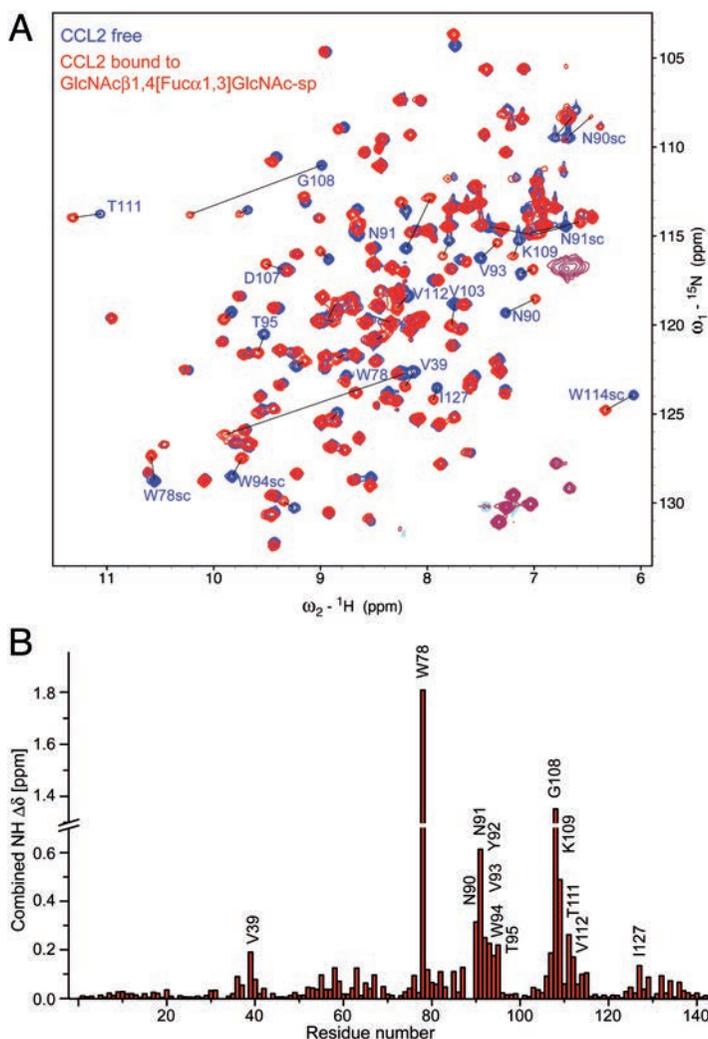
Figure 3.8 Chemical shift deviations verify the accuracy of a complex structure. (A) Overlay of two ^{15}N-HSQC spectra of CCL2 in the presence (red) and absence (blue) of the ligand fucosylated chitobiose. The NH group of each protein residue is represented by one signal whose spectral position is sensitive to the local environment. A change in the proximity of an NH, *e.g.* due to ligand binding, is picked up by this sensitive reporter and a deviation of the signal position is observed. The most prominent deviations are labelled and the corresponding signals connected by a line. (B) Plot of the combined ^{1}H and ^{15}N chemical shift deviations between free and bound CCL2. (Reproduced with permission from ref. 16, Copyright: © 2012 Schubert *et al.* This is an open-access article distributed under the terms of the Creative Commons Attribution License.)

In conclusion, chemical shift changes and expected ring current effects can be used to judge the accuracy and quality of a protein–carbohydrate complex structure. Observables that were not used for the structure determination enable an unbiased verification, similar to the separate set of reflections in X-ray crystallography for calculating R_{free}.

3.5   Conclusions

In summary, NMR spectroscopy is a very powerful method to obtain 3D structures of protein–carbohydrate complexes. Especially in the case of weak carbohydrate interactions, NMR spectroscopy offers significant advantages over X-ray crystallography due to the difficulty in obtaining high quality crystals and blurred or lacking electron density. The development of higher magnetic fields promises a further advance for the study of protein–carbohydrate interactions not only due to better sensitivity, resolution and more favourable NOE build-up, but also by providing another way to counteract detrimental exchange broadening.

With the continuous finding of new functional glycoepitopes and their involvements of protein–carbohydrate recognition interactions in almost all biological functions, the elucidation of the 'glycocode' is an area expected to grow very fast in the near future. Still underestimated in many fields of biology and medicine there is increasing evidence that weak to moderate protein–carbohydrate interactions are involved in central elements of immune function, for example self and non-self recognition, down or up-regulating immune responses. Exploring the molecular basis of such interactions that are so central to immunology and medicine will have a tremendous impact. NMR spectroscopy can make a significant contribution by providing three-dimensional structures of the protein–carbohydrate complexes.

Acknowledgements

References

1. J. D. Marth and P. K. Grewal, *Nat. Rev. Immunol.*, 2008, **8**, 874–887.
2. M. Sperandio, C. A. Gleissner and K. Ley, *Immunol. Rev.*, 2009, **230**, 97–113.
3. K. S. Lau and J. W. Dennis, *Glycobiology*, 2008, **18**, 750–760.

4. Y. Y. Zhao, M. Takahashi, J. G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume and N. Taniguchi, *Cancer Sci.*, 2008, **99**, 1304–1310.
5. F. Lehmann, E. Tiralongo and J. Tiralongo, *Cell. Mol. Life Sci.*, 2006, **63**, 1331–1354.
6. G. R. Vasta, *Nat. Rev. Microbiol.*, 2009, **7**, 424–438.
7. B. Khatua, S. Roy and C. Mandal, *Indian J. Med. Res.*, 2013, **138**, 648–662.
8. A. Tannous, G. B. Pisoni, D. N. Hebert and M. Molinari, *Semin. Cell Dev. Biol.*, 2015, **41**, 79–89.
9. Y. Kamiya, T. Satoh and K. Kato, *Biochim. Biophys. Acta, Gen. Subj.*, 2012, **1820**, 1327–1337.
10. R. D. Cummings, *Mol. BioSyst.*, 2009, **5**, 1087–1104.
11. A. Arda, A. Canales, F. J. Canada and J. Jimenez-Barbero, in *Carbohydrates in Drug Design and Discovery*, ed. J. Jimenez-Barbero, F. J. Canada and M.-S. Sonsoles, The Royal Society of Chemistry, 2015, pp. 1–20.
12. V. Roldos, F. J. Canada and J. Jimenez-Barbero, *ChemBioChem*, 2011, **12**, 990–1005.
13. C. A. Bewley and S. Shahzad-ul-Hussan, *Biopolymers*, 2013, **99**, 796–806.
14. M. D. Fernandez-Alonso, D. Diaz, M. A. Berbis, F. Marcelo, J. Canada and J. Jimenez-Barbero, *Curr. Protein Pept. Sci.*, 2012, **13**, 816–830.
15. A. Clery, M. Schubert and F. H. T. Allain, *Chimia*, 2012, **66**, 741–746.
16. M. Schubert, S. Bleuler-Martinez, A. Butschi, M. A. Walti, P. Egloff, K. Stutz, S. Yan, I. B. H. Wilson, M. O. Hengartner, M. Aebi, F. H. T. Allain and M. Kunzler, *PLoS Pathog.*, 2012, **8**, e1002706.
17. I. Vakonakis, T. Langenhan, S. Promel, A. Russ and I. D. Cambell, *Structure*, 2008, **16**, 944–953.
18. I. Botos, B. R. O'Keefe, S. R. Shenoy, L. K. Cartner, D. M. Ratner, P. H. Seeberger, M. R. Boyd and A. Wlodawer, *J. Biol. Chem.*, 2002, **277**, 34336–34342.
19. H. Attrill, H. Takazawa, S. Witt, S. Kelm, R. Isecke, R. Brossmer, T. Ando, H. Ishida, M. Kiso, P. R. Crocker and D. M. F. van Aalten, *Biochem. J.*, 2006, **397**, 271–278.
20. C. Fotinou, P. Emsley, I. Black, H. Ando, H. Ishida, M. Kiso, K. A. Sinha, N. F. Fairweather and N. W. Isaacs, *J. Biol. Chem.*, 2001, **276**, 32274–32281.
21. T. Lutteke, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2009, **65**, 156–168.
22. A. L. Breeze, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2000, **36**, 323–372.
23. I. M. Robertson, L. Spyracopoulos and B. D. Sykes, *Nato Science Peace S*, 2009, pp. 101–119.
24. C. Dominguez, M. Schubert, O. Duss, S. Ravindranathan and F. H. T. Allain, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2011, **58**, 1–61.
25. R. D. Peterson, C. A. Theimer, H. H. Wu and J. Feigon, *J. Biomol. NMR*, 2004, **28**, 59–67.
26. M. Piotto, V. Saudek and V. Sklenar, *J. Biomol. NMR*, 1992, **2**, 661–665.
27. V. Sklenar, M. Piotto, R. Leppik and V. Saudek, *J. Magn. Reson., Ser. A*, 1993, **102**, 241–245.
28. S. A. Colebrooke, C. D. Blundell, P. L. DeAngelis, I. D. Campbell and A. Almond, *Magn. Reson. Chem.*, 2005, **43**, 805–815.

29. R. Harris, T. J. Rutherford, M. J. Milton and S. W. Homans, *J. Biomol. NMR*, 1997, **9**, 47–54.

30. D. Live, L. A. P. Silks and J. Schmidt, *Methods Enzymol.*, 2001, **338**, 305–319.

31. M. A. Probert, M. J. Milton, R. Harris, S. Schenkman, J. M. Brown, S. W. Homans and R. A. Field, *Tetrahedron Lett.*, 1997, **38**, 5861–5864.

32. C. D. Blundell, P. L. DeAngelis, A. J. Day and A. Almond, *Glycobiology*, 2004, **14**, 999–1009.

33. C. Laguri, N. Sapay, J. P. Simorre, B. Brutscher, A. Imberty, P. Gans and H. Lortat-Jacob, *J. Am. Chem. Soc.*, 2011, **133**, 9642–9645.

34. H. F. Azurmendi, J. Vionnet, L. Wrightson, L. B. Trinh, J. Shiloach and D. I. Freedberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11557–11561.

35. W. Wang, H. J. Sass, U. Zahringer and S. Grzesiek, *Angew. Chem., Int. Ed.*, 2008, **47**, 9870–9874.

36. M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **73**, 1–16.

37. S. Shahzad-ul-Hussan, E. Gustchina, R. Ghirlando, G. M. Clore and C. A. Bewley, *J. Biol. Chem.*, 2011, **286**, 20788–20796.

38. Y. Hargous, G. M. Hautbergue, A. M. Tintaru, L. Skrisovska, A. P. Golovanov, J. Stevenin, L. Y. Lian, S. A. Wilson and F. H. T. Allain, *EMBO J.*, 2006, **25**, 5126–5137.

39. G. Wagner, D. Bruhwiler and K. Wuthrich, *J. Mol. Biol.*, 1987, **196**, 227–231.

40. Z. Cienikova, F. F. Damberger, J. Hall, F. H. T. Allain and C. Maris, *J. Am. Chem. Soc.*, 2014, **136**, 14536–14544.

41. A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger and W. A. Hendrickson, *Nature*, 1998, **393**, 812–817.

42. S. Faham, R. E. Hileman, J. R. Fromm, R. J. Linhardt and D. C. Rees, *Science*, 1996, **271**, 1116–1120.

43. J. Schlessinger, A. N. Plotnikov, O. A. Ibrahimi, A. V. Eliseenkova, B. K. Yeh, A. Yayon, R. J. Linhardt and M. Mohammadi, *Mol. Cell*, 2000, **6**, 743–750.

44. A. Canales, R. Lozano, B. Lopez-Mendez, J. Angulo, R. Ojeda, P. M. Nieto, M. Martin-Lomas, G. Gimenez-Gallego and J. Jimenez-Barbero, *FEBS J.*, 2006, **273**, 4716–4727.

45. M. Allegrozzi, I. Bertini, M. B. L. Janik, Y. M. Lee, G. H. Lin and C. Luchinat, *J. Am. Chem. Soc.*, 2000, **122**, 4154–4161.

46. I. Bertini, C. Luchinat and G. Parigi, *Concepts Magn. Reson.*, 2002, **14**, 259–286.

47. K. Chen and N. Tjandra, *Top. Curr. Chem.*, 2012, **326**, 47–67.

48. M. A. S. Hass and M. Ubbink, *Curr. Opin. Struct. Biol.*, 2014, **24**, 45–53.

49. D. Shishmarev and G. Otting, *J. Biomol. NMR*, 2013, **56**, 203–216.

50. M. Erdelyi, E. d'Auvergne, A. Navarro-Vazquez, A. Leonov and C. Griesinger, *Chem.–Eur. J.*, 2011, **17**, 9368–9376.

51. A. Canales, A. Mallagaray, J. Perez-Castells, I. Boos, C. Unverzagt, S. Andre, H. J. Gabius, F. J. Canada and J. Jimenez-Barbero, *Angew. Chem., Int. Ed.*, 2013, **52**, 13789–13793.

52. T. Yamaguchi, Y. Sakae, Y. Zhang, S. Yamamoto, Y. Okamoto and K. Kato, *Angew. Chem., Int. Ed.*, 2014, **53**, 10941–10944.

53. T. D. Zhuang, H. S. Lee, B. Imperiali and J. H. Prestegard, *Protein Sci.*, 2008, **17**, 1220–1231.

54. A. Canales, A. Mallagaray, M. A. Berbis, A. Navarro-Vazquez, G. Dominguez, F. J. Canada, S. Andre, H. J. Gabius, J. Perez-Castells and J. Jimenez-Barbero, *J. Am. Chem. Soc.*, 2014, **136**, 8011–8017.

55. K. Sorimachi, M. F. LeGalCoeffet, G. Williamson, D. B. Archer and M. P. Williamson, *Structure*, 1997, **5**, 647–661.

56. C. A. Bewley, *Structure*, 2001, **9**, 931–940.

57. N. Aboitiz, M. Vila-Perello, P. Groves, J. L. Asensio, D. Andreu, F. J. Canada and J. Jimenez-Barbero, *ChemBioChem*, 2004, **5**, 1245–1255.

58. S. T. D. Hsu, E. Breukink, E. Tischenko, M. A. G. Lutters, B. de Kruijff, R. Kaptein, A. M. J. J. Bonvin and N. A. J. van Nuland, *Nat. Struct. Mol. Biol.*, 2004, **11**, 963–967.

59. M. I. Chavez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, G. Asensio, M. Muraki, F. J. Canada and J. Jimenez-Barbero, *Chem.–Eur. J.*, 2005, **11**, 7060–7074.

60. T. Schallus, C. Jaeckh, K. Feher, A. S. Palma, Y. Liu, J. C. Simpson, M. Mackeen, G. Stier, T. J. Gibson, T. Feizi, T. Pieler and C. Muhle-Goll, *Mol. Biol. Cell*, 2008, **19**, 3404–3414.

61. S. M. Kumar, H. M. Wang, S. K. Mohan, R. H. Chou and C. Yu, *Biochemistry*, 2010, **49**, 10756–10764.

62. T. Schallus, K. Feher, U. Sternberg, V. Rybin and C. Muhle-Goll, *Glycobiology*, 2010, **20**, 1010–1020.

63. J. Marchant, B. Cowper, Y. Liu, L. Lai, C. Pinzan, J. B. Marq, N. Friedrich, K. Sawmynaden, L. Liew, W. G. Chai, R. A. Childs, S. Saouros, P. Simpson, M. C. R. Barreira, T. Feizi, D. Soldati-Favre and S. Matthews, *J. Biol. Chem.*, 2012, **287**, 16720–16733.

64. M. F. Garcia-Mayoral, A. Canales, D. Diaz, J. Lopez-Prados, M. Moussaoui, J. L. de Paz, J. Angulo, P. M. Nieto, J. Jimenez-Barbero, E. Boix and M. Bruix, *ACS Chem. Biol.*, 2013, **8**, 144–151.

65. M. P. Williamson, M. F. LeGalCoeffet, K. Sorimachi, C. S. M. Furniss, D. B. Archer and G. Williamson, *Biochemistry*, 1997, **36**, 7535–7539.

66. I. Wiedemann, E. Breukink, C. van Kraaij, O. P. Kuipers, G. Bierbaum, B. de Kruijff and H. G. Sahl, *J. Biol. Chem.*, 2001, **276**, 1772–1779.