doi:10.2533/chimia.2012.741 Chimia 66 (2012) 741–746

NMR Spectroscopy: An Excellent Tool to Understand RNA and Carbohydrate Recognition by Proteins

Antoine Cléry, Mario Schubert*, and Frédéric H.-T. Allain*

Abstract: Structural biology plays a key role in understanding how networks of protein interactions with their partners are organized at the atomic level. In this review, we show that NMR is a very efficient method to solve 3D structures of protein–RNA and protein–carbohydrate complexes of high quality. We explain the importance of studying such interactions and describe the main steps that are required to determine structures of these types of complexes by NMR. Finally, we show that X-ray crystallography and NMR are complementary methods and briefly report on advantages and disadvantages of each approach.

Keywords: NMR spectroscopy · Protein–carbohydrate complex · Protein–RNA complex · Protein structure · Specific interaction

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the best methods to determine three-dimensional structures of biological molecules at high resolution. The first protein structures determined by NMR spectroscopy were reported in the mid-1980s.[1–3] Several years later, the first structures of protein–DNA[4] and protein–RNA[5] complexes were solved using this method. More recently, structures of proteins bound to carbohydrates emerged, showing that NMR is an important tool to characterize protein interactions with a variety of other biomolecules at the atomic level. Since then, an increasing number of structures have been solved (Fig. 1) revealing crucial information about the mode of recognition of DNA, RNA and carbohydrates by proteins. In this review we describe the key role played by NMR in understanding protein–RNA and protein–carbohydrate interactions and their involvement in biological processes.

2. Importance of Understanding Protein–RNA and Protein–Carbohydrate Recognition

The interest of scientists for RNA started to grow with the discovery of non-protein coding RNAs and the evidence that this molecule was not only an intermediate between DNA and proteins. RNA is a very sophisticated molecule that plays multiple crucial roles in gene expression.[6] RNA is often post-transcriptionally modified (*e.g.* 2'-O-methylation, pseudouridylation, base editing) (Internet resource: *http://library.med.utah.edu/RNAmods/*), can adopt different secondary and tertiary structures or form base-pairs with specific RNA targets (*e.g.* miRNA, piRNA, siRNA, snoRNA or snRNA)[7] and some even have a catalytic activity (*e.g.* ribozyme).[8] In addition, RNA molecules can interact with proteins orchestrating their specific recruitment to the place where they must act in cells.[9] The activity of RNA molecules directly depends on their structure, their post-transcriptional modifications and their specific interaction with other molecules like RNA binding proteins.[10] It is therefore essential to determine the structures of RNA and protein–RNA complexes to fully understand the mode of action of RNA in cells.

In contrast to protein–RNA interactions that play crucial roles inside cells, protein–carbohydrate recognitions are mainly involved in recognition processes on the cell surface. Protein–carbohydrate interactions are central to a variety of biological processes, including cell–cell interactions,[11] functions in the immune system,[12,13] cancer progression and metastasis,[14,15] and host–pathogen interactions.[16,17] However, intracellular functions like protein folding and trafficking are also known.[18]

Due to the large variety of linking monosaccharide building blocks an enormous number of structures can be generated even with a small number of units. There is growing evidence that this variety of glycans is used by nature to store information whose code is called glycocode or sugar code.[19] This code is presented on the surface of proteins and cells and gives information about the type and status of a certain protein or cell. Glycans can decide about the fate of a protein *e.g.* during quality control in the endoplasmic reticulum[18] and are crucial for the distinction between

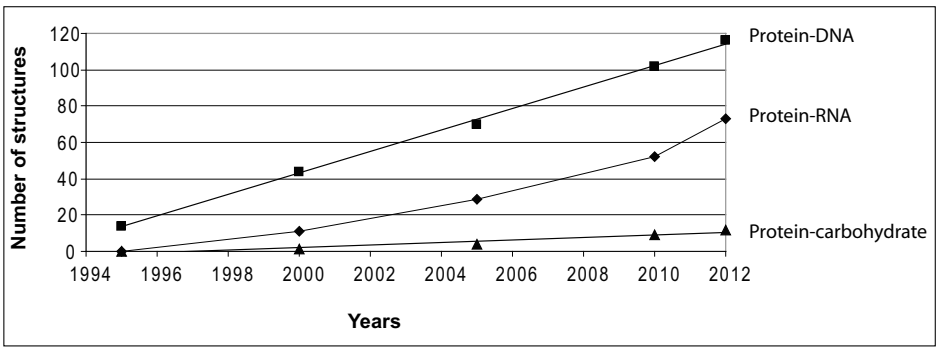

Fig. 1. Evolution of the number of protein–complex structures determined by NMR spectroscopy within the last 15 years. The graph shows the number of protein–DNA (black square), protein–RNA (black diamond) and protein–carbohydrate (black triangle) complex structures deposited in the PDB databank within the last 15 years.

*Correspondence: Prof. Dr. F. H.-T. Allain, Dr. M. Schubert
Institute for Molecular Biology and Biophysics
ETH Zürich
CH-8093 Zürich
Tel.: +41 44 633 3940, +41 44 633 0706
Fax: +41 44 633 1294
E-mail: allain@mol.biol.ethz.ch, schubert@mol.biol.ethz.ch

self and non-self that is central for any immune response.[20] Carbohydrate-binding proteins are the key components in reading the glycocode. Delineating the atomic details of those recognition processes helps us to uncover the glycocode.

The ultimate use of the 3D structure of a complex is to explain the specificity and affinity of a protein domain for its RNA or carbohydrate target. High-quality structures reveal which intermolecular interactions like hydrogen-bonds or methyl-π interactions are responsible for the recognition and sometimes only with the help of a 3D structure one can define a consensus sequences like recently in our group with the discovery of the degenerate motif 5'-(C/G)(C/G)NG-3' (N being any nucleotide) for SRSF2 RRM.[21]

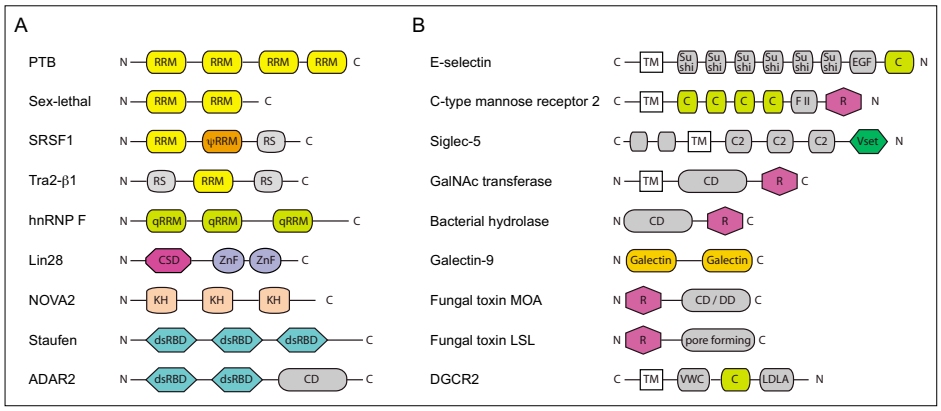

Fig. 2. Domain organization of typical RNA- and carbohydrate-binding proteins (A and B, respectively). Some examples are shown to illustrate the diversity of the domain-architecture of RNA- and carbohydrate-binding proteins. The following abbreviations are used. RRM: RNA recognition motif; ΨRRM: pseudo-RNA recognition motif; qRRM: quasi-RNA recognition motif; RS: arginine-serine rich domain; ZnF: zinc-finger domain; CSD: cold-shock domain; KH: hnRNP K-homology domain; dsRBD: double-stranded RNA binding domain; CD: catalytic domain; TM: transmembrane segment; EGF: epidermal growth factor-like domain; C: C-type lectin domain; FII: fibronectin type 2 domain; Vset: immunoglobulin domain; R: R-type or ricin-type lectin or CBM13 domain; DD: dimerization domain; VWC: von Willibrand factor type C domain; LDLA: low density lipoprotein (LDL) receptor class A domain.

3. Proteins Interacting with RNA and Carbohydrates have typically a Modular Architecture

As illustrated in Fig. 2A, RNA binding proteins (RBPs) typically contain several modular domains spaced by flexible linkers of variable lengths.[22] Small RNA binding domains (RBDs) are crucial for function of these proteins as they dictate their transient or stable interactions with specific RNA targets. The most frequent domains that are found in these proteins are RNA recognition motifs (RRMs), zinc fingers, KH domains and double-stranded RNA binding motifs (dsRBMs). They all have a different fold and use distinct modes of interaction with RNA.[10]

Similarly to RBPs many carbohydrate-binding proteins consist also of multiple domains including individual carbohydrate binding domains (CBDs) as shown in Fig. 2B. The similarity in the modular nature between RNA- and carbohydrate-binding proteins provides a novel perspective on the later. The term 'carbohydrate binding motif (CBM)'[23] is traditionally exclusively used for domains of enzymes that use carbohydrates as substrates. CBMs are categorized in 64 families by the CAZY database,[24] but CBDs of nonenzymatic proteins like lectins are not included. In contrast, the term lectin, standing originally for blood agglutinating proteins, is used today for carbohydrate binding proteins. The domains responsible for carbohydrate binding are sometimes called lectin domains. However, the term lectin is strictly distinguished from enzymes and antibodies that bind also carbohydrates.[25] Such a distinction is not known for protein–RNA interactions, for example dsRBDs are found both in enzymes, *e.g.* ADAR and non-enzymatic proteins like Staufen (Fig. 2A). The same is also true for CBDs, however, a β-trefoil fold domain that recognizes carbohydrates is either categorized as CBM13 family member if found in an enzyme or as a ricin-type (or R-type) lectin domain if part of a lectin (Fig. 2B). We will use the term CBD in the present review that we consider as interchangeable with CBM and lectin domain. Single domain lectins are also known and a particularity of those is their dimerization or tetramerization state that is important for their function as agglutinins.

Proteins that consist of multiple domains linked with flexible linkers are often best approached by studying them as individual domains. This way the specificity, affinity and molecular details of a single binding site can be well characterized. The size of RBD and CBD domains is typically 10–15 kDa, a size range ideally suited for NMR binding studies and 3D structure determination. The affinities of RBDs for RNA cover a wide range (nanomolar to micromolar K_D values) and NMR spectroscopy has demonstrated to be able to determine 3D structures of both weak affinity complexes like SRp20 in complex with 5'-CAUC-3' with a K_D = 18 μM[26] and high affinity complexes like RBMY–S1A RNA with a K_D = 0.6 nM.[27] Interestingly, CBDs display a very similar affinity range for carbohydrates and the few NMR 3D structures of such complexes have K_D values as low as 139 nM[28] and as high as 1.8 mM.[29]

Within the past 15 years, 73 NMR structures of protein–RNA complexes have been solved[30] (Fig. 1). They revealed an unexpected diversity of interactions between these domains and RNA molecules[10] and provided information that allowed a better understanding of RBP functions.[10] The application of NMR to protein–carbohydrate complexes is less established, only 12 structures have been determined and deposited in the PDB database so far, including small hevein domains, ricin-type lectin domains, cyanovirins and the novel fold malectin[28,29,31–40] (Fig. 1). These structures represent less than 0.5% of all deposited protein–carbohydrate complex entries in the PDB illustrating the large potential that NMR spectroscopy could play in the future. The quality of the first structures was often not sufficient to explain the atomic details of specificity for a certain carbohydrate but the most recent structure from our group[38] has a quality comparable to a crystal structure.

4. NMR, A Robust Methodology to Solve Structures of Proteins in Complex with RNA and Carbohydrates

As mentioned above, most of RNA- and carbohydrate-binding proteins contain modular domains that can be studied independently. Typically a single domain or sometimes two domains are used for NMR structure determination staying below the size limit of ~20 kDa required for an efficient methodology. For several domains the structure has already been solved in their free form, indicating that they are soluble and amenable for NMR investigations. Our current methodology for the 3D structure determination of protein–RNA complexes has been described in detail recently.[30]

In the present review, we will focus on the similarities and differences between protein–RNA and protein–carbohydrate structure determination illustrated by two

examples from our laboratory: the structure of Tra2-β1 in complex with the short 5'-AAGAAC-3' RNA sequence[41] and the structure of CCL2 in complex with the trisaccharide GlcNAcβ1,4[Fucα1,3]GlcANcβ.[38] We will cover briefly the most important aspects of the methodology.

To record the multi-dimensional NMR experiments that are needed for protein structure determination, ^{13}C and ^{15}N atoms have to be first incorporated into proteins of interest during their translation in *E. coli*.[30] Optimizing the expression of folded protein in *E. coli* growing on minimal medium is crucial and in most cases successful for RNA-binding domains that are naturally expressed in a reducing environment. In contrast, carbohydrate-binding proteins are often found in an oxidizing environment and expression in *E. coli* can be challenging. However, many CBDs could be successfully expressed in *E. coli* either in the cytoplasm[28,36,37,39] or into inclusion bodies (which then requires an additional refolding protocol *in vitro*).[42]

After the proteins of interest are obtained with isotope labeling, potential binding partners and buffer conditions need to be screened to obtain a protein complex that is suitable for NMR structure determination. For RNA-binding proteins the RNA sequences to be tested first usually come from SELEX[43] and/or CLIP[44] approaches, which provide high-affinity consensus sequences selected in the presence of the protein of interest *in vitro* and *in vivo*, respectively.

The search for the natural target of carbohydrate-binding proteins is fundamentally different, a similar approach to SELEX is missing since carbohydrates are not coded in a linear template that can be randomized at certain positions, neither does a CLIP-like approach exist because there are much fewer aromatic rings at the recognition interface for cross-linking and carbohydrates cannot be amplified. Instead chemically synthesized glycan arrays with several hundred of glycan structures are used to detect binding of a protein of interest that is fluorescently labeled.[45] Finding a good carbohydrate or RNA target for NMR structure determination is not trivial. Several potential targets need to be tested. NMR spectroscopy offers a clear advantage in this regard over other methods since NMR titrations reveal rapidly if a ligand binds and if nice spectra of the complex can be obtained. In the case of the fungal defense lectin CCL2 a glycan array showed that glycans containing either Galβ1,4[Fucα1,3]GlcNAc or GalNAcβ1,4[Fucα1,3]GlcNAc were binding.[38] However, the initial guess that the consensus disaccharide Fucα1,3GlcNAc is sufficient for binding was disproved by

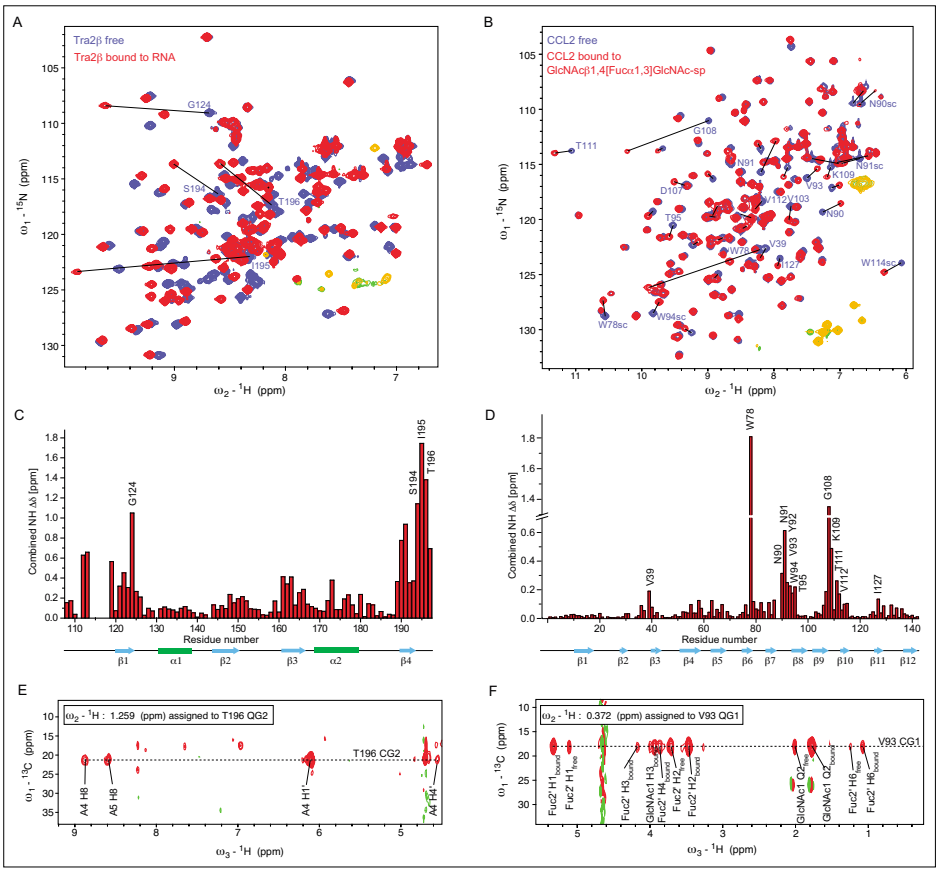

Fig. 3. NMR data revealing the interaction sites between protein and RNA or carbohydrate. (A) Overlay of ^{1}H-^{15}N HSQC spectra of Tra2-β1 protein in its free form (blue) and bound to the RNA 5'-AAGAAC-3' (1:1 ratio) (red). The negative peaks, corresponding to the amides of arginine side chains, in the free and RNA bound states are coloured green and orange, respectively. (B) Overlay of ^{1}H-^{15}N HSQC spectra of the lectin CCL2 in its free form (blue) and bound to its carbohydrate ligand (red). Aliased negative peaks of arginine side chains are coloured green and orange, in the free and carbohydrate bound states, respectively. Representation of the combined chemical shift perturbations ($\Delta\delta = [(\Delta\delta_{HN})^2 + (\Delta\delta_N/5)^2]^{1/2}$) of Tra2-β1 (C) and CCL2 (D) amides upon binding. The secondary structure elements of the domains are displayed at the bottom of each graph. The largest chemical shift perturbations are indicated. (E) Example of intermolecular NOEs observed in a 3D ^{13}C F1-edited F3-filtered HSQC-NOESY spectrum for Tra2-β1 (E) and CCL2 (F). For more experimental details see refs [30,38,41].

NMR titration experiments that did not show evidence for binding. Yet, the trisaccharide Galβ1,4[Fucα1,3]GlcNAc, known as Lewis^X or Le^X, binds with weak affinity (K_D = 0.5 mM) and the trisaccharide GlcNAcβ1,4[Fucα1,3]GlcNAc that was absent on the glycan array binds with high affinity (K_D = 1.4 μM). In other words, the methods mentioned above provide excellent starting points but are rarely sufficient to provide the final sequence of the ligand used in the structure determination of the complex.

NMR is very powerful in screening binding partners and conditions because it can monitor binding interactions on the protein and the ligand side using ^{1}H-^{15}N HSQC fingerprint spectra (Fig. 3A and B) for protein resonances and 2D TOCSY and NOESY correlations at specific RNA chemical shifts. Following NMR titrations using ^{1}H-^{15}N HSQC spectra allows the identification of residues involved in RNA binding. In a ^{1}H-^{15}N HSQC spectrum, each amide of the protein backbone gives a specific ^{15}N-^{1}H cross-peak in the spectrum. Upon RNA binding, amides of residues that are in contact with RNA have a different chemical environment and change therefore to a different location in the spectrum (Fig. 3A,B). These chemical shift deviations between the free and RNA-bound form of the protein can be quantified and mapped for each amide of protein backbone as illustrated in Fig. 3C. In good agreement with the structure of Tra2-β1 RRM bound to RNA, the largest chemical shift perturbations are observed for residues of β-sheet surface and at N- and C-terminal extremities that are involved in RNA interaction. However, large chemical shift perturbations can sometimes result from intra protein-protein interaction induced upon RNA binding and the determination of a 3D structure is still the only way to understand protein-RNA recognition at an atomic level. A very similar chemical shift perturbation plot was obtained for the

protein-carbohydrate complex of CCL2 (Fig. 3D). In both types of complexes, the largest chemical shift deviations were observed for amides forming a direct hydrogen bond to the carbohydrate or the RNA. Monitoring the state of a bound carbohydrate is more complicated since the typical carbohydrate ^{1}H chemical shifts coincide with protein signals and therefore 2D filtered NOESY (see below) and natural abundance ^{1}H-^{13}C HSQC spectra need to be used.

A protein complex suitable for 3D structure determination should give nice line widths of both components and be almost devoid of line broadening or missing signals. The oligomerization state of the protein should be analyzed beforehand since in most cases only monomers give reasonable sharp line widths. In addition an overlay of the ^{1}H-^{15}N HSQC fingerprint spectrum of the individual domain with a comparable spectrum of the full-length protein should make sure that the investigated domain is really independent (peaks fit exactly). A binding stoichiometry of 1:1 is best for 3D structure determination, however 3D complex structures with a stoichiometry of 2:1 and 2:2 were also determined.[46–50] The multivalency of some carbohydrate binding domains like R-type lectins that contain typically three binding sites.[51] stands in contrast to the mainly monovalent RNA-binding domains.

NOE (Nuclear Overhauser Effect) distance restraints can then be extracted from NMR spectra and used for structure calculations.[30] The key technique in obtaining high-precision RNA complex structures is the use of filtering and editing techniques together with a complementary labeling scheme for both components. A 2D ^{13}C filtered-filtered NOESY (in D_2O) of a complex consisting of a ^{13}C labeled protein and unlabeled RNA suppresses all protein signals and results in a 2D NOESY of the unlabeled RNA component.[30] In contrast, a 2D ^{13}C filtered-edited NOESY of the same complex selects for intermolecular NOEs between the protein and the RNA that are of central importance for complex structure determination. Typically 3D ^{13}C edited-filtered NOESY spectra are used to unambiguously assign the observed intermolecular NOE cross-peaks as illustrated in Fig. 3E. The same experiments can be used to obtain intra-carbohydrate and intermolecular NOEs of protein-carbohydrate complexes (Fig. 3F). A reversed labeling scheme can be applied to protein–RNA complexes in which the protein is unlabeled and the RNA $^{13}C/^{15}N$ labeled increasing the amount of unambiguously assigned intermolecular NOEs and thus the quality and precision of the final structure.[30] Whereas labeled RNA are routinely obtained by *in vitro* transcription using labeled NTPs and sometimes chemical synthesis,[52] obtaining isotope labeled carbohydrates is a challenge, since they need to be chemically synthesized from the few available ^{13}C labeled precursors.

For the complex formed between Tra2-β1 RRM and the RNA 5'-AAGAAC-3', an ensemble of 12 structures all consistent with the NMR experimental data was obtained using 93 intermolecular NOEs[41] (Fig. 4A). A particularity of this structure is the involvement of the two regions flanking the RRM for binding RNA. They cross each other and wrap the RNA molecule bound to the β-sheet surface (Fig. 4A). The precision of the structure allowed the identification of hydrogen bonds involved in the interaction, revealing that the protein recognizes specifically the four consecutive nucleotides AGAA[41] (Fig. 4B). The well-defined structure of CCL2 in complex with the trisaccharide GlcNAcβ1,4[Fucα1,3]GlcNAc is shown in Fig. 4C and represents to date the protein–carbohydrate complex determined with most experimental restraints at the interface for such complexes (82 intermolecular NOEs).[38] The carbohydrate binding interface is precisely-defined and the structure allowed the identification of 10 intermolecular hydrogen bonds to all three monosaccharide units (Fig. 4D) together with hydrophobic and methyl-π interactions.

Multiple high-precision structures have already been solved by NMR in the presence of RNA molecules bound to RNA recognition motifs (RRMs), KH domains, zinc fingers and double-stranded RNA binding domains (dsRBD).[10] The methodology is already quite advanced and challenging protein–RNA complexes have

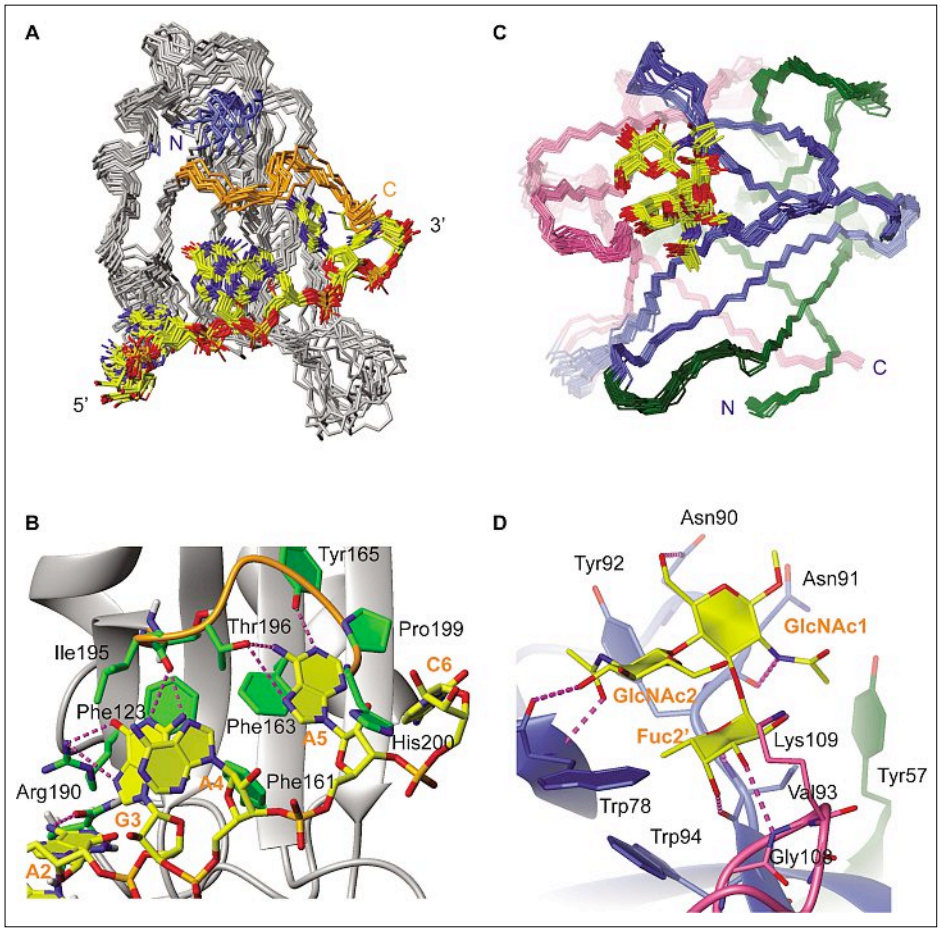

Fig. 4. 3D structures of RNA- and carbohydrate-protein complexes. (A) Overlay of the 12 lowest energy structures of Tra2-β1 bound to the RNA 5'-AAGAAC-3'[41] superimposed on the backbone of the structured protein parts (residues 111 to 201) and heavy atoms of the RNA. The protein backbone is shown in grey and RNA heavy atoms are shown in orange (P atoms), yellow (C atoms), red (O atoms) and blue (N atoms). The N- and C-terminal regions of the RRM are shown in blue and orange, respectively. The N-terminal region of the protein has been truncated in order to not mask the RNA molecule. (B) Molecular recognition of the 5'-AGAA-3' RNA sequence by Tra2-β1 RRM.[41] Intermolecular interactions that are most commonly observed in the structures are shown. Hydrogen bonds are represented by purple dashed lines. (C) Structural ensemble of the 20 best structures of the lectin CCL2 in complex with GlcNAcβ1,4[Fucα1,3]GlcNAcβ (yellow).[38] The three subunits of the pseudo C_3 symmetric β-trefoil fold are colored green, blue and pink. (D) Detailed view of the CCL2 interaction side showing the most representative structure with hydrogen bonds to the trisaccharide.[38] The figures were generated by the program MOLMOL.[76]

already been determined like structures of two domains in complex with RNA,[49,53–57] or a symmetric homo-dimer bound to two RNA stem-loops[46] or to one RNA stem.[48]

5. NMR Spectroscopy is a Competitive Approach for Structure Determination of Protein–RNA and Protein–Carbohydrate Complexes

Several methods have been developed to gain structural information about the mode of interaction of proteins with RNA and carbohydrates. For very large protein complexes (>30 kDa), cryo-electron microscopy and X-ray crystallography are most competitive and almost exclusively applied. However, for smaller systems (<30 kDa) NMR methodology is highly competitive. Here, advantages and limits of solution NMR compared to X-ray crystallography are discussed.

One strength of NMR is the unambiguous localization of binding sites for RNA or carbohydrates on the protein surface. Intermolecular NOE cross-peaks link directly hydrogen atoms of the protein surface with hydrogens of the RNA/carbohydrate that are close in space (<6 Å). Extracting equivalent information from crystal structures is sometimes not trivial. Not so uncommonly, larger oligosaccharides are found to be in contact with two or more neighboring protein molecules[58–63] raising the question which of the interactions is the biological relevant one. Crystal structures obtained by soaking mono- or dissaccharides into protein crystals have the disadvantage that only the region not involved in crystal packing is available for interactions.[64] Those structures should be interpreted with care since the binding surface might be different than the one in solution. Complex structures with different monosaccharides that are bound with many water-mediated hydrogen bonds might just reflect a rather unspecific binding of monosaccharides as in the case of *Sambucus nigra* agglutinin II (SNA-II) that was crystallized with the rather weakly binding ligands galactose, GalNAc, lactose, fucose and xylose displaying dissociation constants between 18 and 680 μM.[65] The unique ability of NMR to localize interaction sites even in larger systems is illustrated using the adhesin TgMIC4 from the parasite *Toxoplasma gondii*. Marchant *et al*. demonstrated that out of the six apple domains only domain 5 interacts with Galβ1,3GlcNAc and they presented an NMR-based HADDOCK model of this interaction.[37]

The crystallization of proteins in the presence of RNA or carbohydrates sometimes fails because of flexible extensions of the ligands that adopt multiple conformations or due to steric hindrance of the ligand that prevents crystal packing. In some cases the protein crystallizes without the binding partner. In addition, NMR structures are less prone to artificial interactions that can occur due to packing in the context of a crystal lattice allowing native domain arrangements to be studied. An example that illustrates well this difference is the recent structure of U2AF65 bound to RNA, which was solved by NMR[53] and X-ray crystallography.[66] The arrangement of the two RRMs in the crystal structure was different and inconsistent with NMR data, which were based on two independent and robust approaches (RDC and PRE).[53,66] This difference comes most likely from crystal packing forces and/or alteration of the linker between the two RRMs, which had to be shortened to enable crystallization.[66]

Another example is a complex of the lectin cyanovirin-N from the cyanobacterium *Nostoc ellipsosporum* that displays antiviral activity, especially against HIV. The carbohydrate interaction of cyanovirin-N was first studied by NMR spectroscopy that revealed two binding sites on the monomeric protein for Manα1,2Manα with K_D values of 140 nM and 1.4 μM.[28] The NMR complex structure was determined with both binding sites occupied. Especially the high affinity site revealed hydrogen bonds to both Man residues that indicated how the two α1,2-linked mannoses are recognized. In contrast a subsequent crystal structure of a complex with $Man_9GlcNAc_2$ revealed a swapped dimer structure with a buffer molecule from the crystallization solution in the high-affinity binding site and clear electron density for a Manα1,2Manα1,2Manα trissaccharide in the second binding site, which was however, slightly distorted by the domain-swapping compared to the solution structure.[67] An NMR investigation of this rather confusing finding revealed that the swapped dimer structure is a kinetically trapped, meta-stable structure that can also exist in solution but is converted to the monomer at slightly elevated temperatures (38 °C).[68]

Finding immediately the correct orientation of a carbohydrate in the binding site is another advantage of NMR spectroscopy. Intermolecular NOE cross-peaks link directly hydrogen atoms of the ligand to the protein surface whereas modeling small oligosaccharides into electron density is more challenging. In some cases this can lead to erroneous linkages being built.[69,70] The lack of electron density for carbohydrates is another complication for X-ray that occurred for example in the case of the norovirus protruding domain, co-crystallized with the histo-blood group antigens Le^A and Le^X-trisaccharides so that the carbohydrate could not be modeled.[71]

In addition, NMR offers the possibility to get information about the flexibility and dynamic of molecules upon complex formation[30] under almost physiological conditions. This is another advantage of NMR spectroscopy that crystallography can not reveal. This type of studies are important as it was previously shown that dynamics of both the RNA[72] and the protein[73] can play a crucial role for RNA–protein recognition. One example for a carbohydrate binding domain is the study of galectin-3 that revealed conformational entropy changes of the lectin upon binding of lactose.[74]

However, there are also limitations of NMR spectroscopy in studying protein complexes by NMR. NMR is still limited by the size of the studied system. Complexes smaller than 20 kDa can be solved using standard NMR experiments, but in the presence of larger systems, NMR spectra become more crowded and the signal to noise ratio tends to decrease making chemical shift assignment more difficult and therefore the structure determination more uncertain. However, measurements at higher temperature and more advanced techniques like protein deuteration and TROSY NMR experiments enabled the structure determination of a few protein-RNA complexes over 35 kDa.[47,54,75]

Another limitation is certainly the requirement of isotope labeling of the protein with ^{15}N and ^{13}C, which has to be close to 100% so that the central filtered-edited NOESY experiments work properly.[30] Signal overlap of the unlabeled ligand, especially in the case of carbohydrates is another limiting factor. This can be partially overcome by using the highest magnetic fields available, *e.g.* a 900 MHz spectrometer, and by testing a range of different buffer conditions and temperatures.

6. Conclusion and Perspectives

In this review, we showed that X-ray and NMR are two complementary methods that bring information at the atomic level for protein–ligand interactions. NMR is a very powerful and efficient method to determine structures of protein–RNA and protein–carbohydrate complexes under physiological conditions as long as their molecular size is below 30 kDa. Above this molecular weight, X-ray crystallography is currently the best method for high resolution structure determination. However, an increasing number of strategies become available to extend the NMR size limit, opening the door to structure determination of large complexes in solution.

One of the next challenges in these area of structural biology will consist of study-

ing how multiple factors assemble on a single RNA molecule to understand how such assemblies are coordinated. NMR is among the most promising approaches to address these questions *in vitro* but also possibly *in vivo*. NMR spectroscopy is also a very promising technique to help reveal the glycocode by the structure determination of more protein-carbohydrate complexes. Since the glycocode is part of the natural language that cells use for communication, understanding this code will certainly help developing novel strategies against diseases, associated with such molecular interactions. An argument that is also valid for the protein–RNA recognition code whose perturbation is at the origin of numerous genetic diseases.

Acknowledgements

The authors would like to thank the Swiss National Science Foundation (No. 3100A0-118118), the SNF sinergia grant CRSII3_127333 and the SNF-NCCR Structural Biology for financial support to FHTA.

Received: August 9, 2012

[1] M. P. Williamson, T. F. Havel, K. Wüthrich, *J. Mol. Biol.* **1985**, *182*, 295.
[2] A. S. Arseniev, V. I. Kondakov, V. N. Maiorov, V. F. Bystrov, *FEBS Lett.* **1984**, *165*, 57.
[3] R. Kaptein, E. R. Zuiderweg, R. M. Scheek, R. Boelens, W. F. van Gunsteren, *J. Mol. Biol.* **1985**, *182*, 179.
[4] J. G. Omichinski, G. M. Clore, M. Robien, K. Sakaguchi, E. Appella, A. M. Gronenborn, *Biochemistry* **1992**, *31*, 3907.
[5] F. H. Allain, C. C. Gubser, P. W. Howe, K. Nagai, D. Neuhaus, G. Varani, *Nature* **1996**, *380*, 646.
[6] L. W. Barrett, S. Fletcher, S. D. Wilton, *Cell. Mol. Life Sci.* **2012**, 69.
[7] M. Blignaut, *Epigenetics* **2012**, *7*, 664.
[8] C. Hammann, A. Luptak, J. Perreault, M. de la Pena, *RNA* **2012**, *18*, 871.
[9] M. Chen, J. L. Manley, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 741.
[10] A. Clery, F. H. Allain, 'From Structure to Function of RNA Binding Domains', Landes Bioscience, Austin, **2012**, p. 137.
[11] M. E. Taylor, K. Drickamer, *Curr. Opin. Cell. Biol.* **2007**, *19*, 572.
[12] J. D. Marth, P. K. Grewal, *Nat. Rev. Immunol.* **2008**, *8*, 874.
[13] M. Sperandio, C. A. Gleissner, K. Ley, *Immunol. Rev.* **2009**, *230*, 97.
[14] Y. Y. Zhao, M. Takahashi, J. G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume, N. Taniguchi, *Cancer Sci*. **2008**, *99*, 1304.
[15] K. S. Lau, J. W. Dennis, *Glycobiology* **2008**, *18*, 750.
[16] F. Lehmann, E. Tiralongo, J. Tiralongo, *Cell. Mol. Life Sci.* **2006**, *63*, 1331.
[17] G. R. Vasta, *Nat. Rev. Microbiol.* **2009**, *7*, 424.
[18] M. Aebi, R. Bernasconi, S. Clerc, M. Molinari, *Trends Biochem. Sci.* **2010**, *35*, 74.
[19] H. J. Gabius, S. Andre, J. Jimenez-Barbero, A. Romero, D. Solis, *Trends Biochem. Sci.* **2011**, *36*, 298.
[20] G. A. Rabinovich, Y. van Kooyk, B. A. Cobb, *Ann. NY Acad. Sci.* **2012**, *1253*, 1.
[21] G. M. Daubner, A. Clery, S. Jayne, J. Stevenin, F. H. Allain, *EMBO J.* **2011**, *31*, 162.
[22] B. M. Lunde, C. Moore, G. Varani, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 479.
[23] A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, *Biochem. J.* **2004**, *382*, 769.
[24] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, *Nucleic Acids Res*. **2009**, *37*, D233.
[25] H. J. Gabius, H. C. Siebert, S. Andre, J. Jimenez-Barbero, H. Rudiger, *Chembiochem* **2004**, *5*, 740.
[26] Y. Hargous, G. M. Hautbergue, A. M. Tintaru, L. Skrisovska, A. P. Golovanov, J. Stevenin, L. Y. Lian, S. A. Wilson, F. H. Allain, *EMBO J.* **2006**, *25*, 5126.
[27] L. Skrisovska, C. F. Bourgeois, R. Stefl, S. N. Grellscheid, L. Kister, P. Wenter, D. J. Elliott, J. Stevenin, F. H. Allain, *EMBO Rep.* **2007**, *8*, 372.
[28] C. A. Bewley, *Structure* **2001**, *9*, 931.
[29] I. Vakonakis, T. Langenhan, S. Promel, A. Russ, I. D. Campbell, *Structure* **2008**, *16*, 944.
[30] C. Dominguez, M. Schubert, O. Duss, S. Ravindranathan, F. H. Allain, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 1.
[31] K. Sorimachi, M. F. Le Gal-Coeffet, G. Williamson, D. B. Archer, M. P. Williamson, *Structure* **1997**, *5*, 647.
[32] M. I. Chavez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, G. Asensio, M. Muraki, F. J. Canada, J. Jimenez-Barbero, *Chemistry* **2005**, *11*, 7060.
[33] N. Aboitiz, M. Vila-Perello, P. Groves, J. L. Asensio, D. Andreu, F. J. Canada, J. Jimenez-Barbero, *Chembiochem* **2004**, *5*, 1245.
[34] A. Canales, R. Lozano, B. Lopez-Mendez, J. Angulo, R. Ojeda, P. M. Nieto, M. Martin-Lomas, G. Gimenez-Gallego, J. Jimenez-Barbero, *FEBS J.* **2006**, *273*, 4716.
[35] S. M. Kumar, H. M. Wang, S. K. Mohan, R. H. Chou, C. Yu, *Biochemistry* **2010**, *49*, 10756.
[36] 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, C. Muhle-Goll, *Mol. Biol. Cell* **2008**, *19*, 3404.
[37] J. Marchant, B. Cowper, Y. Liu, L. Lai, C. Pinzan, J. B. Marq, N. Friedrich, K. Sawmynaden, L. Liew, W. Chai, R. A. Childs, S. Saouros, P. Simpson, M. C. Roque Barreira, T. Feizi, D. Soldati-Favre, S. Matthews, *J. Biol. Chem.* **2012**, *287*, 16720.
[38] M. Schubert, S. Bleuler-Martinez, A. Butschi, M. A. Walti, P. Egloff, K. Stutz, S. Yan, I. B. Wilson, M. O. Hengartner, M. Aebi, F. H. Allain, M. Kunzler, *PLoS Pathog* **2012**, *8*, e1002706.
[39] S. Shahzad-ul-Hussan, E. Gustchina, R. Ghirlando, G. M. Clore, C. A. Bewley, *J. Biol. Chem*. **2011**, *286*, 20788.
[40] T. Schallus, K. Feher, U. Sternberg, V. Rybin, C. Muhle-Goll, *Glycobiology* **2010**, *20*, 1010.
[41] A. Clery, S. Jayne, N. Benderska, C. Dominguez, S. Stamm, F. H. Allain, *Nat. Struct. Mol. Biol.* **2011**, *18*, 443.
[42] N. Dimasi, L. Moretta, R. Biassoni, R. A. Mariuzza, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59*, 1856.
[43] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505.
[44] J. Ule, K. Jensen, A. Mele, R. B. Darnell, *Methods* **2005**, *37*, 376.
[45] O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C. H. Wong, J. C. Paulson, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17033.
[46] M. Schubert, K. Lapouge, O. Duss, F. C. Oberstrass, I. Jelesarov, D. Haas, F. H. Allain, *Nat. Struct. Mol. Biol.* **2007**, *14*, 807.
[47] L. Varani, S. I. Gunderson, I. W. Mattaj, L. E. Kay, D. Neuhaus, G. Varani, *Nat. Struct. Biol.* **2000**, *7*, 329.
[48] Y. Yang, N. Declerck, X. Manival, S. Aymerich, M. Kochoyan, *EMBO J.* **2002**, *21*, 1987.
[49] F. C. Oberstrass, S. D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D. L. Black, F. H. Allain, *Science* **2005**, *309*, 2054.
[50] S. D. Auweter, F. C. Oberstrass, F. H. Allain, *J. Mol. Biol.* **2007**, *367*, 174.
[51] E. M. Grahn, H. C. Winter, H. Tateno, I. J. Goldstein, U. Krengel, *J. Mol. Biol.* **2009**, *390*, 457.
[52] P. Wenter, L. Reymond, S. D. Auweter, F. H. Allain, S. Pitsch, *Nucleic Acids Res.* **2006**, *34*, e79.
[53] C. D. Mackereth, T. Madl, S. Bonnal, B. Simon, K. Zanier, A. Gasch, V. Rybin, J. Valcarcel, M. Sattler, *Nature* **2011**, *475*, 408.
[54] R. Stefl, F. C. Oberstrass, J. L. Hood, M. Jourdan, M. Zimmermann, L. Skrisovska, C. Maris, L. Peng, C. Hofr, R. B. Emeson, F. H. Allain, *Cell* **2010**, *143,* 225.
[55] F. H. Allain, P. Bouvet, T. Dieckmann, J. Feigon, *Embo J.* **2000**, *19*, 6870.
[56] J. M. Perez-Canadillas, *Embo J.* **2006**, *25*, 3167.
[57] B. M. Lee, J. Xu, B. K. Clarkson, M. A. Martinez-Yamout, H. J. Dyson, D. A. Case, J. M. Gottesfeld, P. E. Wright, *J. Mol. Biol.* **2006**, *357*, 275.
[58] S. Perret, C. Sabin, C. Dumon, M. Pokorna, C. Gautier, O. Galanina, S. Ilia, N. Bovin, M. Nicaise, M. Desmadril, N. Gilboa-Garber, M. Wimmerova, E. P. Mitchell, A. Imberty, *Biochem. J.* **2005**, *389*, 325.
[59] C. S. Wright, G. Hester, *Structure* **1996**, *4*, 1339.
[60] 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*, 32274.
[61] A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger, W. A. Hendrickson, *Nature* **1998**, *393*, 812.
[62] S. Swaminathan, S. Eswaramoorthy, *Nat. Struct. Biol.* **2000**, *7*, 693.
[63] V. Notenboom, A. B. Boraston, S. J. Williams, D. G. Kilburn, D. R. Rose, *Biochemistry* **2002**, *41*, 4246.
[64] P. Emsley, C. Fotinou, I. Black, N. F. Fairweather, I. G. Charles, C. Watts, E. Hewitt, N. W. Isaacs, *J. Biol. Chem.* **2000**, *275*, 8889.
[65] L. Maveyraud, H. Niwa, V. Guillet, D. I. Svergun, P. V. Konarev, R. A. Palmer, W. J. Peumans, P. Rouge, E. J. Van Damme, C. D. Reynolds, L. Mourey, *Proteins* **2009**, *75*, 89.
[66] E. A. Sickmier, K. E. Frato, H. Shen, S. R. Paranawithana, M. R. Green, C. L. Kielkopf, *Mol. Cell* **2006**, *23*, 49.
[67] I. Botos, B. R. O'Keefe, S. R. Shenoy, L. K. Cartner, D. M. Ratner, P. H. Seeberger, M. R. Boyd, A. Wlodawer, *J. Biol. Chem.* **2002**, *277*, 34336.
[68] L. G. Barrientos, J. M. Louis, I. Botos, T. Mori, Z. Han, B. R. O'Keefe, M. R. Boyd, A. Wlodawer, A. M. Gronenborn, *Structure* **2002**, *10*, 673.
[69] M. Crispin, D. I. Stuart, E. Y. Jones, *Nat. Struct. Mol. Biol.* **2007**, *14*, 354.
[70] H. M. Berman, K. Henrick, H. Nakamura, J. Markley, *Nat. Struct. Mol. Biol.* **2007**, *14*, 354.
[71] G. S. Hansman, C. Biertumpfel, I. Georgiev, J. S. McLellan, L. Chen, T. Zhou, K. Katayama, P. D. Kwong, *J. Virol.* **2011**, *85*, 6687.
[72] F. C. Oberstrass, F. H. Allain, S. Ravindranathan, *J. Am. Chem. Soc.* **2008**, *130*, 12007.
[73] S. Ravindranathan, F. C. Oberstrass, F. H. Allain, *J. Mol. Biol.* **2010**, *396*, 732.
[74] C. Diehl, S. Genheden, K. Modig, U. Ryde, M. Akke, *J. Biomol. NMR* **2009**, *45*, 157.
[75] V. D'Souza, M. F. Summers, *Nature* **2004**, *431*, 586.
[76] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graph.* **1996**, *14*, 51.