Combining Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) Spectroscopy for Integrative Structural Biology of Protein–RNA Complexes

Alexander Leitner,¹ Georg Dorn,² and Frédéric H.-T. Allain²

¹Institute of Molecular Systems Biology, Department of Biology, Eidgenössische Technische Hochschule (ETH) Zürich, 8093 Zürich, Switzerland

²Institute of Molecular Biology and Biophysics, Department of Biology, ETH Zürich, 8093 Zürich, Switzerland

Correspondence: allain@mol.biol.ethz.ch; leitner@imsb.biol.ethz.ch

SUMMARY

Deciphering complex RNA–protein interactions on a (near-)atomic level is a hurdle that hinders advancing our understanding of fundamental processes in RNA metabolism and RNA-based gene regulation. To overcome challenges associated with individual structure determination methods, structural information derived from complementary biophysical methods can be combined in integrative structural biology approaches. Here, we review recent advances in such hybrid structural approaches with a focus on combining mass spectrometric analysis of cross-linked protein–RNA complexes and nuclear magnetic resonance (NMR) spectroscopy.

Outline

- 1 Introduction
 - 2 MS in structural biology
 - 3 Applications of MS to study RNA and protein–RNA interactions
 - 4 NMR spectroscopy of large protein–RNA complexes: The “divide and conquer” approach and its limitations
 - 5 CLIR–MS/MS as a technique for high-resolution mapping of interactions in protein–RNA complexes
 - 6 Integrative modeling using CLIR–MS/MS-derived distance restraints and structural information derived from NMR and/or available structural information from the PDB
- References

1 INTRODUCTION

1.1 Toward Integrative Structural Biology of Protein–RNA Complexes

With the increasingly recognized role of RNA in gene regulation, it has become clear that defining RNA–protein interactions at the atomic level is of utmost importance for understanding the underlying molecular mechanisms and finding therapies against diseases associated with gene regulation. Yet, solving structures of protein–RNA complexes remains a challenge as few structures of RNA-binding proteins (RBPs) bound to their cognate RNA have been solved in comparison to their abundance (5%–10% of the genome; Castello et al. 2012). High-throughput methodologies in RNA sequencing and mass spectrometry (MS) have recently allowed the description of the genome-wide positioning of individual RBPs (sometimes at the nucleotide resolution) and the identification of the RBPome (and the domains involved), respectively. Nevertheless, it remains largely elusive how RBPs assemble onto an RNA (pre-messenger RNA [mRNA], mRNA, noncoding RNA, etc.), shape its structure, and control its fate. The large gap between large-scale MS analysis and RBP structure is illustrated by the recent structures of the spliceosome in which one would not anticipate such structural complexity solely on the basis of the identification of components.

With the structures of the ribosome, the nuclear pore complex, and now of the spliceosome well characterized even for the most evolved species, the current challenge in the ribonucleoprotein (RNP) field is certainly to understand the regulatory mechanism controlling splicing, translation, decay, mRNA export, and transport. To this end, relying on a single methodology to investigate the structure of these regulatory RNPs is probably not a good avenue considering the expected complexity, dynamics, and versatility of these complexes. For several decades, structural biology relied primarily on two biophysical methods: X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. The recent “resolution revolution” in cryo-electron microscopy (cryo-EM; Kühlbrandt 2014), the emergence of MS, electron paramagnetic resonance (EPR), Förster resonance energy transfer (FRET), small-angle X-ray scattering (SAXS), and small-angle neutron scattering (SANS) for structural applications and ever-increasing computational power have changed the landscape of biophysical methods that can be used in structural biology. Combining multiple biophysical methods in structural biology (so-called “integrative structural biology” or “hybrid approaches in structural biology”; Ward et al. 2013) is now clearly emerging and is likely to become the norm rather than the exception. A few examples of such hybrid approaches used for RNPs will be succinctly described to

illustrate the need to combine multiple structural biology technologies.

There are multiple examples of combining X-ray crystallography and NMR spectroscopy in the literature, but a recent application that comes to mind is work from the Sprangers laboratory on decapping enzymes (Wurm et al. 2017), in which NMR spectroscopy measurements of the enzyme allowed clarification of which crystal structures were functionally relevant and which ones were not. When crystal packing makes the obvious interpretation of a crystal structure doubtful, it should become the norm to confirm the functionality of the structure using solution methods like NMR, SAXS, or SANS. Electron microscopy (EM) and crystallography have been long-term partners, in which—in earlier times—low-resolution 3D EM maps required crystal structures to interpret them—for example, to place the crystal structure of the Sm ring (Kambach et al. 1999) into the U1 small nuclear ribonucleoprotein complex (snRNP) low-resolution EM map (Stark et al. 2001). More recently, despite the EM revolution, several available crystal structures were used to interpret the 6 Å cryo-EM map of the tri-snRNP U4/U6.U5 (Nguyen et al. 2015), illustrating the need for high-resolution crystal structures of individual proteins or complexes to interpret cryo-EM maps at medium resolution. Cross-linking coupled to MS (XL-MS) has also been of great help for interpreting cryo-EM maps, in particular if the electron density is ambiguous despite high resolution and crystal structures of the fragments are not available. XL-MS data greatly improve the confidence of 3D model building. This is exemplified by structures of the mitochondrial ribosome in which MS data allowed the unambiguous placement of closely positioned ribosomal proteins into the ribosome 3D structure (Greber et al. 2014). Similarly, the combination of NMR and SAXS/SANS has been used in a number of RNP structural studies by the Sattler and Carlomagno groups. SAXS and SANS provide information about the 3D envelope, which can be used to restrain conformational space and help improve the precision of the structural ensemble by distinguishing conformations that fit NMR restraints equally well, as seen for the structure determination of U2AF bound to RNA (Huang et al. 2014). SANS and NMR were used to decipher a structural model of the box C/D RNA methylation enzyme bound to its RNA’s substrate (Lapinaite et al. 2013). In that work, SANS enabled the protein shape to be differentiated from the RNA shape in the complex and helped in the modeling of a 3D structure. NMR probes of high-molecular weight (methyl groups in a deuterated environment) allowed probing conformational changes in the protein on complex formation and supported 3D model building. In addition, NMR spectroscopy allowed enzyme catalysis to be followed by observing ^{13}C 2'-O-methyl groups in the RNA.

We also needed to combine NMR spectroscopy with other biophysical methods in some of our recent structural studies, especially when studying fairly large RNPs in the range of 50–80 kDa. Our first hybrid approach was to combine NMR and EPR spectroscopy when studying the 70-kDa complex of the protein ribosomal RNA small subunit methyltransferase E (RsmE) bound to the noncoding RNA RsmZ. RsmZ contains five stem loops (SLs) and can bind cooperatively up to five RsmE dimers (Duss et al. 2014a). Solving the solution structure of RsmZ bound to three RsmE dimers revealed two conformations for the complex (of almost equal population) that resulted from two different assembly pathways of the RNP. EPR data were essential to provide 21 long-range distances within the RNA, thereby revealing the existence of these two conformations. More recently, we also combined NMR, EPR, and single-molecule FRET to decipher the structure of the two double-stranded RNA-binding domains (dsRBDs) of human immunodeficiency virus transactivating response RNA-binding protein (TRBP) bound to a small interfering RNA (siRNA; Masliah et al. 2018). This work also revealed the presence of two different conformations for this complex, which could be unambiguously confirmed by EPR. With single-molecule FRET, the relative population of the two forms could be determined. We have also successfully combined NMR with long molecular dynamics simulations in a number of studies to understand hydration at protein–RNA interfaces (Krepl et al. 2017) and the role of aromatic side-chains in RNA recognition motif (RRM)–RNA recognition (Diarra Dit Konté et al. 2017). Last, but not least, we recently developed a novel approach that combines NMR spectroscopy and XL-MS and allows rapid generation of models for protein–RNA complexes at atomic resolution (Dorn et al. 2017). We believe that this new hybrid structural biology approach has very high potential in terms of speed and sensitivity. More details are given below.

2 MS IN STRUCTURAL BIOLOGY

MS is a method essential for the large-scale identification and quantification of diverse groups of biomolecules, including proteins and metabolites. For those classes of molecules that cannot be amplified before detection, it provides crucial benefits such as high sensitivity and sample throughput. This way, MS has emerged as the key enabling technology for proteomics and metabolomics research. In contrast, MS is not the method of choice for RNA profiling because techniques exist that are faster, more robust, and better suited for automation—although RNA can also be detected and sequenced by MS. In the last decade, MS has increasingly been applied to provide low-resolution structural information on proteins and protein complexes,

including RNPs. Because MS may not be widely known as a structural biology tool, we focus in this section on the fundamentals of biological MS and introduce the concepts of structural MS methods.

2.1 Fundamentals of Biological MS

Ultimately, any mass spectrometer determines the mass-to-charge (m/z) ratio of ions in the gas phase and this information is used to infer the identity of compound(s) of interest. For this purpose, molecules such as peptides, proteins or RNA need to be transferred from a condensed phase (usually a liquid) into the gas phase and be ionized during the process. Nowadays, electrospray ionization (ESI) is the most commonly used ionization technique for large biomolecules (Wilm 2011). ESI generates ions through the formation of charged droplets in the presence of strong electric fields (several kV/cm), and free ions are formed from these droplets after droplet fission events and evaporation of the surrounding solvent. By virtue of their chemical nature, proteins and peptides are more easily converted into cations (protonation of basic residues), whereas oligonucleotides are best ionized as anions (deprotonation of phosphate groups).

The ionization process takes place at atmospheric pressure, and the generated ions need to be transferred into (ultra)high vacuum regions inside the mass spectrometer, where their m/z ratio can be determined. Detection is based on different physical principles, such as the frequency of the ions' movement in trapping devices (such as the Orbitrap analyzer; Eliuk and Makarov 2015) or the flight time of ions (time-of-flight analyzer). Different types of analyzers can be integrated in one instrument, and most mass spectrometers are also equipped with devices (e.g., collision cells) to fragment ions for further structure elucidation (tandem mass spectrometry or MS/MS). This is most commonly performed by collision with inert gas molecules. Bond cleavages between the monomeric units in proteins/peptide (Paizs and Suhai 2005) and in oligonucleotides (Schürch 2016) occur according to predictable rules, so that the sequence can be determined from MS/MS spectra (most often, software-assisted or completely automated).

2.2 Mass Spectrometric Workflows

Depending on the application and the complexity of the sample, different types of chromatographic separation may be performed off-line or in-line with MS detection. Alternatively, a (highly purified) sample is analyzed directly. For an overview of these workflows, see Figure 1.

The most commonly used workflow for high-throughput identification of proteins can be used to characterize a wide variety of sample types (e.g., cell lysates or crudely

Figure 1. Characterization of protein–RNA interactions by mass spectrometry (MS). (A) Typical steps of a liquid chromatography (LC)–MS workflow. Proteins are digested into peptides, the resulting peptides are separated by LC and ionized and their mass-to-charge ratio (m/z) is determined. A subset of peptides is selected for fragmentation, allowing the identity of the peptides, possible modifications, and/or cross-linking (XL) sites to be derived through bioinformatic analysis. Samples can be of different complexity and can be prepared with or without XL steps. (B) Four different MS strategies for profiling protein–RNA interactions. From *top to bottom*: RNA-binding proteins (RBPs) may be identified from complex samples using XL with ultraviolet (UV) light (reflected by the orange star), and subsequent affinity purification. Upon enzymatic digestion, typically only peptides (visualized by horizontal bars on the *right*) that do not contain the RNA-binding site are identified. Samples of lower complexity such as purified complexes can be analyzed using structural MS methods, such as direct analysis by native/nondenaturing MS or protein–protein and protein–RNA XL. From such experiments, crucial information including complex mass, stoichiometry, spatial proximity of subunits, and even precise interaction sites can be obtained.

purified complexes isolated by [immuno]affinity enrichment). This way, RBPs may be profiled (see Sec. 3.2). Here, isolated proteins are first digested into peptides with a specific protease such as trypsin, which cleaves on the carboxy-terminal side of arginine (Arg) and lysine (Lys) residues and generates peptides of suitable length (typically 5–30 residues) for MS. The resulting peptide mixture is chromatographically separated before MS/MS analysis. Mass spectra are analyzed by dedicated software that assigns peptide sequences, and proteins are inferred from the identified peptide pool (Fig. 1A). This so-called “bottom-up” workflow—in which peptides are used as a proxy for the proteins from which they originate—is common to most proteomic methods, including those that provide structural information (see Sec. 2.3). Alternatively, proteins may be directly analyzed by MS without a digestion step (“top-down”), although this is less commonly performed because of

challenges related to protein separation, fragmentation, and bioinformatic analysis.

2.3 Structural MS Methods

To connect MS analysis to structural biology, two different approaches are used. In the first approach, isolated biomolecules such as protein complexes, RNA, or RNPs are directly introduced into the mass spectrometer under near “native” or rather nondenaturing conditions (Leney and Heck 2017). Although the transition from liquid to gas phase and the energetic activation of molecules by the ionization step can lead to some artifacts, under carefully controlled conditions many ternary and quaternary structure elements are known to be preserved. This way, “native MS” may inform about the molecular mass of an intact assembly, the stoichiometry of its constituent subunits,

and—by means of gas-phase dissociation—the composition of subcomplexes.

Most often, however, structural MS methods follow the bottom-up workflow outlined above (Fig. 1). In this case, structural information needs to be “encoded” in the sample before digestion (Leitner 2016a). This is usually performed by one of the following strategies: (1) exchange of labile hydrogens with deuterium; (2) covalent labeling with short-lived radicals (“footprinting”), such as hydroxyl radicals, or other labeling reagents; and (3) XL approaches. All of the above approaches result in a defined mass shift that can be detected in the mass spectrometer, but they provide orthogonal information about the sample. Hydrogen-deuterium (H/D) exchange mainly informs on solvent accessibility and hydrogen bonding, as well as changes in these properties if different samples are combined. Covalent labeling patterns also reflect solvent/surface exposure, whereas XL provides spatial proximity and (low-resolution) distance information. Such information makes XL the most practically useful method among the bottom-up structural MS methods, because it can be most easily integrated into hybrid structural biology workflows (Faini et al. 2016).

2.4 Protein–Protein and Protein–RNA XL

XL-MS can be applied in different ways to study the structure of protein and protein–RNA complexes (Leitner 2016b). Cross-links within or between proteins are most commonly formed using specific XL reagents such as succinimide esters that connect amino groups (Lys, amino terminus). Less frequently, XL approaches that connect carboxyl groups (aspartic acid [Asp], glutamic acid [Glu], carboxyl terminus) with each other or amines with carboxyl groups are used. Cross-linked samples are processed in a manner similar to that described above (proteolysis, chromatographic separation, MS/MS analysis). However, data analysis is more complex, because both of the peptides that become connected to each other by the cross-link need to be identified.

In contrast to the solution chemistry used for protein–protein XL, photochemistry is the preferred approach for forming cross-links between protein and RNA (Schmidt et al. 2012). UV irradiation at 254 nm leads to the formation of covalent links between nucleotides and amino acids, whereby uracil is most reactive on the RNA side, while most amino acids are found to be reactive. The yield of UV XL is fairly low but can be increased by incorporating nonnatural nucleotides such as 4-thio-uracil (4-sU) or 5-iodo-uracil, which react more efficiently at 365 nm. In all cases, both the protein and the RNA chain(s) need to be enzymatically cleaved before MS analysis. The size/length

of the cleavage product on the protein side is again controlled by use of specific proteases such as trypsin, while a nuclease with comparable cleavage properties is not available. Therefore, a cocktail of different nucleases is commonly used to cut the oligonucleotide into remnants of one to four nucleotides. Upon MS/MS analysis, it is relatively straightforward to assign the corresponding peptide and thus protein, but the sequence context on the RNA side is typically lost because the short nucleotide sequence does not allow unequivocal identification of the cross-linked region. To overcome this limitation, we have recently introduced a new approach that is described in more detail in Section 5.

3 APPLICATIONS OF MS TO STUDY RNA AND PROTEIN–RNA INTERACTIONS

Using the experimental strategies outlined in Section 2 above and summarized in Figure 1B, MS has generated many insights into the structure and function of protein–RNA assemblies, as we illustrate below.

3.1 Direct Analysis of RNA by MS

The application of MS to the direct analysis of RNA has remained rather limited, although fundamental work has examined higher-order structures of RNA such as G-quadruplex formation (reviewed in Abi-Ghanem and Gabelica 2014; Schürch 2016). In addition, MS lends itself to the identification of natural and artificial/chemical RNA modifications (Gaston and Limbach 2014; Heiss and Kellner 2017; Limbach and Paulines 2017). Nonetheless, at the moment, these applications remain restricted to a few expert laboratories.

3.2 Identification of RBPs by MS-Based Proteomics

A more common approach related to RNA biology has been the application of proteomic methods to identify RBPs, either in targeted studies by using specific RNA motifs as baits for affinity purification workflows or in unbiased, proteome-wide screens (reviewed by Ascano et al. 2013; Jazurek et al. 2016). The methods for the latter application especially have dramatically improved in recent years. All use some form of photochemical XL to attach interacting proteins covalently to RNA in vivo; although the XL sites are not directly identified, the XL step serves to stabilize the interaction. In 2012, the first two landmark studies of large-scale profiling of RBPs were reported (Baltz et al. 2012; Castello et al. 2012). Both identified more than 800 RNA-interacting proteins in workflows that consisted of a UV XL step, followed by the lysis of cross-linked cells, the isolation of XL products by oligo-deoxythymine (oligo-d[T]) pull-

down, and RNase/protease digestion before liquid chromatography (LC)-MS analysis. More recently, the Hentze laboratory reported an improved version of their RBDmap (mapping of RBDs) protocol that includes two steps of oligo-d(T) capture for enhanced coverage (Castello et al. 2016; Castello et al. 2017). This further refines the spatial localization of interacting sites, and 1,174 putative RNA-binding sites have been identified in HeLa cells using this approach. Recently, Nielsen and co-workers introduced pCLAP (peptide cross-linking and affinity profiling; Mullari et al. 2017), a method that compares favorably with RBDmap in terms of numbers of RBPs identified.

However, all the above approaches rely on the selection of poly(A) motifs in RNA, therefore excluding types of RNA that are not polyadenylated. To circumvent this limitation and taking advantage of the enhanced XL efficiency of 4sU-labeled cells, Bonasio and coworkers introduced their RBR-ID approach (identification of RNA-binding regions; He et al. 2016). RBR-ID relies on reduction of the abundance of peptides spanning RNA-interacting/binding regions after UV XL using quantitative proteomics workflows. Although the method is more generic than the above-mentioned ones, the lack of any dedicated enrichment method makes it less sensitive in comparison. However, the reported data (interaction sites in >800 proteins from mouse embryonic stem cells) still suggest that it is competitive with and complementary to methods that use enrichment. Very recently, two methods that take advantage of bioorthogonal click chemistry approaches on metabolic incorporation of 5-ethynyluridine have been reported and enable poly(A)-independent enrichment steps (Bao et al. 2018; Huang et al. 2018).

Taking the spatial localization of RNA-binding sites on the protein level one step further, Urlaub and coworkers developed a series of methods and computational tools that directly identify peptides containing RNA remnants by MS after protease and nuclease digestion (reviewed in Schmidt et al. 2012). This is in contrast to using non-cross-linked peptides from RBPs for identification, which allows a protein to be categorized only as an RNA interactor. To enable the precise localization of interaction sites, peptides that are modified with short RNA segments (one to three nucleotides) are treated similarly to posttranslational protein modifications, such as phosphorylation, during data analysis. However, the detection of these peptide-RNA adducts is challenging for many reasons, including the low abundance of the products and the diversity of possible RNA attachments. Despite these challenges, the successful application of this strategy has been shown for purified protein complexes and even for whole cells. Notable recent examples include the application of this XL strategy to the spliceosome and clustered regularly interspaced palindromic

repeat (CRISPR) systems (Cretu et al. 2016; Gleditsch et al. 2016). Extending the strategy to the proteome scale has been made possible by the development of an integrated experimental and computational workflow (Kramer et al. 2014; Veit et al. 2016).

3.3 Structural Analysis of Protein-RNA Complexes by XL Coupled to MS

Protein-protein XL has been applied to various large macromolecular assemblies (reviewed by Leitner et al. 2016c) and has provided valuable insights into the architecture of these complexes. A typical example is the use of XL-MS to facilitate subunit positioning in combination with lower-resolution (cryo-)EM maps. Several protein-RNA complexes have been studied in this way. For example, XL was used to position subunits in the large subunit of the mammalian mitochondrial ribosome (Greber et al. 2014) and, later, in the complete 55S mitoribosome (Greber et al. 2015). XL-MS has also been used to investigate ribosomal assembly intermediates such as the yeast pre-60S particle containing numerous assembly factors (Wu et al. 2016) or the ribosome complexed with translation initiation factors (Erzberger et al. 2014; Eliseev et al. 2018). The spliceosome was also studied by a combination of XL-MS and EM in several studies focusing on the human U4/U6.U5 tri-snRNP subcomplex (Agafonov et al. 2016) or on activated states of the yeast (Rauhut et al. 2016) and human complex (Bertram et al. 2017).

The use of protein-protein XL in integrative structural biology has been facilitated by the relatively straightforward localization of the XL sites on the two interacting regions (both peptides). In contrast, site localization at the protein and RNA level for protein-RNA cross-links has been elusive for a long time. Recently, improvements have been made with the help of stable isotope-labeling strategies (Lelyveld et al. 2015; Dorn et al. 2017). In particular, CLIR-MS/MS (cross-linking of segmentally isotope-labeled RNA coupled to tandem MS analysis; Dorn et al. 2017) has been used to identify contacts at single-amino acid and single-nucleotide resolution in an integrated workflow that is described in more detail in Sections 5 and 6.

4 NMR SPECTROSCOPY OF LARGE PROTEIN-RNA COMPLEXES: THE “DIVIDE AND CONQUER” APPROACH AND ITS LIMITATIONS

NMR has clearly become a method of choice to investigate binding of RNA to medium size RBPs or RBDs (Dominguez et al. 2011; Daubner et al. 2013). Solving structures of protein-RNA complexes at atomic resolution is not an easy task as is evident from the low number of protein-RNA

complex NMR structures deposited in the Protein Data Bank (PDB); only approximately 100 of such structures have been released since the first two were published more than 20 years ago (Allain et al. 1996; Battiste et al. 1996). Yet, NMR spectroscopy is a powerful method for detecting RNA binding and localizing the RNA-binding surface. Deuteration of proteins has also allowed the observation of large protein–RNA complexes as in the case of the 600-kDa box C/D methylation enzyme (Lapinaite et al. 2013).

Although solving structures of 10–15-kDa protein–RNA complexes is clearly feasible, this is still a lengthy process that requires two to three years of work by a dedicated investigator. Typically, the first year is needed to identify the correct RNA sequence and ideal buffer conditions for the complex, the second year to assign the resonances of the protein–RNA complex, and the third year to determine a precise structure. The time can double when the protein contains multiple RBDs. Yet, NMR spectroscopy allows structure determination of protein–RNA complexes with weak affinity (mM to μ M K_d), which exist for many RNA regulatory processes such as splicing and translation.

To tackle the structure of larger protein–RNA complexes, additional methods like EPR, SAXS, SANS, and EM are required to complement NMR data if the mass of the complex exceeds 200 kDa. This so-called “divide and conquer” approach is exemplified by our study of RsmE bound to RsmZ (Duss et al. 2014a). NMR spectroscopy was used to solve the structure of several individual RNA SLs bound to RsmE (“divide”) while EPR was crucial to obtain the long-range distances between the subcomplexes to calculate the whole 70-kDa complex (“conquer”; Fig. 2). Whereas solving the structures of RsmE bound to each individual SL was fairly straightforward but tedious (Duss et al. 2014b), considering the similarity with our previously published RsmE–SL complex (Schubert et al. 2007), inserting two spin labels in the different RsmE subcomplexes required preparation of the RNA in several parts, spin labeling them at 4-sU, and ligating the various pieces together (Duss et al. 2014c). For NMR spectroscopy of the 70-kDa complex, we also needed to isotope-label the RNA segmentally (Duss et al. 2010). Collecting 21 EPR distances between the three sub-complexes was then sufficient to calculate two

Figure 2. Examples of the “divide and conquer” approach used to solve the 3D structure of large RNPs in solution. (A) Combining nuclear magnetic resonance spectroscopy (NMR) and electron paramagnetic resonance (EPR) to solve the structure of the 70-kDa RsmE–RsmZ complex (Duss et al. 2014a,b). L and R represent two possible conformations, placing stem loop (SL) 1 (in violet) either to the left (L) or to the right (R) of SL 2 (in green). (B) Combining NMR, cross-linking of segmentally isotope-labeled RNA coupled to tandem MS analysis (CLIR-MS/MS), and other methods to solve the structure of the 85-kDa complex between polypyrimidine tract-binding protein 1 (PTBP1) and the encephalomyocarditis virus internal ribosomal entry site (EMCV IRES) (Dorn et al. 2017).

precise structural ensembles of this protein–RNA complex, because two conformations were observed by EPR.

Using the same approach, we initially considered it feasible to tackle the structure of the 58-kDa polypyrimidine tract-binding protein 1 (PTBP1) bound to an 85-nt RNA from the internal ribosomal entry site (IRES) of the encephalomyocarditis virus (EMCV), overall an 85-kDa complex (Fig. 2). We confirmed early on that RRM1 or RRM2 could bind SL E and F, respectively, but positioning RRM3 and RRM4 proved to be more difficult because of the large size of the complex and because RRM3 and RRM4 interact (Dorn 2017). We therefore asked whether MS could help us find precisely which RRM of PTBP1 binds where in the EMCV IRES. The idea of using segmentally labeled RNA to initially identify the peptides cross-linked to RNA came from our previous work with RsmZ (Duss et al. 2010). This approach proved very fruitful as described in detail below.

5 CLIR-MS/MS AS A TECHNIQUE FOR HIGH-RESOLUTION MAPPING OF INTERACTIONS IN PROTEIN-RNA COMPLEXES

UV-induced protein–RNA cross-linking is widely used for the transcriptome-wide identification of RNA targets of specific bait proteins (Ule et al. 2005; König et al. 2010; Sugimoto et al. 2015; Van Nostrand et al. 2016) and for the identification of new RBPs and their exact RNA interaction sites, respectively, by MS (Castello et al. 2012; Kramer et al. 2014), as discussed above. Although irradiation with UV light induces zero-length cross-links that therefore contain information about the local interfaces of both protein and RNA, single-residue resolution is achieved only for one of them. In particular, the unfavorable fragmentation properties of RNA (Kramer et al. 2014) prevent mass spectroscopic sequencing of cross-linked peptides when large nucleotide stretches are covalently attached; thus, cross-linked RNA moieties must be enzymatically digested. This, in turn, limits the accessible information about the cross-linked RNA to its overall mass, but only up to stretches of four nucleotides. Recently, we developed a new approach called CLIR-MS/MS that combines protein–RNA cross-linking and MS analysis with the use of segmental isotope labeling of RNA to pinpoint protein–RNA interactions (Dorn et al. 2017). The information obtained by the use of heavy isotopes is used to restrict the mapping possibilities for the cross-linked, short RNA stretches detected by MS to the isotopically labeled segment, thereby allowing the identification of the interface at single-amino acid and single-nucleotide resolution simultaneously. CLIR-MS/MS has no general limitation in terms of size, solubility, or crystallizability, but at present

requires homogeneous protein–RNA complexes and depends critically on cross-linking efficiency (Dorn et al. 2017). We used CLIR-MS/MS to pinpoint contacts between PTBP1 and domains D–F of the EMCV IRES and combined these results with structural information derived from the PDB to generate local models for the four RRM s (hereafter, RRM1 for the amino-terminal RRM, etc.) of PTBP1 and their RNA-binding sites.

5.1 Segmental Labeling of RNA

Enzymatic RNA ligation has been used for decades to introduce site-specifically labeled or modified segments that enable insights into the structure and function of the cognate RNA segment in the context of a larger RNA (Moore and Sharp 1992; Xu et al. 1996). Reducing the complexity of NMR spectra by segmental isotope-labeling strategies has turned out to be essential for studying the structures of large RNAs by NMR spectroscopy (Kim et al. 2002; Lukavsky et al. 2003; Lu et al. 2011; Duss et al. 2014a), for instance allowing the determination of a solution structure for the noncoding RNA RsmZ in complex with RsmE protein dimers (Duss et al. 2014a).

RNA segments can be generated by chemical synthesis, which allows single-residue modifications (e.g., for spin-label attachment; Duss et al. 2014c), by in vitro transcription (Milligan et al. 1987; Gurevich et al. 1991; Pokrovskaya and Gurevich 1994; Tzakos et al. 2007; Nelissen et al. 2008; Lebars et al. 2014) or by enzymatic cleavage of a precursor RNA (Xu et al. 1996; Duss et al. 2010). Isotopically labeled nucleotides are commercially available and can be incorporated during transcription (Lu et al. 2010). Special care must be taken to avoid any inhomogeneity at the ligation site. For in vitro transcribed RNA, ribozymatic trimming of the RNA ends or enzymatic and sequence-specific cleavage of a large precursor RNA can yield homogeneous segments.

Sequence-specific cleavage of such precursor RNAs can be achieved using RNase H directed to the desired cleavage site by a chemically synthesized, complementary 2'-O-methyl-RNA/DNA chimera (Duss et al. 2015). Importantly, such chimeras contain about 10–12 residues of 2'-O-methyl-RNA to increase sequence specificity and exactly four deoxynucleotides that form the RNA–DNA hybrid, which in turn is recognized by RNase H. Cleavage occurs across from the 5' end of these four deoxynucleotides (Fig. 3A). The concentration of RNase H and the optimal RNA:chimera ratios are typically optimized in a small-scale reaction to ensure complete cleavage and avoid nonspecific cleavage that can occur in unfavorable cases at certain RNA:chimera ratios (Duss et al. 2015). Recently, Kielpinski et al. characterized the substrate sequence preference of *E. coli* RNase H for efficient cleavage (Kiel-

Figure 3. CLIR-MS/MS analysis of protein-RNA complexes. (A) Design of a 2'-O-methyl-RNA/DNA chimera used to induce RNase H cleavage between nucleotide 319 and 320 of the EMCV IRES RNA. (B) Segmental isotope-labeling strategy to label individual segments of the EMCV IRES RNA (domains D-F) using T4-DNA-ligase in a splinted ligation of RNase H-derived fragments. (C) CLIR-MS/MS analysis workflow to identify protein-RNA cross-links derived from the differentially labeled segment. (Modified, with permission, from Dorn et al. 2017.)

pinski et al. 2017). However, prediction of the optimal RNA:chimera ratio remains difficult as factors like RNA secondary structure might affect cleavage efficiency; in practice, a broad range of ratios should be screened. For segmental isotope labeling of the IRES domains D-F of EMCV and the U1 small nuclear RNA (snRNA), we used ratios ranging from 50:1 to 5:1 (Fig. 3B) or 2:1, respectively, leading to nearly complete RNase H cleavage (Dorn et al. 2017). To generate fragments for segmental isotope labeling of the EMCV IRES domains D, E, the sequence linking E and F (Link), and F, we in vitro transcribed three aliquots of unlabeled RNA (containing their natural abundance isotopes) and one aliquot of $^{13}\text{C}^{15}\text{N}$ -labeled RNA. Each aliquot of unlabeled RNA was cleaved using one of the chimeras, while the isotopically labeled RNA was exposed

to all three chimeras at the same time (Fig. 3B). Cleavage products were then separated by classical RNA purification methods (Edelmann et al. 2014), demanding that cleavage site selection, as well as the design of the precursor RNA, ideally produce optimal separation of fragments. Therefore, we designed cleavage sites such that RNase H digestion using all three chimeras at the same time results in fragments of 36, 17, 10, and 27 nt.

Several strategies are available for RNA ligation that differ specifically in the catalyst for RNA ligation and the use of a DNA splint for fragment association. RNA ligation using T4-DNA-ligase requires a 3'-hydroxyl and 5'-monophosphate group on substrates that are annealed to a DNA splint and has been widely used to incorporate functionalized or labeled RNA segments (Kershaw and O'Keefe

2012). Although the length of the splint influences the ligation efficiency (Kurschat et al. 2005), the splint can also be used to protect unstable RNA sequences (Duss et al. 2015). Ligations catalyzed by T4-RNA-ligase require the same functional groups on the 3' and 5' ends of substrates as for T4-DNA-ligase, but the use of a DNA splint is optional (Lu et al. 2010). Ligation of RNA segments using deoxyribozymes does not require a monophosphate group at the 5' end of the substrate (Purtha et al. 2005). In our study, we used T4-DNA-ligase in a splinted two-piece or three-piece ligation to create samples for complex formation and subsequent UV cross-linking; ligation products were then purified by denaturing anion exchange chromatography (Dorn et al. 2017).

5.2 Strategy for CLIR-MS/MS-Derived High-Resolution Mapping of Protein-RNA Interactions

For CLIR-MS/MS, segmentally isotope-labeled RNA and unlabeled RNA are mixed at equimolar ratios before reconstitution of the protein-RNA complex of interest. UV cross-linking is conducted under conditions preoptimized to yield maximal amounts of cross-linked protein-RNA complex without inducing photodamage (Fig. 3C). Samples are digested with nonspecific RNases and specific proteases to generate peptides modified by the attachment of short RNA stretches (Dorn et al. 2017). As the majority of the resulting digest consists of unmodified peptides and RNA pieces, cross-linked peptides need to be enriched by solid-phase extraction and titanium dioxide affinity chromatography before MS/MS analysis (Leitner et al. 2010). Importantly, cross-links occurring between protein and nucleotides that reside in the labeled RNA segment will generate doublets in the precursor ion mass spectrum. The mass shift then corresponds to composition of the attached, differentially labeled nucleotide(s) and can be predicted for all possible modifications (Dorn et al. 2017). The predicted mass shift is used to identify modified peptides, the modified amino acid, and the modifications themselves using the software xQuest (Rinner et al. 2008; Walzthoeni et al. 2012). xQuest was originally developed to analyze chemically induced protein-protein cross-links and combines mass spectra from differentially isotope-labeled peptides (here due to the RNA adducts) to yield a consensus spectrum that is finally matched against a database of target proteins.

The identification of amino acids modified with differentially isotope-labeled nucleotides provides proof that the particular residue and consequently the corresponding protein domain is in close proximity to the isotopically labeled RNA segment. This information is already valuable for direct future experiments such as tests of deletion mutants, etc. Again, using PTBP1 as an example, we exclusively de-

tected cross-linking for RRM2 and RRM4 to F and Link, respectively, while RRM1 and RRM3 were found to primarily cross-link to E and D, respectively. In our experience, RNase digestion is incomplete and detection of mono-, di-, and trinucleotide modifications can be expected. This infers the composition of a stretch of up to five nucleotides (if partial sequences overlap) and in turn restricts interaction mapping. For example, cross-links detected for PTBP1-RRM3 with D comprised only U and UU, thus restricting mapping to U303-U304 of D as this is the only UU sequence present. Assuming that the protein interacts with the RNA in one predominant conformation, different modifications detected at the same amino acid report the partial nucleotide sequence of the cross-linking site. His457 on PTBP1-RRM4 showed U, C, UC, and UU as modifications (Fig. 4). U and C as single nucleotide modifications indicate the positioning of His457 between these two nucleotides, allowing cross-linking to either of them on the Link sequence. This is consistent with the previously solved NMR structures of PTBP1-RRM34 complexed with single-stranded CU-6mer RNA (Oberstrass et al. 2005). Obtaining UC and UU as dinucleotide modifications indicates a sequence of (U)CUU or UUC(U), of which only UUC is present in the Link sequence. Thus, His457 resides between U342 and C343 (Figs. 3A and 4). This mapping is supported by cross-links of other amino acids at the RNA-binding surface of RRM4.

6 INTEGRATIVE MODELING USING CLIR-MS/MS-DERIVED DISTANCE RESTRAINTS AND STRUCTURAL INFORMATION DERIVED FROM NMR AND/OR AVAILABLE STRUCTURAL INFORMATION FROM THE PDB

As zero-length cross-links, CLIR-MS/MS results can be used as short-distance restraints for integrative modeling. We designed a modeling protocol based on the combined assignment and dynamics algorithm for NMR applications (CYANA) 3.0 (Guntert 2004), which uses both artificial restraints derived from previously published structures (Oberstrass et al. 2005) (RRM1, PDB: 2N3O) and the detected cross-links as ambiguous intermolecular distance restraints (Dorn et al. 2017). Although a detailed manual analysis of the detected modifications leads to a precise assignment of interacting amino acids and nucleotides, we anticipated that fewer data are available on other systems, resulting in ambiguity of the localization of cross-links throughout a coarsely mapped binding site. In contrast to rigid body docking programs, CYANA (Guntert 2004) uses a simulated annealing protocol to generate a bundle of structures that allows individual residues to alter their conformations to accommodate their binding partner.

Figure 4. Single-residue resolution of protein–RNA interactions at the amino acid and nucleotide level by CLIR–MS/MS, exemplified by the identification of one PTBP1 peptide modified with multiple remnants of the EMCV IRES RNA. (Data from Dorn et al. 2017.)

Intra-RRM restraints extracted from the published structures comprise dihedral angle information and atom–atom distance restraints. Intra-RNA restraints were similarly derived from RNA SL structure predictions using MC-Fold and MC-Sym (Parisien and Major 2008). In an initial structure calculation run, we specified these intra-molecular restraints and ambiguous CLIR–MS/MS-derived restraints as the only intermolecular restraints. Because CLIR–MS/MS restraints are sparse, the resulting model was not precise but the ambiguity of CLIR–MS/MS restraints was resolved and binding registers could be extracted at this stage. In subsequent structure calculations, additional unambiguous intermolecular restraints were implemented that represent specific protein–RNA contacts extracted from the available complex structures (Dorn et al. 2017). CYANA 3.0 is a widely used program for NMR structure determination and, thus, the combination of CLIR–MS/MS-derived restraints with classical NMR-derived restraints is straightforward.

To check the quality and accuracy of the 3D model we obtained, we determined independently the structure of RRM2 bound to SL F using CLIR–MS/MS-derived distance constraints and more classically using NMR spectroscopy (Dorn et al. 2017). The results were quite unexpected

because the precision of the structural ensemble was similar and the two structures showed very similar modes of RNA recognition; thus, CLIR–MS/MS appears to be a sufficiently powerful method to produce accurate structures. Note that with CLIR–MS/MS, the structure of RRM2 bound to SL F was determined in the context of the 85-kDa PTBP1–IRES complex, whereas with NMR the structure was determined for an isolated 15-kDa RRM2–SL F complex. Also, about six months of work were required to decipher the structure of RRM2 with CLIR–MS/MS (in parallel to the other three RRMs bound to RNA), whereas about two years were needed to obtain a similar 3D model by NMR. In terms of material needed for both experiments, 100–1000 times less material was needed for CLIR–MS/MS than for NMR. This illustrates the high potential of this new structural approach to decipher protein–RNA interactions. Yet, CLIR–MS/MS as a structural method is highly dependent on other biophysical methods, NMR spectroscopy, and X-ray crystallography. Indeed, while CLIR–MS/MS can provide the intermolecular constraints, the other methods are necessary to generate information on the 3D structure of the protein and the RNA. The other methods will also be needed to distinguish data resulting from less populated

conformational states, which the CLIR-MS/MS method might detect.

In summary, CLIR-MS/MS offers a new avenue for integrative or hybrid structural biology approaches to protein-RNA complexes. Considering the large number of RBP structures determined in their apo state and the ever-increasing knowledge of the sequence specificity of RBPs, CLIR-MS/MS might become a method of choice to investigate rapidly the 3D structure of RBPs in vitro (and perhaps soon in cell extracts or cell culture). Although the method has been applied to only a few RBP complexes, we expect much wider applicability of the method as a very large fraction of the known RBPs can UV cross-link to RNA (Castello et al. 2012).

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (SNF) project 31003A_170130 and the SNF-National Center of Competence in Research (NCCR) RNA and Disease to F.H.-T.A.

REFERENCES

- Abi-Ghanem J, Gabelica V. 2014. Nucleic acid ion structures in the gas phase. *Phys Chem Chem Phys* **16**: 21204–21218.
- Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu W-T, Urlaub H, Lührmann R, Stark H. 2016. Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* **351**: 1416–1420.
- Allain FH-T, Gubser CC, Howe PWA, Nagai H, Neuhaus D, Varani A. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380**: 646–650.
- Ascano M, Gerstberger S, Tuschl T. 2013. Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Curr Opin Gen Dev* **23**: 20–28.
- Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**: 675–690.
- Bao X, Guo X, Xin M, Tariq M, Lai Y, Kanwal S, Zhou J, Li N, Lv Y, Pulido-Quetglas C, et al. 2018. Capturing the interactome of newly transcribed RNA. *Nat Methods* **15**: 213–220.
- Battiste JL, Mao HY, Rao NS, Tan RY, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. 1996. α helix-RNA major groove recognition in an HIV-1 rev peptide RRE RNA complex. *Science* **273**: 1547–1551.
- Bertram K, Agafonov DE, Liu W-T, Dybkov O, Will CL, Hartmuth K, Urlaub H, Kastner B, Stark H, Lührmann R. 2017. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**: 318–323.
- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**: 1393–1406.
- Castello A, Fischer B, Frese CK, Horos R, Alleaume A-M, Foehr S, Curk T, Krijgsvelde J, Hentze MW. 2016. Comprehensive identification of RNA-binding domains in human cells. *Mol Cell* **63**: 696–710.
- Castello A, Frese CK, Fischer B, Järvelin AI, Horos R, Alleaume A-M, Foehr S, Curk T, Krijgsvelde J, Hentze MW. 2017. Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap. *Nat Protoc* **12**: 2447–2464.
- Cretu C, Schmitzová J, Ponce-Salvatierra A, Dybkov O, De Laurentiis EI, Sharma K, Will CL, Urlaub H, Lührmann R, Pena V. 2016. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol Cell* **64**: 307–319.
- Daubner GM, Cléry A, Allain F H-T. 2013. RRM-RNA recognition: NMR or crystallography...and new findings. *Curr Opin Struct Biol* **23**: 100–108.
- Diarra Dit Konté N, Krepl M, Damberger FF, Ripin N, Duss O, Sponer J, Allain FH-T. 2017. Aromatic side-chain conformational switch on the surface of the RNA recognition motif enables RNA discrimination. *Nat Commun* **8**: 654.
- Dominguez C, Schubert M, Duss O, Ravindranathan S, Allain F H-T. 2011. Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* **58**: 1–61.
- Dorn G. 2017. “Structural investigation and hybrid structural modelling of polypyrimidine tract binding protein 1 in complex with the internal ribosomal entry site of encephalomyocarditis virus.” PhD thesis, ETH Zurich. Dissertation no. 24054. doi: 10.3929/ethz-b-000165700.
- Dorn G, Leitner A, Boudet J, Campagne S, Von Schroetter C, Moursy A, Aebersold R, Allain F H-T. 2017. Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat Methods* **14**: 487–490.
- Duss O, Maris C, Von Schroetter C, Allain F H-T. 2010. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucl Acids Res* **38**: e188.
- Duss O, Michel E, Yulikov M, Schubert M, Jeschke G, Allain F H-T. 2014a. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* **509**: 588–592.
- Duss O, Michel E, Diarra Dit Konté N, Schubert M, Allain F H-T. 2014b. Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucl Acids Res* **42**: 5332–5346.
- Duss O, Yulikov M, Jeschke G, Allain F H-T. 2014c. EPR-aided approach for solution structure determination of large RNAs or protein-RNA complexes. *Nat Commun* **5**: 3669.
- Duss O, Diarra Dit Konté N, Allain F H-T. 2015. Cut and paste RNA for nuclear magnetic resonance, paramagnetic resonance enhancement, and electron paramagnetic resonance structural studies. *Methods Enzymol* **565**: 537–562.
- Edelmann FT, Niedner A, Niessing D. 2014. Production of pure and functional RNA for in vitro reconstitution experiments. *Methods* **65**: 333–341.
- Eliseev B, Yeramala L, Leitner A, Karuppasamy M, Raimondeau E, Huard K, Alkalaeva E, Aebersold R, Schaffitzel C. 2018. Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucl Acids Res* **46**: 2678–2689.
- Eliuk S, Makarov A. 2015. Evolution of Orbitrap mass spectrometry instrumentation. *Annu Rev Anal Chem* **8**: 61–80.
- Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CHS, Cimermančič P, Boehringer D, Sali A, Aebersold R, et al. 2014. Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell* **158**: 1123–1135.
- Faini M, Stengel F, Aebersold R. 2016. The evolving contribution of mass spectrometry to integrative structural biology. *J Am Soc Mass Spectrom* **27**: 966–974.
- Gaston KW, Limbach PA. 2014. The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry. *RNA Biol* **11**: 1568–1585.
- Gleditsch D, Müller-Esparza H, Pausch P, Sharma K, Dwarakanath S, Urlaub H, Bange G, Randau L. 2016. Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucl Acids Res* **44**: 5872–5882.
- Greber BJ, Boehringer D, Leitner A, Bieri P, Voigts-Hoffmann F, Erzberger JP, Leibundgut M, Aebersold R, Ban N. 2014. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* **505**: 515–519.

- Greber BJ, Bieri P, Leibundgut M, Leitner A, Aebersold R, Boehringer D, Ban N. 2015. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348**: 303–308.
- Guntert P. 2004. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**: 353–378.
- Gurevich VV, Pokrovskaya ID, Obukhova TA, Zozulya SA. 1991. Preparative in vitro mRNA synthesis using SP6 and T7 RNA polymerases. *Anal Biochem* **195**: 207–213.
- He C, Sidoli S, Warneford-Thomson R, Tatomer DC, Wilusz JE, Garcia BA, Bonasio R. 2016. High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol Cell* **64**: 416–430.
- Heiss M, Kellner S. 2017. Detection of nucleic acid modifications by chemical reagents. *RNA Biol* **14**: 1166–1174.
- Huang J-R, Warner LR, Sanchez C, Gabel F, Madl T, Mackereth CD, Sattler M, Blackledge M. 2014. Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: A combined NMR and SAXS study. *J Am Chem Soc* **136**: 7068–7076.
- Huang R, Han M, Meng L, Chen X. 2018. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proc Natl Acad Sci* **115**: E3879–E3887.
- Jazurek M, Ciesiolka A, Starega-Roslan J, Bilinska K, Krzyzosiak WJ. 2016. Identifying proteins that bind to specific RNAs—Focus on simple repeat expansion diseases. *Nucl Acids Res* **44**: 9050–9070.
- Kambach C, Walke S, Young R, Avis JM, De La Fortelle E, Raker VA, Lührmann R, Li J, Nagai K. 1999. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**: 375–387.
- Kershaw CJ, O'Keefe RT. 2012. Splint ligation of RNA with T4 DNA ligase. *Methods Mol Biol* **941**: 257–269.
- Kielbinski LJ, Hagedorn PH, Lindow M, Vinther J. 2017. RNase H sequence preferences influence antisense oligonucleotide efficiency. *Nucleic Acids Res* **45**: 12932–12944.
- Kim I, Lukavsky PJ, Puglisi JD. 2002. NMR study of 100 kDa HCV IRES RNA using segmental isotope labeling. *J Am Chem Soc* **124**: 9338–9339.
- König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**: 909–915.
- Kramer K, Sachsenberg T, Beckmann BM, Qamar S, Boon K-L, Hentze MW, Kohlbacher O, Urlaub H. 2014. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* **11**: 1064–1070.
- Krepl M, Blatter M, Cléry A, Damberger FF, Allain F H-T, Sponer J. 2017. Structural study of the Fox-1 RRM protein hydration reveals a role for key water molecules in RRM-RNA recognition. *Nucl Acids Res* **45**: 8046–8063.
- Kurschat WC, Muller J, Wombacher R, Helm M. 2005. Optimizing splinted ligation of highly structured small RNAs. *RNA* **11**: 1909–1914.
- Kühlbrandt W. 2014. The resolution revolution. *Science* **343**: 1443–1444.
- Lapinaite A, Simon B, Skjaerven L, Rakwalska-Bange M, Gabel F, Carlungo T. 2013. The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* **502**: 519–523.
- Lebars IB, Vilen B, Bourbigot S, Turek P, Wolff P, Kieffer B. 2014. A fully enzymatic method for site-directed spin labeling of long RNA. *Nucleic Acids Res* **42**: e117.
- Leitner A, Sturm M, Hudecz O, Mazanek M, Småt J-H, Lindén M, Lindner W, Mechtler K. 2010. Probing the phosphoproteome of HeLa cells using nanocast metal oxide microspheres for phosphopeptide enrichment. *Anal Chem* **82**: 2726–2733.
- Leitner A. 2016a. Cross-linking and other structural proteomics techniques: How chemistry is enabling mass spectrometry applications in structural biology. *Chem Sci* **7**: 4792–4803.
- Leitner A. 2016b. Application of chemical cross-linking/mass spectrometry to probe protein structures. In *Encyclopedia of analytical chemistry* (ed. Meyers RA), a9549. Wiley, New York.
- Leitner A, Faini M, Stengel F, Aebersold R. 2016c. Crosslinking and mass spectrometry: An integrated technology to understand the structure and function of molecular machines. *Trends Biochem Sci* **41**: 20–32.
- Lelyveld VS, Björkbohm A, Ransey EM, Sliz P, Szostak JW. 2015. Pinpointing RNA-protein cross-links with site-specific stable isotope-labeled oligonucleotides. *J Am Chem Soc* **137**: 15378–15381.
- Leney AC, Heck AJR. 2017. Native mass spectrometry: What is in the name? *J Am Soc Mass Spectrom* **28**: 5–13.
- Limbach PA, Paulines MJ. 2017. Going global: The new era of mapping modifications in RNA. *WIREs RNA* **8**: e1367.
- Lu K, Miyazaki Y, Summers MF. 2010. Isotope labeling strategies for NMR studies of RNA. *J Biomol NMR* **46**: 113–125.
- Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hiremath A, et al. 2011. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **334**: 242–245.
- Lukavsky PJ, Kim I, Otto GA, Puglisi JD. 2003. Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* **10**: 1033–1038.
- Masliyah G, Maris C, König SLB, Yulikov M, Aeschmann F, Malinowska AL, Mabile J, Weiler J, Holla A, Hunziker J, et al. 2018. Structural basis of siRNA recognition by TRBP double-stranded RNA binding domains. *EMBO J* **37**: e97089.
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**: 8783–8798.
- Moore MJ, Sharp PA. 1992. Site-specific modification of pre-mRNA: The 2'-hydroxyl groups at the splice sites. *Science* **256**: 992–997.
- Mullari M, Lyon D, Jensen LJ, Nielsen ML. 2017. Specifying RNA-binding regions in proteins by peptide cross-linking and affinity purification. *J Proteome Res* **16**: 2762–2772.
- Nelissen FH, van Gammeren AJ, Tessari M, Girard FC, Heus HA, Wijmenga SS. 2008. Multiple segmental and selective isotope labeling of large RNA for NMR structural studies. *Nucleic Acids Res* **36**: e89.
- Nguyen THD, Galej WP, Bai X-C, Savva CG, Newman AJ, Scheres SHW, Nagai K. 2015. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**: 47–52.
- Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, et al. 2005. Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. *Science* **309**: 2054–2057.
- Paizs B, Suhai S. 2005. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* **24**: 508–548.
- Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**: 51–55.
- Pokrovskaya ID, Gurevich VV. 1994. In vitro transcription: Preparative RNA yields in analytical scale reactions. *Anal Biochem* **220**: 420–423.
- Purtha WE, Coppins RL, Smalley MK, Silverman SK. 2005. General deoxyribozyme-catalyzed synthesis of native 3'-5' RNA linkages. *J Am Chem Soc* **127**: 13124–13125.
- Rauhut R, Fabrizio P, Dybkov O, Hartmuth K, Pena V, Chari A, Kumar V, Lee C-T, Urlaub H, Kastner B, et al. 2016. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353**: 1399–1405.
- Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. 2008. Identification of cross-linked peptides from large sequence databases. *Nat Methods* **5**: 315–318.
- Schmidt C, Kramer K, Urlaub H. 2012. Investigation of protein-RNA interactions by mass spectrometry—Techniques and applications. *J Proteomics* **75**: 3478–3494.
- Schubert M, Lapouge K, Duss O, Oberstrass FC, Jelesarov I, Hass D, Allain F H-T. 2007. Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat Struct Mol Biol* **14**: 807–813.

- Schürch S. 2016. Characterization of nucleic acids by tandem mass spectrometry—The second decade (2004–2013): From DNA to RNA and modified sequences. *Mass Spectrom Rev* **35**: 483–523.
- Stark H, Dube P, Lührmann R, Kastner B. 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**: 539–542.
- Sugimoto Y, Vigilante A, Darbo E, Zirra A, Militti C, D'Ambrogio A, Luscombe NM, Ule J. 2015. hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufien 1. *Nature* **519**: 491–494.
- Tzakos AG, Easton LE, Lukavsky PJ. 2007. Preparation of large RNA oligonucleotides with complementary isotope-labeled segments for NMR structural studies. *Nat Protoc* **2**: 2139–2147.
- Ule J, Jensen K, Mele A, Darnell RB. 2005. CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods* **37**: 376–386.
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, et al. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**: 508–514.
- Veit J, Sachsenberg T, Chervov A, Aicheler F, Urlaub H, Kohlbacher O. 2016. LFQProfiler and RNPxl: Open-source tools for label-free quantification and protein–RNA cross-linking integrated into proteome discoverer. *J Proteome Res* **15**: 3441–3448.
- Walzthoeni T, Claassen M, Leitner A, Herzog F, Bohn S, Förster F, Beck M, Aebersold R. 2012. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat Methods* **9**: 901–903.
- Ward AB, Sali A, Wilson IA. 2013. Integrative structural biology. *Science* **339**: 913–915.
- Wilm M. 2011. Principles of electrospray ionization. *Mol Cell Proteomics* **10**: M111.009407.
- Wu S, Tutuncuoglu B, Yan K, Brown H, Zhang Y, Tan D, Gamalinda M, Yuan Y, Li Z, Jakovljevic J, et al. 2016. Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* **534**: 133–137.
- Wurm JP, Holdermann I, Overbeck JH, Mayer PHO, Sprangers R. 2017. Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. *Proc Natl Acad Sci* **114**: 6034–6039.
- Xu J, Lapham J, Crothers DM. 1996. Determining RNA solution structure by segmental isotopic labeling and NMR: Application to *Caenorhabditis elegans* spliced leader RNA 1. *Proc Natl Acad Sci* **93**: 44–48.

Cold Spring Harbor Perspectives in Biology

Combining Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) Spectroscopy for Integrative Structural Biology of Protein–RNA Complexes

Alexander Leitner, Georg Dorn and Frédéric H.-T. Allain

Cold Spring Harb Perspect Biol 2019; doi: 10.1101/cshperspect.a032359

Subject Collection [RNA Worlds](#)

Alternate RNA Structures

Marie Teng-Pei Wu and Victoria D'Souza

Approaches for Understanding the Mechanisms of Long Noncoding RNA Regulation of Gene Expression

Patrick McDonel and Mitchell Guttman

Principles and Practices of Hybridization Capture Experiments to Study Long Noncoding RNAs That Act on Chromatin

Matthew D. Simon and Martin Machyna

Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers

Sarah K. Denny and William J. Greenleaf

Extensions, Extra Factors, and Extreme Complexity: Ribosomal Structures Provide Insights into Eukaryotic Translation

Melanie Weisser and Nenad Ban

Nascent RNA and the Coordination of Splicing with Transcription

Karla M. Neugebauer

Structural Biology of Telomerase

Yaqiang Wang, Lukas Susac and Juli Feigon

Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes

Berthold Kastner, Cindy L. Will, Holger Stark, et al.

What Are 3' UTRs Doing?

Christine Mayr

Single-Molecule Analysis of Reverse Transcriptase Enzymes

Linnea I. Jansson and Michael D. Stone

CRISPR Tools for Systematic Studies of RNA Regulation

Jesse Engreitz, Omar Abudayyeh, Jonathan Gootenberg, et al.

Relating Structure and Dynamics in RNA Biology

Kevin P. Larsen, Junhong Choi, Arjun Prabhakar, et al.

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>

**All Modifications and
Oligo Types Synthesized**
Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

Oligo Modifications?
Your wish is our command.

Combining Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) Spectroscopy for Integrative Structural Biology of Protein–RNA Complexes

Alexander Leitner, Georg Dorn and Frédéric H.-T. Allain

Discovering and Mapping the Modified Nucleotides That Comprise the Epitranscriptome of mRNA

Bastian Linder and Samie R. Jaffrey

Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics

Alexander I. Taylor, Gillian Houlihan and Philipp Holliger

Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast

Clemens Plaschka, Andrew J. Newman and Kiyoshi Nagai

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>

All Modifications and Oligo Types Synthesized

Oligo Modifications?
Your wish is our command.

Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense