From the DEPARTMENT OF MEDICAL BIOCHEMISTRY AND BIOPHYSICS
Karolinska Institutet, Stockholm, Sweden

CONFORMATIONAL DYNAMICS IN MICRORNAS: THE EXAMPLE OF MIR-34A TARGETING SIRT1 MRNA

Lorenzo Baronti

Stockholm 2020

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB

ISBN 978-91-7831-857-5

Conformational dynamics in microRNAs: the example of miR-34a targeting Sirt1 mRNA

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Lorenzo Baronti

Principal Supervisor:
Assoc. Prof. Dr. Katja Petzold
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Structural Biology

Co-supervisor:
Prof. Dr. Gunter Schneider
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Structural Biology

Opponent:
Assoc. Prof. Dr. Kwaku T. Dayie
University of Maryland
Department of Chemistry and Biochemistry

Examination Board:
Prof. Dr. Gerhart Wagner
Uppsala University
Department of Cell and Molecular Biology

Prof. Dr. Lena Mäler
Stockholm University
Department of Biochemistry and Biophysics

Prof. Dr. Björn Högberg
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Biomaterials

To my parents Franco and Rosy

ABSTRACT

In biology, regulatory mechanisms are essential to achieve complex tasks, as virtually every process can be positively or negatively modulated in its outcome, upon different cues.

In humans, microRNAs (miRNAs) constitute a fundamental layer of post-transcriptional gene expression regulation. This class of molecules finely tune protein expression, by downregulating messenger RNAs (mRNA) levels and their translation. The mechanism by which miRNAs find and act upon their targets primarily relies on their nucleotide sequence, relative to the corresponding binding site on the mRNA.

The development of an exhaustive miRNA–mRNA interactome is particularly attractive because of the profound implication for basic biology as well as for diagnostics and therapeutics in human health. However, computational prediction of target sites and associated downregulation levels, using the limited sequence determinants available, is still an outstanding challenge in the field.

In this thesis, we bring forward the hypothesis that modeling of miRNA–mRNA pairs might benefit from considering the inherent structural flexibility of these complexes, at the molecular level.

In the introductory chapter, we present the structural features of RNAs with a focus on their conformational dynamics and NMR spectroscopy as a tool to investigate these motions. The molecular details of miRNA biogenesis and function are later introduced to contextualize the results of Paper I. Finally, the challenges associated with RNA sample preparation are discussed in light of the work presented in Paper II.

In Paper I, we show that a miRNA–mRNA pair involved in a cancer-regulating pathway exploits its flexibility to toggle between lower and higher target repression states. This study shows that suboptimal structures of a given miRNA–mRNA pair, that are overlooked by computational prediction and that often elude experimental detection, can be functionally relevant and are essential to draw a mechanistic picture of miRNA function.

The methods used in Paper I for RNA sample preparation and molecular simulation are described in Paper II and II, respectively. While these methods were essential to achieve the results of Paper I, they also find widespread application in the RNA field.

LIST OF SCIENTIFIC PAPERS

I. **Baronti L.**, Guzzetti I., Ebrahimi P., Friebe Sandoz S., Steiner E., Schlagnitweit J., Fromm B., Silva L., Fontana C., Chen A.A., Petzold K. (2020). Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. *Nature*, 10.1038/s41586-020-2336-3.
Here presented in manuscript format.

II. Karlsson H., **Baronti L.**, & Petzold K. (2020). A robust and versatile method for production and purification of large-scale RNA samples for structural biology. *RNA*, rna-075697.
Here presented in manuscript format.

III. Ebrahimi P., Kaur S., **Baronti L.**, Petzold K., & Chen A. A. (2019). A two-dimensional replica-exchange molecular dynamics method for simulating RNA folding using sparse experimental restraints. *Methods*, 162, 96-107.

Publications not included in this thesis

Baronti L., Karlsson H., Marušič M., & Petzold K. (2018). A guide to large-scale RNA sample preparation. *Analytical and bioanalytical chemistry*, 410(14), 3239-3252.

Feyrer H., Munteanu R., **Baronti L.**, & Petzold K. (2020). One-Pot Production of RNA in High Yield and Purity Through Cleaving Tandem Transcripts. *Molecules*, 25(5), 1142.

LIST OF ABBREVIATIONS

RNA	RiboNucleic Acid
mRNA	messenger RNA
ncRNA	non-coding RNA
miRNA	micro-RNA
nt	nucleotide
WC	Watson-Crick
NMR	Nuclear Magnetic Resonance
RF	Radio Frequency
ppm	parts per million
HSQC	Heteronuclear Single Quantum Coherence
NOESY	Nuclear Overhauser Effect SpectroscopY
SL	Spin Lock
GS	Ground State
ES	Excited State
Ago	Argonaute
pri-miRNA	primary miRNA
pre-miRNA	precursor miRNA
RISC	RNA-Induced Silencing Complex
mSirt1	Sirtuin 1 mRNA
HPLC	High Performance Liquid Chromatography
PAGE	PolyAcrylamide Gel Electrophoresis
REMD	Replica-Exchange Molecular Dynamics
DLR	Dual-Luciferase Reporter

1 INTRODUCTION

Ribonucleic acids (RNAs) possess the formidable, dual capacity of acting as genetic information carriers as well as functional elements. While the first role has been described and accepted as early as the 1960's[1–3], the second, with the exceptions of ribosomal[4] and transfer[5] RNAs, has only been widely recognized in recent years[6,7]. These RNAs, that do not participate in the central dogma of biology as information messengers (mRNAs), are often referred to as non-protein-coding or non-coding RNAs (ncRNAs). With the advent of next-generation sequencing[8], the extent to which these ncRNAs are transcribed in humans[9,10] is becoming apparent, revealing that ncRNAs pervade every aspect of Life.

The ability of ncRNAs to perform tasks that were originally only accredited to proteins (i.e. gene regulation[7], catalysis[11], metabolites binding[12]) largely rely on their ability to adopt complex folds. Therefore, determining the secondary and tertiary structure of ncRNAs is of utmost importance to gain a mechanistic understanding of their function.

Traditional high-resolution structure determination techniques, such as X-ray crystallography, have proven successful for a large number of highly structured RNAs and protein-RNA complexes. However, it is becoming increasingly clear that the majority of ncRNAs retain a certain degree of structural flexibility and that a single-structure snapshot is not sufficient to fully understand how these molecules work[13–15].

In this introductory chapter, we will describe the molecular properties that confer the RNAs the ability to switch between different conformations. We will introduce the basics of Nuclear Magnetic Resonance Spectroscopy (NMR) as a tool to infer the secondary structure of RNAs, as well as the NMR experiments used to characterize different conformational states in solution. Later on, the class of small regulatory ncRNAs called microRNAs (miRNAs), subject of Paper I, will be introduced and their role in eukaryotic post-transcriptional gene expression regulation will be discussed. Finally, we will present the challenges associated with the large-scale RNA sample preparation for structural studies.

1.1 RNA structure and dynamics

RNA is a polymer consisting of four nucleotide types: adenosine (A), guanosine (G), cytidine (C) and uridine (U). Each nucleotide is composed of a nucleobase that confers its chemical identity, adenine and guanine (purines), cytosine and uracil (pyrimidines), a sugar ribose and a phosphate group carrying a negative charge. The nucleotides are connected to one another via a phosphodiester bond between the 3'-OH group of a nucleotide *i* and the

5’-PO_4 group of the following nucleotide *i+1* (Fig. 1a–b). This arrangement gives the polymer an end-to-end directionality, that is conventionally drawn left to right, from the 5’-end to 3’-end.

Unlike DNA, which is commonly found in nature in a double-stranded form, where anti-parallel strands with sequence complementary are brought together via nucleobases-specific hydrogen bonding (G–C and A–T) and stacking, RNA can be found in a variety of different configurations. Perhaps, rather than describing RNA in terms of double- or single-stranded, it is worth noting that often RNAs are composed of alternating structural elements. These *secondary* structural elements can take advantage of the ability of nucleobases to stack and base-pair (hence forming locally double-stranded regions called helical stems), as well as exploiting the many torsion angles (Fig. 1c) and interaction opportunities available in the molecule to form locally single stranded features. This structural heterogeneity, that is the hallmark of RNA, ultimately enables these molecules to fold into intricate three-dimensional conformations and interact with other RNA and protein partners.

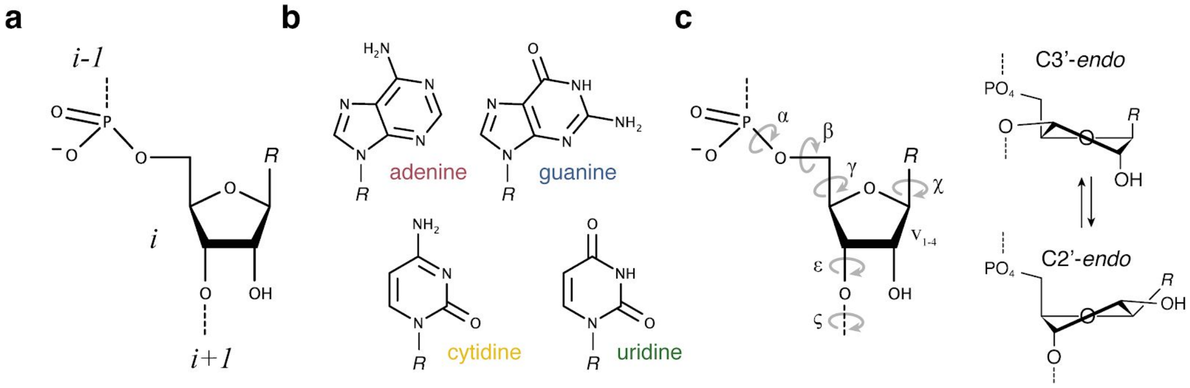


Fig. 1 RNA structure. a, Five carbons ribose sugar and phosphate group arrangement in the oligonucleotide chain. **b,** Four RNA bases. **c,** Sugar-phosphate backbone and glycosidic bond angles (left) and sugar puckers resulting from the constrained pseudorotation angles υ_{1-4} (right).

In the next paragraphs we’ll present a summary of common secondary structural elements and the repertoire of base-pairing configurations available to RNAs. We’ll introduce a few examples where these secondary structural elements and non-canonical base pairs enable the RNAs to adopt multiple alternative conformations and how this *conformational dynamics* confers a functional advantage to these molecules.

1.1.1 Secondary structural elements and folding

Helices. Helices are formed when two strands, with Watson-Crick (WC) sequence complementarity (G–C and A–U) are brought together by stacking and base pair hydrogen bonding, in an antiparallel fashion (Fig. 2a). Whether the strands are from the same molecule (intramolecular) or from two different ones (intermolecular), the helical stem possesses the same geometrical features reminiscent of A-form DNA duplexes, hence named A-form RNA. The riboses are all constrained to adopt 3'-*endo* puckers (Fig. 1c), each full turn of the helix comprises 11 nucleotides and the major and minor grooves are narrow and deep, wide and shallow, respectively[16]. Unlike DNA, the base composition of A-form RNA does not influence its structural properties[17]. At high ionic strength conditions, a slightly different configuration has been observed (A'-form RNA) with 12 nts turn and wider major groove[18]. Helices can accommodate mismatches or non-canonical base pairs, depending on the nature and number of the base pair, different distortions in the geometry of the A-form stem occur[19,20].

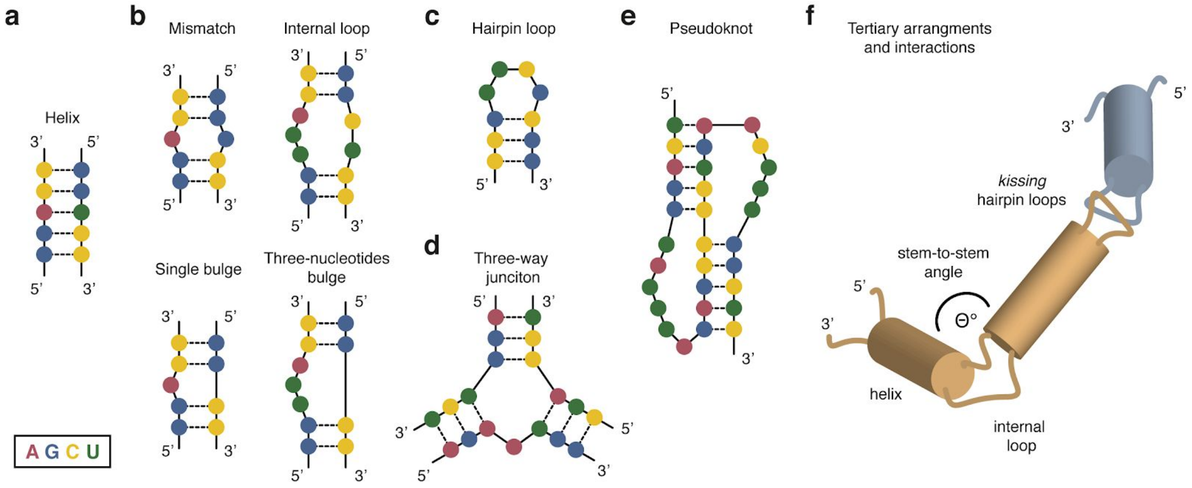


Fig. 2 Secondary structure elements and tertiary arrangements. a, A-form helical element composed of WC base pairs. **b,** Mismatches, internal loops and bulges. **c,** Hairpin loop element. **d,** Three-way junction element. **e,** Pseudoknot. **f,** Tertiary arrangements exemplified as two inter-molecular kissing hairpin loops and intra-molecular stem-to-stem arrangement. Cylinders represent helical elements.

Mismatches, internal loops and bulges. The simplest deviation from the A-form helix occurs when a single nucleotide is mutated so that it cannot form a WC base pair with the corresponding nucleotide on the opposite strand (Fig. 2b). In this configuration the mismatched nucleotides are embedded between two WC helices. Depending on the number of sequential mismatched pairs and their stacking configuration, these secondary structural elements are referred to as internal loops. When the two strands contain a different number of mismatched nucleotides the internal loop is considered asymmetric. A particular case of asymmetric internal loop occurs when all the mismatched nucleotides reside on only one of the two strands. In this case the unpaired nucleotides form a bulge (Fig. 2b).

Hairpin loops. Often in RNAs, helices are formed intramolecularly, within a single transcript. This is possible thanks to looping nts that allow for the two strands to interact (Fig. 2c). The number of looping nts can, in principle, vary indefinitely. Typical loops are sized between 4 and 8 nts and often adopt a well defined three-dimensional structure that confers thermodynamic stability to the stem they are connecting. A well studied class of hairpin loops are tetraloops, that can often be found in the following configurations: UNCG, GNRA and CUUG (where N = A, G, C or U and R = A or G)[21].

Three- and four-way junctions. These are common motifs found in highly structured RNAs where multiple helical domains converge at a common junction (Fig. 2d). Perhaps the most famous example of four-way junction is the one that enables the tRNA to adopt the typical L-shaped configuration in 3D, commonly drawn as a cloverleaf in 2D[22].

Pseudo-knots. Pseudoknots occur when bases in an internal or hairpin loop motif base-pairs with bases outside of these motifs (Fig. 2e). This general definition comprises different types of pseudoknots that have been observed experimentally[23].

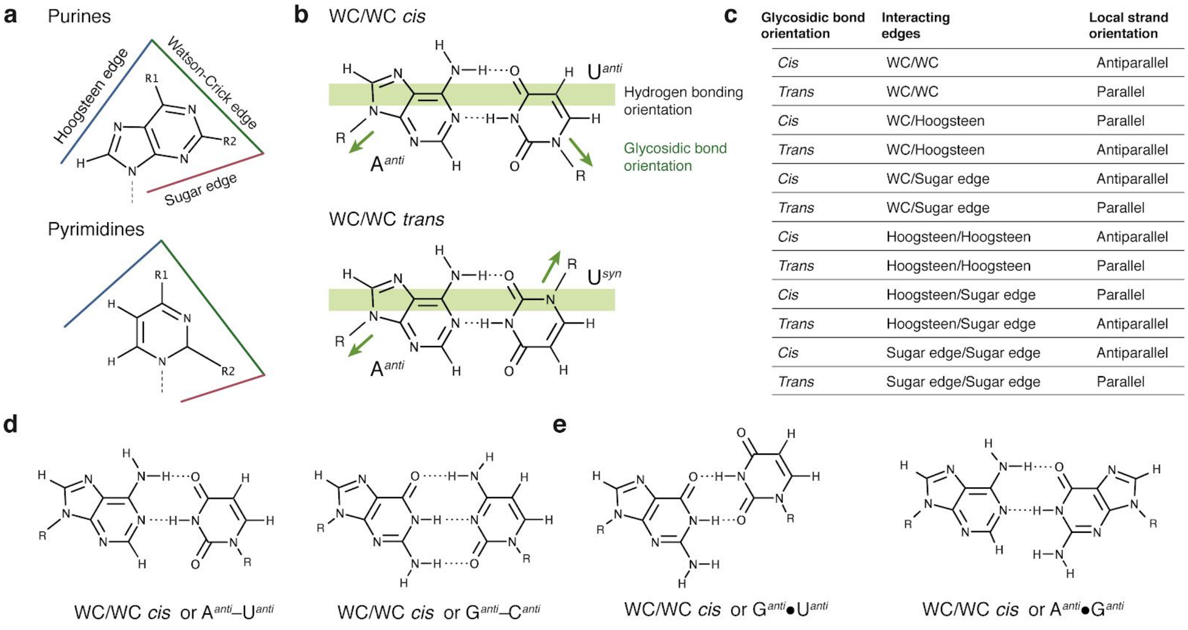
Folding of the secondary structural elements mentioned above depends on the RNA primary sequence. Therefore, *de novo* prediction of structure folding has been a historically outstanding challenge[24]. A set of rules, that combines thermodynamic parameters and nearest-neighbor model, has been described over the years, to unveil how nucleotides impact the 2D folding energetics. This was achieved by systematic measurements of thermodynamic parameters, as a function of primary sequence, using of short RNA duplexes and hairpins typically by means of thermal melting followed by ultra-violet (UV) and circular dichroism (CD) absorption as well as NMR spectroscopy. The result is a set of rules that takes the name from the work of Turner and colleagues and is at the basis of most secondary structure prediction algorithms[25].

It is worth noting that, in the 3D space, each class of secondary elements, depending on their specific nucleotide composition, can give rise to very different arrangements. Furthermore, for internal loops, bulges and junctions, these elements will contribute to the relative orientation of neighboring helical stems (Fig. 2f). This feature, together with long range tertiary contacts (i.e. *kissing* hairpin loops), ultimately determines the final shape of the molecule (Fig. 2f). While secondary structure prediction achieves good results for relatively short and mostly canonically base-paired RNAs, the RNA folding problem is far from being solved for tertiary structures without the help of experimental structural data.

1.1.2 Base pairs

Nucleotides interact with each other via their base, sugar and phosphate components. Inter-nucleotide base-base interactions are of two main types: π-π stacking and hydrogen bonding[26]. While the first type of interaction is of utmost importance for RNA structure and dynamics, for the purpose of this chapter, we will only introduce the fundamental concepts behind base pair hydrogen bonding and the most commonly found examples in ncRNAs. For the reader interested in a comprehensive review of inter-nucleotide interaction modes Sweeney *et al*., compiled an excellent overview on the subject[26].

The directional nature of hydrogen bonding and the geometry of nucleobases, limits base-base interactions to discrete edges of each base in the pair (Fig. 3a). These edges are referred to as 1) the Watson-Crick edge, given the canonical interaction in the DNA double helix[27], 2) the Hoogsteen edge, named after Karst Hoogsteen's observations[28] and 3) the sugar edge (Fig. 3a). Each of these edges present both hydrogen-bonding donors and acceptors groups[26]. In 2001, Leontis and Westhof proposed a unified and systematic base pair nomenclature that is based on the interacting edges and the the glycosidic bonds orientation with respect to the hydrogen bonding orientation (i.e. *cis* or *trans* configuration)[29] (Fig. 3b–c). The edges involved in the base pair and the *cis/trans* configuration determine a local orientation of the strands that can be considered "locally" parallel or anti-parallel. The 12 resulting combinations, or families, of base pairs types presented by Leontis and Westhof, with the associated strand orientation, assume that all the bases are in the most common *anti* rotation state about the glycosidic bond χ (Fig. 3c). If the base rotates 180° about χ (*syn* conformation), the strand orientations are reversed. In analogy, if both bases are in *syn*, the strand orientations are reversed-back to their original state.



Glycosidic bond orientation	Interacting edges	Local strand orientation
Cis	WC/WC	Antiparallel
Trans	WC/WC	Parallel
Cis	WC/Hoogsteen	Parallel
Trans	WC/Hoogsteen	Antiparallel
Cis	WC/Sugar edge	Antiparallel
Trans	WC/Sugar edge	Parallel
Cis	Hoogsteen/Hoogsteen	Antiparallel
Trans	Hoogsteen/Hoogsteen	Parallel
Cis	Hoogsteen/Sugar edge	Parallel
Trans	Hoogsteen/Sugar edge	Antiparallel
Cis	Sugar edge/Sugar edge	Antiparallel
Trans	Sugar edge/Sugar edge	Parallel

WC/WC *cis* or A^{anti}–U^{anti}

WC/WC *cis* or G^{anti}–C^{anti}

WC/WC *cis* or G^{anti}•U^{anti}

WC/WC *cis* or A^{anti}•G^{anti}

Fig. 3 Leontis and Westhof base pairs schematics and nomenclature. Adapted from[29]. **a,** Base edges. **b,** Schematics of *syn* and *anti* glycosidic bond orientation (green arrows), relative to the hydrogen bonding orientation (green bar). **c,** 12 base pair families. **d,** Canonical WC base pairs. **e,** Non-canonical WC/WC *cis* base pairs.

Canonical WC base pairs

WC/WC *cis* G–C, C–G, A–U and U–A base pairs are considered canonical, as they are the most commonly found in nature and are the basis of fundamental biological processes such as the codon-anticodon recognition, during protein synthesis. Canonical base pairs are isosteric as they retain the same C1'-C1' distance, the bases have the same rotation about χ and the hydrogen-bonding network is in the same register[30–32]. The hydrogen bonding network for A–U and G–C is shown in Fig 3d. These features lead to a fundamental property of the canonical A-form RNA by which the helix geometry is independent of its primary sequence composition. This *structural neutrality* has profound implications for the study of evolutionary conservation (i.e. covariation) and 3D structure determination. Isostericity subclasses within families can be described for both canonical and non-canonical base pairs[30–33].

Non-canonical example: the G–U wobble

The most commonly found non-canonical base pair is the WC/WC *cis* G–U "wobble". This base pair takes its name from Crick's "Wobble Hypothesis" for codon-anticodon recognition[34] and was first predicted to be present in the folded yeast tRNA^{Ala}[35]. Beside its role in codon-anticodon recognition, the G–U wobble plays crucial roles in many ncRNAs and often provides a recognition element for protein binding[19,20,36]. From a structural point of view, both bases interact via their WC edges and the register of hydrogen bonding is shifted with respect to a canonical G–C pair, leaving the exocyclic amino group of the guanosine not hydrogen-bonded (Fig. 3e, left). In A-form helices, the G–U wobble results in an uninterrupted patch of negative electrostatic potential in the major groove[20], providing a hub for divalent metal ion binding[37]. The hydrogen bonding register shift results in a difference between the nucleotides' glycosidic bond angles. Unlike canonical pairs, for which χ is about 54° for both nucleotides in the pair, the G–U wobble has a χ of about 40° for the guanosine and 65° for the uridine. This asymmetry in turns leads to the non-isostericity of G–U and U–G wobble pairs[20,36]. Interestingly, 5'-G–U-3' and 5'-U–G-3' have different stacking arrangement with the preceding base pairs when placed at the end of an helix[38], favoring the 5'-G–U-3' over the 5'-U–G-3'[36]. Instead, when the wobble closes a hairpin loop, the 5'-U–G-3' is favored[39].

It must be noted that the wobble is not the only configuration accessible to G and U to pair. "Bifurcated" configurations have been observed where the O2 (or O4) of the uridine forms two hydrogen bonds with the proton in the amino and imino groups of the guanosine. These bifurcated G–U pairs and their occurrence have been recently reviewed by Turner and colleagues[40]. In the same work Berger *et al.*, prompted by the previously observed instability of closing G–U pairs in internal loops[41], used NMR spectroscopy to infer the occurrence of wobble or bifurcated types flanking different 2X2 internal loops. They propose that "*GU pairs with the U 5' of an internal loop are typically wobble pairs*"[40], potentailly due to their increased thermodynamic stability compared to 5' G sequences[40]. While "*GU pairs with the G 5' of an internal loop can form non-wobble pairs*"[40], when preceded 5' by an A–U or U–A pair [40]. These interesting observations are particularly relevant for the result of Paper I, in which we observe the transient formation of a G–U pair, with the G 5' of a 4 nts asymmetric bulge.

Considering the rest

With three interacting edges with multiple hydrogen bond donors and acceptors, the combination of potential for non-canonical base pairs is large. This repertoire can be further expanded as certain groups can change their protonation or tautomeric state, thereby providing new bonding donor or acceptor groups[32,42]. Beyond the G–U wobble, many other non-canonical pairs have been observed and catalogued[29,30,43]. When predicting or validating an RNA structure is therefore important to consider the contribution of these pairs. Among many laboratories developing tools to compute 2D and 3D structures, such as those participating in the RNA-Puzzles[44], the Major group showed that the incorporation of non-canonical pairs in their predictions was crucial to improve accuracy[45]. In particular, MC-Fold was shown to be fundamental in the identification and validation of suboptimal secondary structures[46]. These alternative conformations, often requiring the formation of non-canonical base pairs, and their relation to the "native" state, are the subject of the next paragraph.

1.1.3 Conformational dynamics

The conformational (or energy) landscape of a given molecule, represents the distribution of all possible structural conformations that a molecule can adopt, as a function of their respective free energy, under a given set of experimental (or theoretical) conditions. The conformational landscape is often represented as a 2D or 3D plot, where the y axis is the Gibbs free energy and the x/z axes are convoluted variables representing the parameters that describe the molecular architecture (i.e. torsion angles or atomic coordinates) (Fig. 4a). This can help visualize folding pathways and equilibrium conformational dynamics, where minima represent metastable structures and maxima represent transition-state structures. In this perspective, folding of a certain biopolymer is a defined trajectory across the landscape, while the equilibrium conformational dynamics is the collection of exchange processes between neighboring global and local minima.

For globular protein, as formulated by Anfisen and colleagues in the "Thermodynamic Hypothesis"[47,48], the state of lower Gibbs free energy represents the native state, in physiological conditions. To a first approximation, the landscape around the global minimum is shaped like a funnel[49]. While protein folding is a multifaceted field of research and exceptions to this exist[50,51], the folding funnel is a generally accepted model for globular proteins[52–54]. In RNA instead, due to the high degeneracy of base pairs and stacking energetics and the modular nature of secondary structure elements, the conformational landscape has quite different characteristics[13,14,55–57].

From a folding perspective, RNAs can engage multiple trajectories that lead to different structures, with free energies that are comparable with the natively folded one. In addition, these alternative conformers are often kinetically trapped, as their transition-state energies are too high to be overcome by simple thermal fluctuations, in biologically compatible timescales[58,59]. This features confer the landscape its so-called *rugged* shape. In nature, co-transcriptional folding[60–62] and RNA chaperones[63,64] have evolved to avoid such traps that may, or may not be, of functional relevance. *In vitro* and *in vivo*, comparative structural studies are emerging to address the role of the cellular components that drive the folding trajectories and to identify biologically relevant alternative conformations[65]. A unique feature of RNAs is that their folding occurs in a hierarchical fashion, with secondary structures, such as base pairs, preceding the formation of tertiary contacts[66].

Once the RNA has adopted its native fold (at the equilibrium conditions), it can exchange between neighboring states on timescales and populations that depend on the transition state energies and relative free-energy difference between the two exchanging states (Fig.

4a). These motions collectively describe the conformational dynamics of the RNA. To describe the hierarchy of RNA conformational dynamics, Al-Hashimi and coworkers introduced the concept of *tiers* as different classes of motion that happen on discrete timescales[57]. This classification is useful to conceptualize RNA motions, however, as the authors point out, one must keep in mind that different tiers are correlated to each other and timescale ranges can overlap to a certain extent[57]. An additional layer of complexity, that won't be discussed here, is presented by how different cellular cues and naturally occurring RNA chemical modification can redraw the conformational landscapes inferred *in vitro*. Commonly, the effect of metal ions, in particular divalent cations, such as Mg^{2+}, that are necessary to counteract the negative charges of the phosphate backbone and promote folding, and different pH conditions, are probed *in vitro* for physiologically relevant ranges. However, the throughput of traditional biophysical, biochemical and structural techniques pales compared to the large number of parameters that should, in principle, be systematically tested. Novel methods probing RNA structures in cells[67] as well as techniques that raise the throughput of parallel *in vitro* experiments[68] are showing promising avenues and will likely contribute to developing our understanding of conformational dynamics.

Tier 0. Large secondary structure rearrangements that involve multiple nucleotides (Fig. 4a, left). This tier dynamics occurs within long exchange times (>ms) relative to biological events. A good example of this are riboswitches, a class of *cis*-acting mRNA elements that undergoes a tier 0 conformational transition to regulate mRNA expression, often at the level of translation[69]. Given the high transition state energy involved, protein chaperones, small molecular ligands, divalent cations and other cellular factors are needed to lower the energy barriers between kinetically trapped states and tune the timescale of motion with the processes these RNAs are controlling (i.e. transcription termination, translation initiation, splicing *etc.*)[57,69].

Tier 1. Nucleobase dynamics (Fig. 4a, center). In the *μs-ms* timescale, the hydrogen bonding of canonical base pairs can melt and reform[70–72]. These *breathing* motions provide opportunities to trigger wider transitions or to expose residues that would otherwise be buried in a helical element[57,69], to chaperones[64] and helicases[73]. Ionization and keto-enol tautomerization have also been reported[74], though these events involve a chemical rather than a structural exchange mechanism. All these processes can lead to change in configuration of a given base pair, that propagates to the topology of the backbone and major/minor groove geometry with consequences to partner binding[75] and even tRNA decoding accuracy[74,76].

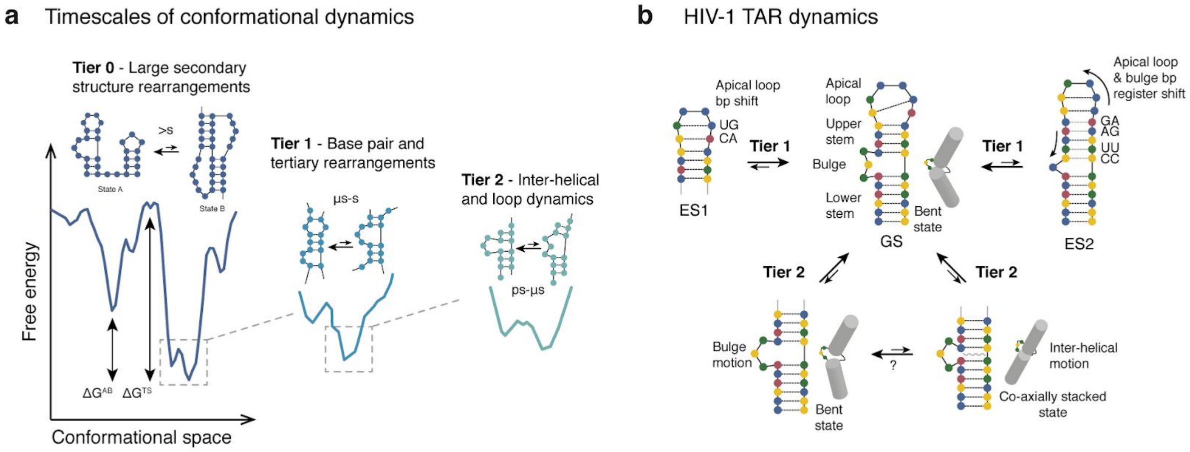


Fig. 4 Conformational dynamics and the example of HIV-1 TAR. a, Structural arrangements occur on different timescales. **b,** HIV-1 TAR undergoes tier 1 and 2 conformational dynamics.

Finally, and perhaps of most relevance for the results presented in Paper I, the multiplicity of available base pair configurations and their energetic degeneracy make nucleotides close to non-helical secondary structure elements prone to spontaneously rearrange and shift the pairing register in this timescale[57].

The Trans-Activation Response (TAR) element of Human Immunodeficiency Virus Type I (HIV-1) is probably the best characterized RNA in terms of conformational dynamics and it has been investigated throughout the years using variety of structure probing techniques, recently summarized by Merriman *et al.*, in the introduction of their work[77]. The minimal construct is 29 nts and is composed of a lower bottom stem (6 WC base pairs), a 3 nts bulge, an upper stem (4 WC base pairs) and a 6 nts apical loop (Fig. 4b, top center). This is the lowest energy secondary structure which populates the largest fraction of the conformational ensemble in physiological conditions and it is often referred to as the ground state (GS).

To exemplify tier 1 and 2 motions of TAR and provide a framework to the results presented in Paper I, we'll briefly discuss the work of the Al-Hashimi group using NMR $R_{1\rho}$ relaxation dispersion experiments to characterize this model RNA. However, it is important to note that our current understanding of HIV-1 TAR RNA structure, dynamics and function comes from the fundamental contribution of many different research groups, whose work, for the sake of brevity, will not be reviewed in this chapter. In particular, for the structure and dynamics by NMR of this molecule, the interested reader is directed to the recent work of the Drobny and Varani groups[78–82].

A fast tier 1 motion involving residues of the apical loop was initially described by Al-Hashimi and coworkers on a truncated TAR construct using ^{13}C NMR $R_{1\rho}$ relaxation dispersion[46] (a technique that will be discussed in the next paragraph). This process occurs with a timescale of exchange of ~ 40 μs and populates about 13% of the ensemble and leads to a secondary structure termed Excited State 1 (ES1) (Fig. 4b, top left). G34, initially paired with C30, shifts to a *trans* U∘G^{syn} wobble base pair with U31 while A35, initially unpaired, forms a A$^+$•C with C30 in ES1. Later on, using $^{13}C/^{15}N$ NMR $R_{1\rho}$, the same research group identified a second, slower conformational exchange process, where the apical loop and bulge motions are coupled[83] (Fig. 4b, top right). In ES2, a register shift occurs in the upper stem leading to four non-canonical base pairs being accommodated in this helical element. The shift drives shortening of the bulge to 1 nts and collapse of the apical part to a compact tetraloop. This second process is characterized by an exchange rate of 2 ms and populates only 0.4% of the ensemble[83]. ES1 and ES2 motions were also shown to be decoupled from each other using insertion and deletion constructs and $^{13}C/^{15}N$ NMR $R_{1\rho}$[77,83] on the 29-mer TAR. Both ESs, albeit at different rates, have been proposed to play a role in modulating the interaction with Tat and Cyclin T1[56,83], proteins which bind TAR at distinct sites but in a cooperative fashion[84] to promote HIV-1 genome transcription in infected cells[85].

For the sake of conciseness, we will only mention that tertiary interactions, such as *kissing* hairpin loops and pseudoknot formation are included in this tier[57].

Tier 2. Inter-helical and loop dynamics (Fig. 4a, right). These motions involve local loop dynamics, such as bases stacking-in or out neighboring helices, backbone distorsions and inter-helical orientations. These motions can vary in amplitude, therefore tier 2 covers a broader range than other tiers, spanning from ps to μs[57].

HIV-1 TAR is also a good model for tier 2 dynamics (Fig. 4b, bottom). In a recent publication[86], Al-Hashimi and coworkers described the tier 2 dynamics that involves the 3 nts bulge and its effect on the relative orientation between the upper-lower stems. This work made use of NMR Residual Dipolar Coupling[87,88], a domain elongation approach[89], spin relaxation and molecular dynamics to describe a conformational ensemble where the major and minor states are quantitatively represented. The results indicate that TAR mostly resides in a bent conformation that transitions in the ns-μs timescale to a coaxially stacked configuration. Furthermore, the bulge residues undergo ps-ns timescale motions that occur within the bent or stacked states[86] and could potentially provide the necessary rearrangement to transition between the two states.

As shown in this paragraph, it is often necessary to describe RNAs as conformational ensembles, rather than single structures. Therefore, integrative approaches that combine molecular simulations with different experimental techniques, probing RNA structure at different levels of resolution, are often necessary to gain a complete picture of the system of study[90]. Among experimental structural methods, NMR is particularly suited to collect conformational dynamics information on a variety of timescales. In the next paragraph we'll discuss the basic experiments necessary to solve secondary structures of RNAs and their motions with this technique.

1.2 Nuclear Magnetic Resonance spectroscopy

NMR spectroscopy is based on the nuclear magnetic resonance effect observed for atoms (with nuclear spin $I \neq 0$) in a constant magnetic field (B_0) and perturbed by a radiofrequency (RF) pulse. The theoretical and practical framework of NMR spectroscopy and its instrumentation requires extensive knowledge in fundamental physics, mathematics and electrical engineering. Nevertheless, NMR is used routinely by researchers and laboratory personnel with different backgrounds and scope of work across the world, from basic analytical chemistry application to cutting edge method development. The popularity of this technique hints that NMR can be approached at different levels of complexity and that deep theoretical understanding and application can be decorrelated, to a certain extent.

For the biochemist interested in using solution-state NMR to characterize structure, dynamics and other physical properties of biomolecules, a common starting point is the understanding of nuclear magnetization through the vector model and progressively approach more complex experiments using the product operator formalism[91]. In parallel, concepts of data acquisition and processing (i.e. fourier transformation) are developed to understand the essential commands at the spectrometer. These two, theoretical and practical, sides of the topic can be engaged with a basic mathematical background that requires familiarity with calculus, trigonometry, complex numbers and matrices. Further understanding of so-called *relaxation* phenomena[91] is required in order to interpret the experiments used for the quantification of conformational dynamics.

The description of NMR basics is beyond the scope of this introductory chapter, however three key concepts, that are often a source of confusion for the general reader, need to be presented in order to understand paragraphs 1.2.1–3 and fully grasp the impact of the results presented in Paper I.

First, likewise any other spectroscopic technique, a 1D spectrum shows intensity of absorption or emission of energy as a function of frequencies in the electromagnetic spectrum. For NMR, the range involved falls into the radio frequencies, typically from 10 to 800 MHz[91]. Unlike other spectroscopies, where the x axis is often presented in frequencies or wave-length units, an NMR spectrum has its x axis expressed as adimensional parts per million (ppm). This has to do with the fact that the absolute value of the frequencies observed depends on the *strength* of the static magnetic field applied B_0. Therefore, being able to compare data coming from different spectrometers (therefore different B_0), requires referencing the x axis by the frequency of a standard compound at B_0, thereby obtaining the unit-less ppm values. The position of each peak in the spectrum, expressed in ppm, specifies the frequency at which a given nucleus resonates relative to a reference compound. This difference is referred to as the *chemical shift* of such a nucleus. Chemical shift values are strong reporters of the chemical and structural environment of a given nucleus.

Second, in biomolecular NMR it is often necessary to increase the dimensionality of the spectra to resolve the overlap between different peaks. Typically, 2D NMR spectra are acquired to obtain a fingerprint of a given biomolecule and are enough to resolve single amino acids or nucleotides in small-sized proteins or RNAs. These fingerprint spectra are commonly Heteronuclear Single Quantum Coherence (HSQC)[92] correlating ^{1}H-^{13}C or ^{1}H-^{15}N couples of specific chemical group (i.e. amide ^{1}H-^{15}N for proteins, aromatic ^{1}H-^{13}C couples in position 6/8 or pyrimidines and purines, respectively). In the simplest terms, a 2D HSQC spectrum is composed of a series of 1D spectra for a given nucleus (i.e. ^{1}H). The second dimension is built up by using RF pulses that *transfer the magnetization* between the first nucleus type covalently bound to the second type (i.e. ^{13}C or ^{15}N). This second, *indirect*, dimension is acquired as a function of the frequency of the second nucleus and is expressed in ppm.

Lastly, the nuclei magnetization induced by B_0 is possible for those nuclei whose nuclear spin number is $I \neq 0$. In RNAs, naturally occurring NMR *active* isotopes are ^{1}H and ^{31}P ($I = +½$), and ^{14}N ($I = +1$) while ^{12}C and ^{16}O are NMR *silent* nuclei ($I = 0$). To overcome the limitations imposed by the quadrupolar nucleus of ^{14}N, the nitrogen ^{15}N isotope ($I = -½$) is generally used. Similarly, given its high gyromagnetic ratio and the abundant presence of carbon atoms in the nucleotides, the NMR active isotope, ^{13}C ($I = +½$) is used. Since these two isotopes are not enriched naturally, several methods, ranging from heterologous expression to chemical synthesis, have been developed to incorporate isotopically labelled nucleotides in RNA samples[93,94].

1.2.1 RNA sequence-specific resonance assignment

Resonances assignment is the first and necessary step for any subsequent investigation using NMR. This is the procedure by which each peak, in a given spectrum, is assigned to the nucleus, in the molecule of study, that gives rise to that peak. For biopolymers this is often achieved by the use of 2D (or higher dimensionality) spectra that correlate nuclei belonging to neighboring monomers[95–97].

Assignment experiments can be loosely classified into two categories, 1) those which correlate nuclei close to each other in space and 2) nuclei connected via covalent bonding. In RNAs, each nucleotide is connected to its neighbor via a phosphodiester bond. Given the poor dispersion of phosphate ^{31}P nuclei chemical shifts, *through-bond J*-coupling based experiments are often insufficient to obtain a complete sequence-specific assignment.

Through-space experiments, instead, such as the Nuclear Overhauser Effect SpectroscopY (NOESY) experiment[98], that exploits the cross-relaxation phenomena, are used to correlate nuclei close in space and obtain a great wealth of structural information. For a typical homonuclear ^{1}H-^{1}H NOESY experiment, ^{1}H nuclei with a up to ~5 Å radius distance from each other give rise to off-diagonal correlation peaks indicating their spatial relationship. In addition, since the intensity of such off-diagonal peaks scales with the distance, NOESY experiments are used, often in combination with other measurables, as semi-quantitative restraints for 3D structure calculation by molecular dynamics. This concept is the foundation of NMR 3D structure determination of biomolecules[99].

For the purpose of this paragraph we'll introduce the experiments that are needed to obtain assignment of the following nuclei: imino ^{1}H-^{15}N 1/3 (Fig. 5), aromatic ^{1}H-^{13}C 2/6/8 and sugar ^{1}H-^{13}C 1' (Fig. 6). These nuclei couples are excellent reporters of base-pairing status (^{1}H-^{15}N) and secondary structure (^{1}H-^{13}C) (i.e. their chemical shifts strongly depend on the conformation adopted by the nucleotide). In addition, the experiments needed to obtain their 2D fingerprint spectra have good sensitivity and their chemical shifts suffer from limited spectral overlap. These nuclei were used as reporters to probe conformational dynamics using $R_{1\rho}$ relaxation dispersion experiments in Paper I. For comprehensive overview of the NMR experiments developed to assign other important nuclei in RNAs, we direct the reader to the following reviews[95–97].

NOESY imino walk: assignment of imino groups and secondary structure determination.

As seen in 1.1.1-2, RNAs secondary structure is intimately correlated with base-pairing patterns. Imino groups, that are present in Gs and Us (Fig. 5a), possess few characteristics that make them ideal reporters of base pairs in NMR. Firstly, ^{1}H chemical shifts of imino groups do not overlap with any other groups in the ^{1}H 1D spectra in RNAs, this property was observed and exploited since the dawn of the field in the 70's[97,100].

Secondly, they possess a labile N-H covalent bond that, if unpaired, can transiently break and reform, allowing for the proton to exchange with the solvent. In this case, due to the exchange rate of the process, imino groups are not detectable using a traditional ^{1}H-^{15}N HSQC or ^{1}H-^{1}H NOESY. On the contrary, if the imino groups are involved in hydrogen bonding, such as the ones in canonical WC base pairs, the exchange rate is significantly slowed down and the imino groups become visible in these spectra[97]. Lastly, ^{15}N chemical shifts of Gs differ from Us, while ^{1}H chemical shifts of Gs and Us base paired in WC configuration differ from non-canonical base pairs (Fig. 5a), making base pair types readily identifiable[97]. In this regard, database surveys of chemical shifts can be exploited for automated assignments and prediction in RNAs for ^{1}H and ^{13}C chemical shifts[101,102]. In Paper I, a systematic search of the Biological Magnetic Resonance data Bank (BMRB) of G–C and G–U base pairs was conducted to validate $R_{1\rho}$-derived chemical shifts.

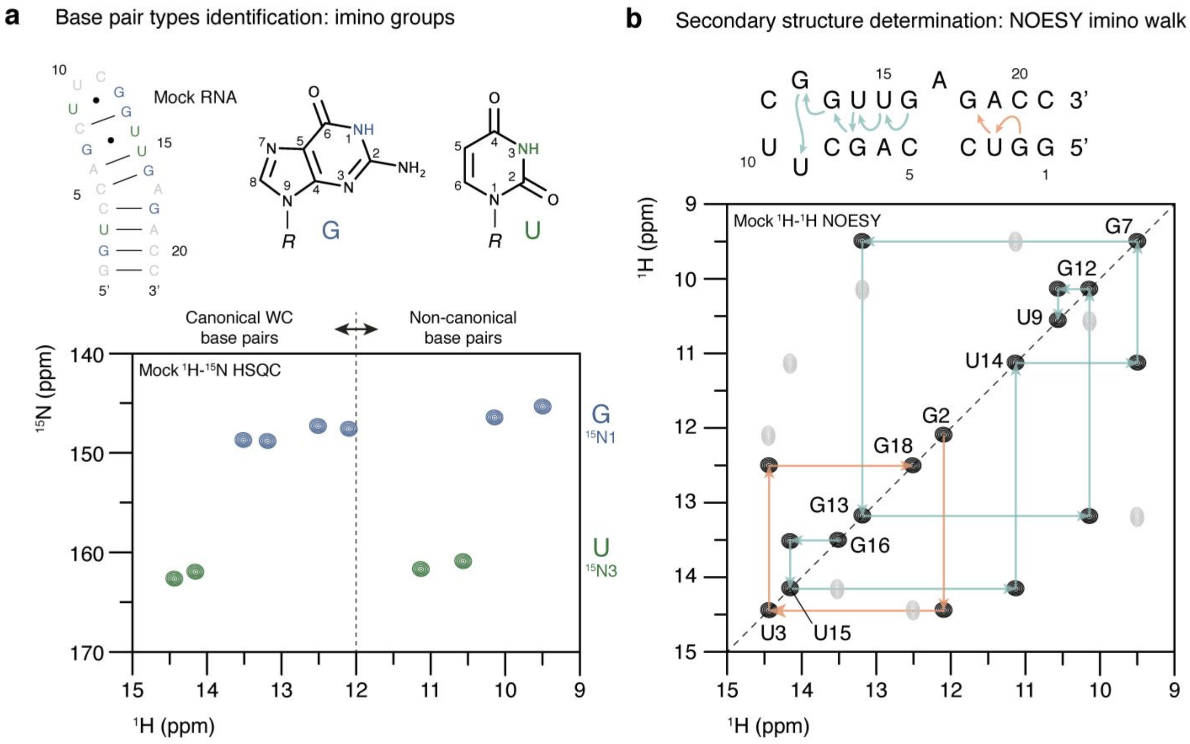


Fig. 5 Base pairs and secondary structure determination by NMR. a, Base pair types are readily identified using ^{1}H-^{15}N HSQC of the Gs and Us imino groups[97]. **b,** The NOESY spectrum of the imino region allows for the identification of imino ^{1}H close in space (off-diagonal peaks), thereby providing a *walk* through the base-pairing pattern. All spectra presented are non-experimental mock diagrams for illustration purposes.

The imino region (14-9 ppm) of the ^{1}H-^{1}H NOESY (Fig. 5b) provides off-diagonal peaks for neighboring base pairs, while the ^{1}H-^{15}N HSQC provide the identity of such base pairs (Fig. 5a). These two experiments, often in combination with COrrelation SpectroscopY (COSY) *through-bond* type of experiments such as the HNN COSY[103] and developments thereof[104], are sufficient to determine the secondary structure of intermediate- and small-sized RNAs. The procedure to assign base pairs to a putative secondary structure, calculated *a priori* by computational means (i.e. using MC-Fold[45]), is depicted in Fig.b. This connectivity can be observed for intra- and inter-strand nts in stable A-form helical segments, while large non-canonical elements and closing base pairs are often too weakly formed to give rise to detectable off-diagonal peaks in NOESY.

NOESY aromatic-sugar walk: assignment of aromatic and sugar peaks and secondary structure elements identification.

The aromatic-sugar walk provides intra- and inter-nts connectivity between aromatic ^{1}H6/8 and sugar ^{1}H1’ (Fig. 6). This region comprises ~7.0–8.0 ppm for the aromatic (Fig. 6a) and

~4.5–6.5 ppm for the sugar nuclei[97] (Fig. 6b). Similarly to the imino walk, this path has stronger and characteristic off-diagonal peaks for nts in A-form helices (Fig. 6c). However, unlike iminos, aromatic and sugar protons do not exchange with the solvent and are detectable in single stranded RNA. Though challenging, non-canonical moites can be assigned if a certain degree of stability is retained (i.e. base-base stacking in internal loops or bulges) and ^{1}H-^{1}H contacts have a sufficiently long lifetime. In addition to the ^{1}H-^{1}H NOESY, *through-bond* experiments are used for intra-nt ^{1}H6/8- ^{1}H1’ couples determination, such as the case of HCN triple resonance experiments[105].

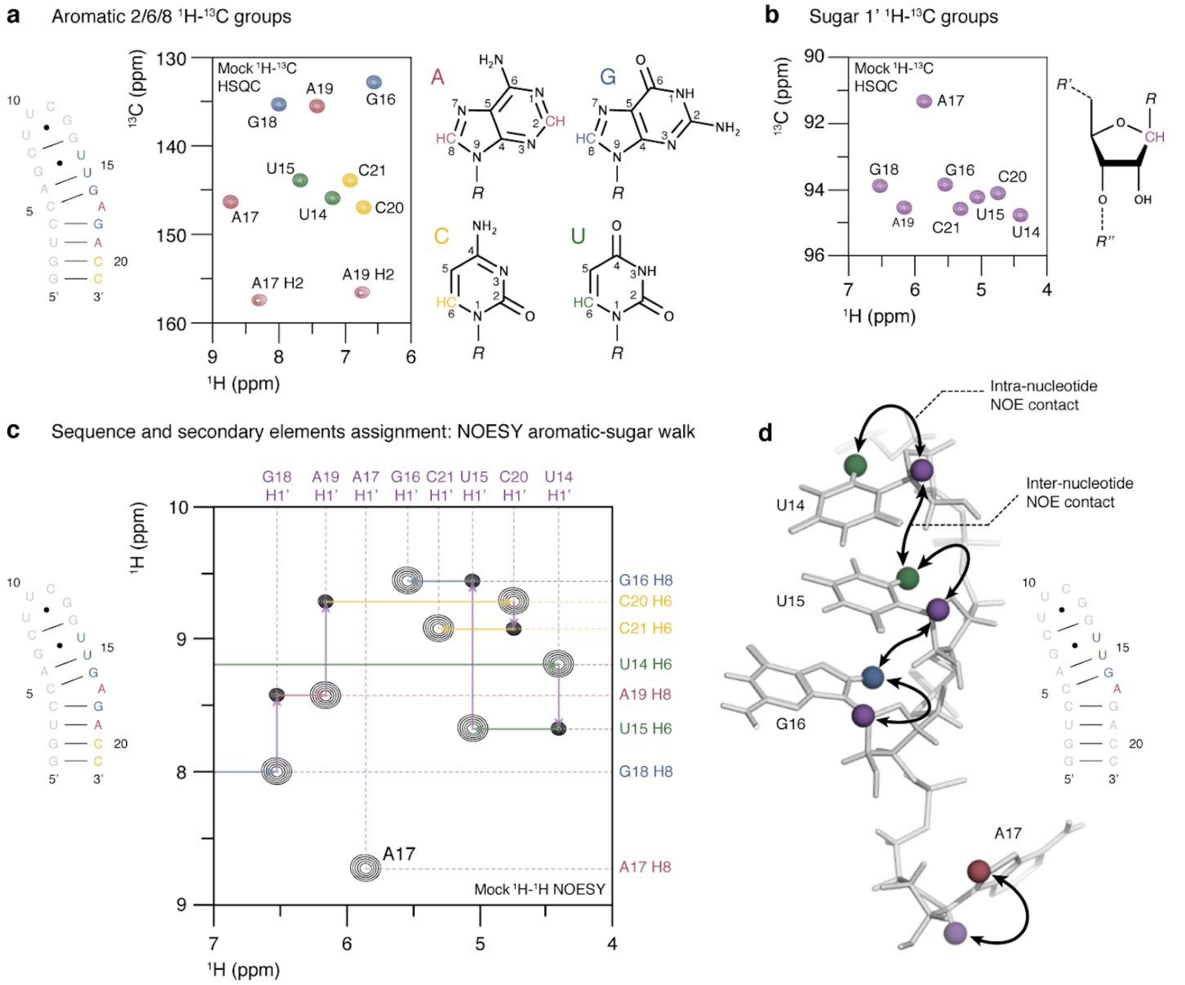


Fig. 6 Sequence specific assignment of aromatic and sugar atoms. a, Aromatic 2/6/8 ^{1}H-^{13}C groups. **b,** Sugar 1’ ^{1}H-^{13}C group. **c,** The NOESY aromatic (7-10 ppm, vertical) and sugar (4-7 ppm, horizontal) region allows for the identification of ^{1}H close in space. Each cross-peak indicates an intra- or inter-nucleotide NOE contact. **d,** 3D model of the example hairpin construct generated using MC-Sym[45]. intra- or inter-nucleotide contacts are indicated by the black arrows. Unpaired nucleotides such as A17, often interrupt the

aromatic-sugar walk, making complete assignment particularly challenging. All spectra presented are non-experimental mock diagrams for illustration purposes.

Ultimately, imino and aromatic-sugar walks are interconnected and the cross-talk between these two paths via Cs aminos, As ^{1}H2 and Cs/Us ^{1}H5 protons is essential to complete the assignment. Ambiguities in assignment that are arising from spectral overlap and chemical/conformational exchange processes can be resolved by means of differential experimental conditions, such as temperature, divalent cation concentration (i.e. Mg^{2+}) or pH. In addition, selective isotope labelling schemes[93], in combination with heteronuclear edited/filter NOESY[106], are often essential for the complete assignment of highly dynamic or larger RNAs. Among others, decrowding by means of homonuclear *J*-coupling, in unlabeled RNAs[107], and ^{13}C direct detection[108] have also shown interesting avenues for the assignment of complex RNAs.

1.2.2 Dynamics probing techniques

As seen in 1.1.3, RNAs conformational dynamics spans across several orders of magnitude in timescales of motion. Similarly, NMR methods have been developed to detect chemical or conformational exchange across a near span. Before introducing these experiments and their application, we need to present a qualitative description of exchange processes in relation to the NMR timescale.

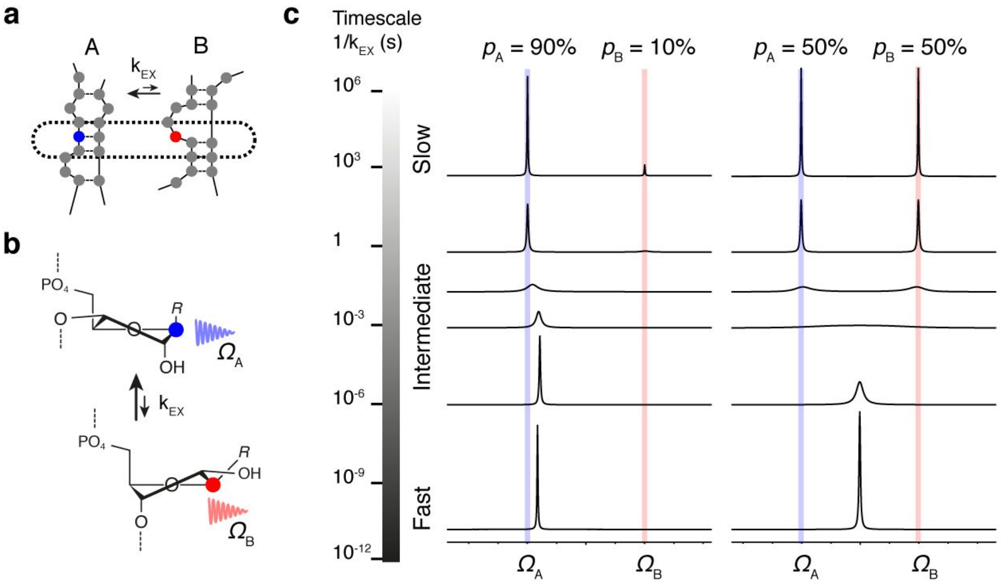


Fig. 7 Conformational exchange and the NMR timescale[109]. a, Tier 1 base pair rearrangement that occurs with a given exchange rate (k_{EX}). The colored nt exchanges

between a paired (blue) and unpaired (red) conformation. **b,** During the exchange process the ribose undergoes a sugar pucker interconversion. In the two states, the C1’ nucleus resonates at two different NMR frequencies Ω_A and Ω_B. **c,** The observed nucleus gives rise to different spectra depending on the relationship between k_{EX} and $\Delta\Omega$, as well as the relative population of the two states (p_A and p_B).

The simplest example is represented by two chemical or structural conformations, exchanging with each other, of a given nucleus (Fig. 7a–b). In this example the parameters describing the resulting NMR spectrum are the difference in chemical shift between state A and B ($\Delta\Omega = \Omega_A - \Omega_B$) and the exchange rate between these two states ($k_{EX} = k_{AB} + k_{BA}$). In addition, the relative free energy difference between A and B determines their relative populations (p_A and p_B), that can be expressed in relation to k_{EX} as $p_B = k_{AB}/k_{EX}$ with $p_A + p_B = 1$. When the populations of these two states are different from each other, the more populated state is referred to as the Ground State (GS), whereas the other is referred to as the Excited State (ES). Depending on the relationship between $\Delta\Omega$ and k_{EX} three NMR exchange regimes can be defined.

For each exchange regime we’ll mention the relationship between $\Delta\Omega$, k_{EX} and the resulting signal, as well as the major NMR experiments used to characterize RNA motions in that timescale. As one might have noticed, these exchange regimes loosely correspond to the motional tiers described in 1.1.3, therefore a comparison can be made with the conformational transition described above with the following experimental probing techniques. The NMR methods and application examples mentioned below were recently reviewed by our group in a comprehensive work on the subject[109].

Slow exchange $k_{EX} \ll \Delta\Omega$ – The exchange process occurs with a timescale that is slower than the NMR frequency between the two states (Fig. 7c, top). The spectrum results in two distinct peaks for A and B, each corresponding to its own chemical environment and with characteristic Ω_A and Ω_B. However, the lineshape of the peaks is generally not affected by the process and the intensities are a direct readout of the relative populations[109]. This regime corresponds to seconds or slower exchange rates and the peaks corresponding to the two states resonating at Ω_A and Ω_B are readily observable in fingerprint spectra. Folding and unfolding, ribozyme catalysis and riboswitch transitions have been followed using real-time NMR, EXchange SpectroscopY (EXSY) and hydrogen/deuterium (H/D) exchange experiments[109].

Intermediate exchange $k_{EX} \approx \Delta\Omega$ – The exchange is comparable to the NMR timescale and occurs in the µs-ms range (Fig. 7c, center). In the intermediate-slow regime, the lineshape of the peaks is affected by the exchange process, resulting in the so-called motional broadening phenomenon. When k_{EX} approaches the coalescence point ($k_{EX} = \Delta\Omega/(2\sqrt{2})$), the individual peaks broaden into one another, resulting in a single broad peak. When k_{EX} surpasses the coalescence point, in the intermediate-fast regime, the motional narrowing phenomenon results in a single narrower peak, whose chemical shift approaches the average chemical shift value in a population-weighed fashion. When populations are highly skewed, the ES broadens beyond detection in the intermediate-slow regime, therefore becoming *invisible* in the typical 1D or 2D fingerprint spectra. Similarly, in the intermediate-fast regime, the visible single peak is only minimally shifted from the GS chemical shift value Ω_A. The experiments described below employ different approaches to directly probe and characterize these ESs that correspond to tier 1 motions (1.1.3).

Chemical Exchange Saturation Transfer (CEST) experiments are used to probe slow to intermediate-slow processes with highly skewed populations[109,110]. CEST experiments make use of a spin-lock (SL) low-power RF pulse that *locks* the nuclei magnetization it is resonant with, during the time it is applied. In the simplest terms, this SL saturates these resonances leading to a loss of signal. Several experiments, where the normalized intensity of the observable GS state is presented as a function of different SL offsets, are recorded. If the SL is *on-resonance* with the visible GS, a global minimum in the CEST profile is observed. Similarly, if the SL is *on-resonance* with the *invisible* ES excited state, a loss of signal is observed for the GS peak. This is due to the exchange process and an *off-resonance* local minimum is observed in the CEST profile. Fitting these curves to Bloch-McConnel equations[111] gives a quantitative description of the exchange process[109,110]. CEST was used to probe fluoride[112–114] and SAM II[115] riboswitches.

Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments probe intermediate exchange processes over the μs-ms range. The main component of a CPMG pulse sequence is a Hahn's spin echo[116] block that *refocuses* the magnetization after a time delay, by means of a 180° pulse. In the case of two exchanging states (i.e. Fig. 7), part of the magnetization is not refocused by the spin echo and results in a loss of signal. This signal loss depends on two phenomena, 1) R_2 relaxation (that will be briefly introduced in the next subparagraph) and 2) the actual exchange between state A and B. Fitting the signal intensity as a function of the number of spin echo blocks to a monoexponential function, allows for determination of the effective R_2 ($R_{2\mathrm{eff}}$) relaxation rate constant. To rule out the R_2 contribution to $R_{2\mathrm{eff}}$, and obtain information on the exchange process, a train of spin echo pulses can be applied with variable frequency (υ_{CPMG}) and a relaxation dispersion curve is obtained by plotting $R_{2\mathrm{eff}}$ as a function of υ_{CPMG}. Subsequent deconvolution of exchange parameters (k_{EX}, populations and $\Delta\Omega$) requires collection of datasets at different B_0 fields and solution of analytical expression derived from the Bloch-McConnel equations [109,117]. Due to the nature of 180° pulses, typical schemes to avoid spectroscopic artifacts cannot be implemented in CPMG experiments, making the analysis of RNAs particularly challenging[109]. However, position-selective labeling of RNAs has been a successful route to overcome these limitations by the Kreutz[118–122], Dayie[113,123,124] and Hoogstraten[125] groups. Of particular interest for the results of Paper I, the Dayie group applied CPMG to study the dynamics of the miR-1224 binding site in the *CCR*5 RNA pseudoknot showing that the pseudoknot samples at least three conformations prior to miR-1224 binding, with one of the minor populated states potentially providing a dynamic recognition site for interaction partner[123].

$R_{1\rho}$ relaxation dispersion experiments, detect exchange in the μs-ms range. In these methods, a SL pulse is used to detect the exchange process in a similar way that the spin echo trains are used in CPMG. The signal intensity of a given nucleus is measured as a function of increasing times in which the SL pulse is applied. The relaxation rate constant $R_{1\rho}$ is obtained by fitting intensity vs. time plot to a monoexponential function. If the SL applied is sufficient, in terms of strength/frequency, to make the difference in effective field for both spins negligible, then the obtained value $R_{1\rho}$ has only a R_2 contribution, that is an intrinsic characteristic of the spin system observed. On the other hand, if SL is not sufficient to cover both Ω_A and Ω_B, then an additional relaxation contribution, this time intrinsic of the exchange process (R_{EX}), is detected on top of R_2. Therefore measured $R_{1\rho}$ will be the sum of R_2 and R_{EX}. Similarly, if no exchange is occurring, there will be no R_{EX} contribution, and every $R_{1\rho}$ measured for each SL strength will be equal to R_2. Typically, $R_{1\rho}$ values are plotted as a function of the SL strength applied, resulting in relaxation dispersion curves. Intuitively, decay in the dispersion curve will indicate the presence of exchange, whereas a flat dispersion profile, will report absence of exchange in the timescale probed. The *on-resonance* relaxation dispersion curves can be fitted to analytical expressions of the Bloch-McConnel equations, depending on the time sub-regime probed, to quantify k_{EX}[109,126,127]. Additional *off-resonance* experiments, where an offset is applied to the SL carrier, are recorded to obtain the populations of the states and their chemical shift difference $\Delta\Omega$[109,126,127]. ^{13}C and ^{15}N $R_{1\rho}$ have been applied to a variety of RNAs, chiefly by the Al-Hashimi group[109] as previously mentioned for tier 1 motions. These experiments, in conjunction with ^{1}H $R_{1\rho}$[107], were also used to characterize the conformational exchange process in a miRNA–mRNA pair in Paper I.

Fast exchange $k_{EX} \gg \Delta\Omega$ – In this regime the exchange rate constants are in the ps-ns range (Fig. 7c, bottom). These exchange processes in tier 2, induce local fluctuation in the magnetic field, that affect the intrinsic longitudinal (R_1) and the transverse (R_2) relaxation rate constants of each spin. These constants dictate the rates by which the spin magnetization relaxes back to the thermal equilibrium, from a non-equilibrium state, on the *z* axis (R_1) and the *xy* plane (R_2), in the vector model. In addition to R_1 and R_2, cross-correlated relaxation phenomena between nuclei close in space, such as NOE and ROE, are affected by these motions. These perturbations happen mainly through dipole-dipole interactions (i.e. dipolar coupling) and chemical shift anisotropy, among others[109]. Fast exchange processes result in a relatively narrow, single peak whose chemical shift is the population-weighted average of $\Omega_A$and Ω_B. Using solution-state NMR it is not possible to directly obtain the exchange parameter k_{EX}, populations and $\Delta\Omega$, as for the methods described above. Rather, R_1, R_2, heteronuclear NOE, and off-resonance ROE can be measured to quantify the amplitude of the motions through an *order parameter* and correlation times[109]. These experiments are generally described as spin relaxation methods and have been widely used to probe fast dynamics in RNAs[109]. In addition, Residual Dipolar Couplings (RDCs) can be measured if a medium is introduced in solution, so that the molecules become partially aligned to B_0 and dipole-dipole bond vectors are not averaged to zero, due to stochastic tumbling. Unlike other observables, RDCs are sensitive on a larger timescale (ps-ms), therefore their interpretation has been cumbersome[109]. RDCs are commonly used for 3D structure calculation, however, as seen in 1.1.3, these observables have been used, in combination with others, to infer inter-helical motions in the HIV-1 TAR RNA[86,128].

1.3 Micro RNAs

A particular group of ncRNA-protein complexes (RNPs) gained prominence during the last two decades due to their important biological functions as well as their potential in therapeutic and biotechnological applications. These are RNPs that use their RNA component as a *guide* to find specific RNA or DNA targets through base pair complementarity[129]. RNA-guided RNPs are present in all three domains of Life, where they are involved in multiple and diverse aspects of cellular functions. Among others, the broad group of RNA-guided RNPs, include small RNAs associated with Argonaute proteins and their homologs[130], bacterial small RNAs[131–133] and CRISPR RNAs[134,135].

In this paragraph we'll introduce the features of a class of RNA guides that act, in complex with their partner Argonaute proteins, as post-transcriptional gene expression regulators: the microRNAs (miRNAs).

Our discussion will be limited to the role of miRNAs in the eukaryotic kingdom of Animalia (i.e. Metazoans)[136] and their relevance to human diseases. However, one must keep in mind that 1) in Animals, other similar RNAs are present (i.e. siRNAs, piwiRNAs)[137], 2) miRNAs are ubiquitous in other eukaryotic kingdoms (i.e. Plantae)[130], 3) that Argonaute-associated small-RNA/DNAs are also found in prokaryotic and archeal organisms[130]. These variations on the theme all possess distinct biological functions, however, a certain degree of structural similarity is conserved among Argonaute-associated small-RNA[130]. Therefore, where relevant, a comparative approach will be taken in discussing the structural similarities and differences between these complexes.

1.3.1 Biogenesis and function

MicroRNAs (miRNAs) are small (~22 nt.), ncRNA molecules that play a central role in post-transcriptional gene expression by regulating mRNA levels and their translation. First discovered in nematodes[138,139] miRNAs were subsequently found to be a widespread class among many organisms, including *Homo sapiens*[140–143] where their genes count for 1-2% of the total number of protein coding genes[144–146]. Many miRNA gene loci are found to cluster in specific genomic regions that can be transcribed simultaneously as poly-cistronic units[147]. Canonical human miRNAs genes are transcribed by Polymerase II to a large pri-miRNA precursor bearing a 5'-end cap and a 3'-end poly(A) tail[148] (Fig. 8a, step 1). These pri-miRNA precursors undergo nuclear processing by the enzyme Drosha that performs a double endo-nucleolytic cleavage that results in a ~70 nts pre-miRNA hairpin[136,149–153] (Fig. 8a, step 2). Pre-miRNAs are then translocated from the nucleus by the Exportin5-RAN-GTP complex to the cytoplasm[154] (Fig. 8a, step 3). Dicer, in complex with transactivation response element RNA-binding protein (TRBP)[153,155,156], further process pre-miRNAs by endo-nucleolytic cleavage to an asymmetric RNA duplex containing the 22 nts guide and a partially complementary passenger strand [136,150–153,157–159] (Fig. 8a, step 4). Depending on their orientation in the pre-miRNA stem, guide and passenger can be referred to as miRNA-5p and miRNA-3p. Characteristic features of guide–passenger duplexes are the phosphorylation at both the 5'-ends, the two 2 nts 3'-end overhangs and the presence of non-canonical base pairs, mismatches and single bulges throughout the central stem. The guide–passenger duplex is subsequently loaded into the Argonaute protein (Ago 1-4) to form the RNA-Induced Silencing Complex (RISC), through a mechanism that is only partly understood[153,160,161] (Fig. 8a, step 5). The last steps toward a mature RISC involves the removal of the passenger strand and was proposed to be accomplished either by a step-wise conformational change of the protein that expels the thermodynamically unstable RNA strand (Ago1, Ago3 and Ago4 isoforms) or through the intrinsic endo-nucleolytic activity of the Ago2 isoform[153,160,161].

Ultimately, the mature RISC interferes with gene expression by binding to partially complementary binding sites in target messenger RNAs (mRNAs), typically located in the 3' UnTranslated Region (UTR) (Fig. 8b). Target repression is achieved by mechanisms mediated by the TriNucleotide Repeat-Containing gene 6 proteins (TNRC6) of the GW182 family. RISC-bound TNRC6, associates with the Poly(A)-Binding Protein (PABPC) and deadenylase complexes PAN2-PAN3 and CCR4-NOT. These complexes lead to irreversible silencing of the transcript by deadenylation and subsequent decapping and 5'-to-3' decay. In addition, TNRC6 recruits effector complexes that mediate translation inhibition of the mRNA translation at the initiation stage[136,162,163] (Fig. 8b).

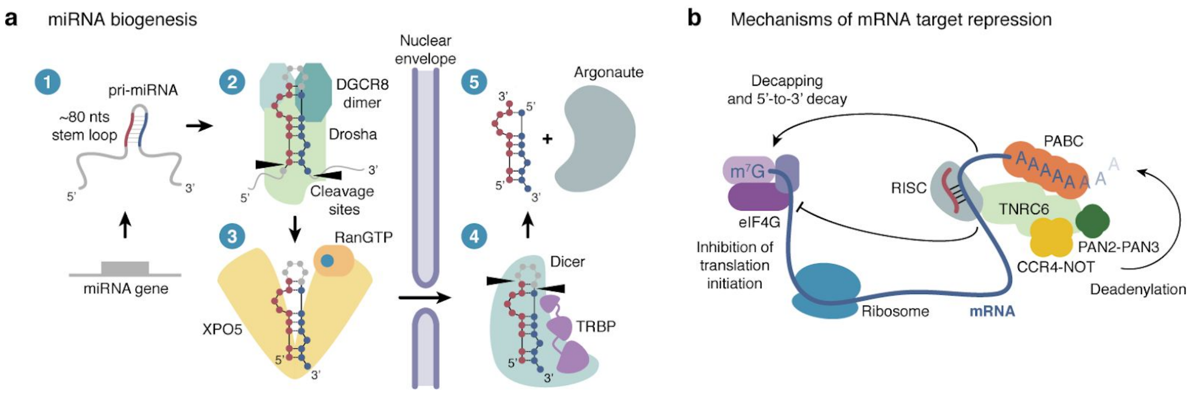


Fig. 8 MiRNAs biogenesis and function. a, 1) The primary miRNA (pri–miRNA) transcript is transcribed. A stem-loop structure contains a 5p (red) and 3p (blue) arm. 2) DGCR8/Drosha complex processes the pri–miRNA to precursor miRNA (pre–RNA). 3). The RanGTP/XPO5 complex translocates the pre–miRNA to the cytoplasm. 4) TRBP/Dicer complex processes the pre–miRNA resulting in the guide–passenger duplex. 5) The duplex is loaded into the Ago protein and the passenger strand is ejected. The mature RNA-Induced Silencing Complex (RISC) is formed by the guide RNA (red) and Ago. **b,** RISC interacts with partly complementary binding sites in the 3'-UTR of a target mRNA and induces its repression through mRNA decay and inhibition of translation initiation.

1.3.2 mRNA binding motifs

Many exceptions to the biogenetic pathway presented above have been reviewed[136,150–153,160,161,164]. However, all of them converge to a picture where the mature RISC embed the guide in such a conformation that the miRNA bases are partially exposed to the solvent and able to base-pair with target mRNA sites[165,166]. Unlike siRNAs, where full Watson-Crick (WC) complementarity is observed between guide and target sequences, miRNA–mRNA complexes retain WC base pairing between nucleotides 2-8, the so-called *seed* region[151]. Outside of the seed, the secondary structure configuration is not perfect and can include non-canonical pairs, mismatches, single bulges and internal loops[151]. Evolutionary conservation and experimental data made it possible to classify binding sites and their efficacy in mRNA repression according to their base pairing patterns [136,151] (Fig. 9a). Canonical sites are considered when the seed region is ≥ of 6 base-paired nts (6mer) and suffice to the silencing activity (Fig. 9a, top). Atypical are instead evolutionary conserved sites where additional base pairing at the 3'-end is present[136] (Fig. 9a, center). The 3'-end pairing is centred between position 13-16 of the guide strand, with a conserved stretch of ≥ 3-4 canonical WC base-paired nts. When the seed sequence is perturbed by a mismatch or a single nucleotide bulge the 3' pairing is considered to be non-canonical *compensatory*[136] (Fig. 9a, bottom). Atypical sites constitute around 5% of the seed-matched sites[136,167] and are proposed to contribute only marginally to binding affinity and target repression[136,168–171], despite evidences of their role in targeting specificity[172,173]. However, the role of 3’-end pairing is just beginning to be uncovered, from a structural perspective[174,175].

Among other features, prediction algorithms, such as TargetScan[170], rely on the seed region stability and 3’-end pairing as the main secondary structure determinants of the miRNA–mRNA pair when predicting binding sites at the genomic level. This approach, albeit proven to be functional, does not take into account the inherent dynamic nature of RNAs and the possible base pair rearrangements that could take place in the context of the RISC, leaving room for improvement.

High resolution structural investigations that go in this direction are represented by the studies of the Plavec group on the miRNA let-7 binding sites on lin-41 mRNA[176,177]. Let-1 Complementary Sites (LCS) LCS1 and LCS2 on lin-41 mRNA are 3’-compensatory, where LCS1 bears a single nucleotide A bulge[177] and LCS2 a G–U wobble base pair[176] in the seed, respectively. Both present an asymmetric internal loop with predicted unmatched stretch 5’-GUU-3’ in LCS1 and 5’-AUU-3’ in LCS2 and 5’-UA-3’ in the let-7 miRNA. Despite the high similarity between the two binding sites, the 3D structure obtained from the solution NMR could reveal some differences, particularly in terms of the conformational stability of the internal loop. Where LCS2 shows a major stable conformation of the central bulge, stabilised through a base triplet[176], LCS1 features the characteristics of a flexible RNA sampling multiple conformers[177]. The secondary structural elements perturbing the seed region do not show any sign of disruption of the A-form RNA helix in the lower stem in both constructs used. Despite the lack of further studies that could shed light on the link between the different structures and variations in functional activity, this seminal work indicates that target sites clustered in the same groups by prediction algorithms can show rather different structural behaviour.

1.3.3 RNA-Induced Silencing Complex

Argonaute (Ago) proteins are part of the P element Induced WImpy testis (PIWI) protein superfamily. These are present in eukaryotes (eAgos) and prokaryotes (pAgos), distributed among different families, with different functions[130]. eAgos involved in the miRNA pathways belong to the Ago-like family[130]. In humans there are eight isoforms, four of which are largely expressed in many tissues and studied in greater detail (hAgo1, hAgo2, hAgo3 and hAgo4)[137]. eAgos are often classified according to their ability to enzymatically cleave the phosphodiester bond of the sugar-phosphate backbone of the target. This enzymatic reaction is often referred to as *slicing* and it is only possible when extended WC base pairing between guide and target RNA is achieved (i.e. siRNA pathways). hAgo2 is the only human isoform to be able to slice the targets, whereas hAgo1, hAgo3 and hAgo4 are not[178].

Historically, pAgos X-ray crystal structures were solved first[179–181], opening to the high resolution investigation of this class of proteins. From a structural point of view, all the members of the PIWI superfamily are characterized by the presence of a PIWI domain and, optionally, the PIWI-Ago-Zwille (PAZ) domain[130]. For the purpose of this paragraph we'll discuss the domain architecture of those Agos that are most similar to hAgo2, studied in Paper I. These Agos are composed of two lobes: the PAZ lobe comprising N, L1 and PAZ domains and the PIWI lobe with the PIWI and MID domains. PAZ and PIWI lobes are connected by a linker region called L2 (Fig. 9b).

The MID domain presents a nucleotide binding pocket involved in the recognition of the 5'-end nucleotide of the guide strand (g1) during loading. Crystal structures showed that the phosphate group of g1 is locked in a stable conformation coordinated by several contacts with conserved amino-acids[182–185]. The base of g1 is also selected with preference of U/A over G/C[185], and locked in a conformation that prevents it from forming base pairing with target bases[186–188]. The PIWI domain presents a RNaseH type of fold where substrate cleavage takes place in slicing Agos thanks to a DEDX catalytic tetrad[187–190]. During RISC loading the N domain acts as a wedge in displacing the passenger strand and obstructs the nucleic acid binding channel that otherwise could accommodate base paired duplex up to the 3'-end of the guide[160,161]. The PAZ domain binds the 3'-end of the guide [191–195] and protects it from exonucleolytic degradation[196].

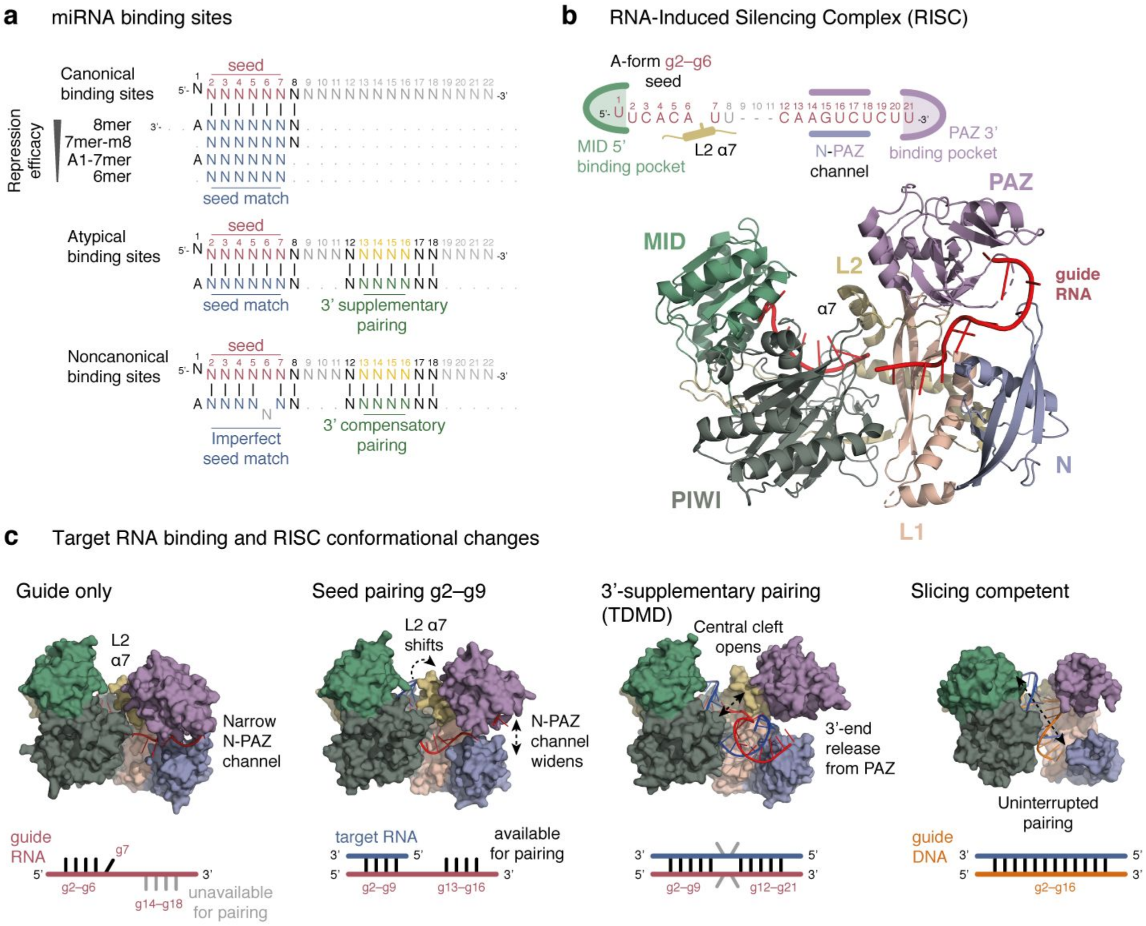


Fig. 9 MiRNA binding sites and structural details of RISC. a, Canonical, atypical and non-canonical binding sites[136]. **b,** Crystal structure of guide-loaded hAgo2 (pdb 4w5n)[188]. The inset shows the constraints imposed by the protein to the RNA in order to ensure efficient target screening. **c,** Different crystal structures of guide and guide–target Ago complexes. All complexes were aligned to the PIWI domain of guide-only structure. Guide only[188]. Seed pairing g2–g9, pairing to the target induces α7 shift and widening of the N-PAZ channel in hAgo2(pdb 4w5o)[188]. 3'-supplementary pairing of TDMD target (bu2, pdb 6mdz)[175]. Propagation of 3'-supplementary pairing induces 3'-end release from the PAZ domain and opening of the central cleft in hAgo2. Slicing competent conformation of *T. thermophilus* pAgo (pdb 3hm9)[197]. Fully complementary guide DNA and target RNA forms an uninterrupted duplex. b–c Few nts were omitted from the cartoon representation here, although present in the original structures.

In hAgo1–2[165,166,189,190] the MID-PIWI lobe constitutes the binding site for the seed region nts g2–g8. The binding takes place between the amino-acids and the guide's phosphates and 2'-OH groups, making this protein-RNA interaction mode sequence independent (Fig. 9b). The WC edges of the bases are exposed to the solvent for base pairing with the target. The topology of this narrow channel positions the seed g2–g6 in a A-form helical conformation. This preorganization of the seed is proposed to reduce the entropic cost of the duplex formation and increase binding affinity[150,165]. A conserved feature of the X-ray structure of eAgos[165,166,187,189,190] is the presence of a isoleucine side-chain of L2 α-helix 7 (α7) inserted between g6 and g7 of the guide RNA. This intercalation disrupts the pre-formed helix of the seed and induces a kink in the backbone of the RNA suggesting that the base pair between guide and target beyond g6 would require α7 to be displaced from that position[130]. g7–g9 are stacked and a kink stabilised by an arginine sidechain between g9 and g10 interrupts this short stretch[166].

Beyond g10, the guide RNA seems to be rather flexible resulting in poor electron density of these crystallographic structures, where transients or no interaction with Ago occur. The 3'-end of the guide, stabilized by its interaction with the PAZ is observed for g21–g22[165,189,190] or g17–g20[166], in hAgos.

These evidences describe a complex machinery that, in principle, can deal with a variety of RNA targets, regardless of their base pairing with the guide beyond the seed. The region between g5 and g10, despite being well characterised in the X-ray structures, elicits some questions on the topology of the ternary complex with target RNAs and how different non-canonical base pair patterns can be influenced or influence this site. These questions are partly answered by the structure of hAgo2 in complex with a guide RNA base-paired to a target up to g9 solved by the MacRae group[188].

hAgo2 interacts with the A-form seed helix through the minor groove, explaining why G–U wobble base pairs in this region are less tolerated[188,198,199]. Instead positions g8–g9 do not contact hAgo2 through the minor groove making non-canonical base pairs and mismatches more tolerated in this position[188]. Interestingly, α7 is observed to shift ~4 Å in the target-bound ternary complex, relaxing the kink between g6 and g7 and allowing extended formation of the A-form helical seed. The authors conclude that hAgo2 uses the g2–g5 segment to initially identify targets. Subsequently, the g6–g7 and α7 conformational changes constitute a check-point where the ternary complex is stabilised if WC base-pairing with the target is present. Pairing at g8 enhances affinity, while extension beyond this point does not[188], suggesting that a conformational roadblock in the central cleft prevents further pairing extension. The role of α7 in target recognition was also further studied by the same group, showing that this element of L2 is essential in modulating the kinetics of the ternary complex[200].

In their landmark paper, Schirle *et al.*[188], were also able to solve the structure of the guide nucleotides in hAgo2 that were missing from previous studies[165,166]. Prior to target binding, g14–g18 reside in a narrow channel in the N–PAZ domains with the WC edges of g15–g18 unavailable for binding with the target[188]. Coupled to the α7 shift, upon target binding, the authors observe a conformational rearrangement of the PAZ domain that leads to an opening of the N–PAZ channel[188] (Fig. 9c, center left). The opening in turn remodels g11–g16 now adopting an A-form configuration with g13–g16 WC edges exposed to the solvent and potentially available for pairing with the target[188].

Consistently with biochemical binding kinetics studies of fly and mouse hAgo2[169,171], RISC pre-organizes the guide RNA into modular units to allow for this stepwise binding mode.

These and other works were recently reviewed by Bartel, who proposed a “unified model for miRNA target recognition and pairing propagation” in which 3’-end pairing nucleates at g13–g16 to then propagate to the rest of the target. The extended 3’-end pairing induces the release of the guide 3’-end from the binding pocket, reaching, for fully complementary sites, a *cleavage compatible*, fully stacked conformation as observed for *T. thermophilus* pAgo[197,201]. As central g9–g12 pairing was shown to contribute little or negatively to binding affinity and kinetics[169,171,188], the formation of such a complex would require a further conformational change of the protein central cleft, to allow for pairing beyond g9[136,171,188].

Additions to this model can be inferred from a few key papers published in the past two year. One paper using next-generation sequencing RNA Bind-n-Seq[202] from the Bartel group[203] and one using a massive parallel imaging technique[68] from the Greenleaf and Zamore groups[204] revealed additional sequence determinants to binding and repression, by probing thousands of different target sequences and dramatically raising the throughput of previous experiments[169,171,188,200,205–208]. For the scope of this paragraph we will only mention that 1) the primary role of canonical sites is confirmed[203], that 2) the 3’-supplementary binding is a prerequisite for subsequent binding of the g9–g12 stretch[203], and that 3) RISC can accommodate large target bulges (up to 7 nts) in the central region to allow for 3'-supplementary pairing, without disrupting binding affinities[204], confirming a previous report on this feature[174].

Similarly, two papers from the MacRae group[174,175] were recently published, shedding more light into the role of the 3’-end pairing from a structural perspective. The X-ray crystallographic structure of the catalytically inactive mutant (D669A) of hAgo2 ternary complex solved by Sheu-Gruttadauria *et al.*[174], shows that the 3’-supplementary helix is accommodated in a supplemental chamber in the N–PAZ channel, where g13–g16 are observed to be paired with the target, maintaining a certain degree of mobility[174]. The loops of L2 and PIWI constitute the conformational roadblock preventing seed and 3’-end helix from connecting[174]. In this region, in the seed chamber g9 is observed in *syn* conformation, g10–g11 are not determined, while g12 is in the supplemental chamber available for pairing but impaired by the L2 loop[174]. The 3’-end of the guide is anchored to the PAZ domain[174]. Interestingly, the target sequence has full WC complementarity to the guide that suggested the formation of a fully stacked conformation[197,201], however this structure is not adopted by the complex[174]. In this work the authors identify L2 and PIWI loops as the second conformational check-point imposed by hAgo2 on 3'-end supplementary pairing, reinforcing the model of 3'-end nucleation and propagation through the opening of the central cleft, coupled to PAZ disengagement[136,174]. The role of g9–g12 pairing remains controversial as the central cleft would require more opening than observed to connect the two helices[174]. In addition, they discuss that since the supplementary chamber can only take up to 5 base pairs, more extended 3'-end pairing would require a yet different conformation of the complex[174]. Indeed, this arrangement was observed in the subsequent paper[175], where Sheu-Gruttadauria, Pawlica *et al.*, studied the mechanism by which extended base-paired target can induce Target-Directed MiRNA Degradation (TDMD), a pathway that leads to miRNA decay mediated by 3'-end tailing and trimming[209,210]. In these structures, the authors identify a third conformational checkpoint that couples 3'-end release of the guide from the PAZ domain to the opening of the central cleft, to form a continuous channel, connecting seed to supplemental chambers[175] (Fig. 9c, center right). The central mismatches in the duplexes allow for a kinked orientation of the two stems with respect to each other, with the 3'-end of the guide exposed for TMDM-mediated modification[175]. Even in this open conformation, a modelled fully extended duplex shows steric clashes with PIWI, L1 and N, indicating that a slicing competent conformation is yet to be determined[175].

1.3.4 p53, Sirt1 and miR-34a feedback loop

In Paper I, we focus on the interaction between miR-34a and the Silent Information Regulator 1 (SIRT1) mRNA. SIRT1 is a NAD-dependent deacetylase enzyme that regulates apoptosis by keeping p53 in its inactive (de-acetylated) form in normal cellular conditions[211]. Upon DNA damage or cell stress, p53 transcription and post-translational modifications are upregulated, including acetylation. The active form of p53 up-regulates the transcription of miR-34a which in turn down-regulates the translation of SIRT1 mRNA (mSIRT1). This negative feedback loop composed of p53–SIRT1–miR-34a was found to be essential for cell proliferation regulation and any disruption of the miR-34a–SIRT1 axis could potentially increase the risk of cancer development[212,213]. The binding site of miR-34a on mSIRT1 has been identified and validated[212] showing that the predicted secondary structure recapitulates the features of 3'-supplementary binding sites, mentioned in 1.3.2.

Transcription of miR-34a is ubiquitous in mice[214] and its transcription has been shown to be under the control of the tumor suppressor transcription factor p53[215–219]. Upon cell stress miR-34a regulates a variety of cancer-related pathways by targeting multiple mRNAs and ultimately leading to cell cycle arrest and apoptosis[220]. MiR-34a has been proposed as a potential new agent in tumor-suppressive therapies[221,222].

1.4 Remarks on RNA sample preparation

Our ability to study biomolecules *in vitro* heavily depends on the quality of the sample reconstituted in the test tube. This is generally true for structural, biochemical and biophysical techniques that aim at characterizing molecules or molecular complexes outside of their native biological milieu, in minimal and controlled conditions. In biomolecular NMR, the *quality* of a sample can be judged by two main criteria: quantity and homogeneity.

1.4.1 Quantity

Being NMR a relatively insensitive technique, the molecular concentration in the test tube is a critical factor to record spectra with sufficient signal-to-noise (SN) ratio, in reasonable experimental time. Typical fingerprint 2D HSQC spectra of small RNAs (<50 nts) require samples ranging from the tens of μM to the low mM in concentration and dissolved in a volume of few hundreds of μl to be recorded in less than 1 h, with sufficient SN ratio and resolution. However, assignment (1.2.1) and dynamics (1.2.2) measurements usually require longer data collection routines and producing a sample at the optimal (and often maximal) concentration drastically reduces the experimental time needed and maximizes the usage of the spectrometer.

1.4.2 Homogeneity

RNA samples can be produced using three major approaches: chemical synthesis, *in vitro* transcription and *in vivo* expression[94]. While each method has its own advantages and disadvantages, all of them fail to produce a completely homogeneous sample in terms of chemical and conformational purity. Therefore, downstream purification methods are essential to achieve the wanted grade of homogeneity. For relatively short RNAs, chemical impurities that require most attention during purification are represented by RNA species that are different from the wanted product by a few nts in length (i.e. +/- 1nt). These species are often the result of the incomplete chemical synthesis or the pitfalls intrinsic of the transcription system used (i.e. T7 RNA polymerase). Conformational heterogeneity, instead, represents an issue for RNAs that cannot be completely refolded to their native state *in vitro*, forming misfolded species, multimeric complexes and aggregates.

In addition to the common hurdles of RNA preparation, NMR samples often require expensive building blocks enriched in ^{13}C and ^{15}N isotopes (i.e. nt triphosphates and phosphoramidites). Therefore, methods to produce RNA samples for NMR spectroscopy are often tailored to minimize sample loss while maintaining the highest standards of purity.

The production of RNA samples in quantity and homogeneity amenable for NMR studies is the subject of Paper II and a recent review[94] from our group. For the reader interested in this subject, in this latter work, we report the standard and most advanced methods to produce and purify RNAs for structural studies.

2 AIMS

Collectively, the papers compiled in this thesis aim at contributing to the understanding of the role of conformational dynamics in RNAs function.

2.1 Paper I

Given the sparsity of structural information on the RNA component of RISC, in miRNA function, we set out to describe a model miRNA–mRNA pair involved in a human tumor-suppressive pathway using NMR spectroscopy. The miR-34a–mSirt1 pair, was anticipated to provide a model for centrally bulged atypical 3'-supplementary binding sites. Considering the high degree of flexibility of bulged RNAs, we aimed at quantitatively describing the motion of the miR-34a–mSirt1 pair using $R_{1\rho}$ relaxation dispersion NMR experiments and translate those information in the context of RISC, testing our hypotheses using molecular simulation, biochemical and cell culture assays.

The results of this work were expected to fill the gaps between the static snapshot of RISC structures available, that often miss the flexible RNA component, and provide a framework to better understand miRNAs function.

2.2 Paper II

In this work we aimed at setting up a pipeline for the production of RNA samples for our newly established laboratory. Given the quantity and homogeneity requirements for NMR samples, we set out to design a comprehensive method that could incorporate the latest advancements of T7 *in vitro* RNA transcription coupled to High Performance Liquid Chromatography (HPLC) purification steps, while maintaining the robustness, speed and simplicity of more traditional protocols.

2.3 Paper III

The aim of this paper was to develop a method to improve accuracy and efficiency of Replica-Exchange Molecular Dynamics (REMD) simulation of 3D RNA folding, using sparse experimental constraints. The method developed here, was anticipated to be used to compute the RNA structures in Paper I.

3 RESULTS AND DISCUSSIONS

3.1 Paper I

Using solution NMR spectroscopy we solved the secondary structure of free miR-34a bound to the binding site of mSirt1 mRNA. The complex is composed of a seed and a 3'-supplementary helix, separated by a 4-nts asymmetric bulge. The seed extends up to g6, only forming a transient base pair in position g7. Accordingly with the established nomenclature[136] and given the crucial role of g7 pairing to overcome the first checkpoint in hAgo2[188,200], we termed this site "weak 7mer-A1". We suggested that, given the weakness of pairing in position g7, the site in this conformation might resemble a 6mer, rather than a 7mer-A1, in binding affinity.

Given the bulge size and the broadening observed for a few peaks in the fingerprint HSQC spectra, we probed a subset of nts using $R_{1\rho}$ NMR relaxation dispersion, to determine whether the central region was undergoing any type of conformational exchange in the μs–ms time regime.

We identified a global process involving gG8H1, gG8N1, gG8C8, tU21C6, tC17C1′, tU20C1′, tA19C8, tA19C2 and tA22C8, taking place with an exchange rate of about 1 ms and populating only about 1% of the whole conformational ensemble. Using the chemical shift differences derived from these and additionally measured nts we could draw a model for this excited state conformation. In summary, gG8 repositions form the lower 3'-supplementary to the seed stem, undergoing a GC–to–GU switch, enhancing the seed to form an 8mer site. In addition, using REMD simulations we showed that the ground and excited states adopt two different 3D conformations. We designed a two-point mutation to trap the excited state conformation (trapped excited state) and we showed that this construct retains the same structural features by NMR and REMD.

To assess whether the excited state had any effect on the mRNA repression, we measured the downregulation levels of wild-type mSirt1 and trapped excited state constructs using a Dual-Luciferase Reporter (DLR) assay, upon co-transfection with miR-34a in human cell lines. While the wild-type and trapped excited states maintain the same biophysical properties (as measured melting temperatures and dissociation binding constants towards free and hAgo2-bound miR-34a), the DLR repression is significantly higher in the trapped excited state.

To provide a structural model of the different conformers, we computed the RNA structures in context of RISC using the X-ray crystal structure of hAgo2[188] and a slow-growth simulation approach. We found that the ground state matches the initial binding phase, where the guide 3'-end is bound to the PAZ domain. Conversely, the excited state adopts a remarkably different configuration, with the seed and 3'-supplementary helices coaxially stacked, with this latter accommodated along the N-PIWI domain, in a conformation similar to the one observed for *T. thermophilus* pAgo[197,201] (Fig. 9c, right).

In addition, we predicted the occurrence of similar binding sites with potential GC–to–GU switch in position g8, for miR-34a in human 3' UTRs, according to sequence and secondary structure determinants. We found a total of 593 predicted GC–to–GU switches and we selected 5 (HEBP1, ADAM22, ATG9A, ANKS1A and CCND1 mRNAs) for experimental validation using the DLR assay. In all cases tested, the trapped excited state showed a significantly higher repression then the wild-type.

Based on these evidence and previous reports (see 1.3.3), we proposed a mechanism by which these centrally bulged sites with 3'-supplementary pairing might adopt an "active" complex exploiting the RNA motions. After initial seed match and nucleation of 3'-end pairing at position g13–g16, the single base pair GC–to–GU switch enables full displacement of α7. This switch is coupled to the helices coaxial stacking, that induce the disengagement of the guide 3'-end from the PAZ pocket. In the "active" complex the guide might be destabilized[223,224] and enhance repression by letting RISC achieve multiple turnovers.

3.2 Paper II

Typical research laboratory protocols for the production of RNA samples in large-scale (i.e. milligrams or higher), rely on *in vitro* transcription using the T7 RNA polymerase, coupled to preparative PolyAcrylamide Gel Electrophoresis (PAGE) under denaturing conditions.

Here, we developed a protocol that combines the latest innovations to improve T7 *in vitro* transcription and we replaced denaturing PAGE with a two-step HPLC purification. Given the high demand of RNA samples and the diversity of users in our laboratory, we designed the protocol to be robust, fast and easy to use.

We detailed the use of previously reported methods to drastically reduce 3'-end inhomogeneity of the transcripts such as C2'-methoxy modification of the 5'-end of the DNA template[225] in conjunction with the use of DMSO as co-solvent[226,227] during transcription. We proposed the use of an EDTA solution to dissolve the magnesium-pyrophosphate precipitates forming during the transcription reaction, as an alternative to more laborious and expensive use of the recombinant inorganic pyrophosphatase enzyme.

We optimized buffers and gradients for a first Ion-Pairing Reverse-Phase (IP-RP) HPLC purification step, to separate the target RNA transcripts from higher and lower molecular weight products, cofactors, NTPs and proteins. This method makes use of tetrabutylammonium hydrogen sulfate as an ion-pairing agent to allow for the hydrophobic stationary phase to interact with the RNA[228]. The analytes are eluted from the column using a linear gradient of acetonitrile.

Subsequently, the target transcript is separated from similar molecular weight RNA species (i.e. +/- 1 nts deletion and addition products) using an Ion-Exchange (IE) HPLC purification step[229] under thermal denaturing conditions. During this phase, we propose that fractionation of the target peak, during the sodium perchlorate elution gradient, is essential to achieve near single-nucleotide purity.

In this work we present the preparation of four samples (22, 29, 46 and 82 nts in length) showing that the protocol can be performed in a few days to produce a diverse set of samples with high-yield and purity for structural biology purposes.

3.3 Paper III

The method presented in this work improves upon existing temperature Replica-Exchange Molecular Dynamics (REMD) simulations of RNA and it exploits the advantage of incorporating a minimal set of experimentally-derived long-range restraints.

These restraints are variable-length base-pairing contacts that can be easily inferred by NMR spectroscopy or chemical probing experiments (i.e. Selective 2′ Hydroxyl Acylation analyzed by Primer Extension (SHAPE)[230]).

Four different RNAs, for which experimental 3D structures were available, were used to benchmark the method. We proposed the use of the human Hepatitis B Virus (HBV) encapsidation signal apical stem-loop and measured its NMR imino walk, comparing it to previously published results[231].

The advantage of this approach resides in the ability to avoid kinetic traps during folding, associated with the formation of alternative secondary structures. The relatively minimal experimental data required and the accuracy inherent of the force-field and algorithm used, makes this method an attractive alternative not only to infer long-range tertiary interactions, but also to probe the conformational preferences of flexible structural elements, otherwise not accessible by high-resolution experimental methods in 3D.

Using traditionally assigned base pairs as well as restraints derived from $R_{1\rho}$ relaxation dispersion, we could compute the 3D RNA structures of the ground, excited and trapped excited state conformers presented in Paper I.

4 CONCLUSIONS AND PERSPECTIVES

We used solution NMR spectroscopy, molecular simulations, *in vitro* assays and previously published high-resolution structural information, to derive a model by which the miRNA–mRNA flexibility drives a functionally relevant conformational change in hAgo2.

While, miR-34a–mSirt1 is an important model, relevant for human health, one must keep in mind that each miRNA works in the context of a large network of intertwined other miRNAs and targets. Based on bioinformatic search and functional assays, we suggested that this mechanism is common to other miR-34a targets and, given the nature of miRNAs, it is to be expected that other miRNA–mRNA pairs, with similar sequence and structural features, will undergo the same process. To validate these assumptions, it will be necessary to develop a method that raises the throughput by which we infer structure and dynamics of miRNA–mRNA pairs.

The measurement of RNA dynamics in hAgo2 by NMR is the logical following step to the results of this thesis. For the researchers taking on this challenge, few technical obstacles will need to be overcome, chiefly the production of a ternary complex in quantity and homogeneity amenable for solution NMR studies. This will open up interesting avenues to answer outstanding questions in the field of miRNA biology as well as siRNA therapeutics.

For example, titration experiments with different target RNA could tell whether the disengagement from the 3’-end of the guide from the PAZ binding pocket follows the current stepwise consensus model derived from crystallographic structures and biochemical studies. Probing the RNA dynamics within hAgo2 could further confirm our models, revealing the degree of mobility of the 3’-supplementary helix, helping understand how the search for additional 3’-end pairing occurs after seed matching and how large bulges are accommodated within RISC. In addition, one could investigate the mechanism behind target cleavage at atomic resolution, with native and chemically modified guides, thereby contributing to the design of better siRNA therapeutics.

5 ACKNOWLEDGEMENTS

This journey started sometime during summer 2014, when I had an interview with a newly established PI in Sweden. She was very excited about the project, the new NMR spectrometer was being installed and the wet lab was growing (as a sign of it, I still remember the wooden pallets lying around in the background of Katja's office during the video call). I barely knew anything about RNAs and I had never been to Stockholm, but both were intriguing for some reason or another.

I'm very happy to acknowledge all the people that I met and worked with during this six-years journey, as well as those who supported me from my life-before-the-phd.

First, I want to thank my supervisor **Katja Petzold** for letting me join her lab. For giving me the opportunity to work on exciting science and for her continuous trust and support in my scientific and personal decisions. For hiring the most talented and fun colleagues ever and for providing an unmatched work environment. For the conferences, meetings and travels around the world, for giving the opportunity to visit places I otherwise would never have. None of these things were taken for granted and I consider it a privilege to have pursued a PhD in your lab under these conditions.

I want to thank my co-supervisor **Gunter Schneider** for welcoming me in the division and for the stimulating discussions. All the members of the Molecular Structural Biology division pre-Biomedicum era: **Ahmad**, **Victoria**, **Robert**, **Bernie**, **Brinda**, **Richard**, **Peter**, **Francesca**, **Dominik**, **Eva Maria** and **Katharina** for the countless discussions, lunches, pubs and dinners we had.

The beautiful people of the Petzold Lab: **Hampus**, for your philosophy, quotes and humor, for your inspiring and genuine scientific curiosity and rigor, for sharing the highs and lows of the student life and for taking all our 'criticism' on swedish culture. **Luca** for being a close friend in and out of the office, for your generosity, for sharing the experience of the italian immigrant and for all the laughs we had (even when there wasn't much to laugh about). **Emilie**, for answering all the dumb questions I had, for the mentoring, for being the social center of our group and for inventing all the best catchphrases and in-jokes. **Judith**, for answering all the other dumb questions I had, for recognizing me on the bus, for always having a moment to talk, for the laughs and for teaching us how to make spätzle. **Ileana**, for your work ethic and countless hours shared in the lab, for your incredible stories and for our fight about the best way to cut zucchini. **Sarah**, for teaching us biology, the pragmatism, the american accent and for never refusing a coffee break. **Maja**, for representing the DNA crowd and for baking the best bread and pastries. **Anurupa**, for all the help and for always sharing some delicious lunch bites. **Hannes**, for the memes and for the inspiring way you supervise students. **Lara**, for coming from so far away and for always taking the chance to try something new out in the lab. **Carolina**, for sharing the first months in the office and helping me out at the spectrometer. All the great students: **Luis**, **Karen**, **Noah**, **Linda**, **Andrea**, **Mateu**, **Raluca**, **Cenk**, **Oana** and **Mona** and the 'new' members: **Magdalena**, **Elnaz** and **Ting**.

Alan Chen and **Parisa** for being such nice collaborators, for the many discussions, drafts exchanges and the incredible work you did. **Emma Andersson** and her group for our meetings and for hosting me occasionally in their lab. Our neighbors in the **Högberg**, **Teixeira** and **Stevens** groups and all the members of the **B9 quarter**. In particular **Giulio** for the help at the bench and the PhD paperwork. **Patrick Lundström** and his group in Linköping for our joint meetings. **Helena**, **Ida** and all **PSF** staff. **Rhiju Das** for introducing me and coordinating my visit to UCSC, to **Michael Stone** and his group for hosting me and **Christina** for teaching me SHAPE and showing me around in the desert. **Sonia Longhi**, **Roberta Pierattelli** and **Isabella Felli** for initiating and inspiring me to pursue a career in this profession.

All the wonderful people I met at KI, particularly **Sara**, **Konstantina** and **Roberta** for always being up for a laugh and welcoming in their social circles. **Giulia**, **Gwladys**, **Mose**, **Simone** and **Charles**, for making the flat we were living in feel like home and after that for being good friends and such nice people to be around.

Finally I want to thank my friends and family for the unconditional love and support during these years, the countless calls and chats, the visits and weekends around Europe and the holidays reunions in Florence. Without you I would be nothing. **Elisa**, **Michele**, **Angela**, **Sofia**, **Simone**, and **Giacomo**. **Sandro**, **Umberto**, **Franca**, **Carmela** and **Monica** and my parents **Franco** and **Rosy**.

6 REFERENCES

1. Brenner, S., Jacob, F. & Meselson, M. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **190**, 576–581 (1961).

2. Gros, F. *et al.* Unstable Ribonucleic Acid Revealed by Pulse Labelling of Escherichia Coli. *Nature* **190**, 581–585 (1961).

3. Cobb, M. Who discovered messenger RNA? *Curr. Biol.* **25**, R526–R532 (2015).

4. Palade, G. E. A Small Particulate Component of the Cytoplasm. *J. Cell Biol.* **1**, 59–68 (1955).

5. Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I. & Zamecnik, P. C. A Soluble Ribonucleic Acid Intermediate in Protein Synthesis. *J. Biol. Chem.* **231**, 241–257 (1958).

6. Cech, T. R. & Steitz, J. A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **157**, 77–94 (2014).

7. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437 (2014).

8. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).

9. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

10. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

11. Müller, S., Appel, B., Balke, D., Hieronymus, R. & Nübel, C. Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? *F1000Research* **5**, (2016).

12. Zhang, J., Lau, M. W. & Ferré-D'Amaré, A. R. Ribozymes and riboswitches: modulation of RNA function by small molecules. *Biochemistry* **49**, 9123–9131 (2010).

13. Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* **18**, 321–329 (2008).

14. S. Marek, M., Johnson-Buck, A. & G. Walter, N. The shape-shifting quasispecies of RNA: one sequence, many functional folds. *Phys. Chem. Chem. Phys.* **13**, 11524–11537 (2011).

15. Bevilacqua, P. C., Ritchey, L. E., Su, Z. & Assmann, S. M. Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* **50**, 235–266 (2016).

16. Saenger, W. *Principles of nucleic acid structure*. (Springer Science & Business Media, 2013).

17. Neidle, S. *Principles of Nucleic Acid Structure*. (Elsevier, 2008). doi:10.1016/B978-0-12-369507-9.X5001-8.

18. Arnott, S., Fuller, W., Hodgson, A. & Prutton, I. Molecular conformations and structure transitions of RNA complementary helices and their possible biological significance. *Nature* **220**, 561–564 (1968).

19. Hermann, T. & Westhof, E. Non-Watson-Crick base pairs in RNA-protein recognition. *Chem. Biol.* **6**, R335–R343 (1999).

20. Varani, G. & McClain, W. H. The G·U wobble base pair. *EMBO Rep.* **1**, 18–23 (2000).

21. Varani, G. Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379–404 (1995).

22. Sigler, P. B. An Analysis of the Structure of tRNA. *Annu. Rev. Biophys. Bioeng.* **4**, 477–527 (1975).

23. Staple, D. W. & Butcher, S. E. Pseudoknots: RNA Structures with Diverse Functions. *PLOS Biol.* **3**, e213 (2005).

24. Moore, P. B. The RNA folding problem. *Cold Spring Harb. Monogr. Ser.* **37**, 381–402 (1999).

25. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure11Edited by I. Tinoco. *J. Mol. Biol.* **288**, 911–940 (1999).

26. Sweeney, B. A., Roy, P. & Leontis, N. B. An introduction to recurrent nucleotide interactions in RNA. *Wiley Interdiscip. Rev. RNA* **6**, 17–45 (2015).

27. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

28. Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16**, 907–916 (1963).

29. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA N. Y. N* **7**, 499–512 (2001).

30. Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30**, 3497–3531 (2002).

31. Nasalean, L., Stombaugh, J., Zirbel, C. L. & Leontis, N. B. RNA 3D Structural Motifs: Definition, Identification, Annotation, and Database Searching. in *Non-Protein Coding RNAs* (eds. Walter, N. G., Woodson, S. A. & Batey, R. T.) 1–26 (Springer, 2009). doi:10.1007/978-3-540-70840-7_1.

32. Westhof, E. Isostericity and tautomerism of base pairs in nucleic acids. *FEBS Lett.* **588**, 2464–2469 (2014).

33. Stombaugh, J., Zirbel, C. L., Westhof, E. & Leontis, N. B. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* **37**, 2294–2312 (2009).

34. Crick, F. H. Codon--anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).

35. Holley, R. W. *et al.* Structure of a Ribonucleic Acid. *Science* **147**, 1462–1465 (1965).

36. Masquida, B. & Westhof, E. On the wobble GoU and related pairs. *RNA N. Y. N* **6**, 9–15 (2000).

37. Allain, F. H. & Varani, G. Divalent metal ion binding to a conserved wobble pair defining the upstream site of cleavage of group I self-splicing introns. *Nucleic Acids Res.* **23**, 341–350 (1995).

38. Mizuno, H. & Sundaralingam, M. Stacking of Crick Wobble pair and Watson-Crick pair: stability rules of G-U pairs at ends of helical stems in tRNAs and the relation to codon-anticodon Wobble interaction. *Nucleic Acids Res.* **5**, 4451–4461 (1978).

39. Giese, M. R. *et al.* Stability of RNA Hairpins Closed by Wobble Base Pairs. *Biochemistry* **37**, 1094–1100 (1998).

40. Berger, K. D., Kennedy, S. D. & Turner, D. H. Nuclear Magnetic Resonance Reveals That GU Base Pairs Flanking Internal Loops Can Adopt Diverse Structures. *Biochemistry* **58**, 1094–1108 (2019).

41. Berger, K. D. *et al.* Surprising Sequence Effects on GU Closure of Symmetric 2 × 2 Nucleotide RNA Internal Loops. *Biochemistry* **57**, 2121–2131 (2018).

42. Kimsey, I. & Al-Hashimi, H. M. Increasing occurrences and functional roles for high energy purine-pyrimidine base-pairs in nucleic acids. *Curr. Opin. Struct. Biol.* **24**, 72–80 (2014).

43. Lemieux, S. & Major, F. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.* **30**, 4250–4263 (2002).

44. Cruz, J. A. *et al.* RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA N. Y. N* **18**, 610–625 (2012).

45. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51 (2008).

46. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-Hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* **491**, 724–728 (2012).

47. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1309–1314 (1961).

48. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230 (1973).

49. Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8721–8725 (1992).

50. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

51. Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing Protein Intrinsic Disorder. *Chem. Rev.* **114**, 6561–6588 (2014).

52. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).

53. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).

54. Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **338**, 1042–1046 (2012).

55. Cruz, J. A. & Westhof, E. The Dynamic Landscapes of RNA Architecture. *Cell* **136**, 604–609 (2009).

56. Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* **482**, 322 (2012).

57. Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.* **83**, 441–466 (2014).

58. Herschlag, D. RNA Chaperones and the RNA Folding Problem. *J. Biol. Chem.* **270**, 20871–20874 (1995).

59. Woodson, S. A. Compact intermediates in RNA folding. *Annu. Rev. Biophys.* **39**, 61–77 (2010).

60. Pan, T. & Sosnick, T. Rna Folding During Transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 161–175 (2006).

61. Lai, D., Proctor, J. R. & Meyer, I. M. On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461–1473 (2013).

62. Liu, S.-R., Hu, C.-G. & Zhang, J.-Z. Regulatory effects of cotranscriptional RNA structure formation and transitions. *Wiley Interdiscip. Rev. RNA* **7**, 562–574 (2016).

63. Russell, R. RNA misfolding and the action of chaperones. *Front. Biosci. J. Virtual Libr.* **13**, 1–20 (2008).

64. Woodson, S. A. Taming free energy landscapes with RNA chaperones. *RNA Biol.* **7**, 677–686 (2010).

65. Leamy, K. A., Assmann, S. M., Mathews, D. H. & Bevilacqua, P. C. Bridging the gap between in vitro and in vivo RNA folding. *Q. Rev. Biophys.* **49**, (2016).

66. Brion, P. & Westhof, E. Hierarchy and Dynamics of Rna Folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113–137 (1997).

67. Mitchell, D., Assmann, S. M. & Bevilacqua, P. C. Probing RNA structure in vivo. *Curr. Opin. Struct. Biol.* **59**, 151–158 (2019).

68. Denny, S. K. & Greenleaf, W. J. Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers. *Cold Spring Harb. Perspect. Biol.* **11**, a032300 (2019).

69. Breaker, R. R. Riboswitches and Translation Control. *Cold Spring Harb. Perspect. Biol.* **10**, a032797 (2018).

70. Snoussi, K. & Leroy, J.-L. Imino Proton Exchange and Base-Pair Kinetics in RNA Duplexes. *Biochemistry* **40**, 8898–8904 (2001).

71. Chen, C., Jiang, L., Michalczyk, R. & Russu, I. M. Structural Energetics and Base-Pair Opening Dynamics in Sarcin−Ricin Domain RNA. *Biochemistry* **45**, 13606–13613 (2006).

72. Rinnenthal, J., Klinkert, B., Narberhaus, F. & Schwalbe, H. Direct observation of the temperature-induced melting process of the Salmonella fourU RNA thermometer at base-pair resolution. *Nucleic Acids Res.* **38**, 3834–3847 (2010).

73. Jankowsky, E. RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* **36**, 19–29 (2011).

74. Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W. & Al-Hashimi, H. M. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **519**, 315–320 (2015).

75. Peterson, R. D. & Feigon, J. Structural Change in Rev Responsive Element RNA of HIV-1 on Binding Rev Peptide. *J. Mol. Biol.* **264**, 863–877 (1996).

76. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256–259 (2012).

77. Merriman, D. K. *et al.* Shortening the HIV-1 TAR RNA Bulge by a Single Nucleotide

Preserves Motional Modes over a Broad Range of Time Scales. *Biochemistry* **55**, 4445–4456 (2016).

78. Olsen, G. L. *et al.* Solid-State Deuterium NMR Studies Reveal μs-ns Motions in the HIV-1 Transactivation Response RNA Recognition Site. *J. Am. Chem. Soc.* **130**, 2896–2897 (2008).

79. Olsen, G. L., Bardaro, M. F., Echodu, D. C., Drobny, G. P. & Varani, G. Hydration dependent dynamics in RNA. *J. Biomol. NMR* **45**, 133–142 (2009).

80. Olsen, G. L., Bardaro, M. F., Echodu, D. C., Drobny, G. P. & Varani, G. Intermediate Rate Atomic Trajectories of RNA by Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **132**, 303–308 (2010).

81. Emani, P. S., Olsen, G. L., Varani, G. & Drobny, G. P. Theory of Nonrigid Rotational Motion Applied to NMR Relaxation in RNA. *J. Phys. Chem. A* **115**, 12055–12069 (2011).

82. Huang, W., Emani, P. S., Varani, G. & Drobny, G. P. Ultraslow Domain Motions in HIV-1 TAR RNA Revealed by Solid-State Deuterium NMR. *J. Phys. Chem. B* **121**, 110–117 (2017).

83. Lee, J., Dethoff, E. A. & Al-Hashimi, H. M. Invisible RNA state dynamically couples distant motifs. *Proc. Natl. Acad. Sci.* **111**, 9485–9490 (2014).

84. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H. & Jones, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451–462 (1998).

85. Karn, J. & Stoltzfus, C. M. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. *Cold Spring Harb. Perspect. Med.* **2**, (2012).

86. Salmon, L., Bascom, G., Andricioaei, I. & Al-Hashimi, H. M. A General Method for Constructing Atomic-Resolution RNA Ensembles using NMR Residual Dipolar Couplings: The Basis for Interhelical Motions Revealed. *J. Am. Chem. Soc.* **135**, 5457–5466 (2013).

87. Zhang, Q. & Al-Hashimi, H. M. Extending the NMR spatial resolution limit for RNA by motional couplings. *Nat. Methods* **5**, 243–245 (2008).

88. Zhang, Q., Stelzer, A. C., Fisher, C. K. & Al-Hashimi, H. M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263–1267 (2007).

89. Dethoff, E. A., Hansen, A. L., Zhang, Q. & Al-Hashimi, H. M. Variable helix elongation as a tool to modulate RNA alignment and motional couplings. *J. Magn. Reson.* **202**, 117–121 (2010).

90. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **20**, 474–489 (2019).

91. Keeler, J. *Understanding NMR spectroscopy*. (John Wiley & Sons, 2011).

92. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189 (1980).

93. Lu, K., Miyazaki, Y. & Summers, M. F. Isotope labeling strategies for NMR studies of RNA. *J. Biomol. Nmr* **46**, 113–125 (2010).

94. Baronti, L., Karlsson, H., Marušič, M. & Petzold, K. A guide to large-scale RNA sample preparation. *Anal. Bioanal. Chem.* **410**, 3239–3252 (2018).

95. Varani, G., Aboul-ela, F. & Allain, F. H.-T. NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spectrosc.* **29**, 51–127 (1996).

96. Wijmenga, S. S. & van Buuren, B. N. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **32**, 287–387 (1998).

97. Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *Chembiochem* **4**, 936–962 (2003).

98. Kumar, A., Ernst, R. R. & Wüthrich, K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6 (1980).

99. Wüthrich, K. Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* **265**, 22059–22062 (1990).

100. Kearns, D. R., Patel, D. J. & Shulman, R. G. High Resolution Nuclear Magnetic Resonance Studies of Hydrogen Bonded Protons of tRNA in Water. *Nature* **229**, 338–339 (1971).

101. Aeschbacher, T. *et al.* Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.* **41**, e172–e172 (2013).

102. Brown, J. D., Summers, M. F. & Johnson, B. A. Prediction of hydrogen and carbon chemical shifts from RNA using database mining and support vector regression. *J. Biomol. NMR* **63**, 39–52 (2015).

103. Dingley, A. J. & Grzesiek, S. Direct Observation of Hydrogen Bonds in Nucleic Acid Base Pairs by Internucleotide 2JNN Couplings. *J. Am. Chem. Soc.* **120**, 8293–8297 (1998).

104. Dallmann, A. *et al.* Efficient detection of hydrogen bonds in dynamic regions of RNA by sensitivity-optimized NMR pulse sequences. *Angew. Chem. Int. Ed.* **52**, 10487–10490 (2013).

105. Sklenář, V., Peterson, R. D., Rejante, M. R. & Feigon, J. Two-and three-dimensional HCN experiments for correlating base and sugar resonances in 15N, 13C-labeled RNA oligonucleotides. *J. Biomol. NMR* **3**, 721–727 (1993).

106. Peterson, R. D., Theimer, C. A., Wu, H. & Feigon, J. New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR* **28**, 59–67 (2004).

107. Schlagnitweit, J., Steiner, E., Karlsson, H. & Petzold, K. Efficient Detection of Structure and Dynamics in Unlabeled RNAs: The SELOPE Approach. *Chem. – Eur. J.* **24**, 6067–6070 (2018).

108. Richter, C. *et al.* 13C-direct detected NMR experiments for the sequential J-based resonance assignment of RNA oligonucleotides. *J. Biomol. Nmr* **47**, 259–269 (2010).

109. Marušič, M., Schlagnitweit, J. & Petzold, K. RNA dynamics by NMR. *ChemBioChem* (2019).

110. Vallurupalli, P., Sekhar, A., Yuwen, T. & Kay, L. E. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. *J. Biomol. NMR* **67**, 243–271 (2017).

111. McConnell, H. M. Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* **28**, 430–431 (1958).

112. Zhao, B., Hansen, A. L. & Zhang, Q. Characterizing Slow Chemical Exchange in Nucleic Acids by Carbon CEST and Low Spin-Lock Field R1ρ NMR Spectroscopy. *J. Am. Chem. Soc.* **136**, 20–23 (2014).

113. Longhini, A. P. *et al.* Chemo-enzymatic synthesis of site-specific isotopically labeled nucleotides for use in NMR resonance assignment, dynamics and structural characterizations. *Nucleic Acids Res.* **44**, e52 (2016).

114. Zhao, B., Guffy, S. L., Williams, B. & Zhang, Q. An excited state underlies gene regulation of a transcriptional riboswitch. *Nat. Chem. Biol.* **13**, 968–974 (2017).

115. Chen, B., LeBlanc, R. & Dayie, T. K. SAM-II Riboswitch Samples at least Two Conformations in Solution in the Absence of Ligand: Implications for Recognition. *Angew. Chem. Int. Ed Engl.* **55**, 2724–2727 (2016).

116. Hahn, E. L. Spin Echoes. *Phys. Rev.* **80**, 580–594 (1950).

117. Sauerwein, A. C. & Hansen, D. F. Relaxation Dispersion NMR Spectroscopy. in *Protein NMR: Modern Techniques and Biomedical Applications* (ed. Berliner, L.) 75–132 (Springer US, 2015). doi:10.1007/978-1-4899-7621-5_3.

118. Kloiber, K., Spitzer, R., Tollinger, M., Konrat, R. & Kreutz, C. Probing RNA dynamics via longitudinal exchange and CPMG relaxation dispersion NMR spectroscopy using a sensitive 13C-methyl label. *Nucleic Acids Res.* **39**, 4340–4351 (2011).

119. Wunderlich, C. H. *et al.* Synthesis of (6-13C)Pyrimidine Nucleotides as Spin-Labels for RNA Dynamics. *J. Am. Chem. Soc.* **134**, 7558–7569 (2012).

120. Wunderlich, C. H. *et al.* Stable isotope-labeled RNA phosphoramidites to facilitate dynamics by NMR. *Methods Enzymol.* **565**, 461–494 (2015).

121. Juen, M. A. *et al.* Excited States of Nucleic Acids Probed by Proton Relaxation Dispersion NMR Spectroscopy. *Angew. Chem. Int. Ed.* **55**, 12008–12012 (2016).

122. Strebitzer, E., Nußbaumer, F., Kremser, J., Tollinger, M. & Kreutz, C. Studying sparsely populated conformational states in RNA combining chemical synthesis and solution NMR spectroscopy. *Methods* **148**, 39–47 (2018).

123. Chen, B. *et al.* CCR5 RNA Pseudoknots: Residue and Site-Specific Labeling correlate Internal Motions with microRNA Binding. *Chem. – Eur. J.* **24**, 5462–5468 (2018).

124. LeBlanc, R. M., Longhini, A. P., Tugarinov, V. & Dayie, T. K. NMR probing of invisible excited states using selectively labeled RNAs. *J. Biomol. NMR* **71**, 165–172 (2018).

125. Johnson, J. E. & Hoogstraten, C. G. Extensive Backbone Dynamics in the GCAA RNA Tetraloop Analyzed Using 13C NMR Spin Relaxation and Specific Isotope Labeling. *J. Am. Chem. Soc.* **130**, 16757–16769 (2008).

126. Miloushev, V. Z. & Palmer, A. G. R(1rho) relaxation for two-site chemical exchange: general approximations and some exact solutions. *J. Magn. Reson. San Diego Calif 1997* **177**, 221–227 (2005).

127. Palmer, A. G. & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106**, 1700–1719 (2006).

128. Zhang, Q. & Al-Hashimi, H. M. Domain-elongation NMR spectroscopy yields new insights into RNA dynamics and adaptive recognition. *RNA N. Y. N* **15**, 1941–1948 (2009).

129. Gorski, S. A., Vogel, J. & Doudna, J. A. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.* **18**, 215–228 (2017).

130. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753 (2014).

131. Wagner, E. G. H. Cycling of RNAs on Hfq. *RNA Biol.* **10**, 619–626 (2013).

132. Santiago-Frangos, A. & Woodson, S. A. Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip. Rev. RNA* **9**, e1475 (2018).

133. Holmqvist, E. & Vogel, J. RNA-binding proteins in bacteria. *Nat. Rev. Microbiol.* **16**, 601–615 (2018).

134. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).

135. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR

Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).

136. Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20–51 (2018).

137. Peters, L. & Meister, G. Argonaute Proteins: Mediators of RNA Silencing. *Mol. Cell* **26**, 611–623 (2007).

138. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).

139. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855–862 (1993).

140. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).

141. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of Novel Genes Coding for Small Expressed RNAs. *Science* **294**, 853–858 (2001).

142. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans. *Science* **294**, 858–862 (2001).

143. Lee, R. C. & Ambros, V. An Extensive Class of Small RNAs in Caenorhabditis elegans. *Science* **294**, 862–864 (2001).

144. Ruby, J. G. *et al.* Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans. *Cell* **127**, 1193–1207 (2006).

145. Landgraf, P. *et al.* A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* **129**, 1401–1414 (2007).

146. Ruby, J. G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res.* **17**, 1850–1864 (2007).

147. Altuvia, Y. *et al.* Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.* **33**, 2697–2706 (2005).

148. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).

149. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).

150. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).

151. Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215–233 (2009).

152. Gebert, L. & MacRae, I. J. Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* (2018).

153. Treiber, T., Treiber, N. & Meister, G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat. Rev. Mol. Cell Biol.* **20**, 5–20 (2019).

154. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).

155. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).

156. Wilson, R. C. *et al.* Dicer-TRBP Complex Formation Ensures Accurate Mammalian MicroRNA Biogenesis. *Mol. Cell* **57**, 397–407 (2015).

157. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**, 23–34 (2001).

158. Hutvágner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).

159. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev.* **15**, 2654–2659 (2001).

160. Nakanishi, K. Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins? *WIREs RNA* **7**, 637–660 (2016).

161. Sheu-Gruttadauria, J. & MacRae, I. J. Structural Foundations of RNA Silencing by Argonaute. *J. Mol. Biol.* **429**, 2619–2639 (2017).

162. Pasquinelli, A. E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.* **13**, 271–282 (2012).

163. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433 (2015).

164. Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* **11**, 228–234 (2009).

165. Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. *Science* **336**, 1037–1040 (2012).

166. Elkayam, E. *et al.* The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell* **150**, 100–110 (2012).

167. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian

mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).

168. Grimson, A. *et al.* MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Mol. Cell* **27**, 91–105 (2007).

169. Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute Divides Its RNA Guide into Domains with Distinct Functions and RNA-Binding Properties.(Report). *Cell* **151**, 1055 (2012).

170. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *elife* **4**, e05005 (2015).

171. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* **162**, 84–95 (2015).

172. Moore, M. J. *et al.* miRNA–target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.* **6**, 1–17 (2015).

173. Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W. & Pasquinelli, A. E. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol. Cell* **64**, 320–333 (2016).

174. Sheu-Gruttadauria, J., Xiao, Y., Gebert, L. F. & MacRae, I. J. Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. *EMBO J.* e101153 (2019).

175. Sheu-Gruttadauria, J. *et al.* Structural Basis for Target-Directed MicroRNA Degradation. *Mol. Cell* **75**, 1243-1255.e7 (2019).

176. Cevec, M., Thibaudeau, C. & Plavec, J. Solution structure of a let-7 miRNA: lin-41 mRNA complex from C. elegans. *Nucleic Acids Res.* **36**, 2330–2337 (2008).

177. Cevec, M., Thibaudeau, C. & Plavec, J. NMR structure of the let-7 miRNA interacting with the site LCS1 of lin-41 mRNA from Caenorhabditis elegans. *Nucleic Acids Res.* **38**, 7814–7821 (2010).

178. Meister, G. *et al.* Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).

179. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science* **305**, 1434–1437 (2004).

180. Yuan, Y.-R. *et al.* Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage. *Mol. Cell* **19**, 405–419 (2005).

181. Rashid, U. J. *et al.* Structure of Aquifex aeolicus Argonaute Highlights Conformational Flexibility of the PAZ Domain as a Potential Regulator of RNA-induced Silencing Complex Function. *J. Biol. Chem.* **282**, 13824–13832 (2007).

182. Ma, J.-B. *et al.* Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**, 666–670 (2005).

183. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663–666 (2005).

184. Boland, A., Tritschler, F., Heimstädt, S., Izaurralde, E. & Weichenrieder, O. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep.* **11**, 522–527 (2010).

185. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).

186. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).

187. Nakanishi, K., Weinberg, D. E., Bartel, D. P. & Patel, D. J. Structure of yeast Argonaute with guide RNA. *Nature* **486**, 368 (2012).

188. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science* **346**, 608–613 (2014).

189. Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J. & Joshua-Tor, L. The Making of a Slicer: Activation of Human Argonaute-1. *Cell Rep.* **3**, 1901–1909 (2013).

190. Nakanishi, K. *et al.* Eukaryote-Specific Insertion Elements Control Human ARGONAUTE Slicer Activity. *Cell Rep.* **3**, 1893–1900 (2013).

191. Yan, K. S. *et al.* Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468–474 (2003).

192. Song, J.-J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat. Struct. Biol.* **10**, 1026–1032 (2003).

193. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426**, 465–469 (2003).

194. Ma, J.-B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).

195. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* **11**, 576 (2004).

196. Hur, J. K., Zinchenko, M. K., Djuranovic, S. & Green, R. Regulation of Argonaute Slicer Activity by Guide RNA 3′ End Interactions with the N-terminal Lobe. *J. Biol. Chem.* **288**, 7829–7840 (2013).

197. Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754 (2009).

198. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of MicroRNA–Target Recognition. *PLOS Biol.* **3**, e85 (2005).

199. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).

200. Klum, S. M., Chandradoss, S. D., Schirle, N. T., Joo, C. & MacRae, I. J. Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition. *EMBO J* **37**, 75–88 (2018).

201. Sheng, G. *et al.* Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci.* **111**, 652 (2014).

202. Lambert, N. *et al.* RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. *Mol. Cell* **54**, 887–900 (2014).

203. McGeary, S. E. *et al.* The biochemical basis of microRNA targeting efficacy. *Science* (2019) doi:10.1126/science.aav1741.

204. Becker, W. R. *et al.* High-Throughput Analysis Reveals Rules for Target RNA Binding and Cleavage by AGO2. *Mol. Cell* **75**, 741-755.e11 (2019).

205. Deerberg, A., Willkomm, S. & Restle, T. Minimal mechanistic model of siRNA-dependent target RNA slicing by recombinant human Argonaute 2 protein. *Proc. Natl. Acad. Sci.* **110**, 17850–17855 (2013).

206. Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J. & Joo, C. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell* **162**, 96–107 (2015).

207. Schirle, N. T., Sheu-Gruttadauria, J., Chandradoss, S. D., Joo, C. & MacRae, I. J. Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. *eLife* **4**, e07646 (2015).

208. Jo, M. H. *et al.* Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. *Mol. Cell* **59**, 117–124 (2015).

209. Ameres, S. L. *et al.* Target RNA–Directed Trimming and Tailing of Small Silencing RNAs. *Science* **328**, 1534–1539 (2010).

210. Cazalla, D., Yario, T. & Steitz, J. A. Down-Regulation of a Host MicroRNA by a Herpesvirus saimiri Noncoding RNA. *Science* **328**, 1563–1566 (2010).

211. Vaziri, H. *et al.* hSIR2SIRT1 Functions as an NAD-Dependent p53 Deacetylase. *Cell* **107**, 149–159 (2001).

212. Yamakuchi, M., Ferlito, M. & Lowenstein, C. J. miR-34a repression of SIRT1 regulates apoptosis. *Proc. Natl. Acad. Sci.* **105**, 13421–13426 (2008).

213. Yamakuchi, M. & Lowenstein, C. J. MiR-34, SIRT1, and p53: The feedback loop. *Cell Cycle* **8**, 712–715 (2009).

214. Lodygin, D. *et al.* Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle* **7**, 2591–2600 (2008).

215. He, L. *et al.* A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).

216. Chang, T.-C. *et al.* Transactivation of miR-34a by p53 Broadly Influences Gene Expression and Promotes Apoptosis. *Mol. Cell* **26**, 745–752 (2007).

217. Raver-Shapira, N. *et al.* Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol. Cell* **26**, 731–743 (2007).

218. Tarasov, V. *et al.* Differential Regulation of microRNAs by p53 Revealed by Massively Parallel Sequencing: miR-34a is a p53 Target That Induces Apoptosis and G1-arrest. *Cell Cycle* **6**, 1586–1593 (2007).

219. Bommer, G. T. *et al.* p53-Mediated Activation of miRNA34 Candidate Tumor-Suppressor Genes. *Curr. Biol.* **17**, 1298–1307 (2007).

220. Hermeking, H. The miR-34 family in cancer and apoptosis. *Cell Death Differ.* **17**, 193–199 (2010).

221. Misso, G. *et al.* Mir-34: A New Weapon Against Cancer? *Mol. Ther. - Nucleic Acids* **3**, (2014).

222. Saito, Y., Nakaoka, T. & Saito, H. microRNA-34a as a Therapeutic Agent against Human Cancer. *J. Clin. Med.* **4**, 1951–1959 (2015).

223. De, N. *et al.* Highly Complementary Target RNAs Promote Release of Guide RNAs from Human Argonaute2. *Mol. Cell* **50**, 344–355 (2013).

224. Park, J. H., Shin, S.-Y. & Shin, C. Non-canonical targets destabilize microRNAs in human Argonautes. *Nucleic Acids Res.* **45**, 1569–1583 (2017).

225. Kao, C., Zheng, M. & Rüdisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3′ terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268–1272 (1999).

226. Chen, Z. & Zhang, Y. Dimethyl sulfoxide targets phage RNA polymerases to promote transcription. *Biochem. Biophys. Res. Commun.* **333**, 664–670 (2005).

227. Helmling, C. *et al.* Rapid NMR screening of RNA secondary structure and binding. *J. Biomol. NMR* **63**, 67–76 (2015).

228. Murray, J. B., Collier, A. K. & Arnold, J. R. A general purification procedure for chemically synthesized oligoribonucleotides. *Anal. Biochem.* **218**, 177–184 (1994).

229. Oliver, R. W. A. *HPLC of macromolecules: a practical approach*. vol. 872 (Oxford

University Press on Demand, 1998).

230. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616 (2006).

231. Flodell, S. *et al.* The apical stem–loop of the hepatitis B virus encapsidation signal folds into a stable tri–loop with two underlying pyrimidine bulges. *Nucleic Acids Res.* **30**, 4803–4811 (2002).