Probing biomolecular dynamics and intercations by combining chemical synthesis and NMR spectroscopy

Emeline Mestdach

Numéro National de Thèse : 2022ENSL0015

THÈSE

en vue de l'obtention du grade de Docteur, délivré par
l'ECOLE NORMALE SUPERIEURE DE LYON

Ecole Doctorale N° 206

Ecole Doctorale de Chimie (Chimie, Procédés, Environnement)

Discipline : Chimie

Soutenue publiquement le 26/09/2022, par :

Emeline MESTDACH

Probing biomolecular dynamics and interactions by combining chemical synthesis and NMR spectroscopy

Etudes de dynamiques et interactions biomoléculaires, en combinant synthèse chimique et spectroscopie RMN

Devant le jury composé de :

Cameron Mackereth	Directeur de Recherche Institut Européen de Chimie et de Biologie, France	Rapporteur
Virginie Gervais	Chargée de Recherche CEA (HDR) Institute for Integrative Biology of the Cell, France	Rapporteure
Catherine Bougault	Maître de Conférences Institut de Biologie Structurale, France	Examinatrice
Eric Ennifar	Directeur de Recherche Institut de Biologie Moléculaire et Cellulaire, France	Examinateur
Loïc Salmon	Chargé de Recherche CNRS (HDR) CRMN de Lyon, France	Directeur de thèse
Guido Pintacuda	Directeur de Recherche CRMN de Lyon, France	Co-encadrant de thèse

Contents

Acknowledgements

First of all, I would like to thank the reviewers, Virginie Gervais and Cameron Mackereth, for their evaluation of this manuscript. I would also like to thank Eric Ennifar and Catherine Bougault, for agreeing to be jury members. To all the committee, thank you for your shared inputs and thoughts on the manuscript, the defence, and of course the projects. Un merci particulier à Catherine, sans qui je n'aurai peut-être pas eu l'opportunité de postuler pour cette thèse.

I would like to thank my PhD supervisor, Loïc Salmon. Cette thèse aura été lancée très rapidement, à quelques jours de la clôture des admissions de l'Ecole Doctorale. Merci de m'avoir donné la chance de travailler au CRMN, encore plus en sachant que c'était le début de ton groupe. Voir son développement a été une superbe expérience; les rencontres que j'ai pu faire aussi bien sur un plan professionnel que personnel.

I also would like to thank Guido Pintacuda. Merci pour les conseils et pour la réflexion scientifique indéniables que tu as pu m'apporter, ainsi que simplement pour les bons moments.

Of course, I would like to thank the other members that made the group as lively as it was. Merci Sirine, pour les rires, et également pour le partage de points de vue. Nos discussions auront grandement contribué à garder un semblant de sanité, spécialement en fin de thèse ;) Merci Landry pour les (longues) discussions; je savais à quelle heure j'arrivais pour te saluer, mais je ne savais jamais à quelle heure je repartirais à mon bureau! Grâce à vous 2, le groupe est clairement devenu un groupe et ça a été un plaisir d'avoir d'autres personnes avec qui travailler, ou juste discuter. Merci Selin pour les rires, pour endurer toutes nos taquineries, pour m'avoir appris à me sortir de toutes les situations grâce à 10 chiffres en turc ;) Et merci Laura, pour avoir le même humour que moi, pour avoir été dépitée, réjouie, avoir de l'espoir et être blasée autant que moi sur nos projets. Merci à tou(te)s pour les moments hors du labo également (Landry on t'attend toujours au bar)!

Along with the group, I would like to thank all the people at the CRMN who made the work environment what it was. Merci Cécile pour ton soutien et tes histoires, ta bonne humeur au CRMN puis en dehors, et merci pour les découvertes culinaires! Merci David pour tout le temps dédié à m'aider sur les spectro mais également pour le reste, notamment de nous avoir aidés à agrandir la culture cinématographique de Selin ;) Merci Tanguy et Olivier, pour vos blagues, votre aide et vos gentilles paroles.

I would like to thank the collaborators I had the opportunity to work with: merci aux Carole pour m'avoir appris (presque) tout ce que je sais de la synthèse chimique et pour avoir comblé le silence des labos, merci à Séraphine et Cécile pour les discussions pertinentes et pour leur entrain presque palpable à chaque réunion, thanks to Pierre Barraud, Guillaume Bouvignies, Guillaume Hoffmann, Madalen Le Gorrec, Stephen Cusack, Malene Ringkjøbing Jensen, Andrés Palencia and Vladimir Ladizhansky.

Un merci spécial à ceux qui m'ont supportée encore un peu plus que les autres: merci Marta pour m'avoir montré la voie et pour ta compagnie apaisante, même dans les pires moments de ta thèse. Merci Georges pour avoir été un ami, pour les encouragements et le soutien, dans les bons comme les mauvais moments, et merci pour Smash Up ;) Merci Thomas pour débattre sur des sujets quels qu'ils soient, les plus vieilles expressions que j'aie jamais entendues, et nous laisser squatter chez toi! Et merci Marie pour avoir enduré un stage à mes côtés sans te plaindre (un exploit!).

Thanks to all of those that joined/left the CRMN during these past 4 years (sorry if anyone is forgotten): Adrian, Arthur, Ribal, Tobias, Théo (ça y est on n'est plus pareil!), Chloé, Ladislav, Alicia, Jan, Charlotte, Quentin, Morgan, Nathalie, Maurane, Camille, Lorenzo, Zhuoran, Jeff, Daniela, Sam, Stuart, Kevin, Alicia, Monu, Jasmine, Giacomo, Francesca, Hubert, Christelle, Maryline, Dorothea, Anne, Andrew, Sami, Isabelle, Sylvie, Gilles, (etc?). And thanks to Judith, for not letting me despairing alone in front of the spectrometer!

Merci à mes amis d'un peu partout en dehors du CRMN. Merci Marianne pour être ma meilleure amie depuis des années, pour m'avoir fait apprécier nos moments peu importe mon état d'esprit, et pour les nombreux autres moments partagés de nos vies à venir. Merci Chloé pour ton amitié depuis le lycée, pour avoir fait le déplacement pour ma soutenance, et ton soutien de tous les jours. Merci Frédérique pour me rappeler que peu importe ce qui se passe, il suffit de chanter un Disney pour que le moment reste positivement gravé dans mon esprit ;) Thank you Krisztina, for being my friend anytime even more than a 1000km apart, for coming to Lyon, for surviving a full holiday with me ;) Merci à mes amis du nord: Olivier, François, Apolline, Emilie, Gautier qui tolèrent mon absence et sont présents à chacun de mes retours. Merci à mes amis de l'ENSCMu, tout particulièrement à Maxime, Léa et Zig.

Evidemment, merci à ma famille: Maman, Papa, mon frère, ainsi que mes grands-parents, Micheline, Suzanne, André et Jean-Marie. Vous voir me redonne de la bonne humeur et de la force à chaque fois. Merci pour vos encouragements dans tout ce que j'entreprends, que je puisse encore vous voir ou non, vous êtes toujours avec moi.

Merci à Michèle, Jean-Marc, Noël et Julien pour m'avoir accueillie dans leur famille si gentiment.

Et finalement, merci à Misha. Nos moments passés comme ceux à venir me remplissent de bonheur, et j'ai hâte de découvrir chaque jour du reste de notre vie ensemble.

List of abbreviations/Acronyms

Ac: Acetyl
ACN: Acetonitrile
Bz: Benzoyl
CEST: Chemical Exchange Saturation Transfer
CMC: Critical Micelle Concentration
CPG: Controlled Pore Glass
CPMG: Carll-Purcell-Meiboom-Gill
CS: Chemical Shift
CSA: Chemical Shift Anisotropy
DCA: DiChloroAcetic acid
DDM: n-dodecyl -D-maltoside
DMT: 4,4-DiMethoxyTrityl
DMSO: DiMéthylSulfOxyde
DNA: DeoxyriboNucleic Acid
E. coli: Escherichia coli
HSQC: HeteroNuclear Single Quantum Coherence
HPLC: High-Performance Liquid Chromatography
iBu: isoButyryl
Llp: lytic conversion periplasmic lipoprotein
miRNA: micro RNA
mRNA: messenger RNA
NMR: Nuclear Magnetic Resonance
NOESY: Nuclear Overhauser Effect SpectroscopY
NOE: Nuclear Overhauser Effect
PS: PolyStyrene
RBP: Receptor Binding Protein
RF:Radio-Frequency
RNA: RiboNucleic Acid
rRNA: ribosomal RNA
siRNA: small interfering RNA
sol-Llp: soluble form of Llp
SPR: Surface Plasmon Resonance
TBDMS: Tert-ButylDiMethylSilyl
TEAA: Triethylammonium acetate
TOCSY: TOtal Correlation SpectroscopY
tRNA: transfer RNA
TREAT: triethylamine trihydrofluoride

TROSY: Transverse Relaxation-Optimised SpectroscopY
UV: UltraViolet
WATERGATE: WATER suppression by GrAdient Tailored Excitation

Résumé

Ma thèse portait sur le développement de méthodologies de RMN biomoléculaire pour caractériser la dynamique et les interactions de molécules biologiques, en explorant de nouveaux horizons de la biologie, de la chimie et de la spectroscopie RMN. Mes travaux ont d'abord porté sur les micro-ARN (miARN), des régulateurs génétiques clés, capables de bloquer la traduction des ARN messagers (ARNm). Le miARN *let-7* et ses cibles ont été synthétisés chimiquement pour incorporer des sondes ^{19}F, permettant de décrire l'interaction entre *let-7* et ses ARNm cibles par RMN. Le deuxième projet était centré sur un ARN de transfert (ARNt), un composant essentiel de la traduction. La modulation de la dynamique de l'ARNt a été étudiée à l'aide de la relaxation de spin ^{15}N, dans le contexte de la maturation de l'ARNt. Les vitesses de relaxation ont ainsi été mesurées à plusieurs champs et analysées pour entrevoir la dynamique rapide de ce système. La troisième étude visait à caractériser une protéine impliquée dans l'infection de bactérie par le bactériophage T5, et à sonder son interaction avec son récepteur membranaire. L'attribution des résonances de la protéine LLP impliquée dans l'infection par le phage a été réalisée, ainsi que la caractérisation de la structure 3D. La relaxation de spin a été utilisée pour entrapercevoir la dynamique de LLP et une titration avec le récepteur bactérien a permis de révéler l'interface de liaison. Un projet complémentaire (non détaillé ici) concernait l'étude dynamique d'une aquaporine, une protéine membranaire jouant le rôle de passeur d'eau, par RMN du solide à plusieurs champs. Tous ces exemples illustrent comment la spectroscopie RMN peut sonder des processus dynamiques complexes dans divers contextes biologiques.

Abstract

My thesis focused on the development of novel biomolecular NMR methodologies to characterise dynamics and interactions of biological macromolecules, by exploring novel horizons of biology, chemistry and NMR spectroscopy. My work first focused on micro-RNAs (miRNAs), key genetic regulators, capable to block messenger RNA (mRNA) translation. The *let-7* miRNA and its targets were chemically synthesized to incorporate ^{19}F probes, allowing to describe the interaction between *let-7* and its target mRNAs by NMR. The second project was centered on a transfer RNA (tRNA), an essential component of the translation. Modulation of tRNA dynamic was probed using NMR ^{15}N spin relaxation, in the context of tRNA maturation. Here, multiple-field relaxation rates, were measured and analysed to capture the fast dynamics of this system. The third study aimed at characterising a protein involved in the T5 bacteriophage infection of bacteria, and probe its interaction with its bacteria receptor protein. Resonance assignment of LLP the T5 protein involved in phage infection has been performed, as well as the 3D structure characterisation. Spin relaxation was used to probe LLP dynamics and titration experiments with the bacteria receptor allowed to reveal the binding interface. An additional project (not detailed) involved the dynamic study of an aquaporin, a membrane protein acting as water channel, by solid-state NMR at multiple fields. All these examples exemplify how NMR spectroscopy can probe complex dynamic processes in various contexts.

Introduction

Structural biology employs a variety of techniques to investigate biomolecules: how they assemble, function and interact. Proteins and oligonucleotides constitute essential parts of the biological processes, and are involved in various fundamental aspects of life, as well as implicated in diseases. The information gathered by structural studies has helped advancing the understanding of the biomolecules structures, that are connected to the completion of their function. However, the snapshots of different conformational states adopted by the biomolecules are not the only factor to be taken into consideration to describe their mechanism: their intrinsic dynamic participate actively in their activity.

During this PhD, several biomolecules were investigated, structurally or dynamically, to better comprehend their actions. As many techniques are available to study the biomolecular structures (e.g. x-ray crystallography and cryo-electron microscopy), ascertain their motional behaviour is a necessity that can be accomplished using NMR spectroscopy.

Nucleic acids

Context

Nucleic acids have been a growing interest of the scientific community over the past decades (see Figure 1). Indeed, progressions in synthesis, analysis and delivery of nucleic acids, coupled with their variety of functions, have rendered them ideal candidates for therapeutics.

Starting with the DeoxyriboNucleic Acid (DNA) structure determination in 1953 by Watson and Crick[1], further progress on identifying RiboNucleic Acids (RNAs), both in terms of structures and functions, have followed. The latest approval of medicines such as AntiSense Oligonucleotides (ASOs) and Short Interfering RNAs (siRNAs) represents a step towards novel medical solutions centered on gene silencing. These past years, the significance of the nucleic acids researches has also been highlighted by the worldwide recognised messenger RNA (mRNA) vaccines, against the Sars-CoV-2 epidemic[2][3].

Structure

Nucleic acids are constituted of a linear succession of nucleotides, forming DNAs and RNAs. These nucleotides encompass three major components, namely a phosphate group, a deoxyribose/ribose sugar and a base[4].

Several recurrent bases have been identified in nucleic acids (see Figure 2), categorised into two types: pyrimidines englobing cytosine (C), uracil (U, in RNAs) and thymine (T,

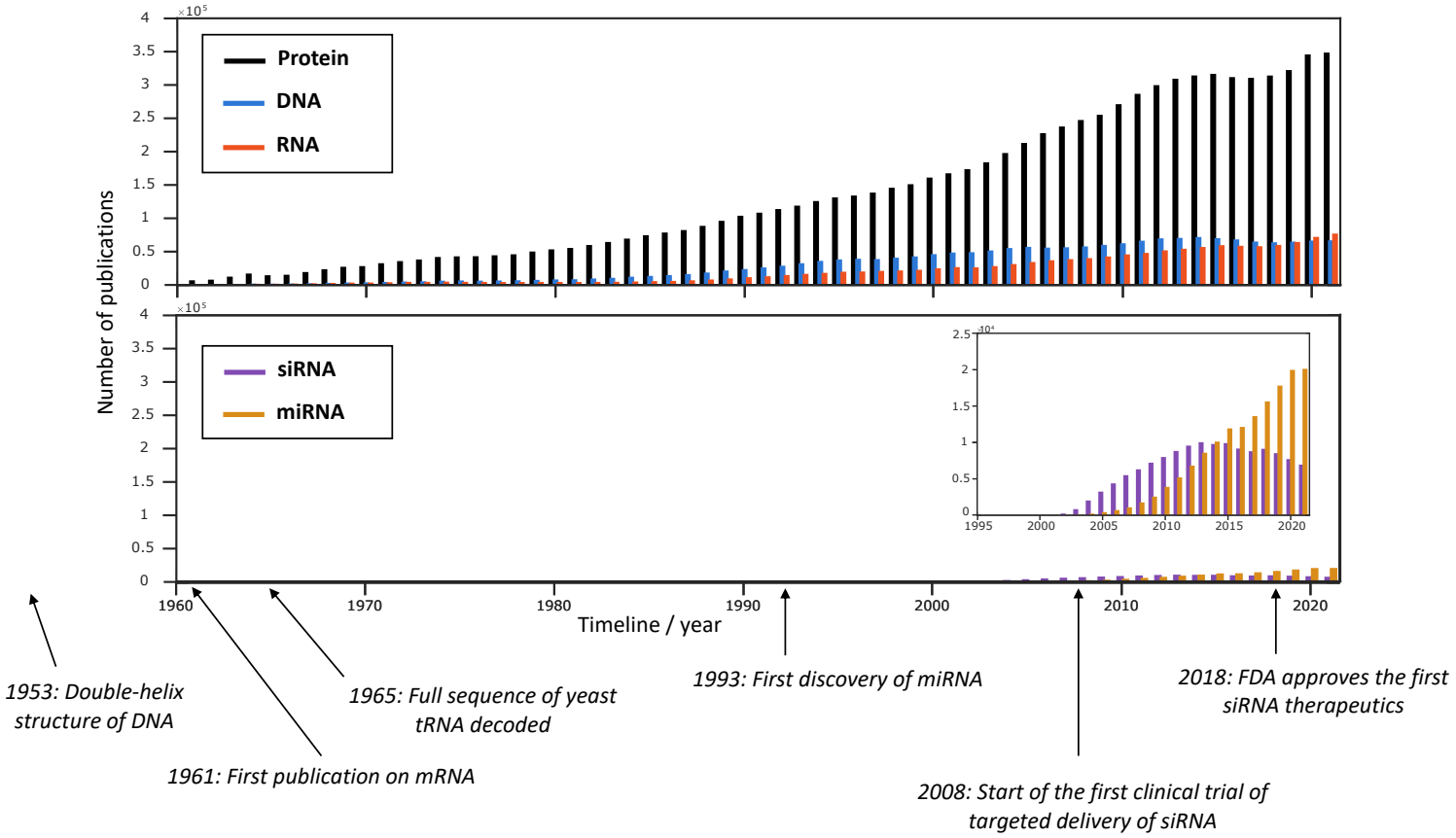

Figure 1: Interest evolution and discoveries on nucleic acids over the past decades.

in DNAs); or purines divided into adenine (A) and guanine (G). Pyrimidines contain heterocyclic groups located on a six-member ring, while purines are composed of a six-member ring connected to a five-member ring.

Those bases are each connected to a sugar, a five-member ring containing either a 2'-H group (deoxyribose, in DNAs) or a 2'-hydroxyl (ribose, in RNAs). In nucleosides, the sugar C_1' is connected to the base through N_1 (in pyrimidines) or N_9 (in purines). Nucleotides are composed of the nucleosides covalently bound to a phosphodiester group. To form the nucleic acid chain (see Figure 3), phosphate groups are connected to 5' and 3'-carbons, thus acting as the oligonucleotide backbone.

Presence of the 2'-OH group makes RNAs more labile than DNAs, because of the possibility to form phosphodiester bonds with both 2' and 3'-OH ribose groups.

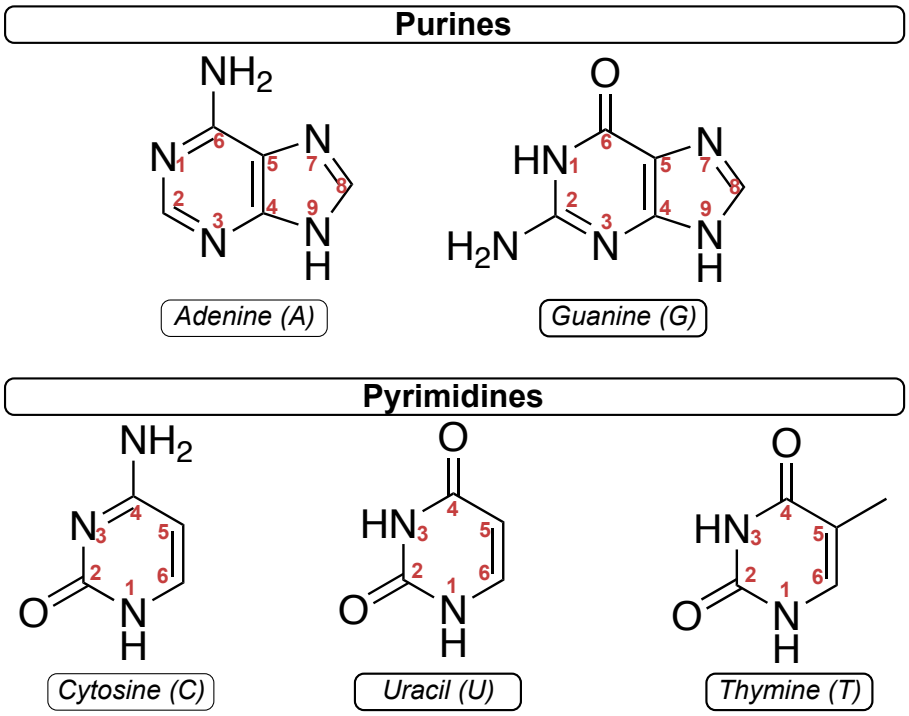

Figure 2: Conventional numbering of nucleic acid nitrogenous bases.

Biomolecules structures can be described in terms of primary, secondary and tertiary structure. In nucleic acids, primary structure refers to the nucleotide sequence, while the secondary structure corresponds to its folding, especially in terms of hydrogens bonds leading to base-pairing. Indeed, the major force driving the secondary structure is coming from hydrogen bondings. The tertiary structure englobes the 3D arrangement of the whole oligonucleotide.

Base-pairing in RNAs tends to favour conformations implying a Watson-Crick conformation but can also adopt various other non-canonical organisations, such as Wobble or Hoogsteen[5][6]. Those are classified according to the bases implied in the base-pairing and the orientation of the base regarding to the ribose. In Watson-Crick base-pairing, bases hybridise in anti-parallel way: one is oriented from 5' to 3' end while the other one is oriented from 3' to 5' end. Watson-Crick base-pairs are usually more stable than non-canonical ones. Purine is there complementary with pyrimidine, due to spatial restriction

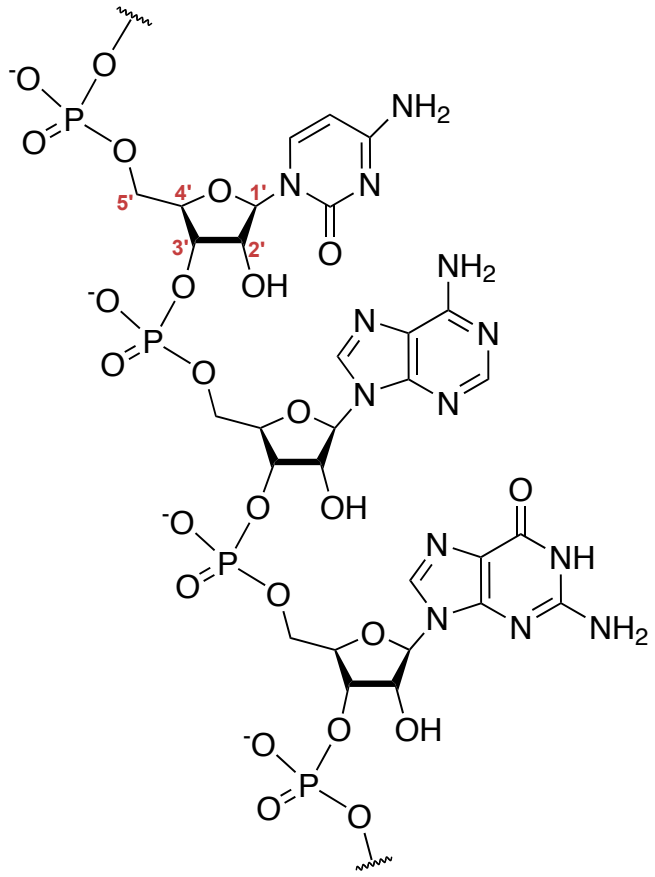

Figure 3: Ribonucleic acid backbone structure.

for hydrogen bonding. Because of the pattern and electronegativity of the different nuclei, the privileged base-pairs imply G-C and A-U (see Figure 4).

Wobble base-pairs tend to appear mostly in tRNAs. G-U and A-C base-pairs are based on tautomerism (respectively keto-enol and amino-imino), meaning a hydrogen atom is relocated, thus changing a covalent double bond[7].

In Hoogsteen scheme, base pairs consist in A-U and G-C[8]. However, the cytosine is protonated on the N3 position, and the purine is rotated by 2π leading to a *syn* conformation.

All these pairing conformations are driven by the opportunity to establish hydrogen bonds. Indeed, a purine is essentially base-paired with a pyrimidine, due to either a lack of space in the case of two purines in a classical helical structure, or by the distance between two pyrimidines in the same structural element leading to the impossibility to form these hydrogen bonds.

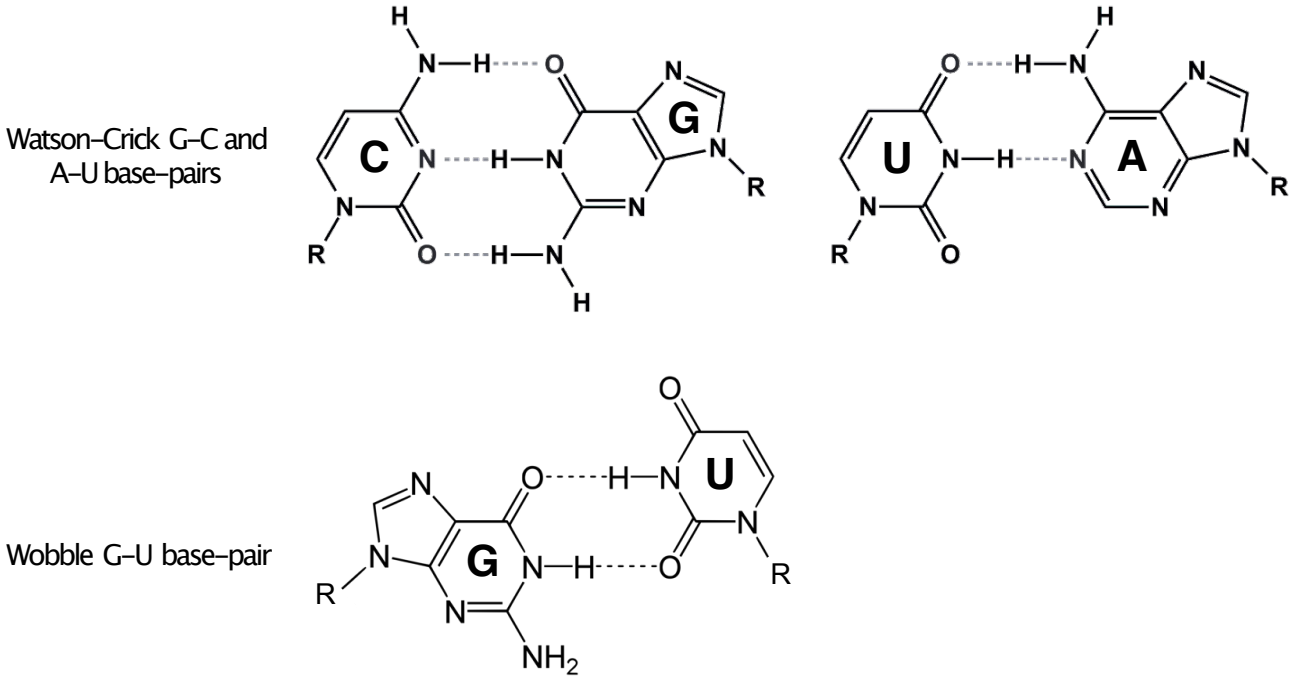

Figure 4: Watson-Crick and Wobble base-pairings. Dotted line represent hydrogen bonds while R emphasizes the ribose bonding.

Secondary structures of RNAs encompass various structural shapes, such as stems constituted of successive base-pairings; bulge, with one or several base coming out of a structured base-paired part of the RNA; and loops, that are often found at the end of strands, or linking helices[9].

Functions

RNAs possess a huge variety of structures and functions, as they are involved in many different cell functions (gene expression, chromosome maintenance, metazoan development, etc.). The main characteristic differentiating them is their involvement in protein coding, however all are implicated in gene regulation, originating the protein synthesis. Coding RNAs and non-coding RNAs are separated into different families: a (non-exhaustive) summary of the main RNA families is described in Figure 5.

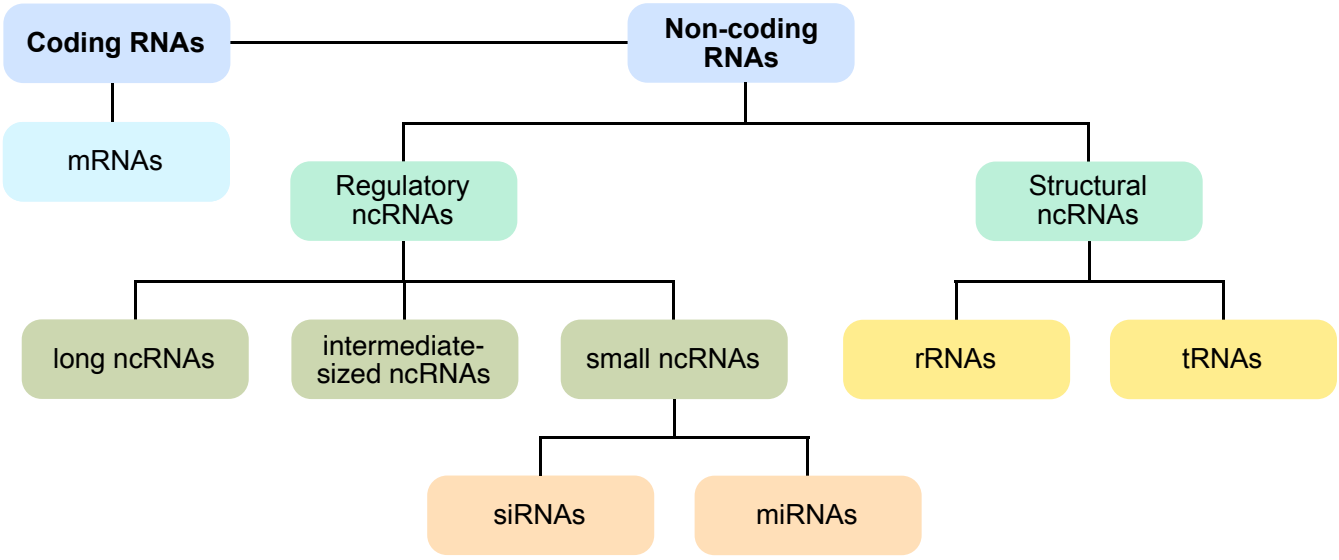

Figure 5: RNA classification: main families of interest are displayed, originally differentiated by their ability to code for proteins. Detailed information is available in main text. The list presented in the scheme above is not exhaustive.

Coding **messenger RNAs** (mRNAs) possess a nucleotide sequence encoding for specific proteins. During the translation process, their *codons* (assemble of three consecutive nucleotides) are read within the ribosome. Ribosomes are large entities composed of a large and a small subunits, regrouping around a mRNA strand. As the ribosome advances along the mRNA sequence, codons are read and interpreted as specific amino acids. Those are assembled into a nascent polypeptide chain until a stop codon is reached, therefore terminating the protein synthesis.

Amongst non-coding RNAs are found the **transfer RNAs** (tRNAs), composed of 70 to 100 nucleotides, that carry the amino acids to the ribosome during the translation of mRNAs into proteins. tRNAs arbor an anticodon, that can bind to various mRNA codons (their overall shape is detailed in further chapter).
The **ribosomal RNAs** (rRNAs) are cell molecules, involved in the catalytic part of protein synthesis. They are loaded onto proteins to form the ribosome subunits, and force the translation process by binding mRNAs and tRNAs.
Small interfering and micro RNAs (siRNAs, miRNAs) are involved in gene regulation, as they can degrade mRNAs or inhibit protein translation. While siRNAs are double-stranded, miRNAs consist of small single strands.

Proteins: structure

Proteins are complex molecules that are major actors of the human body[10]. Proteins have been and are implicated in a wide range of biological functions, such as cycle regulation, nutrient transport, or catalysis.

These biological elements are constituted of a linear succession of amino acids, forming a peptide chain. These protein building blocks are decomposed in two distinct subunits: the peptide backbone and the variable side-chain groups (see Figure 6). The 20 natural amino acid residues are thus differentiable by their side-chain structures.

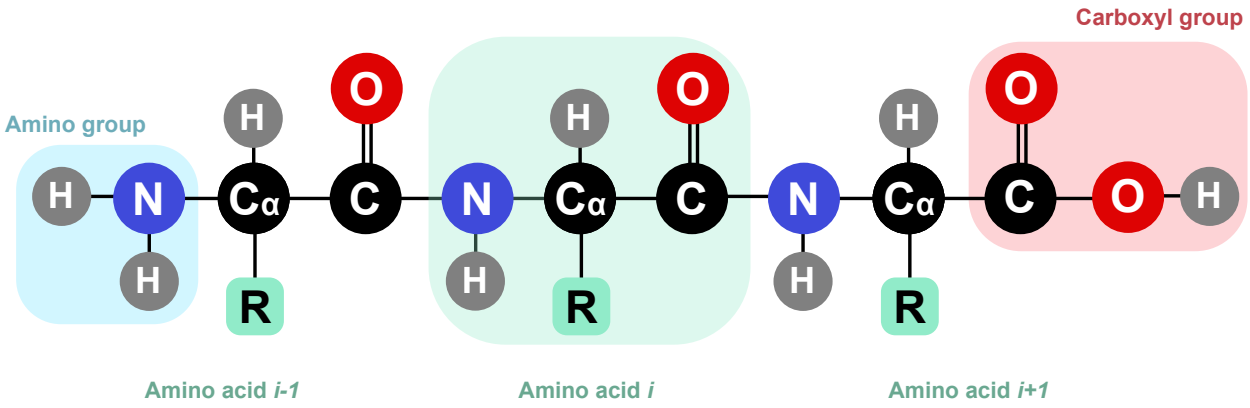

Figure 6: Protein backbone general structure, the R group corresponding to the amino acid specific side-chain.

As well as for RNAs, primary structure refers to the amino acid sequence, while the secondary structure corresponds to its backbone local arrangement. The tertiary structure englobes the 3D structure of a polypeptide chain, and the quaternary structure refers to the assembly of multiple subunits.

NMR spectroscopy

Investigation of biomolecules is fundamental to understand the biological processes originating life, as well as diseases. To get a better insight on molecular mechanisms, studies of proteins and nucleic acids structures is essential[11]. However, their structure is not the only aspect governing their functions. The motions occurring on several timescales by the

whole molecule or its subdomains may strongly impact how its role is completed. The activity of the biomolecules is connected to their environment, and the cellular biomolecules surrounding them. Therefore, the motions and interactions between them are essential to approach a complete understanding of their functions.

Structural biology is an extending area of research focused on the comprehension on these biomolecular mechanisms. As well-established techniques, such as X-ray crystallography and NMR spectroscopy, already exist, instrumentation is still thoroughly developed to extend the techniques limits. The available techniques all possess strong advantages, as well as limitations, that make them undeniably complementary to each other, depending on the biomolecular object, such as its size and physical properties, as well as the focus of the study, such as its structure or interaction with other biomolecules.

Biomolecular investigation may be performed by NMR spectroscopy, especially because of the possibility to label isotopically the nuclei mostly encountered in biomolecules, such as carbons and nitrogens, that can be combined with multidimensional experiments. Both structures and dynamics can be thoroughly investigated by NMR spectroscopy. Indeed, the structure can be studied through specific pulse programs to connect the nuclei chemical shits in a sequential manner while spatial arrangement can also be investigated to comprehend the biomolecule folding.

Biomolecular motions are observable on a wide range of timescales, contributing to the completion of different functions. Changes in the biomolecules structures, commonly referred to as *dynamics*, therefore represent a fundamental step to understand biological mechanisms. NMR spectroscopy is a method of choice regarding the dynamics investigation , as it allows to study the biomolecules in a solution state, thus allowing more motion freedom. The diverse conformations encountered while studying biomolecules can be classified depending on the time window and the domain modifications that are involved. Indeed, dynamics range from local motions, involving a few atoms on a picosecond-nanosecond timescale, to protein folding and domain rearrangements, that can last up to several days.

As structural biology is progressing to investigate a large panel of biomolecules, their structure complexity and multiplicity of interactions unravel. The synthetical, spectroscopic and computational aspects have therefore faced challenges, that require overcoming the limitations and connecting different techniques to completely comprehend the biological meaning.

Thesis

The objective of this PhD has been to establish protocols and investigate RNA dynamics, through different members of their families: tRNAs as well as miRNA:mRNA interactions,

using solid-phase synthesis and liquid-state NMR spectroscopy. Additional projects bring a broaden perspective on biomolecules and NMR techniques, including protein studies, as well as solid-state NMR.

In the first part of this manuscript, I provide NMR spectroscopy background, where an overview of the theoretical expressions and the experimental measurements can be compared to unravel biomolecular dynamics. The emphasis is made on both spin relaxation and various techniques of relaxation dispersion, as these are the methods used during this PhD, on both RNAs and proteins of interest.

In Chapter 2, the focus is set on the methodological aspect of the oligonucleotide solid-phase synthesis. The synthesis steps as well as the post-treatment required to get the final form of the oligonucleotides are presented, and the chemical reactions described, focusing on the methodology used in the laboratory.

In Chapter 3, I present the starting methodology used to synthesize RNAs, the results obtained and the issues faced. The improvements and adaptations in the protocol made to overcome those practical problems are detailed, and the final protocol established. The protocol variations required while incorporating a fluorine into an oligonucleotide are presented. The final part of the chapter grants the starting analysis of the synthesised RNAs and their interactions by interpreting both proton and fluorine NMR experiments.

Chapter 4 provides the application of the Lipari-Szabo *model-free* approach, through the acquisition of ^{15}N relaxation data on a non-modified sample of tRNA.

Finally, Chapter 5 presents the characterisation of a lipoprotein, rising from the infection of bacteria cells by T5 bacteriophage. Acquisition and analysis of data for protein assignment, structure determination, and glance at dynamics by spin relaxation at one field, are detailed. Its interaction with a membrane protein is also investigated using NMR titration.

Chapter 1

NMR relaxation theory applied to biomolecules

1.1 NMR spectroscopy and biomolecular dynamics

Resolving the structures of biomolecules has been the principal aim of structural biology for several decades. However, to unravel the complexity of biomolecules, one also need to take into account their dynamics. NMR spectroscopy is one of the techniques able to characterise motions of systems at atomic level over a wide range of timescales. Several methods allow the detection of particular biomolecular motions, such as interactions, reorientations, and internal motions.

1.1.1 Biomolecular motions

In any biomolecular sample, nuclei undergo random motions, due to thermal agitation. These motions generate a fluctuating local magnetic field, resulting mainly from the electronic environment and the spin-spin coupling. The local field are thus dependent of the nuclei and its environment, and influencing the longitudinal and transversal relaxation rates.

In solution, a biomolecule presents different types of random motions: rotational, translational and distortional, that can occur on a large range of timescales. The whole molecule is characterised by a global correlation time τ_c, also designated as the *tumbling*, affected by temperature and viscosity of the solvent.

Different kinds of motions are therefore usually experimented over different timescales (see Figure 1.1). Typically, local and rotational motions are generally observed on fast timescale (ps-ns), that can be probed by spin relaxation. The standard set of spin relaxation experiments used to investigate those dynamics include longitudinal and transverse relaxation rates, as well as heteronuclear steady-state NOE.

Meanwhile, conformational rearrangements may be identified using relaxation dispersion experiments. Catalysis, ligand binding and allostery are linked to molecular motions occurring on intermediate timescale, ranging from μs to ms. These motions may include an exchange between several states. Variations in transverse relaxation (R2) as well as

rotating-frame relaxation rates (R1ρ) are usual relaxation dispersion experiments.

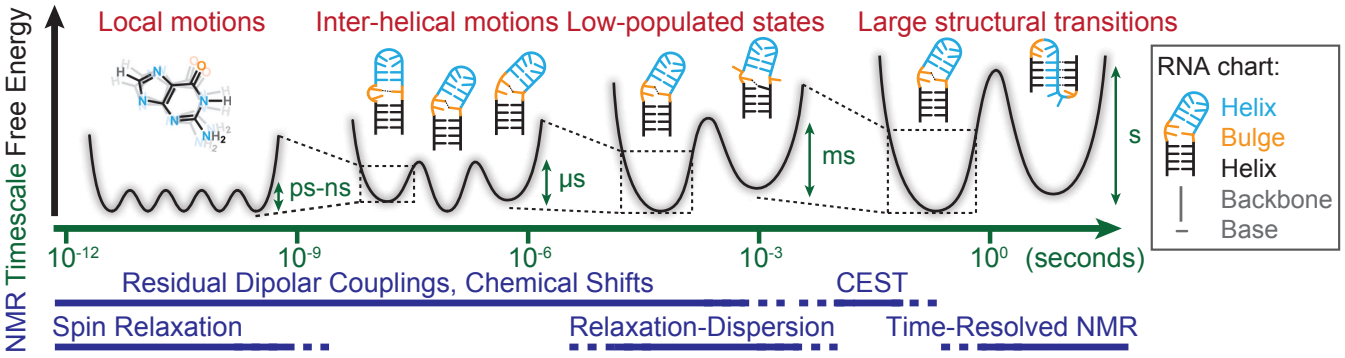

Figure 1.1: Typical motional timescales identifiable by NMR spectroscopy.

Relaxation occurs because of the variations in the local magnetic fields of the nuclear spins, due to thermal motions. ^{15}N relaxation in biomolecules are governed by two interactions: dipolar coupling with the neighbouring ^{1}H (d^2_{NH}) and ^{15}N chemical shift anisotropy (a^2_N)[12]. Other field fluctuations may imply quadrupolar or paramagnetic interactions, that will not be further discussed in this thesis.

1.1.2 Classical description of NMR

In NMR spectroscopy, nuclei are placed in an external field, where the interaction between nuclei and electromagnetic waves inducing spin transitions is probed. In a spectrometer, the magnetisation is oriented along the magnetic field B_0, conventionally placed along the z-axis. The intensity of the magnetisation depends on the spin number, the nuclei gyromagnetic ratio and the B_0 field.

In order to be detectable by NMR spectroscopy, nuclei have to possess a spin quantum number I different from 0. In this thesis, we focus on nuclei which spin is equal to 1/2, essentially ^{1}H, ^{15}N, ^{13}C and ^{19}F. From now on, only spin I=1/2 are considered. The z component of the magnetic moment, designating the interaction between the molecules and the magnetic field, is determined by the magnetic quantum number m, taking $2I$+1 values (-I, -I+1, ..., I-1, I):

$$\mu_z = \gamma m \hbar \tag{1.1}$$

with γ the gyromagnetic ratio of the considered nuclei, and $\hbar$ the reduced Planck constant.

While being placed in a static magnetic field B_0, magnetic moment can adopt two orientations (parallel and antiparallel to the field), which energies can be calculated as following[13]:

$$\begin{aligned} E = -\mu_z B_0 = -\gamma m \hbar B_0 = -mh\nu_0 \\ E_\alpha = -\frac{1}{2}\gamma\hbar B_0 \\ E_\beta = \frac{1}{2}\gamma\hbar B_0 \end{aligned} \tag{1.2}$$

The energy difference between two states is therefore obtained by subtracting the energy in the β and the α states[14]:

$$\Delta E = \gamma\hbar B_0 \tag{1.3}$$

Nuclei magnetic momentum, when placed in an intense magnetic field such as B_0, are animated by a *precession* motion, around the z-axis. The frequency at which this motion is occurring is defined by Larmor relation, either using ω_0 in rad.s^{-1} or ν_0 in Hz[13][14]:

$$\begin{aligned} \omega_0 = -\gamma B_0 \\ \nu_0 = -\frac{\gamma B_0}{2\pi} \end{aligned} \tag{1.4}$$

To rotate the magnetisation, another field B_1, a *radio-frequency (RF)* impulsion, is applied at a π rotation from the z-axis. Once the rotation is completed, the field B_1 is turned off. The magnetisation can therefore be measured in the (xy) plane.

The z-axis is called the longitudinal axis and the (xy) plane is referred to as the transverse plane. During an NMR experiment, the magnetisation possesses both a component parallel to B_0 (longitudinal magnetisation M_z), and a perpendicular component (transverse magnetisation M_{xy}).

The Boltzmann distribution characterises the ratio of spins in the lower and higher energy states at thermal equilibrium:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{E_\beta - E_\alpha}{kT}} \tag{1.5}$$

After being rotated in the transversal plane, the magnetisation is progressively going back to its equilibrium position, along the B_0 field, thus undergoing a phenomenon called *relaxation*. Relaxation mechanisms are the connexion between the theoretical approach of biomolecules dynamics, the experimental measurements that can be performed, and the biologically relevant physical interpretation.

1.1.3 Bloch equations

Magnetisation is the vectoriel sum of all magnetic moments, animated by a precession motion. After application of RF pulses influencing the magnetisation, the relaxation phenomenon brings the magnetisation back to the z-axis. Therefore, when the magnetisation has not returned yet to the z-axis with its full amplitude, the magnetisation is undergoing complex motion constituted of a precession around B_0 and a progressive return to an orientation along B_0. This motion can be described using Bloch equations, with M the magnetisation and B the external magnetic field[15]:

$$\frac{d\overrightarrow{M}}{dt} = \overrightarrow{M}(t) \times \gamma \overrightarrow{B}(t) \tag{1.6}$$

where the first term describes the return to equilibrium by relaxation processes and the second one the precession motion.

This equation can be written as matrices:

$$\frac{d}{dt}\begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} = \gamma \begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ B_0 \end{bmatrix} = \begin{bmatrix} \gamma M_x B_0 \\ -\gamma M_y B_0 \\ 0 \end{bmatrix} \tag{1.7}$$

When applying an *rf* field (designated as $\overrightarrow{B}_1$), the resulting magnetic field $\overrightarrow{B}$ becomes the sum of the static field $\overrightarrow{B}_0$ and this perturbing $\overrightarrow{B}_1$ field. In the laboratory frame, the evolution of $\overrightarrow{M}$(t) is therefore dependent on both the static $\overrightarrow{B}_0$ field and the applied $\overrightarrow{B}_1$ field:

$$\frac{d\overrightarrow{M}}{dt} = \overrightarrow{M} \times \gamma \overrightarrow{B}_0 + \overrightarrow{M} \times \gamma \overrightarrow{B}_1 \tag{1.8}$$

As a return to equilibrium was observed after undergoing perturbation, relaxation phenomena were introduced following two first-order processes: the longitudinal relaxation rate R_1 along the z-axis and the transverse relaxation rate R_2 in the (xy) plane. Denoting

the magnetisation at the equilibrium as M_0, the relaxation processes can be introduced as:

$$\begin{aligned} \frac{dM_x}{dt} &= -R_2 M_x(t) \\ \frac{dM_y}{dt} &= -R_2 M_y(t) \\ \frac{dM_z}{dt} &= R_1[M_0 - M_z(t)] \end{aligned} \tag{1.9}$$

The solution to these equations therefore leads to single exponential solutions:

$$\begin{aligned} M_x(t) &= M_x(0)e^{-R_2 t} \\ M_y(t) &= M_y(0)e^{-R_2 t} \\ M_z(t) &= M_0 - [M_0 - M_z(0)]e^{-R_1 t} \end{aligned} \tag{1.10}$$

Relaxation at a certain moment then depends on the initial magnetisation along each coordinate ($M_x(0)$, $M_y(0)$, $M_z(0)$), the relaxation rate (R_2, R_1), and relaxation delay t.

However, relaxation of each spin is not sufficient to completely comprehend the return to equilibrium. The influence of nearby spins, leading to dipolar relaxation, has to be taken into consideration. This description of a *cross-relaxation* pathway is described using Solomon equations.

1.1.4 Energy levels transitions

Dipole-dipole interaction mathematical theory was first described by Bloembergen, Purcell, and Pound in 1948[16]. It was then extended by Solomon, and applied to electron spin relaxation by Bloembergen and Morgan in 1961.

While considering a two-spin system (I and S), the energy levels $\alpha\alpha$, $\alpha\beta$, $\beta\beta$ and $\beta\alpha$ are characterised by the evolution of the populations P due to the different transitions[17].

With the transition rates between the energy levels denoted as W, the differential equations of each state population P is written using the population deviation from the equilibrium (e.g. $\Delta\ P\alpha\alpha = P\alpha\alpha$ - $P^0\alpha\alpha$)[18][12]:

$$
\begin{aligned}
\frac{d}{dt}\Delta P_{\alpha\alpha} &= -(W_I + W_2 + W_S)\Delta P_{\alpha\alpha} + W_I \Delta P_{\beta\alpha} + W_2 \Delta P_{\beta\beta} + W_S \Delta P_{\alpha\beta} \\
\frac{d}{dt}\Delta P_{\alpha\beta} &= -(W_I + W_0 + W_S)\Delta P_{\alpha\beta} + W_I \Delta P_{\beta\beta} + W_0 \Delta P_{\beta\alpha} + W_S \Delta P_{\alpha\alpha} \\
\frac{d}{dt}\Delta P_{\beta\beta} &= -(W_I + W_2 + W_S)\Delta P_{\beta\beta} + W_I \Delta P_{\alpha\beta} + W_2 \Delta P_{\alpha\alpha} + W_S \Delta P_{\beta\alpha} \\
\frac{d}{dt}\Delta P_{\beta\alpha} &= -(W_S + W_0 + W_I)\Delta P_{\beta\alpha} + W_S \Delta P_{\beta\beta} + W_0 \Delta P_{\alpha\beta} + W_I \Delta P_{\alpha\alpha}
\end{aligned}
\tag{1.11}
$$

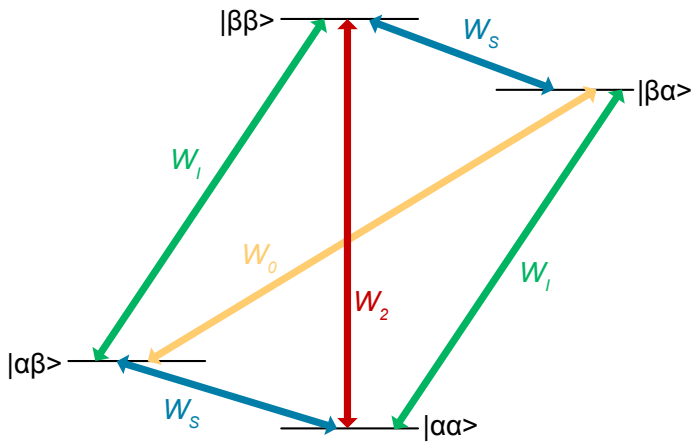

Figure 1.2: Transitions for a two-spin system.

The I (or S) spin magnetisation is then introduced as equal to the population difference across I (or S) spins transitions. Defining $\Delta I_z(\mathrm{t}) = I_z(\mathrm{t})$ - ${I_z}^0$ with ${I_z}^0$ the I spin magnetisation at equilibrium, the magnetisation derivation from equilibrium is given by[12]:

$$
\begin{aligned}
\frac{d}{dt}\Delta I_z &= \frac{d}{dt}\Delta P_{\alpha\alpha} - \frac{d}{dt}\Delta P_{\beta\alpha} - \frac{d}{dt}\Delta P_{\beta\beta} + \frac{d}{dt}\Delta P_{\alpha\beta} \\
\frac{d}{dt}\Delta S_z &= \frac{d}{dt}\Delta P_{\alpha\alpha} + \frac{d}{dt}\Delta P_{\beta\alpha} - \frac{d}{dt}\Delta P_{\beta\beta} - \frac{d}{dt}\Delta P_{\alpha\beta}
\end{aligned}
\tag{1.12}
$$

By combining the two previous equations and the definition of I and S magnetisations[12]:

$$
\begin{aligned}
\frac{d}{dt}\Delta I_z(t) &= -(2W_I + W_0 + W_2)\Delta I_Z(t) - (W_2 - W_0)\Delta S_z(t) \\
\frac{d}{dt}\Delta S_z(t) &= -(2W_S + W_0 + W_2)\Delta S_Z(t) - (W_2 - W_0)\Delta I_z(t)
\end{aligned}
\tag{1.13}
$$

By identifying the relaxation rates, the final form of Solomon equations for a two-spin system is reached[12][19]:

$$\begin{cases} \rho_I = 2W_I + W_0 + W_2 \\ \rho_S = 2W_S + W_0 + W_2 \\ \sigma_{IS} = W_2 - W_0 \end{cases} \\ \frac{d}{dt}\Delta I_z(t) = -\rho_I \Delta I_z(t) - \sigma_{IS}\Delta S_z(t) \\ \frac{d}{dt}\Delta S_z(t) = -\rho_S \Delta S_z(t) - \sigma_{IS}\Delta I_z(t) \tag{1.14}$$

While ρ_I and ρ_S stands for the corresponding spin relaxation (*auto-relaxation*), σ_{IS} denotes the influence of each spin onto the other (*cross-relaxation*).

Solomon equations therefore demonstrate that, along with auto-relaxation rates identifiable in Bloch equations, the return to equilibrium is also influenced by cross-relaxation between the spins. However, to connect experimental measurements and theory, the Wangsness-Bloch-Redfield semi-classical approach is necessary.

1.1.5 Redfield relaxation theory

Redfield relaxation theory, also known as Wangsness-Bloch-Redfield theory, reasons in terms of density operator rather than directly exploiting the energy level populations, therefore not assuming equal probability of transitions between states[20][21].

The laboratory-frame Hamiltonian is divided into two components: the static part $\hat{\mathcal{H}}_0$ and the time-dependent stochastic part $\hat{\mathcal{H}}_1(t)$ in the absence of *radiofrequency (rf)* fields[13]. The static part encompasses the isotropic terms including the Zeeman effect, the isotropic chemical shift and the J-couplings. The stochastic term contains the time dependent dipolar couplings, the chemical shift anisotropy, and the quadrupolar effects[22][23].

$$\begin{cases} \hat{\mathcal{H}}(t) = \hat{\mathcal{H}}_0 + \hat{\mathcal{H}}_1(t) \\ \hat{\mathcal{H}}_0 = \hat{\mathcal{H}}_Z + \hat{\mathcal{H}}_{CS} + \hat{\mathcal{H}}_J \\ \hat{\mathcal{H}}_1(t) = \hat{\mathcal{H}}_D(t) + \hat{\mathcal{H}}_{CSA}(t) + \hat{\mathcal{H}}_Q(t) \end{cases} \tag{1.15}$$

Considering the Liouville-von Neumann equation, the evolution of the time-dependent density operator σ(t) is described by both the fluctuations of $\hat{\mathcal{H}}_1$ due to molecular motions causing relaxation, and the effects of $\hat{\mathcal{H}}_0$:

$$\frac{d\sigma(t)}{dt} = -i[\hat{\mathcal{H}}_0 + \hat{\mathcal{H}}_1, \sigma(t)] \tag{1.16}$$

Transforming this equation into the *interaction frame* where the *spin-lattice* interaction is isolated from the static part, the hamiltonian $\hat{\mathcal{H}}_0$ can be eliminated. The evolution of the time-dependent density operator in the *interaction frame* σ^T(t) then becomes[12][24]:

$$\frac{d\sigma^T(t)}{dt} = -i[\hat{\mathcal{H}}_1^T(t), \sigma^T(t)] \tag{1.17}$$

The solutions to the previous equation are denoted as sums of density operators, that can be integrated over various orders but will be here limited to the first and second-order perturbation theory:

$$\begin{aligned} \sigma^T(t) &= \sigma^T(0) - i\int_0^t [\hat{\mathcal{H}}_1^T(t'), \sigma^T(t')]dt' \\ = \sigma^T(0) - i\int_0^t [\hat{\mathcal{H}}_1^T(t'), \sigma^T(0)]dt' &- i\int_0^t \int_0^{t'} [\hat{\mathcal{H}}_1^T(t'), [\hat{\mathcal{H}}_1^T(t''), \sigma^T(t'')]]dt'dt'' \end{aligned} \tag{1.18}$$

When considering τ = t - t', the Liouville-von Neumann equation becomes[22]:

$$\frac{d\sigma^T(t)}{dt} = -i[\hat{\mathcal{H}}_1^T(t), \sigma^T(0)] - \int_0^t [\hat{\mathcal{H}}_1^T(t), [\hat{\mathcal{H}}_1^T(t-\tau), \sigma^T(t-\tau)]]d\tau \tag{1.19}$$

Since the stochastic process influences different part of the sample in different ways, ensemble average needs to be considered:

$$\frac{\overline{d\sigma^T(t)}}{dt} = -i\overline{[\hat{\mathcal{H}}_1^T(t), \sigma^T(0)]} - \int_0^t \overline{[\hat{\mathcal{H}}_1^T(t), [\hat{\mathcal{H}}_1^T(t-\tau), \sigma^T(t-\tau)]]}d\tau \tag{1.20}$$

Since the average of the fluctuations in the $\hat{\mathcal{H}}_1$ part is equal to 0, the first term of the Liouville-von Neumann equation (2.20) becomes zero, and the whole equation simplifies

to:

$$\frac{d\overline{\sigma^T(t)}}{dt} = -\int_0^t \overline{[\hat{\mathcal{H}}_1^T(t), [\hat{\mathcal{H}}_1^T(t-\tau), \sigma^T(t-\tau)]]} d\tau \tag{1.21}$$

When introducing a thermal equilibrium density operator ${\sigma^T}_{eq}$, the previous equation becomes[25][26]:

$$\frac{d\overline{\sigma^T(t)}}{dt} = -\int_0^t \overline{[\hat{\mathcal{H}}_1^T(t), [\hat{\mathcal{H}}_1^T(t-\tau), \sigma^T(t-\tau) - \sigma_{eq}^T]]} d\tau \tag{1.22}$$

Assuming τ is short, meaning σ^T(t - τ) is approximately σ^T(t), is equivalent to assuming that the relaxation processes are on highly faster timescales than the variations in the density operator. This is the weak collision limit, leading to the following equation:

$$\frac{d\overline{\sigma^T(t)}}{dt} = -\int_0^\infty \overline{[\hat{\mathcal{H}}_1^T(t), [\hat{\mathcal{H}}_1^T(t-\tau)], \sigma^T(t) - \sigma_{eq}^T]} d\tau \tag{1.23}$$

Decomposition of the stochastic Hamiltonian into a sum of random functions of spatial variables ${F^q}_k$(t) and a tensor spin operator ${A^q}_k$, with k the rank of the tensor, can be implemented to go back to the laboratory frame[22][24][27]. In NMR spectroscopy, k is usually equal to one or two.

$$\hat{\mathcal{H}}_1(t) = \sum_{q=-k}^{k} (-1)^q F_k^{-q}(t) A_k^q \tag{1.24}$$

The spin operator ${\mathrm{A}^q}_k$ in the interaction frame becomes[22]:

$$A_k^q = \sum_p A_{kp}^q exp(i\omega_p^q t) \tag{1.25}$$

Combining Equations (1.17), (1.23) and (1.24)[22]:

$$\begin{aligned} \frac{d\sigma^T(t)}{dt} = &-\sum_{q=-k}^{k} \sum_{p,p'} exp[i(\omega_p^q - \omega_{p'}^q)t][A_{kp'}^{-q}, [A_{kp'}^q \sigma^T(t) - \sigma_{eq}^T]] \\ &\times \int_0^\infty \overline{[F_k^q(t) F_k^{-q}(t-\tau) exp(-i\omega_p^q \tau) d\tau]} \end{aligned} \tag{1.26}$$

where the terms $F^q{}_k(t)F^{-q}{}_k(t - \tau)$ are *correlation functions*, and the real part of the integral is the *power spectral density function* $j^q(\omega)$:

$$\begin{aligned} j^q(\omega) &= 2Re\left\{\int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}exp(-i\omega\tau)d\tau\right\} \\ j^q(\omega) &= Re\left\{\int_{-\infty}^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}exp(-i\omega\tau)d\tau\right\} \end{aligned} \tag{1.27}$$

The power spectral density function is thus an even function of τ and ω.

1.2 Spectral density function and relaxation mechanisms

1.2.1 Spectral density function

In the case of isotropic fluids in the high-temperature limit, only one spectral density function has to be calculated[28]:

$$j^q(\omega) = (-1)^q j^0(\omega) = (-1)^q j(\omega) \tag{1.28}$$

Rank 2 tensor operators are then used to describe the relaxation mechanisms of interest: the dipolar interaction and the chemical shift anisotropy, detailed in a further section. Random functions of the real part of the integral in equation (1.26) are factored as[27]:

$$F_2^0(t) = c_0(t)Y_2^0[\Omega(t)] \tag{1.29}$$

The spectral density function can then be written:

$$\begin{aligned} j(\omega) &= Re\left\{\int_{-\infty}^\infty C(\tau)exp(-i\omega\tau)d\tau\right\} \\ C(\tau) &= \overline{c_0(t)c_0(t+\tau)Y_2^0[\Omega(t)]Y_2^0[\Omega(t+\tau)]} \end{aligned} \tag{1.30}$$

where $C(\tau)$ is the correlation function. $Y_2{}^0[\Omega(t)]$ is a second-order spherical harmonic function, that characterises the orientation of a vector in the interaction principal direction, in the laboratory frame.

For the dipolar interaction, the principal direction is the internuclear bond, while for the chemical shift anisotropy, it is the tensor principal axis (e.g. the axis of symmetry in axially symmetric tensor).

The spectral density describes the fluctuations distribution of $Y_2{}^0[\Omega(\mathrm{t})]$ over a frequency range, thus measuring the contribution of the molecular motions, that cause the relaxation processes[28].

For a rigid molecule, the spatial variable is constant ($\mathrm{c_0(t)} = \mathrm{c_0}$) and the orientation spectral density function $J(\omega)$ is given by[29]:

$$\begin{gathered} J(\omega) = Re\left\{\int_{-\infty}^{\infty} C_{00}^2(\tau) exp(-i\omega\tau) d\tau\right\} \\ with C_{00}^2(\tau) = \overline{Y_2^0[\Omega(t)] Y_2^0[\Omega(t+\tau)]} \end{gathered} \tag{1.31}$$

In the case of an isotropic diffusion of a rigid rotor, the orientation correlation function is given by[30]:

$$C_{00}^2(\tau) = \frac{1}{5} e^{\frac{-\tau}{\tau_c}} \tag{1.32}$$

where τ_c is the global correlation time, usually in the order of a few nanoseconds.

The corresponding spectral density function leads to a Lorentzian[29]:

$$J(\omega) = \frac{2}{5} \frac{\tau_c}{(1 + \omega^2 \tau_c^2)} \tag{1.33}$$

However, both overall tumbling and internal motions of nuclei are contributing to the fluctuations of the interactions. If the global and internal motions are considered independent, then the correlation function for isotropic rotational diffusion becomes[31]:

$$C(\tau) = C_0(\tau) C_1(\tau) \tag{1.34}$$

where $\mathrm{C_0(t)}$ is the correlation function of overall motions (calculated as for $\mathrm{C^2{}_{00}(t)}$ previously), and $\mathrm{C_1(t)}$ is the correlation function for internal motions.

^{15}N relaxation rates in biomolecules are influenced by dipolar interaction and chemical shift anisotropy relaxation mechanisms. Relaxation rates of ^{15}N can then be expressed using these two major processes, and the spectral density functions of the amide bond

vector[22][32]:

$$
\begin{aligned}
R_1 &= \frac{d_{NH}^2}{4}\left[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)\right] + a_N^2 J(\omega_N) \\
R_2 &= \frac{d_{NH}^2}{8}\left[4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_N) + 6J(\omega_H + \omega_N)\right] + \frac{a_N^2}{6}\left[4J(0) + J3(\omega_N)\right] + R_{ex} \\
\eta_{NH} &= 1 + \frac{\gamma_N}{\gamma_H}\frac{d_{NH}^2}{4R_1}\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right]
\end{aligned}
\tag{1.35}
$$

where the anisotropy factor depends on the Larmor frequencies of the spins (here, ^{1}H and ^{15}N) and the chemical shift anisotropy of the ^{15}N spin $\Delta\sigma$, while the dipolar interaction factor depends on the gyromagnetic ratio of the two coupled nuclei and the distance r in between[22]:

$$
\begin{aligned}
a_N &= \frac{\omega_N \Delta\sigma_N}{\sqrt{3}} \\
d_{NH} &= \frac{\mu_0 h \gamma_N \gamma_H}{8\pi^2 r_{NH}^3}
\end{aligned}
\tag{1.36}
$$

1.2.2 Dipolar interaction and chemical shift anisotropy

Dipole-dipole interaction is the effect of several spins affecting each other's magnetic field. The strength of this interaction depends on the spin nature, the distance and the angle between the spins, as well as their motions. Indeed, the motions of the spins relative to one another strongly impacts the relaxation rates.

In isotropic solution, dipolar couplings average to zero. However, their impact on spin relaxation is observable through nuclear Overhauser effect (NOE) experiments.

The electronic cloud surrounding the nuclei generates a magnetic field responsible for the chemical shift, that is usually not isotropic. This local electronic environment should therefore be described by a tensor. Because of the molecular motions, this tensor is re-oriented in the laboratory frame, and thus becomes time-dependent. This is the reason for *chemical shift anisotropy* (CSA) relaxation. In liquid-state NMR, the fast tumbling of the molecule averages the anisotropy and one chemical shift is observed. However, it is still impactful in the relaxation contribution.

CSA is particularly important when the biomolecule has an asymmetric structure, and experiences relatively slow motions. Its relaxation effect scales with the square of the external magnetic field. Whereas the CSA relaxation could be negligible for protons in aromatic

systems, it is essential to take it into account while observing heteronuclei.

1.2.3 Lipari-Szabo *model-free* formalism

The investigation of biomolecular dynamics in NMR expresses the necessity to access information about the relaxation processes. In the Lipari-Szabo *model-free* formalism, the spin-lattice relaxation rate R_1, the spin-spin relaxation rate R_2, and the heteronuclear steady-state NOE are analysed at one or various magnetic fields, to extract physically meaningful information[33].

The longitudinal relaxation (also called spin-lattice relaxation, $T_1=1/R_1$) describes the return to equilibrium of the z-magnetisation, while the transverse relaxation (also called spin-spin relaxation, $T_2 = 1/R_2$) describes the decay of the *(x,y)*-magnetisation. Alternatively to the T_2, one can measure the spin-lattice relaxation while applying a spin-lock and thus generating a magnetic field perpendicular to the static field: this is the spin-lattice relaxation in the rotating frame ($T_1\rho = 1/R_1\rho$).

Transverse R_2 and $R_1\rho$ can be calculated from each other using the following equation:

$$R_2 = \frac{R_{1\rho}}{sin^2(\theta)} - \frac{R_1}{tan^2(\theta)} \tag{1.37}$$

with $tan^2(\theta) = \omega_1/\Omega$ where ω_1 is the spin-lock RF field and Ω the offset from the nuclei carrier.

The *steady-state* NOE experiments selectively irradiates a spin resonance to saturate the transition between α and β states, and propagate the NOE. Indeed, if a spin resonance is irradiated (e.g. protons), a variation in signal intensity of other nearby resonances (e.g. nitrogens) can be observed. This effect is linked to the dipolar relaxation mechanism: due to the dipolar coupling influence, the perturbation provoked on the spin I propagates to the spin S, leading to cross-relaxation.

A reference spectrum is acquired without irradiation. Here, the ^{15}N-{^{1}H} NOE is measured using interleaved spectra, alternating with and without saturation of the proton signal, and the NOE values are obtained using the ratio of the resonance intensities observable in the spectrum with and without saturation.

The *model-free* formalism was formulated by Lipari and Szabo in 1982[33][34]. In 1990, Clore extended this methodology, thus leading to the *extended model-free* formalism.

The "classical" way of fitting relaxation data had been to assume a motional model, to define an appropriate correlation function to calculate the spectral density, and fit the experimental data. The motional model is accepted when the fit is acceptable.

With Lipari and Szabo approach, a correlation function is approximated and the spectral density function is calculated, then fitted to the experimental data. The fit complexity

then depends on an increasing number of fitted parameters, each one bringing an additional description of the intramolecular motions. Thus, the spectral density function is compared to the experimental data, without preliminary hypothesis on intramolecular motions.

The internal dynamics can be characterised using several parameters:
- the general order parameter S^2, describing the amplitude of the motion, comprised between 0 and 1. It corresponds to a spatial restriction of the motion: the closest the value is to 1 (0), the most rigid (flexible) the observed vector is;
- the effective correlation time τ_e, describing the timescale of the motion from the picosecond to the global tumbling correlation time τ_c (a few nanoseconds);
- the chemical exchange R_{ex}, stating an exchange between two different states of the biomolecule, on a slower timescale (microsecond-millisecond).

The motion is finally generally assumed as a free diffusion in a cone, which half-angle is designated by Θ.

For isotropic overall tumbling, the correlation function for internal motion is designated by:

$$C_{int}(t) = S^2 + (1 - S^2)e^{\frac{-t}{\tau_e}} \tag{1.38}$$

The corresponding spectral density function therefore is obtained by Fourier transform[35]:

$$\begin{gathered} J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1+(\omega\tau_c)^2} + \frac{(1-S^2)\tau_{e'}}{1+(\omega\tau_{e'})^2}\right) \\ with \frac{1}{\tau_{e'}} = \frac{1}{\tau_c} + \frac{1}{\tau_e} \end{gathered} \tag{1.39}$$

For anisotropic global rotation, the corresponding spectral density function is described as following, with A_r and τ_r depending on the orientation and components of the diffusion tensor:

$$\begin{gathered} J(\omega) = \sum_j A_j \frac{S^2\tau_r}{1+(\omega\tau_r)^2} + \frac{(1-S^2)\tau_{e,r}}{1+(\omega\tau_{e,r})^2} \\ with \frac{1}{\tau_{e,r}} = \frac{1}{\tau_r} + \frac{1}{\tau_e} \end{gathered} \tag{1.40}$$

When introducing motions on different timescales (${S^2}_f$ the fastest one and ${S^2}_s$ the slowest one), the global order parameter is defined by:

$$S^2 = S_f^2 S_s^2 \tag{1.41}$$

The correlation and spectral density functions in an isotropic case then become:

$$\begin{gathered} C_{int}(t) = S_f^2 S_s^2 + (1 - S_f^2) e^{\frac{-t}{\tau_f}} + S_f^2 (1 - S_s^2) e^{\frac{-t}{\tau_s}} \\ J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{(1 - S_f^2)\tau_{f'}}{1 + (\omega \tau_{f'})^2} + \frac{(S_f^2 - S^2)\tau_{s'}}{1 + (\omega \tau_{s'})^2} \right) \\ with \frac{1}{\tau_{f'}} = \frac{1}{\tau_c} + \frac{1}{\tau_f} \\ and \frac{1}{\tau_{s'}} = \frac{1}{\tau_c} + \frac{1}{\tau_s} \end{gathered} \tag{1.42}$$

1.2.4 Increasing complexity of the fit

Different cases presenting various number of parameters to fit are compared with experimental data.

Starting with the simpler model, M1, the spectral density used to describe the relaxation is of the form of Equation 1.39 or 1.40 depending on the diffusion tensor, with τ_e considered negligible due to the speed at which the motion occurs. This approximation assumes that the internal motion is greatly faster than the overall tumbling ($\tau_e << \tau_c$), which is valid for internal motions faster than tens of picoseconds while the global motions is slower than a few nanoseconds, in the case of ^{15}N experiments.

If the previous model is considered not describing the experimental data, the analysis (M2) can be made introducing another parameter: the internal correlation time is no longer assumed as fast compared to the overall tumbling, and fitted instead.

In the third model (M3), the internal correlation time is again supposed negligible, and an exchange term R_{ex} is introduced in Equation 1.35. Indeed, chemical exchanges contribute to the transverse relaxation during the spin-lock used to measure $R_1\rho$ (and thus R_2). The exchange is here approximated to a two-site chemical exchange. Other experimental measurements include longitudinal relaxation in a rotating frame $R_{1\rho}$ on and off-resonance, and transverse relaxation with a spin-echo (see following sections).

In M4, the internal correlation time is incorporated again, along with the exchange rate and the order parameter.

Finally, when all of these analysis failed to describe adequately the experimental data, another motion occurring on a different timescale than the global and internal motions

already present is introduced. The internal motions are thus separated in *fast* and *slow* motions.

Here is a summary of the models and parameters to be fitted generally used for the *model-free* approach:

Model number	Parameters fitted
M1	S^2_s
M2	S^2_s, τ_e
M3	S^2_s, R_{ex}
M4	S^2_s, τ_e, R_{ex}
M5	S^2_s, τ_e, S^2_f

Table 1.1: Model-free formalism, model numbers and parameters to fit, additionally to the global correlation time.

The general protocol to analyse relaxation data starts with estimating the overall correlation time, either using a specific pulse sequence or by calculating the R_2/R_1 ratio, eliminating the nuclei experiencing high mobility (low NOE values) or conformational exchange (higher R_2).

The tensor is then chosen, whether it is isotropic, axially symmetric or fully asymmetric. Local order parameters as well as internal correlation times and presence/absence of exchange can then be determined specifically for each residue.

1.2.5 Methodological advances

If *model-free* analysis is not a new methodology to analyse the site-specific biomolecular dynamics on the nanosecond timescale, this is still the most common used method because of its advantages: the absence of assumption regarding the nature of the motion, and the amount of information obtainable. The order parameters are useful probes to relate to biomolecular functions, as they directly reflect on the flexibility of the domains of interest. Additionally, robust programs have been developed for proteins over the years, that could be adapted for oligonucleotides[36][37][38].

Mostly, dynamics studies result in ^{1}H -^{15}N order parameters, very sensitive to the chemical shift anisotropy of the nitrogens, and the distance between protons and nitrogens. Various studies have been performed to get the dipolar couplings and CSA values in proteins[39][40][41], rendering the *model-free* analysis more robust, but less studies have been performed on oligonucleotides[42], leading to uncertainties on the results obtained. Thus, the *model-free* analysis performed within the range of CSA and radius encountered

in the literature may be useful to investigate, and compare to the results obtained.

One of the first spin relaxation studies of RNA was performed by Akke and co-workers, correlating the structure and the flexibility of imino groups[43]. Application of spin relaxation analysis have ranged from calcium binding[44] to protein folding[45]. *Model-free* analysis has been employed in a multitude of cases, notably for proteins (e.g. conformational dynamics of intrinsically disordered proteins [46]), and is started to be used for oligonucleotides (e.g. characterisation of domain motions combined with RNA elongation [47]).
Additionally, various *model-free* approach can be coupled with other techniques such as Residual Dipolar Couplings (RDCs) and Molecular Dynamics (MD) simulations to cross-validate results, and obtain information on larger timescales[48].

1.3 Experimental approaches of NMR spin relaxation

1.3.1 TROSY effect

In classical ^{1}H -^{15}N HSQC experiments, the decoupling allows the removal of J-coupling effects in both direct and indirect dimensions, leading to a single resonance at a defined frequency.

Due to constructive or destructive interaction between the relaxation mechanisms, the different components constituting the multiplet rising in a non-decoupled experiment have different widths, and the averaged resonance is obtained using decoupling schemes.

For large biomolecules, and even more truly at high fields, the dipolar and the CSA-induced relaxations are the major components of the transverse relaxation.

Instead of applying decoupling to get the average of the four components, the Transverse Relaxation-Optimised SpectroscopY (TROSY)-based experiment selects the most slowly relaxing component, the other three being filtered out[49]. This allows to get the sharpest resonance out of the four components.

This technique becomes more effective than averaging out the resonances using decoupling, for large biomolecules and high fields. Whereas, for small molecules, the broadness of the peak is negligible as the four components arbor sharp peaks, thus selecting one of them would only result in a loss of signal.

1.3.2 Pulse sequences for the determination of ^{15}N relaxation rates

The ^{13}C pulses are used to eliminate the cross-correlation effect between ^{13}C and ^{15}N; as the samples in this thesis have only been labelled ^{15}N, this will not be further detailed.

The temperature compensation block (*T-comp*), which is part of the magnetisation recovery delay t_{recov} between scans ensures the same heating applied to all R_1 and $R_1\rho$ experiments, as a compensation pulse is applied after data acquisition to match the longest spin-lock and RF power applied in the experiments.

The first π pulse eliminates the $^{15}N_z$ magnetisation resulting from the TROSY readout scheme[49].

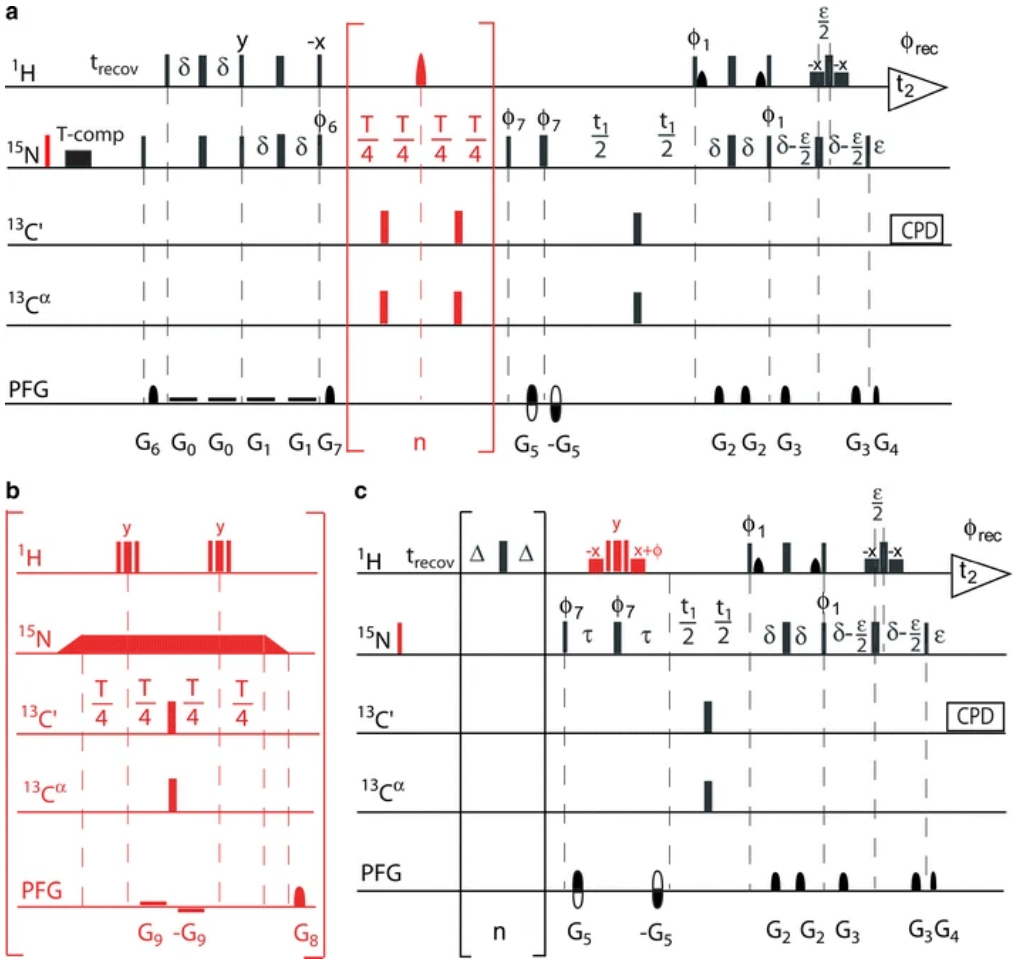

Figure 1.3: Pulse sequences used for the experimental measurements of ^{15}N relaxation rates through TROSY schemes: (a) ^{15}N R_1, substitution of the red bracket leads to (b) ^{15}N $R_1\rho$ and (c) ^{15}N-{^{1}H} NOE sequence. Detailed explanations are available in main text. Reproduced with authorisation from [50].

The starting magnetisation in H_z is transferred via INEPT block and rephased, to obtain N_z magnetisation. A purge pulse is added to remove any residual magnetisation that has not been rephased.

The relaxation delays, to be changed through a list of experiments, influence the magnitude of the remaining magnetisation. The 2π pulses applied during the relaxation delays refocus the magnetisation to avoid coupling effects, and is repeated as this loop an even number of times. In the $R_1\rho$ experiment, a spin-lock is applied during the relaxation time on the ^{15}N channel.

Following the relaxation period of longitudinal ^{15}N magnetisation, the most slowly relaxing component, resulting from the destructive coherence between CSA and dipolar interactions, is selected in both proton and ^{15}N dimensions.

To remove most of the water signal, a WATER suppression by GrAdient Tailored Excitation (WATERGATE) is performed, constituted of a composite pulse surrounded by gradients, that inverts all the resonances except the water signal, which is dephased by the gradients[51].

In ^{15}N-{^{1}H} NOE pulse sequence (Figure 1.3, c), the proton saturation is reached by applying n repetitions of the symmetric delay-2π-delay block, during which the proton carrier is switched to the center of the NH region, and then moved back to the water chemical shift[52].

1.4 Relaxation dispersion

1.4.1 Chemical exchange

The majority of biomolecules are in dynamic equilibrium between various conformers while in NMR samples. Chemical exchange involves the reorganisation of the structure of the biomolecule, including exchange between free and bound forms between biomolecules, conformational rearrangement, and solvent exchange. The dynamic nuclei experiment various environments, thus modifying their chemical shifts. However, the resonances can be absent from the NMR spectra due to the exchange regime. The apparent relaxation R_2 is also altered by an additional exchange rate R_{ex}[53].

Considering a two-site exchange:

$$A \underset{k_{B-A}}{\overset{k_{A-B}}{\rightleftharpoons}} B \tag{1.43}$$

The summed population of each state p_A and p_B is equal to 1, while the chemical shift difference and the precession frequency are equal to the difference between each state. The exchange is characterised by:

$$k_{ex} = k_{A-B} + k_{B-A} \tag{1.44}$$

Depending on this k_{ex} constant, the exchange is categorised into several regimes:

- if $k_{ex} << \Delta\omega_{AB}$: the slow exchange where both signals can be observed in the NMR spectrum;

- if $k_{ex} >> \Delta\omega_{AB}$: the fast exchange where an averaged single chemical shift is observable;

- if none of these conditions is fulfilled ($k_{ex} \approx \Delta\omega_{AB}$), the *coalescence* is reached: the peaks merge and the obtained signal becomes broad.

In the case of slow exchange however, a massive difference in population between states can lead to the observation of a single NMR peak, the other one being absorbed in the experimental noise. These states adopted by the studied biomolecules that are not directly observable through resonances on the NMR spectrum are called *invisible states* and can be investigated through several NMR experiments.

Most common NMR experiments used to probe exchange are Carll-Purcell-Meiboom-Gill (CPMG), Chemical Exchange Saturation Transfer (CEST) and R1ρ experiments, probing exchange occurring on the millisecond to second timescale[54][55][56].

1.4.2 CPMG

CPMG experiments are based on the refocusing of coherent effects such as chemical shifts while dephasing effects occur due to T_2 relaxation[57][58]. The experiments are constituted of a repetitive series of pulses (τ_{cpmg} - π - τ_{cpmg}), at the end of which the transverse magnetisation is analysed.

Over a constant relaxation time, the experiment is repeated with a variable number of π pulses, at a repetition rate ν_{cpmg}, which influence the refocusing of the magnetisation, and allows to probe the presence of an exchange[59].

For very short delays between pulses, the chemical shift is refocused and the exchange contribution to the transverse relaxation is decreased. However, when the pulsing is less frequent, the loss of coherence due to exchange processes increases, and the impact on the transverse relaxation becomes noticeable through an apparent increase.

This method can sample exchange rates going from around tens (10-100) to few thousands (2000-6000) s^{-1}, for populations representing at least 0.5%[60][61].

Relaxation rates $R_{2,eff}$ are monitored extracting resonances intensities as function of the frequency ν_{cpmg}: the frequency of the applied π pulses during the relaxation time delay T_{relax}.

$$\begin{aligned} \nu_{cpmg} &= \frac{1}{4\tau_{cpmg}} \\ R_{2,eff} &= \frac{-1}{T_{relax}} ln\left(\frac{I}{I_0}\right) \end{aligned} \tag{1.45}$$

where I is the intensity of the resonance in the corresponding ^{1}H-^{15}N spectrum, and I_0 the reference intensity.

CPMG experiments applied to ^{1}H, ^{15}N and ^{13}C nuclei are widely used to study interaction between biomolecules, conformational exchange and catalysis.

Information about the exchange kinetics (rate, population) can then be obtained by fitting $R_{2,eff}$ as a function of ν_{cpmg} using theoretical calculations for two- or three-sites exchanges[62].

1.4.3 R1ρ

Rotating frame relaxation rates $R_1\rho$ are measured while applying a spin-lock at specific positions, to probe exchange contributions. The $R_1\rho$ experiments are similar to the CPMG ones, replacing the train of pulses by a spin-lock pulse. Typically, 1 to 6 kHz spin-locks are applied for ^{15}N, and up to 15 kHz for protons and fluorines, using on-resonance and off-resonance pulse sequences.

On-resonance pulse sequences carrier are located on the chemical shift of the resonance to investigate, and the spin-lock strength is varied. For the off-resonance sequences, a spin-lock is applied at various offsets Ω from the frequency of interest. $R_1\rho$ experiments can assess higher frequencies than CPMG experiments, thus investigating faster dynamics. Indeed, this method can sample exchange rates going from around tens to tens of thousands s^{-1}[60][61].

$R_1\rho$ are calculated from single exponential decay of the resonance intensities, obtained at each investigated frequency, for a series of spin-lock durations or offsets.

In the presence of exchange, through the on-resonance experiments, a decrease of $R_1\rho$ is observed while going to higher spin-lock fields, due to the quenching of the exchange leading to the decrease in R_{ex} contribution to the apparent relaxation (see Figure 1.4). Moreover, in the off-resonance experiments, an increase in $R_1\rho$ values will rise at the specific frequency of the exchange conformer, thus allowing to highlight the specific nuclei chemical shift while in another conformational state.

$R_1\rho$ experiments applied to ^1H and ^{19}F nuclei are useful to characterise invisible conformational structures of biomolecules, estimating the kinetics of the exchange in terms of population and rates, as well as probing the chemical shift of the so-called *invisible state*.

For the $R_1\rho$ sequence used for the ^{19}F experiments, starting from a SELOPE pulse sequence for proton $R_1\rho$, a loop was added to convert the experiment in pseudo 2D and a spin-echo to flatten the baseline[63]. Indeed, the Teflon in the probe yields a broad and intense signal, in the -150 to -225 ppm region. Considering the size of the NMR sample resonance compared to the Teflon one, overflow was a recurrent issue, that was overcome by adding a spin-echo before the acquisition.

In the following chapters, these experiments will be applied to different biomolecular systems: microRNA:messenger RNA interaction, transfer RNA and protein dynamics.

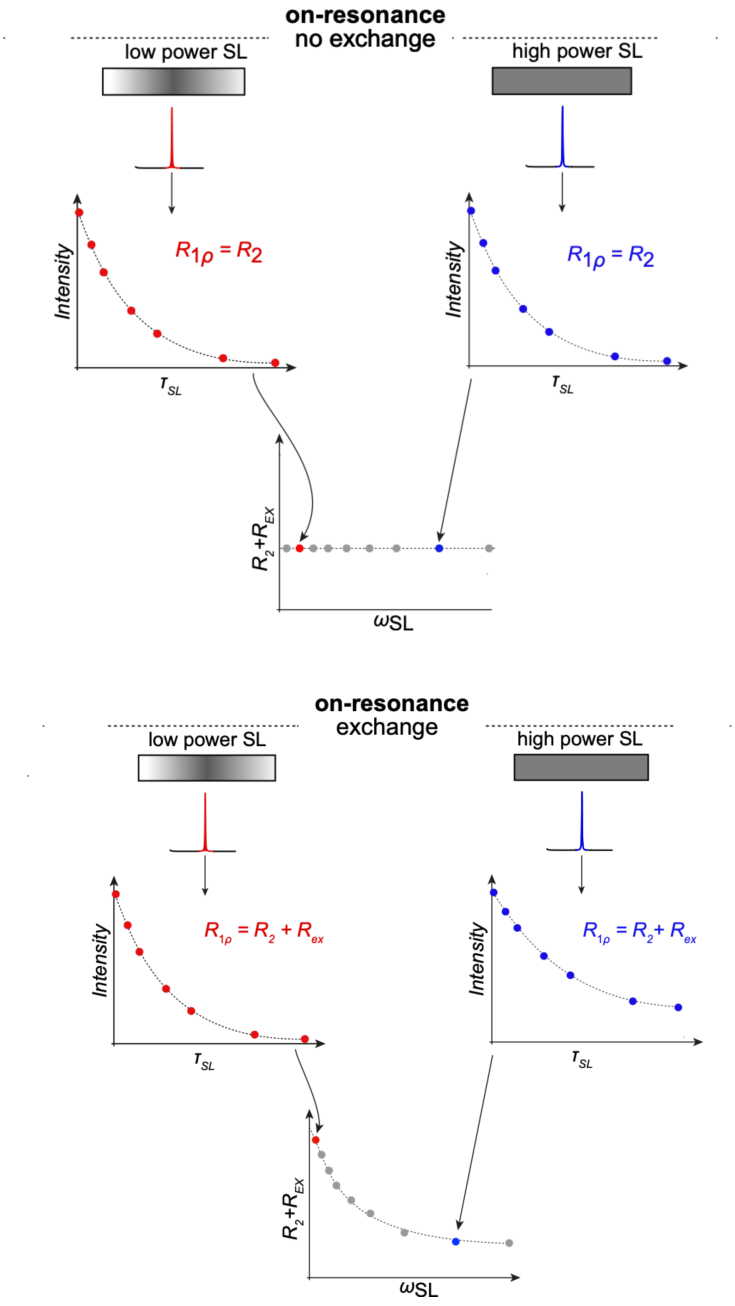

Figure 1.4: $R_1\rho$ on-resonance experiments principle. Several spin-locks ranging from low to high power are applied at a specific resonance, with various durations. The obtained resonance intensity is plotted against the spin-lock duration, and fitted to a single exponential decay. The $R_1\rho$ rate obtained is then plotted against the spin-lock power, and depends on the exchange contribution. Adapted from [55].

Chapter 2

Oligonucleotide chemical synthesis and purification for NMR spectroscopy

2.1 Solid-phase synthesis : global explanation

2.1.1 Study context

Until several decades ago, RNAs were only considered as tools for protein synthesis. However, since the discovery of non-coding RNAs, an interest has been rising to understand the implication and function of those RNAs on the degradation, the inhibition or the translation of the messenger RNAs, their targets.

miRNAs are short RNAs, that are thus ideal for chemical synthesis. Moreover, chemical synthesis allows to introduce specific probes at specific position within the synthesized oligonucleotide sequence, which afterwards grants the possibility to easily connect the information gathered by NMR with its exact position within the RNA. The incorporated probes can be specific nuclei naturally present or not within the molecule (^{19}F) or site-specific labels (^{13}C, ^{15}N)[64][65]. Some of them can be useful for NMR spectroscopy as well as for sample stability, e.g. when introducing -OMe groups in the 2'-position of RNAs[66].

2.1.2 Chemical synthesis overview

Nucleotide phosphoramidites are natural nucleotides, additionally possessing groups protecting their highly reactive sites. These groups can be attached to the base, the phosphorus or the ribose. Chosen types of phosphoramidites, corresponding to the different bases (A, G, C and U/T) are added one after the other to form an oligonucleotide, thus constituted of various bases connected in a desired order.

In solid-phase oligonucleotide chemical synthesis, the synthesis is carried out from the 3' to 5' direction. At the 3'-end of the nucleic acid chain therefore lays a free -OH group on the 3'-carbon of the sugar, while at the 5'-end the 5'-carbon of the ribose is connected to a free phosphate group. A solid support allows the growth of the oligonucleotide. During the synthesis, a cycle of steps is required to connect the nucleotides to each other, remove the unreacted ones, and stabilising the oligonucleotide that is being formed.

The principal advantage of chemical synthesis over biochemical synthesis is that site-specific labelling can be introduced within the oligonucleotide[65][67][68][69].

2.1.3 History

Nucleic acids and their genetic functions have been investigated for decades, especially since the discovery by Watson and Crick of DNAs' double-helix structure in 1953[70].

In the late 1950s, H. Khorana developed solid-phase synthesis of oligonucleotide, interesting himself mainly on 2 topics: stabilising the phosphorylated nucleoside using phosphodiester, and using protections scheme for sequential synthesis[71]. However, the phosphate being unprotected, there were major issues concerning branching of the oligonucleotides, increasing with the size of the oligonucleotide synthesized. The synthesis was then very slow because the oligonucleotide had to be purified between each step to remove reagents.

Khorana's most impactful development was in the area of the protective groups, intending to get rid of these purification steps[72]. In particular, the 5'-hydroxyl protection by 4,4-dimethoxytrityl (DMT) was developed by Khorana's group, and is still used nowadays, as well as some of the bases protections: isobutyryl for guanine, and benzoyl for cytosine and adenine[73].

In 1963, Robert Bruce Merrifield introduced solid-phase peptide synthesis, for which he received the Nobel Prize in Chemistry [74]. Indeed, peptides chemistry was easier to study and reproduce, which lead to the development of solid-phase synthesis using amino acids, instead of phosphoramidites for our oligonucleotides. Peptide automatic synthesisers were born, and became available commercially, the first device being sold by Vega Biotechnologies.

In 1964, solid-phase oligonucleotide synthesis was developed by Robert Letsinger, who was also studying peptide chemical synthesis at the time Merrifield submitted his manuscript [75]. Letsinger then adapted the method for oligonucleotides, introducing at the same time the phosphotriester method, which will be then used by Caruthers in the phosphoramidite method.

Prof. Marvin Caruthers addressed two major issues for chemical synthesis: the swelling of the polymer supports used at the time, and the instability of the intermediate phosphorylated nucleoside[76][77][78]. In the 1980s, Caruthers and Hood created Applied Biosystems Incorporated (ABI), that commercialised the first DNA synthesizer.

2.1.4 Principle

During automated solid-phase synthesis, a solid support is trapped into a column which endings are constituted of two filters. Phosphoramidites and solvents are introduced at the bottom of the column and flow through to reach the top, allowing a better homogeneity and an improved washing of the solvents compared to the reverse way.

Chemical synthesis is usually privileged when synthesising small RNAs or DNAs, of less than 30 nucleotides. This is due to the yields that can be achieved, and the costs that are induced.

The synthesis cycle (see Figure 2.1) is composed of the following steps, detailed in further sections [79][80]:
1/ *Detritylation* of 5'-end of the oligonucleotide being formed (also called *deblocking*);
2/ *Coupling* between the on-going oligonucleotide and the introduced phosphoramidite nucleotide;
3/ *Capping* of the unreacted growing chain on the support;
4/ *Oxidation* of the phosphite triester.

The cycle then goes on until the oligonucleotide is completed.

A series of post-synthesis treatments is then necessary to remove protecting groups, and get the oligonucleotide in its final form.

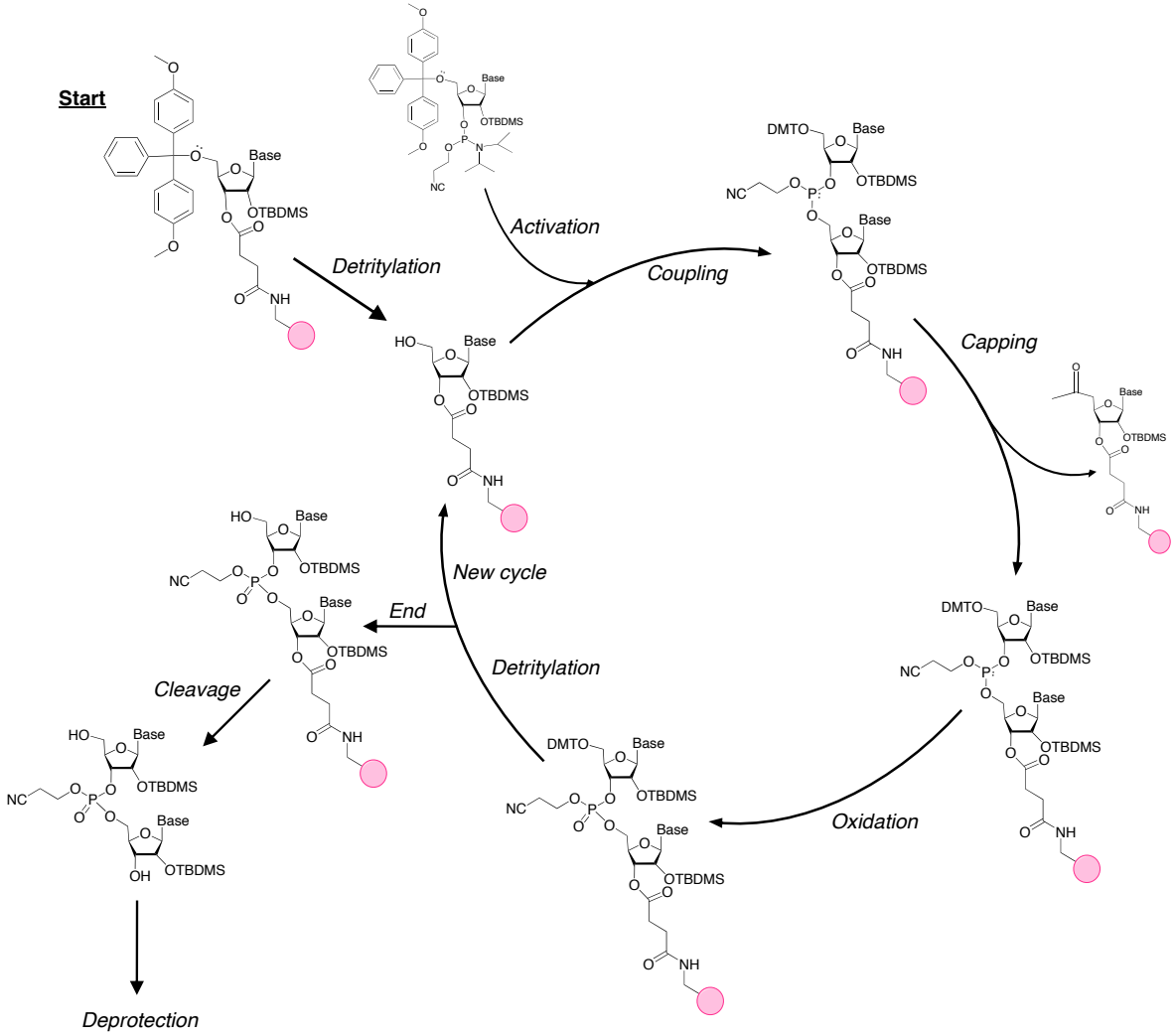

Figure 2.1: Solid-phase oligonucleotide synthesis cycle. The cycle above is demonstrated for a RNA synthesis, with TBDMS protecting group in 2'.
First, the support-attached phosphoramidite is detritylated to unprotect the 5'-end, and is then coupled to the next phosphoramidite (during the coupling step). The unreacted oligonucleotides are capped. The oxidation of the phosphite triester occurs right before the end of the cycle. At this stage, either the oligonucleotide is formed or the cycle undergoes another repetition.

2.1.5 Ribose phosphoramidites for natural RNA synthesis

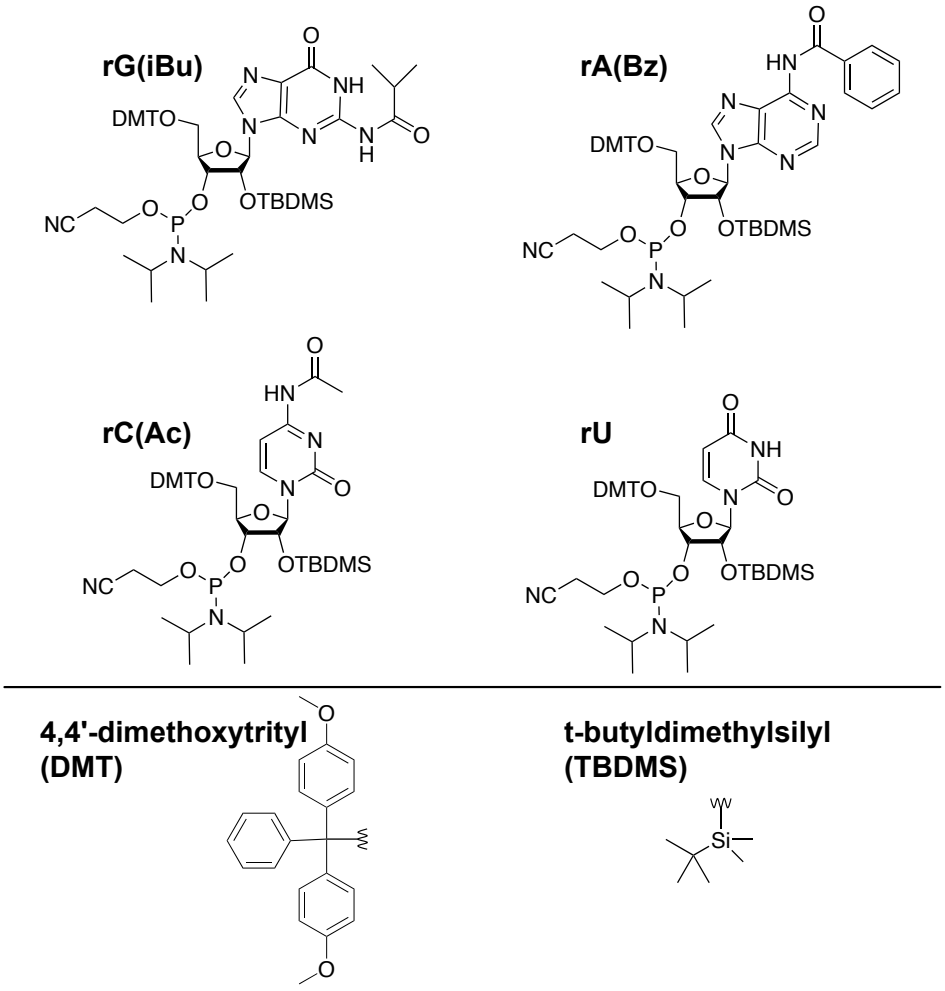

Figure 2.2: Phosphoramidites used for RNA chemical synthesis.
Each of them has a particular base protective group. DMT and TBDMS are common to each phosphoramidite, respectively protecting the alcohol in 5'-end and the 2' on the ribose.

As both amino groups (-NH_2) and hydroxyl groups (-OH) are highly reactive when in contact with the chemicals used during oligonucleotide synthesis, it is needed to use protected nucleotides and therefore prevent unwanted reactions to happen with these groups[81].

Different protecting groups can be used. Commercially available 2'-hydroxyl protecting groups for RNAs include the triisopropyl-silyl-oximethyl (TOM) and the tert-butyldimethylsilyl (TBDMS). For the bases, there are two main types of protection: the classic ones (Bz, Ac, iBu) and the ultramild ones, which require a gentler deprotection of the groups (Ac, PAC).

Here, nucleotide phosphoramidite bases (see Figure 2.2) are protected with:
- an isobutyryl group for guanine: rG(iBu) (MW = 970 g/mol);
- a benzoyl group for adenine: rA(Bz) (MW = 988 g/mol);
- an acetyl group for cytosine: rC(Ac) (MW = 902 g/mol).
Uracil is not protected as there is no exocyclic function: rU (MW = 861 g/mol).

The 5'-hydroxyl group is protected by 4,4'-dimethoxytrityl (DMT), while the phosphite bears both a 2-cyanoethyl group and a diisopropylamine (iPr2N) that is a leaving group allowing the coupling with the next phosphoramidite. As RNA also contains an hydroxyl group on the 2', TBDMS is used to prevent any reaction on this side of the ribose.

2.2 Synthesis steps

The device used in our laboratory is an AKTA oligopilot plus OP10 lab scale oligonucleotide synthesizer, based on a closed circuit encompassing the column, where the phosphoramidites are flowing for a fixed amount of time.

Once the sequence of the desired oligonucleotide is entered into the UNICORN software, several additional inputs are required, including the column parameters, the scale of the synthesis, the coupling and recycling times, the number of equivalences, etc. These variables are detailed in the following sections and in chapter 3.

2.2.1 Solid support

Different kinds of solid-phase supports are available to perform solid-phase synthesis. Standard supports contain a first nucleoside phosphoramidite pre-attached to the support.

Controlled Pore Glass (CPG) is constituted of rigid deep non-swelling 500 Å pores that are mostly used for the chemical synthesis of short oligonucleotides, as the yields tend to

drop when the template is composed of more than 40 bases[82]. This is due to the size of the oligonucleotide comparing to the size of the pores, which with growing templates, reduces diffusion and blocks pores. Large-pore glass (1000 Å or 2000 Å) is more suitable for oligonucleotides but become more fragile[83]. Typical loadings are about 20-50 μmol of nucleoside per gram of support, and can reach 100 μmol/g for high loading varieties.

Polystyrene (PS) beads are performant for the synthesis of small scale oligonucleotide, as it is subject to the phenomenon of swelling: the support expands when solvated, which causes the column back pressure to increase throughout the synthesis. They are typically loaded for 20-30 μmol of nucleoside per gram of resin, but can go up to 350 μmol/g for particularly short oligonucleotides[84].

Other types of support include the universal support, which allows to start oligonucleotide synthesis with any nucleoside[85]. The phosphoramidite that will constitute the 3'-end of the oligonucleotide template is directly coupled to the universal support, then the synthesis goes on as for any other support. The main advantage of this approach is that it can be used for any synthesis, without caring about the first nucleoside. However, if the 3'-terminal nucleoside has a special function or label, the cleavage needed in the case of universal supports may not be appropriate and thus another support will be required.

2.2.2 Detritylation

Before detritylation, the support is washed with acetonitrile (ACN), either to remove the remaining solvents from a previous cycle or to humidify the solid particles.

A protecting group 4,4'-dimethoxytrityl (DMT) has to be removed from the 5'-end of the nucleotide attached to the solid support, to be able to react with the chemicals used in the following steps of the synthesis cycle. The completion of the deprotection is achieved using 3% dichloroacetic acid (DCA) in toluene (see Figure 2.3). It is crucial that the concentration of the solution of DCA is not high, because it would lead not only to detritylation but also to depurination, where purine bases are cleaved from their ribose.

The free DMT carbocation formed absorbs at 495 nm and is therefore coloured in orange while passing through the system tubing. This cation can also be monitored at 380 nm or by conductimetry by the synthesizer. The ratio between the obtained UV peak area and the one of the previous cycle shows a hint about the incorporation yield of the current nucleotide.

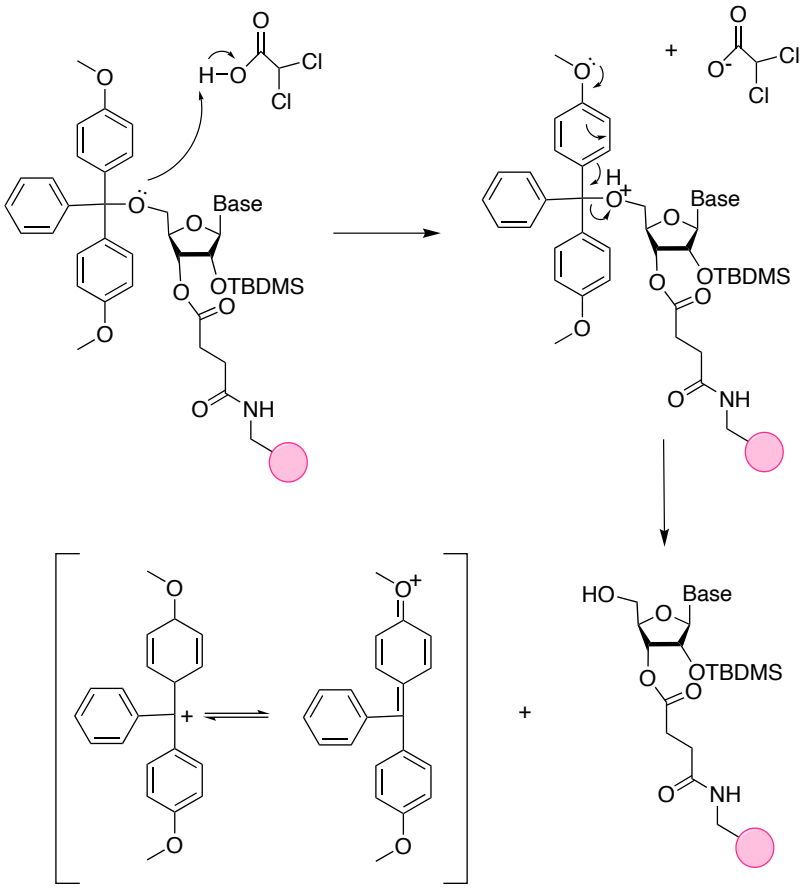

Figure 2.3: DCA mechanism of detritylation over DMT-protected phosphoramidite.

2.2.3 Coupling

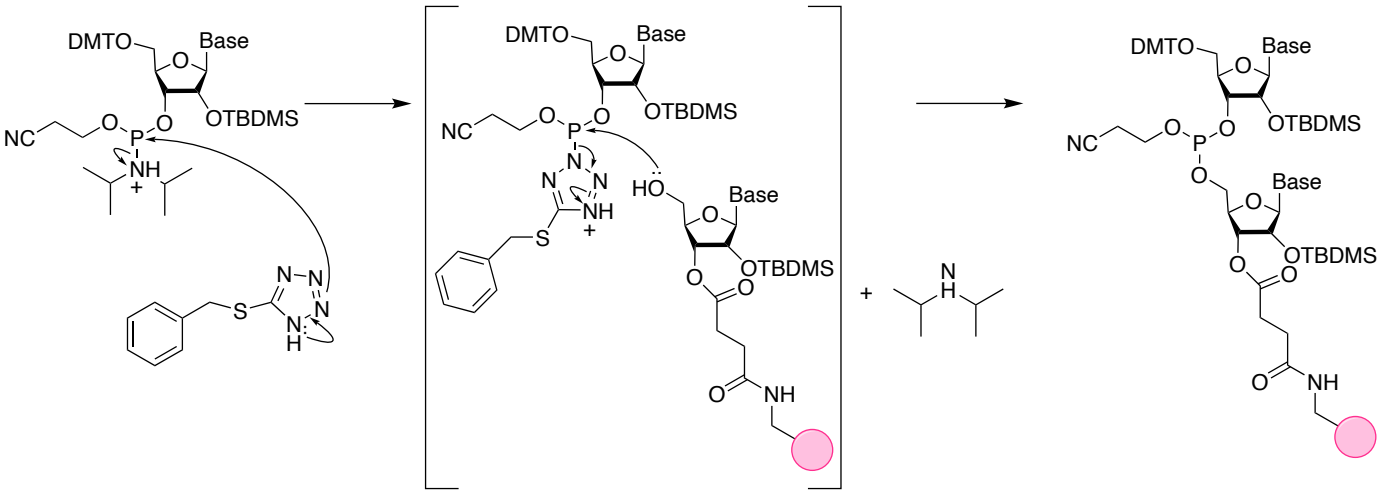

Figure 2.4: 5-benzylthio-^{1}H-tetrazole mechanism: coupling between two subsequent phosphoramidites during a solid-phase oligonucleotide synthesis cycle.

The coupling step consists in incorporating the next phosphoramidite, with the already support-bound ones. The free 5'-hydroxyl function reacts with the added phosphoramidite function.

An activator (in Figure 2.4, the 5-benzylthio-1H-tetrazole), is used to convert the amidite in a tetrazolide derivative. The phosphite group is thus highly reactive with the 5'-hydroxyl of the growing oligonucleotide. The coupling reaction should be accomplished on a timescale of 5 to 15 minutes, leading to coupling efficiencies around 99%.

2.2.4 Capping

The capping step consists in blocking the 5'-hydroxylated functions that have not reacted with the added phosphoramidite in the coupling step (usually about 1 %). An acetyl protecting group is introduced by reaction with acetic anhydride in the presence of N-methylimidazole (see Figure 2.5). This is a crucial point to avoid getting too many aborted templates.

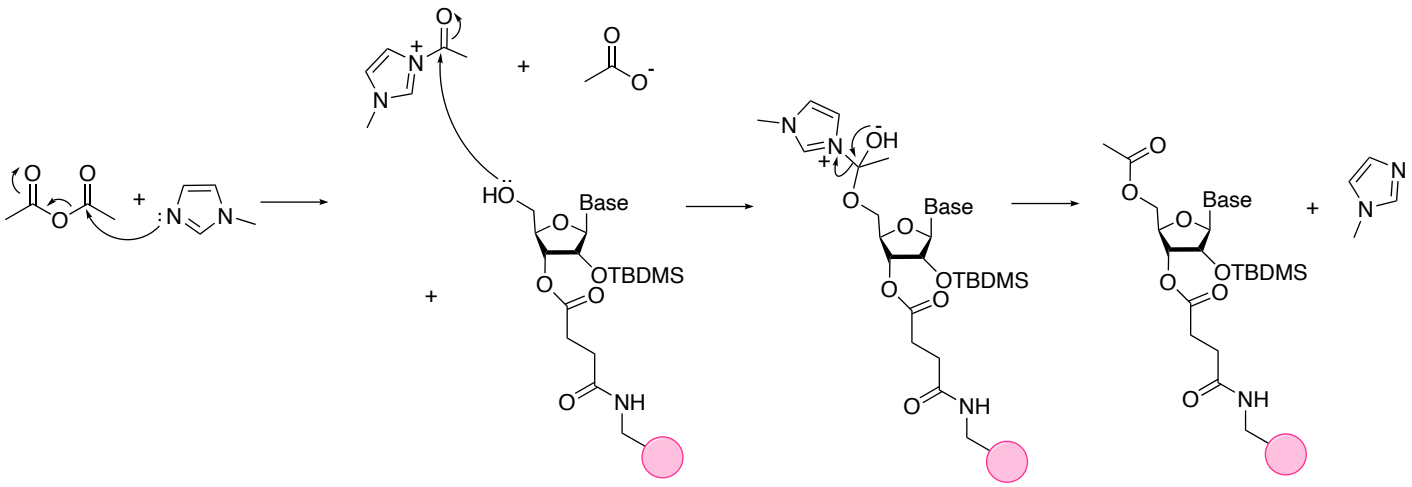

Figure 2.5: Methylimidazole and acetic anhydride capping reaction. The unreacted templates no longer react with the additional phosphoramidites to come.

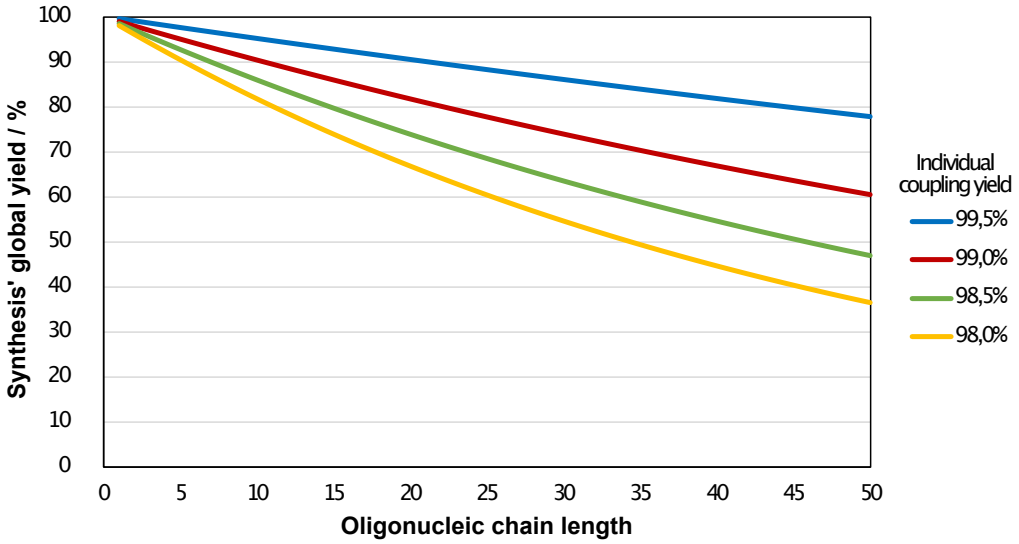

Figure 2.6: Theoretical global synthesis yield.
The global yield depends on phosphoramidite individual coupling yield (colour-coded). The longer the oligonucleotide, the more the yield is dropping.

The capping step is essential: as the DMT group is removed during the detritylation step, the acetyl groups will remain on the aborted templates, inhibiting them from further reacting with the next added nucleotides phosphoramidites which would be pointless and consume phosphoramidites, decreasing the yield and leading to sequential mistakes. Indeed, it is important to avoid getting final oligonucleotides that reach almost the right size (N-1 nucleotides) but with various nucleotide deletion within the sequence, that would highly complicate the purification process.

The yield is a crucial point in chemical synthesis, as it may drop very quickly, especially for longer oligonucleotides, as observable on Figure 2.6. Since the apparition of solid-phase synthesis, typical coupling yields have been improved to reach values between 98 and almost 100%.

For example, to form a 20 base-long oligonucleotide, the global yield synthesis is around 90% / 67% when each phosphoramidite coupling reaches 99.5% / 98% respectively.

2.2.5 Oxidation

Oxidation of phosphite triester (III) in phosphate triester (V) is stabilising the internucleotide linkage. It is achieved using iodine in a water:pyridine 1:9 mixture (see Figure 2.7). Indeed, the phosphite triester is not stable enough to remain in this form during the detritylation step. This step may also be replaced by thiolation to form phosphorothioates.

Phosphate groups are still protected by a cyanoethyl group that prevents unwanted reactions to occur during other steps of the synthesis cycle.

On some synthesizer, the oxidation step is followed by a drying step, in order to make sure the water has been removed from the column, since its presence would radically drop the synthesis yield. Indeed, the degree of degradation will depend on the presence of water, as well as the incorporated phosphoramidites, the exocyclic groups being strongly affected by hydrolysis.

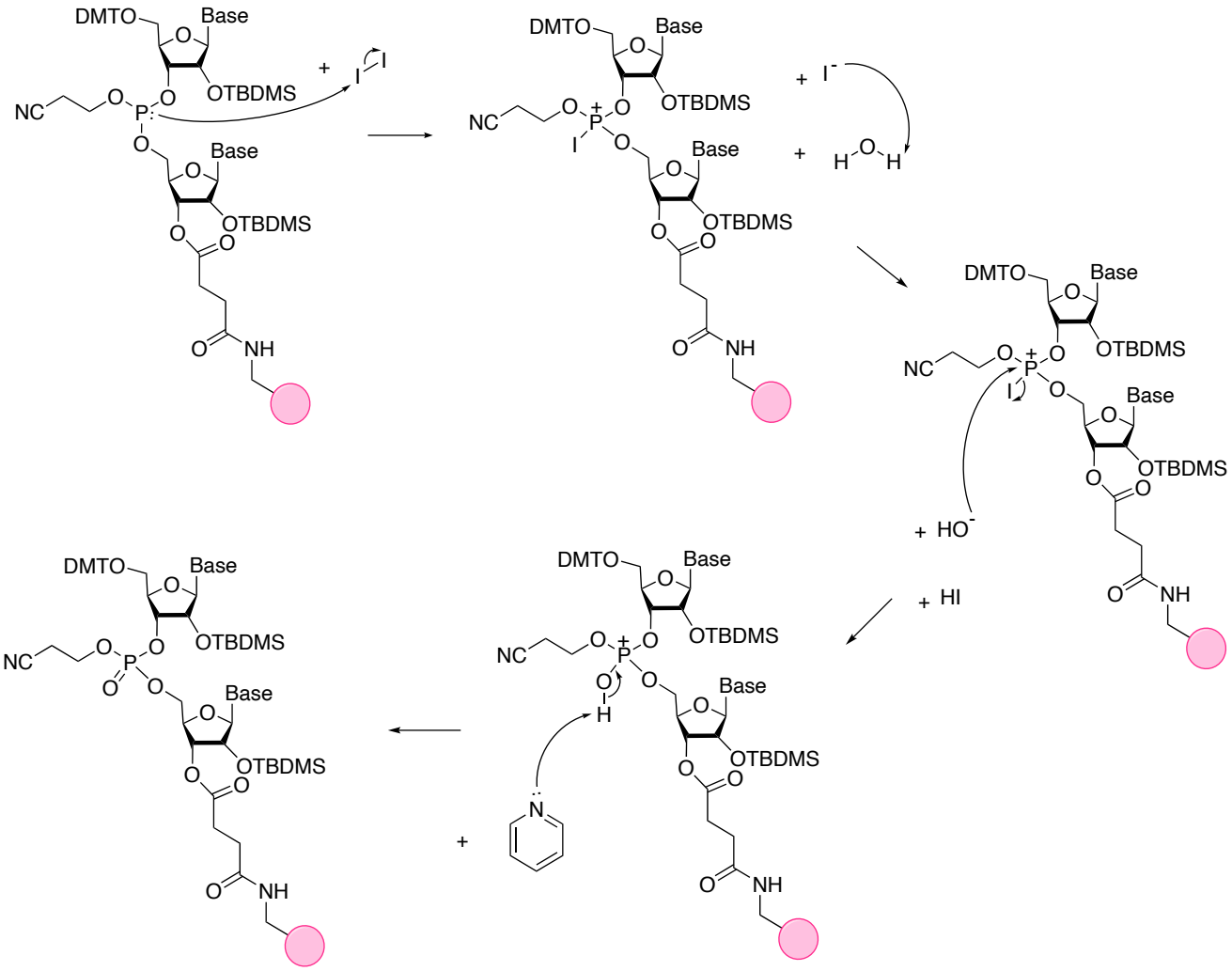

Figure 2.7: Oxidation step using iodine in pyridine and water.

2.3 Post-synthesis treatment

The basic treatment is a key step: on one hand, it separates the biomolecule from its solid support, and on the other hand, it induces the deprotection of the bases and phosphate groups.

2.3.1 Cleavage

To obtain the oligonucleotide synthesized and be able to remove the protecting groups, the template has to be separated from the solid support to which it is attached during the synthesis.[86]

The first step is the transfer of the support, and thus the oligonucleotide, to a container where it can be dried. A solution of methyl-amine in ethanol and ammonium hydroxide is added to the support-bound oligonucleotide. The container is then centrifuged to separate the liquid phase containing the oligonucleotide from the solid phase containing the support.

The post-synthesis treatment can then go on with the deprotection of the oligonucleotide.

2.3.2 Heterocyclic bases deprotection

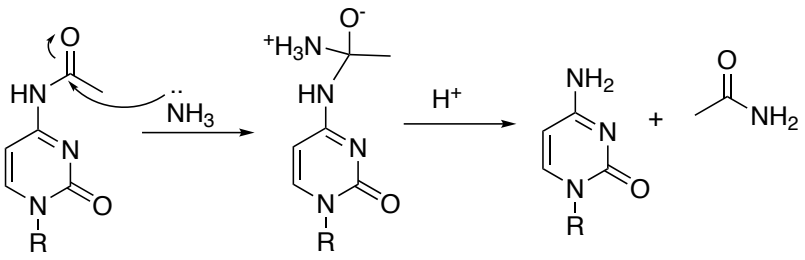

Figure 2.8: Base deprotection.

The exocyclic amino groups present in nucleic bases (except for T and U) are protected with base labile groups (de-protectable in basic conditions), that need to be removed in order to obtain the nucleotides final form. Fortunately, those protecting groups are removed in the treatment realised for the cleavage of the oligonucleotide from the solid support and no additional step is required (see Figure 2.8).

2.3.3 Phosphodiester backbone deprotection

The internucleosidic phosphate or phosphorothioate moieties are protected with 2-cyanoethyl groups which have to be removed to advance towards the final form of the oligonucleotide. Fortunately, as for the bases deprotection, this step is also realised during the cleavage from the support and no additional step is then required (see Figure 2.9).

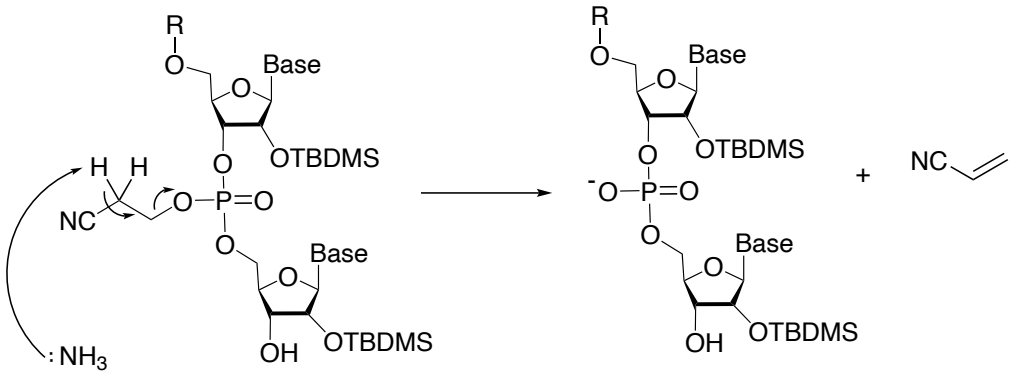

Figure 2.9: Phosphodiester deprotection.

The solution is then dried either by using a rotary evaporator or a speed vacuum system.

2.3.4 Desilylation

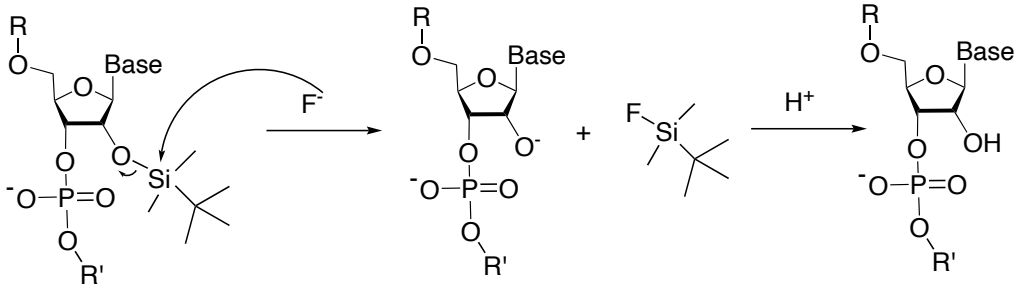

Figure 2.10: Ribose 2' deprotection using TREAT.

The oligonucleotide has to be re-dissolved into the smallest amount of DMSO able to dissolve the whole pellet. It is treated with triethylamine trihydrofluoride (TREAT, see Figure 2.10), which is a particularly hazardous chemical and should be manipulated with

extra care. The solution is then heated at 60°C for about 3h.

To get rid of TREAT and other remaining impurities, the oligonucleotide is then precipitated using cold alcohol (butanol or ethanol) in several steps, and redissolved in a water-based buffer containing Triethylammonium acetate (TEAA) to inactivate the remaining TREAT.

The treated oligonucleotide can then be analysed and purified if necessary.

2.4 Characterisation and purification

2.4.1 HPLC general principle

Chromatography is an analytical technique that was firstly used as a method of separating colours, in the 1900s, hence the name of the method. Mikhail Tswett is a Russian botanist that used the principle of chromatography, involving stationary and mobile phases to purify mixtures of plant pigments[87].

High-performance liquid chromatography (HPLC, or formerly called *high pressure*) is a technique used to analyse and/or purify a mixture by separating its components. It is based on a progressive migration of the molecules in a column that can handle high pressures.

Interactions that influence the migration of each molecule are based on affinities between the compound to be eluted, the mobile phase that is constituted of the buffers, and the stationary phase that is filling the column.

Mobile phases are contained into bottles, and pumped into the system with a flow-rate and for a time depending on the gradient used. The sample is injected through an injector, which can be either manual (loaded through a syringe into a loop) or automatic (a mechanical arm handles the vials where the sample is stored, bringing then to a needle that automatically injects the desired amount into the system).

2.4.2 Stationary and mobile phases

The so-called *normal phase* in HPLC is to have a hydrophilic stationary phase while the mobile phase is hydrophobic.

In our project, the stationary phase is hydrophobic while the mobile phase is hydrophilic: this is called *reversed-phase* HPLC. In this case, the more non-polar the material

is, the longer it will remain in the system as it will be retained by the stationary phase longer.

However, here, *reversed phase* is coupled with *ion-exchange* chromatography[88][89]. The principle is based on the relative ionic strength of the compounds present in the mixture to be separated. The stationary phase is composed of the material charged with the opposite one from the molecule to purify: if the stationary phase is positive, it will retain negatively charge compound.

Here, the longer the RNA, the more negatively charged it is. Hydrophilic or charged analytes are separated on columns using reverse-phase stationary phases, carrying no charges. The mechanism lying behind the reverse-phase ion-pair chromatography is still subject to debates[90]. One hypothesis is that the ion-pairing reagent's alkyl chains (TEAA) interact with the stationary phase, thus leaving the cationic part of the reagent exposed to the analytes that are migrating through the column; and the negatively charged analytes (RNAs) are therefore retained due to their interaction with this modified form of the stationary phase.

2.4.3 Stationary phase: column characterisation

HPLC columns capacity and efficiency is based on their stationary phase (type, particles size, length, diameter, etc.). Whether the normal or reverse-phase is preferred, the nature of the stationary phase will change. Interactions between analytes and stationary phase will depend on the bound phase (C18, C8, C4, etc.), which influence the hydrophobicity and polarity of the stationary phase, thus influencing the retaining of the analyte. The temperature used in a run can also influence the separation obtained.

Columns are characterised by column's plate count (N), a theoretical number depending on the length of the column (L) and the height of a theoretical plate (H):

$$N = \frac{L}{H} \tag{2.1}$$

Smaller plate height implies a larger number of plates in the column, resulting in higher efficiency.

The compound's eluted peak should have a gaussian shape which maximum is the retention time of the compound t_r. The plate count can be calculated using the retention time and the full peak width W_b:

$$N = 16 \times (\frac{t_r}{W_b})^2 \tag{2.2}$$

In the case where the peak width is measured at half-height, the factor 16 is switched to 5.54, which is a more robust method since the full width is hard to differentiate from the noise.

The greater the N number is, the more efficient is the column. To maximise the number of plates, the plate height therefore has to be minimal. The plate height depends on the flow rates, and the optimum flow rate of the column can also be determined for a particular association of mobile phase and analyte. This plates determination is usually done and sent with the column by the manufacturer. It is still possible to perform this test in the laboratory, either to check the material or to ensure that a column is still properly working after it has been used.

The detector, that is located after the column where the separation is happening, is here constituted of a UV-spectrophotometer. The absorption of light of the sample is measured at specified wavelengths after elution (260 nm).

2.4.4 Analytical HPLC: purification optimisation

Several methods are used to analyse and separate the compounds of a mixture analyte.

The composition of the mobile phase can be kept constant through the analysis: this is the *isocratic* method.

Several buffers (most frequently two) with different polarities can also be combined to get the mobile phase. The ratio of the buffers can be modified throughout the analysis. In reversed-phase chromatography, starting from a condition where the mobile phase is (almost) entirely constituted of the more polar buffer, the quantity of the other - less polar - buffer is gradually increased. A buffer *gradient* as described here is often necessary to maintain the separation of the compounds while limiting the time required for the purification. However, this slight gradation between buffers has to be relatively accurate and the equipment has to be able to maintain a flow-rate and a buffer ratio that is precise enough to get reproducible results, which can not be obtained with the lower quality devices.

To optimise a gradient separation by HPLC, the analyte needs to be tested firstly using a wider gradient, which will elute all of the compounds present in the sample. The gradient is then reduced while keeping the stiffness used previously, using the retention time of the peak of the desired compound. The slope of the gradient, depending on the solvent composition, is optimised to get a satisfactory separation of the eluted compounds in a minimum of time. To achieve a better separation, a return to the baseline must be observed between peaks; however it can be practically hard to reach and in that case the edge of the desired peak are collected aside from the main part.

The gradient then starts from a higher non-polar buffer percentage value at the injection time of the sample, which is calculated using the retention time of the large gradient and its composition at elution, shortening the time required for separation. If necessary, the stiffness can then be improved, adapting slightly the buffer composition and the length of the gradient.

It is also possible to optimise the flow-rate used, applying the same gradient optimisation protocol, and compare the different runs using the resolution between the desired compounds and its contaminations. To calculate the resolution between two compounds A and B (A eluting before B), their retention times ($t_{R,A}$ and $t_{R,B}$) and their band width (W_A and W_B) are used:

$$R_s = \frac{2 \times (t_{R,B} - t_{R,A})}{W_B + W_A} \tag{2.3}$$

The higher R_s is, the better the peaks are separated. However, if the compound desired is eluting far away from any other compounds, it may not be necessary to optimise the gradient up to this point, since any resolution greater than one is already enough to separate peaks.

The gradient is usually ending with a purge phase, at a higher composition of non-polar solvent (in reversed-phase chromatography) in order to clean the column from further impurities that could stay attached to the stationary phase once the oligonucleotide of interest has been washed of. For example, the oligonucleotide of right length but that have not been unprotected may stay attached at the composition of elution of the interested oligonucleotide, and therefore limit the stationary phase capacity for the next run.

The column is then reconditioned to the starting condition of the gradient.

2.4.5 Fluorine expectations

The incorporation of fluorine in the RNA 2'-position would suggest a decrease in its hydration properties. This would lead to an increase in hydrophobicity, and therefore increase the retention time compared to the unmodified form of the RNA. However, when incorporating only one fluorine probe, the changes occurring in the RNA properties remain less impactant than the differences between the completed RNA sequence and the aborted ones.

2.4.6 Transfer from analytical to preparative HPLC

Transferring the purification process from analytical to preparative HPLC is necessary for higher quantities of samples, that are required to perform NMR spectroscopy.

The first step of the transfer is to quantify the different volumes in each system: column volume, and dead volume (including dwell volume and volume after the column). These volumes occasion a delay between the gradient implemented in the system and the actual gradient happening during the purification, which may not be negligible when calculating elution composition of peaks, and therefore this determination has to be made before the gradient is done and optimised on the sample.

Concentration and volume of the injected sample can be optimised on analytical HPLC by increasing their amount gradually and observing the effect on the spectrum. Both will tend to broaden the peak when increased, so the resolution can be calculated for any run to make sure the peaks are not overlapping. One should also be careful not to overload the column, and therefore lose sample that will not be retained onto the column due to its maximum capacity reached.

To transfer from analytical to preparative, several parameters have to be adapted: the flow rate and the injection volume, using the following formula:

$$F_{prep} = F_{ana} \times (\frac{D_{prep}}{D_{ana}})^2 \times \frac{D_{part,ana}}{D_{part,prep}} \tag{2.4}$$

with, for either analytical or preparative system, F the flow-rate, D the diameter of the column, and D_{part} the stationary phase particules diameter.

In the case of identical particles in both analytical and preparative columns, the above equation can be simplified to:

$$F_{prep} = F_{ana} \times (\frac{D_{prep}}{D_{ana}})^2 \tag{2.5}$$

Same can be done with the injection volume:

$$V_{inj,prep} = V_{inj,ana} \times \frac{L_{prep,col} \times D^2_{prep,col}}{L_{ana,col} \times D^2_{ana,col}} \tag{2.6}$$

with, for either analytical or preparative system, V_{inj} the volume of injection, L_{col} the length of the column, D_{col} the diameter of the column.

2.4.7 Spectrophotometer principle

To quantify the concentration of the oligonucleotide obtained by chemical synthesis, a spectrophotometer can be used. This instrument allows to measure the concentration of an oligonucleotide in solution, depending on the amount of light absorbed by the sample,

that is placed in a container within the spectrophotometer. A previous blank containing the sample buffer is performed and subtracted from the value obtained using the sample.

Depending on the wavelength, the sample will absorb differently. Light intensity is measured depending on the wavelength of the light sent to the sample, using a diffracting prism on the light beam.

Typically, the absorbance is measured on a large range of wavelength, going from 800 nm to 200 nm, both allowing to quantify the amount of oligonucleotide using its absorbance at 260 nm and checking for contaminants other than aborted templates and if they absorb at other wavelengths. For example, the 260 nm over 280 nm ratio can be used to detect contamination by proteins, which is useful in samples synthesized using biochemical processes.

2.4.8 Concentration determination by UV analysis

To calculate the oligonucleotide concentration from its UV absorbance, optical absorbance at the observed wavelength needs to be taken into account: here, at 260 nm. The appropriate way is to calculate the extinction coefficient exactly for each oligonucleotide, according to its sequence, since the absorbance of each base vary.

$$A = log(\frac{I_0}{I}) = \epsilon \times c \times L \tag{2.7}$$

where A is the absorbance, I_0 and I are the intensities of incident and transmitted light, ϵ is the extinction coefficient, c the oligonucleotide concentration and L the length of the sample container (where the light is passing through).

The oligonucleotide quantified after purification will give a quite accurate result. However, if the concentration is to be calculated before purification (for example to know how much to load on the HPLC column), there will be an error, that will be reduced if the synthesis is clean enough. Indeed, the aborted templates will also absorb at 260 nm and thus distort the UV measurement, increasing the apparent concentration while they will be get rid of in the purification step.

Chapter 3

Investigation of miRNA:mRNA interaction using chemical synthesis and ^{19}F NMR spectroscopy

All the experiments in this chapter have been made in collaboration with Carole Farre and Carole Chaix-Bauvais from ISA (Lyon, France). Marie Papin has also been involved in the targets and *let-7* U6 ^{19}F synthesis and purification processes during her internship. Finally, Laura Troussicot performed the large scale purifications of the targets and the fluorinated *let-7*, in addition to participating in the *let-7* U9 ^{19}F synthesis and troubleshooting.

Concerning RNA characterisation, biochemical (GG-*let-7*) were synthesized and their melting temperatures (along with commercial samples) were measured by Sirine Nouri.

3.1 Study context

The non-coding miRNAs are small highly dynamics RNAs, therefore difficult to analyse exclusively by crystallography or cryo-Electron Microscopy. However, NMR spectroscopy is a method of choice to study both small RNAs, and interactions. Lately, probes have been introduced in biomolecules to get more easily or quickly site-specific information[91]. Here, ^{19}F is a particularly interesting probe, as it spin is 1/2, its gyromagnetic ratio is close to the proton one, and is the only stable fluorine isotope thus having an abundance close to 100 %.

Introducing a specific probe such as ^{19}F, absent from the natural oligonucleotides, is a method of choice to investigate the interaction between a well-identified microRNA, implicated in several forms of cancer, and some mRNA targets, to circumvent the laborious process of assignment, and the expensive cost of ^{15}N/^{13}C labelling.

The chemical synthesis protocol is therefore first established with classical phosphoramidites while synthesising a model RNA and some of the mRNAs targets, then adapted to insert the fluorine probe into the miRNA.

3.1.1 Biogenesis of miRNAs

Over the past decades, the scientific community has developed a strong interest on ribonucleic acids, investigating their structures and behaviours to connect them to their functions. Non-coding RNAs have uncovered a large panel of regulatory functions, linked to various known biological processes [92][93].

Amongst the large variety of ncRNAs, three categories are differentiable according to their size: (1) small ncRNAs, that are constituted of less than 50 nucleotides, such as microRNAs (miRNAs, the focus of this project); (2) long ncRNAs, which template is larger than 500 nucleotides, and (3) intermediate-sized ncRNAs in between[94].

miRNAs are short non-coding RNAs of about 21-25 nucleotides, involved in gene regulation.[96]

In term of biogenesis (see Figure 3.1), they derive from larger precursors that undergo several transformations[97][98][99]. Initially, the large protein RNA polymerase II transcribes DNA strands into primary transcripts of mRNAs and miRNAs (pri-miRNAs)[100]. These pri-miRNAs are then cleaved into about 70 nucleotides-long precursors (pre-miRNAs) by the ribonuclease enzyme Drosha, within the cell nucleus. The precursors are afterwards transferred to the cytoplasm via the transport protein Exportin 5.

There, the Dicer enzyme cleaves the pre-miRNAs into short double-stranded miRNA:miRNA* duplexes, where miRNA is the antisense/guide strand and miRNA* the sense/passenger strand. The miRNA strand is assembled into a larger complex with an argonaute protein (AGO): the RNA-induced silencing complex (RISC)[101][102]. The 5'-end of the miRNAs is both responsible for their stability and loading into the RISC complex.

miRNAs are involved in many development and physiological aspects: apoptosis, proliferation, differentiation and maturation[103][104][105]. miRNAs have a silencing mechanism: by forming complex with their mRNAs targets, they regulate their levels, and thus their functions[106][107]. Their mutation may therefore deregulate cell division and functionalization, leading to the development of diseases, e.g. the proliferation of cells in cancer[108][109].

3.1.2 The *let-7* family

The first miRNAs were discovered in the nematode *Caenorhabditis elegans*, such as *let-7* in the 2000s that encodes a 21-nucleotide long RNA controlling the development from larval last state L4 to adult stage[110][111]. Nowadays several hundreds of miRNAs have been identified[112].

let-7 was since then encountered in several other organisms, and is highly conserved in metazoans, whether they are molluscs or humans, where it is implicated in cell differenti-

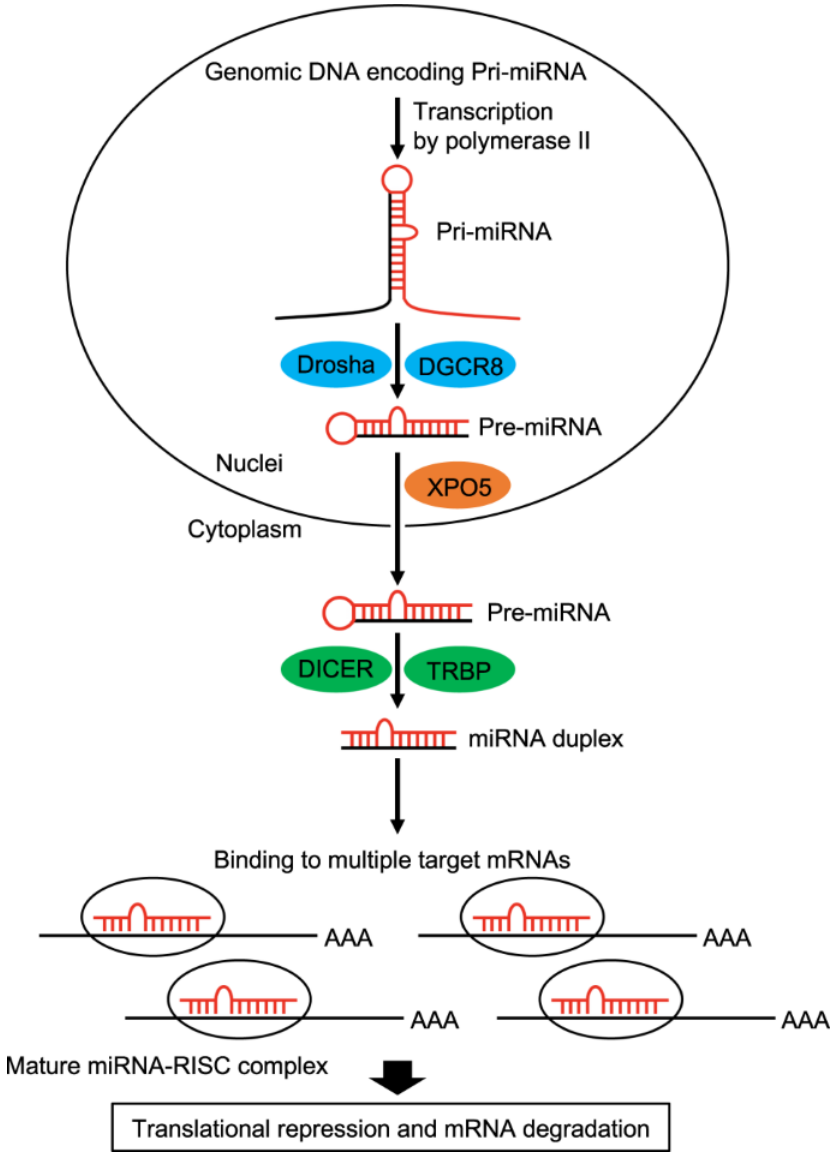

Figure 3.1: miRNA biogenesis: the larger precursors are successively cut and exported to the cytoplasm, where the mature miRNA is loaded onto the argonaute protein to form the RISC complex and interact with its mRNA targets. Reproduced with permission from [95].

ation and division[113][114].

let-7 interacts with several different targets, amongst which the most known include RAS, LIN28, LIN41 and HMGA2. This miRNA binds to the 3'-UnTranslated Region (3'-

UTR) of its targets, inhibiting their translation[115][116]. The mechanism of regulation of *let-7* however is largely unknown.

Human *let-7* have been noticed to be down-regulated when in presence of several cancer forms[117][118][119][120]. It also appears that the reintroduction of *let-7* when in a depletion state prevents the tumorigenesis. *let-7* thus belongs to the tumor suppressor miRNAs, hence promoting the repression or degradation of oncogenic mRNAs.

As miRNAs have multiple targets, a therapy based on them could target a wide range of genes, and potentially increase the efficiency against the spreading or the formation of tumour[121][122].

Mimics of miRNAs could thus silence a broad range of genes, increasing their potential efficiency against the growth and/or dissemination of cancers[123]. Notably, the *let-7* family is a common down-regulated miRNA in several cancers, e.g. lung and colon cancers[124].

Considering their importance in genetic functions, the understanding of the miRNAs behaviour is crucial in order to be able to mimic their approach[125]. Therefore, the use of NMR spectroscopy is a state-of-the-art choice to study their structure, dynamics and interactions.

3.1.3 Interaction with the mRNA targets

miRNAs are associated with mRNAs metabolism[126][127]. The binding of those regulatory miRNAs by their so called *seed* region, constituted of n°2 to 7 nucleotides (n°2 to 8 for the *extended seed*), to the *canonical* site of the mRNA affects its metabolism. miRNAs can have as many as hundreds of targets, to which they can bind either completely or with sparse complementarity, their seed region usually binding perfectly in both cases[128]. This miRNA:mRNA interaction leads to protein synthesis repression or degradation of the mRNA[129][130][131][92].

MicroRNAs are organised in families, mostly divided according to their seed region, that target several mRNAs. One of the first families discovered, which metabolism has been one of the most studied, is the *let-7* family[132]. Indeed, many targets have been identified for the miRNA *let-7*. Three of them were selected in our project: LIN28A, LIN41 and HMGA2.

LIN28A is a RNA binding protein (RBP) encoding for 209 amino acid long protein[133]. LIN28 is implicated in biological processes such as development and glucose metabolism[134]. LIN28/*let-7* pathway is diverse, since LIN28 binds to the terminal loop of pre-*let-7*, and *let-7* is also affecting LIN28[135]. Indeed, LIN28A destabilises the pre-*let-7* by interacting with its loop, thus leading to its degradation in the cytoplasm. It can also keep *let-7* in

the nuclei, therefore preventing the mRNA repression process. LIN28A is implied in many cancer forms, due to its inhibition of *let-7* pathway, and *a fortiori* leading to activation of the tumorigenesis processes.

LIN41, also called TRIM71, is a Trim-NHL protein: TRIpartite Motif (TRIM) meaning it contains three main domains, whereas *NHL* describes the motif located at the C-terminus named after NCL-1/HTA2/LIN41, where it was first observed[136]. Those are involved in the regulation of development processes, such as cell differentiation, and inhibition of cell growth and proliferation.

LIN41 and *let-7* are known to be correlated since the observation of the down-regulation of LIN41 in mouses where *let-7* is up-regulated[137]. Indeed, *let-7*:LIN41 complexes induces the suppression of the expression of the mRNA[116][138].

The High Mobility Group (HMG) proteins are components of chromatin, influencing the transcription of several genes through the interaction with sequences furnished in AT bases. The HMG AT-Hook2 (HMGA2) protein is 108 amino acid-long, and is involved in gene transcription through its impact on chromatin architecture. Indeed, a positively charged part of HMGA2 sequence can bind to oligonucleotides, to get them into an ordered conformation. HMGA2 is involved in various biological processes, such as DNA repairing, apoptosis, and senescence. The over-expression of the mRNA can lead to apparition of malignant tumours, especially lung cancer[117][139]. *let-7* binds to HMGA2 3'UTR, allowing to repress the oncogene, and therefore limiting its tumorigenesis contribution[140].

Some of *let-7* pathways, including the targets described above, are illustrated on Figure 3.2.

To study the interaction between *let-7* and these three targets by NMR spectroscopy, all of the RNAs have to be priorly synthesised. The method chosen is solid-phase chemical synthesis, as it is particularly robust for small RNAs and allow to insert probes at specific positions. However, while the miRNA is to be synthesize entirely, the mRNAs are narrowed to the 3'-UnTranslated Region (3'-UTR) with approximately the same size as the miRNA. Indeed, larger templates would complicate both the synthesis and the NMR analysis, and have not been identified as required for the miRNA:mRNA interaction.

3.2 Starting protocol

3.2.1 Methods: Synthesis

All materials used during the work presented in this section were of synthesis/analytical grade or equivalent if available. All reagents and solvents were supplied by Fisher and

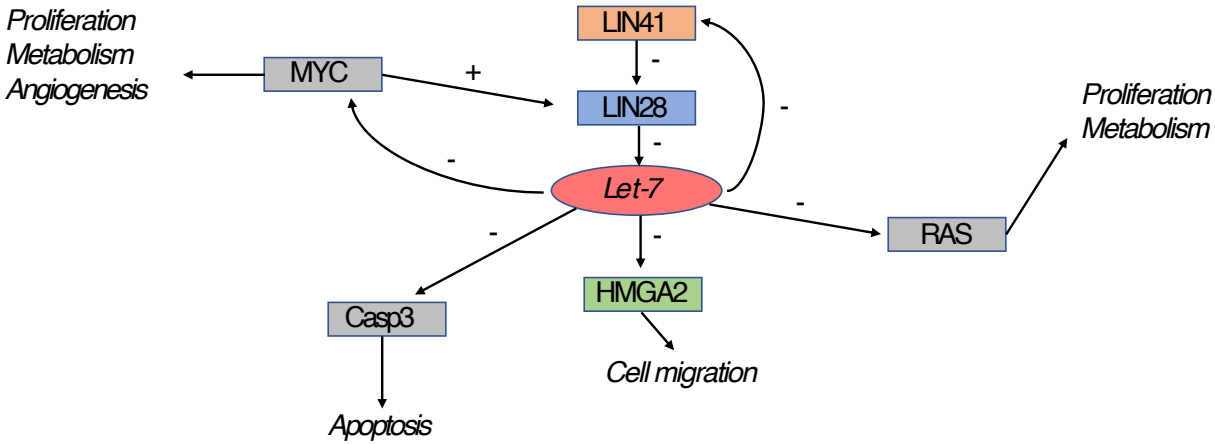

Figure 3.2: Illustration of gene regulation between the miRNA *let-7* and several of its targets. The one of interest in this project are LIN28, LIN41 and HMGA2. RAS, MYC and Casp3 are here to emphasise the diversity of interactions. The + sign denotes an up-regulation, while a - sign designates a down-regulation.

Carbosynth. All reagents were used as provided with no further purification.

Prior to mixing reagents, bottles are cleaned with acetone, dried at 120°C and flushed with argon to get rid of any residual water that could greatly impact the synthesis. Most of the solvents are usually prepared a few days prior to synthesis to minimise the solvent air content due to manipulation. The activator however is mixed right before the start of the synthesis to maximise its efficiency. Details of the solvents preparation can be found in Table 3.1.

Phosphoramidites are dissolved in ACN (99.9%, extra dry) to reach a 0.1 M concentration. Molecular sieve 3Å is added to the phosphoramidites and to the activator once the BTT is dissolved, to ensure their dryness along the synthesis.

Once the solvents and the phosphoramidites are loaded onto the synthesizer, prior to the synthesis, the whole synthesizer is purged, using both automated and manual ways. The tubings have to be filled with the solvents at the adequate concentration to be prepared for the first cycle of the synthesis.

Presence of bubbles in the system is constantly observed at first glance, but successfully removed by the purges. The flow-rate is checked using an automated method and a graduated cylinder: individually for each pump (A and B; 5 min at 10 mL/min each) and

Solvents	CAP A	CAP B	Oxidiser	Detritylation	Activator
N-methyl imidazole	40 mL				
Acetic anhydride		40 mL			
Lutidine		60 mL			
Iodine			2.54 g		
Pyridine			180 mL		
H_2O			20 mL		
DCA				30 mL	
Toluene				970 mL	
BTT					11.55 g
ACN	160 mL	100 mL			200 mL

Table 3.1: Amounts of reagents for each synthesis solvent.

jointly (3 min both at 10 mL/min).

The solid support has to be placed into a stainless steel column (1.2 mL), to be loaded onto the synthesizer. The incorporated mass of support m_{supp} (33-85 mg) is calculated using the scale of the oligonucleotide synthesis s_{synth} (here, 10-25 μmol), and the loading of the support l_{supp} (around 300 μmol per g), using the following equation:

$$m_{supp} = \frac{s_{synth}}{l_{supp}}$$

The support used depends on the oligonucleotide to be synthesized: for example, *let-7* possesses a uracil on its 3'-end, therefore the column is to be filled with Primer Support 5G riboU, that is adequate for oligonucleotide constituted of up to 70 bases starting with a uracil, and for a wide range of scales.

A few parameters can be adjusted within the synthesiser software, *UNICORN*. During the coupling step, a recycling time can be specified: the coupling mixture (4 equivalences of phosphoramidites) is cycled through the column for an amount of time defined by the user, before being redirected to the waste (here, 12 min). Main parameters used for the synthesis can be found in Table 3.2.

As shown on Table 3.2, a series of three synthesis is planned: each synthesis involves a different oligonucleotide and is launched one day after the other. Post-synthesis treatments are started right after completion of each oligonucleotide synthesis.

Synthesis	1. Model RNA	2.*let-7*	3. HMGA2
Sequence (5' to 3')	GCG AGC UUC GGC UCG C	UGA GGU AGU AGG UUG UAU AGU U	GCC AAC GUU CGA UUU CUA CCU CA
Length	16-mer	22-mer	23-mer
Scale / μmol	25	10	10
Column volume / mL	1.2	1.2	1.2
Amidites concentration RNA / M	0.1	0.1	0.1
Amidites equivalences RNA	4	4	4
Recycling time / min	12	12	12

Table 3.2: Main parameters used for the laboratory first RNA synthesis.

3.2.2 Methods: Post-synthesis treatment

Post-synthesis treatments are divided into a series of steps, namely:

1. Dry the support to which the synthesized oligonucleotide is attached;
Once the column has been removed from the synthesizer, its content is dried using a vacuum pump connected to the metallic column. The synthesised oligonucleotide, attached to the solid support, is then transferred into a Wheaton tube (required to handle increasing pressure).

2. Cleave the oligonucleotide from the support, while unprotecting bases and phosphate groups;
Methylamine (33% solution in absolute ethanol) is mixed with ammonium hydroxide in equal ratio (depending on the synthesis scale, 2-3 mL each for 10-15 μmol synthesis), and poured into the tube containing the support-attached oligonucleotide. The solution is placed into the oven at 60°C for 2h, then let under the fume hood to go back to Room Temperature (RT).

3. Separate the chemicals and the solid support from the oligonucleotide;
A sinter filter (porosity N°3, previously washed with H_2O:absolute ethanol 1:1), connected to a vacuum pump, is used to separate the support from the oligonucleotide in the liquid phase. The tube's content is therefore poured onto the filter and the recovered liquid phase is transferred to a 100 mL flask. The remaining of the solid phase on the filter is washed off with a solution of H_2O:absolute ethanol 1:1. Each washing step is followed by the transfer of the liquid phase to the previously used flask.

Since ammonium hydroxide is a corrosive chemical, the equipment can be damaged through contact. To extend the equipment lifetime, the evaporation is therefore firstly

performed by letting the flask under the fume hood overnight, topped with a perforated aluminum foil.

The solution is then placed on a rotary evaporator at RT. A back-flow preventer is added to minimise the bumping of the solution, meaning to prevent the solution within the flask from being drawn into the condensing part of the evaporator. However, even while gently reducing pressure, foam was rising up until the back-flow preventer. The vacuum therefore had to be stopped several times, to recover the liquid phase.

After a few hours, the foam started to decrease. Pressure could then be slowly reduced to about 500 mbar. The remaining solution became cloudy, and the pressure was gradually reduced to 50 mbar. A slightly yellow film, well distributed on the flask, was obtained. The precipitate is removed from the evaporator and topped, to be stored at -20 ° C.

As the first oligonucleotide evaporation (model RNA) showed troubles in foam while evaporating, one of the two remaining templates (*let-7*) was instead separated in eppendorf tubes (2 mL), to be evaporated using a speed vacuum system at RT. Evaporation was controlled at several timings; in the end, total evaporation lasted for a few hours, and slightly yellow crystalline pellets were obtained.

For the last oligonucleotide (HMGA2), the flask was again placed on a rotary evaporator, since lesser amounts of transfers are required, thus reducing risks of contamination and loss. The idea was then to be even more gentle while decreasing the pressure: it was set to 750 mbar for 2 hours to evaporate the volatile gases as the first step, since much less foam is formed at this pressure, followed by another step at 500 mbar. The pressure could then be set to 50 mbar so that the rest of the liquid phase could be evaporated, and a slightly yellow film was formed on the flask.

4. Re-suspend the oligonucleotide;

After evaporation, the flask contains solid oligonucleotides, unprotected on the bases and phosphate groups.

Anhydrous DMSO (0.5 mL) is added to dissolve the precipitate. After transferring the resulting solution into a Falcon tube, the step is repeated twice (0.5 mL then 0.25 mL) to improve the collection of the remaining RNA.

5. Unprotect the 2'-OH;

Triethylamine trihydrofluoride 97% (TREAT, 10 mL) is added to the solution. The tube is placed into the oven at 60°C for 3h30, then let under the fume hood to go back to RT.

6. Remove the used chemicals by precipitation of the oligonucleotide;

The solution is placed at -20°C along with butanol for a few hours, and taken out on ice. Butanol (4 x 5 mL) is added to the solution containing TREAT and the oligonucleotide, at that moment also unprotected on the ribose 2'-hydroxyl group. A white precipitate appears. The solution is centrifuged (5 min, 4°C, 3.0 rcf). A larger white pellet is observed, with a yellow supernatant.

The supernatant is removed from the tube and centrifuged again, making sure to prevent any loss (5 min, 3.0 rcf, 4°C). In the case of the model RNA, a small pellet appeared, which was collected and added to the first one. No noticeable pellet was formed in the other oligonucleotides supernatant second centrifugation.

To wash the pellet, cold butanol (2 x 10 mL) is added. It is necessary to use a vortex to re-disperse the pellet entirely. Following the same methodology, the solution is centrifuged (5 min, 4°C, 3.0 rcf), as well as the obtained supernatant.

Each oligonucleotide pellet is dissolved in triethylammonium acetate (TEAA, 10-20 mL, 1 M) and stored at -20°C.

7. Characterise and purify the oligonucleotide.
To characterise the oligonucleotide, both UV quantification for concentration calculation, and analytical HPLC for purity assessment, are performed.
To assess the accuracy of the concentration determination of the desired oligonucleotide, the purity of the solution has to be characterised, since any aborted oligonucleotide formed during the synthesis could distort the UV analysis.

Reverse-phase analytical HPLC is used. However, to get an approximation of the injection volume, the concentration of the crude oligonucleotide is still measured by UV using the spectrophotometer prior to purification. The HPLC buffers consists in: buffer A) TEAA in H_2O (5 mM) and buffer B) TEAA in ACN (5 mM). The first global gradient is performed from 2% buffer B to 80% buffer B. Refined gradients are performed from 2% buffer B to 15% buffer B.

3.2.3 Results and discussions

Since the scale of the first oligonucleotide was higher (model RNA, 1.5 times more than the other two), higher amounts of methylamine and ammonium hydroxide have to be used for cleavage/de-protection, which could have promoted foam formation during the drying step. This can be countered by letting the ammonium hydroxide evaporate in the fume hood for a longer period, and/or use a larger flask to load on the rotary evaporator.

Considering the synthesis scale, the extinction coefficient for each oligonucleotide, and the absorbance obtained at 260 nm, obtained yields for the first round of synthesis are within the range of 19 % - 38 % (before purification). This is not a sufficient yield regarding the chemical synthesis achievements, but as this is the laboratory first RNA synthesis, it is promising.

However, the yield value is senseless if the sample majorly contains aborted oligonucleotides. To assess the purity of the synthesis, HPLC global and refined gradients have been performed for each synthesised oligonucleotide (see Figure 3.3).

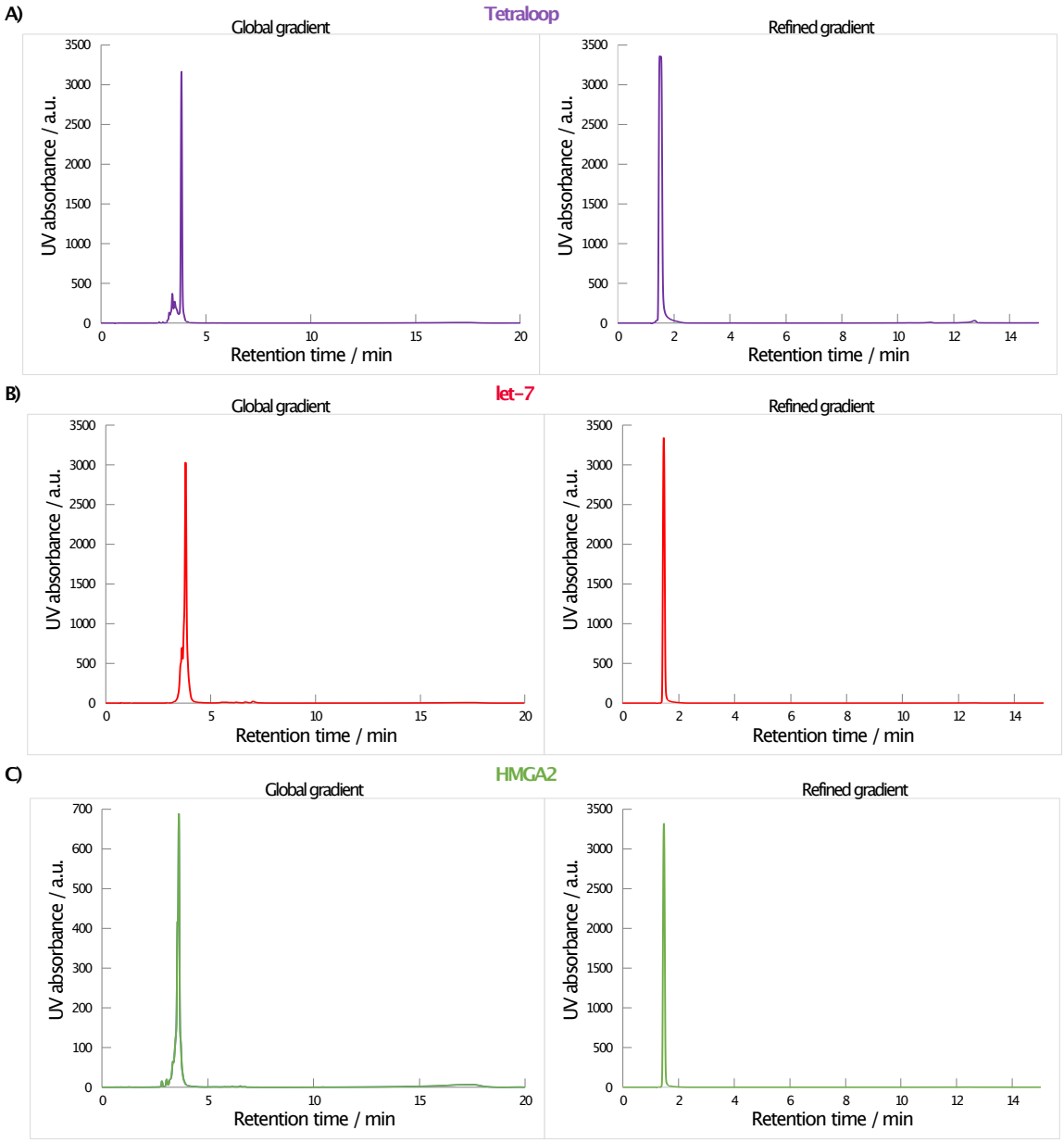

Figure 3.3: HPLC analysis of the first series of chemically synthesized RNAs: A) model RNA, B) *let-7* and C) HMGA2 target, with for each (left panel) global gradient from 2 to 80% buffer B in 20 min and (right panel) refined gradient from 2 to 15% buffer B in 15 min.

Throughout repetitions of the same gradient, each injection induced the same conclusion: the elution time of the oligonucleotide was shortening, until everything contained in the sample eluted in the dead volume. The analytes were not or almost not retained by the static phase in the column.

Considering the fragility of RNAs, it became clear that all of our synthesised samples had undergone degradation. The most probable hypothesis was because of the presence of RiboNuclease (RNase): the oligonucleotides had dissociated into nucleotides, thus explaining why they showed no retention. To make sure the sample was not contaminated from the HPLC needle, other aliquots were tested, and the HPLC needle was cleaned repeatedly with isopropanol and RNAse free water, but the same results were obtained (see Figure 3.3, right panel).

The presence of RNase was tested using a commercial fluorescent RNase alert test kit. The buffer kit has to be in contact with the solution to be tested for RNase for 1h at 37°C. The solution can then be observed: if a fluorescence is observable, the sample contains RNase. Indeed, the commercial buffer contains oligonucleotides with fluorescent markers which react if they are separated from the rest of the template, meaning the oligonucleotide has been cleaved by RNase, attesting of their presence. However, the kit is only usable in certain types of chemical, and there may be false positive while used in uncertified environment.

Here, the kit buffer (5 μL) was added to sample solution (45 μL, each solution containing a chemically synthesised oligonucleotide), and let 1h at 37°C before comparing to positive (containing RNase) and negative (no RNase) controls, as observable in Figure 3.4.

Some of the solutions became fluorescent right upon addition of the buffer. All of the solutions containing the synthesized oligonucleotides exhibited fluorescence. TEAA in RNase free water was tested, in order to eliminate the solution from the contamination possibilities, and it indeed did not react.

The most intense fluorescent sample (*let-7*) had been treated differently from the other ones, as it was dried only using speed vacuum system. This drying step may have brought additional RNase from manipulation and open eppendorfs. However, considering all of the samples were contaminated, the whole post-synthesis treatment has to be investigated in order to remove the sources of contamination, which would otherwise always degrade the synthesized RNAs once 2'-unprotected.

The used chemicals were also tested (methylamine and ammonium hydroxide) and reacted when in contact with kit buffer. However, it is uncertain if this is due to RNase contaminants or if the kit is not functioning properly in this kind of reagent. The latest hypothesis was the most probable so focus was put on modifying other variables in our protocol, and check the results while performing further synthesis.

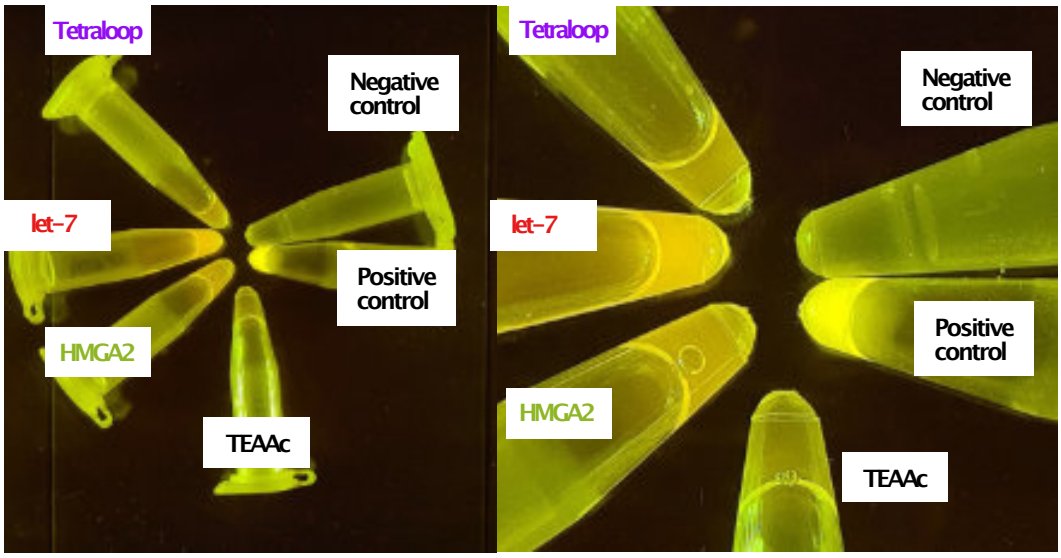

Figure 3.4: RNase detection in chemically synthesised oligonucleotides samples: the more fluorescent the sample, the more contaminated by RNase.

3.3 *let-7* targets conclusive synthesis

3.3.1 Methods: getting a non-degraded sample

The first step to improve the protocol is getting a non-degraded sample, meaning the samples have to be RNase free.

As RNase is a recurrent issue when handling RNA, the protocol has to be adapted to ensure there is no contamination, especially after the 2' deprotection by TREAT. All the glassware used for the synthesis would be priorly put into a bath of water mixed with *RNase away*, an acidic solution degrading the RNase. Every step involving reusable material that could be accomplished using RNase free dispensable equipment instead is to be switched. The remaining glassware is let in a bath of water containing *RNase away* for a few hours, then washed off with RNase free water, and dried at 120°C overnight prior to their use.

The main suspected source of RNase contamination is the sinter filter, as it is hard to clean and making sure it is RNase free, due to its porosity. The filtering method is therefore switched to centrifugation: the solid support is condensed into a pellet while the oligonucleotide remains in the liquid phase, which can then be removed using dispensable pipettes.

During the HPLC analysis of the subsequent oligonucleotide synthesis, the successive runs exhibited variations in retention times of the same oligonucleotides. This rose a

concern, as ongoing degradation could be a probable issue. However, the samples were stable for a few days in solution, thus limiting the possibility of the presence of RNAse. Degradation was not a plausible reason anymore.

As the elution of the analyte depends on the polarity of both the mobile phase and the analyte, the change in polarity of the different injections was considered. Indeed, if some butanol remained with the pellets before the last re-dissolutions, it could greatly affect the retention on the column.

To ensure the butanol removal, the oligonucleotide sample was filtered on an Amicon membrane and washed off with water containing a small percentage of TEAA.

Buffer B was also switched from TEAA in ACN (5 mM) to TEAA in H_2O:ACN 1:1 (5 mM). The analytes separation would thus be improved by softening the polarity variation between buffers.

3.3.2 Methods: increasing the yield

The second step to improve the protocol is increasing the global yield. Several steps where loss of oligonucleotides may have occurred have been identified and switched to more adapted protocols.

A recurrent reason of a decrease in the yield is the incorporation of the rG(iBu) phosphoramidite within the oligonucleotide during the chemical synthesis. This can be overcome by increasing the recycling time of the coupling step for this particular base.

One of the reasons of the material loss is coming directly from the first post-synthesis step: the drying of the support and transfer from the column to a Wheaton container. Indeed, some of the support remained in the metallic column and thus provoke a loss in material. This aspect can be improved, to lose as less as possible, however it can not be avoided.

In order to remove more of the solid support contained into the column, its dryness has to be insured. To improve the dryness of the support connected to the oligonucleotide, absolute ethanol was used in addition to submit the sample to a vacuum pump, which leads to much drier support (easily observable through the colour changes and the easiness of removal).

Different rounds of synthesis have been performed using the centrifugation method after the cleavage/de-protection step. This centrifugation is followed by the transfer of the liquid phase to a flask to evaporate at the rotary evaporator, followed by dissolution in DMSO and transfer to Falcon tube.

It was observed that the transfer from the flask (drying step) to the Falcon tube (desilylation) was a crucial part of the post-synthesis treatment yield drop. Indeed, the size of

the flask compared to the volume of DMSO used to redissolve was not sufficient to ensure the dissolution of the whole material: particles were always observed on the flask. However, keeping on adding DMSO was not an option, as most of the oligonucleotide would then remain in solution instead of precipitating using butanol.

To get rid of this step, the evaporation was performed using a speed vacuum system adapted for Falcon tubes. The drying step followed by the dissolution in DMSO could therefore be performed without additional transfer, thus limiting the losses.

Trying to improve the precipitation after desilylation, cold butanol precipitation was compared to cold ethanol precipitation. However, the butanol one gave better results and was then kept in every other post-synthesis treatment. To further improve the precipitation yield, both the butanol and the sample were placed back at -20°C at each precipitation attempt.

During one of our synthesis, an extraction was performed on the remaining supernatant after the 3 usual rounds of butanol precipitation performed. A significant amount of oligonucleotide was collected, and added to the final yield, significantly increasing it. The detailed can be found in the following ^{19}F *let-7 synthesis* section.

3.3.3 Further issues

Indubitably, another way to improve the yield is to start with a cleaner oligonucleotide synthesis: if the coupling yield is higher, less aborted templates will be present in the final solution before purification, and the desired oligonucleotide will be produced in higher yield.

However, at the time of writing, this issue has been thoroughly investigated and solutions have not been found yet. Indeed, throughout the use of the synthesizer, several issues were faced, to which either solutions have been encountered or some remain problematic. Here is a list of the main issues and examples of when they occurred:

- Evaporation of phosphoramidites;

Evaporation is a huge issue, given that the system is under argon and should not be permissible to air. Indeed, even a small leak can generate a variation in the pressure, leading to drastic effects on the phosphoramidites and solvents used.

During synthesis and pauses in between them, several cases of evaporation in the phosphoramidites bottles connected to the synthesizer were observed, leading either to the complete evaporation/use of a phosphoramidite bottle content, or to the formation of a viscous material at its bottom. This is a problem to be solved immediately, since any leak there would provoke evaporation/pressure variations, leading to variations in the phosphoramidites concentrations and volumes, as well as a regarding the accuracy of the flow-rate

of the device. Indeed, phosphoramidite concentrations and volumes are designated carefully into the synthesizer software, meaning any change would impact the synthesis and its yield. Moreover, due to the instability of the phosphoramidites, it is important to keep them under an inert atmosphere, and a leak would mean that air (containing traces of water) might get in contact with the phosphoramidites, rending them less reactive. Additionally, due to the price of the phosphoramidites, RNA synthesis are always calculated to use as much quantity as possible, thus reaching the volume limit usable within the bottle which, coupled with evaporation, would lead to empty phosphoramidite bottles for the latest couplings, and therefore an incomplete oligonucleotide.

The problem was addressed by removing the equipment keeping the phosphoramidite bottles in place, cleaning the seals and exchanging phosphoramidite bottles. The evaporation of the phosphoramidite bottles can be investigated replacing them with acetonitrile bottles (the storage solvent for the synthesizer) and observe the evolution of the liquid level in the bottles. Further tests using different bottle sizes were made, to reach a point where no evaporation was observed in the phosphoramidites bottles anymore.

An additional parameter to be careful with is the molecular sieve. Indeed, one should be careful as to check that the molecular sieve is in proper conditions, and does no absorb part of the solution, additionally to possible water molecules. In this case, the molecular sieve (cut-off: 3Å) is disposed off, and a new bottle of phosphoramidites is needed.

- Evaporation/Crystallisation of the solvents;

As evaporation was encountered in the phosphoramidite part of the synthesizer, it was also observed in the solvent bottles. As well as problems already mentioned above, this raises a problem regarding the activator. Indeed, as all of the concentrations could be affected, the activator is also subjected to cristallise, in the solvent bottle as well as in the tubing lines.

Purging of the synthesizer using acetonitrile is sufficient to get rid of any visible crystals. Airtightness was improved by changing the seals and adding Parafilm. As the activator is close to the solubility limit, it was placed into a bead bath kept at 23°C to ensure that the temperature was not dropping during nights.

- Pressure and flow-rate issues;

Variations in pressure and flow-rates were observed, on the manometer and performing the routine flow-rates check-ups. This could affect the proper delivery of both phosphoramidites and solvents by modifying the quantities, and negatively impact the synthesis.

As the tubing started to get twisted due to bottle changes, they were modified to ensure proper circulation. Concerning flow-rates, the delivery tests usually lead to expected results but the presence of bubbles affects them on regular basis and purges should be performed. Indeed, the pump butterflies and some part of the tubing are filled with bubbles quickly, which creates a need to purge almost everyday prior to the synthesis. This may alter the flow rates at any time during the synthesis.

Since solvents and phosphoramidites levels dropping (without pumping) were observed even while using the right sized bottles, a pressure issue/an argon flow abnormally high was suspected. A pressure variation would be a problem for a regular flow-rate and the evaporation of the solvents. This is still investigated at the time of writing.

- Synthesis detritylation step.

The most recurrent issue faced is the detritylation UV peak area dropping after only several bases incorporated during the oligonucleotide synthesis. This is not an accurate quantitative analysis tool regarding the synthesis yield, especially considering the detector is often saturated at the beginning of the synthesis, however large drops in the detritylation area are visible when phosphoramidites are not well incorporated into the chain being synthesised, therefore leading to an aborted synthesis due to very low coupling yields.

Many factors could influence the relative detritylation dropping, including the ones quoted above, therefore one issue could not be isolated and taken responsible for it, amongst the previous ones listed. Increasing the equivalences number used during synthesis allowed to perform some synthesis with poor yield but enough to be able to prepare NMR samples. In the meantime, a maintenance has been required in order to help us investigate on all of these issues.

3.3.4 Results

The synthesis parameters are displayed in Table 5.2.

Synthesis	1. LIN28A	2.LIN41	3. HMGA2
Sequence (5' to 3')	GAG UGC ACA GCC AUA UUG AAC UAC CUC A	UUU AUA CAA CCA UUC UGC CUC U	GCC AAC GUU CGA UUU CUA CCU CA
Length	28-mer	22-mer	23-mer
Scale / μmol	10	10	10
Column volume / mL	1.2	1.2	1.2
Amidites concentration RNA / M	0.1	0.1	0.1
Amidites equivalences RNA	4	4	4
Recycling time / min	12	12	12

Table 3.3: Main parameters used for the targets synthesis.

Overall, mRNA targets were successfully synthesized and purified by HPLC (see Figure 3.5), as the quantities were enough to be used in NMR spectroscopy, and the desired oligonucleotide could be separated from the aborted templates.

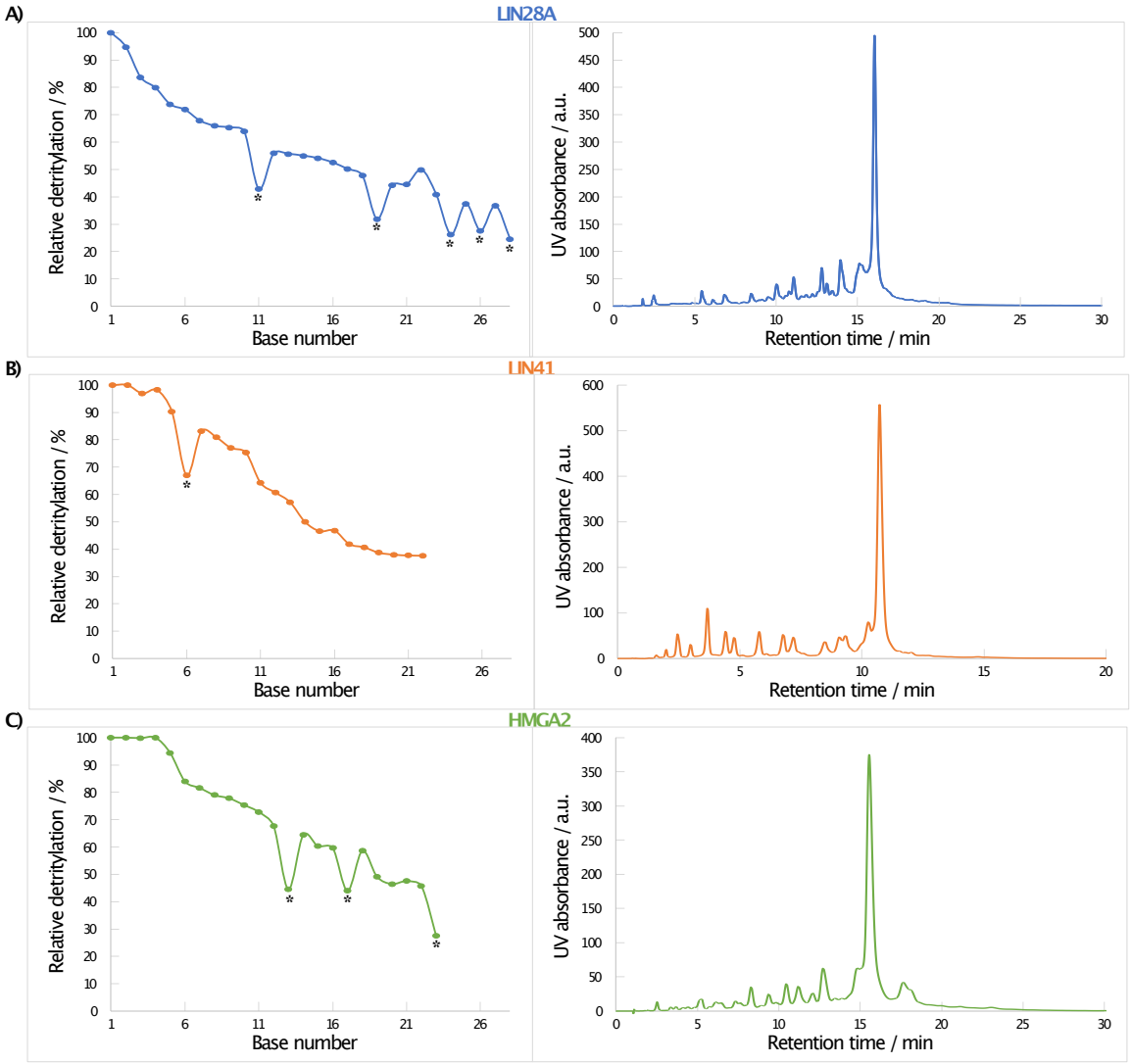

Figure 3.5: (Left panel) Detritylation peak area (in %) comparison between successive phosphoramidite incorporation of targets synthesis. The asterisks* represent the G-bases, known to be harder to incorporate.
(Right panel) HPLC characterisation of targets, gradient going from: (LIN28A) 10 to 25% buffer B (30 min, 1 mL/min); (LIN41) 12 to 22% buffer B (20 min 1 mL/min); (HMGA2) 12 to 22% buffer B (30 min, 1 mL/min).

However, improvements are still required, especially because of the poor final yield and poor coupling yields during synthesis. Indeed, the relative detritylation is decreasing faster

than in the previous synthesis performed within the laboratory, and to what is expectable regarding the literature (see Figure 3.6 and Table 3.4).

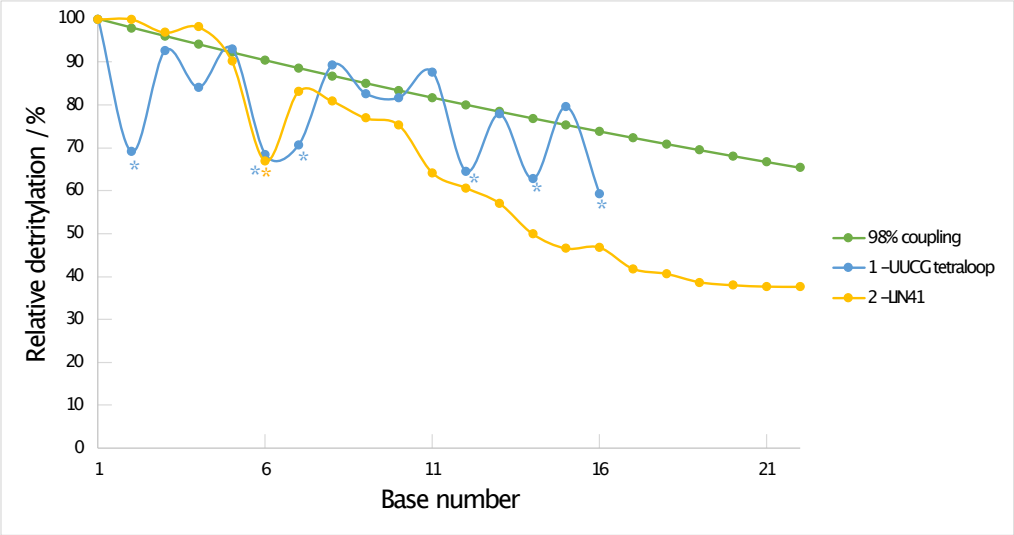

Figure 3.6: Detritylation peak area (in %) comparison between successive phosphoramidite incorporation during several synthesis: (green) theoretical 98% coupling efficiency; (blue) model RNA (yellow) LIN41. The asterisks* represent the G-bases, known to be harder to incorporate.

RNA	Sequence length	Relative detritylation at 15th base incorporation / %	Relative final detritylation / %	Relative final detritylation that is not a rG / %
Model RNA	16	80	59	80
LIN28A	28	54	24	37
LIN41	22	47	38	38
HMGA2	23	60	28	46

Table 3.4: Comparison of the synthesis yield through relative detritylation observation.

Additionally to the target synthesis, *let-7* synthesis were performed afterwards, that kept decreasing the relative detritylation: the penultimate one reached 13% for 22 nucleotides, while the ultimate one had to be stopped at the third base because the detritylation dropped to as little as 3%.

As the solvents and phosphoramidites used are still ordered from the same provider,

and it did not seem to only affect one batch, the problem should come from the manipulation while loading the synthesizer or within the synthesizer itself.

3.4 ^{19}F *let-7* synthesis

As mentioned previously, fluorine probes present various advantages for NMR spectroscopy investigations. The *let-7* synthesis plan is to produce several samples, with a fluorine nucleus incorporated in the uracil ribose 2' position (U-base 2'), located at different places of the RNA sequence. These samples should permit to assess the interaction with the various targets at specific positions, in various structural elements, without going through assignment process and avoiding resonance overlap.

3.4.1 Methods: Protocol adjustments

A few parameters were changed to improve the yield observed in previous synthesis, and to adapt to the fluorine nucleus.

The recycling time for the fluorinated phosphoramidite is reduced to 3 min.

Methylamine (33wt % solution in absolute ethanol, 4mL) was mixed with ammonium hydroxide in 1:1 ratio, and incorporated with the oligonucleotide attached to the solid support. The amount of reagents has been increased relative to the scale, compared to previous synthesis, to properly cleave and unprotect the oligonucleotide. The solution was let under the fume hood for 4h. Deprotection step was carried out at RT because of the 2'-F residue. Indeed, fluorinated oligonucleotides require weaker conditions, as the fluorine is more sensitive and can be degraded in harsher conditions. The centrifugation step if then performed and the liquid phase containing the oligonucleotide is removed, separated from the solid support.

To ensure the full cleavage of the oligonucleotide from the support, this step is repeated using a solution of ammonium hydroxide and absolute ethanol (2:1, 8 mL). The solution is poured into the Wheaton tube still containing the solid support and let under the fume hood for 16 hours at RT.

The rest of the post-synthesis treatment is then performed as previously done for the targets, and the contents of the repeated steps are gathered.

To check if the precipitation is sufficient, or if higher yields can be reached, an extraction is performed on the first precipitation supernatant. A solution of TEAA (1M) is poured on the supernatant. The falcon is stored at -26°C overnight. The next morning, two phases are observable. A second extraction, as well as one performed on the second

precipitation supernatant, did not lead to significant RNA collection.

3.4.2 Results

During the synthesis, the incorporation of the fluorinated phosphoramidite was really poor, which is probably due to the equivalences number (here, 4), that should be increased for further fluorinated oligonucleotides synthesis[141][142][143][66].

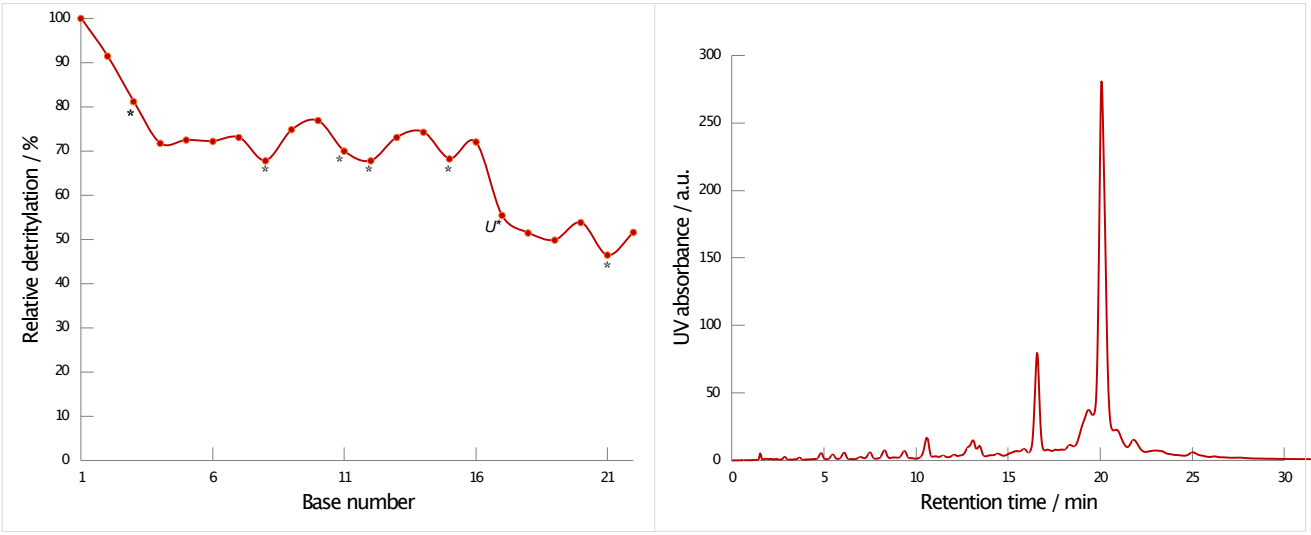

Figure 3.7: Left panel: Detritylation peak area (in %) comparison between successive phosphoramidite incorporation of *let-7* U6F synthesis. The asterisks* represent the G-bases, known to be harder to incorporate, while the *U** emphasizes the incorporation of the fluorinated phosphoramidite.
Right panel: HPLC characterisation of *let-7* U6F, gradient going from 12 to 20% buffer B (30 min, 1 mL/min).

Analysing the oligonucleotide solution using UV spectrophotometer, the yield before purification was estimated around 51%. The global scale of the synthesis was 16.8 μmol (calculated from weighted solid support); 7.65μmol were gathered from the precipitation (after filtration on amicon) and an additional 0.84 μmol was gathered after extraction.

This yield is the highest achieved over all the RNA synthesis performed in the laboratory. This increase is probably due to the improvement in the protocol throughout the series of synthesis performed, the more gentle and repeated cleavage step, and the

extraction performed on the liquid phase after butanol precipitation.

Surely, these amounts should be considered with care, as they are calculated before purification. However, the synthesis coupling yields seemed rather satisfying, except for the fluorinated phosphoramidite, which is a promising result towards better synthesis global yields (see Figure 3.7).

Melting temperatures were measured for *let-7* samples produced by chemical synthesis. They were determined for *let-7* itself, to check if the miRNA was folding onto itself, then in interaction with its chemically synthesized shortened mRNA targets, mixed in a 1:1 ratio.

The methodology used starts with a ramp going from 15 to 80°C in 5 min, then 3 ramps (going back and forth) from 80°C to 15°C by 1°C per minute (see Figure 3.8). The results were compared to melting temperatures previously obtained for biochemically synthesized samples (GG-*let-7*) and commercially available ones.

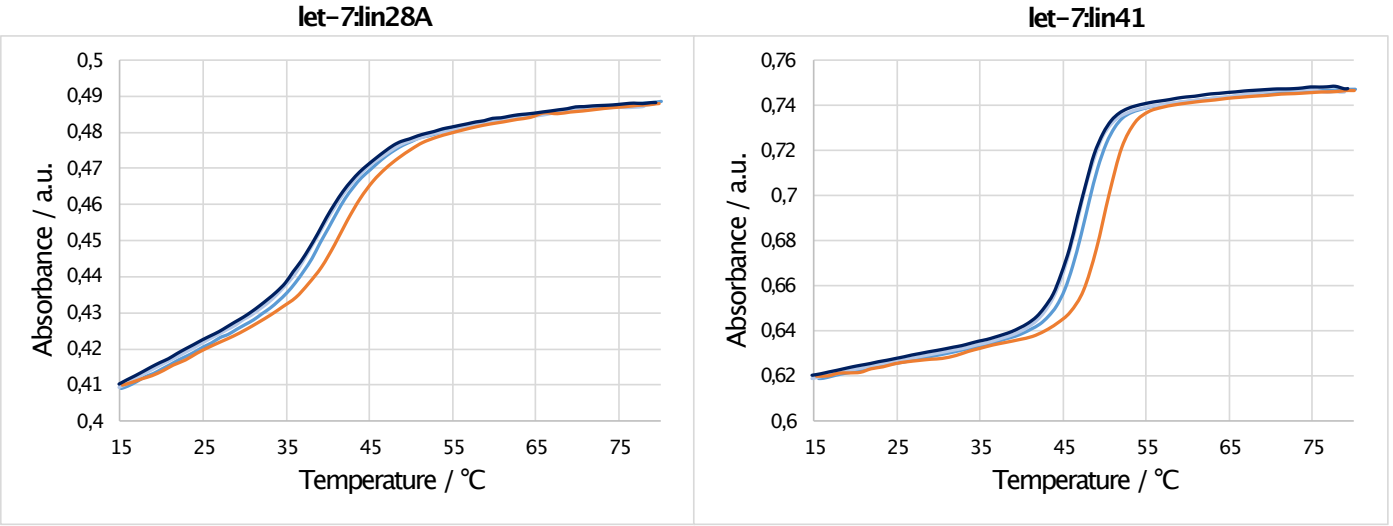

Figure 3.8: T_m determination for chemically synthesized *let-7* U6F, in interaction with chemically synthesised (left) LIN28A and (right) LIN41.
Several colours indicate the repeated ramps used for measurement: (orange) the fastest one, 5 min ramp 15°C to 80°C, (blue shades) from 80 to 15°C by 1°C/min back and forth.

As expected, commercially available and chemically synthesized *let-7* do not show any sign of melting temperature, as they do not fold onto themselves, which on the contrary may happen for the GG-*let-7* (T_m = 58.22 °C).

Along with the folding of biochemically synthesized GG-*let-7*, melting temperatures involving the interaction between either biochemically synthesized RNAs or commercially

available RNAs are clearly distinct, thus suggesting different pairings. This variation between samples comfort the decision to pursue with RNA chemical synthesis, as the GG addition definitely impacted the interactions between the miRNA and its targets. It is here important to notice that the melting temperature of the biochemically synthesized GG-*let-7* is identical to the one where GG-*let-7* is in interaction with a biochemically synthesized GG-LIN41. The interaction is therefore probably prohibited by the self-folding of GG-*let-7*.

Commercially available *let-7* interacting with targets (LIN41 / LIN28A) exhibit T_m close to the GG-*let-7* interacting with those targets. When comparing these values with chemically synthesized *let-7* labelled ^{19}F on the ribose connected to the U6, a small increase is observable, resulting in + 2°C / 2.2°C compared to commercial samples. This can be explained by the fact that the chemically synthesized *let-7* is labeled with a fluorine which, according to the literature, can raise the T_m by 1 or 2°C for each added fluorine[144][142]. Furthermore, the values obtained for both chemically synthesized *let-7* containing fluorine in interaction with each of the targets are comparable, as expected since the fluorine has only a shifted position in the miRNA structure. All the measurements are reported in Table 3.5.

Samples	*let-7* commercially available	*let-7* biochemically synthesized	*let-7* chemically synthesized (2'-F U6)	*let-7* chemically synthesized (2'-F U9)
No target	N.A.	58.22	N.A.	N.T.
+ commercial LIN41	45.09	46.1	N.T.	N.T.
+ biochemical LIN41	N.T.	58.22	N.T.	N.T.
+ chemical LIN41	N.T.	N.T.	47.05	46.94
+ commercial LIN28A	36.77	37.3	N.T.	N.T.
+ chemical LIN28A	N.T.	N.T.	39.01	39.57
+ chemical HMGA2	N.T.	N.T.	32.99	32.62

Table 3.5: Melting temperatures summary, for *let-7* and targets. N.A. denotes no melting temperature when performing the experiment, while N.T. stands for *Not Tested*.

3.5 NMR analysis

NMR data was recorded on samples of *let-7* and targets in 15 mM sodium phosphate (pH 6.5), 25 mM NaCl, 0.1 mM EDTA, 6-8% D2O. Concentrations were in the range 0.06-0.100 mM. More concentrated samples were also produced to reach a few hundreds μM.

3.5.1 Folding predictions

Predictions of the interaction between *let-7* and its targets were performed using Mfold software (see Figure 3.9)[145].

As expected, the seed region of *let-7* binds with perfect complementarity to its targets, whether it is LIN28A, LIN41 or HMGA2. The rest of the miRNA:mRNA pairing is incomplete, leading to bulge and loop formation.

These pairing predictions were used to elaborate a ^{19}F labelling plan for *let-7*. Indeed, 2'-^{19}F labelled rU phosphoramidites were incorporated in the seed region of *let-7* (U6) and a different structure component (U9): (LIN28A) base-paired near a loop; (LIN41) a loop ; (HMGA2) a bulge.

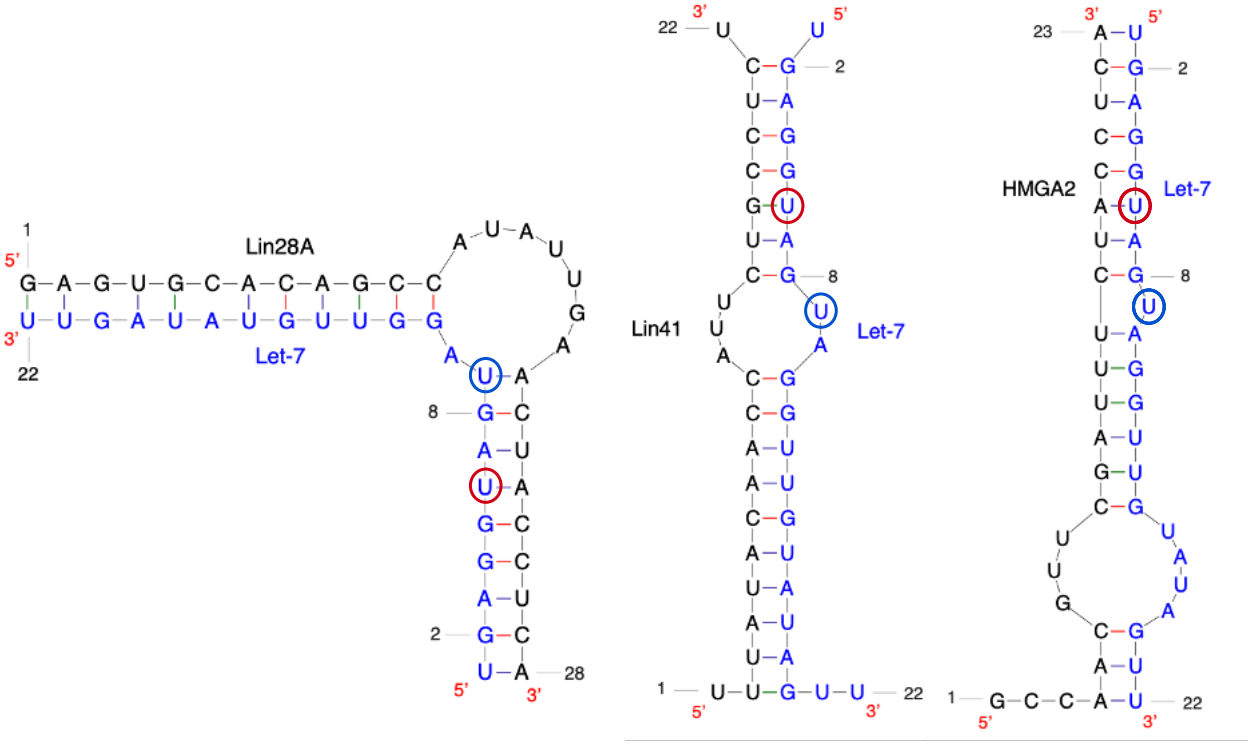

Figure 3.9: Predictions of the interaction between *let-7* and its targets (from left to right: LIN28A, LIN41 and HMGA2) using Mfold software. Encircled in red (blue) is the U6 (U9) position, in the *seed region*, which ribose is labelled in 2' with a fluorine.

3.5.2 Proton NMR

The first step was to analyse *let-7* on its own (see Figure 3.10).

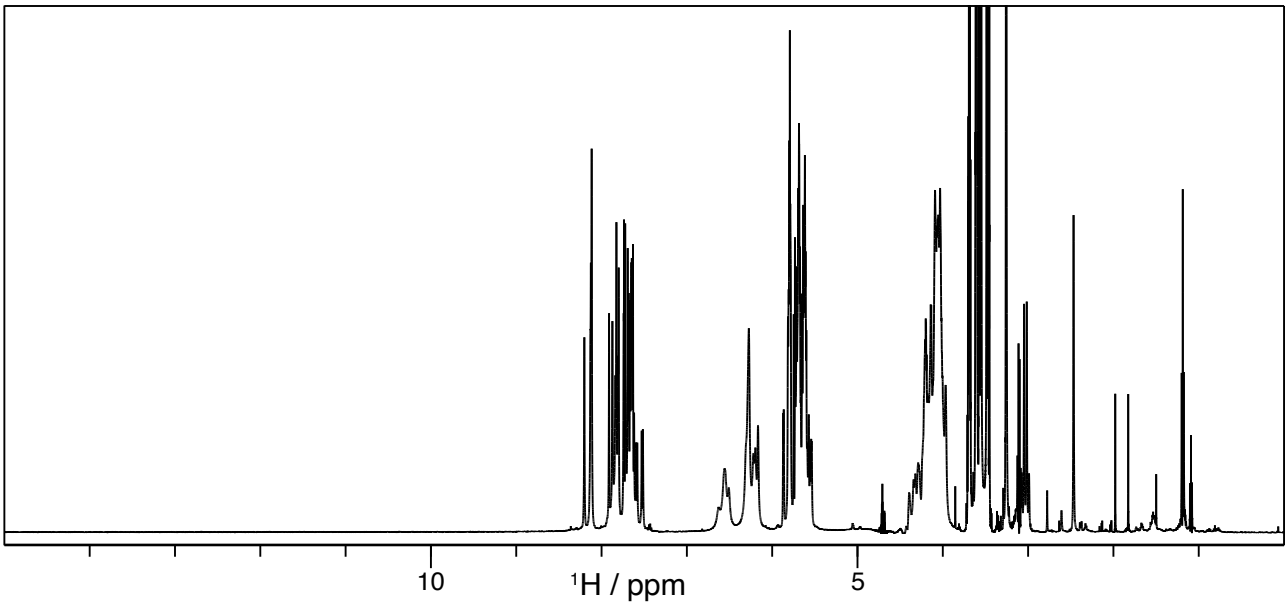

Figure 3.10: ^{1}H NMR spectrum of *let-7*-U6F (600 MHz, 298 K).

Some impurities appear in the more shielded part of the spectrum. These contaminants are easily identified, as they mainly consist of EDTA (present in NMR buffer) and some glycerol (recovering the amicon membrane). Further washing of the membrane should therefore be done before exchanging buffers, for the next samples.

Most importantly, no imino protons are observable, which means they are exchanging with the solvent, and therefore not trapped in a base-paired environment. NMR spectroscopy thus confirms the results obtained by T_m measurement: *let-7* does not fold onto itself.

let-7 is then divided into three samples to be mixed with LIN41, LIN28A and HMGA2. All samples are heated (85°C, 5 min) to trigger the pairing between the miRNA and its target. The same procedure is applied to mix the targets with both *let-7* labelled 2'-F on the U6 and on the U9 sugar. The comparison of the ^{1}H NMR spectra of *let-7* in interaction with its targets is observable on Figure 3.11.

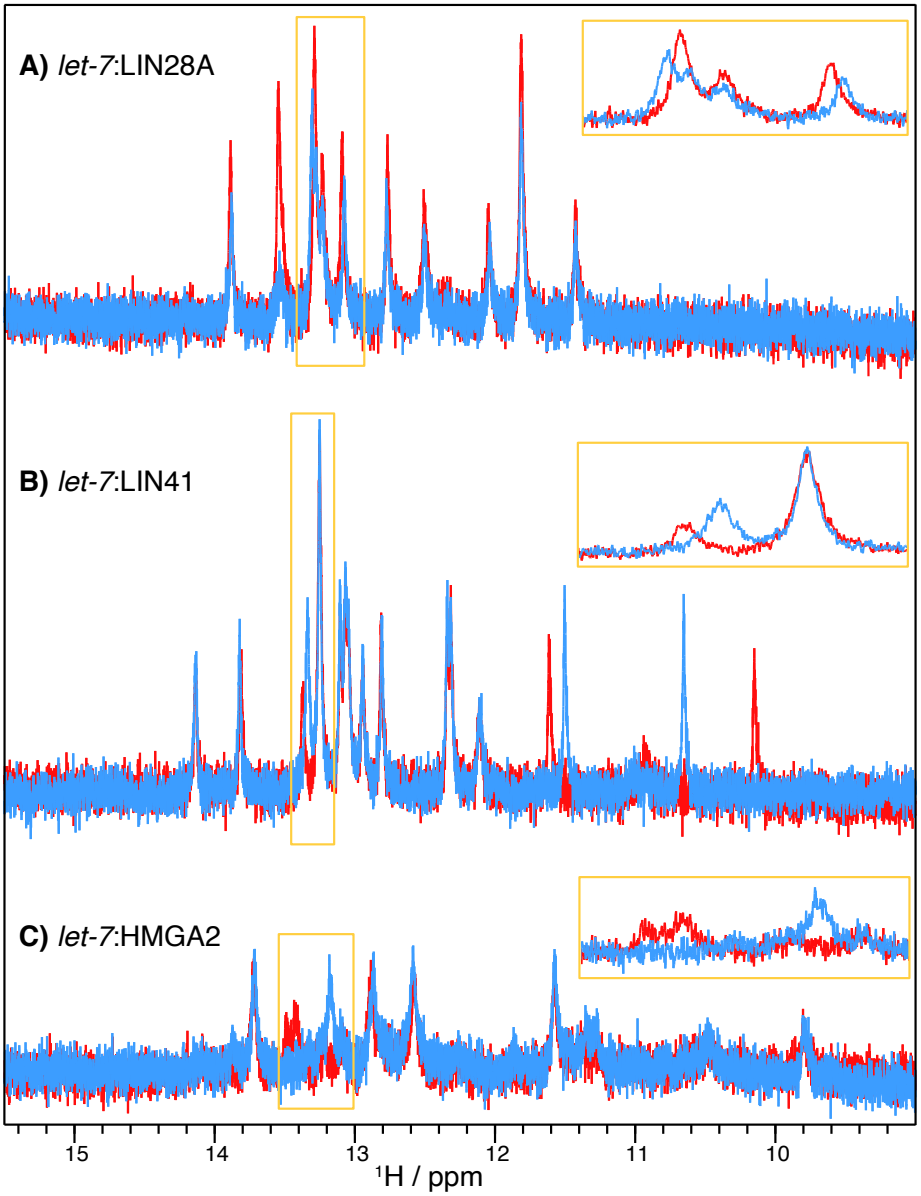

Figure 3.11: ^{1}H imino part of the NMR spectra of *let-7* interaction with targets (A) LIN28A, (B) LIN41 and (C) HMGA2 (600 MHz, 298 K for LIN28A and LIN41, 280K for HMGA2). In each case, red spectra highlight the interaction between targets and *let-7* labeled ^{19}F on 2'-U6, while light blues ones are samples containing ^{19}F 2'-U9 *let-7*.

Imino protons show different signals when *let-7* is paired with the targets, denoting their variable environment. Indeed, as expected considering sequences and the folding prediction, miRNA:mRNA interaction is different for each target. However, it seems that the pairing is either not the one expected from predictions with HMGA2, or only detectable at low temperatures.

It is still important to notice that most parts of the spectra are identical when different labeled *let-7* are put in interaction with the same targets, with however small chemical shifts changes, that should occur for nuclei spatially close to the sites containing fluorine (previous and next base-pairings).

3.5.3 Fluorine NMR

^{19}F is an interesting probe in NMR, as it has a Larmor frequency close but distinctive to the proton, has a 1/2 spin and is at almost 100% natural abundance[146].

Fluorine has already be proven very useful to identify structures and dynamics in proteins, when placed at specific positions within amino acids. It also has been incorporated in nucleic acids to help investigate their structures[147]. Indeed, as fluorine is naturally absent from these nucleic acids, its spectroscopic properties and differentiation from any other nuclei greatly ease the analysis of the experimental results.

Samples of *let-7*:LIN28A are labelled with fluorine at different positions: U6 in the *seed* region, supposedly in a base-paired environment and U9 located, according to the folding prediction, at the edge of a loop.

As observed by Li *et al.* for 2'-labelled RNAs, ^{19}F signal was observed around -200 ppm (see Figure 3.12). An additional signal was also observed around -120 ppm, corresponding to an impurity also observed in the literature[142].

Fluorinated samples showed a poor dispersion (less than 2 ppm), especially when comparing samples labelled on U9 position, that are all, according to the structure predictions, located in loops or bulge. The ^{19}F nuclei implicated in base-pairing in the *seed* region shows a higher chemical shift difference when in interaction with LIN41 compared to the other samples. This can be explained by the fact that the nucleotide U6 is there supposed to be implicated in a U-G base-pair while in a U-A base-pair when interacting with LIN28A and HMGA2.

As mentioned in a previous chapter, ^{19}F frequency reveals a strong Teflon signal rising over the chemical shift range of interest for the miRNA. While in the 1D, phasing has therefore been an issue, the issue has been overcome for the R1ρ experiments using a spin-echo block before the acquisition.

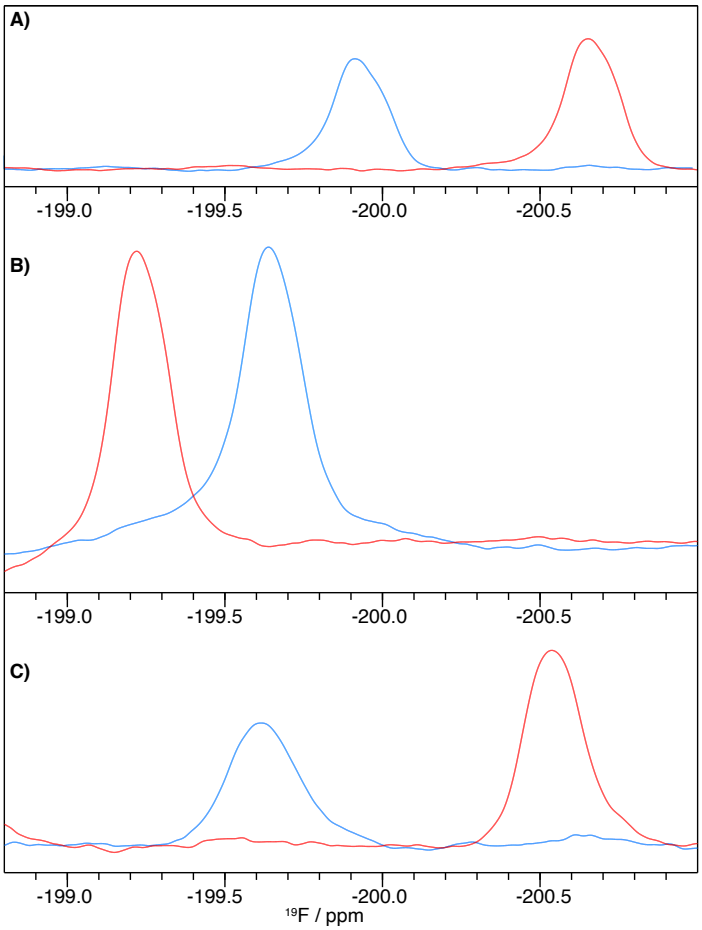

Figure 3.12: ^{19}F NMR spectra of *let-7* in interaction with targets (A) LIN28A and (B) LIN41 (600 MHz, 298K), and (C) HMGA2 (600 MHz, 280 K). In each case, red spectra highlight the interaction between targets and *let-7* labeled ^{19}F on 2'-U6, while light blues ones are samples containing ^{19}F 2'-U9 *let-7*.

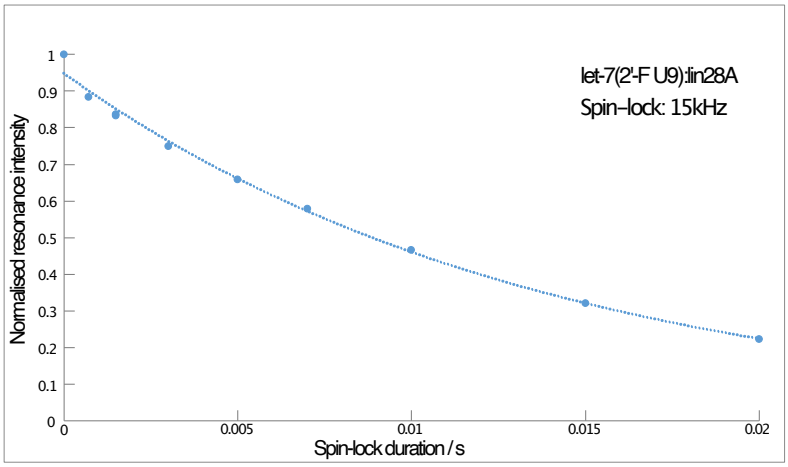

Figure 3.13: ^{19}F on-resonance $R_1\rho$: normalised resonance intensity of *let-7* (2'-F U9) in interaction with LIN28A as a function of spin-lock duration (600 MHz, 298K, 15 kHz).

T_1 have been measured for each sample and determined in the range 350-400 ms.

R1ρ on-resonance experiments have been performed on *let-7*:LIN28A and on *let-7*:LIN41 samples, at several spin-locks (1.5, 2.5, 3, (4), 6, 8, 10, (12), 15 kHz) and spin-lock durations (0.001, 0.007, 1.5, 3, 5, 7, 10, 15 (repeated), 20 ms).

Indeed, the melting temperature of the *let-7*:HMGA2 complex (around 32.8°C) is lower than in the case of the other interactions. The analysis of the interaction thus require to sample lower temperatures, however the signal-to-noise ratio obtained with the sample was not sufficient to conduct experiments in the time allowed. We thus chose to discard the *let-7*:HMGA2 complex in a first step, to focus on *let-7*:LIN28A and *let-7*:LIN41 interactions.

For each spin-lock, the resonance intensity was plotted as a function of the spin-lock duration, and fitted using a single exponential decay (see Figure 3.13). The $R_1\rho$ were calculated for each spin-lock power and plotted against the spin-lock frequencies (see Figure 3.14 and Figure 3.15).

The on-resonance experiments performed on *let-7*:LIN28A (see Figure 3.14) did not reveal any R1ρ rates discrepancy while changing the spin-lock. This would suggest that both fluorine probes incorporated at the U6 and U9 position are located in stable conformers in this time window. Therefore, both U6 and U9 are involved in strong base-pairs that are not exchanging. This result was expected concerning the U6 position, located in the seed region, however according to the structure prediction, the U9 is located at the edge of a loop, and the opening/closing on a few base-pairs around the loop could have been expected.

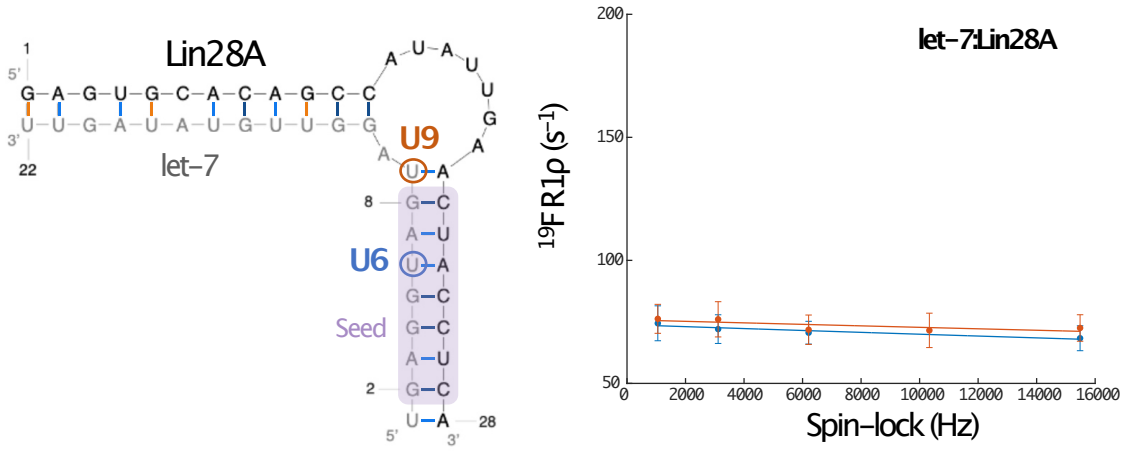

Figure 3.14: ^{19}F on-resonance $R_1\rho$ of *let-7* in interaction with LIN28A as a function of spin-lock power (600 MHz, 298K).

Regarding the *let-7*:LIN41 (see Figure 3.15) interaction, on-resonance experiments demonstrate a clear exchange in the U9 position, contrary to the U6 position where the possible exchange is poorly observable due to noise uncertainty and has to be more accurately investigated.

In the case of the U9 ^{19}F, the $R_1\rho$ values decrease from around 160 s^{-1} to around 110 s^{-1} between 1.5 and 15 kHz spin-locks. The *plateau* has not been reached yet, that may reflect a very slow conformational equilibrium or an abundance of excited states.

The next step consists in performing the off-resonance experiments, by applying low spin-locks at specific offsets of the fluorine frequency, to characterise the exchange through exchange rates, population of the states, and invisible state resonance frequency. The acquisition of these experiments as well as the data analysis is currently ongoing, at the time of writing.

Preliminary results were obtained for the *let-7*:LIN41 interaction (see Figure 3.16), using a series of offset compared to the ^{19}F U9 resonance, and 3 different spin-locks (600 Hz, 900 Hz, 1200 Hz).

Moreover, the spin-lock calibration of the probe was performed. The difference between specified and applied values usually represents less than 5% of the spin-lock power, however, to carry out a more advanced (and quantitative) analysis, the spin-lock calibration is required.

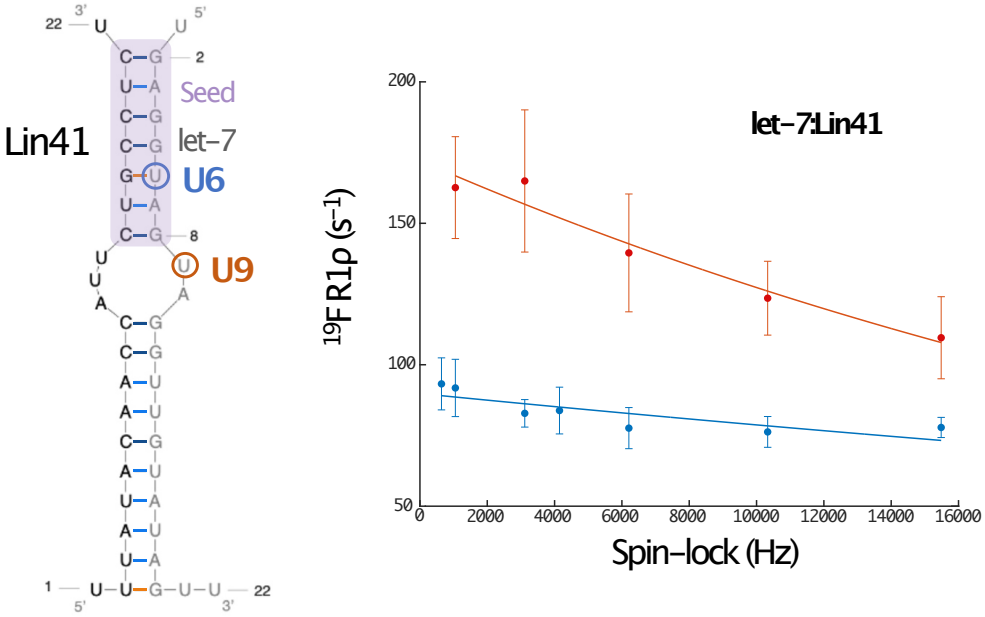

Figure 3.15: ^{19}F on-resonance $R_1\rho$ of *let-7* in interaction with LIN41 as a function of spin-lock power (600 MHz, 298K).

Although the analysis is not advanced yet, the preliminary results are promising. Indeed, separate analysis fitting the relaxation data to Laguerre's equations, using a two-states model and 1000 Monte-Carlo simulations, indicate relatively close results in terms of excited state population (around 20 %) and chemical shift offset (around 1 ppm), as well as approximative order of magnitudes for the exchange rate (see Table 3.6).

It is important to note that the chemical shift of the excited state seem quite close to the observable one. This is expected considering the poor dispersion already obtained between the synthesized samples (all contained in a 2 ppm window).

Parameter	600 Hz	900 Hz	1200 Hz
pB / %	18 +/- 16	20 +/- 46	19 +/- 31
$\delta\omega_B$ / ppm	0.6 +/- 0.2	1.4 +/- 0.7	0.8 +/- 0.3
k_{exAB} / s^{-1}	6e3 +/- 5e3	22e3 +/- 16e3	12e3 +/- 12e3

Table 3.6: Off-resonance exchange characterisation for *let-7*:LIN41 interaction, labelled ^{19}F on the U9 ribose 2' position, using Laguerre's equations and a two-states model.

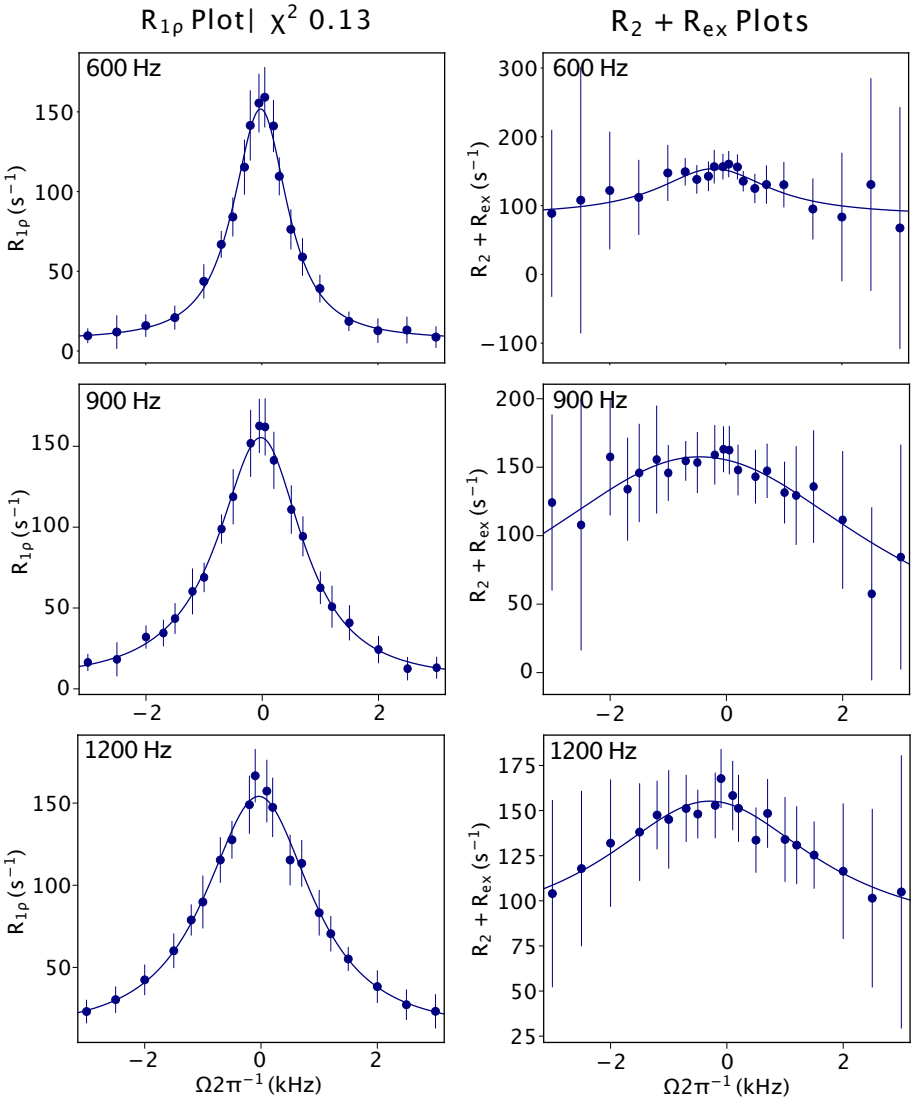

Figure 3.16: ^{19}F off-resonance $R_1\rho$ of *let-7* in interaction with LIN41 as a function of spin-lock offset to the U9 resonance, at 600, 900 and 1200 Hz, 298 K and 600 MHz

3.5.4 Conclusion

Many different synthesis and analysis may be performed to obtain a complete view of the RNAs behaviours. In this case, the introduction of the fluorine probe into the *let-7* sequence by chemical synthesis was successful, which could lead to a multiplicity of samples rendering the assignment obsolete in the ^{19}F dimension. The ^{19}F relaxation dispersion experiments were implemented and preliminary results could be obtained at several position, for *let-7* in

interaction with several targets.The convenient NMR properties of the fluorine nuclei, the directly accessible information, and the price of the phosphoramidites (cheaper than ^{13}C or ^{15}N labelled) make this method a valuable choice regarding RNAs studies. Therefore, combining chemical synthesis with ^{19}F to study interaction and dynamics by NMR is a method that is worth taking a look at for further understanding of the RNA world.

In this project, the absence of (or lower) exchange in the seed region of *let-7* in interaction with either LIN41 or LIN28A denotes the local strong pairing, which could corroborate the fact that the interaction depends on this region of a few nucleotides interacting. The dynamics observable in the non-seed regions also confirm this hypothesis: indeed, the presence or absence of exchange in other parts of *let-7* does not seem to affect the binding between the miRNA and its targets.

Eventually, the characterisation of the interaction would require to synthesize additional samples to be able to probe the interaction in supplementary locations within *let-7* sequence. Additional information is, at the time of writing, being acquired to fully comprehend the *invisible states* conformations.

Chapter 4

Investigation of a transfer RNA dynamics

Samples were all produced by Pierre Barraud (EGM, IBPC, Paris, France).

4.1 Transfer RNA

4.1.1 Biological function

Transfer RNAs (tRNAs) are involved in the translation process, where the mRNA is encoding for a protein *via* series of 3-nucleotides long: the *codons*. These codons are recognised by the *anticodon* part of matching tRNAs, that are bringing the corresponding amino acids loaded onto them to the ribosome. There, the successive delivered amino acids are assembled into the nascent protein chain, until a stop codon is reached.

Each kind of tRNA is matching with one particular amino acid, thus many different types of tRNAs can be encountered. As multiple codons can code for the same amino acid, it is also possible that various tRNAs bear the same amino acid.

4.1.2 Maturation process

Transfer RNAs initially derive from larger precursor (see Figure 4.1 (a)) undergoing a complex maturation process.

From pre-tRNA transcript to their mature form, tRNAs undergo a complex biogenesis where sequence parts might be removed, the 3'-CCA amino acid accepting sequence is added, amino-acylation is performed, and various post-transcriptional chemical modifications are incorporated (see Figure 4.1 (a))[148]. Interestingly, all these steps lead to the functionalization of the tRNAs, and allow the cell not being engorged by non-functioning tRNAs.

During this maturation, a series of reaction occurs, leading to specific base modifications within the tRNA sequence. Enzymes are responsible for the chemical modifications, occurring at specific positions within the tRNAs sequences, such as methylation and pseudouridylation. Among all RNAs, tRNAs are the ones presenting the most chemical modifi-

cations, that can consist on addition or substitutions of functional groups, as well as more complex ones.

It is important to note that these modifications are not all independent, but the variation of one nucleotide can have incidence on the apparition of other specific modifications in specific positions. These modifications are mainly located in the D-arm, the T-arm and the anticodon loop region.

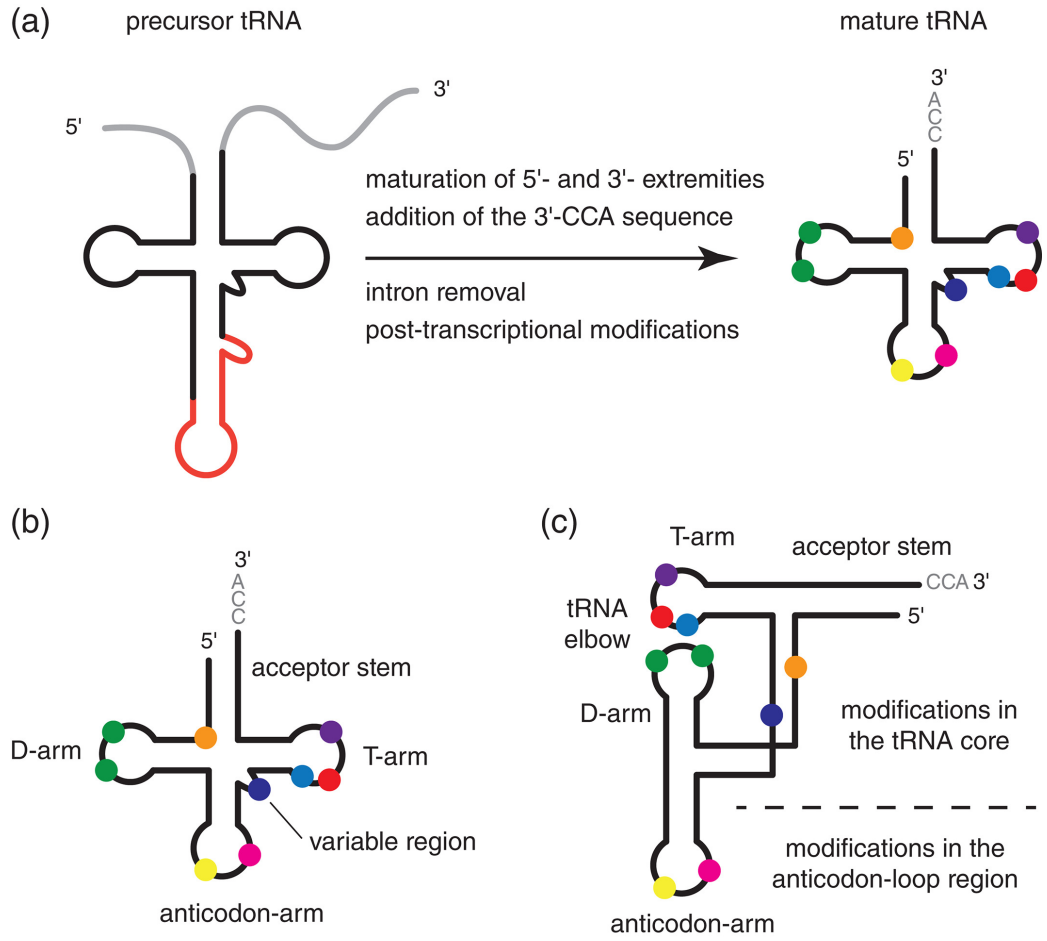

Figure 4.1: (a) Precursor tRNAs undergo a maturation process to turn to its mature form, during which it might undergo partial sequence suppression, addition of 3'-CCA sequence, and chemical modifications of the nucleotides (indicated by coloured dots). (b) tRNA typical mature form and nomenclature. (c) Overall tRNA folding structure to form the recognizable *cloverleaf* structure. Reproduced with authorisation from [149].

It has been shown that the Ψ55 is the initial modification incorporated in the yeast

tRNA^{Phe} is followed by the m^{7}G46 and T54 modifications[150] (see Figure 4.2). These modifications are quite recurrent in tRNAs, and therefore a study of the tRNA^{Phe} with and without modifications could give some insight on those chemical modifications functions[151].

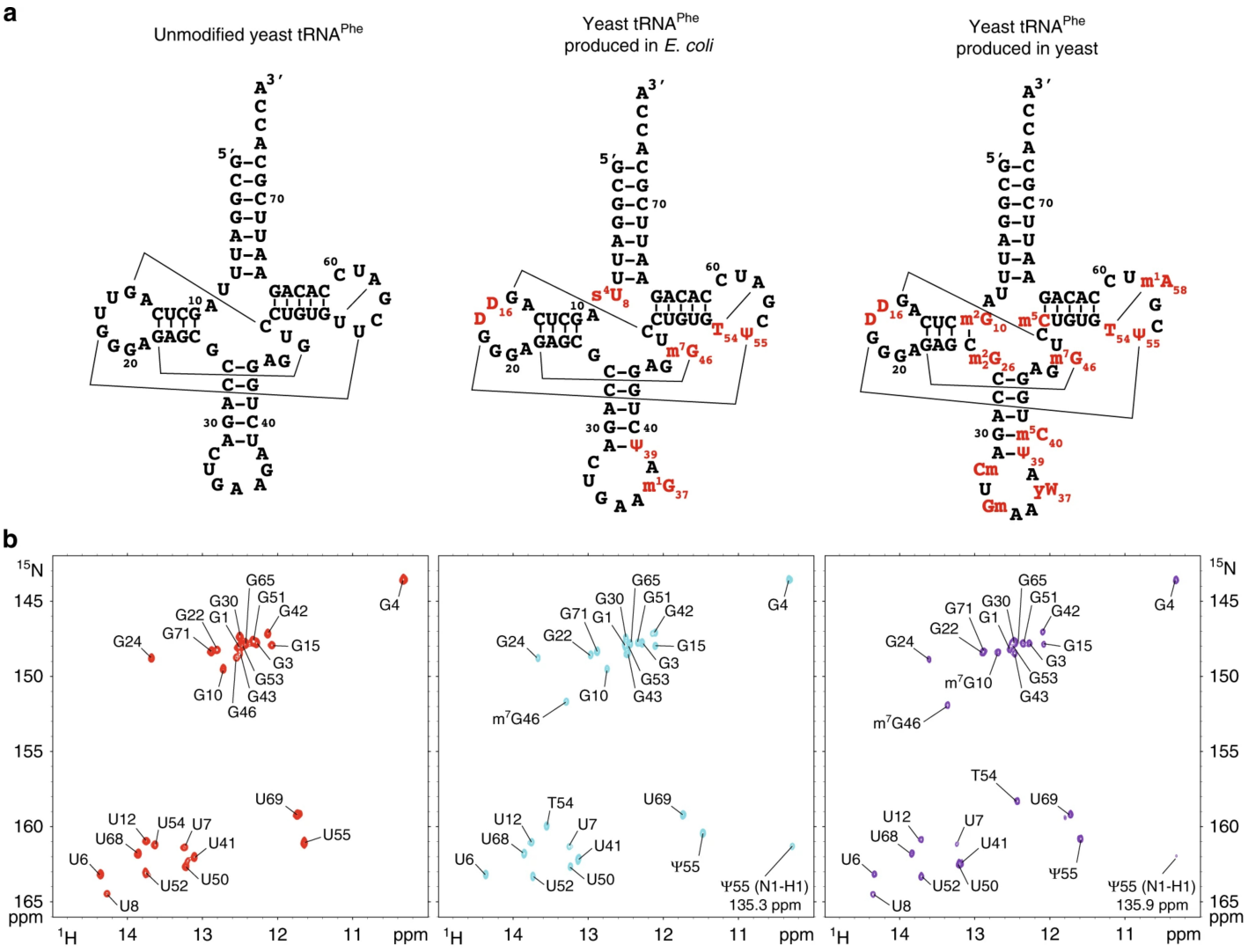

Figure 4.2: a) Structure and b) ^{1}H, ^{15}N-HSQC spectra comparison for unmodified and maturated form of the yeast tRNA^{Phe}. Reproduced with authorisation from [150].

4.1.3 Structure

tRNAs are typically 76 to 90 nucleotides long. tRNAs adopt an overall common shape, designated as a *cloverleaf* (see Figure 4.1 (b))[152].

They are constituted of an acceptor stem composed of a 5'-phosphate ending group, 7 to 9 base-pairs, and a longer 3'-terminus designated as the CCA tail. As nominated, the CCA tail is constituted of a cytosine-cytosine-adenine sequence, where the amino acid is loaded onto the tRNA by aminoacyl synthetases, thus leading to the formation of aminoacyl-tRNAs[153].

The D-arm is a stem usually composed of 4 to 6 base pairs, involved in the tRNAs tertiary structure stability. The anticodon loop region is 5 base pairs long and ending with a loop. Finally, the T-arm is a 4 to 5 base paired-stem involved in the interaction with the ribosome. Between the anticodon region and the T-arm lies a variable loop.

As the most recurrent variations occur in the T-arm, it is also named the TψC arm, after the chemical modifications present in its sequence: thymine, pseudouridine and cytidine[154].

The tRNAs 3D structure is folded so that they adopt the shape of an L (see Figure 4.1 (c)). Their shape allows the tRNA to recognises the characteristic mRNA codon on one end, while being charged with the corresponding amino acid on the other end of their structure.

NMR spectroscopy is a method of choice to understand dynamics variations occurring within the non-maturated and maturated form of the tRNA^{Phe}, by spin relaxation as well as relaxation dispersion, that lead to global and site-specific information.

The tRNA^{Phe} ^{1}H, ^{15}N-HSQC spectra reveal small overlap in the guanine regions, and almost none in the uracil region. The spectra could be folded until reaching a 3 ppm window in the nitrogen dimension, without bringing any additional overlap.

4.2 Spin relaxation

Following spin relaxation data acquisition and analysis, the connection to the physical motions may be performed using a *modelfree* description. The global approach is summarised in Figure 4.3.

Indeed, site-specific spin relaxation have been connected to internal motions that are variating while the biomolecules undergo different processes such as binding, folding and catalysis[155].

Investigating these internal motions in the non-modified tRNA^{Phe} and comparing them with a similar dataset on its modified form could help understand the importance of the

chemical modifications regarding the tRNA^{Phe} functions[156].

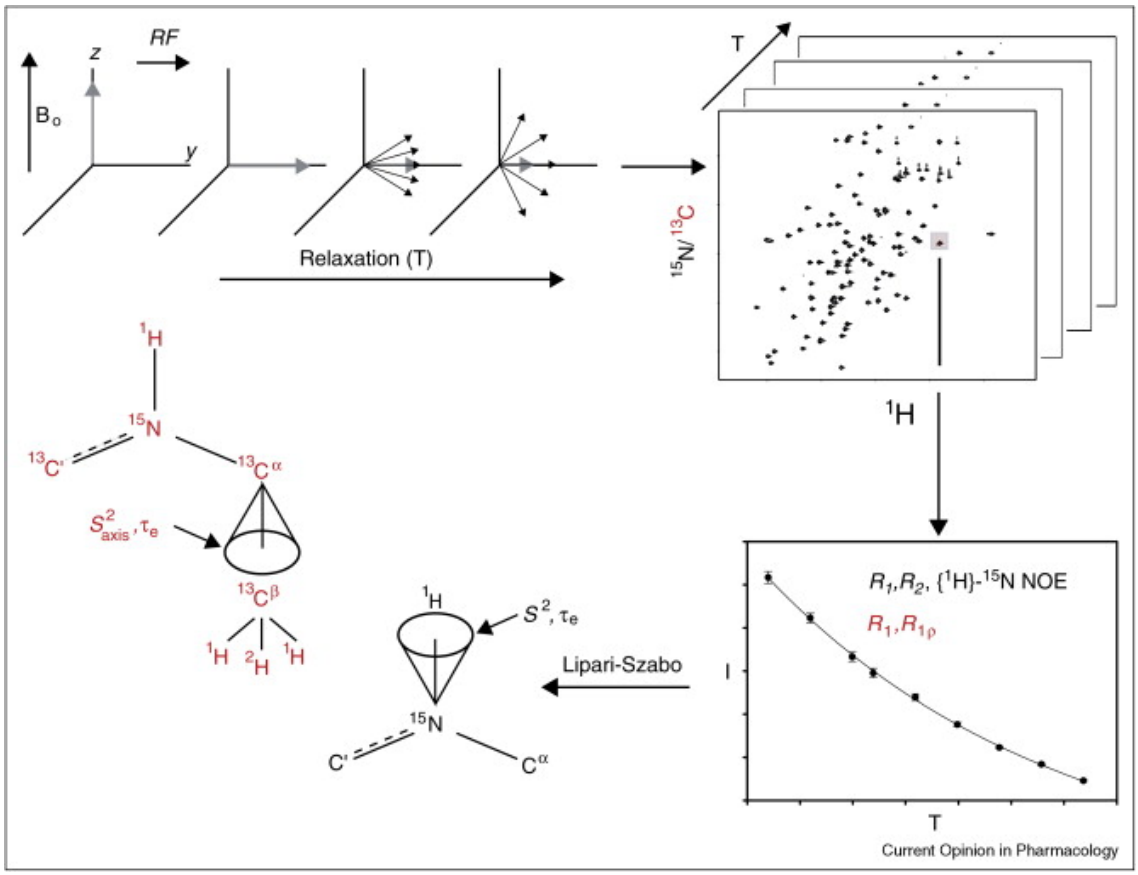

Figure 4.3: Flow-chart of site-specific spin relaxation acquisition and analysis to obtain biomolecular motional information. The relaxation rates are obtained through fitting single exponential decay to resonance intensities extracted from pseudo 3D ^{1}H, ^{15}N-TROSY planes, while the NOE values are obtained through ratio of the resonance intensities in irradiated and non-irradiated spectra. Data is then used to perform *modelfree* analysis, and obtain information on amplitude and timescales of internal motions. Reproduced with authorisation from [157].

4.2.1 Experimental set-up

Data were recorded on a ^{15}N-G and U labelled sample of unmodified yeast tRNA^{Phe}.

Site-specific ^{15}N relaxation experiments were acquired at 310 K, at 14.1 T, 18.8 T and 23.5 T (respectively 600 MHz, 800 MHz and 1 GHz) on Bruker Avance III or Bruker Avance Neo spectrometer, equipped with cryo-probes 5mm HCND or 3mm H/FCNP. Experiments at 800 MHz were acquired at the Institut Européen de Chimie et de Biologie (IECB), through the INFRANALYTICS network. Pulse sequences used were TROSY-based[50].

Different delays were used to record relaxation delays, in a randomised manner, with repetition of one value, in order to asses the reproducibility of the experiments:

- at 1 GHz: 0, 80, 160, 320, 480, 720, 960 (repeated), 1280, 1760 and 2960 ms for R_1 delays; 1, 10, 20, 30, 50 (repeated), 70, 90, 120, 160 and 200 ms for $R_1\rho$;
- at 800 MHz: 0, 80, 200 (repeated), 400, 600, 800 (repeated), 1200, 1600, 2000, 2400, 3200, 4000 ms for R_1 delays; 1, 2, 4, 6, 10 (repeated), 20, 30, 50 and 80 ms for $R_1\rho$;
- at 600 MHz: 0, 80, 240, 480, 720, 960 (repeated), 1280, 1760, 2960 ms for R_1 delays; 1, 10, 20, 30, 50 (repeated), 70, 90, 120, 160 and 200 ms for $R_1\rho$.

For $R_1\rho$ measurements, the RF spin-lock was set at 1.4 kHz. The sample was labelled on ^{15}N only so no ^{13}C decoupling was applied. 128 scans per FID were recorded, with a recovery delay between scans of 2 s.

At 1000 MHz, the spectral dimensions were 21739 Hz (F2) and 375 Hz (F1), corresponding to 21.7 x 3.7 ppm. The spectral width were kept identical for all magnetic fields. The ^{15}N carrier was set at 162 ppm.

The ^{15}N-^{1}H hetNOE and reference spectra were recorded in an inter-leaved approach. Carriers were kept identical to the relaxation experiments. A recovery delay of 5 s was chosen, and an additional delay of 3 s was added in presence or absence of saturation.

4.2.2 Data analysis

Spectra were processed and analysed using the NMRPipe software package. Spectral intensities were extracted with NMRPipe scripts, and fitted using single-exponential decays (plots available in annexe at 600 MHz). Both pulse sequences and NMRPipe processing scripts are available online through Bax's group library. Analysis of relaxation rates and error calculations through Monte-Carlo simulations were performed using Matlab scripts written in-house. Random noise based on experimental spectral noise was added to the data to be fitted, and the error was calculated based on the standard deviation produced by 500 simulations.

As expected, the longitudinal relaxation rates decrease while the transverse relaxation rates increase with the magnetic field. However, the gap between R_2 values for the different fields is especially observable between 600 and 800 MHz for most of the nucleotides. It seems that these ones may experience exchange, as the R_{ex} contribution becomes larger with the magnetic field.

While a few nucleotides arbour lower hetNOE values, the values are contained in a narrow range around the average value. This is expected, as most of them are implicated in base-pairings. All of the NOE values are comprised between 0.7 and 1, which is also unsurprising for this size of RNA, and shows no particular variation between the magnetic fields.

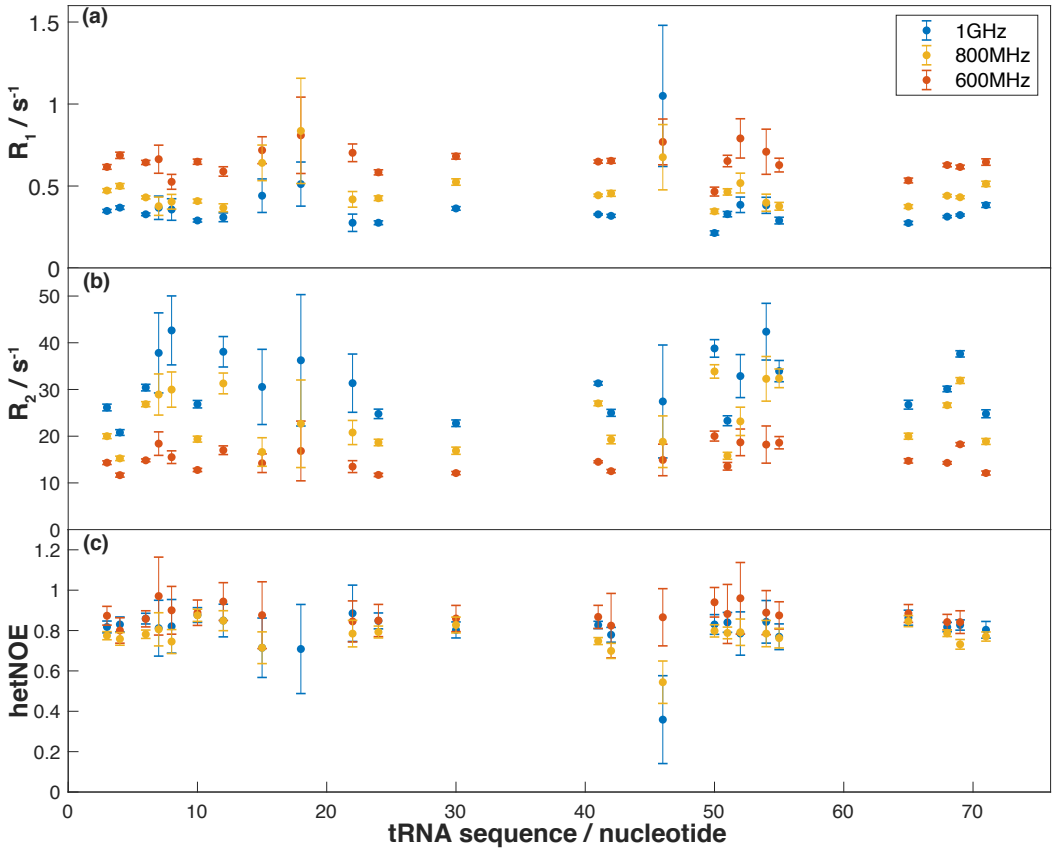

Figure 4.4: tRNA^{Phe} ^{15}N site-specific spin relaxation rates at 600 (red), 800 (yellow) and 1000 (blue) MHz and 310K. (a) Longitudinal relaxation rates R_1; (b) Transverse relaxation rates R_2; (c) ^{15}N-{^{1}H} NOE.

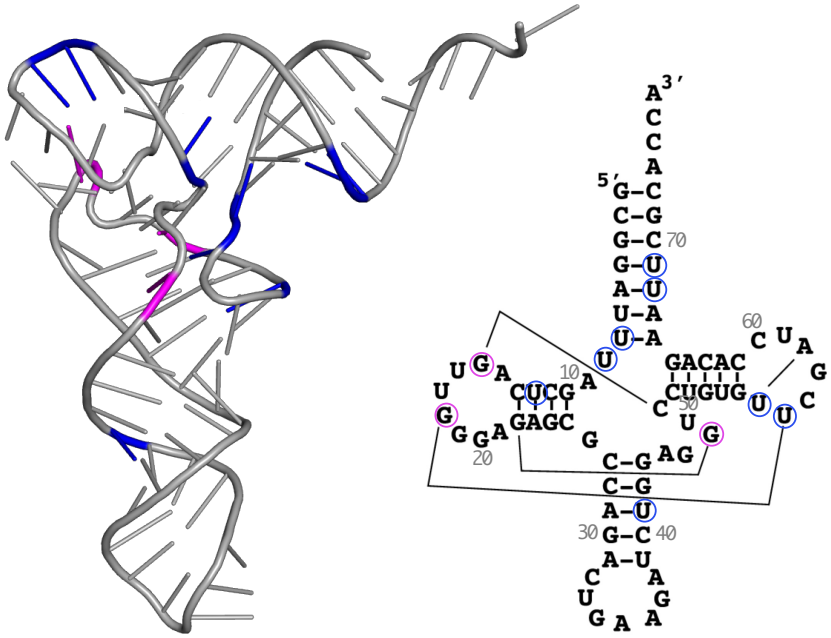

Figure 4.5: Secondary structure highlight of the nucleotides (non-modified tRNA form) involved in conformational exchange. Encircled nucleotides experience a *(Pink)* NOE decrease (thus higher flexibility) or *(Blue)* visible exchange rate (observable through R_2 significant increase along with magnetic field), through first glance at relaxation results.

At first glance, the nucleotides experiencing exchange or higher flexibility are located either in the acceptor stem or the arms. In particular, lower NOE values are observable for nucleotides involved in tertiary interactions, which could affect the folding of the tRNA, whereas eventual exchange in the acceptor stem may affect the amino acid binding. To comprehend better the relaxation data, a more robust analysis is required: the *model-free* one.

4.3 Model-free analysis

4.3.1 Variable parameters, models and selection

The software *ModelFree* developed in Palmer's lab was used to interpret the relaxation data, throughout the Lipari & Szabo formalism[36][158].

To recall the models used in the software are the ones listed below:

Model number	Parameters fitted
M1	S^2_s
M2	S^2_s, τ_e
M3	S^2_s, R_{ex}
M4	S^2_s, τ_e, R_{ex}
M5	S^2_s, τ_e, S^2_f

Table 4.1: Modelfree: model numbers and parameters to be fitted.

The fit of each model is evaluated using sum-squared error for each nucleotide i (*SSE(i)* or Γ_i), which sum over the whole $tRNA^{Phe}$ sequence gives the χ^2 of the fit, and for each magnetic field j. The *SSE(i)* is calculated using the fitted values for R_1, R_2 and NOE, to the power of 2, divided by the experimental error to the power of 2.

$$\Gamma_i = SSE(i) = \sum_{i=1}^{N}\sum_{j=1}^{M} \frac{(R_{1,exp,ij} - R_{1,fit,ij})^2}{\sigma^2_{R1,ij}} + \frac{(R_{2,exp,ij} - R_{2,fit,ij})^2}{\sigma^2_{R2,ij}} + \frac{(NOE_{exp,ij} - NOE_{fit,ij})^2}{\sigma^2_{NOE,ij}} \tag{4.1}$$

with R_{exp} and NOE_{exp} the experimental relaxation data, R_{fit} and NOE_{fit} the corresponding fitted relaxation data, σ the corresponding experimental error, N the number of nucleotides, and M the number of magnetic fields.

The comparison between the 2 models being tested is realised using an F-statistical test, where:

$$F = \frac{d_2}{d_1 - d_2}\frac{SSE_1(i) - SSE_2(i)}{SSE_2(i)} \qquad d = n - m \tag{4.2}$$

with n the number of relaxation parameters and m the number of fitted *modelfree* parameter.

The model selection was performed as detailed in Palmer's *modelfree* manual and Zhang and co-workers publication[47].

Comparison between models parameters fitting is done between two successive models: M1-M2, M2-M3, M3-M4, M4-M5; as well as in between models that are specifically subsets of each other: M1-M3 and M2-M4.

For each model, SSE(i) are compared with the distribution, at a confidence level of $\alpha = 0.10$. Starting from simpler models comparison, if SSE(i) was less than the critical value, then the model is admitted to describing well the dataset. However, the two models are still compared using the F-stat: if the F-stat value is higher than the critical value, then the second model is described even better the experimental datasets. The same procedure was repeated for all models M_xM_{x+1} comparisons. A summary of the selection process is represented on Figure 4.6.

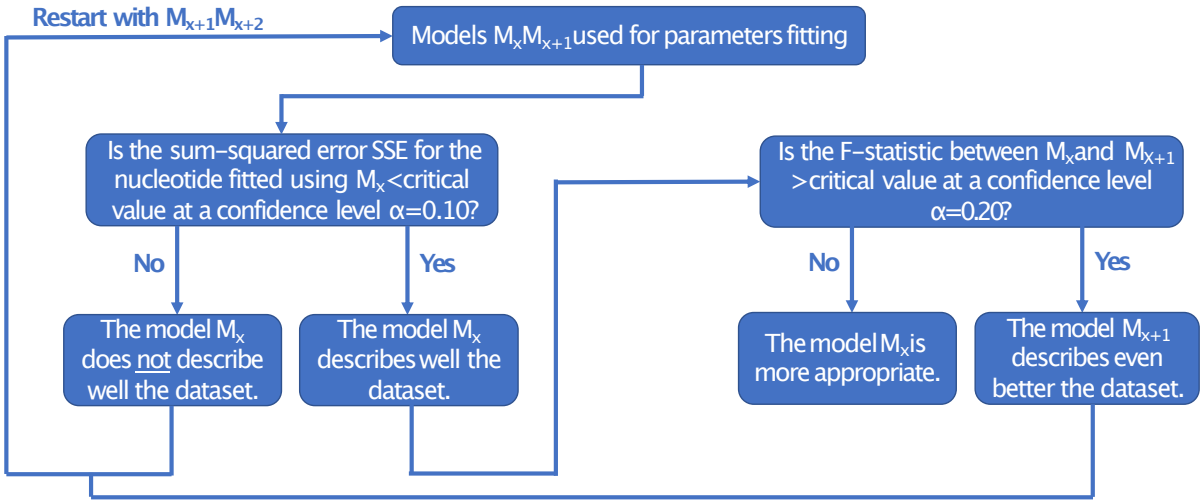

Figure 4.6: Scheme of the model selection by comparison of parameters fitting to the experimental data, in the model-free analysis.

4.3.2 Preliminary parameters determination

The global rotational correlation time τ_c has to be characterised first. This global motion is considered to be independent from the internal motions, and is suspected to occur on tens of nanoseconds while the NH vectors should experience motions around the picosecond-nanosecond timescales.

τ_c was characterised using TRACT experiments, at 14.1 T and 23.5 T (respectively

600 MHz and 1 GHz)[159]. ^{1}H, ^{15}N-TRACT experiment measures the transverse cross-correlated relaxation time between amide protons and nitrogens, to determine τ_c without influence of the dipolar interactions from nearby protons. τ_c was determined to be between 13.2 ns (using -130 ppm value for CSA) and 15.8 ns (using -110 ppm).

The diffusion tensor was determined to be axially symmetric using HYDRONMR software[160]. HYDRONMR evaluates basic hydrodynamic quantities and calculates relaxation for each nuclei taken in a PDB file. The D_x and D_y values obtained are similar (around 8e^6 s-*1*), while the D_z has an increased value (around 1.5e^7 s-*1*), meaning that the diffusion tensor for the tRNA^{Phe} tends to be axially symmetric. The tensor coefficients were converted using Fushman's $D_{//}/D\perp$ ratio, leading to a D_{ratio} around 1.9.

4.3.3 Model-free analysis of tRNA^{Phe} spin relaxation data

Considering the amount of relaxation data available, model-free preliminary analysis was performed using each single field and all together, first using all the datasets available for every nucleotide.

The global correlation time and tensor D_{ratio} were either fixed using HYDRONMR parameters (τ_c = 15 ns and D_{ratio} = 1.9) or optimised as free parameters. Imino bonds length and chemical shift anisotropy values were fixed according to Zhang, Q. et *al.* (Science, 2006), ie r_{NH} = 1.01 Å and CSA(G-N1) = -130 ppm while CSA(U-N3) = -100 ppm in G-C or A-U Watson-Crick baise pairs as well as loops[47].

Secondly, the same analysis was conducted using selected nucleotides, representing the ones without noticeably large increase in R_2 values that would denote of an exchange contribution. This was done comparing the dataset acquired at the different fields, and spotting nucleotides sporting the largest relaxation variations. Again, the analysis was conducted both at each field separately and all together, fixing τ_c and D_{ratio} or optimising them. All analysis were performed using 300 MonteCarlo repetitions.

When the global correlation time and the D_{ratio} were not fixed, the M3 model was selected for most of the nucleotides. However, the parameters optimisation presented a wide variation (a few nanoseconds for τ_c, and an order of magnitude for D_{ratio}).

With the HYDRONMR fixed parameters, none of the conditions were completed, therefore none of the models was considered to fit the datasets. However the global tendency for the SSE(i) value to decrease for the same model as the one selected in the previous non-restricted trials (mostly M3) was observed.

Thus, the global correlation time and the tensor D_{ratio} were fixed using an average of the optimised parameter obtained when running the program: τ_c = 14.8 ns while D_{ratio} = 0.4.

When fixing these parameters, the conditions to consider a model appropriate were

fulfilled, and the same trend towards M3 models appeared, for the exact same nucleotides (see Table 4.2).

Considering the exchange probed in the transfer RNA, it seems in concordance to the fact that the R_2/R_1 ratios are high compared to the global correlation time of the tRNA, due to exchange contribution to a wide range of the tRNA nucleotides (see Figure 4.7).

Nucleotide	Model selected	S^2	$\Delta S^2{}_s$	τ_e (ps)	$\Delta\tau_e$ (ps)	R_{ex} (s^{-1})	ΔR_{ex} (s^{-1})
G3	M3	0.632	0.011	N.F.	N.F.	4.460	0.390
G4	M3	0.712	0.022	N.F.	N.F.	0.951	0.409
U6	M3	0.846	0.021	N.F.	N.F.	1.284	0.447
U7	M4	0.799	0.129	413.487	1163.348	6.162	3.448
U8	M3	0.670	0.089	N.F.	N.F.	5.183	2.626
G10	M3	0.696	0.023	N.F.	N.F.	0.678	0.499
U12	M3	0.773	0.058	N.F.	N.F.	3.839	1.561
G15	M2	0.766	0.107	70.671	314.480	N.F.	N.F.
G22	M1	0.727	0.074	N.F	N.F	N.F	N.F.
G24	M3	0.651	0.021	N.F.	1.127	0.220	0.344
G30	M3	0.690	0.020	N.F.	N.F.	0.192	0.314
U41	M3	0.844	0.015	N.F.	N.F.	1.561	0.301
G42	M3	0.701	0.021	N.F.	N.F.	0.215	0.311
G46	M2	0.725	0.166	44.356	47.154	N.F.	N.F.
U50	M3	0.490	0.030	N.F.	N.F.	12.422	1.006
G51	M3	0.731	0.035	N.F.	N.F.	0.373	0.462
U52	M4	0.933	0.070	1879.324	1811.193	1.993	1.914
U54	M3	0.785	0.120	N.F.	N.F.	8.937	3.368
U55	M3	0.698	0.054	N.F.	N.F.	7.691	1.340
G65	M3	0.520	0.019	N.F.	N.F.	5.706	0.488
U68	M3	0.809	0.019	N.F.	N.F.	2.068	0.383
U69	M3	0.685	0.015	N.F.	N.F.	9.003	0.340

Table 4.2: Modelfree results of unmodified ^{15}N tRNA^{Phe} relaxation data at all fields. N.F. stands for *Not Fitted*. Here, exchange is estimated at 310 K and 600 MHz.

The order parameters globally adopt low values (0.6 - 0.8, see Figure 4.8), that would be expected in flexible parts such as loops. However, additional model-free analysis were performed with modified bond lengths values. Indeed, by taking into account the distances from a similar study by Biedenbänder and co-workers on a tRNA^{fMet}, meaning $r_{NH,G}$ = 1.047 Å and $r_{NH,U}$ = 1.053 Å, as well as CSA values, there is a global increase in the order parameters (see Annex 6), that bring them near values that would be expected in folded

domains of proteins (0.8-1)[161].

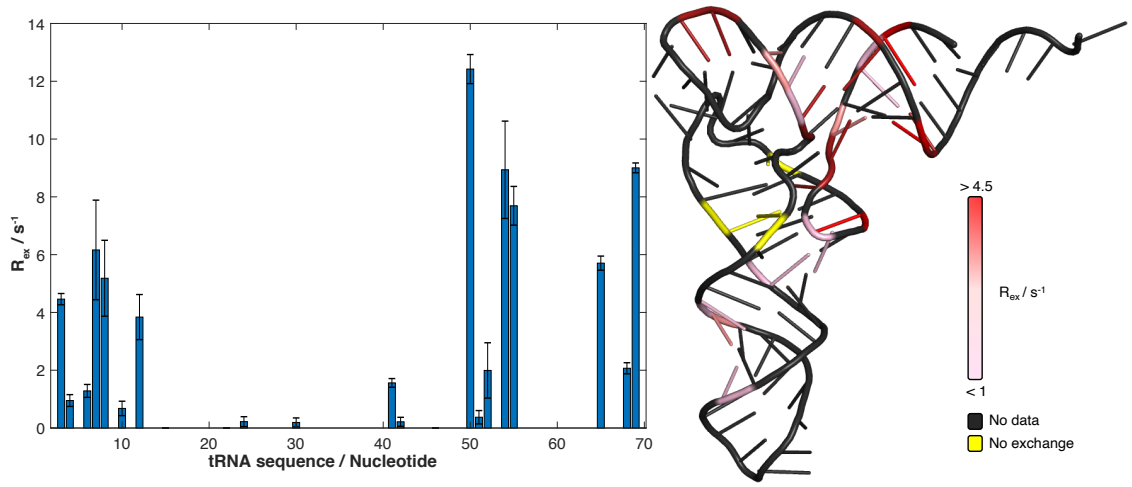

Figure 4.7: Exchange contribution determined by model-free analysis, plotted against the nucleotide sequence, and on structure.

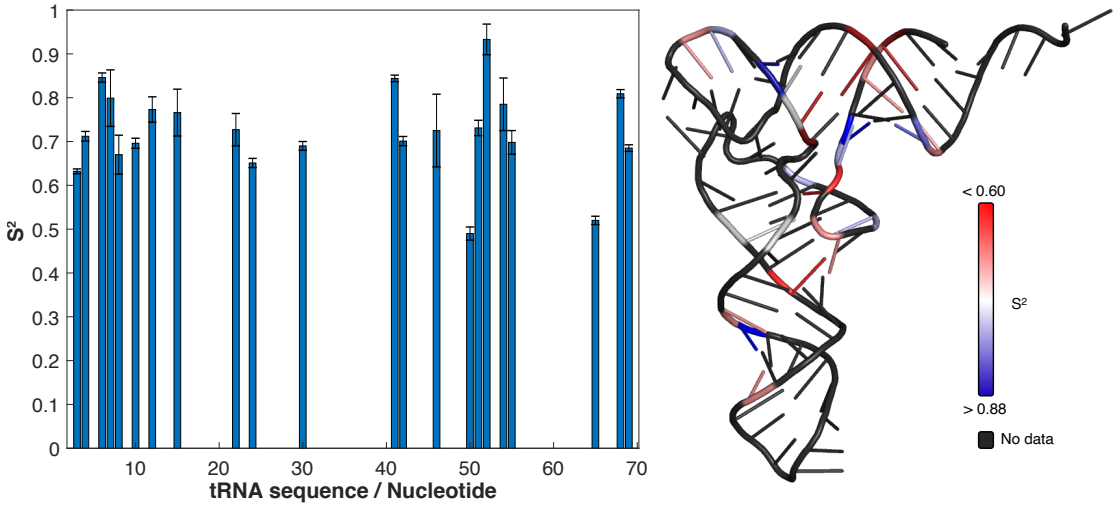

Figure 4.8: Order parameter determined by model-free analysis, plotted against the nucleotide sequence, and on structure.

The relative values between nucleotides are however more trustworthy, since the increase/decrease of the order parameters is global when changing the NH bond lengths. Therefore, the NH vectors with particularly low S^2 values (G3, U50 and G65) result as more flexible than the others for every analysis.

Nonetheless, the variation in diffusion tensor obtained with the HYDRONMR analysis and the optimisation in the *modelfree* software should be (and is at the time of writing) further investigated.

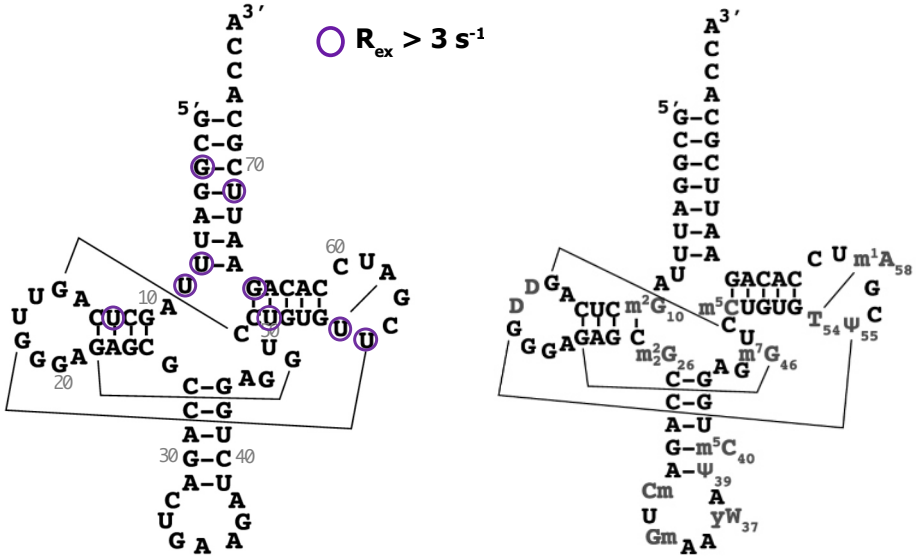

Figure 4.9: Secondary structure highlight of the nucleotides involved in conformational exchange. Encircled nucleotides experience an exchange rate greater than 3 s^{-1} according to the *modelfree* analysis results. *(Left)* Non-modified tRNA structure; *(Right)* Modified tRNA structure (yeast tRNA produced in yeast), for reference.

Interestingly, the highest exchange contributions highlighted using the *modelfree* analysis (see Figure 4.9) are located in the arms and acceptor stems. This exchange between conformers involving arms that should fold onto themselves may have a great impact on the capacity of the tRNA to complete its function by entering the ribosome. It would be interesting to conduct this study on the modified form of the tRNA, to observe whether the chemical modifications stabilise its structure in a particular conformer.

Most of the nucleotides therefore experience exchange at low timescales. This is in accordance with the results obtained on a similar study by Biedenbänder and co-workers on a $tRNA^{fMet}$[161]. They observed this exchanging behaviour for the non-modified $tRNA^{fMet}$

at 298 K, that seemed to vanish when increasing the temperature to 313 K. With the $tRNA^{Phe}$ here, the exchange is still observable at 310 K, but it would be interesting to increase the temperature to see if the same tendency appears. Nonetheless, while increasing the temperature, the expectations are that the exchange would be faster, so these results are particularly surprising.

For the modified $tRNA^{fMet}$, they observed a global stabilisation, both at 298 K and 313 K, especially in the T-arm, where the pseudouridine modification is located.

This could corroborate the hypothesis emitted for the stabilisation of the tRNA to enter the ribosome. Further work on the $tRNA^{Phe}$ would be helpful to extend or contrast with the data obtained for the $tRNA^{fMet}$, and comprehend better the tRNA in a more global vision.

4.4 Relaxation dispersion

4.4.1 CPMG data acquisition

NMR relaxation dispersion was performed on the non-modified tRNA.

At 1000 MHz, the ncyc list was composed of 19 cycles (0 repeated, 1, 2, 3, 4, 5, 6, 7, 8, 10 repeated, 11, 12, 13, 14, 16, 18 and 20), and 14 at 600 MHz (0 repeated, 1, 2, 3, 4, 6,9, 12, 16, 20 repeated and 32. The relaxation delay was set to 32 ms at both fields. The 15N carrier was set at 161.45 ppm. At 1000 MHz, the spectral dimensions were 14423 Hz (F3) and 225 Hz (F1), corresponding to 24 x 3.7 ppm.

4.4.2 Results

The relaxation dispersion curves globally do not reveal any exchange, detectable through CPMG experiments, apart from the nucleotides U52 and U55 (see Annexes 4 and 5).

The program CATIA was used to estimate the kinetics of the exchange, following Bloch-McConnell matrices[162].

Here, the analysis suggests the presence of a minor state population of 9.9 +/- 2.9 % with an interconversion rate between states k_{ex} is around 460 +/- 150 s^{-1}.

The amount of exchange detectable by CPMG is unexpected compared to the exchange determined using *model-free* analysis. However, the results are not incompatible, since the exchange timescale could explain the difference between both results.

An interesting addition to the set of relaxation experiments already performed would be to perform CEST and R1ρ experiments, that could help further characterise the exchange,

confirm the results obtained, and investigate different timescales.

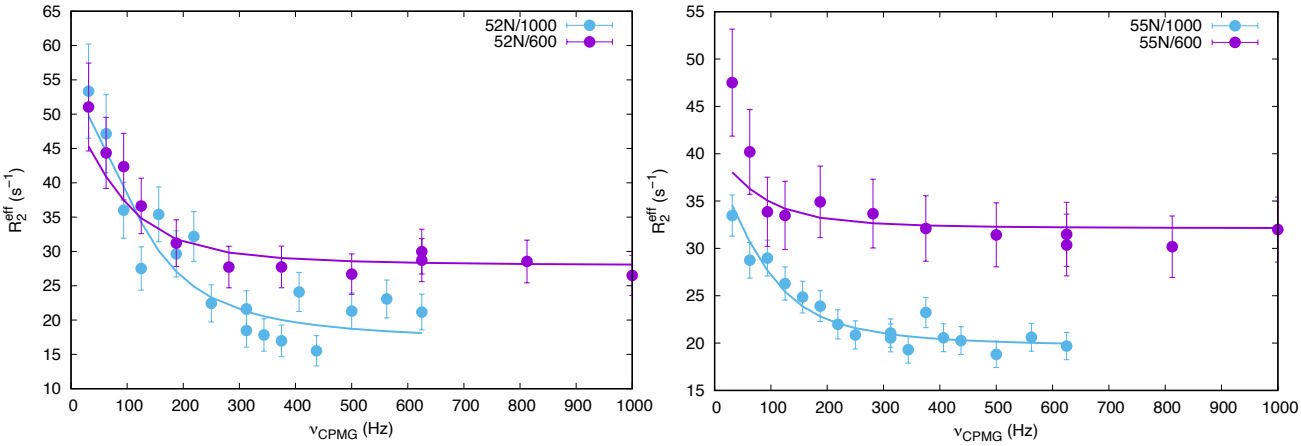

Figure 4.10: CPMG data analysis for two nucleotides experiencing exchange (U52 and U55), using the CATIA program, at two fields (600 MHz and 1000 MHz)[162].

4.5 Conclusions

The non-modified tRNA^{Phe} seems to undergo exchange with other conformer, as determined using ^{15}N spin relaxation. Interestingly, as observed in literature for tRNA^{fMet}, this exchange seems to vanish at higher temperature, as well as when introducing chemical modifications. This behaviour change could explain the importance of the chemical modifications, as they could stabilise the tRNA^{Phe} structure into one conformer, that would be more suitable to enter the ribosome and deliver the amino acid necessary to the protein synthesis.

Interestingly, tRNAs are known to adopt many different states once loaded within the ribosome subunit. Indeed, ribosomes are also highly dynamic entities, and the adaptation of the tRNAs may be a key step towards the tRNA motion resulting in peptide synthesis[163]. However, the exchange between conformers prior to loading onto the ribosome may be a restriction to the entry into the ribosome, or at least changing peptide synthesis by altering the rate of the amino acid addition within the ribosome[164]. Chemical modifications could therefore help control the peptide synthesis rate by influencing the folding of tRNA into their correct structures[165].

The following part of this project would consist in both studying the yeast tRNA^{Phe} while introducing chemical modifications, as well as characterising the exchange variation between the sample using methods such as relaxation dispersion. Indeed, the observations

made through CPMG experiments confirm the presence of an exchange with a minor state population, however the timescale probed only concerned a few nucleotides compared to the ones subject to exchange in *model-free* analysis.

The *model-free* analysis is also being tested with other CSA and NH distances parameters, to analyse their influence on the final results even further.

Chapter 5

Investigating a bacteriophage recognition interference by NMR spectroscopy

5.1 Introduction

Bacteriophages (also called phages) are the most abundant biological entities on the planet.[166][167] These viruses infect specific species of bacteria cells, which makes them adequate candidates to treat bacterial diseases.[168] They are composed of several proteins, enveloping genetic material. Phages require a host to be able to replicate, and propagate in their environment.

5.1.1 History overview

In 1915, Frederick Twort discovered that colonies of bacteria could become cloudy, due to their breakdown.[169] In 1917, Félix D'Hérelle noticed that small organisms were responsible for the recovery of patients suffering from diarrheal illnesses.[170][171] These organisms were able to provoke lysis of bacteria, and were therefore named bacteriophages, a derivation from the word *bacteria* and the greek term *phagein* literally meaning *to devour*. D'Hérelle worked on studying these bacteriophages and their mechanism, focusing on using them as a tool to cure bacterial diseases, with a particular emphasis on treating dysentery.[172]

However, the specific aspect of the bacteriophage methodology was not well understood at that time and, in Western Europe, the use of phages was forgotten and replaced by the antibiotics due to their immediate efficiency and availability. In the 1920s, George Eliava, a Georgian microbiologist, visited the Pasteur Institute, meeting D'Hérelle and taking up an interest in those bacteriophage viruses. Returning to Eastern Europe, and more precisely to Georgia, Eliava then founded an institute (renamed *George Eliava Institute* after his death), where phagotherapy was, and still is, investigated and used.

Nowadays, phagotherapy is a current interest topic considering its alternative and complementary aspect regarding antibiotics and the appearance of resistance strains.

5.1.2 Bacteriophages structures and classification

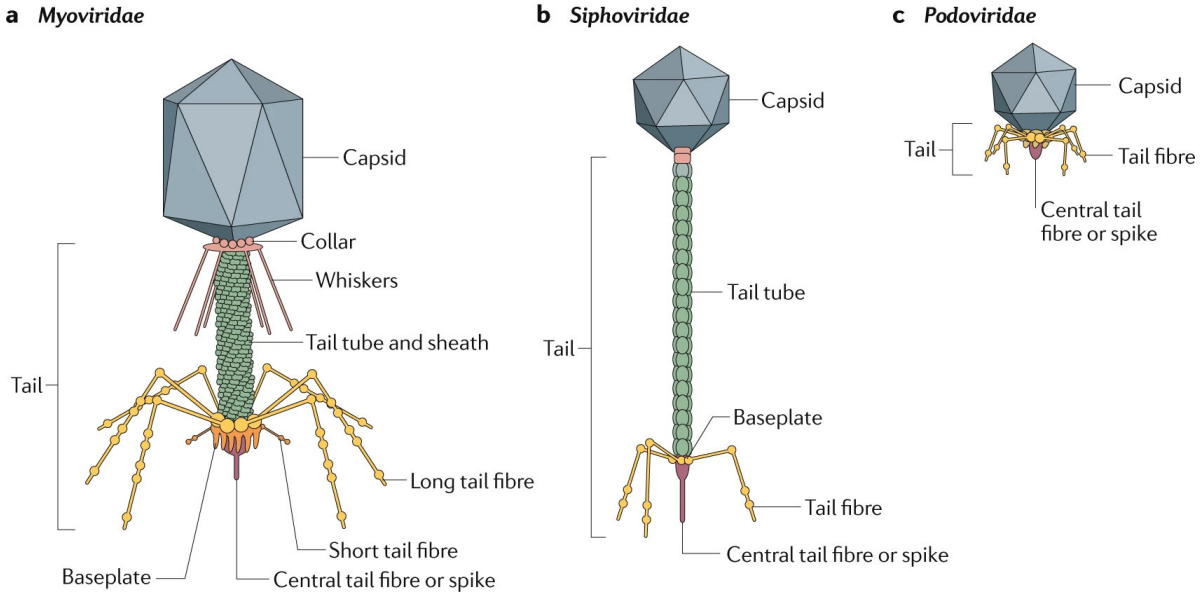

Figure 5.1: Bacteriophages structures within *Caudovirales* family (reprinted with permission from reference[173]).
All the *Caudovirales* phages possess a tail attached to a capsid containing the genome. (a) Phages of the *Myoviridae* family have tails with contractible sheath; (a) (b) Both *Myoviridae* and *Siphoviridae* families posses a tail baseplate with receptor-binding proteins attached; (c) *Podoviridae* phages have no baseplate and shorter tails.

Bacteriophages are classified into 11 families depending on the genetic material they contain and on their structure. Bacteriophages structure can be as varied as being icosahedral, filamentous, or constituted of a head, generally called the *capsid*, and a tail.

The largest family (representing around 95% of known bacteriophages) is the *Caudovirales*, which includes the phages that possess a tail and contains double-stranded DNA (dsDNA). Sub-categories within the *Caudovirales* family are divided depending on the tail structure: *Myoviridae* have long contractible tail, *Podoviridae* small tails and *Siphoviridae*, the one this study is focused on, possess long non-contractible tail (see Figure 5.1)[174]. These tails connect the capsid containing the genome and the spike, which is surrounded by tail fibres.

5.1.3 Mechanism of infection of the T5 bacteriophage

Specificity of infections orchestrated by phages is based on the recognition of cell surface receptors that will interact with the receptor protein located at the phage tail tip: the Receptor Binding Protein (RBP). Once this interaction happens, the tail is triggered which leads the phage to rearrange its conformation to perfore bacteria wall and release its genome contained in the capsid, through the tail to the bacteria.[175]

Infection of a bacterium by a bacteriophage is constituted of different steps, namely:
1/ the fixation of the phage to the host by recognition of the surface receptors on bacteria cell membrane and binding of tail fibres to the cell surface,
2/ the degradation of the cell wall, which consists in a membrane perforation,
3/ finally, the injection of the phage's DNA from the capsid to the gram negative bacteria.

This study is focused on the T5 bacteriophage, within the *Siphorividae* family. Based on the phage viral reproduction cycle, T5 is part of the Type I class: a strict lytic phage, which infects the bacteria *Escherichia coli (E. coli)*. As previously stated, the selectivity of bacteriophage depends on the receptors both located on the cell membrane and on the phage tail. Here, those receptors are respectively the protein pb5 for T5 and FhuA for *E. coli*. As the lytic cycle occurs, T5 injects its DNA into *E. coli* where it is replicated and proteins are synthesized, then assembled into new phages. This production of new entities makes the bacteria burst open, liberating a multitude of new phages ready to infect other bacteria.

5.1.4 T5 lipoprotein Llp

To prevent a surinfection of the cell by other bacteriophages, a protein is encoded by the phage T5 right after the DNA injection. This *lytic conversion* consists in the expression of the lytic conversion periplasmic lipoprotein (Llp).The binding of Llp to the receptor FhuA prevents pb5 recognition/interaction, and thus prevents infection from additional phages.[176][177] Llp inserts in the *E. coli* membrane and binds to FhuA, in order to inhibit the binding of FhuA to other phages through their RBP (see Figure 5.2).[178][179][180]

Llp has a molecular weight of 7.8 kDa and its sequence Q38162 is constituted of 62 residues:
CSTFGPKDIKCEAYYMQDHVKYKANVFDRKGDMFLVSPIMAYGSFWAPVSYFTEG
NTCEGVF.

The lipoprotein mechanism of interaction with FhuA is not known but is probably responsible of conformational rearrangements within the bacteria cell receptor. The aim of this study is therefore to investigate Llp and its interaction with FhuA, which prevents the

binding of additional phages to the already infected bacteria.

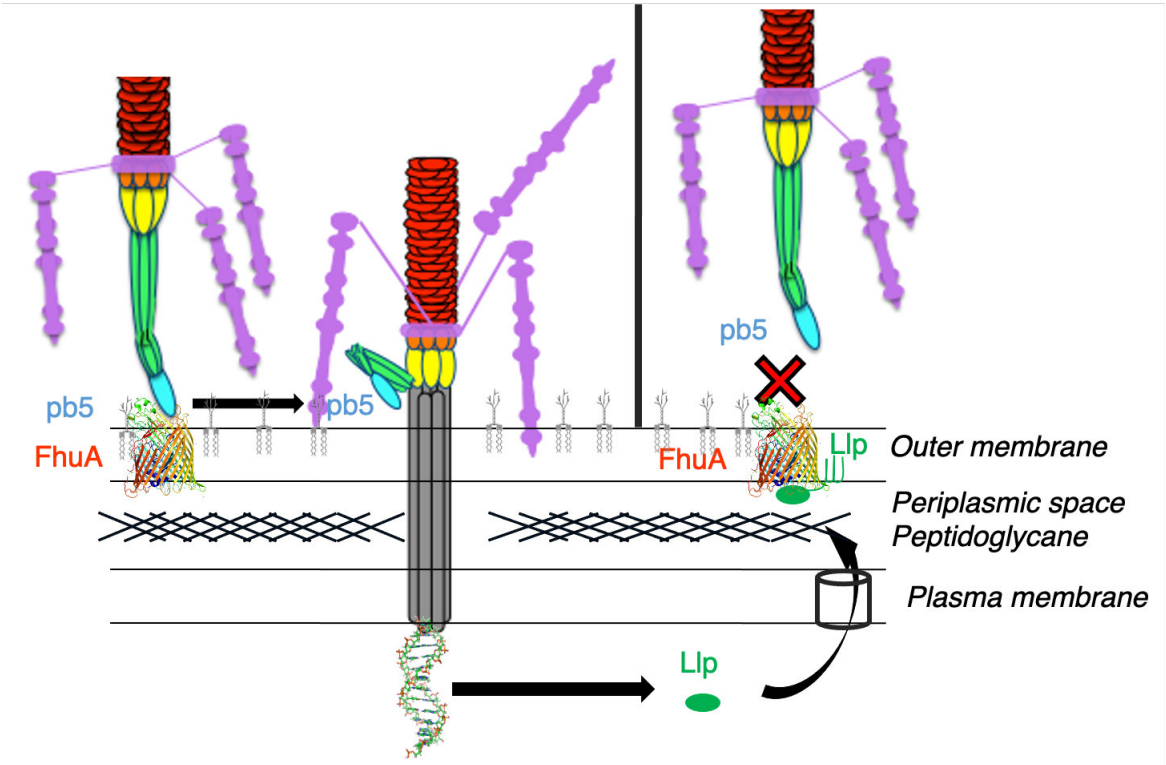

Figure 5.2: Llp-FhuA interaction due to T5 infection.
In the first step of infection, the RBP pb5 recognises the *E. coli* FhuA receptor, which leads to a membrane perforation and the injection of the DNA into the bacteria. This infection triggers the production of several proteins amongst which is Llp. The lipoprotein travels through the membrane in order to bind to other FhuA receptors, inhibiting further infection of the cell. This figure was reproduced with authorisation from Séraphine Degroux (IBS).

5.1.5 *E. coli* receptor FhuA

FhuA is a protein of 78.9 kDa located on *E. coli* outer-membrane, transporting ferrichrome coupled ions. It is the receptor of both some antibiotics as well as some phages.

This receptor structure is composed of 22-stranded β-barrels, extracellular hydrophobic loops, and an additional domain called *the cork*, within the β-barrel (see Figure 5.3). As

investigated by Cécile Breyton's team (IBS) by Differential Scanning Calorimetry (DSC), ferrichrome ions seems to affect the cork and loops part of FhuA by stabilising them. Removal of the cork however had the inverse effect, with no noticeable improvement while adding ferrichrome.

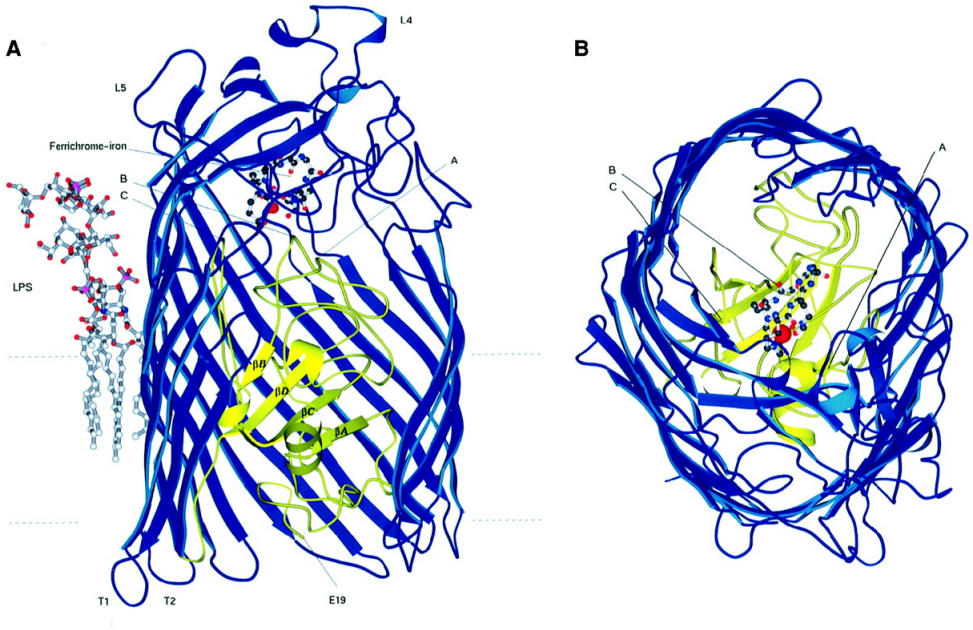

Figure 5.3: FhuA structure, from [178]. Reprinted with permission from AAAS.
The barrel structure (blue) is surrounding the cork (yellow). The transported iron ion is depicted as red sphere.

It has also been shown that both octylglucoside-solubilised FhuA combined with Llp and Llp co-reconstituted into liposomes prevents T5 from binding to FhuA, without knowing any further detail about the underlying mechanism.[181]

To summarise, the RBP pb5, the host receptor FhuA and the lytic conversion protein Llp have all been identified and can be synthesized, which is advantageous to pursue our study. T5 is also representative of one of the most populated families of phages and understanding its mechanism of action would be a great advance towards the comprehension of the bacteriophage world, that may help phagotherapy development.

5.2 Methods

5.2.1 Collaboration

This project is a collaboration between Loïc Salmon's group at CRMN (Lyon, France) and Cécile Breyton's group at IBS (Grenoble, France).

Here is a list of the work presented in this thesis that has been realised by Séraphine Degroux (IBS): synthesis and purification of the proteins, characterisation by CD and MS, NMR samples preparation, investigation of proteins K_D by SPR/MST/BLI, manual assignment of Llp and mutants study. Structure calculation has also been done at the IBS. Descriptions of methods and results are given in the following parts.

5.2.2 Protein synthesis and purification

Llp was synthesized in 2 different versions: acylated (Ac-Llp) and soluble (sol-Llp). While the first cysteine is part of the sequence in the Ac-Llp, it is absent from the soluble form sol-Llp.

To obtain enough soluble and functional Llp protein two strategies were tested. Llp gene was cloned without N-terminal cysteine and signal sequence or fused with MalE genes separated by cleavage site TEV[178]. This gene codes for the periplasmic Maltose-Binding Protein (MBP), which is often used to increase the solubility of proteins. The expression of Sol-Llp in cytoplasm of BL21 was very low and not used. Llp-MBP fusion protein is convenient to prevent aggregation of Llp. This fusion was expressed in sufficient yields to be structurally characterised. MBP trap column was used to purify Sol-Llp. After cleavage of the MBP, pure monomeric and soluble Llp was obtained.

Circular Dichroism (CD) showed that sol-Llp mainly contains β structures, which attests that the proteins indeed folds, while Mass Spectrometry attested the presence of a disulfide bridge between Cysteine 11 and Cysteine 58.
NMR experiments were all recorded on sol-Llp.

Additionally, detergent solubilised proteins were used, as membrane proteins need to be kept in a percentage of detergent superior to their Critical Micelle Concentration (CMC). Here, FhuA sample contains 0.01% DDM.

5.2.3 Preliminary investigations

All NMR experiments were acquired using either Bruker Avance III 600 MHz and 1 GHz spectrometers, equipped with cryogenically cooled triple resonance (^{1}H, ^{13}C, ^{15}N) probes. All NMR data were processed using NMRPipe program. The CCPNMR program was used

to display and assign 2D and 3D NMR spectra.

Data were recorded on a uniformly ^{13}C, ^{15}N-labelled sample of sol-Llp in 25 mM Tris (pH 6.5), 75 mM NaCl buffer, 10 % D_2O.

^{1}H, ^{15}N-HSQC spectra were recorded at 600 MHz over a range of temperatures (293 to 303 K), to determine the most appropriate one.

Generally speaking, the spectral dimensions were around 1522 Hz (F1) and 9615 Hz (F2), corresponding to 25 x 16 ppm, for acquisition times around 84 ms and 106 ms respectively. RF carriers were usually at 4.70 ppm for ^{1}H and 118.5 ppm for ^{15}N, and recycling delay set at 1 s.

The stability of sol-Llp (T_m = 337 K) was tested using Mass Spectrometry (MS) on samples let at 293 K and 277K, over several weeks, from pH 4.9 to 10. The protein expected mass is 7120 Da.

^{1}H, ^{15}N-HSQC were acquired at 303 K at 600 MHz, at 8 days interval (between APSY experiments). Spectral dimensions were identical as previously fixed.

5.2.4 Protein assignment strategy

Resonance assignment of common protein nuclei (^{1}H, ^{13}C, ^{15}N) is the first step towards the structural and dynamic studies by NMR spectroscopy. Different nuclei act as probes inside the protein and allow access to information about the studied molecule itself, but also about its interaction with other molecules.

In NMR experiments, magnetisation is transferred through the excited proton to the nearby atoms through covalent bonds. This magnetisation is then transferred back to protons for detection. The evolution of chemical shifts of each edited nuclei is leading to a dimension in the final spectrum.

Usually, 2D or 3D experiments are acquired to connect the ^{1}H and ^{15}N amino groups to the protein's backbone and side-chains ^{13}C chemical shifts within a residue. This multidimensional approach also allows to associate an amino acid and its adjacent neighbours in the protein sequence. Indeed, the edited ^{1}H, ^{15}N and ^{13}C resonances can belong either to the same (intra) residue and/or to the i-1 (previous) residue.

For example, the magnetisation in HNCα experiment is passed from a proton to its bound ^{15}N and to the Cα by J-coupling. However, since the nitrogen has a high enough coupling with both intra and preceding Cα, the chemical shifts of both carbons are visible on the spectrum. Besides, since the coupling to the intra residue is higher than the one with the one located in the i-1 residue, those intra-residue peaks have a higher intensity than the i-1 ones. In the Cβ experiments, both chemical shifts of Cα and Cβ are evolving

at the same time since their frequency is within the same range, and their resonances are therefore all visible in Cβ spectra.

It is important to notice that the Cβ chemical shifts are more sensitive to the side-chain chemical environment and are therefore useful indicators of the residue type.

Here, a set of experiments has been recorded in order to determine the resonances of the backbone atoms and the side-chains alpha and beta carbons: HNCO, HNcaCO, HNCA, HNcoCA, HNcaCB, HNcocaCB.
Spectral dimensions for ^{1}H (F3) and ^{15}N (F2) were identical as previously fixed, the ^{13}C dimension (F1) was 4228 Hz (28 ppm) for HNCA and HNcoCA; 2114 Hz (14 ppm) for HNCO; 3322 Hz (22 ppm) for HNcaCO; 11326 Hz (75 ppm) for HNcaCB and HNcocaCB. RF carriers were placed at 4.70 ppm for ^{1}H and 118.5 ppm for ^{15}N, and for ^{13}C, at: 54.5 ppm in HNCA and HNcoCA; 173 ppm in HNCO; 176 ppm in HNcaCO; 39 ppm in HNcaCB and HNcocaCB.

5.2.5 Structure determination

Additionally to the resonance assignment, the tertiary structure of the protein can be determined through trough-space correlations. Using NOESY sequences, it can be determined if a proton is in proximity to another one, depending on the mixing time of the experiment, and thus allowing to place protons within a range of distance to one another. All those restraints are then used to calculate the structures of the protein that fit these parameters the best and establish a set of different possible structures for Llp. Structure calculations were done using UNIO software and the following set of NOESY-HSQC: ^{1}H,^{15}N-NOESY-HSQC; ^{1}H,^{13}C-NOESY-HSQC; ^{1}H-aromatic,^{13}C-NOESY-HSQC; and HCCh-TOCSY.

Spectra were recorded at 303K, at 1 GHz. The spectral dimensions were 13889 Hz (F1 and F3) and 2534 Hz (F2), corresponding to 13.9 x 25 ppm in the ^{1}H,^{15}N-NOESY-HSQC; 13889 Hz (F1 and F3) and 20123 Hz (F2), corresponding to 13.9 x 80 ppm in the ^{1}H,^{13}C-NOESY-HSQC and ^{1}H,^{13}C-NOESY-HSQC (aromatic); 13889 Hz (F3) and 20123 Hz (F1 and F2), corresponding to 13.9 x 80 ppm in the HCCh-TOCSY. All NOESY-HSQC experiments were recorded with a mixing time of 120 ms. The scheme for homonuclear ^{13}C decoupling in TOCSY was dipsi3.

5.2.6 Spin relaxation

Data were recorded on a uniformly ^{15}N-labelled sample of sol-Llp in 25 mM Tris (pH 6.5), 75 mM NaCl buffer, 10 % D2O.

^{15}N relaxation spectra were acquired at 303 K on a Bruker Avance III 600 MHz spectrometer, equipped with a TXI Cryoprobe 5mm HCND, to get a glance at Llp dynamics. As previously mentioned, T1, T1ρ, T2 and heteronuclear NOEs give information on motions in the ps-ns timescale.

Different delays were used to record relaxation time measurements, in a randomised manner, with repetition of one value, in order to assess the reproducibility of the experiments: 0, 400, 160, 320, 560, 400, 80, 720, 1120, 880 and 240 ms for R1 delays; 120, 10, 60, 250, 40, 160, 90, 1, 60, 200, 20 ms for R1ρ.

For R1ρ measurements, the radiofrequency (RF) spin-lock was set at 1.4 kHz. Llp was labelled on ^{1}H and ^{15}N only so no ^{13}C decoupling was applied. 16 scans per FID were recorded, with a recovery delay between scans of 2 s.

The ^{15}N-^{1}H NOE and reference spectra were recorded in an inter-leaved approach. Carriers were kept identical to the relaxation experiments, but 96 scans were recorded. A recovery delay of 5 s was chosen, and an additional delay of 3 s was added in presence or absence of saturation.

Spectra were processed and analysed using the NMRPipe software package[182]. Spectral intensities and relaxation rates were extracted with NMRPipe scripts. Both pulse sequences and NMRPipe processing scripts are available online through Bax's group library[50]. Analysis of relaxation rates and error calculations through Monte-Carlo simulations were performed using Matlab scripts written in-house.

The software *Tensor*[37][183] was used to interpret the relaxation data using the model-free formalism from Lipari & Szabo[33]. This study was performed prior to the tRNA one described in a previous chapter. As the *Tensor* software is not fitted for RNA analysis, due to non-accessibility of key parameters required to be changed such as the chemical shift anisotropy, Palmer's *modelfree* software was there preferred.

Tensor is based on relaxation theory, and can fit either an isotropic or an anisotropic tumbling. The rotational diffusion tensor is determined, using residues that are not located near the terminus and do not experience exchange based on R1ρ values observation nor decreased hetNOEs values: usually the residues are located in structured parts of the protein. Once the tensor is fixed, Monte-Carlo simulations are implemented to test the pertinence of the tensor. Different models are used to analyse internal motions depending on the timescales of the motions (and if they can be separated from one another), as well as the exchange contribution that the backbone NH vectors might be experiencing.

Here, the ^{1}H-^{15}N distance was fixed at 1.02 angstroms. An isotropic diffusion tensor was used, calculated using data from residues in β structural parts of the protein, that did not present increased R1ρ values nor decreased hetNOE values (< 0.83). Experimental relaxation data can either be fit using an isotropic tensor, where the rotational diffusion is

independent of the NH vectors orientations, or an anisotropic one.

Order parameters, correlation time and exchange constants are then calculated throughout Llp sequence. Order parameters of ^{1}H-^{15}N vectors are restrained between 0 and 1 values, 1 being the most rigid limit while 0 is the most flexible.

Different models are used to fit the experimental data: from 1 to 5, with an increasing number of parameters. In particular, model 1 fits an order parameter while the internal correlation time of the motion (τ_i) is considered to be under 20 ps (clearly faster than the overall tumbling) while model 2 is the classic Lipari-Szabo analysis and fits this τ_i. Model 3 incorporates an exchange rate R_{ex} to the model 1, while model 4 does the same to model 2. Model 5 includes an additional correlation time on a slower timescale, that can be separated, as in the extended Lipari-Szabo model.

F-tests are then used to determine if a significant improvement rises from adding parameters, and thus increase the model number and the parameters to be fitted to the experimental data.

5.2.7 Titration of the Llp-FhuA interaction

As stated previously, FhuA is a membrane protein that requires detergent to remain in a monomeric soluble form. Here, as DDM is the chosen detergent, its influence on Llp should first be investigated. The samples of sol-Llp in absence of DDM (in the following parts designated as *Llp in buffer*) and sol-Llp in presence of DDM (*Llp in DDM*) have to be compared. ^{1}H,^{15}N-SOFAST-HSQC are recorded to investigate the detergent influence on the lipoprotein.

Moving on to the characterisation of the interaction between Llp and FhuA, a titration is performed, using NMR spectroscopy. This method involves the analysis of chemical shifts (CS) and peak intensity variations between ^{1}H,^{15}N-SOFAST-HSQC spectra on samples with an increasing [Llp]/[FhuA] ratio.

DDM is added to Llp stock solution so that the final % of Llp and the final % of DDM in solution were not evolving independently during the titration: as the amounts of Llp are added into the NMR tube containing the FhuA sample, DDM is also added.

A total of 10 points were recorded, separated into 2 sub-titrations (see Table 5.1). The samples were mixed using stocks of [Llp]=1.2 mM and [FhuA]=0.15 mM.

^{1}H,^{15}N-SOFAST-HSQC were recorded for each Llp addition. Peak intensities were normalised by both number of scans in experiment and Llp concentration.

To characterise an interaction between compounds, one of the parameters that can be monitored is the equilibrium dissociation constant (K_D). It represents the equilibrium between a protein (P) and a ligand (L) in their free and complex (PL) form. The ligand can

V_{Llp} / μL	V_{FhuA} / μL	V_{total} / μL	LLP/FhuA / %
5	110	130	0.36
12	110	137	0.87
20	110	145	1.45
30	110	155	2.18
25	65	130	3.08
32	65	137	3.94
40	65	145	4.92
50	65	155	6.15
60	65	165	7.38

Table 5.1: Titration Llp-FhuA

be, such as in this case, another protein.
Basically, K_D can be described using the proteins respective concentrations and stoichiometry:

$$\begin{aligned} xA + yB &\rightarrow A_xB_y \\ K_D &= \frac{[A]^x \times [B]^y}{[A_xB_y]} \end{aligned} \tag{5.1}$$

This value can be investigated by NMR, as well as many other techniques, for instance Surface Plasmon Resonance (SPR), MicroScale Thermophoresis (MST) and BioLayer Interferometry (BLI).

5.3 Results and discussions

5.3.1 Temperature investigation

The behaviour of the lipoprotein over a temperature range (293 K to 303 K) was checked, and no major change in the spectrum other than the temperature effect on chemical shifts was observable (see Figure 5.4). The following experiments were therefore recorded at 303K. Indeed the tumbling is increased at 303K compared to working at cooler temperatures, thus leading to longer transverse relaxation, and hence narrower (more intense) peaks.

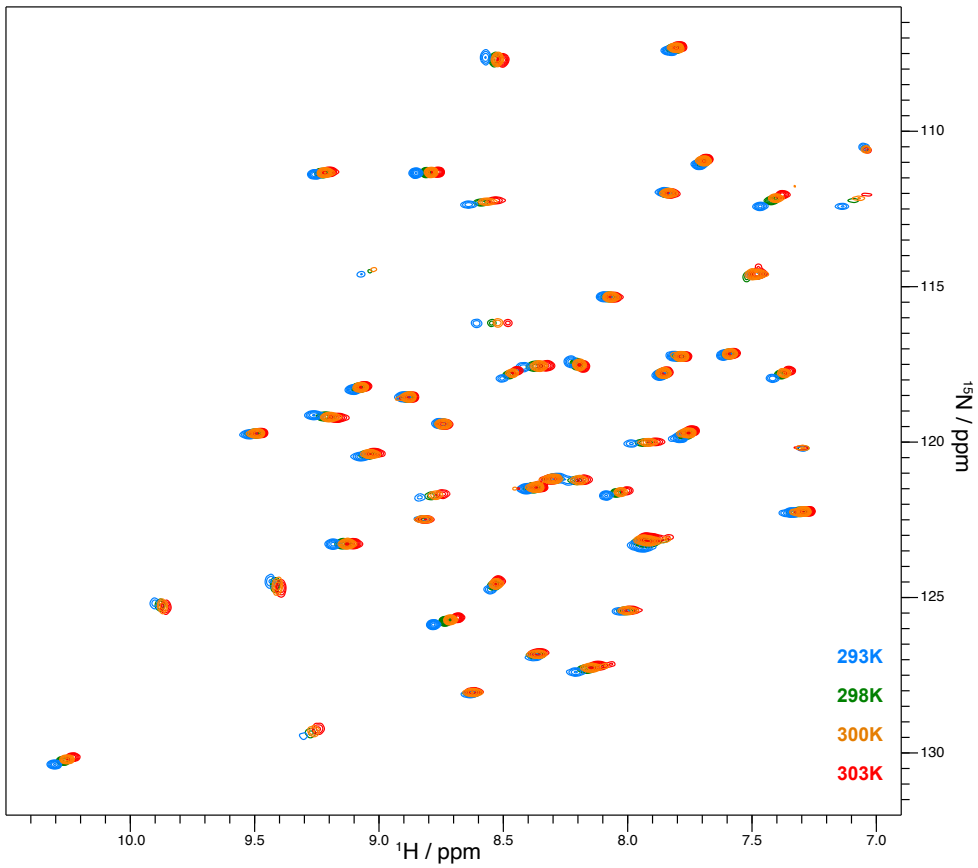

Figure 5.4: Llp temperature influence investigation.
Temperature effect on Llp was checked over a range from 293K to 303K, using ^{1}H, ^{15}N-HSQC spectra recorded at 600 MHz.

5.3.2 Stability study

Mass Spectrometry (MS) analysis on sol-Llp right after its purification and after being stored at 277 K present a mass of 7118.6 Da in both cases, which is the expected value for the soluble protein.

When Llp is conserved at 293 K, MS shows the apparition of 2 peaks at mass of 7060.5 Da (-58.10 Da relative to 7118.6 Da) and 6972.47 Da (-88.02 Da relative to 7060.5 Da). The hypothesis is that the protein is losing a glycine and a serine, which may be the ones located at the N-terminal, or within a non-structured part of Llp.

The soluble form of Llp is therefore rather stable as long as the sample is stored at 253

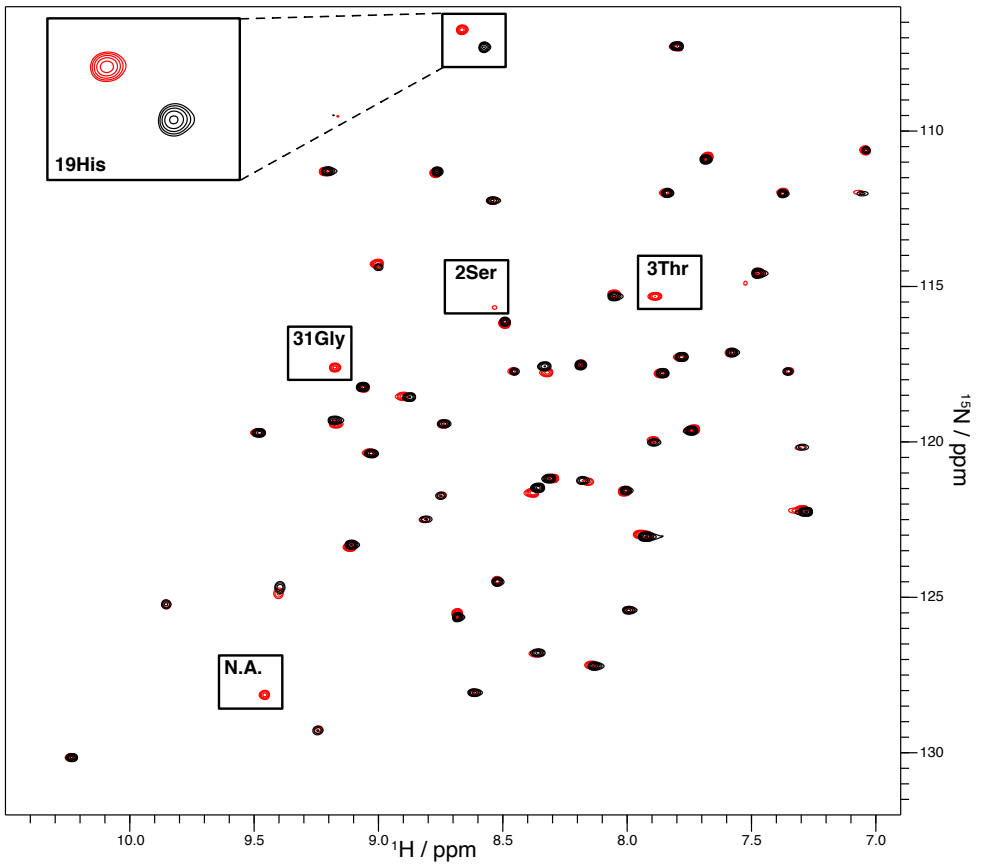

Figure 5.5: Llp stability investigation.
^{1}H, ^{15}N-HSQC were acquired at 303K at 600 MHz: at the start of experiments (black) and after 8 days interval (red), between experiments. Most visible differences are highlighted (residue type and number, N.A. being non-assigned).

K or 277 K, but undergoes some degradation for longer periods at room temperature (RT).

At 8 days interval at 303 K (see Figure 5.5), some chemical shift variations are noticeable as well as some peaks appearance and disappearance in different regions: residues around N-terminal (Ser2 and Thr3) and in the first half-part of the protein (His19 and Gly31). These residues are located in different unstructured parts of Llp.

5.3.3 Llp sequence assignment

As stated previously, the lipoprotein's chemical shifts assignment strategy was based on 2D and 3D spectra of covalently bound atoms. The wide spreading of chemical shifts and the degradation undergone by the protein at temperatures superior to 277 K were two reasons not to edit further dimensions.
Llp assignment results from the connection between the ^{1}H and ^{15}N chemical shifts to the ^{13}C chemical shifts, both intra-residue and to the previous one, as shown and described in Figure 5.6.

The sol-Llp protein contains 58 non proline amino acids. Most of them are visible on 2D ^{1}H,^{15}N-HSQC spectra. 47 of these amino acids have been assigned, resulting in 76% completed assignment (see Figure 5.7).

Additional amino acids may be assigned in the n°38-48 region. However, peaks corresponding to their possible chemical shifts are visible in some of the 3D spectra while absent from some others. Most of these chemical shifts also correspond to really weak peaks in the ^{1}H,^{15}N-HSQC spectrum, that could barely be discerned from the spectrum noise. Moreover, this region also contains adjacent serine (n°44) and glycine (n°43), which could also be the ones involved in the protein degradation. For all of those reasons, we chose to discard them from any further analysis.

5.3.4 Llp structure

Llp structure could be calculated using the set of NMR experiments recorded (see Figure 5.8.

The determined structure resembles a SRC Homology 3 (SH3) domain (see Figure 5.9), with less structured part, especially one of the loops in Llp would correspond to a β strand in SH3.

5.3.5 Llp dynamics

Intramolecular motions are determinant for biomolecules functions. Therefore, Llp dynamics have been investigated using spin relaxation, in order to obtain information on the motion of backbone NH bond vectors.

Intensities of each assigned residue are easily extracted from NMR spectra due to their chemical shift dispersion. Relative intensities are then plotted against relaxation delays

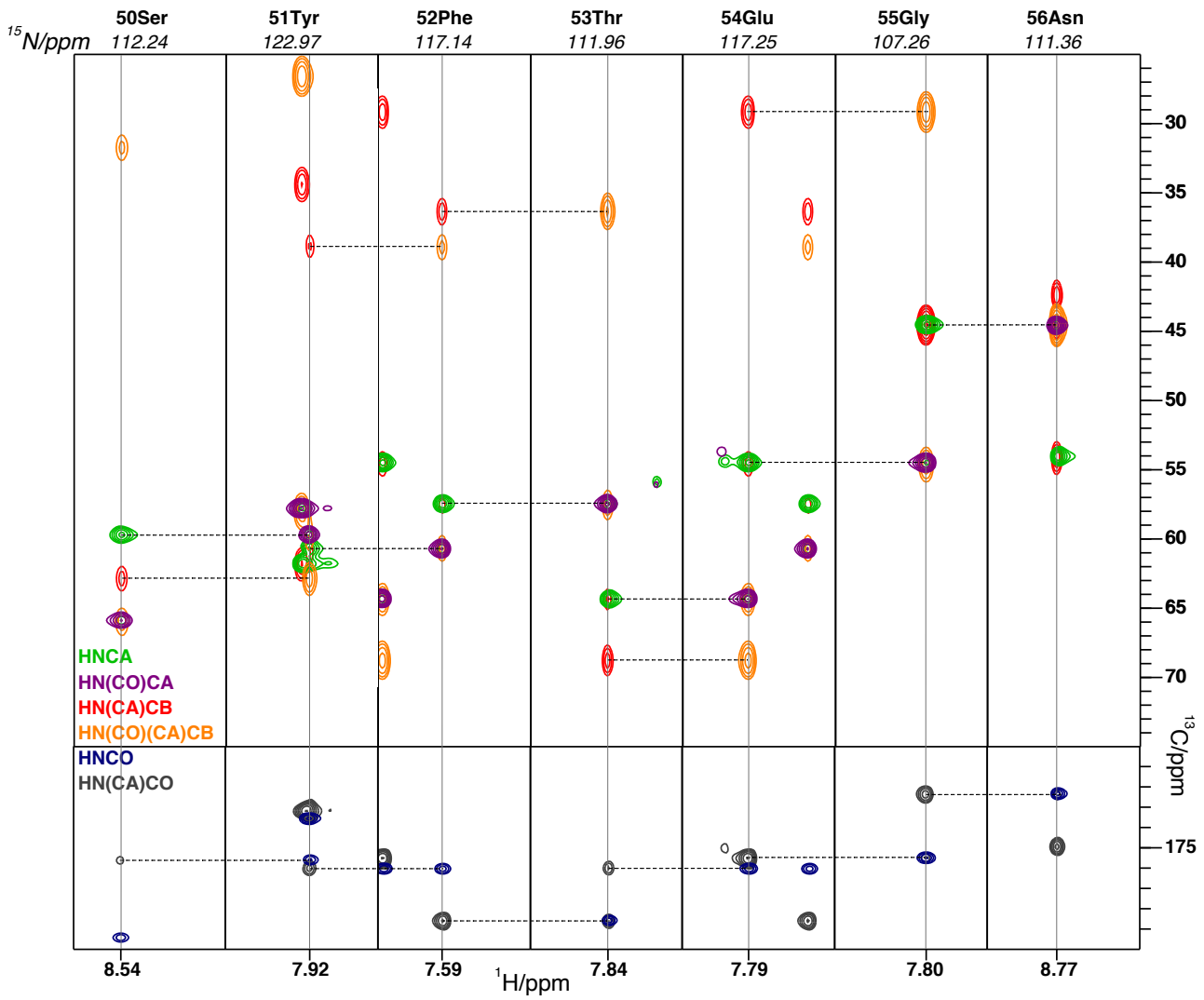

Figure 5.6: Llp assignment strategy for sol-Llp residues S50-N56.
Spectra strips (colour legend on figure) contain inter- and intra-residue Cα, Cβ and CO resonances for each NH groups, which ^{15}N frequency is written in italic. Dotted lines show the connexion between spectra while full vertical grey lines show the corresponding ^{1}H frequency observed in each strip.

during spectra acquisition (see Figure 5.10).

Relaxation rates R_1 and $R_1\rho$ at 600 MHz for Llp exhibit decreasing values around the

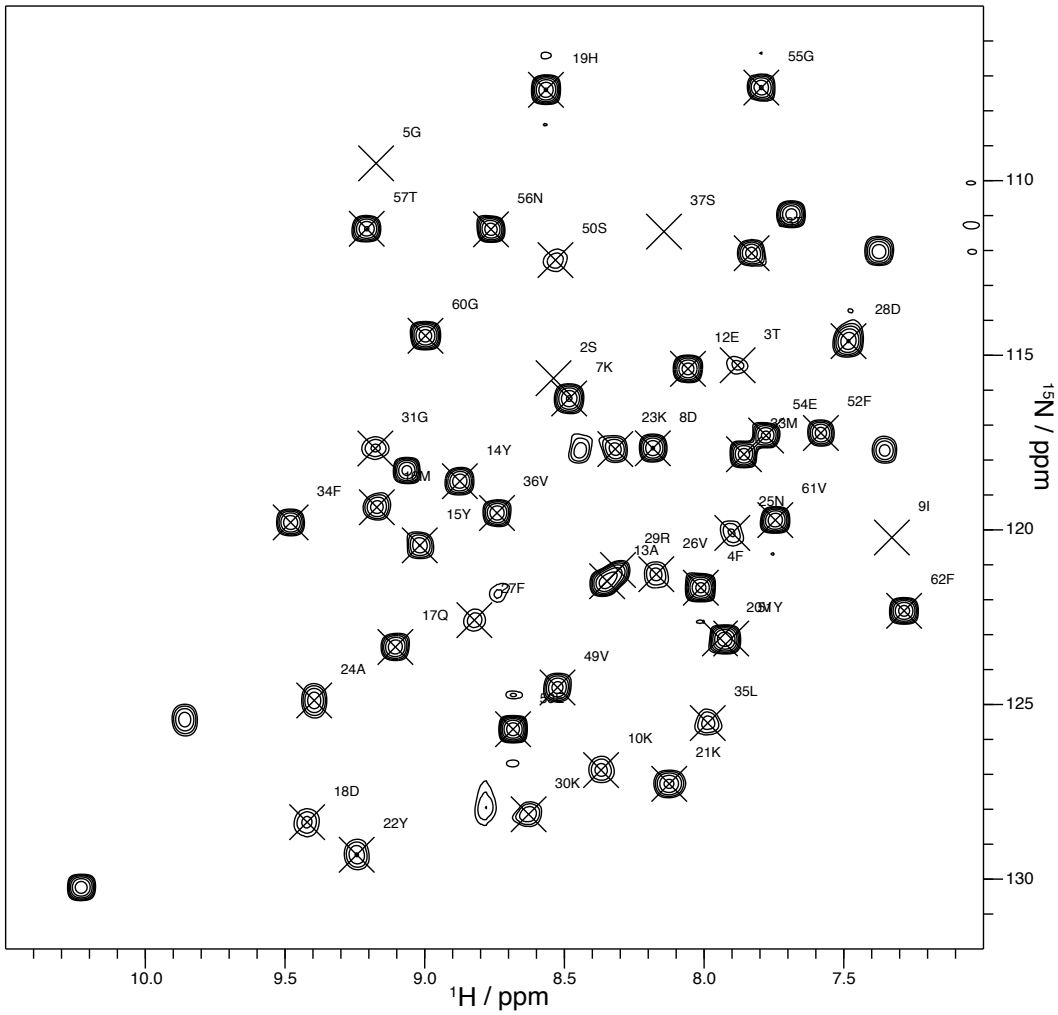

Figure 5.7: Llp ^{1}H,^{15}N-HSQC spectrum, with assignment.
Spectrum was recorded at 600 MHz and 303 K. Each peak position provides ^{1}H and ^{15}N chemical shifts for the protein amino groups. 47 of the 58 non-proline amino acids have been assigned, resulting in 76% completed assignment.

C-terminal and N-terminal parts of the protein (see Figure 5.11 (a) and (b)). Tyr22 and Ala24 seem to undergo exchange, which is noticeable through their increased R1ρ values (and therefore also R2); His19 presents a particularly high R1ρ value, but its interest will not be discussed in details as this residue is in the protein part facing degradation changes.

As expected, Llp terminal parts arbor lower hetNOEs values, due to their extreme flex-

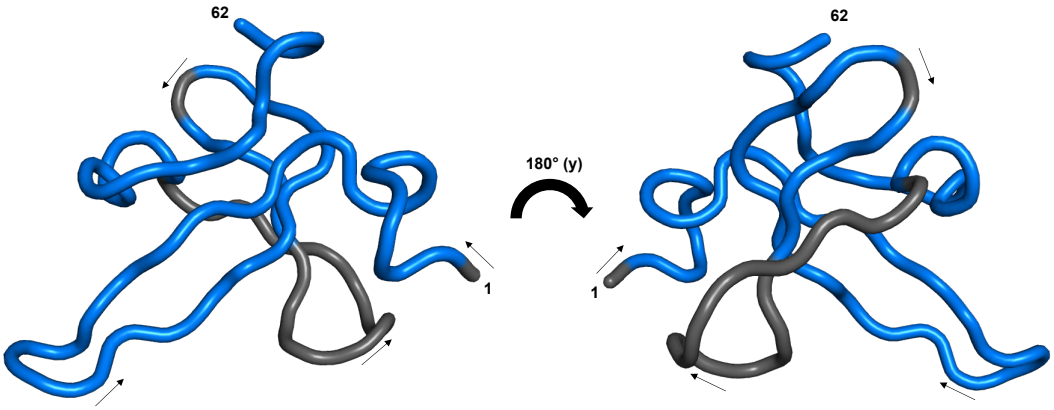

Figure 5.8: Llp structure.
Residues assigned are highlighted (blue) and arrows indicate the direction of increasing residue number in Llp sequence.

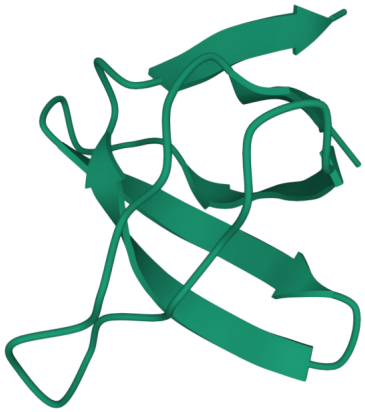

Figure 5.9: SH3 domain.
PDB ID: 1SHG. The structure is composed of β strands arranged in anti-parallel β sheets, connected by short helices.

ibility (see Figure 5.11 (c)). There is a slight decrease in residues n°17 to 25: the average value for residues n°17 to 25 is 0.814 while the average value for residues n°11 to 58 is 0.840 (residues n°17-24 excluded, and purposely removing residues in N- or C-terminus).

The relaxation data was then analysed using *Tensor*. After 1000 Monte-Carlo simulations, the experimental data gives a χ^2 about 18, which is below the χ^2 obtained using

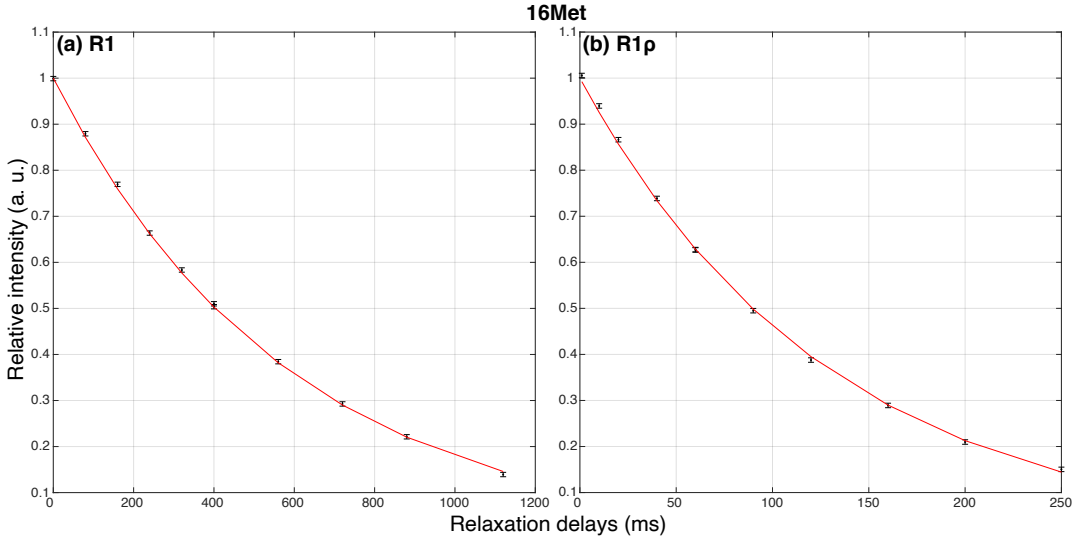

Figure 5.10: Intensity fluctuation of 16Met in ^{1}H,^{15}N-HSQC spectra upon relaxation delay variation, for (a) R1 and (b) R1ρ, at 600 MHz and 303 K.

simulated data (30) using 95% confidence limits: the model is fitting well the data. The anisotropic tensor is still tested for both axial symmetry and full asymmetry. As the F-tests show no significant improvement, the internal mobility analysis can be started, using an isotropic tensor.

The global correlation time of Llp τ_C, calculated using R2/R1 ratios, is about 6.3 ns.

The order parameters average is about 0.88, which is in accordance with the values usually obtained for proteins. Most part of the structure is therefore quite rigid, while the S^2 are decreasing rapidly near the N- and C-terminus due to the flexibility of the chains. A few residues also experience exchange: His19, Tyr22 and Ala24, as previously observable through R1ρ values, as well as Met33.

However, surprising results emerge in the loop regions of Llp. Indeed, relaxation data analysis displays an abnormally lacking flexibility in this non-structured part (residues n°14-24 essentially).

To investigate further the behaviour of this loop, two different aspects could be covered. Residual Dipolar Couplings (RDC) measurements and NOESY-HSQC with different mixing times could be measured on one part to rise the structural accuracy of this region. Complementary relaxation measurements would also be needed in order to get a better overview of the protein's dynamics such as different magnetic fields/temperatures, and/or NMR experiments probing other timescales.

Additional measurements would also be need to characterise the degradation of the

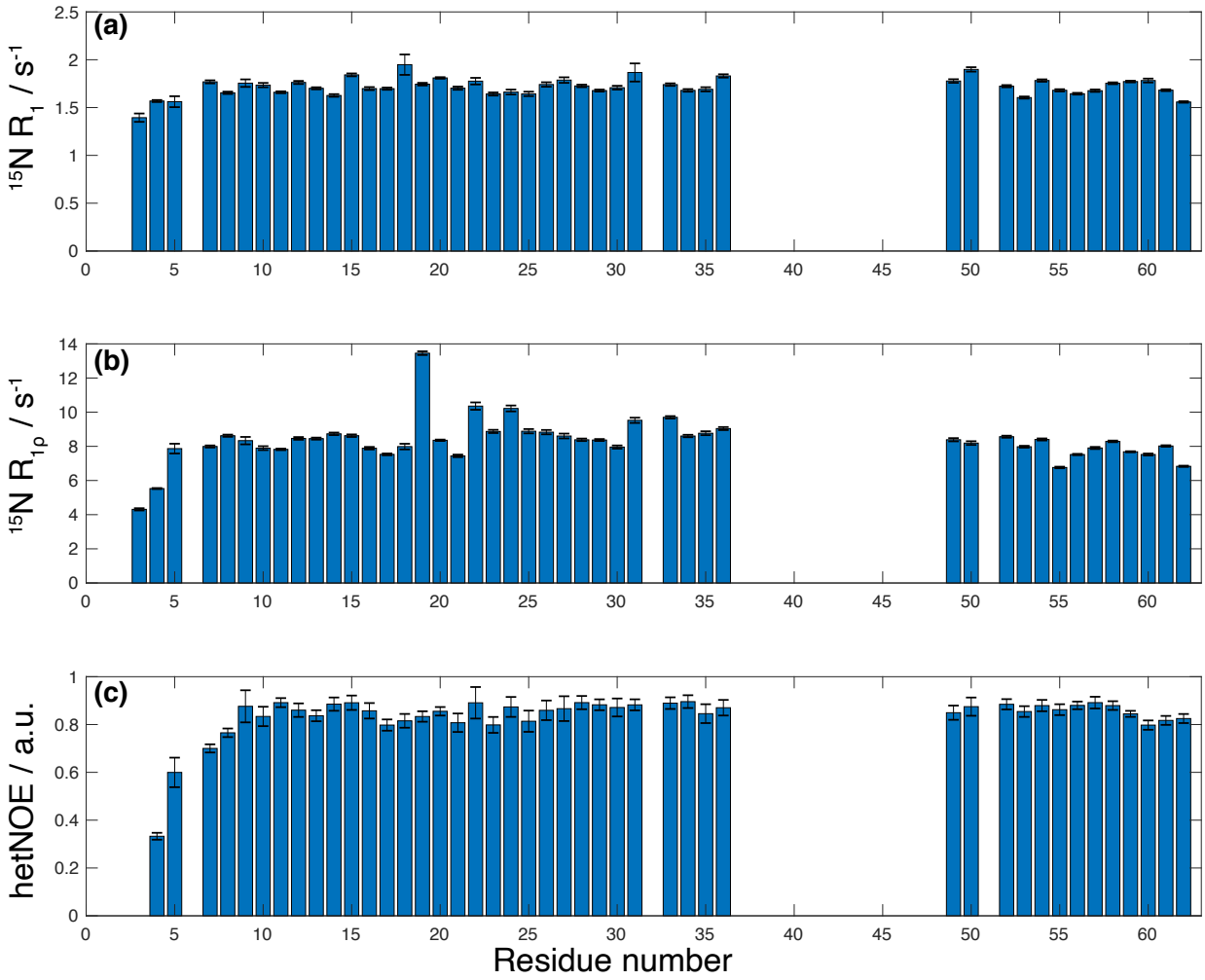

Figure 5.11: Llp site-specific ^{15}N relaxation data at 600 MHz and 303K.
(a) Longitudinal relaxation rates R1; (b) Spin-lattice relaxation rates in the rotating frame R1ρ; (c) Heteronuclear ^{1}H,^{15}N-NOE.

protein, to differentiate its effect from the protein dynamics. However, as the degradation was observed for the Histidine 19, the other residues experiencing exchange are all located in the same region (see5.12 (C)), which indicates a motion on a slower timescale in this part of Llp.

Finally, the uncertainty of the assignment leading to the absence of data in one of the loop region is not unusual when residues experience a high flexibility but is particularly

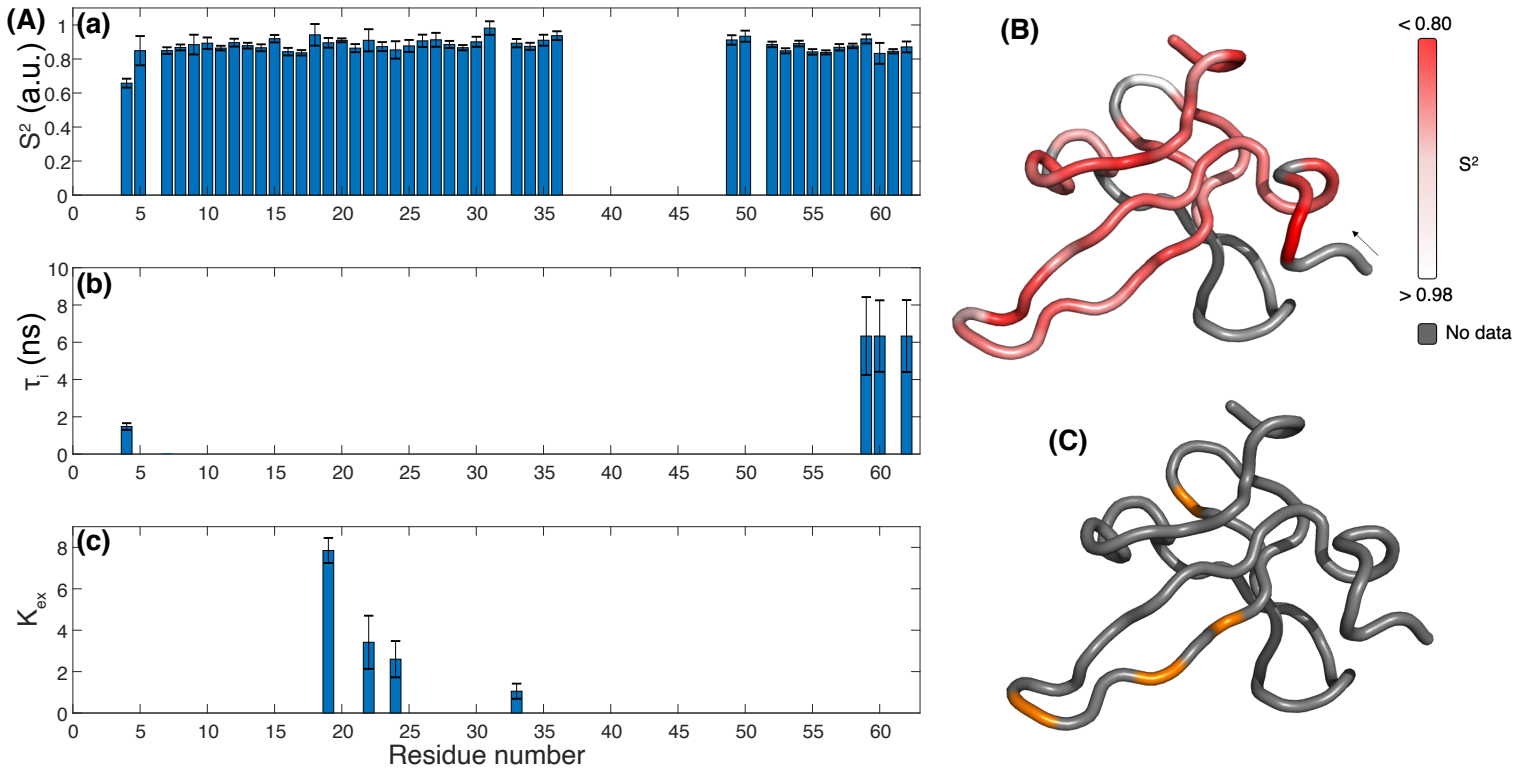

Figure 5.12: (A) sol-Llp model-free analysis using Tensor software: (a) order parameters for each backbone ^{1}H-^{15}N vector; (b) corresponding τ_i when fitted; and for models 3 or 4 (c) the corresponding exchange K_{ex}.
(B) Order parameters repartition on Llp structure, the redder the more flexible is the residue (grey for no data); the arrow indicates the N-terminal.
(C) Residues experiencing exchange are highlighted in orange on the structure.

inconvenient for a protein of this size.

5.3.6 Llp-FhuA interaction

The interaction between Llp and FhuA was firstly probed using biomolecular interaction techniques. K_D was investigated for Ac-Llp with FhuA in DDM (n-dodecyl β-D-maltoside) but the detergents prevented data interpretation.
Concerning the solubilised form of Llp, the K_D was determined to be around 1.5 mM using MST, while signal was also hidden by the detergents using SPR.

Using NMR spectroscopy, determining whether the interaction between Llp and FhuA occurs within the sample is a crucial step before progressing any further, the main task

would then be to identify the protein residues involved in the interaction using chemical shift and intensity variations, and finally, if possible, determine the K_D and compare it to the one obtained by SPR.

Chemical shifts perturbations as well as intensity changes can be observed throughout the titration.

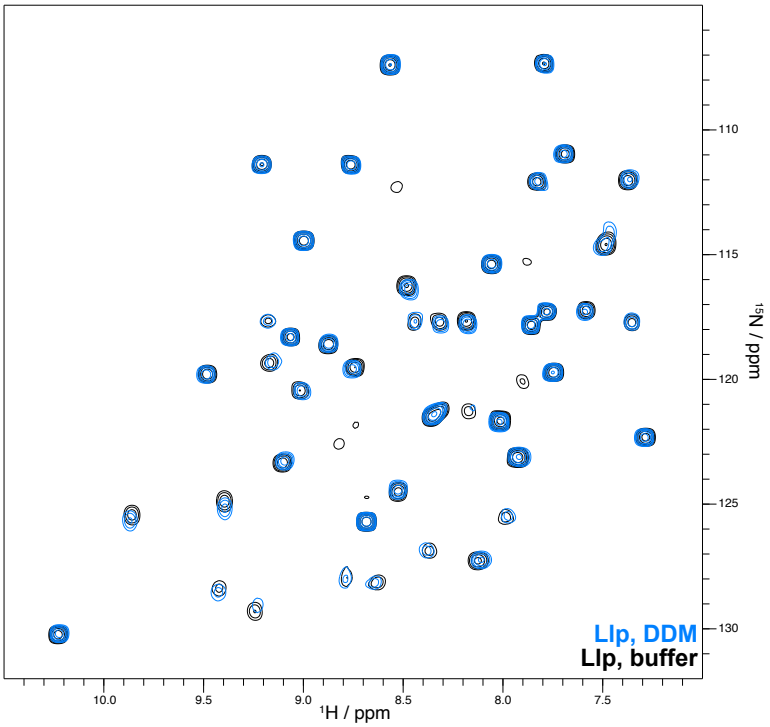

Figure 5.13: ^{1}H,^{15}N-SOFAST-HSQC on Llp: (blue) in DDM added to NMR buffer; (black) in NMR buffer only.

Through direct observation on the spectra (see 5.13) and further analysis exploiting intensities and chemical shifts, the influence of the detergent on Llp can be assessed.

Indeed, as most of the protein seems slightly affected by the presence of detergent (small CS changes and global intensity variation), there are still some regions that may be in direct interaction with the DDM. Figure 5.14 (b) clearly exhibits that many residues in the region from n°22 to 28 experience a significant decrease in intensity. Chemical shifts changes are happening on a small-scale (see Figure 5.14 (a)), for the residues n°20-25, which may be due to an interaction with the detergent or affected by the interaction of nearby residues as observed through intensity analysis. Met16 may also be sensitive to DDM considering

its intensity decay.

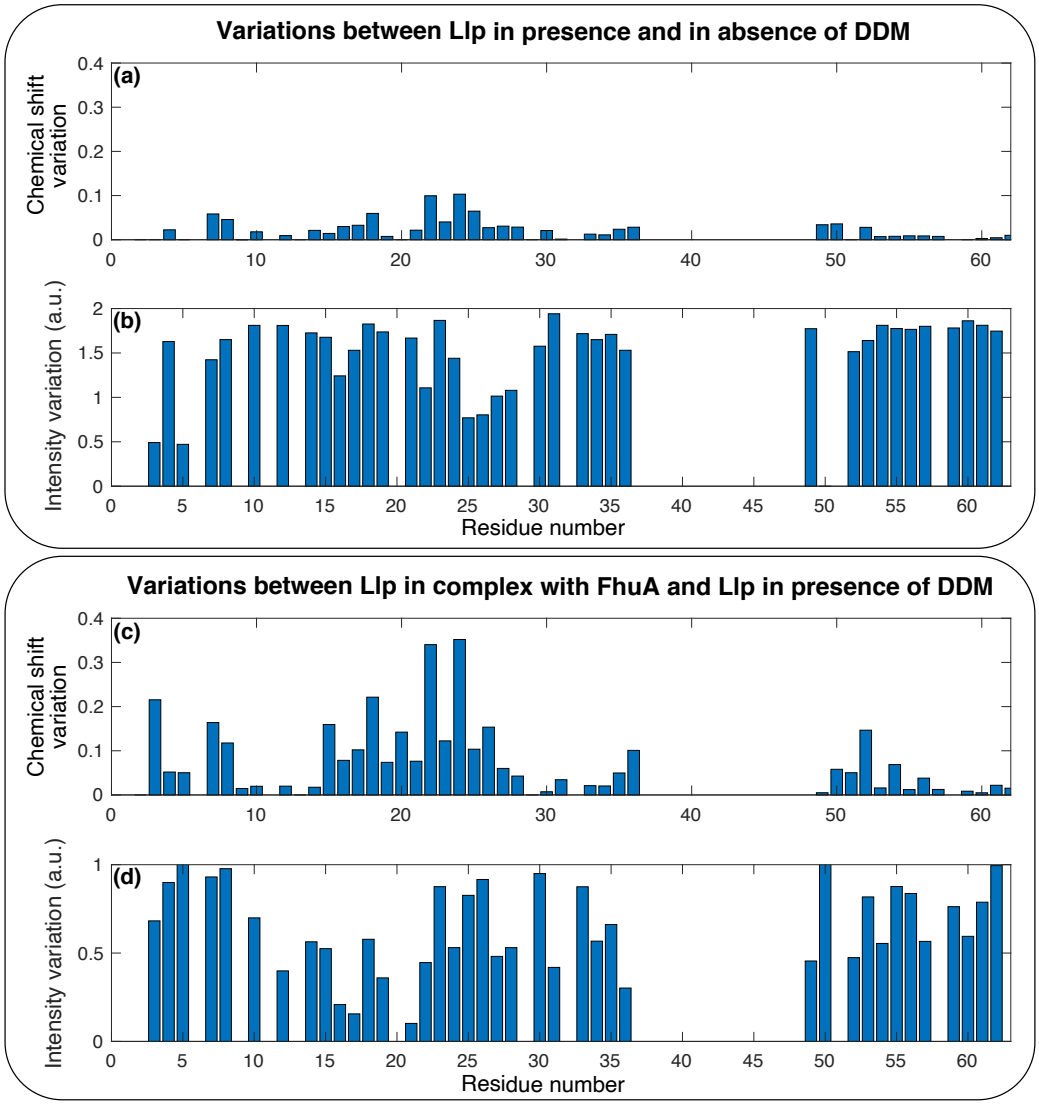

Figure 5.14: Investigation of the DDM influence (above) and FhuA influence (below) over Llp, through ^{1}H,^{15}N-SOFAST-HSQC analysis.
(a) Chemical shifts variations and (b) Intensity variations between Llp in presence and absence of DDM.
(c) Chemical shifts variations and (d) Intensity variations between Llp-FhuA HC and Llp in DDM.

Once the influence of detergent was investigated, the titration of the interaction between Llp and FhuA could be started. From this point, the effect of the complex Llp-FhuA is to be discussed: the titration point with the highest percentage of Llp-FhuA complex is designated as *Llp-FhuA HC* and the lowest percentage of complex by *Llp-FhuA LC*.

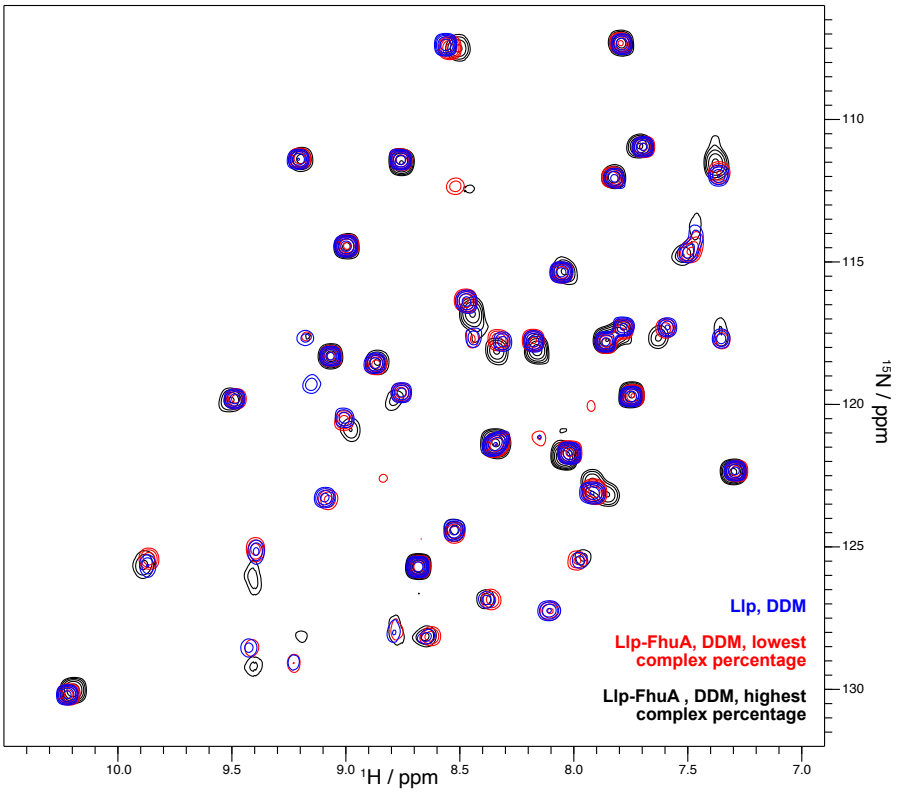

Figure 5.15: ^{1}H,^{15}N-SOFAST-HSQC on Llp samples: in presence of DDM (blue), Llp-FhuA LC (red), and Llp-FhuA HC (black).
There are slight observable differences between spectrum of Llp in DDM and Llp-FhuA LC, but the clearest ones are noticeable when comparing Llp-FhuA HC, especially in the central and bottom parts of the spectrum.

As predictable, the global spectrum of Llp (see Figure 5.15) undergoes a decrease in intensity while in contact with FhuA: the complex formed increase the size of the system, leading to a slower tumbling, which reduces transverse relaxation and broaden peaks on the NMR spectra.

The influence of DDM is still observable, however some regions show even greater in-

tensity reduction and CS variations, such as the residues n°12-22 as well as residues n°51-54 located near the C-terminal region (see Figure 5.14 (c) and (d)). As reminder, the residues n°14-24 form an external loop of Llp, which would be an accessible spot to interact with FhuA.

The CS variations were analysed for each spectrum acquired upon Llp addition during the titration. Considering the well isolated peaks and the slowly graduated variation observable, the chemical shift assignment was rather straight-forward, without the use of re-assigning the affected parts of the protein.

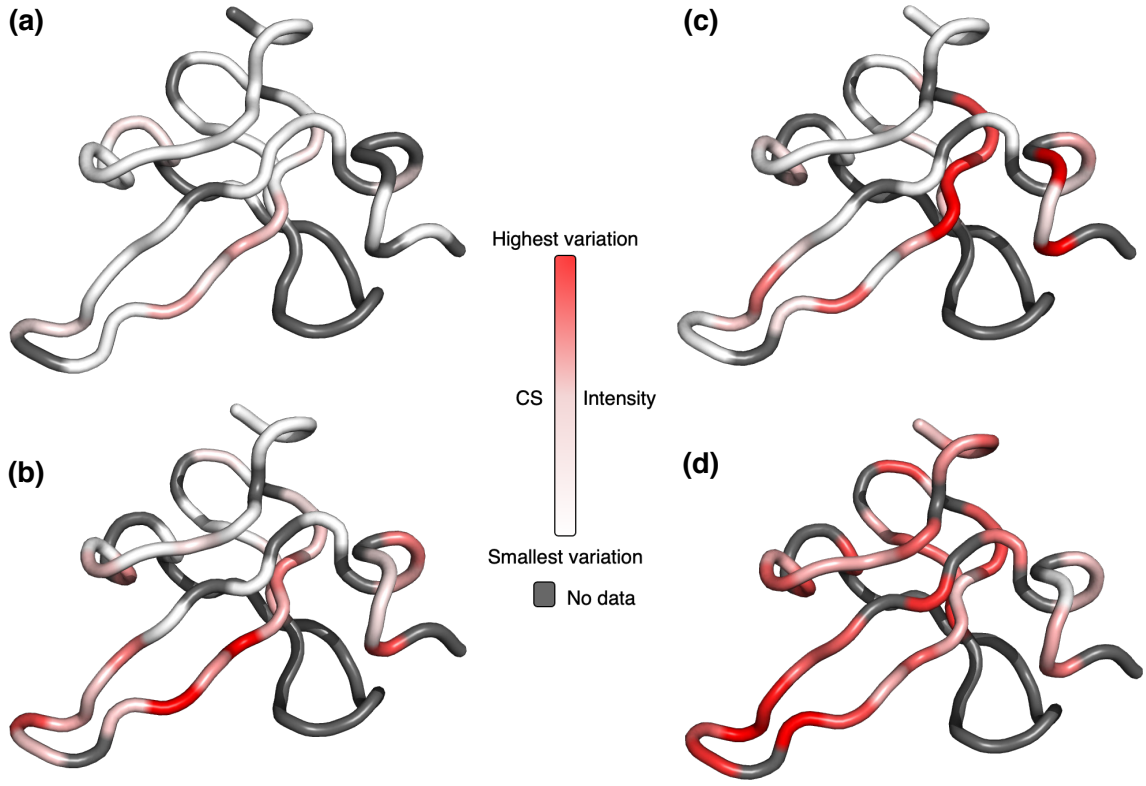

Figure 5.16: Llp CS (left panel) and intensity (right panel) variations between: a) c) Llp in DDM and Llp in buffer; b) d) Llp-FhuA HC and Llp in DDM. Results show that the region affected by DDM is still affected by the titration FhuA-Llp but there are also some additional regions of interest, particularly the rest of the loop (the redder the residue being, the more variations between samples).

In summary (see Figure 5.16), residues n°22-28 show a significant intensity reduction while adding DDM to the protein sample. Residues in the 12-22 range undergo a decrease in intensity observable while comparing spectra of Llp in DDM and Llp interacting with FhuA in presence of DDM (this decrease become more and more obvious with the increasing percentage of complex), which does not seem related only to detergent addition.

In light of our analysis, it seems that the regions interacting with DDM and with FhuA are close but not exactly identical, but considering their proximity it is not clear which impact they could also have on each other. It is important to note that these residues are all located in the same loop and that, as much as the differences between the effects of the detergent and of FhuA are highlighted, the membrane of *E. coli* may also have a role in Llp and FhuA interaction. Therefore the region of the protein sensitive to DDM may also be part of Llp functionality.

Unfortunately, the titration ratios remained quite low: as FhuA is a membrane protein, there is a difficulty to obtain it in high concentration. Indeed, DDM is generally used for solubilising membrane proteins due to its non-denaturing effect but the solubilisation yield is usually poor. The concentrations reachable thus prevented us to get sufficient data to determine an accurate K_D by NMR.

5.3.7 Llp mutants: impact on the interaction with FhuA

Based on NMR observations, a scope of Llp mutants was prepared to observe the impact of the modified residues on the infection.

According to the biological assay (see Figure 5.17), empty *E. coli* C43 strain are infected by phages T5, while C43 strain together with Ac-Llp shows no sign of infection due to Llp-FhuA interaction. These controls are necessary for comparison to the Llp mutants produced in order to assess their impact on Llp and the *E. coli* receptor interaction.

This study shows that most of the mutants alter the interaction between FhuA and Llp, depending on the T5 phages quantity, and therefore the infection is not prevented. The most impactful of the residues seem to be located in the 42-46 and the 14-18 regions. This region was already noticed as affected when comparing intensities and chemical shifts evolution during the titration in NMR. The mutants on residues n°52-54 also have an impact on Ac-Llp-FhuA interaction, which was also observable but to a lesser point, using titration intensity analysis.

When mutating the residues Trp46, Asn28, Arg29 and Lys30, Tyr22 and Lys23, the interaction between Ac-Llp and FhuA seems to occur and therefore prevent phage from entering *E. coli* cells up until some point where there are too many phages to fully prevent the interaction. Lys7 and Asn8 have even less impact on Ac-Llp-FhuA interaction.

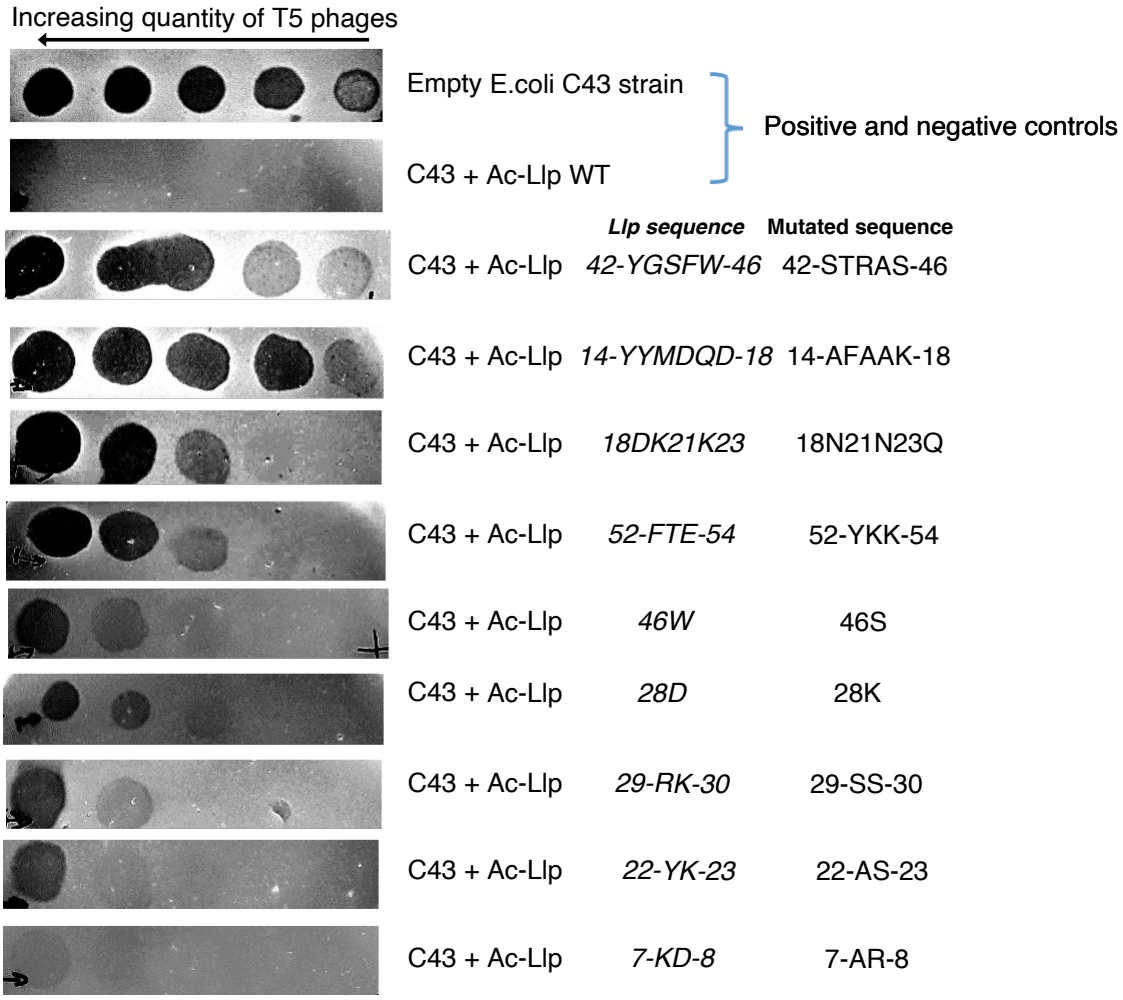

Figure 5.17: Influence of Llp mutations on interaction with FhuA, and therefore inhibition of E. coli C43 strain infection by phage T5, depending on the mutated residues. From left to right: biological assay; ^{1}H,^{15}N NMR spectrum of Llp in DDM (blue), Llp-FhuA LC (red), and Llp-FhuA HC (black); original llp sequence and mutated residues (adapted with permission from Séraphine Degroux).

This dataset is in agreement with the hypothesis formulated upon analysing NMR titration data, and are especially useful regarding n°22-23 and 28 since these residues were suspected to have an interaction in DDM but they do not seem to interfere in the receptor interaction.

As we also observe a high impact on the Ac-Llp-FhuA interaction when the residues

18, 21 and 23 are mutated, it seems the effect is almost absent when the residues 22 and 23 are mutated, which would mean either the association of the residues has an impact or only n°18 and 21 have one.

From the plotted influence of residues n°14-18, 21-23, 52-54 (observed impact on interaction), and n°7-8, 22-23, 29-30 (no/slight impact on interaction) on Llp structure (see Figure 5.18), it appears that the hypothesis made regarding the location of the interaction may be verified again by the spatial proximity between residues 14-18 in the loop region and the n°52-54 located nearby.

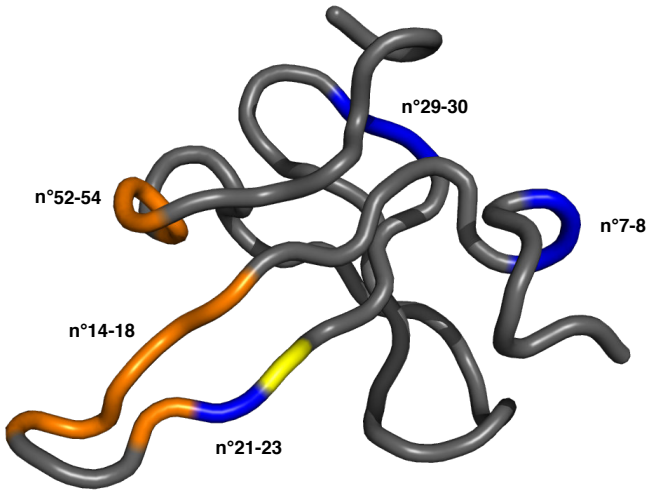

Figure 5.18: Map of the influence of Ac-Llp mutations on interaction with FhuA: (orange) residues that have direct influence on interaction with FhuA; (blue) residues with no/very slight influence on interaction, based on biological assay. Residue n°23 is highlighted in yellow, since it was mutated both in Ac-Llp not and still interacting with FhuA.

5.4 Conclusions

Using a soluble form of the lytic conversion lipoprotein Llp, assignment of most of the chemical shifts could be done using 2D and 3D spectra correlations. As there is some change in the spectrum over time, Llp is probably subject to degradation or at least a rearrangement at some point when conserved at temperatures near or above RT.

Dynamics of sol-Llp seemed mostly in accordance with the usual observations in proteins, namely the extended flexibility of the terminus and the rigidity of most structured parts. However, the surprising rigidity noticeable using relaxation data obtained for Llp

loop region rises questions. Further experiments on other timescales, other temperatures and other fields would be helpful to complete this analysis, that is just at the premises.

Using NMR titration, FhuA and sol-Llp have been shown to interact, while disregarding the influence of the detergent. However, to quantify the interaction, further sampling of the FhuA-Llp interaction has to be performed, mainly by increasing [Llp]/[FhuA] ratios, which could be achieved by concentrating the membrane protein more that our available stock sample was. The main point of this study was achieved considering we identified some regions undergoing CS and intensity variations while sol-Llp was interacting with FhuA, which was then confirmed by the Ac-Llp mutants.

In the end, the results obtained are promising: most of Llp assignment was achieved, some premises on the protein behaviour were studied using spin relaxation, and the interaction between the lipoprotein Llp and the *E. coli* receptor FhuA was visible. Next stage would require more time and experiments to characterise further the interaction between these two proteins, by quantifying the equilibrium constants, and even investigating the dynamic changes between Llp in a free form and in interaction with FhuA.

Conclusion

During this PhD, I contributed to the development and application of both sample synthesis, as well as structural and dynamics analysis by NMR spectroscopy.

The main laboratory work consisted in implementing protocols to synthesize oligonucleotides, starting with their natural form, then inserting probes that are especially useful for NMR spectroscopy.

Non-coding RNAs have been a rising center of interest over these past decades, whereas proteins have been the main investigated field within the cell machinery. Thus, the development of synthesis methodologies to simplify (or complete) the RNAs characterisation is a significant achievement. The synthesis and post-synthesis treatments methodologies have been elaborated through several trials within the laboratory, using various RNA sequences that also are coincident with biological questions.

Indeed, the *let-7* family implicated in various forms of cancer is a particularly interesting miRNA, due to the knowledge already available on its biogenesis and its messenger RNAs targets. Nonetheless, if its influence in several diseases has been demonstrated, the interaction between the miRNA and its targets has not been fully characterised. Therefore, understanding the various interactions may help in designing mimics with therapeutic potential.

Fluorine NMR has also been developed over the past years, considering its undoubtable useful aspect for NMR spectroscopy, and the poor presence in biomolecules, rendering it easily identifiable once placed within a biomolecule sequence.

Here, the chemical introduction of fluorine allowed to directly look into the interaction between *let-7* and some of its targets. Relaxation dispersion experiments have unravelled conformational exchange involving some of the labelled nucleotides (*let-7*:LIN41 U9) while most of them did not seem affected by spin-lock variations. The base-pairing rigidity in the seed region contrasts with the rest of the miRNA sequence, as the presence of exchange elsewhere does not seem to affect the binding with the target. Additional spin-locks are now required to characterise the conformation of the invisible state, as well as further fluorine labelling to obtain supplementary information along *let-7* sequence.

Another objective of my work was to characterise the dynamics of a transfer RNA relaxation processes.

As dynamics has become a widely investigated field in proteins, the methodology is not always straightforward to apply to oligonucleotides. Here, I associated the acquisition of the relaxation data at different fields, as well as their theoretical explanations, to grasp their physical meaning.

While the *modelfree* analysis has been considerably applied to proteins, the implemen-

tation for oligonucleotides has triggered interrogations over parameters (e.g. CSA and N-H distances) implied in the relevance of the fitting, and therefore on the motion interpretation.

An attempt to reconcile the exchange probed in the *modelfree* analysis and relaxation dispersion through CPMG experiments did not lead to particularly connected results. However, the exchange is, at the time of writing, still investigated through additional relaxation dispersion experiments, including CEST.

Along with the NMR methodology, the biological question regarding $tRNA^{Phe}$ and its maturated form was also a point of interest. Indeed, the investigation was also focused on the chemical modifications undergone during the maturation process and their impact on the tRNA dynamics. In this mindset, the characterisation of a tRNA possessing one chemical modification was started using relaxation dispersion, that will then be compared to the non-maturated form.

If RNAs are of growing interest, proteins are still a fundamental dogma of biology. In particular, bacteriophages are a current compelling topic, especially with the rise of resistance strains to antibiotics, in bacterial infections.

The bacteriophage structure itself represents already a huge characterisation investment, but the bacteriophage action requires additional understanding of its infection of the bacteria cells. Thus, the study of the lipoprotein responsible for avoiding the superinfection of the cell due to its interaction with the membrane is fundamental.

The first step was to look into Llp structure using assignment and spatial restrictions, through ^{1}H, ^{13}C and ^{15}N NMR. Once characterised on its own, the interaction between Llp and the bacteriophage receptor FhuA located on the bacteria membrane could be probed. Interestingly, even if the interaction could not be quantified, a binding interface could be determined by interpreting the chemical shift and resonance intensity variations between the spectra of various Llp:FhuA ratio.

Not mentioned in the manuscript, a part of my work also consisted in studying the dynamics in human Aquaporin 1 by solid-state NMR.

The aquaporins are molecules located on the cell membranes, that facilitate the transport of water and other small compounds such as glycerol into the cell. Aquaporins are usually formed by a monomeric unit consisting in 6 membrane-spanning helices, connected by 5 loops, intra- or extra-cellular. Dynamic processes are a key to understand the water permeation mechanism.

^{1}H-^{13}C and ^{1}H-^{15}N relaxation experiments were realised: ^{15}N R_1 and $R_1\rho$ experiments at 800 MHz and 1000 MHz, 15kHz and 280 K, as well as experiments to obtain order parameters (Variable Contact time-Cross-Polarisation, CP-VC and Rotational-Echo DOuble Resonance, REDOR). The relaxation experiments confirmed the rigidity of the helices structural parts, and increased mobility of the loop regions. The second set of experiments were used to obtain order parameters, that were also expected considering the aquaporin structure. However, noticeable increased rigidity was encountered at the end of a loop,

indicating the possible formation of a β-turn, that would lead to a conformational rearrangement providing a correct orientation to the adjacent loop, and facilitating water transport.

During this PhD, I approached various biomolecules, ranging from oligonucleotides to proteins, and many NMR methodologies (spin relaxation, relaxation dispersion, assignment, interaction). NMR spectroscopy has been and remains an important technique in structural biology. The many possibilities it offers to investigate structures and dynamics make a step towards the understanding of various biomolecular functions. As a vast part of the biological kingdom still needs to be explored, NMR spectroscopy is still developing, to adjust the relevance of the biological questions: lately, *in-cell* NMR has further advanced the comprehension of the biomolecules in their native environment. The development is also focused on overcoming its limits, e.g. through the increase in magnetic fields, along with the development of Dynamic Nuclear Polarization (DNP).

Advances in the understanding of biological mechanisms is a necessity to comprehend their dysfunctions, as well as to be able to correct them. Exploring the biomolecular structures, dynamics and interactions, help us highlight the pathological disorders and head towards different therapies. As there still remains a lot of biologically unanswered questions, several strategies are still starting to emerge, including for the lately discovered miRNAs, constituting a thrilling area of research [184].

Annexes

Annex 1: tRNA^{Phe} assignment table

Assignment	^{15}N / ppm	^{1}H / ppm
03	162.25	12.225
04	161.707	10.267
06	162.871	14.302
07	161.072	13.185
08	160.45	14.216
10	160.283	12.665
12	160.631	13.694
15	162.398	12.019
18	161.763	9.27
22	162.703	12.756
24	163.268	13.621
30	161.832	12.445
41	161.713	13.05
42	161.634	12.07
46	163.22	12.491
50	162.36	13.166
51	162.099	12.261
52	162.785	13.701
54	160.915	13.573
55	160.715	11.589
65	162.273	12.368
68	161.493	13.798
69	162.559	11.682
71	162.808	12.824

Table 5.2: tRNA^{Phe} assignment table.

Annex 2: tRNA R_1 at 600 MHz

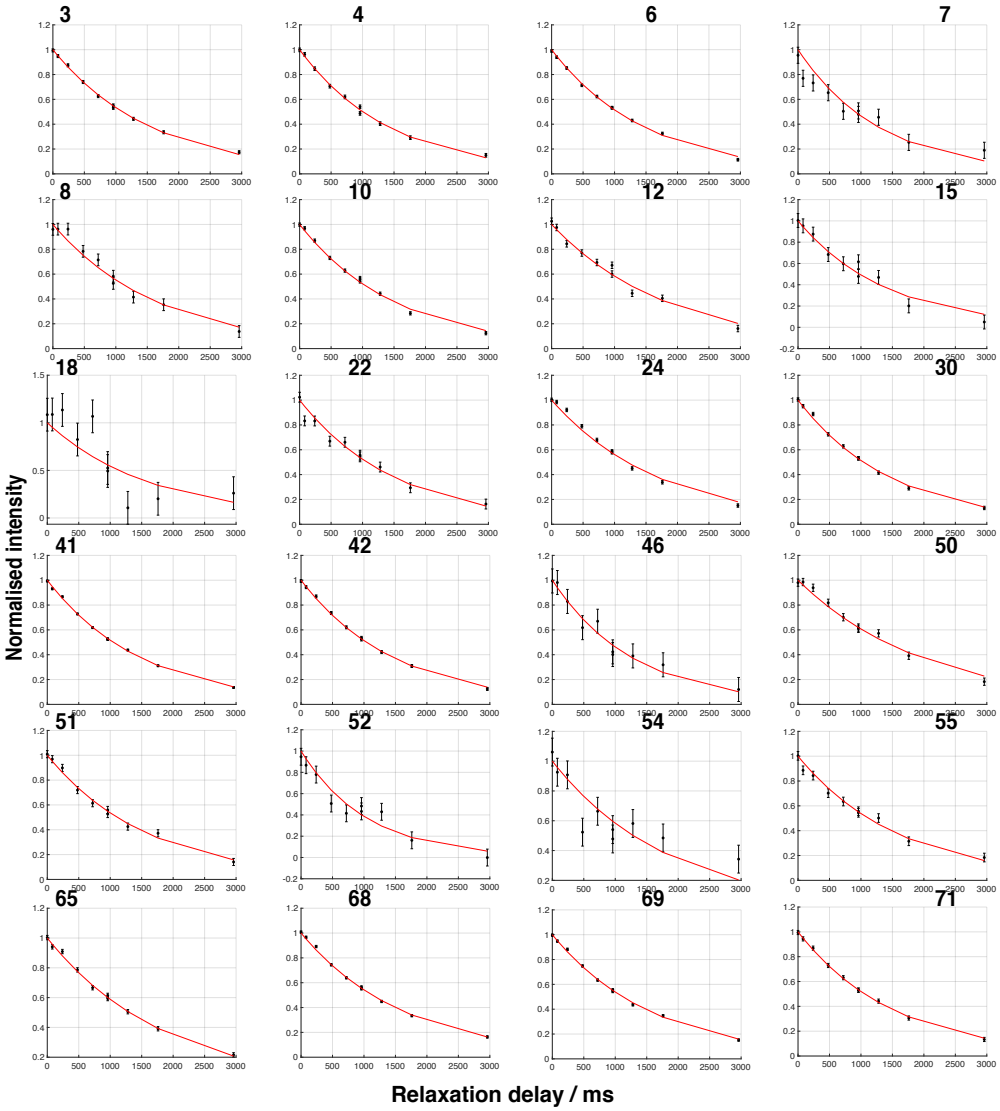

Figure 5.19: tRNA^{Phe} ^{15}N site-specific longitudinal relaxation rates R_1 at 600 MHz and 310K.

Annex 3: tRNA $R_1\rho$ at 600 MHz

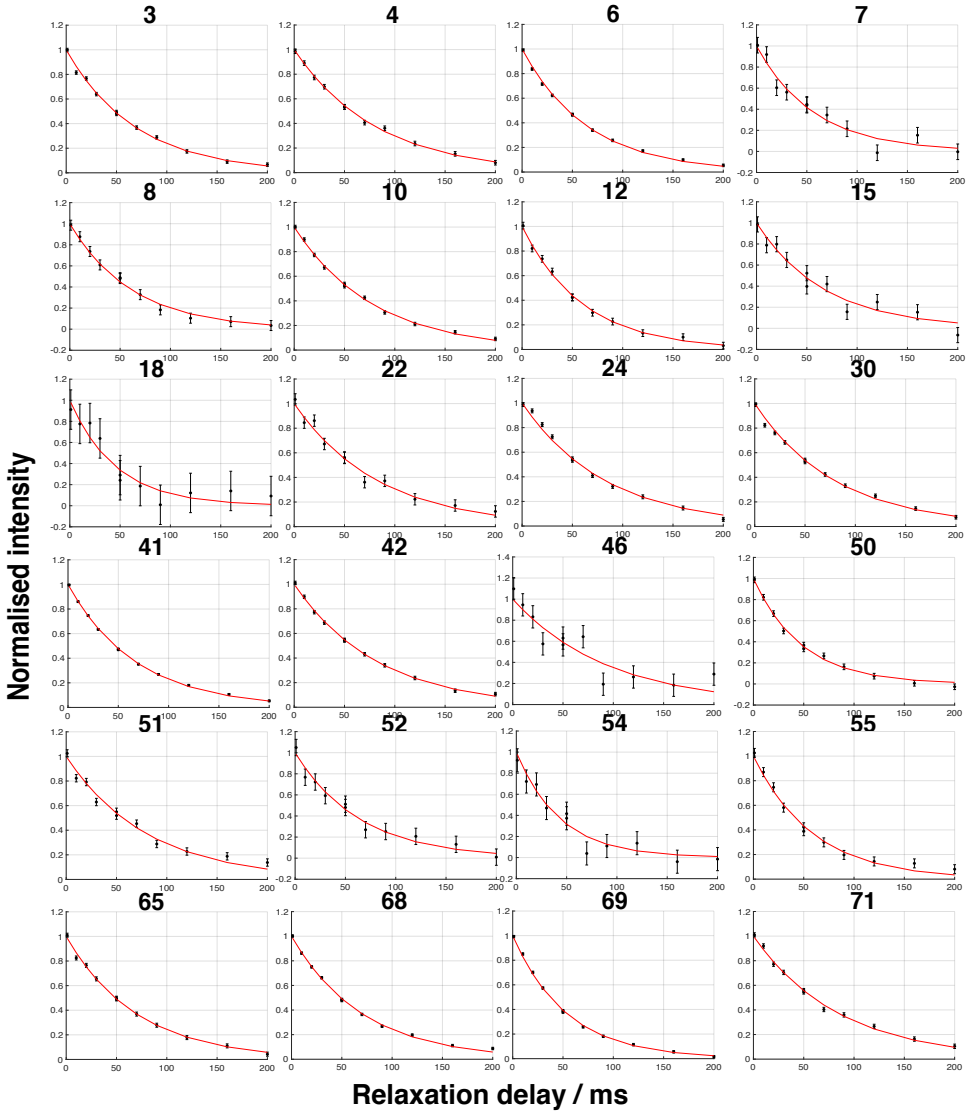

Figure 5.20: tRNA^{Phe} ^{15}N site-specific relaxation rates $R_1\rho$ at 600 MHz and 310K.

Annex 4: tRNA CPMG at 600 MHz

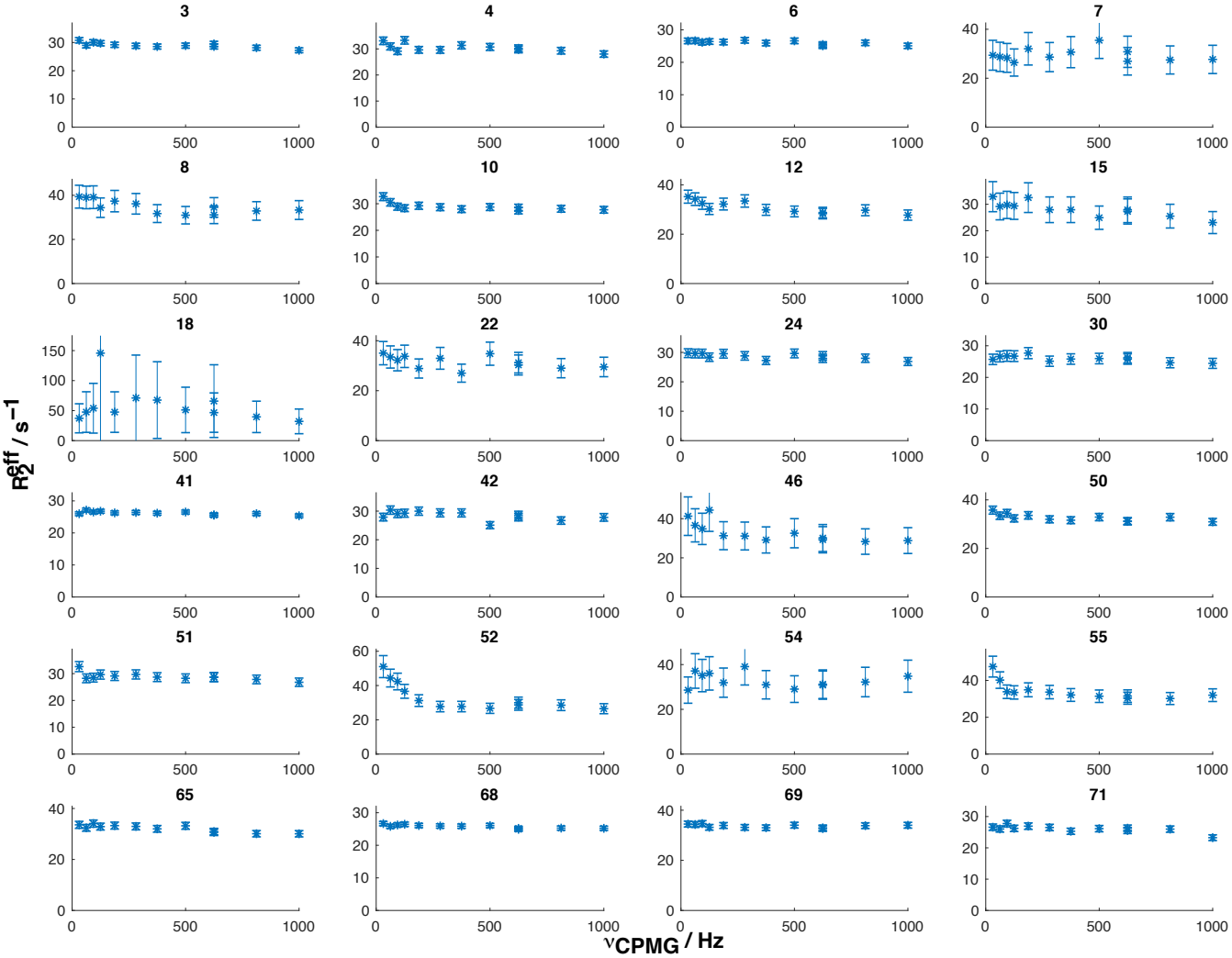

Figure 5.21: tRNA^{Phe} ^{15}N CPMG at 600 MHz and 310K.

Annex 5: tRNA CPMG at 1000 MHz

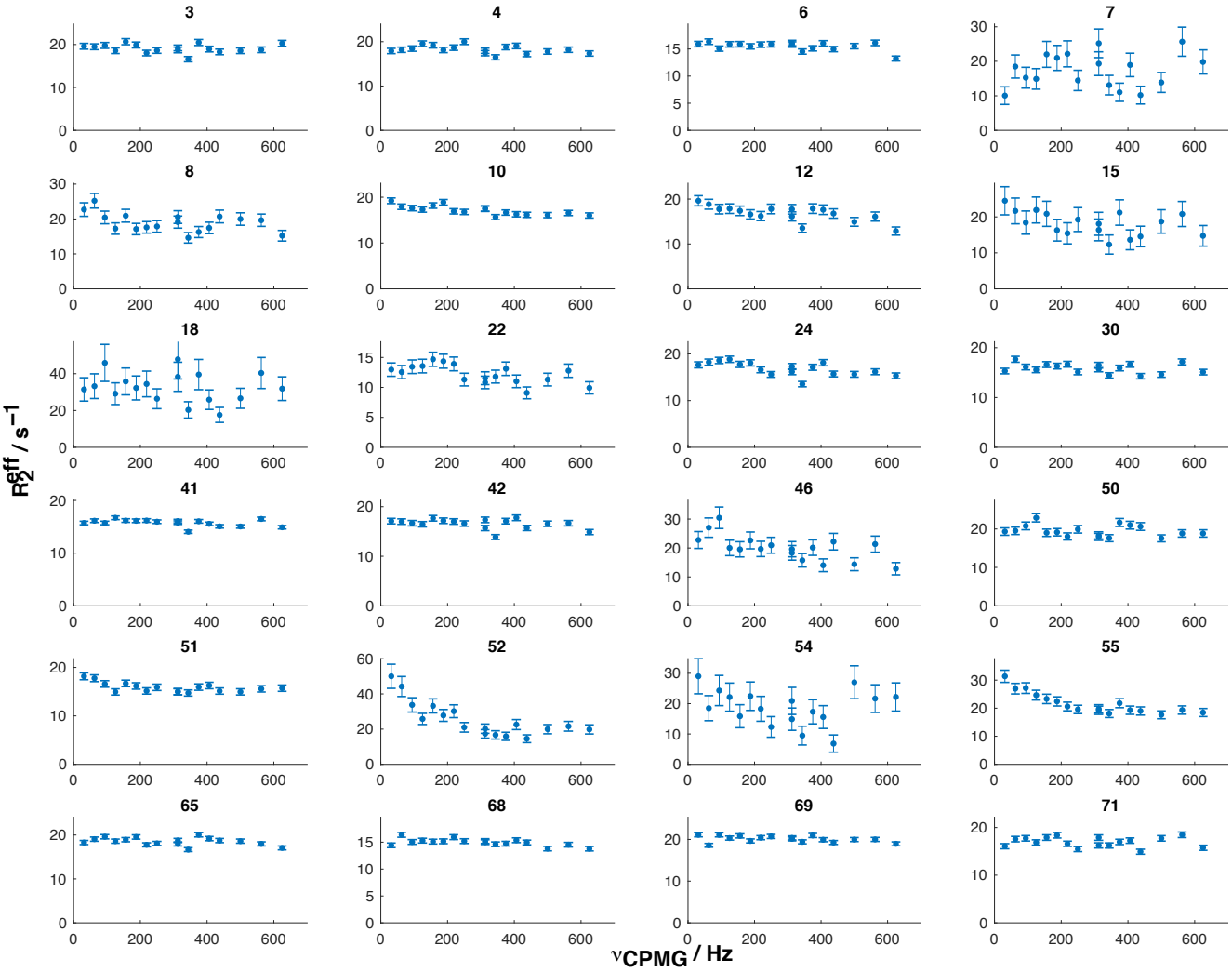

Figure 5.22: tRNA^{Phe} ^{15}N CPMG at 1000 MHz and 310K.

Annex 6: tRNA *model-free* analysis: CSA and r(NH) influence investigation

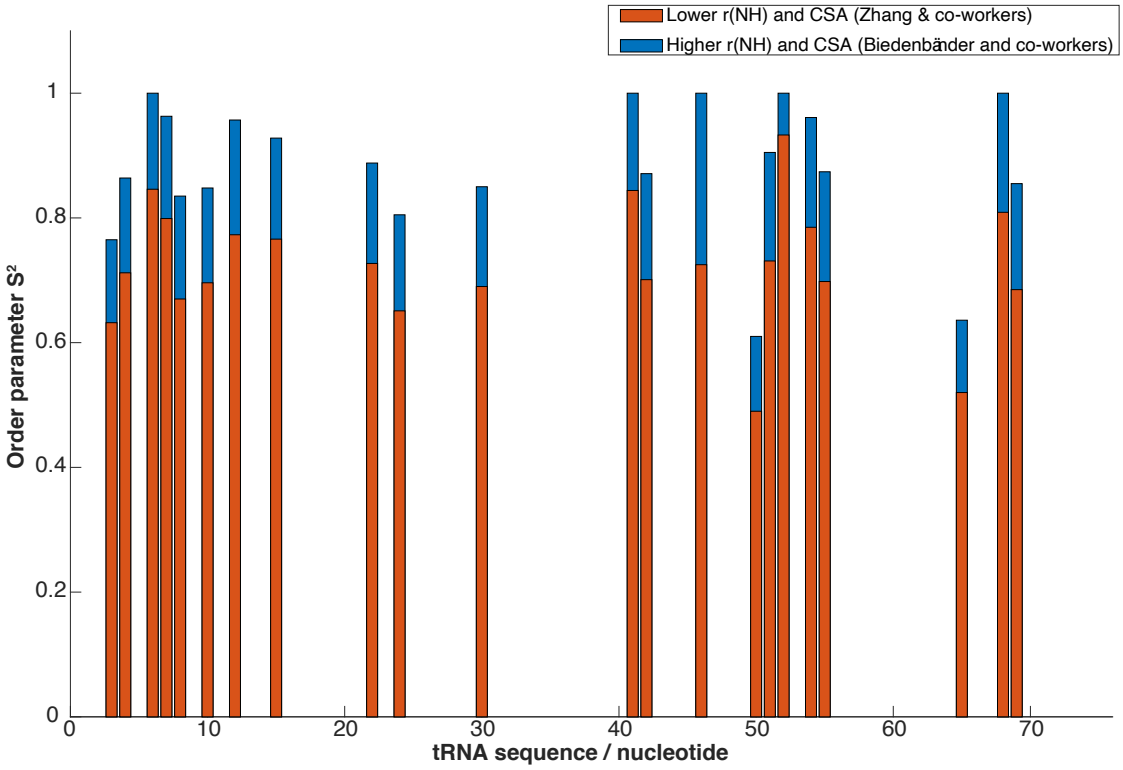

Figure 5.23: $tRNA^{Phe}$ *model-free* analysis using various CSA and r(NH) values from literature: order parameters are plotted for each nucleotide.

Annex 7: Llp assignment table

Assignment	Amino acid	^{1}H / ppm	^{15}N / ppm
2	SER	8.538858872	115.6558688
3	THR	7.888240453	115.2940954
4	PHE	8.014214654	121.576304
5	GLY	9.17398242	109.5042989
7	LYS	8.496598299	116.1632646
8	ASP	8.190838112	117.5176164
9	ILE	7.303190585	120.1669687
10	LYS	8.36863369	126.774504
12	GLU	8.058305789	115.2327434
13	ALA	8.378207723	121.6101623
14	TYR	8.896099138	118.5166974
15	TYR	9.038145825	120.3481126
16	MET	9.179586601	119.3761271
17	GLN	9.11903347	123.3719176
18	ASP	9.452077361	128.1014885
19	HIS	8.654631853	106.7126784
20	VAL	7.944947585	122.9577093
21	LYS	8.145670593	127.1724642
22	TYR	9.249052105	129.260635
23	LYS	8.32875453	117.7415827
24	ALA	9.401388756	124.8447148
25	ASN	7.897336763	119.9218807
26	VAL	8.168523795	121.2667207
27	PHE	8.81454329	122.4967573
28	ASP	7.481573276	114.516902
29	ARG	8.30810044	121.1491178

Assignment	Amino acid	^{1}H / ppm	^{15}N / ppm
30	LYS	8.621664933	128.05622
31	GLY	9.181524197	117.5989529
33	MET	7.864736675	117.7789516
34	PHE	9.482669205	119.686574
35	LEU	7.99792061	125.3929601
36	VAL	8.740753225	119.402574
37	SER	8.151109418	111.3999198
40	MET	8.351241063	118.5140838
41	ALA	9.636806747	129.3053617
42	TYR	8.758778916	120.5847836
46	TRP	8.749054797	121.7111059
47	ALA	9.86231413	125.2233552
49	VAL	8.5285985	124.4565927
50	SER	8.544261229	112.2289273
51	TYR	7.925010619	122.9769541
52	PHE	7.589462295	117.1264381
53	THR	7.84556127	111.9593695
54	GLU	7.785662481	117.2504679
55	GLY	7.801711667	107.261295
56	ASN	8.771275243	111.3328407
57	THR	9.217469034	111.2996312
59	GLU	8.687237533	125.5157988
60	GLY	9.006072549	114.2564055
61	VAL	7.737931521	119.5749511
62	PHE	7.295609868	122.175065

Table 5.3: Llp assignment table.

Bibliography

[1] James Watson and Francis Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[2] Lisa A. Jackson, Evan J. Anderson, Nadine G. Rouphael, Paul C. Roberts, Mamodikoe Makhene, Rhea N. Coler, Michele P. McCullough, James D. Chappell, Mark R. Denison, Laura J. Stevens, Andrea J. Pruijssers, Adrian McDermott, Britta Flach, Nicole A. Doria-Rose, Kizzmekia S. Corbett, Kaitlyn M. Morabito, Sijy O'Dell, Stephen D. Schmidt, Phillip A. Swanson, Marcelino Padilla, John R. Mascola, Kathleen M. Neuzil, Hamilton Bennett, Wellington Sun, Etza Peters, Mat Makowski, Jim Albert, Kaitlyn Cross, Wendy Buchanan, Rhonda Pikaart-Tautges, Julie E. Ledgerwood, Barney S. Graham, and John H. Beigel. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *New England Journal of Medicine*, 383(20):1920–1931, 2020.

[3] Emily Bettini and Michela Locci. SARS-CoV-2 mRNA Vaccines: Immunological mechanism and beyond. *Vaccines*, 9(2):1–20, 2021.

[4] Stephen Neidle and Mark Sanderson. *Principles of Nucleic Acid Structure*. 2022.

[5] Neocles B. Leontis and Eric Westhof. Conserved geometrical base-pairing patterns in RNA. *Quarterly Reviews of Biophysics*, 31(4):399–455, 1998.

[6] Neocles B. Leontis and Eric Westhof. Geometric nomenclature and classification of RNA base pairs. *Rna*, 7(4):499–512, 2001.

[7] Gabriele Varani and William H McClain. The G·U wobble base pair. *EMBO reports*, 1(1):18–23, 2000.

[8] G. Raghunathan, H. Todd Miles, and V. Sasisekharan. Parallel nucleic acid helices with hoogsteen base pairing: Symmetry and structure. *Biopolymers*, 34(12):1573–1581, 1994.

[9] Michael Bon, Graziano Vernizzi, Henri Orland, and A. Zee. Topological Classification of RNA Structures. *Journal of Molecular Biology*, 379(4):900–911, 2008.

[10] Matthew Cobb. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biology*, 15(9):1–8, 2017.

[11] R. Kaptein, R. Boelens, R. M. Scheek, and W. F. Van Gunsteren. Protein Structures from NMR. *Biochemistry*, 27(15):5389–5395, 1988.

[12] John Cavanagh, Wayne J. Fairbrother, Arthur G. Palmer, Nicholas J. Skelton, and Mark Rance. *Protein NMR Spectroscopy*. 2007.

[13] Akiva Feintuch and Shimon Vega. *Spin dynamics*, volume 6. 2017.

[14] James Keeler. *Understanding NMR spectroscopy*, volume 43. 2006.

[15] F. Bloch. Nuclear induction. *Physical Review*, 70(7-8):460–474, 1946.

[16] N. Bloembergen, E. M. Purcell, and R. V. Pound. Relaxation effects in nuclear magnetic resonance absorption. *Physical Review*, 73(7):679–712, 1948.

[17] Vol Um, E 99, Num B Er, and L Soromont. P H YSI CAL REVIEW Relaxation Processes in a System of Two Spins*. *Physical Review*, 953(8):559–565, 1955.

[18] Arthur G. Palmer. Relaxation and Dynamic Processes. pages 1–118, 1997.

[19] I. Solomon. Relaxation processes in a system of two spins. *Physical Review*, 99(2):559–565, 1955.

[20] R. K. Wangsness and F. Bloch. The dynamical theory of nuclear induction, 1953.

[21] F. M. Chen and R. F. Snider. Collision theory of relaxation processes. *The Journal of Chemical Physics*, 46(10):3937–3940, 1967.

[22] A Abragam. *The principles of nuclear magnetism-Oxford, Clarendon Press (1961)*. 1961.

[23] Patrik Westerkull. The ponto bone-anchored hearing system. *Advances in Oto-Rhino-Laryngology*, 71:32–40, 2011.

[24] J. Klafter and I.M. Sokolov. Master equations. *First Steps in Random Walks*, 241:68–79, 2013.

[25] Ranajeet Ghose. Average Liouvillian theory in nuclear magnetic resonance - Principles, properties, and applications. *Concepts in Magnetic Resonance*, 12(3):152–172, 2000.

[26] Malcolm H. Levitt and Lorenzo Di Bari. Steady state in magnetic resonance pulse experiments. *Physical Review Letters*, 69(21):3124–3127, 1992.

[27] Jeffrey W. Peng, V. Thanabal, and Ger Hard Wagner. 2D heteronuclear NMR measurements of spin-lattice relaxation times in the rotating frame of X nuclei in heteronuclear HX spin systems. *Journal of Magnetic Resonance (1969)*, 94(1):82–100, 1991.

[28] Paul S. Hubbard. Some properties of correlation functions of irreducible tensor operators. *Physical Review*, 180(1):319–326, 1969.

[29] Abhishek Gupta, Timothy Stait-Gardner, Minoo J. Moghaddam, and William S. Price. Dipolar relaxation revisited: A complete derivation for the two spin case. *Concepts in Magnetic Resonance Part A: Bridging Education and Research*, 44(2):74–113, 2015.

[30] Lawrence G. Werbelow and David M. Grant. Intramolecular Dipolar Relaxation in Multispin Systems. *Advances in Magnetic and Optical Resonance*, 9(C):189–299, 1977.

[31] E D'Auvergne. *The relax manual*. 2014.

[32] Neil A. Farrow, Ouwen Zhang, Attila Szabo, Dennis A. Torchia, and Lewis E. Kay. Spectral density function mapping using 15N relaxation data exclusively. *Journal of Biomolecular NMR*, 6(2):153–162, 1995.

[33] Giovanni Lipari and Attila Szabo. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society*, 104(17):4546–4559, 1982.

[34] Giovanni Lipari and Attila Szabo. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *Journal of the American Chemical Society*, 104(17):4559–4570, 1982.

[35] Peter Allard, Magnus Helgstrand, and Torleif Härd. The Complete Homogeneous Master Equation for a Heteronuclear Two-Spin System in the Basis of Cartesian Product Operators. *Journal of Magnetic Resonance*, 134(1):7–16, 1998.

[36] Arthur M. Mandel, Mikael Akke, and Arthur G. Palmer. Backbone dynamics of Escherichia coli ribonuclease HI: Correlations with structure and function in an active enzyme. *Journal of Molecular Biology*, 246(1):144–163, 1995.

[37] Patrice Dosset, Jean Christophe Hus, Martin Blackledge, and Dominique Marion. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *Journal of Biomolecular NMR*, 16(1):23–28, 2000.

[38] Sébastien Morin, Troels E. Linnet, Mathilde Lescanne, Paul Schanda, Gary S. Thompson, Martin Tollinger, Kaare Teilum, Stéphane Gagné, Dominique Marion, Christian Griesinger, Martin Blackledge, and Edward J. D'auvergne. Relax: The analysis of biomolecular kinetics and thermodynamics using NMR relaxation dispersion data. *Bioinformatics*, 30(15):2219–2220, 2014.

[39] Yufeng Wei, Angel C. De Dios, and Ann E. McDermott. Solid-state 15N NMR chemical shift anisotropy of histidines: Experimental and theoretical studies of hydrogen bonding. *Journal of the American Chemical Society*, 121(44):10389–10394, 1999.

[40] David Fushman, Nico Tjandra, and David Cowburn. Direct measurement of 15N chemical shift anisotropy in solution. *Journal of the American Chemical Society*, 120(42):10947–10952, 1998.

[41] Andrew L. Lee and A. Joshua Wand. Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. *Journal of Biomolecular NMR*, 13(2):101–112, 1999.

[42] Alexander Grishaev, Lishan Yao, Jinfa Ying, Arthur Pardi, and Ad Bax. Chemical shift anisotropy of imino 15N nuclei in Watson-Crick base pairs from magic angle spinning liquid crystal NMR and nuclear spin relaxation. *Journal of the American Chemical Society*, 131(27):9490–9491, 2009.

[43] Mikael Akke, Radovan Fiala, Feng Jiang, Dinshaw Patel, and Arthur G. Palmer. Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: Correlations with structure and stability. *Rna*, 3(7):702–709, 1997.

[44] Mikael Akke, Rafael Brüschweiler, and Arthur G. Palmer. NMR Order Parameters and Free Energy: An Analytical Approach and Its Application to Cooperative Ca2+ Binding by Calbindin D9k. *Journal of the American Chemical Society*, 115(21):9832–9833, 1993.

[45] Daiwen Yang and Lewis E. Kay. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding. *Journal of Molecular Biology*, 263(2):369–382, 1996.

[46] Nicola Salvi, Anton Abyzov, and Martin Blackledge. Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 102-103:43–60, 2017.

[47] Qi Zhang, Xiaoyan Sun, Eric D. Watt, and Hashim M. Al-Hashimi. Resolving the motional modes that code for RNA adaptation. *Science*, 311(5761):653–656, 2006.

[48] Olivier Fisette, Patrick Lagüe, Stéphane Gagné, and Sébastien Morin. Synergistic applications of MD and NMR for the study of biological systems. *Journal of Biomedicine and Biotechnology*, 2012(2012):1–12, 2012.

[49] Adrien Favier and Bernhard Brutscher. Recovering lost magnetization: Polarization enhancement in biomolecular NMR. *Journal of Biomolecular NMR*, 49(1):9–15, 2011.

[50] Nils Alexander Lakomek, Jinfa Ying, and Ad Bax. Measurement of 15N relaxation rates in perdeuterated proteins by TROSY-based methods. *Journal of Biomolecular NMR*, 53(3):209–221, 2012.

[51] Martial Piotto, Vladimir Saudek, and Vladimir Sklenář. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR*, 2(6):661–665, 1992.

[52] Fabien Ferrage, Amy Reichel, Shibani Battacharya, David Cowburn, and Ranajeet Ghose. On the measurement of 15N-{1H} nuclear Overhauser effects. 2. Effects of the saturation scheme and water signal suppression. *Journal of Magnetic Resonance*, 207(2):294–303, 2010.

[53] F. Massi. Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR R1 relaxation experiments. *Protein Science*, 14(3):735–742, 2005.

[54] Yi Xue, Dawn Kellogg, Isaac J. Kimsey, Bharathwaj Sathyamoorthy, Zachary W. Stein, Mitchell McBrairty, and Hashim M. Al-Hashimi. Characterizing RNA excited states using NMR relaxation dispersion. *Methods in Enzymology*, 558(1):39–73, 2015.

[55] Maja Marušič, Judith Schlagnitweit, and Katja Petzold. RNA Dynamics by NMR Spectroscopy. *ChemBioChem*, 20(21):2685–2710, 2019.

[56] Tairan Yuwen, Guillaume Bouvignies, and Lewis E. Kay. Exploring methods to expedite the recording of CEST datasets using selective pulse excitation. *Journal of Magnetic Resonance*, 292:1–7, 2018.

[57] Carr H. Y . and Purcell E.M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94(3):630, 1954.

[58] S. Meiboom and D. Gill. Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29(8):688–691, 1958.

[59] D. Flemming Hansen, Pramodh Vallurupalli, and Lewis E. Kay. An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *Journal of Physical Chemistry B*, 112(19):5898–5904, 2008.

[60] Pramodh Vallurupalli, Guillaume Bouvignies, and Lewis E. Kay. Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. *Journal of Physical Chemistry B*, 115(49):14891–14900, 2011.

[61] Jameson R. Bothe, Zachary W. Stein, and Hashim M. Al-Hashimi. Evaluating the uncertainty in exchange parameters determined from off-resonance R1rho; relaxation dispersion for systems in fast exchange. *Journal of Magnetic Resonance*, 244:18–29, 2014.

[62] J. P. Carver and R. E. Richards. A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *Journal of Magnetic Resonance (1969)*, 6(1):89–105, 1972.

[63] Judith Schlagnitweit, Emilie Steiner, Hampus Karlsson, and Katja Petzold. Efficient Detection of Structure and Dynamics in Unlabeled RNAs: The SELOPE Approach. *Chemistry - A European Journal*, 24(23):6067–6070, 2018.

[64] J. Kremser, E. Strebitzer, R. Plangger, M. A. Juen, F. Nußbaumer, H. Glasner, K. Breuker, and C. Kreutz. Chemical synthesis and NMR spectroscopy of long stable isotope labelled RNA. *Chemical Communications*, 53(96):12938–12941, 2017.

[65] Owen Becette, Lukasz T. Olenginski, and Theodore K. Dayie. Solid-phase chemical synthesis of stable isotope-labeled RNA to aiD structure and dynamics studies by NMR spectroscopy. *Molecules*, 24(19), 2019.

[66] Maximilian Himmelstoß, Kevin Erharter, Eva Renard, Eric Ennifar, Christoph Kreutz, and Ronald Micura. 2̆'-: O -Trifluoromethylated RNA-a powerful modification for RNA chemistry and NMR spectroscopy. *Chemical Science*, 11(41):11322–11330, 2020.

[67] Karolina Bartosik, Katarzyna Debiec, Anna Czarnecka, Elzbieta Sochacka, and Grazyna Leszczynska. Synthesis of nucleobase-modified RNA oligonucleotides by post-synthetic approach. *Molecules*, 25(15), 2020.

[68] Daniel A. Glazier, Daniel A. Glazier, Junzhuo Liao, Brett L. Roberts, Xiaolei Li, Ka Yang, Christopher M. Stevens, Weiping Tang, and Weiping Tang. Chemical Synthesis and Biological Application of Modified Oligonucleotides. *Bioconjugate Chemistry*, 31(5):1213–1233, 2020.

[69] Glen F. Deleavey and Masad J. Damha. Designing chemically modified oligonucleotides for targeted gene silencing. *Chemistry and Biology*, 19(8):937–954, 2012.

[70] Francis Crick and James Watson. 1953. *Chinese Education*, 11(2-3):76–106, 1978.

[71] M. Smith, D. H. Rammler, I. H. Goldberg, and H. G. Khorana. Studies on Polynucleotides. XIV.1 Specific Synthesis of the C3'-C5' Interribonucleotide Linkage. Syntheses of Uridylyl-(3' → 5')-Uridine and Uridylyl-(3' → 5')-Adenosine. *Journal of the American Chemical Society*, 84(3):430–440, 1962.

[72] Ramamoorthy Belagaje. -c GTAAT99 , GCACCACC ,. *Methods Enzymol.*, 68(1976), 1979.

[73] H. Schaller, G. Weimann, B. Lerch, and H. G. Khorana. Studies on Polynucleotides. XXIV. The Stepwise Synthesis of Specific Deoxyribopolynucleotides (4). Protected

Derivatives of Deoxyribonucleosides and New Syntheses of Deoxyribonucleoside-3' Phosphates. *Journal of the American Chemical Society*, 85(23):3821–3827, 1963.

[74] R. B. Merrifield. Solid phase peptide synthesis. *Excerpta Med., I.C.S.*, No.374:29–39, 1976.

[75] Robert L. Letsinger and V. Mahadevan. Oligonucleotide Synthesis on a Polymer Support1,2. *Journal of the American Chemical Society*, 87(15):3526–3527, 1965.

[76] Marvin H. Caruthers. A brief review of DNA and RNA chemical synthesis. *Biochemical Society Transactions*, 39(2):575–580, 2011.

[77] Marvin H. Caruthers. The chemical synthesis of DNA/RNA: Our gift to science. *Journal of Biological Chemistry*, 288(2):1420–1427, 2013.

[78] K. C. Tsou and K. F. Yip. Synthesis of deoxyoligonucleotides on an isotactic polymer support. *Journal of Macromolecular Science: Part A - Chemistry*, 7(5):1097–1116, 1973.

[79] Marie Flamme, Luke K. McKenzie, Ivo Sarac, and Marcel Hollenstein. Chemical methods for the modification of RNA. *Methods*, 161(March 2019):64–82, 2019.

[80] Serge L. Beaucage and Radhakrishman P. Iyer. Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach. *Tetrahedron*, 48(12):2223–2311, 1992.

[81] S. SomozaÀlvaro. Protecting groups for RNA synthesis: An increasing need for selective preparative methods. *Chemical Society Reviews*, 37(12):2668–2675, 2008.

[82] Cynthia J. Krieger, Peifen Zhang, Lukas A. Mueller, Alfred Wang, Suzanne Paley, Martha Arnaud, John Pick, Seung Y. Rhee, and Peter D. Karp. MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research*, 32(DATABASE ISS.):5253–5372, 2004.

[83] Igor A. Kozlov, Minh Dang, Ken Sikes, Theofilos Kotseroglou, David L. Barker, and Chanfeng Zhao. Significant improvement of quality for long oligonucleotides by using controlled pore glass with large pores. *Nucleosides, Nucleotides and Nucleic Acids*, 24(5-7):1037–1041, 2005.

[84] Sung Won Kim, Yang Il Cho, and Kyeong Eun Jung. Avoiding High-Pressure Problem for Modified RNA-attached Polystyrene Support by Pre-Swelling Using Toluene in the Oligonucleotide Synthesis. *Bulletin of the Korean Chemical Society*, 42(10):1296–1303, 2021.

[85] Vasulinga T. Ravikumar, R. Krishna Kumar, Phil Olsen, Max N. Moore, Recaldo L. Carty, Mark Andrade, Dennis Gorman, Xuefeng Zhu, Isaiah Cedillo, Zhiwei Wang,

Lucio Mendez, Anthony N. Scozzari, Gerardo Aguirre, Ratnasamy Somanathan, and Sylvain Berneès. UnyLinker: An efficient and scaleable synthesis of oligonucleotides utilizing a universal linker molecule: A novel approach to enhance the purity of drugs. *Organic Process Research and Development*, 12(3):399–410, 2008.

[86] GE Healthcare. 1-µmol synthesis of RNA using ÄKTA oligopilot plus 10, 2006.

[87] L. S. Ettre and K. I. Sakodynskii. M. S. Tswett and the discovery of chromatography I: Early work (1899-1903). *Chromatographia*, 35(3-4):223–231, 1993.

[88] Ken Cook and Jim Thayer. Advantages of ion-exchange chromatography for oligonucleotide analysis. *Bioanalysis*, 3(10):1109–1120, 2011.

[89] Alexandre Goyon, Peter Yehl, and Kelly Zhang. Characterization of therapeutic oligonucleotides by liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis*, 182:113105, 2020.

[90] B. A. Bidlingmeyer, S. N. Deming, W. P. Price, B. Sachok, and M. Petrusek. Retention mechamism for reversed-phase ion-pair liquid chromatography. *Journal of Chromatography A*, 186(C):419–434, 1979.

[91] Jan H. Overbeck, Werner Kremer, and Remco Sprangers. A suite of 19F based relaxation dispersion experiments to assess biomolecular motions. *Journal of Biomolecular NMR*, 74(12):753–766, 2020.

[92] Daehyun Baek, Judit Villén, Chanseok Shin, Fernando D. Camargo, Steven P. Gygi, and David P. Bartel. The impact of microRNAs on protein output. *Nature*, 455(7209):64–71, 2008.

[93] Stefan L. Ameres and Phillip D. Zamore. Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology*, 14(8):475–488, 2013.

[94] Yu Meng Sun and Yue Qin Chen. Principles and innovative technologies for decrypting noncoding RNAs: From discovery and functional prediction to clinical application. *Journal of Hematology and Oncology*, 13(1):1–27, 2020.

[95] Jun Inoue and Johji Inazawa. Cancer-associated miRNAs and their therapeutic potential. *Journal of Human Genetics*, 66(9):937–945, 2021.

[96] Lin He and Gregory J. Hannon. MicroRNAs: Small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, 5(7):522–531, 2004.

[97] Michael J. Axtell, Jakub O. Westholm, and Eric C. Lai. Vive la différence: Biogenesis and evolution of microRNAs in plants and animals. *Genome Biology*, 12(4):1–13, 2011.

[98] Shuibin Lin and Richard I. Gregory. MicroRNA biogenesis pathways in cancer. *Nature Reviews Cancer*, 15(6):321–333, 2015.

[99] David P. Bartel. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, 116(2):281–297, 2004.

[100] Yoontae Lee, Minju Kim, Jinju Han, Kyu Hyun Yeom, Sanghyuk Lee, Sung Hee Baek, and V. Narry Kim. MicroRNA genes are transcribed by RNA polymerase II. *EMBO Journal*, 23(20):4051–4060, 2004.

[101] Keita Miyoshi, Tomoko N. Okada, Haruhiko Siomi, and Mikiko C. Siomi. Characterization of the miRNA-RISC loading complex and miRNA-RISC formed in the Drosophila miRNA pathway. *Rna*, 15(7):1282–1291, 2009.

[102] Tomoko Kawamata and Yukihide Tomari. Making RISC. *Trends in Biochemical Sciences*, 35(7):368–376, 2010.

[103] Sergei B. Koralov, Stefan A. Muljo, Gunther R. Galler, Azra Krek, Tirtha Chakraborty, Chryssa Kanellopoulou, Kari Jensen, Bradley S. Cobb, Matthias Merkenschlager, Nikolaus Rajewsky, and Klaus Rajewsky. Dicer Ablation Affects Antibody Diversity and Cell Survival in the B Lymphocyte Lineage. *Cell*, 132(5):860–874, 2008.

[104] Yvonne Tay, Jinqiu Zhang, Andrew M. Thomson, Bing Lim, and Isidore Rigoutsos. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, 455(7216):1124–1128, 2008.

[105] Rui Yi, Matthew N. Poy, Markus Stoffel, and Elaine Fuchs. A skin microRNA promotes differentiation by repressing 'stemness'. *Nature*, 452(7184):225–229, 2008.

[106] Hiro oki Iwakawa and Yukihide Tomari. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends in Cell Biology*, 25(11):651–665, 2015.

[107] Takashi Fukaya and Yukihide Tomari. MicroRNAs Mediate Gene Silencing via Multiple Different Pathways in Drosophila. *Molecular Cell*, 48(6):825–836, 2012.

[108] Sylvia K. Shenouda and Suresh K. Alahari. MicroRNA function in cancer: Oncogene or a tumor suppressor?, 2009.

[109] Aurora Esquela-Kerscher and Frank J. Slack. Oncomirs - MicroRNAs with a role in cancer, 2006.

[110] Bruce Wightman, Ilho Ha, and Gary Ruvkun. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75(5):855–862, 1993.

[111] BJ Reinhart, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, and Ruvkun GB. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403:901–906, 2000.

[112] Rhonda Feinbaum, Victor Ambros, and Rosalind Lee. The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to lin-14. *Cell*, 116(116):843–854, 2004.

[113] Amy E. Pasquinelli, Brenda J. Reinhart, Frank Slack, Mark Q. Martindale, Mitzi I. Kuroda, Betsy Maller, David C. Hayward, Eldon E. Ball, Bernard Degnan, Peter Müller, Jürg Spring, Ashok Srinivasan, Mark Fishman, John Finnerty, Joseph Corbo, Michael Levine, Patrick Leahy, Eric Davidson, and Gary Ruvkun. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–89, 2000.

[114] R. C. Lee and V. Ambros. An extensive class of small RNAs in Caenorhabditis elegans. *Science*, 294(5543):862–864, 2001.

[115] Marc Robert Fabian, Nahum Sonenberg, and Witold Filipowicz. Regulation of mRNA translation and stability by microRNAs. *Annual Review of Biochemistry*, 79:351–379, 2010.

[116] Mirko Cevec, Christophe Thibaudeau, and Janez Plavec. Solution structure of a let-7 miRNA: Lin-41 mRNA complex from C. elegans. *Nucleic Acids Research*, 36(7):2330–2337, 2008.

[117] Madhu S. Kumar, Stefan J. Erkeland, Ryan E. Pester, Cindy Y. Chen, Margaret S. Ebert, Phillip A. Sharp, and Tyler Jacks. Suppression of non-small cell lung tumor development by the let-7 microRNA family. *Proceedings of the National Academy of Sciences of the United States of America*, 105(10):3903–3908, 2008.

[118] Junichi Takamizawa, Hiroyuki Konishi, Kiyoshi Yanagisawa, Shuta Tomida, Hirotaka Osada, Hideki Endoh, Tomoko Harano, Yasushi Yatabe, Masato Nagino, Yuji Nimura, Tetsuya Mitsudomi, and Takashi Takahashi. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Research*, 64(11):3753–3756, 2004.

[119] Charles D. Johnson, Aurora Esquela-Kerscher, Giovanni Stefani, Mike Byrom, Kevin Kelnar, Dmitriy Ovcharenko, Mike Wilson, Xiaowei Wang, Jeffrey Shelton, Jaclyn Shingara, Lena Chin, David Brown, and Frank J. Slack. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Research*, 67(16):7713–7722, 2007.

[120] Benjamin Boyerinas, Sun Mi Park, Annika Hau, Andrea E. Murmann, and Marcus E. Peter. The role of let-7 in cell differentiation and cancer. *Endocrine-Related Cancer*, 17(1):19–36, 2010.

[121] Ramiro Garzon, Guido Marcucci, and Carlo M. Croce. Targeting microRNAs in cancer: Rationale, strategies and challenges, 2010.

[122] Man Lung Yeung and Kuan Teh Jeang. MicroRNAs and cancer therapeutics. *Pharmaceutical Research*, 28(12):3043–3049, 2011.

[123] Simon Hirschberger, Ludwig Christian Hinske, and Simone Kreth. MiRNAs: dynamic regulators of immune cell functions in inflammation and cancer, 2018.

[124] Fengyan Yu, Herui Yao, Pengcheng Zhu, Xiaoqin Zhang, Qiuhui Pan, Chang Gong, Yijun Huang, Xiaoqu Hu, Fengxi Su, Judy Lieberman, and Erwei Song. let-7 Regulates Self Renewal and Tumorigenicity of Breast Cancer Cells. *Cell*, 131(6):1109–1123, 2007.

[125] Janaiah Kota, Raghu R. Chivukula, Kathryn A. O'Donnell, Erik A. Wentzel, Chrystal L. Montgomery, Hun Way Hwang, Tsung Cheng Chang, Perumal Vivekanandan, Michael Torbenson, K. Reed Clark, Jerry R. Mendell, and Joshua T. Mendell. Therapeutic microRNA Delivery Suppresses Tumorigenesis in a Murine Liver Cancer Model. *Cell*, 137(6):1005–1017, 2009.

[126] Jean Hausser and Mihaela Zavolan. Identification and consequences of miRNA-target interactions-beyond repression of gene expression, 2014.

[127] Agnieszka Belter, Dorota Gudanis, Katarzyna Rolle, Monika Piwecka, Zofia Gdaniec, Mirosława Z. Naskre¸t-Barciszewska, and Jan Barciszewski. Mature miRNAs form secondary structure, which suggests their function beyond RISC. *PLoS ONE*, 9(11), 2014.

[128] Amy E. Pasquinelli. MicroRNAs and their targets: Recognition, regulation and an emerging reciprocal relationship. *Nature Reviews Genetics*, 13(4):271–282, 2012.

[129] Shveta Bagga, John Bracht, Shaun Hunter, Katlin Massirer, Janette Holtz, Rachel Eachus, and Amy E. Pasquinelli. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell*, 122(4):553–563, 2005.

[130] John G. Doench and Phillip A. Sharp. Specificity of microRNA target selection in translational repression. *Genes and Development*, 18(5):504–511, 2004.

[131] Matthias Selbach, Björn Schwanhäusser, Nadine Thierfelder, Zhuo Fang, Raya Khanin, and Nikolaus Rajewsky. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63, 2008.

[132] G. M. Innocenti. The time when the "Tomte" of evolution was playing with time. *Behavioral and Brain Sciences*, 24(2):287, 2001.

[133] Jun Yang, Brian D. Bennett, Shujun Luo, Kaoru Inoue, Sara A. Grimm, Gary P. Schroth, Pierre R. Bushel, H. Karimi Kinyamu, and Trevor K. Archer. LIN28A Modulates Splicing and Gene Expression Programs in Breast Cancer Cells. *Molecular and Cellular Biology*, 35(18):3225–3243, 2015.

[134] Elena Piskounova, Christos Polytarchou, James E. Thornton, Robert J. Lapierre, Charalabos Pothoulakis, John P. Hagan, Dimitrios Iliopoulos, and Richard I. Gregory. Lin28A and Lin28B inhibit let-7 MicroRNA biogenesis by distinct mechanisms. *Cell*, 147(5):1066–1079, 2011.

[135] Julien Balzeau, Miriam R. Menezes, Siyu Cao, and John P. Hagan. The LIN28/let-7 pathway in cancer. *Frontiers in Genetics*, 8(MAR):1–16, 2017.

[136] Matyas Ecsedi and Helge Großhans. Lin-41/Trim71: Emancipation of a miRNA target. *Genes and Development*, 27(6):581–589, 2013.

[137] Betsy R.Maller Schulman, Aurora Esquela-Kerscher, and Frank J. Slack. Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. *Developmental Dynamics*, 234(4):1046–1054, 2005.

[138] Monica C. Vella, Eun Young Choi, Shin Yi Lin, Kristy Reinert, and Frank J. Slack. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3˘UTR. *Genes and Development*, 18(2):132–137, 2004.

[139] Christine Mayr, Michael T. Hemann, and David P. Bartel. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science*, 315(5818):1576–1579, 2007.

[140] Sun Lee Yong and Anindya Dutta. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes and Development*, 21(9):1025–1030, 2007.

[141] Barbara Puffer, Christoph Kreutz, Ulrike Rieder, Marc Olivier Ebert, Robert Konrat, and Ronald Micura. 5-Fluoro pyrimidines: Labels to probe DNA and RNA secondary structures by 1D 19F NMR spectroscopy. *Nucleic Acids Research*, 37(22):7728–7740, 2009.

[142] Qiang Li, Jialiang Chen, Marko Trajkovski, Yifei Zhou, Chaochao Fan, Kuan Lu, Pingping Tang, Xuncheng Su, Janez Plavec, Zhen Xi, and Chuanzheng Zhou. 4˘-Fluorinated RNA: Synthesis, Structure, and Applications as a Sensitive 19F NMR Probe of RNA Structure and Function. *Journal of the American Chemical Society*, 142(10):4739–4748, 2020.

[143] Caijie Zhao, Matthew Devany, and Nancy L. Greenbaum. Measurement of chemical exchange between RNA conformers by 19F NMR. *Biochemical and Biophysical Research Communications*, 453(4):692–695, 2014.

[144] Amritraj Patra, Michael Paolillo, Klaus Charisse, Muthiah Manoharan, Eriks Rozners, and Martin Egli. 2$\breve{}$-fluoro RNA shows increased watson-crick H-bonding strength and stacking relative to RNA: Evidence from NMR and thermodynamic Data. *Angewandte Chemie - International Edition*, 51(47):11863–11866, 2012.

[145] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, 2003.

[146] Diana Gimenez, Aoife Phelan, Cormac D. Murphy, and Steven L. Cobb. 19F NMR as a tool in chemical biology. *Beilstein Journal of Organic Chemistry*, 17:293–318, 2021.

[147] Mirko Hennig, Lincoln G. Scott, Edit Sperling, Wolfgang Bermel, and James R. Williamson. Synthesis of 5-fluoropyrimidine nucleotides as sensitive NMR probes of RNA structure. *Journal of the American Chemical Society*, 129(48):14911–14921, 2007.

[148] Heike Betat, Yicheng Long, Jane E. Jackman, and Mario Mörl. From end to end: TRNA editing at 5$\breve{}$-and 3$\breve{}$-terminal positions. *International Journal of Molecular Sciences*, 15(12):23975–23998, 2014.

[149] Pierre Barraud and Carine Tisné. To be or not to be modified: Miscellaneous aspects influencing nucleotide modifications in tRNAs. *IUBMB Life*, 71(8):1126–1140, 2019.

[150] Pierre Barraud, Alexandre Gato, Matthias Heiss, Marjorie Catala, Stefanie Kellner, and Carine Tisné. Time-resolved NMR monitoring of tRNA maturation. *Nature Communications*, 10(1), 2019.

[151] Sarah Kai Leigh Schultz and Ute Kothe. TRNA elbow modifications affect the tRNA pseudouridine synthase TruB and the methyltransferase TrmA. *Rna*, 26(9):1131–1142, 2020.

[152] Mirosława Z. Barciszewska, Patrick M. Perrigue, and Jan Barciszewski. TRNA-the golden standard in molecular biology. *Molecular BioSystems*, 12(1):12–17, 2016.

[153] Stephen Cusack. Aminoacyl-tRNA synthetases. *Current Opinion in Structural Biology*, 7(6):881–889, 1997.

[154] Christian Lorenz, Christina E. Lünse, and Mario Mörl. Trna modifications: Impact on structure and thermal adaptation. *Biomolecules*, 7(2), 2017.

[155] Hashim M. Al-Hashimi. NMR studies of nucleic acid dynamics. *Journal of Magnetic Resonance*, 237:191–204, 2013.

[156] Richard Giegé. Toward a more complete view of tRNA biology. *Nature Structural and Molecular Biology*, 15(10):1007–1014, 2008.

[157] Paul J. Sapienza and Andrew L. Lee. Using NMR to study fast dynamics in proteins: Methods and applications. *Current Opinion in Pharmacology*, 10(6):723–730, 2010.

[158] Charles L. Parsons. The American Chemical Society. *Science*, 54(1383):13, 1921.

[159] Donghan Lee, Christian Hilty, Gerhard Wider, and Kurt Wüthrich. Effective rotational correlation times of proteins from NMR relaxation interference. *Journal of Magnetic Resonance*, 178(1):72–76, 2006.

[160] J. García De La Torre, M. L. Huertas, and B. Carrasco. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *Journal of Magnetic Resonance*, 147(1):138–146, 2000.

[161] Thomas Biedenbänder, Vanessa De Jesus, Martina Schmidt-dengler, Mark Helm, Björn Corzilius, and Johannes Gutenberg-universität Mainz. Supplementary Material RNA modifications stabilize the tertiary structure of tRNA fMet by locally increasing conformational dynamics Extended Figure Caption :. pages 1–31.

[162] D. Flemming Hansen, Pramodh Vallurupalli, Patrik Lundström, Philipp Neudecker, and Lewis E. Kay. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? *Journal of the American Chemical Society*, 130(8):2667–2675, 2008.

[163] Niels Fischer, Andrey L. Konevega, Wolfgang Wintermeyer, Marina V. Rodnina, and Holger Stark. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, 466(7304):329–333, 2010.

[164] Tyler J. Smith, Mehmet Tardu, Hem Raj Khatri, and Kristin S. Koutmou. mRNA and tRNA modification states influence ribosome speed and frame maintenance during poly(lysine) peptide synthesis. *Journal of Biological Chemistry*, 298(6):102039, 2022.

[165] Tao Pan. Modifications and functional genomics of human transfer RNA. *Cell Research*, 28(4):395–404, 2018.

[166] Curtis A. Suttle. Marine viruses - Major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10):801–812, 2007.

[167] Mart Krupovic, David Prangishvili, Roger W. Hendrix, and Dennis H. Bamford. Genomics of Bacterial and Archaeal Viruses: Dynamics within the Prokaryotic Virosphere. *Microbiology and Molecular Biology Reviews*, 75(4):610–635, 2011.

[168] Eric C. Keen. A century of phage research: Bacteriophages and the shaping of modern biology. *BioEssays*, 37(1):6–9, 2015.

[169] Eric C. Keen. A century of phage research: Bacteriophages and the shaping of modern biology. *BioEssays*, 37(1):6–9, 2015.

[170] William C. Summers. The strange history of phage therapy. *Bacteriophage*, 2(2):130–133, 2012.

[171] Nina Chanishvili. *Phage Therapy-History from Twort and d'Herelle Through Soviet Experience to Current Approaches*, volume 83. Elsevier Inc., 1 edition, 2012.

[172] Nina Chanishvili and Richard Sharp. Bacteriophage therapy: experience from the Eliava Institute, Georgia. *Microbiology Australia*, 29(2):96, 2008.

[173] Franklin L. Nobrega, Marnix Vlot, Patrick A. de Jonge, Lisa L. Dreesens, Hubertus J.E. Beaumont, Rob Lavigne, Bas E. Dutilh, and Stan J.J. Brouns. Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology*, 16(12):760–773, 2018.

[174] Cynthia Maria Chibani, Anton Farr, Sandra Klama, Sascha Dietrich, and Heiko Liesegang. Classifying the unclassified: A phage classification method. *Viruses*, 11(2), 2019.

[175] Nahla M Mansour. Bacteriophages are natural gift, could we pay further attention! *Journal Food mircobiology*, 1(1):22–24, 2017.

[176] V. Braun, H. Killmann, and C. Herrmann. Inactivation of FhuA at the cell surface of Escherichia coli K-12 by a phage T5 lipoprotein at the periplasmic face of the outer membrane. *Journal of Bacteriology*, 176(15):4710–4717, 1994.

[177] Katja Decker, Verena Krauel, Anke Meesmann, and Knut J. Heller. Lytic conversion of Escherichia coli by bacteriophage T5: blocking of the FhuA receptor protein by a lipoprotein expressed early during infection. *Molecular Microbiology*, 12(2):321–332, 1994.

[178] L. Plançon, C. Janmot, M. Le Maire, M. Desmadril, M. Bonhivers, L. Letellier, and P. Boulanger. Characterization of a high-affinity complex between the bacterial outer membrane protein FhuA and the phage T5 protein pb5. *Journal of Molecular Biology*, 318(2):557–569, 2002.

[179] Ali Flayhan, Frank Wien, Maïté Paternostre, Pascale Boulanger, and Cécile Breyton. New insights into pb5, the receptor binding protein of bacteriophage T5, and its interaction with its Escherichia coli receptor FhuA. *Biochimie*, 94(9):1982–1989, 2012.

[180] Cécile Breyton, Ali Flayhan, Frank Gabel, Mathilde Lethier, Grégory Durand, Pascale Boulanger, Mohamed Chami, and Christine Ebel. Assessing the conformational changes of pb5, the receptor-binding protein of phage T5, upon binding to its escherichia coli receptor FhuA. *Journal of Biological Chemistry*, 288(42):30763–30772, 2013.

[181] Ivo Pedruzzi, Jurg P. Rosenbusch, and Kaspar P. Locher. Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein. *FEMS Microbiology Letters*, 168(1):119–125, 1998.

[182] Frank Delaglio, Stephan Grzesiek, Geerten W. Vuister, Guang Zhu, John Pfeifer, and Ad Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293, 1995.

[183] Pascale Tsan, Jean Christophe Hus, Michael Caffrey, Dominique Marion, and Martin Blackledge. Rotational diffusion anisotropy and local backbone dynamics of carbon monoxide-bound Rhodobacter capsulatus cytochrome c'. *Journal of the American Chemical Society*, 122(23):5603–5612, 2000.

[184] Danyel Fernandes Contiliani, Yasmin de Araújo Ribeiro, Vitor Nolasco de Moraes, and Tiago Campos Pereira. Micrornas in prion diseases—from molecular mechanisms to insights in translational medicine. *Cells*, 10(7):1–24, 2021.