Setting up a method to detect ligand-macromolecule interaction through 19F NMR

Raphaël Recht

UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé

IGBMC – CNRS UMR 7104 – Inserm U 964

THÈSE présentée par :

Raphaël RECHT

soutenue le : **23 septembre 2016**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Biophysique

Mise au point de méthodes de détection d'interaction ligand-macromolécule par RMN du ^{19}F

Setting up a method to detect ligand-macromolecule interaction through ^{19}F NMR

THÈSE dirigée par :

M. KIEFFER Bruno	Directeur de recherche, Université de Strasbourg
M. YUSUPOV Marat	Directeur de recherche, Université de Strasbourg

RAPPORTEURS :

M. JAHNKE Wolfgang	Senior Research Investigator, Novartis Institues for Biomedical Research
M. MARTINS José	Professor, Ghent University

AUTRES MEMBRES DU JURY :

Mme KELLENBERGER Esther	Professeur, Université de Strasbourg
M. PIOTTO Martial	Responsable des projets collaboratifs, Bruker Biospin

THANKS

A thesis is everything but a straight line. In my case, I would define it as driving a car that cannot steer on its own so the trajectory is corrected with each crash (illustrated in figure 1).

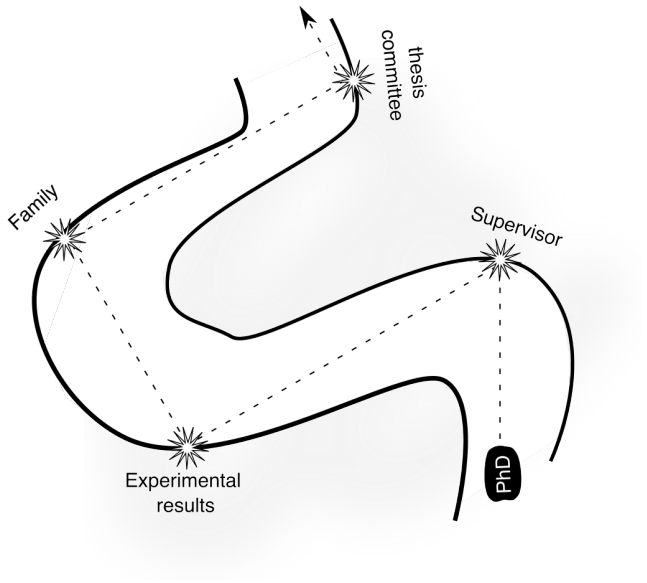


Fig 1. Schematic of the path I took during my PhD.

So I will thank all the people that assisted me in taking the right direction. I must start by thanking my jury members: Dr. Wolfgang Jahnke, Pr. José Martins, Dr. Esther Kellenberger, Dr. Martial Piotto, for accepting to review my work, I feel truly honored by this.

Then I'll thank my PhD supervisors, Pr. Marat Yusupov and Pr. Bruno Kieffer. Marat, thanks for your kindness and your wise advices. Bruno, congratulations on your patience with me. You could have killed me a hundred times and be justified to do so, due to my lack of scientific rigor, yet I am still breathing. So you are either the worst killer ever or the best pedagogue ever. I tend to think that it is the later…

I thank all of my colleagues, that helped me, at a moment or another: Claude Ling (who set up with me record times to install/uninstall cryoprobes), Lionel Chiron (absolute expert in Python), Yves Nominé (unstoppable life-form, helped me especially in maths & stats), Isabelle Lebars (who always had the right advice), Marc-Andre Delsuc (the reference in signal

processing), Emmanuel Oppong (works in the wet lab, but I still like him), Juan Ramirez (for all the help on the PDZ) … And I'm forgetting many many colleagues…

In regards to the other PhD students that I had the chance to work with, I thank Sandra Meyer (who never stops, I was always impressed), Julia Asencio-Hernandez (too shy to admit that she is a talented chemist, and now NMR spectroscopist), Justine Vieville (that lent me the BBFO probe many times!). In the same topic, the NMR Chemistry team (Lionel Allouche, Bruno Vincent…), thanks you for having tolerated my numerous BBFO kidnappings.

Outside of the NMR magnetic field, I spent the last 8 years studying alongside Jérôme Eberhardt so I reward him quickly by a mention, this way, when I'll be a world-famous scientist, he will be able to prove that he knew me. There are also Blum Matthias and Chollet Pierre-Etienne that I know from my Master Thesis and now work at the IGBMC, in company of Valerya Malysheva, Anna Belorusova and Mickael Dumas. I have to thank all the members of the Student and Post-Doc association of the IGBMC, where I was proud to help a little bit. I started at the same time as David Bruchlen and Karima Tazibt and I'll see them at their own thesis defense. I was happy to collaborate with Oyindamola Oladosu, a very fresh PhD student. It is the same for Iskander Khusainov, who was a great partner in attack during our football sessions. Valerio Taverniti and Simone Pellegrino were always nice to me, as were Michel Ney, Perrine Strasser, Justine Mailiot, etc. Anyway, I forget again many peoples and I apologize for it. The key message is that the IGBMC is a big institute, that is an incredible luck for a young student, as many friendships can start and from a scientific viewpoint, it is easy to learn a little bit of everyone's work. I will miss this environment.

I thanks all members of my family, especially my wife (for keeping my daughter busy when I needed to write) and my daughter (for keeping my wife busy when I needed to write).

All of this wouldn't be worth it without them.

*

RESUME DE LA THESE

Mise au point de méthodes de détection d'interaction ligand-macromolécule par RMN du ^{19}F

Recht Raphaël

Superviseurs : Kieffer Bruno et Yusupov Marat

Les interactions biologiques sont régies par des mécanismes complexes, qui mêlent différentes échelles, de temps comme de taille. C'est le cas du ribosome, un complexe nucléoprotéique responsable de la traduction de l'ARNm en protéines, et ce faisant, une cible thérapeutique primordiale. Or la taille du ribosome procaryote 70S (2.4 MDa) rend difficile l'applications des techniques classiques de criblage de ligands. Au cours de ma thèse, j'ai exploré la possibilité d'utiliser la RMN du fluor pour caractériser les interactions entre des ligands et le ribosome procaryote. Cette approche a été motivée par l'apport de nouvelles méthodes de détection pouvant coupler la versatilité de la RMN (Résonance Magnétique Nucléaire) avec les propriétés de l'atome de fluor. L'atome ^{19}F se prête parfaitement à la RMN, avec son rapport gyromagnétique proche du proton et son abondance isotopique naturelle de 100%. De plus, le fluor est bio-orthogonal au Vivant. Enfin, les caractéristiques physico-chimiques du fluor sont bien exploitées dans la pharmacopée (un quart des antibiotiques en possèdent un groupement).

Ces travaux ont été conduits sous la double direction de Bruno Kieffer, pour l'expertise spectroscopique, et Marat Yusupov, pour la production des échantillons de ribosome. La méthode spectroscopique a été mise au point à partir du modèle théorique FAXS (Fluorine chemical shift Anisotropy and eXchange for screening) (Dalvit et al, 2003). Cette méthode repose sur l'observation de la relaxation transversale d'un atome de fluor présent sur un ligand. Cette relaxation est très différente, selon que le ligand soit libre ou lié à sa cible. Les conditions d'étude ont demandé une phase de mise au point significative du fait des particularités du ribosome. Ainsi, le projet a nécessité le développement de molécules témoins négatifs de l'interaction, adapté aux caractéristiques particulières du ribosome. Il a également fallu optimiser les quantités de ligand et de ribosome à mettre en solution. La principale qualité de l'approche envisagée au sein de ce projet est de cribler les

ligands directement sur le ribosome entier. Cela permet de se rapprocher au plus près des conditions in-vivo.

Les molécules testées initialement ont été le Linezolid (famille des oxazolidinones), le Florfenicol (phénicolés) ainsi que la Flurithromycine (macrolides). Le Linezolid est un antibiotique disponible commercialement sous le nom « Zyvox », de Pfizer. Il se lie dans le peptidyl-transferase center (PTC) du ribosome comme le montre la figure 1.

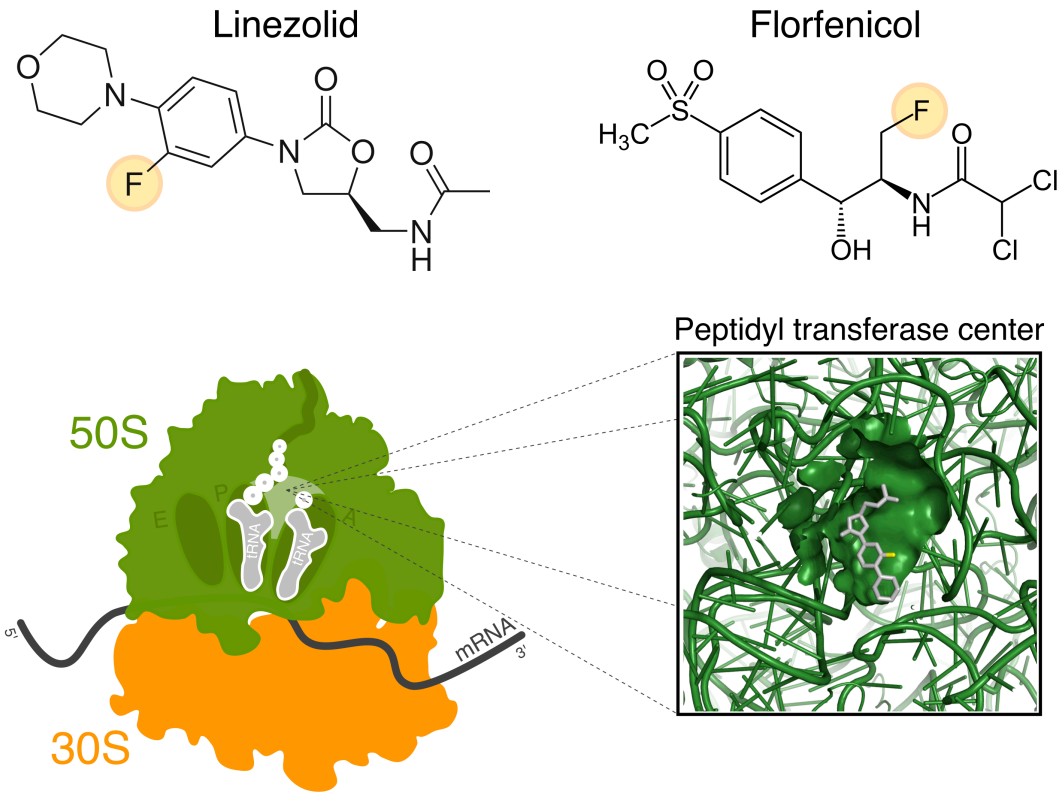


Fig 1. Structures chimiques du linezolid et du florfenicol, avec leurs atomes de fluor mis en évidence. Un schéma de la structure du ribosome bactérien est dessiné en-dessous avec un encart reproduisant le site de liaison du linezolid, au cœur du site de transfert catalytique de la liaison peptidique.

La liaison du linezolid au ribosome a pu être observée dans différentes conditions (température, concentration). La compétition pour son site de liaison a pu être testée avec des ligands connus se liant également dans le PTC, la Lincomycine et le Chloramphénicol. Ces expériences de compétition entre la molécule testée et un ligand connu, dit « espion », ont été complétées avec une molécule ne se liant pas dans le PTC, la Kanamycine. Un lot

d'expériences reprenant des paramètres identiques a pu être conduit sur le Florfenicol, une molécule dont la formule chimique diffère très peu du Chloramphenicol. Les résultats obtenus lors de titrations avec le ribosome ont montré des similarités globales avec le Linezolid mais également des différences fines. Ces distinctions permettent de mettre en évidence les affinités relatives de chacun des deux antibiotiques. La méthode créée dans le cadre de ce projet a également été appliquée sur des composés en phase de développement. Ce fut le cas pour une molécule membre de la famille des 4HPP, une nouvelle classe d'inhibiteurs de la synthèse procaryotique (Ribble et al., 2010). Le site de la molécule étudiée n'est cependant pas encore connu. Les expériences menées suivant la méthode ont permis de caractériser par RMN la structure chimique de la molécule de 4HPP puis de confirmer sa liaison au ribosome.

De manière similaire des peptides antibiotiques fluorés, synthétisés par la compagnie française Nosopharm, ont pu être criblés via notre méthode. Nous avons pu démontrer que le site de liaison de ces peptides se trouvait bien sur le ribosome. En effet, l'addition répétée de doses croissantes de ribosome conduit à une altération distincte des pics de fluor des peptides en solution comme le montre la figure 2. De plus, différentes molécules témoins présentes également dans la solution n'ont elles montrées aucune réaction. Enfin, l'expérience de criblage a été menée sur le ribosome de *Thermus thermophilus* mais également sur celui d'un eucaryote, *Escherichia coli*, avec des résultats identiques.

Les différentes expériences menées au cours de cette étude ont révélé la difficulté à établir le caractère spécifique de l'interaction. En effet, la mise en évidence de l'interaction compétitive est brouillée par la présence d'un grand nombre d'interactions non spécifiques se produisant à différentes échelles d'affinité. Ce phénomène est accentué par la différence de taille entre les interactants (quelques centaines de Dalton pour un antibiotique contre des millions de Dalton pour le ribosome). Afin de finaliser la méthode mise au point durant la thèse, la fin de 3^{e} année ainsi que la 4^{e} ont été consacrées à réduire ce problème de spécificité. Plusieurs paramètres ont pu être testés dont la concentration de ligand ou de ribosome, et la température.

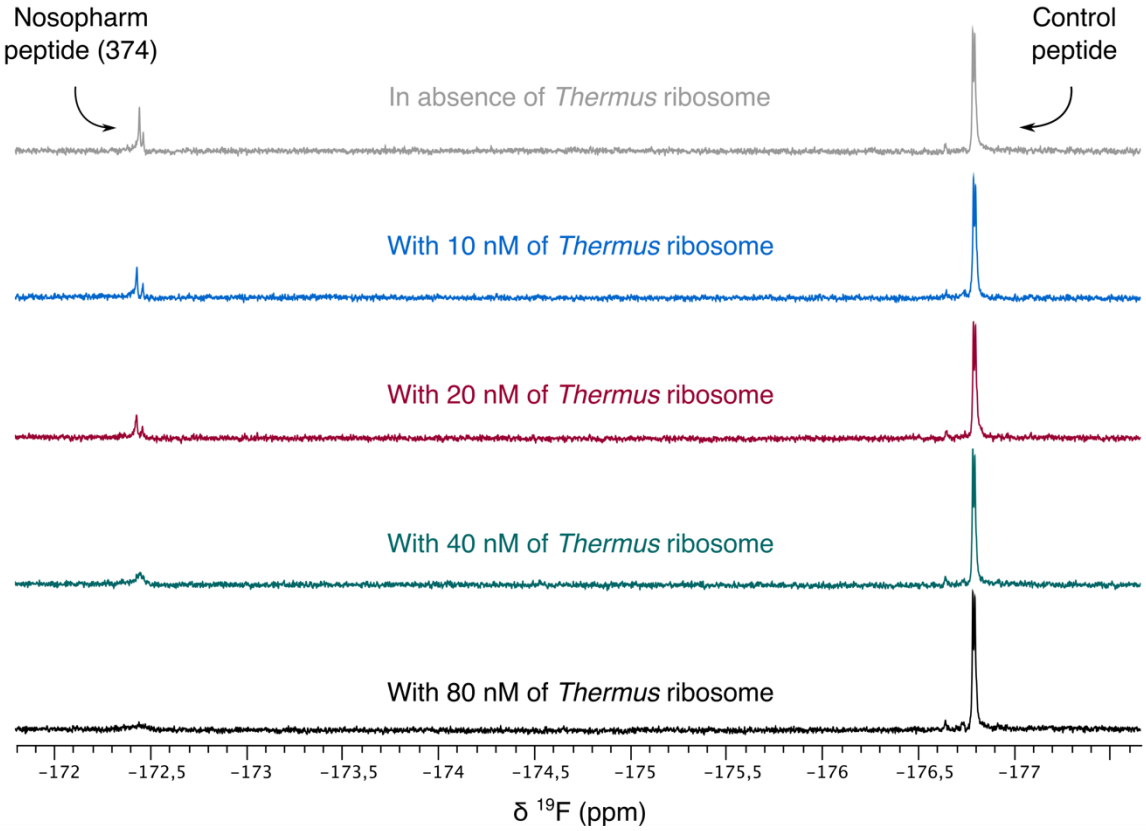


Fig 2. Superposition de spectres RMN issus de la titration de deux peptides (N374 de Nosopharm et un peptide contrôle) par des quantités croissantes de ribosome 70S de *Thermus thermophilus*.

Au cours de ma thèse, j'ai également étudié d'autres problématiques biologiques. Notamment l'interaction entre un domaine riche en prolines situé sur le DBD (DNA Binding Domain) de RARγ (Retinoic Acid Receptor) et le domaine SH3 d'un co-régulateur de la Vinexineβ. Cette interaction est modulée par la phosphorylation d'une sérine située au milieu du domaine riche en proline. Ainsi, des peptides comportant des fluoroprolines construits à partir du DBD de RAR-γ ont été étudiés par RMN du proton mais également du fluor. La figure 3 illustre l'interaction entre un de ces peptides fluorés et le domaine SH3. L'utilisation du noyau fluor permet d'étudier les conséquences de la phosphorylation sur le comportement de la région polyproline et d'expliquer les affinités observées. De plus, les paramètres cinétiques de l'interaction peuvent également être obtenus à partir de l'étude des spectres fluor. Ces développements ont pu être directement utilisés dans le criblage des peptides de Nosopharm, ceux-ci possédant également une fluoroproline.

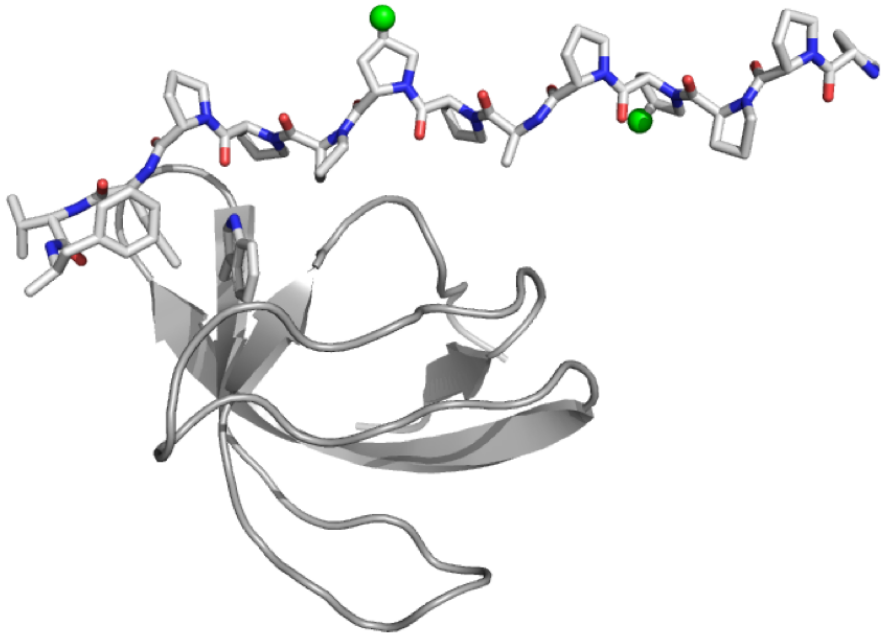


Fig 3. Représentation de la structure du domaine SH33 de la vinexinβ et d'un peptide de séquence $PPP^{19F}PPAP^{19F}PPPPRVYK$. Les atomes de fluor sont représentés en vert et ils sont orientés non pas vers le domaine mais en direction du solvant afin de ne pas perturber l'interaction.

En conclusion, la méthode de criblage de ligands sur le ribosome modèle de *Thermus* bâtie autour de la RMN du fluor a donné des résultats probants sur des ligands antibiotiques connus, ainsi que sur de nouveaux candidats. Différents types d'interactions ont pu mettre en évidence les domaines d'application de la méthode mais aussi ses limites, ouvrant des voies de développement. Ces pistes de recherche sont maintenant complétées par des tests de techniques NMR différentes (Water-Logsy, DEST). Une publication détaillant la méthode ainsi que les résultats obtenus est en cours d'écriture. Au cours du doctorat, deux papiers scientifiques ont pu être publiés, l'un résultant de travaux effectués sur le papillomavirus humain lors du stage de Master 2 et un autre se concentrant sur la méthode optimale pour effectuer un dosage ligand-protéine par RMN.

LISTE DES PUBLICATIONS

Köhler C, Recht R, Quinternet M, De lamotte F, Delsuc MA, Kieffer B. Accurate protein-peptide titration experiments by nuclear magnetic resonance using low-volume samples. Methods Mol Biol. 2015;1286:279-96.

Ramírez J, Recht R, Charbonnier S, et al. Disorder-to-order transition of MAGI-1 PDZ1 C-terminal extension upon peptide binding: thermodynamic and dynamic insights. Biochemistry. 2015;54(6):1327-37.

TABLE OF CONTENTS

INTRODUCTION

1. Structural and functional overview of the ribosome

1.1 Historical discovery

In a few simple words, ribosomes are sub-microscopic -but dense- molecules found in all living cells and they are responsible for the protein production. Ribosomes are nowadays known to be the cellular protein factory and still are the focus of many biological studies. As a testimony to the intense scientific unrest that aimed at deciphering the core mechanism of the protein synthesis, 17 years passed between the first observation of the ribosomes (in 1941) and their final naming (in 1958) (reviewed by (Tissières-1974) and (Bąkowska-Żywicka-2009)). In that timespan, they had different names, adopted from the scientific pioneers of that era: Claude's particles, Palade's particles… The discovery of ribosomes was preceded by two related observations made in the thirties: RNA was found mainly located in the cytoplasm by Brachet as soon as 1933 (Brachet-1933) and Caspersson established a relationship between the amount of RNA and tissue growth in 1938 (Caspersson – 1938). The work on cell fractionation by Albert Claude led to its observation in 1941 of particles (Claude-1941), that he named "small granules" then "microsomes" (from Greek σῶμα, *sôma*, meaning body). Claude was limited in its discovery by its instrumentation, the optic microscope, unable to resolve the cell's interior detail. Ultracentrifugal analysis of microsome fractions (from mice spleen in 1952) by Petermann and Hamilton showed discrete peaks, ranging from 40S to 76S (Petermann-1952). Those fractions after electrophoresis showed high content of RNA as well as aminoacids, hinting at a connection between the microsomes and the protein biosynthesis. The use of an electronic microscope allowed in 1955 George Palade (a student of Claude) to show that the microsomes are fragments of endoplasmic reticulum and that on the surface of those, dense granules are present, in all the tissues examined (Palade-1955). Palade and Claude, in association with de Duve, were recompensed by the 1974 Nobel Prize in Physiology or Medicine for their work. The name "ribosome" was proposed and adopted in the 1958 meeting of the Biophysical Society by Richard Brooke Roberts, pointing out the ribonucleic acid part. At this time, scientists purified ribosomes from many organisms and started to explore their physico-chemical properties. In 1956, Schachmann and Chao isolated stable eukaryotic ribosomes (from yeast extracts) with a sedimentation coefficient of 80S, and noticed that they

dissociate into two portions of 60S and 40S (Chao-1956). Two years later, Alfred Tissières and James Watson characterized the 70S particle of prokaryotes (with *Escherichia coli*), as well as the dissociation in 50S and 30S sub-units (Tissières-1958). The Svedberg coefficient units of the particles stuck as their names.

1.2 Functional representation

Hans Noll, one of the scientists involved in the early stages of the ribosome discovery, remembers a stroll he took with Jim Watson in which Noll challenged his friend to explain how a single ribosome could translate a messenger RNA many times its diameter (Noll-2008). Watson showed Noll that the messenger strand could be read by moving like a tape through a fixed read-out site. While rudimentary, Watson's model is right in its core, an analogy can be made between the ribosome operating mode and a machine, that is a linotype or a factory. The ribosome, formed by the association of two sub-units, will ensure three functions in one cycle: the translation of a genetic information obtained from the nucleus through messenger RNA (mRNA), the collect of amino-acids from the cytoplasm through their transfer RNA (tRNA), and finally the linking between those residues to form a nascent polypeptide and export it to the cytoplasm. The ribosome can join up peptidic bonds at a rate of 10-20 peptide bonds/sec (Kjeldgaard – 1974) with a ~10-4 error rate (Loftfield-1972) but while the energy required for formation of the peptide bond comes from the ester bond of tRNA charged with its cognate amino acid, the ribosome also requires the GTPase activities of two G-protein elongation factors, EF-Tu and EF-G (Vanzi-2003). The three functions of the ribosome can be described as three stages and three operational sites involved in the protein production line (BSCB-2016): the "Initiation", "Elongation" and "Termination" using the sites "A", "P" and "E" in reading order of the mRNA strand. Table 1 resumes the positions and their roles. During the initiation stage, a small ribosomal sub-unit, typically alongside a formylmethionine-tRNA bound to an EF-Tu with GTP, will bind to the AUG codon of the 5' end of a mRNA strand. The ternary complex is completed by a large sub-unit. The start codon is recognized by the ribosome through specific sequences in the mRNA: the Shine-Dalgarno sequence in the prokaryotes and the Kozak box for the eukaryotes.

Binding site	mRNA strand entry site	Biological term	Main processes
A site	1st	**A**minoacyl-tRNA	Admission of codon of mRNA & 'charged' strand of tRNA. Checking and decoding and start of 'handing over' one amino acid molecule.
P site	2nd	**P**eptidyl-tRNA	Peptide synthesis, consolidation, elongation and transfer of peptide chain to site A
E site	3rd	**E**xit	Preparation of 'uncharged' tRNA for exit to the cytoplasm

Tab 1. Binding sites, positions and functions in a ribosome

During the peptidyl-transfer, the catalysis of the peptide bond will involve the C2 hydroxyl of RNA's P-site adenosine. The RNA catalytic activity earned ribosome the term "ribozyme" and in an early RNA world, RNA provided both genetic information and catalytic function, later embellished by additional proteins (Cech-2000). The ribosome brings about a 10^7-fold enhancement in the rate of the peptidyl-transfer reaction compared with the second-order reaction between model substrates in solution, by lowering the entropy of activation. By contrast, enzymes that employ general acid–base or covalent catalysis act by lowering the activation enthalpy of the catalyzed reaction (Rodnina-2007). Other steps in the protein synthesis are caused by changes in protein conformations, like the translocation, the shift of the mRNA/tRNA position by one codon, accompanied by a ratchet-like motion of the ribosomal subunits. The elongation factor EF-G in complex with GTP bind to the ribosome to facilitate the tRNAs translocation to the P and E sites. The release of the EF-G after GTP hydrolysis leaves the ribosome in the post-translocational state. It is now ready to accept a new aminoacyl-tRNA. In case of a match, EF-Tu with the hydrolyzed GTP and the E-site tRNA leave the ribosome. The ribosome can now enter the recycling phase, after which a 70S initiation complex is formed again. Figure 1 help illustrate the functional cycle of the ribosome (Yamamoto and al.-2013), though a supplementary elongation factor is presented, EF4-GTP, that can catalyze a reversal of the translocation, in order to mobilize stalled ribosomes.

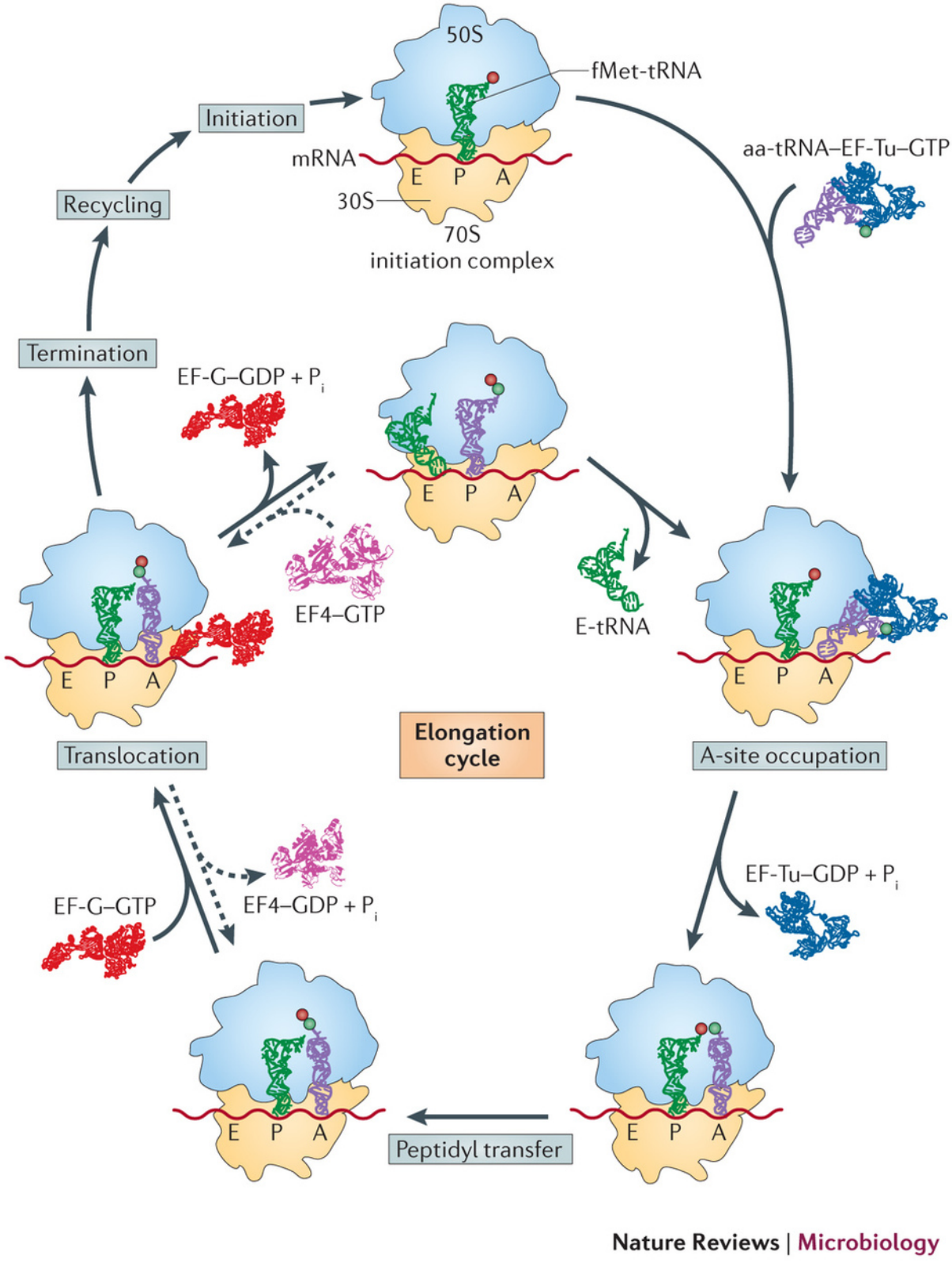


Fig 1. Functional phases of the prokaryotic ribosome during translation (Yamamoto and al.-2013).

To conclude this part as it was started and to make the transition towards the next part, Noll was present when Francis Crick said, during a sailing session, “If you don’t understand function, study structure” that got around as Crick’s second dogma (Noll-2008).

1.3 Structural model

The ribosomal system is one of the hottest focal areas in molecular cell biology, due to its functional role and progresses in the resolution of its structure, despite its large size. Both prokaryotic and eukaryotic ribosomes are made of two sub-units but they differ in their structures. The prokaryotic ribosome is smaller than its eukaryotic counterpart, at around 20 nm of diameter versus 25-30 nm. The size is correlated with the mass, with around 2 MDa in bacteria and 3.2 MDa in eukaryotic organisms. To designate the ribosome and its subunits, their sedimentation coefficient is usually preferred: 70S for the prokaryotic, divided in a 50S large subunit and a 30 small subunit; 80S for the eukaryotic, divided in 60S and 40S (Svedberg units, a measure of the rate of sedimentation in centrifugation, do not add up). Those subunits are composed of a balance of proteins and ribosomal RNAs (rRNA), in a ratio of 35%/65% in the prokaryotic world while the rRNA-to-protein ratio is close to 1 for the eukaryotic world. For the prokaryotes, their small subunit has a 16S RNA (consisting of 1540 nucleotides) bound to 21 proteins. The large subunit is composed of a 5S RNA subunit (120 nucleotides), a 23S RNA subunit (2900 nucleotides) and 31 proteins (Alberts-2002). For the eukaryotes, their 40S subunit has an 18S RNA (1900 nucleotides) and 33 proteins. The large subunit is composed of a 5S RNA (120 nucleotides), 28S RNA (4700 nucleotides), a 5.8S RNA (160 nucleotides) subunits and 46 proteins (Alberts-2002). Those structural properties are resumed in table 2.

		Prokaryotic	Eukaryotic
Ribosome	Diameter	~20 nm	~25-30 nm
	Molecular Mass	~2 MDa	~3.2 MDa
	Sedimentation	70S	80S
Large subunit	Sedimentation	50S	60S
	rRNAs	23S (2839 nucleotides) 5S (122 nucleotides)	28S (3354 nucleotides) 5S (154 nucleotides) 5.8S (120 nucleotides)
	Proteins	33	47
Small subunit	Sedimentation	30S	40S
	rRNAs	16S (1504 nucleotides)	18S (1753 nucleotides)
	Proteins	20	32

Tab 2. Main structural properties of the prokaryotic and eukaryotic ribosomes

The ribosome undergoes several conformational changes and binds to different factors that catalyze specific reactions. Atomic details are crucial to understanding its function. Although the general organization of the ribosome was known since the 1970s, 3D structures of it came thirty years later. The initial structures of ribosomes were obtained through electron microscopy, then cryo-electron microscopy, on eukaryotic ribosomes from the yeast (Verschoor-1998). The resolution on the first structures was still low, in the 30 Å range, but progress in crystallography quickly allowed to reach higher resolutions in the order of a few Å. There was a scientific race to publish the first ribosomal structure at atomic resolution. In the year 2000, the journals *Science*, *Cell* and *Nature* published the ribosomal structures at atomic resolution from respectively the 50S subunit of the archaeon *Haloarcula marismortui* (Ban-2000), the bacterium *Deinococcus radiodurans* (Schluenzen-2000), and the 30S subunit from *Thermus thermophilus* (Wimberly-2000). Nine years later, the Nobel Prize in Chemistry would reward Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath for their work in the structure and function studies of the ribosome. Shortly after those publications, the entire *Thermus thermophilus* 70S structure was published in *Science* by Marat Yusupov team (Yusupov-2001). A decade later, due to difficulties in obtaining crystals of sufficient quality, the team led by Yusupov finally published the first atomic structure of the whole 80S of a eukaryotic organism, *Saccharomyces cerevisiae* (Ben-Shem-2011). Their atomic model (PDB ids: 3U5B to 3U5I) revealed the architecture of eukaryote-specific elements and their interaction with the universally conserved core, and described eukaryote-specific bridges between the two ribosomal subunits. The precise structural data between eukaryotic and

prokaryotic ribosomes allow to compare them and pinpoint their differences, as illustrated in figure 2 (adapted from (Ben-Shem-2011)).

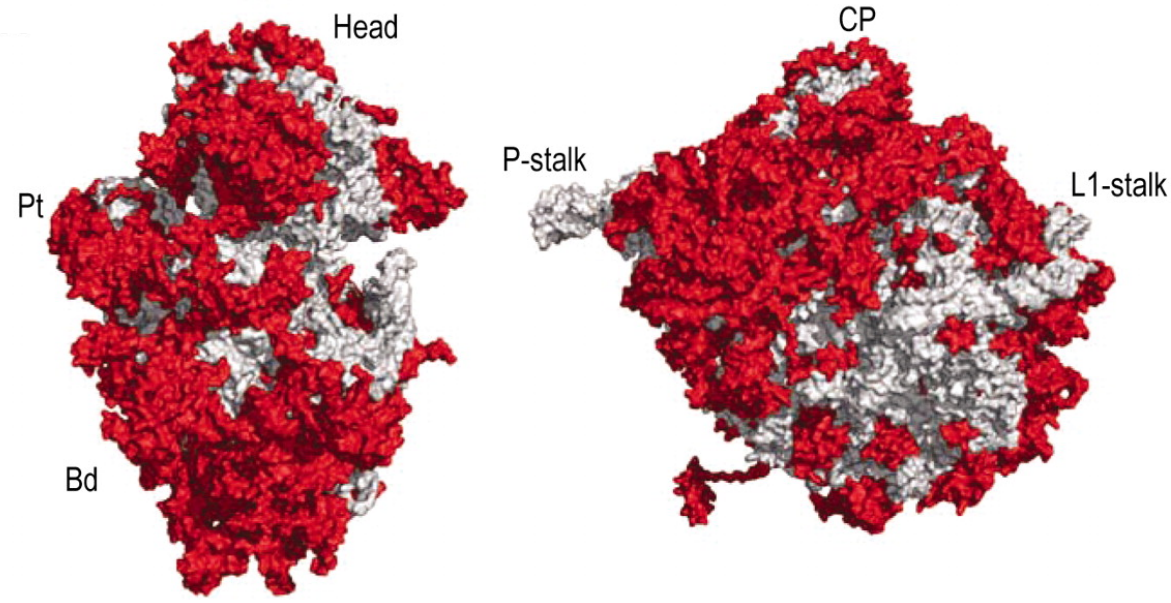


Fig 2. Characteristics of the 80S ribosome surface, the two subunits being viewed from their solvent-side. The eukaryote-specific elements are in red while conserved core elements are in gray. Landmarks include head, body (Bd), and platform (Pt) of 40S as well as central protuberance (CP), P-stalk and L1-stalk on the 60S.

The solvent-exposed sides of the 40S and 60S subunits contain the majority of the eukaryote-specific RNA and protein elements. The subunit interface, as well as important functional regions such as the peptidyl transferase center (in the large subunit) and the decoding site (in the small subunit) are mostly conserved. Eukaryote-specific protein segments engage in a multitude of protein-protein interactions. Long distance interactions are mediated by eukaryote-specific helical extensions of ribosomal proteins, and several eukaryotic ribosomal proteins can form inter-protein beta-sheets (Ben-Shem-2011). This enhanced interactivity of the eukaryotic ribosome can be linked to a more complex regulation of the biosynthesis happening in eukaryotic life forms (Topisirovic-2011). The targeting of newly-synthetized proteins and the translocation machinery are also much more complex in eukaryotes (Bohnsack-2010). Finally, the increased cancer susceptibility that result from mutations in ribosomal protein support such a regulatory role (Stumpf-2011).

2. A therapeutic crossroad

2.1 Ribosomal antibiotics

The differences between the bacterial and eukaryotic ribosomes are exploited by pharmaceutical chemists to create antibiotics that can destroy a bacterial infection without harming the cells of the infected person. Due to the differences in their structures, the bacterial 70S ribosomes are vulnerable to these antibiotics while the eukaryotic 80S ribosomes are not. Even though mitochondria possess ribosomes similar to the bacterial ones, mitochondria are not affected by these antibiotics because they are surrounded by a double membrane that does not easily admit these antibiotics into the organelle (O'Brien-1971). Most antibiotics are microbial secondary metabolites mainly produced by actinomycetes. Under unfavorable growth production, actinomycetes see their metabolic activity slowed down and can release antibiotics to protect themselves. Indeed, by blocking the growth of better adapted competitors, stressed actinomycetes can improve their survival chances (Cundliffe-2010). In term of semantics, "antibiotics" are only natural products, "chemotherapeutics" describes semi or full-synthetic molecules and "antimicrobial" is the term with the widest definition (Nikolay et al., 2016). There are many different antibiotics and also many different targets: a hundred of cellular goals were identified from high-throughput screening campaigns by GlaxoSmithKline (Payne et al., 2006). The target can either be the replication, the transcription, the translation or the cell wall synthesis (with for example, in respective order, nalidixic acid, rifamycin, chloramphenicol or penicillin). Every step of the ribosomal cycle can be targeted by an antibiotic but the functional complexity doesn't really translate in a wild variety of binding sites. The ribosomal sub-units have specific area where many inhibitors are binding, concentrated in functional key sites. Given the fundamental importance of the rRNA, it is not surprising that most inhibitors target the rRNA-rich surfaces on the 30S and 50S subunits (Poehlsgaard and Douthwaite, 2005). Figure 3 illustrates the possibilities of inhibition of the ribosomal cycle through different antibiotics while figure 4 locates those molecules binding sites on the ribosomal structure (both adapted from (Wilson-2013)).

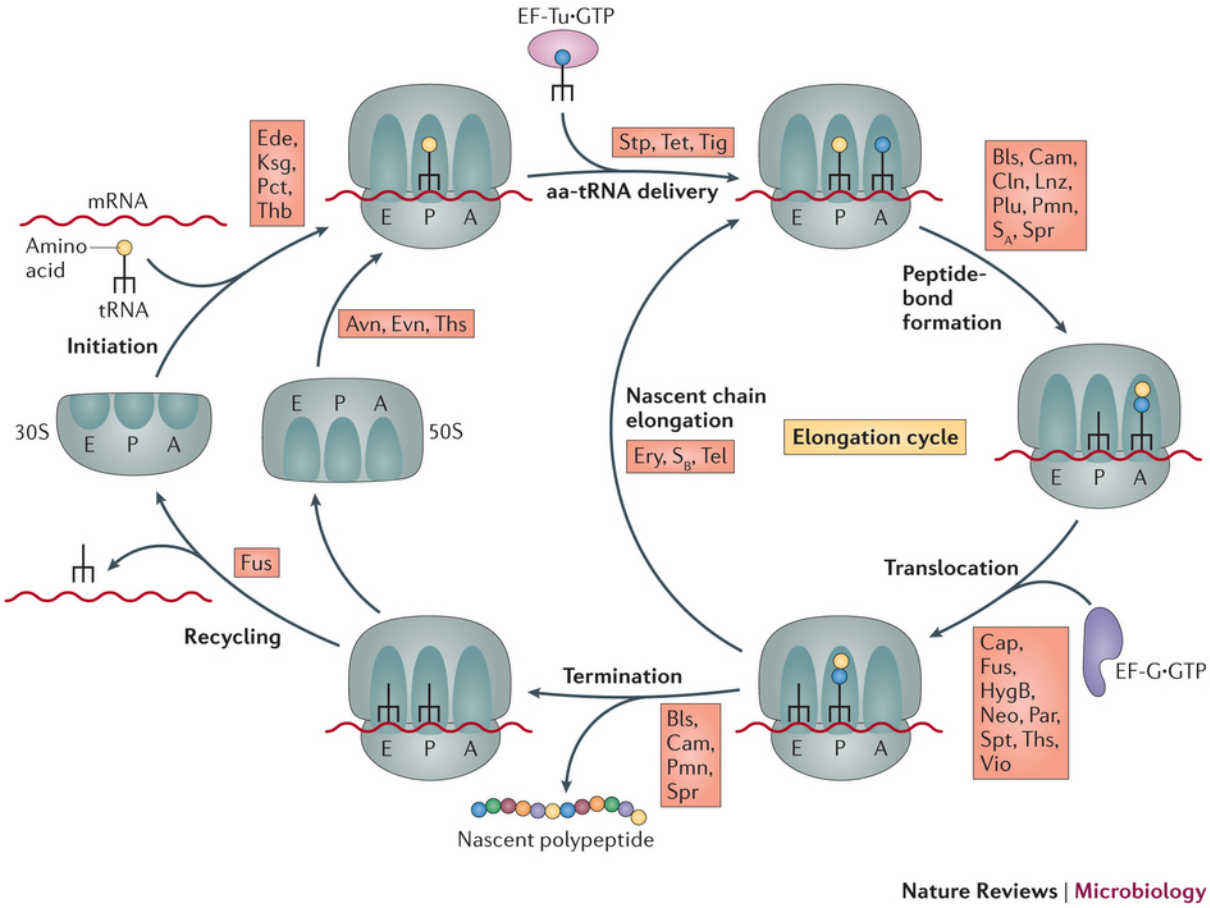


Fig 3. Antibiotic inhibition of the ribosomal cycle. The protein synthesis steps were described in the first part of this chapter. Each process can be blocked through one or more molecules. The antibiotics names are avilamycin (Avn), blasticidin S (Bls), capreomycin (Cap), chloramphenicol (Cam), clindamycin (Cln), edeine (Ede), erythromycin (Ery), evernimicin (Evn), fusidic acid (Fus), hygromycin B (HygB), linezolid (Lnz), kasugamycin (Ksg), neomycin (Neo), pactamycin (Pct), paromomycin (Par), pleuromutilins (Plu), puromycin (Pmn), sparsomycin (Spr), spectinomycin (Spt), streptogramin A (SA), streptogramin B (SB), streptomycin (Stp), telithromycin (Tel), tetracyclines (Tet), tigecycline (Tig), thermorubin (Thb), thiostrepton (Ths), and viomycin (Vio).

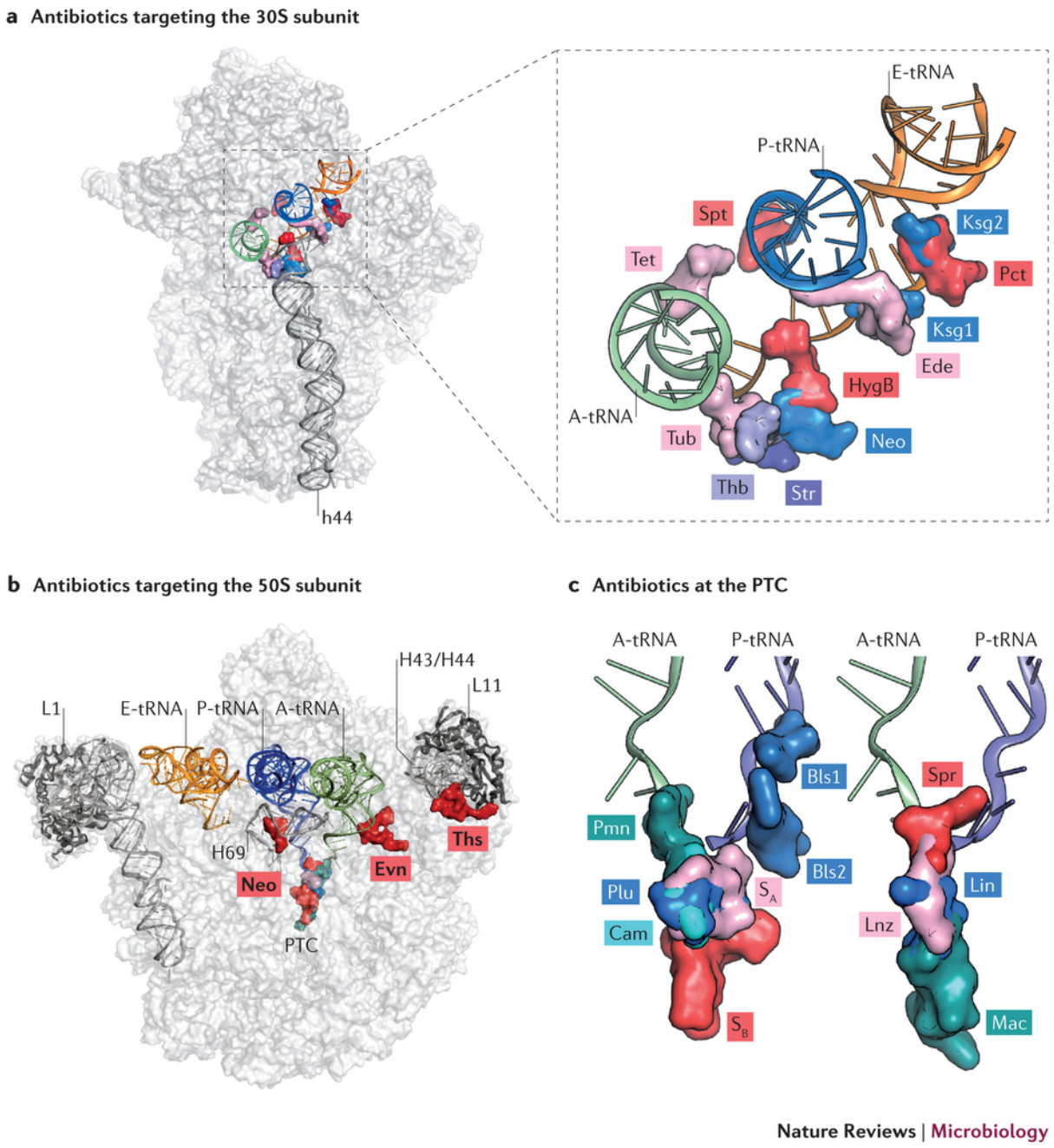


Fig 4. Overview and enlargement of antibiotic binding sites on the 70S ribosome. Both subunits are presented at their binding interface. The enlargements focus on the mRNA binding channel of the 30S subunit and the peptidyl-transferase center of the 50S. Supplementary antibiotic names in this figure concern the lincomycin (Lin), macrolides (Mac) and the tuberactinomycins (Tub).

Antibiotic binding of the small subunit may hinder its principle function of deciphering the genetic information encoded in the mRNA. In the A site area of the decoding center, the group of the aminoglycosides (kasugamycin (Ksg), neomycin (Neo), streptomycin (Stp)) meddles with the aa-tRNAs selection fidelity and before that, the assembly of the subunit (Mehta-2002). The tetracyclines group acts by disrupting aa-tRNAs binding stability but the tetracyclines own binding site is outside of the decoding center. Binding on the 50S subunit

can interfere with its main functions in controlling GTP hydrolysis, the formation of peptide bonds, and channeling the peptide through the subunit tunnel. Regarding the peptidyl-transferase center (PTC), diverse reasons can explain its prevalence as an antibiotic target. It is located at the bottom of a large structural cavity and many anfractuosities allow the binding of diverse small molecules. The PTC is the host of the two principal chemical reactions of protein synthesis: peptide bond formation and peptide release, needing an important structural adaptability. High fitness costs are correlated with the binding of small molecules in the peptidyl-transferase center (Polacek-2005). Some antibiotics, like the fusidic acid, shown in figure 3 but not figure 4, can also act not by binding directly to the ribosome but on associated proteins, like on the second elongation factor G (EF-G).

2.2 Bacterial ways of resistance

Bacteria originated over 3.8 billion years ago and based on the genetic divergence of antibiotic biosynthetic gene clusters, antibiotics are at least hundreds of millions of years old. The resistance mechanisms have been evolving for the same large amount of time and thus, predates the industrial use of antibiotics (Wright and Poinar-2012). The antibiotic properties of the *Penicillium* mold are used by mankind since antiquity but with Fleming 's demonstration and purification of penicillin in 1929, a period of drug innovation and wide implementation in human and animal health started. Antibiotics quickly became mankind's most effective therapeutic, but only three quarters of a century later, a looming antibiotic-resistance crisis might make diseases that were once easily treatable deadly again. The timeline in figure 5 shortly resumes the history of drug discovery, with the prevalent drug discovery models for each era (from (Brown and Wright - 2016)). The roots of the pandemic are multiple and complex. One could cite the misuse and overuse of antibiotics (especially in veterinary applications) that allowed resistant germs to spread through HGT (horizontal gene transfer) but mainly, it is due to a lack of basic knowledge on the biochemical and biophysical mechanisms of resistance (Davies and Davies - 2010). There are two main ways for bacteria to withstand the action of an antibiotic: stop the antibiotic from reaching its target at a high enough concentration or modify/bypass the target that the antibiotic inhibits. Over time, bacteria have evolved many different resistance strategies to accomplish this and some of those are presented in figure 6.

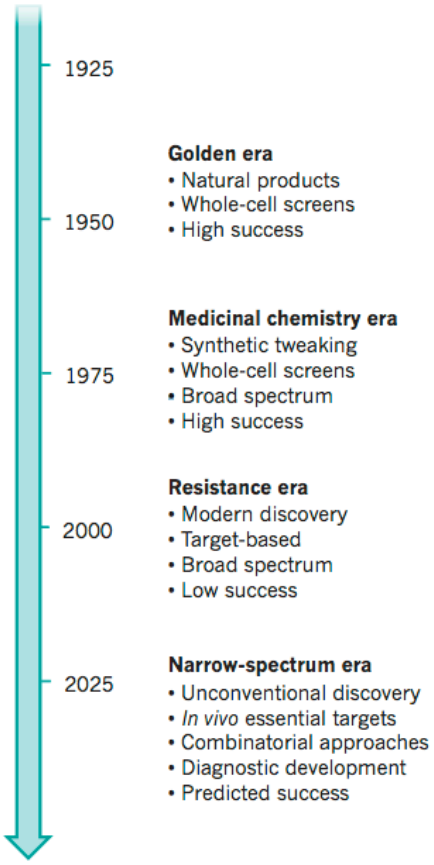


Fig 5. Development of the drug-discovery, divided by era models. The discovery of penicillin and later sulfonamides started the golden era. In subsequent years, termed the medicinal-chemistry era, these scaffolds were modified chemically, but without finding new classes of antibiotics. In the future, a focus on innovative methods must help to create narrow-spectrum agents.

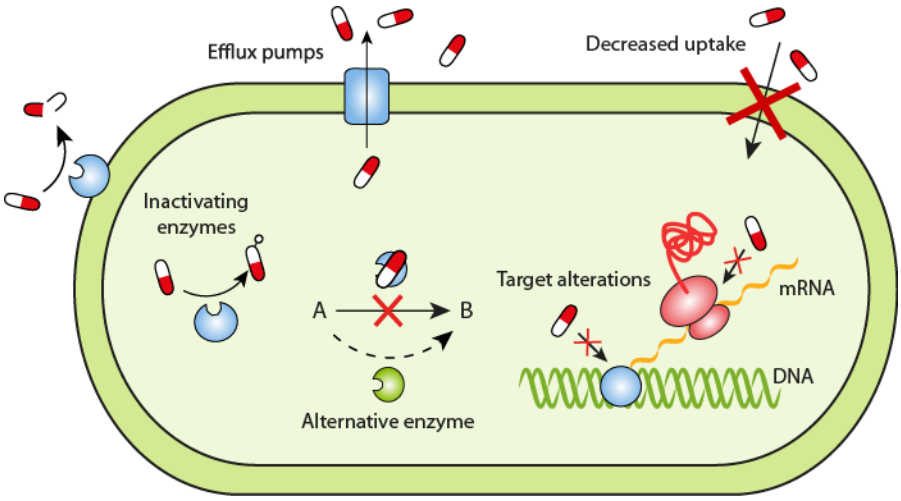


Fig 6. Resistance strategies in bacteria (Courtesy of Dr. E. Gullberg – ReAct group)

To stop the antibiotic from reaching its target, the bacteria may pump the antibiotic out or decrease the permeability of the membrane. Mutations in bacterial DNA can make the bacteria produces more efflux pumps. The antibiotic can be modified or simply degradated, as it this the case with the β-lactamase that destroys the active component (the β-lactam ring) of penicillins. The antiobiotic target can be modified in its structure or composition, or a variant can be produced as it is the case with Vancomycin-resistant bacteria that produce a different cell wall. In MRSA (methicillin-resistant *Staphylococcus aureus*), a new penicillin-binding protein is produced, with a low affinity to β-lactams. In order to pump drugs outside the bacteria, ABC transporters (for ATP-binding cassette) are often used. This is one of the most common bacterial resistance mechanisms. Resistance to many ribosomal drugs was found to be obtained through the ABC-F family, recent studies have shown that the ABC-F proteins can displace drugs from the binding site of the ribosome (Wilson-2016). Mutation rates for bacterial products are in the range of 10^{-6} to 10^{-8}, meaning that one bacterium out of one or a hundred million can be resistant to a specific drug and later transferring that resistance. With 10^{13} bacteria in the human body, in a rough 1:1 ratio with human cells (Sender, fuchs and Milo - 2016), it is therefore obvious that new antibacterial drugs are needed.

2.3 Eukaryotic organisms

The selectivity of antibiotics on prokaryotic inhibition is based on differences between the prokaryotic (i.e, bacterial) and eukaryotic (i.e., protozoan parasite, parasitical worms, animal, human…) ribosomes. In some cases, the eukaryotic and prokaryotic worlds overlap a little. For example, malaria and leishmaniasis are diseases caused by the parasites (*plasmodium* and *leishmania*) that possess a eukaryotic 80S ribosome but also an organelle, an apicoplast, that uses its own 70S-like ribosome. Apicoplasts are similar in some points with mitochondria and chloroplasts, found in plants. Doxycycline, an antibiotic first used in 1967, is also employed as a defensive measure against malaria. Crucial apicoplasts proteins will see their synthesis inhibited through the action of doxycycline on the 70S ribosome (Dahl-2008). The paromycin, known since the 1950s to bind the 30S subunit, is now thought to interact with both the 80S (in the cytosol) and the 70S ribosome in *leishmania* parasites, but with an enhanced affinity for the 70S version (Fernandez et al.-2011). Mitochondria, despite sharing ribosomal similarities with bacteria, are unaffected by antibiotics. Different

reasons can explain this: mitochondria possess a double membrane that protect from some antibiotics, mitochondrial rRNAs differ from bacterial ones, and finally, in the drug-discovery process, if antibiotics had been found to be harmful to human cells through mitochondria inhibition, they would not be used. However, they are reports of side-effects in widely used antibiotics (Kalghatgi et al., 2013). Inhibitors of eukaryotic translation may also hold therapeutic potential for application as antifungal (against Candida or Aspergillus) but also antitumor agents. Indeed, the deregulation of global translation has emerged as an important component of cancer chemo-responsiveness. Some antitumor agents can act as modulators of the eukaryotic ribosome, neglecting the effect of premature STOP codons found in genetic diseases, in order to restore the synthesis of the full and viable protein. This action aims at “tripping” the ribosome, rather than “stopping” it. Other inhibitors of eukaryotic protein synthesis have entered the cancer drug development pipeline since the antitumor activities of elongation inhibitors such as mycalamide B (MycB) or cycloheximide (CHX), that binds to the 60S E-site, have been found in 'in vivo' and 'in vitro' settings (Dang et al. – 2010). Some advanced to Phase I and Phase II clinical trials, but those are plagued by high failure rates due to toxicity problems (Pelletier-2007). The atomic resolution of the human ribosome (Ben-Shem - 2011) enables now the use of structure-based methods for the design of novel therapeutics, and allows the structural differences to the bacterial ribosome to be exploited, improving the selectivity of drugs and therefore reducing adverse effects. More than 16 high-resolution structures of 80S ribosomes from *Saccharomyces cerevisiae* in complexes with 12 eukaryote-specific and 4 broad-spectrum inhibitors were recently obtained by the Yusupov team (Garreau de Loubresse et al.-2014). Figures 7 and 8, in a symmetrical manner to figures 3 and 4, show the steps of protein synthesis in eukaryotes impaired by small molecule inhibitors as well as their binding sites.

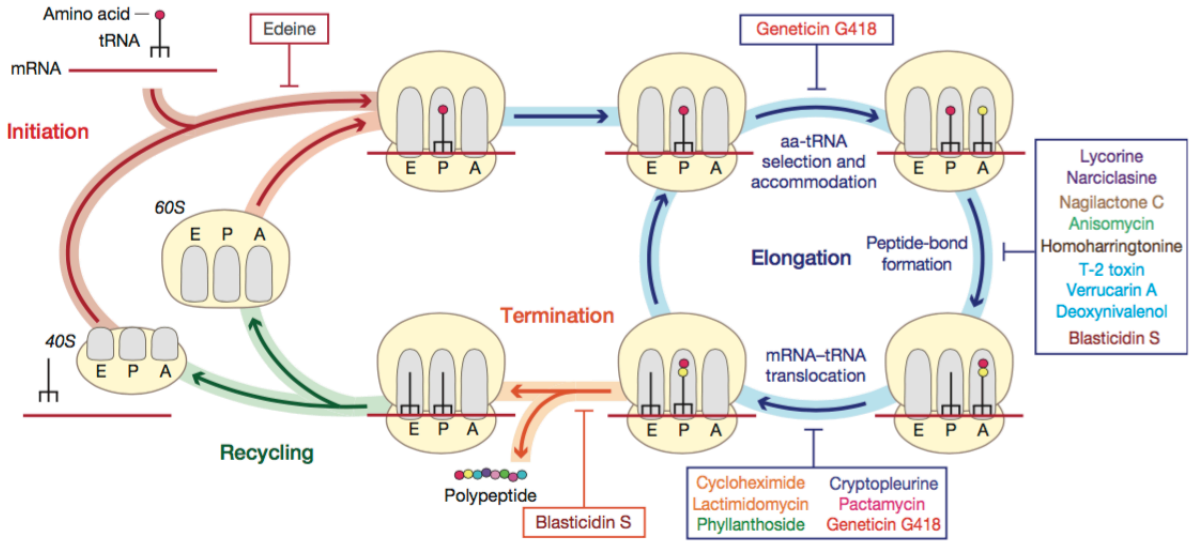


Fig 7. Inhibitors action during protein synthesis in eukaryotes. Schematic showing the steps of protein synthesis in eukaryotes impaired by 16 small-molecule inhibitors (Garreau de Loubresse et al.-2014).

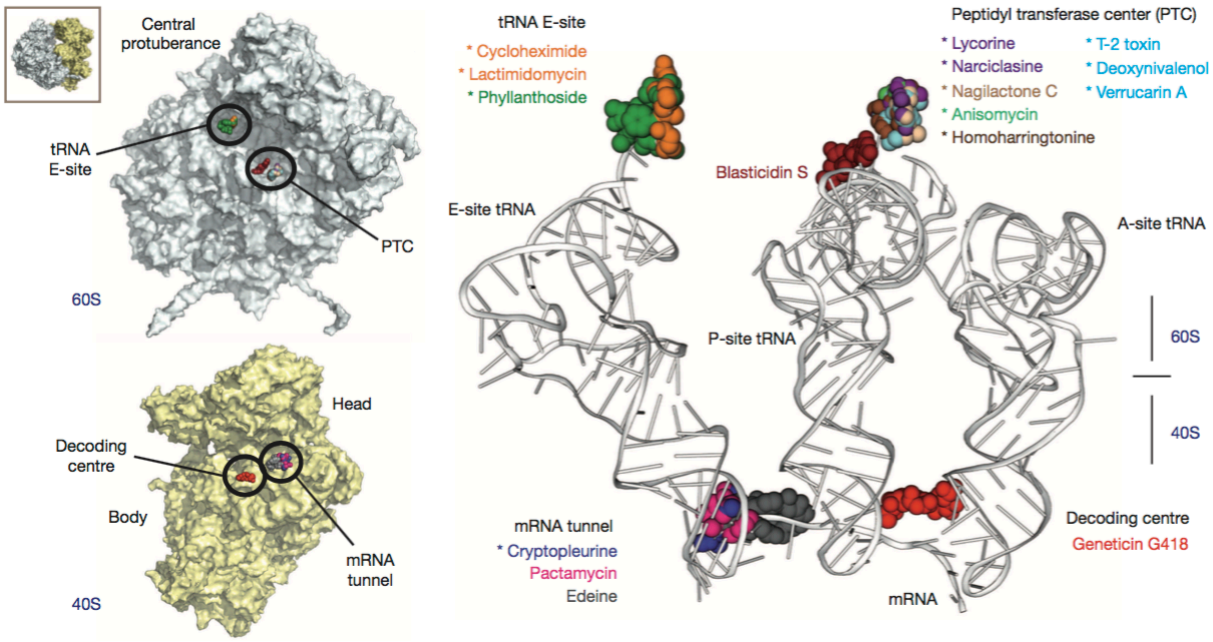


Fig 8. Binding sites of inhibitors on the yeast ribosome. The 80S ribosome is split to show the 60S and 40S subunit interface. Eukaryotic specific inhibitors are marked with an asterisk. (From (Garreau de Loubresse et al.-2014)).

3. Screening on the ribosome

3.1 NMR in Drug discovery

The technological and scientific advancements seen in the late twentieth century (high throughput screening, genomic sequencing, combinatorial chemistry on compound libraries) created new approaches for drug design and discovery to find new drug leads (Hajduk-2007). Nuclear magnetic resonance (NMR) offers several screening possibilities and can be combined with X-ray crystallography and computational modelling (Pellecchia-2008). NMR took then new roles in the drug discovery process, that can be defined as the identification and optimization of drug candidates. Drug discovery is a long (around a decade) and costly (estimated in hundreds of millions of US dollars) process that can be segmented in different stages: the program selection (the disease to remedy), the target identification, the assay development, the lead generation, and the clinical trials. The lead generation step identifies "hits" with binding activity against the target. The hit compounds are used as a starting point for optimization in order to improve potency, selectivity, and pharmacokinetic parameters. NMR is mostly used, through different strategies, in the early stages, from the hit discovery to the lead optimization, as illustrated in figure 9 (from (Pellecchia-2008)).

The hits identification is usually done by high throughput screening (HTS) or fragment based drug design (FBDD), a new approach that emerged as an alternative (Jhoti-2013). HTS involves screening a large chemical library for compounds that bind to the target with a high affinity. Although a large number of compounds can be screened, it is limited by the size and characteristics of the chemical libraries used. FBDD is based on a screening of a smaller number of compounds but trying to find binders that can be expanded to generate a high affinity lead. NMR is well adapted to the screening of compounds with weak to strong binding affinities (with K_d ranging respectively from the mM to the µM) and it can be done by two approaches: receptor-based screening or ligand-based screening.

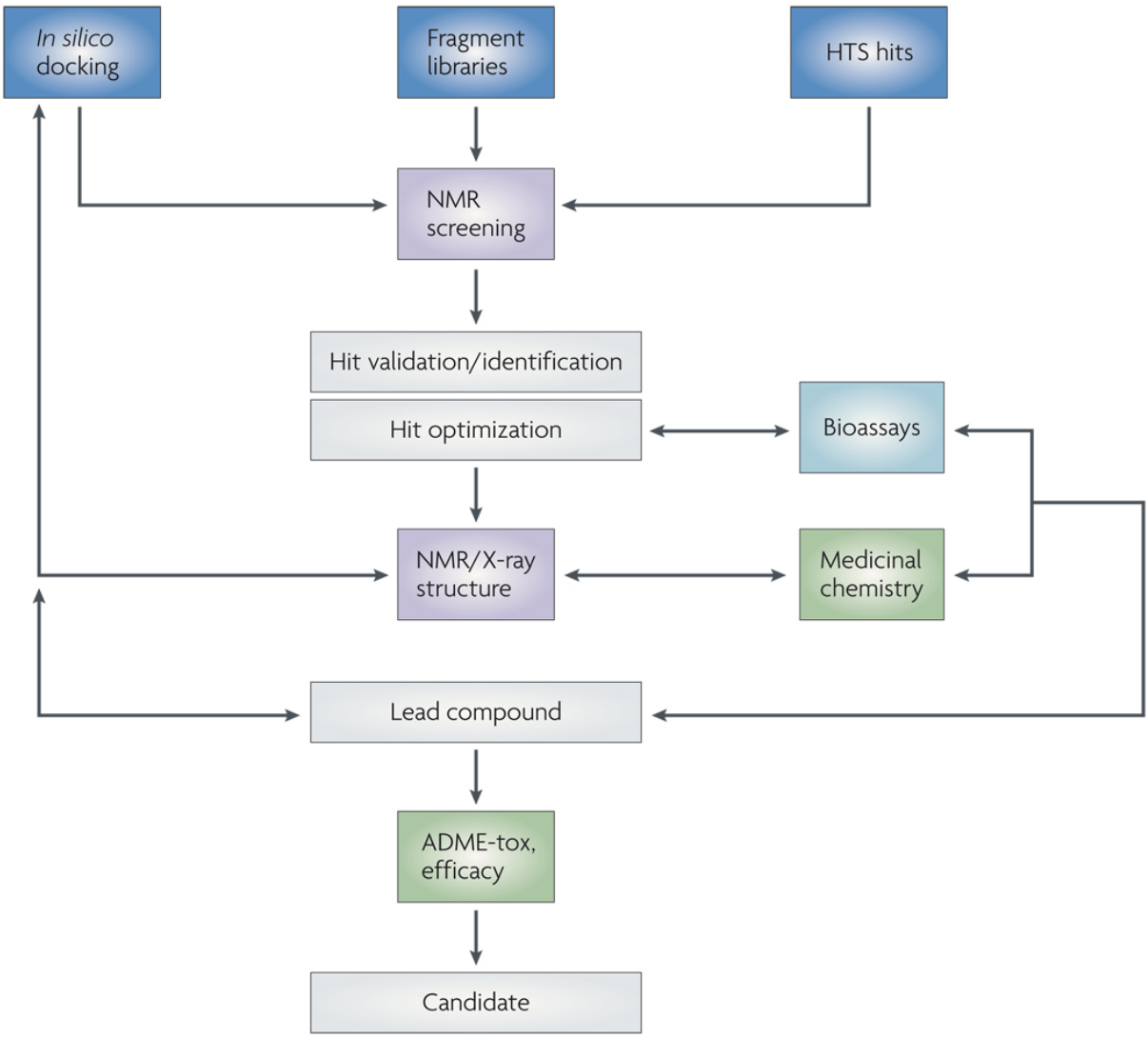


Fig 9. NMR roles in the drug discovery process. NMR can be employed in primary screening, in the validation of the binding site or later in the process, in structural analysis of the complex structure. (ADME-tox stands for Absorption, Distribution, Metabolism, Excretion and TOXicity)(from (Pellecchia-2008)).

3.2 Screening through NMR

NMR is a popular tool in fragment-based techniques. The methods employed can be divided into two groups whether it is the resonances of the target or the ligand that are monitored to detect the binding event. In target or receptor-based screening, one of the oldest and most widely used method to detect binding is the use of chemical shifts perturbations (CSPs). The target chemical shifts are monitored upon the interaction with the ligand and the changes observed give information on the binding parameters. The structure-activity relationship (SAR) by NMR uses the chemical shift mapping for the identification of a high affinity binding ligand on the basis of two ligand fragments that bind to two neighboring binding sites (Shuker-1996). SAR by NMR is carried out in five main steps, first, it starts with

the identification of ligands with a sufficient binding affinity (K_d under the mM range) from a library of compounds by using 2D ^{1}H-^{15}N HSQC, then the optimization of ligands by chemical modification. The operation is repeated for a neighboring site and finished by the linking of the two ligands. Receptor-based techniques usually require relatively high concentration of protein, that may need isotopic labelling. This can lead to technical difficulties and/or high associated costs. The target size is also an important factor since the spectral overlap increases with the molecular weight and the resonances get broader due to reduced tumbling. There have been NMR developments to tackle those issues, namely the increased magnetic fields, cryoprobes, optimized acquisition protocols (Renaud-2009). Thus, NMR experiments on the ribosome have been realized, but due to the molecular weight of the ribosome, sharp resonances are confined to specific regions of the macromolecule (Waudby – 2013). Techniques based on the ligand observation offer complementary advantages.

In ligand-based screenings, the interaction is monitored through the modifications of the ligand signals upon addition of the target, not only its chemical shift difference. Cocktails of ligands can be screened at the same time, if the resonances can be properly attributed. Different techniques have been developed: the saturation-transfer difference (STD), methods based either on the relaxation or the NOE, diffusion-edited screenings… The STD method is often found in NMR screenings. It consists in the selective saturation of a spectral area that belongs only to the receptor (Mayer and Meyer – 1999). If a ligand binds, the saturation will propagate from the receptor protons to other protons of receptor (via spin diffusion) and then the saturation is transferred to binding compounds (by a cross relaxation mechanism) at the ligand-receptor interface. A control experiment must be done by saturating a region of the spectrum that does not contain any signal. The resulting difference spectrum yields only the resonances that have experienced saturation, namely those of the receptor and those of the compound that binds to the receptor. But the receptor is typically present at very small concentrations so its resonances will not be visible, leaving thus only the bound ligand signal. The WaterLOGSY experiment is a variant of the STD where saturation transfer involves bound water molecules instead of the protein so it is the water resonance that will be saturated (Dalvit-2000). Two-dimensional screening experiment exist also among ligand-observed techniques, such as the trNOE (Chen and Shapiro – 1998). Transferred NOE (Nuclear Overhauser Effect) can be used to screen for small molecules that bind to a receptor under conditions of fast exchange. No isotopic labeling of the ligand or

receptor is necessary. Even under conditions where there is only a small percentage of bound ligand in solution, the "memory" of NOEs present within the bound state conformation of the ligand are carried over to the ligand free in solution under conditions of fast exchange and nOes will appear if binding occurs. From the ligand viewpoint, ^{1}H-NMR T_2 (CPMG) measurements have been realized on MLSB antibiotics (macrolides, lincosamides and streptogramin B) but the binding-induced line broadening circumscribed the ligand-target molar ratio to 600 at minima (Verdier – 2000).

Finally, NMR screenings experiments can be conducted by exploiting the properties of the fluorine isotope ^{19}F, which are reviewed in the first chapter of this thesis. ^{19}F-based screening methods are relatively recent (Dalvit-2003) and offer interesting advantages compared to other screening techniques, as listed in table 3 (adapted from (Dalvit-2009)).

Properties and requirements	NMR techniques				Fluorescence Anisotropy	SPR
	Ligand detection			Protein detection		
	STD and WaterLOGSY	$^{H}R_2$	$^{F}R_2$	SAR by NMR		
Response dependence	[EL]	[EL]/[LT]	[EL]/[LT]	[EL]/[ET]	[EL]/[ET]	[EL]/[ET]
Intrinsic sensitivity	Low	Low	Low	Low	High	High
Dynamic range	Large	Large	Very large	Medium-large	Small	Small
Protein size sensitivity	increase with size	increase with size	increase with size	decrease with size	increase with size	unpredictable
Fragment concentration	µM to mM	Low µM	Low µM	µM to mM	µM to mM	µM to mM
Protein concentration	Low µM	Low µM	nM to low µM	µM	nM	nM
Volumes	30-500 µL				5-50 µl	50-100 µl

Tab 3. Biophysical techniques frequently applied to fragment screening (adapted from (Dalvit-2009)). The dynamic range describes the possibility of detecting an interaction between different concentration scales of target and ligand. If the dynamic range of a technique is small, the interaction can be detected only around ligand concentration value similar to the K_d.

3.3 Our research goal

The goal targeted through this thesis project is to develop a method capable of screening ligands on the ribosome using ^{19}F NMR. The screening of chemical libraries is performed on the target (in this case, the whole ribosome) in the presence of a weak- to medium-affinity ligand referred in this role as the "spy" molecule. It is also necessary to experiment in the presence of a control molecule that does not interact with the target. Screening is carried out simply by monitoring the relative intensities of the spy and control in the presence of the target and the various candidate molecules successively tested. If the spy is bound to the target (it is supposed to be at the beginning of the experiment) its NMR signal disappears, only to reappear if the spy is specifically displaced of its binding site by another ligand. This method, originally performed with the proton, can also be used while detecting the resonance of the fluorine atom. Indeed, the use of the signals of only our two fluorinated molecules allows simple and rapid detection. In addition, the properties of fluorine in NMR guarantee a very high contrast between the free and bound states. This very sensible approach is called FAXS (Fluorine chemical shift Anisotropy and eXchange for Screening) and was introduced by the works of Claudio Dalvit (Dalvit - 2003). This method has since then been successfully used by Marc-André Delsuc, now in Bruno Kieffer lab, on a project on the Glucokinase (Assemat et al., 2015).

The development of this method of detection by NMR screening of new ligands on the ribosome will first start on the bacterial ribosome. Indeed, model organisms such as *Thermus thermophilus* exists in the bacterial world in extreme environments, from which a very stable ribosome can be extracted, ideal for testing different experimental conditions. Once experimental parameters will be well-defined (temperature, concentration, signal to noise ratio, etc.) the FAXS experiments will be able to begin. It will be relevant to test different fluorinated antibiotics as "spies" and various antibiotics competitors for each site. It will also be necessary to define a control molecule. Purification of ribosomes is a long and difficult process, but it is much more complicated for eukaryotic than prokaryotic organisms, even with a eukaryotic organism as simple as the yeast Saccharomyces cerevisiae. Thus, the validation of the experimental model with bacterial ribosomes is required before testing on

eukaryotic ribosomes. It is the same for the passage of the yeast ribosome to the human ribosome.

This thesis manuscript will then start by reviewing the properties of fluorine and present how they can be exploited in NMR. The characterization through NMR of the ligands used in our screening campaigns are detailed in the second chapter. The third chapter will present the screening results obtained on the ribosome, that will soon be submitted as a scientific publication. In the fourth chapter is related a collaboration done with a biotech company, Nosopharm, as our screening method was put to the test.

During my thesis, I had the opportunity to work on other projects and two of them are presented in the fifth and sixth chapter of this thesis. One is the observation of a binding event between peptides containing fluorinated polyprolines and a SH3 domain while another project, that I jumped on during my Master thesis, revolves around the dynamic properties of a protein domain and led to a publication.

References

Assemat, O., Antoine, M., Fourquez, J., Wierzbicki, M., Charton, Y., Hennig, P., Perron-Sierra, F., Ferry, G., Boutin, J. and Delsuc, M. (2015). 19F nuclear magnetic resonance screening of glucokinase activators. Analytical Biochemistry, 477, pp.62-68.

Bruce Alberts, et al. (2002). The Molecular Biology of the Cell, fourth edition. Garland Science pg. 342

Bąkowska-Żywicka K, Tyczewska A. (2009). The structure of the ribosome – short history. Biotechnologia; 1:14 – 23

Ban N, Nissen P, Hansen J, Moore P, Steitz T. (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution". Science. 289 (5481): 905–20.

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011). "The structure of the eukaryotic ribosome at 3.0 Å resolution". Science. 334 (6062): 1524–1529.

Bohnsack, Markus T.; Schleiff, Enrico. "The evolution of protein targeting and translocation systems". Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 1803 (10): 1115–1130.

Brachet, J. (1933). Recherches sur la synthase de l'acide thymonucleique pendant Ie developpement de l'oeuf d'oursin. Arch. Biol. 44:519.

“Ribosome.” British Society of Cell Biology. British Society for Cell Biology, n.d. Web. http://bscb.org/learning-resources/softcell-e-learning/ribosome/

Caspersson, T. and J. Schultz. (1938). Nucleic acid metabolism of the chromosomes in relation to gene reproduction. Nature 142:294.

Cech T. (2000). "Structural biology. The ribosome is a ribozyme". Science. 289 (5481): 878–9.

Claude, A. (1941). Particulate components of cytoplasm. Cold Spring Harbor Symp. Quant. Biol. 9:263.

Chao FC S. H. (1956). Arch. Biochem. Biophys., 61, 220-230.

Chen A. and M. J. Shapiro. (1998). NOE pumping: a novel NMR technique for identification of compounds with binding affinity to macromolecules. Journal of the American Chemical Society, 120(39):10258–10259.

Cundliffe, E.; Demain, A.L. (2010). Avoidance of suicide in antibiotic-producing microbes. J. Ind. Microbiol. Biotechnol. 37, 643–672.

Dahl EL and Rosenthal PJ. (2008). “Apicomplast translation, transcription and genome replication: targets for antimalarial antibiotics.” TRENDS in Parasitology 24: 279-284.

Dalvit, C. (2009). NMR methods in fragment screening: theory and a comparison with other biophysical techniques. Drug Discovery Today, 14(21-22), pp.1051-1057.

Dalvit C., P. E. Fagerness, D. T. Hadden, R. W. Sarver, and B. J. Stockman. (2003). Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. Journal of the American Chemical Society, 125(25):7696–7703,.

Dalvit, C., P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti and M. Sundstrom. (2000). "Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water." Journal of Biomolecular Nmr 18(1): 65-68.

Dang, Y., Schneider-Poetsch, T., Eyler, D., Jewett, J., Bhat, S., Rawal, V., Green, R. and Liu, J. (2011). Inhibition of eukaryotic translation elongation by the antitumor natural product Mycalamide B. RNA, 17(8), pp.1578-1588.

Davies, J. and Davies, D. (2010). Origins and Evolution of Antibiotic Resistance. Microbiology and Molecular Biology Reviews, 74(3), pp.417-433.

Hajduk, P. J. and Greer J. (2007). A decade of fragment-based drug design: strategic advances and lessons learned. Nature reviews Drug discovery, 6(3):211–219.

Jhoti H., G. Williams, D. C. Rees, and C. W. Murray. (2013) The ‘rule of three’ for fragment-based drug discovery: where are we now? Nature Reviews Drug Discovery, 12(8):644–644.

Fernandez MM et al. (2011) “Differential effects of paromomycin on ribosomes of Leishmania mexicana and mammalian cells.” Antimicrobial Agents and Chemotherapies 55: 86-92.

Garreau de Loubresse, N., Prokhorova, I., Holtkamp, W., Rodnina, M., Yusupova, G. and Yusupov, M. (2014). Structural basis for the inhibition of the eukaryotic ribosome. Nature, 513(7519), pp.517-522.

Kalghatgi, S., Spina, C., Costello, J., Liesa, M., Morones-Ramirez, J., Slomovic, S., Molina, A., Shirihai, O. and Collins, J. (2013). Bactericidal Antibiotics Induce Mitochondrial Dysfunction and Oxidative Damage in Mammalian Cells. Science Translational Medicine, 5(192), pp.192ra85-192ra85.

Kjeldgaard, N.O. and Gausing, K. (1974). Regulation of biosynthesis of ribosomes. In Ribosomes (eds. M. Nomura et al.), pp. 369–392. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Nikolay, R., Schmidt, S., Schlömer, R., Deuerling, E. and Nierhaus, K. (2016). Ribosome Assembly as Antimicrobial Target. Antibiotics, 5(2), p.18.

Noll, H. (2008). The discovery of polyribosomes. Bioessays, 30(11-12), pp.1220-1234.

Loftfield RB, Vanderjagt D. (1972). The frequency of errors in protein biosynthesis. Biochem J.; 128(5):1353-6.

Mayer, M. and B. Meyer (1999). "Characterization of ligand binding by saturation transfer difference NMR spectroscopy." Angewandte Chemie-International Edition 38(12): 1784-1788.

Mehta, R.; Champney, W.S. (2002). 30S ribosomal subunit assembly is a target for inhibition by aminoglycosides in Escherichia coli. Antimicrob. Agents Chemother. 46, 1546–1549.

O'Brien, T.W. (1971). "The General Occurrence of 55S Ribosomes in Mammalian Liver Mitochondria". J. Biol. Chem. 245: 3409.

Palade, G. E. (1955). A small particulate component of the cytoplasm. J. Biophys. Biochem. Cytol. 1:59.

Pelletier, J.; Peltz, S.W. (2007). "Therapeutic Opportunities in Translation". Cold Spring Harbor Monograph Archive. 48: 855–895.

Petermann, M. L. and M. G. Hamilton. (1952). An ultracentrifugal analysis of the macromolecular particles of normal and leukemic mouse spleen. Cancer Res. 12:373

Poehlsgaard, J. and Douthwaite, S. (2005). The bacterial ribosome as a target for antibiotics. Nature Reviews Microbiology, 3(11), pp.870-881.

Polacek, N.; Mankin, A.S. (2005). The ribosomal peptidyl transferase center: Structure, function, evolution, inhibition. Crit. Rev. Biochem. Mol. Biol., 40, 285–311.

J.-P. Renaud and M.-A. Delsuc (2009). Biophysical techniques for ligand screening and drug design. Current opinion in pharmacology, 9(5):622–628.

Rodnina MV, Beringer M, Wintermeyer W (2007). "How ribosomes make peptide bonds". Trends Biochem. Sci. 32 (1): 20–6.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A (2000). "Structure of functionally activated small ribosomal subunit at 3.3 Åresolution". Cell. 102 (5): 615–23.

Sender, R., Fuchs, S. and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. bioRxiv, [online] p.036103.

Shuker, S. B., P. J. Hajduk, R. P. Meadows and S. W. Fesik (1996). "Discovering high-affinity ligands for proteins: SAR by NMR." Science 274(5292): 1531-4.

Stumpf, CR., Ruggero, D. (2011). "The cancerous translation apparatus". Curr Opin Genet Dev. 21 (4): 474–83. doi:10.1016/j.gde.2011.03.007.

Tissieres A., Watson J. D., (1958), Nature, 182, 778-780.

Tissières, A. (1974). Ribosome Research: Historical Background. Cold Spring Harbor Monograph Archive, 04(0).

Topisirovic, I. Sonenberg, N. (Apr 2011). "Translational control by the eukaryotic ribosome". Cell. 145 (3): 333–4.

Vanzi, F. (2003). Protein synthesis by single ribosomes. RNA, 9(10), pp.1174-1179.

Verdier, L., Gharbi-Benarous, J., Bertho, G., Evrard-Todeschi, N., Mauvais, P., & Girault, J.-P. (2000). Dissociation–equilibrium constant and bound conformation for weak antibiotic binding interaction with different bacterial ribosomes. Journal of the Chemical Society, Perkin Transactions 2, (12), 2363–2371.

Verschoor, A; Warner, JR; Srivastava, S; Grassucci, RA; Frank, J. (1998). "Three-dimensional structure of the yeast ribosome". Nucleic Acids Res. 26 (2): 655–61

Waudby CA, Launay H, Cabrita LD, Christodoulou J. (2013). Protein folding on the ribosome studied using NMR spectroscopy. Prog Nucl Magn Reson Spectrosc.;74:57-75.

Wilson, D. (2013). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. Nature Reviews Microbiology, 12(1), pp.35-48.

Wilson, D. (2016). The ABC of Ribosome-Related Antibiotic Resistance. mBio, 7(3), pp.e00598-16.

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V . (2000). "Structure of the 30S ribosomal subunit". Nature. 407 (6802): 327–39.

Wright, G. and Poinar, H. (2012). Antibiotic resistance is ancient: implications for drug discovery. Trends in Microbiology, 20(4), pp.157-159.

Yamamoto, H., Qin, Y., Achenbach, J., Li, C., Kijek, J., Spahn, C. and Nierhaus, K. (2013). EF-G and EF4: translocation and back-translocation on the bacterial ribosome. Nature Reviews Microbiology, 12(2), pp.89-100.

Yusupov MM, Yusupova GZ, Baucom A, et al. (May 2001). "Crystal structure of the ribosome at 5.5 A resolution". Science. 292 (5518): 883–96.

CHAPTER 1:
Fluorine in NMR

Introduction

Fluorine (symbol F and atomic number 9) is a very peculiar chemical element. Most of the time, fluorine is only known for its dental care applications. But its specific chemical properties lead to its widespread introduction in various molecules used in everyday life. Ranging from refrigerants (Freon) to cookware (Teflon), not mentioning toothpaste, fluorochemical sales are worth 15 billion $ per year (Freedonia Industry study – 2012). The pharmaceutical industry has currently a fifth of its drugs that contain at least one fluorine (Arnston – 2015). The natural ^{19}F isotope is well suited for NMR spectroscopy, although the corresponding number of scientific publications did not raise over recent years at the same pace as fluorine applications (figure 1.1).

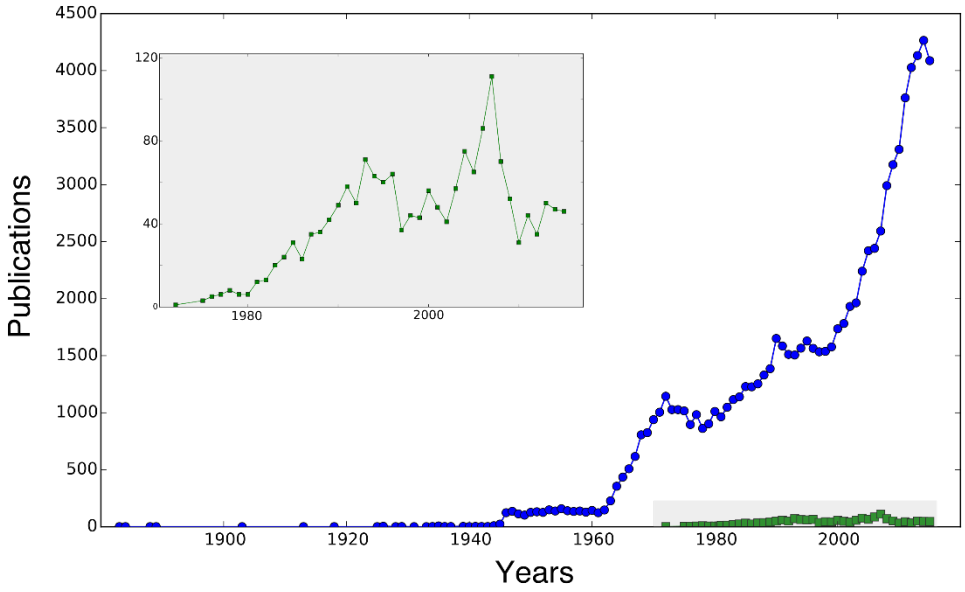


Fig 1.1 – Number of scientific papers published in the PubMed database per years about fluorine applications (blue curve, key word "Fluorine") and ^{19}F NMR (in green, key words "19F NMR"). The insert magnifies the publication number related to NMR.

Figure 1.1 shows the modern history of fluorine, a little bit before and after the twentieth century and highlights the major steps of fluorine applications. World War 2 and the Manhattan Project triggered the development of industrial production of fluorine. Uranium separation was done through the use of gaseous uranium hexafluoride. This gas was so

corrosive and reactive that all valves and seals from the separation plants were coated with Teflon, a polymer made of tetrafluoroethylene. (Takashi – 2009). Public academic work on fluorine raised interest when safe fluorinating agents were available for ordinary laboratories (Müller – 2007). The sixties were favorable to the development of many applications of fluorine chemistry. Teflon was used in cooking pans, but also in fabric, with the Gore-Tex technology, in bulletproof vest and in spaceships shields. Fluorine played also a role in cooling, with the chlorofluorocarbons initially, later replaced by hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (HFCs) for ecological reasons. Nowadays, many commercial applications rely on fluorine compounds. Half of the annual mined fluorite (predicted 2016–18 production figures of 3.5 to 5.9 million tons) goes toward the steelmaking industry. The other half is transformed in organic fluorides, as previously mentioned, or into artificial cryolite (used for aluminium refining). All mentioned applications use the fluorine-19 isotope, but the fluorine-18 isotope has also applications. It is used in the radiopharmaceutical industry, and is primarily synthesized into fluoro-2-deoxy-D-glucose (FDG) to be used in positron emission tomography (PET scans). Its interest is due to both its short half-life and the emission of positrons (Jacobson – 2015). ^{19}F NMR represents only a very small portion of studies reported on fluorine. Initially restricted to the chemical analysis of organofluorine compounds, ^{19}F NMR is currently expanding in biology. ^{19}F NMR finds applications in metabolic studies, in protein-ligand binding assays (supported by the works of Dalvit around 2000) and even in structural analysis of protein and macromolecules.

This chapter aims at presenting how the combination of fluorine and NMR emerges as an efficient technique in biology. We will review the fluorine atom properties and how those lead to its wide use in pharmaceutical compounds. A second part will present how NMR can exploit the peculiar properties of fluorine, for ligand screening studies. Some practical aspects of ^{19}F NMR will be finally presented in the third part of the chapter.

1.1 The fluorine atom

1.1.1 Physicochemical properties of fluorine

Fluorine is the most electronegative element, a halogen sitting in the top right of the periodic table just before the noble gases column. It is extremely reactive and can combine with nearly all elements, even with some from the noble gases family. Fluorine is the 13th element in terrestrial abundance and is primary found as fluorite, a mineral form of calcium fluoride (CaF_2). The first traces of fluorite, the main mineral source for fluorine, already appear in early antiquity. Fluorite was described in 1529 and got its name from the Latin "*fluere"* which means "to flow". Indeed, fluorite was used as an additive to metal ores to lower the melting point. Addition of an acid to fluorite releases toxic fumes. Those fumes, fluoric acid, were used by sixteenth century glassworkers to etch glass. Chemists attempted to identify the material that was capable of etching glass leading to the production of fluorine through an electrolytic process by George Gore in 1869. Unknown to Gore, fluorine combines with hydrogen and the transformed hydrofluoric acid causes deep burns in contact with the skin. Due to this toxicity, it is only in 1886 that a French chemist, Henri Moissan, successfully isolated fluorine. For his work on an element that killed many 19th-century experimenters, Moissan got the Nobel Prize in 1906, paving the way for the current widespread applications of fluorine (Wang 2014). He died two months later, poisoned several times by hydrogen fluoride gas. Fluorine can be compared to hydrogen and to chlorine, a well-known halogen. The fluorine atom can establish hydrogen bonds or interact with alkali metals (Na, K…) through its three non-bonded electron pairs. It can form hydrogen fluoride, a colorless gas or liquid that form hydrofluoric acid in aqueous solution. The comparison with hydrogen chloride (HCl), hydrochloric acid in water, is interesting because HCl is a strong acid (pKa of -8.0) whereas HF is a weak acid (3.15 for HF), despite its stronger electronegativity. This is due to the smaller ion size of fluorine, that leads to a larger concentration of negative charges on the ion and subsequently a destabilization of the F^- negative ion. Cl^- is more stable as the charge is distributed over a larger surface so HCl will ionize more easily in water to give H^+ and Cl^- ions. However, while HF is a weaker acid than HCl, it is actually much more dangerous for the human body (CDC database – 2015). The fluorine physicochemical properties are peculiar and fluorine reactivity cannot be explained

by analyzing single properties, but rather an ensemble. The table 1.1 reports several fluorine physicochemical parameters and compares them to hydrogen and chlorine. One can have a look at the difference in ionization energy and electron affinity between fluorine (-328 kJ/mol) and chlorine (-348 kJ/mol). The ionization energy quantifies the energy required for the formation of positive ions while the electron affinity measures the attraction between an incoming electron and the nucleus. Fluorine is a small atom, therefore a high value of electron affinity is expected however the space being already crowded with electrons induces a significant amount of repulsion. In a molecule, the high electronegativity of fluorine also results in a large electronic effect on the neighbouring atoms, impacting the molecular dipole moment and subsequently its acido-basic properties. The Fluorine Van der Waals radius is only moderately larger than the hydrogen one (147 pm versus 120 pm) resulting in a low steric cost for the substitution of a hydrogen by a fluorine. The carbon-fluorine bond, with a length comprised between the C-H and the C-Cl, is the strongest in organic chemistry (Harper - 2003) with a dissociation energy of 485 kJ/mol (411 kJ/mol for C-H) resulting in an increased thermal stability.

	F	**H**	**Cl**
Electronic configuration	$1s^2 2s^2 2p^5$	$1s^1$	$1s^2 2s^2 2p^6 3s^2 3p^5$
Van der Waals radius (pm)	147	120	175
Electronegativity (Pauling scale)	3.98	2.2	3.16
First ionization energy (kJ/mol)	1681	1312	1251.2
Electron affinity (kJ/mol)	-328	-72	-348
Bond energy of C-X (kJ/mol)	485	411	327
Bond length of C-X (pm)	135	109	177

Tab 1.1 – Physicochemical data for fluorine, hydrogen and chlorine.

1.1.2 Specific aspects of fluorine chemistry

Because carbon-fluorine bonds are difficult to form (and to break), they are seldom found in nature. Although 3700 naturally occurring organohalogens are currently known to exist, only relatively few contain fluorine (Gribble, 2002). There are three main reasons for the dearth of natural fluorinated compounds (Murphy, 2003). First, there is a relatively low bioavailability of fluorine in comparison with the other halogens, since it is found in a largely insoluble form. Thus the concentration of fluoride in seawater (1.3 ppm) is several orders of magnitude lower than that of chloride (19 000 ppm). Second, owing to its high heat of

hydration, the fluoride ion is a poor nucleophile in aqueous solution, thereby limiting its participation in displacement reactions. Finally, the fluorine's high redox potential precludes the haloperoxidase-type mechanism used in the metabolic incorporation of chloride and bromide ions (Dong, 2004). Some species of plants and bacteria found in the tropics produce fluorine-containing metabolites. There are only 6 of those naturally found fluorinated products and they are depicted in figure 1.2.

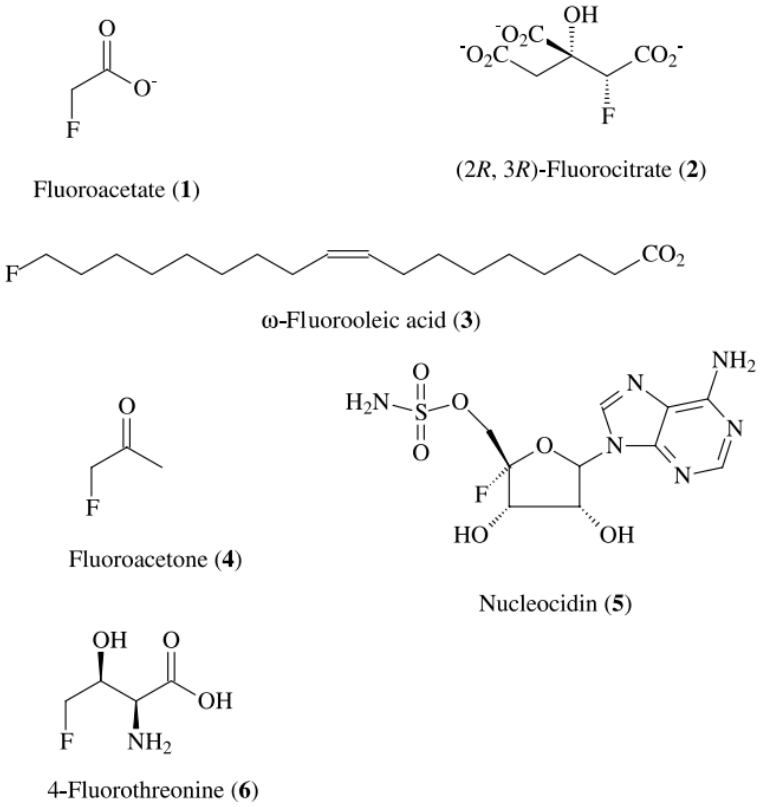


Fig 1.2 – Fluorinated products found in plants (1, 2, 3 and 4) and bacteria (1, 5 and 6) (Murphy, 2003)

The fluoroacetate is a well-known poison, used as a defense against herbivores by different plants (i.e. the gifblaar in south Africa) and as rodenticide in many countries, under the commercial name 1080. The soil bacterium *Streptomyces cattleya* biosynthesizes fluoroacetate and provides a convenient model to understand the mechanism of enzymatic carbon-fluorine bond formation, the strongest single bond in organic chemistry (Murphy, 2003). Its fluorinase enzyme (also known as adenosyl-fluoride synthase) catalyzes the reaction between fluoride ion and the co-factor S-adenosyl-L-methionine (SAM) to generate L-methionine and 5'-fluoro-5'-deoxyadenosine (FDA), which is the first product of the fluoroacetate pathway (illustrated in figure 1.3).

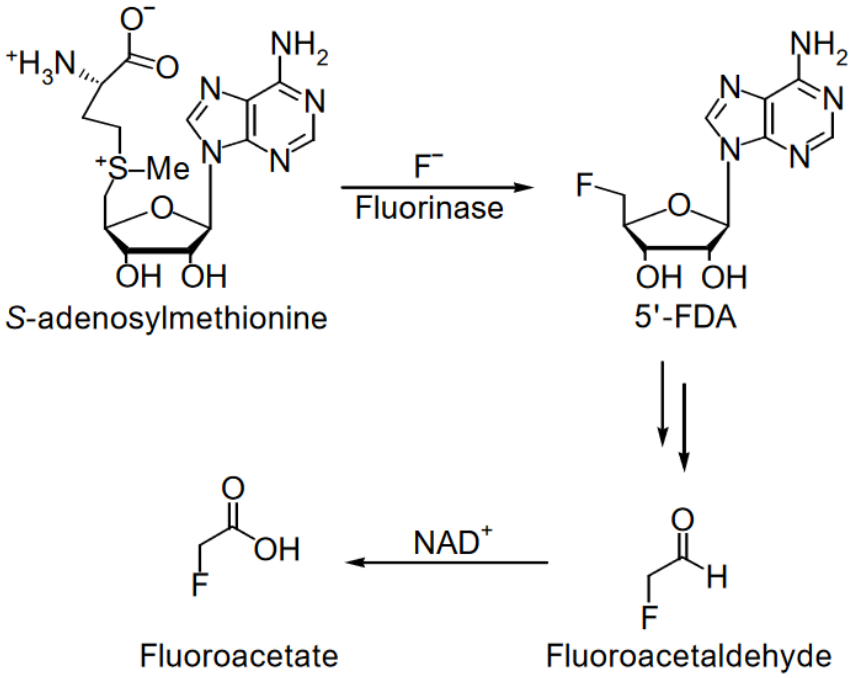


Fig 1.3 – Fluoroacetate biosynthesis pathway.

As seen with those distinctive properties, the introduction of a fluorine atom in a molecule will modify its conformational properties and influence its chemical reactivity. But getting fluorine into a molecule easily, selectively, and safely presents numerous synthetic difficulties. Although the first examples of nucleophilic and electrophilic fluorination reactions were reported in the 19th century, fluorine chemistry is still challenging, lacking general predictability and practicality (Liang – 2013). Due to the fluorine's high electronegativity and high hydration energy of fluoride anion the control of the C-F bond formation presents several difficulties. There are two main approaches: introducing a fluorine moiety at a specific synthesis step or converting an initial molecule that is already fluorinated. Progress was recently achieved with the development of fluorinating agents (DAST, SF4, Selectfluor, Deoxo-fluor…) but those requires specific expertise and can often be expensive. Commercially available starting materials were then developed, tailored for laboratories working on small-scale projects. Mastering when and how to graft this element in the context of a complex organic molecule is essential to elaborate efficient drug candidates since the resulting effects may be somewhat cryptic in nature rather than simply additive (Gillis - 2015). As an example, the fluorine-substituted methyl group, CF_3, leads to a

completely different reactivity when compared to CH_3. A CF_3 substitution on an aromatic ring will have a strong electron withdrawing effect and will direct a second substituent to the meta position (91%, as opposed to only 3% meta for a methyl group, an ortho/para director). With a single fluorine atom, the weaker electron withdrawing effect favors rather the para (86%) (Carey – 2013). In aliphatic systems, the fluorine induced effects produce a strong preference for vicinal groups to be aligned *gauche* to fluorine (figure 1.4), with a stabilization energy that ranks between 2.1 to 24.2 kJ/mol (Gillis – 2015).

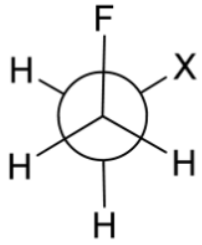


Fig 1.4 – Gauche-isomer of a X-substituted fluoroethane derivative.

For instance, introducing a fluorine in a proline changes the conformation of the pyrrolidine ring due to a favorable *gauche* interaction between the fluorine atom and the N-C=O group, an effect that also impacts the peptide bond. 4-(R)-fluoroproline and 4-(S)-fluoroproline are proline analogs that are used to investigate the mechanisms of 4-(R)-hydroxyproline in enhancing the thermal stability of collagen, the major structural protein in connective tissues. Instead of a network of H-bonds between hydroxyl groups and water molecules, the work realized on fluoroprolines revealed that it is rather the conformation of the pyrrolidine ring, shaped by the electron-withdrawing effect of the hydroxyl substituent, that provides the thermal stability of the collagen helix. In practice, polypeptides of *proline-4(S)fluoroproline-glycine* form a triple helix structure while polymers of *proline-4(R)fluoroproline-glycine* do not (Doi – 2003).

The effect of fluorine on the biological activity of a molecule is usually difficult to predict (Wang 2014), but the statistic accumulation of data allows general prediction and underlines the importance of fluorine in drug design. The physicochemical properties of fluorine transformed it from an unfit candidate for biological applications to a key tool in drug design.

1.1.3 Presence of fluorine in pharmaceuticals

The number of fluorine-containing drugs in 1970 was only around 2% while nowadays fluorine is present in more than a fifth of all pharmaceutical drugs. Three of the top ten best sellers have fluorine atoms: the antiasthma Seretide (GlaxoSmithKline), the proton pump inhibitor Prevacid (Novartis) and the lipid lowering agent Lipitor (Pfizer) that generated more revenues than any other drug ($US 5.9 billion in 2008), before the apparition of generic versions in 2011. Other well-known fluorinated drugs include the anti-depressant Prozac (Lilly) and many antibiotics, starting with the sulfanilamides introduced in the mid-1930s. The chemical structure of some blockbuster drugs are presented in figure 1.5. This situation can be explained by favorable changes in ligand pharmacokinetics and/or pharmacodynamics properties caused by the introduction of a fluorine into lead pharmaceutical compounds (Vulpetti – 2012). These pharmaceutical properties are provided by the shifted electronic, steric and lipophilic parameters. One of the main reason for the strong proportion of fluorine in medicinal chemistry is the modulation of acidity and basicity of the initial compound (Wang – 2014). Lipophilicity is increased when a fluorine group is positioned near basic nitrogens or in an aromatic ring. Adding a fluorine in a ligand can then help it partition into membranes, and thus boosting its rates of absorption and transport of drugs in-vivo (Filler – 2009).

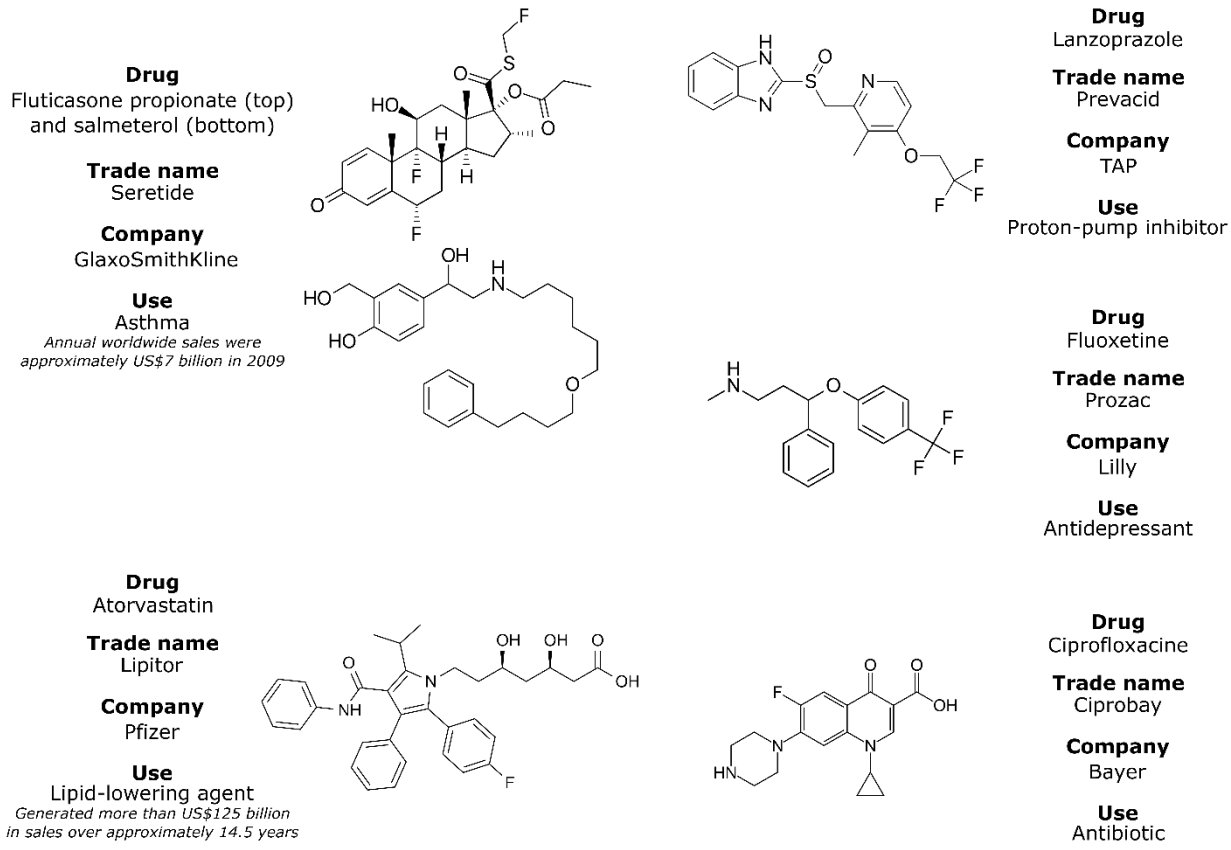


Fig 1.5 – Chemical structure and information regarding well-known fluorinated pharmaceutical drugs.

Several chemical companies offer screening collections and fragment libraries, that include fluorine labelled fragments. The 2016 BIONET Fluorine Fragment library provides an overview of the distribution of the fluorinated groups in such libraries. In figure 1.6 are summarized the ^{19}F NMR shifts of the set of 461 fragments. There are 199 CF_3 (-52.9 to -84.7 ppm), 3 OCF_3 (-57 ppm), 9 OCF_2H (-81.0 to -81.7 ppm), 2 CF_2 (-88 and -104ppm) and 252 F (here -96 to -141 ppm). These data indicate a nearly half-half distribution between the fluorine present on aromatic groups and the CF_3 moieties.

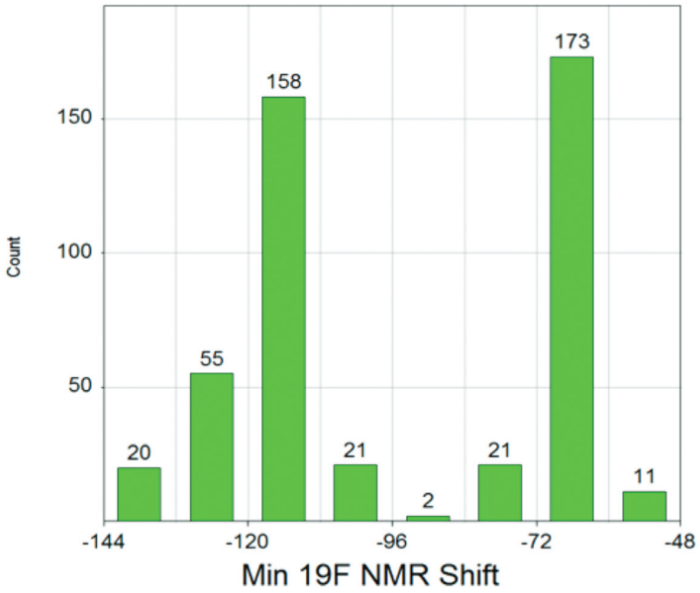


Fig 1.6 – ^{19}F NMR chemical shifts count for the 2016 BioNet chemical library of fluorinated fragments (source: http://www.keyorganics.net/services/bionet-products/fragment-libraries/). The -120ppm corresponds to fluorine in aromatic environment while the -70ppm represents the CF_3 NMR signature.

The properties of amino-acids can also be modulated by fluorine substitution, providing interesting new perspectives to develop drugs based on peptides. Fluorinated amino acids have found applications as biological tracers, enzyme inhibitors and significantly impacted the stability or the reactivity of proteins (Qiu – 2004). Fluorinated amino acids can be divided into three main groups based on the site of fluorination: α, β and cyclic. The (S)-α-fluoromethylhistidine is an irreversible inhibitor for the histidine decarboxylase, and can then be used as an antihistamine. Fluorinated β-amino acids present a greater structural diversity (due to the increased possible fluorination sites) and can also serve as precursors for β-lactam antibiotics. Taxol, a famous anticancer drug owned by Bristol-Myers Squibb, can have one of its side chain, N-Bz-(2R,3S)-3-phenylisoserine, replaced by a fluorinated β-amino acid N-Bz-(2R,3S)-3-CF_3-isoserine, that results in a tenfold increase of its biological activity against cancer (Qiu – 2006). Figure 1.7 summarizes the amino acids fluorinated analogs that have been used in peptide studies and non-canonical genetic code engineering (Merkel – 2012).

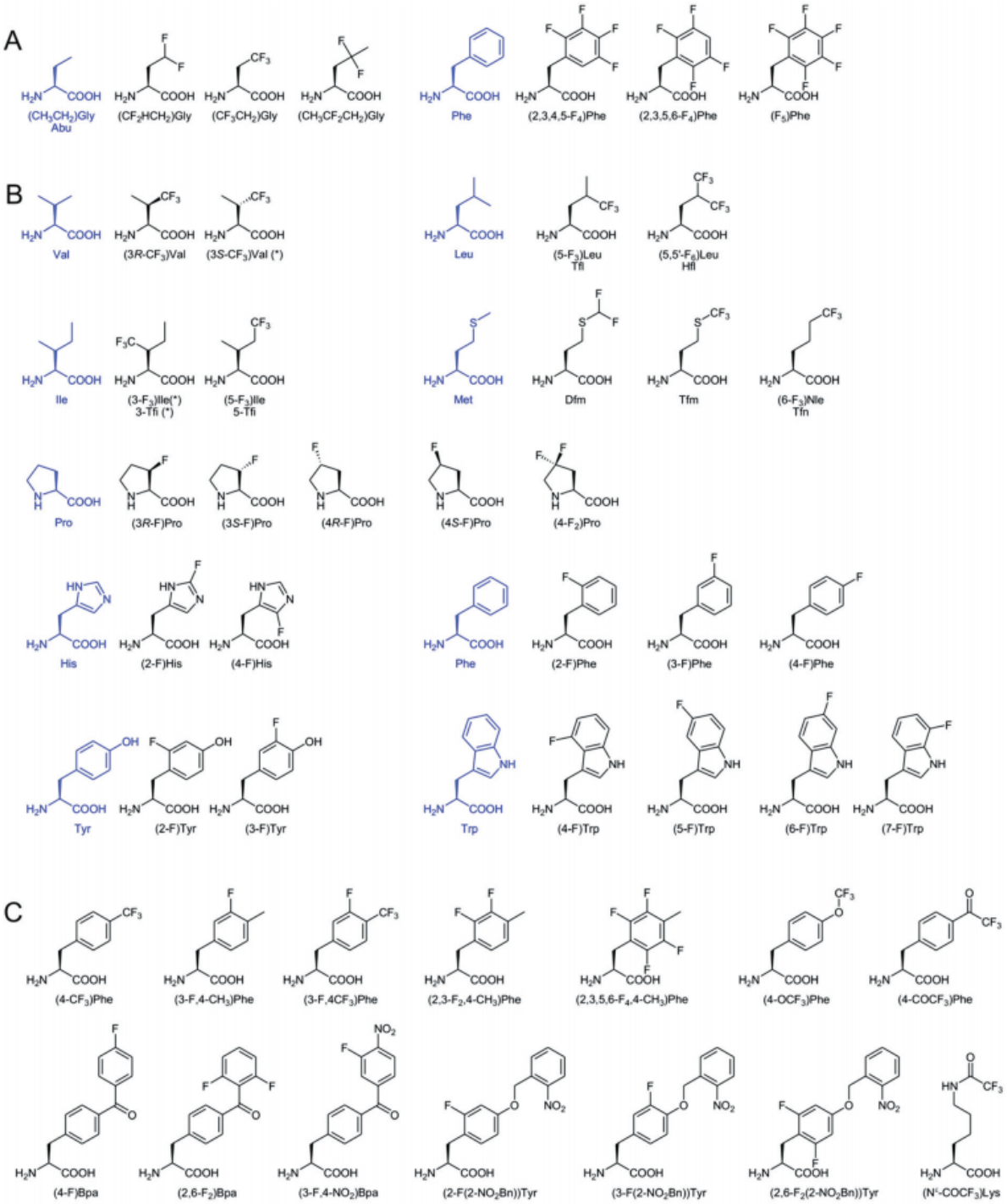


Fig 1.7 – Fluorinated non-canonical amino acids (as reviewed in Merkel, 2012). The blue formulas represent the canonical amino acids. Some scaffolds have been used only for peptide studies (A), other were also used in either genetic code engineering (B) or genetic code expansion (C). The formulas marked with an asterisk could not be incorporated in vivo.

One of the first incorporation of fluorinated amino-acids in protein was reported by Ian Campbell's team in the 90's in order to resolve resonances of macromolecules, such as the 380 kDa ATPase (Kim-1990). Other aromatic amino-acids were employed as probes, such as 5-fluorotryptophan and 3-fluorotyrosine, to detect bromodomain–ligand interactions (Mishra – 2014). 3-fluorophenylalanine was used to measure the lifetime of Calmodulin peptide-bound states through ^{19}F NMR relaxation dispersion measurements (Hoang-2014). The advantages of favoring tryptophan over tyrosine or phenylalanine as a fluorinated probe are its lower abundance (thus fewer signals relative to ^{19}F-Tyr or ^{19}F-Phe) and its frequent occurrence at protein interfaces (Crowley-2012). Fluorotryptophan is also a useful fluorescent probe. Peter Crowley and its team showed that Escherichia coli BL21, the workhorse of protein production, can incorporate fluoroindole for the biosynthesis of proteins containing ^{19}F-tryptophan. Using a tryptophan precursor instead of analogues and inhibitors resulted into a simplified process and lower costs. However, high concentrations of fluoroindole are toxic (60 $mg.L^{-1}$ of 5-fluoroindole led to the same yield than in minimum media but 300 $mg.L^{-1}$ of 5-fluoroindole reduced the total cell mass obtained by nearly a fourth (Crowley-2012)). A greater ease of production of fluorine-labelled proteins is required to promote the advancements in ^{19}F protein NMR and in particular in the field of in-cell NMR spectroscopy. It is currently clear that all advantages and possibilities of fluorinated peptides and proteins have not yet been exploited and further development of chemical and biochemical tools are needed to achieve this goal.

1.2 Fluorine NMR

Nuclear magnetic resonance (NMR) is a well-adapted technique for the study of fluorine's single natural isotope, ^{19}F. The versatility of the NMR analysis can be combined to the remarkable properties of fluorine to create a multi-faceted powerful tool including the experimental characterization of fluorinated drugs, the study of their metabolism, the design of binding assays between those ligands and their targets, and even protein structural studies. The sensitivity of the ^{19}F resonance to its environment can also be exploited to explore the dynamical features of molecular interaction (Danielson – 1996).

1.2.1 ^{19}F NMR properties

Several factors can be listed to explain the power of ^{19}F NMR. First, fluorine has a single stable isotope, ^{19}F, characterized by a ½ spin, favorable to NMR measurements. This isotopic proportion of 100% can be compared to the 1% proportion of ^{13}C or to the 0.36% natural abundance of ^{15}N that both result in major sensitivity penalties. ^{19}F gyromagnetic ratio is only 94% of the proton one ($251.662*10^6$ rad.s^{-1}.T^{-1} versus $267.513*10^6$ rad.s^{-1}.T^{-1}) providing interesting sensitivity properties. The intensity of NMR signals depends upon the magnetic moment, the Boltzmann population and the nuclear precession, all three of them being proportional to the gyromagnetic ratio. This means that a 0.94 ratio in gyromagnetic ratio leads to an 83% sensitivity ratio between ^{19}F and ^{1}H. Table 1.2 summarizes the properties of the most common nuclei in NMR.

Nucleus	**Gyromagnetic ratio** ($*10^6$ rad.s^{-1}.T^{-1})	**NMR Frequency** (MHz) at 14.0954 T	**Natural abundance** (%)	**Relative sensitivity** (constant number of nuclei)	**Chemical shift range** (ppm)
^{1}H	26.7519	600.13	99.9885	1.0	13
^{19}F	25.181	564.686	100.0	0.832	400
^{13}C	6.7283	150.903	1.07	0.0159	250
^{15}N	2.712	60.834	0.364	0.00104	1700
^{31}P	10.841	242.938	100.0	0.0665	530

Tab 1.2 – NMR properties for nuclei routinely used in structural biology (adapted from (Gerig – 2001)).

The lack of background fluorine signal in biological samples provides an additional advantage to the use of ^{19}F NMR to probe molecular interactions in complex media (illustrated in figure 1.8).

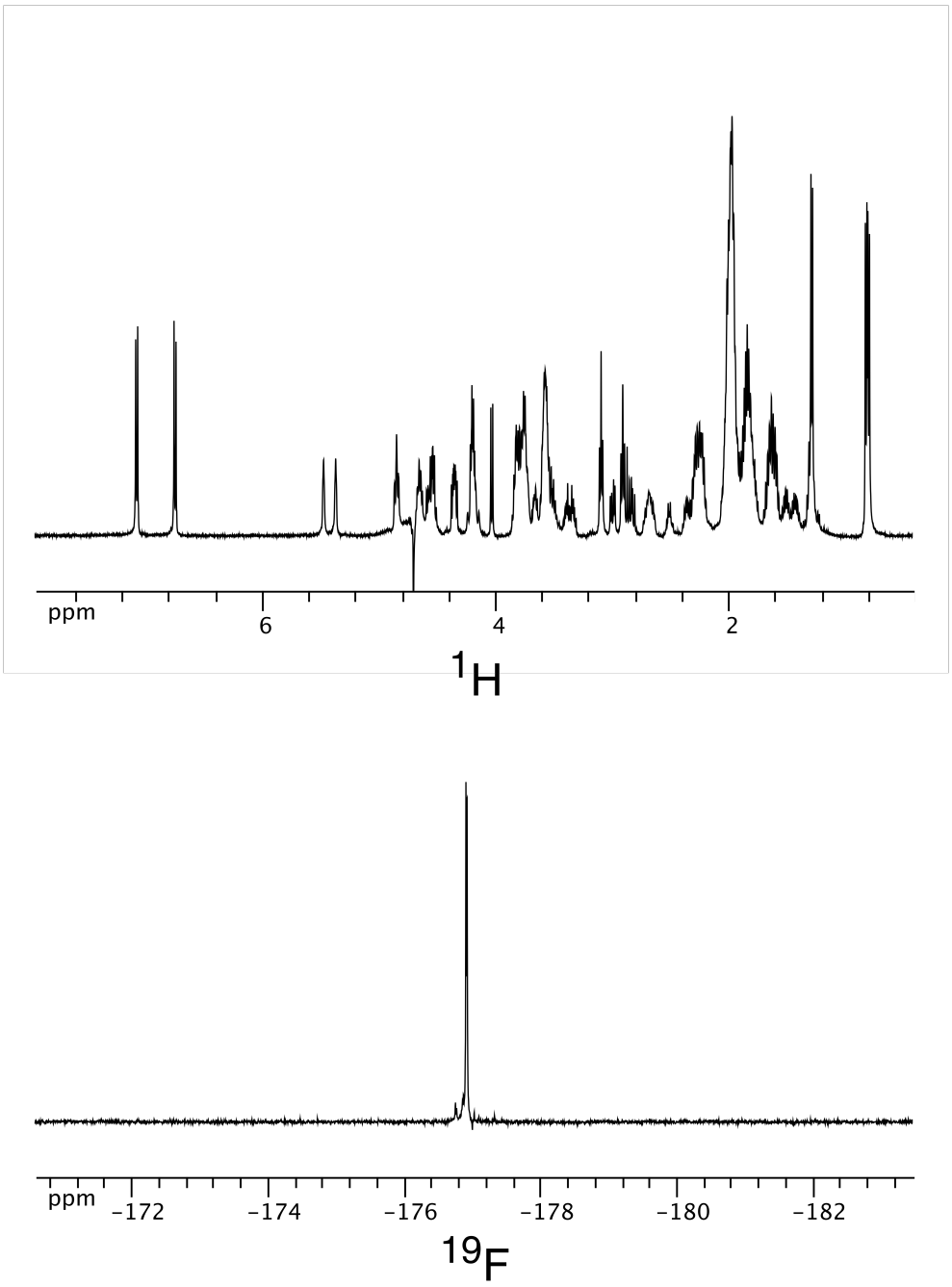


Fig 1.8 – The same mixture, composed of two fluorinated compounds (TFE and linezolid) in a ribosomal buffer, was observed through proton NMR (top spectrum) and fluorine NMR (bottom spectrum). The juxtaposition of the two experiment illustrates the main properties of a 19F NMR spectrum: a ten-fold increased spectral width, no overlapping of the signal and no background noise.

1.2.1.1 Fluorine chemical shifts

The NMR parameters of fluorine are very sensitive to small changes in structure and environment. The effective field strength B_{eff} is related to the external field B_0 by equation 1.1:

$$B_{eff} = B_0 * (1 - \sigma) \quad \textit{[eq 1. 1]}$$

σ is a dimensionless shielding constant that can be divided in three components (equation 1.2):

$$\sigma = \sigma_{dia} + \sigma_{para} + \sigma_i \quad \textit{[eq 1. 2]}$$

The diamagnetic term σ_{dia} corresponds to the opposing field resulting from the effect of the imposed field on the electron cloud surrounding the nucleus (Dolbier, 2009). The paramagnetic term σ_{para} originates from the presence of magnetic dipoles (for instance due to p-electrons) that get oriented in the magnetic field. The induced local magnetic field adds up to B_0 and its impact is opposite to that of diamagnetic shielding. Finally, the term σ_i derives from the effect of neighboring groups, which can increase or decrease the local field at the ^{19}F nucleus position. The main reason explaining the large differences between the proton and the fluorine chemical shift range is the lack of the paramagnetic contribution in the case of the proton, that has only *s* orbital electron. For fluorine, the electron lone-pair weights on the paramagnetic term, making it dominant in determining the relative shielding of fluorine nuclei. The common chemical solvent fluorotrichloromethane ($CFCl_3$) has become the accepted, preferred internal reference for ^{19}F NMR spectra. In practice, ^{19}F frequencies span a chemical shift range of ~1100 ppm and figure 1.9 illustrates the predicted resonance positions of various fluorinated groups. Single carbon-bound F-substituents have a greater sensitivity to their environment than carbon-bound CF_2 or CF_3 substituents. Indeed, single fluorine chemical shifts, which encompass vinylic, aryl, and saturated aliphatic fluorine substituents, are found between −70ppm and −238 ppm, whereas the corresponding range for CF_2 groups is −80 to −130 ppm, and that of the CF_3 group is even smaller, between about −52 and −87 ppm (Dolbier - 2009). The CF_3 group has the advantage -at equal concentration- of a superior resonance intensity and will avoid $^2J_{HF}$ mutual coupling like CF_2H or CH_2F (but CF_3 groups could still have $^3J_{HF}$ couplings).

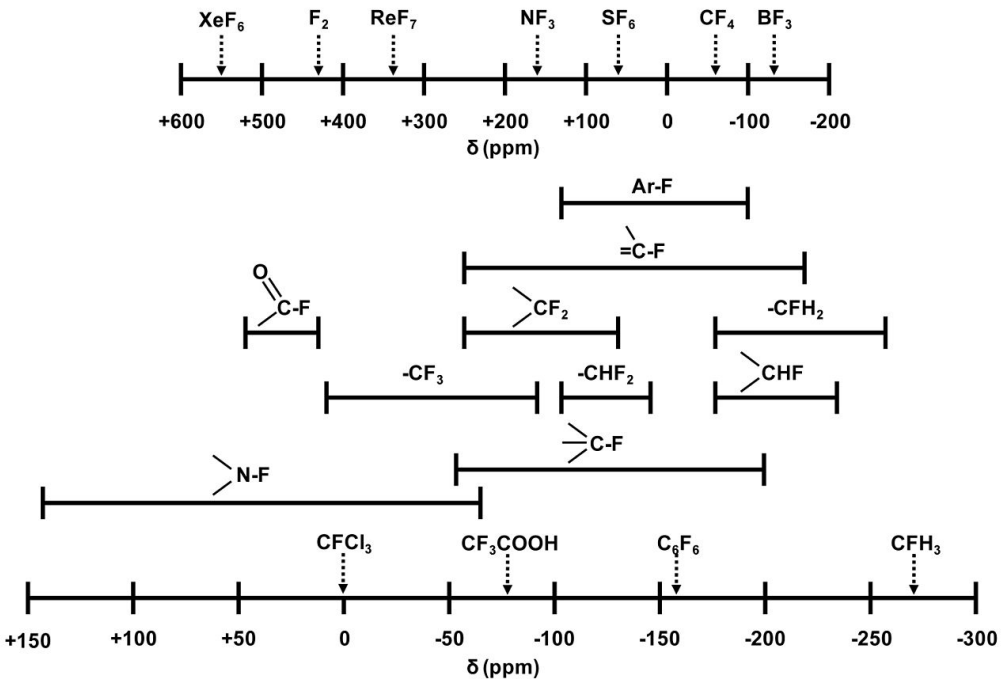


Fig 1.9 – ^{19}F chemical shifts for different fluorochemicals and fluorinated fragments (Wikimedia commons, under a CC-BY-SA-3.0 license)

Since fluorine is the most electronegative element, its introduction in a molecule will have a strong influence on the nearby groups, de-shielding them. Protons chemical resonances can be shifted by a fluorine group as far as six covalent bonds away, as seen in figure 1.10.

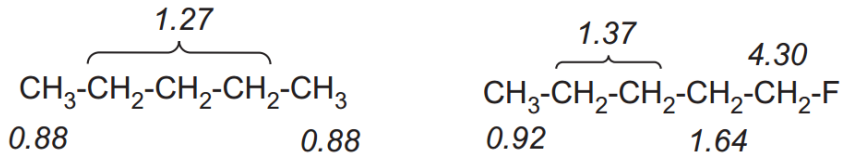


Fig 1.10 – Effect of a fluorine substitution on the proton chemical shifts of a pentane molecule. (Dolbier – 2009).

A similar effect can be observed for the carbon chemical shifts, albeit a bit weaker (figure 1.11).

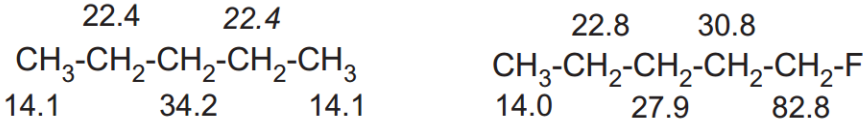


Fig 1.11 – Effect of a fluorine substitution on the carbon chemical shifts of a pentane molecule. (Dolbier – 2009).

With such a wide range for chemical shifts, overlapping of resonance frequencies happen rarely.

1.2.1.2 Fluorine couplings

Fluorine spin-spin coupling constants to other fluorine nuclei, to neighboring proton, and to carbons are highly variable in magnitude and also depends highly on the environment. The number of fluorine substituents on a carbon atom affects significantly the value of vicinal coupling. Fluorine scalar coupling constants are typically much larger than proton one: $^{2}J_{FF}$ usually range between 100 and 200 Hz, $^{3}J_{FF}$ is often less than 15-16 Hz and long range coupling up to $^{7}J_{FF}$ can be observed (Tressaud-2008). Geminal coupling constants between fluorine and proton are around $^{2}J_{FH}$ ~45-60 Hz and vicinal one $^{3}J_{FH}$ ~0-53 Hz, exhibiting large variations. Fluorine carbon couplings are typically large ($^{1}J_{CF}$ > 150 Hz) and could be observed as ^{13}C satellites on fluorine spectra. Fluorine homonuclear vicinal coupling constants are heavily dependent on the *cis* or *trans* conformations. *Trans* coupling constant are much larger than their *cis* counterpart. This can be illustrated with the vicinal fluorines in 1-chloro-1,2-difluoroethenylbenzene (figure 1.12) where the cis-vicinal fluorines also deshield each other.

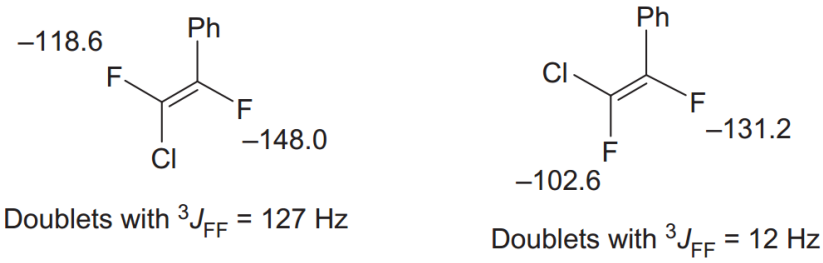


Fig 1.12 – ^{19}F chemical shifts and coupling constants comparison for the trans (left) and cis (right) conformations of 1-chloro-1,2-difluoroethenylbenzene (Dolbier – 2009)

The fluorine signal can have multiplets resulting from fluorine, carbon and hydrogen coupling since those are all ½ spins. The trifluoromethyl group of 2,2,2-trifluoroethanol will be split into a triplet ($^{3}J_{HF}$ = 9.12 Hz) by the neighboring two protons in its fluorine NMR spectra (figure 1.13A). The same CH_2 protons are split into a doublet of doublets by the CF_3 group ($^{3}J_{FH}$ = 9.12 Hz). The coupling pattern of a fluorine nucleus quickly become harder to untangle once several proton nuclei are in the vicinity. As seen in figure 1.13B, in a fluoroproline, the signal of the fluorine nuclei appears as a multi-coupled resonance.

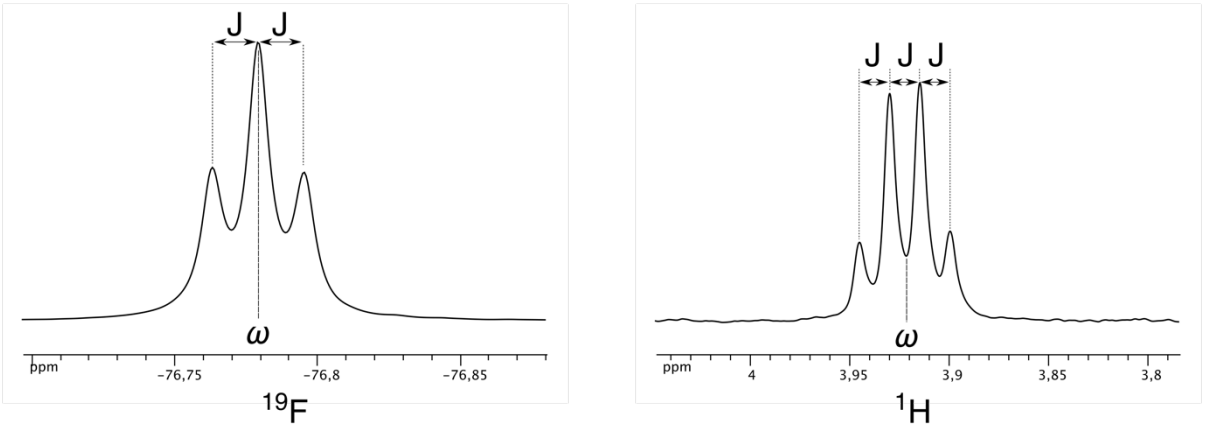


Fig 1.13A – ^{19}F spectrum of the CF_3 group of trifluoroethanol (at -76.77 ppm) and ^{1}H spectrum for the CH_2 group of the same molecule (at 3.92 ppm). Both spectra were recorded at 298K, on a 14.1T spectrometer.

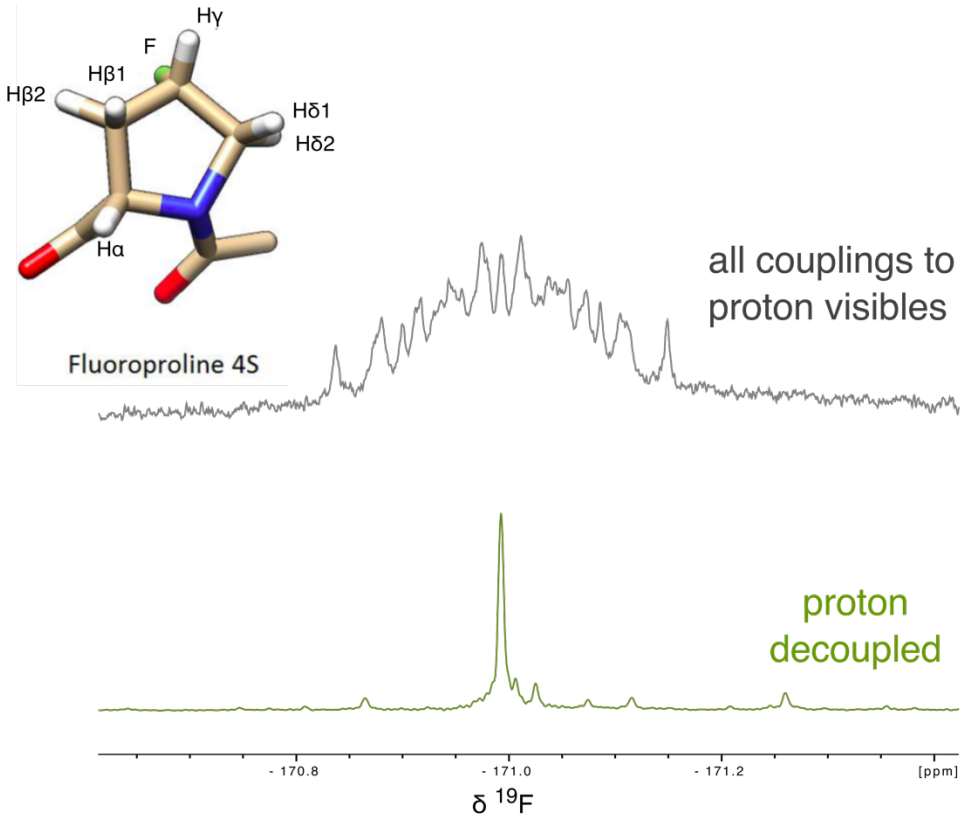


Fig 1.13B – ^{19}F spectra of a fluoroproline, without (top spectrum) and with (bottom spectrum) proton decoupling. Both spectra were recorded at 298K, in 64 scans on a 14.1T spectrometer. The ^{1}H decoupling used a GARP pulse.

The ^{19}F-^{13}C couplings can be evidenced on 1D fluorine spectra by the presence of satellites. In contrast to the ^{1}H-^{13}C satellites visible on proton spectra, the ^{19}F-^{13}C satellites are not

centered around the ^{19}F-^{12}C peak, as shown in figure 1.14. This is due to the large isotopic shift that shifts significantly the resonances of fluorine nuclei bound to a ^{13}C. Approximatively 1% of the signal is from the $^{13}CF_3$-$^{12}COOH$ isotopic isomer (also called isotopomer) giving rise to a doublet with $^{1}J_{F\text{-}C}$ = 292 Hz (shifted of 0.13 ppm from the main resonance line) and approximately 1% of the signal is from the $^{12}CF_3$-$^{13}COOH$ isotopomer resulting in a doublet with $^{2}J_{F\text{-}C}$ = 35.6 Hz (shifted of 0.01 ppm from the main resonance line). The spacing between the satellites is a scalar coupling and is thus field independent (constant value in Hz), while the center of the satellites is the shift of the ^{13}C isotopomer, and the difference between the ^{13}C and ^{12}C isotopomers is field dependent (constant in ppm). It is remarkable that $^{2}J_{F\text{-}C}$ are observed in fluorine spectra, providing an illustration of the sensitivity of fluorine NMR to its environment.

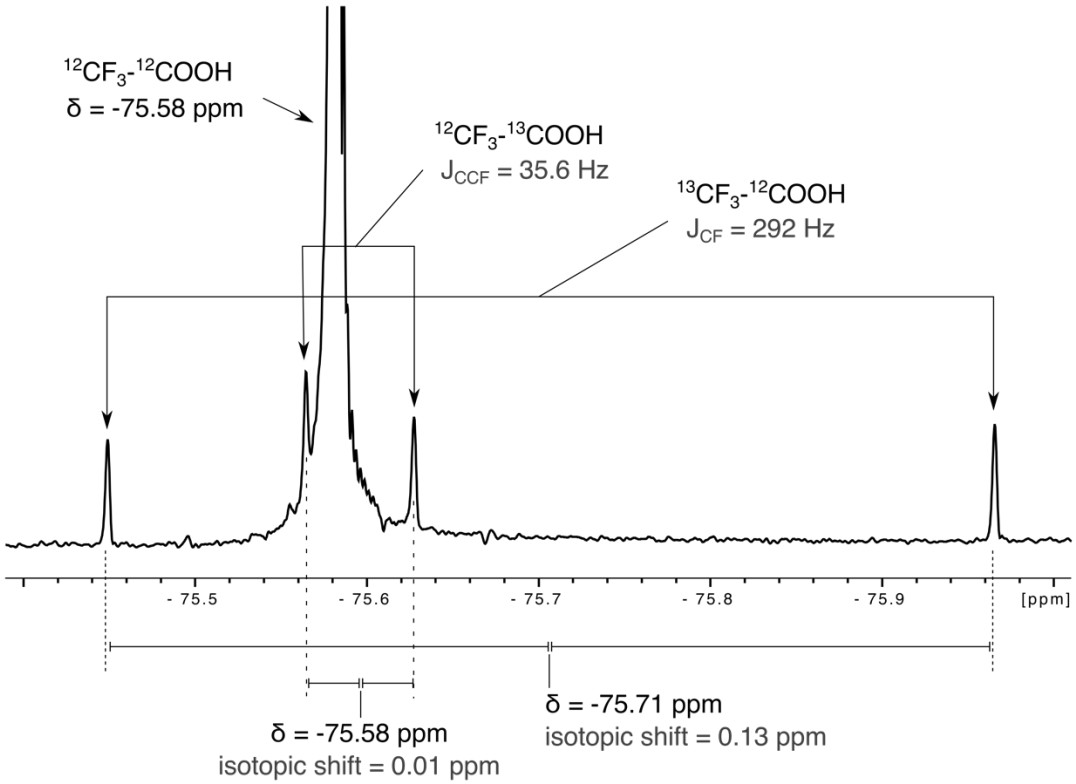


Fig 1.14 – ^{19}F chemical shift of the CF3 group of trifluoroethanoic acid zoomed in to show the carbon satellites resulting from $^{1}J_{FC}$ and $^{2}J_{FC}$ couplings.

The principal factors influencing vicinal F-F and F-H couplings are the torsional angles between the coupled nuclei. There is a Karplus-type dependence on the dihedral angle formed by the nuclei, confirmed empirically (Williamson – 1968) that can be used to deduce

the conformation of molecules. All coupling is transmitted by electrons, whether implied in covalent bonds or in unshared pairs. Independently from the *J* number of covalent bonds linking two nuclei, if their Van der Waals radii overlap (and if at least one of the nuclei has an un-bonded electron pair), they can exchange spin information through space. Figure 1.15 shows this "through space" coupling.

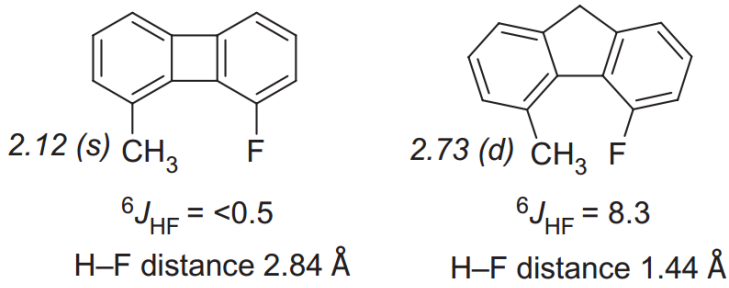


Fig 1.15 – Long range coupling example between a fluorine atom and a methyl group. The fluorine spatial position influences the chemical shift and the peak shape of the methyl group (Dolbier – 2009).

Due to the enhanced chemical shift ranges, fluorine NMR has had a significant impact in fragment based drug discovery where cocktails of ligand accelerate the screening process. For protein studies, the chemical shift dispersion observed in the incorporated fluorinated amino acids (up to nearly 15 ppm (Lian – 1994)) cannot be attributed to variations of chemical structure but rather to differences in local secondary and tertiary structure, solvent exposure, and conformational dynamics (Gerig, 1994). This enhanced sensitivity is revealed through the chemical shifts but it can also be found in the relaxation rates. Indeed, in fluorine NMR, the transversal relaxation rate R_2 is highly sensitive to binding events. This property can be exploited to improve the binding detection.

1.2.2 NMR relaxation of ^{19}F

1.2.2.1 Relaxation mechanisms

Understanding relaxation processes in NMR is crucial for proper measurements and interpretations of spectra. Relaxation is defined as the return to a thermodynamic equilibrium after a RF pulse. In liquid at room temperature and for small molecules, this process takes a time ranging from hundreds of milli-seconds to few seconds. The relaxation is either reported as a time *T* or a rate *R=1/T*. Two types of relaxation can be distinguished: the longitudinal relaxation (R_1, NOE), also called spin-lattice relaxation, and the transverse relaxation (R_2), caused by spin-spin interactions and describing the coherence disappearance between the spin states. While the first one is due to the establishment of thermal population of spin states, the transverse relaxation describes the loss of coherences between these spin states. A fast relaxation can lead to broad peaks, blurring the attribution of the resonances and the J-coupling. Indeed, the R_2 is directly related to the width of a resonance through equation 1.3:

$$LW_{1/2} = \frac{1}{(\pi T_2)} = \frac{1}{(\pi)} R_2 \qquad \textit{[eq 1. 3]}$$

Relaxation is caused by random magnetic field perturbations, the principal source for those being Brownian molecular motions. The molecular motions can affect nuclear relaxation through many different mechanisms: the most common one being dipole-dipole interaction that depends on the distance between the pair of dipoles (spins) but also on their orientation relative to the external magnetic field. The chemical shift anisotropy (CSA) represent an important source of relaxation when the electronic environment is anisotropic. The magnitude of the chemical shift will depend on the molecular orientation relative to the (fixed) external magnetic field. Other mechanisms include the quadrupolar relaxation (QR), the scalar coupling –also called J-coupling modulation- (SC), Spin rotation (SR), the anisotropic relaxation, oxygen relaxation, etc. For the T_1 relaxation of fluorine, the dipole-dipole relaxation mechanism dominates, as for proton, but spin-rotation and CSA effects are enhanced in fluorine, due to the non-paired electron pair.

The spectral density function provides an effective mean to quantify the frequency distribution of the local magnetic field fluctuations that contributes to relaxation processes.

Assuming a random isotropic motion characterized by a correlation time τ_c, the spectral density function is a lorentzian function of a general form:

$$J_q(\omega) = C_q \left[\frac{\tau_c}{(1 + \omega^2 \tau_c^2)} \right] \qquad \textit{[eq 1. 4]}$$

With (q=0,1,2) and C_0=24/15, C_1=4/15 and C_2=16/15.

In a two spins system, with an intramolecular heteronuclear dipole-dipole mechanism of relaxation, equation 1.4 can be used to describe the influence of a proton *H* on the T_1 relaxation of a fluorine spin *F* (in a simplified model: a single correlation time τ_c for the molecule and a fixed r_{FH} distance):

$$R_1^F = \frac{1}{T_1^F} = E_{FH} r_{FH}^{-6} \left[\left(\frac{1}{12}\right) J_0(\omega_F - \omega_H) + \left(\frac{3}{2}\right) J_1(\omega_F) + \left(\frac{3}{4}\right) J_2(\omega_F + \omega_H) \right] \text{ [eq 1.4]}$$

$$where\ E_{FH} = \left(\frac{\mu_0}{4\pi}\right)^2 (\gamma_F \gamma_H \hbar)^2 S_H (S_H + 1)$$

Where in this equation γ_F and γ_H, respectively, are the gyromagnetic ratios of fluorine and hydrogen, $\hbar$ is Planck's constant divided by 2π, ω_F and ω_H are the resonance frequencies for fluorine and hydrogen and S^H the spin number of hydrogen, 1/2. The correlation time τ_c describes the rate of molecular motions through the time taken for a molecule to rotate by an angle of 1 radian. The correlation time can be roughly (for spherical molecules in non-viscous solvents near room temperature) approximated by: $\tau_c = 10^{-12} M_w$ where M_w is the molecular mass in Daltons (Gerig-2001). Figure 1.16 displays the value of the heteronuclear and hom

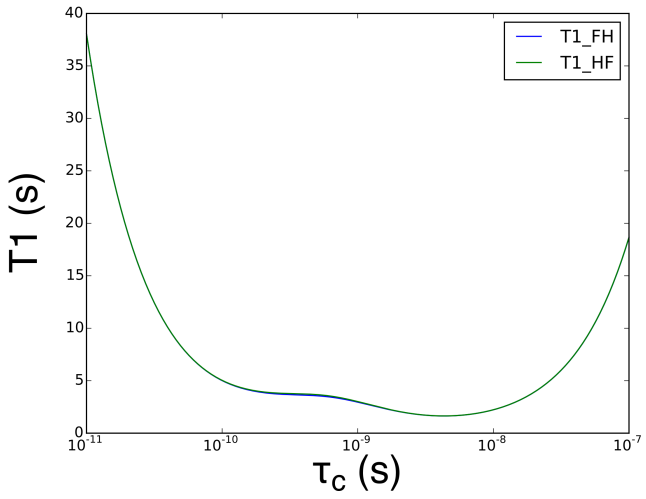


Fig 1.16 – Predicted heteronuclear dipole-dipole contribution to the T_1 relaxation at 564MHz following equation 1.4, in two cases: if fluorine is relaxed by a proton spin and the opposite. Both curves are very similar.

The graph in figure 1.16 shows that the T_1 depends on the tumbling of the system and that the same value of T_1 can be found at different correlation times. Except for τ_c values a little below the nanosecond, the dipole-dipole longitudinal relaxation mechanism is identical between fluorine and proton nuclei. The dipole-dipole interaction affects the transverse relaxation according to equation 1.5:

$$R_2^I = \frac{1}{T_2^I} = E_{IS} r_{IS}^{-6} \left[\left(\frac{1}{6}\right) J_0(0) + \left(\frac{1}{24}\right) J_0(\omega_I - \omega_S) + \left(\frac{3}{4}\right) J_1(\omega_I) + \left(\frac{3}{2}\right) J_1(\omega_S) + \left(\frac{3}{8}\right) J_2(\omega_I + \omega_S) \right] \quad \textit{[eq 1. 5]}$$

The relation between the τ_c and the transversal relaxation with the DD mechanism is once again better understood once displayed as a graph:

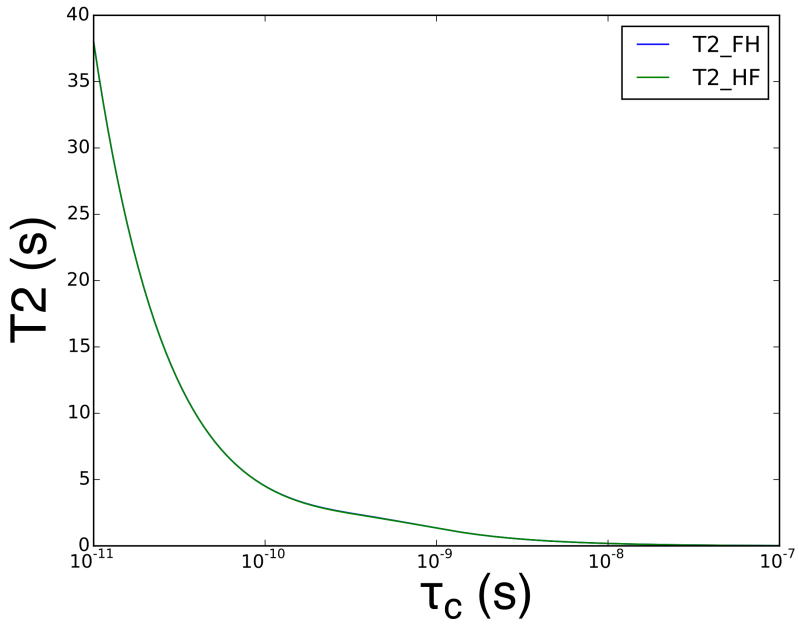


Fig 1.16 – Predicted heteronuclear dipole-dipole contribution to the T_2 relaxation at 564MHz following equation 1.5, in two cases: if fluorine is relaxed by a proton spin and the opposite. Both contributions are nearly identical.

As for longitudinal relaxation, the dipole-dipole perturbation of the transversal relaxation is very similar between a fluorine and a proton. In the case of homonuclear intramolecular dipole-dipole relaxation, there are more visible discrepancies, although the calculated curves still follow the same tendencies:

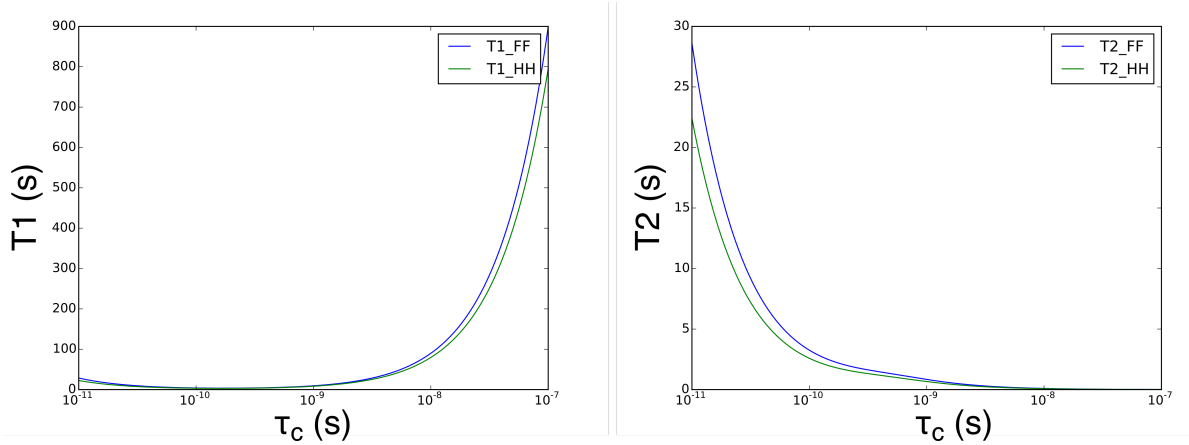


Fig 1.17 – Predicted homonuclear dipole-dipole contribution to the T_1 and T_2 relaxation at 564MHz for the fluorine and proton nuclei.

The equations used in the computing of the figure 1.17 were (with *I* designating either the fluorine or the proton spin):

$$R_1^I = \frac{1}{T_1^I} = E_I r_I^{-6} \left(\frac{3}{2}\right) [J_1(\omega_I) + J_2(2\omega_I)] \qquad \textit{[eq 1. 6]}$$

$$R_2^I = \frac{1}{T_2^I} = E_I r_I^{-6} \left[\left(\frac{3}{8}\right) J_0(0) + \left(\frac{15}{4}\right) J_1(\omega_I) + \left(\frac{3}{8}\right) J_2(2\omega_I)\right] \qquad \textit{[eq 1. 7]}$$

$$where\ E_I = (\frac{\mu_0}{4\pi})^2 (\gamma_I{}^2 \hbar)^2 I(I+1)$$

1.2.2.2 CSA effect on the fluorine relaxation

The movement of electrons is easier in some molecular orientations relative to the B_0 field, leading to an anisotropic shielding parameter σ. This shielding value will be averaged with the molecular motions in solution, but will cause oscillations of the local magnetic field, contributing to the T_1 and T_2 relaxations. The equations 1.9 and 1.10 show the contribution of fluorine CSA to the T_1 and T_2 relaxations using a symmetric ellipsoid chemical shift tensor:

$$R1 = \frac{1}{T1} = \frac{2}{15} . E_{CSA} . (\frac{\tau_c}{(1+\omega^2\tau_c^2)}) \qquad \textit{[eq 1. 8]}$$

$$R2 = \frac{1}{T2} = \frac{1}{90} . E_{CSA} . (8\tau_c + \frac{6\tau_c}{(1+\omega^2\tau_c^2)}) \qquad \textit{[eq 1. 9]}$$

$$where\ E_{CSA} = \gamma^2 B_0^2 \Delta\sigma^2\ and\ \Delta\sigma = (\sigma_\parallel - \sigma_\perp)\ in\ ppm$$

The strong difference between fluorine and proton in relaxation is in the term $(\sigma_{\parallel} - \sigma_{\perp})$ that is squared in the equation. In the case of proton, the CSA is about 5 ppm while values as large as 1000 ppm can be observed for fluorine. The $R_{2,(CSA)}$ is proportional to the square of the field strength, getting much more important at high field strengths up to the point where $R_{2,(CSA)}$ becomes the dominant contribution to the relaxation of fluorine (Brey – 1996). Figure 1.18 displays both the dipole-dipole and CSA relaxation times, as well as the sum, as a function of the frequency, for a molecular tumbling of 0.5 nanoseconds. It is shown that the CSA relaxation is correlated with the field increase, and that it becomes the predominant relaxation mechanism above 200 MHz. With an increased τ_c (i.e. when bound to a macromolecule), the CSA contribution result in even higher T_2 times. The Transverse Relaxation Optimized SpectroscopY experiment (TROSY) was spe-

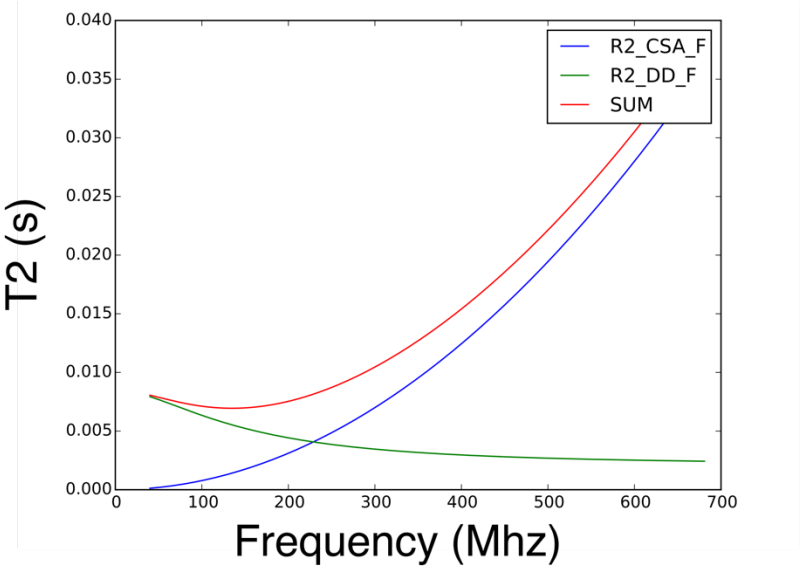


Fig 1.18 – Predicted effects of dipole-dipole and CSA interaction on the fluorine T_2 at increasing ^{19}F frequencies for a correlation times of 0.5 ns.

-cifically designed by Kurt Wüthrich and his team to study large complexes by playing on the constructive use of interference between DD coupling and CSA, leading to a linewidth reduction (Pervushin-1997). Regarding the Nuclear Overhauser Effect (NOE) of fluorine, the same correlation with molecular size as for the proton apply: positive NOE for small molecules with rapid motion and short correlation times; negative NOE in the opposite situation (Brey -1996). This means that there is a zero for the fluorine-proton NOE signal. The motion dependence of NOE is very similar to the one observed for proton due to the similar magnetogyric ratio. In particular, small molecules are characterized by a positive NOE

while large molecules display large negative NOE. The $^{19}F\{^{1}H\}$NOE can be plotted in regard to the correlation time:

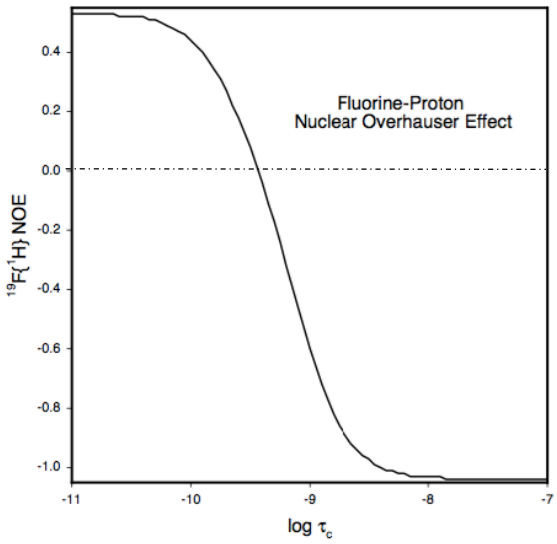


Fig 1.19 – Predicted fluorine proton NOE at 470MHz (Gerig-2001).

The over-broadening of resonances coming alongside the rise in magnet field strength led to fluorine NMR being momentarily deserted for protein studies. ^{19}F NMR was truly revitalized by ligand screening studies since the difference in $R_{2,(CSA)}$ between free and bound states is amplified and thus, even with low bound proportion, the observed transversal relaxation is strongly impacted. The following parts will show how fluorine NMR parameters can be used to monitor macromolecule-ligand interactions.

1.2.3 Fluorine NMR and molecular interaction

1.2.3.1 Thermodynamic and kinetic models of molecular interaction

The first step prior to a biological response is nearly always a binding event (although different triggers exist, like for the photo-receptors). The binding and the later release of a ligand can even be the sole function of a macromolecule, such as myoglobin. Those interactions between molecules in living organisms form complex systems that depend on spatial (cellular compartmentalization) and temporal (kinetic rates) parameters. Molecular interactions are usually represented by simple association/dissociation models that involves on- and off-rates. These simplified models can reproduce accurately most of the measurements made on mixtures of interacting molecules, provided that the system is at equilibrium. For example, the competitive binding of two ligands (a ligand *L* and an inhibitor *I*) for a single binding site on a receptor (an enzyme *E*) can be represented by the following sketch (figure 1.20):

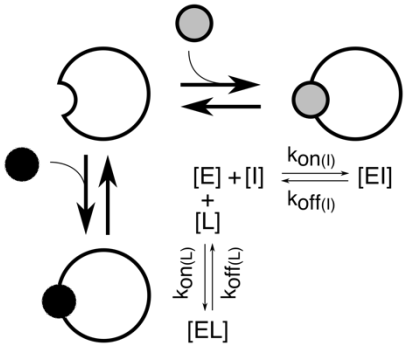


Fig 1.20. Model with a single receptor site for two molecules with different binding affinities.

This model underlies the analysis of competition experiments that are at the basis of the FAXS method proposed by Dalvit et al (Dalvit-2003). The simplest model of an interaction happening between a ligand and a receptor is described by the equation below:

$$E + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} EL \qquad \textit{[eq 1. 10]}$$

Where L stands for the free ligand concentration, E the free enzyme and EL the concentration of the complex formed. k_{on} (in $mol^{-1}.s^{-1}$) and k_{off} (in s^{-1}) are kinetic parameters related to the K_d (mol^{-1}) through:

$$K_d = \frac{k_{off}}{k_{on}} \quad \textit{[eq 1. 11]}$$

The equilibrium dissociation constant K_d correspond to the ligand concentration for which 50% of the binding sites are occupied (once the reaction equilibrium is reached). The K_d also reflects the affinity of a ligand for its target. This affinity is approximated by the ratio between the ligand kinetic rates for association (k_{on}) and dissociation (k_{off}). Indeed, at the equilibrium, the ligand binding rate to the receptor is equal to its dissociation rate:

$$[E] * [L] * k_{on} = [EL] * k_{off} \quad \textit{[eq 1. 12]}$$

And thus, the relation between the concentration of the molecules and the kinetic parameters becomes:

$$\frac{[E] * [L]}{[EL]} = \frac{k_{off}}{k_{on}} = K_d \quad \textit{[eq 1. 13]}$$

A low K_d value (in the µM range or below) corresponds to a strong binding: the EL proportion will be high at any given E and L concentration. On the opposite, if the complex concentration EL is low for the same values of E and L, the binding is said of weak affinity, the K_d value is elevated (superior to a hundred of µM). The association rate k_{on} is limited by the diffusion of the smallest molecule (usually the ligand). This diffusion rate is in the range of 10^{-8} $M^{-1}.s^{-1}$ but can effectively vary over several orders of magnitude. The dissociation rate can be very small, a situation that corresponds to an infinite lifetime of the bound state ($\tau=1/k_{off}$). Depending on the system, NMR may provide information both on the kinetic parameters (k_{on}/k_{off}) and the concentration of the species in solution. The later corresponds to the relative intensities of the different signals while the former impacts their line widths, and more generally their relaxation parameters. The competitive binding can be solved by two different approaches, either by integration of the kinetic equations, or by solving the equilibrium equations.

1.2.3.2 Solving the equilibrium equations

The competition experiment depicted in figure 1.20 leads to the following species in solution:

$$\begin{cases} [E_0] = [E] + [EL] + [EI] \\ [L_0] = [L] + [EL] \\ [I_0] = [I] + [EI] \end{cases} \quad \textit{[eq 1. 14]}$$

with dissociation equilibrium constants defined as:

$$\begin{cases} K_D = \dfrac{[E][L]}{[EL]} \\ K_I = \dfrac{[E][I]}{[EI]} \end{cases} \quad \textit{[eq 1. 15]}$$

Using equations 1.14 and 1.15 leads to the concentrations of the two types of complexes [EI] and [EL]:

$$K_I = \frac{([E_0] - [EL] - [EI])([I_0] - [EI])}{[EI]} \quad \textit{[eq 1. 16]}$$

$$K_I.[EI] = [E_0].[I_0] - [EI].[E_0] - [EL].[I_0] + [EL].[EI] - [EI].[I_0] + [EI]^2 \quad \textit{[eq 1. 17]}$$

$$K_D = \frac{([E_0] - [EL] - [EI])([L_0] - [EL])}{[EL]} \quad \textit{[eq 1. 18]}$$

$$K_D.[EL] = [E_0].[L_0] - [EL].[E_0] - [EL].[L_0] + [EL].[EI] - [EI].[L_0] + [EL]^2 \quad \textit{[eq 1. 19]}$$

$$\begin{cases} [EI]^2 - [EI].([I_0] + K_I + [E_0] - [EL]) - [EL].[I_0] + [E_0].[I_0] = 0 \\ [EL]^2 - [EL].([L_0] + K_D + [E_0] - [EI]) - [EI].[L_0] + [E_0].[L_0] = 0 \end{cases} \quad \textit{[eq 1. 20]}$$

Solving the set of coupled second-order polynomial equations is a complex task requiring some knowledge in conic functions. The solution may be found in several textbooks or publications (Roehrl, 2004). Roerhl and al. wrote an analytical treatment of competitive binding, for fluorescence binding assays. In a ^{19}F NMR paper published by Dalvit in 2005, a simulation illustrates the fraction of bound “spy” molecule as a function of the concentration of a competing molecule for different values of its binding constant. The simulation took into account only a specific interaction between the spy and its competitor. The equation used is partially extracted from the Roerhl article where the solution corresponds precisely to a model of complete competitive binding. The equation is presented below:

$$\frac{[\mathrm{EL}]}{[\mathrm{L_{TOT}}]} = \frac{2\sqrt{(a^2 - 3b)}\cos(\theta/3) - a}{3K_\mathrm{D} + 2\sqrt{(a^2 - 3b)}\cos(\theta/3) - a} \qquad \textit{[eq 1. 21]}$$

where

$$\theta = ar\cos\left[\frac{-2a^3 + 9ab - 27c}{2\sqrt{(a^2 - 3b)^3}}\right]$$

$$a = K_\mathrm{D} + K_\mathrm{I} + [\mathrm{L_{TOT}}] + [\mathrm{I}] - [\mathrm{E_{TOT}}] \qquad \textit{[eq 1. 22]}$$

$$b = \{[\mathrm{I}] - [\mathrm{E_{TOT}}]\}K_\mathrm{D} + \{[\mathrm{L_{TOT}}] - [\mathrm{E_{TOT}}]\}K_\mathrm{I} + K_\mathrm{D}K_\mathrm{I}$$

$$c = -K_\mathrm{D}K_\mathrm{I}[\mathrm{E_{TOT}}]$$

1.2.3.3 Kinetic treatment of the equilibrium

A second approach to find the equilibrium concentration of the different species in a competition experiment consist in integrating the set of differential equations that describes the flow of molecules, assuming a value for the diffusion limited k_{on}^{L} and k_{on}^{I}. The variations of concentrations of [EI] and [EL] in function of time can be expressed as:

$$\begin{cases} \dfrac{d[EL]}{dt} = [E] * [L] * k_{on_L} - [EL] * k_{off_L} \\ \dfrac{d[EI]}{dt} = [E] * [I] * k_{on_I} - [EI] * k_{off_I} \end{cases} \qquad \textit{[eq 1. 23]}$$

To obtain a numerical solution of this differential equation system, the programming language Python can be used, more precisely the “odeint” function of the sub-library “scipy.integrate”. The “odeint” function return a numerical estimate of the solution to the so-called Cauchy problem. A solution to the first order differential equation $Y'(t) = f(Y_{(t)}, t)$ is looked for, while respecting the initial condition $Y(t_0) = Y_0$.

The time dependent evolution of [EI], [EL] is returned by the program, and equilibrium concentrations are obtained when the variation of species concentrations between two consecutive time points is negligible. The python script that we wrote and used for our simulations is available at the following address: https://github.com/rrecht/NMR_Binding_R2_simulation.git. Figure 1.21 presents side-by-side the result from solving equation 1.24 and equation 1.21. The top graphs describe the

impact of different competitor affinities (named K_c) on the bound ligand proportion while the bottom graphs allow to evaluate the bound ligand proportion in function of the ligand affinity K_d, at different ligand total concentrations.

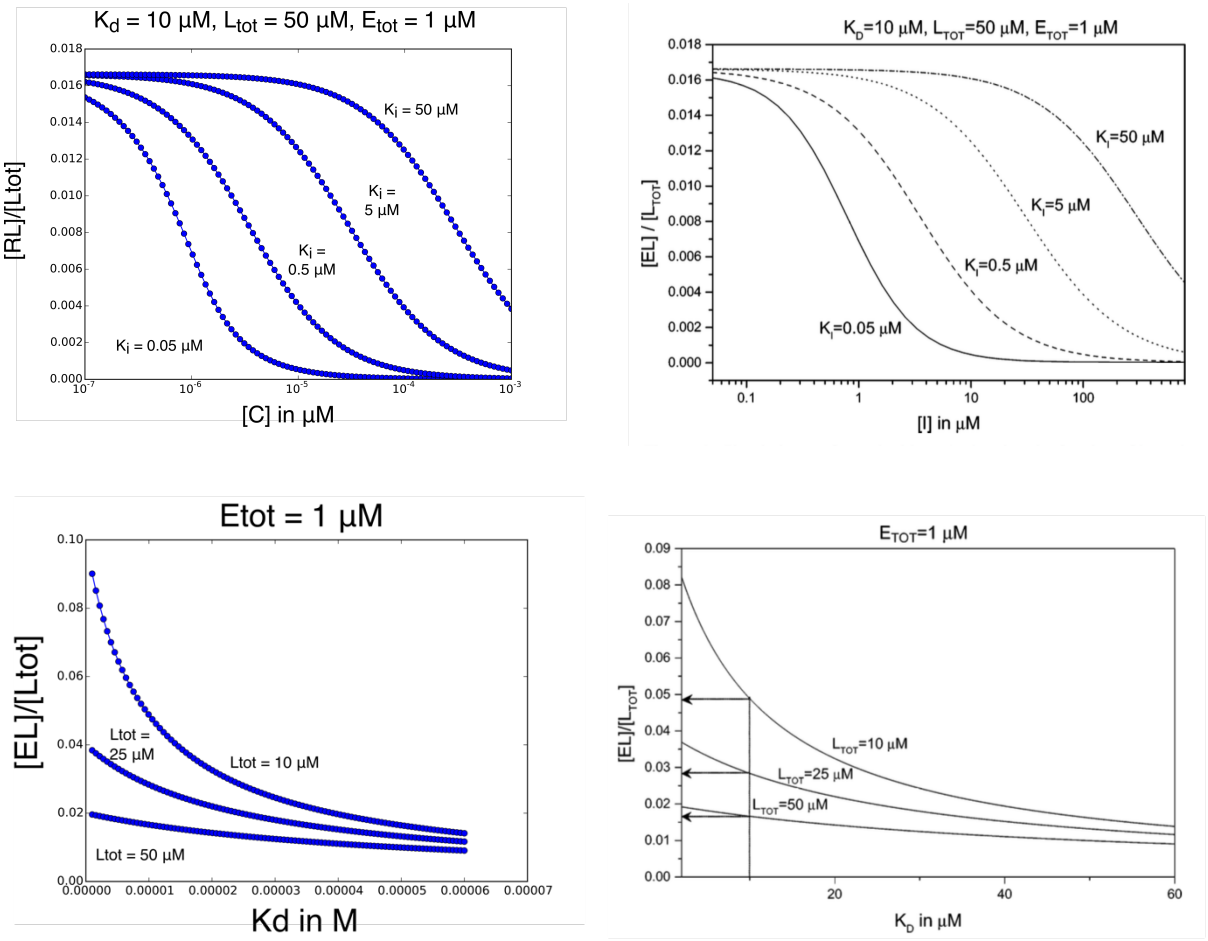


Fig 1.21. Comparison between the simulation obtained with the resolution of the differential equations 1.24 (left graph) and plotted using equation 1.22 (right graph).

The comparison of the results obtained from solving equilibrium equations or kinetic equations indicates that both methods lead to the same results as expected. However, solving kinetic equations is more appropriate and informative to interpret NMR data since relaxation parameters are sensitive to these kinetics parameters. Thus, making hypothesis on k_{on} and k_{off} values to simulate complex interactions experiments may lead to a better and more complete understanding of the NMR observations.

1.2.3.4 Impact of molecular interaction on NMR parameters

When a molecule exchange between a free and a receptor-bound state, the obtained NMR spectrum depends on the kinetic of the exchange, as shown in figure 1.23.

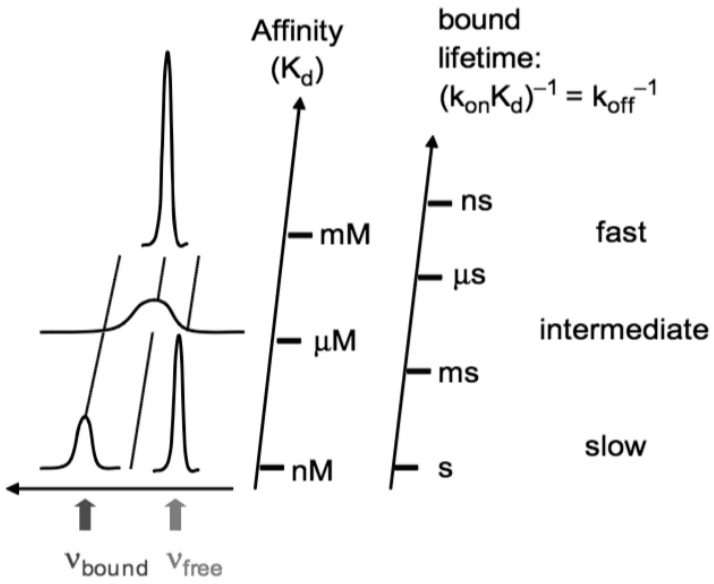


Fig 1.23- *NMR exchange time scales, from a fast to a slow probing of the binding site. The time scales are related to the NMR observation time scale, meaning that the binding event is either happening faster or slower than the sampling of the NMR spectrum. For a fast exchange, this result in an averaging of the frequencies and thus a single peak observable in the NMR spectrum (reproduced from (Kieffer- 2011)).*

For low affinity ligands (in the mM range), the fast off-rate leads to an averaging of chemical shifts between the free and bound states values, while high affinity ligands (in the nM range) allow the observation of the frequencies that are specific to the free and bound states. If the exchange rate between the bound and free states matches the frequency difference between these states, a broadening of the resonance lines is observed, usually described as exchange broadening. This broadening represents an additional contribution to the specific relaxation rates of the free and bound states, $R_{2,ex}$:

$$R_{2,obs} = (1-\alpha)R_{2,free} + \alpha R_{2,bound} + R_{2,ex} \qquad \textit{[eq 1. 23]}$$

In the equation, α is the fraction of bound ligand. In the fast-to-intermediate time scale, the exchange contribution to the observed transversal relaxation can be written as:

$$R_{2,ex} = (1-\alpha)^2 \alpha 4\pi^2 (\delta^{free} - \delta^{bound})^2 \tau_{bound} \qquad \textit{[eq 1. 24]}$$

In the equation 1.25, τ_{bound} is the residency time of the ligand in the binding site. The residency time is linked to the off rate constant k_{off} by $\tau_{bound} = 1/k_{off}$. It is worth noting that equation 1.25 is a simplification of the general equation for small values of τ_{bound}. Since for the fluorine nuclei, the chemical shift anisotropy (CSA) is large, the squared term $\delta^{free} - \delta^{bound}$ amplifies the binding impact on the observed R_2. The FAXS (Fluorine Anisotropy eXchange Screening) (Dalvit -2003) methodology exploits this property by recording a ^{19}F spectrum using a spin-echo sequence. The length of the spin echo is optimized to maximize the peak intensity difference between the free and bound states. Ideally, in the bound state, the resonance must be faint but still visible. If another molecule is competing with the ligand (so α will be lowered) the observed ^{19}F transversal relaxation speed is diminished and so the ligand peak gets thinner again. The method is resumed in the figure 1.24.

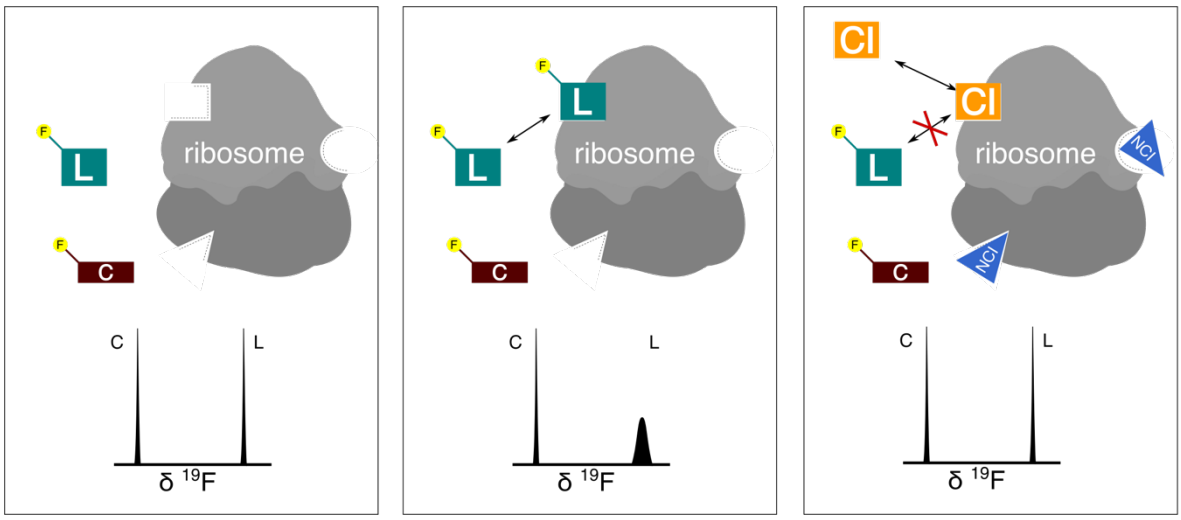


Fig 1.24 - *The binding event of a fluorinated ligand (L) to the ribosome result in a broadening of its ^{19}F NMR signal (in a situation of fast to intermediate exchange) whereas an internal control (C) is unaffected. Several sites are possible on the macromolecule, but the addition of a competitive inhibitor (CI) lead to the release of the ligand, and thus the narrowing of its peak. Non-competitive inhibitor (NCI) can be added in the solution in excess as well, to reduce the possibility of non-specific binding of the ligand.*

For our simulations, $R_{2,obs}$ is the desired output since it will allow to describe the broadening of the ligand resonance. The bound proportion is obtained from the result of the script described above. The τ_{bound} can be approximated by knowing beforehand the K_d value of the ligand (or just an estimation) and using an estimate of the diffusion value for k_{on} (10^{-8} $M^{-1}s^{-1}$). The R_2^{free} and the δ^{free} are simple to obtain experimentally for a ligand with a single ^{19}F NMR signal. Finally, the R_2^{bound} and the δ^{bound} are more difficult to estimate, especially

in the case of a macromolecule where those bound properties might be very far away from those corresponding to the free state of the ligand. The 70S ribosome rotational correlation time was estimated at 2.5 µs at 298 K (Lavallette – 1977). Finally, for the value of δ^{bound} , it depends on several parameters. First, The CSA's of ^{19}F nuclei in aliphatic positions in amino acids are smaller than those of ^{19}F nuclei in aromatic rings (Grage-2008). A ^{19}F NMR publication (Miura – 1988) has shown a 1500 Hz shift between the free and bound resonance of fluorobenzoate in interaction with D-amino acid oxidase. With values within those scales, the squared difference between the chemical shift of the bound and free form in ^{19}F greatly enhance the R_{ex} contribution to the observed relaxation. A simulation of R_2 values with those parameters is shown in figure 1.26, for a ligand of weak affinity (100 µM) in presence of 0.7 and 1.4 µM of ribosome, at different ligand concentrations. The hypothesis was made that only one site was available. In presence of ribosome, the transversal relaxation observed should be roughly a hundred times higher than for the free state for micromolar ligand concentrations. Even with a ligand in the millimolar range of concentration, the relaxation rate in presence of ribosome is ten times faster.

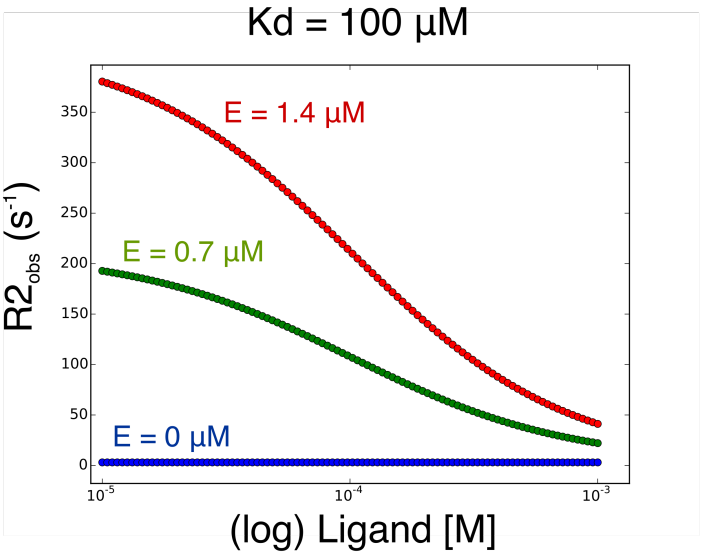


Fig 1.26- Simulation of the $R_{2,obs}$ values of a ligand with a 100 µM K_d in presence or absence of ribosome. The values are computed for ligand concentrations between 10 µM and 1 mM.

In a complementary manner, the script was used to assess the theoretical $R_{2,obs}$ values obtained with ligands whom the affinity is ranked from 1 to 200 µM (figure 1.27). This simulation shows that a reduction in the ligand concentration toward the µM range allows for a better differentiation between the possible K_d values for the ligand.

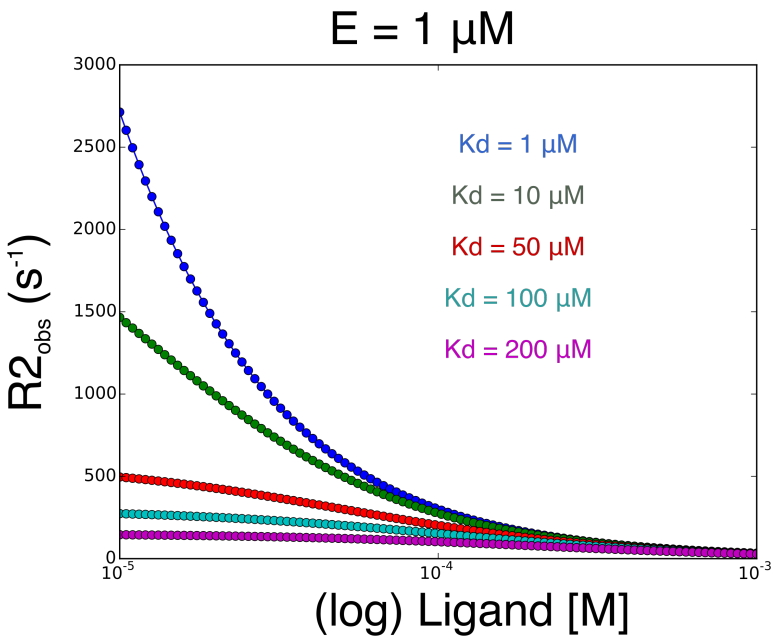


Fig 1.27-. Simulation of the $R_{2,obs}$ values of a ligand with different affinities in presence of 1µM ribosome. The values are computed for ligand concentrations between 10 µM and 1 mM.

1.3 ^{19}F NMR Practical considerations

Describing binding events using ^{19}F NMR has many advantages, such as the readability of the spectra. For instance, the water suppression that complicates the acquisition of proton spectra is avoided in fluorine NMR. Nevertheless, to make the best the fluorine qualities requires the optimization of some parameters.

1.3.1 Fluorine probes

The quality of the probe constitutes the primary instrumental factor to optimize the sensitivity of the NMR experiment. The FAXS methodology is set up to work with ligands concentration between 10^1 to 10^2 µM. The concentration in ligands must be low enough to avoid solubility problems and yet high enough to see its resonance peak. Different probes can be used for the detection of fluorine. The Bruker BBFO (double resonance broad band) probe was used for the main part of this thesis. The inner coil of this versatile probe is tunable over a frequency range that enables observation of nuclei between ^{31}P and ^{15}N, ^{19}F included. The outer coil is optimized for ^{1}H decoupling and observation. However, in the absence of independent high-frequency channels, a X-QNP switchbox has to be used for dual resonance (^{1}H/^{19}F) experiments. This switchbox allows to switch the transmitter output between ^{1}H and ^{19}F channels, within typically 20 µs. Band-selective ^{1}H and ^{19}F-preamplifiers are also needed, as an amplifier, in our case a BLAXH 300/50 (featuring three rf-channels, two for ^{1}H/^{19}F, one for X frequencies). Commonly achievable π/2-pulse widths using a BBFO probe and a BLAXH 300/50 amplifier operating at 600/564 MHz ^{1}H/^{19}F resonance frequencies are 9 µs (^{1}H, ca. 10–30 W) and 20 µs (^{19}F, ca. 30 W). On a side note, fluorocarbon polymers are often found in the design of proton-observed NMR probes, and these can give rise to large, broad, baseline distorting signals in a ^{19}F NMR spectrum (figure 1.28).

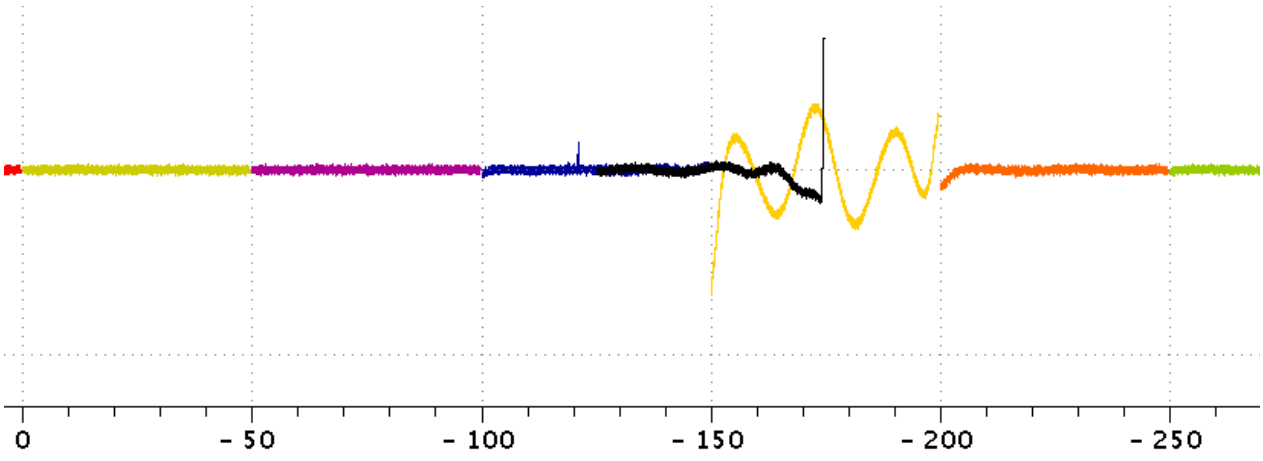


Fig 1.28 – Probe parasitic signal on a ^{19}F NMR spectrum, between -150 and -200 ppm.

A way to improve the sensitivity would be the use of cryogenic probe optimized for ^{19}F detection. The 5-mm QCI-F "inverse" cryoprobe developed by Bruker is equipped with cryogenically cooled preamplifiers for multiple nuclei (^{19}F, ^{13}C, ^{1}H, and ^{2}H) and a ^{15}N channel. Simultaneous proton decoupling is possible on this probe. The sensitivity is said to be increased by at least a factor of 3 compared to hot probes as shown by figure 1.29. Using hot ^{19}F NMR probes such as the Bruker BBFO, spectra can be recorded with a satisfying signal-to-noise ratio in one hour and a half at 50 µM while using a cryogenically cooled probe, like a Bruker Q-CIF, it can be lowered to 15 minutes and only 12 µM of ligand (figure 1.29).

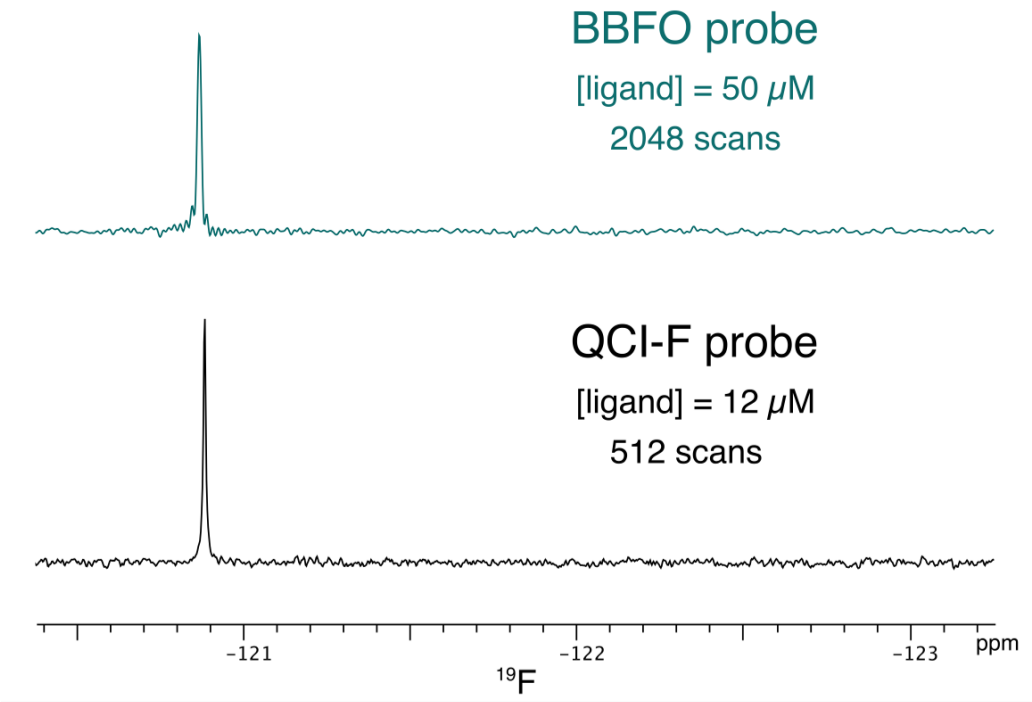


Fig 1.29 – Superposition of two spectra of the same solution, recorded with two different ^{19}F NMR probes. Both experiments were set-up at 298 K, with a recycle delay D1 of 1 second, pulses of 14.25 µs (BBFO) and 13.5 µs (QCI-F). The resulting signal-to-noise ratio were 33.8 (BBFO) and 48.6 (Q-CIF).

1.3.2 NMR tubes

The appropriate choice of NMR tubes represents another critical parameter in order to design interaction experiments. Several choices are possible between 5 mm and lower diameter tubes (3 mm or capillary tubes). For screening experiments using protein samples, the 5 mm tubes are not considered since they require a large sample volume, around 500 µl. Shigemi tubes allow a reduction of the sample volume, while preserving the active volume (that is the sample volume that contributes to the NMR signal). The glass magnetic susceptibility is matched to the water one, a feature that is of little interest for fluorine NMR. However, we found that optimal compromise between the mass of ribosome and the experimental time was provided by the use of Shigemi tubes. The spectrum recorded with the 3 mm tube had a signal-to-noise ratio of 17.1 while the shigemi was at 32.5, a 90% improvement for only 50% of supplementary material.

1.3.3 Samples concentration evaluation

NMR provides an efficient and accurate way to measure precisely the concentration of the different species in solution. The area under a peak in an NMR spectrum is proportional to the number of nuclei in a given chemical environment in a molecule. The concentration of a molecule can be evaluated through the comparison of the signal integrals, provided the assignments of the molecule resonances are known. In order to measure accurately the concentrations of the ligands in a mixture, an internal standard was used (Koehler – 2014). In this method, a small amount of ligand stock solution is mixed with a solution of tryptophan whose concentration is known with high accuracy by UV absorption. The mixture is usually prepared in D_2O to minimize water suppression artifacts and enable the use of presaturation methods for solvent suppression. An example of the concentration measurement of 3-fluorophenol is provided in figure 1.30.

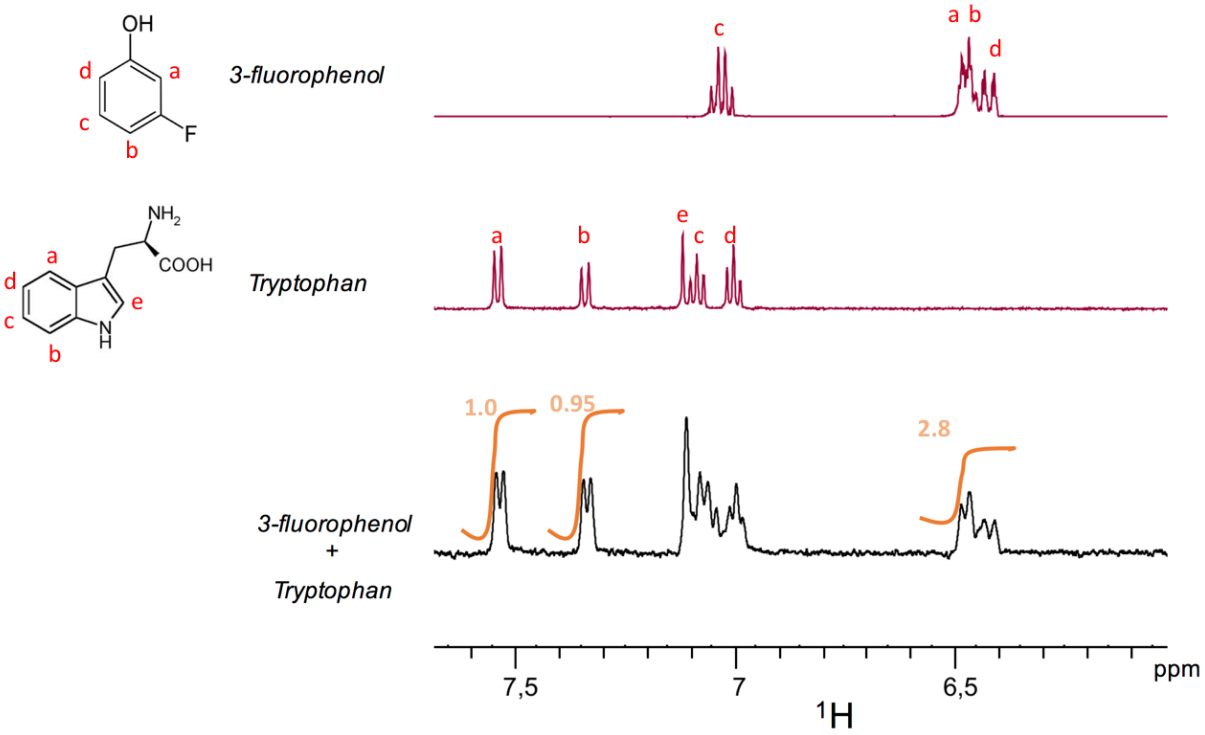


Fig 1.30 – If the 3-fluorophenol is used as a control in a titration experiment, its concentration can be controlled with the integral ratio obtained with peaks coming from a solution of tryptophan. In fluorine, the integral ratio is easier to obtain than in proton due to the absence of overlap. The two spectra on the top were obtained on mM concentrations of individual samples while the bottom one had 3-fluorophenol and tryptophan mixed at concentrations around 50 µM.

In figure 1.30, the ratio of the 3-fluorophenol and tryptophan aromatic resonances can be compared. The peaks around 6.45 ppm account for three protons of 3-fluorophenol while

those at 7.55 and 7.35 ppm represent each one proton from tryptophan. The comparison of the integrals shows that the 3-fluorophenol real concentration is a little bit lower than what was expected (the stock concentration should be adjusted with a 0.93 coefficient). By taking in account the relaxation issues, an accuracy of 5% can be achieved. If a number of sources of errors are properly handled (saturation, digital resolution…), a precision of 1% is possible (Reich – 2016). The delay between pulses must be long enough to ensure an appropriate longitudinal relaxation. A delay five time longer than the T_1 value will allow more than 99% of the magnetization to come back at the equilibrium. The concentration of fluorinated products can be evaluated in ^{19}F NMR by adding trifluoro-ethanol to the solution (the simple proton spectrum of TFE allows for an easy quantification with tryptophan). While this method adds an experimental step, the ^{19}F NMR properties can make it worth it, especially in the case of overcrowded proton spectra.

1.3.4 Spin-Echo sequence optimal echo delay

The FAXS methodology allows to work with low receptor concentrations, below the micro-molar. This is a critical point as some proteins can be expensive, difficult to product or even insoluble at higher concentrations. The R_2 rates of the free ligand and while in presence of the target can be obtained through the recording of spin-echo experiments with increased relaxation delays (sequence type presented in figure 1.31).

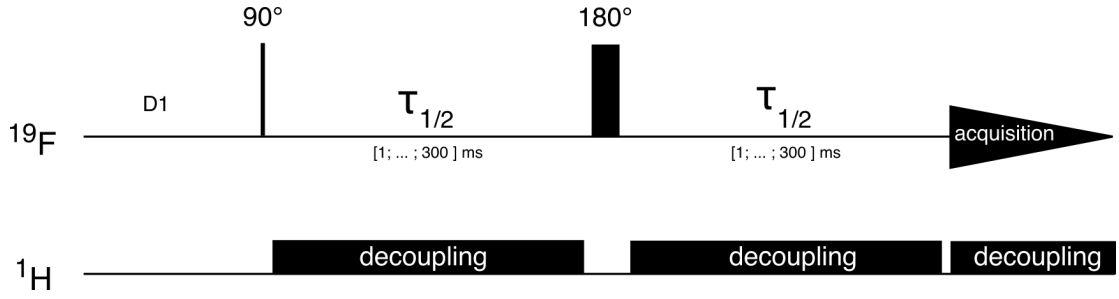


Fig 1.31- Spin-echo sequence used to determine the R_2 values of the compounds. Proton decoupling was applied for the duration of the sequence.

The optimal half-echo time $\tau_{1/2}$ for the FAXS experiment maximizes the difference between the intensities of the signal of the free ligand and the signal measured in the presence of the receptor ($R_{2,obs}$) as well as the intensity of the measurement. For linezolid/ribosome measurements, the relaxation curves are shown in figure 1.32. The corresponding curves equations are:

$$f(t) = e^{-R_{2,free}*t} \quad \textit{[eq 1. 25]}$$

$$g(t) = e^{-R_{2,obs}*t} \quad \textit{[eq 1. 26]}$$

$$h(t) = f(t) - g(t) = e^{-R_{2,free}*t} - e^{-R_{2,obs}*t} \quad \textit{[eq 1. 27]}$$

$$i(t) = f(t) * h(t) = e^{-R_{2,free}*t} * (e^{-R_{2,free}*t} - e^{-R_{2,obs}*t}) \quad \textit{[eq 1. 28]}$$

The optimal time $t = \tau_{1/2}$ is found for:

$$\frac{d(i(t))}{dt} = 0 \quad \textit{[eq 1. 29]}$$

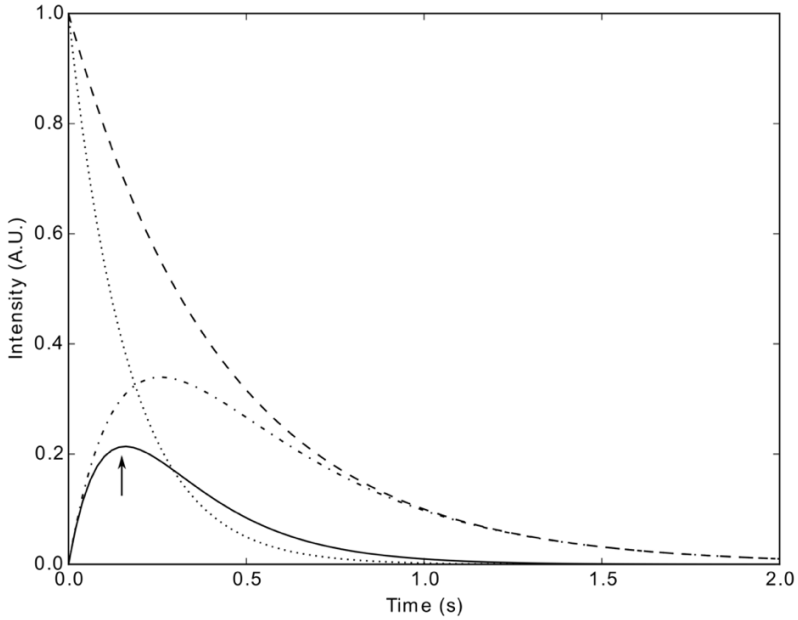


Fig 1.32- Superposition of several relaxation curves allowing to determinate the optimal half-echo time maximizing the differences between the ligand states. The transversal relaxation decay curves follow the equation $M_z = M_0 * e^{-R_2*t}$ with R_2=2.3 s^{-1} for the free ligand (drawn as "---") and R_2=6 s^{-1} in the presence of 0.5 µM of ribosome ("⋯"). The difference between the two curves is plotted ("-·-") and finally, the free curve weighted by the difference is drawn as a solid line. The optimal value, $\tau_{1/2}$=0.15 s, is indicated in the spectrum via an arrow.

1.3.5 Uniform excitation of the wide ^{19}F chemical shift range

Whether to estimate the concentrations or to monitor shim perturbations, a ligand ^{19}F signal must be compared with the ^{19}F signal of an internal control molecule. This control molecule must not bind the ligand's target. The control must be soluble in the protein's buffer to avoid aggregation. The control ^{19}F intensity should ideally stay unmodified during the ligand interaction with the target. As far as possible, the fluorine chemical shift of the control molecule should be similar to the one from the tested ligand. If the chemical shift of the

ligand and the control molecule are too far away to be equally excited, the experiment can be recorded twice with the transmitter frequency each time centered on the resonance of one of the molecules.

A more efficient method consists in using shaped pulses adapted to provide large bandwith for excitation and refocussing pulses. A null in the excitation profile for a pulse with on-resonance angle α = 90° is achieved at an offset frequency from the resonance equal to:

$$\Omega = \pm\sqrt{15} * 2\pi * \omega_1 \quad where \quad \omega_1 = \gamma B_1 \quad and \quad B_1 = \frac{1}{4\tau_{90}} \qquad [eq\ 1.\ 30]$$

Where τ_{90} is the length of the 90° pulse. The chemical shift difference between a CFH_2 fluorine atom (for example, on the florfenicol molecule) and a trifluoromethyl CF_3 (as it can be found in 2,2,2-trifluoroethanol) can be a 150 ppm gap. On a 14.1 Tesla spectrometer, that is a bit less than 90 kHz separation. This situation is more precisely detailed in figure 1.33:

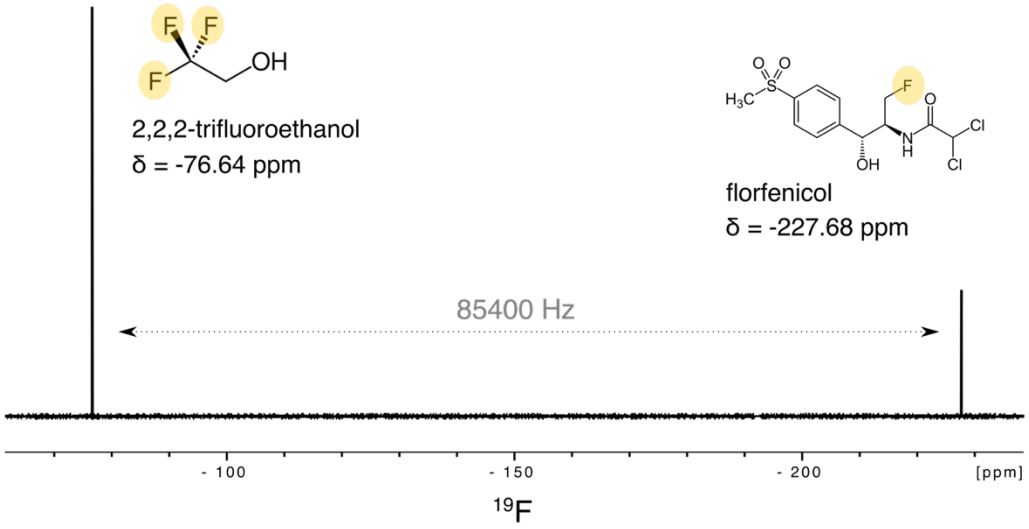


Fig 1.33- Illustration of the spectral width that can be observed between a screened antibiotic (florfenicol) and a control molecule (2,2,2-trifluoroethanol). Both species are separated by more than 150 ppm.

In order to record a spin-echo experiment with quantitative results, both 90° and 180° RF pulses must affect the resonances with the same amplitude (the peak phase can be adjusted). In FT-NMR, the pulse width and the power level determines the bandwidth of frequencies that are excited. Shorter pulses widths will result in broader frequency bandwidth. So for short-duration RF pulses (around 10 µs), typically a wide enough portion of the spectrum is excited simultaneously. In order to verify this assumption in the case of a fluorine spectrum and obtain precisely the excitation profile, an 11.1 µs excitation pulse with a power of 12.11 Watts was swept around the peak resonance of florfenicol (from 200 to -

592 ppm with an 8 ppm step) and its intensity monitored. A 100 µs, 95 mW GARP decoupling pulse scheme was used on the proton channel to keep the florfenicol resonance as sharp as possible. The intensities of the peak after the carrier sweep were collected by an in-house Python script and are observable in figure 1.34:

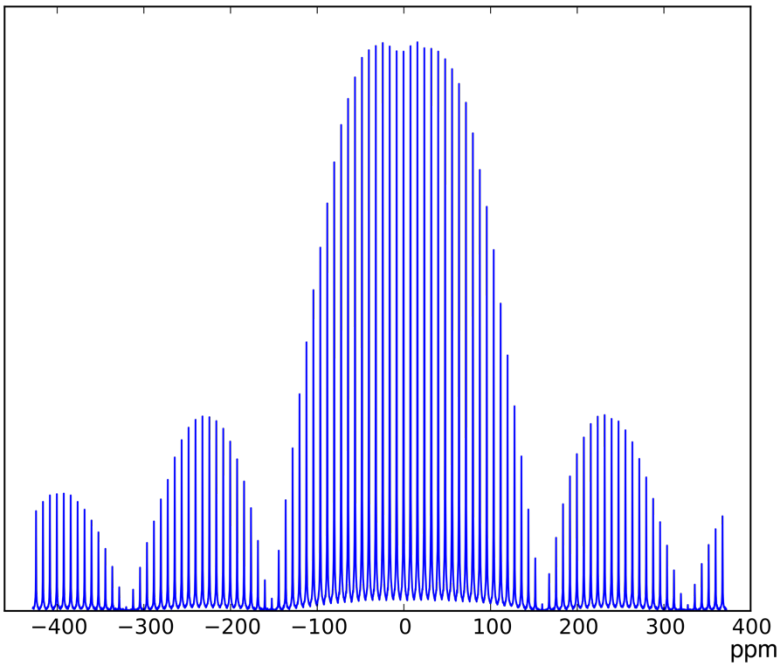


Fig 1.34- Excitation profile of a 90° excitation pulse (11.1 µs with 12.1 W), its frequency swept on an 800 ppm radius around the florfenicol peak resonance.

There is a nearly flat plateau at the top of the excitation profile, as large as 100 ppm, enough for a uniform excitation of the peaks of 2,2,2-trifluoroethanol and florfenicol, granted that the carrier frequency is set in the middle of the two resonances. Now, with the same power, a 180° inversion pulse twice as long as the 90° (so 22.2 µs at 12.11 W) will cover a shorter range of frequencies. To ensure that the signal at the end of the echo originates from a refocussing signal, two z-gradients of the same sign were added (of SINE.100 shape, for 1000 µs and with a 20% Z power), on each side of the 180° pulse. The spectral coverage of this spin echo sequence with gradients was tested in a similar manner as previously, around the florfenicol resonance, and the collected peak intensities are illustrated in figure 1.35:

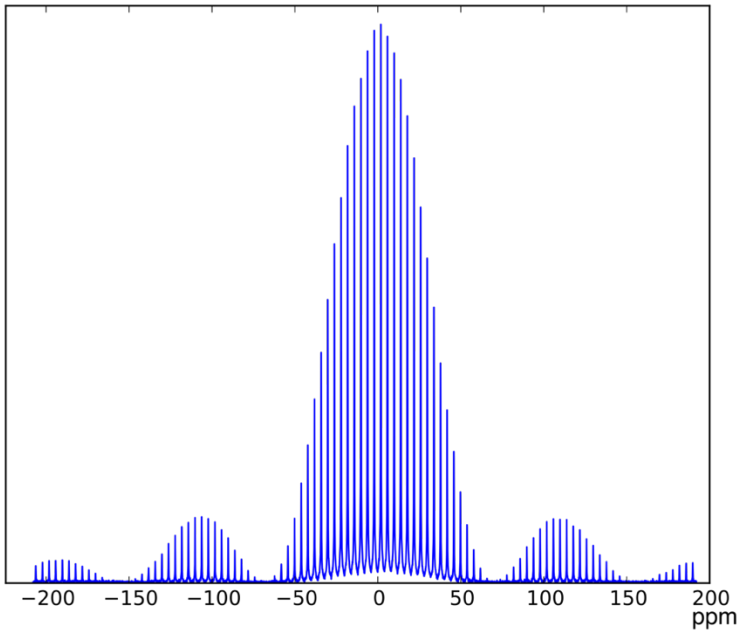


Fig 1.35- Excitation profile of a spin-echo sequence with a 11.1 µs pulse for the 90° excitation and a 22.2 µs pulse for the non-selective 180° inversion (both at 12.1 W), the frequency swept on a 400 ppm radius around the florfenicol peak resonance.

Figure 1.35 shows that the use of a standard spin echo experiment results in a very poor spectral coverage, due essentially to the poor property of rectangular refocussing pulse. In order to circumvent this drawback, we used low power frequency swept pulses, known as chirp pulses. Since the FT is reversible, to obtain a squarely excited spectral region a *sinc* shaped pulse can be used. Many pulsed shapes can be used in FT-NMR, a frequently used one being the Gaussian shaped pulse since the FT of a Gaussian is a Gaussian. Adiabatic pulses can also be chosen, a special type of shaped pulse where either a frequency or phase sweep happens during the pulse duration. The slow passage of a chirped pulse through the resonances invert sequentially the spins. In a square pulse, all spins are inverted simultaneously. Adiabatic shaped pulses are especially suited for 180° rotations since they provide accurate flip angles over large range of RF power levels and thus, allow a minimization of decoupling power and sample heating. Using a standard Bruker 180 adiabatic inversion chirped pulse (Crp60,0.5,20.1 of 500µs) on the florfenicol and 2,2,2-trifluorethanol sample now grants a uniform profile, as seen in figure 1.36:

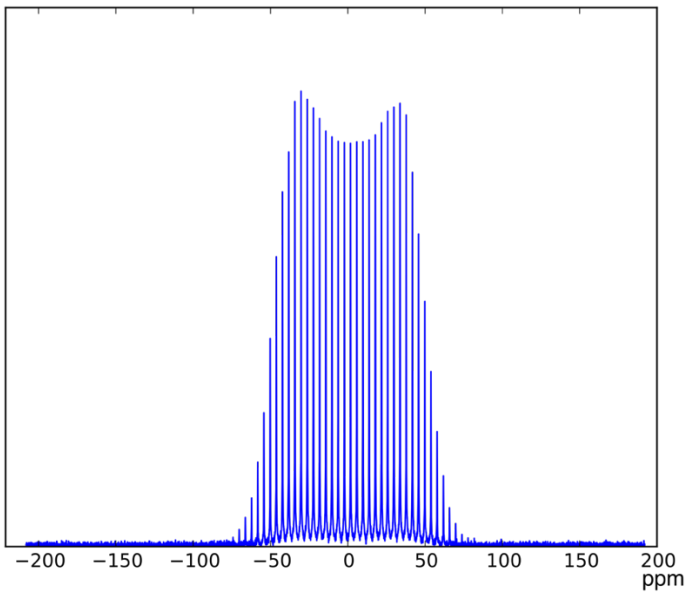


Fig 1.35- Excitation profile of a spin-echo sequence with a 11.1 µs pulse for the 90° excitation and an adiabatic shaped pulse for the 180° inversion, the frequency swept on a 400 ppm radius around the florfenicol peak resonance.

There are other methods to circumvent resonance offset effects, since dealing with wide chemical shift ranges is common in fluorine NMR, the most recent one being the CHORUS pulse sequence (Power-2016).

Conclusion

Screening through ^{19}F NMR present several advantages. Due to the fluorine CSA properties, the binding event can be detected at low protein concentrations. The pulse sequence is simple at its core and easy to implement. The absence of overlapping in the spectra allows to monitor the concentration, solubility and stability of the fluorinated compounds (Dalvit – 2007). With the absence of fluorine signal in biological samples, in-cell ^{19}F NMR could be realized. Many of pharmaceutical compounds are already fluorinated but the experiments are not restricted to ^{19}F-labelled molecules since competitors for the ligand can be devoid of fluorine.

One drawback of focusing on fluorinated products is that the manufacturing process can be challenging. Its attractive properties for drug design are double-edged for the chemistry production. The synthesis of new fluorinated products is done only by fluorination chemistry experts. So before starting our screening project involving fluorine and antibiotics, we looked for fluorinated molecules already available. Luckily, among the ribosomal ligands, there exist fluorinated compounds belonging to different families. Prior to their interaction with the ribosomal unit, full characterizations of the molecules were realized. The methodology as well as the ligand properties obtained are resumed in the next chapter of this thesis.

References

Arntson KE, Pomerantz WC. Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery. J Med Chem. 2015

Brey, Wallace S. and Mary Louise Brey. "Fluorine-19 NMR". *eMagRes* (1996).

Carey F, Giuliano R. Organic Chemistry. McGraw-Hill Education; 2013.

"CDC - The Emergency Response Safety and Health Database: Systemic Agent: HYDROGEN FLUORIDE/ HYDROFLUORIC ACID - NIOSH". www.cdc.gov. Retrieved 2015-12-04.

Crowley PB, Kyne C, Monteith WB. Simple and inexpensive incorporation of 19F-tryptophan for protein NMR spectroscopy. Chem Commun (Camb). 2012;48(86):10681-3.

Dalvit C, Fagerness PE, Hadden DT, Sarver RW, Stockman BJ. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. J Am Chem Soc. 2003;125(25):7696-703.

Dalvit C. Ligand- and substrate-based 19F NMR screening: Principles and applications to drug discovery. Prog. Nuclear Magn. Reson. Spectrosc. 2007;51:243–271.

Danielson MA, Falke JJ. Use of 19F NMR to probe protein structure and conformational changes. Annu Rev Biophys Biomol Struct. 1996; 25:163-95.

Doi M, Nishi Y, Uchiyama S, et al. Characterization of collagen model peptides containing 4-fluoroproline; (4(S)-fluoroproline-pro-gly)10 forms a triple helix, but (4(R)-fluoroproline-pro-gly)10 does not. J Am Chem Soc. 2003;125(33):9922-3.

Dolbier W. Guide to Fluorine NMR for Organic Chemists. 2009.

Dong, Changjiang et al. "Crystal Structure And Mechanism Of A Bacterial Fluorinating Enzyme". *Nature* 427.6974 (2004): 561-565.

Filler, Robert and Rituparna Saha. Future Med. Chem. (2009) 1 (5), 777-791. Fluorine in medicinal chemistry: a century of progress and a 60-year retrospective of selected highlights.

"Freedonia industry study #1555 – Fluorochemicals" (PDF). Freedonia. Retrieved 23 June 2012.

Gerig J. Fluorine NMR. Monograph published online. Available from http://www.biophysics.org/portals/1/pdfs/education/gerig.pdf. 2001:1–35

Gerig J. Fluorine NMR of proteins. Progress in Nuclear Magnetic Resonance Spectroscopy. 1994;26:293-370.

Gillis EP, Eastman KJ, Hill MD, Donnelly DJ, Meanwell NA. Applications of Fluorine in Medicinal Chemistry. J Med Chem. 2015;58(21):8315-59.

Gribble, G. W. (2002). "Naturally Occurring Organofluorines". In Neison, A. H. Organofluorines. The Handbook of Environmental Chemistry 3N. Berlin: Springer. pp. 121–136. doi:10.1007/10721878_5. ISBN 3-540-42064-9.

Harper David B, David O'Hagan and Cormac B Murphy, Fluorinated Natural products: Occurrence and Biosynthesis, Handbook of Environmental Chem., 2003, 3(P), 141 –169

Hoang J, Prosser RS. Conformational selection and functional dynamics of calmodulin: a (19)F nuclear magnetic resonance study. Biochemistry. 2014;53(36):5727-36.

Jacobson, Orit, Dale O. Kiesewetter, and Xiaoyuan Chen. "Fluorine-18 Radiochemistry, Labeling Strategies And Synthetic Routes". *Bioconjugate Chem.* 26.1 (2015)

Kieffer B, S. Homans and W. Jahnke. .Nuclear Magnetic Resonance of Ligand Binding to Proteins. Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection, Measurement and Modelling, Chapter 2, 2011, 15-55

Kim HW, Perez JA, Ferguson SJ, Campbell ID. The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. Application to fluorotryptophan labelling to the H(+)-ATPase of Escherichia coli and NMR studies. FEBS Lett. 1990;272(1-2):34-6.

Kitevski-LeBlanc, J. (2010). Development and Application of 19F NMR of Proteins (Doctoral dissertation, University of Toronto).

Köhler C, Recht R, Quinternet M, De lamotte F, Delsuc MA, Kieffer B. Accurate protein-peptide titration experiments by nuclear magnetic resonance using low-volume samples. Methods Mol Biol. 2015;1286:279-96.

Lian C, Le H, Montez B, Patterson J, Harrell S, Laws D et al. Fluorine-19 Nuclear Magnetic Resonance Spectroscopic Study of Fluorophenylalanine- and Fluorotryptophan-Labeled Avian Egg White Lysozymes. Biochemistry. 1994;33(17):5238-5245.

Liang T, Neumann CN, Ritter T. Introduction of fluorine and fluorine-containing functional groups. Angew Chem Int Ed Engl. 2013;52(32):8214-64.

Merkel, L. and N. Budisa. "Organic Fluorine As A Polypeptide Building Element: In Vivo Expression Of Fluorinated Peptides, Proteins And Proteomes". *Organic & Biomolecular Chemistry* 10.36 (2012): 7241.

Mishra, Neeraj K. et al. "Fluorinated Aromatic Amino Acids Are Sensitive 19 F NMR Probes For Bromodomain-Ligand Interactions". *ACS Chem. Biol.* 9.12 (2014): 2755-2760.

Muller, Norbert and Duane T. Carr. "Carbon-13 Splittings In Fluorine Nuclear Magnetic Resonance". *The Journal of Physical Chemistry* 67.1 (1963): 112-115.

Müller K, Faeh C, Diederich F. Fluorine in pharmaceuticals: looking beyond intuition. Science. 2007;317(5846):1881-6.

Murphy, Cormac D., Christoph Schaffrath, and David O'Hagan. "Fluorinated Natural Products: The Biosynthesis Of Fluoroacetate And 4-Fluorothreonine In Streptomyces Cattleya". *Chemosphere* 52.2 (2003): 455-461.

Okazoe, Takashi (2009). "Overview on the history of organofluorine chemistry from the viewpoint of material industry" (PDF). Proceedings of the Japan Academy, Series B 85 (8): 276–289.

Pervushin K., R. Riek, G. Wider, and K. Wüthrich (1997). "Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution". Proc. Natl. Acad. Sci. USA 94 12366-71

Power JE, Foroozandeh M, Adams RW, et al. Increasing the quantitative bandwidth of NMR measurements. Chem Commun (Camb). 2016;52(14):2916-9.

Qiu, Xiao-Long, Wei-Dong Meng, and Feng-Ling Qing. "Synthesis Of Fluorinated Amino Acids". *Tetrahedron* 60.32 (2004): 6711-6745.

Qiu, Xiao-Long and Feng-Ling Qing. "Recent Advances In The Synthesis Of Fluorinated Amino Acids". *European Journal of Organic Chemistry* 2011.18 (2011): 3261-3278.

Reich, H. Chem 605 (2016)- Structure Determination Using Spectroscopic Methods (https://www.chem.wisc.edu/areas/reich/chem605/index.htm).

Thomas, Colin A., Erach R. Talaty, and James G. Bann. (2009) "3S-Fluoroproline As A Probe To Monitor Proline Isomerization During Protein Folding By 19F-NMR". *Chemical Communications* 23: 3366.

Tressaud, Alain and Günter Haufe. *Fluorine And Health*. Amsterdam: Elsevier Science, 2008. Print.

Vulpetti A, Dalvit C. Fluorine local environment: from screening to drug design. Drug Discov Today. 2012;17(15-16):890-7.

Wang J, Sánchez-roselló M, Aceña JL, et al. Fluorine in pharmaceutical industry: fluorine-containing drugs introduced to the market in the last decade (2001-2011). Chem Rev. 2014;114(4):2432-506.

Williamson , K. L. ; Hsu , Y. - F. L. ; Hall , F. H. ; Swager , S. ; Coulter , M. S. J. Am.Chem. Soc. 1968, 90, 6717 – 6722.

Zheng, Tong-Yuan, Yu-Ju Lin, and Jia-Cherng Horng. "Thermodynamic Consequences Of Incorporating 4-Substituted Proline Derivatives Into A Small Helical Protein". *Biochemistry* 49.19 (2010): 4255-4263.

CHAPTER 2:

NMR characterization of the ligands used in the screening process

Introduction

Several molecules were employed during the set-up of this screening method through ^{19}F NMR. The characterization of the screening target, the ribosome, was a task far too gigantic for this project. Nevertheless, all the other interacting compounds, being antibiotics or small control molecules, fluorinated or not, needed a proper biophysical characterization. The reasons ranged from quality control for the documented molecules to an exploratory step for some new compounds. It is noteworthy that, once tested by NMR, one of the purchased compound was found to be different from what was expected: deprived of its fluorine moiety. The fluorine atom has extreme properties, including its electronegativity and its oxidation potential. Therefore, its incorporation in natural compounds didn't start until the second half of the twentieth century (Wang-2013) and nowadays it can still present a challenge. The presence of a fluorine atom will strongly influence the neighbouring atoms, leading to remarkable chemical shifts and couplings. Those properties were sought out and listed.

2.1 Fluorinated antibiotics

2.1.1 Linezolid

The rising tide of bacterial resistance has created an urgent need for antibiotics belonging to new drug classes. Linezolid N-[[(5S)-3-[3-fluoro-4-(4-morpholinyl)phenyl]-2-oxo-5-oxazolidinyl]methyl] was the first commercially available synthetic oxazolidinone antibiotic, synthesized in 1996 and approved for clinical use in 2000. It represents the first principally new antibiotic platform that has entered medical practice in more than 30 years. The key pharmacophore of linezolid is a substituted oxazolidinone ring (Leach-2006). In addition to the oxazolidinone pharmacophore, linezolid possesses structural groups that enhance its effect. Indeed, the fluorine atom on the aryl group duplicates the *in vitro* and *in vivo* activity (Barbachyn- 2003). The drug acts as antibacterial agent for the treatment of multidrug-resistant gram-positive bacterial infections. Specific applications are skin infections or nosocomial pneumonia, where other antibiotics such as methicillin or penicillin are not

effective due to antibiotic resistance (Maccaroni-2008). Oxazolidinones exhibit their action through a mechanism that inhibits the bacterial protein synthesis at a very early stage, prior to chain initiation. Several of the oxazolidinones were shown to bind to the large (50S) ribosomal subunit of the bacterial ribosome and compete with chloramphenicol and lincomycin, known inhibitors of the ribosomal Peptidyl-Transferase-Center (Shinabarger-1997). Despite being located at the A-site, linezolid is thought to act as an initiation inhibitor by perturbing the position of initiator tRNA in the P-site 56 (Wilson-2013). A study employed native conditions and ^{1}H NMR spectroscopy to measure the binding affinity of four different oxazolidinones to the ribosome under equilibrium conditions, confirming that this new class of antibiotics specifically binds to 50S and 70S but not to 30S subunits. The K_d of 195 ± 40 μM determined for the oxazolidinones compounds in this study was nearly 10 times greater than that obtained with a denaturation assay (Zhou-2002). When linezolid was approved in 2000, it became the reference oxazolidinone compound to improve upon going forward. Despite research activities at more than 30 companies for over 20 years, linezolid remains the only FDA approved oxazolidinone. Linezolid is marketed by Pfizer under the trade names Zyvox (in the US and UK) or Zyvoxid (in Europe). Worldwide sales of Zyvox have exceeded $1 billion per year since 2008, making it one of few antimicrobial agents to have achieved this level of commercial success (patent expiated in May 2015) (Shaw 11). Linezolid is then a recent, widely distributed and commonly used antibiotic. The linezolid chemical structure is displayed in figure 2.1.

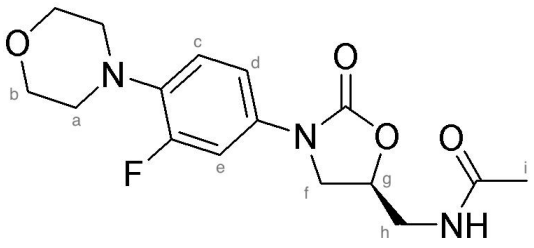


Figure 2.1 – Linezolid molecule (C16H20FN3O4), with each proton-wielding carbon labelled arbitrarily.

5 mg of Linezolid were bought from Sigma-Aldrich (CAS: 165800-03-3) and dissolved in 2 ml of D_2O (solubility of linezolid in water: 3 mg/ml, logP of 0.55) (Herrmann-2008). The molecular mass of Linezolid is 337.3 g/mol, giving then a concentration of 7.4 mM. For characterization, 500 µl of this solution were put in a 5 mm NMR tube, at 298K. A Bruker Avance 500 MHz spectrometer was used for the experiments, equipped with a BBFO probe. The attribution was realized using one-dimensional pulse sequences (with and without decoupling) as well as two-dimensional homonuclear (COSY, TOCSY, ROESY) and heteronuclear (HMQC, HMBC, HOESY) pulse sequences.

Once dissolved in D_2O, a linezolid sample was put in the spectrometer to assess its nature. Nine peaks were immediately identified. To simplify an NMR attribution process, it is wiser to start with the peaks having extreme chemical shifts values. For instance, on the far right of the spectra in figure 2.2, there is a sharp singlet around 2 ppm. This is a methyl group. On the far left, between 7 and 8 ppm, several peaks can be seen and those are aromatic protons. There are on the spectra less peaks than the number of protons in the structure. This can be briefly explained by the fact that protons in similar environments will have similar chemical shift perturbations (Jacobsen 07) and thus, be superimposed. In order to determine the chemical shift of each proton, the area under the peaks as well as their shape need to be observed. The various couplings of the nuclei impact the NMR peak structure. Besides, in NMR the area under each peak is directly proportional to the quantity of nucleus present in the solution. This property combined with the shape of the peak allows to ease the attribution of the peaks. Table 2.1 collects the peaks of the spectrum with their chemical shift and their areas. The sum of the areas equals 19 (referenced at 3 for the methyl area value) which is the number of non-exchanging protons that are present on the linezolid. The big peaks around 3 and 4 ppm, with an area of 4 protons each, correspond to the protons of the morpholine ring. The difference in area between the peaks in the aromatic region allows to distinguish between the « c », « d » and « e » protons as the two first are superimposed.

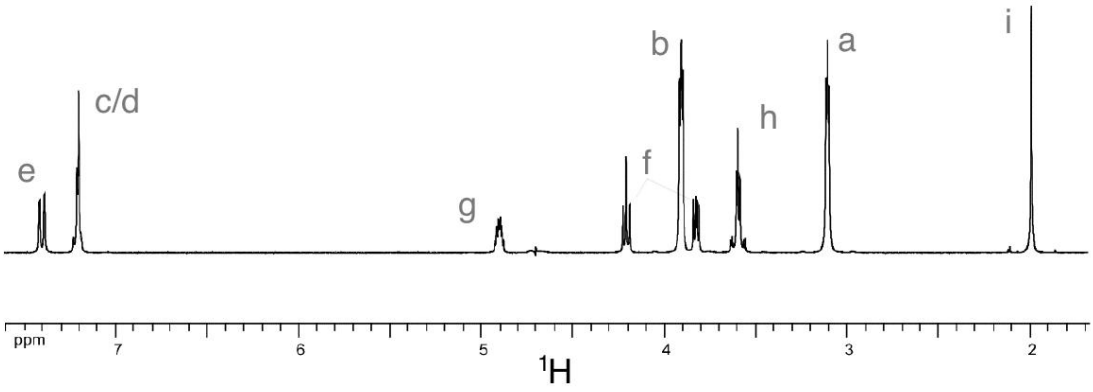


Figure 2.2 – 1D proton spectrum of linezolid in D_2O, no decoupling of ^{19}F.

The names of the corresponding protons are displayed next to their peaks, as named in figure 2.1.

Label	e	c/d	g	f	b	f	h	a	i
^{1}H (ppm)	7.41	7.21	4.9	4.21	3.91	3.83	3.6	3.11	1.99
Shape	*d*	*s*	*m*	*t*	*t*	*m*	*t*	*t*	*s*
Peak area	1	2	1	1	4	1	2	4	3

Table 2.1 – ^{1}H spectra data of the linezolid peaks, at 298K, in D_2O.

To go further in the attribution, 2D-resolved experiments are necessary. A COSY experiment produces a homonuclear correlation spectrum where a peak appears between two protons separated by a J-coupling of less than three bonds. In figure 2.3 is the COSY obtained with the linezolid. The COSY allows to determine whose peaks correspond to the protons named « f », « g » and « h » on the figure 2.1. The « f » protons, while geminal, are split at different frequencies. Those two protons establish a ^{3}J correlation with the « g » one, but not with the « h » ones, too far to be observed in COSY.

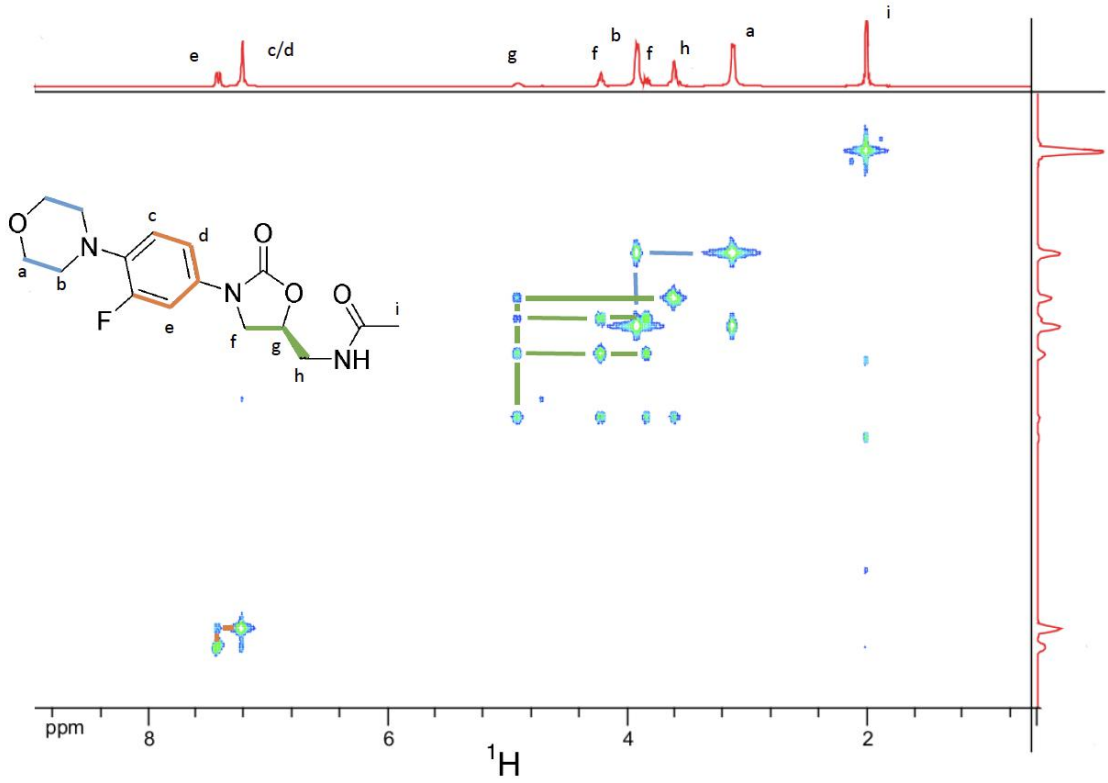


Figure 2.3 – ^{1}H COSY spectrum of Linezolid, with the 1D traces on axes, 298 K, 100% D_2O.

This set of NMR experiment is sufficient to resolve the attribution of the linezolid. Nevertheless, heteronuclear experiments, based either on carbon or fluorine, can confirm the attribution and provide useful supplementary information.

A fluorine spectrum was recorded, with and without proton decoupling, and is displayed in figure 2.4. Since only one fluorine atom is present in the solution, the decoupled spectrum has only one peak, at -121 ppm. In the non-decoupled spectrum, a multiplet appears at the same frequency.

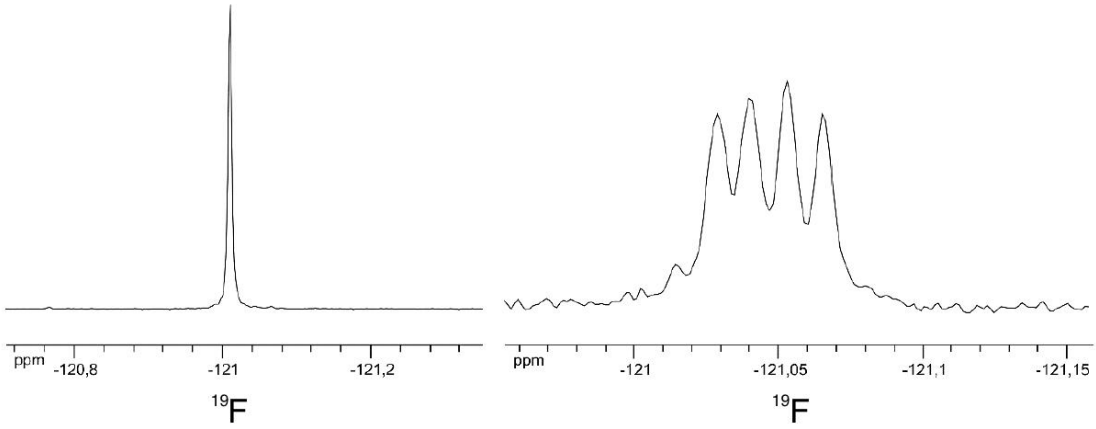


Figure 2.4 – ^{19} F spectra of Linezolid, 298K, 100% D_2O. The spectrum on the left is proton decoupled.

A HMQC is a highly resolved inverse detected heteronuclear experiment where plots of correlation appear between two dissimilar nuclei attached to each other (single bond coupling). The ^{13}C -^{1}H HMQC obtained with linezolid is available in figure 2.5. The carbon-proton correlation shows that the "f" proton signals are indeed split, but they belong to a single carbon. The unique "c/d" peak in proton is also split into two carbon signals but they cannot be properly attributed. Table 2.2 resumes the ^{13}C chemical shifts of linezolid.

Label	e	c/d	g	f	b	f	h	a	i
δ ^{13}C (ppm)	108.7	120.1 116.1	72.9	48.3	66.2	48.3	41.7	50.9	21.7

Table 2.2 – ^{13}C spectra data of the linezolid proton-wielding carbon, at 298K, in D_2O.

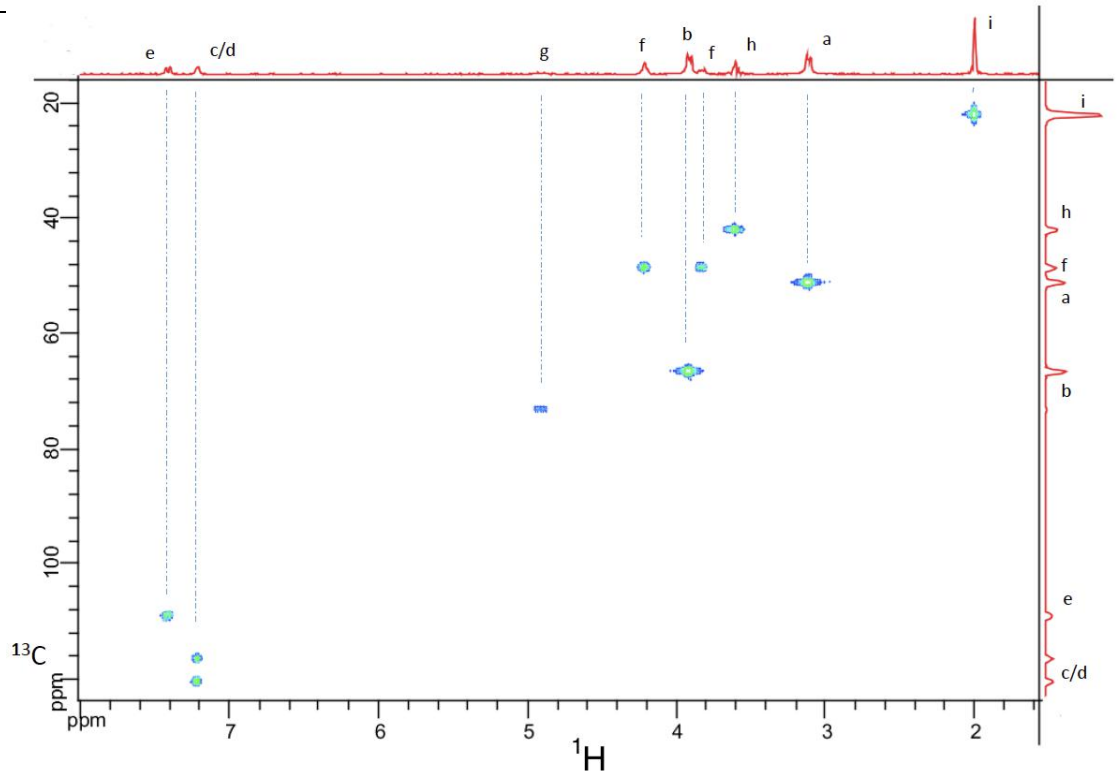


Figure 2.5 – HMQC correlation spectrum of Linezolid with carbon and proton traces on axis.

2.1.2 Florfenicol

Florfenicol is a synthetic, fluorinated analogue of chloramphenicol which lacks chloramphenicol's associated human health risk (the development of aplastic anaemia in humans) (Sams-1995). It has been used in Asia for aquaculture since the 1980's. In early 1996, an injectable formulation of florfenicol was approved for the treatment of bovine respiratory disease in the United States (Keyes-2000). Metabolism of florfenicol has been studied in rats and cattle. Florfenicol is bacteriostatic, and its mechanism of action is similar to that of chloramphenicol (Cannon-1990). Their inhibition of bacterial protein biosynthesis is mainly due to the prevention of peptide chain elongation. This bacteriostatic activity is based on a reversible binding to the peptidyltransferase center at the 50S ribosomal subunit of 70S ribosomes. Both can efficiently block the protein biosynthesis in bacteria. However, chloramphenicol resistance, through chloramphenicol acetyl-transferase (CAT), appeared in different pathogens found in human and veterinary medicine. The substitution of a fluorine atom for the hydroxyl group at C-3 in florfenicol prevents acetylation by CAT. Only a small number of the known CAM resistance genes grants also resistance to florfenicol (Schwarz 04). Florfenicol chemical structure is displayed in figure 2.6.

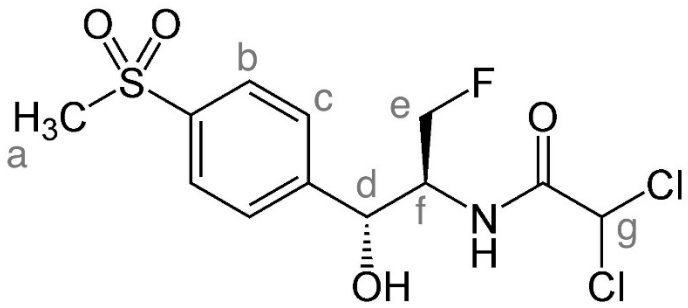


Figure 2.6 – Florfenicol molecule ($C_{12}H_{14}Cl_2FNO_4S$), with each proton-wielding carbon labelled arbitrarily.

1 g of florfenicol were bought from Tokyo Chemical Industry (CAS: 73231-34-2) and a few mg were dissolved in 500 μl of deuterated DMSO (solubility of florfenicol: 100 mM). The molecular mass of florfenicol is 358.2 g/mol. The stock concentration employed was of 63 mM.

A proton spectrum of florfenicol is available in figure 2.7 with the name of the corresponding protons of the chemical structure. Table 2.3 collects the peaks of the spectrum with their chemical shift and their areas. The sum of the areas is 11 which is close to the number of non-exchanging protons that are present on the florfenicol. The difference is due to some peaks overlapping the water signal and thus are partially suppressed. 2D experiments were also conducted on florfenicol, in order to complete the attribution: a COSY, a TOCSY and a HMQC. The COSY is displayed in figure 2.8 while a zoom on the region of the « d », « e » and « f » carbons in the COSY and the TOCSY spectra is available in figure 2.9. ^{1}H-^{1}H TOCSY (TOtal Correlated SpectroscopY also known as HOHAHA – HOmonuclear HArtmann HAhn) is useful for dividing the proton signals into groups or coupling networks. In the case of florfenicol, the TOCSY help completing the attribution done through the COSY.

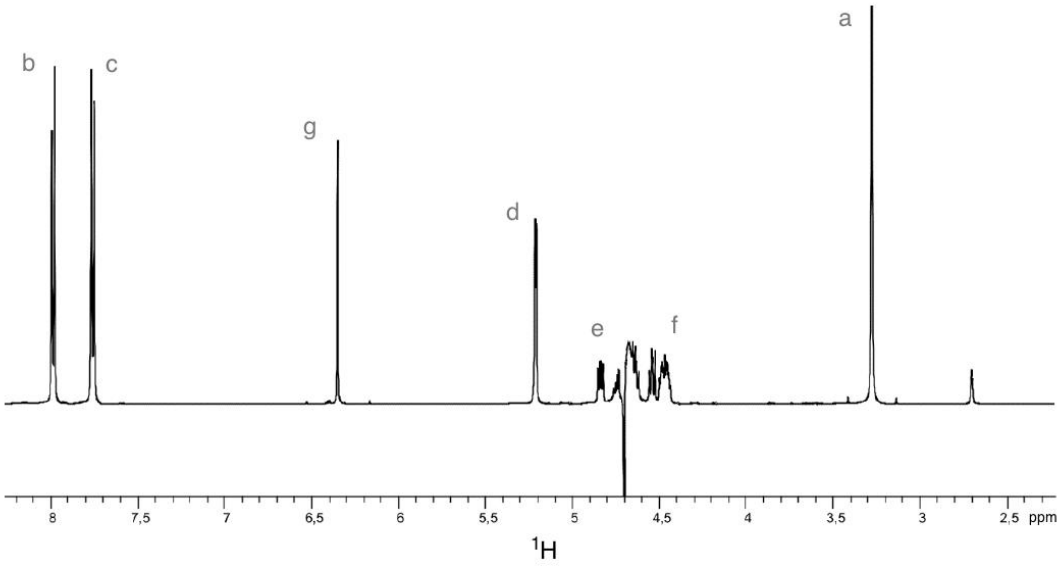


Figure 2.7 – 1D proton spectrum of florfenicol, no decoupling of ^{19}F. The proton peak around 2.8 ppm is due to the residual non-deuterated DMSO.

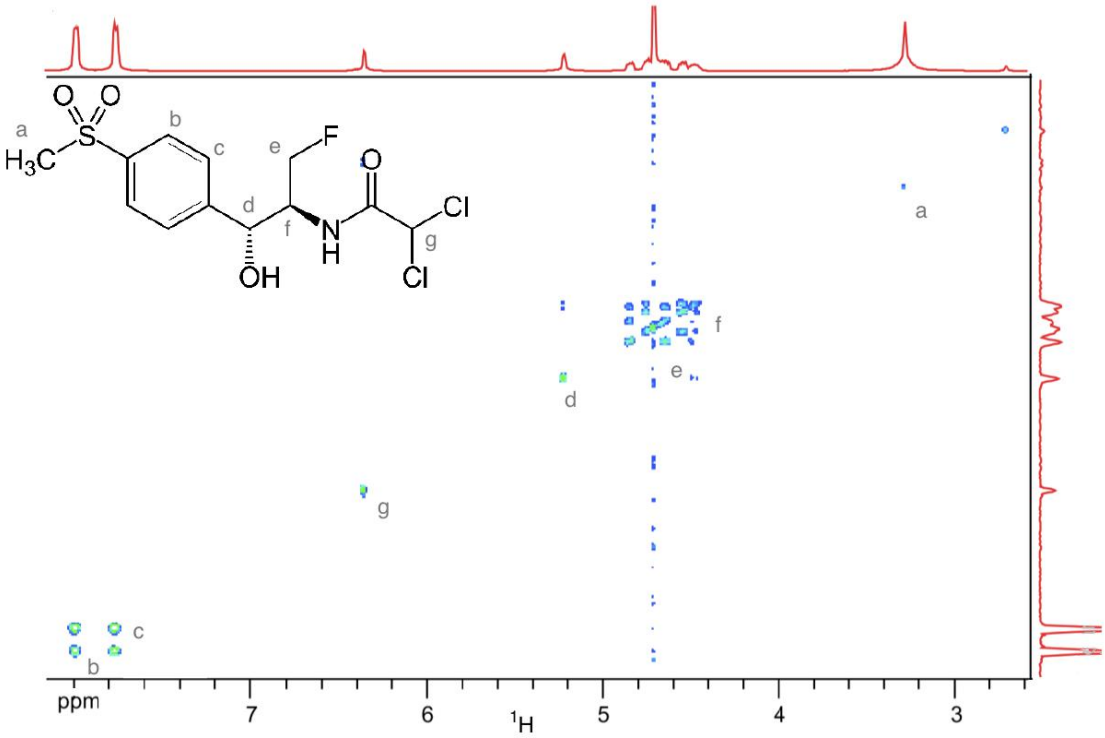


Figure 2.8 – ^{1}H COSY spectrum of Florfenicol, with the 1D traces on axes, 298 K, 100% DMSO-d6.

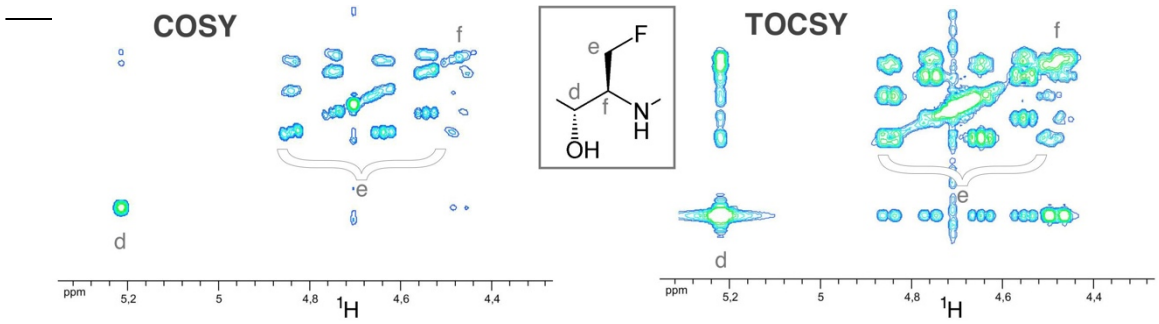


Figure 2.9 – COSY and TOCSY zoomed region of Florfenicol, 298K, DMSO-d6.

A fluorine spectra was recorded as well, with and without proton decoupling, illustrated in figure 2.10. Since only one fluorine atom is present in the solution, the decoupled spectrum has only one peak, at -224 ppm. In the non-decoupled spectrum, a triplet of doublet appears at the same frequency, due to the J-coupling with the 2 « e » protons. The ^{13}C- ^{1}H HSQC obtained with florfenicol is available in figure 2.11. The « e » and « f » proton-wielding carbon are easily distinguished. Table 2.4 resumes the ^{13}C chemical shifts of florfenicol.

Label	b	c	g	d	e	f	a
δ H (ppm)	7.99	7.76	6.35	5.21	4.69	4.46	3.28
Shape	*d*	*d*	*s*	*d*	*m*	*d*	*s*
Peak area	2	2	0.5	1	1.5	1	3

Table 2.3 – ^{1}H spectra data of the florfenicol peaks, at 298K, in DMSO-d6.

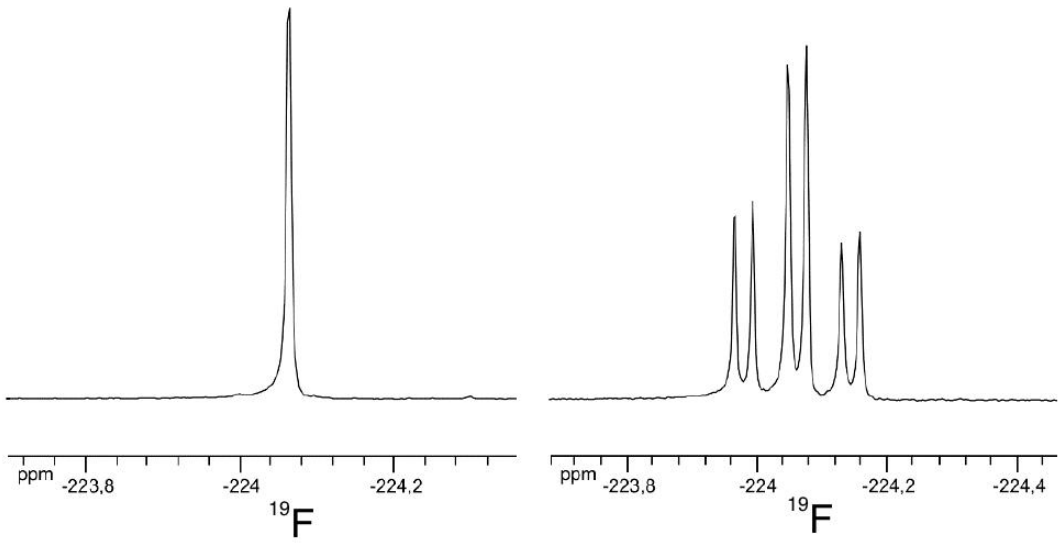


Figure 2.10 – ^{19}F spectra of florfenicol, 298K, 100% DMSO-d6. The spectrum on the left is proton decoupled.

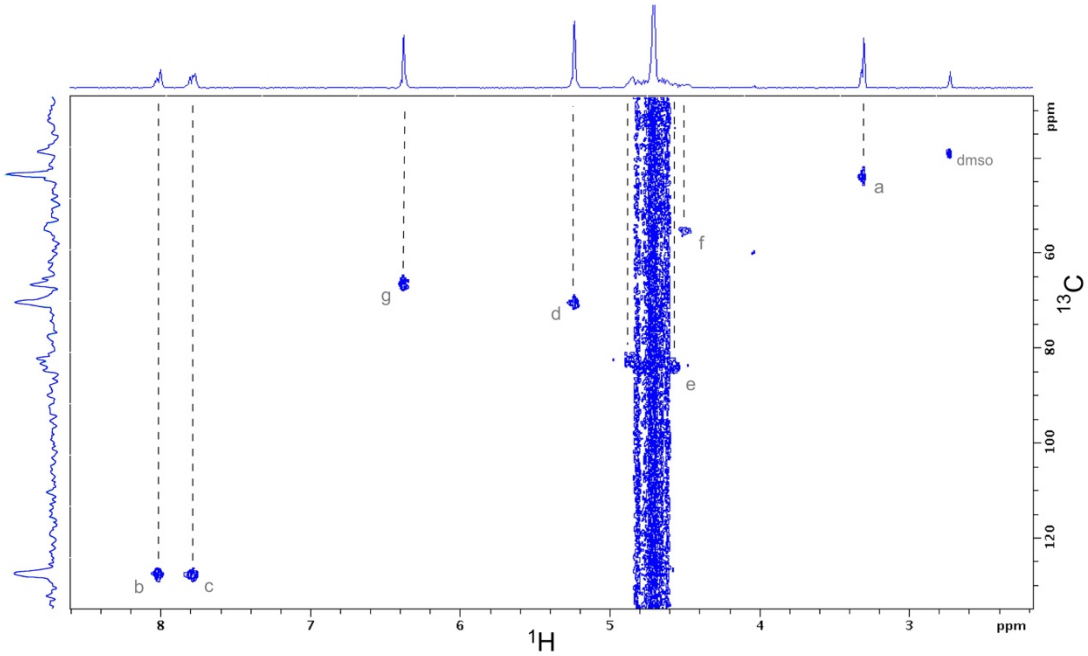


Figure 2.11 – ^{13}C HSQC correlation spectrum of florfenicol with carbon and proton traces on axis.

Label	b/c	g	d	e	f	a
δ ^{13}C(ppm)	127.4	66.4	70.3	83.2	55.1	43.6

Table 2.4 – ^{13}C spectra data of the florfenicol proton-wielding carbon, at 298K, in DMSO-d6.

2.1.3 Compound 323230 from the 4HPP family

Resulting from the urgency to address the increase in bacteria resistance, the 4H-pyridopirimidine family was only recently discovered. The development of a coupled aminoacylation/translation (A/T) system resulted in a screening for compounds that inhibit protein synthesis in bacteria in a high-throughput format (Ribble-2010).

This system screened a small-molecule compound library containing 2,100 compounds, and identified a series of 4H-pyridopyrimidine analogs that are bacterial protein synthesis inhibitors. Through the use of a set of assays to determine the effects of the 4H-pyridopyrimidines on the activities of the ribosomal ligands, it was determined that the likely target of the inhibitors is the ribosome. The activities of EF-Tu, EF-Ts, EF-G, and PheRS showed no sign of a reduction when assayed in the presence of test compounds, whereas in A/T assay mixtures containing ribosomes, protein synthesis was reduced in a dose-dependent manner. In addition, the 4H-pyridopyrimidines did not block protein synthesis in eukaryotic cells. Inhibition of protein synthesis in bacteria as a mode of action with the ribosome as the target and specificity for bacteria makes the 4H-pyridoprymidines candidates for drug development against pathogenic bacterial infections. Initial IC50's in aminoacylation/translation (A/T) assays ranged from 3 to 14 µM. The optimized compounds inhibited the growth of *E. faecalis, M. catarrhalis, H. influenzae, S. pneumoniae, S. aureus* and *E. coli tolC*. 4H-pyridopyrimidine analogs demonstrated broad-spectrum inhibition of bacterial growth and modification of the compounds established SAR (Guiles-1911). The compound of interest in this project belongs to those analogs and is illustrated in picture 2.12.

Working in collaboration with a Colorado University team led by James Bullard, 2.6 mg of powder of compound 323230 were obtained and dissolved in 250 μl of DMSO. The stock solution was estimated at 30 mM (MW = 354 g/mol) and stocked at - 20°C. The stock

solution has a faint yellow color. A proton spectrum of 323230 is available in figure 2.13 with the name of the corresponding protons of the chemical structure. Table 2.5 collects the peaks of the spectrum with their chemical shift and their areas. The DMSO peak as well as the water peak are present in the spectra, the water peak overlapping the signal of the « g » protons. Several 2D experiments were conducted on the 4HPP family member 323320 in order to assess its attribution: COSY, TOCSY, ROESY and HMQC.

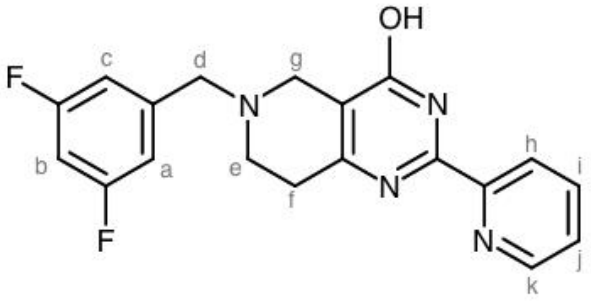


Figure 2.12 – 323230 molecule ($C_{19}H_{16}F_2NO$), with each proton-wielding carbon labelled arbitrarily.

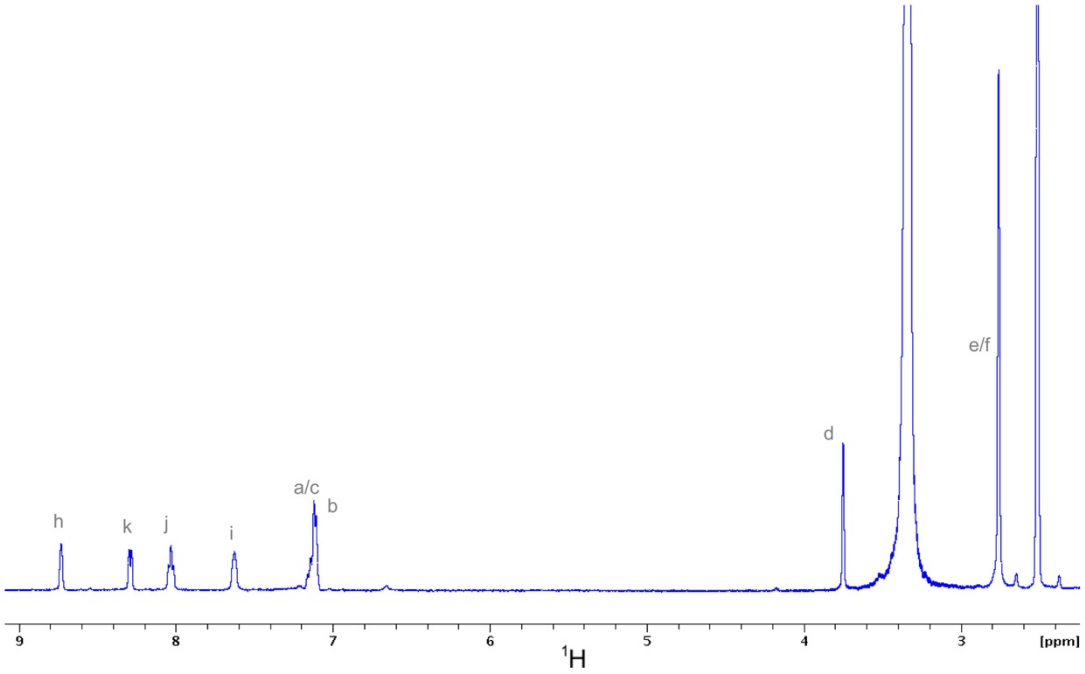


Figure 2.13 – 1D proton spectrum of 323230, no decoupling of ^{19}F, in water, at 298 K.

Label	h	k	j	i	a/c/b	d	g	e/f
1H (ppm)	8.72	8.28	8.01	7.62	7.10	3.74	3.32	2.75
Shape	*d*	*d*	*t*	*t*	m	*s*	-	m
Peak Area	1	1	1	1	3	2	-	4

Table 2.5 – ^{1}H spectra data of the 323230 peaks, at 298K, in water.

A superposition of the three homonuclear experiment is available in figure 2.14, explaining how the attribution can be verified. Indeed, the « g » protons are hidden by the water peak but they establish a NOE contact with the « d » protons and at the same time a TOCSY correlation with the « e » and « f » protons of the same cycle.

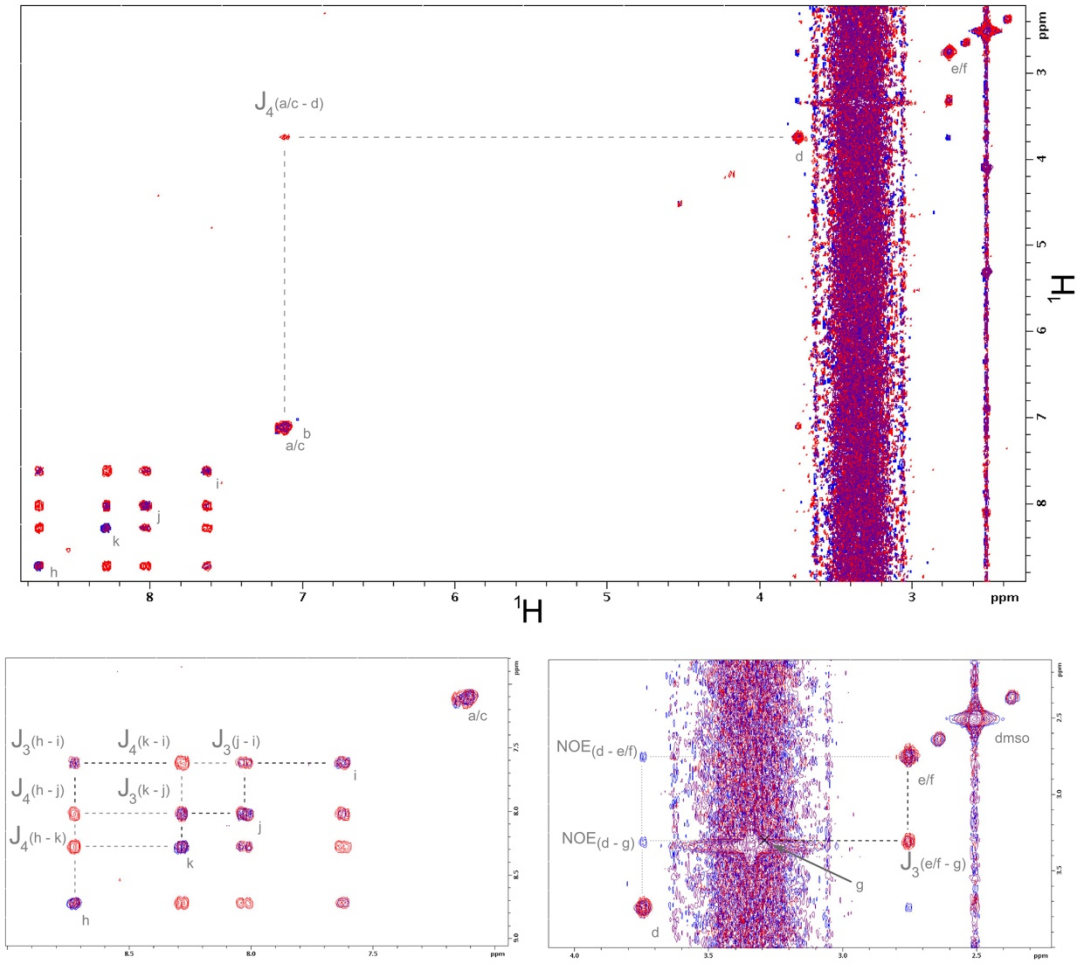


Figure 2.14 – COSY (violet), TOCSY (red) and ROESY (blue) spectra superposition obtained on the 323230 compound, mixture Water - DMSO, 298K

A ^{19}F spectrum was obtained with the sample 323230 and while 2 fluorine atoms are present in the molecule, they are in symmetrical environment, thus giving only one signal in the spectrum in figure 2.15.

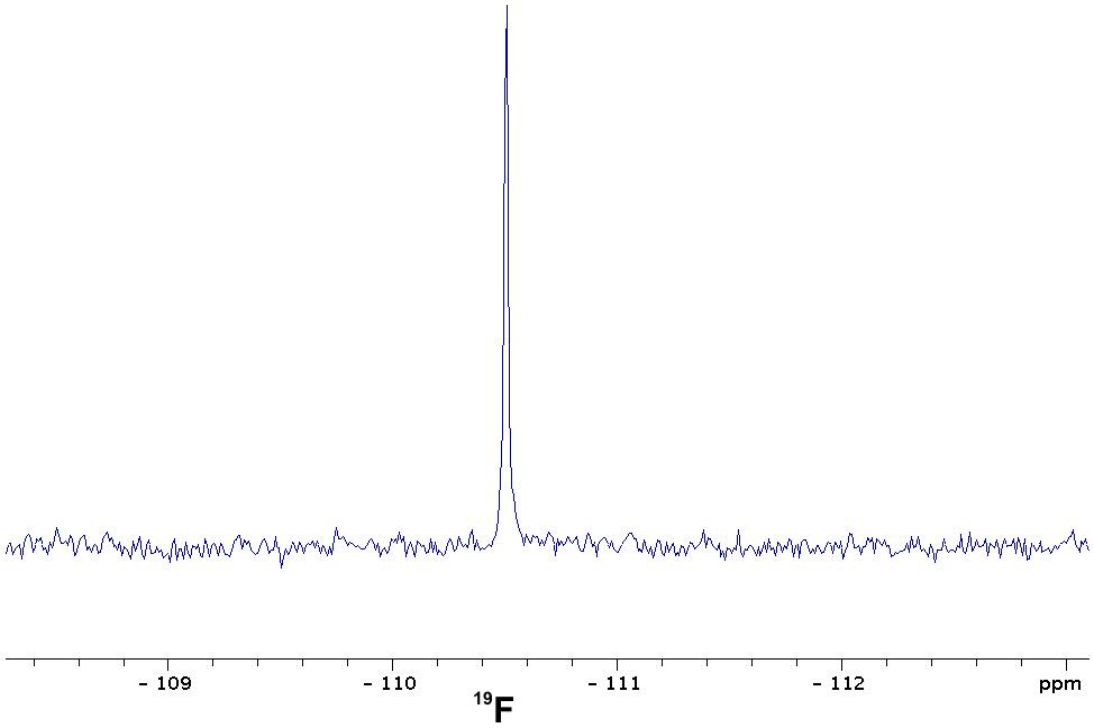


Figure 2.15 ^{19}F spectra of 323230, 298K, in water. The spectrum is proton decoupled.

A carbon homonuclear experiment was conducted as well, confirming the « g » protons chemical shifts. A zoomed portion of this spectrum is available in figure 2.16 and the ^{13}C chemical shifts of the 323230 compounds are resumed in table 2.6.

Label	h	k	j	i	a	c	b	d	g	e/f
^{13}C	101.3	122.8	112.8	117.4	102.9	110.9	-	60.6	49.4	31.5 ; 49.4

Table 2.6 – ^{13}C spectral data (values in ppm) of the 323230 proton-wielding carbon, at 298K, in water.

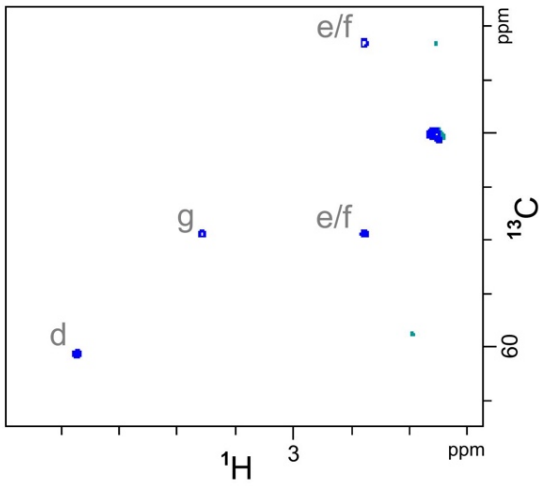


Figure 2.16 – ^{13}C HMQC correlation spectrum of 323230 zoomed on the region of the « d », « g », « e » and « f » proton-carbon correlation peaks. The peak on the far right of the spectra belongs to the DMSO.

2.1.4 Flurithromycine

Flurithromycin (Systematic (IUPAC) name (3R,4S,5S,6R,7R,9S,11R,12R,13S,14R)-6-[(2S,3R,4S,6R)-4-dimethylamino-3-hydroxy-6-methyloxan-2-yl]oxy-14-ethyl-9-fluoro-7,12,13-trihydroxy-4-[(2R,4R,5S,6S)-5-hydroxy-4-methoxy-4,6-dimethyloxan-2-yl]oxy-3,5,7,9,11,13-hexamethyl-1-oxacyclotetradecane-2,10-dione) is a fluorinated derivative of erythromycin, a well know member of the macrolides family. Its chemical structure can be found in figure 2.17, which only differs from erythromycin by a proton replaced by a fluorine atom. Erythromycin is a macrolide antibiotic which is effective against a wide range of pathogenic bacteria. Erythromycin is useful for treatment of patients with penicillin allergies but more importantly is used to treat an array of infections like bronchitis, otitis media, sinusitis and pneumonia caused by a Gram-negative pathogen, *Haemophilus influenza*. If left untreated those infections can result in systemic diseases such as meningitis, septicemia and septic arthritis. But erythromycin is unsuitable for the treatment of another Gram-negative bacterium, *Helicobacter pylori*, which causes gastritis, as the drug decomposes under the acidic conditions of the stomach. Flurithromycin (owned by the laboratories Pharmacia) launched in 1997, it is a fluorinated analogue of erythromycin developed with the aim of improving stability under acid conditions. In the treatment of gastritis, a condition which can lead to peptic ulcers, flurithromycin has a longer biological half live, better bioavailability and reaches higher tissue concentrations than erythromycin in vivo (Purser-2008).

On a structural basis, flurithromycin shows a preferential targeting of 50S subunit formation. It is an inhibitor of protein synthesis with an IC50 of 6 μg/mL and has a stronger effect on 50S ribosomal subunit formation with an IC50 of 8 μg/mL (Mabe-2004).

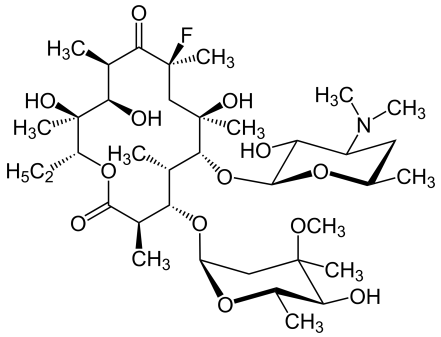


Figure 2.17 – Flurithromycin ($C_{37}H_{66}FNO_3$)

Flurithromycin (MW = 751 g/mol) is highly insoluble in water. Samples of 5 mg of powder were bought from an UK company, Key Organics Ltd, dissolved in DMSO and analyzed by NMR. A proton spectrum of flurithromycin is illustrated in figure 2.18. The high number of proton gives a crowded spectrum. In ^{19}F, Flurithromycin gives a much simpler spectrum, observable in figure 2.19, although experimentally the sample decayed rapidly (a few weeks at most) and the fluorine peak observed at -118 ppm disappeared. A second sample was ordered but no significant ^{19}F peak could be detected (the ligand concentration was of 1 mM), in a spectral range going from 50 to -300 ppm (figure 2.20). Since the only difference between Flurithromycine and Erythromycine is the replacement of a hydrogen atom by a fluorine one, the sample was checked by mass spectrometry. The resulting spectrum identified the compound as Erythromycin (MW = 734 g/mol). A third sample was ordered from the manufacturer and firstly tested in ^{1}H, the spectrum obtained is in figure 2.21. The spectrum being vastly different from the expected pattern of a macrolide antibiotic, the solution was considered faulty. The use of Flurithromycin in the project, despite the interesting developments possible with a fluorinated macrolide, was put to a halt due to the difficulties of obtaining a stable and correct sample.

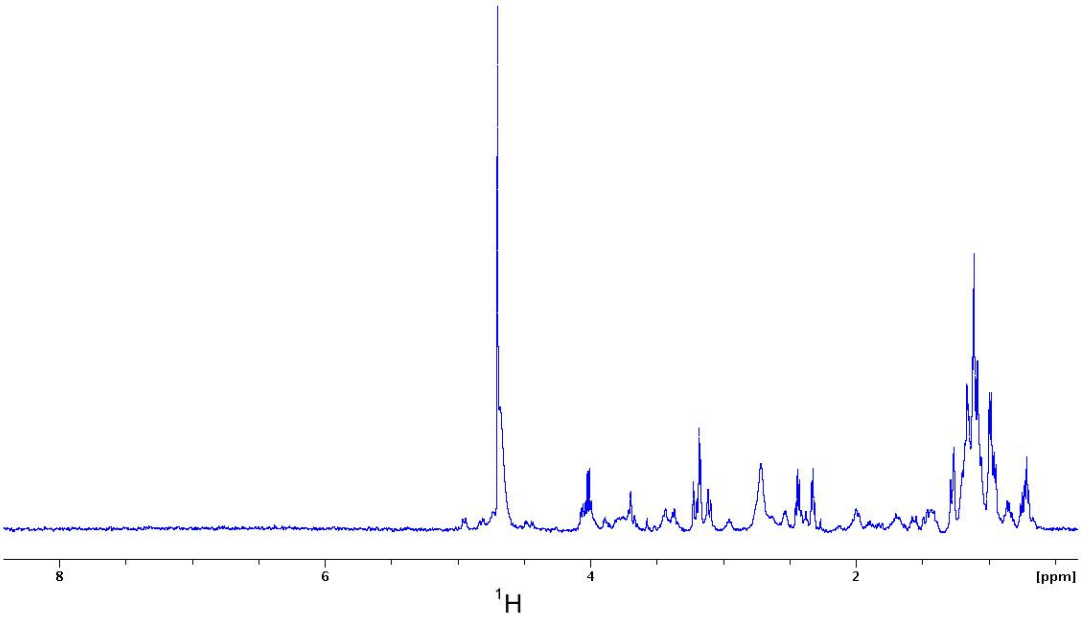


Figure 2.18 – 1D proton spectrum of Flurithromycin, no decoupling of ^{19}F, in water, at 298K.

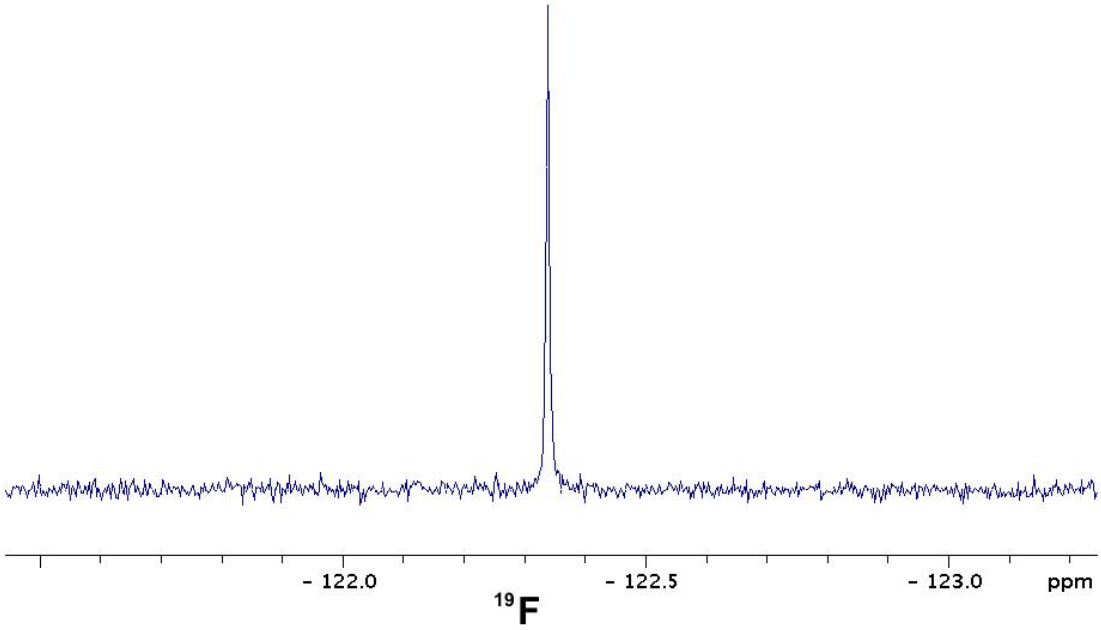


Figure 2.19 – ^{19}F spectrum of the initial peak detected in the solution of Flurithromycin in water, 298K.

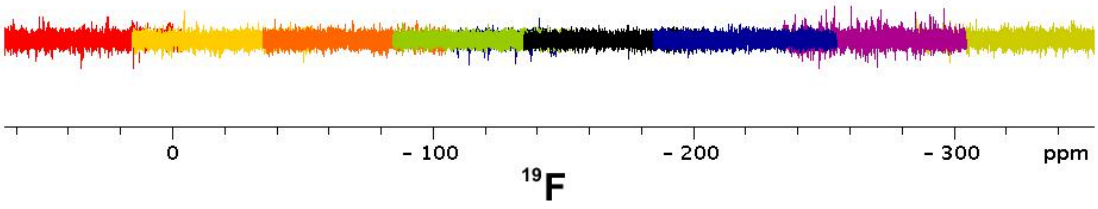


Figure 2.20 – Juxtaposition of several one-dimensional ^{19}F spectra of the second sample of Flurithromycin, in DMSO-d6, at 298 K. The spectral range scanned is 350 ppm (200 kHz). Only noise is recorded by the probe.

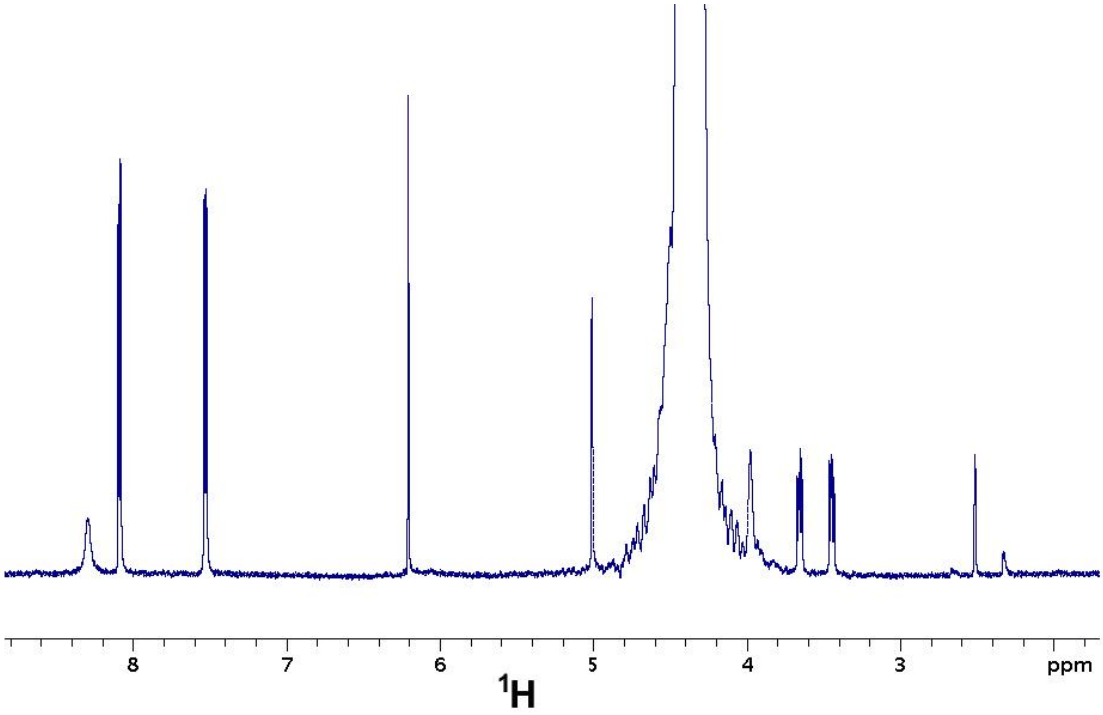


Figure 2.21 – 1D proton spectrum of the third sample of supposed Flurithromycin.

2.2 Control molecules

2.2.1 3-Fluorophenol

This small molecule (MW = 112.1 g/mol) was initially chosen as an internal control for the monitoring of the interaction between the antibiotics and the *Thermus* 70S ribosome. Its simple chemical structure is displayed in figure 2.30 and its fluorine spectra in figure 2.31. The fluorine environment is very similar to the linezolid (a fluorine moiety on an aromatic ring), resulting in a ^{19}F chemical shift close to the linezolid one with roughly 10 ppm between the two peaks in D_2O at 298K.

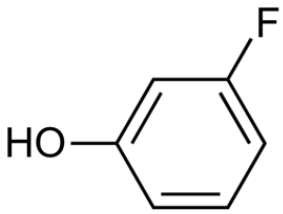


Figure 2.30 – Chemical structure of 3-Fluorophenol (C_6H_4FOH)

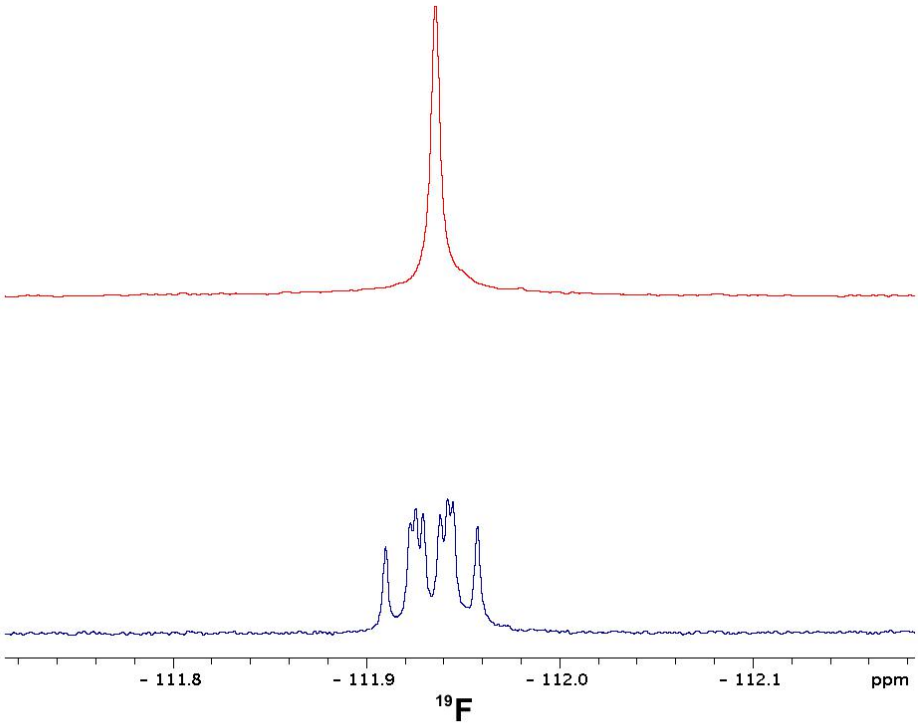


Figure 2.31 – ^{19}F spectra of 3-Fluorophenol in D_2O, 298 K. The top spectrum is proton-decoupled while the bottom one wasn't decoupled.

2.2.2 2,2,2-Trifluoroethanol

Another small molecule used as an internal control, TFE was preferred to 3-fluorophenol because the aromatic ring was suspected of establishing to many non-specific interactions on the ribosome, mainly by stacking with the ribonucleic acid bases. TFE chemical structure is illustrated in figure 2.32.

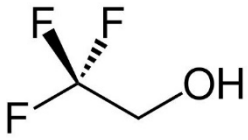


Fig. 2.32- Chemical structure of Trifluoroethanol (CF_3CH_2OH)

Molecule	Linezolid	Florfenicol	Flurithromycin	4HPP	3FP	TFE
δ ^{19}F *(ppm)*	-121	-230	-124	-110.5	-112	-77

Table 2.21 – Summary of the ^{19}F chemical shifts (values in ppm) of the ligands used in the screening project. Data given at 298K, in 90% H_2O, 10% D_2O *solutions.*

Non fluorinated ligands of the ribosome were also used in the project, as competitors for the fluorinated spies. No extensive attribution of those molecules were done, since they are all well-known through the literature. Nevertheless, 1D proton spectra were collected in order to quickly identify and repertory those molecules.

2.2.3 Lincomycin

Lincomycin is a member of the lincosamide family of antibiotic. They bind in the peptidyl-transferase center as linezolid and florfenicol and thus lincomycin was used as a specific competitor of those two fluorinated spies. The molecular structure of Lincomycin can be observed in figure 2.33 and its proton spectra in water is in figure 2.34.

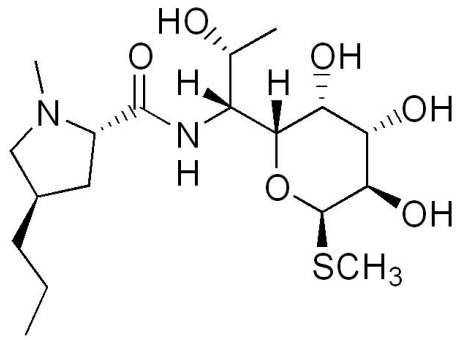


Figure 2.33 – Chemical structure of Lincomycin ($C_{18}H_{34}N_2O_6S$)

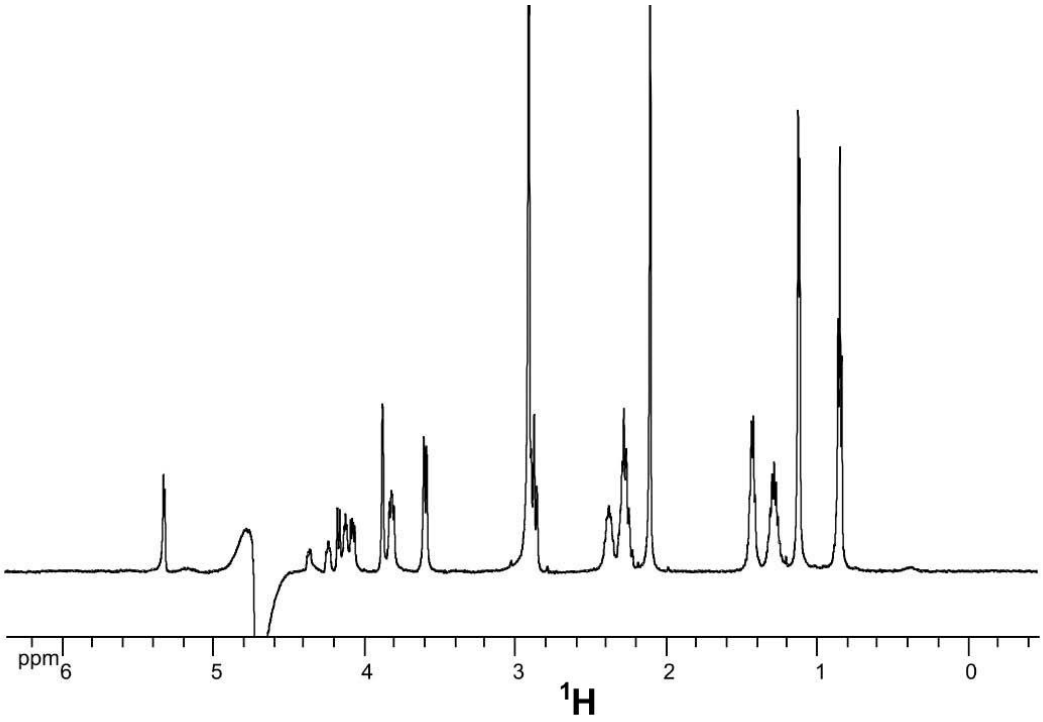


Figure 2.34 – 1D proton spectrum of a Lincomycin solution of 2 mM, (90% H_2O, 10% D_2O) 298K.

2.2.4 Kanamycin

Kanamycin is an aminoglycoside antibiotic. Its binding site is located on the 30S subunit, far from the PTC and subsequently was chosen as a non-specific competitor for linezolid and florfenicol. The molecular structure of Kanamycin A is displayed in figure 2.35 and a proton spectrum recorded on the stock solution of Kanamycin used in the experiments is in figure 2.36. Lincomycin and Kanamycin spectra were just identified, not assigned as it was not necessary for the screening experiments.

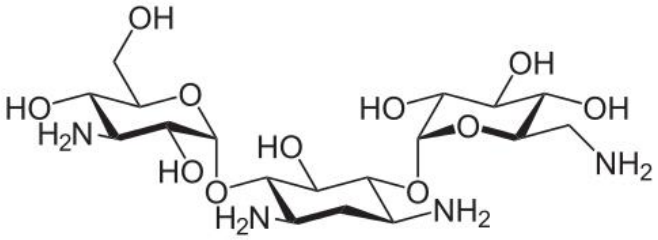


Figure 2.35 – Chemical structure of Kanamycin ($C_{18}H_{36}N_4O_{11}$)

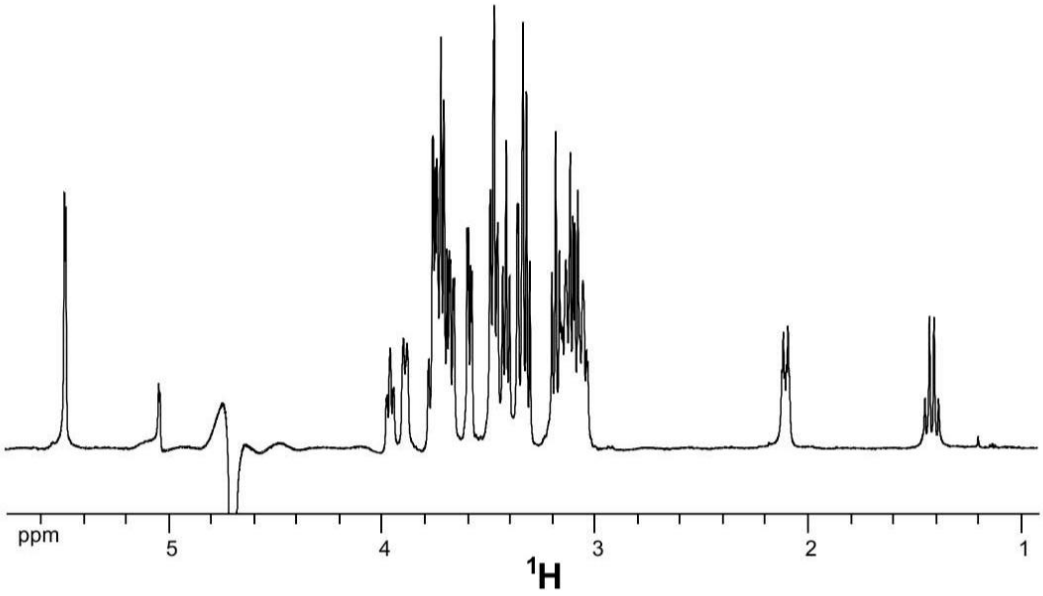


Figure 2.36 – 1D proton spectrum of a Kanamycin solution of 6 mM, (90% H_2O, 10% D_2O*) 298K.*

2.3 Ribosomes and buffers

The 70S ribosome from *T. Thermophilus* were purified with the following protocol (Fechter – 2009): add 1 ml of *T. Thermophilus* HB8 stock culture to 1 L of *T. Thermophilus* medium (0.5% (w/v) polypeptone; 0.2% (w/v) yeast extract; 0.2% (w/v) NaCl; vitamin-mineral solution (USBiological, Swampscott, MA), 1 mL/L media. The pH is adjusted with KOH to 7.2–7.4) and grow at 75°C overnight. Continue growth in 100 L fermentor and collect cells at OD_{600} = 1. Wash cells (100 g) with 1 L of buffer AT (150 mM $MgCl_2$, 500 mM NH_4Cl, 40 mM Tris–HCl of pH 7.5, 1.5 mM EDTA, 1 mM DTT), resuspend the pellet in 100 mL of the same buffer. Add DNaseI RNase free (Roche Diagnostics GmbH, Germany) and PMSF (Phenylmethanesulphonylfluoride : dissolve in ethanol. Handle under hood with extreme care.) to 1 U/mL and 1 μg/mL, respectively. Disrupt cells in French Press (Thermo Spectronic) and remove debris by centrifugation for 30 min at 30,000 × g at 4°C. Layer 29 mL supernatant on top of 7 mL Cushion 1T (1.5 M sucrose, 0.68 M CsCl, 150 mM $MgCl_2$, 20 mM Tris–HCl of pH 7.5, 1.5 mM EDTA, 1 mM DTT) and centrifuge in SW28 rotor at 100,000 × g for 20 h at 4°C. Collect a 5-mL fraction from the bottom of the cushion and dilute it three times with buffer BT (50 mM $MgCl_2$, 150 mM NH_4Cl, 20 mM Tris–HCl of pH 7.5, 0.5 mM EDTA, 1 mM DTT). Layer 29 mL ribosomes on 7 mL Cushion 2T (1.8 M sucrose, 0.8 M CsCl, 150 mM $MgCl_2$, 20 mM Tris–HCl of pH 7.5, 1.5 mM EDTA) and centrifuge in SW28 rotor at 100,000 × g for 40 h at 4°C. Collect a 4-mL fraction from the bottom of cushion and dialyse against buffer CT (Buffers CT_0, $CT_{0.8}$ and CT_1: 10 mM $MgCl_2$, 400 mM NaCl, 20 mM Tris–HCl of pH 7.5, 0.5 mM EDTA, 1 mM DTT and 0.0, 0.8, and 1.0 M $(NH_4)_2SO_4$, respectively). Load ribosomes (900 mg) on 200 mL Toyopearl Butyl 650S column equilibrated in buffer CT_0, and CT_1. Wash the column by 2 volumes of buffer CT 0.8. Elute ribosomes by 900 mL reverse gradient of ammonium sulphate (from 80 to 40%) in buffers CT_0 /CT_1 using the flow rate 6 mL/min and the fraction volume 12 mL. The 70S peak is collected and ribosomes are dialysed against buffer BT.

The end stock had a concentration of 9 mg/ml.

The buffer used in the interaction experiments is standard. Its composition and the role of each component are described in table 2.22. Since the buffer is a water based saline solution, several compounds described in this chapter were barely soluble in it. To solubilize the ligands, DMSO was added to the mixture, without exceeding a few percent.

Component	Concentration	Role
Hepes/KOH	5 mM	Maintaining enzymes structure and function
KCl	50 mM	Ionic strength
NH_4Cl	10 mM	Ionic strength
$Mg(OAc)_2$	10 mM	Prevents sub-unit dissociation
DTT	1 mM	Reduces disulfid bonds formation

Table 2.22 – Composition of the ribosomal buffer, pH 7.5 at room temperature

Conclusion

Characterizing through NMR the molecules employed in the screening experiments brought many information. As evidenced by the flurithromycine samples, it is a mandatory step to verify the molecule structure. Knowledge of the one-dimensional proton spectra of a molecule allow to quickly assert its presence in solution. Other nuclei can fill the role, such as ^{19}F. The fluorine spectra recorded on the ligand used in the project showed narrow peaks and no background noise, as expected. The fluorine chemical shifts observed were heavily influenced by the neighbouring chemical environments, allowing for a clear separation between the ligands (table 2.21). The amplitude of ^{19}F chemical shifts also mean that the pulses sequences must have specific methods (such as adiabatic shapes) to excite evenly spectral windows larger than tens of kHz. The fluorine experiments have shown that the splitting patterns due to ^{1}H-^{19}F coupling are numerous and that proton-decupling is necessary to preserve most of the ^{19}F peaks signal-to-noise ratio. This is especially true in a screening context with a macromolecule whose concentration can hardly be increased and thus, a ligand concentration that may need to be brought just above the spectrometer detection threshold.

References

Barbachyn Michael R & Charles W Ford. (2003). Oxazolidinone structure-activity relationships leading to linezolid. Angewandte Chemie International Edition, vol. 42, no. 18, pages 2010–2023.

Cannon M, S Harford & J Davies. (1990). A comparative study on the inhibitory actions of chloramphenicol, thiamphenicol and some fluorinated derivatives. Journal of Antimicrobial Chemotherapy, vol. 26, no. 3, pages 307–317.

Fechter P, Chevalier C, Yusupova G, Yusupov M, Romby P, Marzi S (2009). Ribosomal initiation complexes probed by toeprinting and effect of trans-acting translational regulators in bacteria. Methods Mol Biol.;540:247-63.

Guiles Joseph W, Andras Toro, Urs A Ochsner & James M Bullard. (2011). Development of 4H-pyridopyrimidines: a class of selective bacterial protein synthesis inhibitors. Organic and Medicinal Chemistry Letters, vol. 2, no. 1, pages 5–5.

Herrmann David J, William J Peppard, Nathan A Ledeboer, Melissa L Theesfeld, John A Weigelt & Bryan J Buechel. (2008). Linezolid for the treatment of drug-resistant infections. Expert review of anti-infective therapy, vol. 6, no. 6, pages 825–848.

Jacobsen Neil. (2007). E. NMR Spectroscopy Explained. Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology. John Wiley & Sons.

Keyes K, C Hudson, J J Maurer, S Thayer, D G White & M D Lee. (2000). Detection of florfenicol resistance genes in Escherichia coli isolated from sick chickens. Antimicrobial Agents and Chemo-therapy, vol. 44, no. 2, pages 421–424.

Leach Karen L, Steven M Swaney, Jerry R Colca, William G McDonald, James R Blinn, Lisa M Thomasco, Robert C Gad-wood, Dean Shinabarger, Liqun Xiong & Alexander S Mankin. (2006). The Site of Action of Oxazolidinone Antibiotics in Living Bacteria and in Human Mitochondria. Molecular Cell, vol. 26, no. 3, pages 393–402.

Mabe Susan, Jessica Eller & W Scott Champney. (2004). *Structure-Activity Relationships for Three Macrolide Antibiotics in Haemophilus influenzae*. Current Microbiology, vol. 49, no. 4, pages 248–254.

Maccaroni E, E Alberti, L Malpezzi, N Masciocchi & C Vladiskovic. (2008). *Polymorphism of linezolid: A combined single-crystal, powder diffraction and NMR study*. International Journal of Pharmaceutics, vol. 351, no. 1-2, pages 144–151.

Purser Sophie, Peter R Moore, Steve Swallow & Véronique Gouverneur (2008). *Fluorine in medicinal chemistry*. Chemical Society Re-views, vol. 37, no. 2, pages 320–330.

Ribble Wendy, Walter E Hill, Urs A Ochsner, Thale C Jarvis, Joseph W Guiles, Nebojsa Janjic & James M Bullard (2010).. *Discovery and analysis of 4H-pyridopyrimidines, a class of selective bacterial protein synthesis inhibitors.* Antimicrobial Agents and Chemotherapy, vol. 54, no. 11, pages 4648–4657,.

Rubini Marina, Martin A Schärer, Guido Capitani, and Rudi Glockshuber. (2013) (4R)- and (4S)-fluoroproline in the conserved cis-prolyl peptide bond of the thioredoxin fold: tertiary structure context dictates ring puckering. Chembiochem: a European journal of chemical biology, 14(9): 1053–1057.

Sams R A. (1995). Florfenicol : Chemistry and metabolism of a novel broad-spectrum antibiotic. Tieraerztliche Umschau (Germany).

Schwarz Stefan, Corinna Kehrenberg, Benoit Doublet & Axel Cloeckaert. (2004). *Molecular basis of bacterial resistance to chloramphenicol and florfenicol.* FEMS Microbiology Reviews, vol. 28, no. 5, pages 519–542.

Shaw Karen Joy & Michael R Barbachyn. (2011). *The oxazolidinones: past, present, and future.* Annals of the New York Academy of Sciences, vol. 1241, pages 48–70.

Shinabarger D L, K R Marotti, R W Murray, A H Lin, E P Melchior, S M Swaney, D S Dunyak, W F Demyan & J M Buysse. (1997). *Mechanism of action of oxazolidinones: effects of linezolid and eperezolid on translation reactions.* Antimicrobial Agents and Chemotherapy, vol. 41, no. 10, pages 2132–2136.

Wang Jiang, María Sánchez-Roselló, José Luis Aceña, Carlos del Pozo, Alexander E. Sorochinsky, Santos Fustero, Vadim A. Soloshonok, and Hong Liu. (2014). Fluorine in Pharmaceutical Industry: Fluorine-Containing Drugs Introduced to the Market in the Last Decade (2001–2011). Chemical Reviews 114 (4), 2432-2506

Wilson Daniel N. (2013). *Ribosome-targeting antibiotics and mechanisms of bacterial resistance.* Nature Reviews Microbiology, vol. 12, no. 1, pages 35–48.

Zhou Casey Chun, Steven M Swaney, Dean L Shinabarger & Brian J Stockman. (2002). *1H nuclear magnetic resonance study of oxazolidinone binding to bacterial ribosomes.* Antimicrobial Agents and Chemotherapy, vol. 46, no. 3, pages 625–629.

CHAPTER 3

Fluorine NMR study of ligand binding to 70S *Thermus thermophilus* Ribosome

(Publication under redaction, will be submitted to *Journal of Biomolecular NMR*)

Raphael Recht, Nicolas Garreau de Loubresse, Marat Yusupov, Gulnara Yusupova, Martial Piotto, and Bruno Kieffer[1]

[1]Département de Biologie et de Génomique Structurales, Institut de Génétique et de Biologie Moléculaire et Cellulaire (Université de Strasbourg / CNRS UMR 7104, INSERM U964), 1 rue Laurent Fries, 67404 Illkirch, France.

1. Introduction

Ribosome is an essential component of living organisms and, as such, constitutes a major target for antibiotics. Recently, our knowledge of ribosome has been considerably increased by the determination of the three-dimensional structure of prokaryotic and eukaryotic ribosomes (Ben-Shem-2011). These structural data provide a unique source of information to understand the fundamental mechanisms of the translation process and to suggest novel routes for the design of future antibiotic and anti-cancer therapeutic strategies. This involves the structure determination of a number of complexes between ribosome and known ribosome ligands. While a number of studies have been conducted in the past to probe ribosome-ligand interactions using ligand-observed NMR methods, there is a current lack of established methods to probe specific interactions between small ligands and the ribosome in solution. The ribosome is the largest and most complex enzyme, comprising more than fifty proteins and several RNA components (Poehsgaard et al.-2005). Several families of antibiotic molecules are targeting different sites on the ribosome such as the mRNA decoding site on the 30S subunit or the Peptidyl Transferase Center (PTC) on the 50S subunit. Most of the antibiotics have been discovered and characterized between 1930 and 1970 and only very few molecules targeting the ribosome have been designed since then. During the twenty-first century, only two novel families of molecules were designed, namely the oxazolidinones and the lipopeptides. The three-dimensional structure of the

oxazolidinone linezolid bound to the *Deinococcus radiodurans* 50S ribosomal subunit was determined. The linezolid was shown to prevent the peptide bond formation at the peptidyltransferase A site pocket by stabilizing the distinct conformation of the universally conserved 23S rRNA nucleotide U2585 (Wilson-2013). It is one of the most recent FDA approved antibiotic (2000), widely available to purchase under the name Zyvox (Pfizer). Together with a number of other pharmaceutical drugs, the linezolid has a fluorine atom that was added to the aryl cycle in order to enhance its potency (Lin-1997). The fluorine atom possesses a half spin that grants it interesting properties for Nuclear Magnetic Resonance applications. In particular, its high intrinsic sensitivity to small changes in environment has been exploited to elaborate efficient methods to detect interactions between small ligands bearing a fluorine atom and macromolecules (Dalvit-2003). Thus, the Fluorine chemical shift Anisotropy and eXchange for Screening (FAXS) method exploits the enhanced relaxation of the fluorine nucleus upon binding to its receptor to detect the interaction. The competition between a fluorinated ligand (defined as a spy ligand) and any other molecule for the same binding site reveals the interaction between this later and the target. Additionally, the general lack of fluorine atoms in biological molecules makes it particularly well adapted to study molecular interactions within complex media. Indeed, a fluorine NMR spectrum recorded in media as complex as a cell extracts will be devoid of background signals originating from proton, carbon and nitrogen atoms. This holds true when studying the interaction between a ligand and a complex target such as the ribosome. In this work, we investigated the possibility to use the FAXS method to detect ligand binding to the ribosome of *Thermus thermophilus*. Several antibiotics targeting the ribosome bearing a fluorine atom were available to setup this experiment including flurithromycin, florfenicol and linezolid. We show that this later provides an efficient way to monitor specific interactions with the peptidyl transferase center of the *Thermus thermophilus* ribosome and discuss the general applicability of this method.

2. Material and Methods

70S ribosome from *T. Thermophilus* was purified with the following protocol (Fechter – 2009): add 1 ml of *T. Thermophilus* HB8 stock culture to 1 L of *T. Thermophilus* medium (0.5% (w/v) polypeptone; 0.2% (w/v) yeast extract; 0.2% (w/v) NaCl; vitamin-mineral solution (USBiological, Swampscott, MA), 1 mL/L media. The pH is adjusted with KOH to 7.2–7.4) and grow at 75°C overnight. Continue growth in 100 L fermentor and collect cells at OD_{600} = 1. Wash cells (100 g) with 1 L of buffer AT (150 mM $MgCl_2$, 500 mM NH_4Cl, 40 mM Tris–HCl of pH 7.5, 1.5 mM EDTA, 1 mM DTT), resuspend the pellet in 100 mL of the same buffer. Add DNaseI RNase free (Roche Diagnostics GmbH, Germany) and PMSF (Phenylmethanesulphonylfluoride : dissolve in ethanol. Handle under hood with extreme care.) to 1 U/mL and 1 µg/mL, respectively. Disrupt cells in French Press (Thermo Spectronic) and remove debris by centrifugation for 30 min at 30,000 × g at 4°C. Layer 29 mL supernatant on top of 7 mL Cushion 1T (1.5 M sucrose, 0.68 M CsCl, 150 mM $MgCl_2$, 20 mM Tris–HCl of pH 7.5, 1.5 mM EDTA, 1 mM DTT) and centrifuge in SW28 rotor at 100,000 × g for 20 h at 4°C. Collect a 5-mL fraction from the bottom of the cushion and dilute it three times with buffer BT (50 mM $MgCl_2$, 150 mM NH_4Cl, 20 mM Tris–HCl of pH 7.5, 0.5 mM EDTA, 1 mM DTT). Layer 29 mL ribosomes on 7 mL Cushion 2T (1.8 M sucrose, 0.8 M CsCl, 150 mM $MgCl_2$, 20 mM Tris–HCl of pH 7.5, 1.5 mM EDTA) and centrifuge in SW28 rotor at 100,000 × g for 40 h at 4°C. Collect a 4-mL fraction from the bottom of cushion and dialyse against buffer CT (Buffers CT_0, $CT_{0.8}$ and CT_1: 10 mM $MgCl_2$, 400 mM NaCl, 20 mM Tris–HCl of pH 7.5, 0.5 mM EDTA, 1 mM DTT and 0.0, 0.8, and 1.0 M $(NH_4)_2SO_4$, respectively). Load ribosomes (900 mg) on 200 mL Toyopearl Butyl 650S column equilibrated in buffer CT_0, and CT_1. Wash the column by 2 volumes of buffer CT 0.8. Elute ribosomes by 900 mL reverse gradient of ammonium sulphate (from 80 to 40%) in buffers CT_0 /CT_1 using the flow rate 6 mL/min and the fraction volume 12 mL. The 70S peak is collected and ribosomes are dialysed against buffer BT.

The end stock had a concentration of 9 mg/ml. The buffer used for the experiments was: 5 mM Hepes/KOH, pH 7.5; 50 mM KCl; 10 mM NH_4Cl; 10 mM $Mg(OAc)_2$; 1 mM DTT. 5 mg of linezolid were bought from Sigma-Aldrich (CAS: 165800-03-3) and dissolved in D_2O. 1 g of florfenicol was bought from Tokyo Chemical Industry (CAS: 73231-34-2) and part of the powder was dissolved in DMSO-d6. NMR experiments were conducted in 5 mm Shigemi

tubes with a volume of 230 µl, 10% D_2O. The experiments were recorded at 298K and 318K. The spectrometer used was a Bruker Avance III 600 (564 MHz Larmor Frequency for ^{19}F). Two different probes were employed, allowing the use of different ligand concentrations. With a 5 mm BBFO probe, ligand and control molecules were at 50 µM (and the spectra were recorded in 2048 scans) while when using a QCI-F cryoprobe, ligand concentration could be lowered at 12 µM (and the spectra recorded in 512 scans). The fluorine spectra were recorded with a transverse relaxation filter: non-selective 90° pulse of 11.7 µs followed by a spin-echo with a 180° refocusing shape-pulse. A WALTZ-16 proton decoupling was applied during the echo and acquisition. The spin-echo theme had a $\tau/2$ of 75 ms. The data were multiplied with an exponential window function with 5 Hz line broadening prior to Fourier transformation.

3. Results

3.1. Ligand binding to the ribosome

The FAXS experiment requires the choice of an appropriate spy ligand as well as a non-interacting control molecule. Two spy molecules were used in this study, the linezolid (LNZ) and the florfenicol (Figure 1).

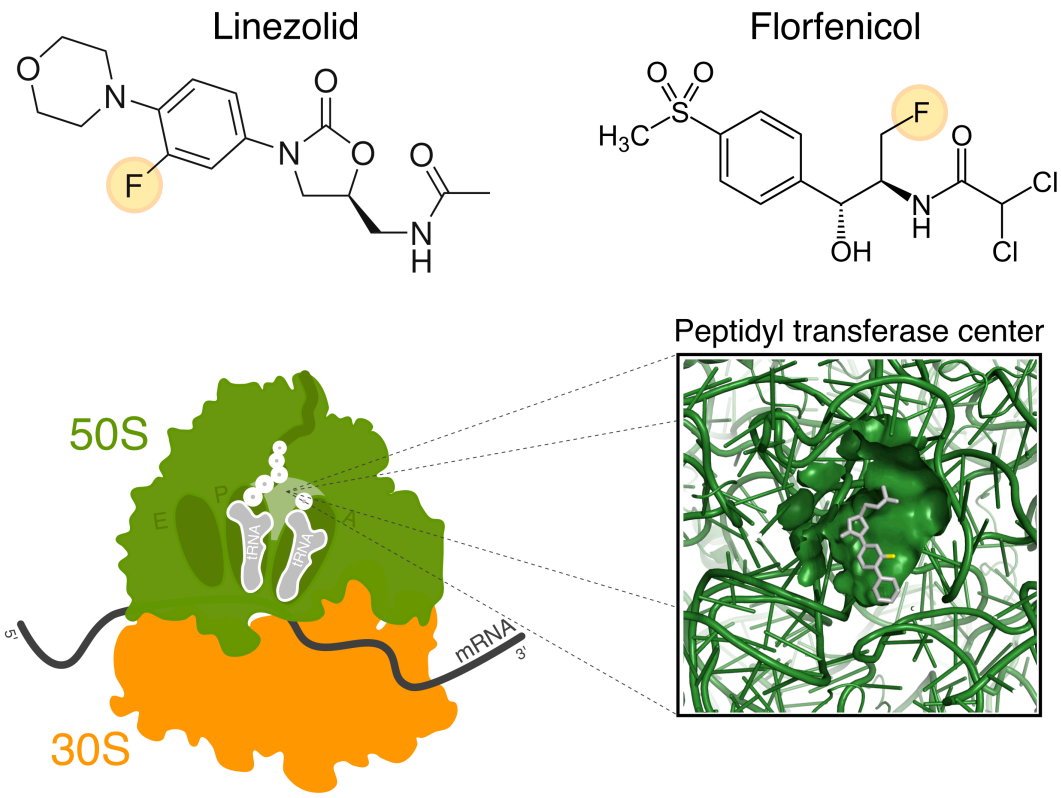


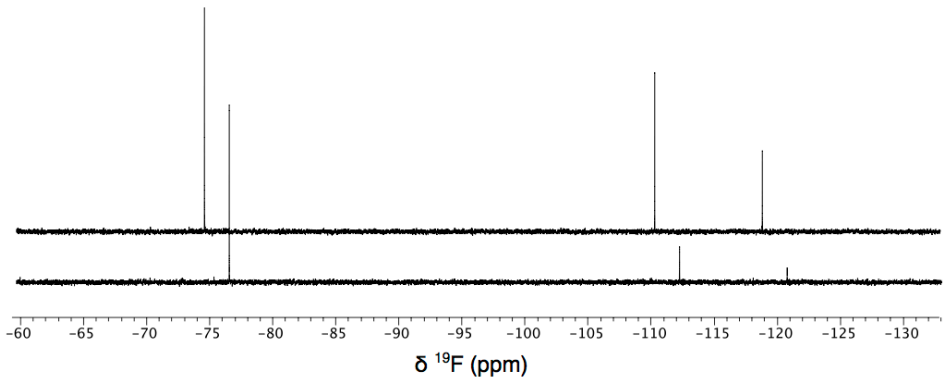
Fig. 1 The chemical structures of linezolid and florfenicol are drawn on top. In linezolid the fluorine atom is attached to a phenyl group while for florfenicol it is an aliphatic carbon in an sp_3 hybridization. Below is displayed an illustration of prokaryotic ribosome in the translation mechanism. Linezolid and florfenicol are disrupting the ribosome activity through their binding in the peptidyl-transferase center (PTC). The binding pocket of linezolid on Deinococcus radiodurans is added to the picture (PDB id : 3DLL). This zoom shows the structural complexity of the ribosome and the ribonucleotidic nature of the PTC.

Both molecules are targeting the peptidyl transferase center with similar affinities (K_d ~ 20 µM and K_d ~ 2–6 µM for the linezolid and the florfenicol, respectively, (Contreras and Vazquez 1977) (Lin et al. 1997)). The chemical environment of the fluorine atom is different in the two molecules. The aromatic environment of the fluorine in the linezolid leads to a chemical shift of -121 ppm while a shift of -220 ppm is found for the fluorine atom of the

florfenicol. Two different molecules were considered as control: 3-fluorophenol (3FP) and trifluoroethanol (TFE). The first one has a fluorine in a similar environment to linezolid with a fluorine chemical shift of -112 ppm. TFE has a trifluoromethyl group resonating at -77 ppm. The interaction experiments were performed by adding 20 µL of a ribosome stock solution to a solution containing 24 µM of 3FP and TFE and 12 µM of LNZ. The ribosome stock solution was obtained from *Thermus thermophilus* using established protocols enabling both 50S and 30S subunits to be obtained as a 70S single complex devoid of mRNA and tRNA (Fechter – 2009). *T. thermophilus* ribosome was chosen for this study for its thermophilic properties and the wide range of functional and structural information available on this system (Cava-2009) (Yusupov-1991) (Garber 1992). The resulting ribosome concentration was 1.4 µM. The ribosome addition leads to a sharp reduction of peak intensities for 3FP and LNZ molecules, while the intensity of TFE is marginally impacted (figure 2A and 2B). This indicates that the binding of both 3FP and LNZ to the ribosome occurs with off-rates in the range of µs to ms, leading to enhanced fluorine transverse relaxation rates.

The unexpected binding of 3FP to the ribosome prevented its usage as a negative control and TFE was retained for subsequent experiments. In order to avoid prohibitive relaxation times between individual scans, the inter-scan delay (2.5 s) was nevertheless set to a value significantly lower than the T_1 of TFE (5.4 s at 298K). The T_1 on LNZ at this temperature is 650 ms, which allowed a full recovery of the magnetization for the spy molecule. To account for the large frequency difference between the TFE and the LNZ (24.8 kHz), the refocussing 180° pulse was performed using a 2 ms chirp pulse with a 60 kHz frequency sweep width and a B_1 field intensity of 1.2 kHz, allowing pure phase resonance peaks to be obtained for the two signals. The comparison of T_2 relaxation times measured for of LNZ fluorine signal in presence or in absence of 1.3 µM of ribosome was used to optimize the echo delay of the FAXS experiment. An optimal delay of 150 ms was found to maximize the intensity difference between the two samples (430 ms and 170 ms for LNZ in absence or presence of ribosome, respectively), while retaining a sufficient signal/noise ratio. Similar binding experiments were conducted with the florfenicol leading to dose-dependent line broadening indicative of an interaction with the ribosome. However, the broadening is significantly lower than the one observed for LNZ.

A/



B/

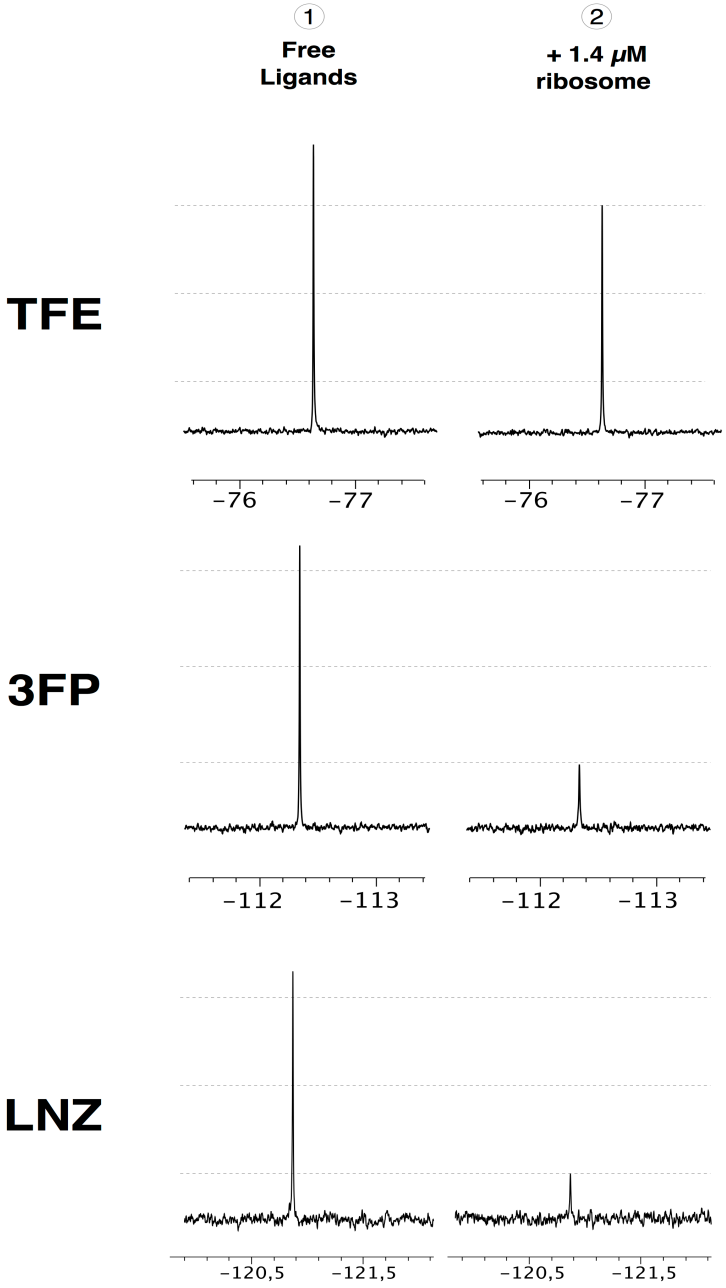


Fig. 2. A/ Superposition of three ^{19}F NMR spectra. Trifluoroethanol, 3-fluorophenol and linezolid were mixed together in a buffered solution at 298 K, with concentrations respectively of 24, 24 and 12 µM. The spectra were recorded with a 90° pulse followed by a spin-echo sequence of 150 ms. ^{1}H decupling was constant through the sequence. **B/** Zoom on the peak shapes of the molecules in the different conditions.

3.2. Effect of temperature and ligand concentration

Aiming at optimizing the conditions to observe the binding of the fluorinated spy molecule to the ribosome and the competition with other ligands, we tested the FAXS experiment over a range of temperatures from 283K to 313K (Figure 3A).

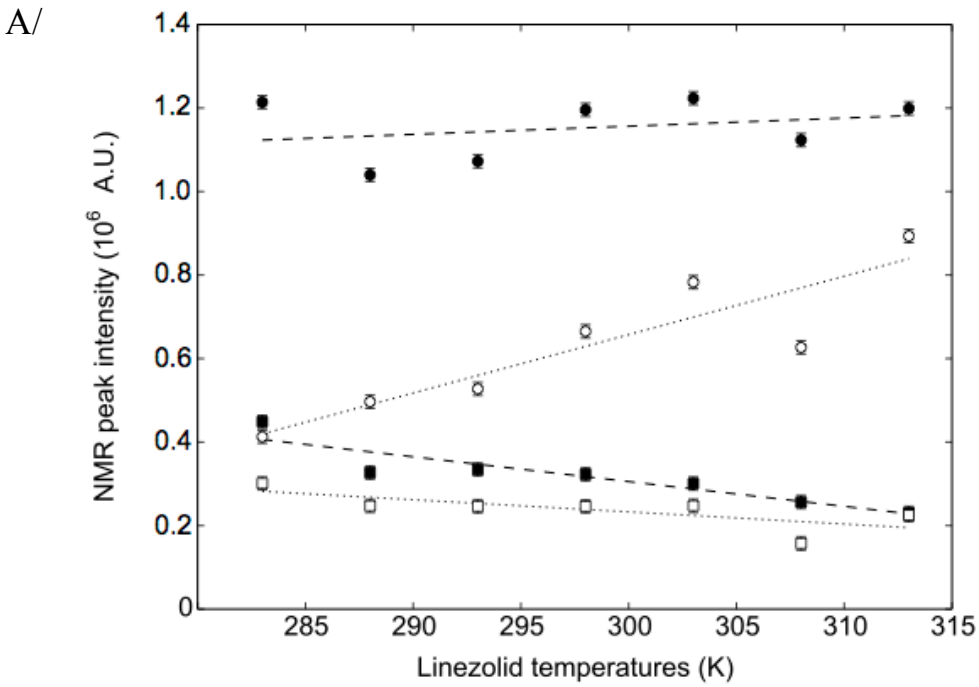


Fig. 3 A/ Impact of the temperature on the NMR peak intensity of linezolid and TFE in two solutions, with or without 0.5 µM of ribosome present. The TFE concentration was 250 µM, in presence (□) and in absence (■) of ribosome. The linezolid concentration, in presence (○) and in absence (●) of ribosome, was 1 mM. The temperature of the system was shifted from 283K to 288K, 293K, 298K, 303K, 308K and 313K. Linear fit of the data points are displayed, for the solutions devoid of ribosome (--) and for those with of ribosome (··). The NMR pulse programs used a spin-echo sequence of 150 ms and a D1 relaxation delay of 3 s.

While temperature has little effect on the fluorine signal intensity of the free LNZ, its intensity increases significantly as the temperature raises from 283 to 313 K in the presence of 1.3 µM of ribosome. This observation suggests that LNZ is in fast exchange between a free and a ribosome-bound state. Thus, a larger contrast between the LNZ signals intensities measured in absence and presence of ribosome is expected at low temperatures. The intensity of TFE resonance is slightly decreasing with the temperature. This is due to the long T_1 of the tri-fluoromethyl group which leads to lower steady state magnetization since the relaxation delay was kept constant for each temperature. The fast exchange regime allows to work in large excess of ligand. The effect of LNZ concentration on its signal intensity was therefore probed in presence and absence of ribosome (Figure 3B).

B/

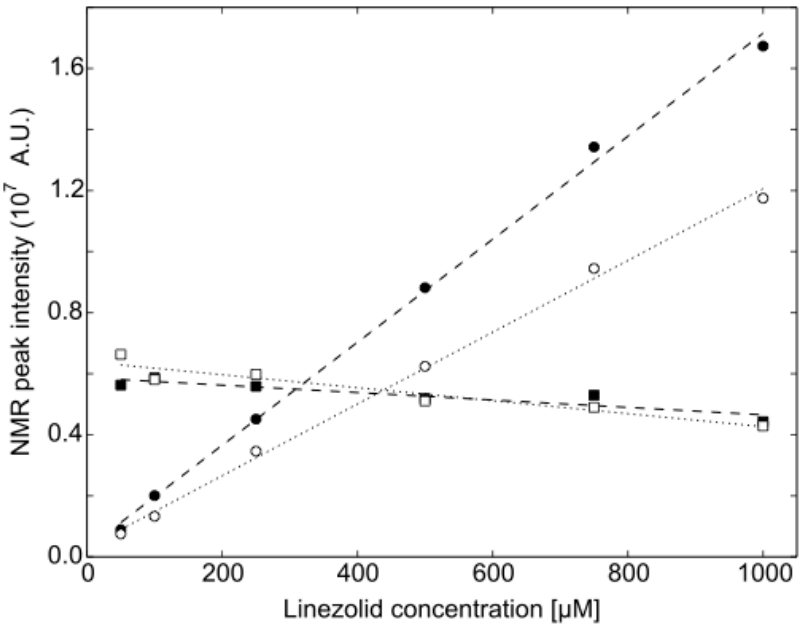


Fig 3.B/ Effect of ribosome addition on various ligand concentrations monitored through NMR peak intensities of linezolid and TFE in two buffered solutions. One of the solutions included 0.5 µM of ribosome. The TFE concentration was 250 µM, in presence (□) and in absence (■) of ribosome. The linezolid concentration, in presence (○) and in absence (●) of ribosome, was initially at 50 µM and, through consecutives additions, brought to 100, 250, 500, 750 and 1000 µM. Linear fit of the data points are displayed, for the solutions devoid of ribosome (--) and for those with of ribosome (··). The titrations were done at 298 K and the NMR pulse programs used a spin-echo sequence of 150 ms, with a D1 relaxation delay of 3 s.

The difference between the LNZ peak intensities measured in presence and in absence of 1.3 µM of ribosome is steadily growing with raising concentrations of LNZ. Noteworthy, the absence of a plateau value for this difference in peak intensities suggest that the interaction is far from the ligand saturating condition and subsequently a affinity of the LNZ for its target in the range of few hundred of µM.

3.3 Competition experiments

Due to its large size, the ribosome is a peculiar target for small ligands with many anfractuosities and binding pockets (David-eden - 2010). The FAXS experiment provides a mean to probe the competition of different ligands for the same binding site(s). The comparison of several competitors may be used to get insights on the specific nature of an interaction. Since the sole ligand peak is monitored by fluorine NMR, the competitor doesn't need to be fluorinated. The oxalidinone family has been described as competing specifically with lincomycin (Lin – 1997). Their common binding site is described in the peptidyl-transferase center, with a partial overlap in the crystallographic structures (Wilson – 2013).

In order to test whether the competition between LNZ and lincomycin could be detected using the FAXS experiment, a 20-fold excess of the latter ligand was added to a sample containing 50 µM of LNZ and 1.3 µM of ribosome (Figure 3A).

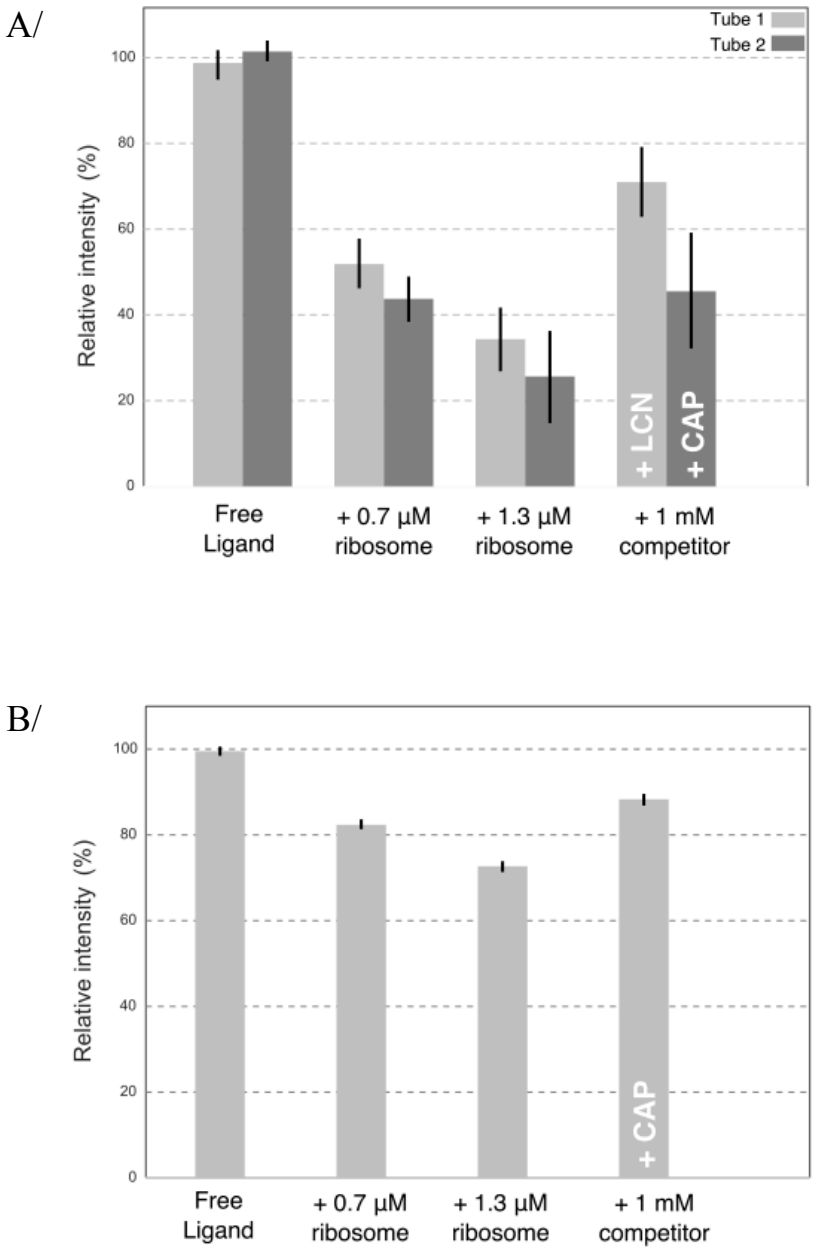


Fig. 3 Intensity of the ^{19}F NMR peak of two ligands, linezolid and florfenicol, monitored through target and competitor addition. **A/** The experiments including linezolid were conducted in parallel in two tubes, one had a final lincomycin (LCN) addition while the other had chloramphenicol (CAP). Lincomycin had a stronger effect on the linezolid peak than chloramphenicol. **B/** Florfenicol was tested with chloramphenicol as a competitor. Florfenicol has a similar titration response to linezolid, although less contrasted. Chloramphenicol competition was more effective on florfenicol than on linezolid. Both ligands were at 50 µM and the NMR pulseprogram consisted of a 90° impulsion followed by a spin-echo sequence of 150 ms. The error bars displayed correspond to the intrinsic spectral noise, the experimental repeatition accuracy can be evaluated through comparison of adjacent bars. The experiments were recorded at 298 K.

This addition led to a partial recovery of the LNZ signal from 35 % to 70 % of the normalized intensity measured in absence of ribosome. A competition experiment was also performed using Chloramphenicol, a known binder of the peptidyl-transferase site (Lin – 1997). The addition of 1 mM of Chloramphenicol to a sample containing 50 µM of LNZ and 1.3 µM of ribosome led to a partial recovery of the LNZ signal from 25 to 45 % of the normalized intensity measured in absence of ribosome. We then tested the ability of Chloramphenicol to displace the florfenicol spy molecule (figure 4B). As for the LNZ, only a partial recovery of the florfenicol signal was observed upon addition of Chloramphenicol (from 70 % to 90 % of the normalized intensity). It is worth noting that in both cases, the normalized intensity of the spy molecule is restored to a level slightly above the one obtained after the first addition of ribosome. Competition experiments were performed using lincomycin as a competitor of LNZ at 278K and 318K. We observed that the amount of recovered intensity of the spy molecule observed after addition of the competitor was systematically lower at low temperature (data not shown). The lack of complete signal recovery of the spy molecule in competition experiments as well as the temperature effect suggest that binding to peripheral sites may contribute to the signal measured by fluorine T_2 experiments.

3.4 Transversal relaxation measurements

More precise (but time-consuming) recordings of full transversal relaxation experiments were conducted on increased ligand concentration values, in absence and presence of different ribosomal concentrations. The end goal was the definition of an interaction model and to do so, the combination of different concentration values had to be tested. On a BBFO probe, the ligand concentration ranged from 1 mM to 50 µM. With a ribosomal concentration value of 0.5 µM, the R_2 values stagnated at a plateau value around 6 s^{-1}, above the free ligand relaxation value, 3 s^{-1}. The presence of a plateau even at high ligand concentration values is incompatible with a single binding site model. In the end of Spring 2016, our lab made the acquisition of a cryogenically cooled fluorine probe Bruker Q-CIF. This probe with an improved sensitivity allowed us to record R_2 values of the linezolid in presence of ribosome at ligands concentration going as low as 10 µM. With a ribosome concentration of 1.5 µM, the plateau is around 9 s^{-1} but the data points corresponding to ligand concentration values under 50 µM show initial relaxation values above the plateau value and swiftly decreasing. The fact that the curve does not go down can be explained by

the presence of many sites with weak affinities. Working at low ligand concentrations allow to detect another kind of interaction, of higher affinity. The number of points on the curve does not allow for a precise determination of the parameter but can help determining that there is more than one binding site on the ribosome. The collected relaxation data are presented in figure 5. The hypothesis of a two-states binding event could be reminded here (Barber - 1991), with a low-affinity pre-inhibition binding event and a high-affinity level step where the proper inhibition of the ribosome machinery occurs. The experimental data were able to be fitted by a model with a single binding site with an affinity of 10^{-7} M and 10^{4} sites of lower affinity (10^{-3} M).

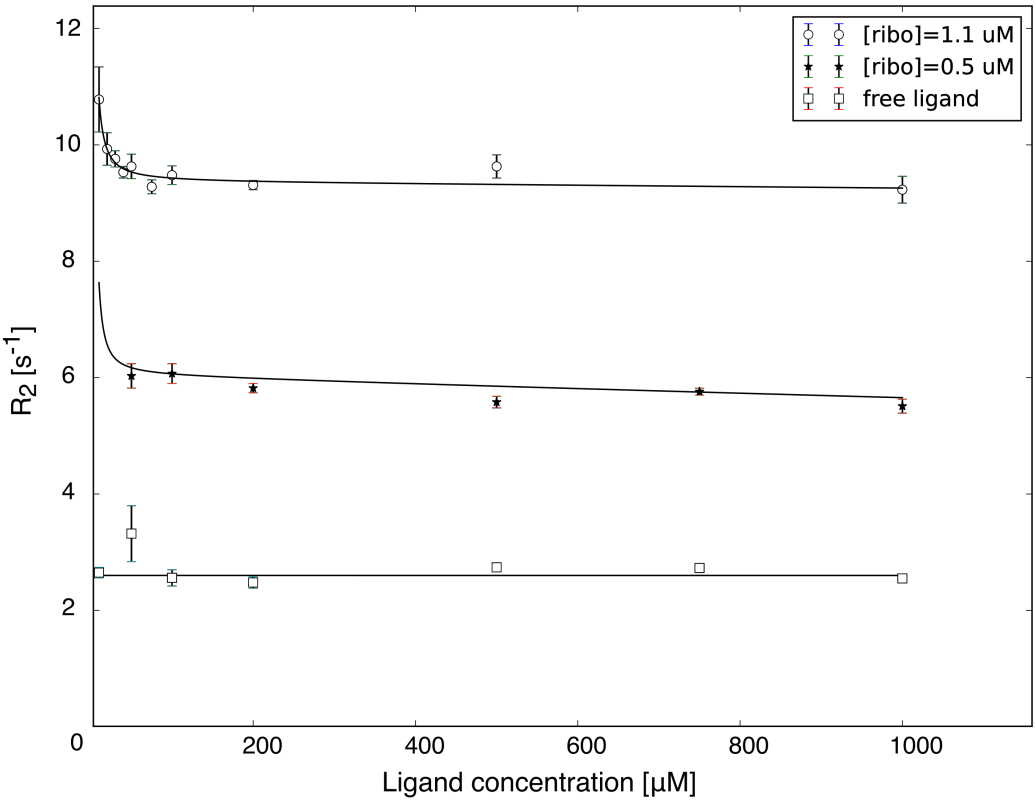


Fig. 5 Transversal relaxation values recorded for increasing concentrations of linezolid (10, 20, 30, 40, 50, 75, 100, 200, 500, and 1000 µM when available) while in absence or in presence of 0.5 and 1.1 µM of ribosome. The superimposed lines were obtained by adjusting a binding model using more than one site (one high affinity site and more than a thousand low affinities sites, a K_d going from the micromolar range up to millimolar values)

4. Conclusion

The drug discovery process is a long and convoluted pathway. Drug screening done through ^{19}F NMR has many advantages (Pellechia – 2008) and was already used on proteins, from the glucokinase (53 kDa) (Assemat – 2015) to the HAS (66.5 kDa) (Tengel - 2004). This is the first time that a screening is realized by ^{19}F NMR and directly performed on a macromolecule. The combination of those two conditions provides interesting hindsight into the binding event. NMR study on the ribosome can be realized, but due to the molecular weight of the ribosome, sharp resonances are confined to specific regions of the macromolecule (Waudby – 2013). From the ligand viewpoint, ^{1}H-NMR T_2 (CPMG) measurements have been realized on MLSB antibiotics but the binding-induced line broadening circumscribed the ligand-target molar ratio to 600 at minima (Verdier – 2000). Using ^{19}F NMR and 70S ribosome, it is possible to screen at a ligand-target ratio of 50 using a BBFO probe and as low as 12 with a cryogenically cooled Q-CIF probe. Using cryogenically cooled ^{19}F probes, NMR is able to sample different scales of binding affinity on the ribosome, including high affinity (K_d < 50 µM in this case). Those data suggest a model where numerous mM affinity pockets are involved in antibiotics binding to ribosome.

Additional experiments are needed to get a quantitative description of antibiotics binding to ribosome. However, since the screening is ligand-based, that the size of the target is not limiting, nor the buffer, and that the ribosome consumption is relatively low, cell-extracts systems and eukaryotic models may also be considered. Screening for ligands on eukaryotic ribosomes allows the discovery of new anti-parasitic drugs but also anti-cancer drugs (Garreau d.Loubresse – 2014). The method presented here could speed up the drug discovery process in complex systems.

REFERENCES

Assemat O, Antoine M, Fourquez JM, et al. ^{19}F nuclear magnetic resonance screening of glucokinase activators. Anal Biochem. 2015;477:62-8.

Ball V, Maechling C. Isothermal microcalorimetry to investigate non specific interactions in biophysical chemistry. Int J Mol Sci. 2009;10(8):3283-315.

Barber Jill, Jeffrey I. Gyi and David A. Pye. Specific, weak binding of erythromycin A (ketone) and chloramphenicol to 50S subunits of E. coli ribosomes: a 1H NMR study, J. Chem. Soc., Chem. Commun., 1991, 1249-1252.

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011). "The structure of the eukaryotic ribosome at 3.0 Å resolution". Science. 334 (6062): 1524–1529.

Cava F, Hidalgo A, Berenguer J. Thermus thermophilus as biological model. Extremophiles. 2009;13(2):213-31.

Cavanagh J., Fairbrother W.J., Palmer, A.G, Rance M., Skelton N.J. (2007) Protein NMR Spectroscopy: Principles and Practice, p21, Elsevier

David-eden H, Mankin AS, Mandel-gutfreund Y. Structural signatures of antibiotic binding sites on the ribosome. Nucleic Acids Res. 2010;38(18):5982-94

Dalvit C, Fagerness PE, Hadden DT, Sarver RW, Stockman BJ. Fluorine-NMR experiments for high-throughput screening. J Am Chem Soc. 2003;125(25):7696-703.

Discovering High-Affinity Ligands for Proteins: SAR by NMR. Science. 1996;274(5292):1531.

Fechter P, Chevalier C, Yusupova G, Yusupov M, Romby P, Marzi S. Ribosomal initiation complexes probed by toeprinting and effect of trans-acting translational regulators in bacteria. Methods Mol Biol. 2009;540:247-63.

Garber MB, Agalarov SC, Eliseikina IA, et al. Ribosomal proteins from Thermus thermophilus for structural investigations. Biochimie. 1992;74(4):327-36.

Garreau de loubresse N, Prokhorova I, Holtkamp W, Rodnina MV, Yusupova G, Yusupov M. Structural basis for the inhibition of the eukaryotic ribosome. Nature. 2014;513(7519):517-22.

Lavalette D, Amand B, Pochon F. Rotational relaxation of 70S ribosomes by a depolarization method using triplet probes. Proc Natl Acad Sci USA. 1977;74(4):1407-11.

Lin, A. H., Murray, R. W., Vidmar, T. J., & Marotti, K. R. (1997). The oxazolidinone eperezolid binds to the 50S ribosomal subunit and competes with binding of chloramphenicol and lincomycin. Antimicrobial Agents and Chemotherapy, 41(10), 2127–2131.

Pellecchia M, Bertini I, Cowburn D, et al. Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov. 2008;7(9):738-45.

Syriopoulou VP, Harding AL, Goldmann DA, Smith AL. In vitro antibacterial activity of fluorinated analogs of chloramphenicol and thiamphenicol. Antimicrob Agents Chemother. 1981;19(2):294-7.

Tengel T, Fex T, Emtenas H, Almqvist F, Sethson I, Kihlberg J. Use of 19F NMR spectroscopy to screen chemical libraries for ligands that bind to proteins. Org Biomol Chem. 2004;2(5):725-31.

Verdier, L., Gharbi-Benarous, J., Bertho, G., Evrard-Todeschi, N., Mauvais, P., & Girault, J.-P. (2000). Dissociation–equilibrium constant and bound conformation for weak antibiotic binding interaction with different bacterial ribosomes †. Journal of the Chemical Society, Perkin Transactions 2, (12), 2363–2371.

Vladimir I. Bakhmutov: NMR spectroscopy in liquids and solids (2015). 407(29), 8625–8626. doi:10.1007/s00216-015-9061-z

Waudby CA, Launay H, Cabrita LD, Christodoulou J. Protein folding on the ribosome studied using NMR spectroscopy. Prog Nucl Magn Reson Spectrosc. 2013;74:57-75.

Wilson D., Ribosome-targeting antibiotics and mechanisms of bacterial resistance. (2013). Nature Reviews Microbiology. 12(1), 35–48.

Yusupov MM, Garber MB, Vasiliev VD, Spirin AS. Thermus thermophilus ribosomes for crystallographic studies. Biochimie. 1991;73(7-8):887-97.

CHAPTER 4
NMR structural characterization and ribosome screening test of a new class of antimicrobial peptides.

4.1 Introduction

Nosocomial infections caused by bacteria affects more and more million people yearly, with multifarious ramifications on a global scale, leading to initiatives from the health organizations (from the US Centers for Disease Control and its EU counterpart, as well as the World Health Organization), including antimicrobial research. Multi-resistant nosocomial infections are caused by Gram-negatives bacteria such as *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumonia, Acinetobacter baumanii, Pseudomonas aeruginosa* and *Enterobacter spp*. Those bacteria, acronymically dubbed the "ESKAPE" pathogens, are capable of "escaping" the biocidal action of antibiotics and mutually represent new paradigms in pathogenesis, transmission and resistance (Pendleton-2013). This situation calls for the continuous development of novel class of antimicrobial compounds, inspired by the observation of natural host-pathogen interactions. A compelling example of such complex interaction is provided by the life cycle of two particular bacterial genii: *Xenorhabdus* and *Photorhabdus,* two Gram-negative bacteria, from the Enterobacteriaceae family. These bacteria live in symbiosis with soil entomopathogenic nematodes from the genii *Steinernema* and *Heterorhabditis*. Thanks to the help of the bacteria, the nematode will infect an insect host to reproduce. Once in the insect body, the bacteria are involved in a pathogenic process, killing the host and then in a mutualistic process with the nematode, by producing antibiotic compounds to prevent the growth of antagonistic microorganisms. The lifecycle is drawn and explained in figure 4.1. The nematode can't establish within his insect host without the bacteria (Ogier-2014) and no free-living forms of Xenorhabdus have been isolated outside of a nematode.

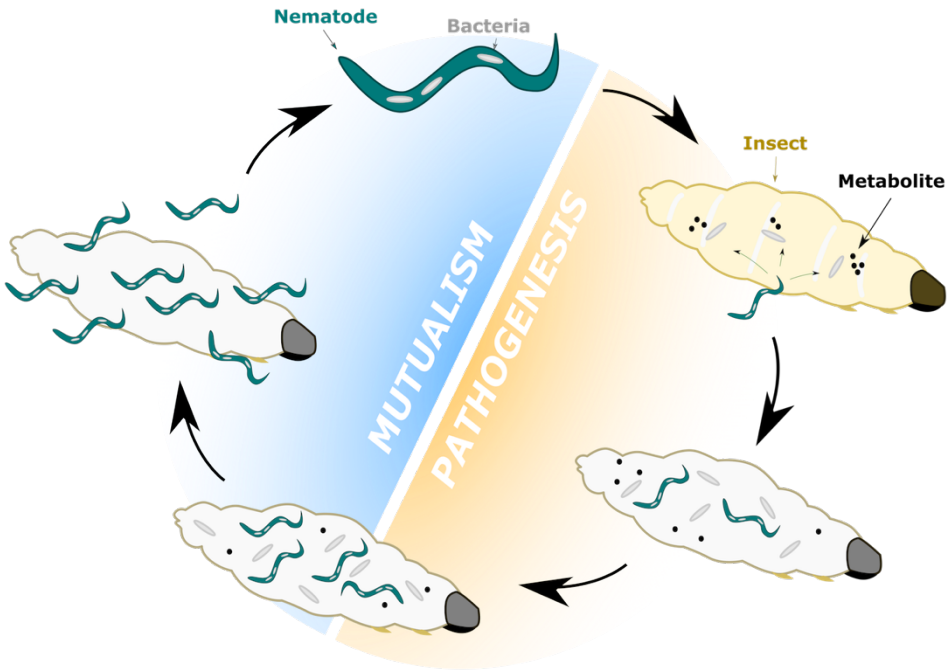


Fig 4.1 Tripartite Xhenorhabdus-nematode-insect interaction. At the beginning of the life cycle, the bacteria are living inside the intestine of the nematode. Once the nematode invades the circulatory system of an insect host, the bacteria are released, suppressing the insect immunity and killing the host. The bacteria then produce antibiotic compounds killing exogenous microorganisms, allowing nematode reproduction. Once the insect biomass is depleted, Xhenorhabdus or Photorhabdus re-colonize the nematode that will now emerge.

A great diversity of bio-active metabolites has been isolated from *Xenorhabdus*, due to an abundance of PKS (polyketide synthases) and NRPS (nonribosomal peptide synthetases) genes in its genome. Among them, the Odilhorhabdines have been identified as a family of peptides that inhibit the protein biosynthesis activity of multi-resistant Gram negative bacteria, including *Escherichia coli*, through a yet unknown mechanism. Nosopharm is a French biotechnology company, created in 2009 and based in Nîmes, France, with the aim to further characterize the mechanisms underlying the biological activity of Odilhorhabdines and develop a novel class of antibacterial compounds. In order to gain insight on the exact target of their synthetized compounds, Nosopharm decided in 2014 to start ribosome co-crystallization screening assays in collaboration with Marat Yusupov's ribosome group at the IGBMC. Sequences of odilorhabdine peptides are shown in figure 4.2.

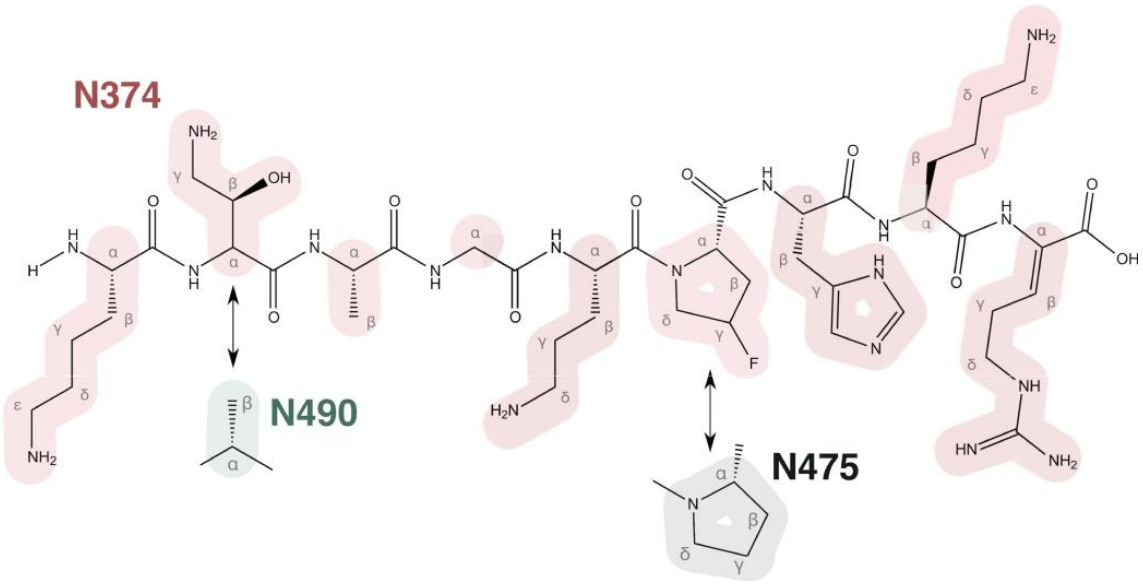


Fig. 4.2- Chemical structure of the N374 peptide, whose sequence is Lysine$_1$-Aminothreonine$_2$-Alanine$_3$-Glycine$_4$-Ornithine$_5$-Fluoroproline$_6$-Histidine$_7$-Lysine$_8$-DihydroArginine$_9$. The N490 peptide differs from N374 by substituting the Aminothreonine$_2$ by an Alanine$_2$. The N375 peptide differs from N374 by having a proline devoid of fluorine.

In order to probe the interaction between the peptides and the ribosome from *Thermus thermophilus* and *Escherichia coli*, fluorine was introduced on the proline residue, a position that was not critical for the interaction. The peptides were provided by the Nosopharm company. These compounds are based on a peptide scaffold of 9 amino-acids, some of them having chemical modifications in their side-chains. Based on biochemical assays, it was found that N374 and N495 are able to inhibit protein translation, but not N490. The amino-threonine in position 2 seems then crucial to the bactericide action. N374 and N490 have a fluorinated proline in their sequence allowing their interaction. The properties of the peptide are resumed in table 4.1. This chapter presents the full characterization of these three peptides by NMR spectroscopy and the study of their interaction with ribosomes from *Thermus thermophilus* and *Escherichia coli*.

Name	Residue X_2	Residue X_6	Bacterial synthesis inhibition
N374	Amino-Threonine	Fluoro-Proline	Yes
N490	Alanine	Fluoro-Proline	No
N495	Amino-Threonine	Proline	Yes

Tab 4.1- *Names and specific residues of the Nosopharm peptides used in this study. The canonical sequence of the 3 peptides was: Lysine1 - X2 - Alanine3 - Glycine 4 - Ornithine5 - X6 - Histidine7 - Lysine 8 - Dihydro-arginine9*

4.2 Material and Methods

The peptides N490, N374 and N495 were obtained in solid form, as samples of 2 mg of powder. The lyophilized peptides were dissolved in 1 ml of 90% H_2O mQ - 10% D_2O, leading to stock solutions with concentrations estimated in the millimolar range. 180 µl of the solutions were sampled in 3 mm NMR tubes. The assignments were realized at 298 K, on a 700 MHz Bruker Avance III spectrometer equipped with a TCI cryoprobe. Homonuclear 1D experiments were first performed in ^{1}H and ^{19}F to distinguish the peptides. Homonuclear 2D experiments were then conducted to obtain the full assignments: COSY (running time: 3 h), TOCSY (mixing time of 80 ms, 3 h) and ROESY (300 ms spinlock, 7h30). Heteronuclear experiments completed the assignment: TOCSY-^{13}C HSQC (mixing time 22 ms, 12 h), ^{13}C HSQC (5 h) and 15N HSQC (7 h).

Ribosome interactions were performed in a buffer of 10 mM MgAc2, 50 mM KCl, 10 mM NH4Cl, 5 mM HEPES pH 7.5 and 1 mM DTT for *Thermus thermophilus* 70S ribosome. For *Escherichia coli* 70S ribosome, the buffer used was 10 mM MgCl2, 50 mM NH4Cl, 20 mM TRIS pH 7.5, 0.1 mM EDTA and 6 mM betamercapto-ethanol. The experiments were realized at 298K, on a 11.7 T and a 14.1T spectrometers equipped with BBFO probes, and recorded in 1024 scans (30 minutes).

4.3 Results

4.3.1. Structural characterization of the peptides

A complete assignment of the peptides was realized, to collect all the structural similarities between them. The differences between the 3 peptides start in their sequences, with the residues 2 and 6. The second residue is an amino-threonine in N374 and N495 or an alanine in N490; the sixth residue is a proline in N495, fluorinated on the Cγ (4S conformation) N374 and N490. The differences in chemical structures can be easily assessed by the 1D spectra. N490 has an additional methyl group due to the alanine at position 2, leading to an additional resonance at 1.2 ppm (figure 4.3). The presence of a fluoroproline for peptides N374 and N490 is evidenced by the presence of a doublet at 5.3 ppm that corresponds to the Hγ of the fluoroproline.

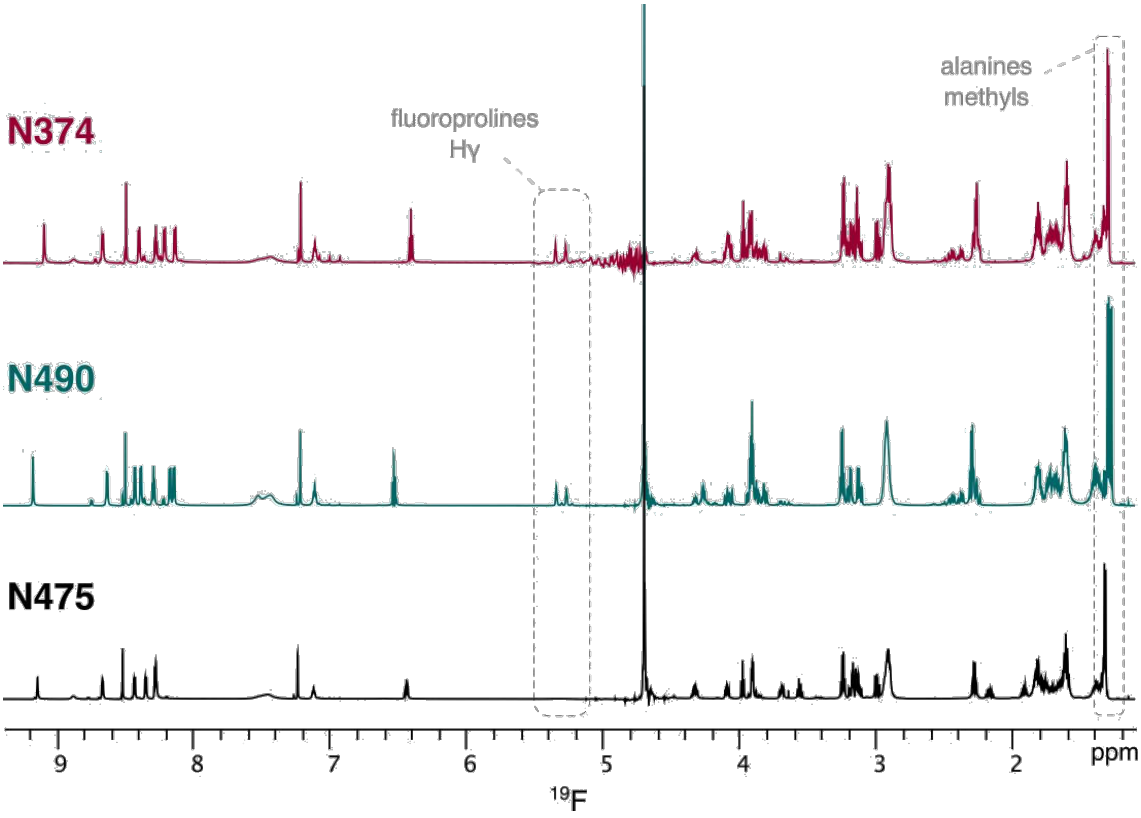


Fig. 4.3- Superposition of 1D proton spectra recorded on the 3 peptides, in 10% D_2O, at 298K.

The N374 ^{1}H chemical shifts are collected in table 4.2. For N490 and N495, it is respectively in table 4.3 and table 4.4. The ^{13} C chemical shifts helped to confirm the proton assignments and therefore are presented in table 4.5 for N374, table 4.6 for N490 and table 4.7 for N495. Finally,^{15} N-^{1}H HSQCs completed the mapping of all magnetic nuclei in the peptides and are then resumed in table 4.8 for N374, table 4.9 for N490 and table 4.10 for N495.

	K_1	Amino-T_2	A_3	G_4	Orn_5	19F-P_6	H_7	K_8	Dihydro-R_9
Hα	3.97	4.46	4.32	3.92	4.52	4.57	4.58	4.33	-
Hβ	1.82	4.08	1.3	-	1.60 1.68	2.26-2.28 2.38-2.44	3.11 3.19	1.72 1.82	-
Hγ	1.33	2.99 3.15	-	-	1.68 1.74	5.27 5.35	-	1.39	2.26
Hδ	1.61	-	-	-	2.93	3.81-3.87 4.06-4.11	-	1.61	3.23
H	2.89	-	-	-	-	-	-	2.91	-
NH	-	8.88	8.66	8.28	8.4	-	8.14	8.21	9.09
2H	-	-	-	-	-	-	7.22	-	-
4H	-	-	-	-	-	-	8.5	-	-
NH_2	7.39	-	-	-	7.51	-	8.72	7.44	(NH) 6.41 (HN) 7.11

Tab. 4.2: ^{1}H NMR (values in ppm) spectral data for the peptide N374, 298K (10% D_2O)

	K_1	A_2	A_3	G_4	$Ornithine_5$	$^{19}F\text{-}P_6$	H_7	K_8	Dihydro-R_9
Hα	3.91	4.27	4.26	3.9	4.5	4.57	4.57	4.32	-
Hβ	1.81	1.28	1.30	-	1.60 1.68	2.25-2.27 2.37-2.44	3.11 3.19	1.82	-
Hγ	1.37	-	-	-	1.68 1.74	5.26 5.34	-	1.36	2.3
Hδ	1.61	-	-	-	2.93	3.81-3.86 4.06-4.10	-	1.62	3.24
Hε	2.91	-	-	-	-	-	-	2.92	-
NH	-	8.63	8.43	8.29	8.38	-	8.14	8.17	9.18
2H	-	-	-	-	-	-	7.22	-	-
4H	-	-	-	-	-	-	8.5	-	-
NH_2	7.47	-	-	-	7.53	-	-	7.44	(NH) 6.53 (HN) 7.11

Tab. 4.3: ^{1}H NMR (values in ppm) spectral data for the peptide N490, 298K (10% D_2O)

	K_1	Amino-T_2	A_3	G_4	Orn_5	P_6	H_7	K_8	Dihydro-R_9
Hα	3.97	4.47	4.32	3.9	4.62	4.32	4.59	4.33	-
Hβ	1.82	4.08	1.32	-	1.64 1.75	1.77 2.16	3.11 3.18	1.71 1.81	-
Hγ	1.33	2.99 3.15	-	-	1.61 1.66	1.83 1.91	-	1.38	2.28
Hδ	1.60	-	-	-	2.93	3.55 3.69	-	1.61	3.23
Hε	2.90	-	-	-	-	-	-	2.9	-
NH	-	8.88	8.66	8.27	8.28	-	8.43	8.35	9.14
2H	-	-	-	-	-	-	7.24	-	-
4H	-	-	-	-	-	-	8.52	-	-
NH_2	7.47	-	-	-	7.51	-	-	7.44	(NH) 6.44 (HN) 7.12

Tab. 4.4: ^{1}H NMR (values in ppm) spectral data for the peptide N495, 298K (10% D_2O)

	K_1	Amino-T_2	A_3	G_4	Orn_5	$^{19}F\text{-}P_6$	H_7	K_8	Dihydro-R_9
Cα	52.9	-	49.87	42.18	51.66	-	52.6	53.74	-
Cβ	30.38	67.5	16.84	-	27.02	35.68	25.95	30.68	-
Cγ	21.17	41.78	-	-	23.12	-	-	21.93	27.05
Cδ	26.37	-	-	-	39.14	53.95	-	23.12	39.8
Cε	39.2	-	-	-	-	-	-	39.41	-
C_{2H}	-	-	-	-	-	-	57.61	-	-

Tab. 4.5: ^{13}C NMR (values in ppm) spectral data for the peptide N374, 298K (10% D_2O)

	K_1	A_2	A_3	G_4	Orn_5	$^{19}F\text{-}P_6$	H_7	K_8	Dihydro-R_9
Cα	52.9	49.63	49.63	41.95	51.7	?	52.42	53.78	-
Cβ	30.44	16.48	16.94	-	27.04	35.69	25.86	30.64	-
Cγ	21.07	-	-	-	?	32.55 33.56	-	21.9	27.12
Cδ	26.37	-	-	-	39.18	53.96	-	26.41	39.71
Cε	39.42	-	-	-	-	-	-	39.18	-
C_{2H}	-	-	-	-	-	-	57.62	-	-

Tab. 4.6: ^{13}C NMR (values in ppm) spectral data for the peptide N490, 298K (10% D_2O)

	K_1	Amino-T_2	A_3	G_4	Orn_5	P_6	H_7	K_8	Dihydro-R_9
Cα	53.04	-	49.81	42.23	51.27	60.60	52.53	53.66	-
Cβ	30.43	67.46	16.848	-	27.49	29.41	26.32	30.65	-
Cγ	21.19	41.76	-	-	23.20	24.28	-	21.98	27.05
Cδ	26.39	-	-	-	39.14	47.84	-	26.39	39.76
Cε	39.18	-	-	-	-	-	-	39.18	-
C_{2H}	-	-	-	-	-	-	57.47	-	-

Tab. 4.7: ^{13}C NMR (values in ppm) spectral data for the peptide N495, 298K (10% D_2O)

	K_1	A-T_2	A_3	G_4	Orn_5	$^{19}F\text{-}P_6$	H_7	K_8	DihydroR_9
NH	-	-	130.28	107.92	121.65	-	117.18	122.93	125.81
NH_2	112.8	-	-	-	112.8	-	-	112.8	124.03

Tab. 4.8: ^{15}N NMR (values in ppm) spectral data for the peptide N374, 298K (10% D_2O)

	K_1	A_2	A_3	G_4	Orn_5	$^{19}F\text{-}P_6$	H_7	K_8	Dihydro-R_9
NH	-	126.95	124.65	107.89	121.51	-	117.04	122.85	124.29
NH_2	112.54	-	-	-	112.54	-	-	112.54	123.86

Tab. 4.9: ^{15}N NMR (values in ppm) spectral data for the peptide N490, 298K (10% D_2O)

	K_1	Amino-T_2	A_3	G_4	Orn_5	P_6	H_7	K_8	Dihydro-R_9
NH	-	116.73	130.27	108.13	120.97	-	118.79	123.22	125.32
NH_2	112.5	-	-	-	112.5	-	-	112.5	123.90

Tab. 4.10: ^{15}N NMR (values in ppm) spectral data for the peptide N495, 298K (10% D_2O)

		(N374 - N490) Aminothreonine$_2$ -> A$_2$			(N374-N495) ^{19F}P$_6$ -> P$_6$		
		K$_1$	A$_3$	G$_4$	O$_5$	H$_7$	K$_8$
^{1}H	Hα	+0.06	-0.06		-0.1		
	Hβ	+0.01	-0.02				
	Hγ	+0.04					
	Hδ	+0.01					
	NH		-0.23		-0.1	-0.3	-0.15
^{13}C	Cα		+0.2	+0.2	+0.3	±0.1	
	Cβ		-0.1		-0.5	-0.7	
	C$_{2H}$					+0.1	
^{15}N	NH		+6		-0.5	+1.7	+0.3

Tab. 4.11: Chemical shift differences in ppm observed between peptides.

The 3 peptides have similar spectra yet the modifications on the residues 2 and 6 result in some localized chemical shift differences. A first insight on the possible consequences of the chemical modifications on the peptide conformations is provided by the pairwise comparison of chemical shifts done in table 4.11. The substitution of the amino-threonine by an alanine at position 2 leads essentially to a large shift of the ^{15}N resonances of the following alanine (A$_3$) due to the direct neighbouring effect. Beside the effect on the A$_3$ resonances, no significant shifts are observed indicating the lack of large conformation changes. Interestingly, the introduction of a fluorine in the proline 6 leads to chemical shifts perturbations over a larger region, as shown by the shift of K$_8$ amide nitrogen by 0.3 ppm.

4.3.1.1 Conformational analysis

Further insights on the peptide conformations were obtained by analysis of inter-proton NOEs (figures 4.4 and 4.5). While sequential NOEs were used to confirm the sequential assignment obtained from the analysis of the TOCSY spectra (the figure 4.4 shows the backbone walk in the fingerprint region of N374 with a superposition of a TOCSY and a ROESY experiment), the comparison of their intensities in the ROESY was used to detect possible changes of the peptide backbone conformation (the values are listed in table 4.12). In particular, the NOE $H_\alpha^i - H_N^{i+1}$ provides a sensitive probe of the local conformation. Similar ratios between the intra residue NOE $H_N^i - H_\alpha^i$ and the inter-residue NOE $H_\alpha^i - H_N^{i+1}$ observed for the three peptides is indicative of a shared conformation.

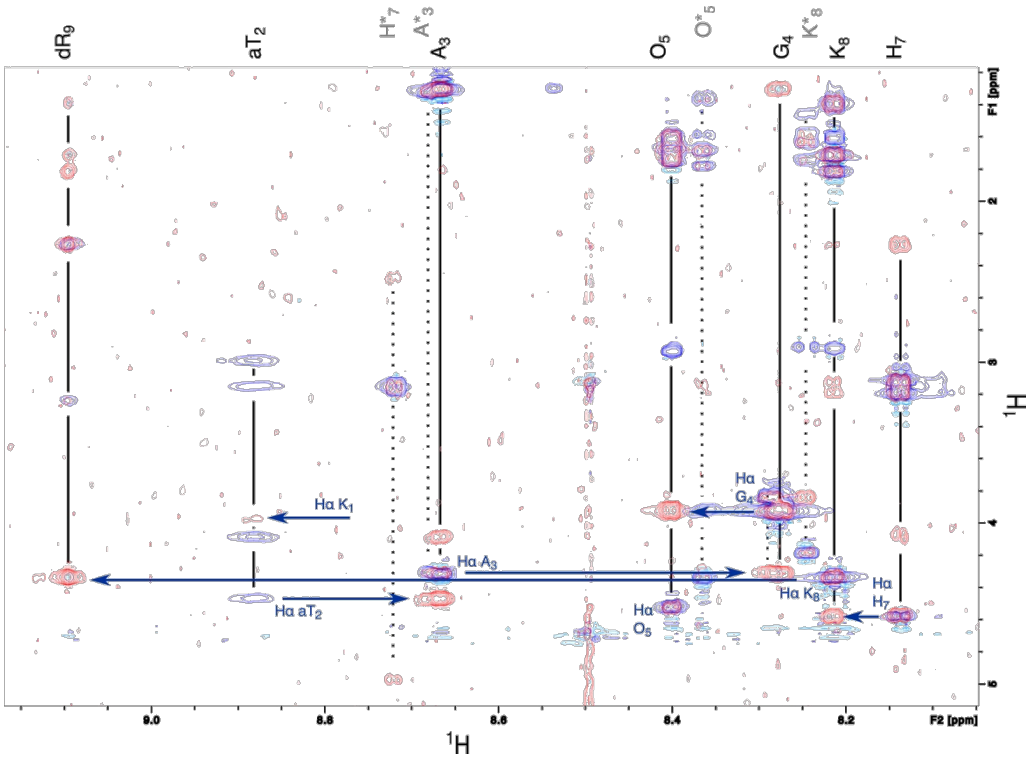


Fig. 4.4: Superposition of TOCSY and ROESY spectra for the peptide N374. The spin systems of the major form are indicated by a black bar while those of the minor form are marked with a dashed line. For the major form of the peptide, the assignment walk is drawn as dark blue lines, following the HN_X-$H\alpha_{(X-1)}$ NOEs contacts.

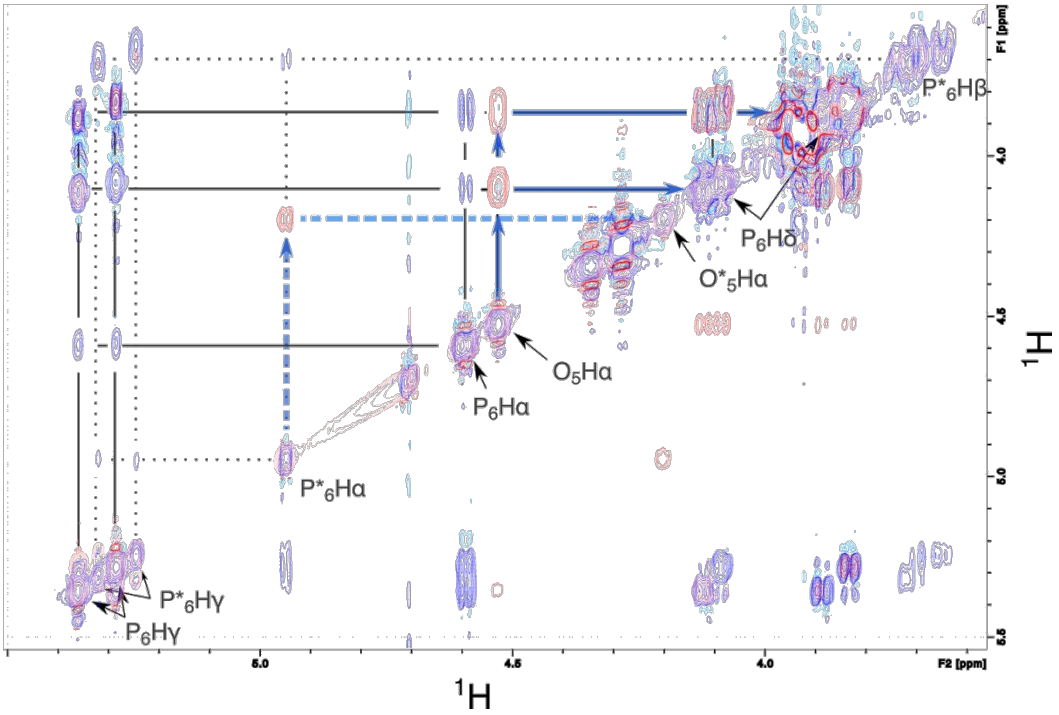


Fig. 4.5: Superposition of TOCSY and ROESY spectra for the N490 peptide. The spectra are zoomed on the region where NOE cross peaks are observable for the two conformations of the $Proline_6$. NOEs peaks can be seen between the Hα of the $Ornithine_5$ and the $H\delta_{1,2}$ of the $Proline_6$ trans and a single NOE peak is present between the Hα of the $Ornithine_{5*}$ and the Hα of the $Proline_{6*}$.

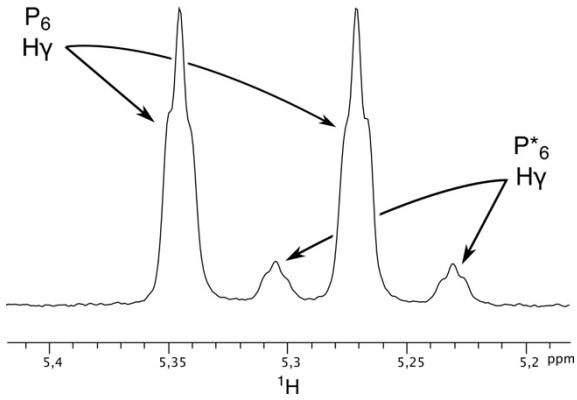


Fig. 4.6- Zoom in a ^{1}H 1D spectrum of N490 on the Hγ region of the fluoroproline. The coupling constant between the two peaks of a same form is due to the coupling with the geminal ^{19}F nuclei and worth 52 Hz.

	A_2	A_3	G_4	O_5	H_7	K_8
N374	-	0.32	0.78	0.15	-	0.93
N490	0.19	-	0.73	0.13	-	1.13
N495	-	0.31	-	-	0.08	1.35

Tab. 4.12: Ratio of the NOEs peaks intensities ($HN_i - H\alpha_i$)/(HN_i-$H\alpha_{(i-1)}$) for the peptides. For several residues, this ratio couldn't be computed since the NOEs peaks were overlapping or absent.

4.3.1.2 K_5-P_6 cis-trans conformation

The X-Pro peptide link may adopt a *cis* conformation with higher frequency than other peptide bonds due to the cyclic nature of the proline amino-acid (figure 4.7).

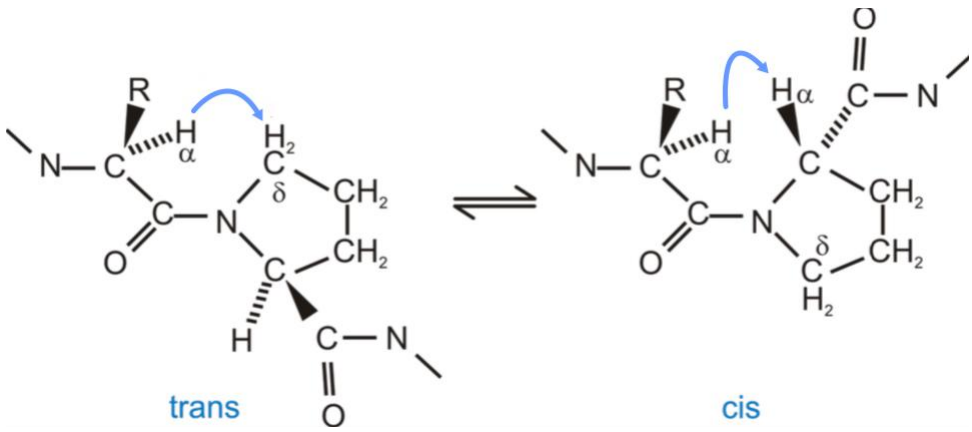


Fig. 4.7- Illustration of the prolyl cis/trans isomerization. A NOE bond of a proton from the prolyl ring with the previous residue can reveal the conformation of the proline.

These two conformations lead to specific inter-residues NOEs: while the *trans* form favors a $H_\alpha^i - H_\delta^{i+1}$, the kinked cis conformation lead to a strong $H_\alpha^i - H_\alpha^{i+1}$ NOE. Furthermore, the addition of a fluorine on the proline ring induces conformational changes of the pyrrole ring

due to the gauche effect of fluorine, and may favor *cis* or *trans* conformation, depending on the fluorine chiral position. A minor form of the Nosopharm peptides was detected in 1D (figure 4.6) and 2D spectra (figure 4.5) of the three peptides (shown with an asterisk). These additional resonances are also linked together by NOEs contacts. While the relative intensities (obtained from the ROESY) are different for the 3 peptides, they are more intense for the N490 and N374 peptides, both of them having a fluoroproline at position 6. The assignment of the minor forms is provided in table 4.13.

		A^*_3	G^*_4	O^*_5	P^*_6	H^*_7	K^*_8
N374	Hα	4.30	3.84	4.18	4.98		4.34
	Hβ	1.32		1.45-1.57	2.47	3.15	1.78
	Hγ			1.63-1.74	5.29-5.22		1.36
	Hδ			2.90	3.67-3.69		1.58-1.68
	Hε						2.92
	NH	8.68	8.28	8.24		8.72	8.36
N490	Hα	4.28	3.81	4.18	4.92		
	Hβ	1.29		1.44-1.54	2.48-2.58	3.15	
	Hγ			1.61-1.74	5.23-5.30		
	Hδ			2.90	3.63-3.70		
	Hε						
	NH	8.65	8.30	8.22		8.75	
N495	Hα	4.30	3.85	4.15		4.58	4.33
	Hβ	1.33		1.40-1.52	1.61-2.24	3.15	1.79
	Hγ			1.57-1.65	1.84-1.99		1.38
	Hδ			2.88	3.41		1.70
	Hε						
	NH	8.69	8.28	8.19		8.77	

Tab. 4.13- ^{1}H chemical shifts of the residues observed in the second form of the peptides

In the aliphatic region, a strong NOE appears between the Hα of the Ornithine 5* and the Hα of the Proline 6* (seen in figure 4.5), meaning that the secondary form is due to the isomerization of the Proline 6 in a *cis* conformation. The *cis/trans* isomerization is kinetically slow allowing distinct resonances to be observed in NMR spectra. The *cis* form is usually less frequent in proteins (Rubini-2013). The proportions of the two forms were estimated by measuring the peaks intensities of different protons. The HN of O_5 and O^*_5 was chosen for the 3 peptides, since it gave a strong signal and wasn't overlapping with other peaks. For the peptides containing a fluoroproline, the relative values of the conformations were double-

checked by measuring the intensity of the P_6 Hγ, an explicit marker of the secondary form (as seen in figure 4.5). The relative amount of *cis* and *trans* forms are resumed in table 4.14. The values indicate a stabilization of the *cis* form in the peptides N490 and N374, where the $Proline_6$ has a ^{19}F in a 4S conformation. We found that the insertion of a fluorine at the 4S position of the proline ring shifts the relative amount of cis form from 4% (N495) to 12%, an effect that was described earlier (Rubini et al. - 2013). It is worth noting that the similar biological activity measured for N495 and N374 is in agreement with the fact that proline is not directly involved in this activity.

	HN O_5	HN O^*_5	Hγ P_6	Hγ P^*_6
N374	1.291e6	2.001e5	5.515e5	8.778e4
Conformations proportion	trans 87%	cis 13%	trans 86%	cis 14%
N490	4.462e5	5.656e4	1.079e6	1.506e5
Conformations proportion	trans 88%	cis 12%	trans 87%	cis 13%
N495	1.160e6	4.652e4	-	-
Conformations proportion	trans 96%	cis 4%	- -	- -

Tab. 4.14- Absolute intensities of the HN and Hγ peak for the residues O_5, O^*_5, P_6, P^*_6 and the relative proportion of the two conformations. The values were obtained from the ROESY spectra and 1D ^1H for the fluoroprolines, with a long relaxation delay (D1=5s).

4.3.2 Interaction of Nosopharm peptides with ribosome

The interaction between the fluorinated peptides (N374 and N490) and ribosomes from *Escherichia coli* and *Thermus thermophilus* was probed using ^{19}F NMR. Both peptides encompass a fluorinated proline, however, only N374 is capable of bacterial growth inhibition. The ^{19}F 1D spectra of both peptides are shown in figure 4.8.

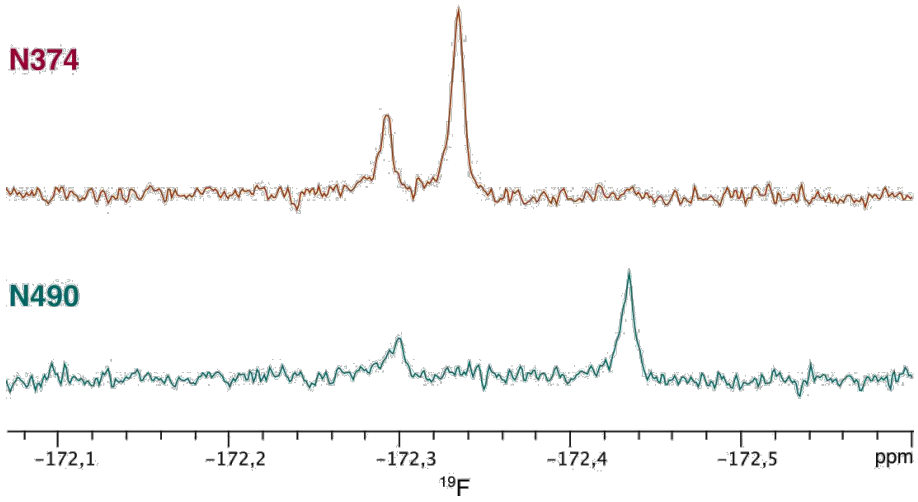


Fig. 4.8- Superposition of the ^{19}F NMR spectra of the N374 and N490 peptides, in 10% D_2O at 298K.

The position of a fluorine at position 4S of the proline results in a resonance at -172.3 ppm for the two peptides. A secondary peak corresponding to the *cis* form is clearly visible for both peptides, confirming the observation made on proton spectra. Remarkably, the amino-threonine substitution by an alanine lead to a shift of fluorine resonances of 0.1 ppm, a significant number considering that the relative site is located 3 residues away from the proline one.

4.3.2.1 Control experiments

Two controls were used to assess the binding of Nosopharm peptides to ribosomes. First, a fluorinated peptide was employed, PK240. The PK240 peptide of sequence PPP-$^{19F}P_{(2S,4R)}$-PAP-$^{19F}P_{(2S,4R)}$-PPPRVYK, held similar characteristics to N490 and N374: the presence of fluorinated prolines (hence the neighbouring chemical shifts) and positively charged amino-acids. Furthermore, this peptide contains two 4R fluoroprolines, that provides resonance frequencies at positions distinct from those of 4S fluoroprolines. Second, the possibility of non-specific binding to RNAs by electrostatic interactions was tested using polyuridylic acid. Polyuridylic acid (Poly(U)), a homopolymer of uridine (figure 4.9), is used as a single-stranded RNA model to determine if the Nosopharm peptide would non-specifically bind to RNA regions of the ribosome.

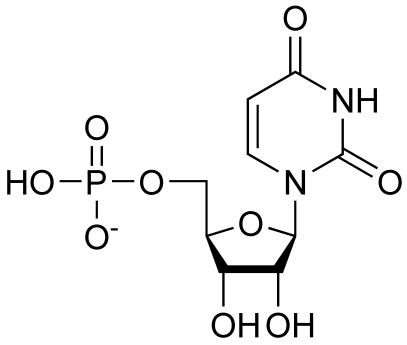


Fig 4.9 Skeletal formula of protonated uridine monophosphate, formed by a phosphate group, a pentose sugar ribose and the nucleobase uracil.

The Sigma datasheet of the product used (CAS Number: 28086-43-3) indicates that the molecular weight of the product is highly variable (from 100 kDa to 1000 kDa), meaning that the poly(U) chains will form different scales of polymers providing an approximate model of the ribosome complexity. Another interesting property of a poly(U) chain is the negative charge brought by the phosphate groups. Since the Nosopharm peptide sequence contains arginine and lysine residues, positively charged, mixing them with poly(U) could test the existence of an ionic effect. 10 nM, 100 nM and finally 400 µM of poly(U) were added to a solution of 200 µM of N490 and N374 and no alteration of the Nosopharm peptide peaks were observed, as illustrated in figure 4.10.

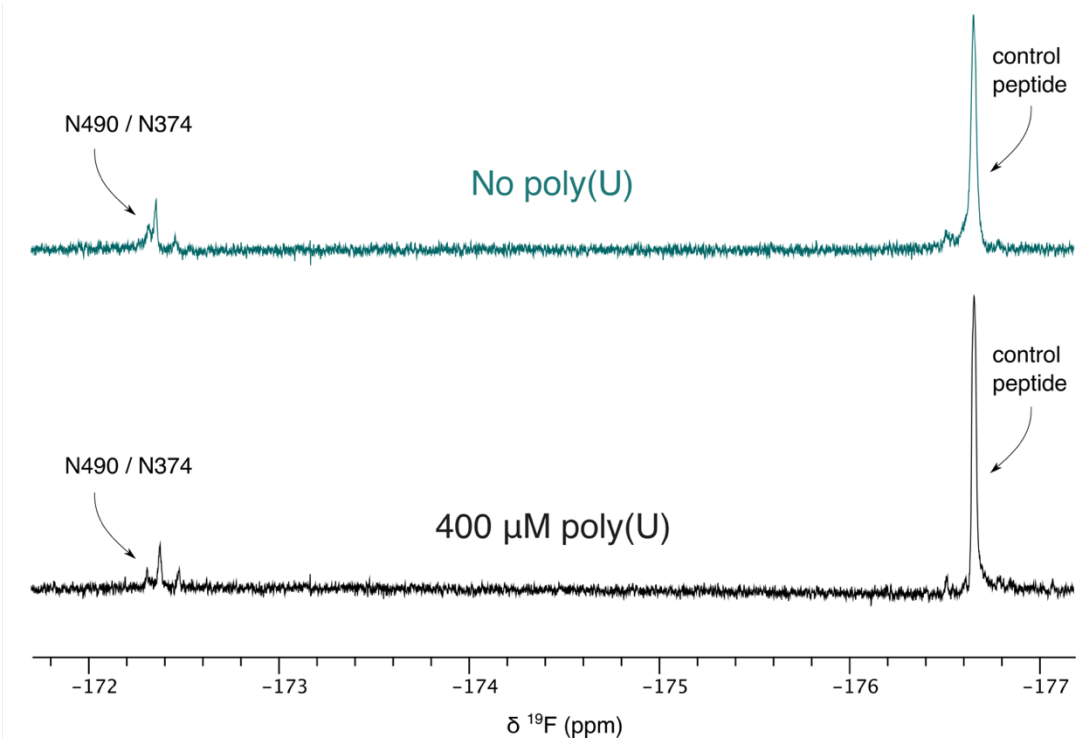


Fig 4.10 Juxtaposition of two ^{19}F (^{1}H decoupled) 1D NMR spectra obtained from a solution of 200 µM of N490 and N374 (with a control peptide) first devoid of poly(U) and then with the final step of the titration, with 400 µM of poly(U). No significant difference in the peak shapes was observed.

Since no interaction with the poly(U) chain was detected, an interaction between the fluorinated Nosopharm peptides and ribosome was subsequently set up.

4.3.2.2 Titration with ribosome

The interaction between the Nosopharm peptides N374 and N490 with the ribosome were tested under different conditions. First, small amounts of *Thermus thermophilus* 70S ribosome were added to each peptide, separately. Then a cocktail of the two peptides was mixed with the same ribosome, allowing a better comparison of the peaks broadening. The results of those screening assays are presented in the following figures:

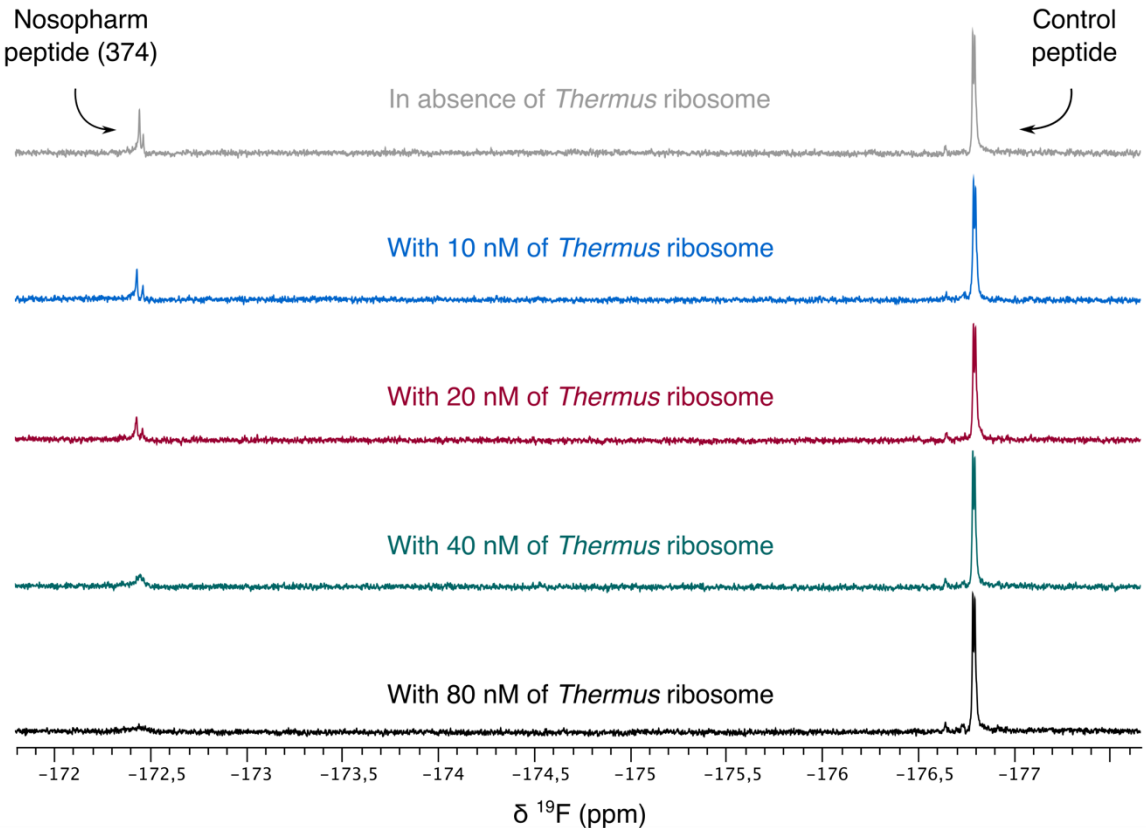


Fig 4.11 Superposition of NMR titration experiments involving two peptides at 200 µM: NOSO-374 and the control peptide PK240 with several additions of 70S *Thermus thermophilus* ribosome.

A distinct broadening of the NOSO-374 peaks can be observed, even at a concentration ratio of 1:10^4 while the resonance of the PK240 control peaks remain unchanged. Those results indicate that NOSO-374 is interacting with the ribosome. The same experiment was conducted with N490.

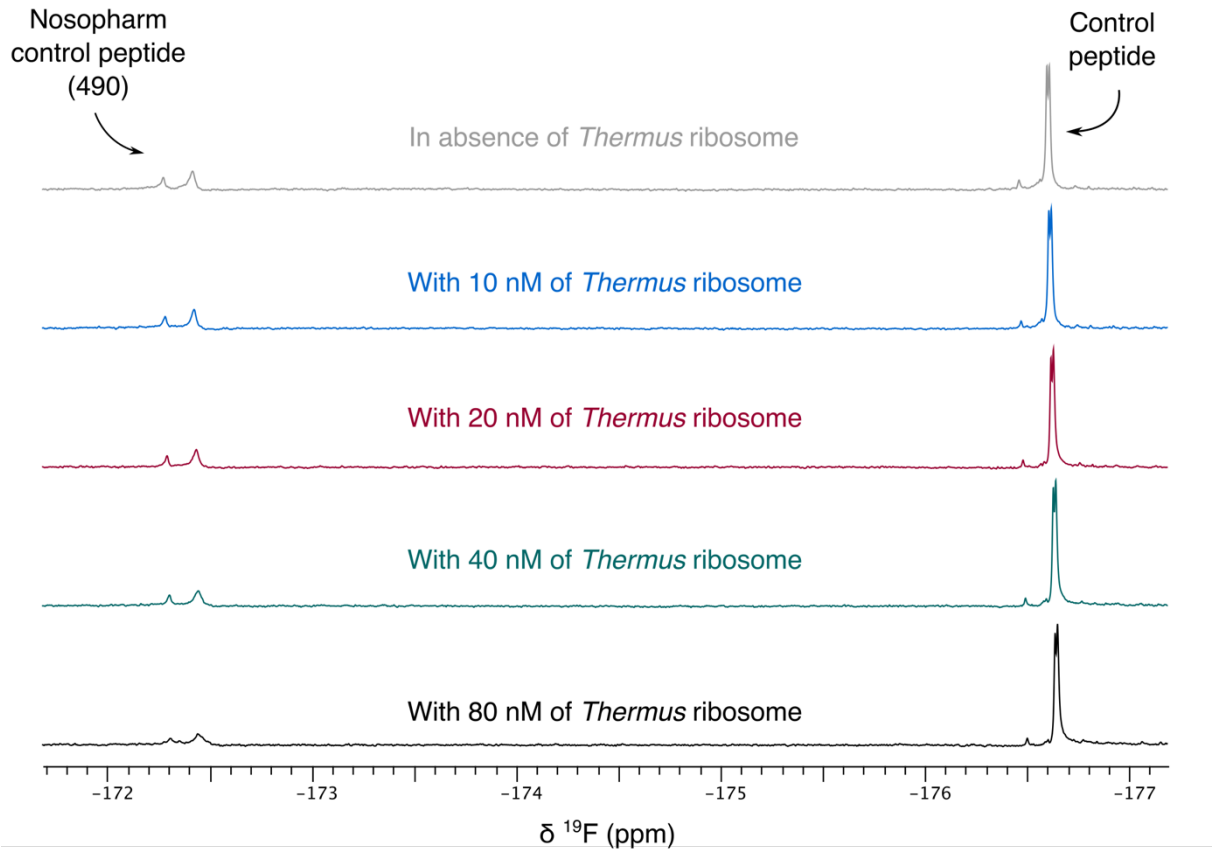


Fig 4.12 Superposition of NMR titration experiments involving two peptides at 200 µM: NOSO-490 and the control peptide PK240 with several additions of 70S *Thermus thermophilus* ribosome.

The figure 4.12 shows a broadening of N490 resonances as for N374. To better compare the two situations, the spectra can be zoomed on the peptides resonances, as illustrated in figure 4.13.

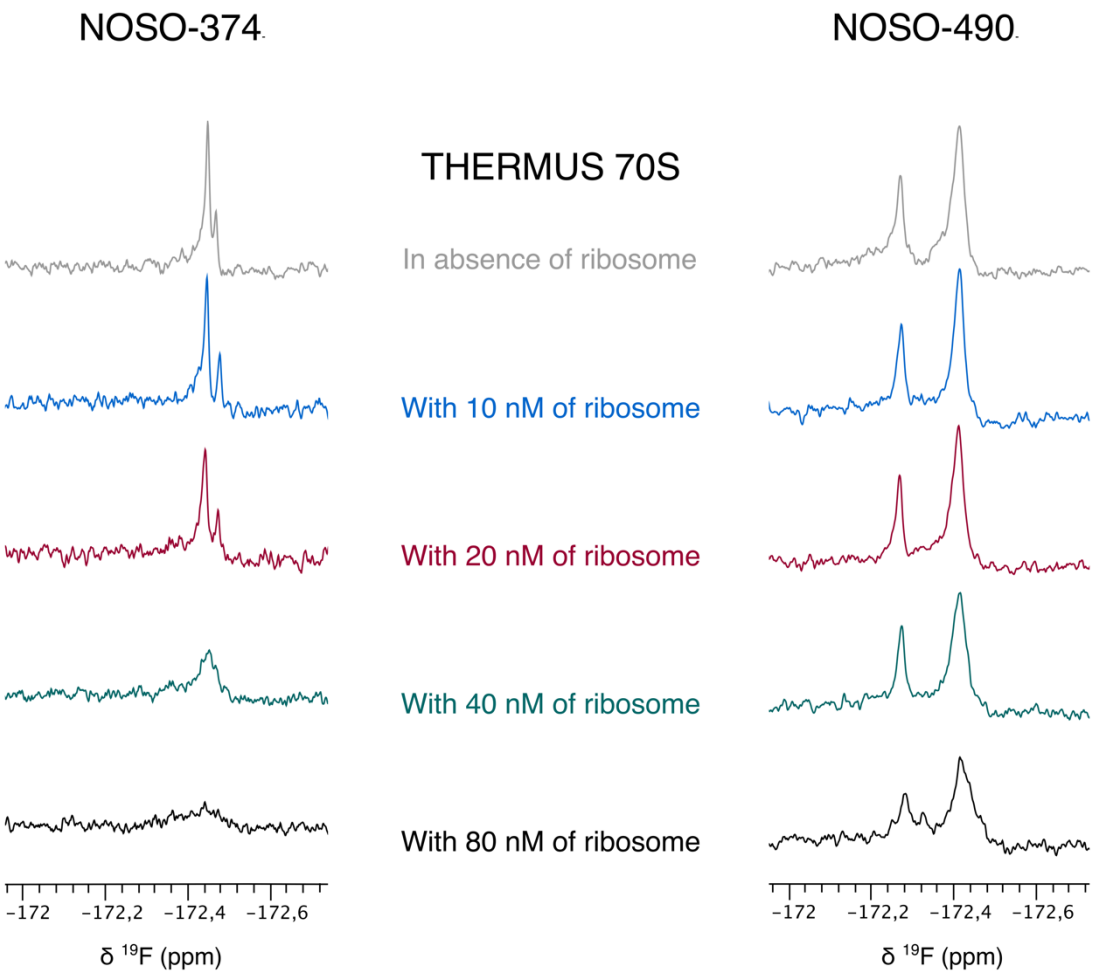


Fig 4.13 Zoom on the spectra obtained from the titration experiments described in figure 4.16 and 4.17. The peaks broadening can be better appreciated when viewed side-by-side.

It is worth noting that resonances corresponding to both *cis* and *trans* conformations are broadened upon increasing additions of ribosome, indicating that *cis-trans* equilibrium does not affect the interaction with the ribosome. Through the side-by-side comparison, it appears that the broadening of the N490 resonances, while present, is less important than for N374. In order to confirm a differential binding of the two peptides, a sample containing a mixture of both peptides was prepared.

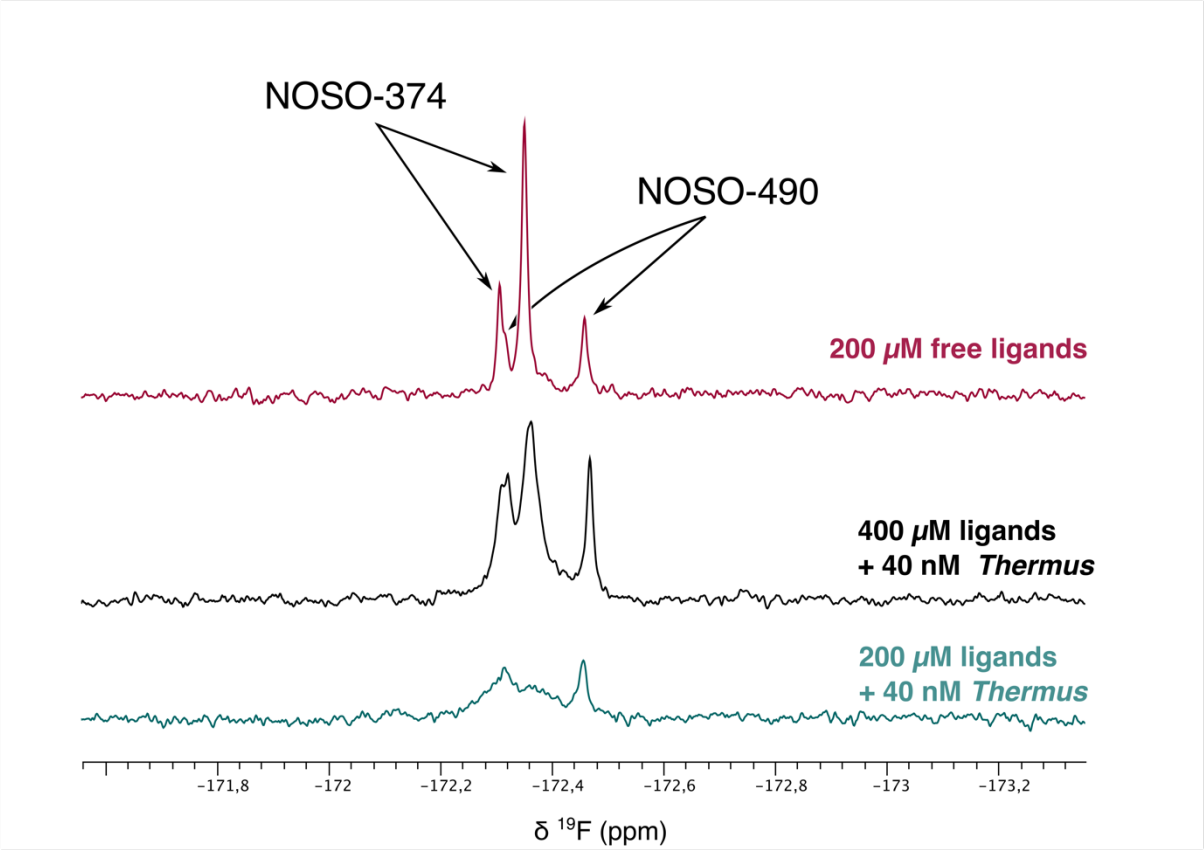


Fig 4.14 ^{19}F 1D NMR spectra of the cocktail of the ligands in different ribosomal conditions.

The experiment described in the figure 4.14 shows that N374 is more strongly impacted by the addition of ribosome than N490, although both ligands are interacting with the macromolecule. This observation correlates with the biochemical assays conducted by Nosopharm, where they found that N374 was a bacteriostatic agent but not N490. However, Nosopharm results were based on screening assays conducted on *Escherichia coli*, our NMR binding experiments used *Thermus thermophilus* as the ribosomal source as a model. The structural differences between those two prokaryotic ribosomes are small but they exist. For example, the thermophilic 16S rRNA of *Thermus thermophilus* has a denser interior than the mesophilic one from *Escherichia coli*, likely a signature of thermostability (Mallik-2013). The screening experiments were repeated with *Escherichia coli* 70S ribosome.

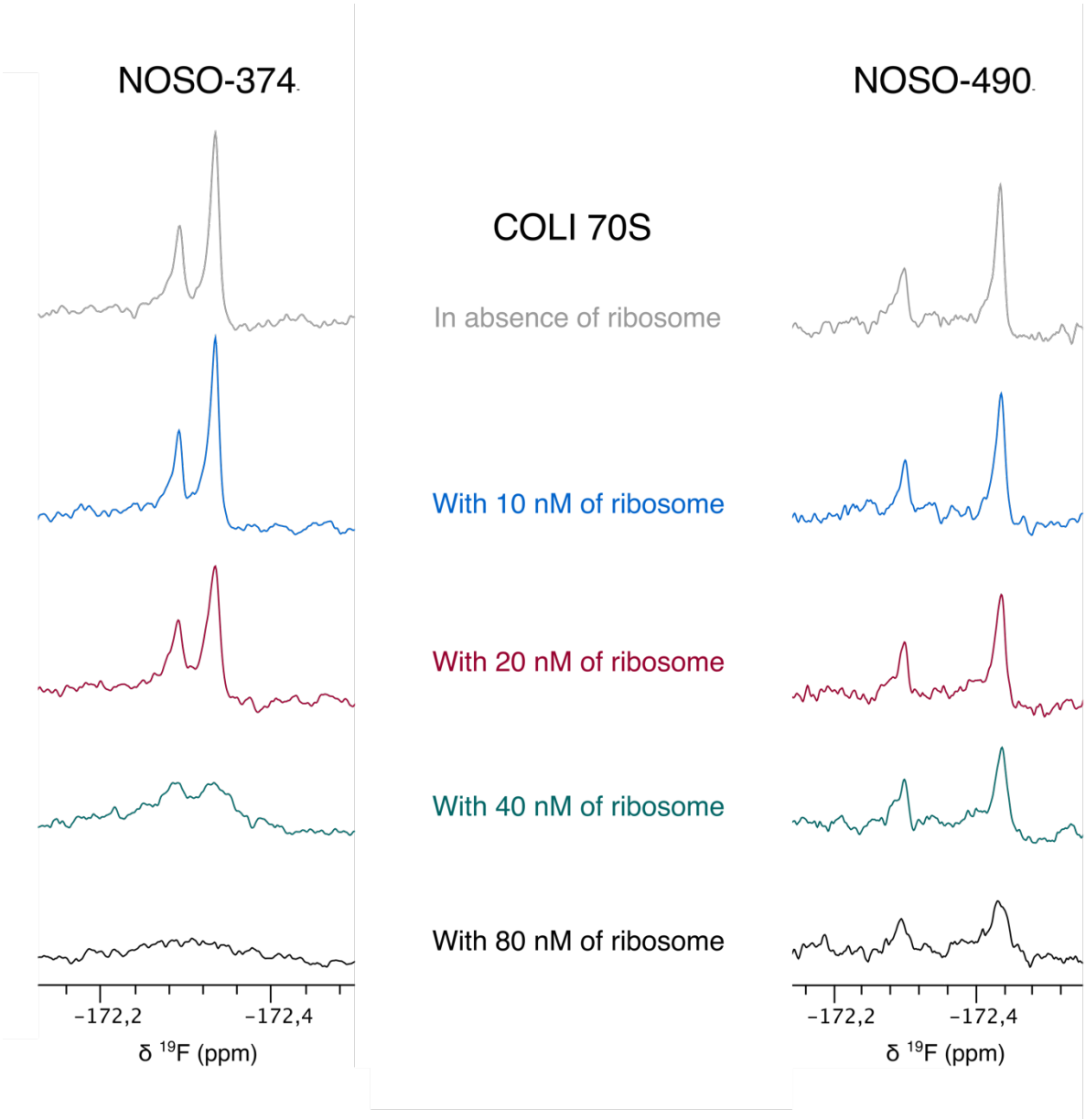


Fig 4.15 Zoom on the spectra obtained from the titration experiments described in figure 4.20 and 4.21. The peaks broadening can be better appreciated when viewed side-by-side.

The addition of ribosome to either a tube containing the N374 or N490 peptide led to the broadening of the proline fluorine resonance. However, as for the *Thermus thermophilus* ribosome, the broadening effect is more pronounced in the case of the N374 peptide.

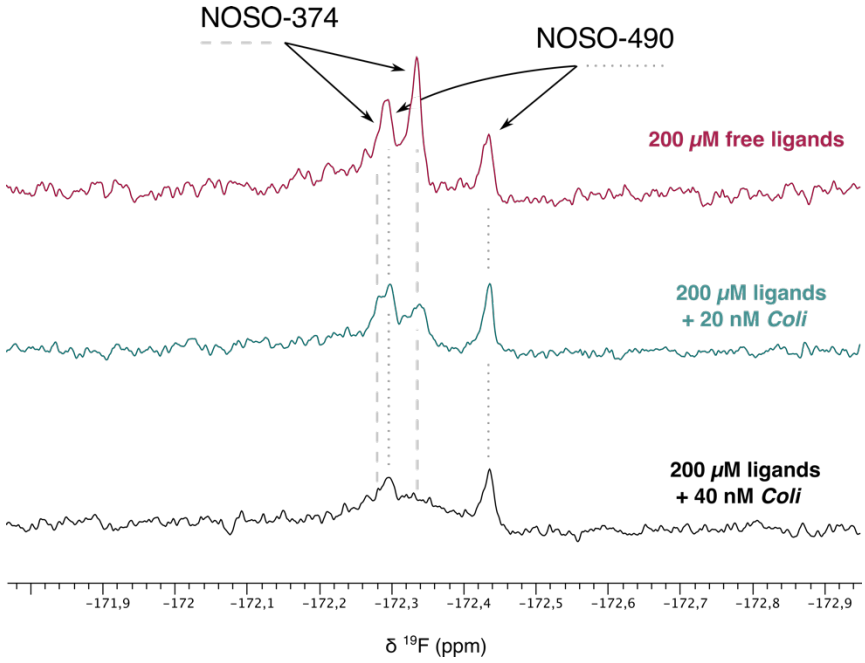


Fig 4.16 ^{19}F 1D NMR spectra of the cocktail of the ligands in different ribosomal conditions.

A cocktail of the two peptides with *Escherichia coli* ribosome was realized, and the two ligands react identically to ribosome from this bacterium or from *Thermus thermophilus*. The two peptides bind to the macromolecular target, but with a stronger impact on the ^{19}F resonances peaks for the NOSO-374 peptide. For the ^{1}H resonances, 1D experiments were also recorded but the overcrowding of peaks makes it much harder than in fluorine to conduct a binding experiment in the same concentration ranges, as illustrated in figure 4.17.

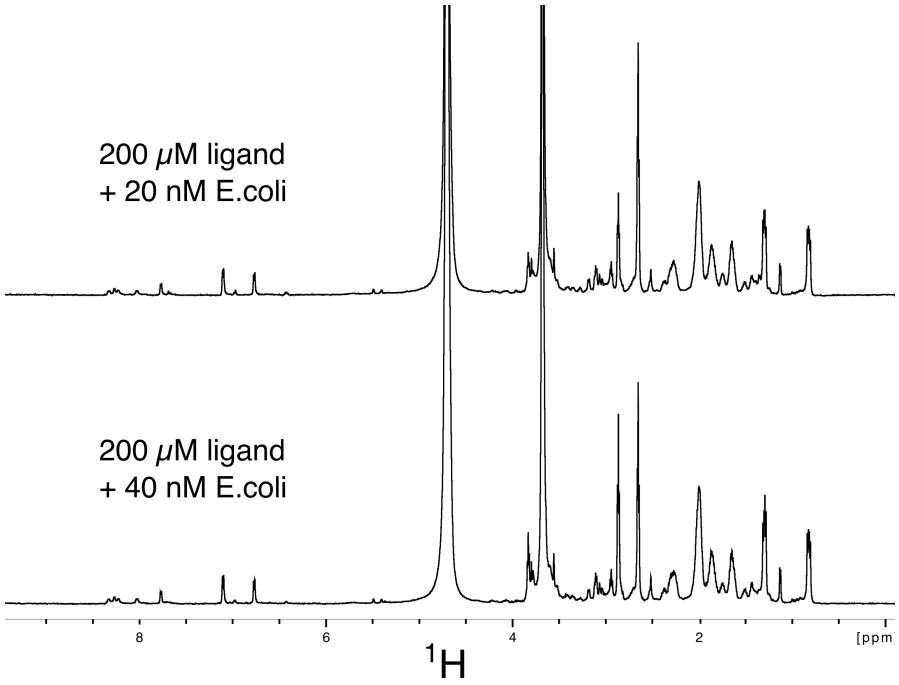


Fig 4.17 ^{1}H 1D NMR spectra of the cocktail of the ligands in two ribosomal conditions.

4.4 Conclusion

Three antimicrobial peptides developed by the Nosopharm company were studied by NMR, in order to probe their ability to bind ribosomes of *Thermus thermophilus* and *Escherichia coli*. It was found that the incorporation of a fluorine atom in the proline residue does not affect the structure of the peptides, but enables the binding studies with the ribosome. The interaction was observed to exist even with very low ribosome concentrations (in the nM range). This result may suggest a very strong affinity or a high number of binding sites. The N374 and N490 peptides, that differ at position 2, display slightly different affinities with both the ribosome of *Escherichia coli* and *Thermus thermophilus*, in agreement with their distinct biological activities. This observation suggests that the antibacterial activity of Odilhorhabdines peptides may be related to their ability to bind nucleosomes.

It should however be stressed that the competition experiments with different antibiotics had no impacts on the spectra and that the differential binding is rather weak. This could be explained by the fact the biological activity may not be directly with the ribosome, it might target associated factors.

This study demonstrated the use of fluorine NMR to detect binding to macromolecular complexes through the use of modified amino-acids. In this case, the use of fluoro-prolines proves to be particularly appropriate to detect peptide binding events.

References

Chaston JM, Suen G, Tucker SL, et al. The entomopathogenic bacterial endosymbionts Xenorhabdus and Photorhabdus: convergent lifestyles from divergent genomes. PLoS ONE. 2011;6(11):e27909.

Mallik S, Kundu S. A comparison of structural and evolutionary attributes of Escherichia coli and Thermus thermophilus small ribosomal subunits: signatures of thermal adaptation. PLoS ONE. 2013;8(8):e69898.

Ogier JC, Pagès S, Bisch G, et al. Attenuated virulence and genomic reductive evolution in the entomopathogenic bacterial symbiont species, Xenorhabdus poinarii. Genome Biol Evol. 2014;6(6):1495-513.

Pendleton JN, Gorman SP, Gilmore BF. Clinical relevance of the ESKAPE pathogens. Expert Rev Anti Infect Ther. 2013;11(3):297-308.

Rubini M, Schärer MA, Capitani G, Glockshuber R. (4R)- and (4S)-fluoroproline in the conserved cis-prolyl peptide bond of the thioredoxin fold: tertiary structure context dictates ring puckering. Chembiochem. 2013;14(9):1053-7.

CHAPTER 5

NMR study of the interaction between a proline-rich ligand from RARγ and a SH3 domain of the human Vinexinβ

5.1 Introduction

5.1.1 Regulation of transcriptional activity of the Retinoic Acid Receptor by phosphorylation

The Retinoic Acid Receptor gamma (RARγ), that belongs to the nuclear receptor superfamily, is involved in the regulation of numerous genes during development (Rochette-Egly – 2007). Interferences in the proper operation of RARγ can cause serious damages in skin and mucous membranes as well as eyesight related problems. RARγ binds retinoic acid through its Ligand Binding Domain (LBD) as well as a specific sequence of DNA through its DNA Binding Domain (DBD) (see figure 5.1). In addition to the activation of transcripts through RA binding to RARγ, it was shown that phosphorylation of a number of amino-acids at the receptor surface is also involved in the regulation of the transcription. This phosphorylation cascade leads ultimately to the phosphorylation of a serine residue located in a proline-rich region of the N-terminal Domain (NTD), located close to the DBD. The team of Rochette-Egly identified the protein vinexinβ as a co-regulator involved in the phospho-dependent regulation of gene transcripts by RARγ.

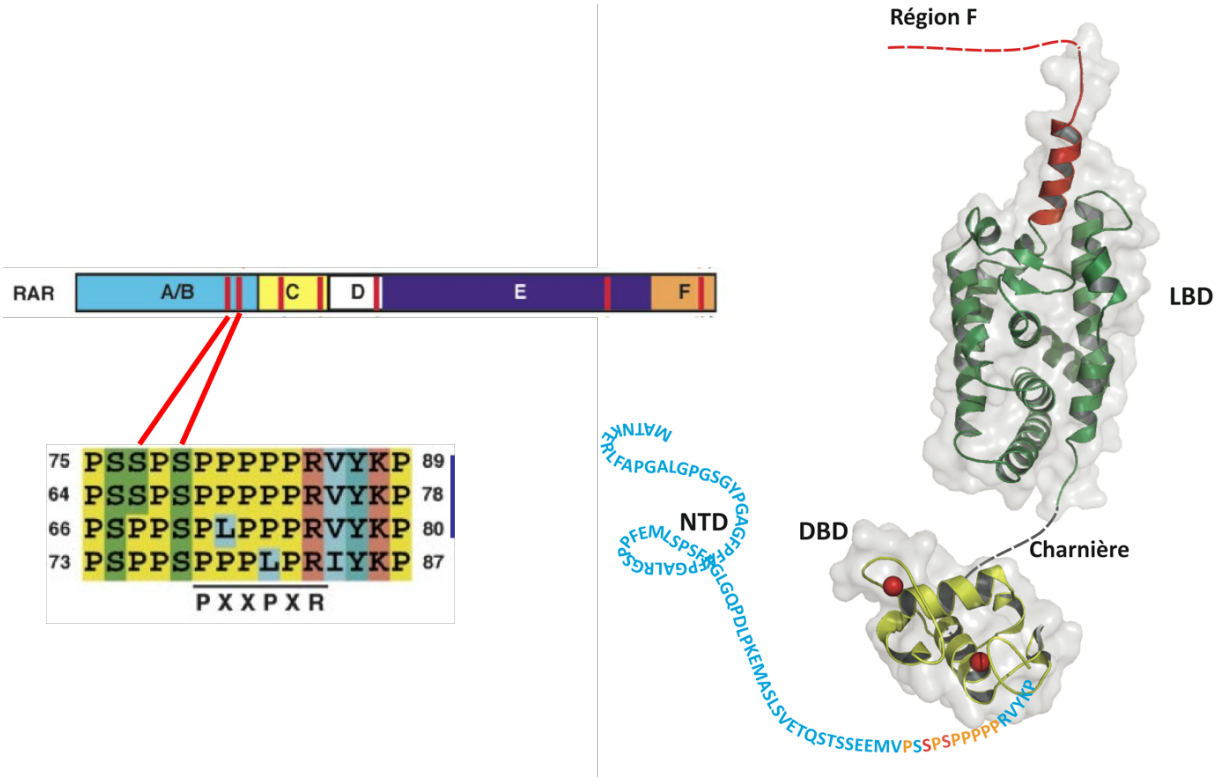


Fig 5.1. Structural organization of the RARγ nuclear receptor (credits to C.Rochette-Egly).

The vinexinβ contains three SH3 domains and the third one binds to the proline rich domain (PRD) of RARγ with an affinity that depends on the serine phosphorylation state. NMR experiments, used to measure the affinity between the vinexinβ SH3 domain and RARγ peptides from the polyproline region, showed that the serine phosphorylation decreases the affinity by a factor of 4 (K_d~120 µM to 450 µM) (Lalevée et al,.-2010). The 3D structure of the complex between the SH3 domain and these peptides was solved by NMR and did not allow to understand the molecular basis of this affinity modulation process. Instead, it was found that the phosphorylated form of the peptide should be favoured by electrostatic interactions.

In order to further our understanding, proline rich model peptides were designed to investigate further the binding mechanism. These model peptides of sequence P_5SP_5RVYK are able to bind the vinexinβ SH3 domain and this binding is modulated by the phosphorylation in a similar way than for wild-type sequences. After an introduction on the structural properties of proline residues, I will present the binding studies of the fluoro-peptides.

5.1.2 Structural properties of polyproline peptides

Prolines are peculiar amino-acids since their side chain form a cycle including the nitrogen atom of the peptidic bound. Due to this structural property, the conformations are limited and the dihedral angle phi is restricted to values around -65°, impacting then the previous amino-acid in the sequence (Kay – 2000). Prolines can also form a *cis* or *trans* bond with the preceeding residue whereas most of the amino-acids form nearly exclusively *trans* peptidic bonds. Sequences of polyproline can adopt two kinds of secondary structure based on their peptidic bond conformation: polyprolines helix type I (PPI) are compact and based on *cis* conformations while polyprolines helix type II, more common, are based on *trans* conformations for exactly 3 residues per turn. PPII helixes are often found in protein parts and play a role in the folding of those (Chiang – 2009). There is an inner rigidity in the PPII helix and this property was used once as “molecular rulers” to measure atomic distances ranging from 10 to 60 angströms (Stryer – 1967). Despite the PPII rigidity, prolines are dynamic residues due to the possible isomerization of their peptidic bond, but also due to the multiples conformations that can be adopted by its ring. The conformations are distributed between two preferential states named “up” (or Cγ-endo) and “down” (or Cγ-

exo). The free energy difference between the two conformations can be explained by the fact that, in the plane conformation of the cycle, the angle Cγ-Cδ-N is pushed toward a non-optimal value (Ho – 2005). The proline peptidic bound isomerization plays also a primordial part in proteins folding. The figure 5.2 illustrates the atomic structure of a proline as well as the carbon numeration of its pyrrolidine ring while the figure 5.3 adds the different conformations and the figure 5.4 shows the two canonical polyproline helices.

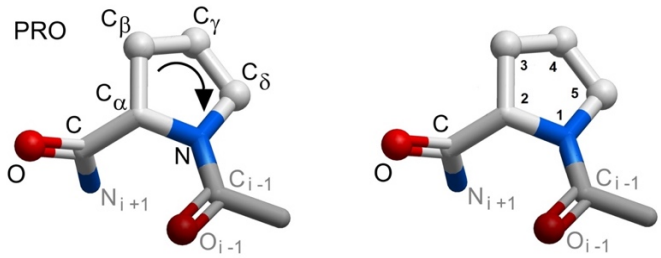


Fig 5.2 Proline atomic structure with the carbon annotations.

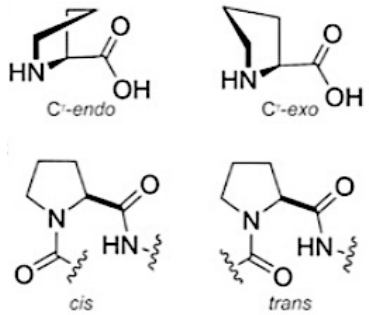


Fig 5.3 Conformational attributes of proline residues, the pyrrolidine ring puckers and the peptide bond isomers (Newberry – 2016).

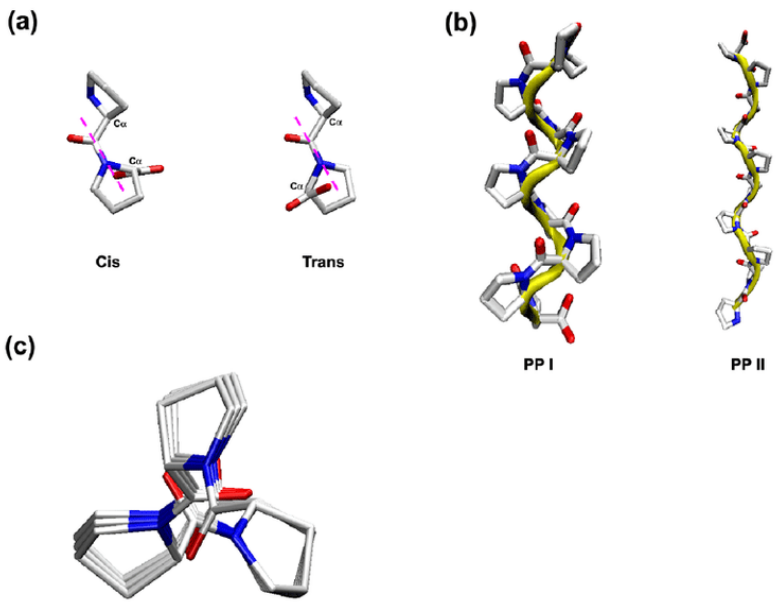


Fig 5.4 Structure of (a) the cis and trans proline residues; (b) the PPI and PPII helices; (c) the PPII helix viewed along the helical axis (Wei - GNU GPL license - 2016)

The *cis* and *trans* forms of the peptidic bound for a proline are nearly iso-energetic but the energetic barrier is elevated, effectively limiting the conversion rate between the two forms. The proline *cis-trans* isomerization process is slow, on the minute to hour scale, compared to a protein folding, meaning that the *cis* or *trans* conformation of a peptidic bond with a proline is already fixed once the protein folds. Since a proline synthetized in a protein is in a *trans* conformation, a protein that may need a *cis* proline somewhere in its structure can have a delayed folding (Wedemeyer - 2002). Some enzymes are specialized in lowering the energetic barrier between the conformations and thus speeding the *cis-trans* conversion, they are called peptidyl prolyl cis-trans isomerases. Proline rich sequences can interact with specific protein regions, including the SH3 domains (Src Homology 3) that are formed by 50 to 70 residues and often found in eukaryotic cytoskeletal proteins or signal transduction proteins. Those domains interact with the Pro-x-x-Pro pattern where x can be any residue. Proline rich sequences often contain a serine or a threonine that can be phosphorylated and thus modulating a binding interaction. For example, it was shown that the phosphorylation of a Pro-x-x-Pro pattern in the WASP protein (Wiscott-Aldrich Syndrome Protein) hindered this region interaction with a SH3 domain of an associated protein (Kay – 2000).

5.1.3 Using fluoroprolines to study the RARγ PRD and vinexinβ $SH3_3$ interaction

4-fluoroprolines are commercially available amino-acids that have been used to explore a number of molecular mechanisms involving proline residues. For instance, fluoroprolines have been used to decipher the molecular mechanism of the formation of amyloid fibers by the β2-microglobulin protein (Li et al. - 2013). The insertion of a fluorine atom at position 4 of the pyrrolidine ring induces a strong ring conformation, due to the gauche effect of the fluorine atom depending on the chirality of the position. A 4R-fluoroproline will have a *trans* peptidic bound and a "up" conformation while a 4S-fluoroproline will exit a *cis* peptidic bound and be in a "down" conformation (Koskinen-2005). In addition to its conformational effect, fluorine may be used to observe the different states of the peptide by NMR. A new family of P_5XP_5RVYK peptide were therefore generated with fluoroprolines at position 4 and 8.

The interaction of these peptides with the $SH3_3$ of vinexinβ was monitored in NMR, through ^{1}H-^{15}N HSQCs for protein based experiments and ^{19}F spectra for ligand based observations. A methodological publication partially derived from the HSQCs recorded was written and published in 2015. This paper is presented here as the final part of this chapter.

5.2 Results

5.2.1 Fluorinated NTD peptides

Five peptides were synthetized, based on the canonical sequence found in the NTD domain, PPPPPSPPPPPRVYK. The prolines 4 and 8 of all peptides were fluorinated on their gamma carbon position, in (4S) or (4R) conformations based on the fluorine position. The table 5.1 list the specificities of the different peptides used, the figure 5.5 shows the full unfolded structure of the peptides and the figure 5.6 display the peptide in interaction with the domain:

Name	6^{th} residue	CIP conformation of the fluorine found in prolines 4 and 8
PL105	Serine	4S
PL106	Phosphorylated serine	4S
PL107	Serine	4R
PL108	Phosphorylated serine	4R
PK240	Alanine	4R

Tab 5.1 Name, residue type and CIP conformation specific to each peptide

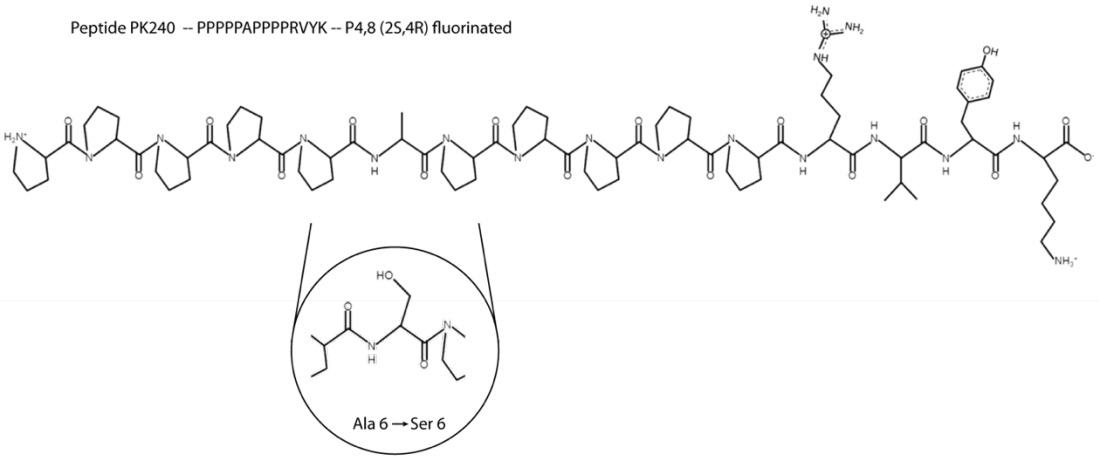


Fig 5.5 Skeletal structure of the PK240 peptide, with an illustration of the replacement of an alanine by a serine.

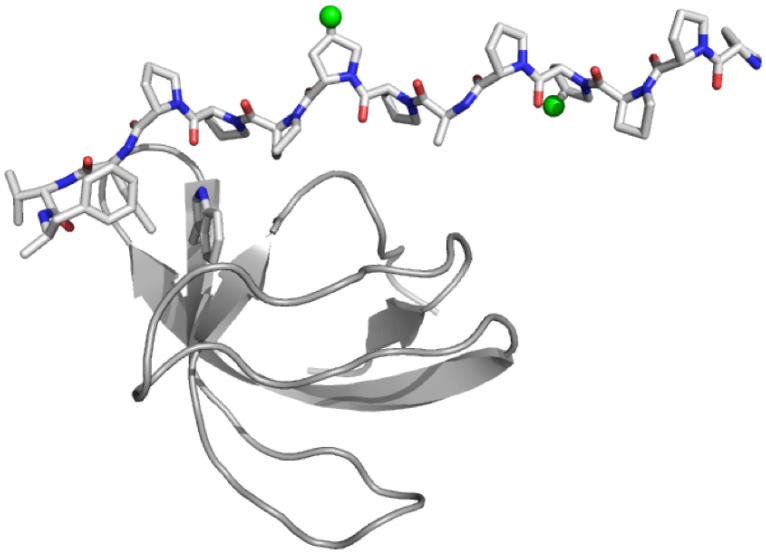


Fig 5.6 Pymol representation of the interaction between the PK240 peptide and the $SH3_3$ domain

The figure 5.5 illustrates how the sixth residue is situated in a repetitive region. The figure 5.6 shows that the position of fluorination for the prolines were chosen so they would be away from the $SH3_3$ domain and facing the solvent, in order to not interfere with the binding. The peptides share a very similar structure and that is reflected in their proton spectra. The proton spectra and fluorine spectra of the peptides is shown in figures 5.6 and 5.7. For the prolines 4 and 8, their environment is identical at n+2 residues, and it is only in the ^{19}F NMR spectra that all the variations of the peptides can be observed.

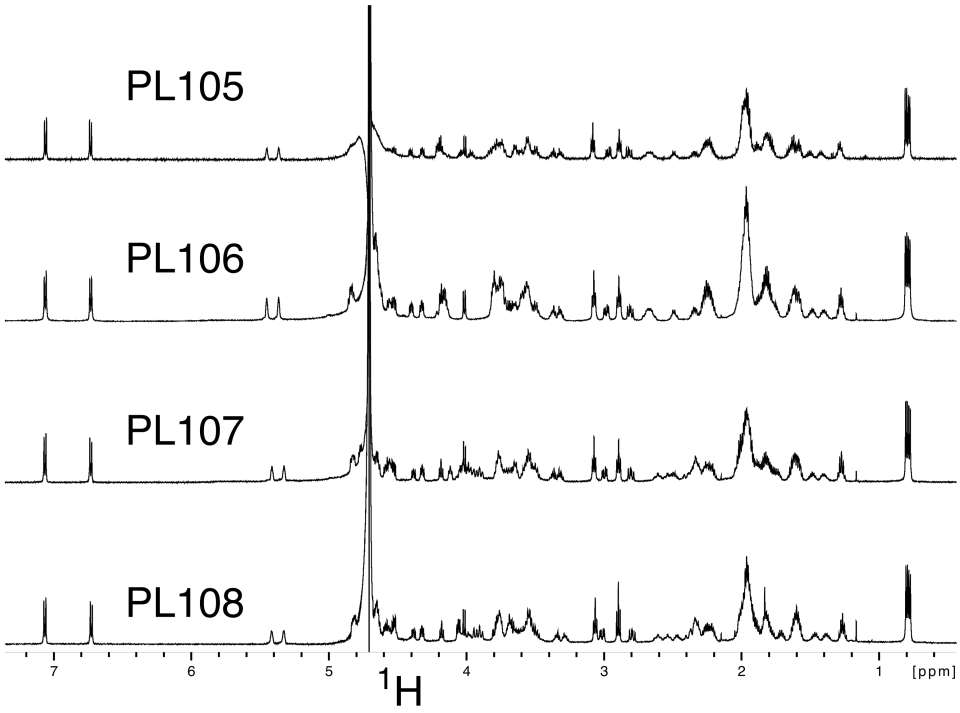


Fig 5.6 Superposition of ^{1}H 1D NMR spectra of 4 peptides PPPPSPPPPRVYK, the serine being phosphorylated or not. The P4 and P8 in the sequence have a ^{19}F atom in position 4 in either a R or S conformation. Spectra recorded at 298K, in a 11.7T magnet with 3 mm tubes in D_2O.

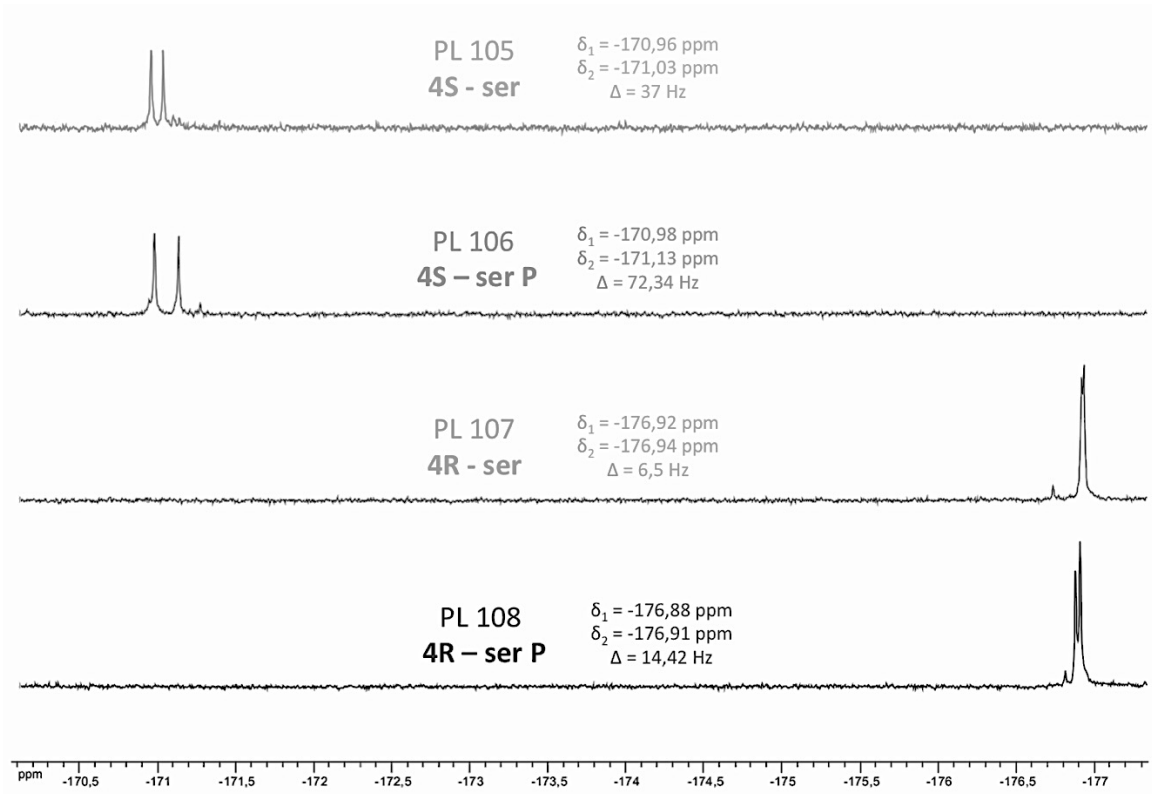


Fig 5.7 Superposition of ^{19}F 1D NMR spectra of 4 peptides PPPPSPPPPRVYK, the serine being phosphorylated or not. The P4 and P8 in the sequence have a ^{19}F atom in position 4 in either a R or S conformation. Spectra recorded at 298K, in a 11.7T magnet with 3 mm tubes in D_2O.

The fluorine conformation in the pyrrolidine ring lead to a major 6 ppm difference between the 4R or 4S peptides, respectively around -177 ppm and -171 ppm. The fluorine moieties are sensitive to the phosphorylation of the serine residue as the two fluorinated proline resonances can be distinguished and the peak gap is higher (roughly twice wider, in Hz in the figure) in presence of a serine phosphorylated. If the peptides would adopt different structural conformations, it could lead to differential tumbling rates of the peptides and thus impact their relaxation properties. T_2 and T_1 determination experiments were then conducted on the peptide through ^{19}F NMR, resulting in similar values for the peptides: around 450 ms of T_1 and 50 ms of T_2. Those values seem to indicate that the spatial folding of the four peptides must be quite the same.

5.2.2 Binding impact on the $SH3_3$ domain of Vinexinβ

In order to evaluate the role of the serine and its phosphorylation in the recognition mechanism with the $SH3_3$ domain of vinexinβ, the binding affinity of different non-fluorinated peptides with the sequence *PPPPPxPPPPPRVYK* was obtained from ^{1}H-^{15}N HSQCs, following the titration protocol published and presented in this chapter. The resulting K_d values are presented in table 5.2.

Residue x	**Affinity (in µM)**
Proline	33 ±3
Alanine	45 ±4
Valine	53 ±3
Serine	120 ±11
Phosphorylated Serine	408 ±27

Tab 5.2 List of calculated affinities of different P_5xP_5RVYK peptides for the $SH3_3$ domain of vinexinβ.

While it appears that serine is not the residue granting the strongest affinity to the interaction, the K_d difference with the phosphorylated form may allow for a fine tuning mechanism, switching between two affinity scales. The phosphorylation of the serine induces a weaker binding, by a factor of four. The 4 peptides presented in the first part of this chapter were used in titrations of the $SH3_3$ domain, and the chemical shift displacement observed was used to compute their respective binding affinity constant K_d. A ^{1}H-^{15}N HSQC of the apo version of the $SH3_3$ domain of vinexinβ is presented in figure 5.8. The peaks in the HSQC of the domain are well dispersed, indicating that the domain is completely folded. The resulting K_d obtained for the fluorinated peptides, as well as the titration fit curves, are displayed in figure 5.9. For this figure, the experimental Chemical Shift Displacement (CSD) for each titration step is marked as a black dot. The grey cloud of point and the grey line are extracted from an in-house Python script used to fit the data with a Monte-Carlo estimation of errors and to extract a K_d value. The second line of graphs use a semi-logarithmic representation to emphasize on the first titration points.

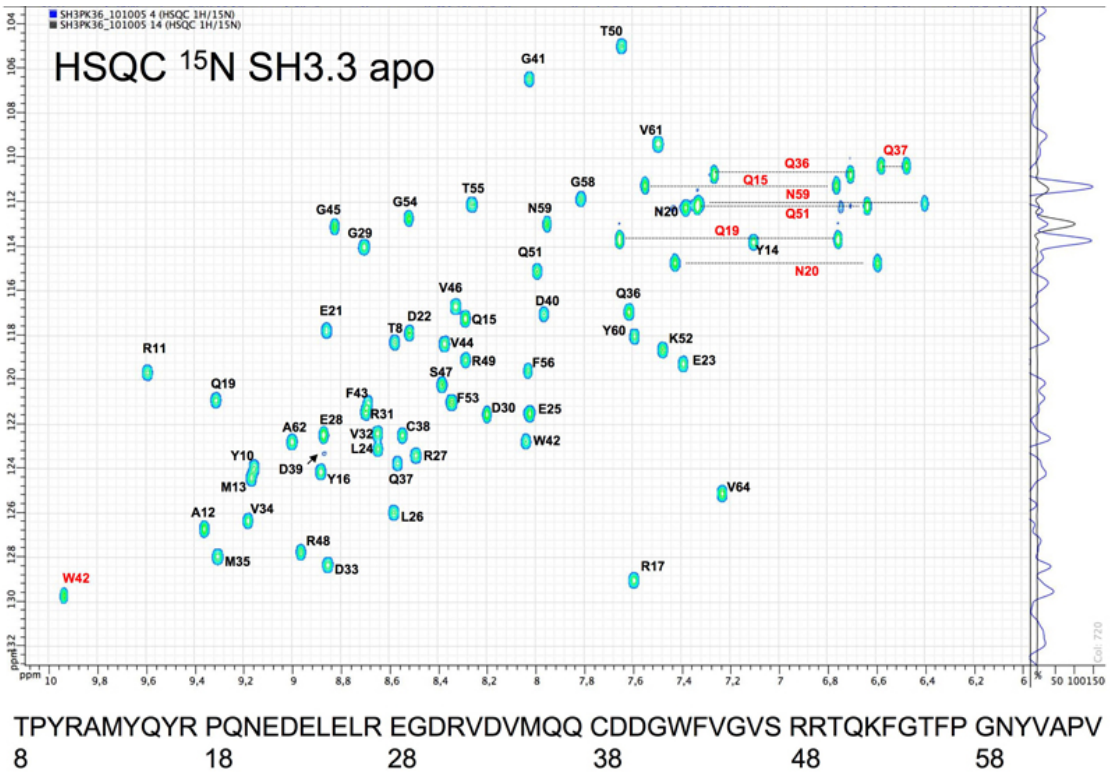


Fig 5.8 HSQC ^{1}H-^{15}N of the apo $SH3_3$ domain of vinexinβ. Recorded on a 700MHz spectrometer, at 298K, in a phosphate buffer 40 mM, pH 7.0 + DTT.

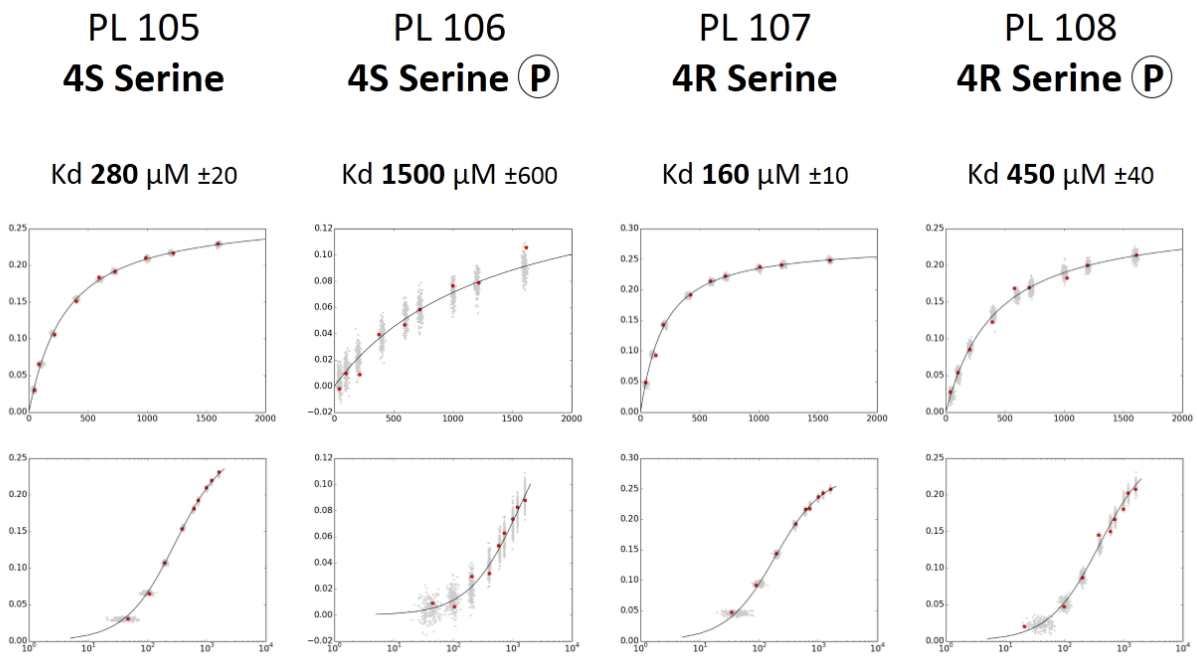


Fig 5.9 Titration curves of the fluorinated peptides on the SH3 domain. The peptides concentrations ranged from 40 to 1600 µM while the SH3 domain concentration was kept at 40 µM. The experiments ran on a 700 MHz, at 298K.

The recorded affinities are in agreement with the previous data: the phosphorylation of the serine lowers the affinity of the peptide for the domain. It is interesting to see that each titration curve is different and that in the case of PL106 (phosphorylated serine with a 4S fluoroproline), the domain is far from being totally bound by the peptide. Indeed, in this case, the K_d value is really high, more than the milli-molar. However, for the peptides with the 4R fluorinated prolines, the K_d values with and without a phosphorylated serine are in the same range than for the non-fluorinated peptide (resumed in table 5.3). This mean that the introduction of fluorine does not impact peptide binding.

	P_5SP_5RVYK (4R)	$PPP^{19}FPPSP^{19F}PPPPRVYK$ (4R)
Non-phosphorylated serine	K_d= 120 ±11	K_d= 160 ±10
phosphorylated serine	K_d= 408 ±27	K_d= 450 ±40

Tab 5.3 Resume of the affinities values observed for the peptides in a 4R conformation, with or without fluorine in their pyrrolidine ring.

However, a different conformation of the polyprolines will influence the binding since with 4S fluoroprolines, that will force the proline *cis* form, the binding affinities values are sensibly different. In order to observe the structural perturbations on the domain upon peptide binding, the chemical shift differences observed between the holo form of the $SH3_3$ domain (for the last peptide addition) and its apo form in the ^{1}H-^{15}N HSQCs were plotted alongside the sequence, for each peptide, in figure 5.10.

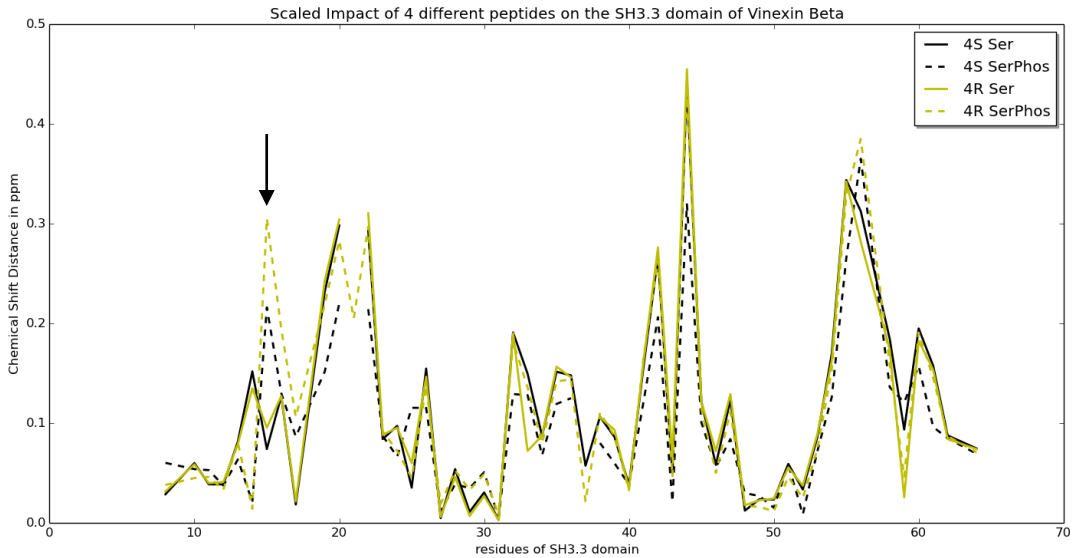


Fig 5.10 Superposition of the Chemical Shift Differences observed on the $SH3_3$ domain between the apo form and the holo form at the last peptide addition of the titration.

The CSD graph shows a difference around the 15th residue (marked by an arrow) of the domain between the phosphorylated and non-phosphorylated peptides. This 15th residue is a glutamine. There is no strong CSD difference between the 4S and 4R versions of the peptides.

Those data correlates with the K_d values observed, the phosphorylation of the serine has a measurable impact on the binding event. The CSD values were mapped on the $SH3_3$ domain surface to see how the CSD difference translate in structural terms. The computed map of the CSD on the domain is shown in figure 5.11, following a yellow to red gradient proportional to the CSD value.

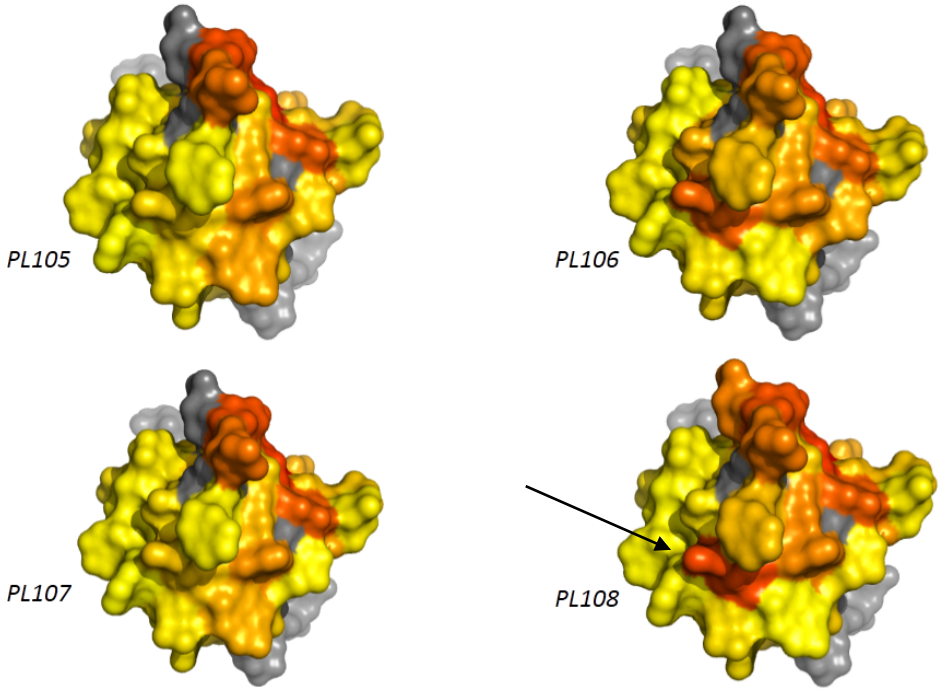


Fig 5.11 Mapping of the Chemical Shift Differences on the $SH3_3$ domain surface (PDB id: 2CT3) (PL105: 4S-Ser, PL106: 4S-phospho-Ser, PL107: 4R-Ser, PL108: 4R-phospho-Ser).

A red spot is common to only the phosphorylated peptides, PL106 and 108, and shown by an arrow on the PL108 results. The red spot corresponds to the Q15 residue. This 3D representation shows that the binding imprint does not change with a different fluorine conformation but rather with the serine phosphorylation.

5.2.3 The binding event through ligand-observed ^{19}F NMR

The interaction between the $SH3_3$ domain and the fluorinated peptides can be monitored by observing the ^{19}F nuclei. There are several advantages to this method, the first one being that isotopic labelling of the protein is not necessary. Then, the sensibility of the ^{19}F NMR experiments and the ease of reading of the spectra allows the monitoring of the peaks broadening. As explained in the chapter 1 of this thesis, from the peak shapes can be extracted kinetic information (k_{on}, k_{off}) complementing the thermodynamic equilibrium information obtained from the HSQCs, the K_d. The ^{19}F NMR 1D spectra obtained during the titration of the PK240 peptide (PPP-$^{19F}P_{(2R,4R)}$-PAP-$^{19F}P_{(2R,4R)}$-PPPRVYK) with increased concentrations of $SH3_3$ domain is presented in figure 5.12.

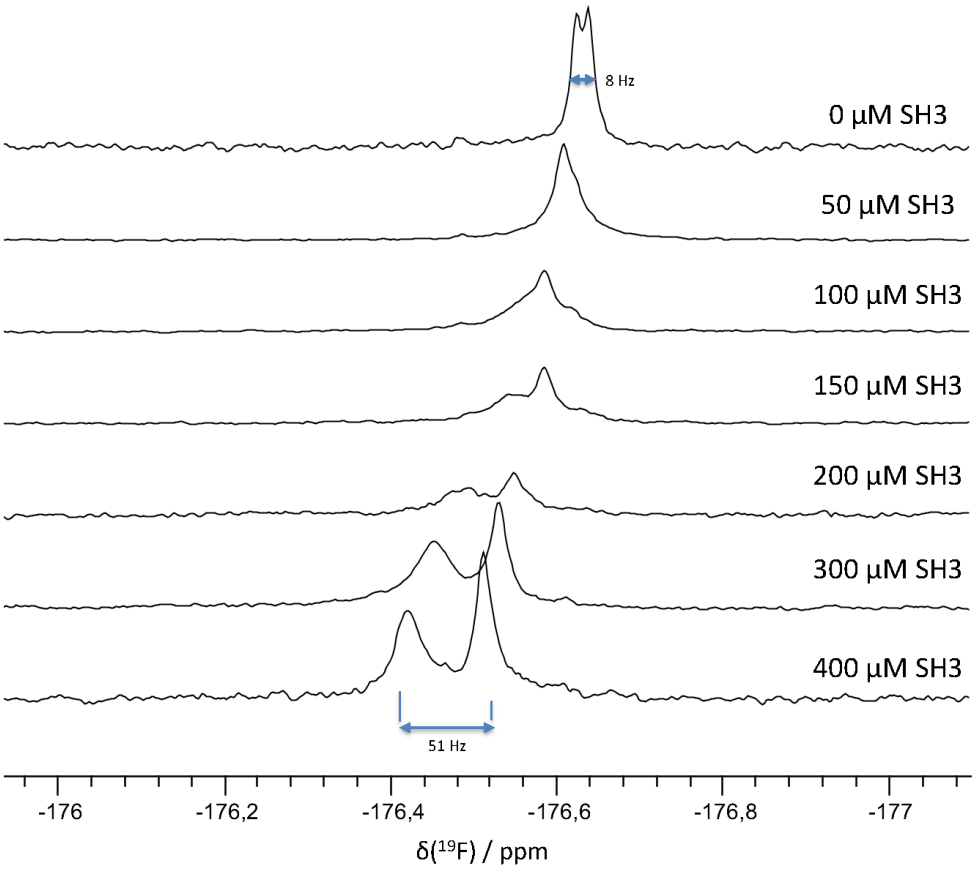


Fig 5.12 ^{19}F 1D NMR spectra of the PK240 peptide (at 200 µM) comporting two fluorinated prolines in different conditions of $SH3_3$ domain concentrations. The spectra were recorded on a 14.1T magnet, with a BBFO probe, at 298K. The spectra used proton decupling, and were recorded in 4096 scans, in a protein buffer (phosphate buffer 40 mM, pH 7.0 + DTT).

The fluorine spectra of the PK240 peptide shows that the two fluorinated peaks can be distinguished, with only an 8 Hz difference between them, due to the similar local environment of both 4R-FPro residues. Upon the addition of $SH3_3$ domain, the line width

changes report on the binding dynamics. Around 200 µM, the peaks are both very broads. The degree of fluorine chemical shift reports on molecular changes happening upon binding. While it is not possible to attribute the resonances, it appears that one of the fluorine is more sensitive to the domain addition, it could be the fluorine closer to the domain in the sequence (P_6). The analysis of the chemical shift displacement as a function of protein concentration allowed the extraction of a K_d of 50µM, in agreement with what was recorded for a non-fluorinated version of the peptide and obtained through HSQCs. However, the K_d value is not sufficient to describe the peak broadening. NMR is sensitive to the affinity (K_d) and the kinetic of an interaction (k_{on} and k_{off}). The result of Python simulations describing the interaction of a ligand and its target with a same K_d but with different k_{on} and k_{off} values is displayed in figure 5.13.

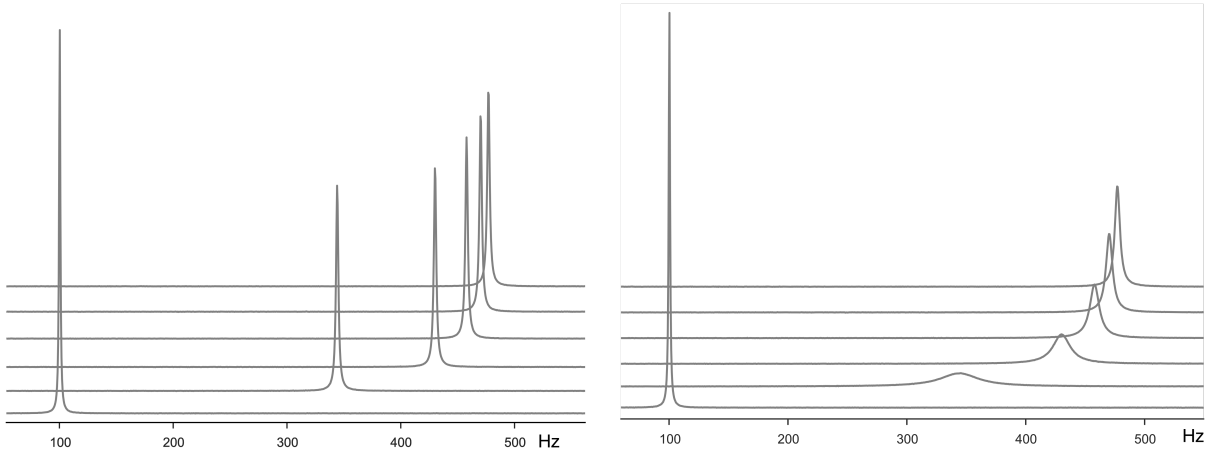


Fig 5.13 Python simulations of 6 spectra of a ligand interacting with a target (with a ligand concentration of 200 µM, a receptor concentration ranging between 0, 200, 400, 600, 800 and 1000 µM, a $R_{2,free}$ of 4 s^{-1}, a $R_{2,bound}$ of 8 s^{-1}, a free ligand resonance set at 100 Hz while the bound resonance was fixed at 500 Hz). In both cases, the K_d was kept at 50 µM but in the left example the k_{on} was 10^8 $M^{-1}.s^{-1}$ and the k_{off} 5.10^{-3} s^{-1} while in the right picture the k_{on} was 10^6 $M^{-1}.s^{-1}$ and the k_{off} 50 s^{-1}.

So in order to extract kinetic information from the peak shapes of the titration, the parameters in a Python simulation of the titration were adjusted to conform to the experimental results. The closest fit is presented in figure 5.14. The simulation was realized with a model using one binding site, using the experimental conditions of the titration (ligand and receptor concentrations). The experimental results could be reproduced with the K_d of 50 µM (obtained from the HSQCs) for both peaks but with different k_{on}: 1.10^8 $M^{-1}.s^{-1}$ and 5.10^5 $M^{-1}.s^{-1}$.

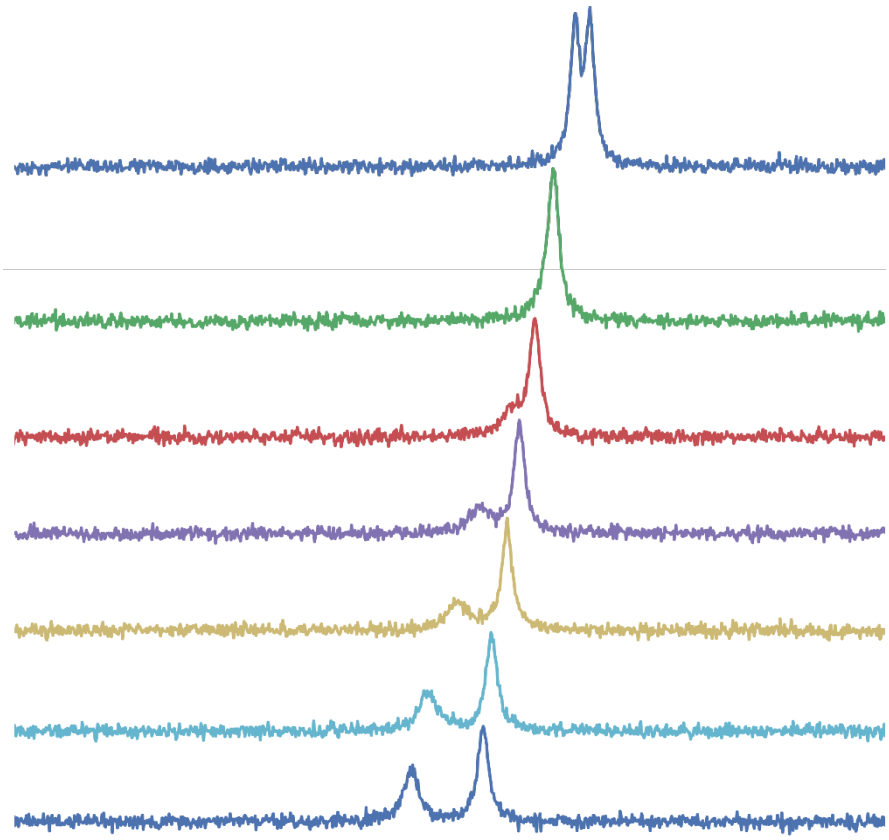


Fig 5.14 Python simulations aiming at reproducing the experimental results described in figure 5.12: the chemical shift displacement as well as the peak broadening.

5.3 Conclusion

The serine in a PRD is a conserved pattern in the RAR sequences and the phosphorylation of the serine has a strong effect on the binding affinity with the $SH3_3$ domain of vinexinβ. Adding a fluorine on the proline and thus, forcing the conformations, allowed to observe the peptide properties with and without the receptor. The phosphorylation of the serine has a strongest impact on the interaction affinity but the prolines conformation had also a measurable effect. The analysis of linewidths suggests a model for this interaction more complex than a direct 1-to-1 interaction. A hypothetical model is drawn in figure 5.15 that could explain the difference observed in the two fluorine peaks: the peptide could either immediately bind the domain in a high affinity region (close to the “RVYK” sequence) or there could be the formation of a pre-interaction complex of lower affinity where the poly-proline helix slides alongside the SH3 fold.

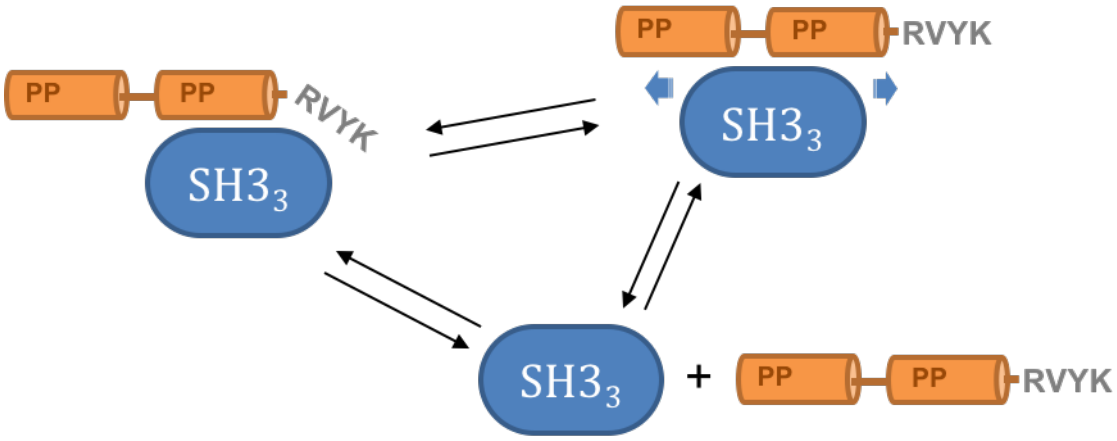


Fig 5.15 Ligand-domain interaction model

This model has to be completed. Our laboratory is planning to record and model ^{19}F titration experiments with the peptides containing a serine (the peptides PL105 to 108) since it allows for the extraction of kinetic data. Moreover, we recently obtained from José C. Martins lab, in Gent, fluorinated peptides with both (4R,4S) and (4S,4R) prolines. This was done to see if the peptide conformations are similar and also in order to distinguish which proline was more affected by the binding. The assignment of fluorine peaks are needed as well as interaction models to decipher how phosphorylation impacts the network of interaction between Vinexinβ and peptides.

References

Chiang, Y.-C., Lin, Y.-J., and Horng, J.-C. (2009). Stereoelectronic effects on the transition barrier of polyproline conformational interconversion. Protein Sci. Publ. Protein Soc. 18, 1967–1977.

Ho, B.K., Coutsias, E.A., Seok, C., and Dill, K.A. (2005). The flexibility in the proline ring couples to the protein backbone. Protein Sci. Publ. Protein Soc. 14, 1011–1018.

Hoffmeier, Wikimedia commons: Image under Public Domain, accessed at the url: https://commons.wikimedia.org/wiki/File:Nuclear_Receptor_Structure.png

Kay, B.K., Williamson, M.P., and Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 14, 231–241.

Koskinen, A.M.P., Helaja, J., Kumpulainen, E.T.T., Koivisto, J., Mansikkamäki, H., and Rissanen, K. (2005). Locked conformations for proline pyrrolidine ring: synthesis and conformational analysis of cis- and trans-4-tert- butylprolines. J. Org. Chem. 70, 6447–6453.

Lalevee, S., Bour, G., Quinternet, M., Samarut, E., Kessler, P., Vitorino, M., Bruck, N., Delsuc, M., Vonesch, J., Kieffer, B. and Rochette-Egly, C. (2010). Vinexin, an atypical "sensor" of retinoic acid receptor signaling: union and sequestration, separation, and phosphorylation. The FASEB Journal, 24(11), pp.4523-4534.

Li, D., Furukawa, H., Deng, H., Liu, C., Yaghi, O. and Eisenberg, D. (2013). Designed amyloid fibers as materials for selective carbon dioxide capture. Proceedings of the National Academy of Sciences, 111(1), pp.191-196.

Newberry, R. and Raines, R. (2016). 4-Fluoroprolines: Conformational Analysis and Effects on the Stability and Folding of Peptides and Proteins. Topics in Heterocyclic Chemistry.

Rochette-Egly, (2007). Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). Nuclear Receptor Signaling, 4.

Stryer, L., and Haugland, R.P. (1967). Energy transfer: a spectroscopic ruler. Proc. Natl. Acad. Sci. U. S. A. 58, 719–726.

Wedemeyer, W.J., Welker, E., and Scheraga, H.A. (2002). Proline cis-trans isomerization and protein folding. Biochemistry (Mosc.) 41, 14637–14644.

Wei - GNU GPL license, accessed at the url: http://gepard.bioinformatik.uni-saarland.de/research/protein-protein-interaction/poly-proline-peptides

Chapter 22

Accurate Protein–Peptide Titration Experiments by Nuclear Magnetic Resonance Using Low-Volume Samples

Christian Köhler, Raphaël Recht, Marc Quinternet, Frederic de Lamotte, Marc-André Delsuc, and Bruno Kieffer

Abstract

NMR spectroscopy allows measurements of very accurate values of equilibrium dissociation constants using chemical shift perturbation methods, provided that the concentrations of the binding partners are known with high precision and accuracy. The accuracy and precision of these experiments are improved if performed using individual capillary tubes, a method enabling full automation of the measurement. We provide here a protocol to set up and perform these experiments as well as a robust method to measure peptide concentrations using tryptophan as an internal standard.

Key words Affinity measurements, Protein–peptide interactions, NMR, Equilibrium binding constants

1 Introduction

Nuclear Magnetic Resonance (NMR) provides a powerful tool to study protein–ligand and protein–protein interactions at atomic resolution [1]. Among many other possibilities, NMR can be used to measure very accurately the equilibrium constant of the interaction, provided that its equilibrium dissociation constants (K_d) is in the range of 10 μM or above, a value that corresponds to the study of rather weak interactions. Several methods have been developed to measure protein–ligand dissociation constants, and they are usually classified in two main classes: the "ligand-observed" and the "protein-observed" methods. While "ligand-observed" methods, such as Saturation Transfer Difference (STD) or WaterLogsy share common principles with other biophysical approaches, the "protein-observed" approach is unique to NMR for its ability to deliver site-specific information [2, 3]. Thanks to these properties, NMR is now an established tool in pharmaceutical industry where it is used in drug discovery strategies, essentially at

Senta Reichelt (ed.), *Affinity Chromatography: Methods and Protocols*, Methods in Molecular Biology, vol. 1286,
DOI 10.1007/978-1-4939-2447-9_22,

the hit-to-lead step, where low to medium affinity ligands are gradually optimized into potent ligands [4]. The classical approach to study ligand–protein interactions relies on the measurement of protein chemical shift perturbations (CSP) induced by the binding of the ligand. This is generally performed using proteins that are enriched with magnetically active isotopes such as nitrogen 15 or carbon 13 and the prior knowledge of the protein resonance assignments that links a measured nucleus frequency to the corresponding molecular site. The chemical shift perturbations are then monitored using heteronuclear correlation spectra upon successive addition of increasing amounts of ligand. This approach is applicable to very large protein complexes such as the proteasome or the nucleosome, provided that appropriate labeling strategies are used such as the selective labeling of methyl groups [5]. It has been recently shown that this approach is also applicable with non-labeled protein samples thanks to the latest progress in NMR spectrometer sensitivity and the use of relaxation optimized pulse sequences such as Methyl SOFAST [6]. For proteins with molecular weights of less than 20 KDa, the common approach relies on the cost-effective production of ^{15}N labeled samples and the use of highly sensitive ^{1}H–^{15}N HSQC correlation spectra to monitor CSP. Here, we present a protocol enabling the equilibrium dissociation constants between a binding peptide and a small protein to be measured with high precision and accuracy. The method relies on the use of several low-volume samples, an approach that provides better accuracy when compared to the classical sequential titration method [7]. The protocol takes advantage of the ability to quantify precisely the amount of ligand present in the different samples as an accurate knowledge of the active concentrations of the interacting partners determines the reliability of the final result. The practical aspects of these measurements are illustrated using the interaction between the third SH3 domain of Vinexinβ and a model polyproline peptide from the N-terminal domain (NTD) of the Retinoic Acid Receptor γ (RARγ) as a prototypal case (Fig. 1). In this particular study, both accurate and precise measurements of K_d values for different peptides are needed to understand the molecular basis of the affinity modulation by the phosphorylation of the RARγ NTD [8].

2 Materials

2.1 Protein Production

The protein is obtained using heterologous expression in *E coli* according a protocol that depends on the system under study. Produce 4–5 mg of purified ^{15}N labeled protein using adapted expression and purification protocols (*see* **Note 1**).

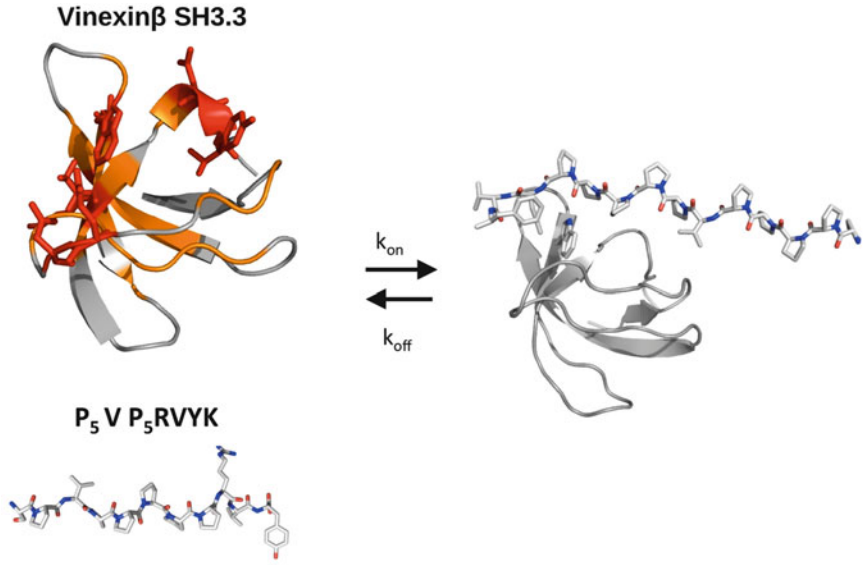


Fig. 1 The titration protocol presented here is illustrated with data originating from an interaction study between a model peptide from the proline-rich region of the RARγ NTD and the third SH3 domain of the human Vinexinβ [8]. The residues highlighted in orange and red show Chemical Shift Perturbation (CSP) of their $^{1}H–^{15}N$ correlation peaks upon addition of increasing amounts of peptide, indicating the location of the binding site on the protein surface. The CSP of *red highlighted* residues were used to fit the equilibrium dissociation constant K_d

2.2 Peptide Synthesis

Peptides are obtained from the peptide synthesis platform at IGBMC using an ABI 443A synthesizer adapted to FMOC chemistry. Purify the crude peptide products by reverse phase high performance liquid chromatography (HPLC) before a second chromatographic purification step in a migration column containing a cluster of resin balls (stable phase). Check the purity (95 % or better) of the resulting product by examining the HPLC elution profile, and by analyzing the peptide by mass spectrometry and NMR (*see* **Note 1**).

2.3 Capillary System

Use 1.7 mm outer-diameter capillary system for NMR measurements. This system is composed of 75 mm long capillaries capped with a teflon tube which is placed into a sample holder. Use a sample volume of 50 μL, which produces a filling height of 40 mm that was tested to be sufficient. The sample holders have a standard 5 mm outer diameter upper section with a transition to a 3 mm outer diameter (60 mm long) stem. The sample holder is reusable and fits all conventional 5 mm rotors. Fill the space between the capillary and the sample holder with 50 μL of D_2O (deuterated water) for the external lock. The system was purchased from "New-Era" (Vineland, NJ, USA).

2.4 NMR Measurements

The NMR measurements should be performed using a high-field (above 600 MHz) NMR spectrometer equipped with a triple resonance cryogenic probe. Set the acquisition parameters to keep the measurement time within reasonable limits of 1–2 h per titration point. If available, use a sample changer to run the experiment unattended overnight (*see* **Note 2**).

2.5 Theoretical Aspects of K_d Measurements from NMR Frequencies

The binding of a ligand peptide (L) to a protein (P) to form a peptide–protein complex (PL) is described by the following equilibrium:

$$P + L \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} PL \tag{1}$$

The dissociation equilibrium constant K_d is defined as:

$$K_d = \frac{k_{\text{off}}}{k_{\text{on}}} = \frac{[P]\,[L]}{[PL]} \tag{2}$$

Where $[P]$, $[L]$ and $[PL]$ are the concentrations of the free protein, the free ligand and the complex respectively and k_{on} and k_{off} the association and dissociation rates respectively. The ability to determine the value of the dissociation constant from chemical shift measurements depends on the exchange kinetic between free and bound species, defined as:

$$k_{\text{exc}} = k_{\text{off}} + k_{\text{on}}[L] \tag{3}$$

For k_{exc} values significantly larger than the NMR frequency difference $2\pi\left(\nu_i^{\text{bound}} - \nu_i^{\text{free}}\right)$ between the bound and free states of the protein, the observed frequency, ν_i is a weighted average between the frequencies of the free and bound states:

$$\nu_i = x_1\nu_i^{\text{bound}} + (1 - x_i)\nu_i^{\text{free}} \tag{4}$$

$x_i \in [0, 1]$ is the occupancy of a given binding site i within the protein. This averaging situation occurs when k_{off} is rather fast, which corresponds to ligands of weak affinity (in the micromolar to millimolar range). Assuming that the frequency change of a given nucleus within the protein is essentially due to local perturbations, its value provides therefore a direct measurement of the occupancy of the binding site localized in its vicinity using:

$$x_i = \frac{\nu_i - \nu_i^{\text{free}}}{\nu_i^{\text{bound}} - \nu_i^{\text{free}}} \tag{5}$$

The subscript i highlights the unique ability of NMR spectroscopy to measure site-specific affinity binding constants. The value of the site-specific dissociation constant, K_d^i, is subsequently obtained using a nonlinear fit of the following equation:

$$x_i^2 - x_i\left(1 + \frac{[L]_0}{[P]_0} + \frac{K_d^i}{[P]_0}\right) + \frac{[L]_0}{[P]_0} = 0 \tag{6}$$

with: $[L_0] = [L] + [PL]$ and $[P_0] = [P] + [PL]$

$K_d{}^i$ and $\nu_i{}^{\text{bound}}$ are adjustable parameters to minimize the value of the target function:

$$f\left(K_d^i, \nu_i^{\text{bound}}\right) = \frac{1}{N}\sum_{j=1}^{N}\left(\nu_{i,j}^{\text{calc}} - \nu_{i,j}^{\text{obs}}\right)^2 \tag{7}$$

$\nu_{i,j}{}^{\text{calc}}$ is a frequency calculated for a given total concentrations of protein $[P]_{0,j}$ and ligand $[L]_{0,j}$, using equations (Eqs. 4 and 6) while $\nu_{i,j}{}^{\text{obs}}$ is the corresponding measured frequency. The subscript j identifies each single titration point from the total number of N different mixtures of protein and ligand.

The protein frequencies are usually measured using ^{15}N or ^{13}C labeled proteins and heteronuclear correlation spectra. For small proteins, such as a SH3 domain, ^{1}H–^{15}N correlation spectra provide an inexpensive and accurate way to monitor the chemical shift perturbations induced by the binding of a ligand. Both nitrogen and its bound amide proton frequencies are reported using a composite chemical shift (frequency) usually defined as:

$$\delta_{\text{comp}} = \sqrt{\delta_{^{15}\text{N}}^2 + \left(\frac{\gamma_{\text{H}}}{\gamma_{\text{N}}}\delta_{^{1}\text{H}}\right)^2} \tag{8}$$

3 Methods

3.1 Design of the NMR Titration Experiment

1. The feasibility of the affinity measurement by NMR will depend on the K_d value and the ability to get the protein and the peptide at concentrations that are compatible with NMR measurements. The minimal protein concentration required to acquire ^{1}H–^{15}N heteronuclear correlation spectra varies between 10 and 100 μM, depending on the available NMR spectrometer. Check with classical methods (UV, DLS, …) whether the protein of interest can be concentrated up to these values using a non-labeled protein sample.
2. Check the quality of the ^{15}N labeled sample by recording a ^{1}H–^{15}N HSQC spectrum of your stock protein solution at its highest concentration. Standard large volume NMR tubes (5 or 3 mm tubes) can be used for this purpose. Check the stability of the protein sample at the planned measurement temperature by recording a ^{1}H–^{15}N HSQC spectrum after a few days at this temperature. The appearance of a subset of sharp peaks is indicative of protein degradation (*see* **Note 3**).

3. Desalt the peptide and transfer it to the buffer used for the protein. Both steps could be done at once using a gel filtration column such as the Superdex Peptide 10/300 GL (*see* **Note 4**).
4. Since the method presented here is only applicable when the protein–peptide interaction leads to a so-called "fast exchange regime," it is important to check whether this condition holds true for the system of interest at an early stage of the study. This could be done by preparing an initial sample with approximately stoichiometric concentrations of protein and peptide and by recording a $^{1}H–^{15}N$ HSQC spectrum of this sample. Four distinct situations may be encountered:
   - The correlation map of the mixture is identical to the one obtained for the sole protein, indicative of an absence of interaction.
   - The spectrum displays broader correlation peaks and several peaks are missing. This case corresponds to more complex situations where the protein undergoes an intermediate time-scale exchange between two (peptide-bound and free) or more states, preventing K_d measurements.
   - A second set of correlation peaks is observed. This is indicative of a "slow exchange regime" corresponding to tight interactions between the protein and the peptide. No quantitative measurement of the K_d will be possible using chemical shift measurements.
   - The correlation map of the mixture contains the same number of peaks, but several of these peaks have different frequencies when compared to the peptide-free spectrum of the protein. This situation will allow the measurement of the K_d.

3.2 Measurement of Peptide and Protein Concentrations

Several factors do affect the accuracy and precision of equilibrium constant measurements by NMR, the most important one being inaccurate estimations of protein and ligand concentrations (*see* **Note 5**). While the protein concentration may be measured with reasonable accuracy using its absorption at 280 nm, this is not the case for the peptides, in particular when they lack tryptophan or tyrosine residues. It is therefore essential to ensure an accurate measurement of protein and peptide concentrations. We report hereafter a simple method that provides reasonable accuracy for peptide concentration measurements by NMR (below 10 %) (*see* **Note 6**).

1. Prepare a stock solution of tryptophan by weighting about 6 mg of L-Tryptophan (MW: 204.23 g/mol). Dissolve the powder in 5 mL of D_2O 99.9 %.

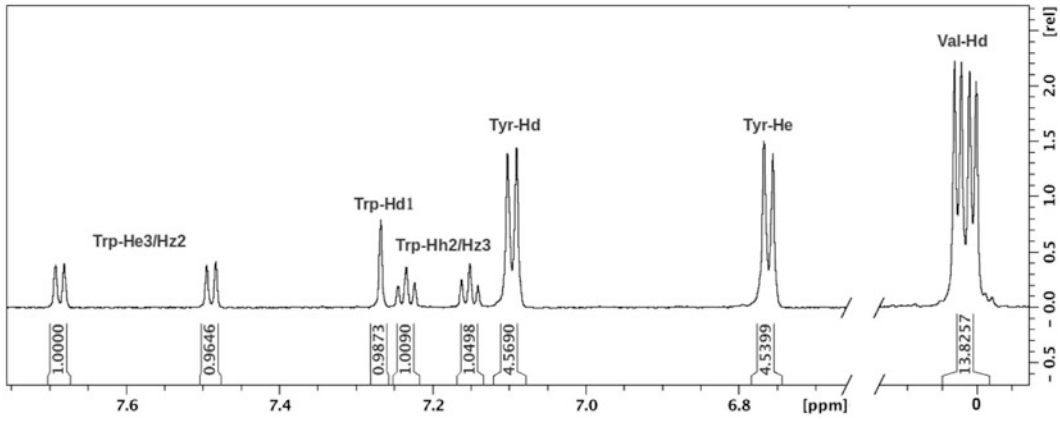


Fig. 2 1D proton spectrum of a mixture between a model peptide (sequence P_5VP_5RVYK) corresponding to the proline-rich region of the RARγ NTD and the tryptophan solution of known concentration. The amount of peptide required for this concentration measurement was 15–20 μg. The ratio between the averaged integrals of the tryptophan peaks and those of the peptide indicated that the peptide was 2.3 times more concentrated than the tryptophan. Given the concentration of the tryptophan standard, this led to concentration of 4.5 ± 0.2 mM for the peptide stock solution. The relative uncertainty on the peptide concentration using this method was 4.4 %

2. Measure the concentration of the L-Tryptophan stock solution (5–6 mM) by measuring the absorption at 280 nm (ε_{280} = 5,690 $mol^{-1} \cdot cm^{-1}$) (*see* **Note 7**).
3. Prepare a NMR sample by mixing a small volume (10–20 μL) of peptide (whose stock solutions are usually available at millimolar concentration) with (5–20 μL) of L-Tryptophan stock solution. Complete with D_2O to get a total sample volume of 150–170 μL, suitable for a 3 mm tube.
4. Record a 1D proton NMR spectrum of the sample with water pre-saturation for solvent signal suppression. Adjust the number of scans to get a reasonable signal-to-noise ratio according the sensitivity of your spectrometer. A long relaxation delay (10–15 s) should be used to account for the long T1 of the tryptophan aromatic protons (about 3 s) (Fig. 2).
5. Perform a baseline correction and integrate the signals of the tryptophan aromatic protons as well as one or few isolated resonance peaks of the peptide (we often use methyl groups resonances). Compute the ratio between the areas (normalized by the number of protons resonating at the corresponding frequency) measured for the peptide and the tryptophan to get the concentration of the peptide stock solution $[L]_0$ using:

$$[L]_0 = \frac{A_L N_w}{A_w N_L} \frac{DF_L}{DF_w} [W]_0 \quad (9)$$

Where A_L is the areas measured under one or several peaks corresponding to N_L proton resonances of the peptide. A_w and

N_w are the corresponding values obtained for the tryptophan resonances. DF_L and DF_w are the dilution factors used to prepare the sample from the peptide and the tryptophan stock solutions, respectively. $[W]_0$ is the concentration of the tryptophan stock solution determined in **step 2**.

6. Measure the protein concentration using its absorption at 280 nm.

3.3 NMR Capillaries Preparation and NMR Acquisition

1. Prior the titration experiment, the protein concentration needed to achieve a reasonable signal-to-noise (S/N) ratio on the heteronuclear $^{1}H–^{15}N$ HSQC spectra should be adjusted. On a 700 MHz equipped with a cryoprobe, a protein concentration (the SH3.3 domain of Vinexin β) of 50 to 80 μM in a 1.7 mm capillary tube provides good quality spectra. This will highly depend on the available NMR equipment as well as on the system under study. The use of NMR capillary tubes is of particular interest when titration experiments have to be performed in high salt concentrations (*see* **Note 8**). As an example, the comparison of relative sensitivity measured on SH3 samples using standard 5 mm, 3 mm tubes and capillary tubes at 700 MHz is provided in Table 1. Despite the apparent reduced signal-to-noise ratio observed for low-volume samples, the relative sensitivity (sensitivity per amount of material) is significantly increased, up to a factor of 3 with capillaries as shown in Table 1 (*see* **Note 9**).
2. Prepare the different protein–peptide mixtures in Eppendorf tubes. Adjust the sample volume according the capacity of the chosen capillaries. For 1.7 mm capillaries, the volume is adjusted to 75 μL using the protein buffer (*see* **Note 1**). Fill the capillaries using a stretched Pasteur pipette or a Hamilton syringe. Add 50 μL D_2O in the capillary holder for external lock. After capping the capillaries, insert them into the capillary holder as shown in Fig. 3. As an example, we provide here a sample preparation table (Table 2) that was used to measure the

Table 1
Experimental sensitivities per amount of protein, relative to a 5 mm (550 μL) NMR tube

Sample geometry	550 μL 5 mm tube 9 % D_2O in sample	180 μL 3 mm tube 9 % D_2O in sample	50 μL capillary 9 % D_2O in sample	50 μL capillary no D_2O in sample
Ratio of protein material	1	0.33	0.09	0.1
HSQC S/N	763	569	179	241
Relative sensitivity	**1**	**2.26**	**2.61**	**3.16**

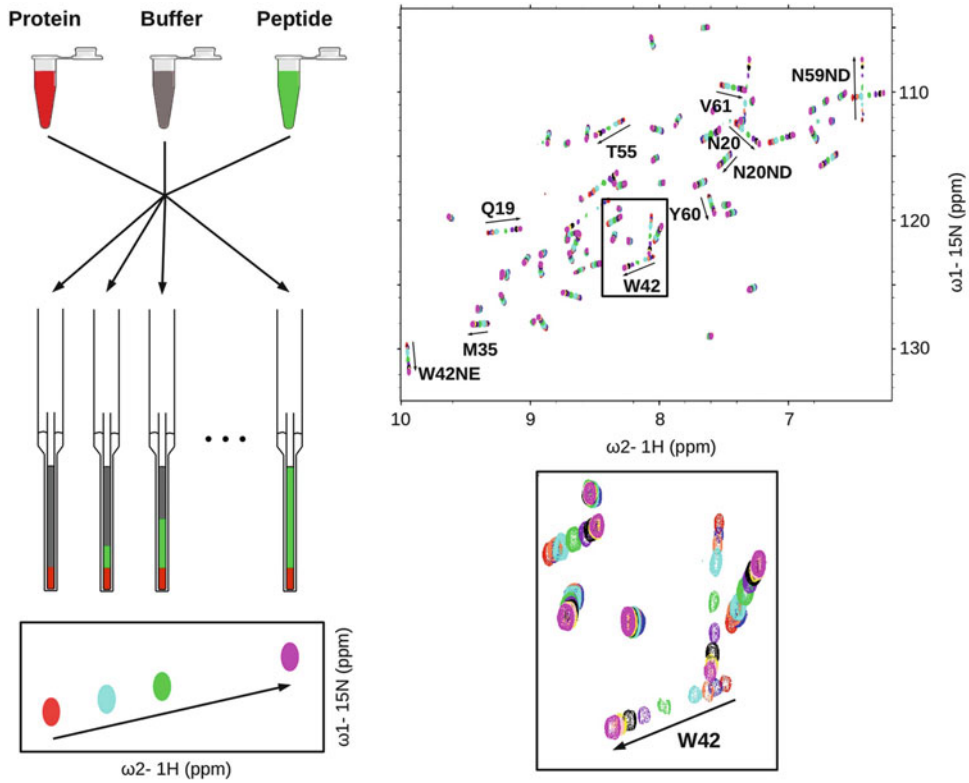


Fig. 3 Preparation of capillary tubes (*left*) for $^{1}H–^{15}N$ HSQC measurements (*right*). The insert shows a close-up on the effect of increasing amounts of peptide on the cross peak corresponding to the backbone amide proton of Tryptophan 42, which is located within the binding site (*see* Fig. 1)

affinity of SH3.3 domain of Vinexin β to a proline rich peptide from the RARγ NTD (*see* **Note 10**).

3. For each sample, record a $^{1}H–^{15}N$ HSQC heteronuclear spectrum with sufficient acquisition time and resolution to allow a precise measurement of nitrogen and proton frequencies.
4. The processed spectra should be superposed in order to identify the $^{1}H–^{15}N$ correlation peaks that are subjected to the largest frequency shifts upon addition of the peptide. Perform a peak-picking on each spectrum in order to compute a composite chemical shift perturbation using:

$$\Delta\delta_{\mathrm{comp}} = \sqrt{(\Delta\delta_{\mathrm{N}})^2 + \left(\frac{\gamma_{\mathrm{H}}}{\gamma_{\mathrm{N}}}\Delta\delta_{\mathrm{H}}\right)^2} \qquad (10)$$

where $\Delta\delta_{\mathrm{N}}$ and $\Delta\delta_{\mathrm{H}}$ are the difference between the nitrogen and proton chemical shifts measured with a given amount of peptide and those measured in absence of peptide. γ_{H} and γ_{N} are the gyromagnetic ratios of the proton and the nitrogen respectively (*see* **Notes 11** and **12**).

Table 2
Composition of samples used for the titration of the C-terminal SH3 domain of human Vinexin β with the P_5VP_5RVYK peptide

Sample N°	Conc. Peptide stock (μM)	Volume SH3 (μL)	Volume peptide (μL)	Volume buffer (μL)	Conc. SH3 (μM)	Conc. peptide (μM)	Stoichiometric ratio
1	45	15	0	60	64.4	0	0
2	45	15	18	42	64.4	10.8	0.17
3	450	15	3	57	64.4	18	0.28
4	450	15	6	54	64.4	36	0.56
5	450	15	15	55	64.4	90	1.40
6	4,500	15	3	57	64.4	180	2.80
7	4,500	15	6	54	64.4	360	5.59
8	4,500	15	12	48	64.4	720	11.18
9	4,500	15	16	44	64.4	960	14.91
10	4,500	15	30	30	64.4	1,800	27.95
11	4,500	15	50	10	64.4	3,000	46.58

3.4 Data Analysis and Error Estimates

The first step of the analysis consists in estimating the number of peptide binding site on the protein surface. (1) a single binding site and one step binding mechanism are characterized by a linear trajectory of the peak in the $^{1}H–^{15}N$ HSQC series [6, 7, 9, 20]. This should be carefully checked, as the K_d is only defined under these conditions. (2) Further check can be performed by mapping the location of the corresponding amino acids on the protein structure, if both the structure and the HSQC assignment are known (*see* **Note 13**). (3) A last insight is provided by the numerical analysis of chemical shift data. The fitting procedure described below may first be applied using individual $^{1}H–^{15}N$ correlations first to extract local K_d values. Their convergence to an identical dissociation constant provides a strong indication that these $^{1}H–^{15}N$ sites monitor the peptide occupancy of the same binding site (*see* **Note 14**).

1. Find the values of K_d and $\Delta\delta_{\mathrm{comp}}^{\mathrm{max}}$ that leads to a minimal value of Eq. 7. This could be performed using least-square fitting procedures available in CcpNmr or other protein NMR software packages. We recommend using Python scripts which offers more flexibility in data analysis and plotting (*see* **Note 15**). Average the Chemical shift changes of Amide groups that belong to the same binding site in order to increase the precision of the binding site occupancy measurement. In case of the Vinexinβ SH3.3 domain, an average chemical shift

perturbation was calculated from 10 $^{1}H–^{15}N$ correlations corresponding to residues Q19, N20, N20Nδ, M35, W42, W42Nε, T55, N59Nδ, Y60, and V61 (highlighted in Fig. 3).

2. Estimate the uncertainty on the resulting K_d values. This is done using a Monte Carlo simulation where synthetic datasets are generated and subsequently fitted. These synthetic datasets are generated using a Gaussian distribution of $\Delta\delta_{comp}$ using the values calculated from the first fit as the mean and the root-mean square deviation (the square root of Eq. 7) as the standard deviation. The uncertainties on protein and peptide concentrations are taken into account by generating distributions of peptide and protein total concentrations around the initial values. The width of the distribution is given by the uncertainties on the concentrations (*see* **Note 16**). As concentration values can't be negative, the Log-normal distribution is chosen to generate the distribution of concentration values [10]. The distribution width is then directly given by the relative uncertainties on the measured concentrations (*see* **Notes 17** and **Note 18**).

4 Notes

1. The protocol used to purify the C-terminal SH3.3 domain of human Vinexinβ (REFSEQ: NP 001018003) was a classical two steps purification protocol (Glutathione affinity and gel filtration) that is described in ref. 8. Alternatively, ^{13}C, ^{15}N double-labeled proteins are also suitable for titration experiments. The final buffer was a low salt phosphate buffer with 20 mM sodium phosphate at pH 7.0, 100 mM NaCl.
2. We used a BRUKER Avance III 700 MHz spectrometer equipped with a TCI cryoprobe and a BACS60 sample changer. $^{1}H–^{15}N$-HSQC spectra were recorded with 32 scans and 128 data points in the indirect dimension resulting in a total acquisition time of 90 min per sample.
3. Several precautions may be used to prevent, or at least slow down protein degradation. Antiproteases are usually added to the final sample as well as sodium azide (NaN_3) (0.01 % w/v) used as an antibacterial. If the protein sequence contains free cysteines, we usually add reducing agents such as Dithiothreitol (DTT) or TCEP (Tris(2-carboxyethyl)phosphine). In that case, all used buffers should be carefully degassed and oxygen removed from the sample by Helium or Argon bubbling.
4. Protocols used for peptide synthesis and purification lead to the presence of significant amount of trifluoro acetic acid (TFA) salts in dry peptide samples. NMR provides an accurate method

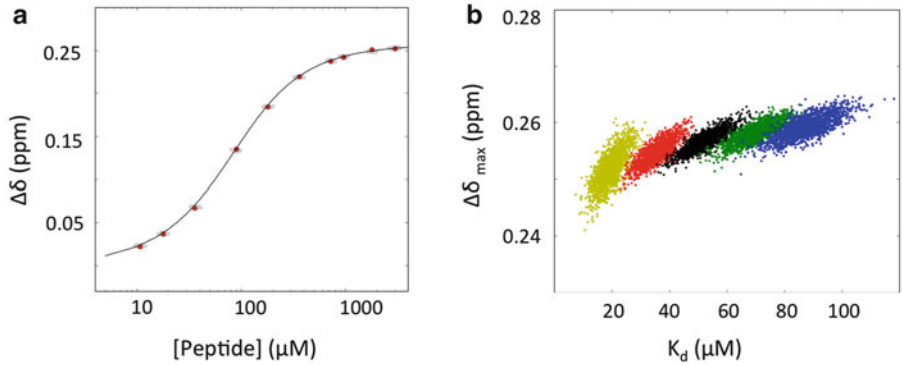


Fig. 4 Least-square fit of the chemical shifts perturbation data measured for the interaction between the P_5VP_5RVYK model peptide and the Vinexinβ SH3.3 domain. (**a**) Semi-log plot of the composite chemical shifts computed from ten residues of SH3.3 as a function of peptide concentrations. Pseudo experimental points generated for the Monte Carlo estimate of the uncertainty on the K_d value are shown in *gray*. These points are distributed according to a gaussian distribution for the Δδ values and according to a log-normal distribution for the peptide and protein concentrations. (**b**) Distribution of the two fitted parameters after the Monte Carlo procedure. The concentration uncertainties were estimated to be 10 % for the SH3.3 protein and between 4 and 5 % for the peptide. The calculations were performed for a peptide stock solution whose concentration was either underestimated by a factor of 0.6 and 0.8 (*yellow* and *red*), or overestimated by 1.2 and 1.4 (*green* and *blue*). The *black points* reflect the effect of pure random noise of the fitting procedure as the concentration of peptide stock solution is considered to be accurate

to check both the efficiency of the desalting procedure and the purity of the final peptide solution by recording ^{1}H and ^{19}F 1D spectra of the stock peptide solution. Depending on the peptide sequence, we found that the gel filtration desalting method may leave significant amounts of residual trifluoroacetate salts in the final sample. In this case, more efficient protocols should be considered [11].

5. Over or underestimated values of the peptide stock concentration have a dramatic impact on the K_d values resulting from the fit of Eq. 7. This effect can be evaluated by performing Monte Carlo simulations with systematically biased values of ligand concentrations (20 or 40 % above or below the true value, as shown in Fig. 4 and Table 3). The results obtained indicate that a concentration of ligand peptide that is underestimated by 20 % leads to an overestimation of the affinity by a factor of 30 % (The apparent K_d value is 36 μM instead of 52 μM). This large effect is due to the high correlation that exists between the different measurement points since the corresponding protein–peptide mixtures are usually prepared from the same peptide stock solution.
6. A method has been recently proposed to compute the molar absorptivity of a protein or peptide at 205 nm from its amino acid sequence, providing an alternative for quantifying peptides

Table 3
Average values and standard deviations of dissociation constants (K_d) and chemical shift perturbations ($\Delta\delta_{max}$) values computed from Monte Carlo calculations

	Relative uncertainties (one standard deviation) on peptide concentrations				
	10 %	**20 %**	**30 %**	**40 %**	**50 %**
K_d (μM)	52.3 ± 5.6	52.0 ± 9.9	52.5 ± 14.6	53.9 ± 19.8	52.3 ± 22.3
$\Delta\delta_{max}$ (ppm)	0.257 ± 0.002	0.256 ± 0.003	0.256 ± 0.005	0.256 ± 0.006	0.255 ± 0.007

	Ratio between measured and real peptide concentrations				
	0.6	**0.8**	**1.**	**1.2**	**1.4**
K_d (μM)	19.6 ± 3.8	35.6 ± 4.2	52.1 ± 5.4	69.8 ± 6.1	87.4 ± 7.3
$\Delta\delta_{max}$ (ppm)	0.252 ± 0.003	0.255 ± 0.002	0.257 ± 0.002	0.258 ± 0.002	0.259 ± 0.002

Experimental chemical shifts were obtained from the interaction of the P_5VP_5RVYK peptide with the Vinexinβ SH3.3 domain. The uncertainty of the SH3.3 protein concentration was estimated to be 10 %. The fitted values are reported for different uncertainties of the peptide concentrations (*upper panel*) or for a systematic error on peptide stock solution (*lower panel*).

lacking tryptophan or tyrosine residues [12]. Combining this measurement with the quantitative evaluation of peptide concentration by NMR provides an interesting way to get robust estimates of concentrations. Other methods have been proposed for protein concentrations measurements by NMR, such as PULCON for instance [13].

7. In order to increase the precision of this OD measurement, we usually perform several $OD_{280\ nm}$ measurements with targeted absorption values of 0.8, 0.4, 0.2, and 0.1. The linear regression of this series of measurements is used to provide an estimation of the uncertainty on the Tryptophan stock solution concentration.
8. The Signal-to-Noise ratio (S/N) in NMR may be written as:

$$S/N \propto \frac{M_0 B_1}{\sqrt{P_s(T_a + T_s) + P_c(T_a + T_c)}} \tag{11}$$

where M_0 is the spin magnetization, B_1 the radio-frequency (RF) field intensity applied to the sample, and P_c and P_s are the RF power absorbed by the coil and by the sample, respectively. T_c and T_s are the temperature of the coil and the sample, respectively, while T_a is the noise temperature of the preamplifier [14–16]. Recent progress in NMR probe development, most notably the development of cryogenic probes, improved the S/N by lowering T_c and T_a down to 10–25 K and by reducing P_c by optimizing the coil quality factor (*see* ref. 16).

There remains room for S/N optimization on the P_s term, which is mostly dependent on the sample itself because of dielectric losses. It is known that the RF power dissipated in the sample depends on the dielectric constant of the medium which is very much dependent on the type of solvent and on the ionic strength when working in H_2O. Thus, the P_s term depends on the distribution of the electric field within the sample geometry and on the strength of the RF irradiation (expressed as its angular frequency ω_1) with:

$$P_S \propto \omega_1^2 \tag{12}$$

Because of this dependency, P_s losses become more prominent with increasing fields. On a given probe, reducing the internal diameter of the NMR tube with a capillary system has two opposite effects on the overall sensitivity of the measurement. First, reducing the sample volume at a given concentration results in a loss of signal due to a proportional reduction of sample quantity. However, the power dissipated within the sample P_s is also reduced and so is the noise, leading to a potential improvement of the S/N. The balance between these two effects strongly depends on the nature of the sample itself, and the amount of the overall effect is not directly predictable. Finally, it should be mentioned that the use of capillary tubes centers the sample in the inner volume of the coil where the electric field is minimum and the impact on P_s and thus on the noise is maximum. This effect has been studied [17] and it was shown that in high salt conditions it is actually beneficial in terms of S/N to reduce the NMR tube diameter while keeping all concentrations constant.

9. This gain results from several factors. First, the signal noise arising from RF losses in the sample itself is minimized in small diameter tubes due to a lower value of P_s, the RF power dissipated within the sample (*see* **Note 8**). This effect will be of increasing importance if high salt concentrations are required for the protein buffer and if a cryogenically cooled probe is used. A second source of sensitivity gain originates from a more optimal use of the sample volume as only about 30 % of the sample volume is outside the RF coil. On 5 mm tubes, susceptibility matched NMR tubes or plugs (Shigemi tubes) are usually used to compensate this effect, allowing doubling the relative sensitivity. Though the handling of these systems is cumbersome, the susceptibility matched approach can also be applied on capillary tubes, with a potential further 43 % gain in relative sensitivity. Finally, the use of an external lock implies that there is no need to add deuterium into the sample itself which otherwise leads to an additional loss of signal due to deuterium exchange of the amide protons. Notably, the

capillary sample lacking 9 % D2O enables another 21 % of gain in relative sensitivity.

10. In our example, the concentration of the protein is constant while the peptide concentration varies. It has been shown that an optimal sampling is achieved when both the protein and peptide concentrations are varied together [18].
11. Peak picking is usually performed using the software packages dedicated to protein NMR spectra analysis such as SPARKY (http://www.cgl.ucsf.edu/home/sparky), CcpNmr Analysis (http://www.ccpn.ac.uk) or CARA (http://cara.nmr.ch). Peak tracking can be performed with algorithms such as described in [19] for instance.
12. The ratio γ_H/γ_N is a weighting factor that compensates the difference of chemical shift ranges between proton and nitrogen frequencies. Its precise value is of little importance and there are also other weighting factors described in the literature.
13. The resonance assignment of a variety of proteins can be obtained from the Biological Magnetic Resonance Data Base (BMRB) at http://www.bmrb.wisc.edu/.
14. The knowledge of the resonance assignments is not required to identify two binding sites if their affinity are different and if this difference could be resolved by NMR titration experiments as shown in [6].
15. The set of Python script used to analyze the interaction between the Vinexinβ SH3.3 domain and the P_5VP_5RVYK RARγ model peptide is available at http://zenodo.org (doi: 10.5281/zenodo.11663).
16. The propagation of uncertainties of volume measurements follows the general law:

$$u^2(y) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\mathrm{cov}(x_i, x_j) \tag{13}$$

where $u(y)$ is the uncertainty on the concentration that depends on several variables ($y = f(x_i)$) depending on the specific scheme that is used for sample preparation. The covariance ($y = f(x_i)$) was set to 1 for volumes if the same pipette was used twice, and for concentrations when the same solution was used. The calculation of uncertainty propagation used for the Vinexinβ work is available at the following address: http://zenodo.org (doi: 10.5281/zenodo.11663).

17. Two main types of uncertainties have to be distinguished: an erroneous estimation of the peptide stock solution will lead to a

Table 4
Comparison of the uncertainties on ligand concentrations for sequential or parallel titration experiments

		Absolute (μM) and relative ligand concentration uncertainties	
Sample number	**Peptide concentration (μM)**	**Sequential titration scheme**	**Parallel titration scheme**
0	0.0		
1	10.8	4.8 % (0.52)	4.9 % (0.53)
2	18.0	3.7 % (0.66)	4.9 % (0.89)
3	36.0	4.1 % (1.49)	4.8 % (1.74)
4	90.0	5.4 % (4.84)	4.8 % (4.28)
5	180.0	5.8 % (10.5)	4.8 % (8.66)
6	360.0	6.7 % (24.3)	4.7 % (16.9)
7	720.0	7.4 % (53.5)	4.6 % (33.3)
8	960.0	6.4 % (61.9)	4.6 % (44.3)
9	1800.0	7.0 % (126.3)	4.6 % (82.7)
10	3000.0	5.1 % (152.7)	4.5 % (135.4)
	Max uncertainty:	7.4 %	4.9 %

systematic bias in the resulting K_d values, while pipetting errors will introduce random noise on the measurements. We have simulated both effects and the resulting uncertainties on fitted parameters are shown in Table 3. While a random noise of 20 % on the peptide concentration leads to a resulting relative uncertainty of 20 % on the K_d value, a 20 % underestimation of the peptide concentration leads to overestimation of the affinity by more than 30 % (36 μM instead of 52 μM). This emphasizes the importance of having the most accurate peptide concentration values before undertaking affinity measurements by NMR or by any other methods.

18. In order to provide a quantitative estimation of these effects, we performed formal calculations to compute the uncertainties on the protein and peptide concentrations for each point of the titration that arise from the uncertainties of volume measurements. These later values were taken from the specifications provided by the pipette manufacturer (Gilson Inc.). The resulting absolute and relative uncertainties on the ligand concentrations together with their impact on the resulting K_d are reported in Table 4. The parallel titration protocol leads to

maximal relative error on ligand concentrations of 4.9 %, a value that is lower than the one obtained (7.4 %) if the experiment would have been performed using a regular sequential addition of ligand to the same tube. It is worth noting that this calculation is probably underestimating the uncertainty associated with the sequential titration protocol as the multiple manipulations of the same tube will lead to unavoidable losses of sample volume, in particular when susceptibility matching tubes are used.

Acknowledgments

This work was supported by the ANR program VINRAR ANR-09-BLAN-0297, the Institut National du Cancer [grant number INCa-PL09-194], the Ligue Regional contre le cancer and by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01, as part of the European Strategy Forum on Research Infrastructures (ESFRI) and through national members agreements. The authors thank Claude Ling (IGBMC) for technical support, the chemical peptide synthesis service at IGBMC and Yves Nominé for critical reading of the manuscript.

References

1. Kieffer B, Homans S, Jahnke W (2011) Nuclear magnetic resonance of ligand binding to proteins. In: Podjarni A, Dejaegere A, Kieffer B (eds) Biophysical approaches determining ligand binding to biomolecular targets. RSC, Cambridge, UK, pp 15–55
2. Fielding L (2007) NMR methods for the determination of protein–ligand dissociation constants. Prog Nucl Magn Reson Spectrosc 51:219–242
3. Dalvit C (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. Drug Discov Today 14:1051–1057
4. Pellecchia M, Bertini I, Cowburn D, Dalvit C, Giralt E, Jahnke W, James TL, Homans SW, Kessler H, Luchinat C, Meyer B, Oschkinat H, Peng J, Schwalbe H, Siegal G (2008) Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov 7:738–745
5. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445:618–622
6. Quinternet M, Starck JP, Delsuc MA, Kieffer B (2012) Unraveling complex small-molecule binding mechanisms by using simple NMR spectroscopy. Chem Eur J 18:3969–3974
7. Bourry D, Sinnaeve D, Gheysen K, Fritzinger B, Vandenborre G, Van Damme EJ, Wieruszeski JM, Lippens G, Ampe C, Martins JC (2011) Intermolecular interaction studies using small volumes. Magn Reson Chem 49:9–15
8. Lalevee S, Bour G, Quinternet M, Samarut E, Kessler P, Vitorino M, Bruck N, Delsuc MA, Vonesch JL, Kieffer B, Rochette-Egly C (2010) Vinexinß, an atypical "sensor" of retinoic acid receptor gamma signaling: union and sequestration, separation, and phosphorylation. FASEB J 24:4523–4534
9. Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc 73:1–16
10. Limpert E, Stahel WA, Abbt M (2001) Log-normal distributions across sciences: keys and clues. Bioscience 51:341–352
11. Roux S, Zekri E, Rousseau B, Paternostre M, Cintrat JC, Fay N (2008) Elimination and exchange of trifluoroacetate counter-ion from cationic peptides: a critical evaluation of different approaches. J Pept Sci 14:354–359

12. Anthis NJ, Clore GM (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. Protein Sci 22:851–858
13. Wider G, Dreier L (2006) Measuring protein concentrations by NMR spectroscopy. J Am Chem Soc 128:2571–2576
14. Hoult DI, Lauterbur PC (1979) The sensitivity of the zeugmatographic experiment involving human samples. J Magn Reson 34:425–433
15. Hoult DI (1996) Sensitivity of the NMR experiment. In: Grant DM (ed) Encyclopaedia of nuclear magnetic resonance. Wiley, New York, NY
16. de Swiet TM (2005) Optimal electric fields for different sample shapes in high resolution NMR spectroscopy. J Magn Reson 174:331–334
17. Voehler MW, Collier G, Young JK, Stone MP, Germann MW (2006) Performance of cryogenic probes as a function of ionic strength and sample tube geometry. J Magn Reson 183:102–109
18. Markin CJ, Spyracopoulos L (2012) Increased precision for analysis of protein-ligand dissociation constants determined from chemical shift titrations. J Biomol NMR 53:125–138
19. Ravel P, Kister G, Malliavin TE, Delsuc MA (2007) A general algorithm for peak-tracking in multi-dimensional NMR experiments. J Biomol NMR 37:265–275
20. Kovrigin EL (2012) NMR line shapes and multi-state binding equilibria. J Biomol NMR 53:257–270

CHAPTER 6

^{15}N NMR relaxation study of the dynamic of a protein domain upon ligand binding

(Published work)

Introduction

The work presented in this chapter, now published in Biochemistry in 2014, aims at deciphering the binding mechanisms between a protein found in human cellular tight junctions and another protein, belonging to a high-risk papilloma virus. A dual-technique approach was chosen, combining ITC and NMR, to explain how the unusual thermodynamic signature of the binding could be linked to the folding of an initially disordered protein region. During my Master Thesis and the early part of my PhD Thesis, I contributed to the project through protein production, the NMR characterization of a mutant with ^{1}H-^{15}N HSQCs and other 2D to 3D experiments, the monitoring of the binding with ^{15}N heteronuclear relaxation measurements and the processing of those data through programming scripts. The technique and knowledge acquired for this scientific task were determinant in the shape of my thesis and thus, the resulting effort is presented as the closing chapter in the thesis manuscript.

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Disorder-To-Order Transition of MAGI-1 PDZ1 C-Terminal Extension upon Peptide Binding: Thermodynamic and Dynamic Insights

Juan Ramírez,[†] Raphaël Recht,[‡] Sebastian Charbonnier,[†] Eric Ennifar,[§] R. Andrew Atkinson,[‡,∥] Gilles Travé,[†] Yves Nominé,*[,†] and Bruno Kieffer[‡]

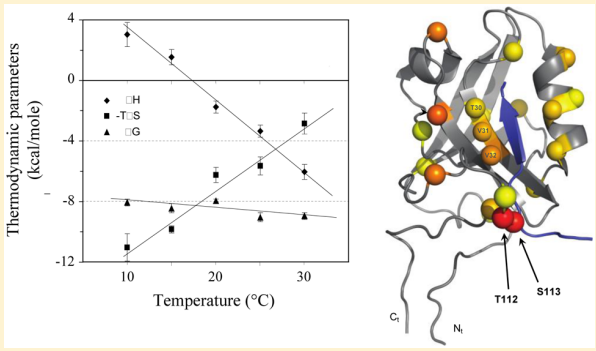
[†]Equipe Oncoprotéines, Ecole Supérieure de Biotechnologie de Strasbourg, Biotechnologie et Signalisation Cellulaire, UMR 7242, CNRS, Université de Strasbourg, Boulevard Sébastien Brandt, BP 10413, 67412 Illkirch cedex, France

[‡]Département de Biologie Structurale Intégrative, Institut de Génétique et de Biologie Moléculaire et Cellulaire, UMR 7104/INSERM U964, CNRS, Université de Strasbourg, 1 rue Laurent Fries, 67404 Illkirch, France

[§]Equipe Biophysique et Biologie Structurale, UPR 9002, CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg, France

Supporting Information

ABSTRACT: PDZ domains are highly abundant protein−protein interaction modules commonly found in multidomain scaffold proteins. The PDZ1 domain of MAGI-1, a protein present at cellular tight junctions that contains six PDZ domains, is targeted by the E6 oncoprotein of the high-risk human papilloma virus. Thermodynamic and dynamic studies using complementary isothermal titration calorimetry and nuclear magnetic resonance (NMR) ^{15}N heteronuclear relaxation measurements were conducted at different temperatures to decipher the molecular mechanism of this interaction. Binding of E6 peptides to the MAGI-1 PDZ1 domain is accompanied by an unusually large and negative change in heat capacity (ΔC_p) that is attributed to a disorder-to-order transition of the C-terminal extension of the PDZ1 domain upon E6 binding. Analysis of temperature-dependent thermodynamic parameters and ^{15}N NMR relaxation data of a PDZ1 mutant in which this disorder-to-order transition was abolished allows the unusual thermodynamic signature of E6 binding to be correlated to local folding of the PDZ1 C-terminal extension. Comparison of the exchange contributions observed for wild-type and mutant proteins explains how variation in the solvent-exposed area may compensate for the loss of conformational entropy and further designates a distinct set of a few residues that mediate this local folding phenomena.



Protein interactions between domains of large globular proteins and short linear motifs are involved in many cellular mechanisms, including signaling pathways, protein trafficking, and post-translational modifications. These interactions are often mediated by small protein domains such as SH3, WW, and PDZ, which are very abundant in the proteomes of higher eukaryotes. Of these, the PDZ domain, which is a small globular domain of ∼90 residues, is one of the most abundant with nearly 270 occurrences distributed over 150 human proteins.[1] Many PDZ domain-containing proteins are located at the interface between the cytoskeleton and the cellular membrane where they are involved in the formation of cellular junctions such as synapses or adherens and tight junctions. They commonly form complexes associated with the regulation of cell growth, cell polarity, maintenance of cell adhesion, and signal transduction pathways.[2]

PDZ domains interact specifically with proteins containing short sequence motifs located in most cases at the C-terminus with dissociation constants in the micromolar range. Over the past 15 years, significant research efforts have been devoted to deciphering the rules that govern the specificity of these interactions, a goal that has proven more demanding than initially expected. Indeed, this research has revealed an important role for regions flanking PDZ domains in the molecular recognition process,[1,3] and numerous recent studies have highlighted subtle molecular mechanisms in which the binding of a given peptide to a PDZ domain is modulated by the protein environment.[4−7]

Because of the importance of PDZ domains in cell regulation processes, viruses, such as hepatitis B, adenovirus, influenza, and papillomavirus, target proteins containing PDZ domains as part of their host cell hijacking strategies. Interestingly, the number of PDZ-containing proteins targeted by viral proteins is small and seems to be common among several different viruses. One of these proteins, membrane-associated guanylate kinase with inverted domain structure-1 (MAGI-1), is found in tight

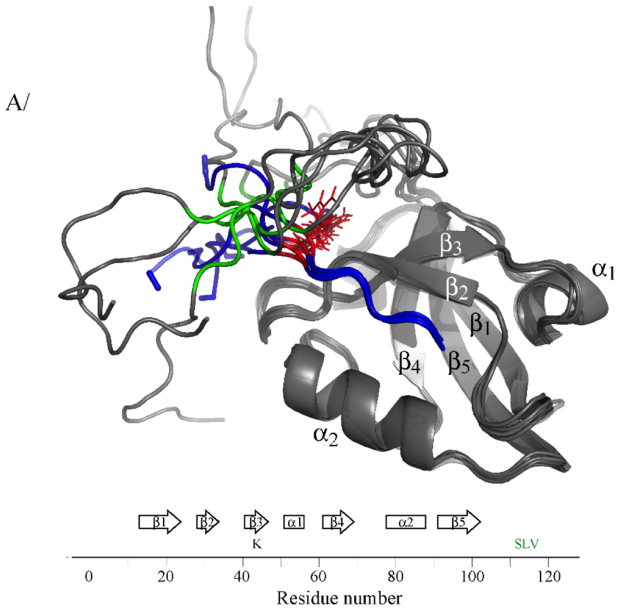
Received: July 9, 2014
Revised: December 1, 2014

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

junctions of most epithelial cells, where it is involved in the assembly and regulatory pathways of these cell−cell junctions.[8] MAGI-1 is a multidomain protein encompassing two WW, one guanylate kinase-like, and six PDZ domains numbered from PDZ0 to PDZ5.

PDZ-binding motifs targeting the PDZ domains of the MAGI-1 protein are found in several viral proteins: Tax from human T-cell leukemia virus type 1 (HTLV-1),[9] NS1 from the influenza virus,[10] E4-ORF1 from adenovirus type 9, and E6 from the human papilloma virus (HPV).[11] In particular, MAGI-1 has been recently identified as a major PDZ-containing target of high-risk HPV type 16 and 18.[12] The E6 protein of these viral strains harbors a C-terminal PDZ-binding motif that targets the second PDZ domain of MAGI-1[13] (hereafter called MAGI-1 PDZ1), leading to the degradation of MAGI-1.[11] The loss of MAGI-1 might favor the development of invasive cervical cancer by disrupting tight junctions and inducing pathologic epithelial hyperplasia. Such cell growth is not induced by low-risk HPVs encoding an E6 protein devoid of a PDZ-binding motif or by high-risk E6 mutants in which this motif has been deleted.[14]

To investigate the molecular mechanisms underlying the specificity of MAGI-1 targeting by HPV E6 proteins, we have previously studied complexes between the PDZ1 domain of MAGI-1 and C-terminal sequences of the E6 protein.[15,16] These studies highlighted the role of protein sequences flanking the canonical region of PDZ1 in the binding process, adding to the growing body of evidence on the functional importance of flanking regions in molecular interactions mediated by PDZ domains.[6,7] The solution structure of free MAGI-1 PDZ1 (PDB: 2KPK) revealed a highly flexible C-terminal tail (residues Y_{101}−P_{125}), whereas the same tail undergoes partial ordering in the presence of a peptide derived from the C-terminus of HPV16 E6[16] (Figure 1A). In the present study, we further investigate the thermodynamic and dynamic characteristics of MAGI-1 PDZ1 at different temperatures using isothermal titration calorimetry (ITC) and nuclear magnetic resonance (NMR) backbone dynamics measurements to delineate the contribution of this disorder-to-order transition to the binding mechanism.



B/

High-risk HPV	-10	-5	0
HPV16_E6ct	R	S	RETQL
HPV18_E6ct	E	R	RETQV
HPV45_E6ct	E	R	RETQV
HPV31_E6ct	C	R	RETQV
HPV51_E6ct	Q	R	RETQV
HPV58_E6ct	C	R	RETQV
Low-risk HPV			
HPV06_E6ct	C	W	TTCMEDMLP
HPV11_E6ct	C	W	TTCMEDLLP
HPV42_E6ct	C	R	GQCVERRLP
HPV44_E6ct	C	W	TSCMETILP
Peptides used in the present work			
16E6ct L0	R	S	RETQL
16E6ct L0|V	R	S	RETQV
16E6ct R-5|A	R	S	ARETQV

Figure 1. Global view of the MAGI-1 PDZ1 domain and the $16E6_{ct}$ peptides used in the present study. (A) Ribbon view of an ensemble of five representative structures of the wild-type MAGI-1 PDZ1 (residues 456−580, gray color) in complex with the $16E6_{ct}$ L_0|V peptide (blue color) on the basis of NMR structures of bound MAGI-1 PDZ1 (PDB: 2KPL). Arrows and boxes indicate the position of secondary structure elements β-strands and α-helices, respectively, and letters indicate the position of mutations. The $S_{113}L_{114}V_{115}$ residues are highlighted in green in the structures, and the side chains of the R_{-5} residues are indicated by red sticks. All structure images were made using PyMOL.[55] For the sake of clarity, residues 456−580 of full-length MAGI-1 PDZ1 are numbered from 1 to 125 according to the assignment previously published.[56] (B) Sequence alignment of ten E6 C-terminal sequences representative of high- and low-risk types of HPV. The peptides are numbered backward from 0 to −10. Below that are the sequences for the three $16E6_{ct}$ peptides used in the present study.

MATERIALS AND METHODS

Preparation of MAGI-1 PDZ1 Samples. Samples of unlabeled, ^{15}N-labeled, and ^{15}N-,^{13}C-labeled proteins (wild-type or SLV|GGG mutant) encompassing the PDZ1 domain (456−580) of MAGI-1 were expressed and purified as described previously[17] in 20 mM phosphate buffer (pH 6.8) with 200 mM NaCl and 2 mM TCEP. This same buffer was used for both the ITC and NMR experiments.

Synthetic Peptides. The peptide sequences used in this study were RSSRTRRETQL, RSSRTRRETQV, and RSSRTARETQV for $16E6_{ct}$ L_0, $16E6_{ct}$ L_0|V, and $16E6_{ct}$ R_{-5}|A, respectively (Figure 1B). The peptides were synthesized by the Chemical Peptide Synthesis Service at IGBMC using Fmoc chemistry and an Applied Biosystems 430A automated solid-phase peptide synthesizer. The peptides were purified by reverse-phase HPLC using a C-18 preparative Luna column, and their molecular weights were verified by electrospray ionization−mass spectrometry (ESI−MS) performed with an ESI/TOF Bruker MicrOTOF Focus instrument.

Prior to use, the peptides were dissolved in neutralized Milli-Q water (pH 7) with 0.1 M NaOH and passed through a Sephadex G-25 column (PD-10 desalting columns, GE Healthcare). The desalted fractions containing the peptide were lyophilized and dissolved in the same buffer as PDZ1. ^1H 1D NMR spectroscopy was employed to estimate the concentrations of PDZ1 and peptide stock solutions by comparing the proton signal from the sample to the signal obtained with a known tryptophan solution added to the NMR sample.[18]

Isothermal Titration Calorimetry Experiments. Isothermal titration calorimetry experiments were performed with a MicroCal ITC200 instrument (GE Healthcare). All concentrated stocks of protein and peptide samples were prepared

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

Table 1. Temperature Dependence of Thermodynamic Parameters for the Binding of Several Combinations of MAGI-1 PDZ1 or SLV|GGG and 16E6$_{ct}$ Peptides As Characterized by ITC in 20 mM Sodium Phosphate (pH 6.8), 200 mM NaCl, and 2 mM TCEP[a]

partners	T (K)	K_D (μM)	n	ΔH (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})	ΔG (kcal mol^{-1})	c_{min}/c_{max}
PDZ1 WT	283	0.76 ± 0.32	0.8 ± 0.2	4.4 ± 0.7	−12.3 ± 0.5	−7.9 ± 0.3	14/37
+ E6 L_0	288	1.72 ± 0.18	0.9 ± 0.1	2.1 ± 0.2	−10.2 ± 0.5	−8.0 ± 0.1	21
	293	1.34 ± 0.32	1.1 ± 0.1	0.5 ± 0.1	−8.1 ± 0.2	−7.6 ± 0.1	30/37
	298	1.73 ± 0.11	0.9 ± 0.1	−1.0 ± 0.1	−7.0 ± 0.1	−8.0 ± 0.2	21
	303	0.69 ± 0.05	0.8 ± 0.1	−3.4 ± 0.2	−5.2 ± 0.3	−8.6 ± 0.1	25/27
PDZ1 WT	283	0.49 ± 0.12	0.7 ± 0.1	3.0 ± 0.8	−11.1 ± 0.9	−8.1 ± 0.2	26/34
+ E6 L_0|V	288	0.54 ± 0.41	0.7 ± 0.1	1.5 ± 0.5	−9.9 ± 0.2	−8.5 ± 0.3	14/60
	293	0.99 ± 0.23	0.7 ± 0.1	−1.8 ± 0.4	−6.3 ± 0.5	−8.0 ± 0.1	9/15
	298	0.20 ± 0.05	0.9 ± 0.3	−3.4 ± 0.4	−5.7 ± 0.6	−9.1 ± 0.3	28/56
	303	0.33 ± 0.08	0.9 ± 0.2	−6.1 ± 0.5	−2.9 ± 0.7	−9.0 ± 0.2	34/74
PDZ1 SLV|GGG	283	1.70 ± 0.13	0.8 ± 0.1	−1.4 ± 0.3	−6.0 ± 0.1	−7.5 ± 0.1	9/41
+ E6 L_0|V	288	1.60 ± 0.14	0.9 ± 0.1	−2.3 ± 0.2	−5.4 ± 0.1	−7.6 ± 0.1	15
	293	1.52 ± 0.12	0.8 ± 0.1	−3.2 ± 0.3	−4.6 ± 0.1	−7.8 ± 0.1	14
	298	1.05 ± 0.10	1.0 ± 0.1	−4.0 ± 0.3	−4.1 ± 0.1	−8.2 ± 0.1	26
	303	0.88 ± 0.11	0.9 ± 0.1	−4.4 ± 0.2	−3.9 ± 0.1	−8.4 ± 0.1	21/81
PDZ1 WT	283	4.1 ± 1.6	0.9 ± 0.1	1.7 ± 0.9	−8.5 ± 0.5	−6.8 ± 0.8	9
+ E6 R_{-5}|A	288	2.0 ± 0.3	0.8 ± 0.1	0.9 ± 0.3	−8.4 ± 0.2	−7.5 ± 0.3	8
	293	7 ± 9	1.8 ± 1.0	−0.1 ± 0.4	−6.8 ± 0.8	−6.9 ± 0.9	7
	298	5.6 ± 0.7	0.9 ± 0.1	−1.6 ± 0.4	−5.6 ± 0.2	−7.2 ± 0.3	7
	303	5.8 ± 1.3	0.9 ± 0.1	−2.7 ± 0.9	−4.6 ± 0.4	−7.3 ± 0.7	6/8

[a]Values reported here result from independent experiments performed singly, in duplicate, or in triplicate with different protein or peptide batches. Protein concentrations vary between 20 and 40 μM, whereas the peptide concentration is fixed at 400 μM. The last column contains the minimal and maximal values of the ITC c parameter calculated as $c = n[\mathrm{PDZ1}]/K_D$.

and dialyzed in the target buffer and then diluted with the same buffer to achieve the desired concentrations. A typical experiment was performed as follows: 200 μL of PDZ1 (with concentrations ranging from 20 to 40 μM) was introduced into the cell, and 40 μL of peptide solution (at 0.4 mM) was loaded into the syringe. The stirring speed was set at 1000 rpm, and the power of the reference at 3−5 μcal/s. After temperature equilibration was reached, each titration began by the addition of 0.3 or 0.7 μL (not used when fitting the data) followed by 29 or 19 additions of 1 or 2 μL during 2 s at intervals of 180 or 120 s, respectively. Most of the time, the ITC experiments were performed in duplicate or triplicate using different PDZ1 and peptide batches. For all titrations, the parameter c ($c = n[\mathrm{PDZ1}]/K_D$) ranged between 6 and 84 (Table 1), close to the optimal range of c values ($10 < c < 100$) required for the accurate determination of K_D.[19,20]

The extraneous heats of mixing and dilution were corrected by conducting control experiments[21] through titration of the peptide into the buffer (dilution of the peptide) and titration of the buffer into the PDZ1 sample (dilution of PDZ1). MicroCal Origin 7.0 was used to correct for the heats of mixing and dilution and to analyze the data. The thermodynamic parameters (enthalpy change (ΔH), entropy change (ΔS), affinity (K_D), and stoichiometry (n)) were obtained using a one-binding site model.

NMR Experiments. Samples of unlabeled, ^{15}N-labeled, and ^{15}N,^{13}C-labeled MAGI-1 PDZ1 (wild-type and mutant) were prepared in 20 mM phosphate buffer (pH 6.8) with 200 mM NaCl and 2 mM TCEP at protein concentrations between 300 and 600 μM. The MAGI-1 PDZ1/16E6$_{ct}$ L_0|V complex was prepared by adding a 3-fold excess of the synthesized 11-mer 16E6$_{ct}$ L_0|V peptide.

NMR experiments were performed at 283 and 303 K on a Bruker DRX 600 MHz spectrometer equipped with a triple-resonance cryoprobe with z-gradient. A set of three-dimensional triple-resonance experiments (HN(CO)CA, HNCA, and CBCA(CO)NH) was recorded to propagate backbone resonance assignments obtained in the wild-type protein to the SLV|GGG mutant. All spectra were processed using NMRPipe[22] and analyzed using Sparky3.[23]

The composite chemical shift perturbation per residue n was calculated using ^{1}H and ^{15}N resonances assigned in the spectra of both MAGI-1 PDZ1 wild-type and MAGI-1 PDZ1 SLV|GGG, as

$$\Delta\delta_n = \sqrt{\sum_{i\in\{^1\mathrm{H},^{15}\mathrm{N}\}} \left(\frac{\delta_{i,\mathrm{SLV|GGG}} - \delta_{i,\mathrm{wt}}}{\sigma_i}\right)^2} \quad (1)$$

where σ_i represents the standard deviation in the chemical shift of atom i (0.6 and 3.75 for ^{1}H and ^{15}N, respectively) as determined using the BMRB Web site (http://www.bmrb.wisc.edu).[24]

^{15}N R_1 and R_2 relaxation rates were measured using a series of ^{1}H−^{15}N heteronuclear single quantum coherence (HSQC)-type spectra as previously described.[25] For ^{15}N R_1 relaxation, intensities were extracted from a set of 14 spectra recorded with relaxation delay values ranging from 10 to 2000 ms. For ^{15}N R_2 relaxation, intensities were extracted from a set of 12 spectra recorded with relaxation delay values ranging from 0 to 160 ms with ^{15}N 180° pulses applied every 1.2 ms at a field strength of 4.2 kHz to average the chemical shielding anisotropy (CSA)-dipolar cross-correlated relaxation. R_1 and R_2 relaxation rates were obtained by fitting the measured peak intensities with a 2 parameter single exponential using the Python Numpy non-linear optimization toolbox. Confidence levels were estimated using 100 Monte Carlo calculations. Heteronuclear {^{1}H}−^{15}N nuclear Overhauser effects (NOEs) were measured from two experiments with and without proton saturation. Proton saturation was achieved by applying a train of ^{1}H 180° pulses at 2 ms intervals for 4 s prior to the first ^{15}N excitation pulse.

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

Order parameters (O^2) and exchange contributions to transverse relaxation values (R_{ex}) and its uncertainties (δR_{ex}) were calculated using Tensor2 software.[26] Correlation times were obtained using a fully anisotropic rotational diffusion model for the protein's global motions.

Solvent accessible surface areas (ASA) were estimated with the PyMol `get_area` function using settings for solvent radius and solvent density of 1.4 Å and 4, respectively. The ASA for the complex was calculated using the structures of PDZ1 MAGI-1 in complex with $16E6_{ct}$ (PDB: 2KPL), whereas the ASA for the free peptide and protein was calculated assuming either a partially folded or fully extended conformation for the C-terminal PDZ domain. The difference between these calculations was used to provide an estimate of the range of the ΔASA differences between the free and bound molecules.

RESULTS

Thermodynamic Parameters of $16E6_{ct}$ L_0|V Peptide Binding to MAGI-1 PDZ1. Isothermal titration calorimetry experiments were conducted to measure the thermodynamic parameters of binding between various peptides encompassing the last 11 residues from HPV E6 proteins and the MAGI-1 PDZ1 domain (Figure 2A). We first used the $16E6_{ct}$ wild-type peptide and a $16E6_{ct}$ peptide in which the C-terminal leucine residue of the HPV16 E6 sequence is replaced by a valine residue (hereafter referred to as $16E6_{ct}$ L_0|V), as is found in the HPV18 strain. This mutation resulted in a slight increase in the binding affinity at 303 K from 0.69 ± 0.05 μM for $16E6_{ct}$ to 0.33 ± 0.08 μM for $16E6_{ct}$ L_0|V (Table 1), which is in agreement with affinities measured by surface plasmon resonance (SPR).[17]

The averaged thermodynamic parameters obtained from several isotherms recorded for the binding of $16E6_{ct}$ L_0|V to MAGI-1 PDZ1 indicate that the binding energy at 303 K is provided by both favorable enthalpic and entropic contributions of −6.1 ± 0.5 and +2.9 ± 0.7 kcal mol^{-1}, respectively. Further insight into the binding was obtained by measuring the thermodynamic parameters as a function of temperature (Figure 2B). The binding enthalpy of $16E6_{ct}$ L_0|V displays significant variation from a slightly endothermic to a strongly exothermic reaction when the temperature is varied from 283 to 303 K, respectively. This gain in binding enthalpy is almost fully compensated for by a loss of binding entropy over this range of temperatures, leading to a roughly constant K_D value of 0.5 ± 0.2 μM (Figure 3). The linear variation of binding enthalpy (ΔH) with temperature suggests that the measured thermodynamic parameters result purely from the binding event and not from any additional temperature-induced conformational change. The change of heat capacity (ΔC_p) due to peptide binding was estimated from the slope of the linear regression using the definition of heat capacity change at constant pressure.

$$\Delta C_p = \frac{\partial H}{\partial T} \quad (2)$$

This analysis gives a heat capacity change of −470 ± 45 cal mol^{-1} K^{-1} upon binding of $16E6_{ct}$ L_0|V to the MAGI-1 PDZ1 domain.

Thermodynamic Contributions from the Flanking Sequences of MAGI-1 PDZ1 and E6 Peptide. Our previous work on the solution structure of both the apo and $16E6_{ct}$ L_0|V-bound forms of the MAGI-1 PDZ1 domain highlighted the role of the C-terminal extension of the PDZ1 domain in the peptide recognition process.[16] The observation of weak intermolecular NOEs among three residues of the C-terminal extension, namely S_{113}, L_{114}, and V_{115}, and the well conserved R_{-5} residue of the E6 peptide (Figure 1B) suggested an active contribution of this region of the PDZ1 domain to peptide recognition and prompted us to design a mutant in which these three residues were replaced with glycines (hereafter named SLV|GGG).[16] Glycine was chosen instead of alanine, the common neutral substitution, to both disrupt potential hydrophobic interactions and introduce higher conformational flexibility within this region. The $^1H-^{15}N$ HSQC spectra of the mutant and wild-type proteins are almost superimposable with the exception of the mutated residues, indicating that the mutation did not alter the protein fold (Figure S1A in the Supporting Information). Conversely, the binding of a peptide derived from $16E6_{ct}$ L_0|V with the arginine residue R_{-5} mutated to an alanine (named $16E6_{ct}$ R_{-5}|A) to MAGI-1 PDZ1 was also studied. The SLV|GGG mutation on the PDZ1 domain led to a loss of affinity for $16E6_{ct}$ L_0|V by a factor of ∼3, resulting from a loss in ΔH that is partially compensated for by a gain in the entropic contribution at 303 K (Figure 4A). Disruption of the interaction through a R_{-5}|A mutation on the peptide led to an increase of the dissociation constant by a factor of 20 at 303 K (Table 1), resulting mostly from a loss of 3.4 kcal mol^{-1} in the binding enthalpy.

Differences in ΔC_p changes induced by the protein and peptide mutations were estimated by measuring thermodynamic parameters between 283 and 303 K. Notably, both mutations led to less negative values of the ΔC_p changes [from −470 ± 45 cal mol^{-1} K^{-1} for the wild-type complex to −240 ± 40 and −150 ± 20 cal mol^{-1} K^{-1} for the R_{-5}|A peptide and SLV|GGG protein mutations, respectively (Figure 4B)]. It has been proposed that the hydrophobic effect is the dominant contributor to ΔC_p, and large negative values of ΔC_p changes are often considered as a thermodynamic signature of processes that remove apolar molecular surfaces from water.[27−29] The reduction in ΔC_p values observed for the SLV|GGG mutant is significantly larger than expected if only local hydration changes of the mutated leucine and valine residues are considered. Furthermore, circular dichroism measured at 208 nm as a function of temperature indicates that the SLV|GGG mutation does not affect the melting temperature (around 338 K) and shows that the secondary structure content of both proteins is preserved within the temperature range used in this study (283−303 K) (Figure S2 in the Supporting Information). Altogether, our data suggest that the unusually large change in the ΔC_p values observed upon binding of E6 peptide to the wild-type MAGI-1 PDZ1 domain results from the disorder-to-order transition of its C-terminal extension. These observations prompted us to further investigate changes in the dynamic behavior experienced by the PDZ1 domain upon peptide binding using ^{15}N heteronuclear relaxation measurements at 283 and 303 K.

Changes of Protein Backbone Motion upon $16E6_{ct}$ L_0|V Peptide Binding. Heteronuclear ^{15}N transverse (R_2) and longitudinal (R_1) relaxation rates were measured for the PDZ1 domain of MAGI-1 in both free and peptide-bound states at 283 and 303 K (Figure 5 and Table S1 in the Supporting Information). At 283 K, binding of $16E6_{ct}$ L_0|V induces major changes in the R_2/R_1 ratios (Figure 5A). The most striking differences are located within the C-terminal part of the domain, encompassing residues $F_{105}-E_{124}$, where the R_2/R_1 values of the peptide-bound protein are systematically larger than those measured for the free protein. Furthermore, large R_2 relaxation rates were measured at several locations in the core of the

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

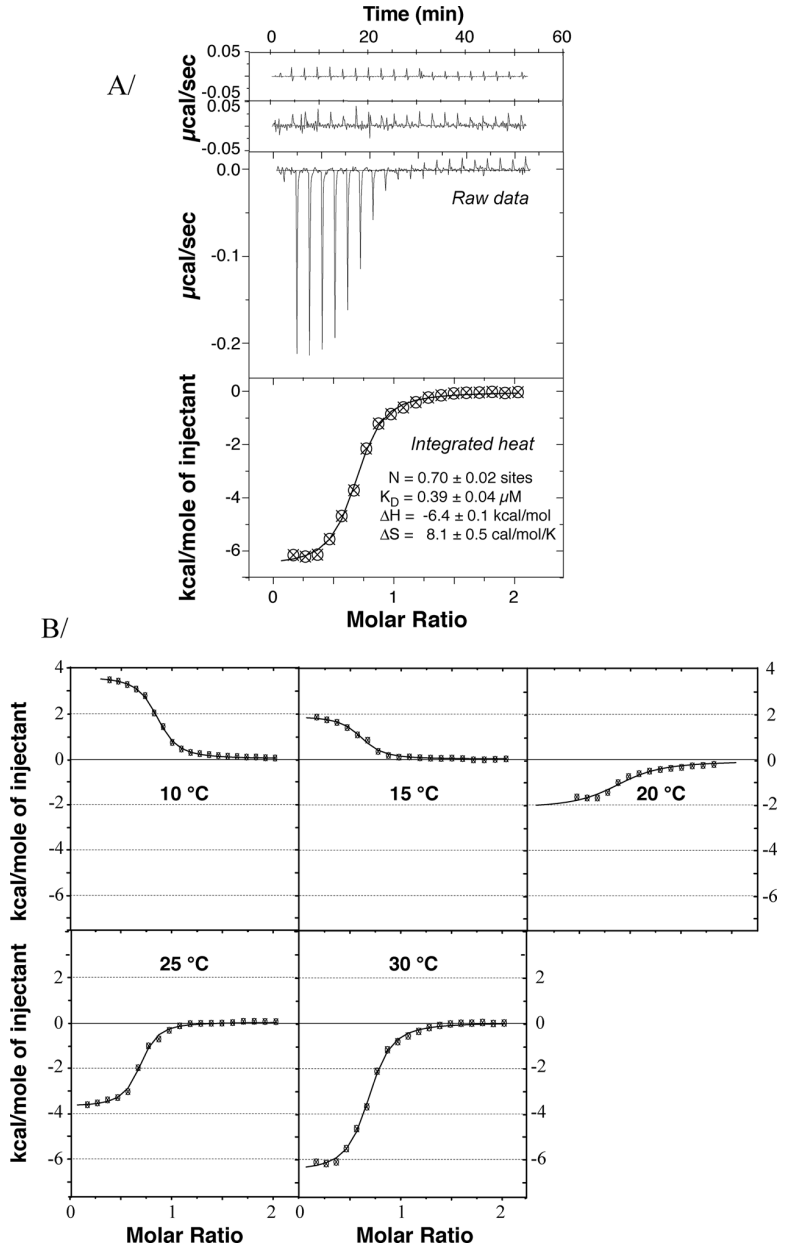


Figure 2. ITC measurements for MAGI-1 PDZ1−E6 peptide interactions. (A) Typical ITC experiments of 16E6$_{ct}$ L_0|V with the MAGI-1 PDZ1 domain at 303 K in 20 mM sodium phosphate buffer, pH 7.5, 200 mM NaCl, and 2 mM TCEP. Shown from top to bottom are the raw titration data of buffer injected into MAGI-1 PDZ1, 16E6$_{ct}$ L_0|V injected into buffer, and 16E6$_{ct}$ L_0|V injected into MAGI-1 PDZ1, and the integrated heat measurements for the titration of MAGI-1 PDZ1 with 16E6$_{ct}$ L_0|V corrected from the reference, respectively. The first injection peak was systematically discarded because of the backlash effect of the syringe plunger. The thermodynamic parameters (affinity constant (K_D), enthalpy change, entropy change, and stoichiometry number (n)) obtained for one typical data set by fitting a standard 1:1 interaction model are reported with the associated standard deviations determined by nonlinear least-squares analysis. (B) Representative integrated heat measurements for the titration of MAGI-1 PDZ1 with 16E6$_{ct}$ L_0|V at temperatures ranging from 283 to 303 K are reported. Each titration was recorded in at least duplicate.

peptide-bound protein, leading to R_2/R_1 values significantly larger than the average ratio within the core of PDZ1. This line broadening, observed for residues located within the C-terminal tail and for residues T_{30}, V_{31}, V_{32}, K_{44}, S_{45}, D_{49}, G_{74}, H_{77}, and S_{86} within the core domain, reflects exchange contributions in these regions. Most of these contributions vanish at 303 K, except those located at the C-terminal part of the domain, suggesting that the associated motions are occurring on a rather fast to intermediate (μs) time scale (Figure 5B).[30]

When the temperature is changed, the average R_2/R_1 ratio is decreased as a result of the lower correlation time expected at higher temperature. The observed decrease in correlation time from 13.2 to 7.0 ns for the unbound form (and from 14.2 to 7.6 ns for the bound form) when the temperature is increased from 283 to 303 K corresponds to a ratio of correlation times of 1.89 (1.87 for the bound form), which is close to the expected ratio (1.75) calculated using the Stokes−Einstein equation.[31] The R_2/R_1 ratios of the C-terminal tail for the unbound form

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

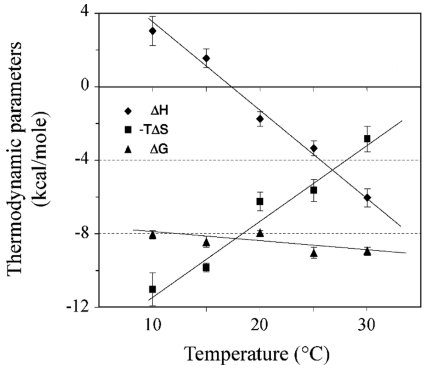


Figure 3. Effect of temperature on the thermodynamic parameters of 16E6$_{ct}$ L$_0$|V binding to the MAGI-1 PDZ1 domain as determined by ITC. 16E6$_{ct}$ L$_0$|V and MAGI-1 PDZ1 were in 20 mM sodium phosphate (pH 6.8), 200 mM NaCl, and 2 mM TCEP. Average values of $\Delta H°$ (◆), $\Delta G°$ (▲), and $-T\Delta S$ (■) were determined from three distinct experiments and plotted as a function of temperature. Error bars represent the standard deviation. The slope of the linear regression fit was used to derive a ΔC_p value of −470 ± 45 cal mol^{-1} K^{-1}.

are distinctively below those of the core of the PDZ1 domain, indicating that this C-terminal region is mostly in a disordered state.

Apart from the observed changes in the exchange contributions to ^{15}N line broadening, there is no evidence of backbone stiffening upon peptide binding except for residue L$_{54}$, which in the complex consistently displays R_2/R_1 ratios that are higher than in the free protein. Changes in the high frequency (ps−ns time scale) motions upon peptide binding were revealed by the analysis of {^{1}H}−^{15}N heteronuclear NOEs, a measurement free of exchange contributions (Figure 5C and Table S2 in the Supporting Information). The profiles recorded for the free and peptide-bound PDZ1 domains at 283 K show that residues F$_{105}$−E$_{124}$ become more ordered upon peptide binding. It is worth noting that this ordering is not complete and that residual motions on this time scale are still present in the complex. Interestingly, the most ordered region of the C-terminal tail of bound PDZ1 corresponds to residues S$_{113}$, L$_{114}$, and V$_{115}$ that interact directly with the 16E6$_{ct}$ L$_0$|V peptide.

Backbone Dynamics of the SLV|GGG Mutant of MAGI-1 PDZ1. The SLV|GGG mutation did not lead to significant changes in backbone chemical shifts of the free PDZ1 domain; observed differences were restricted to the vicinity of the mutation (Figure S1B in the Supporting Information). As for the wild-type domain, exchange between free and bound forms of PDZ1 SLV|GGG occurs in the slow-exchange regime, as expected from its similar affinity for 16E6$_{ct}$ L$_0$|V (Table 1). However, at high temperature, the sample half-life of the mutant was found to be shorter than for the wild-type protein, preventing NMR relaxation experiments from being conducted at 303 K. The impact of the SLV|GGG mutation on the dynamic properties of the MAGI-1 PDZ1 domain was therefore investigated at 283 K using heteronuclear ^{15}N relaxation rate measurements conducted for both the free and peptide-bound states of the mutated protein. As for the free wild-type protein, the C-terminal tail of the free SLV|GGG mutant is mainly disordered (compare Figures 4A and 5A). The dynamic behaviors of the free and bound states of PDZ1 SLV|GGG are highly similar (Figure 6A and B and Table S1 in the Supporting Information), indicating that the distribution of high frequency

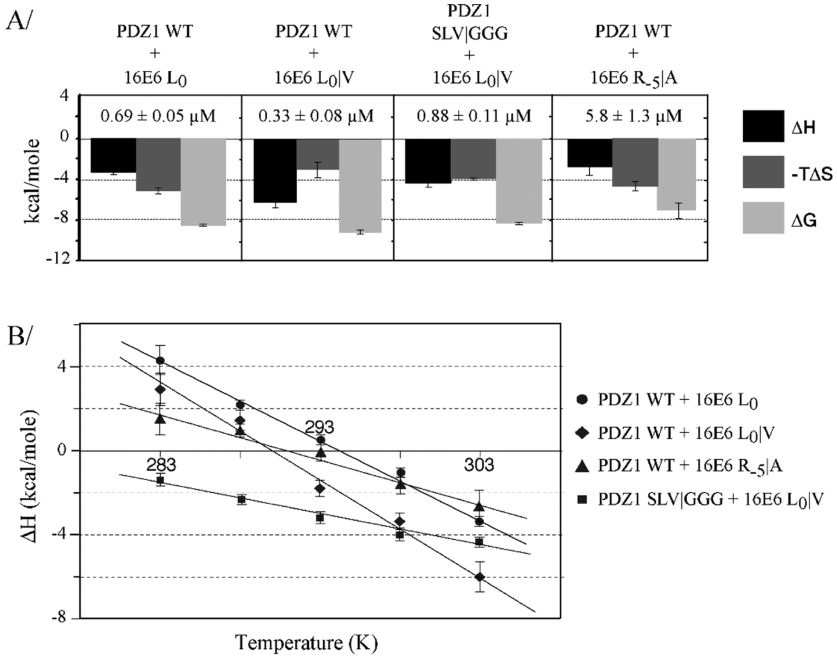


Figure 4. Effect of PDZ1 and E6 peptide mutations on thermodynamic parameters. (A) Values of ΔH, $-T\Delta S$, and ΔG (from dark to light gray, respectively) are plotted for different combinations of PDZ domains and 16E6$_{ct}$ peptides recorded at 303 K. Values are determined by averaging at least two distinct experiments, and error bars represent the standard deviation. Data are summarized in Table 1. (B) ITC experiments were performed at different temperatures for several combinations of PDZ domains and peptides, and the temperature dependence of the enthalpy was used to derive the ΔC_p values assuming ΔC_p to be temperature-independent: MAGI-1 PDZ1 vs 16E6$_{ct}$ L$_0$ (●, −350 ± 25 cal mol^{-1} K^{-1}), MAGI-1 PDZ1 vs 16E6$_{ct}$ L$_0$|V (◆, −470 ± 45 cal mol^{-1} K^{-1}), MAGI-1 PDZ1 vs 16E6$_{ct}$ R$_{-5}$|A (▲, −240 ± 40 cal mol^{-1} K^{-1}), and MAGI-1 PDZ1 SLV|GGG vs 16E6$_{ct}$ L$_0$|V (■, −150 ± 20 cal mol^{-1} K^{-1}).

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

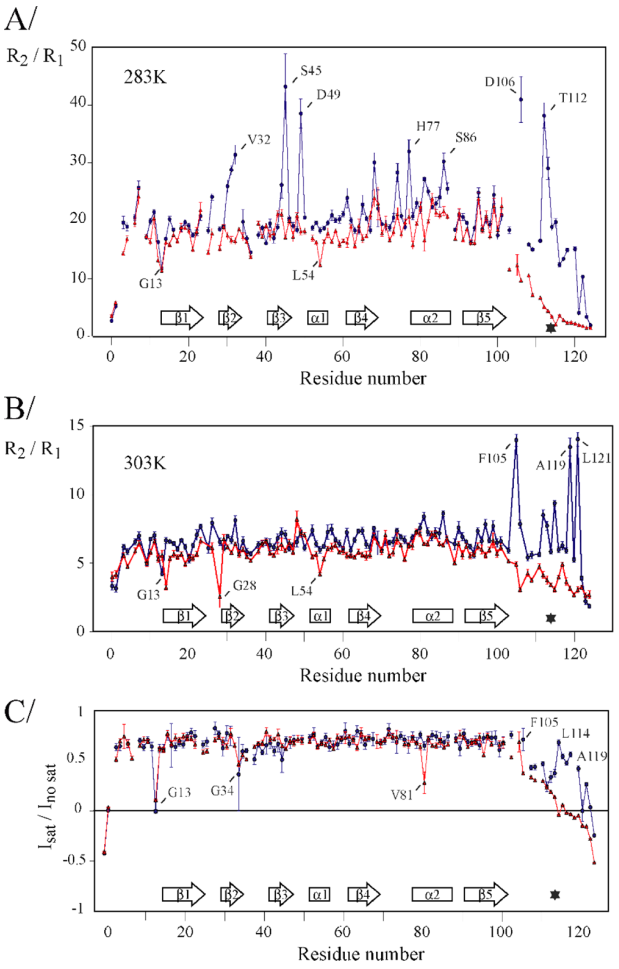


Figure 5. Wild-type PDZ1 temperature-dependent heteronuclear ^{15}N relaxation data. Plots of R_2/R_1 ratios as a function of residue number for the wild-type PDZ domain in the free state (red line) and in complex with 16E6 L_0IV (blue line) at 283 K (A) and 303 K (B). Relaxation data were measured at a 600 MHz ^1H frequency. Residues displaying significant changes in their relaxation parameters upon peptide binding are labeled. (C) $\{^1H\}-^{15}N$ heteronuclear NOE values measured for the free PDZ1 domain (red line) and its complex with 16E6 L_0IV (blue line) at 283 K. The asterisk indicates the position of the SLV_{115}IGGG mutation.

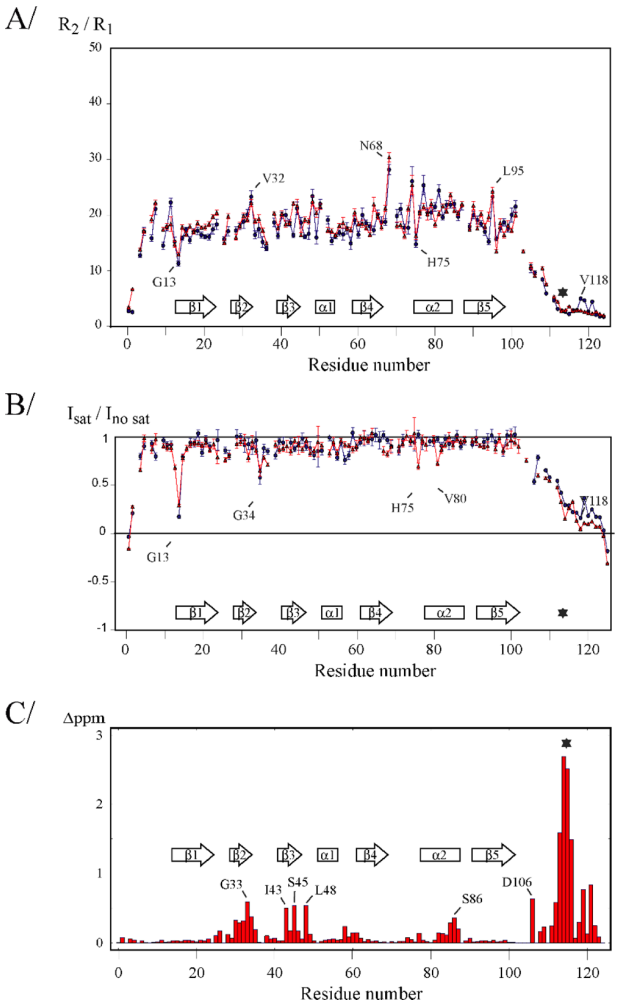


Figure 6. NMR characterization of the SLVIGGG mutant at 283 K in 20 mM phosphate buffer (pH 6.8), 200 mM NaCl, and 2 mM TCEP. (A) Plot of R_2/R_1 ratios against residue number for the apo SLVIGGG mutant PDZ domain (red line) and for its complex with the 16E6 L_0IV peptide (blue line) at 283 K. The asterisk indicates the position of the SLVIGGG mutation. The scale from Figure 5A has been retained to facilitate comparison. (B) $\{^1H\}-^{15}N$ heteronuclear NOE values measured for the apo SLVIGGG mutant (red line) and its complex with 16E6 L_0IV (blue line) at 283 K. (C) Composite chemical shift differences between 16E6 L_0IV bound forms of wild-type MAGI-1 PDZ1 and the SLVIGGG mutant. Aside from the mutation point indicated by an asterisk, residues displaying the largest differences are indicated.

motions throughout the protein backbone is not significantly altered upon peptide binding. In particular, the C-terminal tail in the complex remains largely disordered as it is in the free mutant form. Furthermore, the SLVIGGG mutation led to abolition of the widespread exchange contributions observed at 283 K for the complex between wild-type PDZ1 and the $16E6_{ct}$ L_0IV peptide.

Thus, the relaxation parameters measured for the peptide-bound form of the mutant at 283 K indicate that the SLVIGGG mutation induced a distinct dynamic behavior of the protein. This is supported by significant differences in chemical shifts between the bound states of the wild-type and mutant proteins (Figure 6C). In addition to the region flanking the site of mutation, significant differences were also observed around residue G_{33} in the β_2 strand, I_{43} and S_{45} in the β_3 strand, L_{48} in the $\beta_3-\alpha_1$ loop, S_{86} in the α_2 helix, and several residues in the C-terminal region, including D_{106}.

DISCUSSION

The specific dynamic behavior of the C-terminal region of the MAGI-1 PDZ1 domain upon binding to the viral HPV16 E6 oncoprotein observed in our previous study[16] prompted us to measure quantitative thermodynamic and dynamic data at different temperatures using complementary ITC and NMR relaxation spectroscopy approaches to gain further insights into the mechanisms underlying this molecular recognition process. In the present work, the dissociation constants obtained by ITC are consistent with those previously determined from surface plasmon resonance experiments[16] in which the SLVIGGG mutation at positions 113−115 led to a slight loss of affinity for $16E6_{ct}$ L_0IV at 298 K (K_D increased from 0.23 ± 0.02 to

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

1.17 ± 0.12 μM). These values closely match those determined by ITC at 298 K (K_D of 0.20 ± 0.05 and 1.05 ± 0.10 μM for the wild-type and mutant, respectively). Temperature-dependent ITC experiments further indicated that peptide binding to the MAGI-1 PDZ1 domain is characterized by a heat capacity change (ΔC_p) of −470 ± 45 cal mol^{-1} K^{-1}, which is significantly larger than values reported for other PDZ−peptide interactions. Indeed, the interaction between the PDZ10 domain of MUPP1 and the VSLRVSSV peptide was characterized by a ΔC_p value of −140 cal mol^{-1} K^{-1},[32] whereas a value of −160 cal mol^{-1} K^{-1} was found for the interaction between the PDZ3 domain of PSD95 and the YKQTSV peptide.[33] These reported ΔC_p values are similar to those of other systems involving the binding of a peptide to a protein domain, such as Grb2-SH2 binding to a phosphotyrosine peptide (ΔC_p = −146 cal mol^{-1} K^{-1}).[34]

It has been suggested that ΔC_p changes for protein−peptide interactions are dominated by hydration contributions ($\Delta C_{p,\mathrm{solv}}$) resulting from changes in the solvent-exposed area of the interacting molecules, whereas protein contributions ($\Delta C_{p,\mathrm{prot}}$) are less important.[28,35] Consequently, small ΔC_p values have been attributed to a lack of significant protein conformational changes upon complex formation,[36] whereas large variations in ΔC_p values are associated with folding upon binding events.[37] In the latter case, the change in protein hydration associated with protein folding leads to the exclusion of highly ordered water molecules from exposed hydrophobic areas.[29,36] The large negative ΔC_p values observed upon binding of E6 peptides to the PDZ1 domain of MAGI-1 suggest that this interaction involves partial ordering of the domain. In contrast, ΔC_p values measured for the SLV|GGG mutant are significantly reduced and closer to those reported for the PDZ domains of MUPP1 and PSD95 proteins, suggesting that this ordering event is abolished when the sequence of the C-terminal extension of the domain is altered. In addition, the sharp reduction in ΔC_p measured for the binding of an HPV16 E6 peptide in which the arginine at position −5 has been replaced by an alanine residue further suggests that this position in the ligand is crucial for this ordering process.

We previously found that the C-terminal extension of the PDZ1 domain of MAGI-1 undergoes a disorder-to-order transition upon binding to the 16E6$_{ct}$ L$_0$|V peptide.[16] This observation prompted us to examine whether this transition could account for the large change in heat capacity measured by ITC. From the solution structure ensemble of PDZ1 in complex with 16E6$_{ct}$ L$_0$|V (PDB: 2KPL), we extracted the coordinates for the C-terminal tail as well as for the complex formed by the PDZ1 core domain and the peptide. Depending on the C-terminal tail conformation used for calculations, we estimated that surfaces between 700 and 1000 Å^2 for polar residues and between 1000 and 1300 Å^2 for apolar residues were buried by the presence of the C-terminal tail of the PDZ1 domain (residues Y$_{101}$−P$_{125}$). The large fluctuations on these surfaces reflect the residual disorder that remains within the C-terminal extension in the PDZ−peptide complex. Nevertheless, these values allowed us to estimate the hydration heat capacity change ($\Delta C_{p,\mathrm{solv}}$) from the change of the polar ($\Delta \mathrm{ASA}_{\mathrm{pol}}$) and nonpolar ($\Delta \mathrm{ASA}_{\mathrm{ap}}$) solvent accessible surface areas using the empirical expression

$$\begin{aligned}\Delta C_{p,\mathrm{solv}} &= \Delta C_{p,\mathrm{solv,pol}} + \Delta C_{p,\mathrm{solv,ap}} \\ &= \lambda_p \Delta \mathrm{ASA}_{\mathrm{pol}} + \lambda_a \Delta \mathrm{ASA}_{\mathrm{ap}}\end{aligned} \quad (3)$$

where $\Delta C_{p,\mathrm{solv,pol}}$ and $\Delta C_{p,\mathrm{solv,ap}}$ are the polar and nonpolar contributions to the protein heat capacity with λ_a and λ_p equal to 0.45 ± 0.02 and −0.26 ± 0.03 cal mol^{-1} K^{-1} Å^{-2}, respectively.[38]

According to eq 3, C-terminal tail ordering would result in a $\Delta C_{p,\mathrm{solv}}$ change of about −300 ± 120 cal mol^{-1} K^{-1}, which compares remarkably well with the difference in ΔC_p between the wild-type and SLV|GGG mutant PDZ1 domains (320 ± 50 cal mol^{-1} K^{-1}) measured from ITC experiments. The corresponding gain in solvent entropy ΔS_{solv} was estimated from the heat capacity change using

$$\Delta S_{\mathrm{solv}} = \Delta C_{p,\mathrm{solv,pol}} \ln(T/T^*_{\mathrm{S,pol}}) + \Delta C_{p,\mathrm{solv,ap}} \ln(T/T^*_{\mathrm{S,ap}}) \quad (4)$$

where $T^*_{\mathrm{S,pol}}$ = 335.15 K and $T^*_{\mathrm{S,ap}}$ = 385.15 K are the polar and nonpolar reference temperatures, respectively, at which the hydration entropy is equal to zero.[38] This resulted in a ΔS_{solv} value of ∼30 kcal mol^{-1} at 283 K, indicating that burial of the PDZ domain surface by its C-terminal tail provides a significant amount of the peptide binding energy. This energy is partially counterbalanced by a loss of conformational entropy associated with the disorder-to-order transition. A rough estimate of this latter contribution can be obtained from the change in order parameters (ΔO^2) derived from the analysis of relaxation data using the Lipari and Szabo formalism (Table S3 in the Supporting Information).[39] It has been shown that the assumption of a simple motional model for ps−ns time scale bond vector dynamics allows conformational entropy changes to be derived from ΔO^2.[40] For the C-terminal region of the PDZ1, $\Delta O^2_{\mathrm{bound-free}}$ values are systematically positive with values ranging from 0.2 to 0.7, leading to a conformational entropy change of about 12 kcal mol^{-1} at 283 K. Although this value reflects the sole contribution of backbone motions and may therefore represent only a fraction of the conformational entropy change, it nevertheless provides an interesting indication that gain of solvent entropy (ΔS_{solv}) can significantly offset the entropic cost of ordering to favor formation of the complex.

Although weak intermolecular NOEs were observed between the SLV residues at position 115 of the C-terminal extension and the well conserved R$_{-5}$ residue of the high-risk HPV E6 peptide, their respective mutations do not lead to similar effects. The R$_{-5}$|A mutant induces drastic changes in the ^{1}H−^{15}N HSQC spectrum with large peak broadening (data not shown), and the large change in the dissociation constant observed for the R$_{-5}$|A peptide may result from the loss of additional interactions, probably electrostatic interactions between R$_{-5}$ and the negatively charged residues of the proximal DEPDE loop. In contrast to the difficulties characterizing the effect of peptide R$_{-5}$|A mutation by NMR, the SLV|GGG mutant enabled a complete description of the effect of the mutation at the residue level. Although the ^{1}H−^{15}N HSQC spectra of the free form of both domains are almost identical, when the mutant and wild-type PDZ1 domains are bound to the E6 peptide, differences arise in their chemical shifts (Figure 6C) and in their relaxation parameters, including the presence of exchange contributions to the transverse nitrogen relaxation rates at 283 K (Figures 5A and 6A). Altogether, these data suggest that the SLV|GGG mutation abolished the partial ordering of the C-terminal tail on the PDZ core domain surface that occurs upon peptide binding.

Despite the residual disorder affecting the C-terminal region of the PDZ domain bound to 16E6$_{ct}$ L$_0$|V, the identification of

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

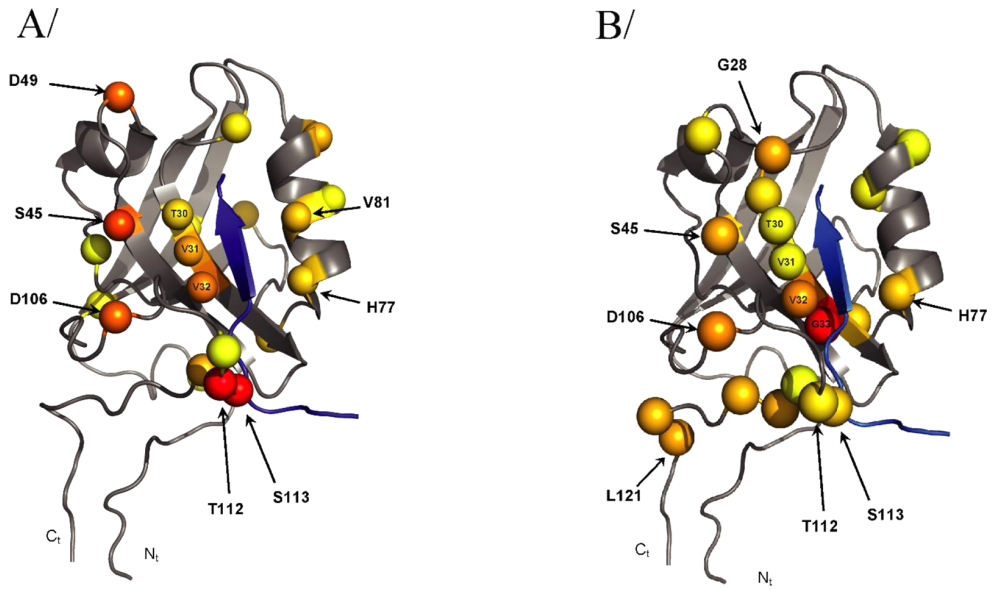


Figure 7. Wild-type PDZ1 surface residues involved in the disorder-to-order transition of the C-terminal extension. (A) Chemical exchange contributions ($R_{ex} \pm \delta R_{ex}$) obtained from ^{15}N heteronuclear relaxation data measured at 283 K and interpreted using Tensor2. To get a significant representation, residues with positive values of the difference ($R_{ex} - 3\delta R_{ex}$) are represented by spheres centered at the C_α positions and colored from yellow (0.0 s^{-1}) to red (30 s^{-1}). (B) Composite chemical shift differences observed for PDZ1 upon binding to 16E6$_{ct}$ L_0|V calculated according to eq 1. Residues with composite chemical shift differences larger than 0.5 ppm are represented by spheres centered at C_α positions and colored from yellow (0.5 ppm) to red (2 ppm). The ribbon representations of the complex were derived from PDB entry 2KPL[55] with the MAGI-1 PDZ1 domain in gray and 16E6 L_0|V in blue.

residues affected by exchange contributions upon peptide binding provides interesting indications about the position of this region relative to the core of the PDZ domain (Figure 7A and Table S4 in the Supporting Information). Indeed, these residues highlight a single side of the PDZ domain encompassing the β_2 and β_3 strands together with α_1 and α_2 helices. This surface patch overlaps with the peptide-binding groove that is formed by the β_2 strand and the α_2 helix. Thus, the combined effect of the peptide binding and the C-terminal tail ordering explains the unusual extent of PDZ chemical shift perturbations that are observed upon peptide binding (Figure 7B). Partial ordering of the PDZ1 C-terminal extension on this surface buries several hydrophobic residues, such as V_{32} and T_{30} on the core of the domain and L_{114} within the C-terminal domain, favoring complex formation through favorable entropic contributions. The strict conservation of these solvent-exposed hydrophobic residues among orthologous sequences of MAGI-1 suggests that the observed disorder-to-order transition of the C-terminal region of the PDZ1 domain results from evolutionary selection. Notably, the surface patch delineated by both large exchange contributions and large frequency differences between apo and peptide-bound proteins encompasses K_{44}, a residue located on the β_2 strand that has been shown to be important for binding specificity.[17] More recently, Kranjec et al. demonstrated that this residue is critical for recognition and degradation processes by HPV16 E6 of MAGI-1 in a full-length context.[41] Chemical shift differences between the free wild-type and a K_{44}|A mutant indicate that the K_{44}|A mutation significantly affects the frequencies of several residues of the C-terminal extension, particularly those around D_{106} (Figure S3 in the Supporting Information). Because a direct electrostatic interaction between K_{44} and the glutamate peptide at position −3 was observed,[17] the strong chemical shift perturbation of D_{106} upon K_{44}|A mutation suggests a close correlation between C-terminal extension ordering on the core PDZ surface and the peptide recognition mechanism.

A growing set of experimental observations combining structural information with data about the dynamics and thermodynamics of PDZ-containing proteins indicate that the PDZ scaffold has evolved into a complex allosteric machine linking the peptide binding event within the core domain to remote regions of the protein.[6] Several experimental and computational studies have highlighted the role of dynamic properties of PDZ domains in achieving these allosteric mechanisms, and networks of dynamically coupled residues have been identified.[42−46] The growing number of studies related to PDZ domains and, more recently, the availability of data on tandem PDZ domains has led to the recognition that PDZ flanking regions and linker sequences have functional roles.[6,7] Coupling between the flanking region and the core PDZ domain allows correlations between multiple signaling pathways to be established, as shown for the PSD-95 protein in which phosphorylation of a PDZ domain extension modulates peptide binding affinity.[47] Mutagenesis of the short and conserved linker sequence between the PDZ1 and PDZ2 domains of PSD-95 allowed the role of concerted interdomain rearrangements in modulating ligand binding affinity to be defined.[48] It is expected that interactions between different PDZ domains belonging to the same protein, and/or interactions between linker sequences and PDZ domains, will provide a general mechanism for encoding complex signaling networks within a single protein.[6,49,50]

Complete or partial folding of PDZ flanking regions and linkers has been observed in a number of cases and represents the most visible manifestation of these interdomain communications.[51−53] A theoretical framework has been proposed to rationalize the role of disorder-to-order transition events in optimizing allosteric coupling in proteins.[54] Our study of the

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

thermodynamic and dynamic aspects of HPV16 E6 protein binding to the PDZ1 domain of MAGI-1 suggests that such a mechanism has been encoded on the surface of this PDZ domain by a limited number of conserved residues. Currently, efforts are underway to assess the effect of the observed disorder-to-order transition on the binding properties of neighboring PDZ domains.

■ ASSOCIATED CONTENT

Supporting Information

A figure demonstrating the effects of the SLV|GGG mutation on the $^{1}H-^{15}N$ HSQC spectra and on the composite chemical shift differences (Figure S1), a figure showing the thermal denaturation recorded by circular dichroism of the wild-type PDZ1 and SLV|GGG mutant domains (Figure S2), a figure showing the effects of the K_{44}|A mutation on the composite chemical shift differences (Figure S3), a table of R_1 and R_2 values for PDZ1 wild-type and SLV|GGG mutant for apo and holo forms recorded at 283 and 303 K (Table S1), a table of NOE values for PDZ1 wild-type and SLV|GGG mutant for apo and holo forms recorded at 283 K (Table S2), a table of O^2 values for PDZ1 wild-type for apo and holo forms recorded at 283 and 303 K (Table S3), and a table of k_{ex} values for PDZ1 wild-type for apo and holo forms recorded at 283 and 303 K (Table S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: yves.nomine@unistra.fr. Phone: (+33)3.68.85.47.25. Fax: (+33)3.68.85.47.18.

Present Address

[‖]R.A.A.: Centre for Biomolecular Spectroscopy and Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1UL, United Kingdom.

Funding

This work was supported by grants from the Association pour la Recherche contre le Cancer (ARC) (Grant 3171), the Agence Nationale de la Recherche (ANR-MIME-2007 EPI-HPV-3D), the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-05-01), the ANR program PDZ-E6 (ANR-06-BLAN-0404), the National Institutes of Health (NIH) (R01CA134737), the Ligue contre le Cancer, Instruct, which is part of the European Strategy Forum on Research Infrastructures (ESFRI), and national member agreements. J.R. was supported by ANR Grant MIME-2007 EPI-HPV-3D and NIH Grant R01CA134737.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Claude Ling (IGBMC, France) for technical support and Pascal Eberling for the chemical peptide synthesis (IGBMC, France).

■ ABBREVIATIONS

HPV, human papillomavirus; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; MAGI, membrane-associated guanylate kinase with inverted domains; MAGI-1 PDZ1, the second of the six PDZ domains in the sequence of human MAGI-1 (residues 456−580 of human MAGI-1 (GenBank accession no.: AF401656) when referring to the specific polypeptide used in this study); TCEP, tris(2-carboxyethyl)-phosphine; $16E6_{ct}$, an 11-residue peptide derived from the C-terminal sequence of oncoprotein E6 from HPV 16 (RSSRT-RRETQL); $16E6_{ct}$ L_0|V, the same 11-residue peptide in which the C-terminal leucine residue is replaced by valine (RSSRT-RRETQV).

■ REFERENCES

(1) Luck, K., Charbonnier, S., and Travé, G. (2012) The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett. 586*, 2648−2661.

(2) Nourry, C., Grant, S. G., and Borg, J. P. (2003) PDZ domain proteins: plug and play! *Sci. Signaling 2003*, re7.

(3) Luck, K., Fournane, S., Kieffer, B., Masson, M., Nominé, Y., and Travé, G. (2011) Putting into practice domain-linear motif interaction predictions for exploration of protein networks. *PLoS One 6*, e25376.

(4) Chi, C. N., Bach, A., Stromgaard, K., Gianni, S., and Jemth, P. (2012) Ligand binding by PDZ domains. *BioFactors 38*, 338−348.

(5) Nomme, J., Fanning, A., Caffrey, M., Lye, M., Anderson, J., and Lavie, A. (2011) The Src homology 3 domain is required for junctional adhesion molecule binding to the third PDZ domain of the scaffolding protein ZO-1. *J. Biol. Chem. 286*, 43352−43360.

(6) Wang, C., Pan, L., Chen, J., and Zhang, M. (2010) Extensions of PDZ domains as important structural and functional elements. *Protein Cell 1*, 737−751.

(7) Ye, F., and Zhang, M. (2013) Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem. J. 455*, 1−14.

(8) Laura, R. (2002) MAGI-1: a widely expressed, alternatively spliced tight junction protein. *Exp. Cell Res. 275*, 155−170.

(9) Makokha, G. N., Takahashi, M., Higuchi, M., Saito, S., Tanaka, Y., and Fujii, M. (2013) Human T-cell leukemia virus type 1 Tax protein interacts with and mislocalizes the PDZ domain protein MAGI-1. *Cancer Sci. 104*, 313−320.

(10) Liu, H., Golebiewski, L., Dow, E. C., Krug, R. M., Javier, R. T., and Rice, A. P. (2010) The ESEV PDZ-binding motif of the avian influenza A virus NS1 protein protects infected cells from apoptosis by directly targeting Scribble. *J. Virol. 84*, 11164−11174.

(11) Glaunsinger, B. A., Lee, S. S., Thomas, M., Banks, L., and Javier, R. (2000) Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. *Oncogene 19*, 5270−5280.

(12) Kranjec, C., and Banks, L. (2011) A systematic analysis of human papillomavirus (HPV) E6 PDZ substrates identifies MAGI-1 as a major target of HPV type 16 (HPV-16) and HPV-18 whose loss accompanies disruption of tight junctions. *J. Virol. 85*, 1757−1764.

(13) Thomas, M., Glaunsinger, B., Pim, D., Javier, R., and Banks, L. (2001) HPV E6 and MAGUK protein interactions: determination of the molecular basis for specific protein recognition and degradation. *Oncogene 20*, 5431−5439.

(14) Nguyen, M. L., Nguyen, M. M., Lee, D., Griep, A. E., and Lambert, P. F. (2003) The PDZ ligand domain of the human papillomavirus type 16 E6 protein is required for E6's induction of epithelial hyperplasia in vivo. *J. Virol. 77*, 6957−6964.

(15) Charbonnier, S., Stier, G., Orfanoudakis, G., Kieffer, B., Atkinson, R. A., and Travé, G. (2008) Defining the minimal interacting regions of the tight junction protein MAGI-1 and HPV16 E6 oncoprotein for solution structure studies. *Protein Expression Purif. 60*, 64−73.

(16) Charbonnier, S., Nominé, Y., Ramirez, J., Luck, K., Chapelle, A., Stote, R. H., Travé, G., Kieffer, B., and Atkinson, R. A. (2011) The structural and dynamic response of MAGI-1 PDZ1 with noncanonical domain boundaries to the binding of human papillomavirus E6. *J. Mol. Biol. 406*, 745−763.

(17) Fournane, S., Charbonnier, S., Chapelle, A., Kieffer, B., Orfanoudakis, G., Travé, G., Masson, M., and Nominé, Y. (2011) Surface plasmon resonance analysis of the binding of high-risk mucosal

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

HPV E6 oncoproteins to the PDZ1 domain of the tight junction protein MAGI-1. *J. Mol. Recognit. 24*, 511−523.
(18) Wider, G., and Dreier, L. (2006) Measuring protein concentrations by NMR spectroscopy. *J. Am. Chem. Soc. 128*, 2571−2576.
(19) Broecker, J., Vargas, C., and Keller, S. (2011) Revisiting the optimal *c* value for isothermal titration calorimetry. *Anal. Biochem. 418*, 307−309.
(20) Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem. 179*, 131−137.
(21) Schwarz, F., Reinisch, T., Hinz, H. J., and Surolia, A. (2008) Recommendations on measurement and analysis of results ontained on biological substances using isothermal titration calorimetry (IUPAC Technical Report). *Pure Appl. Chem. 80*, 2025−2040.
(22) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.
(23) Goddard, T. D., and Kneller, D. G. *Sparky3*, University of California, San Francisco, CA.
(24) van Ingen, H., van Schaik, F. M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A. C., Kruijzer, J. A., Liskamp, R. M., Timmers, H. T., and Boelens, R. (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure 16*, 1245−1256.
(25) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33*, 5984−6003.
(26) Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR 16*, 23−28.
(27) Baldwin, R. L. (1986) Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069−8072.
(28) Sturtevant, J. M. (1977) Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236−2240.
(29) Spolar, R. S., Ha, J. H., and Record, M. T., Jr. (1989) Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc. Natl. Acad. Sci. U.S.A. 86*, 8382−8385.
(30) Palmer, A. G., 3rd, Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol. 339*, 204−238.
(31) Einstein, A. (1905) Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys. (Berlin, Ger.) 17*, 549−560.
(32) Sharma, S. C., Rupasinghe, C. N., Parisien, R. B., and Spaller, M. R. (2007) Design, synthesis, and evaluation of linear and cyclic peptide ligands for PDZ10 of the multi-PDZ domain protein MUPP1. *Biochemistry 46*, 12709−12720.
(33) Saro, D., Li, T., Rupasinghe, C., Paredes, A., Caspers, N., and Spaller, M. R. (2007) A thermodynamic ligand binding study of the third PDZ domain (PDZ3) from the mammalian neuronal protein PSD-95. *Biochemistry 46*, 6340−6352.
(34) McNemar, C., Snow, M. E., Windsor, W. T., Prongay, A., Mui, P., Zhang, R., Durkin, J., Le, H. V., and Weber, P. C. (1997) Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2. *Biochemistry 36*, 10006−10014.
(35) Prabhu, N. V., and Sharp, K. A. (2005) Heat capacity in proteins. *Annu. Rev. Phys. Chem. 56*, 521−548.
(36) Spolar, R. S., and Record, M. T., Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science 263*, 777−784.
(37) Gomez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins. *Proteins 22*, 404−412.
(38) Murphy, K. P., and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem. 43*, 313−361.
(39) Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc. 104*, 4546−4559.
(40) Yang, D. W., and Kay, L. E. (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol. 263*, 369−382.
(41) Kranjec, C., Massimi, P., and Banks, L. (2014) Restoration of MAGI-1 expression in HPV positive tumour cells induces cell growth arrest and apoptosis. *J. Virol. 88*, 7155−7169.
(42) Fuentes, E. J., Gilmore, S. A., Mauldin, R. V., and Lee, A. L. (2006) Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *J. Mol. Biol. 364*, 337−351.
(43) Petit, C. M., Zhang, J., Sapienza, P., Fuentes, E. J., and Lee, A. (2009) Hidden dynamic allostery in a PDZ domain. *Proc. Natl. Acad. Sci. U.S.A. 106*, 18249−18254.
(44) Dhulesia, A., Gsponer, J., and Vendruscolo, M. (2008) Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. *J. Am. Chem. Soc. 130*, 8931−8939.
(45) Kong, Y., and Karplus, M. (2009) Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. *Proteins 74*, 145−154.
(46) Gerek, Z., Ozkan, S., and Nussinov, R. (2011) Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS Comput. Biol. 7*, e1002154.
(47) Zhang, J., Petit, C., King, D., and Lee, A. (2011) Phosphorylation of a PDZ domain extension modulates binding affinity and interdomain interactions in postsynaptic density-95 (PSD-95) protein, a membrane-associated guanylate kinase (MAGUK). *J. Biol. Chem. 286*, 41776−41785.
(48) Wang, W., Weng, J., Zhang, X., Liu, M., and Zhang, M. (2009) Creating conformational entropy by increasing interdomain mobility in ligand binding regulation: a revisit to N-terminal tandem PDZ domains of PSD-95. *J. Am. Chem. Soc. 131*, 787−796.
(49) van den Berk, L. C., Landi, E., Walma, T., Vuister, G. W., Dente, L., and Hendriks, W. J. (2007) An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry 46*, 13629−13637.
(50) Zhang, J., Lewis, S., Kuhlman, B., and Lee, A. (2013) Supertertiary structure of the MAGUK core from PSD-95. *Structure 21*, 402−413.
(51) Tochio, H., Hung, F., Li, M., Bredt, D. S., and Zhang, M. (2000) Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-95. *J. Mol. Biol. 295*, 225−237.
(52) Whitney, D. S., Peterson, F. C., Kovrigin, E. L., and Volkman, B. F. (2013) Allosteric activation of the Par-6 PDZ via a partial unfolding transition. *J. Am. Chem. Soc. 135*, 9377−9383.
(53) Karlsson, O. A., Chi, C. N., Engstrom, A., and Jemth, P. (2012) The transition state of coupled folding and binding for a flexible β-finger. *J. Mol. Biol. 417*, 253−261.
(54) Hilser, V. J., and Thompson, E. B. (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. U.S.A. 104*, 8311−8315.
(55) Delano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.
(56) Charbonnier, S., Coutouly, M. A., Kieffer, B., Travé, G., and Atkinson, R. A. (2006) ^{13}C, ^{15}N and ^{1}H resonance assignment of the PDZ1 domain of MAGI-1 using QUASI. *J. Biomol. NMR 36* (Suppl. 1), 33.

DOI: 10.1021/bi500845j
Biochemistry XXXX, XXX, XXX−XXX

CONCLUSION AND PERSPECTIVES

Through this thesis, several projects were presented and all of them shared a common theme: observing a biological event through NMR. The work presented here has a spectroscopy overtone but in each project the biological background was exposed, as often terrible diseases are the reason science projects are born. When possible, the fluorine nucleus was used, since it brought many advantages. Both NMR as a technique and fluorine as a tool are not new, but they share the fact that they were used in a different manner once, were adapted and now, both NMR and fluorine are reborn (Barrett-2013).

The main project of this thesis concerned the ribosome and the set up of a method to try to detect ligand interaction in a quick and easy way. To do so, fluorine NMR was employed on existing fluorinated drugs. The original angle of the work was to try to screen directly on the ribosome, a molecule thousands of times bigger than the ligands used (>1 MDa vs <500 Da). This was done in order to get the closest possible to in-vivo conditions of the molecule, rather than using a truncated region. By observing the fluorine nuclei, NMR is easily able to follow the ligand signal and its interaction with the target. Indeed, the only fluorine signal in the solution is coming from the ligand. However, the observed ligand relaxation properties showed that a high number of non-specific interactions were happening, scrambling the signal of a ligand displacement by a specific competitor. A significant portion of the thesis was spent tuning experimental parameters to increase the signal-to-noise ratio in the competition assays. On a technical point, our lab did not have permanently access to a fluorine probe. Only in the last month of the thesis a cryoprobe with a fluorine channel was bought, as a testament to the fluorine potential. In only one experimental campaign, it was possible to quickly record many relaxation experiments and at ligand concentrations in the µM range. The results backed up our hypothesis regarding the ribosome: there is a continuum of sites with lower affinities than the sole high affinity binding site (illustrated in figure 1). The work on this part is still ongoing as models could be fitted on the relaxation titration results, to extract the K_d values of the ligands.

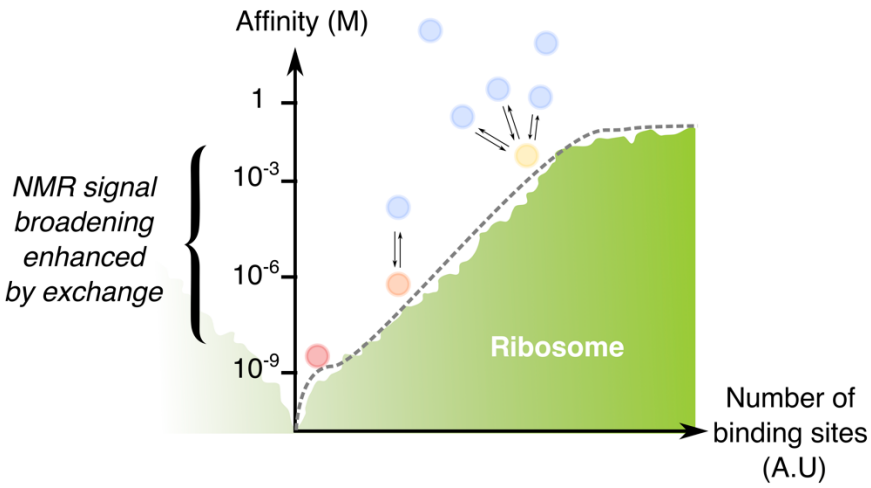


Fig 1. Illustration of a continuum of binding site presents on the ribosome

Another major part of the thesis was spent working on fluoroprolines as they are very peculiar residues. The fluorine moiety, once incorporated into the pyrrolidine ring, has a double advantage: it can of course act as a probe with all the fluorine NMR properties but it can also force the ring conformation. The studies realized on the SH3 domain used both those properties by screening together different fluorinated forms of the peptides. The fluorine sensitivity in NMR allowed to detect the coupling to several neighbouring residues and thus assess the local conformation. A k_{on} slower than anticipated was also extracted. The work on this project is also on-going, with new versions of the fluorinated polyproline peptides. The end goal being the determination of the kinetic constants k_{on}, k_{off} between peptides mimicking the RARγ PRD and the $SH3_3$ domain of vinexinβ.

The high sensitivity of ^{19}F NMR open many perspectives to these preliminary works: we could provide more quantitative insights into the mechanisms of binding of antibiotics to improve the selectivity of screening tests. The kinetic studies of polyproline interaction will be continued: the conformational properties of all polyprolines could be measured through various NMR methods (coupling constants, orienting media). The uses of the ^{19}F nuclei provides many opportunities to get novel insights in biological molecular mechanisms that were only started to be explored during this thesis.

Barrett, P., Chen, J., Cho, M., Kim, J., Lu, Z., Mathew, S., Peng, D., Song, Y., Van Horn, W., Zhuang, T., Sönnichsen, F. and Sanders, C. (2013). The Quiet Renaissance of Protein Nuclear Magnetic Resonance. Biochemistry, 52(8), pp.1303-1320.

Raphaël RECHT

Mise au point de méthodes de détection d'interaction ligand-macromolécule par RMN du ^{19}F

Résumé

Les interactions biologiques sont régies par des mécanismes complexes, qui mêlent différentes échelles, de temps comme de taille. C'est le cas du ribosome, un complexe nucléoprotéique responsable de la traduction de l'ARNm en protéines, et ce faisant, une cible thérapeutique primordiale. Or la taille du ribosome procaryote 70S (2.4 MDa) rend difficile l'applications des techniques classiques de criblage de ligands. Au cours de ma thèse, j'ai exploré la possibilité d'utiliser la RMN du fluor pour caractériser les interactions entre des ligands et le ribosome procaryote. Cette approche a été motivée par l'apport de nouvelles méthodes de détection pouvant coupler la versatilité de la RMN (Résonance Magnétique Nucléaire) avec les propriétés de l'atome de fluor. L'atome ^{19}F se prête parfaitement à la RMN, avec son rapport gyromagnétique proche du proton et son abondance isotopique naturelle de 100%. De plus, le fluor est bio-orthogonal au Vivant. Enfin, les caractéristiques physico-chimiques du fluor sont bien exploitées dans la pharmacopée (un quart des antibiotiques en possèdent un groupement).

Mots-clés : RMN, Fluor, Criblage, Interaction, Ligand, Macromolecule, Ribosome

Resume in English

Biological interactions are under the control of complex mechanisms, across different scales, in time of in size. It is particularly true for the ribosome, a nucleoprotein responsible for the mRNA translation into proteins, and thus, a primary therapeutic target. The size of the prokaryotic 70S ribosome (2.4 MDa) is a problem for the application of classical ligand screening method. During my thesis, I explored using fluorine NMR to characterize the interaction between ligands and the prokaryotic ribosome. This strategy was motivated by new detection approaches that can combine NMR (Nuclear Magnetic Resonance) versatility with the fluorine atom properties. The ^{19}F atom is perfectly suited for NMR, with its gyromagnetic ratio close to the proton one and its isotopic abundance of 100%. Moreover, the fluorine is absent from natural compounds. Finally, the physico-chemical characteristics of fluorine are well exploited in the pharmacopeia (a fourth of all antibiotics has a fluorine moiety).

Keywords: NMR, Fluorine, Screening, Interaction, Ligand, Macromolecule, Ribosome