# UNIVERSITÉ DE STRASBOURG

*ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé*

**UMR7104 – U964 Institut de Génétique et Biologie Moléculaire et Cellulaire**

# THÈSE présentée par :


## Julia ASENCIO HERNANDEZ


soutenue le : **24 Septembre 2015**

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

Discipline/ Spécialité : Sciences de la Vie et de la Santé / Biophysique et biologie structurale

# Novel approaches in NMR and biophysics for the study of complex systems.

## Application to the N-terminal domain of the Androgen Receptor

**THÈSE dirigée par :**

| | |
|---|---|
| **M. DELSUC Marc-André** | Directeur de Recherches, université de Strasbourg |

**RAPPORTEURS :**

| | |
|---|---|
| **M. ASHKENASY Gonen** | Professeur, Ben-Gurion university of the Negev |
| **M. REZAEI Human** | Directeur de Recherches, Inra – Centre Jouy-en-Josas |

---

**EXAMINATEURS :**

| | |
|---|---|
| **M. KIEFFER Bruno** | Professeur, université de Strasbourg |
| **M. DUMAS Philippe** | Professeur, université de Strasbourg |
| **M. LOQUET Antoine** | Chargé de recherches, université de Bordeaux |

# UNIVERSITÉ DE STRASBOURG

*ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé*

**UMR7104 – U964 Institut de Génétique et Biologie Moléculaire et Cellulaire**

# THÈSE présentée par :


## Julia ASENCIO HERNANDEZ


soutenue le : **24 Septembre 2015**

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

Discipline/ Spécialité : Sciences de la Vie et de la Santé / Biophysique et biologie structurale

# Nouvelles approches en RMN et en Biophysique pour l'étude de systèmes complexes.

## Application au domaine N-terminal du récepteur aux androgènes

**THÈSE dirigée par :**

| | |
|---|---|
| **M. DELSUC Marc-André** | Directeur de Recherches, université de Strasbourg |

**RAPPORTEURS :**

| | |
|---|---|
| **M. ASHKENASY Gonen** | Professeur, Ben-Gurion university of the Negev |
| **M. REZAEI Human** | Directeur de Recherches, Inra – Centre Jouy-en-Josas |

---

**EXAMINATEURS :**

| | |
|---|---|
| **M. KIEFFER Bruno** | Professeur, université de Strasbourg |
| **M. DUMAS Philippe** | Professeur, université de Strasbourg |
| **M. LOQUET Antoine** | Chargé de recherches, université de Bordeaux |

# Acknowledgments

This PhD summarizes three years of scientific research, but mainly, it summarizes three years of my life. Three years abroad, in a neighbor country, with a new language, with new friends...

I would like to start by thanking the members of my thesis committee, Prof. Bruno Kieffer, Dr. Antoine Loquet, Prof. Philippe Dumas and specially Dr. Human Rezaei and Prof. Gonen Ashkenasy for accepting to be my referees. I wish you all enjoy this thesis as I did.

I would like to thank the NMRTEC company and all the team, to give me this opportunity, and specially the ITN ReAd network, it has become a small scientific family for me during these three years. And many many thanks Aude, for your kindness when I arrived to Strasbourg, for all your help during these three years and for the affection you gave me.

I want also to thank the NMR team at the IGBMC. Thanks to everybody that have been part of the team along these years, but specially Claude, for always being willing to help. Raphaël, always the most friendly of the team, thank you for being patient and for your help!. I want to thank also both Justine, for the good and friendly moments!. I would like to thank Lionel for helping me to draw NMR sequences! and for your kindness during these three years. Bruno, thank you for accept to be part of my jury, I am really grateful for your scientific help, for your availability to explain and to discuss any topic.

Sandra! what can I say to you? Thank you very much for your friendship and your support, thanks for your generosity and your help! You have your home in Spain.

Finally, I would like to THANK Marc-André, for being my supervisor, for giving me the opportunity to do this amazing job, for believing in me when I didn't do it, for your help, for your NMR knowledge but mainly for being such a great person.

Gracias a mis amigas, por esos *Bondia* a 1600 km, por poder contar siempre con vosotras, incondicionalmente, y por ser *les millor 6*. Vos vuic.

Gracias a mi familia, a mi abuela, gracias a vosotros papás por vuestro apoyo incondicional, por estar siempre ahí aunque estemos lejos. A mi hermana, gracias tata por escucharme y por apoyarme cuando lo necesito. Soy afortunada por teneros, os quiero.

Y gracias a ti Luis, porque si no te hubieses mudado esto no habría sido lo mismo, por compartir todo tipo de aventuras, buenas y malas. Gracias por cuidarme, por cocinarme! pero sobre todo gracias por hacerme sonreir y por hacerme feliz. ¿Preparado para la próxima aventura?. Jag älskar dig.

Merci à tous pour ces trois années.

# Contents

# List of Abbreviations

| | |
|---|---|
| $\delta$ | Chemical Shift |
| ***D*** | Diffusion coefficient |
| ***p*–DOSY** | permutated–DOSY |
| **AF1** | Activation Function 1 |
| **AF2** | Activation Function 2 |
| **AIS** | Androgen Insensitive Syndrome |
| **AR** | Androgen Receptor |
| **BPA** | Bisphenol A |
| **CD** | Circular Dichroism |
| **CRAZED** | COSY revamped by asymmetric Z-gradient |
| **CRPC** | Castration Resistance Prostate Cancer |
| **DBD** | DNA Binding Domain |
| **DHT** | $5\alpha$-dihydrotestosterone |
| **DMSO** | Dimethylsulfoxide |
| **DNA** | Deoxyribonucleic acid |
| **DOSY** | Diffusion Ordered SpectroscopY |
| **EM** | Electron Microscopy |
| **EPI-001** | Bisphenol A (3-chloro-2-hydroxypropyl) (2,3-dihydroxypropyl) ether |
| **EXSY** | EXchange SpectroscopY |
| **FID** | Free Induction Decay |
| **HMBC** | Heternonuclear Multiple Bond Correlation |
| **HSP** | Heat Shock Protein |
| **IDP** | Intrinsically Disordered Protein |
| **LBD** | Ligand Binding Domain |
| **LED** | Longitudinal Eddy-current Delay |
| **MS** | Mass Spectrometry |
| **NLS** | Nuclear Localization Signal |
| **NMR** | Nuclear Magnetic Resonance |
| **NOE** | Nuclear Overhauser Enhancement |
| **NR** | Nuclear Receptor |
| **NTD** | N-Terminal Domain |
| **PFG** | Pulse Field Gradient |
| **RDC** | Residual Dipolar Coupling |
| **RE** | Response Element |
| **RF** | Radiofrequency |
| **RNA** | Ribonucleic acid |
| **ROESY** | Rotating frame Overhauser Effect SpectroscopY |
| $\mathbf{R}_1$ | Longitudinal relaxation rate |
| $\mathbf{R}_2$ | Transversal relaxation rate |
| **SAXS** | Small Angle X-ray Scattering |
| **SBMA** | Spinal Bulbar Muscular Atrophy |
| **SDS** | Sodium dodecyl sulfate |
| **SE** | Spin-Echo |
| **SPE** | Solid Phase Extraction |
| **STD** | Saturation transfer difference |
| **STE** | Stimulated Echo |
| **TCEP** | tris(2-carboxyethyl)phosphine |
| **THF** | Tetrahydrofuran |
| **ThT** | Thioflavine T |
| **TOCSY** | TOtal Correlation SpectroscopY |
| **WaterLOGSY** | Water-ligand Observation via Gradient SpectroscopY |

**Nouvelles approches en RMN et en biophysique pour l'étude de systèmes complexes**,
Application au domaine N-terminal du récepteur aux androgènes.

***Novel approaches in NMR and biophysics for the study of complex systems***,
*Application to the N-terminal domain of the Androgen Receptor.*

Julia ASENCIO HERNANDEZ
*Directeur de thèse*, Marc-André DELSUC

Mon sujet de thèse vise à développer de nouvelles méthodologies RMN (Résonance Magnétique Nucléaire) pour l'étude des systèmes complexes. Ce projet de thèse est une collaboration entre la société NMRTEC (Illkirch) et l'équipe de RMN biomoléculaire de l'IGBMC (Strasbourg).

Ce travail a été porté par la société NMRTEC, et réalisé dans le cadre du projet européen ITN-ReAd. Ce projet européen regroupe une dizaine de laboratoires, et deux entreprises, sur l'étude des systèmes chimiques complexes, tel que les systèmes auto-réplicatifs, ou le développement de nouveaux senseurs. Dans ce consortium, le rôle de l'entreprise est de développer de nouvelles approches d'analyse de ces systèmes.

Dans ce cadre, mon projet vise à réaliser une étude approfondie sur le développement de méthodologies pour l'analyse de mélanges complexes. Cela peut comprendre l'étude des grandes bibliothèques chimiques, des systèmes auto-assemblage, les interactions ligand-protéine et les systèmes biologiques désordonnés. Les méthodes développées recouvrent principalement la RMN (1D, 2D, et multinucléaire), la mesure de diffusion par RMN (DOSY) mais également d'autres techniques telles que la spectrométrie de masse, le dichroïsme circulaire (CD), la microscopie électronique (EM) et diffusion des rayons X aux petits angles (SAXS).

Ainsi, pour mener à bien cette étude, nous avons commencé à explorer les expériences de RMN pour l'étude des systèmes hors d'équilibre. La spectroscopie RMN est un bon outil pour la surveillance des réactions chimiques in situ. En particulier, la mesure DOSY est bien adaptée pour caractériser les espèces transitoires par la détermination de leurs tailles. Cependant, nous rencontrons une difficulté dans les expériences DOSY effectuées sur des systèmes hors d'équilibre. Dans ces systèmes, l'évolution de la concentration des espèces interfère avec le processus de mesure, et crée un biais sur la détermination du coefficient de diffusion qui peut conduire à des interprétations erronées. Nous montrons qu'une permutation aléatoire des intensités des gradients utilisés pendant l'expérience DOSY permet d'éliminer ce biais (Figure 1) par moyennage.

Cette approche, que nous avons dénommé *p*–DOSY ne nécessite pas de changements dans les séquences d'impulsions, ni dans le logiciel de traitement, et restaure complètement la précision totale de la mesure. Cette technique est illustrée sur le suivit de la réaction d'anomérisation de $\alpha$- en $\beta$-glucose. Elle est parfaitement adaptée pour l'analyse des réactions chimiques et des systèmes hors d'équilibre et peut être utilisée pour surveiller le taux de la cinétique de réaction, pour caractériser les espèces transitoires *in-operando* et mesurer leur coefficient de diffusion, ou pour l'étude de phénomènes d'organisation moléculaires.

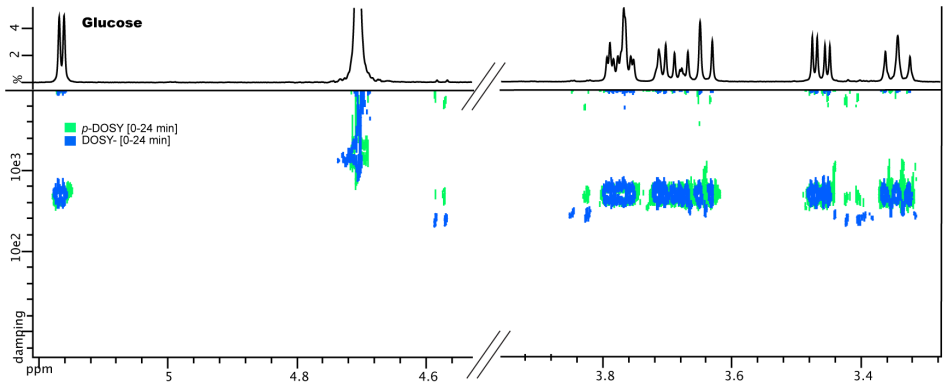



FIGURE 1 – Superposition de la DOSY standard (couleur bleue) et de la *p*–DOSY (couleur verte), acquises dans la période hors d'équilibre (0,00 à 25,00 minutes). On voit le biais de la mesure de DOSY entre les formes $\alpha$ et $\beta$ pour la DOSY standard, corrigé par la mesure en *p*–DOSY.

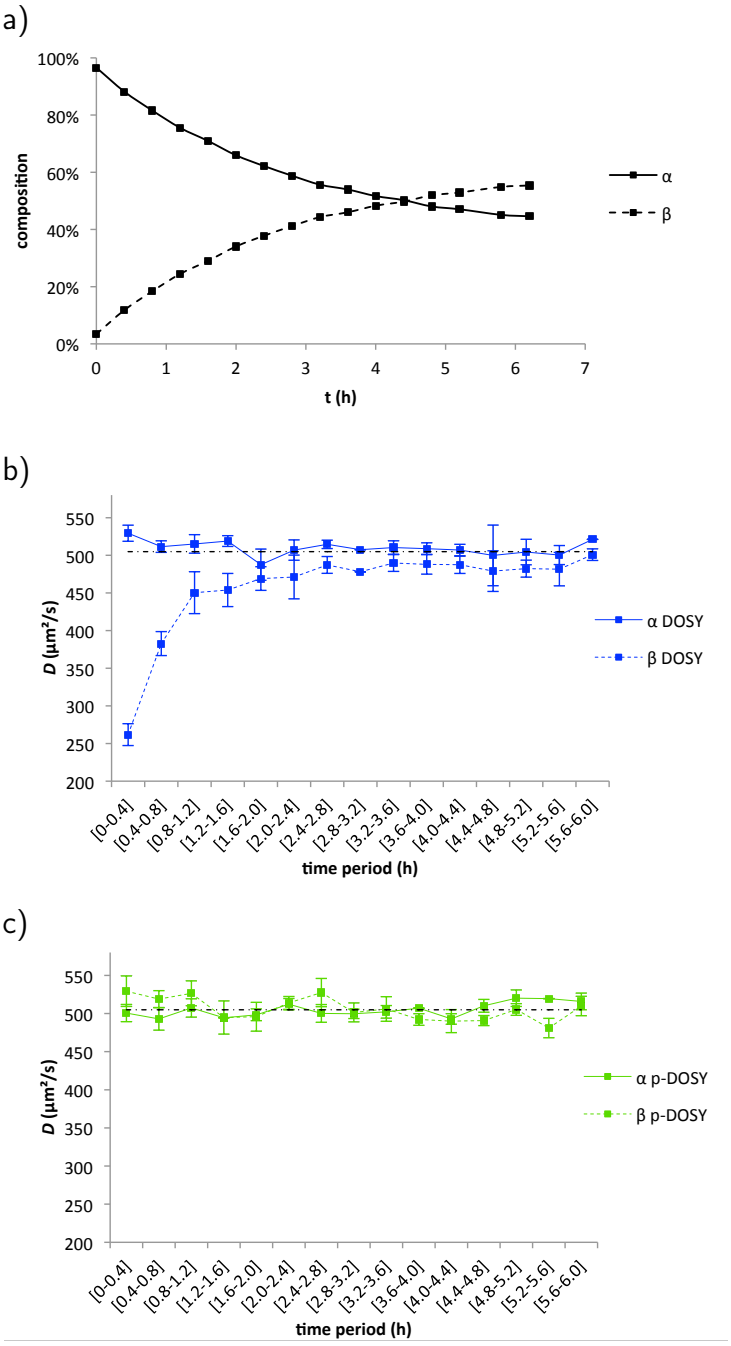



FIGURE 2 – a)lévolution de la concentration de le $\alpha, \beta$-glucose au cours du temps expérimental, mesuré comme l'intensité du pic du proton anomérique à partir d'expériences 1D intercalées entre chaque experience *p*-DOSY. b)l'évolution des coefficients de diffusion apparent de $\alpha$ et $\beta$-glucose au cours du temps expérimental basé sur une série de mesures séquentielles de DOSY classiques. c)même que b, obtenu avec l'expérience *p*–DOSY. Dans b) et c), la ligne horizontale est situé au coefficient de diffusion du $\alpha$-glucose, mesuré comme la valeur moyenne de la valeur *p*–DOSY sur toute la cinétique.

Dans une deuxième partie de ma thèse, la partie N-terminale du récepteur des androgènes (AR) est utilisée comme un système complexe. D'après la littérature, il est connu que cette région joue un rôle important pour l'activité du récepteur, et elle est également décrite comme étant intrinsèquement désordonnée. De ce fait, la technique de spectroscopie RMN est la seule méthode spectroscopique pour appréhender ce domaine au niveau structural. Les résultats que j'ai acquis durant la thèse m'ont permis d'identifier une petite région de ce domaine, impliquée dans la formation réversible de fibres amyloïdes, par modulation des conditions d'oxydo-réduction du milieu.

L'auto-assemblage est un processus spontané par lequel des molécules s'organisent en structures pluri-moléculaires ordonnées. L'une des structures couramment présentées par des peptides et des protéines est la formation de fibrilles connues comme de fibres amyloïdes. Ces fibres peuvent spontanément s'agréger et produire des dépôts dans les tissus, et sont alors impliquées dans différentes pathologies dégénératives, tel que Alzheimer et Parkinson. Dans le domaine N-terminal de AR (NTD-AR), il existe une séquence primaire, unique à cette protéine dans l'ensemble du génome humain et qui est parfaitement conservée chez tous les vertébrés (Figure 3).

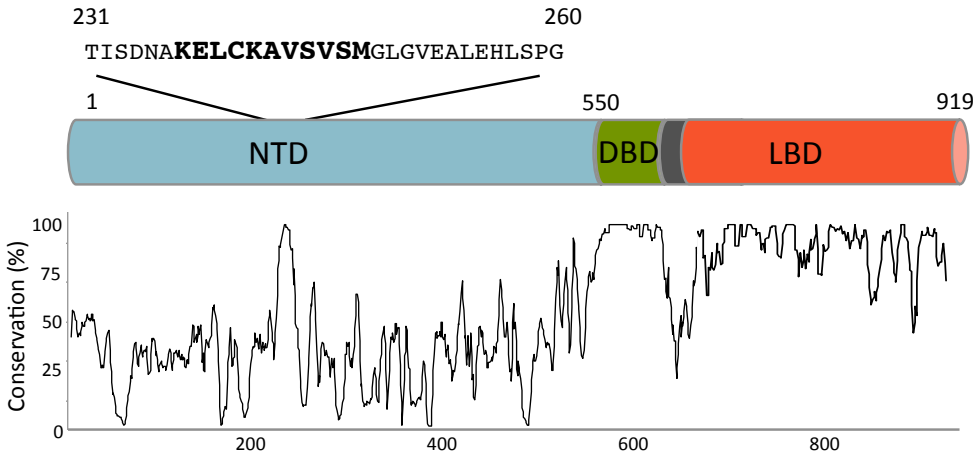



FIGURE 3 – Representation des differentes domains du récepteur des androgènes et la conservation de la séquence chez tous les vertébrés. Zoom sur la séquence du NTD parfaitement conservée.

Cette séquence contient un résidu cystéine et apparaît parfaitement soluble et non structurée dans son état natif. Lors de l'oxydation de la fonction thiol par le DMSO (diméthyl-sulfoxyde), un dimère covalent se forme. Ce dimère s'auto-assemble spontanément en fibres amyloïdes d'un diamètre de 6 nm. L'analyse cinétique de ce phénomène a été suivie par RMN ainsi que par CD, fluorescence, EM et SAXS. Les fibrilles présentent toutes les caractéristiques de fibres amyloïdes, telle que la réponse de fluorescence à la thioflavine, une structure secondaire en feuillet-$\beta$ et un diamètre constant. En outre, l'agrégation est réversible par l'ajout d'un agent réducteur. La découvert de cette propriété ouvre de nouvelles possibilités en termes d'organisation supramoléculaire et d'auto-assemblage, et pourrait être généralisée à d'autres systèmes.

L'originalité des résultats révèle un aspect inconnu du mécanisme de AR et nous permet d'expliquer l'observation de structures amyloïdes dans des biopsies de vésicules séminales, déjà décrite dans la littérature. L'étude du lien fonctionnel entre le stress oxydatif et l'activité de régulation de la transcription du récepteur des androgènes pourrait ouvrir des perspectives inédites pour le traitement du cancer de la prostate.

En raison de la nouveauté du système, une étude plus approfondie est nécessaire pour améliorer la compréhension de son role dans le mécanisme d'activité de AR.

De nos jours, dans la literature le domaine N-terminal du récepteur des androgènes est présenté comme un potentiel cible pour le développement de nouveaux médicaments pour le cancer de la prostate, particulièrement lorsque le cancer de la prostate resistant à la castration (CRPC) apparaît et la sensibilité à la thérapie anti-androgène est perdue. Dans les dernières années de nouvelles études ont été publiées, en présentant la molécule EPI-001 (Figure 6) et certains dérivés, comme un nouveau médicament pour la régression du cancer de la prostate[Andersen 2010, McEwan 2011]. Ces molécules sont un dérivé de l'un des perturbateurs endocriniens le plus connu, le Bisphénol A (BPA) (Figure 6). Récemment, d'autres nouvelles molécules ont fait leur apparition dans la littérature[Bañuelos 2014].

Pour toutes ces raisons, la stratégie est de montrer que les fibres amyloïdes, formées lors de l'oxydation du peptide provenant du NTD-AR, peut être un site cible pour toutes ces nouvelles molécules. En outre, les fibres analysées dans cet travail pourraient présenter un rôle biologique, pour cette raison, des hormones impliquées dans l'activité

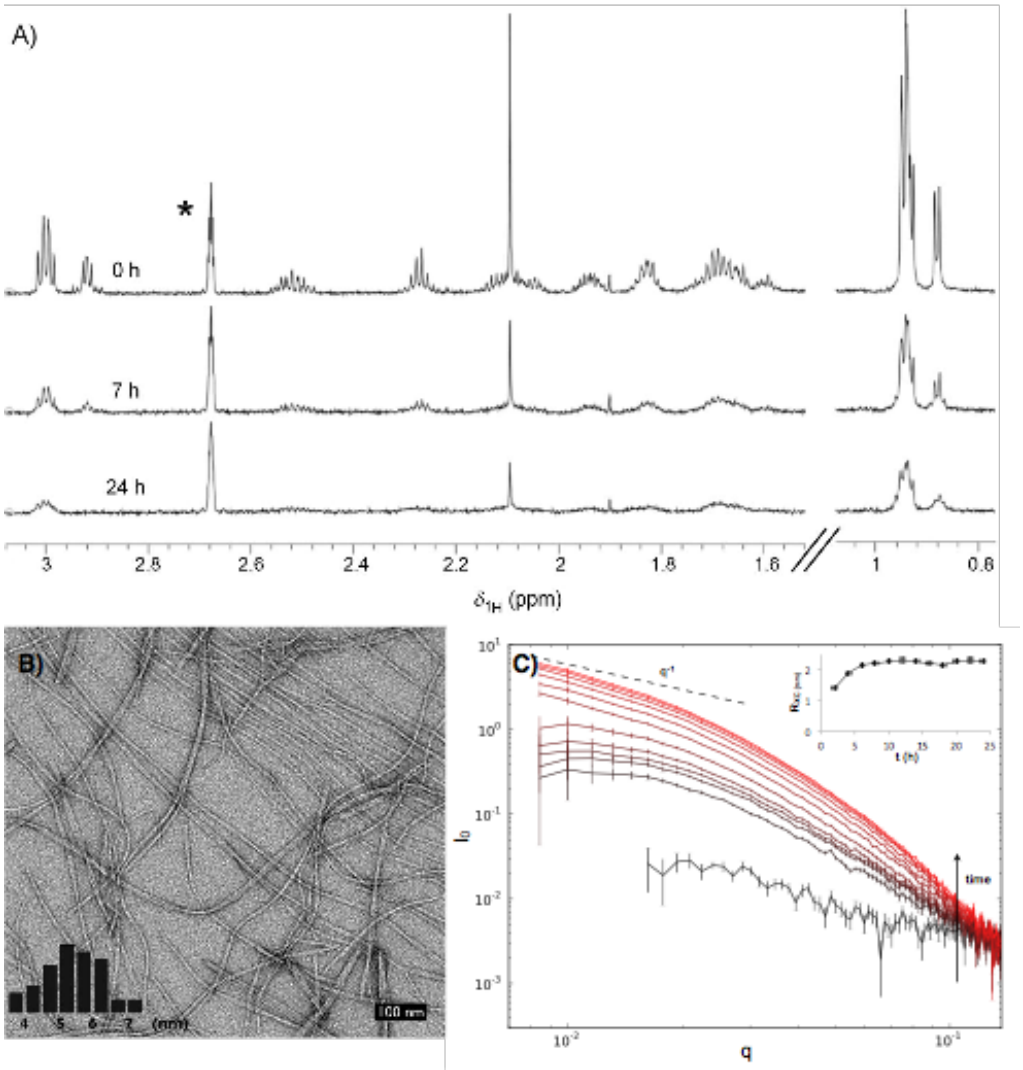



FIGURE 4 – A) Cinétique d'agrégation des fibrilles suivie par RMN. B) Image obtenu par microscopie électronique de les fibres amyloïdes avec la distribution de le diamètre de le fibrilles obtenu d'après l'analyse des images. C) Cinétique d'agrégation des fibrilles suivie par SAXS avec le radius de les fibres calculé.

de AR ont été étudiés. Ont été associés à cette étude les polyphénols en raison de leurs propriétés anti-oxydants et leurs effets sur l'activité de récepteur aux androgenes et de son expression.

Avec cet objectif, nous avons développé une étude d'interaction entre les fibrilles amyloïdes et différentes molecules décrites dans la literature comme possible nouveaux

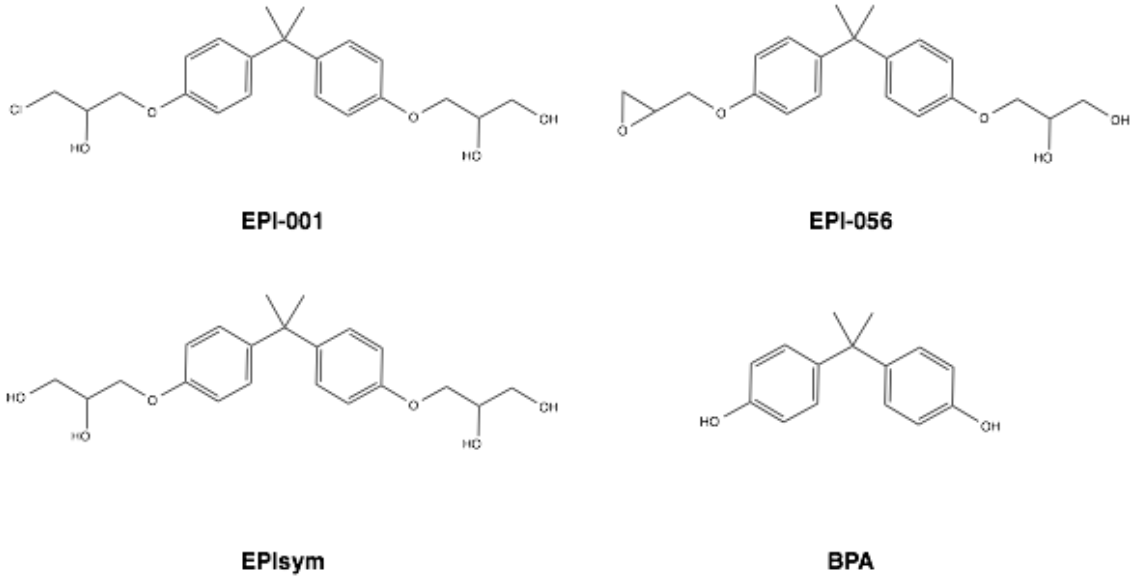



FIGURE 5 – Molecular structure of EPI-001 and Bisphenol A

médicaments pour traiter le cancer de la prostate.

Ici, nous avons montré que les dérivés de BPA famille de petites molécules accélère la cinétique de l'agrégation de fibres amyloïdes. En outre, le BPA a été rapporté dans la littérature pour accélérer la formation d'amyloïde toxique [Gong 2013]. Ce résultat va dans le même sens que dans l'étude présentée ici et peut ouvrir de nouvelles voies de recherche pour contrôler l'agrégation amyloïde.

Enfin, l'interaction de EPI-001 au sein de la NTD-AR, est prouvée sur le cadre de cette étude. Une interaction covalente est caractérisée par HPLC/MS, liaison covalente formé par la Cys du peptide et le EPI-001. Tout ensemble, il a été démontré dans cet étude que, séquence hautement conservée de NTD-AR non seulement s'aggrège sous forme de fibres amyloïdes dans des conditions oxydantes, mais pourrait aussi être un site cible potentiel pour de nouveaux médicaments.

Jusqu'à ce point, dans la thèse, nous avons mis l'accent dans les systèmes complexes d'un point de vue biologique. Mais, les systèmes complexes sont présents tout autour de nous. Au sein de ces grands systèmes avec multiples composantes, où de nombreuses parties interagissent de manière non-linéaire et les interactions faibles entre leurs parties sont cruciales pour le développement et l'émergence de nouvelles propriétés. Dans cette

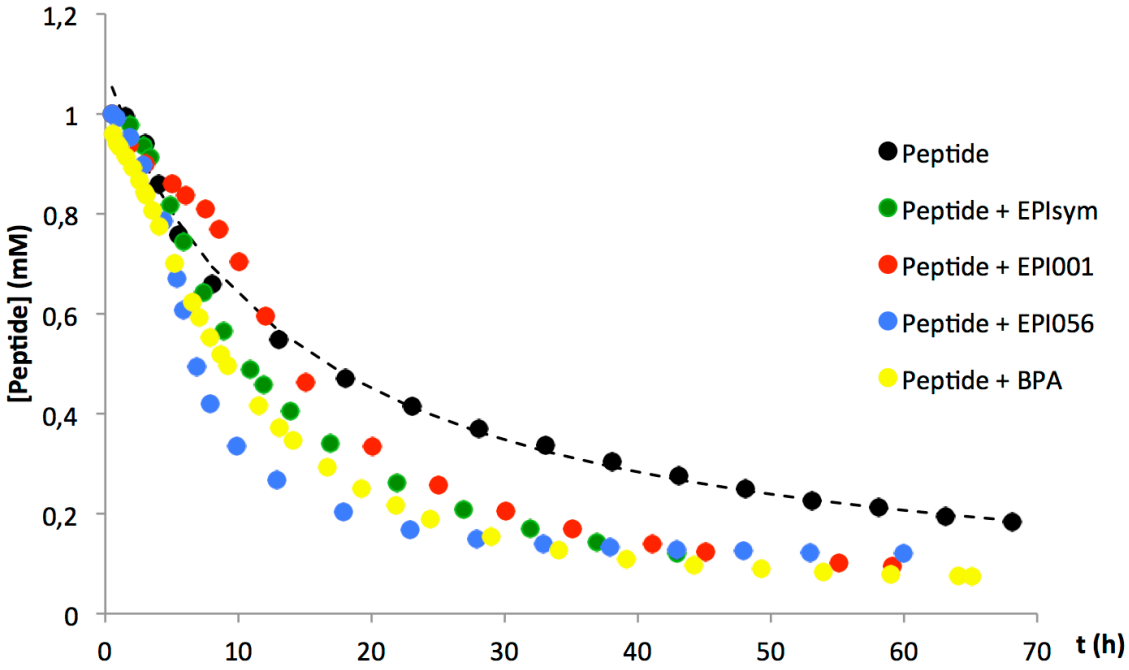



FIGURE 6 – Variation de la concentration de peptide suivie par RMN, en présence de BPA (jaune), EPI-001 (rouge), EPI-056 (bleu) et EPI-sym (vert). En noir la cinétique de le peptide sans molécules, comme contrôle, avec une cinétique de second ordre (ligne pointillée).

organisation hiérarchique de molécules nous nous concentrons maintenant sur une organisation simple en chimie, tels que des solvants. Ceux-ci, a priori, des systèmes simples peuvent élaborer et adopter différentes conformations complexes sous différentes conditions qui pourraient affecter une analyse plus approfondie des résultats expérimentaux. Pour cette raison, nous étudions ici le mélange acétonitrile-eau, qui représente le mélange le plus utilisé comme solvants en chimie et affiche un comportement mal connu à basse température, et aujourd'hui, encore difficile à interpréter.

Ce travail présente l'analyse par RMN d'un comportement particulier d'un mélange simple, liquide-liquide. Alors que l'anomalie du mélange acétonitrile-eau à basse température était déjà connue et caractérisée, son étude par spectroscopie RMN n'avait pas encore été entrepris. La structure particulière de la séparation des phases liquid-liquid qui

a lieu dans ce système conduit à de grandes perturbations dans les spectres RMN 1D (Figure 7 et 8), indiquant de fortes anisotropies dans l'interphase liquide-liquide. D'autres expériences de RMN effectuées (DOSY, EXSY) montrent que les micro-domaines structurés ont certainement une longue durée de vie, avec des propriétés physiques très différentes (fort contraste dans les coefficients de diffusion macroscopiques). L'expérience CRAZED produit des spectres qui montrent que beaucoup d'ordre de spin à longue portée sont présents dans le système, ce qui conduit à de nombreux pics multiples quantiques intermoléculaires. En outre, l'expérience STRANGE, introduit dans cette étude, présentent des caractéristiques , encore inexpliquée.

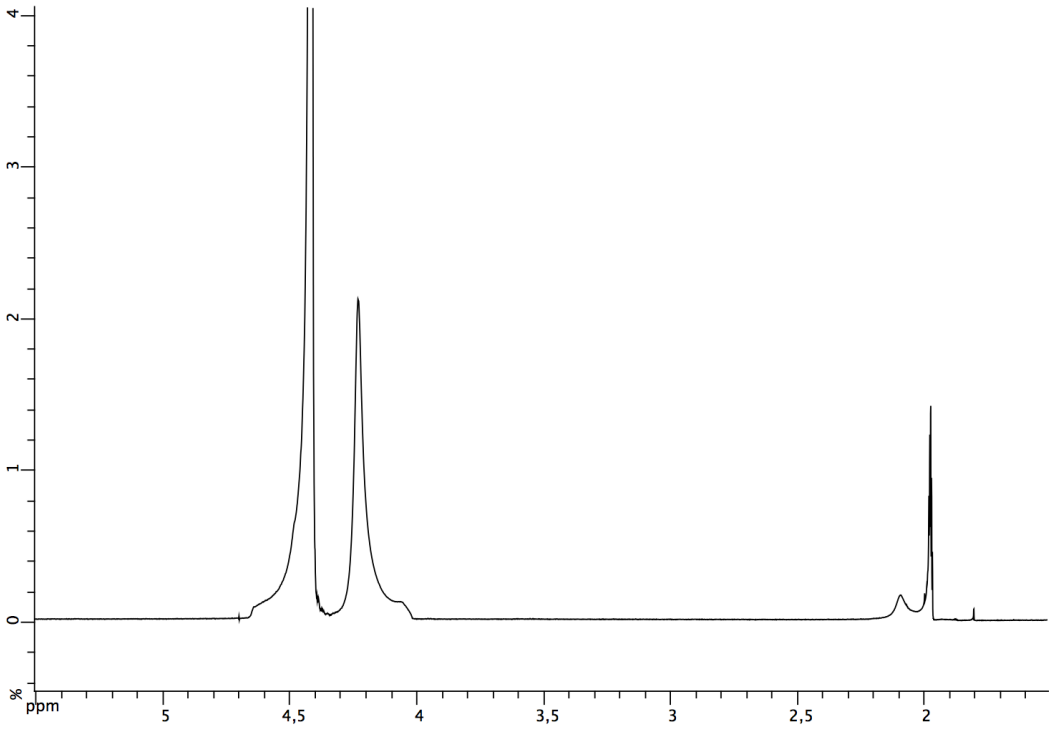



FIGURE 7 – Spectre de RMN 1D $^{1}$H du mélange acétonitrile-eau à 6°C

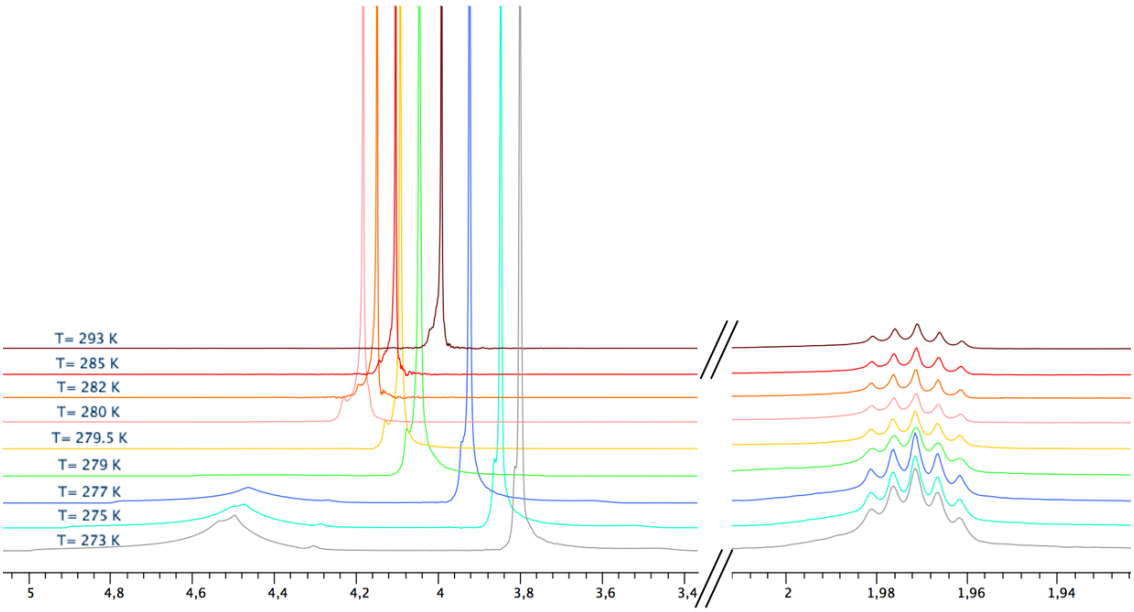



FIGURE 8 – Evolution du spectre RMN 1D $^1$H du mélange acétonitrile-eau ($\chi_{AN} = 0.45$) à partir de 293 K à 273 effectué dans un tube de 5 mm. Le signal de l'eau est sur la gauche, l'acétonitrile est sur la droite. L'instrument est verrouillé sur le signal de l'acétonitrile.

Ce travail a donné lieu à deux publications en premier auteur, un dans *ChemBioChem* [Asencio-Hernández 2014] sur la formation de fibrilles amyloïdes dans le récepteur aux androgènes, ainsi qu'un article sur la méthode *p*–DOSY dans *J.Magn.Reson* [Oikonomou 2015].

# Références

[Andersen 2010] Raymond J Andersen, Nasrin R Mawji, Jun Wang, Gang Wang, Simon Haile, Jae-Kyung Myung, Kate Watt, Teresa Tam, Yu Chi Yang, Carmen A Bañuelos, David E Williams, Iain J McEwan, Yuzhou Wang and Marianne D Sadar. *Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor*. Cancer Cell,

vol. 17, no. 6, pages 535–546, June 2010.

[Asencio-Hernández 2014] Julia Asencio-Hernández, Christine Ruhlmann, Alastair McEwen, Pascal Eberling, Yves Nominé, Jocelyn Céraline, Jean-Philippe Starck and Marc-André Delsuc. *Reversible amyloid fiber formation in the N terminus of androgen receptor.* ChemBioChem, vol. 15, no. 16, pages 2370–2373, November 2014.

[Bañuelos 2014] Carmen A Bañuelos, Aaron Lal, Amy H Tien, Neel Shah, Yu Chi Yang, Nasrin R Mawji, Labros G Meimetis, Jacob Park, Jian Kunzhong, Raymond J Andersen and Marianne D Sadar. *Characterization of Niphatenones that Inhibit Androgen Receptor N-Terminal Domain*. PLoS ONE, vol. 9, no. 9, page e107991, September 2014.

[Gong 2013] Hao Gong, Xin Zhang, Biao Cheng, Yue Sun, Chuanzhou Li, Ting Li, Ling Zheng and Kun Huang. *Bisphenol A Accelerates Toxic Amyloid Formation of Human Islet Amyloid Polypeptide : A Possible Link between Bisphenol A Exposure and Type 2 Diabetes*. PLoS ONE, vol. 8, no. 1, page e54198, January 2013.

[McEwan 2011] Iain J McEwan. *Intrinsic disorder in the androgen receptor : identification, characterisation and drugability*. Mol. BioSyst., vol. 8, no. 1, page 82, 2011.

[Oikonomou 2015] Maria Oikonomou, Julia Asencio-Hernández, Aldrik H Velders and Marc-André Delsuc. *Journal of Magnetic Resonance*. J. Magn. Reson., vol. 258, no. C, pages 12–16, September 2015.

Chapter 1

# Complex Systems



## 1.1 Complex Systems and their features

Complex systems exist all around us, from chemical mixtures to galaxy dynamics, from metabolic pathways to eco-systems, from stock markets to social network. All these systems share in common a few typical characteristics that are the topic of the new science of complex systems, and have been intensively scrutinized.

The most commonly put forward feature for complex systems, is **emergence**. Emergence describes the occurrence in a system, of a new property or behavior which can not be extrapolated from the characteristic of the many simple elements which compose the system. Emergence of planets from the interstellar gas, emergence of the organism from the genome, emergence of bubbles (or crashes) from financial market rules, are few examples.

Emergence is the most striking characteristic, however it is based on a series of fundamental properties that are found in many complex systems [Barabasi 1999].

Emergence is by definition related to the fact that "the whole is more than the sum of its parts", which is in other words, the fact that complex systems do not satisfy the superposition principle. This is intrinsically related to the **nonlinearity** of the interactions taking place in the system.

A large multiple-component system, with many parts interacting in a non-linear manner, displays a complex dynamic behavior, with the absence of long term coherence, but rather pseudo cyclic features - named attractors - local (in time or in space). This property is usually described as **chaotic**.

The chaotic dynamic of the system, which may evolves in one direction or another depending on small variations of the initial conditions, leads to the impossibility to predict the long term behavior of the system [Stavrinides 2013]. This creates a non-proportionality of the cause and on its effect, creates a **non predictability** of the system, even though each elementary dynamic step in the system is acting in a purely deterministic manner [Crutchfield 2011].

The classical example of a chaotic non-predictable system is the weather variation developing in the earth atmosphere.

Of course, this implies that the system is in an **out-of-equilibrium** state. In these conditions, higher structures may even appear. These **dissipative** structures display generally a **multi-scale** aspect, with the shorter scales (in space or in time) dominated by the simple inter-part interactions, while larger scale organized structures, unrelated to the formers, may appear [Sales-Pardo 2007].

The development of such systems is also dependent on **weak interactions** between the parts. Strong interactions usually reduce the number of possible degrees of freedom, and do not allow the development of complex behaviors.

All these characteristics, if present in a system, lead to complex behaviors, with the possibility of the emergence of larger structures, depending in a complex and fully non-linear manner on the external conditions, leading to some **adaptation** capabilities of the system.

Science in general, as a human activity, and Biology in particular is all about how, from the interaction of the molecules and atoms, new structures may emerge, leading to life, to organism interaction, to consciousness, to social interactions, and finally to activities such as the process of writing a thesis on complex systems in biology [Barroso 2015].

## 1.2 Molecular Complex systems

Molecular complex systems start with a simple atom bond formation. Complexity comes from the basis of chemical arrangement to form molecules, these molecules can assemble into macromolecules with a function in biological organisms [Ross 2009]. This exemplifies the concept of emergence, going from chemistry to life by placing complex interactions within bionetworks or bioassemblies.

The characterization of complex systems covers most scientific disciplines [Liu 2011], al-

though the quantitative description of a complex system is intrinsically limited by the experimental accessibility. Weak interactions, including van der Waals force, H-bonding and $\pi$ electron interactions, play an important role in many elemental processes in chemistry, physics, materials science and biology. Despite the fact that the energies of weak interactions are lower than ionic interactions or covalent bonds, their cumulative effect can be more relevant. The presence of these weak interactions together with the dynamics of complex systems, imply the presence of out-of-equilibrium systems. The out-of-equilibrium systems arise as objects of study in different contexts such as interacting particle systems, reaction-diffusion processes, on slow dynamics conditions and the possibility of undergoing phase transitions. This dynamics leads complex systems to be highly sensitive to initial conditions with not proportionality between cause and effect. Small perturbations in initial conditions derivate in a widely number of diverging outcomes, leading a none predictable situation [Crutchfield 2011], as nucleation. Nucleation is the first step in the formation of either a new structure or phase, via self-assembly or self-organization. This process will occur on a different ways for identical systems, resulting a stochastic process. The nonlinearity can be exemplified by allostery, the process by which biological macromolecules when they bind at one site, the effect is transferred to a different active site. This phenomenon has been associated to the intrinsically disordered proteins (IDPs), a challenge to the structure–function paradigm [Motlagh 2014].

From a molecular point of view, the study of complex systems in chemistry has changed in the last years due to the development of analytical tools and the changes in research of molecular networks. The concept of complex systems in chemistry is related to the concept of Supramolecular chemistry. Introduced by Jean-Marie Lehn [Lehn 1993], this concept aims to develop chemical systems with highly complexity from components that interact via *weak* intermolecular forces. The principal exponents of these areas are self-replicating systems (kinetically controlled) and dynamic combinatorial libraries, that represent thermodynamically controlled systems. A molecule in a dynamic combinatorial library can bind and copy itself, this situation may lead to self-assembly into nanoscale structure [Carnall 2010]. Within this topic, the peptides self-assembly and self-replicating systems have been used to understand the organizational principles of biological systems with a final aim of understanding the origin of life [Rubinov 2009]. Also self-assembled nanomaterials can be built based on aromatic peptide amphiphiles [Abul-Haija 2013] for developing of cell-instructive biomaterials.

Contrary to what happens in Chemistry, historically, the study of complex systems has been better stablished in Biology. The hierarchy of complexity levels of organization in proteins and the consequence in its function, exemplifies perfectly the concept of complex system in Biology. The presence of nonlinearity related to small modifications or differences in its primary structure

can conclude in an allosteric effect, or functionality disruption. Nevertheless the absence of the structure also plays an important role in the complexity network of biology.

As said above, complex systems include a wide range of different topics, research approaches, complex molecular networks and different molecular self-organization. Within this huge range of possibilities, it exists some *key concepts* that require a special mention and they will be the main actors in this thesis research.

### 1.2.1 Self-assembly

Self-assembly is a key concept in complex systems, specially in supramolecular chemistry, but also in biology. Self-assembly is a process in which a disordered system spontaneously organizes itself into an ordered structured system or pattern, as a consequence of specific local interactions without external contributions. The nanostructure builds itself.

Self-assembly from a chemist point of view implies order and weak interactions. From the point of view of biology, molecular self-assembly refers mainly to intermolecular self-assembly, while folding refers the intramolecular analog.

Molecular self-assembly underlies the construction of biologic macromolecular assemblies in living organisms, and so is crucial to the function of cells. It is exhibited in the self-assembly of lipids to form the membrane, the formation of double helical DNA through hydrogen bonding of the individual strands, and the assembly of proteins to form quaternary structures. Molecular self-assembly of incorrectly folded proteins into insoluble amyloid fibers is responsible for infectious prion-related neurodegenerative diseases [Noinville 2008].

### 1.2.2 Proteins structures

Complexity, organization and function are concepts that form part of the definition of protein. Proteins are commonly described in terms of four hierarchical levels of organization. Proteins are biological molecules or macromolecules, formed by atoms, atoms that are organized into amino acid residues. These 20 different amino acids, act as monomeric building blocks of proteins and covalent bound in a polypeptide chain. This organization represents the *primary structure* of a protein.

In a next level, the polypeptide chains can adopt different spatial arrangements, *secondary structure*. This secondary structure, without any stabilizing interaction, displays a random coil organization. However, when the stabilization appears as hydrogen bonds between certain residues, different secondary structures can be adopted, as $\alpha$-helix or $\beta$-sheet. These structures were suggested, before any experimental evidence, in 1951 by Linus Pauling and coworkers

[Pauling 1951a, Pauling 1951b]. Apart from these, there are also other secondary structure such as $3_{10}$-helix, $\pi$-helix, polyproline-helix, $\alpha$-sheet and $\beta$-turn.

*Tertiary structure* refers to the three-dimensional structure of a protein, and to its geometric shape, the three-dimensional arrangement of all the amino acids residues. In contrast to secondary structure, which is stabilized by hydrogen bonds, tertiary structure is stabilized by hydrophobic interactions between the nonpolar side chains and, in some proteins, by disulfide bonds. Leading an hydrophobic core and an hydrophilic surface that are the driving forces of the tertiary structure, as hydrogen bounds for the secondary structure and chemical bounds for the primary structure.

Finally, multimeric proteins present a *quaternary structure*. It is the three-dimensional structure of a multi-subunit protein and the way that the subunits fit together. In this context, the quaternary structure is stabilized by the same non-covalent interactions and disulfide bonds as the tertiary structure.

In a similar manner to the hierarchy of structures that make up a protein, proteins themselves are part of a hierarchy of cellular structures. Proteins can associate into larger structures, macromolecular assemblies. Besides this hierarchical organization, the concept of nonlinearity is crucial, in other words, mutations as genetic factors of human diseases. The simple variation in the side chain of an amino acid, for instance, from a -SH to a -OH group (when an Cys is mutated into Ser), that occurs in the Androgen Receptor C240S, can have dramatic consequences, inhibiting amyloid fibrillation [Asencio-Hernández 2014]. Mutations can affect protein folding and stability [Piana 2008, Davies 2008], protein expression, protein function and protein–protein interactions [Luscombe 2002, Lapouge 2008]. Mutations in proteins have a major role in the onset and development of cancer [Brooke 2009, Papa 2014], their special role is determined by the diversity of their impact on molecular function, in a *multi-scale* manner.

### 1.2.3 Intrinsically Disordered Proteins (IDPs)

Within the family of proteins, there is a *class* that does not present a tertiary structure named Intrinsically Disordered Proteins (IDPs). IDPs have been overlooked for many years, however a third of the proteins in higher eukaryotes is expected to be disordered. The lack of three dimensional structure together with their plasticity makes them to play a crucial role in cell signaling and regulation [Myers 2013]. Furthermore, a high percentage of oncoprotein contain intrinsically disordered regions.

Intrinsic disorder is common in proteins associated with many human genetic diseases and its role in many different pathologies make them important target for the research and development of new therapies [Marasco 2015, Guharoy 2015].

The absence of structure of the IDP makes particularly difficult the understanding of the relationship between the different transient folding that they can adopt and their function [Fuxreiter 2014], and new methodological development have to be made. Decoding the hot spots in their sequence to find and to analyze them, is crucial for this proposal. Approaches such as a phylogenic analysis to find non expected conserved regions, and other bioinformatic analysis are essential to map possible binding sites.

The study of IDPs requires to work in solution state, the lack of structure makes impossible the use of X-ray crystallography and difficult the use of electron microscopy (EM). Nowadays, Nuclear Magnetic Resonance (NMR) is the best technique to study and analyze these interactions, as well as Small Angle X-ray scattering (SAXS), interesting not at atomic level, but giving shape information and changes in conformation up to binding.

### 1.2.4 Amyloids

Many of the well-known neurodegenerative diseases are related to protein missfolding. These disorders arise from the failure of a protein to adopt its native functional state. As consequences of miss-folding, protein aggregates into fibers, which leads to loss of function and increase its toxic function. There are proteins that present an intrinsic tendency to assume a pathologic conformation, this conformation becomes evident with high concentrations or aging. Paradoxically, for structured proteins the first step in fibrillation is the partial unfolding whereas for IDPs is their partial folding.

Many neurodegenerative diseases are amyloidoses. The term *Amyloid* defines the deposition of insoluble protein aggregates, associated with different diseases as Alzheimer's disease, Parkinson's disease, Huntington's disease, also non-degenerative diseases as type II diabetes and prion diseases as Creutzfeldt-Jakob Disease (CJD) [Sipe 2000, Knowles 2014]. These protein aggregates share a specific structural traits, they are composed of a $\beta$-sheet structure in a characteristic croos-$\beta$ conformation [Geddes 1968], although they have different primary structures [Pastor 2008]. Amyloid formation is a complex process that takes place through several intermediates of various size and morphologies.

Historically, the term amyloid has been strictly used to describe the extracellular deposition of proteins fibrils [Westermark 2005] with a characteristic pattern under the electronic microscope (linear and unbranched), a X-ray diffraction pattern of $\beta$-strands and $\beta$-sheets and with an affinity for Congo red with concomitant green birefringence. But in many cases aggregates accumulate intracellularly forming similar amyloid structures, that are called *amyloidlike* [Li 2009]. However, the term *Amyloid* is common used in the scientific literature, expanding the definition to include any aggregates of proteins, as well as synthetic peptides, that possess

some but not necessarily all of the above characteristics [Westermark 2005].

Generally, the proteins associated with amyloid diseases used to be classified into two groups, globular proteins and natively unfolded proteins [Li 2009]. An example of the first group, is the prion protein family, on which amyloid formation requires a partial unfolding step, forming non stable oligomers-intermediates that after self-assembly into $\beta$-sheet aggregates. The other group, that includes amyloid $\beta$-protein (A$\beta$), tau and $\alpha$-synuclein, maybe the three best known proteins related with amyloid-diseases.

Initially, the toxic species were believed to be the fibrils deposition in tissue, responsible for cell death, but nowadays there are new evidences indicating that the intermediate oligomers, are the cytotoxic species on the amyloid pathway [Kayed 2003, Simoneau 2007].

To characterize and to study the formation and structural attributes of the amyloid fibrils, regarding their assembly state, we can use different biophysical methods and techniques. For the atomic structure, using solution-state NMR for the monomeric or oligomeric state, or solid-state NMR or X-ray for the fibril state. To study the secondary structure, the most common technique used is Circular dichroism (CD).

## 1.3 Novel approaches for the biophysics study of complex systems

The aim of this thesis is to develop novel approaches in Nuclear Magnetic resonance (NMR) and biophysics for the study of complex systems. Within this topic, this thesis aims to conduct a study on the development of methodologies for the analysis of complex mixtures. This can include the study of large chemical libraries, self-assembly systems, protein-ligand complex interactions and disordered biological systems. The best methods for this purpose are NMR, the diffusion measurement by NMR (DOSY) but also other techniques such as mass spectrometry, circular dichroism (CD), electron microscopy (EM) and small angle X-ray scattering (SAXS).

In order to achieve this objective, a first part of the thesis will be focused in the technique development, specifically on NMR. With the purpose, a complex system has to be set as object of study. The study of *out-of-equilibrium* systems is a good example of complexity. NMR spectroscopy is a good tool for monitoring chemical reactions. In particular, the DOSY experiment is well suited to characterize transient species and to separate them by the differences in diffusion coefficient. However, in these systems, the evolution of the concentration of different species in solution, interferes with the measurement process, and creates a bias in the determination of the diffusion coefficient which can lead to misinterpretation. We try to solve this difficulty by introducing a new experiment named *p*–DOSY.

In the second part of my thesis, the N-terminal domain of the Androgen Receptor (AR) is studied as a complex system. The Androgen Receptor belong to the superfamily of Nuclear Receptors (NRs). Human genome encodes 48 different nuclear receptors. NRs are a good example of the non-linearity in biology, only 48 different proteins are in charge of the control of the all human gene expression. Furthermore, these nuclear receptors share a common structure with high degree of conservation in some domains among the phylogenetic organisms. According to the literature, the N-terminal domain of the receptor plays an important role in the receptor activity, and is also described as being intrinsically disordered. This characteristic makes NMR the best technique to study it at the structural level. As said above, an IDP can assemble into amyloid aggregates. This phenomenon of fibers formation, will be shown within this work, can be regulated with a redox control of the environmental conditions. In addition, aggregation is reversible by the addition of a reducing agent. The discovery of this property opens up new possibilities in terms of supramolecular organization and self-assembly and could be generalized to other systems. The originality of the results reveals an unknown aspect of the receptor mechanism. And this region could be consider as a target site for drug development and allosteric regulation, if it could be demonstrate that it plays an important role for the receptor activity.

In addition, the presence of amyloid fibers can be used as a target site for weak interactions. The interaction with small molecules can disturb is kinetics aggregation, these weak interactions could cause the inhibition of the fibers formation and furthermore be used as amyloid inhibiting therapies for different amyloid diseases. On the other hand, they can cause the opposite effect and to accelerate the fibril formation.

As a result, a complete study of different examples of complex systems is carried out. This project covers many different aspects, from technical characterization of complexity to complex biological mechanisms.

# Bibliography

[Abul-Haija 2013] Yousef M Abul-Haija, Sangita Roy, Pim W J M Frederix, Nadeem Javid, Vineetha Jayawarna and Rein V Ulijn. *Biocatalytically Triggered Co-Assembly of Two-Component Core/Shell Nanofibers*. Small, September 2013. (Cited on pages 15 and 100.)

[Asencio-Hernández 2014] Julia Asencio-Hernández, Christine Ruhlmann, Alastair McEwen, Pascal Eberling, Yves Nominé, Jocelyn Céraline, Jean-Philippe Starck and Marc-André Delsuc. *Reversible amyloid fiber formation in the N terminus of androgen receptor.* ChemBioChem, vol. 15, no. 16, pages 2370–2373, November 2014. (Cited on pages 17 and 89.)

[Barabasi 1999] AL Barabasi and R Albert. *Emergence of scaling in random networks*. Science, vol. 286, no. 5439, pages 509–512, October 1999. (Cited on page 13.)

[Barroso 2015] Gustavo V Barroso and David R Luz. *On the limits of complexity in living forms.* J. Theor. Biol., vol. 379, pages 89–90, August 2015. (Cited on page 14.)

[Brooke 2009] G N Brooke and C L Bevan. *The role of androgen receptor mutations in prostate cancer progression*. Current genomics, 2009. (Cited on page 17.)

[Carnall 2010] Jacqui M A Carnall, Christopher A Waudby, Ana M Belenguer, Marc C A Stuart, Jérôme J.-P. Peyralans and Sijbren Otto. *Mechanosensitive Self-Replication Driven by Self-Organization*. Science, vol. 327, no. 5972, pages 1502–1506, March 2010. (Cited on pages 15, 100 and 101.)

[Crutchfield 2011] James P Crutchfield. *Between order and chaos*. Nature Physics, vol. 8, no. 1, pages 17–24, December 2011. (Cited on pages 14 and 15.)

[Davies 2008] P Davies, K Watt, S M Kelly, C Clark, N C Price and Iain J McEwan. *Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain*. Journal of Molecular Endocrinology, vol. 41, no. 5, pages 301–314, September 2008. (Cited on pages 17, 27 and 33.)

[Fuxreiter 2014] Monika Fuxreiter, Ágnes Tóth-Petróczy, Daniel A Kraut, Andreas T Matouschek, Roderick Y H Lim, Bin Xue, Lukasz Kurgan and Vladimir N Uversky. *Disordered Proteinaceous Machines*. Chem. Rev., vol. 114, no. 13, pages 6806–6843, July 2014. (Cited on page 18.)

[Geddes 1968] A J Geddes, K D Parker, E D T Atkins and E Beighton. *"Cross-$\beta$" conformation in proteins*. Journal of Molecular Biology, vol. 32, no. 2, pages 343–358, March 1968. (Cited on page 18.)

[Guharoy 2015] Mainak Guharoy, Kris Pauwels and Peter Tompa. *SnapShot: Intrinsic StructuralDisorder*. Cell, vol. 161, no. 5, pages 1230–1230.e1, May 2015. (Cited on page 17.)

[Kayed 2003] R Kayed. *Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis*. Science, vol. 300, no. 5618, pages 486–489, April 2003. (Cited on page 19.)

[Knowles 2014] Tuomas P J Knowles, Michele Vendruscolo and Christopher M Dobson. *The amyloid state and its associationwith protein misfolding diseases*. Nature Publishing Group, vol. 15, no. 6, pages 384–396, June 2014. (Cited on pages 18 and 148.)

[Lapouge 2008] Gaëlle Lapouge, Gemma Marcias, Eva Erdmann, Pascal Kessler, Marion Cruchant, Sebastian Serra, Jean-Pierre Bergerat and Jocelyn Céraline. *Specific properties of a C-terminal truncated androgen receptor detected in hormone refractory prostate cancer.* Adv. Exp. Med. Biol., vol. 617, pages 529–534, 2008. (Cited on pages 17 and 34.)

[Lehn 1993] J M Lehn. *Supramolecular chemistry.* Science, vol. 260, no. 5115, pages 1762–1763, June 1993. (Cited on page 15.)

[Li 2009] Huiyuan Li, Farid Rahimi, Sharmistha Sinha, Panchanan Maiti, Gal Bitan and Kazuma Murakami. *Amyloids and protein aggregation—analytical methods*. Encyclopedia of Analytical Chemistry, 2009. (Cited on pages 18 and 19.)

[Liu 2011] Yang-Yu Liu, Jean-Jacques Slotine and Albert-ászló Barabási. *2012-Nature-Liu*. Nature, vol. 473, no. 7346, pages 167–173, May 2011. (Cited on page 14.)

[Luscombe 2002] N M Luscombe and J M Thornton. *Protein-DNA interactions: Amino acid conservation and the effects of mutations on binding specificity*. Journal of Molecular Biology, vol. 320, no. 5, pages 991–1009, 2002. (Cited on page 17.)

[Marasco 2015] Daniela Marasco and Pasqualina Scognamiglio. *Identification of Inhibitors of Biological Interactions Involving Intrinsically Disordered Proteins*. IJMS, vol. 16, no. 4, pages 7394–7412, April 2015. (Cited on page 17.)

[Motlagh 2014] Hesam N Motlagh, James O Wrabl, Jing Li and Vincent J Hilser. *The ensemble nature of allostery*. Nature, vol. 508, no. 7496, pages 331–339, April 2014. (Cited on page 15.)

[Myers 2013] Edmund G Myers. *The most precise atomic mass measurements in Penning traps*. International Journal of Mass Spectrometry, vol. 349-350, pages 107–122, September 2013. (Cited on page 17.)

[Noinville 2008] Sylvie Noinville, Jean-François Chich and Human Rezaei. *Misfolding of the prion protein: linking biophysical and biological approaches.* Vet. Res., vol. 39, no. 4, page 48, July 2008. (Cited on page 16.)

[Papa 2014] Antonella Papa, Lixin Wan, Massimo Bonora, Leonardo Salmena, Min Sup Song, Robin M Hobbs, Andrea Lunardi, Kaitlyn Webster, Christopher Ng, Ryan H Newton, Nicholas Knoblauch, Jlenia Guarnerio, Keisuke Ito, Laurence A Turka, Andy H Beck, Paolo Pinton, Roderick T Bronson, Wenyi Wei and Pier Paolo Pandolfi. *Cancer-Associated PTEN Mutants Act in a Dominant-Negative Mannerto Suppress PTEN Protein Function*. Cell, vol. 157, no. 3, pages 595–610, April 2014. (Cited on page 17.)

[Pastor 2008] M Teresa Pastor, Nico Kümmerer, Vanessa Schubert, Alexandra Esteras-Chopo, Carlos G Dotti, Manuela López de la Paz and Luis Serrano. *Amyloid toxicity is independent of polypeptide sequence, length and chirality.* Journal of Molecular Biology, vol. 375, no. 3, pages 695–707, January 2008. (Cited on page 18.)

[Pauling 1951a] L Pauling and R B Corey. *The pleated sheet, a new layer configuration of polypeptide chains.* Proc. Natl. Acad. Sci. U.S.A., vol. 37, no. 5, pages 251–256, May 1951. (Cited on page 17.)

[Pauling 1951b] L Pauling, R B Corey and H R Branson. *The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain.* Proc. Natl. Acad. Sci. U.S.A., vol. 37, no. 4, pages 205–211, April 1951. (Cited on page 17.)

[Piana 2008] Stefano Piana, Alessandro Laio, Fabrizio Marinelli, Marleen Van Troys, David Bourry, Christophe Ampe and José C Martins. *Predicting the Effect of a Point Mutation on a Protein Fold: The Villin and Advillin Headpieces and Their Pro62Ala Mutants*. Journal of Molecular Biology, vol. 375, no. 2, pages 460–470, January 2008. (Cited on page 17.)

[Ross 2009] John Ross and Adam P Arkin. *Complex systems: from chemistry to systems biology.* Proceedings of the National Academy of Sciences, vol. 106, no. 16, pages 6433–6434, April 2009. (Cited on page 14.)

[Rubinov 2009] Boris Rubinov, Nathaniel Wagner, Hanna Rapaport and Gonen Ashkenasy. *Self-Replicating Amphiphilic $\beta$-Sheet Peptides*. Angew. Chem. Int. Ed., vol. 48, no. 36, pages 6683–6686, August 2009. (Cited on pages 15 and 100.)

[Sales-Pardo 2007] M Sales-Pardo, R Guimera, A A Moreira and L A N Amaral. *Extracting the hierarchical organization of complex systems*. Proceedings of the National Academy of Sciences, vol. 104, no. 39, pages 15224–15229, September 2007. (Cited on page 14.)

[Simoneau 2007] Steve Simoneau, Human Rezaei, Nicole Salès, Gunnar Kaiser-Schulz, Maxime Lefebvre-Roque, Catherine Vidal, Jean-Guy Fournier, Julien Comte, Franziska Wopfner, Jeanne Grosclaude, Hermann Schätzl and Corinne Ida Lasmézas. *In vitro and in vivo neurotoxicity of prion protein oligomers.* PLoS Pathog., vol. 3, no. 8, page e125, August 2007. (Cited on page 19.)

[Sipe 2000] Jean D Sipe and Alan S Cohen. *Review: History of the Amyloid Fibril*. Journal of Structural Biology, vol. 130, no. 2-3, pages 88–98, June 2000. (Cited on page 18.)

[Stavrinides 2013] S G Stavrinides, S Banerjee and H Suleyman. *Chaos and Complex Systems*. Springer Complexity, 2013. (Cited on page 14.)

[Westermark 2005] Per Westermark. *Aspects on human amyloid forms and their fibril polypeptides.* FEBS J., vol. 272, no. 23, pages 5942–5949, December 2005. (Cited on pages 18 and 19.)

Chapter 2

# Androgen Receptor and its role in Prostate Cancer



A good example of complexity is the nuclear receptors (NR) super family. Nuclear receptors form a superfamily of phylogenetically related proteins, there are 48 NR in the human genome [Robinson-Rechavi 2001]. They play a crucial role in eukaryotic differentiation, development and metabolic homeostasis [Ribeiro M D 1995]. The nuclear receptor proteins are composed as a single polypeptide chain with five different domains. A disordered N-terminal domain (NTD), which contains the activation function 1 (AF1) whose action is independent of the presence of ligand. A highly conserved DNA-binding domain (DBD) followed by a flexible hinge. The ligand binding domain (LBD) which contains the activation function 2 (AF2) and finally the C-terminal domain [Huang 2010]. Nuclear receptors according to their families, can form interactions with DNA as a monomer, homo- or heterodimer.

The superfamily is sub-divided into different classes. A first class is the thyroid/retinoid family, including the thyroid receptor (TR), vitamin D receptor (VDR), the retinoic acid receptor (RAR) and the peroxisome proliferator-activated receptor (PPAR). A second class defined as retinoid X receptor family, with the retinoid X receptor (RXR) or the testicular receptor (TR2-4). A third class is the steroid receptor family, including the estrogen receptor (ER), the progesterone receptor (PR), the androgen receptor (AR), the glucocorticoid receptor (GR) and the mineralocorticoid receptor (MR). Finally, a last class of nuclear receptor is known as the orphan receptor family, including a set of proteins sharing significant sequence homology to other known nuclear receptors but with unknowns ligands.

## 2.1 Androgen Receptor

The androgen receptor (AR) belongs to the steroid hormone group of nuclear receptors, (NR3C4, subfamily 3, group c and gene 4). AR is located in the X chromosome, in the region q11-12 [Migeon 1981]. The AR gene presents 8 exons that encodes a 110kDa protein consisting of 919 amino acids and occurs as a single form in mammalian cells. The exon 1 encodes the NTD, the DBD is encoded by the exon 2 and 3 that also encode respectively the zing finger allowing the DNA recognition and the zing finger used as dimerization interface. The exons from 4 to 8, encode the hinge and the LBD [Gelmann 2002].

The Androgen Receptor is a ligand-dependent transcription factor. AR is activated by binding of the androgenic hormones, $5\alpha$-dihydrotestosterone (DHT) or testosterone. Upon activation, AR is translocated into the nucleus, where it recognizes specific DNA sequences and regulates genes expression [van de Wijngaart 2011]. However, AR is also responsible for the initiation of male sexual differentiation, and to the development and maintenance of male reproductive tissues [Li 2009, Matsumoto 2013].

AR presents the architecture in 5 domains common to all Nuclear Receptor: a disordered N-terminal region (NTD), which presents the intrinsic activation function (AF-1), a DNA-binding domain (DBD) followed by a short hinge region, a ligand-binding domain (LBD) which presents the ligand dependent activation function (AF-2), and finally a short C-terminal region.

### The N-terminal domain

AR has the peculiarity of presenting a very large NTD (Figure 2.1) that represents more than half of the entire protein, around 60% of the AR size (residues 1-555). This region contains the major transactivation function 1 (AF1), which contains transactivation units Tau-1 and Tau-5 (see Figure 2.1) important for AR transactivation. Deletion of the Ligand Binding Domain results in a constitutively active AR, implying that AF1 contributes all the activity of the receptor [Bevan 1999]. AF1 appears to be the predominant activation region in AR and depends on androgen binding for activation [Simental 1991].

The NTD-AR contains an FxxLF motif (x is any amino acid) that has been described to have an important role for interactions with the LBD [He 2000]. This represents a key interaction for receptor function, stabilizing bound hormone [He 1999] and behaving as competitor for AF2 surface with certain co-regulators, providing an alternative mechanism for AR transcriptional activity [He 2002, Li 2009]. Furthermore, the FQNLF motif in the AR N-terminal domain is also involved in the receptor dimerization [van Royen 2012].

This NTD of AR is a flexible domain with a high degree of intrinsic disorder characteristics of

an Intrinsically Disordered domain (IDP) [Lavery 2008, McEwan 2011]. Several homoresidues tracts of Glutamine (polyQ), Proline (polyP), Glycine (polyG) and Alanine (polyA) are present in the NTD-AR (Figure 2.1). It also presents numerous hydrophobic and aromatic residues Methionine (M), Tryptophan (W), Phenylalanine (F) or Tyrosine (Y), usually absent in disordered regions.

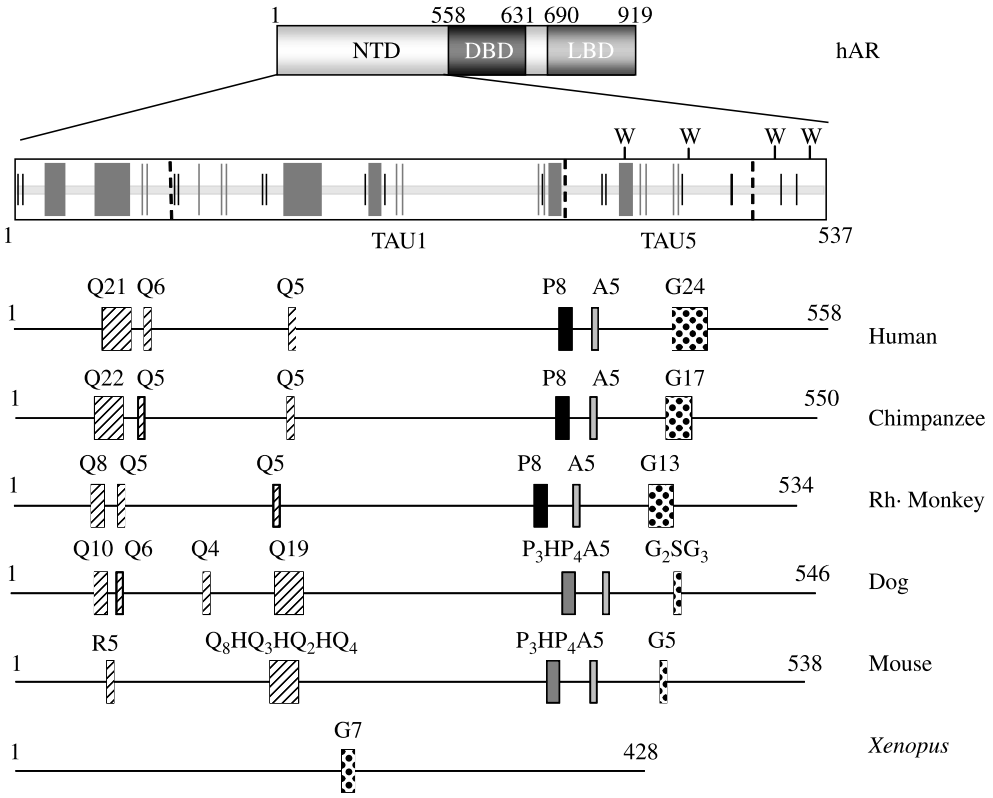



Figure 2.1: NTD-AR organization, presenting the location of the tryptophanes and of the homoresidues regions Q21, Q6, Q5, P8, Q5 and G24. L4 not shown here is positioned just Nter of Q21 (Figure from [Davies 2008]).

The AR polyQ track length has been used as a marker in human population genetics and is involved in the conformation and folding of the androgen receptor amino-terminal domain [Davies 2008]. The presence of a repeat track polyQ with a length of more than 40 glutamine residues results in Kennedy's disease or adult onset spinal bulbar muscular atrophy [La Spada 1991, Tan 2014]. And a shorter polyQ repeat length (less than 21 glutamine residues) increases the risk of prostate cancer [Zhou 1994]. The AR polyQ length does not present the same conservation among the mammals (see Figure 2.1). Contrary to the human sequence, many species present a short glutamine repeat in the same position than the Human

polyQ, but remarkably, they present a glutamine repeat expansion slightly further where is located the human Q5 (see Figure 2.1) [Tan 1988, Lubahn 1988].

The limited structure of the NTD-AR is thought to require interactions with other proteins to assume correct folding for further protein-protein interactions including interactions with bridging factors such as CBP and the basal transcriptional machinery for active transcription. [McEwan 2011, Kumar 2011] In a study reported in the literature, [Reid 2002] it was shown by circular dichroism (CD) and controlled proteolysis, that some partial structure is present in the NTD-AR and is enhanced by the presence of osmolyte such as TFE or TMAO. The same work showed by fluorescence that the environment of the tryptophan residues located in the 400-530 region are affected by the length of the polyQ located in the 60-80 region, thus more than 300 residues away.

### The DNA binding domain

The DNA binding domain (DBD), is the central region of the AR, a cysteine-rich region from residue 556 to 623. This domain is highly conserved among the nuclear receptors family. DBD function is the recognition of response elements (REs), DNA regulation sites present in the promoter of target genes [Pawlak 2012]. The DBD mediates the AR-DNA binding as a dimer, and each monomeric form is composed of two zinc fingers, each zinc ion being coordinated to four cysteine residues [Shaffer 2004](PDB:1R4I). This domain is highly structured, compact and globular, composed by two perpendicular $\alpha$-helix stabilized by the 2 zinc-fingers, a short $\beta$-sheet is also present as well as some residues without secondary structure [Pawlak 2012]. The first $\alpha$-helix of the N-terminal zinc finger, interacts directly with the DNA major groove, nucleotides from the hormone response element.

The proximal box (P-box) is a short motif, CxxCK, located in the first zinc finger, responsible for the DNA specific binding. The helix in C-terminal contributes to the stabilization of the whole complex by establishing a weak non-specific binding to DNA. Additional residues at the second zinc finger forming the D box are involved in receptor dimerization and also contribute to the stabilization of the complex.

### The hinge

The DNA-binding domain is contiguous to a hinge region that spans amino acid residues 624–665 and links with LBD. The nuclear localization signal (NLS) (residues 617-633), forms part of the DBD and the hinge and is responsible for nuclear import of the receptor. The NLS is a nuclear targeting signal composed of two clusters of basic amino acids separated by

10 residues. This region presents the 629-RKLKKLGN-636 motif, also involved in the DNA binding as a coactivator recruitment and in the N/C interaction [Tan 2014].

### The ligand binding domain

Similarly to the DBD, the ligand binding domain (LBD) is a highly conserved domain and its structure has been well characterized. The LBD composes the last C-terminal part of the androgen receptor (residues 666-919). The LBD allows ligand-dependent activation of the receptor, receptor homodimerization and presents binding sites for chaperones interactions [Shaffer 2004]. The LBD architecture presents eleven $\alpha$-helices (H1-H12) and two antiparallel $\beta$-sheets. This tridimensional structure, as for the DBD, is similar in all the members of the nuclear receptors family but in the androgen receptor the helix H2 is not present, and is replaced by a flexible linker. The LBD contains a ligand-dependent transactivation domain, AF2 (activation function 2), and mutations or deletion of this region drastically reduces transcriptional activation as response to a ligand. Moreover, Ligand binding induces a conformational change which stabilizes the helix H12, allowing the interaction with the NTD-AR [Li 2009, Pawlak 2012].

## 2.2 Androgen Receptor Ligands

A large diversity of small molecules interact with the androgen receptor, naturals or synthetics. These molecules can be classified as agonists or antagonists, regarding their ability to activate or inhibit the transcription of target genes and to modulate AR function.

### AR agonists

Testosterone and Dihydrotestosterone (DHT), are the most important endogenous androgens that exists (Figure 2.2). Testosterone is secreted by the testicles in males and in the ovaries in females, and it was first characterized in 1935 by Prof Ernst LAQUEUR. DHT was discovered later, when it was observed that a fraction of the testosterone is metabolized to DHT by a 5$\alpha$-reductase [Tan 2014].

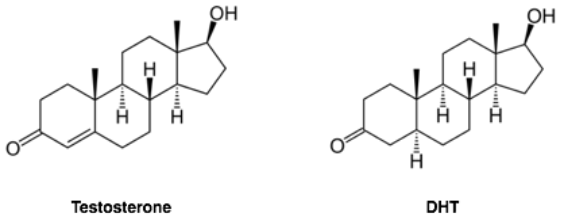



Figure 2.2: Structural representation of AR agonists. Testosterone and Dihydrotestosterone (DHT).

Testosterone differs from DHT by the presence of a single double bond ($C_4$-$C_5$) on the ring A. The absence of this double bond increases two to three times the affinity of DHT for the receptor and the dissociation rate ($k_{off}$) is five times slower than testosterone [Grino 1990, Tan 2014]. Which results in a affinity difference of the specific functions associated with the DHT.

## AR antagonists

The antagonist ligands of the androgen receptor are known as antiandrogens, steroidal or non-steroidal, and they antagonize the androgens actions by competing for the binding sites in AR.

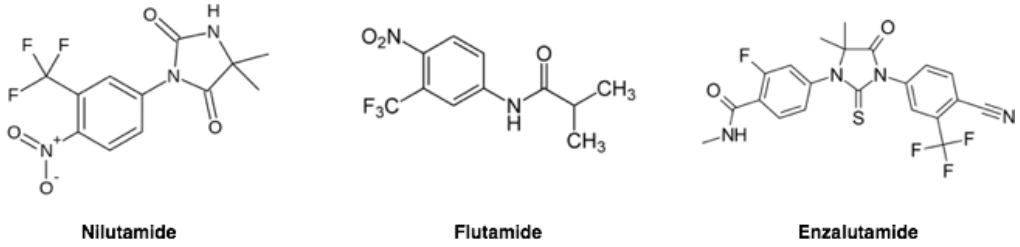



Figure 2.3: Structural representation of antiandrogens. Nilutamide and flutamide, both toluidide derivates. Enzalutamide a second-generation of antiandrogens.

Steroidal antiandrogens competitively inhibit the binding of testosterone or DHT to the Androgen receptor and alter gene expression. Due to undesired side effects they have limited clinical applications. Non-steroidal antiandrogens, toluidide derivates such as nilutamide and flutamide (Figure 2.3), counter androgens and have no steroidal effects. The lack of androgenic effects makes them acceptable for the treatment of prostate cancer [Singh 2000]. Antiandrogens inhibit circulating androgens by blocking androgen receptors, suppressing androgen synthesis, or acting in both ways. In 2012, a second-generation of AR antagonist was

developed, enzalutamide (Figure 2.3), a more potent antagonist that binds to the AR ligand-binding pocket with higher affinity than the first generation molecules, and also restrain the receptor translocation into the nucleus [Tran 2009].

## 2.3 Androgen Receptor mode of action

Androgen Receptor is a hormono-inducible DNA-binding transcription factor that can adopt different modes of action (Figure 2.4). In classical AR signaling, in the absence of ligand, AR is located in the cytosol and forms a complex with heat shock proteins (HSP), chaperone and co-chaperone proteins with the objective to maintain AR in a ligand-binding competent state [Cano 2013].

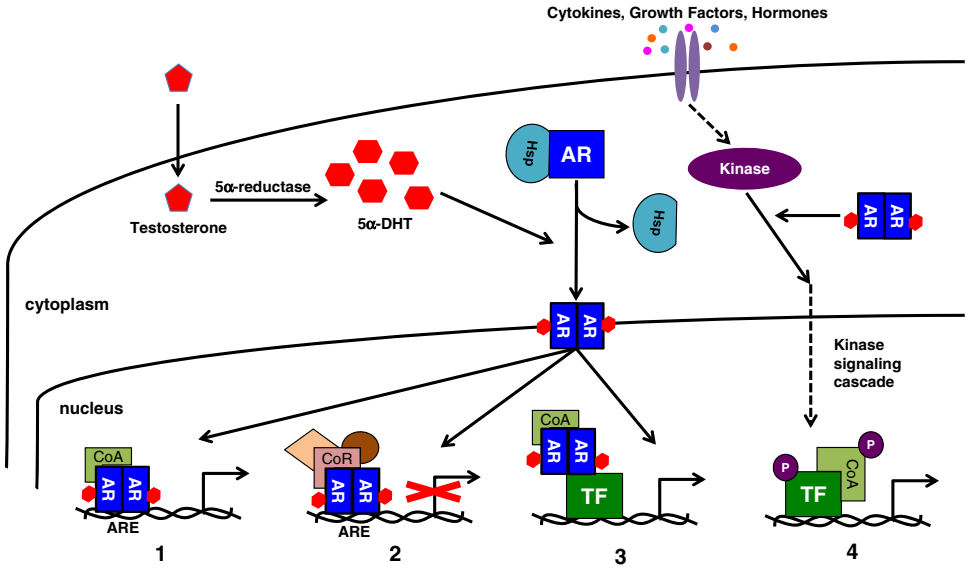



Figure 2.4: Structural representation of the different mechanisms described for the mode of action of AR (Figure from [Shafi 2013]).

Testosterone is synthesized in the testes in men, and in the ovaries in women. In men, once it is produced, the free form enters prostate cells. In the prostate, testosterone is converted to 5$\alpha$-DHT by the enzyme 5$\alpha$-reductase, which is the high affinity ligand for AR and promotes the growth and survival of prostate cells. Upon hormone binding, HSP are displaced from AR allowing hormone-bound AR to dimerize and to translocate to the nucleus. Then, different modes of action can occur. In the best characterized and the most common case (see Figure 2.4 **1**), AR binds to AREs (androgen response elements) on the DNA and interacts with different transcription co-regulators to regulate gene expression. These different co-activators

complexes remodel the chromatin allowing general transcription factors and RNA Polymerase II to begin the process of transcription [Shafi 2013, Matsumoto 2013].

In a second less well characterized mechanism (Figure 2.4 **2**) AR recruits co-repressors and can inhibit transcription [Shafi 2013]. Moreover, in a third mechanism (Figure 2.4 **3**), AR also regulates transcription through a tethering mechanism by binding to other transcription factors (TFs) and recruiting additional co-activators to target genes. Finally, in addition to classical AR signaling, AR can initiate extranuclear signaling (Figure 2.4 **4**). The activated complex formed by hormone-bound AR, dimerizes and can interact with kinases. As a result of this interaction, an activation of a kinase cascade occurs. The resulting signaling cascade can activate other TFs in the absence of direct AR binding to the gene [Shafi 2013].

## 2.4 Androgen Receptor and different related pathologies

The androgen receptor has a fundamental role in androgen signaling for differentiation of male sexual characteristics. Its role in male physiology and the reproductive system, is highlighted by the hereditary disorder of the *Androgen Insensitivity Syndrome* (AIS). AIS can be defined as the partial or complete inability of cells to respond to androgens [Hughes 2012]. In the AR, missense[1] and nonsense[2] mutations cause a large number of abnormalities in male development. The effects of these mutations can result from a light virilization to a total reverse phenotype male-female [Matsumoto 2013]. Nowadays, around 1000 different mutations have been identified in the human AR gene, described in the Androgen Receptor Gene Mutations Database World Wide Web Server (http://androgendb.mcgill.ca/AR23C.pdf) [Gottlieb 1998].

Androgens also regulate *spermatogenesis* via the expression of AR in somatic cells of the testes [Verhoeven 2010]. They could also have an effect mediated by androgens in the development of the *brain masculinization* but this is still controversial [Arnold 1984]. Age-related *osteoporosis* is a poorly known phenomenon; although the reduction of the use of androgens is closely correlated with a wide variety of conditions associated with aging including reduced muscle mass, decrease in bone mass and reduced strength [Kang 2008]. AR would have a direct role in the maintenance of bone through the matrix synthesis and their mineralization [Matsumoto 2013].

Furthermore, an other well known androgen-related disease is the *Spinal Bulbar Muscular Atrophy* (SBMA) or Kennedy's disease. The SBMA is a progressive neurodegenerative disease

[1] a point mutation where a single nucleotide is changed to cause substitution of a different amino acid

[2] a point mutation in a sequence of DNA that results in a premature stop codon resulting in aberrant splicing of AR mRNA

that appears when there is a polyQ expansion in the N-terminal domain of the androgen receptor. Polymorphism repeated CAG normally consists of 9 to 36 repetitions, expansion beyond 40 causes neurotoxicity [La Spada 1991, Davies 2008].

Androgens have been always associated with male and the men reproductive system development, even if AR is located on the human chromosome X.

Despite the fact that AR is located on the human chromosome X, little is known about androgens actions in female physiology, but the detection of hyperandrogenism in *polycystic ovarian syndrome* (PCOS) which causes infertility with abnormal menstrual cycles suggests that the androgen signaling plays an important role in women [Matsumoto 2008].

As has been described above, AR is involved in many different pathologies. Nevertheless, prostate cancer it remains the best known and most studied disease related with the androgen receptor.

## 2.5 Androgen Receptor's role in prostate cancer

The androgen receptor activity is directly related with prostate cancer, which is a leading cause of death of men in the industrialized world. It is the most commonly diagnosed cancer in men in U.S.A and the second leading cause of cancer deaths. With an increasing incidence, 71.500 new cases per year in France in 2010, remains a major public health problem. Around 159 of the AR mutation described, predispose males to develop prostate cancer [Tan 2014]. In addition, a shorter polyQ (<21 residues) influences prostate cancer risk in men [Davies 2008].

The prostate gland is under a strict hormonal control. Prostate cancer cells, as normal prostate cells, are androgens dependent for their growth and their survival. The role of androgens in prostate cancer has been described by Huggins and Hodges in 1941, they demonstrated the reduction of prostate tumors during androgen deprivation. Charles Huggins received the Nobel Prize in Medicine and Physiology in 1966 for the treatment of cancer from hormone therapy. As prostate cancer develops initially as an androgen-dependent disease, this explains the initial effectiveness of anti-androgens therapies for advanced prostate cancers. However, in most cases, a resistance to androgen deprivation develops in one to three years and prostate cancer cells undergo recurrent growth despite low levels of circulating androgens and the presence of synthetic androgen antagonists [Matsumoto 2013].

Castration or androgen deprivation therapy is the standard treatment for metastatic prostate cancer local recurrence after radiotherapy. Inevitably, an exhaust androgen suppression is observed after a median of 12 to 18 months. Many factors contribute to the activation of AR despite a low level of androgens. Alterations in androgen signaling pathway have been as-

sociated with castration resistance such as amplification of the androgen receptor gene, the emergence of mutations in the ligand binding domain which extend the receiver sensitivity spectrum to other steroids that androgens or expression of constitutively active forms, without the ligand binding domain [Funderburk 2009]. Relapsed prostate cancer, known as Castration Resistant Prostate Cancer (CRPC), is associated with a high mortality due to the lack of effective longterm treatment strategies. It is a complex disease that develops during the course of anti-androgen therapy but nonetheless maintains a dependence on AR for its metastatic potential.

The mechanism for the CRPC development remains unclear, furthermore different possibilities have been proposed: an increased sensitivity of the AR to its agonists; some mutations that makes AR reactive to non-androgen ligands; activation of the AR independent of ligands or an other mechanism AR independent [Tan 2014].

Raising evidences support the concept that the development of CRPC is causally related to continued transactivation of AR. All current therapies that target the AR are dependent on the presence of the C-terminal LBD and its activation function 2 (AF2). Consequently when the CRPC is developed, and constitutively active AR splice variants that lack the LBD ($\Delta$LBD) appear [Lapouge 2008], these therapies fail. Then, the activity of the receptor is regulated by the N-terminal domain and its activation function 1 (AF1). Functionally, AR N-terminal domain (NTD) plays the primary role in regulating target gene transcription activation nevertheless the absence of the AF2 makes the receptor as constitutively active. Therefore modulation of NTD function can be considered as an efficient strategy to target AR action.

## 2.6 New strategies to target the Androgen Receptor

The fact that current therapies for the prostate cancer (based on androgem ablation or antiandrogens) fail, put in evidence the necessity of developing new drugs to reduce androgens levels as well as more potent AR antagonists. As explained above, most antiandrogens are designed to target the ligand binding domain. However, when CRPC is developed, the AR splice variants $\Delta$LBD appear and antiandrogens targeting the LBD become useless, without effect in the receptor activity. Then, the NTD-AR with its AF1 becomes the responsible of the receptor activity and appears as a new target strategy for drug development.

The N-terminal domain, as has been introduce previously, is a wide domain with 550 residues and with the absence of structure. These characteristics make it difficult to study it and it has been poorly studied. This structural plasticity of the NTD allows an allosteric regulation of protein folding and function in response to hormone, DNA response element

architecture, and coregulatory protein binding partners. This aspect was recently investigated by Kumar and McEwan in a comprehensive review on the allosteric modulation of steroid hormone receptors [Kumar 2012]. NTD-AR is also known to be the site of many cancer related mutation, either as genetic predisposition for prostate cancer or as somatic mutations, usually specific to the castration-resistant tumor [Matsumoto 2013].

Recent studies [Sadar 2011] have shown that several small molecules (see Figure 2.5) interact with NTD-AR and modulate its activity, for instance by blocking its AF-1 activity, thus showing a potential new therapeutic route. Chromatin immunoprecipitation analyses showed that EPI-001 reduced the interaction of AR with the response elements of the target genes, including PSA and TM-PRSS2. EPI-001 binds to the NTD of AR and inhibits the recruitment of the transcriptional cofactors CBP and RAP74, suggesting that EPI-001 functions by blocking protein–protein interactions at the NTD. EPI-001 reduced androgen-dependent and androgen-independent proliferation of LNCaP prostate cancer cells that express AR and had no effect on cells that do not require AR activation for growth [Andersen 2010, McEwan 2011, Myung 2013]. Furthermore, EPI-001 is a derivate of Bisphenol-A (BPA) molecule which is a well-known endocrine disruptor. A recent study [Delfosse 2012] shows that BPA act as partial agonists of estrogen receptors and also present an antiandrogen activity [Paris 2002]. The NTD-AR has long been known to have an important role in the transcription activity, and was shown to interact with the AR natural action, when over-expressed as a decoy protein in cancer cells and present an antitumor activity.

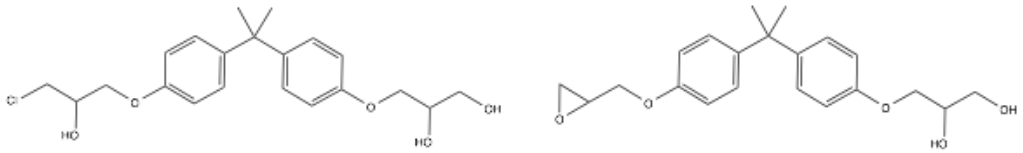



Figure 2.5: Chemical structure of molecules EPI–001 and EPI–056. Molecules reported as new potential drugs for prostate cancer (See [Sadar 2011])

All these recent results explain the increasing interest in the study of the action of the NTD of AR, but also the need for a more molecular investigation of the mechanism of action of NTD-AR [Quayle 2007].

# Bibliography

[Andersen 2010] Raymond J Andersen, Nasrin R Mawji, Jun Wang, Gang Wang, Simon Haile, Jae-Kyung Myung, Kate Watt, Teresa Tam, Yu Chi Yang, Carmen A Bañuelos, David E Williams, Iain J McEwan, Yuzhou Wang and Marianne D Sadar. *Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor*. Cancer Cell, vol. 17, no. 6, pages 535–546, June 2010. (Cited on pages 35 and 123.)

[Arnold 1984] A P Arnold and R A Gorski. *Gonadal-Steroid Induction of Structural Sex-Differences in the Central Nervous-System*. Annu. Rev. Neurosci., vol. 7, pages 413–442, 1984. (Cited on page 32.)

[Bevan 1999] C L Bevan, S Hoare, F Claessens, D M Heery and M G Parker. *The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1.* Mol. Cell. Biol., vol. 19, no. 12, pages 8383–8392, December 1999. (Cited on page 26.)

[Cano 2013] Laia Querol Cano, Derek N Lavery and Charlotte L Bevan. *Mini-review: Foldosome regulation of androgen receptor action in prostate cancer.* MOLECULAR AND CELLULAR ENDOCRINOLOGY, vol. 369, no. 1-2, pages 52–62, April 2013. (Cited on page 31.)

[Davies 2008] P Davies, K Watt, S M Kelly, C Clark, N C Price and Iain J McEwan. *Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain*. Journal of Molecular Endocrinology, vol. 41, no. 5, pages 301–314, September 2008. (Cited on pages 17, 27 and 33.)

[Delfosse 2012] V Delfosse, M Grimaldi, J L Pons, A Boulahtouf, A le Maire, V Cavailles, G Labesse, W Bourguet and P Balaguer. *Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes*. Proceedings of the National Academy of Sciences, vol. 109, no. 37, pages 14930–14935, September 2012. (Cited on page 35.)

[Funderburk 2009] Sarah F Funderburk, Liubov Shatkina, Sigrun Mink, Qun Weis, Susanne Weg-Remers and Andrew C B Cato. *Specific N-terminal mutations in the human androgen receptor induce cytotoxicity*. Neurobiology of Aging, vol. 30, no. 11, pages 1851–1864, November 2009. (Cited on page 34.)

[Gelmann 2002] E P Gelmann. *Molecular Biology of the Androgen Receptor.* Journal of Clinical Oncology, vol. 20, no. 13, pages 3001–3015, July 2002. (Cited on page 26.)

[Gottlieb 1998] B Gottlieb, H Lehvaslaiho, L K Beitel, R Lumbroso, L Pinsky and M Trifiro. *The Androgen Receptor Gene Mutations Database.* Nucleic Acids Res., vol. 26, no. 1, pages 234–238, January 1998. (Cited on pages 32 and 84.)

[Grino 1990] P B Grino, J E Griffin and J D Wilson. *Testosterone at high concentrations interacts with the human androgen receptor similarly to dihydrotestosterone.* Endocrinology, vol. 126, no. 2, pages 1165–1172, February 1990. (Cited on page 30.)

[He 1999] B He, J A Kemppainen, J J Voegel, H Gronemeyer and E M Wilson. *Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH(2)-terminal domain.* J. Biol. Chem., vol. 274, no. 52, pages 37219–37225, December 1999. (Cited on page 26.)

[He 2000] B He, J A Kemppainen and E M Wilson. *FXXLF and WXXLF Sequences Mediate the NH2-terminal Interaction with the Ligand Binding Domain of the Androgen Receptor.* Journal of Biological Chemistry, vol. 275, no. 30, pages 22986–22994, July 2000. (Cited on pages 26 and 84.)

[He 2002] Bin He, Lori W Lee, John T Minges and Elizabeth M Wilson. *Dependence of selective gene activation on the androgen receptor NH2- and COOH-terminal interaction.* J. Biol. Chem., vol. 277, no. 28, pages 25631–25639, July 2002. (Cited on page 26.)

[Huang 2010] Pengxiang Huang, Vikas Chandra and Fraydoon Rastinejad. *Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics.* Annu. Rev. Physiol., vol. 72, pages 247–272, 2010. (Cited on page 25.)

[Hughes 2012] Ieuan A Hughes, John D Davies, Trevor I Bunch, Vickie Pasterski, Kiki Mastroyannopoulou and Jane MacDougall. *Androgen insensitivity syndrome.* Lancet, vol. 380, no. 9851, pages 1419–1428, October 2012. (Cited on page 32.)

[Kang 2008] Hong-Yo Kang, Chih-Rong Shyr, Chiung-Kuei Huang, Meng-Yin Tsai, Hideo Orimo, Pei-Chun Lin, Chawnshang Chang and Ko-En Huang. *Altered TNSALP Expression and Phosphate Regulation Contribute to Reduced Mineralization in Mice Lacking Androgen Receptor.* Mol. Cell. Biol., vol. 28, no. 24, pages 7354–7367, 2008. (Cited on page 32.)

[Kumar 2011] R Kumar, H Atamna, M N Zakharov, S Bhasin, S H Khan and R Jasuja. *Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy*. Life Sciences, vol. 88, no. 13-14, pages 565–571, March 2011. (Cited on pages 28 and 101.)

[Kumar 2012] Raj Kumar and Iain J McEwan. *Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation.* Endocr. Rev., vol. 33, no. 2, pages 271–299, April 2012. (Cited on pages 35 and 101.)

[La Spada 1991] A R La Spada, E M Wilson, D B Lubahn, A E Harding and K H Fischbeck. *Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy.* Nature, vol. 352, no. 6330, pages 77–79, July 1991. (Cited on pages 27 and 33.)

[Lapouge 2008] Gaëlle Lapouge, Gemma Marcias, Eva Erdmann, Pascal Kessler, Marion Cruchant, Sebastian Serra, Jean-Pierre Bergerat and Jocelyn Céraline. *Specific properties of a C-terminal truncated androgen receptor detected in hormone refractory prostate cancer.* Adv. Exp. Med. Biol., vol. 617, pages 529–534, 2008. (Cited on pages 17 and 34.)

[Lavery 2008] Derek N Lavery and Iain J McEwan. *Structural Characterization of the Native NH 2-Terminal Transactivation Domain of the Human Androgen Receptor: A Collapsed Disordered Conformation Underlies Structural Plasticity and Protein-Induced Folding* †. Biochemistry, vol. 47, no. 11, pages 3360–3369, March 2008. (Cited on pages 27 and 101.)

[Li 2009] Jin Li and Farook Al-Azzawi. *Mechanism of androgen receptor action.* Maturitas, vol. 63, no. 2, pages 142–148, June 2009. (Cited on pages 26 and 29.)

[Lubahn 1988] D B Lubahn, D R Joseph, M Sar, J Tan, H N Higgs, R E Larson, F S French and E M Wilson. *The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate.* Mol. Endocrinol., vol. 2, no. 12, pages 1265–1275, December 1988. (Cited on page 28.)

[Matsumoto 2008] Takahiro Matsumoto, Hiroko Shiina, Hirotaka Kawano, Takashi Sato and Shigeaki Kato. *Androgen receptor functions in male and female physiology*. The Journal of Steroid Biochemistry and Molecular Biology, vol. 109, no. 3-5, pages 236–241, April 2008. (Cited on page 33.)

[Matsumoto 2013] Takahiro Matsumoto, Matomo Sakari, Maiko Okada, Atsushi Yokoyama, Sayuri Takahashi, Alexander Kouzmenko and Shigeaki Kato. *The Androgen Receptor*

*in Health and Disease*. Annu. Rev. Physiol., vol. 75, no. 1, pages 201–224, February 2013. (Cited on pages 26, 32, 33 and 35.)

[McEwan 2011] Iain J McEwan. *Intrinsic disorder in the androgen receptor: identification, characterisation and drugability*. Mol. BioSyst., vol. 8, no. 1, page 82, 2011. (Cited on pages 27, 28, 35 and 123.)

[Migeon 1981] B R Migeon, T R Brown, J Axelman and C J Migeon. *Studies of the locus for androgen receptor: localization on the human X chromosome and evidence for homology with the Tfm locus in the mouse*. Proc. Natl. Acad. Sci. U.S.A., vol. 78, no. 10, pages 6339–6343, October 1981. (Cited on page 26.)

[Myung 2013] Jae-Kyung Myung, Carmen A Bañuelos, Javier Garcia Fernandez, Nasrin R Mawji, Jun Wang, Amy H Tien, Yu Chi Yang, Iran Tavakoli, Simon Haile, Kate Watt, Iain J McEwan, Stephen Plymate, Raymond J Andersen and Marianne D Sadar. *An androgen receptor N-terminal domain antagonist for treating prostate cancer.* J. Clin. Invest., vol. 123, no. 7, pages 2948–2960, July 2013. (Cited on page 35.)

[Paris 2002] Françoise Paris, Patrick Balaguer, Béatrice Térouanne, Nadège Servant, Caroline Lacoste, Jean-Pierre Cravedi, Jean-Claude Nicolas and Charles Sultan. *Phenylphenols, biphenols, bisphenol-A and 4-tert-octylphenol exhibit alpha and beta estrogen activities and antiandrogen activity in reporter cell lines.* MOLECULAR AND CELLULAR ENDOCRINOLOGY, vol. 193, no. 1-2, pages 43–49, July 2002. (Cited on page 35.)

[Pawlak 2012] Michal Pawlak, Philippe Lefebvre and Bart Staels. *General molecular biology and architecture of nuclear receptors.* Current Topics in Medicinal Chemistry, vol. 12, no. 6, pages 486–504, 2012. (Cited on pages 28 and 29.)

[Quayle 2007] Steven N Quayle, Nasrin R Mawji, Jun Wang and Marianne D Sadar. *Androgen receptor decoy molecules block the growth of prostate cancer.* Proc. Natl. Acad. Sci. U.S.A., vol. 104, no. 4, pages 1331–1336, January 2007. (Cited on page 35.)

[Reid 2002] J Reid. *Conformational Analysis of the Androgen Receptor Amino-terminal Domain Involved in Transactivation. Influence of structure-stabilizing solutes and protein-protein interactions*. Journal of Biological Chemistry, vol. 277, no. 22, pages 20079–20086, March 2002. (Cited on page 28.)

[Ribeiro M D 1995] Ralff C J Ribeiro M D, Peter J Kushner Ph D and John D Baxter M D. *The nuclear hormone receptro gene superfamily*. Annu. Rev. Med., vol. 46, no. 1, pages 443–453, February 1995. (Cited on page 25.)

[Robinson-Rechavi 2001] M Robinson-Rechavi, A S Carpentier, M Duffraisse and V Laudet. *How many nuclear hormone receptors are there in the human genome?* Trends Genet., vol. 17, no. 10, pages 554–556, October 2001. (Cited on page 25.)

[Sadar 2011] Marianne D Sadar. *Small molecule inhibitors targeting the "achilles' heel" of androgen receptor activity.* Cancer Research, vol. 71, no. 4, pages 1208–1213, February 2011. (Cited on page 35.)

[Shaffer 2004] Paul L Shaffer, Arif Jivan, D Eric Dollins, Frank Claessens and Daniel T Gewirth. *Structural basis of androgen receptor binding to selective androgen response elements.* Proc. Natl. Acad. Sci. U.S.A., vol. 101, no. 14, pages 4758–4763, April 2004. (Cited on pages 28 and 29.)

[Shafi 2013] Ayesha A Shafi, Aihua E Yen and Nancy L Weigel. *Androgen receptors in hormone-dependent and castration-resistant prostate cancer.* Pharmacol. Ther., vol. 140, no. 3, pages 223–238, December 2013. (Cited on pages 31 and 32.)

[Simental 1991] J A Simental, M Sar, M V Lane, F S French and E M Wilson. *Transcriptional activation and nuclear targeting signals of the human androgen receptor.* J. Biol. Chem., vol. 266, no. 1, pages 510–518, January 1991. (Cited on page 26.)

[Singh 2000] S M Singh, S Gauthier and F Labrie. *Androgen receptor antagonists (antiandrogens): structure-activity relationships.* Curr. Med. Chem., vol. 7, no. 2, pages 211–247, February 2000. (Cited on page 30.)

[Tan 1988] J A Tan, D R Joseph, V E Quarmby, D B Lubahn, M Sar, F S French and E M Wilson. *The rat androgen receptor: primary structure, autoregulation of its messenger ribonucleic acid, and immunocytochemical localization of the receptor protein.* Mol. Endocrinol., vol. 2, no. 12, pages 1276–1285, December 1988. (Cited on page 28.)

[Tan 2014] MH Eileen Tan, Jun Li, H Eric Xu, Karsten Melcher and Eu-leong Yong. *Androgen receptor: structure, role in prostate cancer and drug discovery*. Nature Publishing Group, vol. 36, no. 1, pages 3–23, June 2014. (Cited on pages 27, 29, 30, 33 and 34.)

[Tran 2009] C Tran, S Ouk, N J Clegg, Y Chen, P A Watson, V Arora, J Wongvipat, P M Smith-Jones, D Yoo, A Kwon, T Wasielewska, D Welsbie, C D Chen, C S Higano, T M Beer, D T Hung, H I Scher, M E Jung and C L Sawyers. *Development of a Second-Generation Antiandrogen for Treatment of Advanced Prostate Cancer*. Science, vol. 324, no. 5928, pages 787–790, May 2009. (Cited on page 31.)

[van de Wijngaart 2011] Dennis J van de Wijngaart, Hendrikus Jan Dubbink, Martin E van Royen, Jan Trapman and Guido Jenster. *Androgen receptor coregulators: Recruitment via the coactivator binding groove*. Molecular and Cellular Endocrinology, pages 1–13, September 2011. (Cited on page 26.)

[van Royen 2012] Martin E van Royen, Wiggert A van Cappellen, Carola de Vos, Adriaan B Houtsmuller and Jan Trapman. *Stepwise androgen receptor dimerization.* J. Cell. Sci., vol. 125, no. Pt 8, pages 1970–1979, April 2012. (Cited on page 26.)

[Verhoeven 2010] Guido Verhoeven, Ariane Willems, Evi Denolet, Johannes V Swinnen and Karel De Gendt. *Androgens and spermatogenesis: lessons from transgenic mouse models.* Philos. Trans. R. Soc. Lond., B, Biol. Sci., vol. 365, no. 1546, pages 1537–1556, May 2010. (Cited on page 32.)

[Zhou 1994] Z X Zhou, C I Wong, M Sar and E M Wilson. *The androgen receptor: an overview.* Recent Prog. Horm. Res., vol. 49, pages 249–274, 1994. (Cited on page 27.)

Chapter 3

# Biophysical methods to characterize different peptide regions of an IDP



The biophysical study and structural characterization of an IDP can be addressed trough different strategies. The project can be raised to study the entire protein complex or domain, or to study different small peptides constructs from the protein sequence. The choice on these different strategies depends on the behavior of the protein in solution, the availability of the different techniques and our expertise, or the objectives we want to achieve.

Using as an example the N-terminal domain of the Androgen Receptor the same two strategies can be adopted. As explained in Chapter 2, the NTD-AR is a large non-structured domain with 550 residues implying the difficulty of the production and its low stability in solution. A good alternative to characterize this domain is to study it by choosing different small peptide regions which could be easier to manipulate and to stabilize.

The flexible nature of IDPs precludes their study by crystallographic methods. For this reason, the best strategy to achieve this study is to characterize the object in solution. A wide range of biophysical techniques can be used, as fluorescence spectroscopy, circular dichroism (CD), mass spectrometry (MS), small angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR). Although the electron microscopy (EM) is not considered a solution technique, it preserves the specimen in a snapshot of its solution state. For that, it can be also used for the study of an IDP.

The biophysical methods addressed in this thesis, for the peptide study of the NTD-AR, are explained below.

## 3.1 Fluorescence spectroscopy

Fluorescence application in protein and peptide analysis is mainly a probe sensing changes in the fuorophore local environment. Fluorescence is a popular spectrosocopy technique due to its sensitivity and simplicity of data acquisition. The tryptophan residue is used as a intrinsic fluorescent probe for the study of protein folding. And also, for the interaction studies, it can be used to estimate the nature of microenvironment changes up to binding. Fluorescence is a dynamic phenomenon and the variation of the lifetime of the excited state is sufficient to detect a variety of chemical reaction [Ladokhin 2006]. The anisotropy decay is also sensitive to the rotational diffusion of the molecule, and its variation is a sensor to molecular interactions. The characteristic time of fluorescence ranges in the nanosecond and is a good complement to other spectroscopic techniques sensitive to dynamics, such as NMR. Proteins that lack aromatic residues can be coupled to fluorophore to performed these experiments, or an specific dye can be also used. As example the Thioflavine T for the characterization of amyloid fibers [Nilsson 2004].

## 3.2 Circular Dichroism (CD)

Circular dichroism (CD) spectroscopy measures the difference in the absorption of left- and right- circularly polarized light. This occurs at wavelengths of light that can be absorbed by a chiral molecule. At these wavelengths left-and right-circularly polarized light will be absorbed to different extents due to the molecule chirality. The CD spectrum is obtained when the dichroism is measured as a function of wavelength. The units of CD results are reported in degrees of ellipticity ($\theta$, mdeg), molar ellipticity (circular dichroism corrected for concentration).

CD spectroscopy has a wide range of applications in many different fields. A primary use of UV CD is to investigate the secondary structure of proteins [Brahms 1980]. This application can be extended to study molecular interactions or changes due to environmental conditions. Reaction kinetics and thermodynamic information can be also obtained [Kelly 2005].

The far-UV CD (180-250 nm) spectrum is used for the study of the secondary structure of proteins (Figure 3.1). In this aspect, CD gives the information to estimate the fraction of a molecule that is in the alpha-helix conformation, the beta-sheet conformation, the beta-turn conformation, or random coil. The near-UV CD spectrum of proteins provides information

on the tertiary structure [SYu 1994]. The signals obtained in the 250–300 nm region are due to the absorption, dipole orientation and the nature of the surrounding environment of the phenylalanine, tyrosine, cysteine (or S-S disulfide bridges) and tryptophan amino acids.

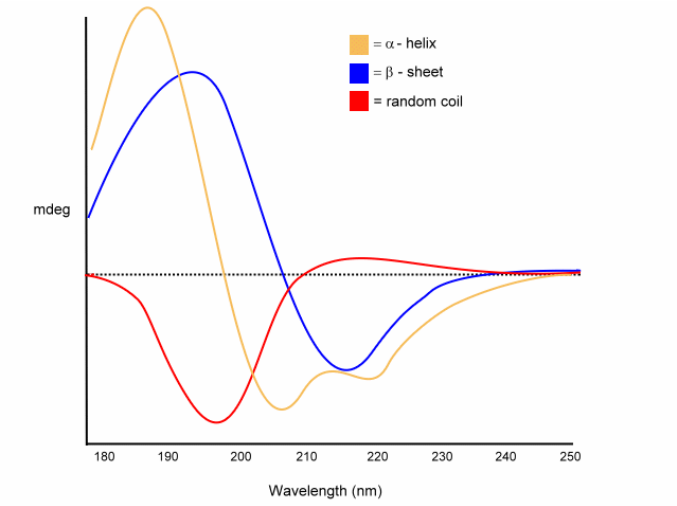



Figure 3.1: Far UV CD spectra of different secondary structure patterns. $\alpha$-helix (yellow), $\beta$-sheet (blue); and random coil (red).

## 3.3 Mass Spectrometry (MS)

Mass spectrometry (MS) is an important technique in proteomics and systems biology. MS is an analytical technique used to identify the mass of the species present in solution, by measuring the mass/charge ratio (m/z). It allows high-throughput analysis and requires small amounts of material (nM–fM range). The study of biological samples by MS has been revolutionized by the ESI ionization technique which allows to transfer to gas phase biological molecules with virtually no limit in the molecular size. As proteins and nucleic acids are naturally ionized, it is straightforward to measure MS spectra of complete proteins or of their fragments. This open the possibility to directly characterize the proteins present in the solution. The precise determination of the mass allows the determination of the present species, and to determine their eventual covalent modification (post-translational modification, adduct formation, etc).

Modern applications of high-resolution mass spectrometry to the metabolism of drug or to study complex aspects of metabolism. Or quantitative analysis of proteins on proteomics [Guengerich 2011].

## 3.4 Electron Microscopy (EM)

The use of electron microscopy (EM) is appropriated for the study of large molecular assemblies. The best approach to study biological samples is cryo-microscopy, however for a rapid, and less precise, identification of the object size and shape, the transmission electron microscopy (TEM) can be a good option. TEM is a microscope technique in which a beam of electrons is transmitted through an ultra-thin specimen, and interacting with it while is passing through Uranium staining. As a result, an image is formed and focused onto a fluorescence screen.

On the other hand, if a three-dimensional study of the large molecular assemblies is desired, the use of cryo-EM is needed. The method detects projections of a frozen hydrated specimen. For a 3D reconstruction of the particle shape, a large number of such projections from particles in all orientations are collected, averaged and analyzed [Göbl 2014]. High structural homogeneity and integrity is required for high resolution reconstructions. EM requires less material and tolerates more impurities than SAXS or NMR, and is a good technique to solve molecular structures of large molecules in combination with these methods.

## 3.5 Small Angle X-ray Scattering (SAXS)

Small Angle X-ray Scattering (SAXS), is a fundamental tool for the study of macromolecules in solution. SAXS is a good complementary technique to high resolution techniques such as nuclear magnetic resonance or X-ray crystallography. Furthermore it can be used in dynamic systems or disordered systems which are limited in X-ray crystallography, in the sense that it requires good crystals formation. In these situations, SAXS becomes an interesting alternative.

SAXS gives information about the shape and the size of the object in solution, consequently is a useful technique to obtain structural information in disordered systems.

The sample preparation requires an homogeneous dilute solution. This is necessary in order to obtain a monodisperse system of particles in solution with a random orientation. As a result, a spherical averaging of the particle scattering will be obtained.

The experiment consists in shining an X-ray beam through the solution sample, and to measure the X-rays scattered at small angles by the scattering particles present in the solution. The scattered X-rays form a pattern which is then detected at the detector. The scattering pattern contains the information on the structure of the sample. This equation summarize the information we can obtain with a SAXS experiments.

$$I(q) = c \int_0^\infty n(r)[f(qr)]^2 S(qr) dt$$

scale factor | size distribution | structure factor

Where the scattering vector is defined by $q = \frac{4\pi \sin(\theta)}{\lambda}$.

The scattering curve can be analyzed at several levels, depending on the domain of the curve which is considered (Figure 3.2):

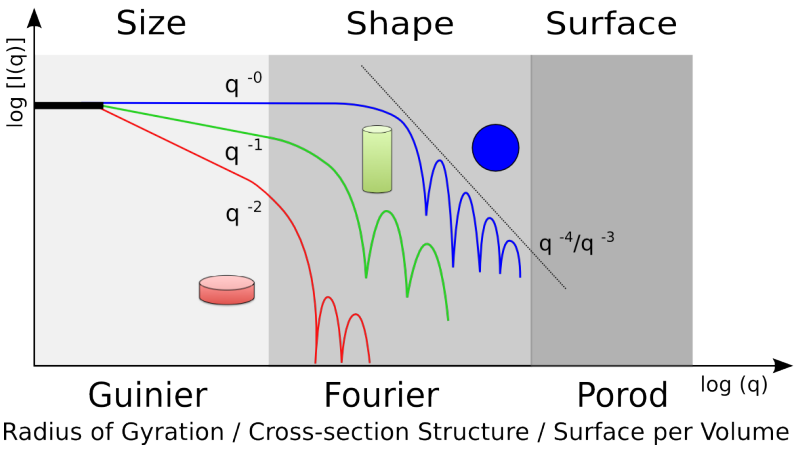



Figure 3.2: In a double logarithmic plot an initial slope of 0, -1 or -2 indicates globular, cylindrical or lamellar shape, respectively. If the slope is steeper than that then the particles are larger than the resolution limit and the Porod region is the only part of the form factor that can be observed.

For ideal monodisperse systems, the Guinier plot ($ln(I(q))$ versus $q^2$) should be a linear function, whose intercept gives $I_0$ and the slope yields the radius of gyration $Rg$ of the scattering particles (see Figure 3.2 and 3.3). The Kratky plot, which analyses the large q domain, is useful to rapidly identify if the system is folded or unfolded (see Figure 3.3). And finally in the same Figure 3.3 is shown how the Pair distance distribution function provides information about the shape of the object.

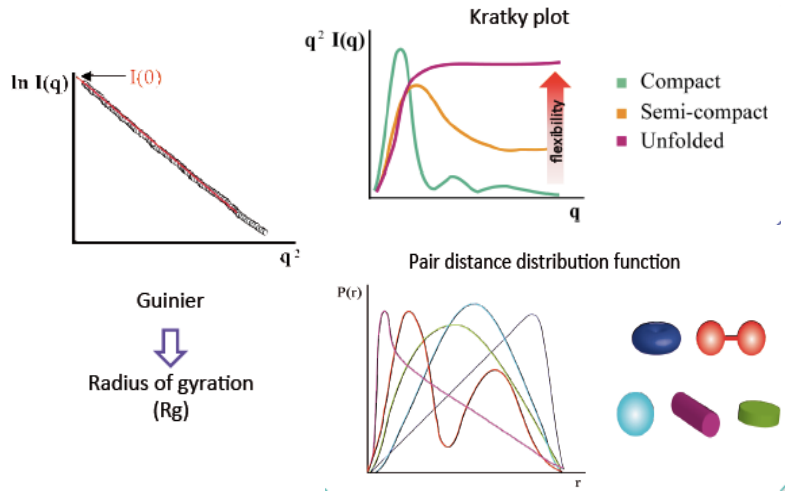



Figure 3.3: Different information domains of a particle size and shape from a SAXS plot. Figure from [Svergun 2003].

## 3.6 Nuclear Magnetic Resonance (NMR)

The origins of the Nuclear Magnetic Resonance (NMR) were formulated in the late 1930s and early 1940, but it was only in the 1980s when NMR starts to be used to obtain information about the structure and dynamics of proteins, and also nucleic acids [Wüthrich 1982] and its use for three-dimensional triple-resonance experiments with $^{13}C$ and $^{15}N$ enriched proteins [Kay 1990].

NMR involves the quantum mechanical properties of the spin of the central nucleus of the atom. These properties depend of the molecular environment at the local level, and their measurement provides a map of the chemical bond between the atoms, how close they are in space and their movement rate with respect to each other.

Most biological samples are studied in water by solution state NMR, but solid state NMR can be also used with solid samples.

The most common nuclei studied by NMR are those with S=1/2 (where S is the spin quantum number). These nuclei include the naturally occurring isotopes $^{1}H$ and $^{31}P$ found in high-abundance (close to 100%) in proteins and nucleic acids; and others such as the low-abundance $^{13}C$ (1.1%) and $^{15}N$ (0.4%), requiring isotopic enrichment through biochemical techniques for optimal sensitivity [Greenbaum 2001]. The engineering development for the isotopic labelling of the proteins and other macromolecules, helped to the development and improvement of the multidimensional NMR experiments in order to achieve the atomic assignment of the signals for further structural calculations. Currently the protein complexes can be

studied routinely up to a molecular weight of about 40 kDa. Recent developments in hardware and pulse programs have extended the study of proteins with a molecular weight around 100 kDa [Tugarinov 2004].

The nuclear magnetic resonance is also used for the description of the dynamics of biological macromolecules. Relaxation measures can be also used to measure molecular motion to help to identify potentially more flexible regions of importance to the function of the protein. Currently it is the second most used method in the determination of high-resolution structure after the X-ray diffraction but with the advantage that by NMR the macromolecule study is dynamic and you can observe the transient different conformations adopted.

### 3.6.1 Peptide structural characterization

Coming back to the strategy to study short peptides regions of the NTD-AR, a set of different NMR experiments is needed for their characterization.

#### $^1$H 1D experiment

All NMR characterization starts performing a $^1$H 1D experiment (Figure 3.4). Where RF signal is transmitted during few $\mu s$ (90° pulse). In order to move the magnetization vector from the z-axis to the x,y-plane. The signal that evolves due to precession of the magnetization vector is measured after the pulse. At the end of the process, the vector returns to equilibrium on the z-axis. The process is called free induction decay (FID) of the magnetization. The Fourier transform is performed to the recorded FID, in order to obtain the NMR spectrum.

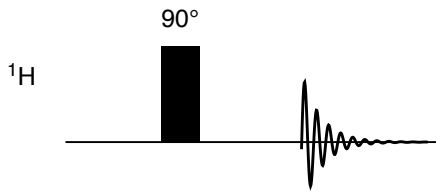



Figure 3.4: NMR sequence of the $^1$H 1D.

This simple experiment gives rapidly, the information of the nature of the sample. Figure 3.5 shows the spectral range for the proton chemical shift ($\delta$), which helps to identify the different amino acids that form the peptide sequence and to detect the presence of possible impurities.

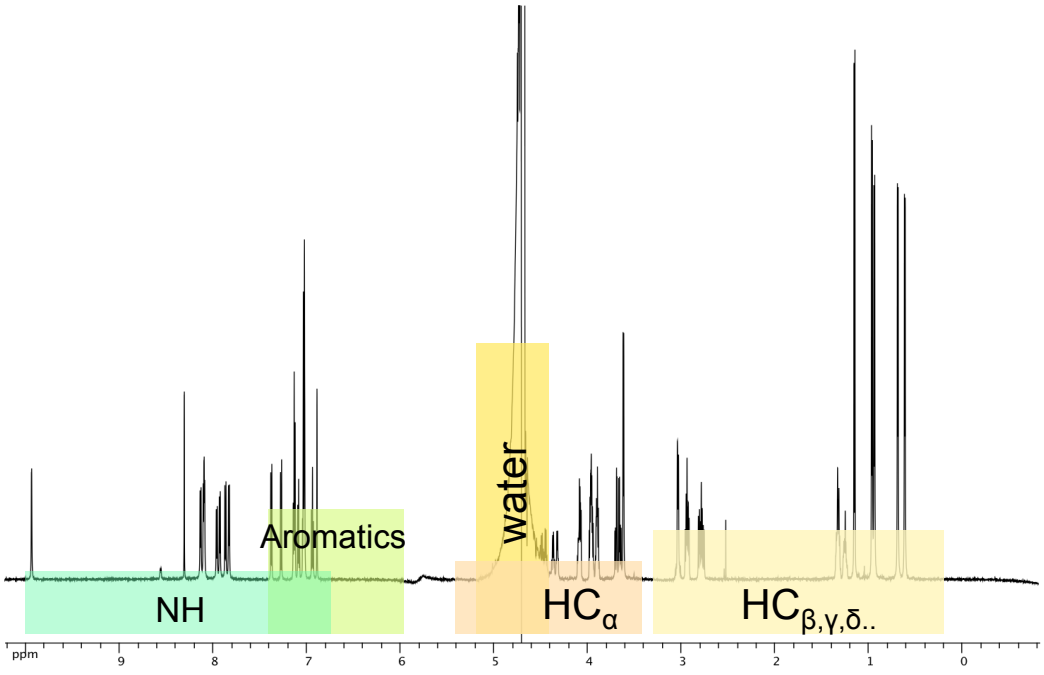



Figure 3.5: $^{1}$H 1D spectrum of a peptide. The characteristics $\delta$ for the different H are highlighted in the plot spectrum.

### $^{1}$H-$^{1}$H 2D experiments

To succeed in the peptide assignment it is necessary to identify the amino acid side chains and the neighboring residues in the peptide sequence. The combination of both give the primary structure assignment of the amino acid sequence. The contacts with the residue neighbors allows to also obtain information about the secondary structure of the residues chain. This objective can be achieved performing TOCSY and ROESY experiments.

### TOCSY (TOtal Correlation SpectroscopY)

This method is used to observe the effect of scalar couplings, which appears between $^{1}$H spin located at a distance of 2 to 3 chemical bounds. TOCSY is used to assign each amino acid individually. Figure 3.6 shows the NMR sequence for the TOCSY experiment. Firstly, the 90° pulse rotates the magnetization from z to the xOy plane, where they evolve during $t_1$. After that, a spin lock on the x axis is applied of a duration time $t_m$ (mixing time), and then the spectrum is acquired.

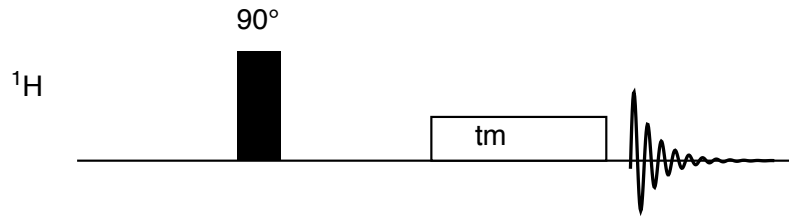



Figure 3.6: NMR sequence for the TOCSY experiment.

### The NOE effect

An important consequence of dipolar-dipolar relaxation is the NOE effect. The Nuclear Overhauser Enhancement (NOE), results from the cross relaxation between two neighboring nucleus within the same molecule or between two nuclei belonging to distinct molecules within a complex (< 6 Å). The irradiation of a spin population of a nuclei A, will cause a change on the population of its neighbor nuclei B (Figure 3.7 A). This occurs by dipole-dipole interaction contrary to J-coupling, which occurs through-bond connection. The NOE is defined as the relative change of intensity after a given irradiation time and the sign of the magnetization transfer depends on the frequency of the molecular motions [Kieffer 2011]. Molecular motion ($\tau_c$) is crucial for the NOE effect. Positive transfer rates are observed for small molecules due to the tumbling with a fast correlation time. The opposite will occur for large values of $\tau_c$, corresponding to larger macromolecules (Figure 3.7 B).

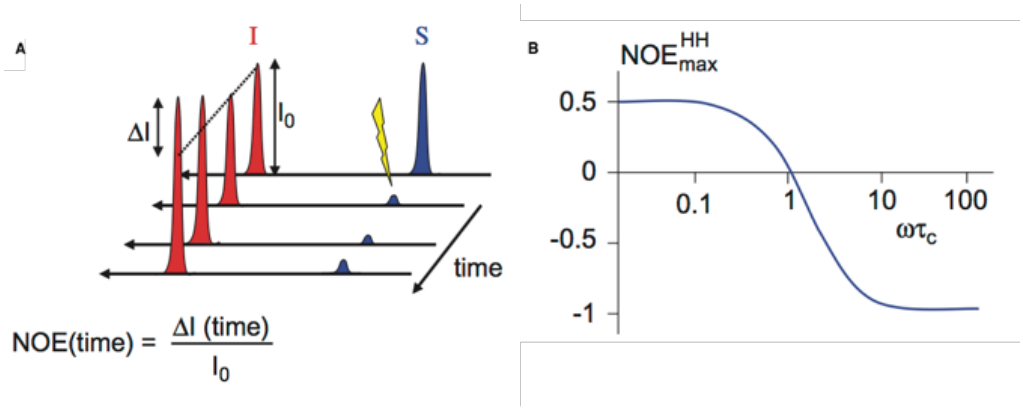



Figure 3.7: A) Nuclear Overhauser Enhancement (NOE) B) Graphical representation of the NOE as function of $\tau_c$. Figure from [Kieffer 2011]

When working with peptides a problem arises. NOE depends strongly on correlation time and goes through zero for $\omega\tau_c = 1$ which corresponds to medium-sized molecules ($M_W$ range

700–1200 Da for a 600-700 MHz spectrometer). This corresponds to most of the peptides studied here. However, in these situations an alternative is to use the ROE effect (Rotating frame Overhause) to detect correlation between nucleus that are spatially close. In this case, a spin-lock is employed to ensure that correlation between spins, which are close in space but have zero NOE, are seen.

### ROESY (Rotating frame Overhauser Effect SpectroscopY)

The aim of this experiment is to observe through space dipole-dipole coupling. The result of a ROESY experiment is a 2D spectrum in which the out off diagonal signals arise from the Overhauser enhancement effect between space coupled protons. ROESY is useful for determining which signals arise from protons that are close to each other in space even if they are distant in the molecular topology, and is used to to identify the position of the amino acids within the sequence.

The sequence is equivalent to the one in Figure 3.6 but the difference resides in the strength/efficiency of the field during $t_m$. In TOCSY, an efficient decoupling of the spins is desired, for this reason, a strong RF field ( 10kHz) and an efficient composite decoupling scheme (MLEV, DIPSI, etc...) are applied. In ROESY, the magnetization is locked in the plane, and the RF field not too strong to avoid the observation of the TOCSY effect.

The correlation peaks obtained from the ROE effect show an opposite sign compared to the diagonal. On the other had, the correlation peaks coming from the chemical exchange present the same sign than the diagonal, allowing to distinguish easily between dipolar interactions and chemical exchange effects.

Figure 3.8 outlines the information that can be obtained with either a TOCSY or a ROESY experiment. In a TOCSY, all spins are seen, for each individual residue. The ROESY will have both intra-residue correlations, as well as inter-residue correlations, which allows us to find which residue is next to which in the peptide chain.

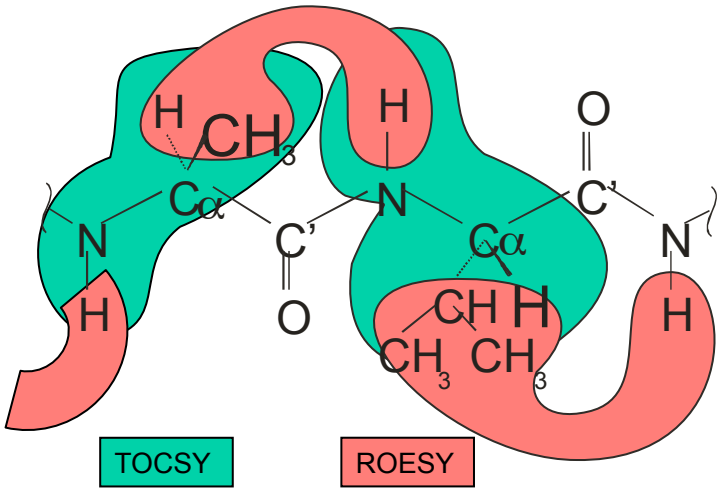



Figure 3.8: Difference between the information obtained with TOCSY and ROESY experiments.

# Bibliography

[Brahms 1980] S Brahms and J Brahms. *Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism.* Journal of Molecular Biology, vol. 138, no. 2, pages 149–178, April 1980. (Cited on page 44.)

[Göbl 2014] Christoph Göbl, Tobias Madl, Bernd Simon and Michael Sattler. *NMR approaches for structural analysis of multidomain proteins and complexes in solution.* Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 80, pages 26–63, July 2014. (Cited on page 46.)

[Greenbaum 2001] Nancy Greenbaum and Ranajeet Ghose. Nuclear Magnetic Resonance (NMR) Spectroscopy: Structure Determination of Proteins and Nucleic Acids. John Wiley & Sons, Ltd, Chichester, UK, May 2001. (Cited on page 48.)

[Guengerich 2011] F P Guengerich. *Thematic Minireview Series on Biological Applications of Mass Spectrometry*. Journal of Biological Chemistry, vol. 286, no. 29, pages 25417–25417, July 2011. (Cited on page 45.)

[Kay 1990] Lewis E Kay, Mitsuhiko Ikura, Rolf Tschudin and Ad Bax. *Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins*. Journal of Magnetic Resonance (1969), vol. 89, no. 3, pages 496–514, October 1990. (Cited on page 48.)

[Kelly 2005] Sharon M Kelly, Thomas J Jess and Nicholas C Price. *How to study proteins by circular dichroism.* Biochim. Biophys. Acta, Proteins Proteomics, vol. 1751, no. 2, pages 119–139, 2005. (Cited on pages 44 and 94.)

[Kieffer 2011] B Kieffer, 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, 2011. (Cited on pages 51 and 118.)

[Ladokhin 2006] Alexey S Ladokhin. Fluorescence Spectroscopy in Peptide and Protein Analysis. Applications, Theory and Instrumentation. John Wiley & Sons, Ltd, Chichester, UK, September 2006. (Cited on page 44.)

[Nilsson 2004] Melanie R Nilsson. *Techniques to study amyloid fibril formation in vitro.* Methods, vol. 34, no. 1, pages 151–160, September 2004. (Cited on pages 44, 93 and 101.)

[Svergun 2003] D I Svergun and MHJ Koch. *Small-angle scattering studies of biological macromolecules in solution*. Reports on Progress in Physics, 2003. (Cited on page 48.)

[SYu 1994] Venyaminov SYu and K S Vassilenko. *Determination of protein tertiary structure class from circular dichroism spectra.* Anal. Biochem., vol. 222, no. 1, pages 176–184, October 1994. (Cited on page 45.)

[Tugarinov 2004] Vitali Tugarinov, Peter M Hwang and Lewis E Kay. *Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins.* Annu. Rev. Biochem., vol. 73, pages 107–146, 2004. (Cited on page 49.)

[Wüthrich 1982] K Wüthrich, G Wider, G Wagner and W Braun. *Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance.* Journal of Molecular Biology, vol. 155, no. 3, pages 311–319, March 1982. (Cited on page 48.)

Chapter 4

# Permutated DOSY (*p*-DOSY) for the study of complex systems



Nuclear magnetic resonance (NMR) spectroscopy has been extensively used in chemistry, biology and medicine for the mixtures analysis. Among all the different NMR experiment that exist, Diffusion Ordered SpectroscopY (DOSY) is in particular a widely applied technique in life sciences for complex mixtures analysis [Barjat 1998, Lucas 2002, Gilard 2011], reactive intermediates [Schlörer 2002, Giuseppone 2008] and molecular organization [Šmejkalová 2008, Cohen 2012, Nonappa 2015].

The analysis of the diffusion coefficients is a powerful tool to gain insight on molecular sizes and shapes [Wilkins 1999, Gomez 2009, Floquet 2009], and can also be used to investigate molecular interactions [Lucas 2004, Cohen 2012, Nonappa 2015], polymer polydispersity [Stchedroff 2004, Viéville 2011], or to characterize reactive intermediates [Schlörer 2002, Li 2009, Nguyen 2009, García-Álvarez 2011].

A spectrum of the different components in the mixture can be separated based on the apparent diffusion coefficients by measuring the NMR signal decay. Furthermore, DOSY experiments setups and methodology can be easily standardized and automated.

Translational diffusion from an atomic point of view can be defined as the result of the random walk of the diffusing particles, where the molecules are self-propelled by thermal energy. The random walk of small molecules in suspension in a fluid was first described by Robert Brown in 1827, and this phenomenon is known as *Brownian motion*. The diffusion of particles observed in this phenomenon can be measured as a translational coefficient or as hydrodynamic radius $r_H$, which are inversely proportional, as described by the Stokes-Einstein eq. 4.1 for spherical shaped molecules.

$$D = \frac{k_B T}{6\pi\eta r_H} \tag{4.1}$$

where $k_B$ is the Boltzmann's constant, $T$ is the absolute temperature, and $\eta$ is the viscosity. Other shapes can be handled either with extended analytical models [Evans 2013] or by a statistical description [Augé 2009, Hernandez Santiago 2015].

## 4.1 Diffusion Ordered SpectroscopY (DOSY)

NMR can be used to measure the diffusion coefficient ($D$) of molecules in solution by using special sequences that use pulsed-field gradients (PFGs) along the sample axis (z). In order to understand how the experiment works, we should start by defining the Larmor frequency $\omega$ (radians $s^{-1}$), $\omega = \gamma B_0$, where $\gamma$ is the gyromagnetic ratio of the nucleus and $B_0$k is the magnetic field. When a PFG is applied along z ($G(z)$), $\omega$ becomes spatially dependent.

$$\omega = \gamma(B_0 + G(z)z) \tag{4.2}$$

This provides the basis for further measurements of $D$. Considering the pulse sequence proposed by Stejskal and Tanner in 1965 [Stejskal 1965] represented in Figure 4.1 a). A series of spin echoes with a increasing gradient field strength is measured. The mean molecule translational diffusion resulting from the Brownian motion, is acquired at the end of the second delay. A pulse gradient of duration $\delta$ and magnitude $G$ is applied between the first 90° and the 180° pulses, it causes an additional dephasing of the magnetization that is now proportional to $\delta$, $G$, and dependent on the $z$ spatial position of each spin. Then the magnetization is inverted by the 180° pulse at time $\tau$, and a second pulse gradient, identical to the first one, is applied. If we have an hypothetic static system, that does not display diffusion, a complete

refocusing of the magnetization is obtained. Nevertheless, in a real sample, spins move during the delay between the two gradient pulses ($\Delta$), as consequence of the self-diffusion. The molecular movement along the helix will scramble the magnetization if the mean path is larger than the helix pitch. This situation leads to an incomplete refocusing, and an attenuation of the signals is observed depending on the diffusion rate of the molecules. It can be described by an exponential law (eq. 4.3 for STE sequence - Figure 4.1 b):

$$I = I_0 e^{-D\gamma^2 g^2 \delta^2 (\Delta - \frac{\delta}{3})} \tag{4.3}$$

Where $D$ is the diffusion coefficient, $\gamma$ the gyromagnetic ratio, and $g$ and $\delta$ the intensity and the duration of the PFG, respectively. $I_o$ is the signal intensity obtained in the absence of field gradient and $I$ in the presence of a gradient of strength $g$; $\Delta$ is the effective diffusion delay defined as the time between the start of the bipolar gradient pairs; see [Sinnaeve 2012] for details.

From which $D$ can be obtained after the acquisition of a series of free induction decays (FID) varying the gradient strength.

The result of a DOSY experiment is a bidimensional NMR spectrum with the diffusion coefficient $D$ as vertical axis. This representation was proposed by Morris and Johnson [Morris 1992] in 1992, and it has been reported in the literature by different reviews [Johnson 1999, Antalek 2002]. The main features of a DOSY experiment are, that it is an NMR technique not based on the time dependence of the signal, only the intensities of coding and decoding gradients are varying and all delays are kept constant. It is also remarkable that in contrary to most NMR experiments that use the Fourier transform for spectral analysis, the DOSY experiment is using the Laplace transform because the signal variations follows an exponential law.

The different NMR experiments used in DOSY acquisition are reported in Figure 4.1.

### 4.1.1 Spin-echo (SE)

The original spin-echo (SE) experiment Figure 4.1a) is the simplest pulse sequence that can be used for diffusion measurement and it is based on the analysis of the decay of the spin-echo signal intensity with increasing PFG intensities. A 90° pulse is applied and the magnetization is aligned on the xOy plane perpendicular to the applied static magnetic field. This static field is then momentarily perturbed by the first PFG with a length of $\delta$. When the PFG is applied, the field intensity varies linearly along the main axis of the sample. As a result of this, a dephasing of the bulk NMR signal is introduced due to a spatial phase encoding which depends on the spin position along the z-axis. A magnetization helix is observed at the end of

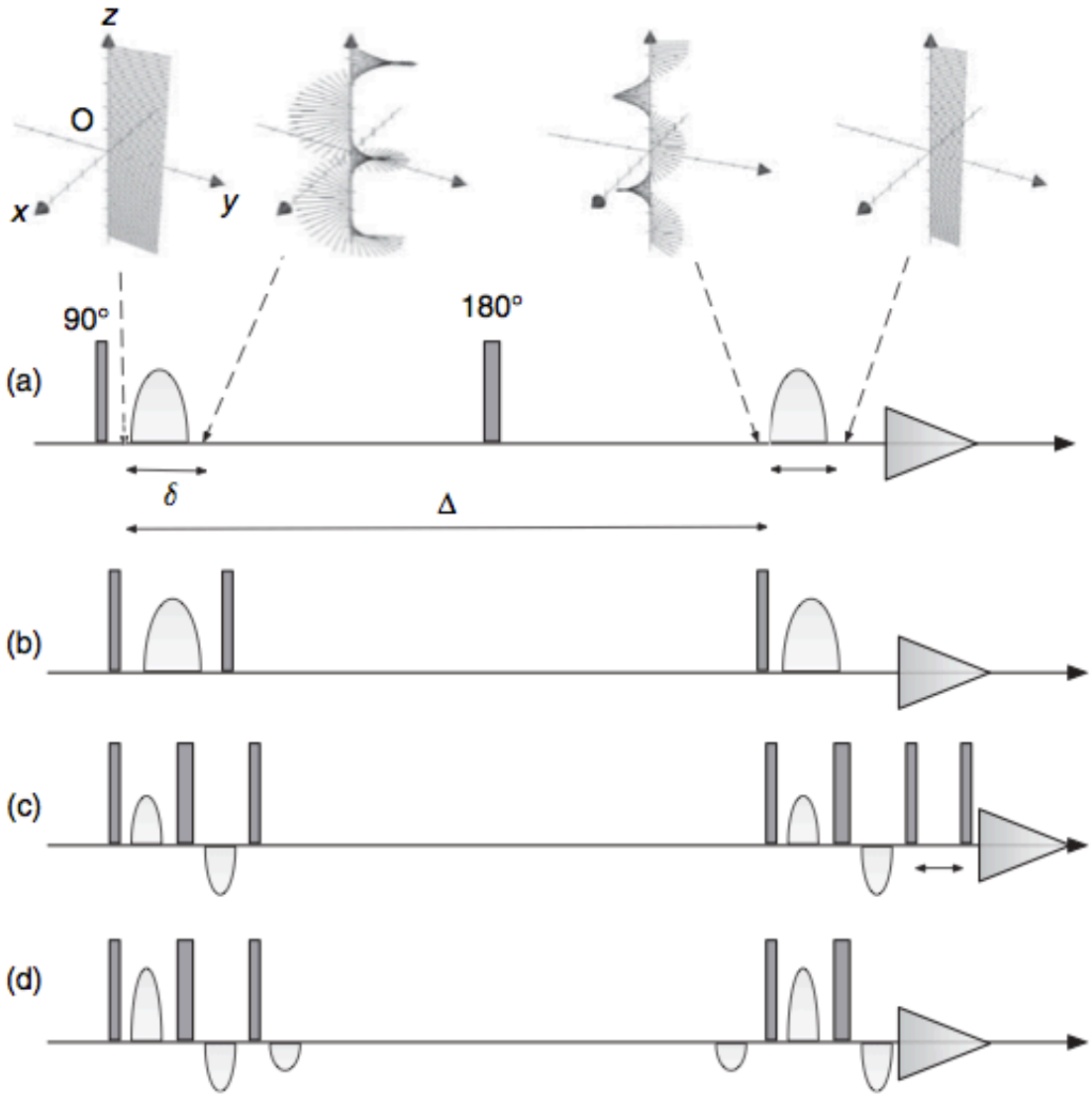



Figure 4.1: DOSY NMR sequences, figure used from [Gilard 2011] publication. a) The original Spin Echo (SE) sequence b) The Stimulated Echo pulse sequence (STE) c) The bipolar pulse pair-stimulated echo-longitudinal eddy current delay (BPP-STE-LED) d) The OneShot experiment.

the PFG. The 180° in the middle of the experiment inverse the magnetization and it creates a Hahn echo that removes any contribution of the chemical shift to the evolution. The final PFG applies an opposite dephasing to the magnetization, allowing the signal restore to its initial state. However, the diffusion takes place during the delay between the two PFG. If the mean molecular displacement is on the order of the pitch of the magnetization helix, this helix is blurred and the intensity of the signal is decreased.

### 4.1.2 Stimulated Echo (STE)

The Stimulated Echo pulse sequence (STE) experiment Figure 4.1b) differs from the SE experiment by the the modification of the 180° pulse into two 90° pulses. Here, the second 90° pulse is applied directly after the coding gradient and it builds a spatial modulation of the longitudinal magnetization along the sample z-axis rather than a transverse modulation as in SE. Being on the z-axis, this modulation does not evolve any further, except for diffusion and $T_1$ decay. The final 90° pulse restores along the y-axis the magnetization, which then produces a stimulated echo once refocused by the second PFG. The advantages of the STE sequence are that spins relaxation depends mainly on $T_1$ instead of $T_2$, on the other hand the sensitivity is reduce by a factor of 2.

### 4.1.3 Bipolar pulse Pair-Stimulated Echo-Longitudinal Eddy current delay (BPP-STE-LED)

The bipolar pulse pair-stimulated echo-longitudinal eddy current delay (BPP-STE-LED) experiment (Figure 4.1 c) is one of the most common DOSY experiment applied nowadays. Here, the STE experiment is further optimized with different improvements. The PFG are made bipolar, that means that they are applied in two opposite pulses, sandwiching a 180° pulse. Also, an extra final delay (LED) is added during which the magnetization is stored along the z-axis to let all the generated fluctuations die away before the acquisition. These two optimizations have the same purpose that is to reduce the intensity of the eddy current generated by the PFG and to minimize its impact on the observed signal.

### 4.1.4 OneShot experiment

The OneShot experiment (Figure 4.1 d) brings further improvements to the standard STE pattern. The additional selection gradients allow the reduction of the phase program. In consequence, it presents a better stability against the experimental fluctuations and allows a

simplification of the measure. Less signal averaging is required permitting a faster measure when the S/N ratio is not a limiting factor.

Up to this point, the classical NMR sequences for DOSY experiments have been summarized, nevertheless new experiments have been developed to improve resolution. One of the main problems of a DOSY experiment is to extract the correct diffusion coefficient when signals overlap, and this commonly occurs in the study of complex mixtures.

Spectral resolution can be considerably improved by using 3D DOSY experiments, where diffusion weighting is added before or after to a 2D experiment. But also by suppressing the multiplet structure in the proton spectrum to give a *pure shift* experiment.

### 4.1.5 3D DOSY

As said previously, the most common solution to the overlap problem in DOSY, is to add a diffusion dimension to a 2D experiment and for this reason, many different pulse sequences for 3D DOSY have been published [Barjat 1998, Lucas 2002, Bradley 2005]

A 3D DOSY sequence can be achieved with different strategies: appending the diffusion encoding (X-DOSY), prepending the diffusion encoding (DOSY-X) or incorporating it internally (IDOSY)

The result of this extra diffusion dimension incorporated into a classical 2D experiment is shown in Figure 4.2, where the $D$ dimension coupled with the HMQC helps the spectral elucidation.

As the DOSY-HMQC [Barjat 1998] experiment shown as example of a 3D DOSY many other sequences have been reported in the literature, as DOSY-TOCSY [Birlirakis 1996], DOSY-COSY [Wu 1996, Balayssac 2009] or the DOSY-TROSY [Didenko 2010] more recently.

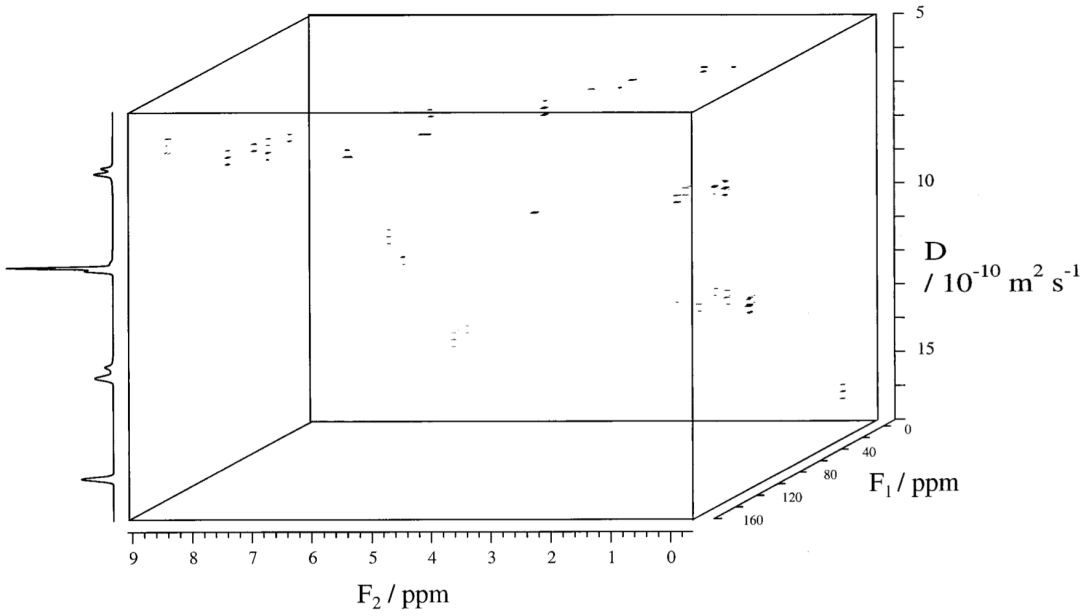



Figure 4.2: DOSY-HMQC spectrum, figure used from [Barjat 1998] showing the result of a 3D DOSY-HMQC experiment with the diffusion separation of the different components of a quinine mixture.

### 4.1.6 IDOSY

The IDOSY sequence is a 3D DOSY experiment where the diffusion encoding is added internally into the 2D sequence chosen for adding an extra diffusion dimension. The first sequence describing a iDOSY experiment was 2DJ-IDOSY by [Nilsson 2004], which improves the signal-to-noise ratio and reduces the time of acquisition compared with the previous 3D experiments described. Furthermore, with this sequence no phase cycling is needed for the diffusion encoding.

Other iDOSY experiments have been also published applying the same strategy, as the COSY-IDOSY [Nilsson 2005], DQF-COSY-IDOSY [Newman 2007] or constant-time-HSQC-IDOSY [Vitorge 2006].

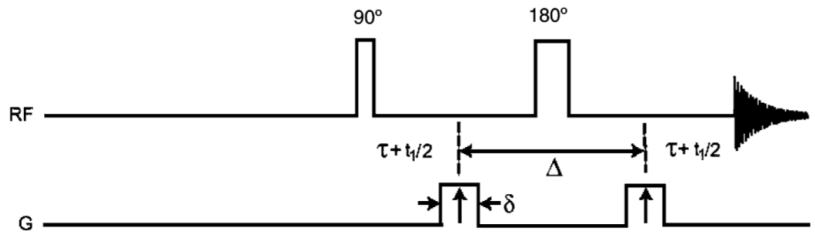



Figure 4.3: 2DJ-IDOSY sequence, figure used from [Nilsson 2004] showing radio-frequency (RF) and PFG pulses, with vertical arrows indicating the gradient levels changed to vary the diffusion encoding; a delay, $\tau$, allowed for gradient stabilization after each gradient pulse.

### 4.1.7 *Pure shift* proton DOSY

*Pure shift* proton DOSY [Nilsson 2007] was intended to be an alternative to increasing spectral dimension, 3D DOSY experiments, and to use the existing ones but simplified. The principle of this simplification is the suppression of multiplet structure in the proton spectrum in order to obtain a $^{1}$H-homodecoupled experiment, then the problem with signal overlap in DOSY is significantly diminished.

The sequence used in the *pure shift* DOSY in shown in Figure 4.4. They show that *pure shift*-DOSY is obtained by the Zangger-Sterk method [Zangger 1997], replacing the original soft 270° pulse with a corresponding 180° pulse.

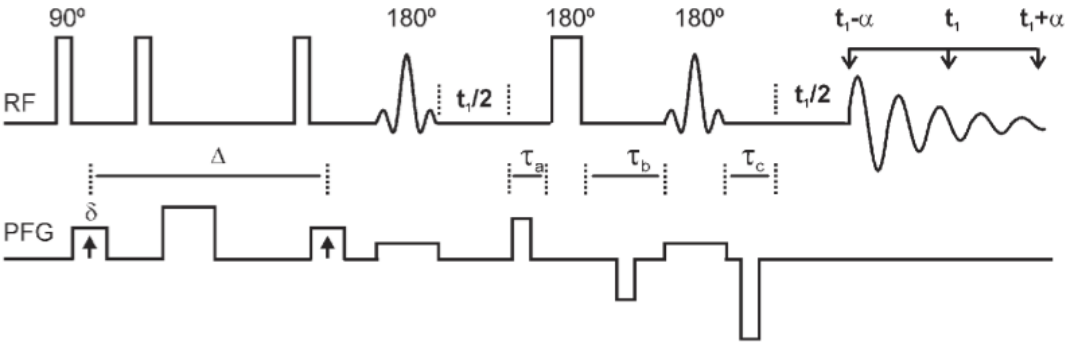



Figure 4.4: *Pure shift* proton DOSY sequence, figure used from [Nilsson 2007] showing radio-frequency (RF) and PFG pulses.

This method achieves a good resolution for 1H DOSY comparable to the one obtained using 3D DOSY methods such as DOSY-HMQC, but without the need to involve low abundance heteronuclei. However, the pure-shift advantage comes to the price of a much lowered sensitivity method.

## 4.2 DOSY on evolving systems

This chapter presents the special case of measuring DOSY experiment on an evolving system. In this case, a difficulty arises. According to the classical Stejskal-Tanner equation, the intensity at each gradient strength depends on the intensity $I_o$ with no gradient applied Eq. 4.3. Usually, $I_o$ is measured at the beginning of the DOSY experiment and considered constant throughout the measurement. However, in out-of-equilibrium systems, $I_o$ is not constant anymore but varies continuously, due to evolution of the concentrations. These variations cause the intensity $I$ to depend not only on the applied gradient but also on a varying $I_o$. Consequently, throughout the DOSY measurement, the signal intensity measured for a given species, depends on the concentration evolution as well as the increasing gradient strength, and appears to decay either too rapidly or too slowly, depending on the evolution of a concentration decreasing or increasing over the time of the measurement, creating a bias in the analysis.

The kinetic rates of the out-of-equilibrium system has to be taken into account and be compared to the timeframe of the DOSY measurement. Rapid DOSY measurements monitoring systems with slow kinetic rates are not affected to a significant degree, because only minor concentration changes occur over the experimental time. On the other hand, long DOSY measurements monitoring systems with fast kinetic rates are not affected either, but fail to record dynamic information of the system such as reactive intermediates.

The effect is dramatic for systems with intermediate kinetic rates monitored by DOSY measurements performed in a time similar to the kinetic constant of the chemical systems. In this case, the DOSY spectra display an apparent shift of the peaks along the diffusion coefficient axis leading to wrong results.

Nilsson et al.[Nilsson 2009] proposed the use of trilinear multivariate analysis in order to characterize both the reaction kinetics and the diffusion coefficient, in the analysis of a set of diffusion experiments measured during the chemical reaction. In this case the concentration variations is used as a third dimension in the multiway analysis. However in this approach, the DOSY measurements are supposed to be rapid compared to the kinetic rates, and there is no compensation for the unavoidable bias in the diffusion coefficient measure. To our knowledge, the DOSY studies performed so far on out-of-equilibrium systems [Schlörer 2002, Li 2009, Nguyen 2009, García-Álvarez 2011, Nilsson 2009] have not taken into account the analytical bias caused by concentration variations.

We propose here to measure, within a single DOSY measurement, a series of stimulated echo spectra with pulse gradient strengths sampled in a permutated manner, as it was already proposed in order to reduce $F_1$ artifacts in 2D NMR spectroscopy [Bowyer 1999]. From the

sequential list of pulse field gradient strengths, a random permutation is computed, and the DOSY is measured with a new list of gradient strengths in permutated order. This protocol has been called permutated–DOSY (*p*–DOSY). To test the effect of the permutated acquisition on the accuracy of the measure, the diffusion coefficient of glucose was monitored during the equilibration from pure $\alpha$-glucose to the anomer equilibrium of $\alpha$ and $\beta$-glucose [Gurst 1991] (see Figure 4.5).

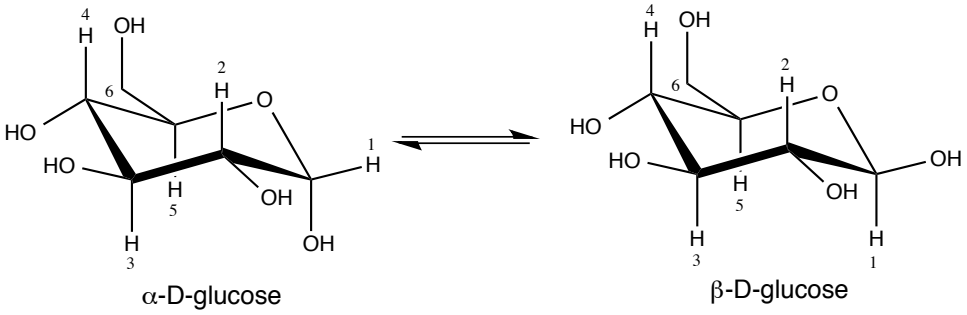



Figure 4.5: Anomerization of $\alpha$-Glucose to $\beta$-glucose with proton numbering

## 4.3 *p*–DOSY: application on evolving systems

The *p*–DOSY approach, and its capacity to restore accuracy in the determination of diffusion coefficients was tested using glucose anomerization as a test case for out-of-equilibrium systems. Glucose crystalizes in the $\alpha$ form, as a consequence, this form is dominant in the solution just after solubilization.

The $\alpha$,$\beta$-D-glucose was studied by NMR and the assignment of the molecule was done in order to be able to characterize the $\alpha$ and $\beta$ protons for further analysis. This assignment is shown in the 1D spectrum in Figure 4.6 and it corresponds to the numbering of both anomeric molecules in Figure 4.5.

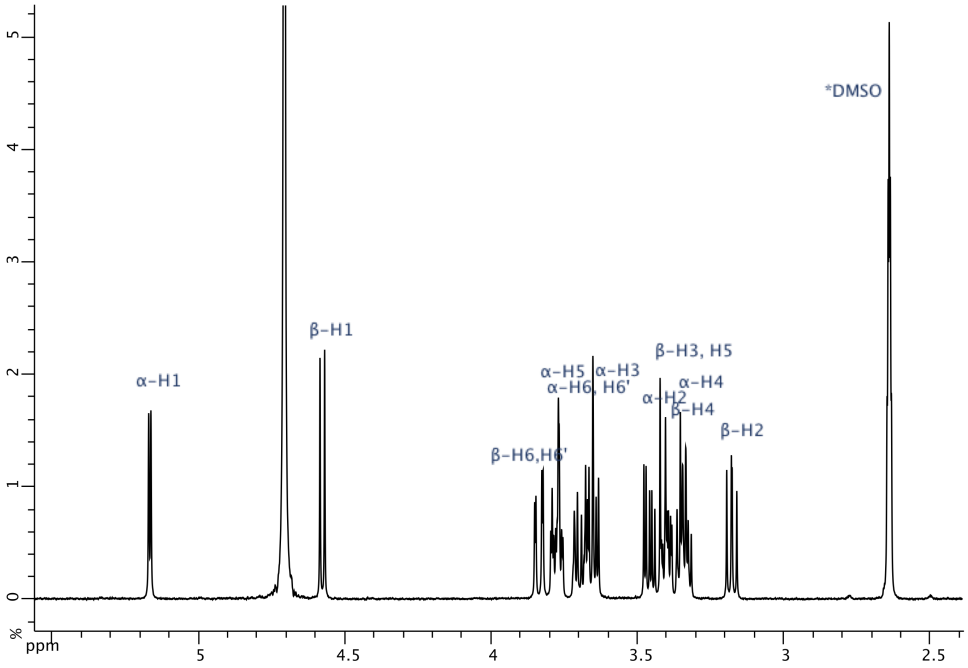



Figure 4.6: Proton assignment of a mixture of $\alpha$,$\beta$-D-glucose (42.7% $\alpha$-D-glucose and 57.4% of $\beta$-D-glucose) at equilibrium.

The anomerization of $\alpha$- to $\beta$-glucose was followed by acquiring DOSY measurements for approximately 6.0 hours, with $\alpha$-D-glucose varying from 96.6% to 44.6%. The same experiment was performed twice, the first time with the standard DOSY method, and the second time using the *p*–DOSY approach. While the anomerization was followed by DOSY measurments, a 1D kinetics Figure 4.7 was also acquired by recording a 1D $^{1}$H spectrum between each DOSY experiment. This allows to follow easily the $\alpha$ signal appearing while the $\beta$ peaks are disappearing.

The diffusion coefficients in water of both species $\alpha$ and $\beta$-glucose are expected to be very similar, as they have very similar shapes. However, they appear with quite different diffusion coefficients when measured with a standard DOSY experiment, acquired in a sequential manner, as can bee seen in figures 4.8 and 4.11. In particular, the signals from the $\beta$-glucose are found initially at a marked slower diffusion coefficient and recover the normal value over the course of the experiment. This is the result of the progressive increase of concentration of $\beta$-glucose over the duration of a single DOSY measurement, which creates a bias in the exponential analysis of the signal. At the end of the 6 hours experiment, the recovery is not complete.

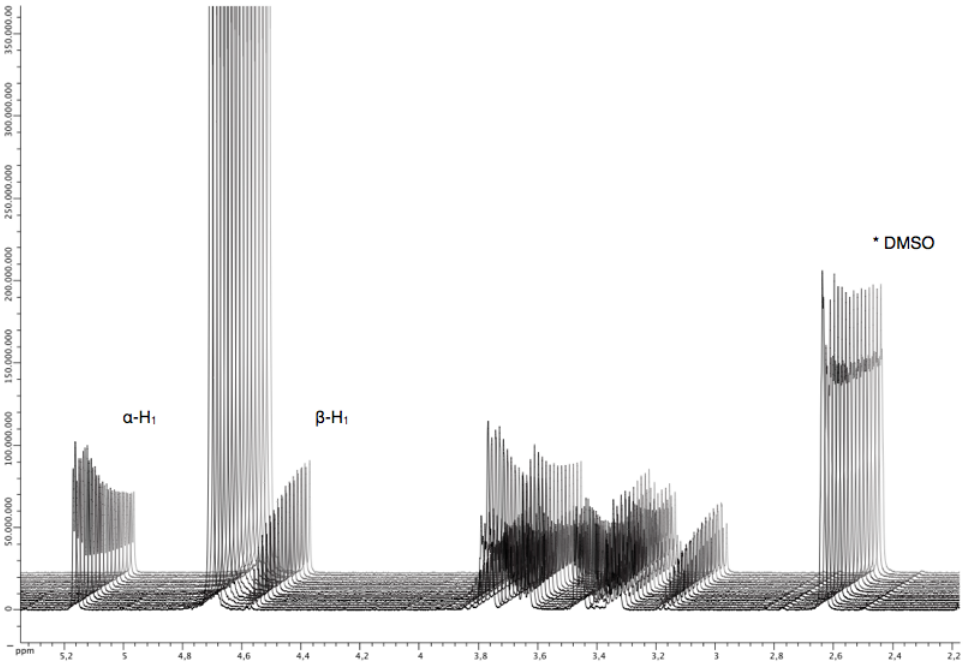



Figure 4.7: 1D $^{1}$H experiments from 0.00-6.00 hours, it means from 96.6% to 44.6% of $\alpha$-glucose.

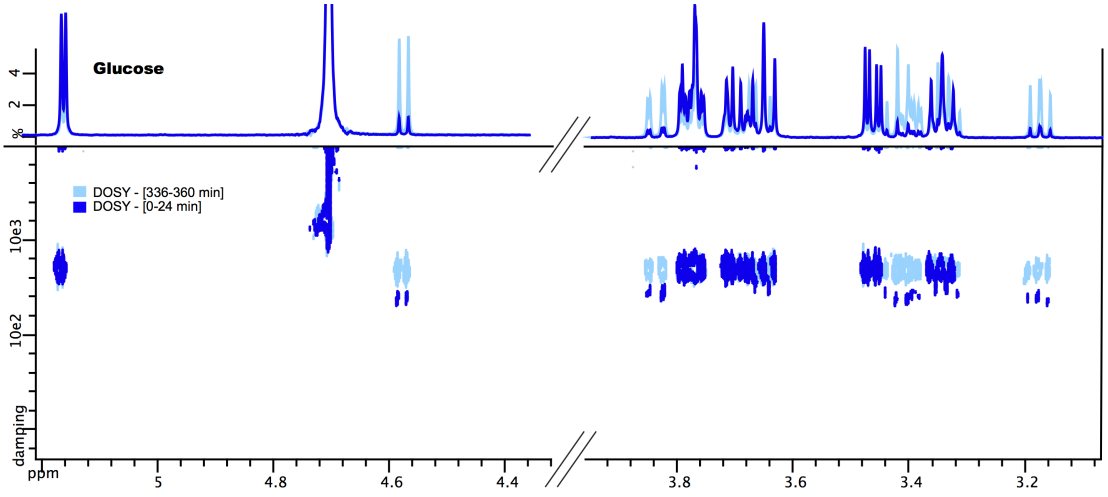



Figure 4.8: Overlay of two standard DOSY experiments, acquired between: a) out-of-equilibrium (0.00-25.00 minutes) (dark blue) and b) at equilibrium (5.41-6.05 hours) (light blue).

On the other hand, this bias is completely absent when the anomerization equilibrium is monitored using the *p*–DOSY measurements, run in the same conditions (see figure 4.9).

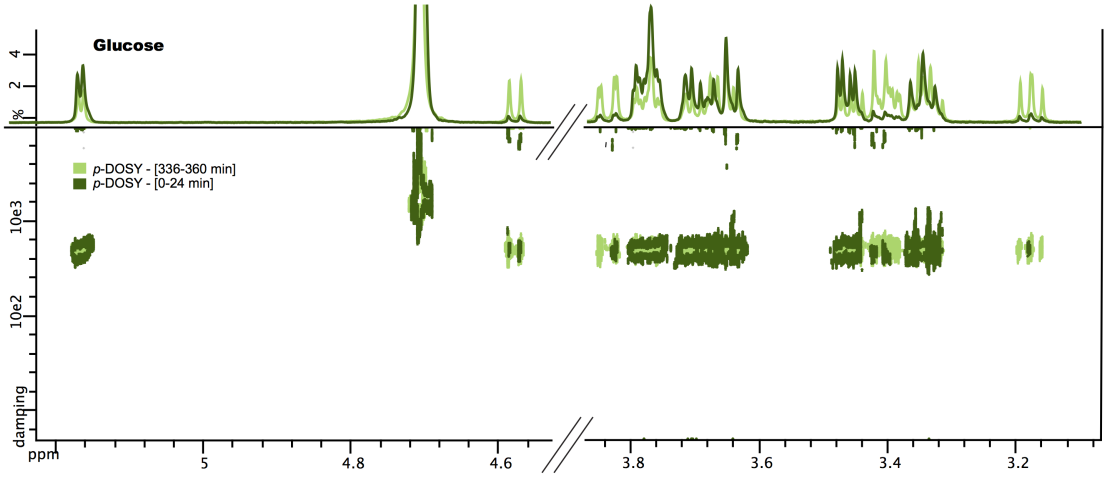



Figure 4.9: Overlay of two *p*–DOSY experiments, acquired between: a) out-of-equilibrium (0.00-25.00 minutes) (dark green) and b) at equilibrium (5.37-6.00 hours) (light green).

This improvement in the diffusion coefficient is clearly shown in Figure 4.10 where at the beginning of the $\alpha$,$\beta$-glucose anomerization the standard DOSY experiment display a bias on the diffusion value for the beta protons whereas the *p*–DOSY experiment withdraw the bias.

In both experiments, the diffusion coefficients of water and DMSO display the standard values for all experiments run during the kinetics, indicating that the error on $\beta$-glucose is not due to an experimental bias but uniquely to the variation of the concentration. Figure 4.11 presents the evolution of the DOSY signal intensities and of the apparent diffusion coefficients over the course of the experiment. It shows that the initial $\beta$-glucose concentration is very low, and that the two populations equilibrate slowly, so that 6 hours are not sufficient to reach the complete equilibrium. When measured with the standard DOSY experiment, the apparent $\beta$-glucose diffusion coefficient, varies in an anomalous manner before converging toward the correct value, while the *p*–DOSY measure presents stable measurements for both species.

The poor accuracy of the diffusion coefficient of $\beta$-glucose is attributed to miscalculation of the analysis of the reference intensity, $I_o$. The diffusion coefficient of $\beta$-glucose consequently, appears to have an incorrect low value. This effect is stronger in the first DOSY measurements since there is a higher kinetic rate. Similarly, the $\alpha$-glucose that evolves in the opposite direction, displays an apparent diffusion coefficient slightly larger than expected in the first

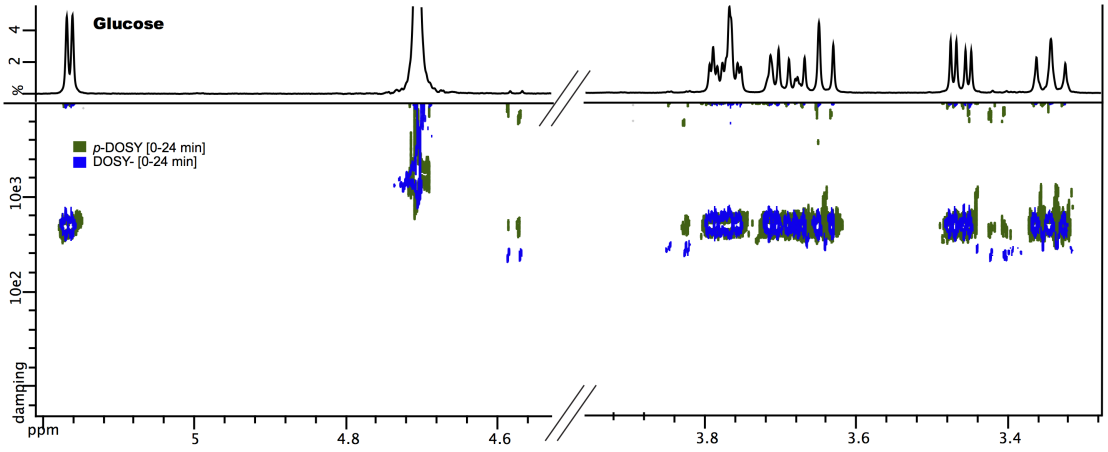



Figure 4.10: Overlay of the standard DOSY (blue color) and the *p*–DOSY (green color) experiments, acquired both at out-of-equilibrium stage (0.00-25.00 minutes).

DOSY measurements. The effect is less marked because the relative concentration variation of the $\alpha$-glucose is less important than in the case of $\beta$-glucose.

On the other hand, the experiment performed with a series of *p*–DOSY measurements does not display this systematic error, and accurate values are determined over the whole experiment. Because of the dispersion effect of the random permutation operation, the systematic evolution of the concentration is transformed here into random errors on the $I_o$ value. This added random noise has no effect on the analysis, but disperses in a random manner the concentration variation over the gradient list, thus providing a way to average out bias. However, because of the added noise, a slightly less precise determination of the diffusion coefficients is expected. This effect, while not very strong, can be observed on Figure 4.11 by the evolution of the error bars on the *p*–DOSY determined diffusion coefficients. The error bars present a somewhat larger spreading in the beginning of the kinetic and converge toward a minimal value.

In a nutshell, one can say that the *p*–DOSY experiment provides on evolving chemical systems, a slightly less precise, but much more accurate determination of the diffusion coefficient that the standard DOSY experiment.

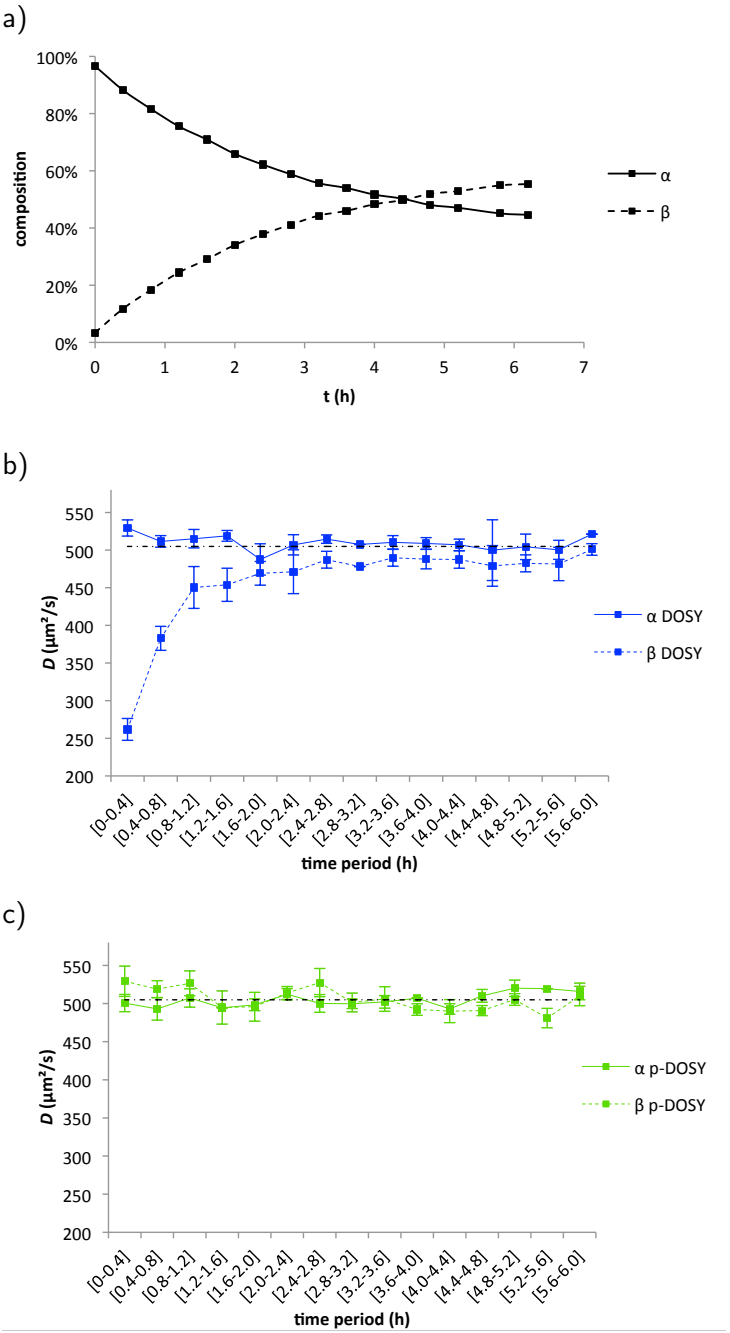



Figure 4.11: a) concentration evolution of $\alpha, \beta$-glucose over experimental time, measured as the intensity of the anomeric proton peak based on 1D experiments interleaved during the *p*–DOSY experiment series; b) evolution of the apparent diffusion coefficients of $\alpha$ and $\beta$-glucose over experimental time based on a series of sequential conventional DOSY measurements; c) same as b, obtained with the *p*–DOSY experiment. In b) and c) the horizontal doted line is located at the diffusion coefficient of $\alpha$-glucose, measured as the mean value of *p*–DOSY value over the whole kinetics.

## 4.4 Conclusions

The accurate estimation of diffusion coefficients is challenging in systems away from equilibrium. An inherent characteristic of out-of-equilibrium systems is the evolution of the concentration of the different components during a DOSY NMR experiment. Since the basis of a single DOSY measurement is the analysis of exponential decay of the signal over an increasing gradient strength, concentration evolution, within the characteristic measurement time, creates a bias, which can lead to large errors in the diffusion coefficient estimation. This was observed for the test case of the anomerization of glucose. Randomizing the list of gradient strength arrays within a single DOSY measurement, decouples the gradient evolution to the concentration evolution and removes the bias in the analysis, thus restoring the accuracy of the measure. The concentration evolution is still there though, as a noise source, and may have an impact on the precision of the measurement compared to a static sample, although its impact is minimum. Depending on the spectrometer software, this experiment requires little or no modification of the standard procedure, and the produced data-set can easily be analyzed using regular processing programs.

Altogether, the DOSY experiment with a randomized list of gradients (the permutated DOSY or *p*–DOSY) is an experiment exempt of bias in presence of an evolving system, in contrast with conventional DOSY. It is perfectly suited for the analysis of chemical reactions and out-of-equilibrium systems and can be used to monitor the rate of reaction kinetics, to characterize *in-operando* transient species and measure their diffusion coefficient, or to observed molecular organization phenomena and dynamics effects.

The robustness of *p*–DOSY can also be used to protect diffusion results from any experimental artifact, such as temperature shift or spectrometer drift. As it does not add any burden in the acquisition step nor at the processing step, we recommend spectroscopists to use *p*–DOSY systematically in diffusion and DOSY experiments.

## 4.5 Publication

The work presented in this chapter resulted in a publication in *Journal of Magnetic Resonance* (JMR), Volume 258, pages 12-16, September 2015.

# Bibliography

[Antalek 2002] B Antalek. *Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results*. Concepts Magn. Reson., vol. 14, no. 4, pages 225—-258, 2002. (Cited on page 59.)

[Augé 2009] Sophie Augé, Pierre-Olivier Schmit, Christopher A Crutchfield, Mohammad T Islam, Douglas J Harris, Emmanuelle Durand, Martin Clemancey, Anne-Agathe Quoineaud, Jean-Marc Lancelin, Yann Prigent, Francis Taulelle and Marc-André Delsuc. *NMR measure of translational diffusion and fractal dimension. Application to molecular mass measurement.* J. Phys. Chem. B, vol. 113, no. 7, pages 1914–1918, February 2009. (Cited on pages 58, 150 and 151.)

[Balayssac 2009] Stéphane Balayssac, Saleh Trefi, Véronique Gilard, Myriam Malet-Martino, Robert Martino and Marc-André Delsuc. *2D and 3D DOSY 1H NMR, a useful tool for analysis of complex mixtures: Application to herbal drugs or dietary supplements for erectile dysfunction*. Journal of Pharmaceutical and Biomedical Analysis, vol. 50, no. 4, pages 602–612, November 2009. (Cited on page 62.)

[Barjat 1998] H Barjat, G A Morris and A G Swanson. *A Three-Dimensional DOSY–HMQC Experiment for the High-Resolution Analysis of Complex Mixtures*. Journal of Magnetic Resonance, 1998. (Cited on pages 57, 62 and 63.)

[Birlirakis 1996] N Birlirakis and E Guittet. *A new approach in the use of gradients for size-resolved 2D-NMR experiments*. J. Am. Chem. Soc., vol. 118, no. 51, pages 13083–13084, 1996. (Cited on page 62.)

[Bowyer 1999] P J Bowyer, A G Swanson and G A Morris. *Randomized acquisition for the suppression of systematic F1 artifacts in two-dimensional NMR spectroscopy.* J. Magn. Reson., vol. 140, no. 2, pages 513–515, October 1999. (Cited on page 65.)

[Bradley 2005] Scott A Bradley, Krish Krishnamurthy and Haitao Hu. *Simplifying DOSY spectra with selective TOCSY edited preparation*. Journal of Magnetic Resonance, vol. 172, no. 1, pages 110–117, January 2005. (Cited on page 62.)

[Cohen 2012] Yoram Cohen, Liat Avram, Tamar Evan-Salem, Sarit Slovak, Noam Shemesh and Limor Frish. *Diffusion NMR in Supramolecular Chemistry and Complexed Systems*. In Analytical Methods in Supramolecular Chemistry, pages 197–285. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, March 2012. (Cited on page 57.)

[Didenko 2010] T Didenko, R Boelens and S G D Rudiger. *3D DOSY-TROSY to determine the translational diffusion coefficient of large protein complexes*. Protein Engineering Design and Selection, vol. 24, no. 1-2, pages 99–103, December 2010. (Cited on page 62.)

[Evans 2013] Robert Evans, Zhaoxia Deng, Alexandria K Rogerson, Andy S McLachlan, Jeff J Richards, Mathias Nilsson and Gareth A Morris. *Quantitative Interpretation of Diffusion-Ordered NMR Spectra: Can We Rationalize Small Molecule Diffusion Coefficients?* Angew. Chem. Int. Ed., vol. 52, no. 11, pages 3199–3202, January 2013. (Cited on page 58.)

[Floquet 2009] Sébastien Floquet, Sébastien Brun, Jean-François Lemonnier, Marc Henry, Marc-André Delsuc, Yann Prigent, Emmanuel Cadot and Francis Taulelle. *Molecular weights of cyclic and hollow clusters measured by DOSY NMR spectroscopy.* J. Am. Chem. Soc., vol. 131, no. 47, pages 17254–17259, December 2009. (Cited on page 57.)

[García-Álvarez 2011] P García-Álvarez, Robert E Mulvey and John A Parkinson. *"LiZn(TMP)$_3$", a Zincate or a Turbo-Lithium Amide Reagent? DOSY NMR Spectroscopic Evidence*. Angew. Chem. Int. Ed. Engl., vol. 123, no. 41, page 9842–9845, 2011. (Cited on pages 57 and 65.)

[Gilard 2011] V Gilard, S Trefi, S Balayssac, M A Delsuc, T Gostan, M Malet-Martino, R Martino, Y Prigent and F Taulelle. *DOSY NMR for drug analysis*. NMR Spectroscopy in Pharmaceutical Analysis, page 269, 2011. (Cited on pages 57 and 60.)

[Giuseppone 2008] Nicolas Giuseppone, Jean-Louis Schmitt, Lionel Allouche and Jean-Marie Lehn. *DOSY NMR experiments as a tool for the analysis of constitutional and motional dynamic processes: implementation for the driven evolution of dynamic combinatorial libraries of helical strands.* Angew. Chem. Int. Ed. Engl., vol. 47, no. 12, pages 2235–2239, 2008. (Cited on page 57.)

[Gomez 2009] M Victoria Gomez, Javier Guerra, Aldrik H Velders and Richard M Crooks. *NMR characterization of fourth-generation PAMAM dendrimers in the presence and absence of palladium dendrimer-encapsulated nanoparticles.* J. Am. Chem. Soc., vol. 131, no. 1, pages 341–350, January 2009. (Cited on page 57.)

[Gurst 1991] Jerome E Gurst. *NMR and the structure of D-glucose*. J. Chem. Educ., vol. 68, no. 12, page 1003, December 1991. (Cited on page 66.)

[Hernandez Santiago 2015] Adrian A Hernandez Santiago, Anatoly S Buchelnikov, Maria A Rubinson, Semen O Yesylevskyy, John A Parkinson and Maxim P Evstigneev. *Shape-independent model (SHIM) approach for studying aggregation by NMR diffusometry.* J. Chem. Phys., vol. 142, no. 10, page 104202, March 2015. (Cited on page 58.)

[Johnson 1999] C S Johnson. *Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications* . Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 34, pages 203–256, 1999. (Cited on page 59.)

[Li 2009] Deyu Li, Ivan Keresztes, Russell Hopson and Paul G Williard. *Characterization of reactive intermediates by multinuclear diffusion-ordered NMR spectroscopy (DOSY).* Acc. Chem. Res., vol. 42, no. 2, pages 270–280, February 2009. (Cited on pages 57 and 65.)

[Lucas 2002] Laura H Lucas, William H Otto and Cynthia K Larive. *The 2D-J-DOSY Experiment: Resolving Diffusion Coefficients in Mixtures*. Journal of Magnetic Resonance, vol. 156, no. 1, pages 138–145, May 2002. (Cited on pages 57 and 62.)

[Lucas 2004] Laura H Lucas and Cynthia K Larive. *Measuring ligand-protein binding using NMR diffusion experiments*. Concepts Magn.Reson., vol. 20A, no. 1, pages 24–41, 2004. (Cited on page 57.)

[Morris 1992] Kevin F Morris and Charles S Johnson Jr. *Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy*. J. Am. Chem. Soc., vol. 114, no. 8, pages 3139–3141, 1992. (Cited on page 59.)

[Newman 2007] Jacob M Newman and Alexej Jerschow. *Improvements in complex mixture analysis by NMR: DQF-COSY iDOSY.* Anal. Chem., vol. 79, no. 7, pages 2957–2960, April 2007. (Cited on page 63.)

[Nguyen 2009] Rémi Nguyen, Lionel Allouche, Eric Buhler and Nicolas Giuseppone. *Dynamic combinatorial evolution within self-replicating supramolecular assemblies.* Angew. Chem. Int. Ed. Engl., vol. 48, no. 6, pages 1093–1096, 2009. (Cited on pages 57 and 65.)

[Nilsson 2004] Mathias Nilsson, Ana M Gil, Ivonne Delgadillo and Gareth A Morris. *Improving Pulse Sequences for 3D Diffusion-Ordered NMR Spectroscopy: 2DJ-IDOSY*. Anal. Chem., vol. 76, no. 18, pages 5418–5422, September 2004. (Cited on pages 63 and 64.)

[Nilsson 2005] Mathias Nilsson, Ana M Gil, Ivonne Delgadillo and Gareth A Morris. *Improving pulse sequences for 3D DOSY: COSY-IDOSY*. Chem. Commun., no. 13, page 1737, 2005. (Cited on page 63.)

[Nilsson 2007] Mathias Nilsson and Gareth A Morris. *Pure shift proton DOSY: diffusion-ordered 1H spectra without multiplet structure.* Chem. Commun. (Camb.), no. 9, pages 933–935, March 2007. (Cited on page 64.)

[Nilsson 2009] Mathias Nilsson, Maryam Khajeh, Adolfo Botana, Michael A Bernstein and Gareth A Morris. *Diffusion NMR and trilinear analysis in the study of reaction kinetics.* Chem. Commun. (Camb.), no. 10, pages 1252–1254, March 2009. (Cited on page 65.)

[Nonappa 2015] Nonappa, David Šaman and Erkki Kolehmainen. *Studies on supramolecular gel formation using DOSY NMR.* Magn. Reson. Chem., vol. 53, no. 4, pages 256–260, January 2015. (Cited on page 57.)

[Schlörer 2002] Nils E Schlörer, Eurico J Cabrita and Stefan Berger. *Characterization of reactive intermediates by diffusion-ordered NMR spectroscopy: a snapshot of the reaction of $^{13}CO_2$ with [$Cp_2Zr(Cl)H$].* Angew. Chem. Int. Ed. Engl., vol. 41, no. 1, pages 107–109, January 2002. (Cited on pages 57 and 65.)

[Sinnaeve 2012] Davy Sinnaeve. *The Stejskal–Tanner equation generalized for any gradient shape—an overview of most pulse sequences measuring free diffusion - Sinnaeve - 2012 - Concepts in Magnetic Resonance Part A - Wiley Online Library*. Concepts Magn. Reson., vol. 40A, no. 2, pages 39–65, March 2012. (Cited on page 59.)

[Šmejkalová 2008] Daniela Šmejkalová and Alessandro Piccolo. *Aggregation and Disaggregation of Humic Supramolecular Assemblies by NMR Diffusion Ordered Spectroscopy (DOSY-NMR)*. Environ. Sci. Technol., vol. 42, no. 3, pages 699–706, February 2008. (Cited on page 57.)

[Stchedroff 2004] Marc J Stchedroff, Alan M Kenwright, Gareth A Morris, Mathias Nilsson and Robin K Harris. *2D and 3D DOSY methods for studying mixtures of oligomeric dimethylsiloxanes*. Phys. Chem. Chem. Phys., vol. 6, no. 13, pages 3221–3227, 2004. (Cited on page 57.)

[Stejskal 1965] E O Stejskal and J E Tanner. *Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient*. J. Chem. Phys., vol. 42, no. 1, page 288, 1965. (Cited on page 58.)

[Viéville 2011] J Viéville, Matthieu Tanty and Marc-André Delsuc. *Journal of Magnetic Resonance*. J. Magn. Reson., vol. 212, no. 1, pages 169–173, September 2011. (Cited on page 57.)

[Vitorge 2006] Bruno Vitorge, Damien Jeannerat and Damien Jeanneat. *NMR diffusion measurements in complex mixtures using constant-time-HSQC-IDOSY and computer-optimized spectral aliasing for high resolution in the carbon dimension.* Anal. Chem., vol. 78, no. 15, pages 5601–5606, August 2006. (Cited on page 63.)

[Wilkins 1999] D K Wilkins, S B Grimshaw, V Receveur, Christopher M Dobson, J A Jones and L J Smith. *Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques*. Biochemistry, vol. 38, no. 50, pages 16424–16431, 1999. (Cited on page 57.)

[Wu 1996] D Wu, A Chen and C S Johnson Jr. *Three-Dimensional Diffusion-Ordered NMR Spectroscopy: The Homonuclear COSY–DOSY Experiment*. Journal of Magnetic Resonance, 1996. (Cited on page 62.)

[Zangger 1997] Klaus Zangger and Heinz Sterk. *Homonuclear broadband-decoupled NMR spectra*. J. Magn. Reson., vol. 124, no. 2, pages 486–489, 1997. (Cited on page 64.)

# Appendix A

## A.1 Molecules, solvents and sample preparation

Crystalized $\alpha$-D-glucose (anhydrous, 96% purity) was purchased to Sigma-Aldrich (USA). Deuterated solvents DMSO-d6 (99.96%) and $D_2O$ (99.9%) were purchased to Euriso-Top (France). All samples were measured in 5.0 mm tubes. A stock solution of 100.0 $mM$ $\alpha$-D-glucose in DMSO was prepared. 100 $\mu$L of the glucose stock solution were added in 900 $\mu$L $D_2O$ resulting in final concentration of 10.0 $mM$.

## A.2 NMR experiments

DOSY NMR measurements were carried out on a Bruker 500 Avance I NMR spectrometer, operating at 500.137 MHz for $^1$H, equipped with a 5 mm TXI probe. The pulse sequence used was a longitudinal eddy current delay bipolar gradient pulse (`1edbpgp2s`) An exponential gradient list of 32 values was created by using the standard AU program `dosy`. For the *p*–DOSY experiment, the list was permutated by an ad-hoc python script, and the permutated list was used for acquisition and processing. Experiments were acquired with 8 scans, $\delta/2$ of 3.0 ms and $\Delta$ of 150 ms, for an overall time of 23:09 min. A series of 15 spectra was acquired at a temperature of 300 $K$ for a total duration of 6.0 hours. Before and after every DOSY spectrum, an 8 scans proton spectrum was recorded. The whole experiment was duplicated once with regularly increasing series of gradients, and once with a permutated series.

The DOSY spectra acquired on Bruker were processed with the NMRnotebook program, using the DOSY MaxEnt implementation (version 2.7, NMRTEC France). No modification had to be made to the computation protocol despite the permutated entries in the DOSY file.

The final diffusion coefficients were calculated from the average position of the anomeric signal for each species. Error bars were estimated as the standard deviation along the diffusion axis of the different signals. by averaging the values measured for all major peaks in each proton signal. Minor viscosity variations were compensated by normalizing all measurements to the DMSO diffusion coefficient.

## A.3 *Permute* script

Here is presented the "Permute" script to permutate the Difframp and difflist for acquisition and processing the *p*–DOSY experiments. The original Difframp, created by TopSpin when a DOSY experiment is launched, has to be replaced by the permutated Difframp before running the *p*–DOSY experiment. The same procedure has to been followed for the difflist before processing the spectrum.

```
"""
Randomizes a Difframp list
Created by DELSUC Marc-Andre and Julia Asencio on 2013-12-18.
"""

import random

N=64 # Adapt to your system

lperm = random.sample(range(N-1), N-1)
perm = [0] + [i+1 for i in lperm]

print N, "points :", perm

filelist = ['difflist', '/opt/topspin/exp/stan/nmr/lists/gp/user/Difframp']

for fname in filelist:
      F = open(fname,'r')
      FF = open(fname+'P','w')
      ll = []
      for l in F:
            if not l.startswith('#'):
                  ll.append(l)
            elif l[0:6] != '##END=':
                  FF.write(l)
      F.close()
      if len(ll) != N:
            raise Exception( 'This prgm is meant for %d points'%N )
      for i in range(N):
            FF.write(ll[perm[i]])
      FF.write('##END=')
      FF.close()
```

Figure A.1: python `permute.py` script to permutate the Difframp and difflist **files** for acquisition and processing of the *p*–DOSY experiments.

Chapter 5

# Biophysical characterization of a high conserved peptide from the N-terminal Domain of the Androgen Receptor



The Androgen Receptor as a complex system and its role in prostate cancer have been introduced in Chapter 2. Within this topic, the receptor and specially its N-terminal domain is presented as the main actor when the castration resistance prostate cancer is developed. The constitutively active AR splice variants that lack the LBD appear and the current therapies (based on antiandrogens) become inefficient. Then, the NTD-AR and its activation function AF1 are responsible for the receptor activity. Consequently, the NTD-AR represents a new target for drug development.

Within this topic, our strategy is to find these potential target regions from the NTD-AR sequence and to characterize them. In order to achieve this objective, we propose a peptide study instead of the entire domain study. For that, an extensively bioinformatic study is carried out to find the hot spots within the sequence then, these peptide regions are going to be characterized using different biophysical methods.

## 5.1 Bioinformatic analysis of the NTD-AR

In order to address the objective of this study, to characterize and to find partial secondary structure in the N-terminal domain of the Androgen Receptor, a bioinformatic study of this domain was carried out.

For this reason, the disorder tendency using metaPrDOS have been studied within the NTD-AR sequence in order to spot the most ordered predicted regions. These regions, with higher order tendency shown in Figure 5.1, give a hint for finding the sequences more likely to present a secondary structure.

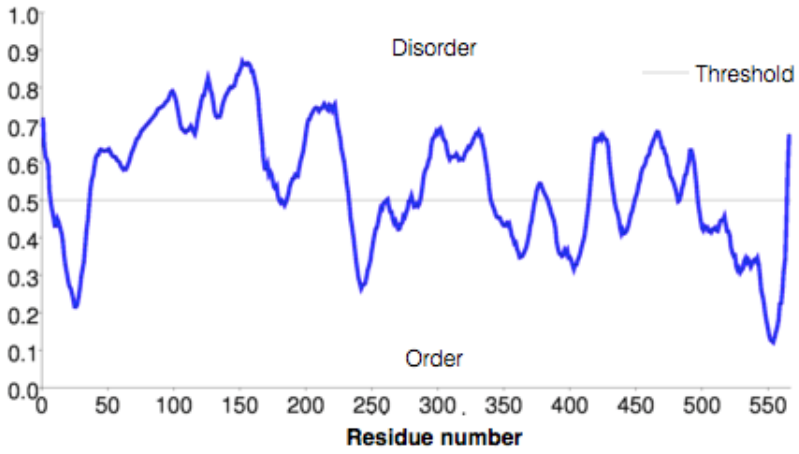



Figure 5.1: Disorder tendency prediction obtained with the metaPrDOS analysis performed on the NTD-AR sequence. Residues below the 0.5 threshold are predicted as ordered, residues over 0.5 are predicted as disordered.

To complement this information, other studies are performed using different bioinformatic tools. MeDor is used to establish the secondary structure prediction on the NTD-AR based on the hydrophobic clusters formed in the amino acid sequence. As a result, a prediction of potential structured regions is stablished, see Figure 5.2.

The results of secondary structure prediction using the different bioinformatic tools explained in Appendix B (**MeDor**, **NPS@**, **PAT**) are summarize in Table 5.1. From this study a list of five peptides are selected from the higher consensus between the different methods, showing a high propensity to present a secondary structure. All these peptides sequences belong to the AF1 region.

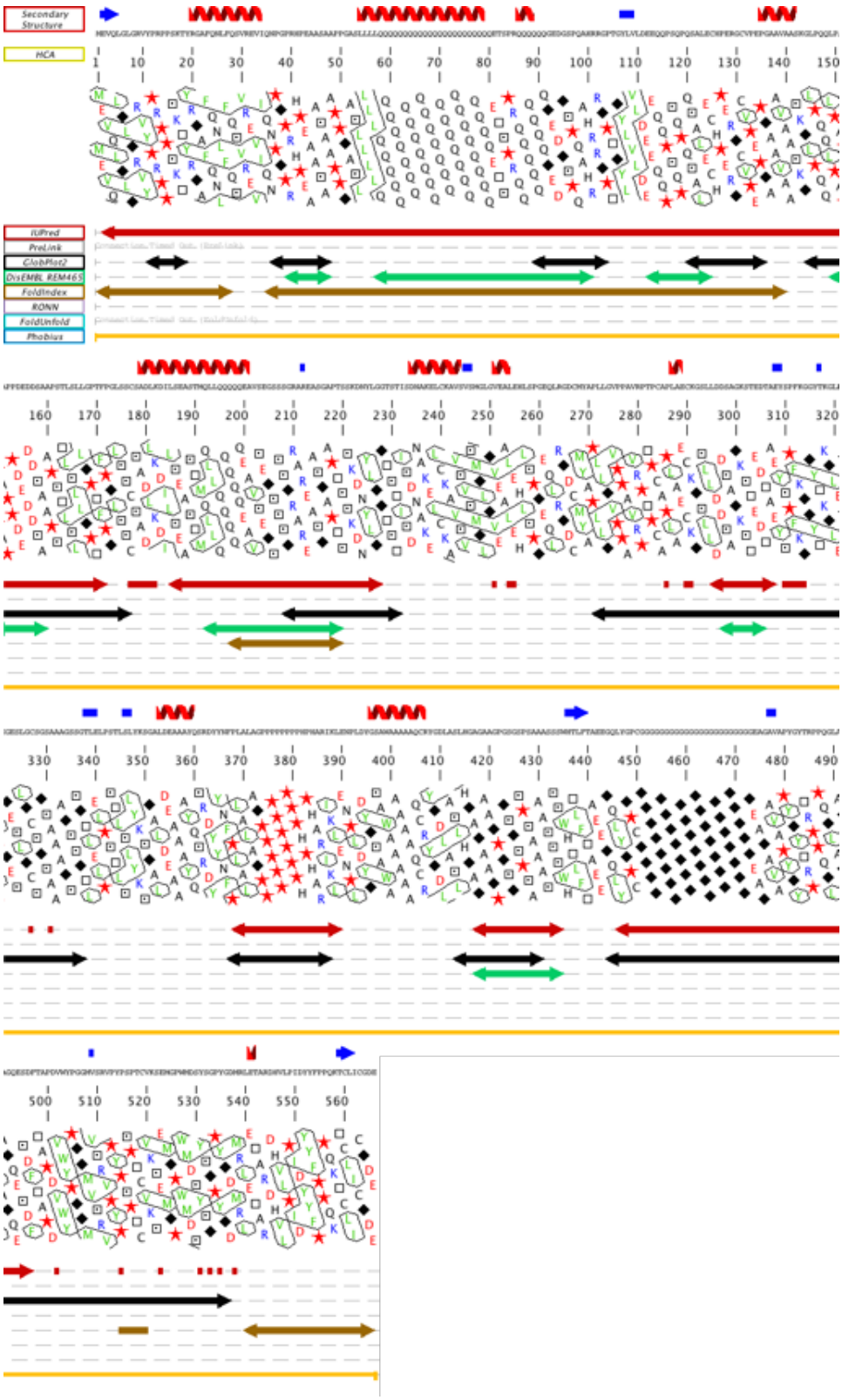



Figure 5.2: Results obtained with MeDor analysis. The figure shows the different secondary structure predicted along the NTD-AR domain and it shows in the top of the sequence tracks with higher consensus to adopt a $\beta$-sheet (blue arrow) or helix (red helix) conformation.

Table 5.1: Secondary structure prediction. A selection of five peptides is presented in this table, with the human-AR numbering, the residues sequence, the secondary structure prediction for the different methods: H (helix), $\beta$ ($\beta$-sheet) and C (coil).

| | Position | Peptide | MeDor | NPS@ | PAT | Remarks/Rational | | |
|---|---|---|---|---|---|---|---|---|
| **1** | [179-194] | ADLKDILSEASTMQLL | H | H | H | L194R | K182R | LxxLL |
| **2** | [237-247] | KELCKAVSVSM | H/$\beta$ | H/$\beta$ | H | | | |
| **3** | [339-355] | LELPSTLSLYKSGALDE | H | H | H | S343* | P342L | |
| **4** | [433-441] | SSWHTLFTA | C | H | H/$\beta$ | W | W435L | WxxLF |
| **5** | [526-537] | PWMDSYSGPYGD | C | C | C | W | | |

Besides these results, also other rationals are taken into consideration. The first aspect is to look for somatic mutations present in prostate cancer using the androgendb database [Gottlieb 1998, Gottlieb 2012], which url address is: http://androgendb.mcgill.ca/, some mutations are found in the five sequences shown in Table 5.1

The second aspect that is taken into consideration is the presence of recognition motifs, as LxxLL or WxxLF that are present in two of the selected peptides. LxxLL has been reported in the literature as a motif in cofactor proteins for interactions with hormone dependent nuclear receptors [Heery 1997], as the androgen receptor [Dubbink 2004], but also in other transcriptional regulation protein–protein interactions and a novel interaction implicated in leukemia [Garber 2005]. And the WxxLF motif, involved in interaction of the N-terminal domain with the ligand binding domain of the androgen receptor [He 2000].

There are also other important peculiarities in the sequence of the human NTD-AR, as the presence of numerous hydrophobic and aromatic residues usually absent in disordered regions. Is the case of the presence of alternates hydrophobic and hydrophilic residues in peptide **2** (KELCKA**VSVS**M), a signature of $\beta$-sheet structure. Or the presence or several tryptophan, tyrosine or phenylalanine in the ending part of the N-terminal domain, close to the beginning of the DNA binding domain represented in the peptides **4** and **5**.

In order to complete the study of the primary sequence of the NTD-AR for selecting the peptides for a further biophysical study, an alignment of the AR was performed among twelve vertebrate species.

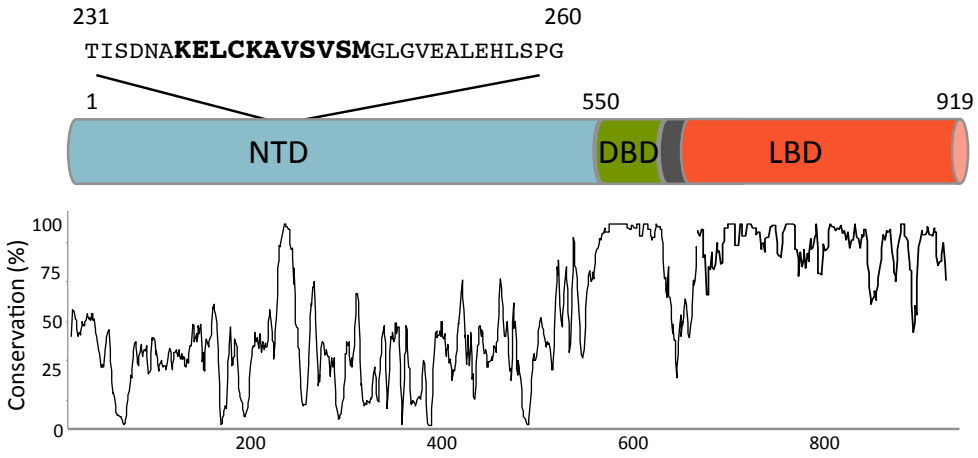



Figure 5.3: Androgen Receptor showing the NTD, DBD, a hinge (gray) and LBD. Numbering is for human AR. Below is a quantitative representation of the primary sequence conservation among vertebrates. The DBD, the LBD, and a short region in the NTD show a high level of conservation, nearly 100%.

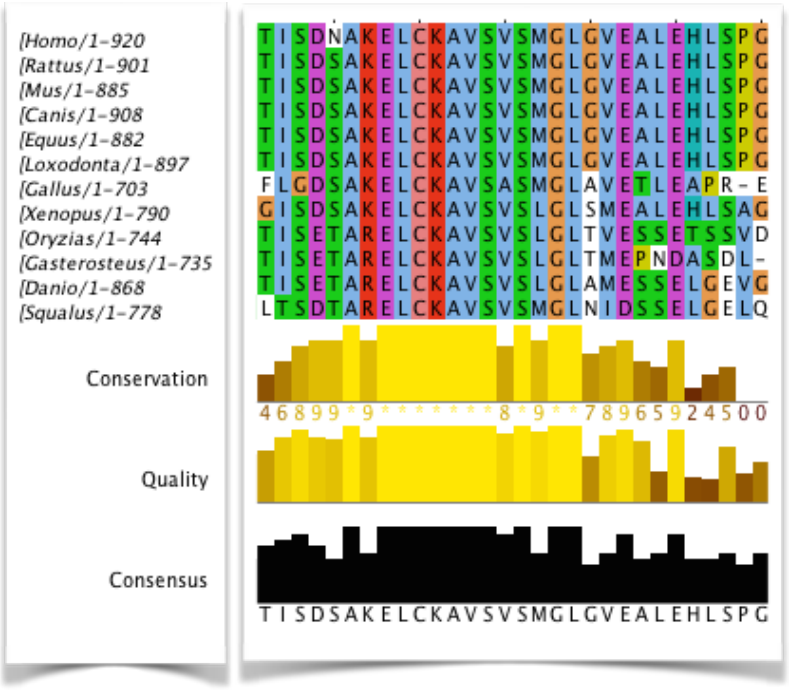



Figure 5.4: Zoom in the highest conserved region of the NTD-AR. Twelve vertebrates species used for the phylogenetic study.

As shown in Figure 5.3, the DBD and LBD presents a high conservation as expected for well known structured regions. On the other hand, the NTD displays a low level of conservancy

(as expected for disordered regions), except for a short region of eleven residues [233-255] (corresponding to the peptide **2** obtained above with the different predictors of structure) that is highly conserved across all vertebrates, at a level equivalent to that of the DBD.

## 5.2 Biophysical characterization of different peptides from the NTD-AR

After completion of the bioinformatic analysis of the NTD-AR, a series of different peptides were chosen to carry out a structural characterization using different biophysical methods. The peptides numbering presented below is referred to the Table 5.1.

### 5.2.1 Peptides study by NMR: A first structural identification.

Firstly, the five peptides have been studied using NMR, which allows an initial characterization. And the possible structure displayed for the peptides can be stablished.

- **1D and 2D NMR: peptide assignment**

  For an initial and rapid characterization of the possible structure present in the peptides, a series of equivalent NMR experiments were performed for each of them. The numbering used in this study corresponds to the one used in Table 5.1.

  **Peptide 4 (SSWHTLFTA)**

  As example of the NMR study carried out for each one of the peptides, here is shown the assignment of the peptide **4**. The analysis starts by recording a 1D $^{1}$H, Figure 5.5. By itself this experiment does not give us a lot of information about the peptide structure and it has to be complemented by a series of 2D experiments. A series of TOCSY and ROESY homonuclear experiments were recorded.

  As a result of these experiments the $^{1}$H assignment Figures 5.5 of the peptide **4** was performed. Together with these results, no NOE signal was obtained.

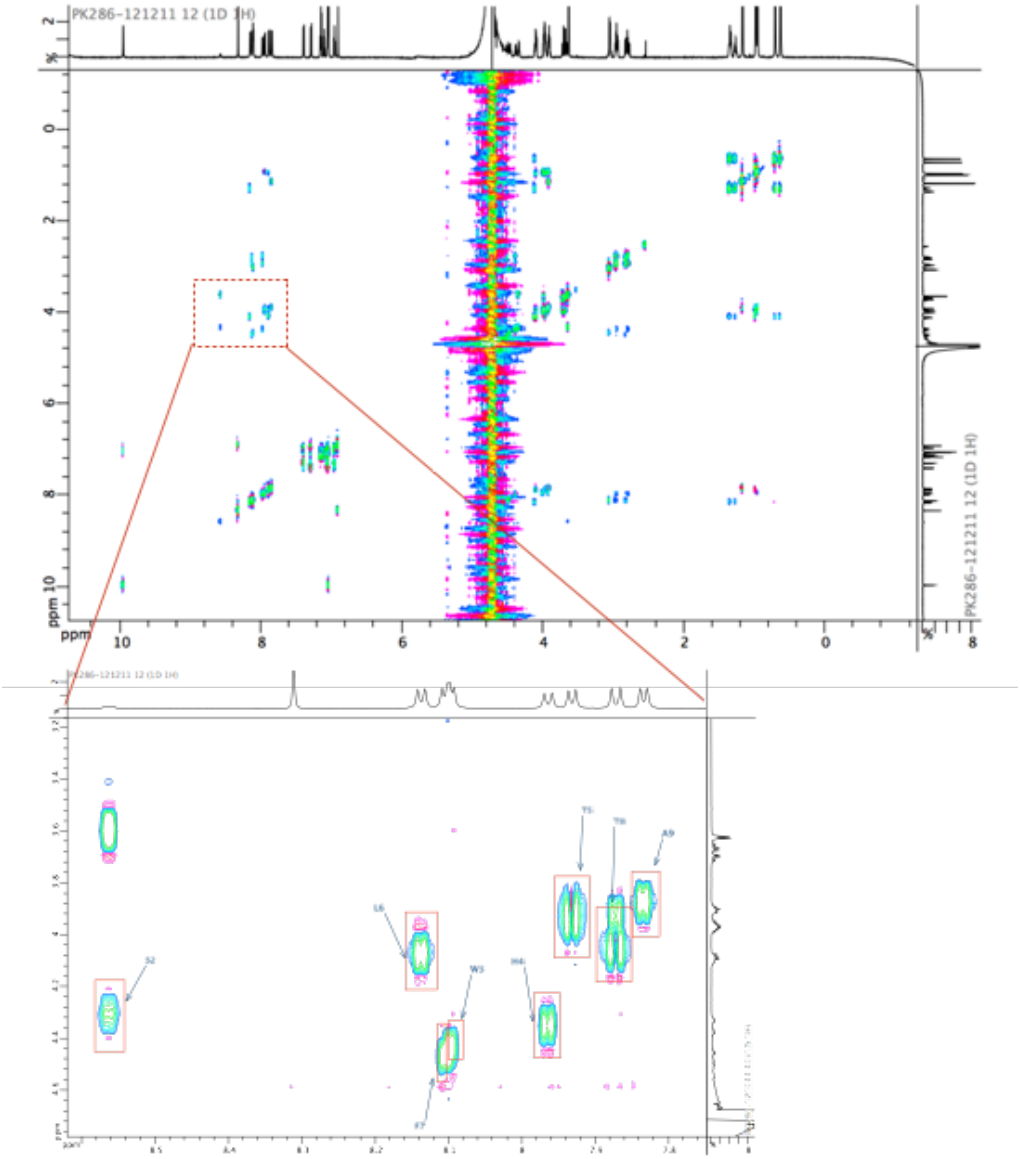



Figure 5.5: The 1D $^{1}$H spectrum of peptide **4** using a zgpr sequence in shown on the top of the 2D TOCSY spectrum. A zoom on the NH region shows the peptide assignment.

### Peptide 1 (ADLKDILSEASTMQLL), 3 (LELPSTLSLYKSGALDE) and 5 (PWMDSYSGPYGD)

The equivalents experiments were performed for peptides 1, 3 and 5 with the aim of finding any structuration sign, nevertheless the results obtained were equivalent to the peptide **4**, no extra information was achieve apart from the $^{1}$H assignment of the residues sequences.

### Peptide 2 (KELCKAVSVSM)

Contrarily to the other peptides studied above, peptide **2** displays a special behavior. While acquiring the NMR experiments to perform the assignment of the peptide over night, $^{1}$H signal disappeared. To understand this strange behavior, a series of $^{1}$H spectra was recorded (Figure 5.6) in order to monitor the reproducibility of the signal disappearance.

Looking to the Figure 5.6 the broadening of the water signal draws immediately our attention, this signature indicates the appearance of soluble mesoscale objects in the sample.

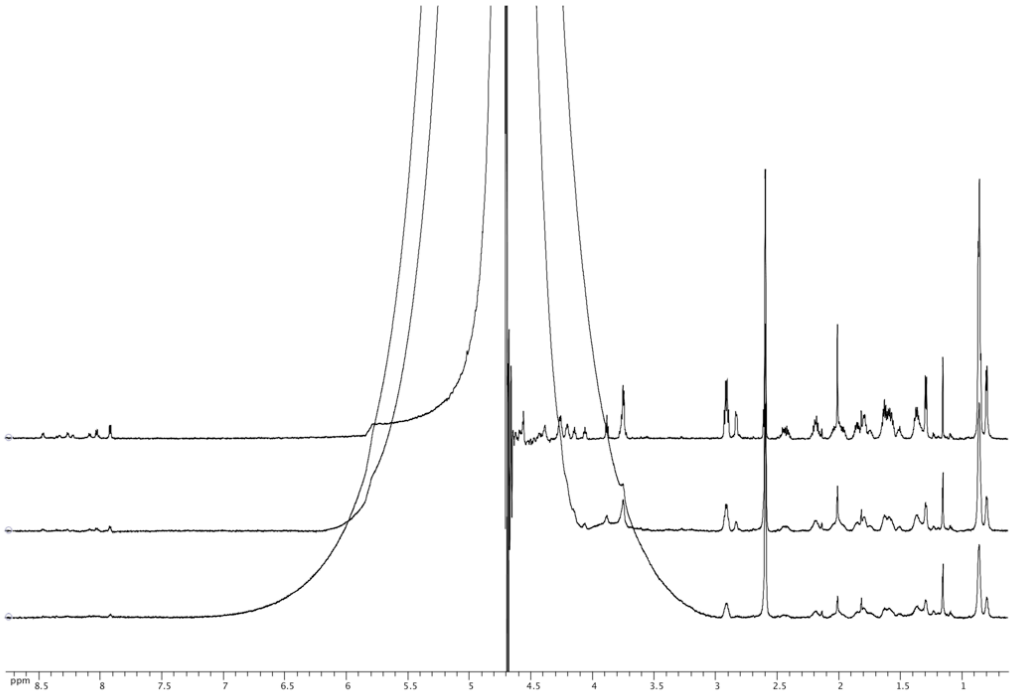



Figure 5.6: NMR spectra obtained during the kinetics with 10% DMSO, time 0.5 h (top), 6 h and 23 h (bottom).

#### 5.2.1.1 DMSO influence

In the peptides study, all samples were prepared with 10% of dimethyl sulfoxide (DMSO) in order to have to control experiments with the same % of DMSO that we use afterwards to solubilize small molecules for future interactions studies peptide with small molecules.

It is well known in chemistry, sometimes forgotten in biology, that DMSO is an efficient mild oxidant for disulfide bridge formation in peptides containing a cysteine [Tam 1991]. Indeed, in peptide **2** there is a Cys present in the amino acid sequence (KEL**C**KAVSVSM).

This was the starting point for a kinetics study. This study aims to follow the signal disappearance of peptide 2, upon the addition of different % of DMSO. Different kinetics were followed by 1D $^{1}$H NMR after adding 1%, 5% and 10% of DMSO to a final solution with 0.5 mM of peptide **2** in phosphate buffer 20 mM pH=6.4 . Figure 5.9 shows the differences in the kinetics rates for the different % of DMSO, the signal disappearance is faster when increasing the % DMSO, the reaction rates are nearly proportional to the concentration of DMSO (see [Asencio-Hernández 2014]). In the absence of DMSO, slow aggregation kinetics is also observed, but in an irreproducible manner, due to no control of the presence of oxygen in the air.

In all cases, the signal intensity decays to a plateau at around 0.1 mM, and the initial part of the curve shows a characteristic second-order kinetics, characterized by a hyperbolic decay (Eq. 5.1).

$$[P](t) = \frac{[P]_o}{1 + 2[P]_o kt} \tag{5.1}$$

Fit of the second-order kinetics equation was performed by a linear analysis of the inverse of the concentration vs time.

$$\frac{1}{[P]}(t) = \frac{1}{[P]_o} + 2kt \tag{5.2}$$

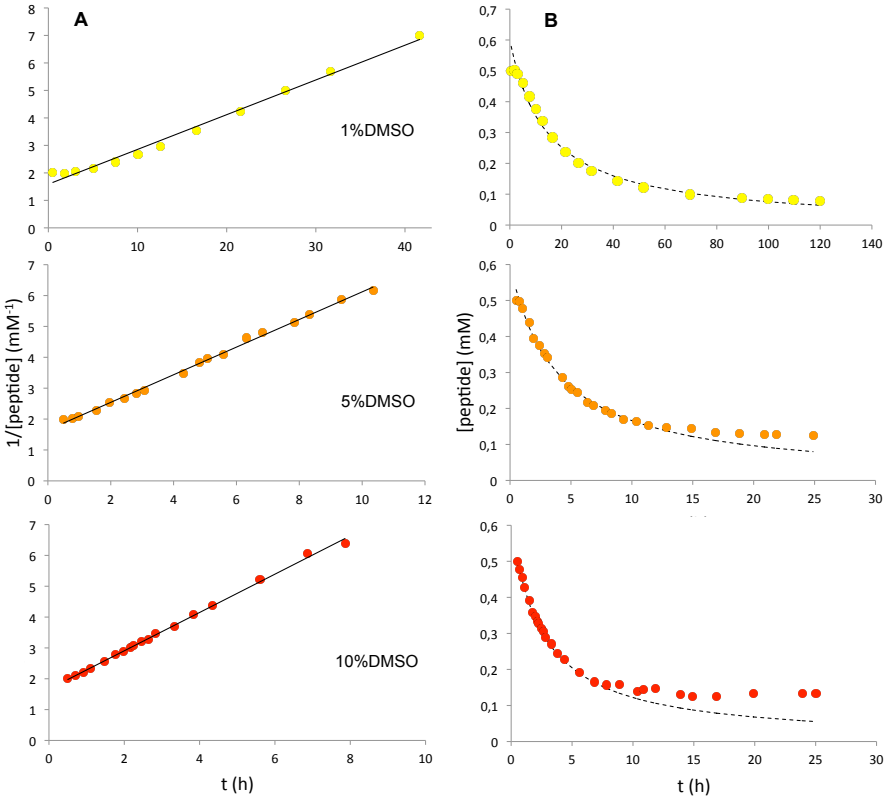



Figure 5.7: Kinetics with 1%, 5%, and 10% DMSO, A) second order kinetics fit. For the fit we excluded the last points of the kinetics, where the peptide concentration remains constant; B) peptide concentration disappearance monitored by NMR with the result of the second-order kinetics fit as a dashed line.

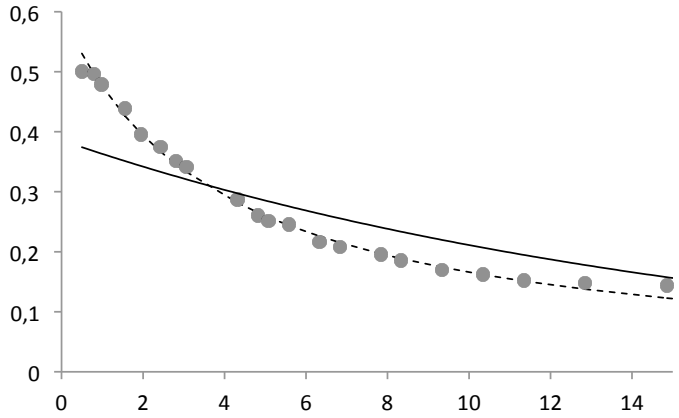


Figure 5.8: Exponential and second-order reaction fitting for the 5% DMSO conditions with the comparison of a fit of the kinetic with a second-order (dashed line) and a first-order kinetic as the exponential (solid line).

Once the DMSO influence was proven with the results shown above. It is necessary to complete the oxidation study, for that we analyze the peptide signal evolution by adding a reducing agent tris(2-carboxyethyl)phosphine (TCEP). The presence of TCEP inhibits the signal disappearance, and the peptide remaines stable in solution (Figure 5.9).

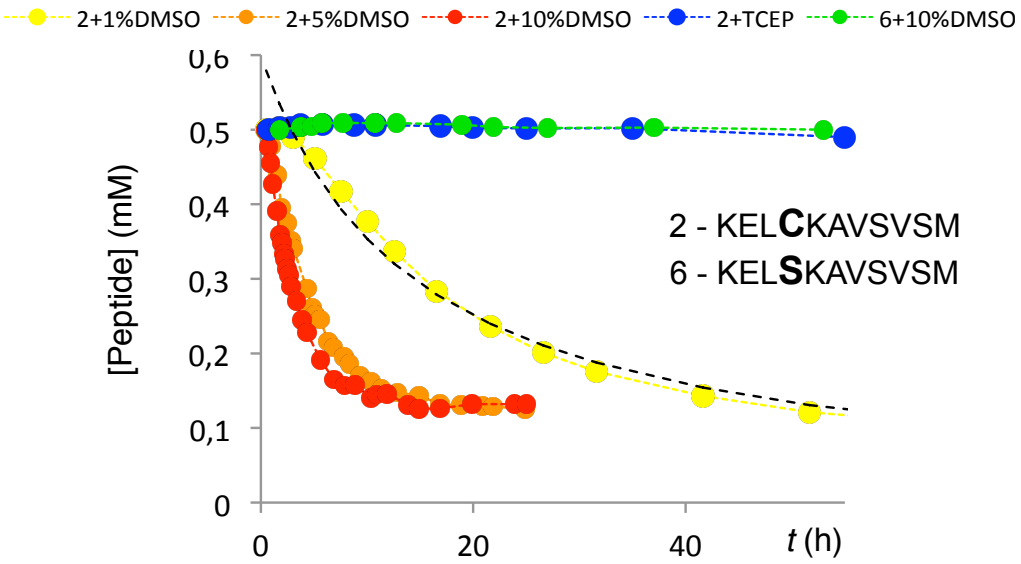



Figure 5.9: Evolution of NMR signal intensities for peptide **2** and **6** under various oxidation conditions. **2**+1% DMSO (yellow), **2**+5% DMSO (orange), **2**+10% DMSO (red), **2**+TCEP (blue) and **6**+10% DMSO (green).

Peptide **6** (KELSKAVSVSM), with a cysteine mutated to a serine, is used as a control to prove the implication of the cysteine in the oxidative process and the further NMR disappearance. This peptide **6**, with a hydroxyl in place of the thiol group, is perfectly stable in solution (Figure 5.9) after the addition of DMSO.

Finally, after all these experiments we conclude that the cause of the NMR signal disappearance is the DMSO, in other words, an oxidation process within the peptide, with a most than probably, dimerization of the peptide through a disulfide bride formation. For this reason, a final simple experiment was to test the reversibility of the reaction. Firstly, a 72 h kinetics at 25°C with 2% DMSO, as oxidative conditions, was performed. The reversibility of the reaction was recorded by adding 0.25 mg (5 mM = 10 equivalents) of TCEP to this solution, then the sample was warmed up to 50°C and sonicated. Figure 5.10 shows the recovery of the 100% of the NMR signal of the monomeric peptide after the addition of TCEP after the kinetics of disappearance, confirming the reversibility of the redox reaction.

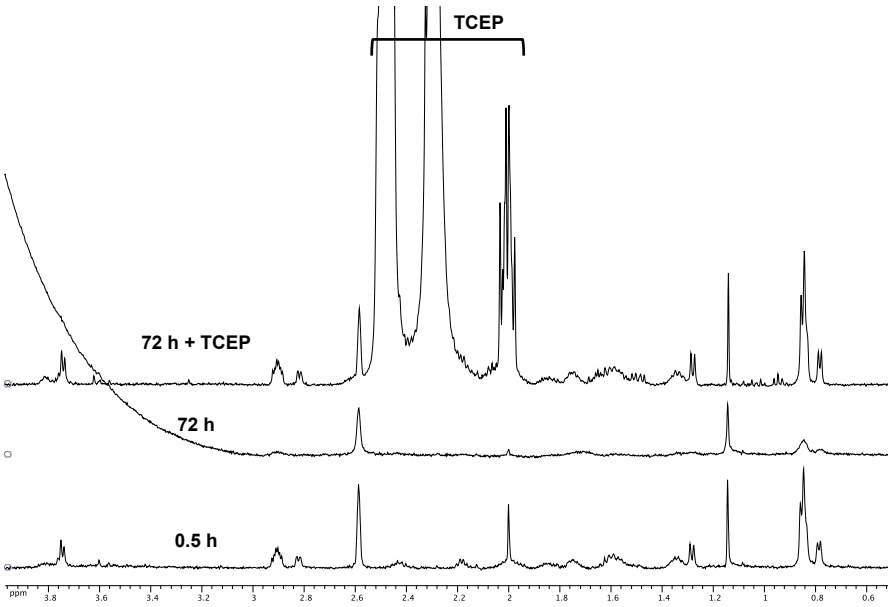



Figure 5.10: Reaction reversibility after the addition of the reducing agent TCEP.

Putting it all together, the previous results show that it is not only a matter of oxidation, or disulfide bridge formation in the peptide **2**. The dimerization by itself should not be enough to justify the signal disappearance from the spectrum, and specially the broadening of the water signal observed in Figure 5.6 is a signature of the appearing of soluble mesoscale objects in the sample. Moreover, a slight opalescence and eventually a light precipitate was appearing in the NMR tube, characteristic of oligomerization process.

A special attention deserves the fact that the second-order kinetics data suggest that the initial step of the oligomerization is the formation of a peptide dimer, by the presence of DMSO acting as oxidant, although more complex mechanism might also be present, as reported in literature [Morris 2009]. The dimer further aggregates into supramolecular structures, and thus disappears from the liquid-state NMR spectrum. This hypothesis is supported by the observation that the peptide with Cys mutated to Ser did not display any oligomerization under the same conditions.

All these results together put us on the track that as larger objects, aggregates, were appearing in solution, the nature of which could be amyloid fibers.

## 5.3 Amyloid fibers characterization

Amyloid fibers have been introduced in Chapter 1 and present a particular biophysical signature. In order to classify an object as amyloid or *amyloidlike* fibers a detailed study of the structure and morphology of the objects have to be performed. Amyloid fibrils should have a fibrillar morphology with a presence of $\beta$–sheet secondary structure. They have a positive response for the Thioflavin assay or Congo Red birefringence assay [Nilsson 2004]. Here we present a detailed study of the peptide **2** aggregation, in order to determine the formation of amyloid fibers in solution.

### Mass Spectrometry

With all these informations and with the aim to characterize if the objects appearing in solutions could be defined as amyloid fibrils, we decided to perform a first mass spectrometry (MS) analysis to determine if the peptide dimer was present.

In this purpose, the precipitate from the NMR tube was dissolved in 1:1 (v/v) Acetonitrile-Water with 2% formic acid and analyzed by MS. The result of this analysis is shown in Figure 5.11, where the main specie is the dimer assemble by disulfide bridge formation. The mass is calculated using the odd m/z (red circles), to avoid miscalculations with the monomeric form.

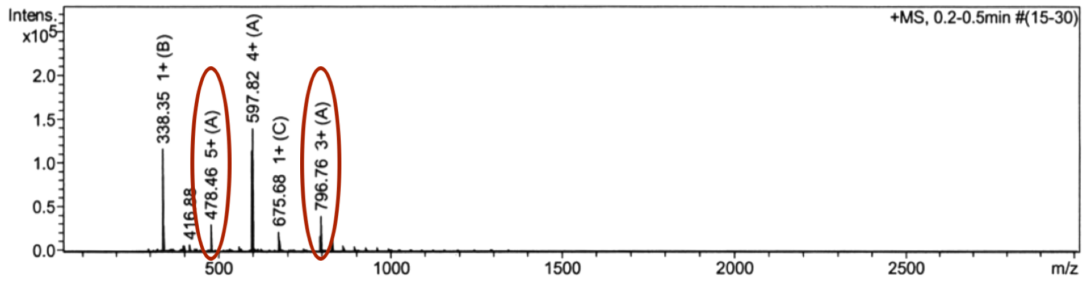



Figure 5.11: The main component (A) corresponds to the molecular mass of the peptide dimer: measured mass 2385.25 $\pm$ 0.010 g/mol (expected mass, 2385.2442 g/mol).

### Thioflavin fluorescence

The thioflavin T (ThT) test [Nilsson 2004], is a simple test for detecting the presence of amyloid fibers. An increase in fluorescence intensity is observed when the amyloid-specific dye ThT binds to the aggregate, thus indicating amyloid fibril formation.

The results obtained are shown in Figure 5.12, a clear increase of the fluorescence intensity is displayed for peptide **2** after the binding of ThT whereas no increase is observed for peptide

**6**, with the Cys to Ser mutation, under the same conditions, where the peptides **2** and **6** incubated during 1 day with 10% DMSO to induce the fibril formation.

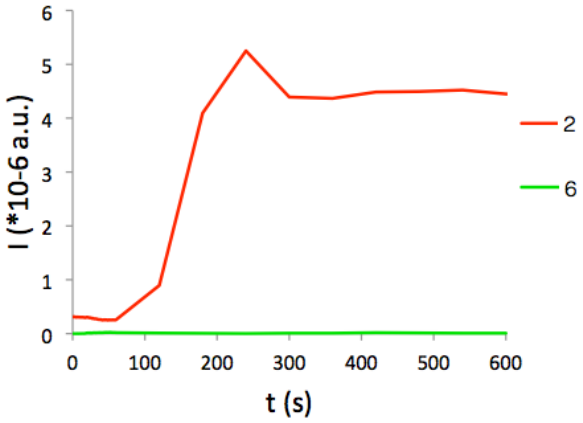



Figure 5.12: Time evolution of the fluorescent signal after substraction of the buffer signal, for peptide **2** and peptide **6**.

### Circular Dichroism

Far UV Circular Dichroism (CD) was used to study the presence of secondary structure [Kelly 2005]. Obtained results are shown in figure 5.13, the CD spectra showed markedly different profiles for peptide **2** and **6** under the same conditions Figure 5.13. The spectrum for **2** shows the typical profile of amyloid fibrils, (high $\beta$-sheet content) [Cascio 2014]. However, the spectra for **6** and **2**, in presence of TCEP, correspond to disordered peptides, a random coil structure.

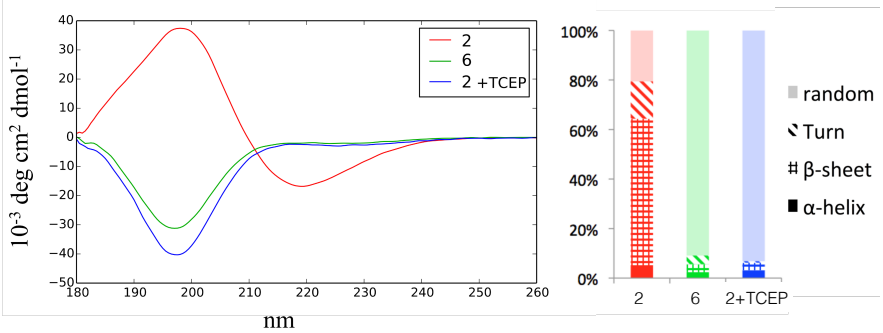



Figure 5.13: A) CD spectra of peptides **2**, **6** and of peptide **2** under reducing conditions; B) Secondary Structure deconvolution for the CD spectra. All experiments measured at 25°C in water, at neutral pH.

### Transmission Electron Microscopy

The aggregated structure was analyzed by electron microscopy (TEM). In Figure 5.14, the images clearly show that the solid appearing in the solution under oxidative conditions is composed of long, thin, cylindrical and unbranched fibers. From these pictures (Figure 5.14 and 5.15), an analysis is performed of the fibers diameter, obtaining a value of $5.4 \pm 1.6\,\text{nm}$. A featureless cylinder is typically observed (Figure 5.14 A-C) but in many instances fibers tend to join together into larger assemblies, as if several fibers join laterally to form a larger ribbon-like structure (Figure 5.14 D and E). Under oxidative conditions, but with agitation during the incubation period, the fibers appear less elongated, as broken, with larger lateral aggregates (Figure 5.14 B).

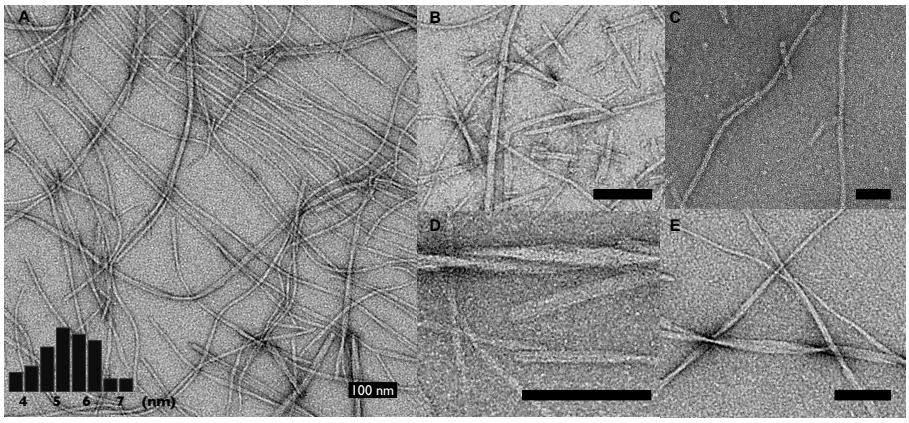



Figure 5.14: EM photographs of the fibrils. All pictures present a 100 nm scale bar. A) fibrils obtained from peptide **2** in resting conditions, and B) under agitation; C) fibrils obtained from peptide **7**; D-E) ribbon-like structures observed for peptide **2**; inset: distance histogram measured on several images of peptide **2**.

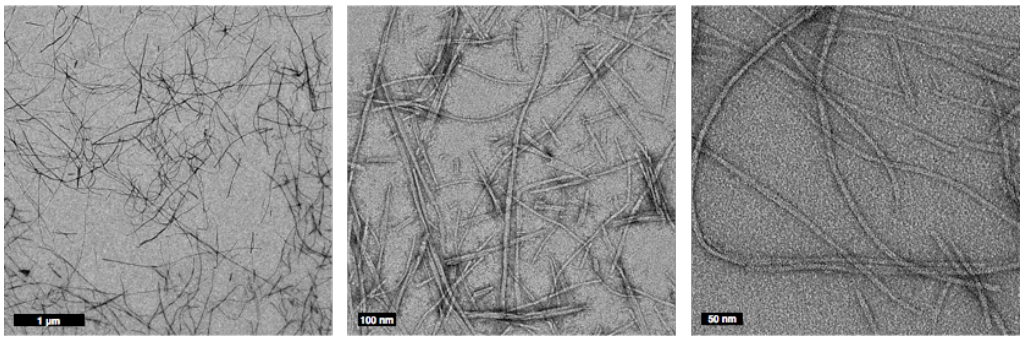



Figure 5.15: Additional EM views. left, wider view showing fiber length; center, fibers obtained after shaking; right, image at higher resolution

### Small Angle X-ray Scattering

The aggregation process was also monitored by small-angle X-ray scattering (SAXS). Figure 5.16 show the evolution of the SAXS curves over time indicating that large objects accumulate in the sample during the experiment. The curves present a typical pattern for an elongated cylinder as indicated by the $q^{-1}$ slope in the intermediate $q$ range.

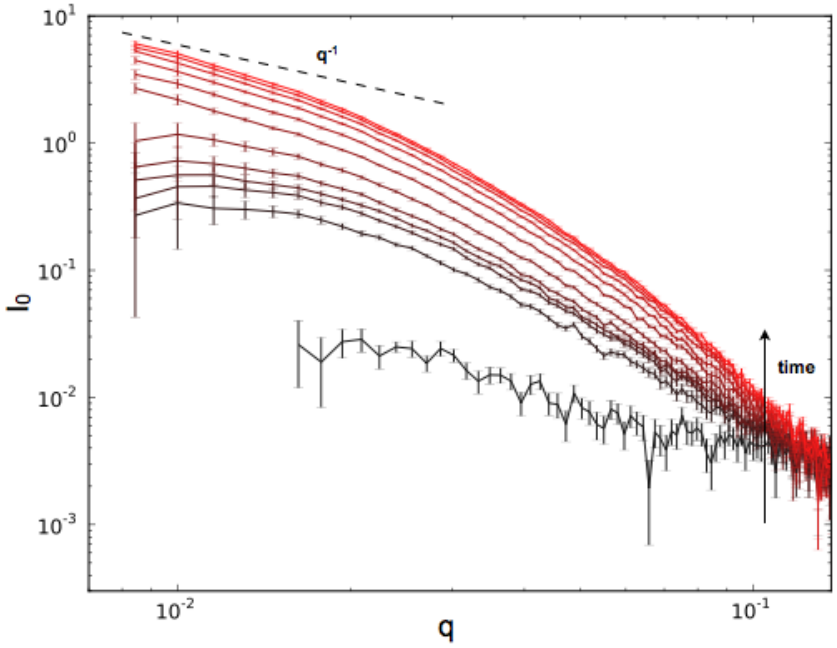



Figure 5.16: Fibril formation kinetics followed by SAXS on a 1 mg/ml peptide solution after 5% DMSO addition. Two hours exposures were collected sequentially during 24 hours and the solvent background was subtracted.

Data were analyzed using the program package PRIMUS [Konarev 2003]. The radius of gyration ($R_g$) and the cross-section radius of gyration ($R_{XC}$) were evaluated using the Guinier approximation, see table 5.2 and Figure 5.17. Proper range in $q$ for the $R_g$ determination could not be used because of low signal to noise and the size of the considered objects. The cross section $R_{XC}$ can be estimated in the intermediate $q$ region, the evolution of this value is given in the inset. The cross section radius of gyration of a homogeneous cylinder of radius $r$ is $\frac{\sqrt{2}}{2}r$, in consequence, the measured value $R_{XC} = 2.26 \pm 0.1$ nm indicates a cylinder of a diameter $6.4 \pm 0.3$ nm in agreement with EM measurements considering the precision provided by the picture analysis.

| Number of experiment | t (h) | Rg (Å) | Io | qRg limits | | Rxc (nm) | qRxc limits | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 41.1 | 0.034 | 1.239 | 1.995 | 1.4 | 0.831 | 1.082 |
| 2 | 4 | 67.2 | 0.388 | 0.671 | 1.923 | 1.89 | 0.951 | 1.597 |
| 3 | 6 | 68.3 | 0.551 | 0.576 | 1.954 | 2.16 | 1.087 | 1.590 |
| 4 | 8 | 69.8 | 0.661 | 0.589 | 1.889 | 2.23 | 0.984 | 1.676 |
| 5 | 10 | 75.2 | 0.884 | 0.634 | 1.801 | 2.26 | 1.173 | 1.664 |
| 6 | 12 | 80 | 1.36 | 0.678 | 1.676 | 2.32 | 1.204 | 1.816 |
| 7 | 14 | 89.6 | 2.42 | 0.756 | 1.868 | 2.27 | 1.319 | 1.847 |
| 8 | 16 | 91.1 | 3.27 | 0.768 | 1.899 | 2.22 | 1.428 | 1.945 |
| 9 | 18 | 93.2 | 4.21 | 0.786 | 1.943 | 2.17 | 1.463 | 1.901 |
| 10 | 20 | 97.3 | 5.08 | 0.972 | 1.878 | 2.27 | 1.425 | 2.094 |
| 11 | 22 | 106 | 6.42 | 0.894 | 1.881 | 2.33 | 1.498 | 1.932 |
| 12 | 24 | 107 | 7.21 | 0.902 | 1.899 | 2.27 | 1.425 | 2.059 |

Table 5.2: $I_o$, $R_g$ and $R_{XC}$ values obtained for SAXS analysis. $qR_g$ and $qR_{XC}$ limits used for the Guinier determination are indicated.

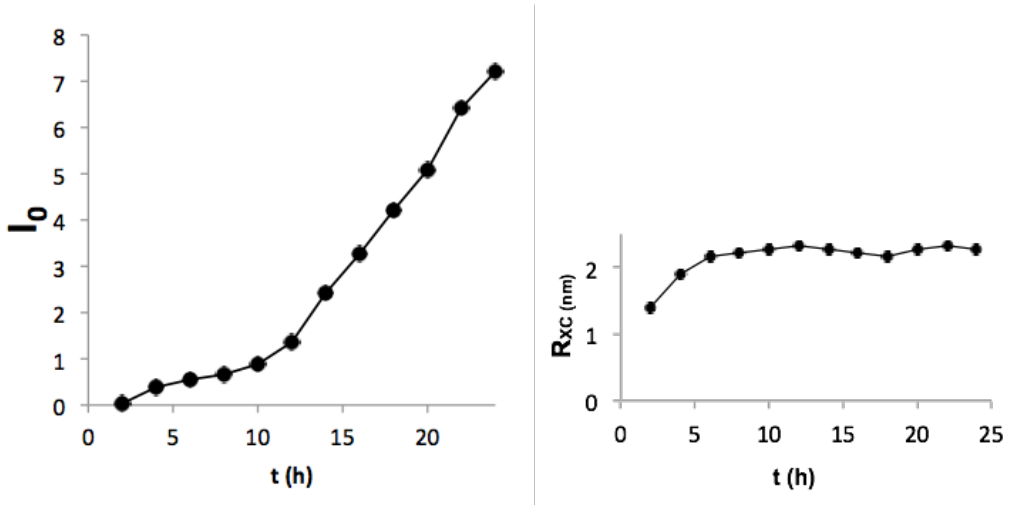



Figure 5.17: $I_o$ evolution (a.u.) vs time and $R_{XC}$ vs time , as determined from the Guinier analysis presented in table 5.2.

These results show unequivocally that the peptides generated from this conserved sequence of the NTD-AR undergo reversible amyloid formation, whereas the free peptides are mostly random. Aggregation is triggered by oxidative formation of a disulfide dimer, thus dimerization of peptides through a disulfide bond is the key event for further fibril formation.

### 5.3.1 Influence of the oxidation conditions to the kinetics aggregation

It is important to study the effect caused on the kinetics rate of the aggregation by using different oxidant conditions. For this purpose, two different conditions were tested, the first

one changing the pH to pH = 8, and the second one also at pH 8.0 but using $H_2O_2$ + NaI. as oxidant. The aim of the pH increasing is to favor the thiolate formation for the Cys residue. $P-SH \xrightarrow{pH8} P-S^-$

A thiol with $pK_a$10 presents 0.1% thiolate form at pH 7, whereas 1% of thiol is in thiolate form at pH 8 [Singh 1993]. The thiolate is a stronger nucleophile than the thiol group and it favors the disulfide bridge formation. This will allow us to see the pH influence on the dimerization step and globally in the kinetics aggregation. With the same purpose we added $H_2O_2$ + NaI, that represents stronger oxidative conditions than using DMSO.

### pH 8.0 + 10% DMSO

No big differences compared with pH 7 were obtained for the aggregation kinetics, as shown in Figure 5.18, with 10% DMSO it is obtained a disappearance of the signal in around 24h.

a) b)

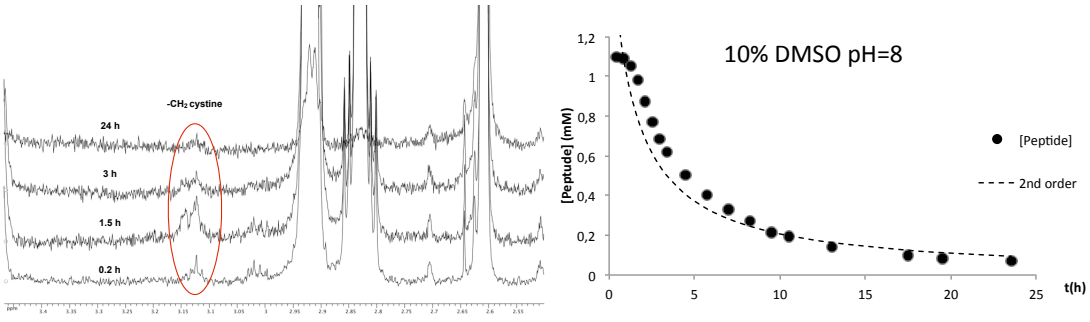



Figure 5.18: a) 1D $^1$H spectrum at 0, 1.5, 3 and 24 h after adding 10% DMSO and pH 8.0 b) Kinetics representations of the signal disappearance from the NMR spectrum and the second-order fit.

While looking into the kinetics we don't observe big differences with the standard conditions used and described above. However, the appearance of the cystine $H_\beta$, circled in red in Figure 5.18, is remarkable. This proton is shifted when the disulfide bridge is formed for the peptide dimerization. On the other hand the kinetics does not fit perfectly to a second-order kinetics. A nucleation step can be also observed at the beginning of the curve. In this case, we can conclude that the increasing of the pH affects the kinetics behavior even while no effect is shown on the global duration of the aggregation. The differences in the kinetics are also emphasized with the fact that the dimer formation is observed by NMR, meaning that the dimerization is faster here allowing to see the dimer appearing before a later disappearance after aggregation.

### pH 8.0 + $H_2O_2$

In the same purpose, DMSO was substituted by $H_2O_2$ + NaI. $H_2O_2$ was used to activate the oxidation process in order to favor the dimerization and further fibrillation. For the reaction, 1 eq$H_2O_2$ + 0.01 eqNaI were used as described in [Kirihara 2007].

a) b)

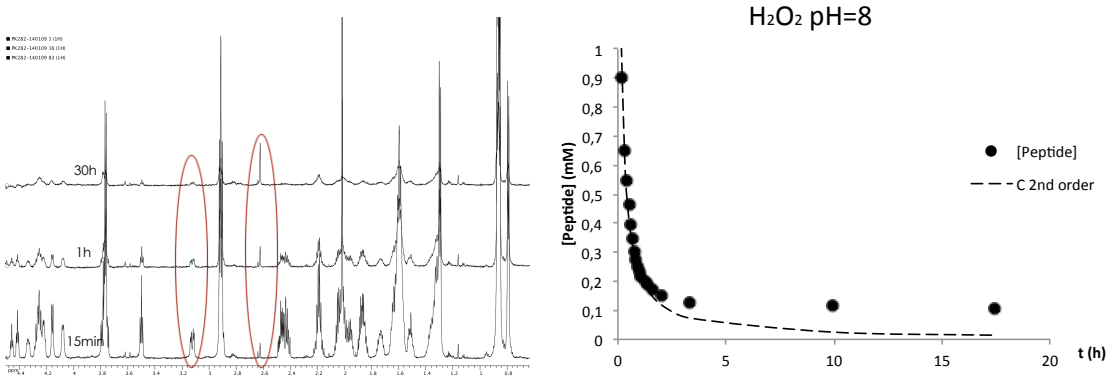



Figure 5.19: a) 1D $^1$H spectrum at 15 min, 1, and 30 h after adding $H_2O_2$. b) Kinetics representations of the signal disappearance from the NMR spectrum and the second-order fit.

Figure 5.19 shows clearly the difference in this kinetics study. A faster kinetics rate is observed, in around 5 h the peptide concentration has reached the plateau of 0.1 mM (Figure 5.19 b). Looking into the NMR spectra, apart from a *immediate* dimer appearance, the oxidation of the Met can also be observed with the appearing of a $^1$H signal at 2.65 ppm (circled in red in Figure 5.19 a). This oxidative conditions appear to be too strong for the peptide fibrillation.

## 5.3.2 Other constructs studied

After the first results on the kinetics aggregation for peptide **2**, and looking into the conservation study performed in the N-terminal domain of AR shown in Figure 5.4. It was considered appropriate to study in detail the influence of the sequence length for the aggregation process, we wondered if there is a peptide shorter than peptide **2** that could also aggregates under oxidative conditions or if on the opposite, there is a maximum on the length of the native conserved region that could not further aggregate.

For this reason, four peptides with different lengths around the residue C240 (from the highest conserved region in the NTD-AR), where chosen to study their behavior by NMR

under oxidative conditions.

Table 5.3: Aggregation study of different length sequences from the highest conserved residues of the N-terminal domain of the Androgen Receptor.

| | Peptide | Dimerization | Aggregation |
|---|---|---|---|
| **2** | KELCKAVSVSM | | X |
| **6** | KELSKAVSVSM | – | – |
| **7** | KELCKAVSVSMGLGVEALEHLS | | X |
| **8** | KELCKAV | X | |
| **9** | KELCKAVSV | X | |
| **10** | TISDNAKELCKAVSVSMGLGVEALEHLS | | X |

Table 5.3 summarizes the results obtained for the different peptides chosen for this study, into *dimerization* or *aggregation*, depending of the behavior displayed for the different amino acids sequences. The classification into *aggregation* also implies dimerization through disulfide bridge formation.

All peptides shorter than peptide **2** did not show any aggregation, whereas all longer peptides exhibited aggregation, with kinetics equivalent to the one studied for peptide **2**. The mutated sequence **6** Cys to Ser, didn't show any effect as expected where removing the cysteine and no disulfide bridge can be formed. The shortest peptides, **8** and **9**, only dimerized in the presence of DMSO, and the dimer remained stable in solution after more than 24 hours.

## 5.4 Conclusions

The results described in this chapter show unequivocally that the peptides generated from this conserved sequence of the NTD-AR undergo reversible amyloid formation. Peptide aggregation is triggered by the oxidative conditions, where the peptide dimerization, through a disulfide bridge formation, is the key point for the fibrils formation.

Chemical triggered fibril self-assembly has already been reported, either by changing the hydrophobic behavior [Abul-Haija 2013] or by using a template strategy [Rubinov 2012], but to our knowledge, not by oxidation. Aggregations of fiber structure, in response to disulfide formation, are known in the literature [Rubinov 2009, Carnall 2010], however the mechanism involved here seems different. The primary sequence (KELCKAVSVSM), alternates hydrophobic and hydrophilic residues, which is common in proteins prone to fibril formation. Looking into the extended primary sequence of peptide **10** (TISDNAKELCKAVSVSMGLGVEALEHLS),the hydrophobic-hydrophilic alternance continues. It is probably the position of an active site.

It was shown that the mutation of some hydrophobic residues within this sequence have a dramatic effect on the receptor activity [Betney 2003].

The kinetics displays a second-order behavior with respect to peptide concentration suggesting that dimerization is the limiting step. However, pseudo second-order kinetics are also known in sorption processes not involving dimerisation [Ho 1999], and a more complex fibrillation mechanism cannot be ruled out as disulfide bridge self-assembling systems are known to display complex evolution patterns [Carnall 2010]. The mechanism of proteins aggregation into amyloid fibrils usually involves a "lag phase" or nucleation step before the aggregation process starts. However the absence of a "lag phase" can be observed due to the presence of pre-formed fibrils [Nilsson 2004]. Additionally, the decay levels-off to a plateau around 0.1 mM suggests a possible equilibrium between the fibril and the free dimer.

The results obtained for peptides **8** and **9**, where only a dimerization process occurs, suggest that the presence of the Methionine in peptide **2**, could have an important role in the aggregation mechanism. The precise molecular structure of the dimer and the fibril is not yet determined and is currently under investigation and is included in the perspective work for this project.

The NTD of AR plays the central role for the activity of the receptor. It contains the activation domain AF1, which is necessary for full in vivo transcriptional activity of the receptor [Lavery 2008a]. As for many transcription factors, it has been shown to be in an intrinsically disordered conformation [Lavery 2008b], and to undergo allosteric disorder/order transition through its interactions with specific coregulatory proteins [Kumar 2012, McEwan 2012].

In humans, a PolyQ longer than 40 residues leads to androgen insensitivity and Spinal Bulbar Muscular Atrophy (SBMA); while a PolyQ shorter than 20 residues increases the risk of prostate cancer [Kumar 2011]. The PolyQ is known to be involved in AR aggregates [Palazzolo 2008]. Nevertheless, this study open new possibilities and the hypothesis propose here is that this conserved sequence could regulate the activity of the androgen receptor by redox control. Presenting a reduced active formed, and an oxidate less active

On the other hand, this conserved motif from the NTD-AR has been studied previously and it is involved on the interaction with the heat-shock protein HSP70 [He 2004]. Finally, AR has already been reported to form disulfide dimers in cells, and to be responsive to redox treatment [Liao 1999].

## 5.5 Publication

The work presented in this chapter resulted in a publication in ChemBioChem, Volume 15, Issue 16, pages 2370–2373, November 3, 2014.

# Bibliography

[Abul-Haija 2013] Yousef M Abul-Haija, Sangita Roy, Pim W J M Frederix, Nadeem Javid, Vineetha Jayawarna and Rein V Ulijn. *Biocatalytically Triggered Co-Assembly of Two-Component Core/Shell Nanofibers*. Small, page doi:10.1002/smll.201301668, September 2013. (Cited on pages 15 and 100.)

[Asencio-Hernández 2014] Julia Asencio-Hernández, Christine Ruhlmann, Alastair McEwen, Pascal Eberling, Yves Nominé, Jocelyn Céraline, Jean-Philippe Starck and Marc-André Delsuc. *Reversible amyloid fiber formation in the N terminus of androgen receptor.* ChemBioChem, vol. 15, no. 16, pages 2370–2373, November 2014. (Cited on pages 17 and 89.)

[Betney 2003] R Betney and Iain J McEwan. *Role of conserved hydrophobic amino acids in androgen receptor AF-1 function.* Journal of Molecular Endocrinology, vol. 31, no. 3, pages 427–439, December 2003. (Cited on page 101.)

[Carnall 2010] Jacqui M A Carnall, Christopher A Waudby, Ana M Belenguer, Marc C A Stuart, Jérôme J.-P. Peyralans and Sijbren Otto. *Mechanosensitive Self-Replication Driven by Self-Organization*. Science, vol. 327, no. 5972, pages 1502–1506, March 2010. (Cited on pages 15, 100 and 101.)

[Cascio 2014] Michael Cascio, Paul A Glazer and B A Wallace. *The secondary structure of human amyloid deposits as determined by circular dichroism spectroscopy.* Biochem. Biophys. Res. Commun., vol. 162, no. 3, pages 1162–1166, January 2014. (Cited on page 94.)

[Dubbink 2004] Hendrikus J Dubbink, Remko Hersmus, Chandra S Verma, Hetty A G M van der Korput, Cor A Berrevoets, Judith van Tol, Angelique C J Ziel-van der Made, Albert O Brinkmann, Ashley C W Pike and Jan Trapman. *Distinct Recognition Modes of FXXLF and LXXLL Motifs by the Androgen Receptor*. Molecular Endocrinology, vol. 18, no. 9, pages 2132–2150, September 2004. (Cited on page 84.)

[Garber 2005] Peter M Garber, Genevieve M Vidanes and David P Toczyski. *Damage in transition*. Trends in Biochemical Sciences, vol. 30, no. 2, pages 63–66, February 2005. (Cited on page 84.)

[Gottlieb 1998] B Gottlieb, H Lehvaslaiho, L K Beitel, R Lumbroso, L Pinsky and M Trifiro. *The Androgen Receptor Gene Mutations Database.* Nucleic Acids Res., vol. 26, no. 1, pages 234–238, January 1998. (Cited on pages 32 and 84.)

[Gottlieb 2012] Bruce Gottlieb, Lenore K Beitel, Abbesha Nadarajah, Miltiadis Paliouras and Mark Trifiro. *The androgen receptor gene mutations database: 2012 update*. Hum. Mutat., vol. 33, no. 5, pages 887–894, March 2012. (Cited on page 84.)

[He 2000] B He, J A Kemppainen and E M Wilson. *FXXLF and WXXLF Sequences Mediate the NH2-terminal Interaction with the Ligand Binding Domain of the Androgen Receptor*. Journal of Biological Chemistry, vol. 275, no. 30, pages 22986–22994, July 2000. (Cited on pages 26 and 84.)

[He 2004] B He, S Bai, A T Hnat, R I Kalman, J T Minges, C Patterson and E M Wilson. *An Androgen Receptor NH2-terminal Conserved Motif Interacts with the COOH Terminus of the Hsp70-interacting Protein (CHIP)*. Journal of Biological Chemistry, vol. 279, no. 29, pages 30643–30653, July 2004. (Cited on page 101.)

[Heery 1997] D M Heery, E Kalkhoven, S Hoare and M G Parker. *A signature motif in transcriptional co-activators mediates binding to nuclear receptors.* Nature, vol. 387, no. 6634, pages 733–736, June 1997. (Cited on page 84.)

[Ho 1999] Yuh-Shan Ho and G McKay. *Pseudo-second order model for sorption processes*. Process Biochem., vol. 34, no. 5, pages 451–465, 1999. (Cited on page 101.)

[Kelly 2005] Sharon M Kelly, Thomas J Jess and Nicholas C Price. *How to study proteins by circular dichroism.* Biochim. Biophys. Acta, Proteins Proteomics, vol. 1751, no. 2, pages 119–139, 2005. (Cited on pages 44 and 94.)

[Kirihara 2007] Masayuki Kirihara, Yasutaka Asai, Shiho Ogawa, Takuya Noguchi, Akihiko Hatano and Yoshiro Hirai. *A Mild and Environmentally Benign Oxidation of Thiols to Disulfides*. Synthesis, vol. 2007, no. 21, pages 3286–3289, November 2007. (Cited on page 99.)

[Konarev 2003] Petr V Konarev, Vladimir V Volkov, Anna V Sokolova, Michel HJ Koch and Dmitri I Svergun. *PRIMUS: a Windows PC-based system for small-angle scattering data analysis*. J. Appl. Crystallogr., vol. 36, no. 5, pages 1277–1282, 2003. (Cited on page 96.)

[Kumar 2011] R Kumar, H Atamna, M N Zakharov, S Bhasin, S H Khan and R Jasuja. *Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy*. Life Sciences, vol. 88, no. 13-14, pages 565–571, March 2011. (Cited on pages 28 and 101.)

[Kumar 2012] Raj Kumar and Iain J McEwan. *Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation.* Endocr. Rev., vol. 33, no. 2, pages 271–299, April 2012. (Cited on pages 35 and 101.)

[Lavery 2008a] Derek N Lavery and Iain J McEwan. *Functional Characterization of the Native NH 2-Terminal Transactivation Domain of the Human Androgen Receptor: Binding Kinetics for Interactions with TFIIF and SRC-1a*. Biochemistry, vol. 47, no. 11, pages 3352–3359, March 2008. (Cited on page 101.)

[Lavery 2008b] Derek N Lavery and Iain J McEwan. *Structural Characterization of the Native NH 2-Terminal Transactivation Domain of the Human Androgen Receptor: A Collapsed Disordered Conformation Underlies Structural Plasticity and Protein-Induced Folding*. Biochemistry, vol. 47, no. 11, pages 3360–3369, March 2008. (Cited on pages 27 and 101.)

[Liao 1999] M Liao, Z X Zhou and E M Wilson. *Redox-dependent DNA binding of the purified androgen receptor: evidence for disulfide-linked androgen receptor dimers*. Biochemistry, vol. 38, no. 30, pages 9718–9727, July 1999. (Cited on page 101.)

[McEwan 2012] Iain J McEwan. *Nuclear hormone receptors: Allosteric switches*. MOLECULAR AND CELLULAR ENDOCRINOLOGY, vol. 348, no. 2, pages 345–347, January 2012. (Cited on page 101.)

[Morris 2009] Aimee M Morris, Murielle A Watzky and Richard G Finke. *Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature*. BBA - Proteins and Proteomics, vol. 1794, no. 3, pages 375–397, March 2009. (Cited on page 92.)

[Nilsson 2004] Melanie R Nilsson. *Techniques to study amyloid fibril formation in vitro.* Methods, vol. 34, no. 1, pages 151–160, September 2004. (Cited on pages 44, 93 and 101.)

[Palazzolo 2008] Isabella Palazzolo, Alessandra Gliozzi, Paola Rusmini, Daniela Sau, Valeria Crippa, Francesca Simonini, Elisa Onesto, Elena Bolzoni and Angelo Poletti. *The role of the polyglutamine tract in androgen receptor.* Journal of Steroid Biochemistry and Molecular Biology, vol. 108, no. 3-5, pages 245–253, February 2008. (Cited on page 101.)

[Rubinov 2009] Boris Rubinov, Nathaniel Wagner, Hanna Rapaport and Gonen Ashkenasy. *Self-Replicating Amphiphilic $\beta$-Sheet Peptides*. Angew. Chem. Int. Ed., vol. 48, no. 36, pages 6683–6686, August 2009. (Cited on pages 15 and 100.)

[Rubinov 2012] Boris Rubinov, Nathaniel Wagner, Maayan Matmor, Oren Regev, Nurit Ashkenasy and Gonen Ashkenasy. *Transient Fibril Structures Facilitating Nonenzymatic Self-Replication*. ACS Nano, vol. 6, no. 9, pages 7893–7901, September 2012. (Cited on page 100.)

[Singh 1993] R Singh and G M Whitesides. *Thiol-disulfide interchange*. Supplement, 1993. (Cited on page 98.)

[Tam 1991] James P Tam, Cui Rong Wu, Wen Liu and Jing Wen Zhang. *Disulfide bond formation in peptides by dimethyl sulfoxide. Scope and applications*. J. Am. Chem. Soc., vol. 113, no. 17, pages 6657–6662, 1991. (Cited on page 88.)

# Appendix B

## B.1 Bioinformatic study

### Disorder and Secondary Structure prediction

Here are described the different bioinformatic programs that were used for the disorder and secondary structure analysis.

- **metaPrDOS** meta Protein DisOrder prediction System, metaPrDOS is a meta server to predict natively disordered regions of a protein chain from its amino acid sequence [Ishida 2008]. metaPrDOS returns disorder tendency of each residue as prediction results.

  metaPrDOS does not predict disordered regions from amino acid sequence directly, but predicts disordered regions by integrating the results of different prediction methods, using eight prediction systems to predict disordered regions:

  PrDOS [Ishida 2007]

  DISOPRED2 [Ward 2004]

  DisEMBL $^{TM}$[Linding 2003]

  DISPROT (VSL2) [Peng 2006]

  DISpro [Cheng 2005]

  IUpred [Dosztanyi 2005]

  POODLE-S [Shimizu 2007]

  DISOclust [McGuffin 2008]

- **NPS@** Network Protein Sequence @nalysis, NPS@ is an interactive Web server dedicated to protein sequence analysis that can be used for secondary structure prediction.

- **MeDor** MeDor [Lieutaud 2008] provides, among other things, Hydrophobic Cluster Analysis (HCA). MeDor is conceived to provide a global overview of various predictions, and to speed up the disorder prediction.

### Alignment

The Androgen Receptor protein sequence alignment was obtained using Jalview [Waterhouse 2009] among 12 vertebrate species: *Homo sapiens, Rattus norvegicus, Mus musculus, Canis lupus, Equus caballus, Loxodonta africana, Gallus gallus, Xenopus laevis, Oryzias latipes, Gasterosteus aculeatus, Danio rario, and Squalus acanthias*. Conservation value was computed with Jalview, and corresponds to the Quality value returned by the program (normalized to 100% for no mutation, and smoothed with a moving average of five residues).

## B.2 Biophysical characterization

## Peptides

Peptide Synthesis was performed locally on a 433A Peptide Synthesizer (ABI) using Fmoc chemistry. Crude peptides were purified by reverse phase HPLC on a 250 X 21.2 mm Luna C18 100 Å, 5 $\mu$m column from Phenomenex, and controlled by Mass Spectrometry (ESI/TOF micrOTOF II Bruker).

The peptides were desalted with $H_2O$ using a GF Superdex 10/300 column in an AKTA system, to remove the TFA (trifluoroacetic acid) coming from the purification process and lyophilized after that.

The peptides were stored as lyophilized powder at -20 °C. Stock solutions were prepared with a final concentration of 5 mM with $H_2O$ Milli Q and 1 mM of $NaN_3$.

## NMR experiments

All NMR studies (except the TCEP recovery) were performed on a Bruker Avance III 700 MHz spectrometer equipped with a Z-gradient cryoprobe at 25 °C. The spectra were processed and analyzed using NMR-notebook® (NMRTEC Illkirch France). $^1$H 1D spectrum, TOCSY and ROESY 2D experiments were performed in order to assign the peptide spectrum. A series of $^1$H spectra were recorded after the addition of 1%, 5%, and 10% dimethyl sulfoxide (DMSO) to the NMR solution. The sample consisted of 180$\mu$l of 0.5 mM peptide in 20 mM phosphate buffer pH = 6.4 with 10% $D_2O$ in 3 mm tubes. Concentration of the soluble peptide forms during the kinetics was measured by the integration of the methyl region ($[0.8\text{ppm} - 1.0\text{ppm}]$) of the NMR spectrum. The TCEP spectra were performed on a Bruker Avance III operating at 500 MHz, using a TBI probe.

The sequence used for the 1D $^{1}$H spectra was `zgpr`. The sequence used for 2D $^{1}$H-$^{1}$H TOCSY spectra was `dipsi2phpr` with a mixing time of 80 ms at 12.3 dB. The sequence used for 2D $^{1}$H-$^{1}$H ROESY spectra was `croesyphpr` with a mixing time of 300 ms at 20 dB.

The 90° pulse for all the experiments presented in Chapter 5 is comprised between 8-11 $\mu$s at 3.2 dB.

The peptide quantification was performed using the Tryptophan concentration as reference as explained in [Wider 2006]. The concentration of the stocksolution of Tryptophan is 6.3 mM, dissolved in water, it was measured by optical density ($\varepsilon_{280(TRP)}$= 5600 cm$^{-1}$M$^{-1}$ at 298 K).

The experiments were conducted at 298K, on 3mm tubes filled with 180 $\mu$l of sample and inserted in a 700 MHz Bruker spectrometer equipped with a TXI cryo-probe. The sample was prepared in water with 10% $D_2O$. The concentration of Tryptophan in the solution was again checked by DO, resulting at 1.2 mM. The 1D proton experiment employed used a D1 relaxation delay of 5 s (T1 of tryptophan = 3 s) and a 90° pulse of 9.75 $\mu$s at 3.2 dB.

## Mass Spectrometry

The precipitate from the NMR tube was dissolved in 1:1 (v/v) Acetonitrile-Water with 2% formic acid and analyzed by Mass Spectrometry (ESI/TOF micrOTOF II Bruker).

## Circular Dichroism

Far UV Circular Dichroism (CD) experiments were recorded on a Jasco J-815 spectropolarimeter (Easton, MD) equipped with an automatic 6-position Peltier thermostated cell holder. The instrument was calibrated with 10-camphorsulphonic acid. CD data were collected in the 180-260 nm range, by 0.5 nm steps using a 0.1 mm path-length cell (Quartz-Suprasil, Hellma UK Ltd) at 25.0 $\pm$ 0.1 °C. Spectra were acquired using a continuous scan rate of 100 nm/min and are presented as an average of 10 successive scans. The response time and the bandwidth were 1.0 sec and 1 nm, respectively. All samples were studied in pure water, at a concentration of 1 mM. As DMSO cannot be used in CD experiments, all samples were incubated for 24 hours at 37 °C. The spectra were corrected by subtracting the solvent spectrum obtained under identical conditions. The signal is expressed in mean residue ellipticity (deg cm$^2$ dmol$^{-1}$).

Secondary structure content was estimated with the CDPro suite [Sreerama 2000] using the CONTIN/LL method. CONTIN/LL fits the CD spectra by comparison to a linear combination of the spectra of a large database of proteins with known conformations. Several databases were tested, and the SDP48 database was chosen, as it provided the smaller RMSD. Obtained

results are given in table B.1, and shown in figure B.1

The spectra have been deposited in the Protein Circular Dichroism Data Bank website (http://pcddb.cryst.bbk.ac.uk) with the accession codes CD0004480000 to CD0004482000.

Table B.1: deconvolution of the CD spectra

| | random | turn | $\beta$-sheet | $\alpha$-helix | rmsd |
|---|---|---|---|---|---|
| **2** | 0.10 | 0.44 | 0.40 | 0.05 | 0.065 |
| **2**+TCEP | 0.93 | 0.02 | 0.02 | 0.03 | 0.016 |
| **6** | 0.94 | 0.02 | 0.02 | 0.03 | 0.01 |

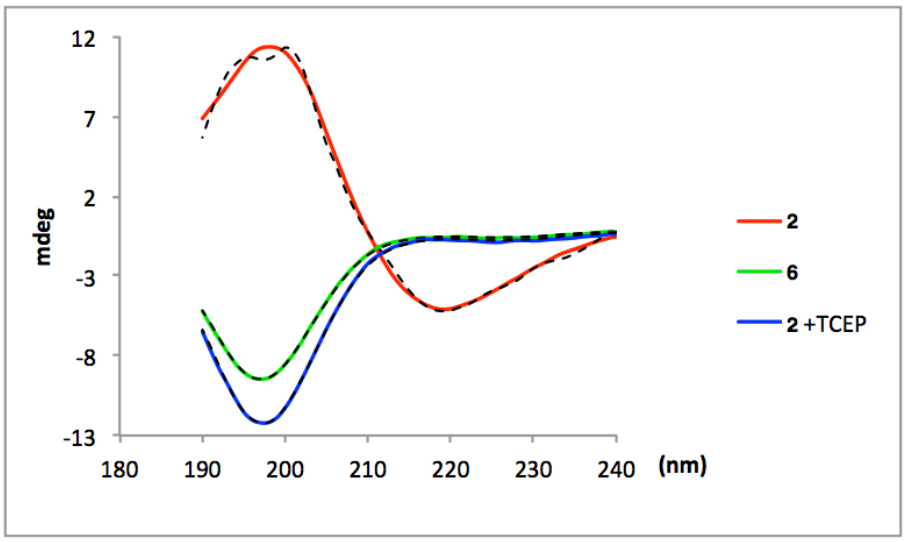



Figure B.1: Experimental CD curves (solid lines) and reconstructed curves (dashed lines) from the deconvolution, showing the goodness of fit.

## Thioflavin fluorescence

The experiments were recorded on a Fluorolog Jobin Yvon spectrofluorometer, without shaking, with excitation at 440 nm and emission at 482 nm. The samples for peptide **2** and **6** were prepared with the same conditions than the for the CD analysis. The measurements were performed during 600 s.

## Electron Microscopy

The same samples used for the NMR studies were analyzed under the Electron Microscope. 5 $\mu$l were deposited onto a glow-discharged, carbon-coated copper electron microscopy grid, stained with 1% uranyl acetate and observed under a Philips CM120 transmission electron microscope equipped with a LaB6 filament and operating at 100 kV.

## Small Angle X-ray Scattering

Small Angle X-ray Scattering (SAXS) data were collected on a BioSAXS-1000 system (Rigaku) coupled to a MicroMax-007HF X-ray generator (Rigaku) using Cu $K_\alpha$ radiation with $\lambda = 1.5418$ Å and a Pilatus 100 K detector (Dectris). 30 $\mu$l of water with 5% DMSO were injected into a sample holder fitted with a 0.1 mm thin wall quartz capillary and a two hour long exposure was collected. DMSO was added to the peptide at 1 mg/ml and 30 $\mu$l were injected into the same sample holder and twelve exposures were collected sequentially.

# Bibliography

[Cheng 2005] J L Cheng, M J Sweredoski and P Baldi. *Accurate prediction of protein disordered regions by mining protein structure data*. Data Mining and Knowledge Discovery, vol. 11, no. 3, pages 213–222, November 2005. (Cited on page 107.)

[Dosztanyi 2005] Z Dosztanyi, V Csizmok, P Tompa and I Simon. *The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins*. Journal of Molecular Biology, vol. 347, no. 4, pages 827–839, 2005. (Cited on page 107.)

[Ishida 2007] Takashi Ishida and Kengo Kinoshita. *PrDOS: prediction of disordered protein regions from amino acid sequence.* Nucleic Acids Res., vol. 35, no. Web Server issue, pages W460–4, July 2007. (Cited on page 107.)

[Ishida 2008] Takashi Ishida and Kengo Kinoshita. *Prediction of disordered regions in proteins based on the meta approach.* Bioinformatics, vol. 24, no. 11, pages 1344–1348, June 2008. (Cited on page 107.)

[Lieutaud 2008] Philippe Lieutaud, Bruno Canard and Sonia Longhi. *MeDor: a metaserver for predicting protein disorder*. BMC Genomics, vol. 9, no. Suppl 2, page S25, 2008. (Cited on page 107.)

[Linding 2003] R Linding, L J Jensen, F Diella, P Bork, T J Gibson and R B Russell. *Protein disorder prediction: Implications for structural proteomics*. Structure/Folding and Design, vol. 11, no. 11, pages 1453–1459, November 2003. (Cited on page 107.)

[McGuffin 2008] Liam J McGuffin. *Intrinsic disorder prediction from the analysis of multiple protein fold recognition models*. Bioinformatics, vol. 24, no. 16, pages 1798–1804, 2008. (Cited on page 107.)

[Peng 2006] Kang Peng, Predrag Radivojac, Slobodan Vucetic, A Keith Dunker and Zoran Obradovic. *Length-dependent prediction of protein intrinsic disorder.* BMC Bioinformatics, vol. 7, page 208, 2006. (Cited on page 107.)

[Shimizu 2007] Kana Shimizu, Shuichi Hirose and Tamotsu Noguchi. *POODLE-S: web application for predicting protein disorder by using physicochemical features and reduced amino acid set of a position-specific scoring matrix*. Bioinformatics, vol. 23, no. 17, pages 2337–2338, 2007. (Cited on page 107.)

[Sreerama 2000] Narasimha Sreerama and Robert W Woody. *Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set*. Anal. Biochem., vol. 287, no. 2, pages 252–260, December 2000. (Cited on page 109.)

[Ward 2004] J J Ward, J S Sodhi, L J McGuffin, B F Buxton and D T Jones. *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life*. Journal of Molecular Biology, vol. 337, no. 3, pages 635–645, 2004. (Cited on page 107.)

[Waterhouse 2009] A M Waterhouse, J B Procter, D M A Martin, M Clamp and G J Barton. *Jalview Version 2–a multiple sequence alignment editor and analysis workbench*. Bioinformatics, vol. 25, no. 9, pages 1189–1191, April 2009. (Cited on page 108.)

[Wider 2006] Gerhard Wider and Lars Dreier. *Measuring Protein Concentrations by NMR Spectroscopy*. J. Am. Chem. Soc., vol. 128, no. 8, pages 2571–2576, March 2006. (Cited on page 109.)

Chapter 6

# Small molecules interactions study with a high conserved peptide of the N-terminal domain of the Androgen Receptor



## 6.1 NMR methodologies for the study of protein-ligand interactions

Nuclear magnetic resonance (NMR) has been widely used for the study of molecular structure, dynamics and protein-ligand interactions [Pellecchia 2002b, Maurer 2005, Markwick 2008]. Specially, for monitoring ligand interactions new NMR approaches has been developed and studied by pharmaceutical industry in drug discovery and screening [Wishart 2005, Carlomagno 2005, Harner 2013].

In order to detect ligand-protein interactions, two basic different approaches have been developed. NMR can be used to observe changes in the ligand side (small molecules) when it binds to the protein, or to observe changes in the protein side up to binding.

### 6.1.1 Protein-ligand interactions: protein based experiments

The binding of the ligand to a protein results in a local environmental effect on the binding surface, but also in the protein's structure and dynamics. These changes will be displayed as alterations in the NMR properties as chemical shifts, relaxations rates and line-shapes. For that reason, these parameters can be used as indicator of ligand-protein interactions.

#### Chemical shift perturbation

The chemical shift perturbation (CSP) experiment, also known as chemical shift mapping, is one of the most widely used in this domain to probe ligand-protein interactions [Williamson 2013]. The aim of this experiment is to follow the changes in the chemical shifts ($^{1}$H, $^{15}$N or $^{13}$C) of a protein during a ligand titration. The amino acid residues that show a progressive shift in $ppm$ can be located in the binding surface or in the domain that suffers a structural change. The dissociation constant ($K_d$) can be calculated from this experiment, derived from chemical-shifts variations as function of the ligand concentration. A good example of the application of the CSP experiment are [Quinternet 2012] and [Köhler 2015].

#### Protein dynamics

Dynamics have an important role for understanding protein function and for structure-based drug design [Sapienza 2010]. Ligand binding can lead to changes in the protein three dimensional structure, and these modifications imply changes to its dynamics. NMR relaxation methods are extremely powerful for the characterization of protein motions with atomic resolution over a different times scales. These approaches can be used to study the fast (ps–ns) and slow ($\mu$s–ms) time-scale dynamics [Ertekin 2007]. Other approaches as non covalent grafting of proteins is used to study protein-ligand interactions using the HR-MAS (High Resolution Magic Angle Spinning) technology [Viéville 2014]. When the nuclear environments are modified slowly regarding the NMR timescale, the spectrum will show two different resonances for the free and bound state. Whereas if the nuclear environment changes rapidly will produce a single resonance as an average of the chemical shifts for the free and bound species.

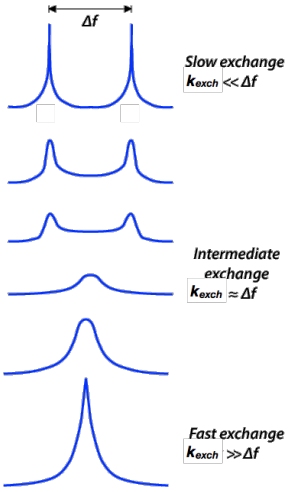



Figure 6.1: Simulation of the spectrum for slow exchange with separate signals for bound and free states, intermediate exchange and fast exchange showing only one signal as a weighted average frequency of both states.

### 6.1.2 Protein-ligand interactions: ligand based experiments

As in the case of protein based experiments, the ligand-based methods are established taking into account that upon binding to a protein, the individual tumbling rates of ligands in solutions will be altered. Small molecules display slow relaxations ($T_1$, $T_2$) and fast diffusion. Upon binding to a protein, this properties dramatically change and decrease. These changes in the ligand properties can be observed in the NMR signals by different methods.

#### Relaxation experiments

One of the most used experiments for NMR binding assays is the comparison of the ligand's relaxation rate, in the absence and the presence of the protein target. Relaxation, that implies the movement of the bulk nuclear magnetization towards equilibrium, can occur through two different mechanisms (Figure 6.2). Thermal equilibrium, is the mechanism of the recovery of the nuclear spin exchanges energy with its lattice (environment) to return to its basal state. In that case the relaxation rate mechanism $R_1$ (longitudinal relaxation rate) occurs at a similar rate for small and large molecules. On the other hand, the selective $R_1$ occurs when a restricted set of protons is saturated or inverted using a selective frequency (as in the STD experiment). In this case, $R_1$ present a $\omega\tau_c$ dependence and is a sensitive probe of binding.

The second mechanism, depends on the size of the molecule such that relaxation becomes more rapidly with the increasing size of the molecule. The relaxation rate via a spin-spin mechanism is called $R_2$ (transverse relaxation rate) and occurs faster for large molecules (slowly

tumbling) [Skinner 2008]. Figure 6.2 shows that for small molecules $R_1$ and $R_2$ have similar values, whereas for large molecules, the transverse relaxation becomes larger and the longitudinal relaxation decreases.

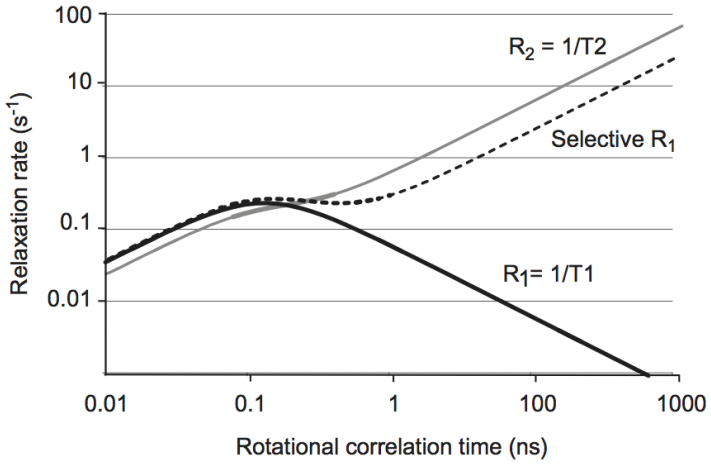



Figure 6.2: Values of $R_1$ and $R_2$ as function of motional rotational correlation times. Selective $R_1$, represents when only one proton is perturbed and the other remains at equilibrium. Figure from [Kieffer 2011].

The shape of the Free Induction Decay (FID), is controlled by the transverse relaxation, a large $R_2$ results in a FID that decays rapidly. Consequently, faster relaxing species produce broader NMR signals, less resolved peaks appear in the spectrum. This is the effect that can be observed when a small molecule binds to a protein target, the small molecule adopt the NMR properties of the large molecule, assumes fast transverse relaxation rates and the consequence is the spectral line broadening of the small molecule signal [Pellecchia 2002a].

### Diffusion experiments

The diffusion coefficient can be measured by NMR using DOSY experiments (see Chapter 4). This measurement can be a convenient way for the study of a ligand-protein interaction [Skinner 2008].

When recording a DOSY experiment for small molecule alone and after adding the target protein, notable differences are observed. The ligand bound to the protein presents a slower diffusion coefficient and their NMR signals intensities are less attenuated in the DOSY experiment for high intensity gradients. In the case of strong binding, ligand and protein signals present the same apparent diffusion coefficient after binding. By selecting a suitable gradient

strength, it is possible to filter out the unbound ligands, because of their fast diffusion, and the bound ligands will be selectively detected from the mixture. However, fast $R_2$ makes this approach difficult to implement. In the case of weak binding, because of the equilibrium between the free and the bound species, the diffusion of the ligand is modified, and the diffusion coefficient reflects the protein diffusion, in proportion to the amount of bound ligands.

Diffusion experiments follow the same basic principles as relaxation experiments but rely on differences in translational instead of rotational motions.

### Exchange Transferred NOEs

The nuclear Overhauser effect (NOE) has been introduced in Chapter 3 The use of the NOEs depends on the ability of the ligand to adopt the relaxation properties of the large molecule. The 1D and 2D exchange transferred NOE experiments (etNOE) are used to determine the specific conformation of weak to medium affinity ligands when bound to a protein target [Post 2003].

The ligand protons undergo additional dipole-dipole interaction with protein's protons, upon binding to a large molecule. As said before, the ligand adopts the tumbling properties of the large molecule and here the binding is detected by the NOEs developed intra ligand upon interaction, when a strong dipole-dipole interaction occurs. The negative NOE cross peaks characteristic of small molecules become positive, indicating the ligand binding with the protein target [Skinner 2008].

### Saturation transfer difference (STD)

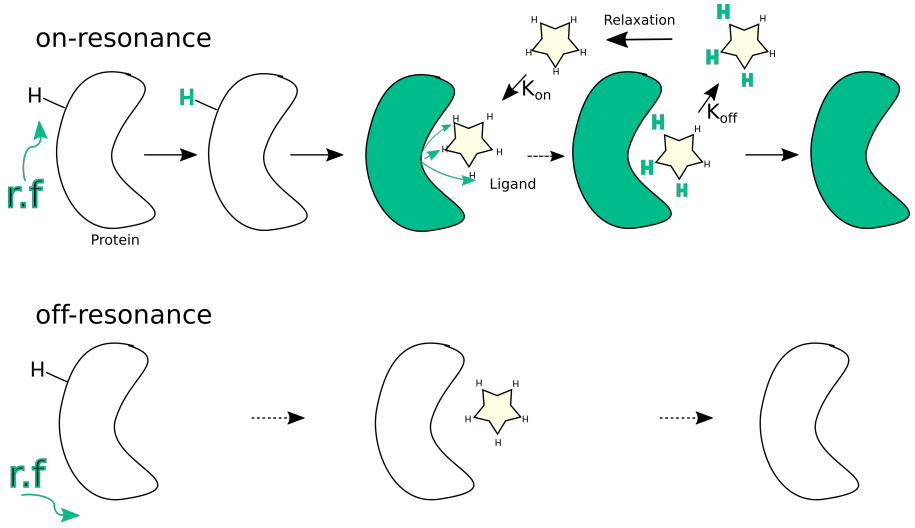
The STD experiment has its basis in the intermolecular magnetization transfer between the ligand and the large molecule through the NOEs as occurs in the NOE experiment. In the STD experiments the protein target resonances are saturated selectively, the protein is saturated with radio frequency pulses (to reach a saturated steady state). At that point, ligand spins that are close to the saturated protein, receive part of this saturation through intermolecular $^1$H-$^1$H cross relaxation pathways, and the ligand signal intensity decrease in the 1D NMR spectrum [Mayer 2001] probing the ligand-protein binding.

As shown in Figure 6.3, a selected $^1$H spectral region of the protein is saturated in the "on" resonance experiment (methyl region near 0 ppm). After that, a second region far from the protein $^1$H spectral range is saturated as "off" resonance experiment. The difference between the resulting spectra shows the magnetization transfer from protein to ligand protons.

The optimal conditions for the STD experiment consists in a complex with a $k_{off}$ on the same order of magnitude as the intermolecular saturation transfer rate, and to optimize the

ligand, protein concentration and the saturation time to fit in this $k_{off}$ range.

a)




b)

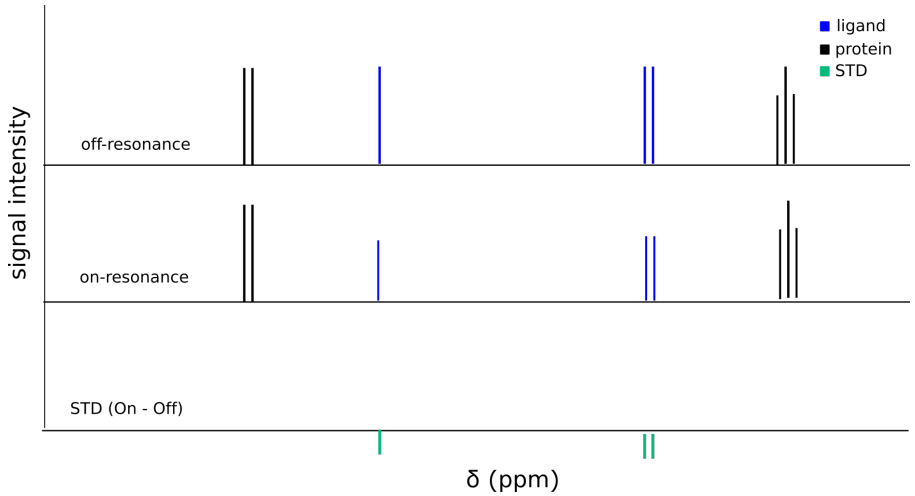



Figure 6.3: a) Mechanism of saturation transfer difference (STD). In the "on"-resonance spectrum, the protein resonances are saturated at a frequency where no ligand resonates (methyls or aromatics protons of the protein). This saturation is transferred to the ligand when it binds to the protein. Thus protons show an enhanced signal in the STD spectrum. In the "off"-resonance spectrum, the selective radio frequency is selected where neither protein nor ligand resonates, resulting in a normal NMR 1D $^{1}$H spectrum. b) NMR spectra resulting of the STD experiment.

The STD amplification factor [Mayer 2001], was defined after a study showing that the STD saturation factor increases with the ligand concentration until it reaches a plateau. The

difference between both points is the STD amplification factor. (STD amplification factor = $[(I_0 - I_{sat})/I_0]$ x ligand excess).

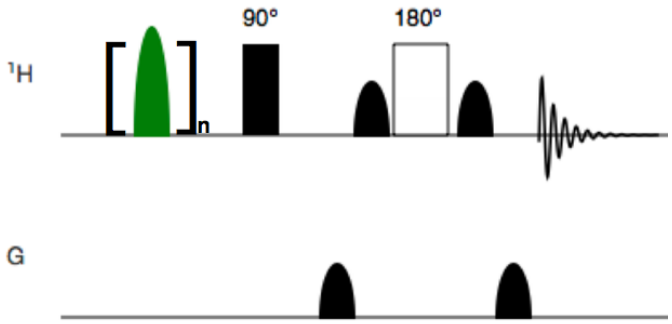



Figure 6.4: Pulse sequence diagrams for the STD NMR experiments with water suppression by WATERGATE method. The solid bars and boxes represent hard 90° and 180° pulses, respectively. For STD after the initial relaxation delay, proton magnetization is saturated either on or off resonance by a train of Gauss pulses (green shaped pulse). The last 180° pulse between two shaped pulses, represents the WATERGATE, canceling the water signal, because of the selective pulses on the water signal.

### Water-Ligand Observation via Gradient SpectroscopY (WaterLOGSY)

The WaterLOGSY experiment is an adaptation of the STD method. In both experiments the objective is to observe the transfer magnetization to the bound ligand and to compare the differences with the spectrum of the unbound ligand. In the case of the WaterLOGSY experiment [Dalvit 2000, Dalvit 2001], the transfer occurs indirectly through the excitation of the bulk water magnetization. The excited water molecules interact with the ligand-protein complex and then transfer the magnetization. This magnetization will be selectively retained by the free ligand, after dissociation (Figure 6.5 a). In the WaterLOGSY experiment the determination of an optimal "on" resonance is less important than in the STD experiment [Antanasijevic 2014].

An alternative important mechanism for explaining magnetization transfer from water to the complex, is the transfer by the labile groups (-NH and -OH) through chemical exchange.

However, in both cases the pathway for the transfer magnetization conserves the same sign of the bulk water, and the non-bound ligands will appear with the opposite sign compared to the water signal (Figure 6.5 b).

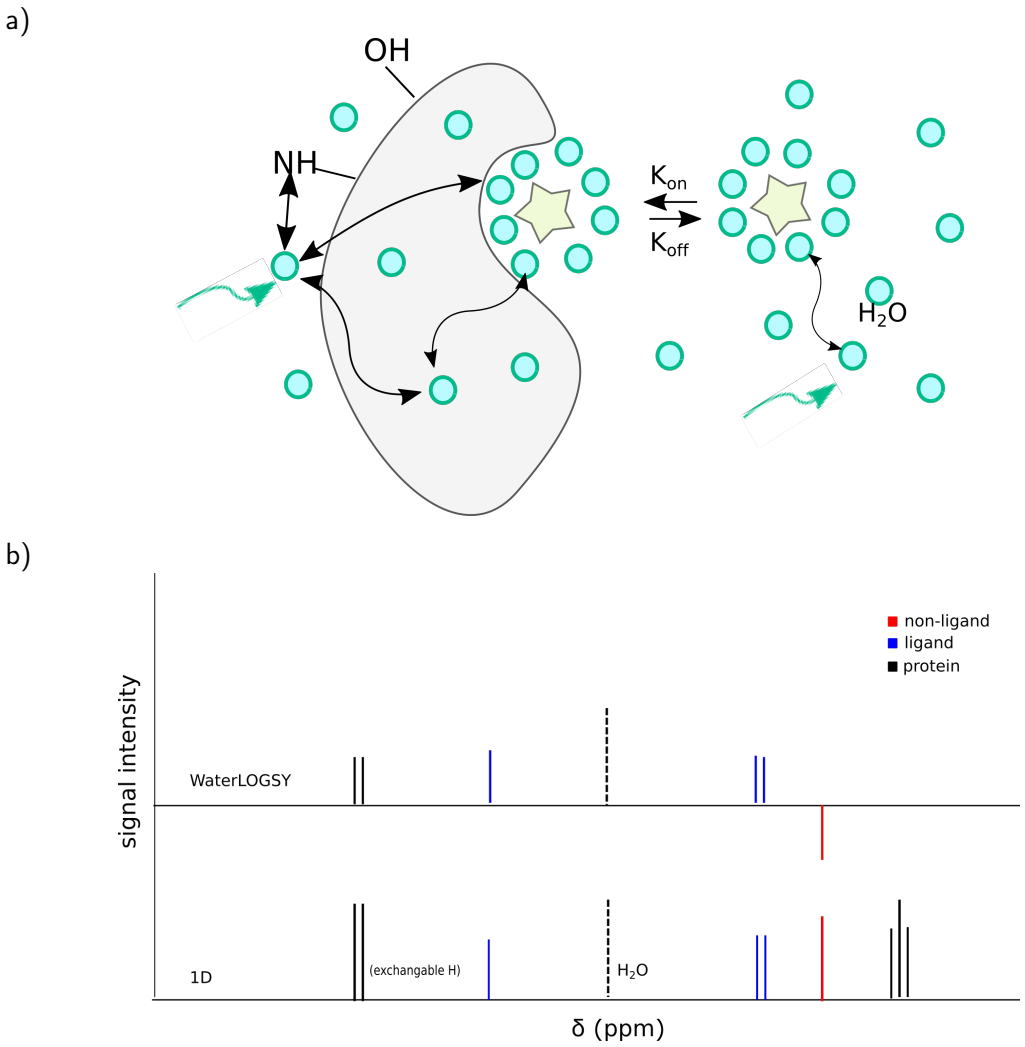



Figure 6.5: a) Mechanism of WaterLOGSY. The bulk water (blue dots) is excited by r.f. (green arrow) and by different magnetization transfer pathways (black arrows) is transferred to the rest of the water molecules in the buried cavities or in the binding site. This magnetization is transferred to the ligand through the active binding site. The ligand is shown in the bound and free states. b) NMR spectra resulting of the WaterLOGSY experiment.

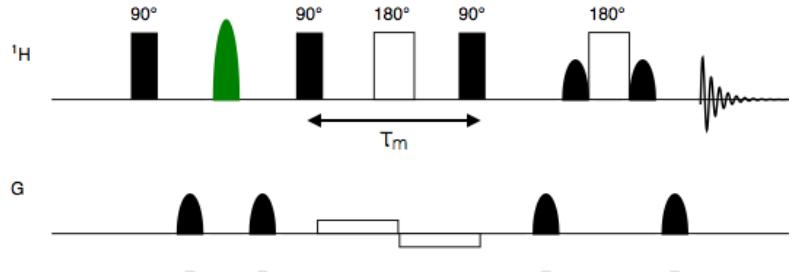



Figure 6.6: Pulse sequence diagrams for WaterLOGSY NMR experiment with water suppression by WATERGATE method. The solid bars and boxes represent hard 90° and 180° pulses, respectively. The green shaped pulse selectively inverts water magnetization. During the mixing time $\tau_m$, a weak bipolar gradient is applied to suppress radiation damping. The last 180° pulse between two shaped pulses, represents the WATERGATE, canceling the water signal, because of the selective pulses on the water signal.

### 6.1.3 Potential targets for the NTD-AR

Nowadays the N-terminal domain of the Androgen Receptor is presented as a potential target site for the development of new drugs for prostate cancer, specially when the castration resistance prostate cancer (CRPC) appears and the sensitivity to anti-androgen therapy is lost. As discussed in Chapter 2, in the last years news studies have been published [Andersen 2010, Godbole 2011, McEwan 2011], presenting the molecule EPI-001 (Figure 6.7) and some derivates, as a potential new drug that present regression of prostate cancer. These molecules are a derivate of the one of the most well known endocrine disruptor, Bisphenol A (BPA) (Figure 6.7). Recently, other new molecules have been appearing in the literature [Bañuelos 2014], indicating that is an emergent field in continuous development.

For all these reasons, the strategy is to show that the amyloid fibers, formed upon oxidation of peptide **2**, can be a target site for all these new molecules. Moreover, the fibers analyzed in this work could present a biological role, for this reason some other related hormone involve in the AR activity are studied. As well as, some polyphenols due to their antioxidants properties and their effects in the androgen receptor activity and expression [Ren 2000, Ferruelo 2014].

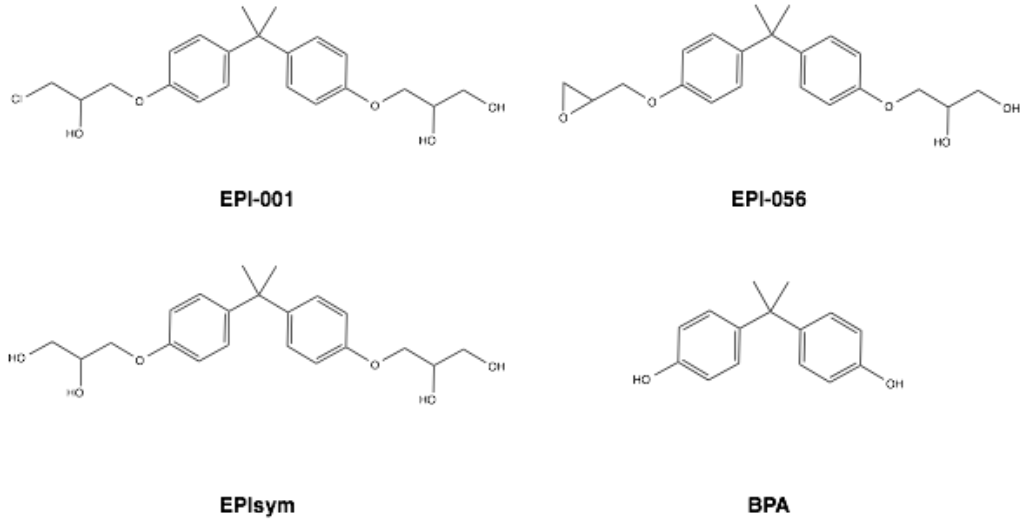



Figure 6.7: Different EPI molecules, all derivates of Bisphenol A.

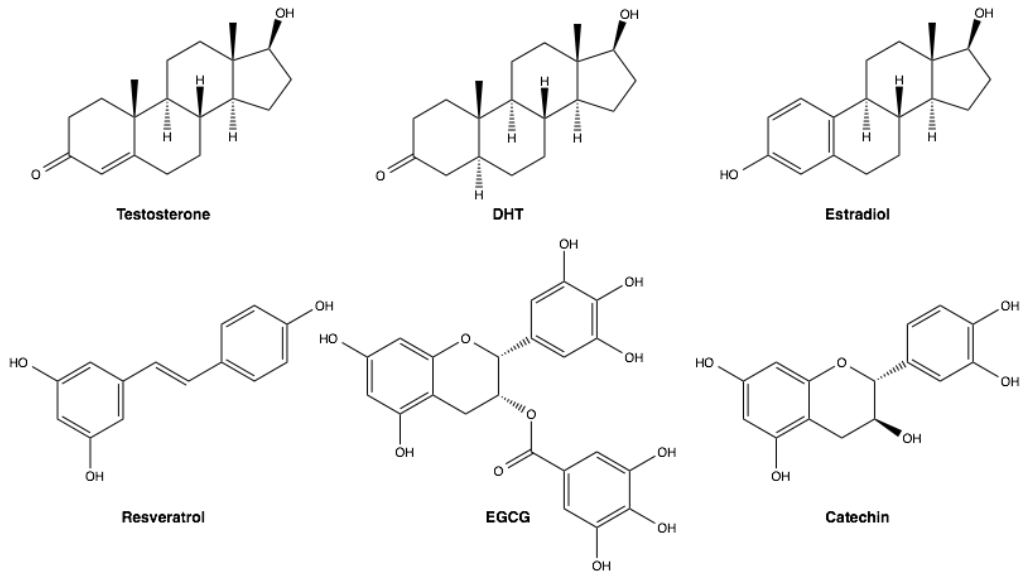



Figure 6.8: Hormones and polyphenols used in this study.

## 6.2 Fibers-molecules interactions

### 6.2.1 Molecules influence on the kinetics aggregation

A first step of the interaction study is to characterize the influence of the BPA derivates molecules, in the kinetics of the fibril formation. In order to check if the presence of the small molecules had an impact in the aggregation, inhibiting the fibril formation or accelerating this

phenomenon, the NMR signal of the mixture (peptide **2** + small molecule) was monitored during 70 h. The reactions were performed in presence of 5% DMSO.

Four different molecules were tested, Bisphenol A, EPI-001, EPI-056 and EPI-sym. As it can be observed in Figure 6.9, in all cases the presence of a small molecule alters the aggregation rate. The kinetics were compared with a peptide aggregation kinetics without the presence of molecules in solution (Figure 6.9 black dots). It is remarkable that, not only the signal disappearance occurs faster when adding the molecules, but also the aggregation does not fit well a second-order kinetics anymore.

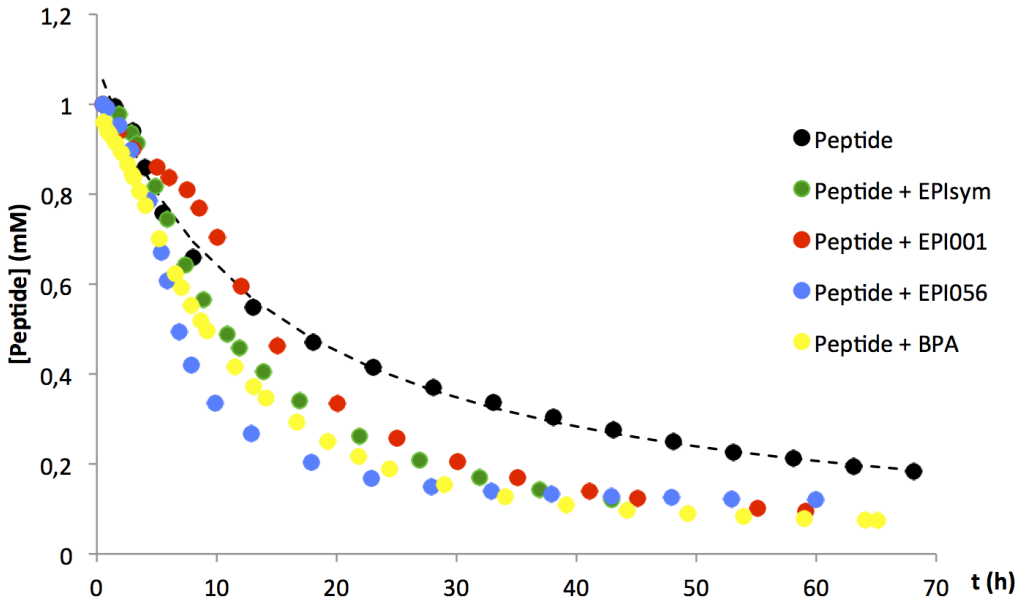



Figure 6.9: Peptide concentration variation followed by NMR, in the presence of BPA (yellow), EPI-001 (red), EPI-056 (blue) and EPI-sym (green). In black the kinetics of the peptide without molecules, as a control. The dashed line shows the second-order kinetics fitting for the control.

The four molecules affect in the same way to the kinetics of the reaction. The major effect is found when adding EPI-056, the signal disappearance display a faster decay. The presence of the epoxy in this molecule could be the factor that makes the molecule to react faster with the peptide. And this effect could alter in a major way the aggregation.

The other molecule that presents a different behavior is EPI-001. With a similar effects regarding the kinetics rate modification, in contrast the peptide in presence of this molecule displays a longer and marked nucleation step. In addition EPI–056 and EPI–001 have been shown to be active against CRPC [Myung 2013].

### 6.2.2 Saturation transfer difference (STD)

We aim to show interaction between the small molecule EPI-001 and the fibrils formed in solution. The STD experiment was only tested with the molecule EPI-001. As explained in the introduction, this method presents less sensitivity than the WaterLOGSY experiment [Antanasijevic 2014] and requires a precise "on" resonance. The difficulty of using the STD experiment in this system comes from the fact that we are not dealing with a standard globular and structured protein. It lacks the aromatic moiety which produces large shifts in the spectrum. We do not observe the fibrils in the NMR spectrum, for this reason the frequencies of the soluble peptide are used as reference. Nevertheless, the methyl groups presented in peptide **2** appear at 0.85 ppm, no so far of the methyl groups of the EPI-001, making it difficult to find an optimal "on" resonance.

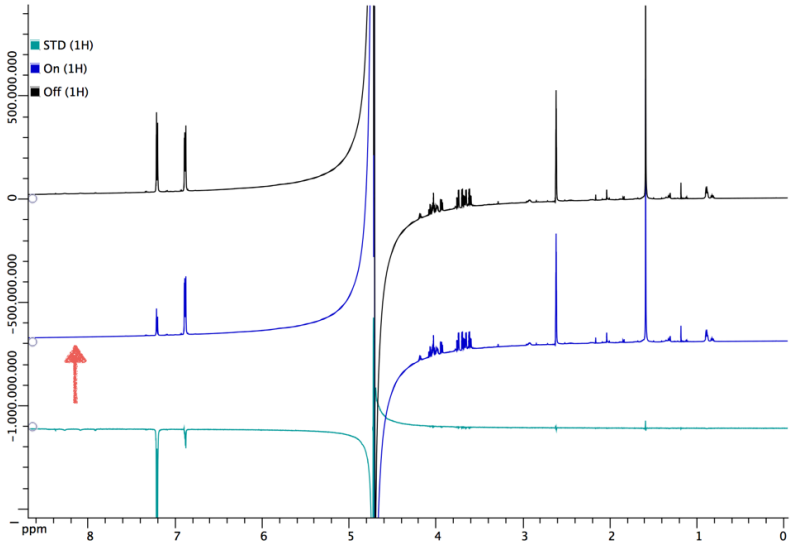



Figure 6.10: NMR spectra resulting of the STD experiment, when the NH region is saturated (red arrow). Off-resonance spectrum (black), On-resonance spectrum (blue) and STD spectrum (difference between On-Off spectra) in light-blue.

We could not obtain any clear result after performing the experiment (Figure 6.10, 6.11 and 6.12). We performed the experiment by saturating three different regions, the NH region (Figure 6.10), the methyl region (Figure 6.11) and finally an offset region at 0ppm (Figure 6.12). An effect is observed on the two first experiments. Nevertheless, when we apply the saturation to the fibrils we are also irradiating the molecule, by proximity. This leads to an unclear result from which no conclusion can be drawn.

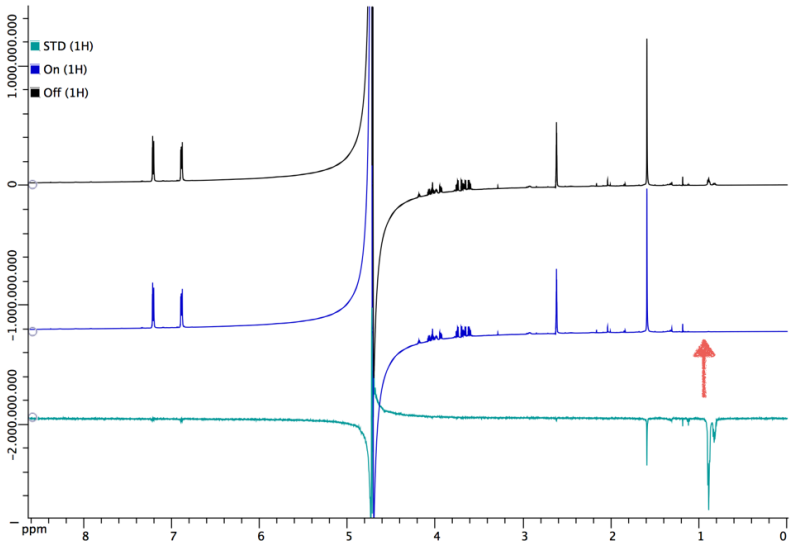



Figure 6.11: NMR spectra resulting of the STD experiment, when the Methyl region is saturated (red arrow). Off-resonance spectrum (black), On-resonance spectrum (blue) and STD spectrum (difference between Off-On spectra) in light-blue.

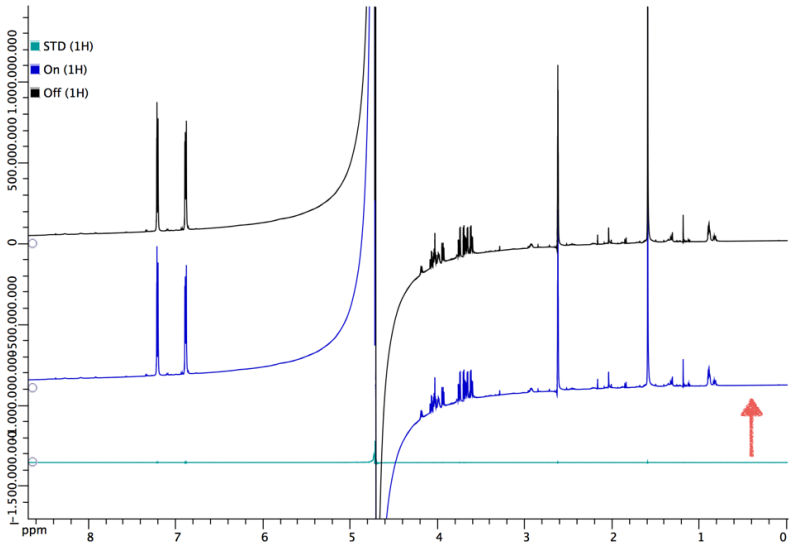



Figure 6.12: NMR spectra resulting of the STD experiment, when the 0ppm region is saturated (red arrow). Off-resonance spectrum (black), On-resonance spectrum (blue) and STD spectrum (difference between Off-On spectra) in light-blue.

### 6.2.3 WaterLOGSY (WL)

The WaterLOGSY (WL) is the alternative to the STD to find interactions between the fibers and different small molecules that could have an important role with the androgen receptor regulation and prostate cancer new therapies. We can organize them in three groups: BPA derivates, hormones and polyphenols.

The first step was to perform a WL control-study of the molecules (See Appendix C). The results of these experiments show that all molecules alone in solution (with 5% DMSO) exhibit a negative (–) response (or 0), as expected, except for EPI-001 that shows a positive (+) response in the WL spectrum. This "false" positive could be associated with the low solubility of the molecule in water and the formation of colloids in the solution can induce to misleadings. A decrease in the concentration of this molecule lead to a spectrum with 0 response proving the hypothesis above (See Figure 6.14).

In a second step, a study was performed by adding the different molecules to a solution with the peptide after oxidation. All the molecules tested displayed a (+) response in the WL spectrum except Chloramphenicol (–) (see Figure 6.13). Chloramphenicol is an aromatic antibiotic that was chosen as negative control for this interaction study, and the results confirmed that.

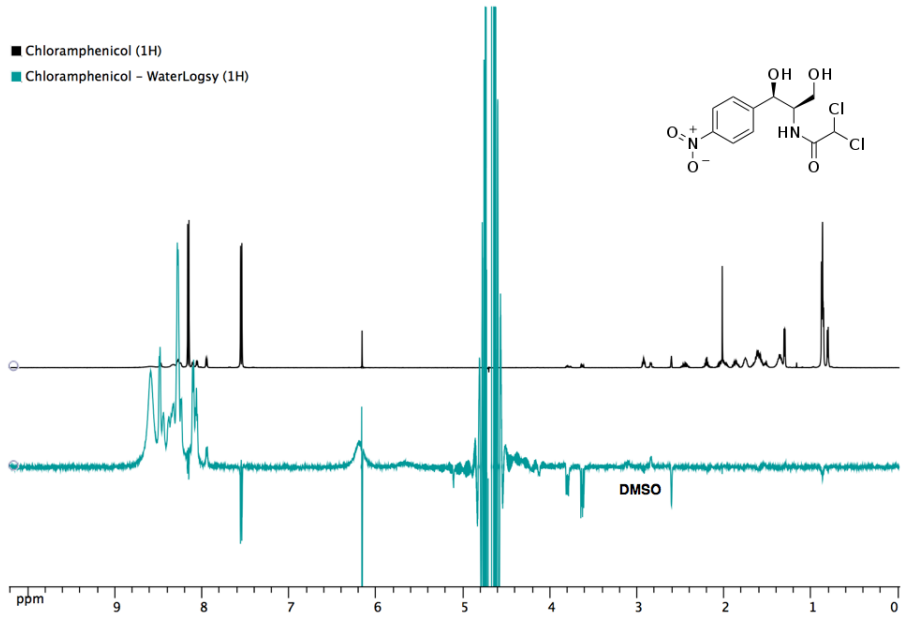



Figure 6.13: Peptide **2** after oxidation + chloramphenicol $^{1}$H spectrum (black) and WL spectrum (blue).

As EPI-001 is the molecule of interest in our research project due to its role as potential new therapeutic drug for prostate cancer. Here we show the main results proving the interaction and affinity of the molecule with the fibers formed in solution.

Figure 6.14 shows the control study of the molecule EPI-001 alone in solution.

In Figure 6.15, the WL response of the molecule is shown once in solution with the fibers formed after peptide oxidation (24 h). In this case we obtain a (+) signal, indicating the interaction between fibrils and molecule. After that, the sample was recovered from the NMR tube and centrifuged at 6000 G during 20 min, this step was performed twice. The supernatant was put back into the tube and the WaterLOGSY experiment was performed, this time without the presence of the fibers in solution. The WL signal of the molecule became again **0**, indicating that no interaction was occurring now in solution.

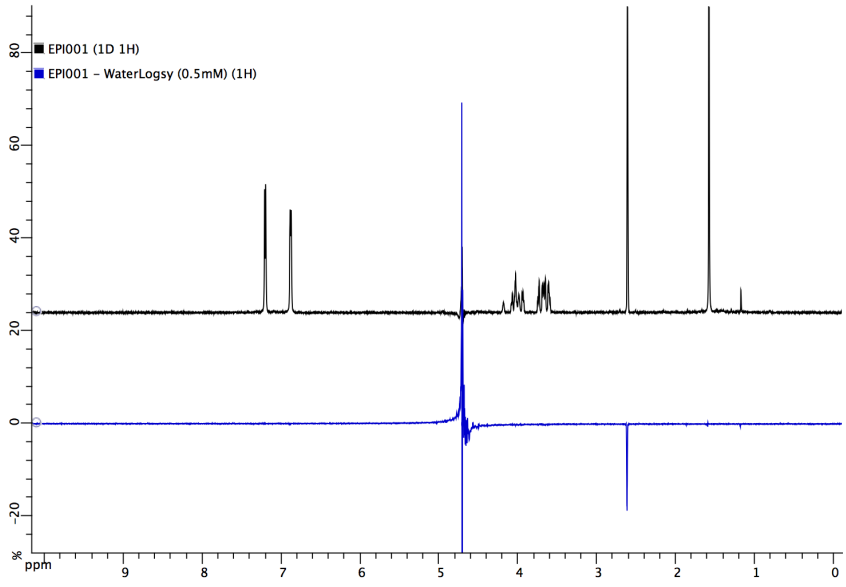



Figure 6.14: Molecule EPI-001 in solution, $^{1}$H spectrum (black), WL spectrum (blue), WL spectrum after dilution (1/10) (green).

In Table 6.1 are summarized the results obtained for all the molecules tested in this study. The results are compared between the control experiment with the different molecules alone in solution or in the presence of amyloid fibers from peptide **2**. Is remarkable the positive response of all molecules in the presence of fibrils in solution, with the exception of chloramphenicol and DMSO. For this reason DMSO is used in all the samples as a internal control. This situation indicates that more (–) response molecules are needed.

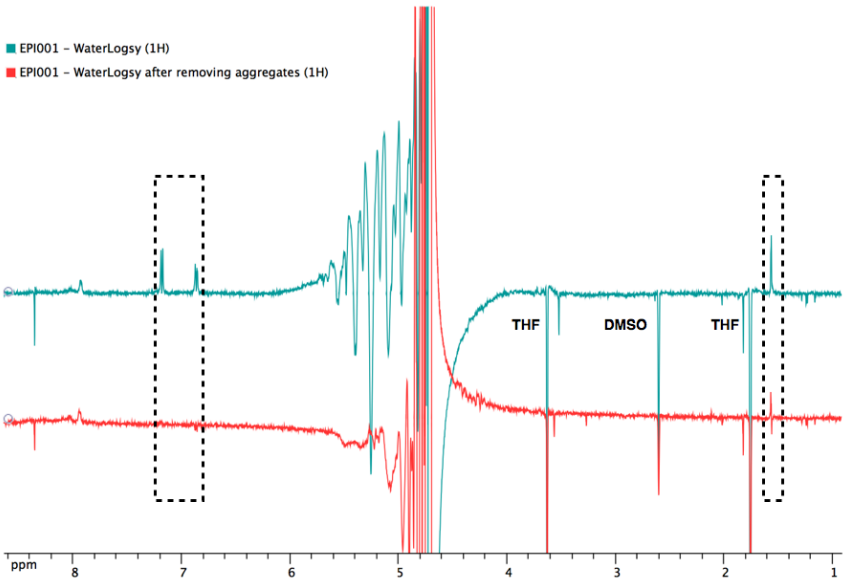



Figure 6.15: WL spectrum of the mixture peptide **2** + EPI-001 (blue) and WL after removing the fibers by centrifugation (red). Some of the peaks corresponding to the molecule (methyls and aromatics $^{1}$H are marked by dashed lines).

Table 6.1: WaterLOGSY results. (+) positive response,(−) negative response or (0).

| Molecule | control | +fibers |
|---|---|---|
| **DMSO** | (−) | (−) |
| **Chloramphenicol** | (−) | (−) |
| **EPI–001** | (0) | (+) |
| **EPI-056** | (0) | (+) |
| **EPIsym** | (−) | (+) |
| **BPA** | (−) | (+) |
| **Catechin** | (−) | (+) |
| **EGCG** | (−) | (+) |
| **Resveratrol** | (−) | (+) |
| **DHT** | (−) | (+) |
| **Testosterone** | (−) | (+) |
| **Estradiol** | (0) | (+) |

## 6.3 EPI-001 + KELCKAVSVSM: covalent bond formation

A different strategy can be used to study the possibles interactions between the new CRPC drugs and the highly conserved peptide from the NTD-AR. As has been proposed in the literature [Myung 2013], EPI–001 probably interacts with the NTD-AR through a covalent interaction. The mechanism proposed involves a first fast reversible interaction between EPI-001 and the AF1 region of AR to place the secondary alcohol of the molecule close to a basic site in the NTD-AR. In a second step, the base can remove the proton of the secondary alcohol of the chlorohydrin with the consequently epoxide formation, through the loss of a chloride. The highly reactive epoxide reacts with a nucleophilic site on an amino acid side chain in the AF1 to form an irreversible covalent bond.

In order to characterize this interaction and to search for a possible adduct formation between the peptide **2** and EPI–001, a coupled LC/MS/NMR experiment was thought as an effective strategy to characterize the reaction.

### 6.3.1 LC/MS/NMR

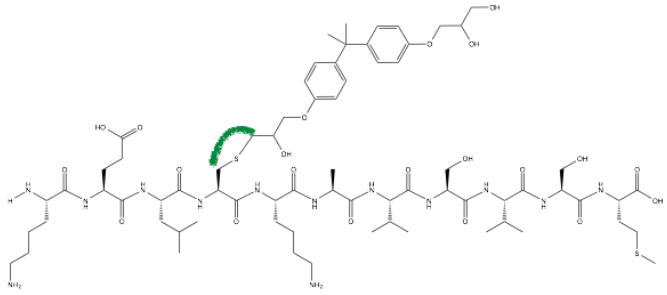
The use of an integrated LC/MS/NMR system can provide additional information and structural identity of the mixture we are analyzing. LC-SPE NMR and mass spectroscopy go especially well together, provide the advantages of each one of the techniques and furthermore it is a closed system. LC/MS/NMR coupled system allows an ultra-sensitive screening in complex mixtures, followed with an NMR structural characterization. The SPE (solid phase extraction) interfaced between LC and NMR allows the usage of completely non-deuterated solvents for the elution of the chromatic fractions collected into the NMR capillary. This is a good system to avoid the pollution or the loss of sample.

This technique has a wide range of applications but is mainly used in pharmaceutical and biotechnology industries for the study of natural product extracts, combinatorial peptides mixtures, food extracts or polymers.

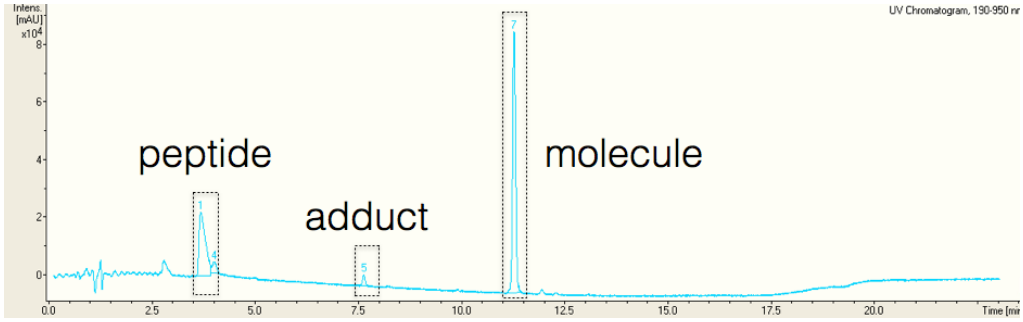
We applied this method to search and to characterize the covalent interaction between EPI–001 and peptide **2**. The sample was prepared in basic conditions, pH = 8.0, to favor the thiolate formation and to accelerate the covalent bound formation. The hypothesis is that the peptide and the molecule can form an adduct as shown in Figure 6.16 a). The sample to analyze was prepared with a final concentration of peptide **2** of 3 mM and EPI–001 5 mM. The molecule EPI–001 was previously solubilize in THF (tetrahydrofuran).

As a first step the three components of the sample (peptide **2**, EPI–001 and adduct) were identified and separated by HPLC. Figure 6.16 b) shows the chromatogram obtained, where

a)



b)




c)

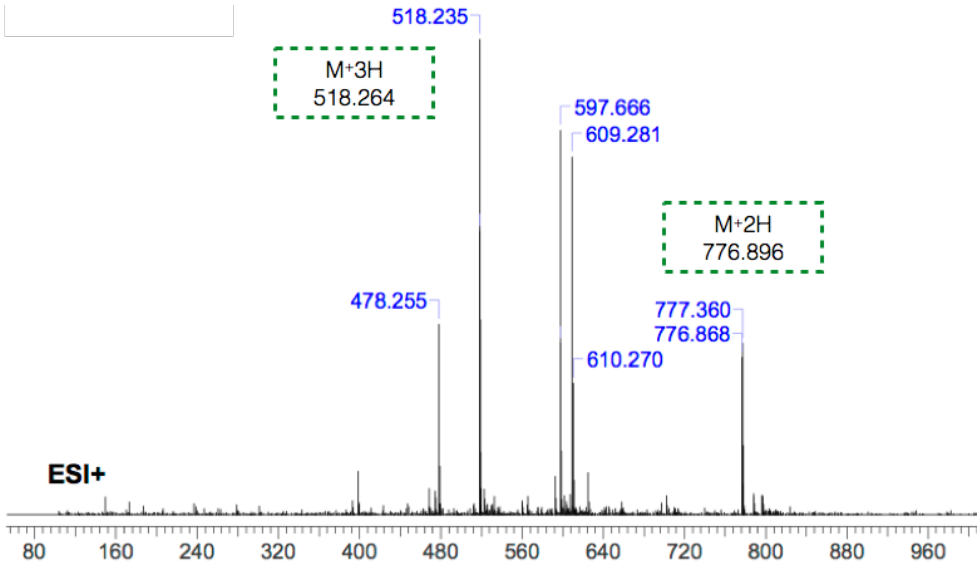



Figure 6.16: a) Covalent bond formation between EPI–001 + peptide **2**. b) HPLC chromatogram resulting of the mixture separation. c) MS spectrum of the adduct corresponding fraction.

we can easily distinguish the three species in solution. In parallel, a MS analysis was performed for each one of the peaks. In Figure 6.16 c) is shown the MS results obtained for the adduct

peak. Peptide ($C_{50}H_{91}N_{13}O_{16}S_2$) and the EPI–001 ($C_{21}H_{27}ClO_5$) interact forming a covalent bound between the -SH of the cysteine and the chlorohydrin group from the molecule (see adduct structure in Figure 6.16 a)). This bound formation results in an adduct with formula $C_{71}H_{117}N_{13}O_{21}S_2$, it corresponds to the loss of chlorine. The attended mass for this adduct is 1551.79 Da. The analysis of these results show the corresponding mass for the double-charge ion $[M+2H]^{2+}$ (518.235 Da) and for the triple-charge ion $[M+3H]^{3+}$ (776.868 Da), for the adduct formed between the peptide **2** and EPI–001.

For a complete characterization, as a final step the sample was collected in the SPE cartridge (solid phase extraction). The cartridge was dried and posteriorly it was eluted to the NMR spectrometer capillary with acetonitrile-$d_3$. Unfortunately, the NMR information could not be obtained due to the low concentration of the adduct in solution and the sensibility limitations of the equipment. Nevertheless, an additional experiment was performed by using the 700 MHz cryo-probe TXI. The sample from the NMR capillary was recovered and used on the 700 MHz. To prove the covalent interaction we need to find a correlation peak between the cystine-$H_\beta$ and the C of the EPI–001 directly attached to the cysteine, see Figure 6.16 a) marked in green. The NMR experiment required to find this correlation, is the HMBC experiment (Figure 6.17), that gives the information of long range heteronuclear correlation. However, it is a low sensitivity method and together with the low adduct concentration, make that the starting conditions are not the ideal to perform the experiment. The HMBC spectrum did not show any peak at cystine-H$\beta$ frequency.

HMBC

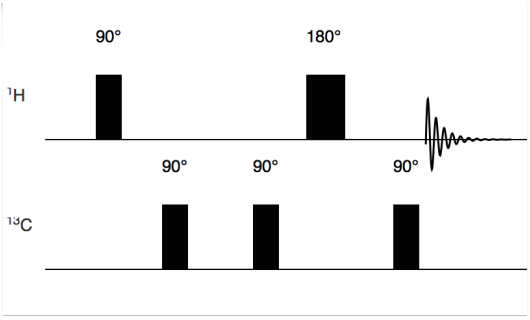



Figure 6.17: Simplified representation of the HMBC (Heteronuclear Multiple Bond Correlation) sequence.

## 6.4 Conclusions

Endocrine disruptors alters androgen receptor function [Luccio-Camelo 2011]. This study aimed to study the influence of small molecules in the amyloid fibril formation of peptide **2** and their possible interactions. EPI–001, the new potential drug for prostate cancer, is a derivate of the well known endocrine disruptor Bisphenol A (BPA). Here we have shown that the BPA derivates family of small molecules accelerates the kinetics aggregation of the amyloid fibers. Furthermore, BPA have been reported in literature to speed-up the toxic amyloid formation [Gong 2013]. This result goes in the same direction than the study presented here and can open new lines of research for controlling amyloid aggregation.

WaterLOGSY is a higher sensitivity method compared with STD [Antanasijevic 2014]. We present a system with special difficulties compared to standard globular proteins. Specially with the absence of an aromatic moiety in the fibers (KELCKAVSVSM), where the optimal saturation frequency is not easy to determine. Using the WaterLOGSY experiment we show that the amyloid fibers formed after oxidation of the peptide **2**, interact with three different categories of molecules: hormones, polyphenols, and EPIs, the new potential drugs for the prostate cancer. These latter molecules result as derivates of a well known endocrine disruptor, Bisphenol A, which also display a positive response as result of the interaction with the fibrils.

Finally, the mechanism of action of the EPI–001 within the NTD-AR, is proven in part with this study. A covalent interaction is characterized by HPLC/MS. The covalent interaction, formed through the Cys and the EPI–001, was not characterized by NMR. The low sensitivity of the HMBC experiment together with the low concentration of the adduct precluded its characterization. A perspective on this study is to carry out new experiments, such as MS fragmentation, to complete the characterization of the adduct formed between the peptide **2** and the EPI-001.

All together, in this chapter has been shown that, the high conserved sequence from the NTD-AR does not only aggregates into amyloid fibers under oxidative conditions, this could also be a target site for the new potential drugs.

# Bibliography

[Andersen 2010] Raymond J Andersen, Nasrin R Mawji, Jun Wang, Gang Wang, Simon Haile, Jae-Kyung Myung, Kate Watt, Teresa Tam, Yu Chi Yang, Carmen A Bañuelos, David E Williams, Iain J McEwan, Yuzhou Wang and Marianne D Sadar. *Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor*. Cancer Cell, vol. 17, no. 6, pages 535–546, June 2010. (Cited on pages 35 and 123.)

[Antanasijevic 2014] Aleksandar Antanasijevic, Benjamin Ramirez and Michael Caffrey. *Comparison of the sensitivities of WaterLOGSY and saturation transfer difference NMR experiments*. J Biomol NMR, vol. 60, no. 1, pages 37–44, September 2014. (Cited on pages 121, 126 and 134.)

[Bañuelos 2014] Carmen A Bañuelos, Aaron Lal, Amy H Tien, Neel Shah, Yu Chi Yang, Nasrin R Mawji, Labros G Meimetis, Jacob Park, Jian Kunzhong, Raymond J Andersen and Marianne D Sadar. *Characterization of Niphatenones that Inhibit Androgen Receptor N-Terminal Domain*. PLoS ONE, vol. 9, no. 9, page e107991, September 2014. (Cited on page 123.)

[Carlomagno 2005] Teresa Carlomagno. *Ligand-target interactions: what can we learn from NMR?* Annu Rev Biophys Biomol Struct, vol. 34, pages 245–266, 2005. (Cited on page 115.)

[Dalvit 2000] C Dalvit, P Pevarello, M Tatò, M Veronesi, A Vulpetti and M Sundström. *Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water.* J Biomol NMR, vol. 18, no. 1, pages 65–68, September 2000. (Cited on page 121.)

[Dalvit 2001] C Dalvit, G Fogliatto, A Stewart, M Veronesi and B Stockman. *WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability*. J Biomol NMR, vol. 21, no. 4, pages 349–359, December 2001. (Cited on page 121.)

[Ertekin 2007] Asli Ertekin and Francesca Massi. Understanding the Role of Conformational Dynamics in Protein-Ligand Interactions Using NMR Relaxation Methods. John Wiley & Sons, Ltd, Chichester, UK, March 2007. (Cited on page 116.)

[Ferruelo 2014] A Ferruelo, I Romero, P M Cabrera, I Arance, G Andrés and J C Angulo. *Effects of resveratrol and other wine polyphenols on the proliferation, apoptosis and*

*androgen receptor expression in LNCaP cells.* Actas Urol Esp, vol. 38, no. 6, pages 397–404, July 2014. (Cited on page 123.)

[Godbole 2011] Abhijit M Godbole and Vincent C O Njar. *New Insights into the Androgen-Targeted Therapies and Epigenetic Therapies in Prostate Cancer.* Prostate Cancer, vol. 2011, no. 1–6, pages 1–13, 2011. (Cited on page 123.)

[Gong 2013] Hao Gong, Xin Zhang, Biao Cheng, Yue Sun, Chuanzhou Li, Ting Li, Ling Zheng and Kun Huang. *Bisphenol A Accelerates Toxic Amyloid Formation of Human Islet Amyloid Polypeptide: A Possible Link between Bisphenol A Exposure and Type 2 Diabetes.* PLoS ONE, vol. 8, no. 1, page e54198, January 2013. (Cited on page 134.)

[Harner 2013] Mary J Harner, Andreas O Frank and Stephen W Fesik. *Fragment-based drug discovery using NMR spectroscopy.* J Biomol NMR, vol. 56, no. 2, pages 65–75, June 2013. (Cited on page 115.)

[Kieffer 2011] B Kieffer, 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, 2011. (Cited on pages 51 and 118.)

[Köhler 2015] Christian Köhler, Raphaël Recht, Marc Quinternet, Frederic de Lamotte, Marc-André Delsuc and Bruno Kieffer. *Accurate protein-peptide titration experiments by nuclear magnetic resonance using low-volume samples.* Methods Mol. Biol., vol. 1286, pages 279–296, 2015. (Cited on page 116.)

[Luccio-Camelo 2011] Doug C Luccio-Camelo and Gail S Prins. *Disruption of androgen receptor signaling in males by environmental chemicals.* The Journal of Steroid Biochemistry and Molecular Biology, vol. 127, no. 1-2, pages 74–82, October 2011. (Cited on page 134.)

[Markwick 2008] Phineus R L Markwick, Thérèse Malliavin and Michael Nilges. *Structural Biology by NMR: Structure, Dynamics, and Interactions.* PLoS Comput Biol, vol. 4, no. 9, page e1000168, September 2008. (Cited on page 115.)

[Maurer 2005] Till Maurer. *NMR Studies of Protein–Ligand Interactions.* In Protein-Ligand Interactions, pages 197–214. Humana Press, New Jersey, April 2005. (Cited on page 115.)

[Mayer 2001] M Mayer and B Meyer. *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor.* J.

Am. Chem. Soc., vol. 123, no. 25, pages 6108–6117, June 2001. (Cited on pages 119 and 120.)

[McEwan 2011] Iain J McEwan. *Intrinsic disorder in the androgen receptor: identification, characterisation and drugability*. Mol. BioSyst., vol. 8, no. 1, page 82, 2011. (Cited on pages 27, 28, 35 and 123.)

[Myung 2013] Jae-Kyung Myung, Carmen A Bañuelos, Javier Garcia Fernandez, Nasrin R Mawji, Jun Wang, Amy H Tien, Yu Chi Yang, Iran Tavakoli, Simon Haile, Kate Watt, Iain J McEwan, Stephen Plymate, Raymond J Andersen and Marianne D Sadar. *An androgen receptor N-terminal domain antagonist for treating prostate cancer*. J. Clin. Invest., vol. 123, no. 7, pages 2948–2960, 2013. (Cited on pages 125 and 131.)

[Pellecchia 2002a] M Pellecchia, D S Sem and K Wüthrich. *NMR in drug discovery*. Nature Reviews Drug Discovery, 2002. (Cited on page 118.)

[Pellecchia 2002b] Maurizio Pellecchia, David Meininger, Qing Dong, Edcon Chang, Rick Jack and Daniel S Sem. *NMR-based structural characterization of large protein-ligand interactions.* J Biomol NMR, vol. 22, no. 2, pages 165–173, February 2002. (Cited on page 115.)

[Post 2003] Carol Beth Post. *Exchange-transferred NOE spectroscopy and bound ligand structure determination.* Curr. Opin. Struct. Biol., vol. 13, no. 5, pages 581–588, October 2003. (Cited on page 119.)

[Quinternet 2012] Marc Quinternet, Jean-Philippe Starck, Marc-André Delsuc and Bruno Kieffer. *Unraveling complex small-molecule binding mechanisms by using simple NMR spectroscopy.* Chemistry, vol. 18, no. 13, pages 3969–3974, March 2012. (Cited on page 116.)

[Ren 2000] F Ren, S Zhang, S H Mitchell, R Butler and C Y Young. *Tea polyphenols down-regulate the expression of the androgen receptor in LNCaP prostate cancer cells.* Oncogene, vol. 19, no. 15, pages 1924–1932, April 2000. (Cited on page 123.)

[Sapienza 2010] Paul J Sapienza and Andrew L Lee. *Using NMR to study fast dynamics in proteins: methods and applications.* Curr Opin Pharmacol, vol. 10, no. 6, pages 723–730, December 2010. (Cited on page 116.)

[Skinner 2008] Andria L Skinner and Jennifer S Laurence. *High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands*. J.

Pharm. Sci., vol. 97, no. 11, pages 4670–4695, November 2008. (Cited on pages 118 and 119.)

[Viéville 2014] J M P Viéville, S Charbonnier, P Eberling, J P Starck and M A Delsuc. *A new NMR technique to probe protein-ligand interaction.* Journal of Pharmaceutical and Biomedical Analysis, vol. 89, pages 18–23, February 2014. (Cited on page 116.)

[Williamson 2013] Mike P Williamson. *Progress in Nuclear Magnetic Resonance Spectroscopy*. Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 73, no. C, pages 1–16, August 2013. (Cited on page 116.)

[Wishart 2005] D Wishart. *NMR spectroscopy and protein structure determination: Applications to drug discovery and development*. Curr Pharm Biotechnol, vol. 6, no. 2, pages 105–120, April 2005. (Cited on page 115.)

## C.1 LC/MS/NMR system

The sample for the LC/MS/NMR experiments was prepared in $H_2O$ pH = 8.0, with a final concentration of peptide **2** of 3 mM and 5 mM of EPI–001 dissolved previously in Tetrahydrofuran (THF) in a 100 mM stock. We did not use DMSO in this experiment due to difficulties that involve its use for the mass spectrometry analysis.

10 $\mu$L of the sample were injected into a C18 ProntoSIL eurobond 125x4 mm (5.0 $\mu$m), in a HPLC Agilent 1200. The sample was analyzed after the HPLC column on a Mass spectrometer Bruker HCT IonTrap, ESI source. 10 runs were performed, different fractions from the HPLC chromatogram were selected and collected in a SPE (solid phase extraction) cartridge (GP and HLB). The cartridges were dried over night using $N_2$. After that, the sample was eluted from the SPE, using deuterated Acetonitrile directly to the 30 $\mu$L capillary in a 500 MHz Bruker LC-SEI.

The chromatographic method used for the HPLC analysis was:

A: H2O 90% (0 min) 31.8% (15 min) 0% (20 min)

B: AcCN

C18 column and UV detection 215 nm

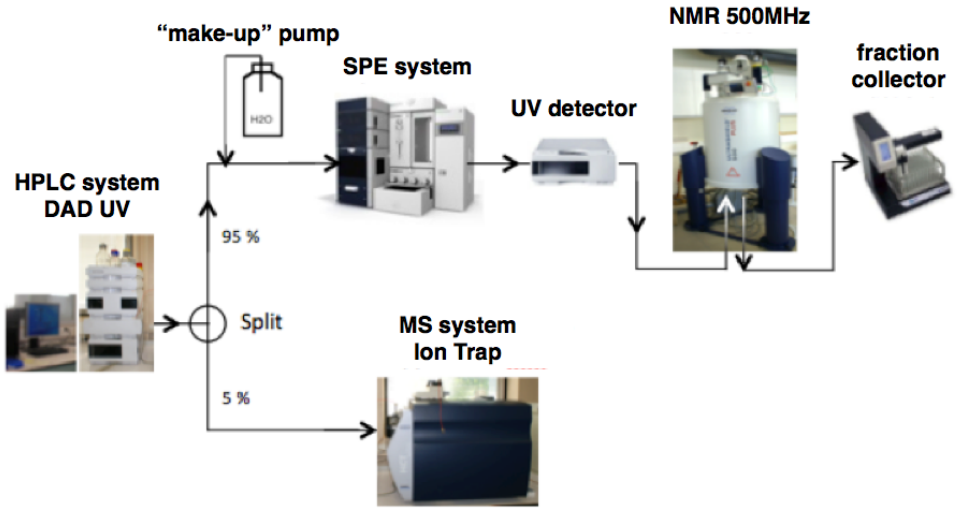



Figure C.1: Schematic description of the LC/MS/NMR system used for the study of a covalent interaction between EPI–001 and peptide **2**.

## C.2 Kinetics, STD and WaterLOGSY experiments

All the experiments were performed in a 3 mm tube, using 180 $\mu$L of sample. The final concentration of the molecules was 5 mM and 1 mM for the peptide **2**, in $H_2O$ with a final pH = 7.0 and 5% DMSO-d6 and 10%$D_2O$. All the molecules were solubilize previously in DMSO-d6. The experiments were performed on a Bruker Avance III 700MHz spectrometer equipped with a Z-gradient cryoprobe at 25°C. The spectra were processed and analyzed using NMR-notebook® (NMRTEC Illkirch France).

The NMR sequence used for the STD experiments, was modified an a noediff sequence was used (Figure C.2). Where the STD a cw pulse is included instead of gaussian pulses but it follows the same saturation transfer difference (on-off resonance) principle than the STD.

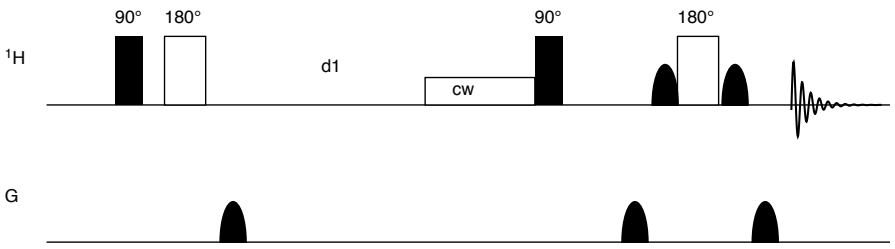



Figure C.2: noediff sequence used for the STD experiments.

Here are presented the WaterLOGSY results for BPA, EPI–056 and EPIsym (Figure C.3);

catechin, EGCG and resveratrol (Figure C.5); DHT, testosterone and estradiol (Figure C.4).

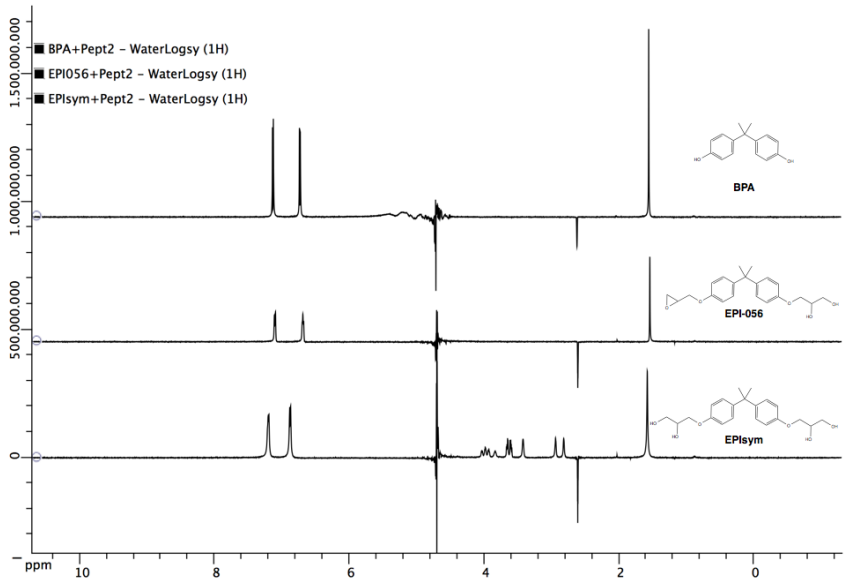



Figure C.3: WaterLogsy spectra for BPA, EPI–056 and EPIsym in the presence of fibrils in solution. Positive response for the three molecules.

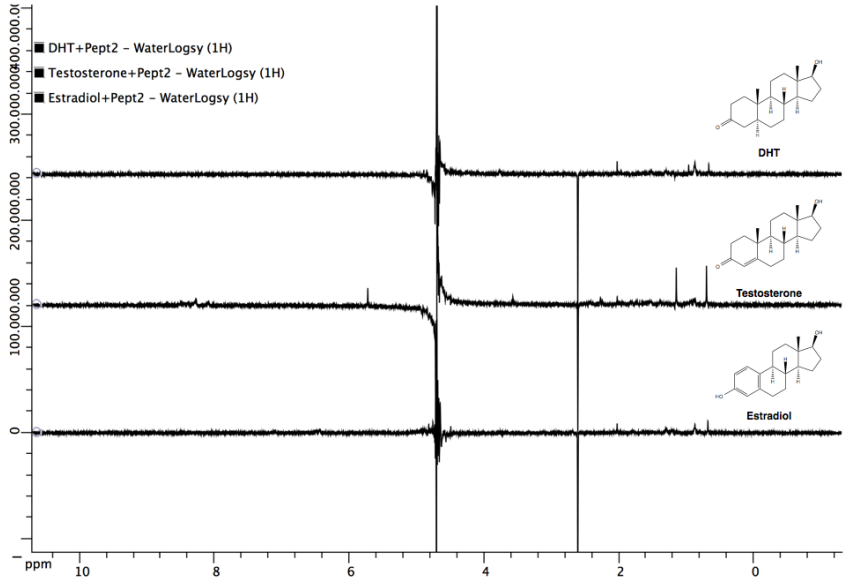



Figure C.4: WaterLogsy spectra for DHT, testosterone and estradiol in the presence of fibrils in solution. Positive response for the three hormones.

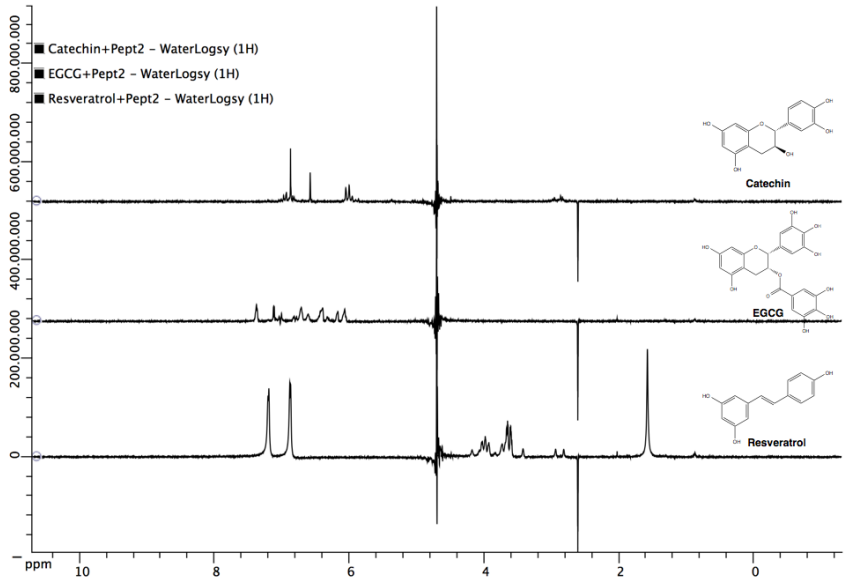



Figure C.5: WaterLogsy spectra for Catechin, EGCG and Resveratrol in the presence of fibrils in solution. Positive response for the three molecules.

The molecules were studied alone in solution in order to have a negative control of the WaterLOGSY response. Each molecule is compared with its corresponding 1D $^{1}$H. The DMSO signal at 2.65 ppm is used also as a negative control. All the molecules tested for this study displayed a negative WaterLOGSY signal without the presence of fibers in solution. (Figures C.6-C.11)

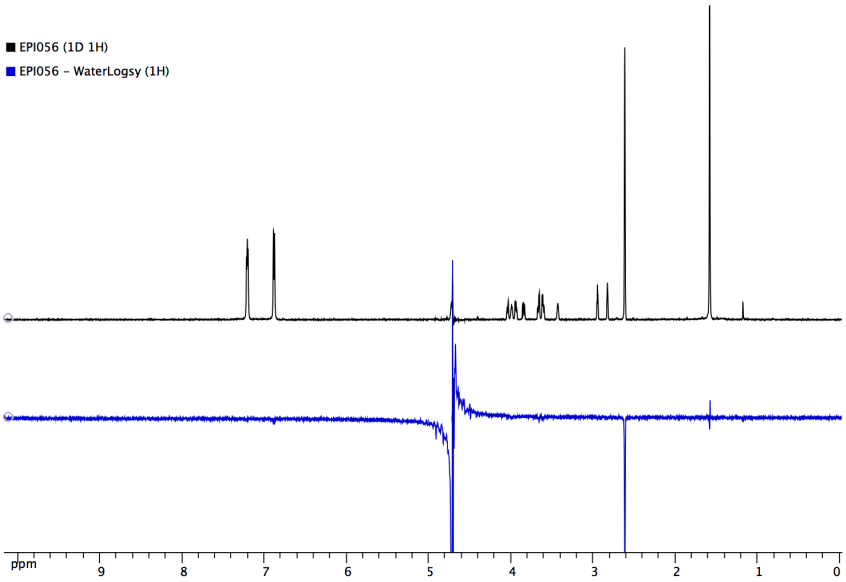



Figure C.6: WaterLogsy control for EPI–056.

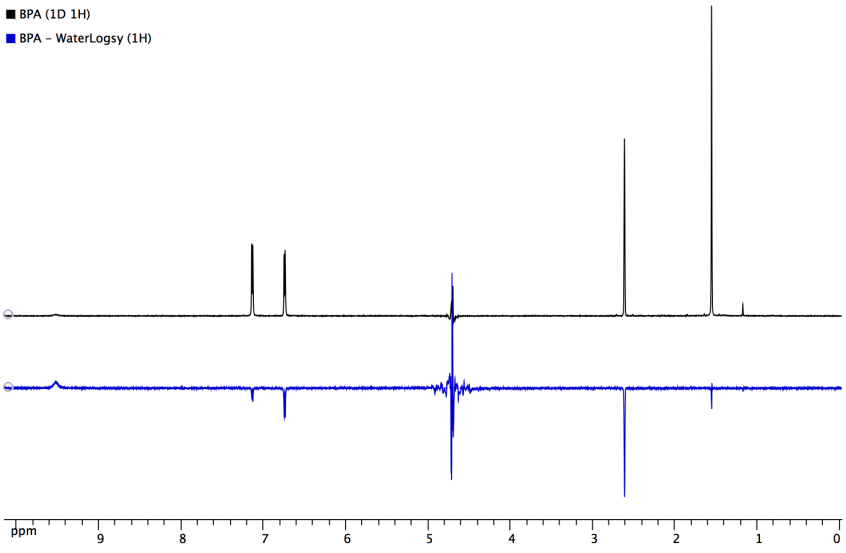



Figure C.7: WaterLogsy control for BPA.

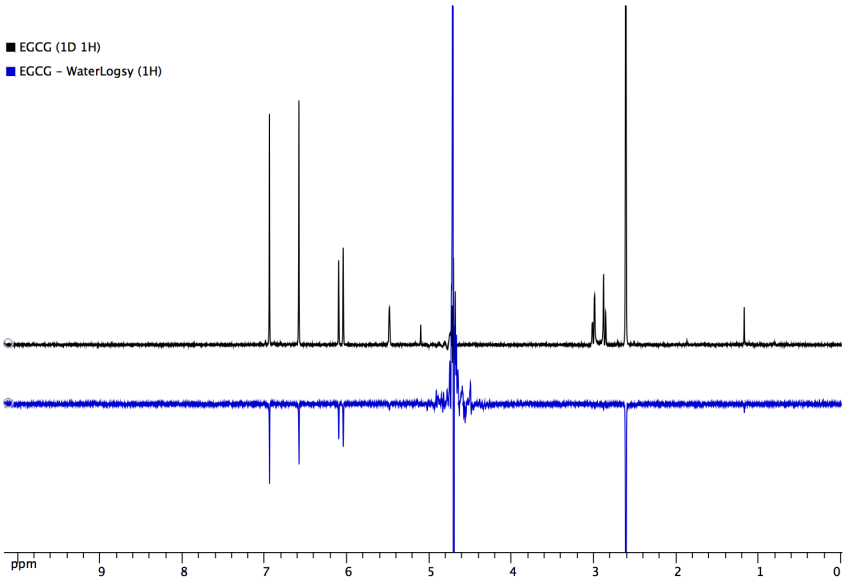



Figure C.8: WaterLogsy control for EGCG.

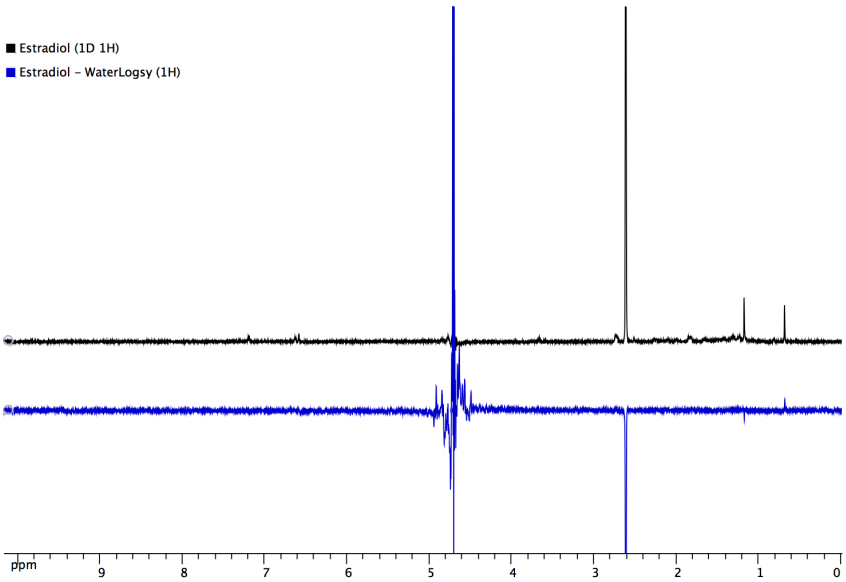



Figure C.9: WaterLogsy control for estradiol.

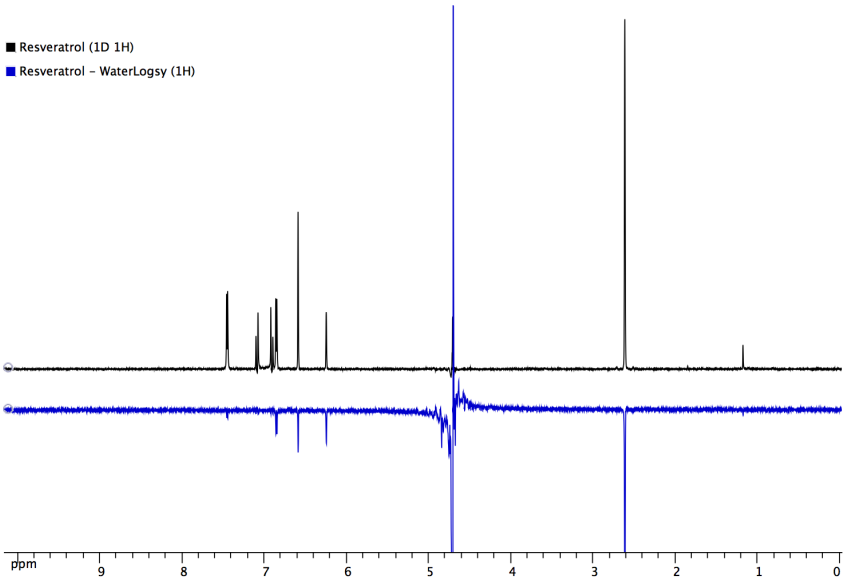



Figure C.10: WaterLogsy control for resveratrol.

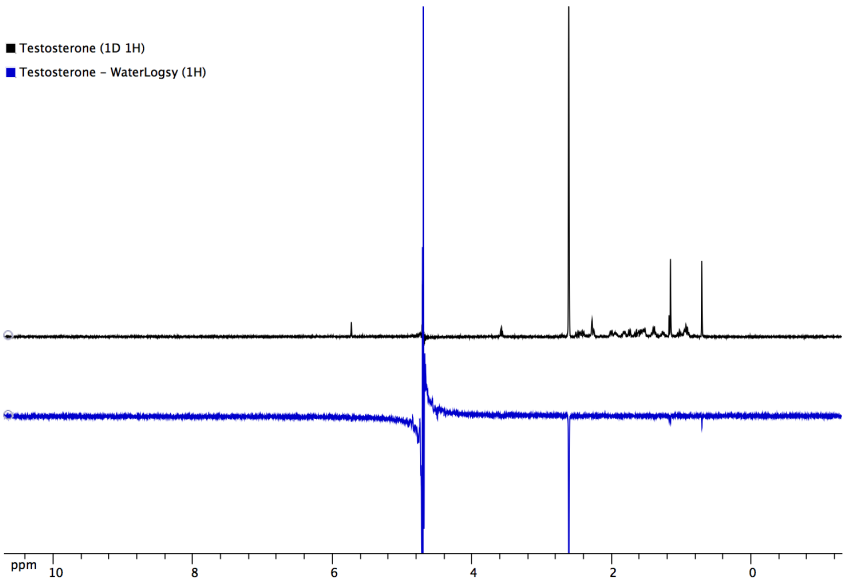



Figure C.11: WaterLogsy control for testosterone.

Chapter 7

# Amyloid fibers formation followed by *p*–DOSY



## 7.1 Amyloid fibers formation: kinetics and soluble oligomers intermediates

Amyloid fibers formation requires the ordering of the monomeric peptides into oligomers that later organize into fibrils. This ordering implies the formation of $\beta$-sheets secondary structures. Fibrillation in Amyloid-diseases have been widely studied. Inhibiting amyloid fibril formation is the primary therapeutic strategy for amyloid-related diseases such as Alzheimer [Estrada 2007, Peng 2015].

Fibrillation seems to occur *via* prefibrillar conformers. Two different nature of these intermediates have been proposed, fibrillation from denatured proteins state [Hamley 2007] or partially folded intermediates (usually rich in $\beta$-sheet conformation) [Uversky 2004, Gerum 2009].

The fibers formation process that has been proposed, commonly for most of amyloid-systems, is fibrillation after a lag phase. This mechanism suggest a nucleation step followed by a rapid growth process. On the other hand, this fibrillation kinetics can be modified with the addition of seed. This action can eliminate the lag time in the fiber formation process as seen in Figure 7.1.

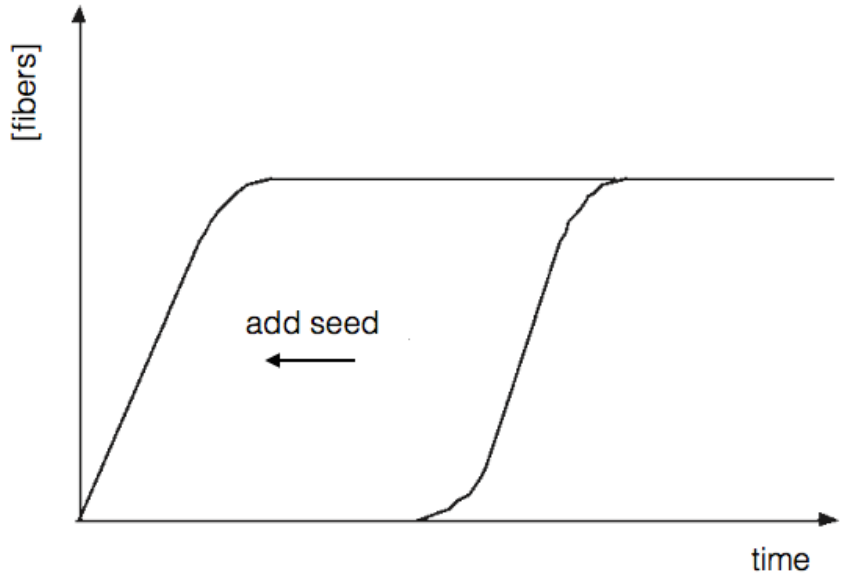



Figure 7.1: Seeding can eliminate the lag phase in the fibrillation kinetics.

A three-stage mechanism: protein misfolding, nucleation, and fibril elongation is usually proposed for amyloid fibril formation with a commonly observed sigmoidal-shaped kinetics [Lee 2007a]. Kinetics fibrillation implies complicated and different mechanism of aggregation, first-order kinetics have been also proposed for A$\beta$40 protein, with a rate proportional to the monomeric concentration for the elongation step [Hamley 2007].

Many other factors, such as pH, temperature, stirring or initial monomeric concentration may affect the fibrillation kinetics. It has been shown that agitation and sonication can affect, not only the kinetics of the fibers formation, but also to its morphology and its toxicity [Petkova 2005, Lee 2007b, Knowles 2014].

The prefibrillar aggregates are soluble oligomer that appear to be the major toxic agents in several degenerative disorders [Lee 2007b]. The structural plasticity associated with amyloid-proteins could lead the appearance of different oligomer structures [Glabe 2008]. This inherently misfolded nature could be the origin of its toxicity, as they display on different chemical surfaces that under normal physiological conditions [Campioni 2010], that can cause undesired interactions with new partners. Specially, their highly hydrophobic surfaces can induce aberrant interactions with other cellular components with its consequence damage. As aggregation produces an array of different oligomers, various types of toxic species can be expected [Knowles 2014].

## 7.2 Aggregation kinetics followed by *p*–DOSY

Previously, in Chapter 5 the kinetics aggregation of the peptide KELCKAVSVSM followed by 1D $^1$H NMR has been characterized. In that case, the signal disappearance from the 1D spectrum was measured. The structural information that was obtained from that study was the dimerization of the peptide through a disulfide bridge formation triggering the formation of amyloid fiber.

In order to better characterize the aggregation phenomenon and to study the possible presence of larger oligomers in solution before the aggregation, and further disappearance from the NMR the kinetics of oligomerization was followed by *p*–DOSY. This experiment, presented in Chapter 4 allows the precise determination of diffusion coefficients for out-of-equilibrium systems. This experiment allows the detection and characterization of the appearance of the dimer and different oligomers in solutions, with different diffusion coefficient.

The use of the *p*–DOSY experiment instead of traditional DOSY is to avoid possible artifacts or bias in the diffusion coefficient due to the concentration variation of the monomeric peptide form while acquiring the experiment, and as consequent misinterpretation of the results (as seen in Chapter 4).

The kinetics was monitored during 50 hours, with 2.5 h *p*–DOSY experiments alternated with 1D $^1$H spectra zgpr.

The results of the evolution on the peptide diffusion coefficient ($D$) during the kinetics aggregations are shown in Figure 7.2. A decreasing of the apparent diffusion coefficient is monitored during the firsts 40 hours. This result indicates that new bigger species are appearing in solution.

Looking into the spectra, a difference on $D$ can be easily observed for the Cys residue. While all protons signals evolve with the same rate and with the same diffusion, the peaks at 2.86 ppm and 3.15 ppm evolve in a different manner. It can be observed in Figure 7.2, where the signal marked in dashed green, corresponds to the Cys–H$_\beta$ of the monomer and the signal marked in dashed orange refers to the Cys–H$_\beta$ of the dimer. This shows that the monomer and the dimer are stable during the kinetics, and do not form complexes of intermediate size. The other signals from the peptide are showing a varying $D$, it comes from the fact that they are measured as the superimposition of the signals of the monomer and the dimer. Due to this superposition, and the small difference between the $D$ of both species, the apparent diffusion coefficient which is measured on this superposition is a weighted sum of the abundance of both species, and thus "weights" the ratio monomer to dimer.

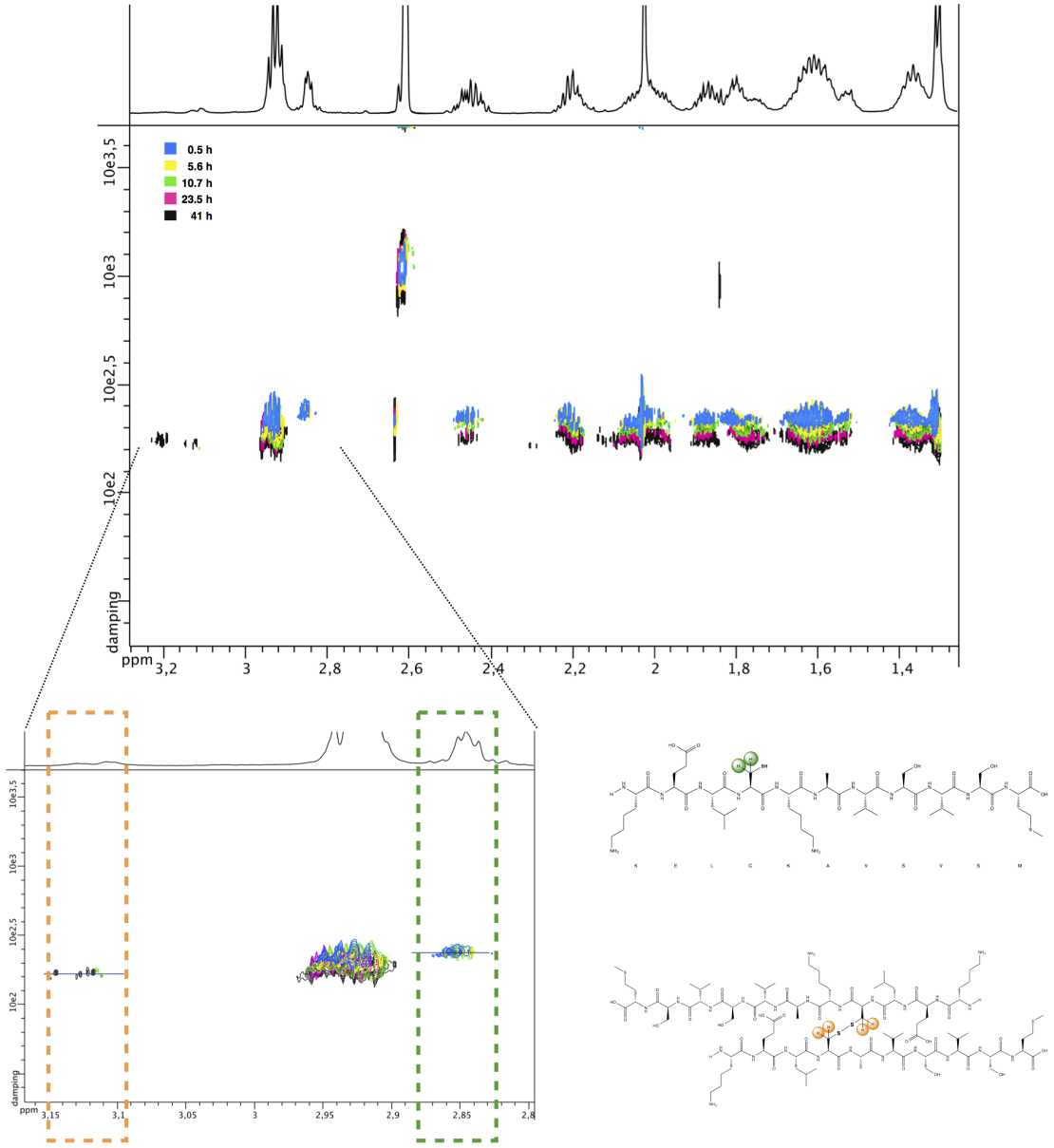



Figure 7.2: *p*–DOSY spectra monitoring peptide **2** kinetics. Zoom of the Cys–$H_\beta$ protons, difference between monomeric and dimeric forms in $\delta$ and $D$.

## 7.3 Analysis and fitting of the Diffusion coefficient ($D$) evolution

The diffusion coefficient of an objet is related to its molecular mass. $D \propto M^{1/\alpha}$ where $\alpha$ is the fractal dimension [Augé 2009]:

$$\frac{D_1}{D_2} = \left(\frac{M_2}{M_1}\right)^{1/\alpha} \tag{7.1}$$

Where $D_1$ and $D_2$ are the different diffusion coefficient of the molecules compared and $M_1$ and $M_2$, their respective molecular masses.

This eq 7.1 is related to the Flory theory for polymers where the gyration radius, $R_g = M^{\delta}$ and $\delta$ is the inverse of the fractal dimension.

Flexible random chains have an alpha comprised between 3 (compact chain), and 5/3 (largest possible extension of a random walk), with 2 for a purely gaussian random walk. In the absence of structural information, an alpha around 2 is usually chosen. The fractal dimension can be comprised between: $5/3 < \alpha < 3$.

To analyze the variation of the apparent diffusion coefficient it is important to consider a correction for viscosity variation due to the appearance of fibrils in the solution. For this reason, a correction of $D$ is performed regarding the diffusion coefficient of DMSO. DMSO is experiencing the same viscosity changes in solution nevertheless it is not experiencing any other interaction. From the analysis of the $D$ evolution by time, a 30% modification of the peptide diffusion between the first and last point of the kinetics is found. This effect was analyzed to estimate the difference in size of the specie in solution. And the result is that a dimer is appearing in solution. This corresponds to a mass increment of 1.7 for $\alpha = 2$ ($1.7^{(1/2)} = 1.3$) which is not feasible. The formation of the dimer corresponds to $\alpha$=2.6 which is a bit high, but corresponds nicely to the value found for structured proteins [Augé 2009]. In any case, the formation of a larger soluble oligomer can be fully ruled out.

The kinetics curve (Figure 7.3) obtained for $D$ is fitted with an exponential law, as follow:

$$D(t) = D_{mono} exp(-k_{oligo} t) + D_{oligo}$$

The results of the fit are shown in Figure 7.4 and are:

k = 0.088 $h^{-1}$

$D_{mono}$ = 242.7 $\mu m^2 s^{-1}$

$D_{olig0}$ = 189.7 $\mu m^2 s^{-1}$

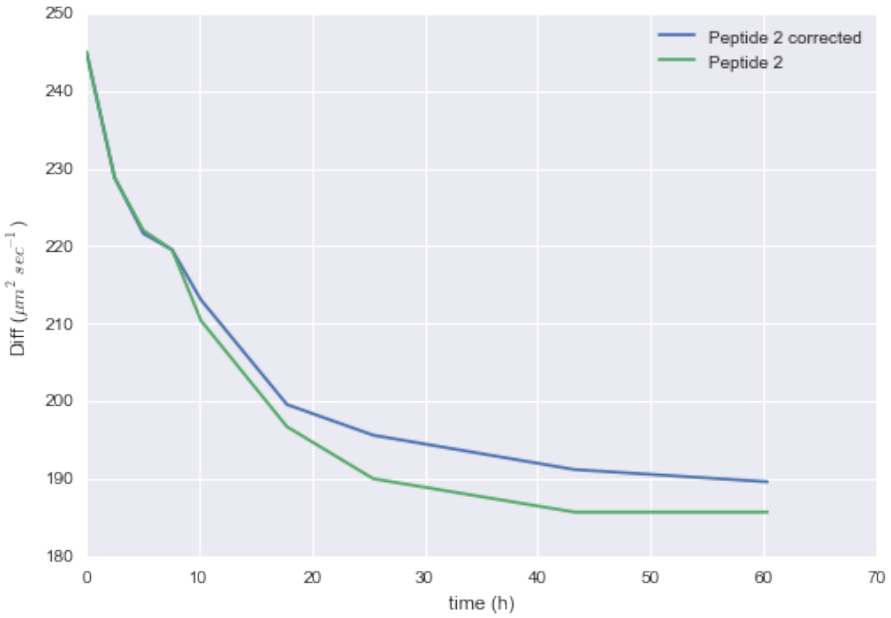



Figure 7.3: Evolution of the diffusion coefficient of the peptide by time. The green line shows the diffusion coefficient value measured and the blue line shows the diffusion coefficient corrected with the solvent diffusion value.

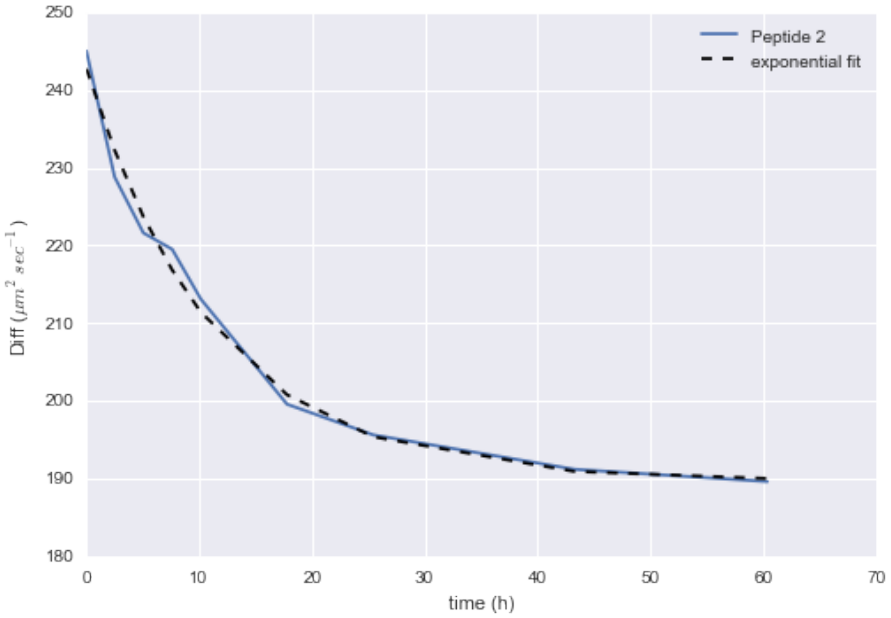



Figure 7.4: Evolution of the diffusion coefficient of the peptide by time. The blue line shows the diffusion coefficient corrected with the solvent diffusion value. The dashed line shows the exponential fit of the curve.

These results put in evidence that the kinetics aggregation recorded by *p*–DOSY (1st order kinetics) and the kinetics aggregation followed by standard 1D $^{1}$H (2nd order kinetics) does not display the same behavior.

## 7.4 Peptide vs Cysteine kinetics

In order to complete the characterization of the different kinetics laws. The 1D spectra recorded after each *p*–DOSY experiment were analyzed. From them was extracted the kinetics for the peptide concentration disappearance by time (using the methyls integrals as in Chapter 5 and 6). But also the kinetics of the disappearance of the Cys–H$_\beta$ signal (Figure 7.5).

In Figure 7.6 is shown the evolution of the soluble forms of the peptide concentration during the 50 hours of kinetics. The curve follows a second-order law and the kinetics does not follow an exponential law. These results are equivalent to the ones shown in Chapter 5. Nevertheless, Figure 7.5 shows the evolution of the cysteine protons, close to the disulfide bridge formation, during the kinetics. This time, the curve does not follow a second-order kinetics and fits perfectly to a first-order reaction with an exponential law.

The results of the fit are shown in Figures 7.5 and 7.6 and are:

**Exponential fit**

k = 0.088 $h^{-1}$

$[Conc]_0$ = 1.07 mM

**Second-order fit**

k = 0.036 $mM^{-1}h^{-1}$

$[Conc]_0$ = 0.98 mM

These results demonstrate that the phenomenon characterized with the *p*–DOSY experiment is the same that the one displayed by the Cys peak in the 1D $^{1}$H experiment. Both kinetics display an exponential profile with the same constant rate $k = 0.088$. Indicating that we observe the dimerization process by *p*–DOSY and with the disappearance of the Cys signal.

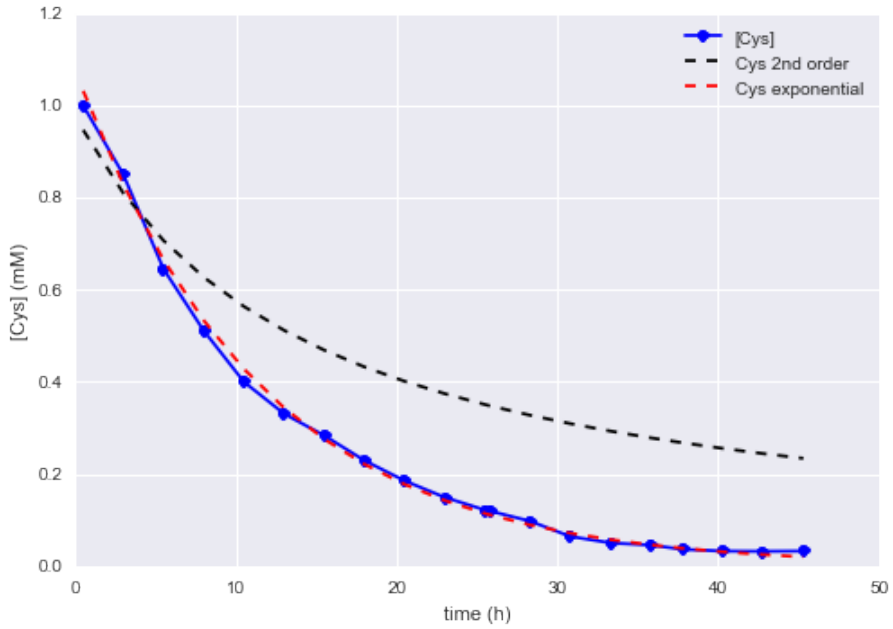



Figure 7.5: Evolution of the Cys–$H_\beta$ concentration by time. The blue line shows the diffusion coefficient corrected with the solvent diffusion value. The dashed red line shows the exponential fit and the dashed black line shows the second-order fit of the curve.

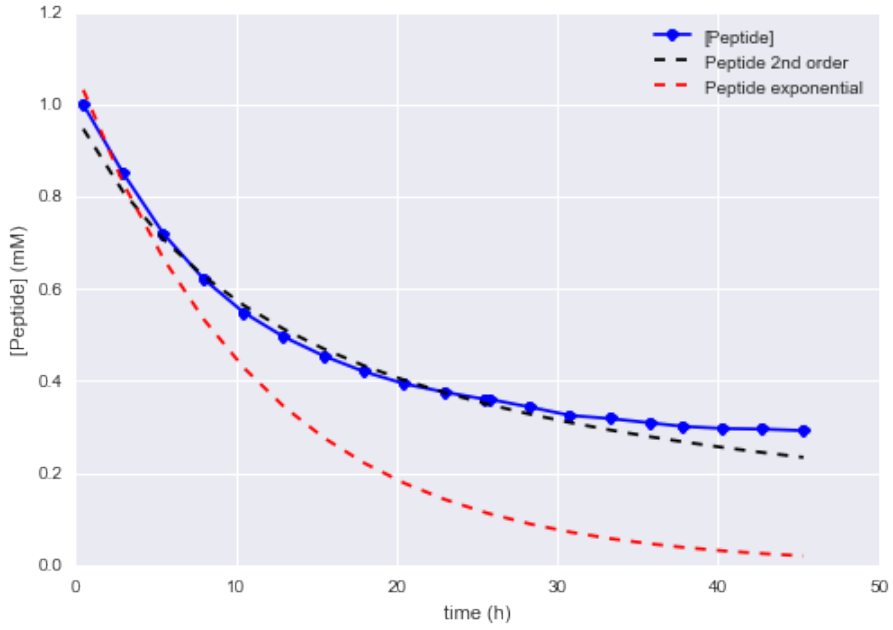



Figure 7.6: Evolution of the peptide concentration by time. The blue line shows the diffusion coefficient corrected with the solvent diffusion value. The dashed red line shows the exponential fit and the dashed black line shows the second-order fit of the curve.

## 7.5 Conclusions

This chapter put in evidence the coexistence of two independent mechanisms. The fact that the *p*–DOSY experiment, where is followed the appearance of bigger objects in solution, displayed a exponential kinetics. Together with the same behavior found for the Cys signal in the $^1$H 1D experiment, where the concentration of the peptide is followed by time. These two results enhance the existence of a different mechanism for they dimerization step where the cysteine is involved through the disulfide bridge formation. Whereas the aggregation of the monomeric form into fibers, follow the second-order kinetics described until now in this study. Finally, the evolution of the diffusion coefficient indicates that the bigger object that is formed in solution before aggregation is the dimer and no soluble large oligomers are detectable by NMR. This means that either there are none, or that there are large and rigid oligomers, with $T_2$ too short to be detectable by DOSY/NMR. The DOSY sequence, with the spin echoes, acts as a $T_2$ filter, filtering out the larger lines, so linewidths (and corresponding $T_2$) of a few dozen Hz will be enough to give vanishing small signals in DOSY. For an observation at 700MHz, these linewidths correspond to molecular systems in the 10-20kD range and larger, excluding tetramer and larger soluble oligomers. This give new crucial information for the development of a model for the fibrillation of this peptide.

# Bibliography

[Augé 2009] Sophie Augé, Pierre-Olivier Schmit, Christopher A Crutchfield, Mohammad T Islam, Douglas J Harris, Emmanuelle Durand, Martin Clemancey, Anne-Agathe Quoineaud, Jean-Marc Lancelin, Yann Prigent, Francis Taulelle and Marc-André Delsuc. *NMR measure of translational diffusion and fractal dimension. Application to molecular mass measurement.* J. Phys. Chem. B, vol. 113, no. 7, pages 1914–1918, February 2009. (Cited on pages 58, 150 and 151.)

[Campioni 2010] Silvia Campioni, Benedetta Mannini, Mariagioia Zampagni, Anna Pensalfini, Claudia Parrini, Elisa Evangelisti, Annalisa Relini, Massimo Stefani, Christopher M Dobson, Cristina Cecchi and Fabrizio Chiti. *a causative link between the structure of aberrantprotein oligomers and their toxicity*. Nature Chemical Biology, vol. 6, no. 2, pages 140–147, January 2010. (Cited on page 148.)

[Estrada 2007] L Estrada and C Soto. *Disrupting $\beta$-Amyloid Aggregation for Alzheimer Disease Treatment*. Current Topics in Medicinal Chemistry, vol. 7, no. 1, pages 115–126, January 2007. (Cited on page 147.)

[Gerum 2009] Christian Gerum, Robert Silvers, Julia Wirmer-Bartoschek and Harald Schwalbe. *Unfolded-State Structure and Dynamics Influence the Fibril Formation of Human Prion Protein*. Angew. Chem. Int. Ed., vol. 48, no. 50, pages 9452–9456, October 2009. (Cited on page 147.)

[Glabe 2008] Charles G Glabe. *Structural classification of toxic amyloid oligomers.* J. Biol. Chem., vol. 283, no. 44, pages 29639–29643, October 2008. (Cited on page 148.)

[Hamley 2007] Ian W Hamley. *Peptide Fibrillization*. Angew. Chem. Int. Ed., vol. 46, no. 43, pages 8128–8147, 2007. (Cited on pages 147 and 148.)

[Knowles 2014] Tuomas P J Knowles, Michele Vendruscolo and Christopher M Dobson. *The amyloid state and its associationwith protein misfolding diseases*. Nature Publishing Group, vol. 15, no. 6, pages 384–396, June 2014. (Cited on pages 18 and 148.)

[Lee 2007a] Chuang-Chung Lee, Arpan Nayak, Ananthakrishnan Sethuraman, Georges Belfort and Gregory J McRae. *A three-stage kinetic model of amyloid fibrillation.* Biophysj, vol. 92, no. 10, pages 3448–3458, May 2007. (Cited on page 148.)

[Lee 2007b] S Lee, E J Fernandez and T A Good. *Role of aggregation conditions in structure, stability, and toxicity of intermediates in the Abeta fibril formation pathway*. Protein Sci., vol. 16, no. 4, pages 723–732, February 2007. (Cited on page 148.)

[Peng 2015] Jian Peng, Yunjing Xiong, Zhiqin Lin, Liping Sun and Jian Weng. *Few-layer bismuth selenides exfoliated by hemin inhibitamyloid-*. Sci. Rep., pages 1–17, May 2015. (Cited on page 147.)

[Petkova 2005] Aneta T Petkova, Richard D Leapman, Zhihong Guo, Wai-Ming Yau, Mark P Mattson and Robert Tycko. *Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils.* Science, vol. 307, no. 5707, pages 262–265, January 2005. (Cited on page 148.)

[Uversky 2004] Vladimir N Uversky and Anthony L Fink. *Conformational constraints for amyloid fibrillation: the importance of being unfolded*. Biochim. Biophys. Acta, Proteins Proteomics, vol. 1698, no. 2, pages 131–153, May 2004. (Cited on page 147.)

Chapter 8

# NMR study of spontaneous Acetonitrile-Water demixing



This chapter marks a turning point in this thesis. This study started as a collaboration with Jean-Marie Lehn who brought us the opportunity to characterize the special behavior that presents the mixture acetonitrile-water at low temperature. This different topic, represents in a perfect way what a complex system is, and how from an apparent simple system we can observe interesting and difficult results. I decided to include this work as a final chapter of my thesis, not only because we have invested a lot of time on it, but also because it has been a passionate work the fact of searching new ideas, hypotheses or experiments that could elucidate or model the phenomenon we are observing. Along this three years we have asked experts from different area to collaborate and to help us to figure out the phenomenon we were observing. Hervé Desvaux, Daniel Gruker, François Xavier Courdet and Marc-André Delsuc, whom have contributed with ideas, solutions or mathematical descriptions and together with we present this results as a manuscript for a future publication.

Up to this point, in the thesis we have been focused in complex systems by a biological point of view. As has been discussed in Chapter 1, complex systems are present all around us. Within these large multiple-component systems, many parts interact in a non-linear manner and the weak interactions between their parts are crucial for the development and the emergence of new properties.

In this hierarchic of molecules organization we focus now on a *simple* organization in chemistry, such as solvents.

These, *a priori*, simple systems can develop and adopt different complex conformations under different conditions that could affect to further analysis of experimental results. For this reason, we investigate here the acetonitrile-water mixture, which represents the most used mixture of solvents in chemistry and displays a non-well known behavior at low temperature, nowadays, still difficult to interpret.

## 8.1 Introduction

Acetonitrile is one of the most important solvents for all the chromatographic methods and specially in pharmaceutical industry. So much so that its recovery from waste has been investigated [McConvey 2012]. It is perfectly miscible with water and this mixture is commonly used at ambient temperature. This systems is supposed to be well known and well biophysically characterized for its wide used in chemistry, nevertheless it displays a strange and special behavior at low temperature.

Running NMR experiment in a deuterated acetonitrile-water mixture, one of us noticed that temperature variations below circa 10°C could lead to the loss of the deuterium lock and to strong spectral artifacts. This strange phenomenon could be attributed to a known anomaly of the acetonitrile-water mixture at low temperature. The acetonitrile-water mixture has been described to undergo a phase transition around 0°C, for a range of acetonitrile mole fraction $x_{AN}$ from 0.2 to 0.6, characterized by a spontaneous phase separation below an upper critical solution temperature (UCST). UCST was shown to raise by several degrees in deuterated solvents and the complete phase diagram has been determined [Szydłowski 1999].

This transition has been investigated by Small Angle X-Ray and Neutron Scattering [Takamuku 1998, Takamuku 2001, Takamuku 2007], spectroscopic techniques, [Bertie 1997] and computer modeling [Mountain 1999] (see [Marcus 2013] for a comprehensive review). However, none of this studies showed an effect as massive as the one observed by NMR, for this reason, we decided to study further this behavior using NMR spectroscopy.

In order to characterize this phase transition and the possible structure of the mixture at

low temperature a series of different NMR experiments are acquired.

## 8.2 Results

### 8.2.1 1D $^1$H experiments and Isotopic effect

In order to characterize this phase transition and the possible structure of the mixture at low temperature a series of different NMR experiments have been acquired.

Above the transition temperature, the acetonitrile-water mixture presents a very simple NMR spectrum composed of two singlets, with the water signal located around 4 ppm and the acetonitrile signal around 2 ppm, the details of the spectrum depending on the molar ratio and on the temperature at which the experiment is recorded. The $^1$H spectrum of the deuterated species shows the residual protonated species, and the acetonitrile displays a 1–2–3–2–1 coupling pattern, typical of the coupling of the proton with 2 deuterons.

When gradually lowering the temperature, at the transition temperature, a large modification of the spectrum appears, with a splitting of the two solvent lines and patterns appearing, reminiscent to solid-state powder patterns. The experiment was recorded for the mixture of the protonated and the deuterated solvents, and are shown in Figure 8.1. These figures present the $^1$H spectra, but the $^2$D spectra recorded in the same conditions are equivalent.

The splitting of the signal shows that the system is separated in at least 2 independent phases. The powder pattern observed for these lines is the signature of an orientational anisotropy. There are presented with opposite forms, and the width of the pattern was observed to extend up to about 250 Hz for the lower temperature. The fact that both water and acetonitrile present the same patterns could indicate the system is a liquid-liquid phase separation, where each liquid is a mixture of water and of acetonitrile.

The patterns obtained are not fully reproducible, and depend not only on the molar ratio but also on the spectrometer, the tube diameter and the speed at which the temperature is cooled down. It is probably related to the homogeneity of the temperature over the sample volumes. After several trials, the best results were obtain on 3 mm NMR tubes, in a regular TXI 5 mm probe, operated on a 500 MHz instrument. When varying the temperature several time below and above the transition, it was observed that the system could retain at high temperature some memory of the low temperature details. This memory could easily be destroyed by pulling the sample tube out of the spectrometer and shaking it before putting it back.

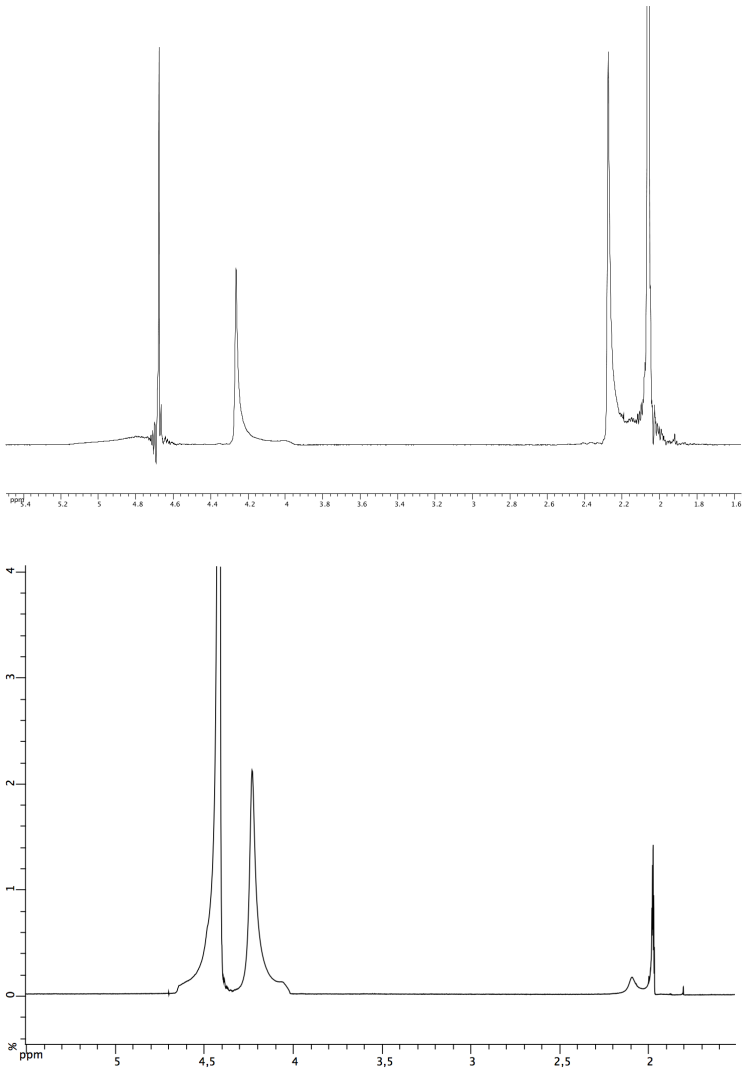



Figure 8.1: 1D $^{1}$H spectra of the acetonitrile-water mixture; left: protonated acetonitrile-water 70% - 30% vv at 0 °C; right: deuterated acetonitrile-water 70% - 30% vv at 6 °C.

Figure 8.2 presents the variation of the $^{1}$H spectrum when varying slowly the temperature, and crossing the transition temperature for a deuterated acetonitrile-water mixture. Above the transition, the water signal is shifted to the left, and present some small distortions, probably due to slight temperature variation in the sample tube. At the transition, both water and acetonitrile (not visible on this figure) present a splitting which then increase sharply for lower temperatures.

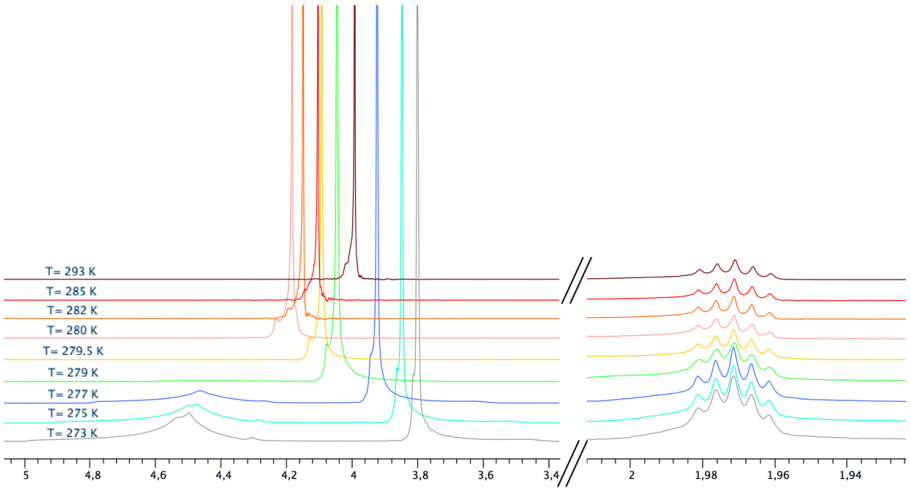



Figure 8.2: Evolution of the NMR 1D $^1$H spectrum of the acetonitrile-water mixture ($x_{AN} = 0.45$) from 293 K down to 273 K performed in a 5 mm tube. Water signal is on the left, acetonitrile is on the right (notice the difference in horizontal scale). The instrument is locked on the acetonitrile signal. The stable coupling pattern of the deuterated acetonitrile is indicating a stable shim condition.

### 8.2.2 Phase diagram

The transition observed in the NMR spectra is very characteristic and can be followed carefully. Figure 8.3 shows the position of the different NMR signals over temperature. The pattern is characteristic of a phase transition, and the temperature transition seems rather independent on the ratio of solvents used.

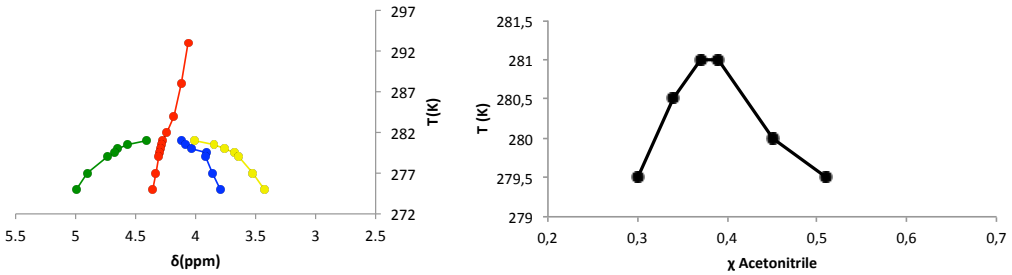



Figure 8.3: Left, evolution of the position (in ppm) of the water line with the temperature (in K), for the 65%-35% (v/v) deuterated acetonitrile-water mixture; the 4 points correspond to the 4 characteristics of the water lines: extremities of the patterns (green and yellow dots) and internal sharp lines (red and blue dots). Right, observed transition temperature for various acetonitrile-water ratios.

The position of the transition temperature has been carefully measured, and is reported in Figure 8.3. The values obtained are very close to the values reported by Szydiowski et al [Szydłowski 1999]. In this study, the authors measured the phase diagram of the transition for protonated and deuterated samples, and showed that the deuterated liquids present a transition temperature quite higher than the protonated ones. We recovered the same results.

### 8.2.3 EXSY

The EXSY (EXchange SpectroscopY) experiment is a 2D NMR experiment which measures the presence of chemical exchange in a system. The sequence is very similar to the NOESY experiment, and contains a mixing period over which the chemical exchange is sampled.

Only one peak could be clearly seen on the EXSY spectrum Figure 8.4, connecting the two longitudinal singularities of the water patterns.

The absence of signal around the powder pattern indicates that there is no reorientation of the cause of the spectral anisotropy for a time period on the order of the second. This rules out the possibility of short range anisotropy, and indicate that the anisotropies are probably on long range distance, which are not perturbed by the (slow) convection found in a 3 mm tube at low temperature. This correlates well with the memory effect, mentioned in the first paragraph.

However, in this context, the exchange peak is difficult to explain, without even a simple molecular model of the two phases, and of their interphase.

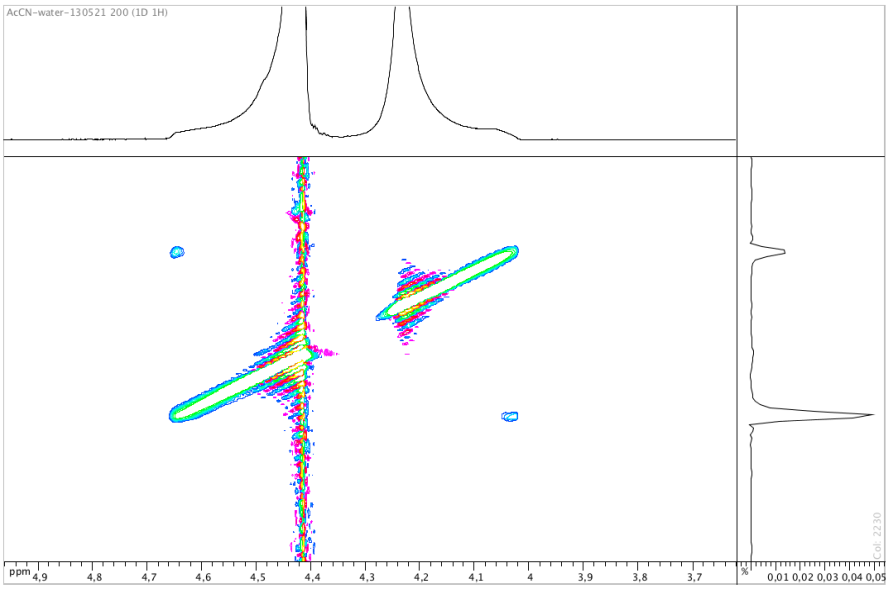



Figure 8.4: The EXSY spectrum of the acetonitrile-water mixture (65%-35% v/v) , here run with a 1 sec mixing time, showing a exchange peak between the singularities of the powder patterns of the two phases

### 8.2.4 DOSY

The DOSY experiment is used for measuring diffusion coefficients of the species in solution. The experiment is not very easy to perform on broad lines, nevertheless three components can be observed in the DOSY, with the same pattern in the water and acetonitrile lines.

Figure 8.5 shows the DOSY spectrum obtained for the acetonitrile-water mixture (65%$_v$-35%$_v$). We can distinguish three different behaviors, one narrow line with a large diffusion coefficient, comparable to the free fluid; and two different coefficients for the left and right "horns" of the powder pattern for both water and acetonitrile, with values corresponding to much more viscous fluids.

We could not detect signs of non-exponential decay in this experiment, which you hint to diffusion taking place in a volume restricted to the mean displacement path explored by this experiment, on the order of 5 $\mu$m for the slow coefficients and 10 $\mu$m. This indicates that the phase domains are quite smaller than this order of magnitude.

With the display of these three different diffusion coefficients, this experiment tends to proves that three different liquids are present in the sample, two complex mixed phases, and a free bulk isotropic phase. Again, the exchange between the three media is slow, as the diffusion coefficients, measured here over a diffusion delay of $\Delta = 50$ msec are not blurred by exchange

over this period.

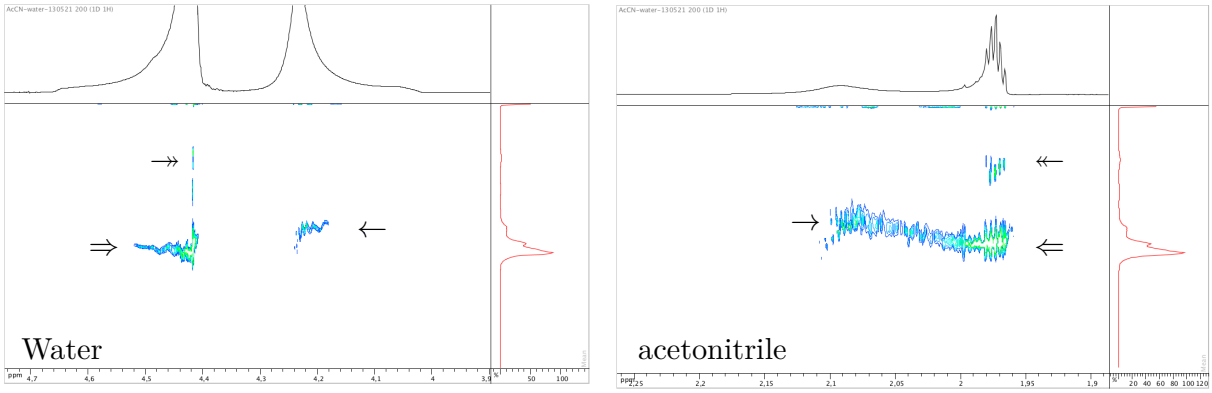



Figure 8.5: DOSY spectrum, showing the rapid (↠, $D_{AcCN} = 1797.7\mu m^2 s^{-1}$ $D_{H_2O} = 2255\mu m^2 s^{-1}$), medium fast (→, $D_{AcCN} = 701.8\mu m^2 s^{-1}$ $D_{H_2O} = 644.9\mu m^2 s^{-1}$) and medium slow (⇒, $D_{AcCN} = 483.2\mu m^2 s^{-1}$ $D_{H_2O} = 443.8\mu m^2 s^{-1}$) components. Notice the inverted pattern between water and acetonitrile .

### 8.2.5 CRAZED

It is known that in a structured heterogeneous sample, the effect of dipolar field created my the induced magnetization through the sample does not reduces to zero. This so-called distant dipolar field (DDF) creates non-linear effect through the sample, which are very sensitive to local organization of matter. These effects can lead to additional echoes and are sometimes called intermolecular multiple-quantum coherences (iMQC).

This has been studied, in particular by W.S.Warren since 1990, who has developed the CRAZED (COSY revamped by asymmetric Z-gradient) approach, and the full understanding of this effect is still under discussion. It has been used as a source of contrast in MRI ([Richter 2000, Branca 2009]) and can lead to diffraction-like patterns in the NMR experiments (see [de Sousa 2004]). CRAZED experiment was performed as detailed in Figure 8.6.

In the CRAZED experiment, multi-spin states, coupled through space by DDF create multiple-order $n$-order quantum resonance after the first $P_{90}$ pulse. The second $P_{90}$ pulse converts the multi-order states to detectable single quantum states. The detection being sensitive only to $-1$ spin-order, the ratio of the two gradients selects which coherences are sampled during the $t_1$ delay. This delay is regularly incremented, and the Fourier transform allows measuring the frequencies of the otherwise undetectable multiple order states.

On the experiment with a gradient ratio $G_2/G_1$ of 2, the stronger signals are located on

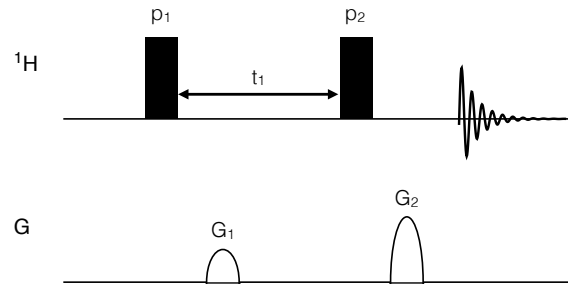



Figure 8.6: NMR pulse sequence for the CRAZED experiment. The $t_1$ delai is regularly incremented, and a Fourier transform is applied along the $F_1$ axis. Pulse phases are for both $P_{90}$ pulses : {X X $\bar{\text{X}}$ $\bar{\text{X}}$ Y Y $\bar{\text{Y}}$ $\bar{\text{Y}}$}, $Rec$ : {X X $\bar{\text{X}}$ $\bar{\text{X}}$ Y Y $\bar{\text{Y}}$ $\bar{\text{Y}}$} Gradients are fixed, with ratio presented in table 8.1.

Table 8.1: Gradient intensities used for the CRAZED experiments

| exp. label | $G_1$ | $G_2$ | $G_2/G_1$ |
|---|---|---|---|
| a) | 15% | 30% | 2 |
| b) | -15% | 30% | -2 |
| c) | 10% | 30% | 3 |

the line $\omega_1 = +2\omega_2$. corresponding to 2-spin order. Surprisingly, signals are also observed on the other diagonals. In the case of the experiment with a gradient ratio $G_2/G_1$ of -2, only signals located on the line $\omega_1 = -2\omega_2$ can be observed. The experiment with a gradient ratio $G_2/G_1$ of 3 is similar to the one with a ratio of 2, but the larger peaks are now located on the $\omega_1 = 3\omega_2$ line. The experiment performed with both gradient inverted for the same ratio, are equivalent to the one presented here.

These spectra are the signature of multiple spin order created through the DDF effect, are characteristic of a medium with a spatial structuration. No signals are observed at positions corresponding to a mixing of the frequencies of water and acetonitrile, indicating the absence of intermolecular spin order.

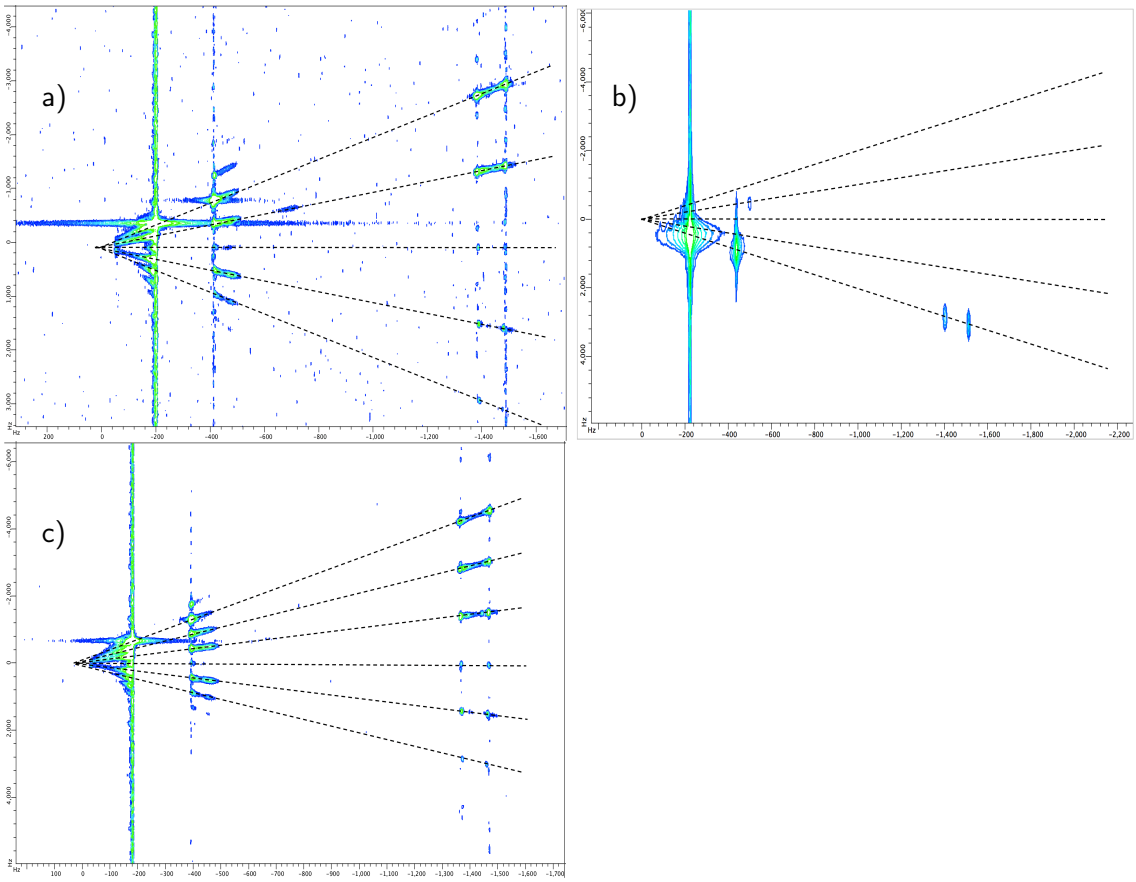



Figure 8.7: spectrum obtained for the CRAZED experiments. The experiments a-c correspond to the conditions listed in the table 8.1. On each spectra, the position of the $\omega_1 = 0$, $\omega_1 = \pm\omega_2$, and $\omega_1 = \pm 2\omega_2$, have been noted. On c), we noted also $\omega_1 = +3\omega_2$

### 8.2.6 STRANGE

Looking to the DOSY acquisition we noticed a strange behavior, associated to the gradient strength. We investigated this aspect by running the sequence presented in Figure 8.8, that we named STRANGE. This sequence, where the gradient intensity is varied is reminiscent to the coding module of the DOSY experiment, and can be seen as a minimum phase encoding imaging sequence. It should have no effect on an homogeneous sample, except a rapid decrease of the NMR signal with increasing gradient intensities.

This sequence, on a homogeneous sample, should produce a complete signal extinction as soon as the gradients are strong enough. Here, when varying the gradient intensity, even to high values, we obtained strong modulation of the line shape (see Figure 8.9). This modulation is very different from what is described in the literature for the CRAZED effect, and presents much more structure.

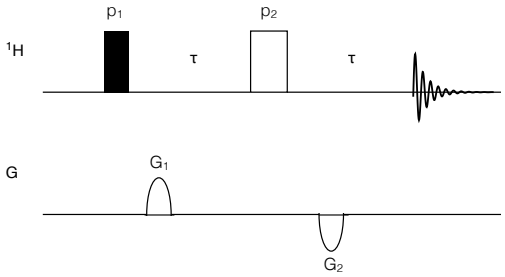



Figure 8.8: NMR pulse sequence for the STRANGE experiment. The delay $\tau$ is kept constant, and the gradient $G_1$ and $G_2$ are of inverted and of equal absolute intensity, and linearly increased from 0 to the final value. Pulse phases are $P_{90}$ : {X X $\bar{\text{X}}$ $\bar{\text{X}}$ $\bar{\text{Y}}$ $\bar{\text{Y}}$ Y Y}, $P_{180}$ : {X $\bar{\text{X}}$}$_4$, $Rec$ : {X X $\bar{\text{X}}$ $\bar{\text{X}}$ Y Y $\bar{\text{Y}}$ $\bar{\text{Y}}$}

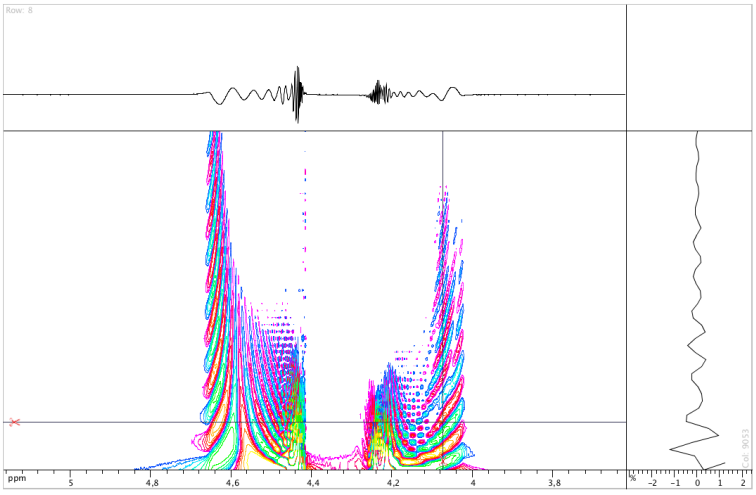



Figure 8.9: modulation of the signal with gradients varying linearly from $G = 0$ to $\approx 25\,G/cm$ applied during $d = 2 \times 1\,msec$. All delays are kept constant. The spectrum was obtained after a simple Fourier transform of the FID along the $F_2$ (horizontal) dimension, no processing was applied on the vertical axis. The top and side spectra correspond respectively to a slice of the 2D at $G = 5\,G/cm$ and to $\delta = 4.07\,ppm$

Doing a phase modulated Fourier transform of the gradient axis produces the images presented in Figure 8.10. These STRANGE spectra are perfectly reproducible, the experiment was run many times, on different days, with varying parameters (for instance to optimize the vertical "field of view"). It was also repeated several time overnight with no visible modification

of the patterns.

Looking to this picture, several striking features appear. In this spectrum, the structure of the powder pattern in both massifs is much easier to decipher. The splitting for water and acetonitrile appear equivalent, as well as the shape present in the STRANGE spectra, with simply an inversion of the position. The system is very regular and organized, and the linewiths are very narrow, for instance, the acetonitrile coupling pattern is fully visible every where - whereas the 1D spectrum is indistinctly broad.

There are also non modulated signals (the only one that would appear in an homogeneous sample) pointed with an arrow symbol in the figure. This homogeneous medium seems to appear only in one of the two phases.

It seems to show that there are several superposed structures, and the same features, that we we call "trajectories" can be found similar on both water and acetonitrile. The same tree different "trajectories" on the left pattern of water and right for the acetonitrile can be isolated. Also, two different "trajectories" on the right pattern of water and left for the acetonitrile are seen, and a possible additional signal around 4.05 ppm for the water (1.90 ppm for the acetonitrile)

On the right (left) pattern of water (acetonitrile) it seems to be two additional patterns: *i)* the homogeneous line *ii)* a kind of "comma". In contrast with the other features, these two additional patterns appear unchanged in the positive and negative gradient experiments (see below).

We do not have yet any explanation for the pattern appearing in the STRANGE experiment. However, if the chemical shift (horizontal axis) is modulated by some angle, and if the STRANGE effect (vertical) is modulated by the local structure, then this experiment contains information on the structure of the phases of the mixtures.

The inversion of the coding gradient modifies completely the pattern (Figure 8.11), as if we were looking at two faces of the same object. On the inverted gradient experiment, three narrow lines appear at a negative frequency in the water region (nothing in the acetonitrile) (pointed with the single arrow in Figure 8.11)

Changing the second pulse of the STRANGE experiment from 180° to 90° does not modify the shape of the spectrum, but only lowers the overall intensities. Complete suppression of this second pulse and one of the gradients, results in an empty spectrum, where the STRANGE effect disappears. On the other hand, pulsing on the deuterium channel, in any manner, has no effect on the STRANGE spectra.

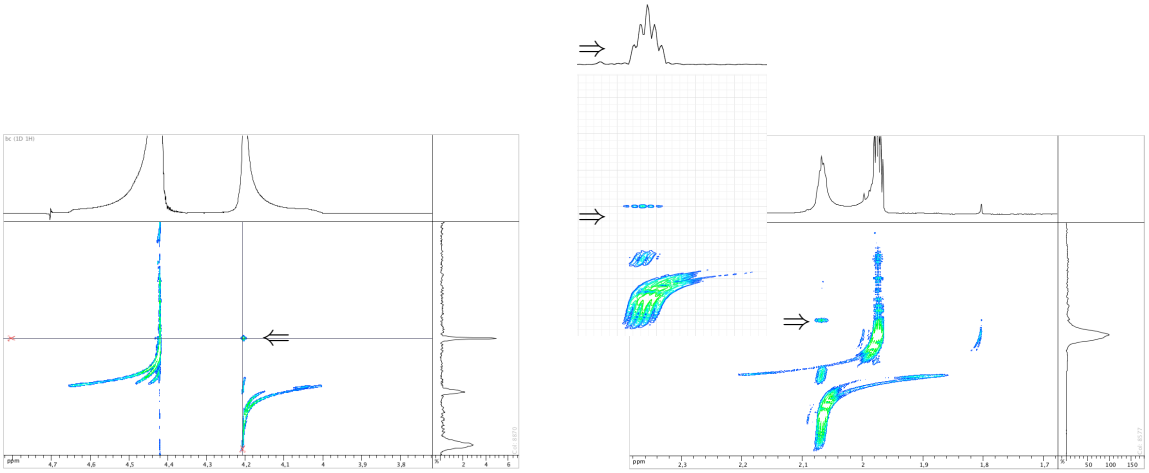



Figure 8.10: STRANGE spectrum, obtained with gradients varying linearly from 0 to $\approx 11\,G/cm$, in 250 points applied during $2 \times 1\,msec$. Inset shows a zoom on the left "horn" of the acetonitrile pattern, showing the underlying 1-2-3-2-1 pattern of the deuterated acetonitrile. Top-inset shows the 1D extract in front of the arrow. The $\Rightarrow$ and $\Leftarrow$ symbols point to the non modulated signals. Again, notice the inverted pattern between water and acetonitrile.

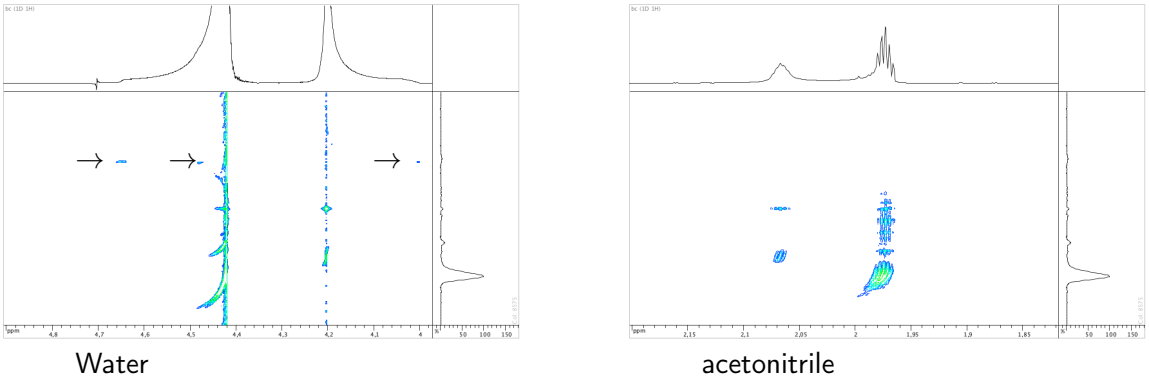



Figure 8.11: STRANGE spectrum, obtained as in 8.10 but with an inverted coding gradient pair.

### 8.2.7 Unsuccessful experiments

Several other experiments were tested, but produced no result.

#### 8.2.7.1 RDC

Suspecting a structured bi-phasic liquid mixture, we tried to investigate the presence of a partial orientation of the system. For that, we tried to measure Residual Dipolar Coupling (RDC) of progesterone molecule dissolved in the acetonitrile-water mixture. The residual dipolar coupling between two spins in a molecule occurs if the molecules in solution exhibit a partial alignment leading to an incomplete averaging of spatially anisotropic dipolar couplings. No RDC values were found, this could be due to the weakness of the coupling or because is only chemical shift anisotropy.

#### 8.2.7.2 Moment law

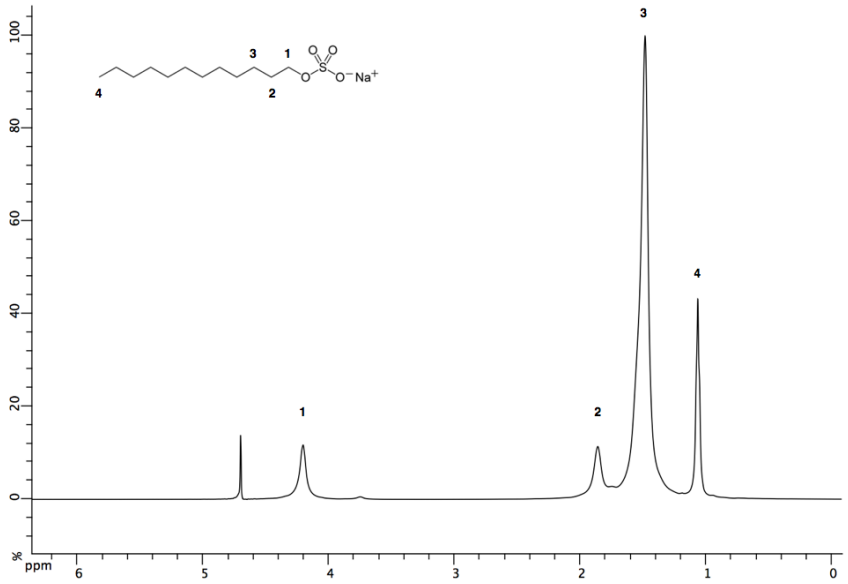
We tried to extract a more precise phase diagram by crossing the transition temperatures with the intensities of the NMR signal, applying the moment law. The precision of the NMR integrals, measured on the wide powder patterns, was not sufficient to perform this analysis in a satisfactory manner.

### 8.2.8 SDS

To test whether the STRANGE effect could be observed on another structured solution, we decided to test it on a high concentration of SDS in water. Sodium dodecyl sulfate (SDS $CH_3(CH_2)_{11}OSO_3Na$), is an anionic surfactant used in many cleaning and hygiene products. It readily adopts a micelle organization in water, above the critical micellar concentration (cmc), which is on the order of 8 mM. SDS micelles are organized with the sulfate group on the outside, and the alkyl chain packed in the interior of the micelle. This results in two independent liquid phases, the intermicellar volume, mostly constituted of water, and the intramicellar volume which can be considered as a oily liquid. SDS micelles are known to be quite monodisperse, constituted of about 50 molecules, for a large part independent on the global SDS concentration. In consequence, a high concentration of SDS in water may constitute a medium in which the effect as observed on the acetonitrile-water mixture may appear. For this reason, we are going to use a 30% solution of SDS in water as a control for the CRAZED and STRANGE experiments. Figures 8.12 and 8.13 show the results of such experiments.

The same effects can be observed on the sample, albeit weaker. In particular the solid-state pattern observed in the 1D spectrum of the acetonitrile-water mixture cannot be observed directly in the 1D spectrum. However, the STRANGE spectrum presents kind of trajectories, on the $\alpha$-methylene (noted 1) and on the $\omega$-methyl (noted 4).

a)




b)

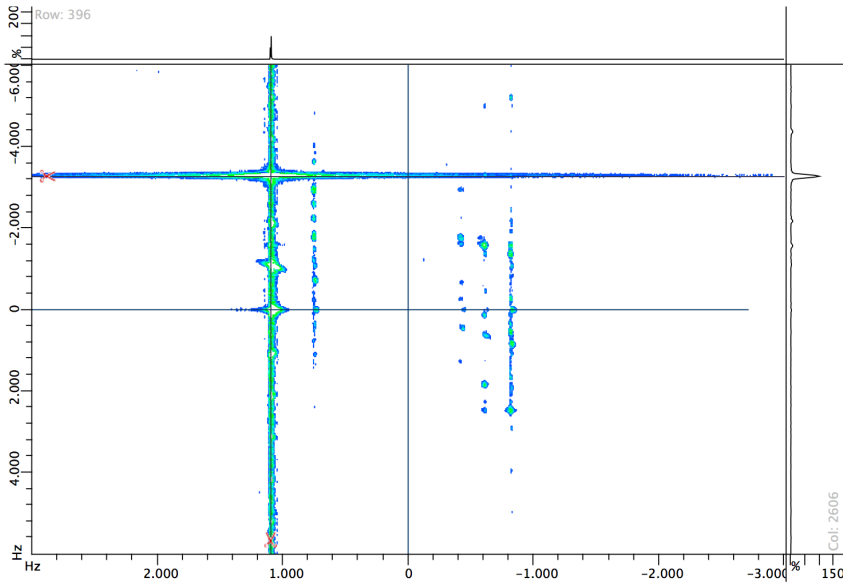



Figure 8.12: a) SDS 1D $^1$H spectrum. b) SDS CRAZED spectrum with $G_2/G_1$=-3, with position of $\omega_1 = \pm\omega_2$ (in black) $\omega_1 = 3\omega_2$ in blue.

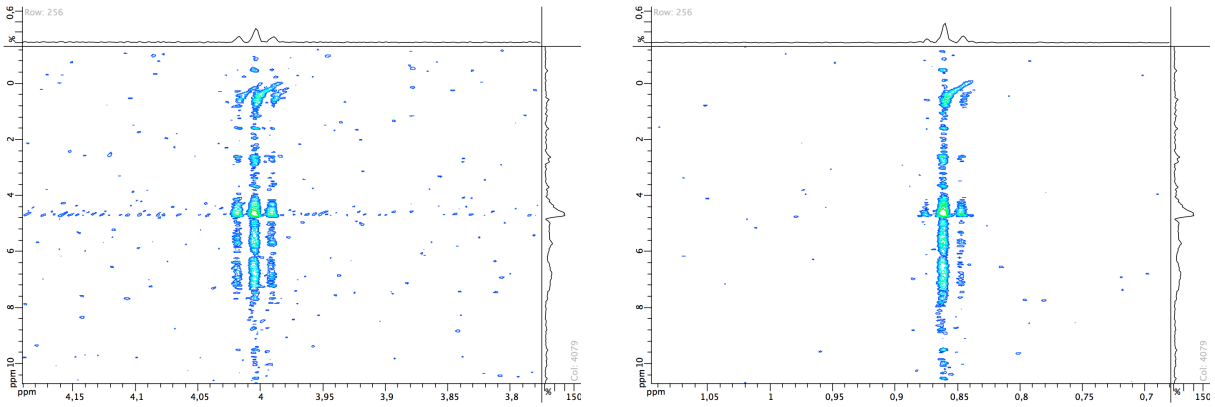


Figure 8.13: Two different regions from the SDS STRANGE spectrum that show the same pattern than the water-acetonitrile mixture.

## 8.3 Conclusion

This work presents the analysis by NMR of a peculiar behavior of a simple liquid-liquid mixture under near room–conditions. While the anomaly of the acetonitrile-water mixture at low temperature was already known and qualified, its study by the mean of the NMR spectroscopy had not been undertaken yet. The particular structure of the liquid-liquid phase separation which takes place in this system leads to large perturbations in the 1D NMR spectra, indicating strong anisotropies in the liquid-liquid interphase. The other NMR experiments performed (DOSY, EXSY) show that the structured micro-domains are certainly long-lived, with very different physical properties (strong contrast in macroscopic diffusion coefficients). The CRAZED experiment produces spectra which show that many long-range spin-order are present in the system, leading to many intermolecular multiple quantum peaks. Moreover the STRANGE experiment, first introduced in this study, present striking features, yet unexplained. When tested on the SDS-water mixture, well known to form well characterized micelles, the same features were observed, with a much smaller amplitude.

The STRANGE experiment provide beautiful patterns, which have still to be interpreted. These patterns are certainly due to couplings between the orientation of intermolecular vectors with respect to the magnetic field, and the chemical shift anisotropy tensor, the high resolution of the spectra showing that the local structure are very regular. The interpretation of this coupling might provide information on the local micro-structure present in the liquid. However the full interpretation of this effect requires the complete modeling of the quantum dynamic

of a very large spin system, a task which is beyond the scope of this work.

## 8.4 Materials & Methods

### Solvents

Acetonitrile-$d_3$ and deuterated water were purchased to Euriso-Top (France).

### NMR experiments

The samples were prepared with different volume-ratio of Acetonitrile-$d_3$ (70-65-62.5-60-55 $\%_v$) and $D_2O$ (30-35-37.5-40-45 $\%_v$), with a final volume of 180 $\mu$L. The protonated sample was prepared using 65 $\%_v$ acetonitrile-$d_3$, 25 % $D_2O$ and 10 % $H_2O$.

The experiments were carried out on a 500 MHz Bruker spectrometer equipped with a 5mm TXI probe. All $^1$H 1D experiment were performed using 16 scans, with a 90° of 9 $\mu$s at 0 dB. A period of 30 min for temperature stabilization was stablished after reaching each temperature step. In most cases, the instrument was locked on the acetonitrile deuterium signal.

Experiments were performed either using 3 mm tubes or 5 mm tubes. 3 mm tubes allowed a better temperature equilibration over the sample volume, and presented well characterized spectra and sharper transitions. 5 mm tubes lead to less spectral reproducibility, but allowed higher signal-to-noise ratio, often necessary when working on the deuterated samples, considering the width of the observed spectral patterns.

# Bibliography

[Bertie 1997] J E Bertie and Z Lan. *Liquid water-acetonitrile mixtures at 25 C: the hydrogen-bonded structure studied through infrared absolute integrated absorption intensities*. J. Phys. Chem. B, 1997. (Cited on page 160.)

[Branca 2009] Rosa T Branca, Yuming M Chen, Vladimir Mouraviev, Gigi Galiana, Elizabeth R Jenista, Challa Kumar, Carola Leuschner and Warren S Warren. *iDQC anisotropy map imaging for tumor tissue characterization in vivo.* Magn Reson Med, vol. 61, no. 4, pages 937–943, April 2009. (Cited on page 166.)

[de Sousa 2004] Paulo Loureiro de Sousa, Daniel Gounot and Daniel Grucker. *Observation of diffraction-like effects in Multiple Spin Echoe (MSE) experiments in structured samples*. Comptes Rendus Chimie, vol. 7, no. 3-4, pages 311–319, March 2004. (Cited on page 166.)

[Marcus 2013] Yizhak Marcus. *The structure of and interactions in binary acetonitrile + water mixtures*. J. Phys. Org. Chem., vol. 25, no. 12, pages 1072–1085, January 2013. (Cited on page 160.)

[McConvey 2012] Ian F McConvey, Dean Woods, Moira Lewis, Quan Gan and Paul Nancarrow. *The Importance of Acetonitrile in the Pharmaceutical Industry and Opportunities for its Recovery from Waste*. Org. Process Res. Dev., vol. 16, no. 4, pages 612–624, April 2012. (Cited on page 160.)

[Mountain 1999] Raymond D Mountain. *Molecular Dynamics Study of WaterAcetonitrile Mixtures*. J. Phys. Chem. A, vol. 103, no. 50, pages 10744–10748, December 1999. (Cited on page 160.)

[Richter 2000] W Richter, M Richter, W S Warren, H Merkle, P Andersen, G Adriany and K Ugurbil. *Functional magnetic resonance imaging with intermolecular multiple-quantum coherences.* Magn Reson Imaging, vol. 18, no. 5, pages 489–494, June 2000. (Cited on page 166.)

[Szydłowski 1999] Jerzy Szydłowski and Marek Szykuła. *Isotope effect on miscibility of acetonitrile and water*. Fluid phase equilibria, vol. 154, no. 1, pages 79–87, 1999. (Cited on pages 160 and 164.)

[Takamuku 1998] Toshiyuki Takamuku, Masaaki Tabata, Atsushi Yamaguchi, Jun Nishimoto, Midori Kumamoto, Hisanobu Wakita and Toshio Yamaguchi. *Liquid Structure of Ace-*

*tonitrileWater Mixtures by X-ray Diffraction and Infrared Spectroscopy*. J. Phys. Chem. B, vol. 102, no. 44, pages 8880–8888, October 1998. (Cited on page 160.)

[Takamuku 2001] Toshiyuki Takamuku, Atsushi Yamaguchi, Daisuke Matsuo, Masaaki Tabata, Midori Kumamoto, Jun Nishimoto, Koji Yoshida, Toshio Yamaguchi, Michihiro Nagao, Toshiya Otomo and Tomohiro Adachi. *Large-Angle X-ray Scattering and Small-Angle Neutron Scattering Study on Phase Separation of AcetonitrileWater Mixtures by Addition of NaCl*. J. Phys. Chem. B, vol. 105, no. 26, pages 6236–6245, July 2001. (Cited on page 160.)

[Takamuku 2007] Toshiyuki Takamuku, Yasukuni Noguchi, Masaru Matsugami, Hiroki Iwase, Toshiya Otomo and Michihiro Nagao. *Heterogeneity of acetonitrile–water mixtures in the temperature range 279–307 K studied by small-angle neutron scattering technique*. Journal of Molecular Liquids, vol. 136, no. 1-2, pages 147–155, November 2007. (Cited on page 160.)

Chapter 9

# General Conclusions

My thesis project have been a collaboration between the NMRTEC company and the Biomolecular NMR team at the IGBMC (Institut de génétique et de biologie moléculaire et cellulaire) under the supervision of Marc-André DELSUC. This work was supported by the NMRTEC society, conducted as part of the European project ITN-ReAd. This European project brings together ten laboratories, and two companies, on the study of complex chemical systems, such as self-replicating systems, or the development of new sensors. In this network, the role of the company is to develop new approaches for the analysis of complex systems.

In this context, my project was focused on the development of methodologies for the analysis of complex systems and the biophysical characterization of these systems. This include the study of self-assembly systems, protein-ligand interaction, disordered biological systems and solvents mixtures. A wide range of biophysical methods were used for this purpose. Mostly, Nuclear Magnetic Resonance (NMR) but also other techniques such as mass spectrometry, circular dichroism (CD), electron microscopy (EM) and small angle X-ray scattering (SAXS).

In a first part of my thesis, an out-of-equilibrium system ($\alpha$- to $\beta$-glucose anomerization) is studied using NMR. The evolution of the concentration during the DOSY acquisition let a bias in the diffusion coefficient value than can lead to misinterpretations. We show that a random permutation of the intensities of the gradients used for the DOSY experiment allows to eliminate this bias. This approach, that we named *p*–DOSY requires no changes in the NMR sequence and improve the accuracy of the measurement.

In a second part of my thesis, the N-terminal Domain of the Androgen Receptor is studied as example of a complex system. This region plays an important role in receptor activity, specially when the castration resistance prostate cancer (CRPC) is developed. And it is also described as being intrinsically disordered. In the N-terminal domain of AR (AR-NTD), there is a short primary sequence that is unique to this protein in all the human genome and which is perfectly conserved in all vertebrates. This sequence contains a cysteine residue and appears perfectly soluble and non structured in its native state. Upon the oxidation of the thiol group by DMSO, a covalent dimer is formed. This dimer spontaneously self-assembles into amyloid fibers with a diameter of 6 nm. Kinetic analysis of this phenomenon was followed by NMR

and CD, fluorescence, EM and SAXS. The fibrils have all the features of amyloid fibers, as the fluorescence response ThT, a secondary structure in $\beta$-sheet and a constant diameter. In addition, aggregation is reversible by the addition of a reducing agent. These results open up new possibilities in terms of supramolecular organization and self-assembly, and could be generalized to other systems. The originality of the results reveals an unknown aspect of the AR mechanism and allows us to explain the observation of amyloid structures in the biopsies of seminal vesicles, described in the literature [Caballero Martinez 2003][1]. The study of the functional link between oxidative stress and the activity of transcriptional regulation of the androgen receptor could open new perspectives for the treatment of prostate cancer. In addition, we carried out interactions studies within this region and the new potential drugs for prostate cancer, such as EPI-001. The results of this part of the thesis conclude that this region could be consider a new potential target site for the development of new therapeutic small molecules to target AR.

In a third part, a study of the acetonitrile-water mixture was performed. This common used mixture of solvents presents an anomaly behavior at low temperature which had not been yet explored by NMR. The particular structure of the liquid-liquid phase separation which takes place in this system leads to large perturbations in the 1D NMR spectra, indicating strong anisotropies in the liquid-liquid interphase. Other NMR experiments have been performed on the mixture, showing the presence of structured micro-domains. However the full interpretation of this effect requires the complete modeling of the quantum dynamic of a very large spin system, a task which is beyond the scope of this work.

Finally, these three different projects have in common the way of scientific and experimental approaches have been developed. *A priori* simple systems, can adopt or develop a non expected complex behavior. This situation illustrates the fact that no assumption can ever be done about the complexity of a system, increasing the difficulty of the scientific approach. From an unclear result or behavior, you build small hypothesis that become evidences and allow you step by step to draw and to understand (not always) a complete story.

The general conclusion, of the work performed during these three years, is that scientific research is amazing, surprising and not always easy but it is worth it.

[1] Caballero Martinez 2003, Arch. Esp. Urol., vol. 56, no. 4, pages 431–434, May 2003.

Julia Asencio Hernández

# Novel approaches in NMR and biophysics for the study of complex systems.


## Résumé

Cette thèse vise à réaliser une étude approfondie sur le développement de méthodologies pour l'analyse de systèmes complexes. Cela comprend l'étude des systèmes hors d'équilibre, des systèmes d'auto-assemblage, et les systèmes biologiques désordonnés. Les méthodes développées recouvrent principalement la RMN, tel que la mesure de diffusion (DOSY) mais également d'autres techniques telles que la spectrométrie de masse, le dichroïsme circulaire (CD), la microscopie électronique (EM) et diffusion des rayons X aux petits angles (SAXS). La partie N-terminale du récepteur des androgènes (AR) est utilisée comme un système complexe. D'après la littérature, il est connu que cette région joue un rôle important pour l'activité du récepteur, et elle est également décrite comme étant intrinsèquement désordonnée. Les résultats que j'ai acquis durant la thèse m'ont permis d'identifier une courte région de ce domaine, impliquée dans la formation réversible de fibres amyloïdes, par modulation des conditions d'oxydo-réduction du milieu. Les résultats révèlent un aspect inconnu du mécanisme de AR.


**Mots-clés** : systèmes complexes, AR, RMN, amyloïdes, auto-assemblage, *p*-DOSY


## Abstract

My PhD project was focused on the development of methods for the analysis of complex systems and their biophysical characterization. This includes the study of large chemical libraries, self-assembly systems, protein-ligand interaction studies and disordered biological systems. A wide range of biophysical methods were used for this purpose. Specially, Nuclear Magnetic Resonance (NMR) but also other techniques such as mass spectrometry, circular dichroism (CD), electron microscopy (EM) and small angle X-ray scattering (SAXS). The N-terminal Domain of the Androgen Receptor is studied as an example of a complex system. This region plays an important role in receptor activity, and is also described as being intrinsically disordered. The results obtained during my thesis shown a short conserved region involved in the amyloid fibers formation under oxidative conditions. These results open new possibilities to understand the mechanism of the AR activity.


**Keywords**: complex systems, AR, NMR, amyloid, self-assembly, *p*-DOSY