

***Development of fast NMR-based
investigational methods for biological systems***

A thesis submitted to attain the degree of

DOCTOR OF SCIENCE OF ETH ZURICH
(Dr. Sc. ETH Zurich)

presented by

Félix Torres

*Pharm.D. Paris 5 University
M.Eng. Chimie ParisTech
M.Sc. Paris 6 University*

born on *24.09.1991*

citizen of France

accepted on the recommendation of:

<i>Prof. Dr. Roland Riek</i>	<i>(Examiner)</i>
<i>Prof. Dr. Gunnar Jeschke</i>	<i>(Co-examiner)</i>
<i>Ass. Prof. Dr. Julien Orts</i>	<i>(Co-examiner)</i>
<i>Prof. Dr. Karl-Heinz Altmann</i>	<i>(Chairman)</i>

«...»

*Singulière fortune où le but se déplace,
Et, n'étant nulle part, peut être n'importe où !
Où l'homme, dont jamais l'espérance n'est lasse,
Pour trouver le repos court toujours comme un fou !*

...»

Charles Baudelaire, Le voyage.

Acknowledgements

I would like to thank my thesis supervisor Prof. Roland Riek, his guidance along these last years has been decisive in my capacity to pursue cross-disciplinary projects. He ensures an open minded, friendly and stimulating environment in the lab, a necessary nevertheless complex equilibrium which enables to reach scientific excellence. Moreover, his support and input in all kind of projects have been critical to my scientific fulfillment.

My gratitude goes also to Ass. Prof. Julien Orts, who co-supervised me. His kindness and expertise helped me to go through the long journey of a Ph.D. student, and learn so much about NMR-based drug discovery. His newly appointment in the University of Vienna will be, I am sure, the start for a tremendous lab, and I look forward to reading the articles which will come out of it.

I thank Prof Gunnar Jeschke for kindly accepting to be my co-examiner, and the kind discussions on radical chemistry throughout my Ph.D. project.

I thank Prof. Karl-Heinz Altmann the chairperson during my examination.

I would also like to thank Dr. Dhiman Ghosh for teaching me recombinant protein expression and purification. I really enjoyed working together, in many projects, in particular brichos.

I would like to express my gratitude to Dr. Alois Renn, for his help on the photo-CIDNP project. He taught me pretty much everything I know about lasers and optics, and I always enjoy learning from him. I hope we can go biking once together, even though I am afraid that I would slow you down.

All the Riek lab members for constant help, and the friendly and stimulating environment:

I thank Dr. Jason Greenwald for the many inputs especially in the photo-CIDNP project, and all the help in the lab. Juan and Natalia for all the conversations in Spanish, it always feels a bit like home thanks to you guys! Saroj and Radek, if I did not break the HPLC, it is in big part thanks to your help! Hari for the assistance with NMR and the friendly lunch discussions. Dima and Dean for being great office mates. Peter and Witek for the assistance with CYANA and all the other software. I also would like to thank Matthias, Yannick, Janina, Marie José, Aditya, Piotr, Lukas, and Gabriela.

I want to express my gratitude to my family, in particular my parents: Rolando Torres Martin and Sophie Hubiche; my brothers: Baltazar and Mahé; my gran-mothers: Danielle Chardon and Dolores “Lola” Martin Manzano; my gran-fathers Antonio Torres del Castillo and Gabriel du Roure. They have been a cause in each of my achievements, and their continuous love and support filled my soul leaving no place for any kind of doubt or sorrow.

Finally, I thank my beloved partner, Faiza Warsame, for completing me in every aspect of my life and taking good care of the most beautiful gift one can receive.

Abstract

Nuclear magnetic resonance (NMR) has established itself over the years as a major biophysical tool. Its inherent versatility makes it possible to study and characterize different aspects of a biological system, such as its dynamics, its structure, or the kinetics of interaction with a partner system. Because one observes the transitions of a nuclear spin between two or more energy levels, this gives NMR atomic-scale resolution, as well as excellent data quality. Many proteins refractory to crystallization have had their structure determined by NMR. Also, the first fragment-based drug design campaigns were carried out by NMR, due to its ability to study transient interactions. Despite certain advantages, the technique suffers from several problems concerning its routine application in an industrial research context. Indeed, the sensitivity of the NMR is low which results in a low signal-to-noise ratio. Therefore, the acquisition of the signal usually requires several hours or even several days for three-dimensional spectra or more and the concentrations used are most often of the order of a few hundred micromolar, well above the physiological concentrations of most biomolecules. The second major problem is the analysis of the signal, in the case of a structural calculation, it is necessary to attribute the major part of the signals, which often number in the thousands. This step is long, tedious, and most often requires the involvement of an experienced spectroscopist for a period of up to several months, where X-ray crystallography is largely automated.

A solution to the second problem was developed by Dr. Julien Orts in 2016, drawing inspiration from the principle of molecular replacement in X-ray crystallography. This solution, called NMR Molecular Replacement (NMR^2) is free from the step of attribution of the signal necessary for the calculation of a protein-ligand complex using a “simulated annealing” type algorithm. This thesis presents the work after this invention, in particular the application to new targets as well as the introduction of additional restraints facilitating and accelerating the calculation of structures of protein-ligand complexes (Chapter 2). In addition, the use of NMR^2 for the calculation of protein-fragment structures is demonstrated, and general guidelines to ensure the success of such a project have been established (Chapter 3). A more classic use of NMR in a drug discovery context is presented in Chapter 4, where the effort made in the fight against the Covid-19 pandemic is presented, emphasizing the importance of the screening process upstream of the use of NMR in structural biology. While NMR^2 is a powerful asset for medicinal chemistry, the main flaw in NMR, namely sensitivity, has been a clear drag in our research effort against SARS-CoV-2.

To solve the problem of sensitivity, several methods of hyperpolarization are available. However, we considered more appropriate the use of photo-CIDNP, which is the only technique that can be carried out in solution, at room temperature, and at a lower cost than for dissolution DNP, for example. However, although the theoretical aspect of the phenomenon is well documented, the method is limited to the analysis of a small number of molecules, including tryptophan and tyrosine. This thesis reports the polarization of new molecules resulting from tryptophan degradation reactions as well as their interdependence with the photosensitizer resulting in divergent polarization intensities (Chapter 5). Finally, an extension of the chemical diversity of polarizable molecules is proposed in Chapter 6 as well as an adaptation of the parameters to be taken into account for the optimization of polarization in the context of low concentrations.

Résumé

La résonance magnétique nucléaire (RMN) s'est imposée au fil des années comme un outil biophysique majeur. La versatilité qui lui est propre, permet d'étudier et de caractériser différents aspects d'un système biologique, tels que sa dynamique, sa structure, ou encore la cinétique d'interaction avec un système partenaire. Du fait que l'on observe les transitions d'un spin entre deux ou plus niveaux d'énergie, cela vaut à la RMN une résolution à l'échelle atomique, ainsi qu'une excellente qualité de données. De nombreuses protéines réfractaires à la cristallisation ont vu leur structure déterminée par RMN. Aussi, les premières campagnes de design de médicament basées sur le screening de fragments ont été réalisées par RMN, du fait de son incroyable capacité à étudier les interactions transitoires. Malgré des avantages certains, la technique souffre de maux rédhibitoires quant à son application en routine dans un contexte de recherche industrielle. En effet, la sensibilité de la RMN est faible ce qui résulte en un faible ratio signal sur bruit. De ce fait, l'acquisition du signal requiert le plus souvent plusieurs heures voire plusieurs jours pour les spectres à trois dimensions ou plus et les concentrations usitées sont le plus souvent de l'ordre de quelques centaines de micromolaires, bien au-dessus des concentrations physiologiques de la plupart des biomolécules. Le second problème majeur est l'analyse du signal, dans le cas d'un calcul de structure, il faut attribuer la majeure partie des signaux, qui sont bien souvent au nombre de plusieurs milliers. Cette étape, est longue, fastidieuse et requiert le plus souvent l'implication d'un spectroscopiste expérimenté pour une durée allant jusqu'à plusieurs mois, là où la cristallographie par rayons X est en immense partie automatisée.

Une solution au second problème a été développée par le Dr. Julien Orts en 2016, s'inspirant du principe de remplacement moléculaire en cristallographie par rayons X. Cette solution, nommée NMR Molecular Replacement (*NMR*²) s'affranchi de l'étape d'attribution du signal nécessaire au calcul d'un complexe protéine-ligand en utilisant un algorithme de type « recuit simulé ». Cette thèse présente les travaux subséquents à cette invention, avec notamment l'application à de nouvelles cibles ainsi que l'introduction de paramètres de restrictions supplémentaires facilitant et accélérant le calcul de structures de complexes protéine-ligand (Chapitre 2). De plus, l'utilisation de *NMR*² pour le calcul de structures protéine-fragment, est démontré et des directives générales pour garantir le succès d'un tel projet ont pu être établies (Chapitre 3). Une utilisation plus classique de la RMN dans un contexte de découverte du médicament est présentée dans le chapitre 4, ou est présenté l'effort fait dans la lutte contre la pandémie du Covid-19, soulignant l'importance du processus de screening en amont de l'utilisation de la RMN en biologie structural. Si *NMR*² est un atout de poids pour la chimie médicinale, le défaut principal de la RMN, à savoir la sensibilité a été un frein clair dans notre effort de recherche contre le SARS-CoV-2.

Pour solutionner le problème de la sensibilité, plusieurs méthodes d'hyperpolarisation sont disponibles. Cependant, nous avons jugé plus adéquate l'utilisation de photo-CIDNP, qui est la seule technique pouvant être conduite en solution, à température ambiante et à moindre cout. Cependant, bien que l'aspect théorique du phénomène soit bien documenté, la méthode se cantonne à l'analyse d'un nombre restreint de molécules, y-compris le tryptophane et la tyrosine. Cette thèse, rapporte la polarisation de nouvelles molécules issues de réactions de dégradation du tryptophane ainsi que leur interdépendance avec le photosensibilisateur aboutissant à des intensités de polarisation divergentes (Chapitre 5). Finalement, une extension de la diversité chimique des molécules polarisables est proposée dans le chapitre 6 ainsi qu'une adaptation des paramètres à prendre en compte pour l'optimisation de la polarisation dans le contexte de basses concentrations.

Table of contents

Acknowledgements.....	5
Abstract.....	6
Résumé.....	7
Table of contents.....	9
Table of abbreviations.....	13
Chapter 1: Introduction.....	17
The paradigm evolution towards structure-based drug design.....	18
State-of-the-art in structure-based drug discovery.....	20
Nuclear magnetic resonance in structure-based drug discovery.....	21
NMR molecular replacement to reduce time-to-structure.....	23
Polarization and photo-CIDNP.....	25
Overview of the state-of-the-art NMR used in the present work.....	27
Chapter 2: <i>NMR</i> ² method development and application to the bromodomain BRD4-BD2.....	31
Introduction.....	32
NMR molecular replacement.....	32
Introduction of new features into <i>NMR</i> ² structure calculation.....	36
Application to bromodomains.....	36
<i>NMR</i> ² method development.....	37
Interleaved filtered-NOESY.....	37
Specific methyl labeling.....	38
Anti-NOE.....	38
Application to the <i>NMR</i> ² of BRD4-BD2-iBET762 complex structure.....	40
Conclusion.....	42
Chapter 3: <i>NMR</i> ² for fragment-based drug design.....	47
Introduction.....	48
State-of-the-art methods for fragment-based drug design.....	49
Benchmarking <i>NMR</i> ² for protein-fragment complex structure calculations.....	51
Application to PIN1-fragment complex structures calculation.....	51
<i>NMR</i> ² benchmark for protein-fragment complexes.....	52

SERAPHic dataset	52
Effect of the distance restraints number on NMR ² protein-fragment complex structure calculations.....	54
Effect of the distance restraints error on NMR ² protein-fragment complex structure calculations.....	55
Partial assignment	59
Effect of the ligand topology on NMR ² protein-fragment complex structure calculations	61
Calculation time	64
Application to PIN1-fragment complex structures calculation	65
Fragments' overview.....	65
Calculation of the effective correlation time.....	66
Structure calculation	69
Time-to-structure and outlook	73
Conclusion	74
Chapter 4: NMR-based drug discovery for SARS-CoV-2	77
Introduction	78
The severe acute respiratory syndrome coronavirus-2 and relevant targets for drug discovery	78
On the NMR study of the M ^{pro} protein.....	80
Screening	82
Literature cross-checking	82
MiniFrag library screening.....	84
FDA library screening.....	85
Outlook on the M ^{pro} -ligand complex structure resolution by NMR.....	93
Conclusion.....	94
Chapter 5: Exploration of the photo-CIDNP performances investigating new photosensitizers and chemical modifications of tryptophan	95
Introduction	96
Photo-CIDNP polarization theory.....	97
Dye investigation	100
Chemical tuning of photo-CIDNP performances.....	101
Dye investigation.....	102

Chemical tuning of photo-CIDNP performances	105
Characterization of HOPI and synthesis	106
Photo-CIDNP performances in the light of magnetic parameters	110
Photo-CIDNP of ¹³ C resonances of HOPI	112
g-factors differences (Δg).....	112
Hyperfine couplings.....	115
Diastereomeric-dependent reaction kinetics	118
Bleaching of the CIDNP effect with cis/trans-HOPI compounds.....	122
On the AT12-HOPI pair – an Outlook.....	123
Conclusion.....	124
Chapter 6: Basis for the empirical exploration of the photo-CIDNP performances.....	125
Introduction	126
Alternative to the calculation of magnetic parameters.....	126
Chemical tuning of g-factor.....	129
1.1 On the Kaptein rule of the oxidocyclization product of tryptophan α -hydroxypyroloindole.....	129
Screening of the tryptophan and tyrosine chemical space.....	130
Conclusion.....	137
Conclusion and outlook	139
Material and methods.....	143
Chapter 2: NMR ² method development and application to the bromodomain BRD4-BD2	143
Chapter 3: NMR ² for fragment-based drug design.....	144
Chapter 4: NMR-based drug discovery for SARS-CoV-2.....	145
Chapter 5: Exploration of the photo-CIDNP performances investigating new photosensitizers and chemical modifications of tryptophan	146
Chapter 6: Basis for the empirical exploration of the photo-CIDNP performances	147
Appendix 1: Structure calculation of the proSpc-Brichos domain	148
Introduction	148
Results and discussion.....	148
Structure calculation of the wt-brichos trimer	148
Design and selection of the brichos mutant	152
Conclusion and outlook.....	156

Appendix 2: <i>NMR</i> ² structure deposition onto the PDB	158
PDB submission and CNS water refinement.....	158
Water refinement in CNS	158
Ligand parametrization	158
CNS compatible complex structure preparation	159
Preparation of the water refinement.....	159
Running the water refinement scripts	159
PDB submission	160
Preparation of the ligand	160
Getting prepared for the deposition	162
Deposition	162
Literature.....	163
Curriculum vitae	Error! Bookmark not defined.

Table of abbreviations

AT12	Atto thio 12
CIDNP	chemically induced dynamic nuclear polarization
CSP	chemical shift perturbation
CW	continuous wave
diH-TRP	dihydro-tryptophan
EPR	electron paramagnetic resonance
ET	electron transfer
FBDD	fragment-based drug design
FDA	food and drug administration (USA)
FID	free induction decay
FMN	flavin mononucleotide
FRET	fluorescence resonance energy transfer
GST	glutathione S-transferase
HFCC	hyperfine coupling constants
HIV	human immunodeficiency virus
HOMO	highest occupied molecular orbital
HOPI	3 α -hydroxypyrrroloindole
HSQC	heteronuclear single quantum coherence
ctHSQC	constant-time HSQC
IAA	indole acetic acid
IC ₅₀	half maximal inhibitory concentration
IPA	indole propionic acid
ITC	isothermal calorimetry
INEPT	insensitive nuclei enhanced by polarization transfer
K _d	dissociation constant
M ^{pro}	main protease of SARS-CoV-2
MW	molecular weight
NMR	nuclear magnetic resonance
NMR ²	NMR molecular replacement

NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
Nsp	non-structural protein
PA	partial assignment
PCET	proton coupled electron transfer
PDB	protein data bank
PIN1	peptidylprolyl <i>cis/trans</i> isomerase, NIMA-interacting 1)
Pp1a	polypeptide 1a
Pp1b	polypeptide 1b
RB	rose bengal
RBD	receptor binding domain
RF	radio frequency
RMSD	root mean square deviation
SAR	structure activity relationship
SBDD	structure-based drug design
SN	signal-to-noise
SNE	SN enhancement
SOMO	singly occupied molecular orbital
SPR	surface plasmon resonance
STD	saturation transfer diffusion
TCBP	3,3',4,4'-tetracarboxybiphenol
TEV	tobacco etch virus
TF	target function
TFC	TF contrast
TOCSY	Total correlation spectroscopy (HCCH is the magnetization pathway)
TR	time resolved
TROSY	transverse relaxation optimized spectroscopy
TRP	tryptophan
TRPA	tryptamine
TSA	thermal shift anisotropy
TYR	tyrosine

TYRA

tyramine

Chapter 1: Introduction

The paradigm evolution towards structure-based drug design

Medicine can be traced back as long as it is possible to trace human civilizations. The consumption of plants, and the practice of surgical acts such as trepanation has been observed from different archeological studies. The combination of trial and failure, serendipity and mysticism, provided over the first 100000 years of human history many drugs which later constituted the traditional medicine, and can be found in the important pharmacopeia of different civilizations (Ayurveda, Arabic medicine, Chinese medicine, shamans), witnessing from the effort of the first men to cure each-others. In these different pharmacopeias, the concept that the raw material can be “cooked” to improve their efficiency is found. The cooking recipes can be decoction, infusion, alcoholic extraction, distillation, and many more... The contemporary paradigm shifted towards the rationalization down to the atomistic level the mechanism of action of a drug on its target. This radical change has been enabled by the tremendous technological improvements made during the last century; the consequence is that modern drugs are characterized molecular entities and not plants or extracts as before. Thirty years ago, the advent of structural biology and its application to drug discovery has been at the origin of the success story of anti-HIV drugs, which have been rationally designed based on structural data.¹ What follows describes with more details the drug discovery paradigm evolution along human history, through the example of aspirin.

The use of white willow (*Salix alba*) to alleviate pain has been evidenced for civilizations as early as Sumer. The Greek physician Hippocrates (5 B.C.) described as well the use of the bark of white willow in infusion. It is only in 1825, when salicylin was isolated from *Salix alba* for the first time.² This corresponds also at the moment, in the early 19th century, when the atom theory started to be well established in the scientific community and the idea that substances are composed by molecules which are the essence of the medicinal activity. The 19th century marks also the mastering at a more global scale of chemical reactions, and the idea that the molecular structure can be modified under certain conditions to serve a purpose was adopted. The human empowerment to modify the matter yielded to the synthesis of salicylic acid from the oxidation of salicylin. Salicylic acid was identified to be more active against pain than salicylin, and it could conveniently bio-sourced from the meadowsweet (*Filipendula ulmaria*) where it is present in important proportions. While the exact date of the first acetylation of salicylic acid into acetylsalicylic acid (aspirin) is subject to controversy, Bayer[®] introduced aspirin to the market, in 1898, after demonstrating the reduction of intestinal bleeding, the main adverse effect from salicylic acid. This history witnesses from the early days of medicinal chemistry and structure activity relationship (SAR): chemical structure of the drug defines its activity, and chemical modifications can be used to fine tune the pharmacodynamics. Nonetheless, it must be noted that the target was not considered yet at this stage in a molecular perspective. Then on, medicine have been approached with the eyes of the biologist, whom guided by the cultural heritage of traditional medicine, experimentation and a bit of serendipity, paved the way for modern drug discovery. In 1971, the improvement in the field of biochemistry and molecular biology enabled the identification of prostaglandin synthase I and II (COX-1 and 2) to be inhibited by aspirin, enlightening the target at a molecular level.³⁻⁵ This discovery awarded Vane the 1982 Nobel prize in conjunction with Bergstrom and Samuelson for their “discoveries on prostaglandins and related biologically active substances”. The mastering of radioactivity was determinant as well in understanding the mechanism of action of the aspirin: In 1980, the acetylation of a serine in the COX was evidenced as the mode of action of aspirin rather than enzymatic inhibition.⁶

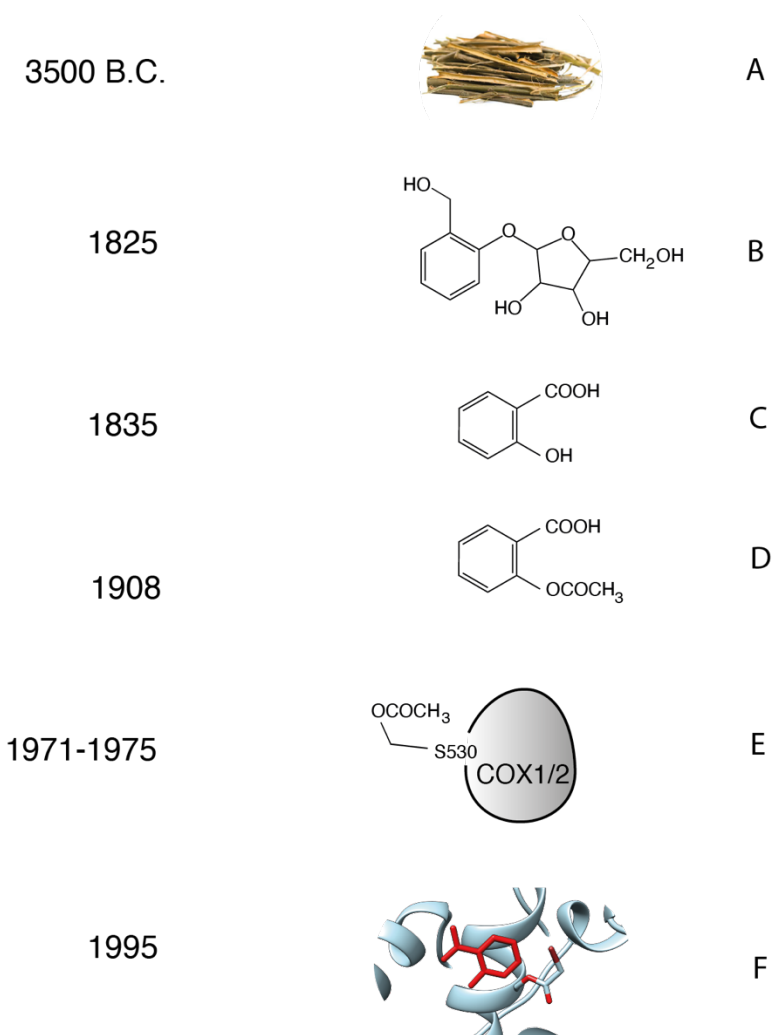


Figure 1.1: chronology of the discovery of salicylin, salicylic acid, and acetyl salicylic acid. A) *Salix alba* bark. B) Salicylin. C) Salicylic acid. D) Aspirin. E) COX-1/2 serine acetylated. F) X-ray structure of COX-1 aspirin complex (pdb 3n8y).

The resolution of the DNA structure, its sequencing and soon after the production and structure determination of recombinant proteins with X-ray crystallography was the starting point for the field known today as structure-based drug design (SBDD). Thus, in 1995, almost hundred years after its commercialization, the first structure of COX-1 in complex with aspirin was solved by X-ray crystallography showing the acetylation of the serine 503.⁷ The history of aspirin illustrates well the evolution of the drug discovery over the ages. It is important to realize that most of the

evolution in the approach, from a plant to a characterized molecule, and from a biological system to a specific target defined at atomistic resolution, are strongly conditioned by the state-of-the-art technologies available. This story emphasizes the need of parallel effort, as much in the pharmaceutical and medical knowledge themselves, as in the technological and methodological development necessary for the further development of SBDD.

State-of-the-art in structure-based drug discovery

SBDD is a cornerstone of rational drug design. It provides information at an atomistic scale about the interaction between a ligand and its target, resulting in the richest SAR possible. From these data, the molecular structure of the ligand can be modified to improve its affinity and specificity to the target. SBDD is punctuated by different steps to identify the molecule binding to the target (Screening), characterize the binding kinetics of such a molecule (Affinity), and determine the structure of the target-ligand complex. In the context of this thesis the only targets considered are proteins. However, the techniques described apply as well in the case of RNA or DNA targets.^{1, 8,9}

The screening of compound libraries is the starting step, identifying the molecules binding to the target of interest, called hits. Among the plethora of techniques developed we mention a few such as fluorescent reporter systems, particularly attractive for the screening of gigantic libraries of millions of molecules. Nonetheless, fluorescence assay often relates on the enzymatic activity of the target and for each target a new assay needs to be adjusted.¹⁰⁻¹¹ Such an assay is also an indirect measure and admits high false positive rates. Surface plasmon resonance (SPR), another screening gold standard assumes a fairly high throughput compatible with fragment libraries, and the affinity of the ligand can be also derived from the SPR spectra.¹² Nevertheless, SPR requires the immobilization of the target onto a surface, a tedious process, and operating the machine needs a sufficient level of expertise. X-ray crystallography assumes crystals diffraction at high throughput enabling the screening of fragment libraries soaked onto apo-crystals.¹³ However, crystal soaking requires high fragment concentrations (> mM) in aqueous buffer, some crystallized binding sites are impervious to soaking and the soaked crystal structure provide no information about the affinity of the fragments. Nuclear magnetic resonance (NMR) is also used to perform small molecules screening with saturation transfer difference (STD) experiments.¹⁴⁻¹⁵ Yet the poor sensitivity of NMR does not allow to measure at high throughput, rather a couple of dozens of molecules can be screened every day.

Regarding the hit validation, the combination of X-ray soaking with SPR is often complemented by NMR, which is the gold standard technique for hit validation, and is mostly performed by ¹⁵N-Heteronuclear Single Quantum Coherence spectroscopy (HSQC) where the chemical shift perturbations (CSP) sign unambiguously for a ligand binding event.¹⁶⁻¹⁷ Furthermore, the dynamics of the evolution of the CSP upon increasing ligand concentration provides the dissociation constant K_d .¹⁷⁻¹⁸

The binding kinetics are essential to characterize the preferential interaction of a ligand with its target, it is also used as a resource for SAR by catalogue. SPR provides the on/off-rate from which is derived the K_d for protein-ligand complexes. The K_d can also be derived from isothermal calorimetry (ITC),¹⁹ fluorescence anisotropy²⁰ or NMR experiments (HSQC, or STD). Interestingly, ITC can also measure thermodynamic parameters such as enthalpy, free enthalpy and indirectly, the entropy of binding.¹⁹

The current gold standard for protein-ligand complex structure determination is X-ray crystallography. The impressive progress in the field towards the full automation of the process and the fast crystal diffraction render X-ray crystallography particularly attractive.^{13,21} However, despite the improvement of crystallization methodology some proteins remain impossible to crystallize, and weak affinity binders such as fragments are often resistant to co-crystallization.²²⁻²³ This latter barrier was partly overcome with crystal soaking, but many issues are still to be addressed due to the experimental conditions.²³ Furthermore, lead molecules or molecules in general showing more molecular complexity than fragments are often not possible to soak. The main alternative to X-ray crystallography is NMR, which can be performed in solution overcoming the problems related to crystallization. However, as it is discussed below NMR faces speed issues rendering its application in routine medicinal chemistry unsuitable.

Nuclear magnetic resonance in structure-based drug discovery

The protons and neutrons constituting the atom nucleus possess an intrinsic angular momentum called spin, which value is $\pm 1/2$. The value of these spins combines to form the nuclear spin, I . As of Pauli principle, when a nucleus bears an even number of both protons and neutrons, $I = 0$; nuclei with an even mass number and both protons and neutrons are odd, have an integer spin; nuclei with an odd mass number have half integer spins, $I = (1 + 2S)/2$, with S an integer. Nuclei with a non-zero spin also have a nuclear magnetic moment, μ , colinear with the spin angular momentum. Such a nucleus when it is placed in a static magnetic field, B_0 , will have its spin precessing and orienting with B_0 . The spin angular momentum is quantized such as if considering its cartesian z-component (along B_0) $I_z = \hbar m$, and $m = (-I, -I+1, \dots, I-1, I)$.²⁴ The presence of the static magnetic field removes the degeneracy of the energy levels which are thus occupied according to the Boltzmann distribution. In the context of biological NMR (bioNMR), most of nuclei considered have a $I = 1/2$ such as ^1H , ^{13}C and ^{15}N . Therefore, two spin states are to be considered $\alpha = +1/2$ and $\beta = 1/2$, and the energy gap between them corresponds to their precession frequency around the z-axis, defined by the B_0 axis. The application of a radiofrequency (rf) pulse at the frequency matching $E_\beta - E_\alpha$ induces coherent spin transitions. Therefore, the magnetic moment resulting from the bulk and oriented initially along z can be rotated according to the right-hand rule with regard to the axis from where is applied the rf-pulse. NMR records the free induction decay (FID) caused by the bulk spin's magnetic component in the xy-plane. Each spin experiencing a different magnetic environment, because of the small magnetic moment from other atoms surrounding it, acting as a Faraday cage, will have a different precession frequency in the xy-plane. After Fourier transformation of the FID, the different frequencies are the signature of the different spins, and the integral of each frequency line is proportional to the amount of these spins. Further description of NMR theory and related experiments is available in Cavanagh et al. *Protein NMR spectroscopy*,²⁵ or Keeler *Understanding NMR spectroscopy*.²⁶

The development of multidimensional NMR paved the way for signal assignment and structure determination of small organic molecules and biomolecules. Moreover, the advent of isotopically labeled samples,²⁷⁻²⁸ heteronuclear NMR experiments,²⁹⁻³¹ transverse relaxation-optimized spectroscopy³²⁻³³ (TROSY), and partial or perdeuterated sample enabled to assign the signals and solve the structures of complex systems, reach molecular weights (MW) up to the 100 kDa range.³⁴⁻³⁵ After 50 years of development, biological NMR is a versatile technique offering of broad range of methods that can be used to detect the interaction of a ligand with a target, to measure the affinity of the ligand with its target or to solve the structure of the target in complex with a ligand. The principal ones are briefly described below.

The ligand screening is classified into two types of techniques, ligand observed and protein observed experiments. Ligand observed experiments have the advantage that they do not require labeling and they relate on 1D proton NMR which is relatively fast to measure. The most common experiments of this type are STD which detects the saturation transfer from a target to a ligand upon interaction;³⁶ WATER-LOGSY selectively saturates bulk water resonance to polarize protein and ligand through NOE and chemical exchange and takes advantage of the sign difference between protein-ligand NOE (negative, strong), and ligand-water NOE (positive, weak) to identify protein-ligand interaction;³⁷⁻³⁸ relaxation-based methods observe the reduction of the signal relaxation time (T_1 , T_2 or $T_{1\rho}$) upon binding to a target of higher molecular.¹⁵ Nevertheless, they mostly detect weak binders. The second problem of these experiments is that they are prone to false positives and they need to be duplicated to obtain a control for each ligand. The protein observed experiments, typically HSQC,³⁹ which observe either the ^{15}N - ^1H vectors (^{15}N -HSQC) or the ^{13}C - ^1H vectors (^{13}C -HSQC) chemical shifts. In routine, the ^{15}N -HSQC is preferred as ^{15}N -labeling is cheaper than ^{13}C . Chemical shifts perturbations (CSPs) that originate from changes in the surrounding magnetic shielding of a spin, upon ligand binding, is an unambiguous evidence of protein-ligand interaction,¹⁷ and if the assignments are available the binding site can be identified. Furthermore, only one control experiment consisting of the apo-protein spectrum is needed. However, HSQC is a bi-dimensional NMR experiment and thus requires relatively long recording time. Moreover, as it is a protein observed experiment, it requires fairly high concentration of protein (> 0.1 mM). These limitations are critical in the context of a small molecules screening campaign, therefore the HSQC are often realized only to confirm a hit or identify the binding site.

Once the hits from the screening are validated, the affinity of the ligand for its target assessed through the calculation of the dissociation constant, K_d , is often needed to rank the ligands and establish the first structure-activity relationship features. The affinity of the ligand can be obtained measuring the CSP from HSQC experiments.¹⁷ The affinity can be also estimated from STD experiments,⁴⁰ the ^{15}N -HSQC lineshapes.⁴¹ The calculation of K_d from ^{13}C -HSQC CSPs is described in Chapter 3.

The classic NMR protein-ligand structure calculation methodology takes advantage of the nuclear overhauser effect (NOE) cross-relaxation rates (σ) which are proportional to r^{-6} , where r is the distance between two nuclei. Hence, NOE is rather suitable for relatively short distance restraints (1 to 5–6 Å) and thus perfectly adapted for ligand-binding pocket restraints.^{18, 25, 42-44} Nevertheless, the structure calculation requires non-ambiguous distance restraints implying the assignment of most of the protein resonances, and of the ligand resonances. The protein resonances represent several hundreds to a couple of thousands of assignments while the ligand resonances are in the order of a dozen of assignments. The protein signal assignment involves the measurement of a more or less large set of tridimensional spectra, e.g.: HNCa, HN(CO)Ca, HNCO, HN(Ca)CO, HNCaCb, HCCH-TOCSY.^{27-28, 45} The measurement and assignment take at least 1-2 months for an experienced spectroscopist, and the assignment of the NOE cross-peaks of the NOESY spectra containing the distance information usually adds a few more weeks before the obtention of the first acceptable structure. This long, and tedious process is discouraging for many of us, but more importantly it is absolutely not aligned with the medicinal chemistry timeline. An illustration of the classical structure calculation procedure is detailed in the Appendix 1 which describes the investigation of the structure activity relationship a chaperone domain, BRICHOS.

Other methodologies based on different restraints have been developed with the ultimate goal to shorten the time-to-structure. Such methods rely on the ranking of docking results using either

ambiguous or unambiguous restraints. Ambiguity is an important concept for structural restraints. An NMR restraint between a pair of defined atoms is said to be 'unambiguous' when each atom belongs to a group assigned to a single resonance. Inversely, 'ambiguous' restraints are defined between two atoms from not assigned groups, thus allowing several combinations.⁴⁶ The CSP-based structure calculations use ambiguous restraints. If this ambiguity seems to be acceptable for protein-protein interactions, for protein-ligand complexes it is difficult to obtain a unique orientation due to the low resolution of the CSP restraints. NOE restraints are considered as not ambiguous as both of the atom signals can be assigned. Several techniques have been developed with the ultimate goal to reduce the time-to-structure. Some of those relate on ambiguous distance restraints originating from different phenomena such as ring current induced CSP (j-surface)⁴⁷⁻⁴⁸, or saturation transfer (INPHARMA⁴⁹, SOS-NMR⁵⁰). The latter originates from the NOE but the protein resonances do not need to be assigned as only the ligand signal is observed. Unambiguous distance restraints can be NOE but also pseudocontact shifts (PCS) which are not ambiguous since they originate from an identified nucleus, Ln^{3+} , and provide accurate restraints similar to the NOE.⁵¹ It is worth noticing that methods relying on CSP such as j-surface or PCS need resonance assignment. Automated resonance assignment procedures might speed up this step, but still require the measurement of a set of 3D NMR experiments. All the techniques discussed previously refine the complex structure by docking and scoring approaches. Even if some docking software enable to run data driven docking such as HADDOCK, a force-field may bias the generated poses.⁵² While it was shown that protein-protein interactions are well defined by restraint docking even with ambiguous restraints thanks to their large interaction surface, for small molecule the energy penalty due to a wrong orientation is often not large enough to exclude these structures since they do not originate from high resolution restraints. In summary, NMR benefits from a wide toolbox for SBDD. Nonetheless, these tools are in large part not routinely used. The derivation of complex structures should be done as simply as possible and quantitative restraints such as NOE or PCS should be prioritized over semi-quantitative ones like CSP. Moreover, since ligand is most likely small molecules, low resolution restraints should be avoided as they will not provide sufficient structural information and will be dampened by the modeling. Ranking of docking poses is the most common way to derive protein-ligand structures but docking tools still suffer from force field inaccuracies or incorrect scoring functions. Therefore, docking-scoring methods should be avoided as much as possible. We think that effort should be put into developing structure calculation methods integrating current available structural data from the Protein Data Bank with quantitative NMR restraints such as PCS and semi-ambiguous NOE. NMR is a versatile technique that when synergistically combined with other approaches provides new structures of very high quality.⁴⁶

NMR molecular replacement to reduce time-to-structure

The problem of signal assignment has been addressed so far using a combination of docking and ambiguous NMR restraints. Two recent methods, NMR molecular replacement⁴³ (NMR^2) and NOE-matching,⁵³ derive the complex structure with NOE-based distance restraints between an assigned ligand and a non-assigned-protein, thus skipping the long and tedious protein signal assignment step. The difference between the two algorithms is that NMR^2 calculates each time the complex structure, while NOE-matching is a docking-scoring method. NMR^2 uses the experimental NOE cross-peaks intensities to generate semi-ambiguous intermolecular distance restraints for the calculation of protein-ligand complex structures. The term semi-ambiguous refers to the fact that only the ligand resonances are assigned, and not the protein resonance.

Therefore, the protein-ligand distance restraints are ambiguous on the protein side, and unambiguous on the ligand side. While unambiguous restraints require the time-consuming assignment step, semi-ambiguous restraints seem to be a good compromise as they conveniently do not require the receptor assignment but retain discrimination capabilities. In this case, NOE restraints are between an assigned ligand resonance and an unassigned, but unique, protein resonance. This protocol is used in NMR^2 , thus avoiding signal assignment for protein and limiting the level of ambiguity. The ligand assignments are readily obtained, and often available from the medicinal chemistry quality control. Furthermore, quantitative analysis of NOE restraints is well established for distance calculations. NMR^2 success has been demonstrated for different systems including strong/weak binders, drug-like/fragment molecules/peptides. While the method requires that the structure of the protein is known, it was successfully performing with structures such as apo-protein, complexes with other ligands, conformers, and even homologous proteins.^{43-44, 54} This versatility makes NMR^2 a serious candidate to complement X-ray crystallography in the structural biology toolbox at the disposition of the medicinal chemist. As described previously a protein-ligand structure calculation based on NOE distance restraints requires the acquisition of several tridimensional spectra such as HNCa, HN(CO)Ca, HNCaCb, HCCH-TOCSY, and a series of 3D-NOESY spectra.⁵⁵⁻⁵⁶ This set of experiments requires at least three weeks of measurement on a high field spectrometer and depending on the complexity of the system, more experiments may be necessary. NMR^2 only needs the acquisition of a series of filtered 2D-NOESY which can be recorded within a few days. Hence, NMR^2 considers the intermolecular distance restraints between the ligand assigned signal and the non-assigned methyl signals of the protein. Protein methyls are borne by alanine, leucine, isoleucine, valine, methionine or threonine, they are present in a very specific chemical shift range (1.5 to -2 ppm) and they benefit from good signal to noise (SN) due to the fact they carry three protons and they relax slower than other protein's protons. Furthermore, the methyls containing amino-acids are well represented in the binding-site of the proteins which often is based on a hydrophobic pocket. Therefore, the ligand to protein's methyl cross-peaks are the only retained for NMR^2 analysis. The semi-ambiguous distance restraints network can be approached in the same fashion as a global positioning system (GPS), as the positions of the ligand's protons are triangulated as the methyls are identified. This is done with a simulated annealing algorithm, generating methyl assignments and calculating the corresponding structure. The calculation of different methyl assignments structures results in the disambiguation of the assignments. Indeed, the produced structures violate more or less to the anonymous distance restraints networks. The importance of the clashes is evaluated thanks to the target function (TF), and the structures with the least amount of clash, the lowest TF, is accepted as the actual structure. The NMR^2 methodology, which overall workflow is resumed in Figure 1.2, is able to obtain a protein-ligand complex structure within a week instead of the usual two to three months. A more detailed description of the NMR^2 workflow is provided in the corresponding Chapter 2: *NMR^2 method development and application to the bromodomain BRD4-BD2*. In this chapter, also the new features to the NMR^2 method are introduced. Furthermore, the structure of BRD4-BD2, a bromodomain infamous for being impervious to co-crystallization, in complex with its inhibitor, iBET762, is described.⁵⁷⁻⁵⁹ This achievement illustrates nicely the potential of NMR^2 for being an alternative to X-ray and keeping alignment with a medicinal chemistry agenda. Chapter 3: *NMR^2 for fragment-based drug design*, reports the result of a NMR^2 benchmark for the application of the method to protein-fragment complex structures. Since NOE-based distance restraints deal with an error of +/- 20%, NMR^2 performs poorly when it comes to structure calculation of very small molecules such as fragment. Clear guidelines, including the use of partial assignment (PA), enable the successful calculation of most of the complexes in the benchmark. Finally, the teachings from the benchmark

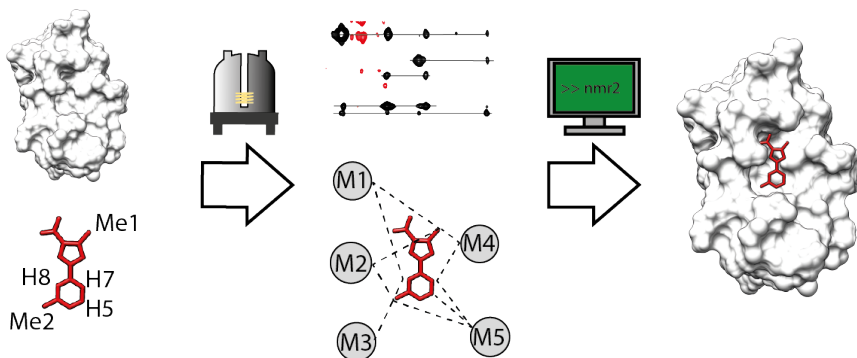


Figure 1.2: NMR² general workflow. The sample is prepared mixing the double labeled [¹³C,¹⁵N]-protein and the ligand. The ligand signals are assigned. A filtered NOESY experiment is measured in a high field NMR spectrometer, and the cross-peaks are used to generate a semiambiguous distance restraints network between the ligand and the protein methyls (M_n). The software NMR² uses simulated annealing to disambiguate the network, and calculate the correct complex structure.

are applied to the calculation of PIN1-fragment complexes, demonstrating the potential of NMR² in the field of fragment-based drug design. Chapter 4: *NMR-based drug discovery for SARS-CoV-2*, reports the attempt of the lab during the Covid-19 pandemic to fuel medicinal chemistry with structural data in short delays using NMR² structure calculation. This later chapter describes the upstream process to structure calculation, screening, and discusses its limitations.

Polarization and photo-CIDNP

NMR signal intensity is proportional to the polarization of the spin distribution. The polarization is defined as the difference between the populations of spins α and β , which is defined according to the Boltzmann population distribution law:

$$\frac{N_{\alpha}}{N} = \frac{e^{-E_{\alpha}/kT}}{e^{-E_{\alpha}/kT} + e^{-E_{\beta}/kT}} \quad (1.1)$$

The polarization is defined by the equation:

$$P = \tanh\left(\frac{\hbar\gamma B}{2kT}\right) \quad (1.2)$$

Since $\hbar\gamma B \ll 2kT$, the polarization is on the order of ppm. For this reason, the sensitivity of NMR is extremely low. In order to achieve sufficient SN the experiments are repeated and added to each other. The signal increases linearly with the number of scans, and the noise follows a square root dependency.²⁵ Hence, the SN only increases by square root of two when the number of scans is doubled. Therefore, the measurement time in NMR is an important limitation. When looking at the Equation (1.2), polarization can be increased by cooling the sample, which is not compatible with biological aimed NMR, or increasing the magnetic field. The technological improvements

over the last 50 years lead to the conception of high-field spectrometers, reaching up to 28.2 Tesla. Nevertheless, such an apparatus is expensive and only brings a polarization improvement of the order 5.7-fold as compared to a 600 MHz spectrometer.²⁵ The introduction of cryo-probes in the 90's improved again the signal-to-noise by a factor 3-fold. However, none of these technological efforts could yield to sufficient sensitivity for the fast analysis of low micromolar or nanomolar samples. A promising alternative to hardware improvement is the generation of out-of-Boltzmann polarization, resulting in hyperpolarization and enhanced signal-to-noise. Several methods have been developed such as dynamic nuclear polarization (DNP),⁶⁰⁻⁶¹ *para*-hydrogen induced polarization (PHIP),⁶² or signal amplification by reversible exchange (SABRE).⁶³⁻⁶⁴ However, these methods have to front many barriers when it comes to applications in the biological context. Dissolution DNP, the solution state NMR version of DNP, needs to be generated in a gyrotron in the solid-state at low temperature (1 K) and transferred to the spectrometer for dissolution. Such conditions require expensive and complex machinery, and the temperature ramp during the experiment (1 K to ~298 K) is incompatible with number of systems.⁶⁰⁻⁶¹ PHIP is successfully applied *in vivo* for a number of molecules, and in some field such as metabolomics. It is important to note that PHIP operates through the hydrogenation of unsaturated carbon chains, and thus requires the chemical modification of the substrate to obtain polarization.⁶² SABRE uses catalyst with more or less specificity to polarize the small molecules, and is therefore particularly useful in material science, and in imaging application with long lived nuclei.^{63, 65} We recommend the excellent review of Kovtunov et al.,⁶⁶ to the reader seeking an extended overview on these techniques. Another hyperpolarization technique is chemically induced dynamic nuclear polarization (CIDNP)⁶⁷⁻⁷², which can be triggered by exciting a photosensitizer (D) which ends up in a triplet state bottle neck and can react with another molecule (M) to form a radical pair, as pictured in the Figure 1.3.⁷³ Photo-CIDNP takes advantage of the interplay between nuclear spin and the electronic recombination of the radical pair to perform spin sorting and bring the polarization out of the Boltzmann equilibrium (Figure 1.3). Because the radical pairs can be generated using light, in solution, and at room temperature, photo-CIDNP is a suitable hyperpolarization technique for bio-NMR. Ideally, if the photo-CIDNP cycle would non-destructive, and could be repeated as many times as necessary.⁷⁰⁻⁷² However, the photosensitizer bleaches subsequently to irradiation, and the intensity of the photo-CIDNP effect decreases accordingly. Photo-CIDNP does not yield to signal enhancement as strong as it is the case for the other hyperpolarization techniques. However, it is simple to implement on any spectrometer as light source are always cheaper, and the light can be simply carried *in situ* thanks to an optic fiber. Even in the case of coherent light sources such as lasers, photo-CIDNP is the most affordable of the hyperpolarization techniques.

This thesis describes the efforts to explore new photosensitizers, and reports the finding of Atto Thio 12 (AT12) which was found to perform especially good for tyrosine hyperpolarization.⁷⁴ Moreover, AT12 study yielded to the serendipitous discovery of an oxidocyclization product of tryptophan with unprecedented photo-CIDNP signal-to-noise enhancement.⁷⁵ The impact of such chemical modification paves the way for the chemical space exploration of tryptophan and tyrosine analogues. The photo-CIDNP performances are known to be modulated by the magnetic parameters: g-factors and hyperfine coupling constants (HFCCs), which are related to the aromatic system.⁷⁰⁻⁷² Yet the magnetic parameters are not sufficient to explain the polarization performance, and we demonstrate the influence of the radical pair reaction rate on the polarization performances. The reaction rates are known to be modulated by the electrostatic interaction between the dye and the molecule in time resolved photo-CIDNP, where nanoseconds irradiation pulses are applied. Hence, when the two molecules carry opposite charges, the reaction rate is faster yielding to higher hyperpolarization.⁷⁶ Interestingly, when irradiation is applied for

seconds, in order to maximize the polarization at low temperature, we describe different reaction rates for two diastereoisomers that are equally charged. Moreover, we demonstrated the importance of sidechain hydrophobicity and the positive effect of the presence of some chemical moieties such as carboxylates. Finally, the study of more tryptophan analogues extends the chemical space of photo-CIDNP active molecule which is sparsely described so far. In Chapter 5 of this thesis: *Exploring the performances of photo-CIDNP through the investigation of new photosensitizers and chemical modifications of tryptophan*, the dye investigation effort and the discovery and characterization of the oxidocyclization of tryptophan, 3 α -hydroxypyrroloindole (HOPI), is reported. It draws also the basics of the influence of non-magnetic parameters based on the study of the HOPI diastereoisomers. Chapter 6: *Basis for the empirical exploration of the photo-CIDNP performances*, reports a simple technique to rank and evaluate the magnetic parameter g-factor of the hyperpolarized molecules, using the Kaptein's rule. Moreover, the effect of sidechain on the photo-CIDNP performances is also investigated.

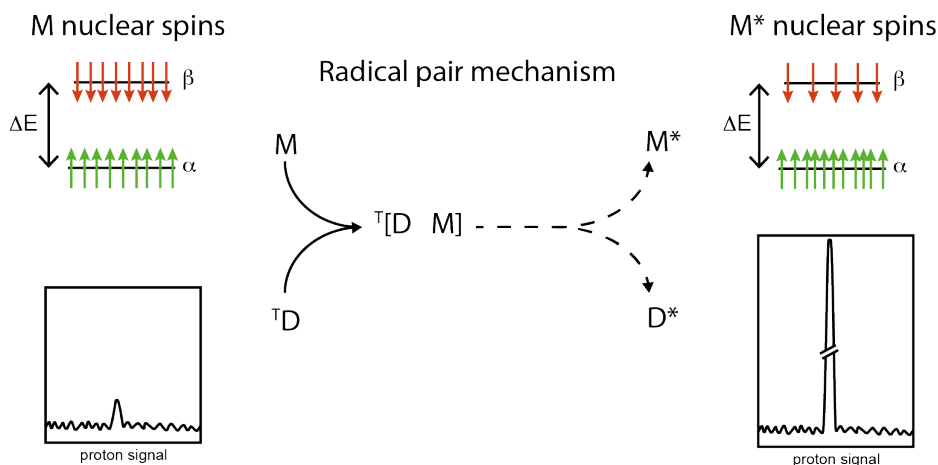


Figure 1.3: General mechanism of photo-CIDNP hyperpolarization. The photosensitizer (D) can form a radical pair by electron transfer with a molecule of interest (M), the nuclear spin dependent recombination of the radical pair to its ground state creates a disequilibrium in the nuclear spin populations of D and M , resulting in enhanced polarization.

Overview of the state-of-the-art NMR used in the present work

In the present thesis, the screening experiments have been performed using STD NMR,³⁶ the hit validation or the affinity titration using $^{15}\text{N}/^{13}\text{C}$ -HSQC, in the case of protein with MW > 15 kDa, the transverse relaxation optimized spectroscopy (TROSY) version of the 15N-HSQC was used.^{32-33, 39} The distance restraints were derived from 2D F_1 -[$^{13}\text{C}, ^{15}\text{N}$]-filtered-[$^1\text{H}, ^1\text{H}$]-NOESY.⁴² These experiments are explained in more detail in the following section.

STD NMR takes advantage of the fact that protein resonances have a broader frequency range than the ligand resonances (Figure 1.4A). Therefore, by applying a saturation pulse specifically in the region -0.5 to -5.0 ppm, only protein methyl proton spins are hit by the rf saturation pulse.

The spin saturation propagates through the protein by spin diffusion, especially because the pulse train is usually applied during 1-2 s (Figure 1.4A a). In the case a ligand binds transiently to the protein, the saturation can be transferred by dipolar coupling to the ligand protons in close proximity to the saturated protein protons. The STD experiments consist into an alternation of experiments with on-resonance (region -0.5 to -5.0 ppm) and off-resonance saturation pulses (>20 ppm). When the difference between the on-resonance and off-resonance spectra is not null, this witnesses to the interaction of the ligand with the target.^{16, 36} The detection pulse (Figure 1.4A b) can be followed by a short (20 ms) spinlock (Figure 1.4A c) to filter out the protein signals.

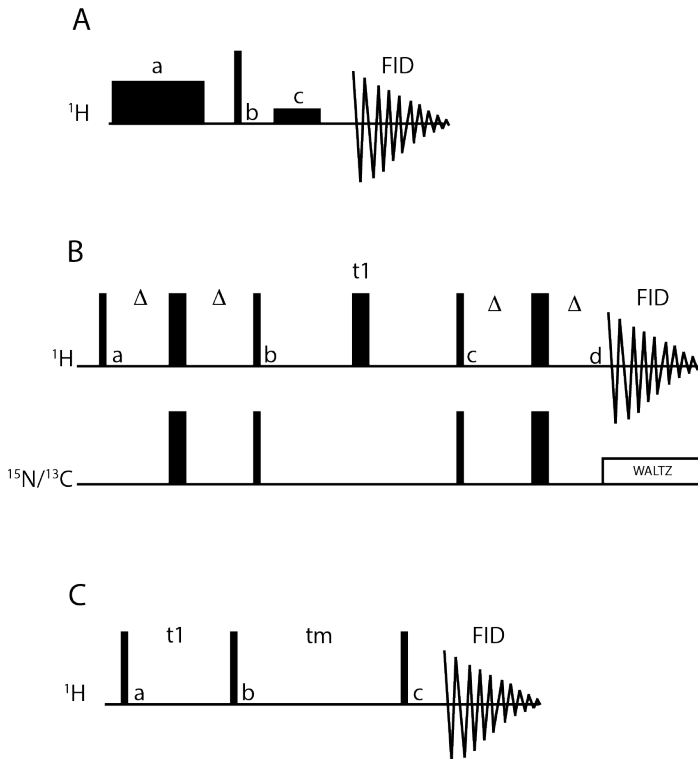


Figure 1.4: Classic NMR pulse sequences used in bio-NMR. A) STD. B) HSQC. C) 2D NOESY. The narrow bars represent 90° pulses and the wide bars 180° pulses. The delays, t_1 , and t_m stand for the indirect dimension and the mixing time respectively.

HSQC is one of the most recorded 2D experiment in the bioNMR field (Figure 1.4B). The magnetization is transferred from proton to heteroatom (^{13}C , ^{15}N) with an insensitive nuclei enhanced by a polarization transfer (INEPT, Figure 1.4B a to b) building-block taking advantage of the scalar coupling between the two atoms (J_{HX}).^{25, 39} The indirect dimension records the heteroatom time dependency (Figure 1.4B b to c) and the polarization is brought back to the proton with a reverse INEPT pulse scheme (Figure 1.4B c to d), and the free induction decay is recorded. In the case of proteins bigger than 15 kDa it is interesting to record the transverse relaxation optimized spectroscopy (TROSY) version of the HSQC pulse sequence, where the J_{XH}

coupling is maintained during the indirect time dependency (180° pulse in the b to c delay) acquisition splitting the signal component of the resonance by $J_{XH}/2$ upfield and downfield in each dimension.³²⁻³³ Hence, the component with the best relaxation properties can be selected by adapting the pulse sequence, which results in narrower linewidth, and therefore better resolution.

The preferred restraints in NMR for structure calculation are NOEs. As we shall see later it is convenient to filter out the protein signal to simplify the analysis of the spectra by observing solely ligand-ligand and ligand-protein cross-peaks. To do so, F_1 - $[^{13}\text{C}, ^{15}\text{N}]$ -filtered- $[^1\text{H}, ^1\text{H}]$ -NOESY is recorded where two half χ -filters eliminate the protons linked to ^{13}C or ^{15}N .^{42, 77-78} Nevertheless, in its simplest form, the 2D NOESY (Figure 1.4C) experiment contains an indirect dimension which is recorded on the proton channel (Figure 1.4C, a to b), and the spins are flipped along the z -axis to enable cross-polarization during the so-called mixing time (Figure 1.4C, b). Finally, the polarization is flipped to the xy -plane to record the FID (Figure 1.4C, c).²⁵

Chapter 2: *NMR*² method development and application to the bromodomain BRD4-BD2

This chapter is an adaptation from unpublished work: Torres F. & Walser R., Kaderli J., Rossi E., Bobby R., Packer M. J., Walker J., Hitchin J. R., Milbradt A. G., and Orts J., NMR Molecular Replacement provides new insights into binding modes to Bromodomains of BRD4 and TRIM24.

Author's contribution: F.T. participated in the protein-ligand structure calculation and refinement in water. F.T. participated in the introduction of the new *NMR*² protocols.

Introduction

Structure-based drug discovery largely relies on structural information from X-ray crystallography.⁷⁹⁻⁸⁰ NMR is in principle able to provide atomic resolution structure as well, in the absence of an X-ray structure. Although it is a reliable technique that can derive both structure and dynamics of a protein-ligand complex, its low throughput is a major drawback and usually does not match the expected timelines of a medicinal chemistry project. As described in the introduction chapter, NMR structure calculation requires the assignment of several hundreds of resonances and of several thousands of cross-peaks. This relies on the measurement of complementary experiments for a duration of up to three weeks and the extensive analysis of the data, up to a couple of months of full-time focus.²⁵ Most of the assignments concern the protein resonances, critical to reconstruct the protein structure in view of obtaining the protein-ligand complex structure. Nonetheless, most of the targets currently under investigation have at least one structure deposited in the protein data bank (PDB). Taking inspiration from the success of molecular replacement for X-ray crystallography, a similar approach has been established recently; it is known as *NMR*², which calculates the ligand pose into its binding site, using a known structure of the protein as a starting point instead of calculating it.

NMR molecular replacement

*NMR*² skips the protein resonances assignment to calculate the structure of a protein-ligand complex, treating NOE-based distance restraints as semiambiguous. This shortens the analysis time to a couple of hours in comparison to the weeks or months required in classical NMR protein-ligand complex structure calculation. The *NMR*² technique was developed in 2016 to rapidly determine protein-ligand complex structures at the binding site at atomic resolution by making use of published structures (apo or holo) of the target protein and treating all observed NOEs as semi-ambiguous restraints.⁸¹⁻⁸⁴ Clear advantages of the method are that it bypasses the long and tedious protein resonance assignment step and it harnesses synergies with other structure determination techniques such as X-ray crystallography. The *NMR*² technique calculates complex structures in a fully automatic way using unassigned sparse NOE data and a structural model of the apo state of the receptor.

General workflow

The NOEs necessary for the distance restraints used by *the NMR*² calculation are measured by a series of a 2D F₁-[¹³C-¹⁵N] filtered-[¹H, ¹H]-NOESY of a sample containing the double labeled ¹³C,¹⁵N-protein and the ligand (Figure 2.1A). The intra-ligand cross-peaks and the ligand-protein cross-peaks are peak-picked, manually or automatically. The ligand resonances are assigned, and the protein resonances are defined as anonymous. Only the intermolecular cross-peaks in the region corresponding to the protein methyls are considered. The cross-peaks intensities are then used to calculate the distance restraints with the Fullmatrix approach (Figure 2.1B), which approach is describe in more details below in the section *distance restraints calculation (eNORA)*.⁸⁵ The semi-ambiguous distance restraints network (Figure 2.1C) is disambiguated by the *NMR*² software which will calculate the different structures possible according to the different protein assignment combinations and select the structure corresponding to the right ligand pose (Figure 2.1D).

Protein expression

The protein expression is a critical step of the NMR² method, a non-uniformly labeled protein will result in poorly suppressed background signal which can interfere with the signal analysis. Therefore, it is important to ensure the production of strictly > 99% labeled material. To do so, the bacteria are transformed normally and plated onto an Agar-Agar-lysogeny broth (LB) media plate, containing the corresponding antibiotics for selection. After the first selection plating is done, a second growing phase on the plate is performed, one plate is prepared per 1 L of media. The second plating phase lasts for 9-12 hours so that the bacteria can grow significantly. Finally, the Bacteria are transferred from the plate directly to M9 minimal media containing ¹⁵N-ammonium chloride and ¹³C-glucose. Doing so avoids the contamination with LB media which can yield to a few percent of unlabeled material and perturbs the protein signal suppression.

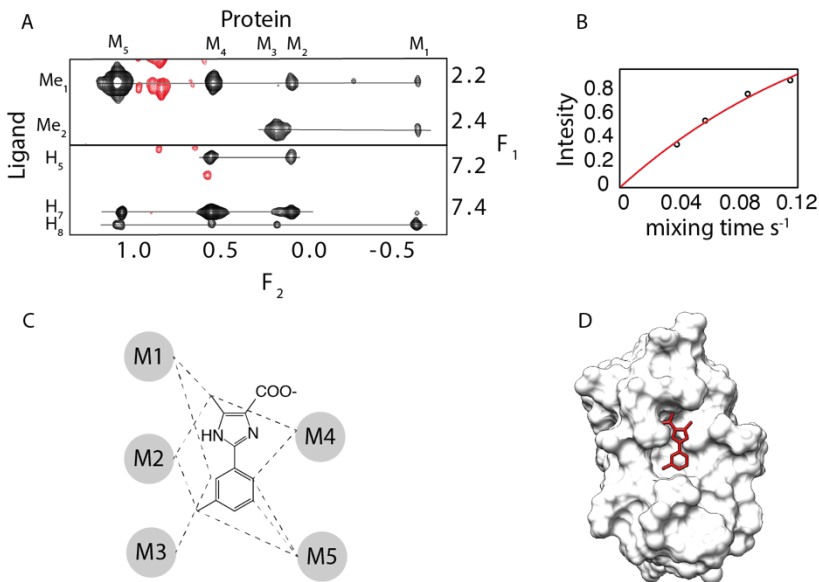


Figure 2.1: Main steps of the NMR² workflow. A) F₁-[¹³C,¹⁵N]-filtered-[¹H,¹H]-NOESY B) Normalized NOE build-up curves C) Distance restraints network D) Final structure.

NOE measurement

As mentioned above, the NOEs, which will be later used for the calculation of the distance restraints, are obtained by measuring F₁-[¹³C,¹⁵N]-filtered-[¹H,¹H]-NOESY (Figure 2.2). The acquisition is done on a high field spectrometer, i.e., 600 MHz and above. As this is a rather insensitive experiment, the higher the field the better so that the SN is improved. In a first instance the magnetization from the proton attached to ¹³C or ¹⁵N is filtered thanks to two consecutive x-filters (point a to b, then b to c). The delays Δ1 and Δ2 are optimized for the suppression of the ¹H-¹³C scalar coupling of 110 Hz and 140 Hz, respectively. These delays are optimized in concert with the carbon shaped pulses sp1 and sp2 which are designed to excite the broad range of carbon frequencies, according to the spectrometer magnetic field following the methodology described earlier by Zwahlen et al.⁴² Only the protons linked to ¹²C or ¹⁴N, i.e., the ligand's protons, are

retained and the protons linked to ^{13}C or ^{15}N , i.e., the protein's protons, are filtered out. The filtered-NOESY is normally recorded for a sample prepared in 100% D_2O so that the contribution of the amide protons is minim, therefore the elimination of the ^{15}N - ^1H (90 Hz) coupled protons is sufficient in practice, despite the non-optimal delays. Then, the indirect dimension is recorded by varying the t_1 delay (d to e) and only the ligand proton magnetization is marked, for a length ranging from 360 (18.3 ms) to 512 (26 ms) points. Afterwards, the spins are flipped back along the z -axis (e) to allow dipolar cross-relaxation; this is the mixing time (t_m , e to f). The magnetization is transferred from the ligand's protons to the other protons close in space; the mixing time is kept between 15 and 120 ms to avoid spin diffusion, and several NOESY with different t_m are recorded. Finally, the free induction decay (FID) is recorded and will be transformed into the direct dimension (F_2) by Fourier transform; as no filter is applied after the mixing time, all the protons are observed regardless of the isotopic labeling. The time dependency of the indirect dimension is later processed into the frequency domain (F_1) by Fourier transform, only the protons bound to unlabeled carbons are observed. The magnetization path can be summarized as ^{12}C - ^1H , ^{14}N - ^1H (F_1) \rightarrow $^{12/13}\text{C}$ - ^1H , $^{14/15}\text{N}$ - ^1H (F_2). The diagonal signals correspond to the magnetization that is not exchanged during the mixing, and they are used to obtain the decay plots from which is obtained the autorelaxation parameter ρ . The off-diagonal signals are the cross-peaks originating from the dipolar cross-relaxation. The intensity of the cross-peaks depends on the amount of magnetization exchanged during the cross-relaxation. These intensities are a function of the duration of the mixing time, and of the distance between the nuclei. The intermolecular distance restraints obtained from the cross-peaks are assigned on the ligand side and anonymous on the protein side. As mentioned in the introduction, only the intermolecular cross-peaks with protein methyls are considered.

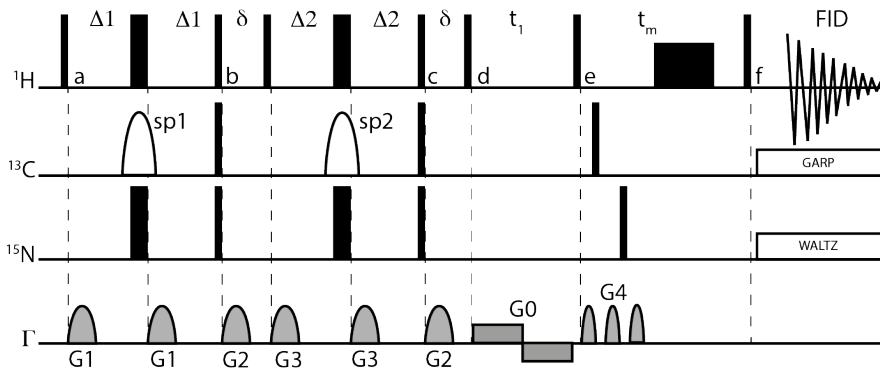


Figure 2.2: F_1 - $[^{13}\text{C},^{15}\text{N}]$ -filtered- $[^1\text{H},^1\text{H}]$ -NOESY pulse sequence. The sharp dark rectangles are for 90° hard pulses and the broad dark rectangles are for 180° hard pulses. $sp1$ and $sp2$ design shaped pulse optimized to excite the broad range of carbon frequencies, they are both Wurst pulses optimized for the desired magnetic field according to Zwalen et al. At a field of 700 MHz, $sp1$ is 1.425 ms; $sp2$ is 356 ms. Where the delays are at the fixed values optimized for proton-carbon coupling frequency selection. E.g., for 700 MHz spectrometer: $\Delta 1=0.685$ ms, $\Delta 2=0.785$ ms, $\delta=0.200$ ms. Γ are the field gradients along the z -axis, with $G0=2$ ms, $G1=0.5$ ms, $G2=1$ ms, $G3=0.5$ ms, $G4=1$ ms. The pulse during the mixing time t_m is a trim pulse for water suppression at same power level as hard pulses and lasts 1 to 2 ms. The indirect dimension acquisition is designated as t_1 .

The spectra are transformed and processed in Topspin[®] to achieve optimal resolution and SN; the spectra are then analyzed in CcpNMR[®]; the intra and intermolecular cross-peaks are picked; the peak lists are exported in the CYANA format using the format converter from CcpNMR[®], and the cross-peaks intensities are used to calculate the distances using the program eNORA.⁸⁵⁻⁸⁶

Distance restraints calculation (eNORA)

The initial slope (σ) of normalized NOE cross-peak intensities build-up curves for different mixing times has a distance dependence in the order of r^{-6} . The distance between two protons is calculated from the NOE intensities using the program eNORA, and is summarized below. The reader is invited to refer to the original publications for more extensive description of the theoretical frame.⁸⁵⁻⁸⁶ The distance restraints are derived from NOE build-up curves using a simple two-spin system model (i,j) and following the established protocol.⁸⁶⁻⁸⁹ The NOE cross-peak intensities, $\Delta M_{ij}(t)$, are normalized by the diagonal intensity at time zero, $\Delta M_{ii}(0)$. The build-ups curves are based on the normalized NOE intensities. While the term $\Delta M_{ij}(t)$ is measured from the cross-peak intensities, the initial magnetization $\Delta M_{ii}(0)$ is not appearing directly in the spectra. Nevertheless, the terms $\Delta M_{ii}(t)$ are simply the intensities of the diagonal peaks, and can be measured from the spectra. Hence, the autorelaxation rates, ρ_i , and initial magnetizations, $\Delta M_{ii}(0)$, are determined using a mono-exponential decay function, $\Delta M_{ii}(t) = \Delta M_{ii}(0) \exp(-\rho_i t)$. Once the build-up curves set, the cross-relaxation rates, σ_{ij} , are fitted following a two-spin system approximation model for the protein-ligand NOEs, $\Delta M_{ij}(t)$, equation (2.1).

$$\frac{\Delta M_{ij}(t)}{\Delta M_{ii}(0)} = -\frac{\sigma_{ij}}{\lambda_+ - \lambda_-} (e^{-\lambda_- t} - e^{-\lambda_+ t}) \quad (2.1)$$

$$\lambda_{\pm} = \frac{\rho_i + \rho_j}{2} \pm \sqrt{\left(\frac{\rho_i - \rho_j}{2}\right)^2 + \sigma_{ij}^2} \quad (2.2)$$

The corresponding distances, r_{ij} , were derived from the cross-relaxation rates, σ_{ij} , defined in equation (2.3):

$$\sigma_{ij} = \frac{b^2}{r_{ij}^2} (6J(2\omega) - J(0)) \quad (2.3)$$

$$J(\omega) = \frac{2}{5} \left(\frac{\tau_c}{1 + (\omega\tau_c)^2} \right) \quad (2.4)$$

$$b = \frac{1}{2} \frac{\mu_0}{4\pi} \hbar \gamma_H^2 \quad (2.5)$$

where μ_0 is the permeability of free space, \hbar the reduced Planck constant and γ_H the gyromagnetic ratio of the nucleus and τ_c the rotational correlation time of the protein. The distances obtained from the set of equations above are forming the semiambiguous distance restraints network between the ligand and the protein methyls which needs to be disambiguated to obtain the protein-ligand complex structure.

Simulated annealing algorithm

The restraints network disambiguation is done by the simulated annealing algorithm of NMR².⁴³ In the first round, a subset of the anonymous methyls is assigned according to all the possible combinations, and the structures for each combination are calculated. The structures, and the respective assignment combinations, for which the distance restraint network violations

exceeding a TF of 20 are rejected. In the further rounds, other anonymous methyls are gradually introduced for disambiguation. Finally, the remaining structures are ranked according to their TF, which value is proportional to the number of restraint violations. A TF contrast (TFC), or the difference between the TF values of two different structures, above 0.2 is considered to be significant enough to consider the structure with the lowest TF to be the best one. Further insights into the importance of the TF and the TFC are provided in the next chapter, *NMR² for fragment-based drug design*.

Introduction of new features into NMR² structure calculation

The introduction of state-of-the-art NMR techniques and their impact on the performances of the *NMR²* software are discussed in this chapter. The first, is the use of interleaved filtered-NOESY pulse program; it is useful to obtain bidirectional NOE build-up curves, which improves the accuracy of the distance restraints. Furthermore, it facilitates the analysis of the NOESY by splitting the spectrum into four different spectra while processing. Such an experiment is well established, and its use in the context of an *NMR²* is attractive as of the density of information it contains. Specific methyl labeling is also a well-established technique in NMR, and is explored in this chapter. It is used to obtain long-range methyl-methyl NOE in big molecular entities such as proteins > 25 kDa.³⁴ The protein perdeuteration reduces the spin diffusion yielding to errors in the distance restraints calculation and reduces the transverse relaxation rate so that the SN is increased. The benefits of methyl labeling and the justification of its application are discussed later. *NMR²* uses distance restraints to perform structure calculation, and also to assess the relevance of the calculated structures. Indeed, the TF increases when the distance restraints are violated, enabling the identification of the right pose, with the lowest TF, which matches with the distance restraints network. Nevertheless, while the presence of NOE cross-peaks witnesses from the spatial proximity between two protons, the absence of NOE cross-peaks is eventually the sign that two protons are spaced by a distance sufficient to avoid magnetization transfer. The use of this information is a promising improvement to constrain the ligand poses even more by generating anti-NOE distance restraints that impose a minimal distance between two protons and contribute to the TF in case of violation.

Application to bromodomains

Bromodomains (BD) defines the protein domains able to recognize lysine residues acetylated on their sidechain amino group. Such kind of residues is particularly found at the N-terminal tails of histones, and therefore bromodomains are involved in transcriptional regulation.⁹⁰⁻⁹⁴ The human genome encodes 46 proteins, spread over eight families, that contain more than 60 different BDs involved in epigenetic regulation and various diseases such as inflammation, diabetes, neurological diseases and cancer through the deregulation of transcription factors, a hallmark of cancer.⁹⁵⁻¹⁰⁴ Bromodomain and extra terminal (BET) form a sub-class family among the bromodomains. BRD4 is a member of the BET family; it is physiologically important during the embryogenesis and cell mitosis. It is involved in cancer through its role in the transcriptional regulation of oncogenes. The inhibition of BET, and among them, BRD4, is a promising approach in the fight against cancer. Several BET inhibitors have now entered into clinical trials. These findings make epigenetic targets, and bromodomains in particular, a relevant field of research for cancer therapeutics.^{96, 100-102, 105-106} Previous work already reported the Bromodomain and Extra-Terminal (BET) proteins being directly involved in cancer and validated the BET proteins as targets for chemotherapies.^{96-98, 104, 107-109} Moreover, the implication of BRD4 in non-transcriptional cellular processes has been recently proposed such as DNA damage repair or checkpoint activation.

The bromodomain-containing protein 4 (BRD4) has two bromodomain reader modules that specifically recognize the acetylation state of lysine side chains on histones. The discovery of iBET762 and other classes of small molecules demonstrated that BRD4 can be targeted by blocking the acetyl lysine binding pocket of its bromodomains with small molecules.¹¹⁰ Structure-based drug discovery has been extensively used to guide the design of these bromodomain inhibitors with over 300 BRD4 BD1-compound complexes deposited in the RCSB. However, there is a paucity of structural information for BRD4 BD2 as highlighted by only a few X-ray-derived entries in the public domain, suggesting BD2 might be less amenable to X-ray diffraction. Therefore, the second bromodomain of BRD2 has often been used as a structural surrogate as BRD2 BD2 has a higher homology to BRD4 BD2 than BRD4 BD1 has to BRD4 BD2.

This chapter reports the structure of the Bromodomain BRD4-BD2 in complex with a small molecule ligand and a new protocol of NMR² structure determination using specifically labelled methyl groups and anti-NOEs. For the calculation of the BRD4-BD2-iBET762 complex, the method was further developed to improve its speed and accuracy by including PA and anti-NOEs. NMR² method is shown to be both able to rationalize SAR data as it is commonly generated in a SBDD project and to allow such insights within the timeframes encountered in a typical hypothesis-synthesis-testing cycle of a medicinal chemistry project.

NMR² method development

Interleaved filtered-NOESY

The preferred experiment to record NOE cross-peaks is the 2D F¹-[¹³C,¹⁵N]-filtered-[¹H,¹H]-NOESY, described in the introduction, which only encodes for the ligand-ligand and the ligand-protein distance restraints. Moreover, the two types of NOE cross-peaks are present in the same 2D spectrum, complicating the analysis. Yet the possibility of measuring interleaved spectra is attractive considering the free induction decays are latter split into four spectra, each encoding for the ligand-ligand, ligand-protein, protein-ligand, protein-protein distance restraints. The analysis is then facilitated and provides bidirectional cross-peak build-ups yielding to high quality distance restraints. Additionally, the ligand-ligand cross-peaks are used for the ligand signal assignment and structure calculation, and the protein-protein distances can help to identify the peaks that belong to the same residue. Indeed, the methyl-methyl distance is fixed for the different residues at 2.5 Å (Val, Leu) and 3.5 Å (Ile). The pulse sequence can be conveniently found in the built-in pulse sequences from Topspin® (noesygpplx19). Nevertheless, the presence of the Watergate building block, to attenuate the water signal, at the end of the sequence is also attenuating the signals close to the carrier in the direct dimension, so that the intensities between 6 and 3.5 ppm are not trustworthy. The correction of this inconvenience is straightforward as the sample is prepared in D₂O, and therefore the pulse sequence can be conveniently modified to remove to Watergate final building block. Importantly, the interleaved filtered-NOESY yields artifacts while measuring double labeled [¹³C,¹⁵N]-protein samples; so far it provided satisfying spectra only for specifically methyl labeled samples. In addition, the pulse sequence takes four times longer to measure than the classic filtered NOESY, and it provides only twice as much as information. Hence, the interleaved filtered NOESY, should be reserved for challenging systems, and not be applied systematically.

Specific methyl labeling

The contribution of spin diffusion to the intensity of the NOE cross-peaks is a major source of bias during the distance restraints calculation. Spin diffusion is defined as the magnetization transfer from a spin to another through remote spins. Therefore, the intensity of the cross-peak is increased by the contributions of the remote protons, and the distance is underestimated (Figure 2.3A). Several methods have been developed to control or correct for spin diffusion.¹¹¹⁻¹¹³ The program eNORA can correct the spin diffusion and correct the intensities to obtain exact NOE distance restraints. Nevertheless, this approach requires the complete assignment of the cross-peaks and a preliminary knowledge of the structure, two unmet conditions in NMR², which therefore does not include spin diffusion correction. Instead, the buildup plots showing clear evidence of spin diffusion contamination are discarded. The spin diffusion is recognized when a second order contribution to the intensity build-up is observed. Nevertheless, it can be necessary to avoid spin diffusion in order to obtain more accurate distance restraints, and also include some more potentially useful distance restraints. Spin diffusion is mostly caused by the protein's protons as they represent most of the protons of the protein-ligand complex. The relevant protons for the NMR² structure calculation are the ligand's protons and the protein methyls' protons. A straight forward approach is to remove surrounding protons by perdeuteration (Figure 2.3C). Specific labeling schemes with re-protonated methyl groups in a deuterated background have been extensively used to study large molecular complexes.¹¹⁴ Methyl groups are particularly interesting because of their sharp resonance lines, even spread throughout the protein structure and abundance in the proteome, included within the hydrophobic environment of the binding sites. The specific labeling of protein with ¹³C,¹H-methyls while the backbone and rest of the sidechains are perdeuterated and with natural abundance carbon isotope (¹²C), as already been described, in principle for the calculation of very large protein.^{34, 115} Generating proteins with such labelling pattern has become straight-forward for *E. coli* based expression systems through the use of commercially available precursors and even complete labelling kits.¹¹⁶ The replacement of the protein background protons by deuterium, considerably reduces the spin diffusion, as the contribution of protons external to the ligand-protein methyls system to the NOE cross-peak intensities is suppressed (Figure 2.3C). Furthermore, the background signal corresponding to the protein protons that were not suppressed by the double x-filter, is drastically reduced.

Moreover, the proton density around the protein methyl participates in the transverse relaxation, T_2 . Hence, the perdeuteration of the protein reduces the transverse relaxation rate, $R_2 = 1/T_2$ (Figure 2.3B and D). Smaller R_2 yields narrower linewidth and improved SN. For these reasons, the production of perdeuterated specifically labeled methyl protein is particularly attractive for the measurement of filtered-NOESY of protein-ligand complexes when the protein is of big molecular weight, i.e., MW > 25 kDa.¹¹⁵ The expression of specifically labeled protein has been previously described and applied to many different proteins. As of the costly specifically labeled precursor, the specific labeling strategy is to be adapted to the methyl present in the binding site. However, different labeling schemes are commercially available and sold as kits.

Anti-NOE

The structure calculation of the protein-ligand complex classically relies on the presence of distance restraints obtained from the NOE cross-peaks.¹¹⁷ These distance restraints constrain the distance between two protons into a range comprising a minimum and a maximum distance. The concept of anti-NOE relies on the absence of NOE between two protons, and therefore sets a distance restraint with only a minimum spacing between the atoms. The effect of the anti-NOE on the TF is to act in the same fashion as an exclusion sphere: if the ligand adopts a pose which

does not fulfill either the NOE or the anti-NOE restraints, the TF will increase. The figure 2.4 is a schematical representation of a NOESY spectrum for the spin system $H_{1,2,3}$ and $M_{1,2}$; the anti-NOEs would then be defined from the absence of cross-peaks, i.e.: H_1-M_2 and H_3-M_1 .

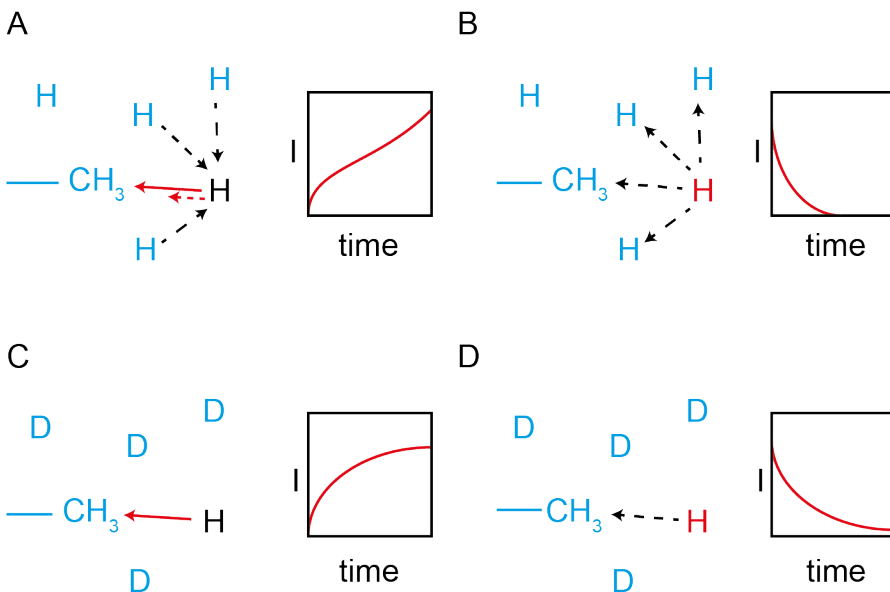


Figure 2.3: A) non deuterated protein, build-up plot. The red solid arrows represent the cross-relaxation between the proton and the methyl, and the red dashed arrow is the second order cross-relaxation resulting from the spin diffusion. B) non deuterated protein, decay plot. The dashed arrows are the magnetization lost by the diagonal term due to transverse relaxation. C) perdeuterated specifically labeled protein, build-up plot. The red solid arrows represent the cross-relaxation between the proton and the methyl. D) perdeuterated specifically labeled protein, decay plot. The dashed arrow is the magnetization lost by the diagonal term due to transverse relaxation.

The input corresponding to the spin system of the figure 2.4 would be:

```
>> [Distances]
>> H1 M1 = 1.6 2.4.      ### NOE-based distance restraints
>> H2 M2 = 2.4 3.6.      ### NOE-based distance restraints
>> H2 M1 = 1.6 2.4       ### NOE-based distance restraints
>> H3 M2 = 2.4 3.6       ### NOE-based distance restraints
>> H1 M2 = 3.0 15.0      ### anti-NOE-based distance restraints
>> H3 M1 = 3.0 15.0      ### anti-NOE-based distance restraints
```

While several reasons can cause an NOE cross peak to disappear, such as dynamics induced line broadening or spectral artifacts, we follow conservative guidelines to make use of them. A missing NOE cross peak is exploited only if both protons involved exhibit visible NOE cross-peaks with other partners. The absence of an NOE can therefore unambiguously be attributed to a large distance between the two nuclei, and other effects such as conformational exchange processes, which can lead to peak broadening below the noise level, can be excluded. The anti-NOE approach needs to define the lower limit used for the anti-NOE based distance restraints. Indeed, if the NOE effect can generate cross-peaks for protons separated by up to 6 Å distance, the maximal distance observed depends on the mixing time and also on the correlation time of the system. The safest way, in our opinion, is to set of the anti-NOEs at a minimal value of 3 Å and 3.6 Å lower limit distance restraints when one methyl or two methyl groups are involved, respectively. The upper limit of the anti-NOE should be carefully set to approximately the protein radius, assuming the protein studied is globular.

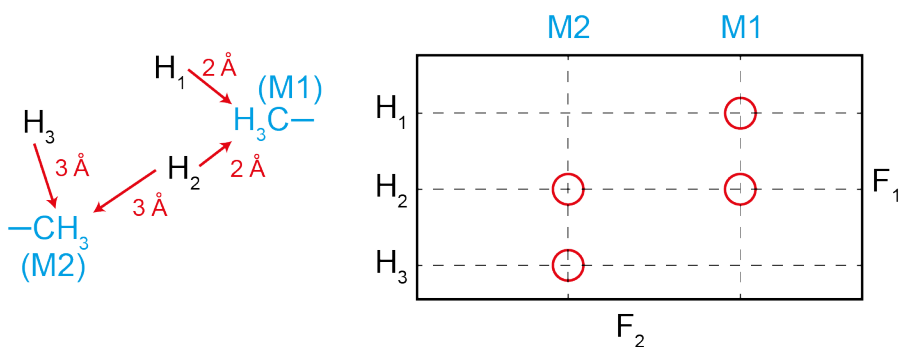


Figure 2.4: Schematical representation of a ligand-protein spin system. The protons in black are of the ligand and in blue of the protein. The red circles represent the NOE cross-peaks. The intersections where no cross-peaks are represented define the anti-NOEs.

Application to the NMR² of BRD4-BD2-iBET762 complex structure

The ligand, iBET762, was assigned simply using the interleaved filtered NOESY corresponding to the intra-ligand pathway, i.e., $[^{12}\text{C}, ^{14}\text{N}]^{-1}\text{H} \rightarrow [^{12}\text{C}, ^{14}\text{N}]^{-1}\text{H}$ (Figure 2.5). The assignments were based on the chemical shifts, characteristic for aromatics and methyls, and on the presence of cross-peaks witnessing from space proximity. Furthermore, distances obtained from the cross peaks validated the assignments and enabled the calculation of the bound-ligand conformation. For the BRD4 BD2-iBET762 complex, 6 intra-ligand and 39 inter-molecular NOEs could be measured, from the ligand-ligand, protein-ligand, and the ligand-protein filtered NOESY spectra (Figure 2.6A). Among the inter-molecular NOEs build-up curves of sufficient quality that were collected, 24 exhibit an excellent fit to the isolated two spin system model (Figure 2.6C). In the cases where the intermolecular NOE build-up curves showed a slight deviation from the two-spin build-up curve model, they were used with a large tolerance for the upper limit restraint of 5.5 Å when a single ligand proton and 6.5 Å when two methyl groups were involved in the NOE, and

no lower limit was applied. NOE build-up curves exhibiting a poor fit, due to spectral noise or strong spin diffusion, were discarded. The quality of the fit is judged visually by observing the build-up plots, and therefore the experience of the user remains an important variable.

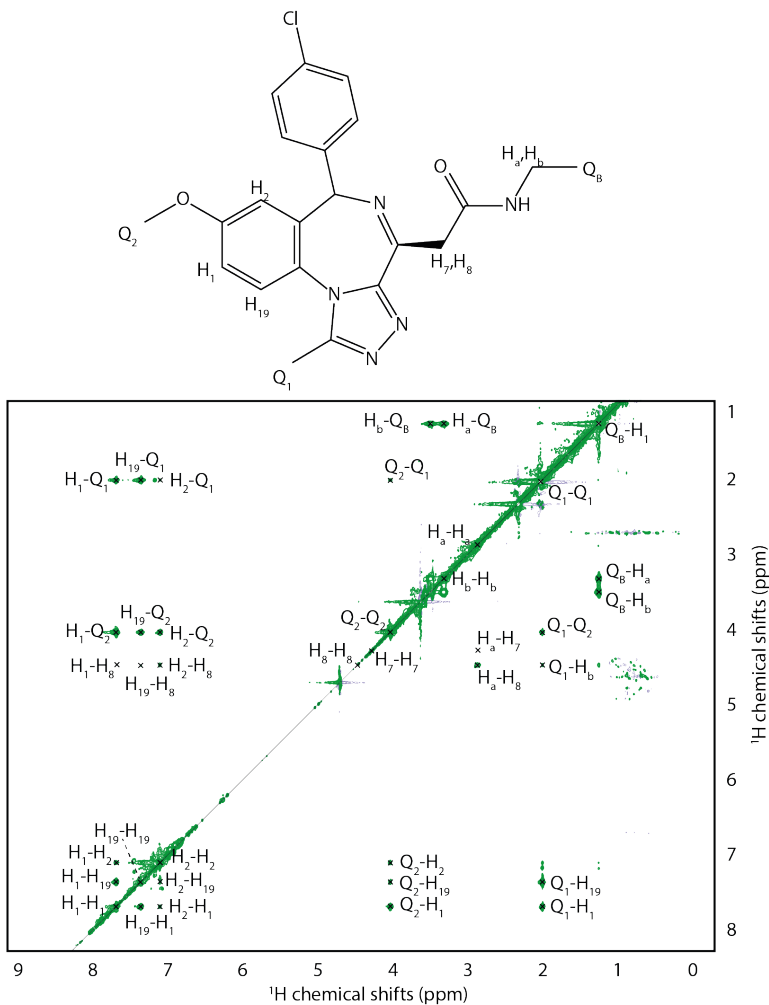


Figure 2.5: Ligand-ligand filtered-NOESY spectrum (mixing time 150 ms) and ligand's proton assignment.

Although the traditional approach to derive structures by NMR spectroscopy does not consider missing NOE cross-peak as usable information, in principle it nonetheless contains structural information, namely that two protons are sufficiently far from each other, and can support a structure calculation in the form of lower distance limits. Conservative guidelines, for the reasons aforementioned were applied: 3 Å and 3.6 Å for a proton-methyl and methyl-methyl missing NOE,

respectively. Additionally, anti-NOE restraints were used when an NOE cross-peak was missing but the two proton groups exhibited inter-molecular NOE(s) with other moieties.¹¹⁸ Anti-NOEs contributed to an additional 37 lower limits restraints. Protein methyl-methyl distances were also observed in the half-filtered NOE experiments and the pairing of the prochiral methyl groups of the valine and leucine amino acid residues could be derived thereof.

NMR^2 structure calculations were performed using intra-ligand and inter-molecular derived distances, anti-NOEs distance restraints and prochiral methyl groups pairing information. The total NMR^2 structure calculation time was ~12 minutes on a single machine (MacBook pro 2.7GHz Intel Core i7).

The capabilities of NMR^2 to provide structural information with a turnaround time suitable for medicinal chemistry were further explored, using the complex BRD4-BD2-iBET762. The specific methyl $^{13}C, ^1H$ -labeling strategy described before, was applied to the BRD4-BD2 domain on the methyl groups of the isoleucine, leucine, and valine amino acid residues, in an otherwise fully $^{12}C, ^1H$ background (Figure 2.6B). This labeling strategy provided beautiful NOESY spectra without residual background signal. In addition, since spin diffusion is reduced by the chosen labelling scheme, the NOESY spectra are nearly free of noise even at long mixing times and distal protons do not exhibit NOE cross-peaks.

Overall, we could add another 37 anti-NOEs restraints raising to a total of 76 protein-ligand restraints, improving convergence and speed of the NMR^2 calculations. Furthermore, since the total amount of structural information increased, we could also increase the tolerance on the calibration of the NOE restraints as shown from the distance calibrations on the bromodomain system. This improves the robustness of the NMR^2 method by reducing errors that could stem for example from non-optimal NMR experiment parameters, bad cross-peak integration or wrong estimation of the complex correlation time. Finally, the half-filtered NOESY experiment provides the possibility to group the prochiral methyl groups of the Leu and Val amino acid residues based on the strong intra-residue prochiral methyl-methyl NOE observed in the protein-protein NOESY spectrum. The presented NMR^2 structure of BRD4-BD2-iBET762 exhibits the conserved interactions that have been reported for this class compound sharing the same scaffold (Figure 2.7). Namely, the hydrophobic interactions are embracing the whole ring system and the hydrogen bond between the Kac mimicking epitope and the BRD4 N433 are present (Figure 2.7A). The fast-computational time, ~12 min., to derive the crucial structural features and the relevant pharmacophore makes NMR^2 an interesting method for SBDD.

Conclusion

NMR^2 is an established technique, which has been successfully demonstrated for different targets. The methodology has so far relied on the acquisition of the two-dimensional $F^1-[^{13}C, ^{15}N]$ -filtered- $[^1H, ^1H]$ -NOESY experiment, where the target is a $^{13}C, ^{15}N$ -labeled protein. Ways to lower the time-to-structure and to allow the more general applicability of the NMR^2 method were investigated.

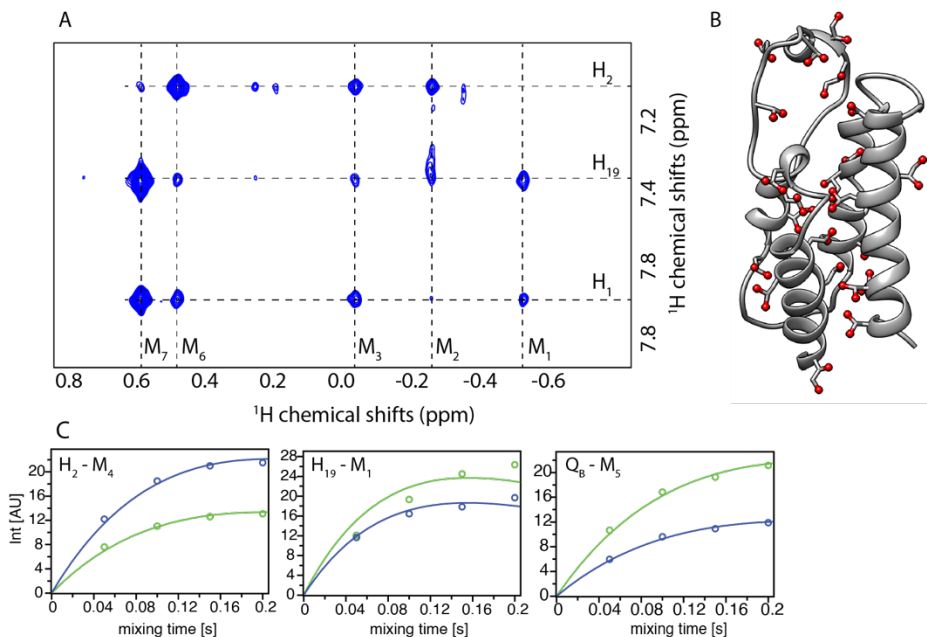


Figure 2.6: (a) Methyl-aromatic region of the $^{15}\text{N},^{13}\text{C}$ -filtered- $^1\text{H},^1\text{H}$ -NOESY spectrum of BRD4-BD2 ($350\ \mu\text{M}$) with the ligand (1:1) measured on an 800 MHz spectrometer at 200 ms mixing time. (b) Ribbon representation of BRD4-BD2. The amino acids residues Ile, Leu, Val are depicted with red sticks and the methyl groups are depicted with spheres. (c) NOE build-ups for the intermolecular NOE cross-peaks measured at mixing times 50, 100, 150 and 200 ms and from both sides of the NOESY spectra diagonal. The build-up curves were fitted using equation (1). The proton names for the ligand are reported in Figure 2.3 and the methyl groups of the protein are arbitrarily called M#.

This chapter reports the possibility of using an interleaved experiment that records four spectra, each encoding for the ligand-ligand, ligand-protein, protein-ligand, protein-protein distance restraints, respectively. The advantage is the possibility to assign the ligand and the possibility to derive the methyl distance restraints to determine the methyls that belong to the same residue. In addition, this approach provided a dense network of distance restraints which is favorable for the algorithm convergence. The limitation of using such interleaved spectra is the low SN hindering its use to systems that are not specifically $^{13}\text{C},^1\text{H}$ -methyl labeled. We therefore tested the applicability of the NMR² method to selective methyl-protonated systems in an otherwise perdeuterated background. Such labelling schemes can extend the accessible molecular weight range for NMR experiments to beyond 1 MDa.^{114, 119} As such it presents in our view the most general method to make the NMR² method applicable to targets of all sizes.

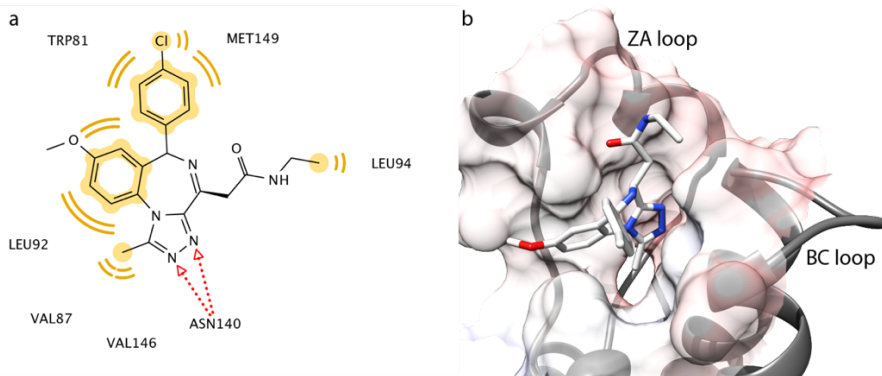


Figure 2.7: (a) Stick representation of compound 2 and its interactions with BRD4-BD2. Yellow coloring represents the hydrophobic interactions between ligand and protein. The red arrows depict intermolecular H-bonds. (b) Ribbon representation of BRD4-BD2 in complex with compound 2. The protein ribbons are depicted in grey, the ligand is represented with sticks and the protein surface is shown as semitransparent solid and shaded according to the electrostatic potential.

We introduced anti-NOEs for the NMR Molecular Replacement method in order to gather more structural restraints driving the structure calculations. The absence of an NOE between two nuclei is not strictly speaking proof that the two nuclei are not close to each other in space. However, in the case where two nuclei are indeed far from each other in space, this information contains valuable information that could be harnessed in structure calculation. This has been realized by other groups and has found different applications, e.g., in automated resonance assignment generation.^{53, 118} We sought to make use of such anti-NOE restraints in a very conservative fashion. We only accepted the use of an anti-NOE where both individual resonances are clearly visible, and both individual resonances show NOEs to other resonances. Anti-NOEs contain a wealth of information and have the potential to significantly improve both the convergence and the accuracy of structure calculation especially when the restraints are ambiguous like in NMR^2 . While the information content of an anti-NOE is lower compared to a visible NOE, the number of anti-NOEs is by far more important. Finally, the use of anti-NOEs will reach its full potential in cases when the ligand can freely sample a large conformational space within the binding site, or when the ligand has a low molecular weight, for instance, fragments.

Finally, the NMR^2 derived structure is the first report of an BRD4-BD2-iBET762 complex; hitherto only structures of iBET762 in complex with BRD4 BD1 (3p5o) or BRD2 BD1 (2yek) have been published. Although iBET762 has been shown to bind to various members of the BET family, published data suggests that the interaction of iBET762 with BRD4 BD2 has one of the highest affinities among members of the BET,^{107, 120} underlining the relevance of this interaction. Furthermore, a subtle feature of the NMR^2 structure is the presence of a dual hydrogen bond between the strictly conserved Asn 433 and the triazolyl ring of iBET762 (Figure 2.7A) and a concomitant slight movement of the compound out of the pocket.

In summary, the speed and robustness of the NMR^2 method were improved by the introduction of ^{13}C , 1H -methyl specific labeling and anti-NOE distance restraints to the structure calculation. The computation time could be reduced to a few minutes on a single desktop machine, making the

method better suited for the short design, synthesis, test cycles encountered in drug discovery. We envision that NMR^2 will become an important tool in structure-based drug design that can be used by a broad population of users, that are not NMR specialists. NMR^2 has the potential to become a new standard in structure-based drug design competing or complementing the current gold standard, X-ray crystallography, or the newly emerging field of cryo-electron microscopy. Finally, the structure obtained could be deposited in the PDB, after CNS water refinement. The procedure is described in the Appendix 2.

Chapter 3: NMR² for fragment-based drug design

This chapter is adapted from the published article: *Felix Torres, Dhiman Ghosh, Dean Strotz, Celestine N. Chi, Ben Davis, Julien Orts, Protein-fragment complex structures derived by NMR Molecular Replacement, RSC Med. Chem., 2020, 11, 591-596*;¹²¹ and from the unpublished manuscript *Felix Torres and Julien Orts, Hallmarks of NMR² protein-fragment complex structure calculation.*

Author contribution: F.T. participated in the expression of the double labeled ¹³C,¹⁵N-PIN1 protein, measured all the NMR experiments, performed the data analysis and the structure calculations.

Introduction

Fragment-based drug design (FBDD) is a powerful method used to identify chemical starting scaffolds and optimize them to obtain lead molecules. FBDD provides a clear structure activity relationship with simple chemical building blocks that can be relatively easily functionalized and cover a significant amount of chemical diversity. The common approach starts with screening a library of thousands of fragments, ca < 300 Da, onto a validated target. Keeping the compounds' size small enables to probe a larger chemical space as it would be the case with more complex molecules.^{14, 122} Considering that the full chemical space for small molecules (< 1000 Da) is 10^{60} , a library of 10^7 molecules would represent only an infinitesimal portion of it.¹²³⁻¹²⁴ However, the chemical space of small fragments (< 160 Da) is estimated to be around 14 million compounds,¹²⁵ therefore a library of even 5k compounds is covering a higher percentage of the chemical space as compared to a library of bigger molecules. This higher coverage guarantees, if the library is properly designed, to find more diverse starting points for synthetic routes to obtain lead compounds, theoretically improving the chances of success of the drug discovery campaign. The fragments identified to bind to a target, can be evolved by growing, meaning the addition of supplementary chemical functions to improve the efficiency. Alternatively, two fragments binding to the same target can be linked together to obtain a compound of higher affinity and specificity. The growing/linking of the fragment into a more complex molecule can increase the affinity from mM to nM, with a few modifications, a key feature of the FBDD (Figure 3.1). It is common to observe that the linking of two fragments together, or the addition of a simple chemical function provides a new molecule with 3-4 orders of magnitude higher affinity. Structural data on the protein-ligand complex is critical to rationalize the fragment evolution and orient the combinatorial chemistry strategy.

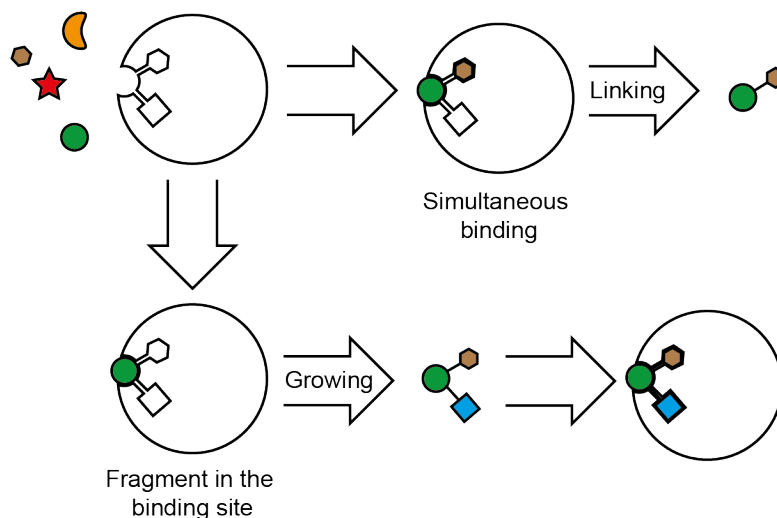


Figure 3.1: Schematic representation of fragment-based drug discovery approach. The fragment that bind simultaneously can be linked when their relative position is known. When a fragment binds to the protein, the structural information about the binding site enables to grow the fragment to build new interactions.

State-of-the-art methods for fragment-based drug design

The screening can be performed with different biophysical techniques such as X-ray crystallography,¹²⁶ surface plasmon resonance¹²⁷ or more recently DNA encoded libraries.¹²⁸ The hit molecules collected from this first cycle need to be validated, often performing NMR ¹⁵N-HSQC experiments.¹⁴ Optimally, the final step, before engaging the hits into the medicinal chemistry pipeline, is to obtain structural information. The structure of the protein-ligand complex helps to understand the key features of the interactions and facilitates the design of a lead compound by growing or linking the initial fragments (Figure 3.1). The gold-standard method for protein-fragment structure characterization is X-ray crystallography.^{122, 129} However, due to the weak affinity of most fragments, co-crystallization is particularly challenging. To address this issue, fragments are soaked into the crystallized protein.²³ Nevertheless, some systems can be incompatible with this approach. This can be the case when the protein does not crystallize or the crystal is impervious to soaking. Moreover, the crystallized protein often adopts an energetically minimized landscape which can significantly differ from the solution state. Consequently, some key interactions between the ligand and the protein can be hindered depending on the crystal form.¹³⁰ In addition, the crystallization buffer used for the soaking maintains the crystal integrity, and this could be incompatible with the required fragment concentration (>10 mM).¹³¹

On the other hand, NMR spectroscopy is particularly suitable for the observation and the structural characterization of weak interacting systems, such as fragments, in solution. While NMR is a highly versatile structural technique providing structural as well as dynamic information, its application in the context of structure-based drug design remains infrequent. This is mainly due to the laborious process of determining macromolecular three-dimensional structures by NMR. The complexity of spectra acquisition and interpretation as well as considerable time required for resonances assignments, derivation of distance restraints and structure calculations remains a cumbersome and time-consuming procedure compared to the equivalent protocols used in X-ray crystallography.

Recently the *NMR*² method has been established to solve the structure of the interaction site of protein-ligand complexes in fully automated manner reducing the timespan from months to days of analysis to calculate protein-ligand complex structures.⁴³ It uses a simulated annealing algorithm that calculates the complex structures from semi-ambiguous intermolecular distance restraints obtained from NMR. The restraints are defined as semi-ambiguous because the ligand's signals are assigned while the protein methyls' signals are not. Moreover, opposed to other fast NMR based drug design methods, such as chemical shift perturbations, it is not a modelling approach because it fully relies on experimental data.⁴⁶ The *NMR*² workflow is extensively described in the introduction of the previous chapter. While the method was successfully applied for ligands with strong and weak binding properties including small molecules and peptides, its applicability on libraries of fragment ligands remains to be shown. Thanks to its reduced timeline, *NMR*² is a serious alternative to X-ray crystallography for the structural determination of protein-ligand complexes, especially in the case of weak binders such as fragments. However, due to the small size of the fragments, ca < 300 Da, the number of distance restraints involved is lower than for the drug sized molecules, and therefore can hinder the algorithm convergence. Hence, the structures of the protein-fragment complexes are particularly challenging to achieve. The *NMR*² software has been benchmarked with the SERAPhic fragment library in order to define the technical challenges to address in protein-fragment NMR structure calculation. Moreover, a straightforward strategy is proposed to guarantee the success of the complex structure calculation. The presented work provides hence clear guidelines for the use of *NMR*² in a fast track for protein-

fragment complex structure resolution as an alternative to X-ray crystallography. NMR^2 disambiguates the distance restraints network obtained from the filtered-NOESY spectra, using a simulated annealing approach. All the assignment combinations compatible with the binding site (defined by the user) are generated, and the structure for each combination is calculated. The latter step results in thousands of structure calculations with different assignment combinations. The correct structure is the one calculated with the right methyl assignments and does not violate, or very less, the distance restraints network. Therefore, the nearly absence of violations is expressed by a low TF, close to 0. Interestingly, some assignment combinations that are wrong can provide the right structure; this happens often when the methyls borne by a same residue (Ile, Leu, Val) are swapped, and therefore the target function increase is small. At the end of the NMR^2 calculation, the ten best complex structures, with the lowest TF, are outputted. In the case if two or more different ligand poses are observed among these ten best structures, it is critical to validate the ligand pose corresponding to the lowest TF as the correct one. To confirm the validity of the best ranked complex structure, the TF contrast (TFC) with the first significantly different ligand pose (root mean square deviation (RMSD) $> 2 \text{ \AA}$) is calculated. The best ranked structure is validated if the TFC is above 0.2.⁴³ The choice of the minimum TFC value 0.2 was based on the true positive and on the false positive rates (Figure 3.2). With a minimum TFC value lower than 0.2, false positives results start to appear and with a minimum TFC value higher than 0.2 some true positives are not detected anymore. In the case of fragments, as a consequence of the molecular size, the scarce number of distance restraints can result in a low amount of restraints violations. This results in a low TF of the wrong ligand pose as compared to the TF of the right ligand pose, and therefore the TFC is below 0.2, hindering the recognition of the right ligand poses.

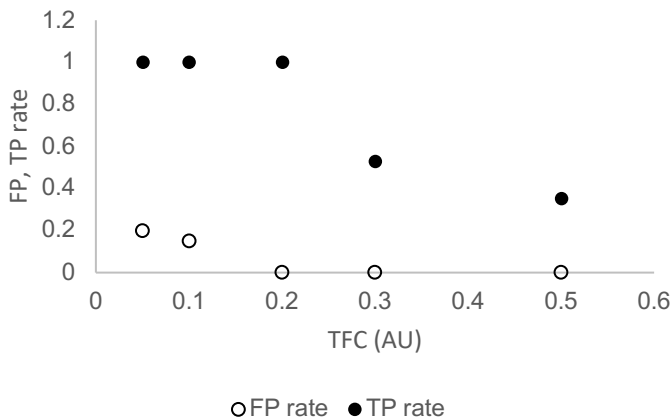


Figure 3.2: Evolution of the false positive (FP) and true positive (TP) rates at different TFC values. The threshold value of 0.2 is justified by a maximum number of TP and a minimum number of FP. The rates are calculated for the 37 structure calculations with 20% of error applied to the distance restraints.

Benchmarking NMR² for protein-fragment complex structure calculations

In an attempt to establish a general methodology for the use of NMR² in the protein-fragment complex structure calculation, the NMR² algorithm has been benchmarked with the SERAPhic library.¹³² The SERAPhic benchmark library contains high quality protein-fragment X-ray structures. The structural heterogeneity of the fragment is ensured by maintaining the Tanimoto similarity coefficient below 0.85. Similarity coefficients are usually defined as $similarity = \frac{1}{1+distance}$, where the distance metrics are in this case calculated according to the Tanimoto methodology.¹³³ This test calculates the distance of two binary vectors characterizing the structures of the molecules. The binary vectors transcribe the substructures as bit into bit encoding for atoms types and bonding, the combination of the binary encoded substructures finally englobes the full molecule structure. When this distance is close to 0, the similarity is close to 1. This test is popular in the medicinal chemistry field and shows robust performances,¹³⁴ ensuring the chemical diversity of the SERAPhic library. Furthermore, the properties of the fragments such as number of H-bond donors/acceptors, number of torsionals, number of rings, log P were shown to overlap well with the molecules in the DrugBank and KEGG COMPOUND repositories.

The present work provides insights into the complex interplay between the experimental error, the amount of information contained in a sparse NMR dataset, the topology of the binding site and the result of NMR² calculation. Finally, we propose a strategy to standardize the data analysis of structural data and reinforce success in FBDD. Finally, we apply the teachings from the benchmark to the study of peptidylprolyl *cis/trans* isomerase, NIMA-interacting 1 (PIN1) protein in a complex with three fragment isosteres, and another fragment of different scaffold. We anticipate that this approach will find a broad application in fragment-based lead discovery. Indeed, the time-to-structure is now under a delay of a week using NMR², which is in line with the medicinal chemistry timeline.

Application to PIN1-fragment complex structures calculation

PIN1 is a peptidyl-prolyl *cis/trans* isomerase that recognizes phosphor-serine/threonine-proline motifs, and a critical modifier of multiple signaling pathways. It is overexpressed in several cancers and its activity contributes to tumor initiation and growth.¹³⁵⁻¹³⁶ Several studies reported inhibitors of PIN1 but no drug has yet reached the market. One of the first developed inhibitors, Juglone, did not lead to a drug candidate due to the lack of selectivity.¹³⁷ Furthermore, the attempts of modification of juglone or of medicinal chemistry from peptide analogues of the phospho-threonine/serine-proline motifs all failed.¹³⁸ More recently, a number of phenylimidazole inhibitors were identified in the context of an X-ray crystallography-based fragment screening campaign (Figure 3.3).¹³¹ These complexes are typically obtained by soaking, and the X-ray structures were obtained for several of these analogues. The three fragment isosteres have been identified from the phenylimidazole campaign and the structure of fragment 2 in complex with PIN1 has already been solved by X-ray crystallography and deposited in the protein data bank (2xp6).¹³¹ A fourth benzindole fragment, shown in Figure 3.3, sharing common features with the phenylimidazole isosteres has also been calculated in complex with PIN1. The structures of PIN1 in complex with these four fragments could be successfully calculated by applying the teaching from the NMR² benchmark. The total analysis time can be accounted in hours and the final calculations could be performed in less than 10 minutes. This application showcases well the potential of NMR² in the context of FBDD, where a set of different protein-fragment complex structure are often sought for further optimization.

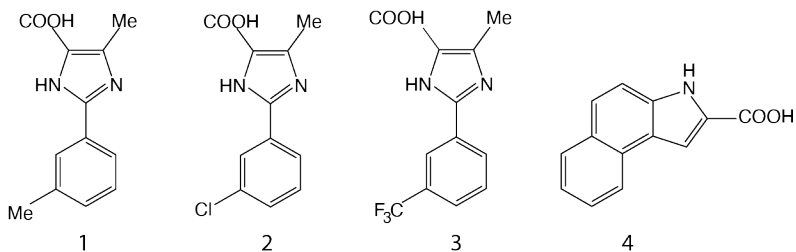


Figure 3.3: Structure of the four different fragments binders to PIN1.

NMR² benchmark for protein-fragment complexes

SERAPHic dataset

Prior to conduct the benchmark of the NMR² software, the intermolecular distances between the protons of the ligands and the methyls present in the binding site of the proteins were measured *in silico*. The identities of the ligand's protons were maintained while the protein's methyls were anonymized as M_n, where n is an arbitrary number attributed to each methyl. These semi-ambiguous distances constitute the intermolecular distance restraints network of each complex. Finally, the structures were calculated using NMR². The used dataset is composed by 39 monomeric protein-ligand complex structures providing a variety of ligand and binding pocket sizes, and representing a wide range of different pocket/ligand size ratios. This variety is important to ensure that the NMR² method is versatile for different topologies. The Table 3.1 provides an overview about the different metrics of the dataset. Interestingly, the median distance restraints length is not correlating necessarily with the pocket/ligand ratio,¹³² resulting from the fact that only the distances with methyl containing residues are assessed. However, the amount of proton anchors, i.e. the ligand protons involved in at least one distance restraints, shows overall proportionality to the size of the ligand. The respective impact of the amount of distance restraints a ligand is involved in and the size itself of the ligand are discussed later in this chapter.

Table 3.1: Protein-ligand complex structure metrics relevant to the benchmark. The pocket size, ligand size and pocket/ligand ratio are adapted from *ref1*.

<i>pdb code</i>	<i>Protons anchors</i>	<i>Distance median</i>	<i>Pocket size A³</i>	<i>Ligand size A³</i>	<i>Pocket/ligand</i>
<i>2f6x</i>	7	6.1	385	110	3.5
<i>2bkx</i>	10	5.3	200	200	1.0
<i>2cix</i>	6	4.8	135	75	1.8
<i>1ynh</i>	6	5.3	295	130	2.3

<i>1wog</i>	8	6.0	140	140	1.0
<i>1k0e</i>	7	4.4	340	150	2.3
<i>1m2x</i>	11	8.0	375	155	2.4
<i>2ff2</i>	5	4.7	250	195	1.3
<i>2i5x</i>	7	4.9	160	140	1.1
<i>1h46</i>	5	6.1	770	210	3.7
<i>2iba</i>	3	6.0	580	90	6.4
<i>1tku</i>	5	3.8	145	150	1.0
<i>1m3u</i>	3	4.1	215	105	2.0
<i>1mlw</i>	6	4.2	250	165	1.5
<i>3c0z</i>	9	7.2	300	90	3.3
<i>1r5y</i>	8	4.6	265	120	2.2
<i>1y2k</i>	7	4.1	240	220	1.1
<i>2aie</i>	8	4.2	285	165	1.7
<i>2q6m</i>	6	5.2	310	225	1.4
<i>2zvj</i>	6	4.9	140	195	0.7
<i>2uy5</i>	9	6.0	235	170	1.4
<i>2j5s</i>	11	4.1	210	115	1.8
<i>2v77</i>	10	4.1	295	155	1.9
<i>1f5f</i>	20	4.2	350	235	1.5
<i>2gg7</i>	7	7.5	350	165	2.1
<i>2rdr</i>	3	5.6	505	90	5.6

<i>lofz_b</i>	8	4.5	490	105	4.7
<i>lofz_a</i>	6	5.1	490	105	4.7
<i>lpwm</i>	9	5.2	350	190	1.8
<i>2gvv</i>	11	3.9	440	165	2.7
<i>2hdq_a</i>	3	6.5	460	90	5.1
<i>2plo</i>	9	6.3	480	115	4.2
<i>luwc</i>	7	4.4	240	150	1.6
<i>lyv5</i>	6	5.8	240	185	1.3
<i>le2i_b</i>	9	4.1	290	145	2.0
<i>lx07</i>	3	5.1	280	150	1.9
<i>2fgq</i>	4	4.0	110	85	1.3
<i>ls5n</i>	5	4.8	450	100	4.5
<i>2brt</i>	8	4.8	190	195	1.0

Effect of the distance restraints number on NMR² protein-fragment complex structure calculations

Given the low molecular weight of the fragments (MW < 300 Da), such molecules usually contain a reduced number of non-exchangeable protons, i.e. bound to a carbon, as compared to lead or drug candidate molecules. These protons are involved in the NOE that yield to the cross-peaks from which are obtained the distance restraints. In consequence, protein-fragment complexes are characterized by less intermolecular distance restraints, and therefore can hinder the algorithm convergence. Calculations with fewer distance restraints are expected to yield to lower TFC since fewer violations are possible. This decrease in the number of distance restraints is also thought to increase the degree of freedom of the possible ligand pose, which would also give rise to smaller TFC. Nonetheless, this was not observed among the different calculation benchmark structure calculations (Figure 3.4A), which did not evidence a clear dependency of the TFC on the distance restraints number. In order to characterize further the effect of the number of the distance restraints onto the algorithm convergence, the boxplot of the distance restraint number distribution for successful and failed runs have been plotted (Figure 3.4B). The boxplots in the Figure 3.4B, show no difference of number of restraints between the successful and the failed structure calculations. To summarize, despite the well-known relation between the low number of restraints and the failure of NMR structure calculation, no direct correlation between the number of restraints and the TFC or the calculation success could be drawn. While this does not exclude the effect of the

number of the distance restraints on the low success rate of the NMR² structure calculations, this suggests that other parameters are at play, smoothing the sole influence of the number of restraints.

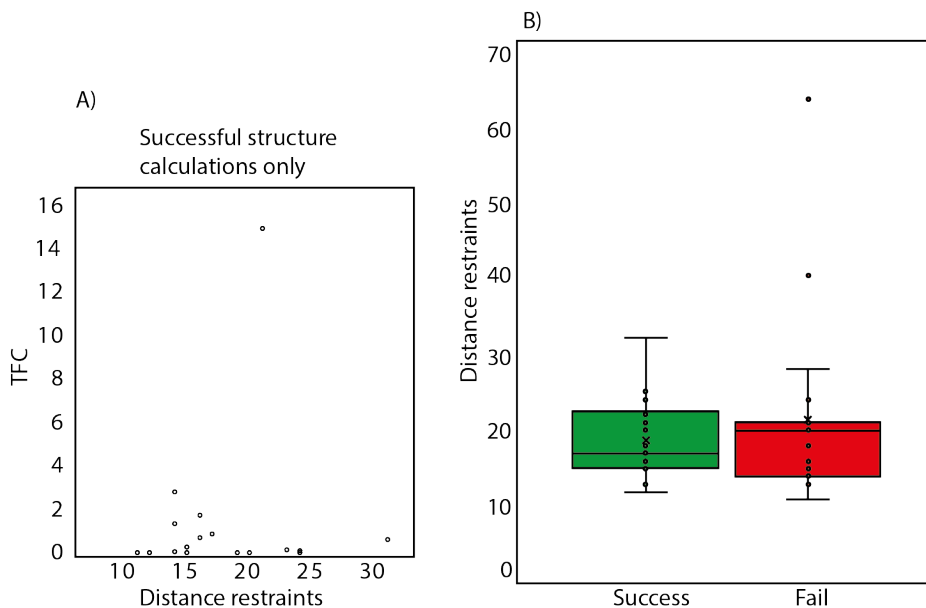


Figure 3.4: A) distance dependency of the TFC for the structures that could be successfully calculated. B) Boxplot distribution of the distance restraints number for each structure that could be successfully calculated (green) or failed to converge (red). For both plots the calculations were performed with the introduction of 20% error to the distance restraints (see below).

Effect of the distance restraints error on NMR² protein-fragment complex structure calculations

In a first instance, no error was applied to the distance restraints, in order to estimate how efficient is the algorithm when exact distances are provided. This first round of calculation provided 35 correct ligand poses with a TFC > 0.2 out of 37, corresponding to 95% of success. The initial dataset was composed by 39 structures, and two structures were discarded for the reasons detailed below. The average heavy atoms RMSD to the X-ray reference structure was of 0.88 Å which is significantly below the consensual value of 2 Å used to admit a correct ligand pose in modelling.¹³⁹⁻¹⁴⁰ Moreover, the two complex structures which presented a TFC < 0.2 and were therefore considered as failed (PDB codes: 2fgq and 2p1o), still ranked the right ligand pose as the best of ten. The TFC of the two structures could be significantly increased by introducing partial assignments (PA) to the NMR² calculations This include two methyls that belong to the same residue, and the identification of alanine, threonine, and methionine methyls thanks to their unambiguous patterns.^{141,142} The introduction of PA is discussed later in this chapter. With the PA the TFCs of 2fgq and 2p1o increase from 0.13 to 1.31 and from 0.16 to 3.4, respectively. The two discarded structures (PDB codes: 2brt and 1s5n) could not find the right pose for the ligand.

The detailed examination of the binding site for these two structures revealed particular binding site topologies, which appear to be incompatible with NMR^2 . Indeed, each binding site presents a topology incompatible with the symmetry breaking necessary for an unambiguous structure determination. The 1s5n complex binding site carries the three methyl containing residues of its binding site at the same side of the binding pocket (Figure 3.5A). The 2brt complex contains only two residues that carry methyls (figure 3.5B). The consequence of these two topologies is a set of unresolved ligand poses, and therefore, 2brt and 1s5n were excluded from the rest of the NMR^2 calculations. In conclusion, unless the binding site carries at least three methyl containing residues homogeneously distributed among the binding pocket, careful examination has to be done prior to judge if it is suitable for NMR^2 calculation. This was already suggested in the first publication regarding the method, where the analogy with the global positioning system (GPS) is made.⁴³

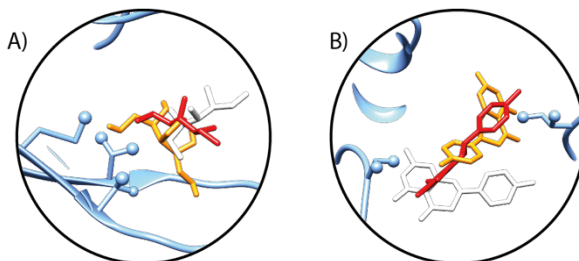


Figure 3.5: ligand binding pose produced from NMR^2 calculations. In white the reference X-ray structure (pdb: 2xp6), in orange and red the two best ranked poses. A) 1s5n B) 2brt

In a second stage, white noise, with a standard deviation of 10% and 20%, was applied to the distance restraints to mimic the error on the NMR^2 input distance restraints. Experimentally, distance restraints are directly obtained from the normalized NOE cross-peak intensities build-up curves.⁸⁵ The distance restraints error can cause the degeneracy of methyl assignments combinations, by increasing the adaptability of the distance restraints network to different assignments options. Hence, the introduction of error has an expected deleterious effect on the efficacy of the calculations, which is observed by the TFC collapse upon error introduction in the figure 3.6A. The consequence of this general TFC value decrease is observed in the figure 3.6B where the success rate drops from 95%, when no error is applied, to 46%, when +/-20% error is applied. This is explained as follows: when the error applied to the distance restraints increases, the TFC of each complex drop, and in some cases below the 0.2 threshold value, hindering the unambiguous recognition of the right poses (Table 3.2). Hence, the TFC decline is what causes the decrease in the success rate, this can be assessed in more details in the Table 3.2. Interestingly, the introduction of PA, rose the individual TFC values and consequently the success rate (Figure 3.6A and B). This is readily explained by looking at how the PA operates (see next section). By fixing some of the assignment possibilities, the number of degrees of freedom decreases and the TF penalty in the case of a wrong pose increases. The Table 3.2 reports the increase of the TFC upon introduction of PA into the calculation. Therefore, the use of PA is highly recommended for the NMR^2 structure calculations of protein-fragment complexes.

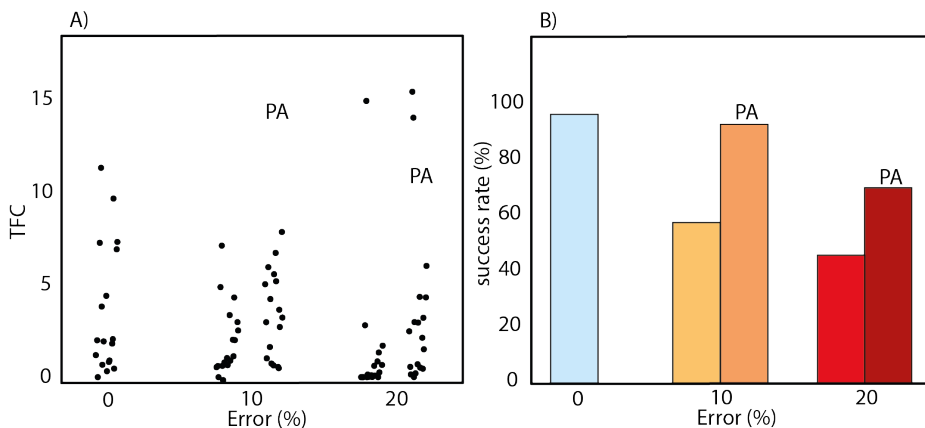


Figure 3.6: A) TFC evolution with increasing error. B) Success rate at different levels of error. PA stands for structure calculation performed with PA.

Alternatively, the effect of the distance restraints error on the heavy atoms RMSD has been evaluated. The distribution of the individual RMSD values of the structures that were successfully calculated at each level of error, is plotted in the figure 3.4B and do not draw any clear tendency driven by the error. Surprisingly, this distribution at 10% error is less scattered than at 0% and 20% error, and has an average RMSD of 0.74 ± 0.29 Å versus 0.87 ± 0.43 Å and 0.97 ± 0.32 Å at 0% error and 20% error, respectively. This apparently lower value of the RMSD has been assessed with a Student statistical t-test for appaired populations, and interpreted at risk $\alpha = 0.05$. The RMSD distributions for the calculations run at 0% and 10% showed no significant difference (p-value = 0.2032). The RMSD distributions of the calculations at 10% and 20% showed significant difference (p-value = 0.0063) suggesting an increase of the RMSD with higher amount of noise (Table 3.2). However, no significant difference was shown while comparing the RMSD at 0% and 20% of error (p-value = 0.4381). The Student t-test have been conducted to probe the difference between the RMSD distributions. The only distributions exhibiting statistically significant difference at risk $\alpha = 0.05$ were 10% error and 20% error (10-20%). This is explained by the relatively small standard deviation of the differences between the RMSD values. Nevertheless, regardless of the amount of error applied to the distance restraints, the average value of the RMSD (Table 3.2) is significantly below 2 Å which is the value admitted by consensus in the modelling community to accept a pose as correct. Moreover, the introduction of the PA to the calculation at 10% and 20% error did not bring major changes in the individual RMSD, and hence in the average RMSD, meaning that the PA have an influence on the TFC and the scoring of the output structures but not on the structure calculation itself. Furthermore, the complex structures calculated with or without PA were overlaid for visual evaluation, and revealed almost identical ligand poses.

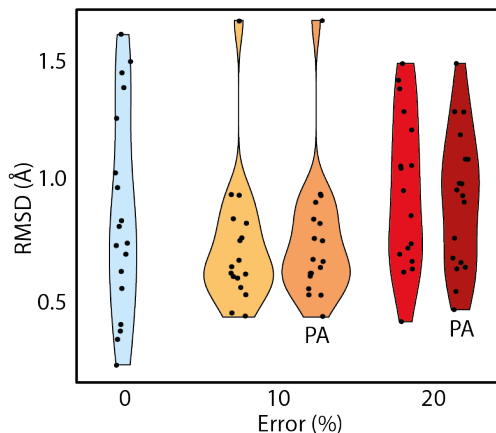


Figure 3.7: violin plot of the RMSD distribution at different levels of error, with or without the introduction of PA.

Table 3.2: Benchmark overview with different conditions. The scores are given after exclusion of the pdb structures 1s5n and 2brt. * Values calculated among the population of structures that converges for all conditions.

error	Success	Fail	RMSD (Å)	TFC	Success: r_l / r_w	Fail: r_l / r_w
0%	95%	5%	0.88+/- 0.42 0.87+/- 0.43*	2.56+/-3.00 3.68+/- 3.50*	1.55+/- 0.58 0.91+/- 0.32	0.98+/- 0.33 0.84+/- 0.16
10%	57%	19%	1.05+/- 0.53 0.74+/- 0.29*	1.58+/-1.99 2.16+/- 1.93*	1.63+/- 0.58 0.95+/- 0.33	1.16+/- 0.41 0.78+/- 0.21
20%	46%	24%	0.96+/- 0.27 0.97+/- 0.32*	0.98+/-2.85 1.50+/- 3.48*	1.79+/- 0.44 1.00+/- 0.26	1.25+/- 0.58 0.82+/- 0.33
10% + PA	92%	5%	0.85+/- 0.34	3.20+/-3.80 4.51+/- 4.50*	1.57+/- 0.56	0.88+/- 0.29

			0.76+/- 0.26*		0.94+/- 0.28	0.56+/- 0.50
20% + PA	70%	15%	1.09+/- 0.51	2.29+/-3.73 3.63+/-4.45*	1.74+/- 0.49	0.99+/- 0.39
			0.93+/- 0.29*		0.99+/- 0.26	0.72+/- 0.34

Despite the introduction of increasing levels of error to the distance restraints, the quality of the output structures was preserved. However, the identification of the right ligand poses among the wrong ones, using the TFC, have been dramatically hindered. This is summarized in the Table 3.2 which reports the results for the different structure calculation conditions.

Table 3.3: Student *t*-test results for the RMSD distribution at different level of error.

Comparison	<i>n</i>	<i>t</i>	<i>p</i> -value	IC 95%	Diff mean	Diff SD
0/10%	17	1.3268	0.2032	[-0.0762,0.3313]	0.31	0.26
10/20%	17	-3.1404	0.0063	[-0.0734,0.3784]	0.29	0.23
0/20%	17	-0.7952	0.4381	[-0.3607, 0.1639]	0.39	0.33

Partial assignment

While NMR² by-passes the protein signal assignment step, a high number of methyls or certain binding site topologies can lengthen the calculation time. In such cases, the algorithm convergence takes up to several days. An obvious solution to speed up the calculation is to assign some of the methyls to reduce the number of methyls to assign, and to decrease the number of degrees of freedom of the ligand pose. This considered, such assignments need to be provided in alignment with the NMR² timeline. For this purpose, an accessory toolbox of two state-of-the-art NMR experiments, which can be measured within a day and analyzed in a couple of hours, was implemented. The constant time (ct) ¹³C-HSQC can be recorded within an hour and takes advantage of the carbon-carbon scalar coupling, $J(^{13}\text{C}-^{13}\text{C})$, which ends up modulating the sign of the ¹³C-¹H signal in the spectrum.^{25, 121} As the methyl of the methionine residues is attached to a sulfur the $J(^{13}\text{C}-^{13}\text{C})$ is absent, and the methionine methyls are of opposite sign as compared to the other amino-acid's methyl (Figure 3.8). This sign alternation is observed with a constant-time of 13.3 ms to record the indirect dimension, as $J(^{13}\text{C}-^{13}\text{C}) = 30\text{-}35$ Hz.^{25, 142} The second experiment is the HCCH-TOCSY, and it can conveniently be measured overnight. In this tridimensional NMR spectrum, the protons of the amino-acid side chains are correlated together after a spin-lock transferring the magnetization through the $J(^{13}\text{C}-^{13}\text{C})$.¹⁴¹ The spectrum is analyzed by going through the carbon planes and identifying in the diagonal the proton signals involved in the NOESY cross-peaks. The HCCH-TOCSY patterns of the different amino-acids are specific enough to unambiguously identify many of them (Figure 3.8).

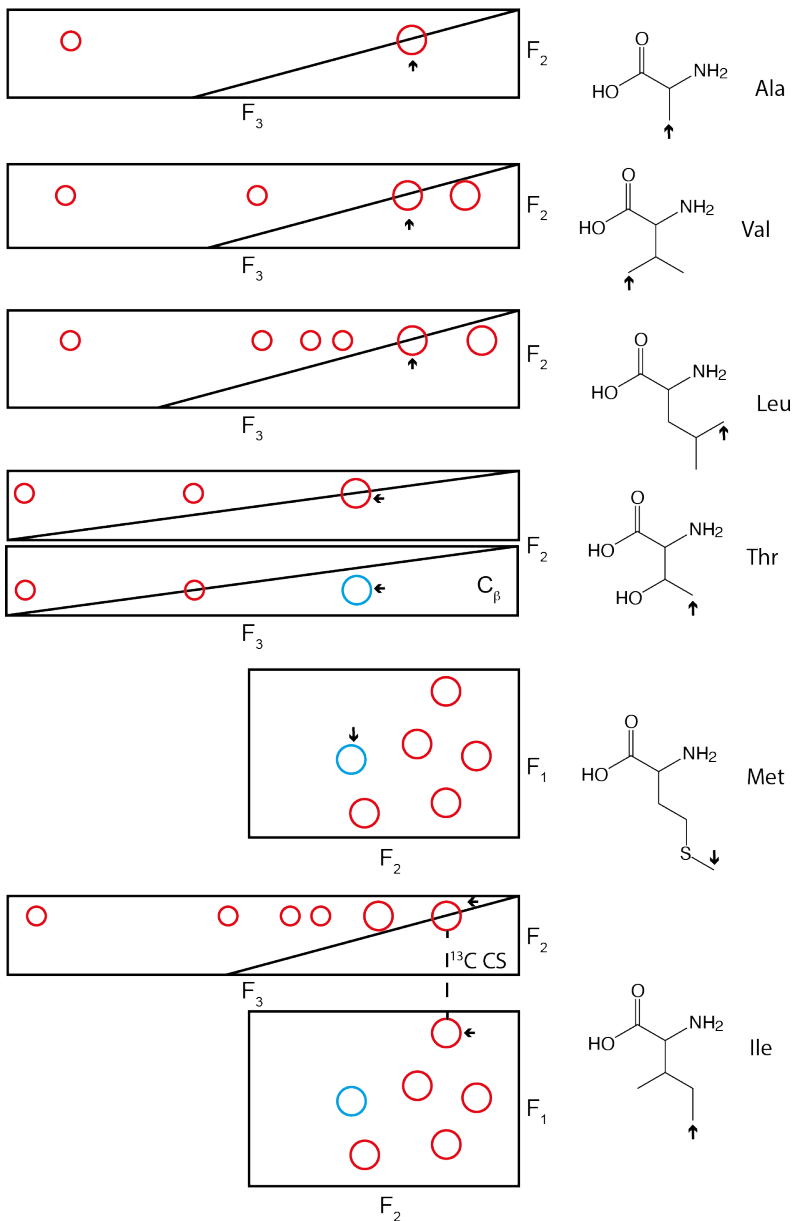


Figure 3.8: HCCH-TOCSY (F_3 , F_2) and ^{13}C -ctHSQC (F_2 , F_1) specific patterns for the different amino-acids. F_1 , F_2 , F_3 are the frequencies of carbon, proton, and proton nuclear spins in ppm, respectively.

For instance, alanine shows only a strong correlation between the methyl and the H_{α} . Valine, is in theory identified because of the presence of only one β -proton as compared to isoleucine or leucine. However, in practice the distinction between leucine and valine is not trivial, especially in the case where considerable signal overlap is present. However, it is still possible to link the two methyls belonging to a leucine or a valine together in the NMR^2 input (see below). Isoleucine can be differentiated from valine and leucine because of its δ -methyl which is upfield in the carbon dimension, and is readily observed in the ^{13}C -ctHSQC. Finally, threonine methyls can be recognized by observing the methyl sign in the C_{β} -plane where the H_{β} - H_{γ} cross-peak is of 180° phase (i.e., negative sign). Altogether, these patterns are used to provide PA inputs to the NMR^2 calculations, that can be included in the different sections of the input file.

```
>> [Assigned]
>>  $M_n = Res$  Type
>> [Filter]
>>  $same\_res = M_i M_j$ 
>>  $M_n = AA$ 
```

Where i , j , and n are the methyls indexes, Res designs the residue number in the sequence, Type is the type of methyl (e.g.: QG2 for a γ_2 -methyl), and AA is the amino-acid type (e.g.: MET for a methionine). The subsection [Assigned] accepts unambiguous assignments. The subsection [Filter] is used to include low level assignments such as $same_res$, indicating that two methyls belong to the same residue, or $M_n = AA$ indicating that the methyl is carried by a residue which is a defined amino-acid type.

The PA introduced in the benchmark are always the ones that could be obtained from the analysis of an HCCH-TOCSY and a ^{13}C -ctHSQC.¹⁴³

Effect of the ligand topology on NMR^2 protein-fragment complex structure calculations

If the increasing level of error is readily identified as a deleterious parameter for the structure calculations, it still needs to be defined what is the cause for the low success rate of structure identification in the case of fragments. Different parameters were suspected to be influencing the result of the NMR^2 calculations when the studied molecules are smaller. An obvious consequence of decreasing the ligand size, is the diminishing of the protons number and therefore the diminishing of the amount of distance restraints. Additionally, topology parameters were introduced in order to address the characteristic geometry of the ligand in the binding pocket considering its length, width, and the intermolecular distance restraints average. Here, the length is defined as the longest distance vector that can be drawn between two protons belonging to the ligand, and the width is the longest vector orthogonal to the length vector that can be drawn between two protons belonging to the ligand. The mean distance needs to be considered since longer distance restraints will contain more error as compared to shorter distance restraints. Therefore, the uncertainty is directly linked to the length of the intermolecular distance restraints. The topology parameters are describing the ratio of the characteristic internal distance of the ligand onto the average distance restraints length:

$$r_l = \frac{T_l}{\langle d \rangle} \quad (3.1)$$

$$r_w = \frac{T_w}{\langle d \rangle} \quad (3.2)$$

With T_l the maximal distance between two protons within the ligand, or the length; T_w the maximal proton-proton distance orthogonal to the length vector, within the ligand. $\langle d \rangle$ is the average intermolecular distance restraints value for the complex. The parameters r_l and r_w provide a rudimentary but efficient insight into the binding site topology. The figure 3.9 provides visual representation of the ligand length, T_l , and the ligand's width, T_w .

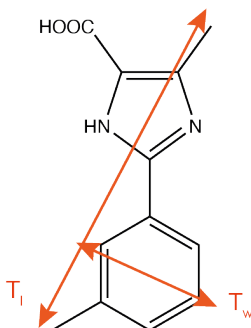


Figure 3.9: schematic representation of the topology parameters T_l and T_w . The length T_l is defined first, and the width T_w is taken as the longest orthogonal proton-proton distance. The exchangeable protons are not considered as they most likely do not yield to observable NOE cross-peaks.

When plotting the successful and unsuccessful calculations according to their amount of distance against the sum of the topology parameters (equation 3.1 and 3.2), only the latter axis dimension was showing a discrimination (Figure 3.10). Figure 3.11 shows in more details the effect of distance restraints number and ligand topology on the TFC, for two structures arbitrarily chosen (pdb codes 1h46 and 2ff2). In order to keep an easily interpretable dimensionality, the topology parameters are aligned on the same axis using the sum of these two parameters: $r_l + r_w$. As observed in Figure 3.11A, when the distance restraints number decreases without affecting the ligand topology, it does not perturb the TFC of the structure calculations of 1h46. In the case of 2ff2, the TFC decreases but stays at a minimum value of 1.7, which is well above the 0.2 threshold used for TFC based structure discrimination (Figure 3.11D). However, when the same amount of distance restraints is removed and the sum of the ligand topology parameters decreases, as it is illustrated in the Figures 3.11C and 3.11F, the TFC drops dramatically. The TFC of 1h46 reaches 0.7 (Figure 3.11B) and the TFC of 2ff2 reaches 0.4 (Figure 3.11E). This suggests that the ligand topology is more critical to the pose discrimination based on the TFC than is the distance restraints number.

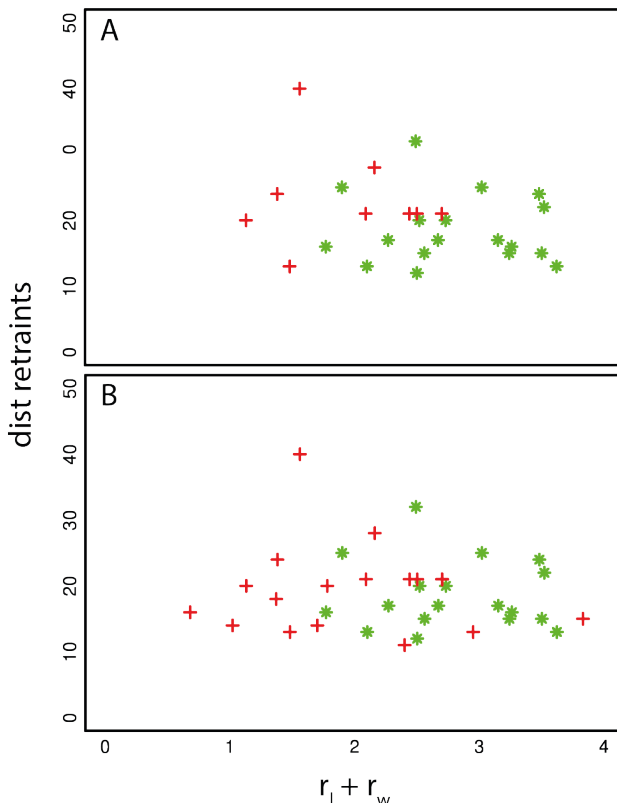


Figure 3.10: Scattered plot of the successful and failed structure calculation along the number of distance restraints and the sum of the topology parameters r_l and r_w . A) 10% error. B) 20% error.

Since the topology parameters exhibit discrimination between the positive and negative results (Figure 3.10), and their influence on the TFC of the structure calculations has been demonstrated (Figure 3.11), their effect has been investigated further. The topology parameters were used to assess the relationship between the fragment – binding site topology, and the success of the calculation at different levels of error. The Figure 3.12 draws the scatter plot of the r_l and r_w parameters at different levels of error with or without PA. For the structures calculated with an error of 10% on the distance restraints, a clear correlation between the high values of topology parameters, r_l or r_w , and the success of the NMR^2 calculations is observed. This correlation is emphasized when PA are introduced. The scatter plot of the calculations at 20% of error showed a trend in the r_l dimension to cluster the successful and unsuccessful structure calculations. However, the introduction of the PA to the calculations, enabled to distinguish better the trend for success when the topology parameters are high. Moreover, the average topology values for successful jobs reported in the Table 3.2 is higher with increasing level of distance restraints error. These values drop again after the introduction of PA in the calculations. These data suggest that the error on the distance restraints is less critical for the determination of complex structures with high ligand topology parameters, r_l and r_w .

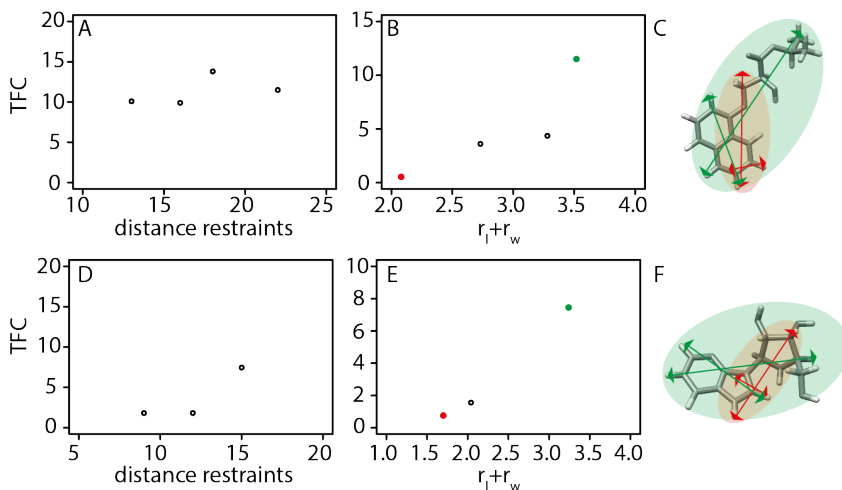


Figure 3.11: Comparison of the influence of different parameters on the TFC for two different structures, 1h46 (A, B, C) and 2ff2 (D, E, F). The first plots are the dependency of the TFC on the number of distance restraints (A, D), and the second the dependency on the topology parameters (B, E). The red and green dots in B and E, correspond to the topologies represented in C and F, respectively.

Furthermore, the use of PA improves the tolerance of complex structures with lower r_l and r_w to this error. It is important to remember that each point corresponds to a different structure with a unique binding site topology. In the case of a same binding site, as it might be in a FBDD campaign, we expect this analysis to yield to a clearer clustering of successful and unsuccessful calculations. Furthermore, in the case of a set of different fragments for which the complex structures with a common target binding site would be engaged in a calculation process, the average intermolecular distance length is expected to be smaller for the bigger fragments as compared to the smaller fragments. Hence, it is possible in a first instance, to evaluate which fragments have the best calculation properties thanks to their intrinsic geometrical parameters, L_{length} and L_{width} .

Calculation time

In the context of an FBDD campaign, the number of fragment hits for which the structure of their complex with the target is sought, can reach several dozens. Therefore, the timespan for the protein-fragment complex calculation is a critical feature of the NMR² methodology. The calculations were performed on 10 CPU, this number was maintained low due to the high amount of different structures at different conditions (see material and methods). However, the calculation time varied from a few minutes to a maximum of 96 hours with a median calculation time of 11 minutes, and was considered satisfactory with regards to the analysis time in classic NOE-based structure calculation which can last several months. No relationship could be established between the calculation time and the amount of methyls or restraints. However, the use of PA significantly decreased the calculation time with a maximum calculation time of 38 hours with a median calculation time of 10 minutes.

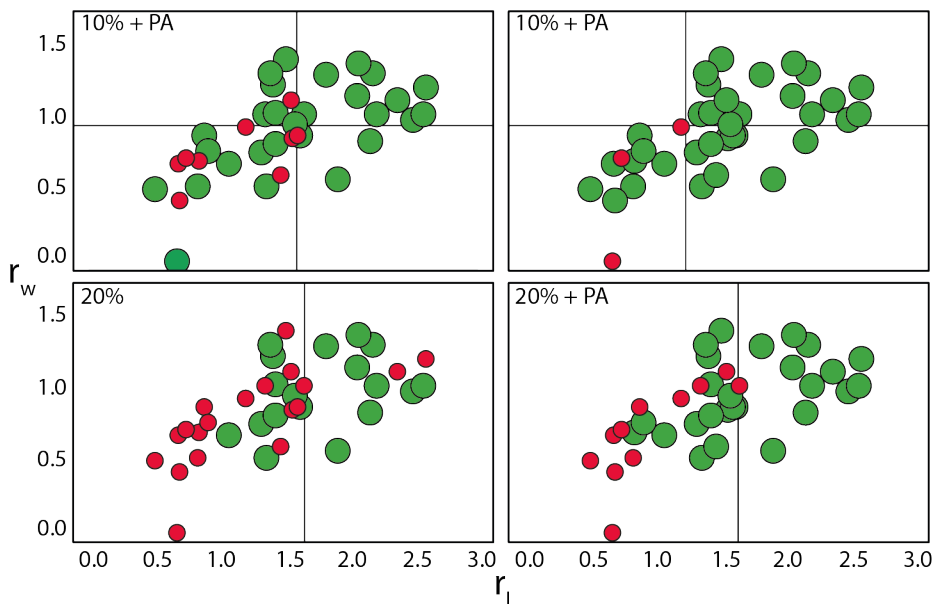


Figure 3.12: Scatter plot of the topology parameters of the fragments constituting the benchmark. In green the successful structure calculations, in red the failed ones, and different level of errors, with or without PA.

NMR^2 reduced the analysis time to obtain a structure to a few hours as compared to several months before. While other methodologies have been developed,^{47-49, 53} none so far is exclusively based on sole experimental restraints. Therefore, NMR^2 presents the advantage of derisking the medicinal chemistry rational design by providing experimentally supported structures and not models. As the analysis is highly automated, and the TF-based structure selection and validation is well established, NMR^2 is a promising method that could be applied in routine. Finally, the structure analysis time is competitive with the analysis time offered by X-ray crystallography.

Application to PIN1-fragment complex structures calculation

Fragments' overview

In comparison to the protein-ligand complex structures calculated in the previous chapter and in prior studies,⁴³⁻⁴⁴ fragments contain less protons and provide less of intermolecular distance restraints. For drug-like ligands, on average 19.6 intermolecular distance restraints⁴³⁻⁴⁴ were used to calculate the complex structures while only 12.3 in this case, for fragments. As demonstrated earlier in this thesis, the reduction in the number of restraints associated with the high experimental error (20%) is responsible for challenging algorithm convergence. Moreover, the fragments have small topologies, and they are therefore more difficult to converge than ligands

that would have the same amount of restraints but bigger topologies. The Table 3.4 summarizes some of the critical properties of the fragments, and provides as a comparison the values expected for drug candidate sized molecules.

Table 3.4: Summary on the fragments size, topology and number of intermolecular distance restraints engaged. A comparison with the average values for drug like ligand is shown, the values are the average from the previous NMR² studies.

fragment	1	2	3	4	<drug like>
MW (g/mol)	216	236	270	211	500-800
Distance restraints	14	11	12	13	20-35
r_l	2.3	1.7	2.0	2.0	3.9
r_w	1.2	1.0	1.2	1.4	3.0

The application of NMR² for fast protein-fragment complex structure calculation is evaluated through the case study of three phenylimidazole isosteres previously identified to bound PIN1. The phenylimidazole ligands considered in a first instance, compounds 1, 2 and 3, share a common phenylimidazole scaffold and exhibit structural differences located at the phenyl-substituent where the isosteres CH₃ (radius ~ 2 Å), CF₃ (radius ~ 2.5 Å), and Cl (radius ~ 1.8 Å) are present. All three are fragment-sized with MW₁ = 216 Da, MW₂ = 236 Da, MW₃ = 270 Da, resulting in low number of protons that can be involved in intermolecular distance restraints. Indeed, fragments 2 and 3 carry four aromatic protons on the phenyl moiety and a methyl on the imidazole ring, and fragment 1 has a supplementary methyl on the aromatic ring. These features justify the use of these three isosteres to showcase the use of NMR² for protein-fragment structure calculation. The fragment 4 is a benzindole flanked with a carboxylate, and deviates in terms of structure from the fragments 1-3. Moreover, it bears only aromatic protons while 1-3 have at least one methyl.

The fragment conformations were found to be planar in agreement with the intra-NOE derived distances measured in their bound states (see experimental section), as well as ab-initio calculations using the software Gaussian.

Calculation of the effective correlation time

The initial rate σ of the normalized NOE build-up curves from which are calculated the distance restraints is defined according to equation 3.3:

$$\sigma_{ij} = \frac{1}{2} \frac{\mu_0}{4\pi} \frac{\hbar \gamma_H^2}{r_{ij}^6} (6J(2\omega) - J(0)) \quad (3.3)$$

And the spectral density is defined as in equation 3.4:

$$J(\omega) = \frac{2}{5} \left(\frac{\tau_c}{1+(\omega\tau_c)^2} \right) \quad (3.4)$$

Where μ_0 is the permeability of the vacuum, γ_H the proton gyromagnetic ratio, \hbar the reduced Planck constant, r_{ij} the distance between the proton i and j , ω the spectral frequency, and τ_c the correlation time. While the initial rate σ is obtained from the Fullmatrix⁸⁶ relaxation treatment of the NOE intensities, the correlation time τ_c is still to be determined in order to calculate the distance restraints. The rotational correlation time of the system is governed by the protein as it is tumbling much slower than the ligand so that $\tau_{c,protein} \gg \tau_{c,ligand}$. The τ_c can be calculated from the spin-lattice (R_1) and the spin-spin (R_2) relaxation rates. Indeed, the ratio R_2/R_1 for the most rigid amide vectors of the protein is independent from the internal motions and the order parameters, and it is uniquely dependent on the rotational correlation time.¹⁴⁴ The τ_c is then fitted to the R_2/R_1 using the software Tensor2,¹⁴⁵ which takes for input the tridimensional coordinates of the protein and the relaxation rates of the backbone amide vectors. The amide vectors relaxation rates were measured using the ¹⁵N- R_1 -TROSY, and ¹⁵N- $R_{1\rho}$ -TROSY, modified for non-perdeuterated proteins.¹⁴⁶ The R_2 relaxation rate was obtained using the relationship:

$$R_2 = R_{1\rho}/\sin^2\theta - R_1/\tan^2\theta \quad (3.5)$$

Where $\tan\theta = \omega_1/\Omega$ and ω_1 is the spin-lock frequency and Ω the offset of the carrier in ¹⁵N dimension. Recalculating R_2 from the $R_{1\rho}$ and R_1 , has the advantage that the exchange contributions are not present as it is the case in the R_2 rates measured with ¹⁵N-CPMG-TROSY. Finally, the τ_c of PIN1 at 298 K was obtained and equal to 8.9ns. The rotational correlation times of the fragments were assumed to be below 0.1ns, which seems to be a safe estimation as of the small size of the molecules.

Fragments are usually binding weakly to their target, i.e., $> 100 \mu\text{M}$. Therefore, the fragment population is distributed among free and bound forms, with the correlation times of the ligand and of the protein, respectively. The titration by ¹³C-HSQC NMR confirmed the fast exchange between the bound and free populations. Therefore, it is necessary to assess the effective correlation times, $\tau_{c,eff}$, defined from the population averaged correlation times of the free and bound fragments (equation 3.6):

$$\tau_{c,eff} = \frac{PL}{P} \tau_c \quad (3.6)$$

The binding affinity of the fragments 2 and 3 were derived using protein methyl group chemical shift perturbations upon fragment titration (Figure 3.13). In the case of fast exchange, the CSP is the expression of the PL/P ratio, with PL the protein-ligand complex and P the total amount of protein. The CSP were calculated according to equation (3.7):

$$CSP = \frac{PL}{P} = \sqrt{\alpha\delta(^{13}\text{C}) + \delta(^1\text{H})} \quad (3.7)$$

With $\alpha=0.1$ an arbitrary scaling parameter, $\delta(^{13}\text{C})$ and $\delta(^1\text{H})$ the chemical shift difference between the apo and holo spectra in the carbon and proton dimension, respectively (Hz). The CSP are then fitted to the equation (3.8) to obtain the dissociation parameters [ref Orts Gossert, Methods, 2018]:

$$[PL] = \frac{[L]_{tot} + [P]_{tot} + K_d - \sqrt{([L]_{tot} + [P]_{tot} + K_d)^2 - 4[L]_{tot} \cdot [P]_{tot}}}{2} \quad (3.8)$$

The dissociation constants of the fragments 2 and 3 were found by least square fit with an in-house python script: $K_{d2} = 760 \mu\text{M}$, and $K_{d3} = 6710 \mu\text{M}$ (Table 3.5).

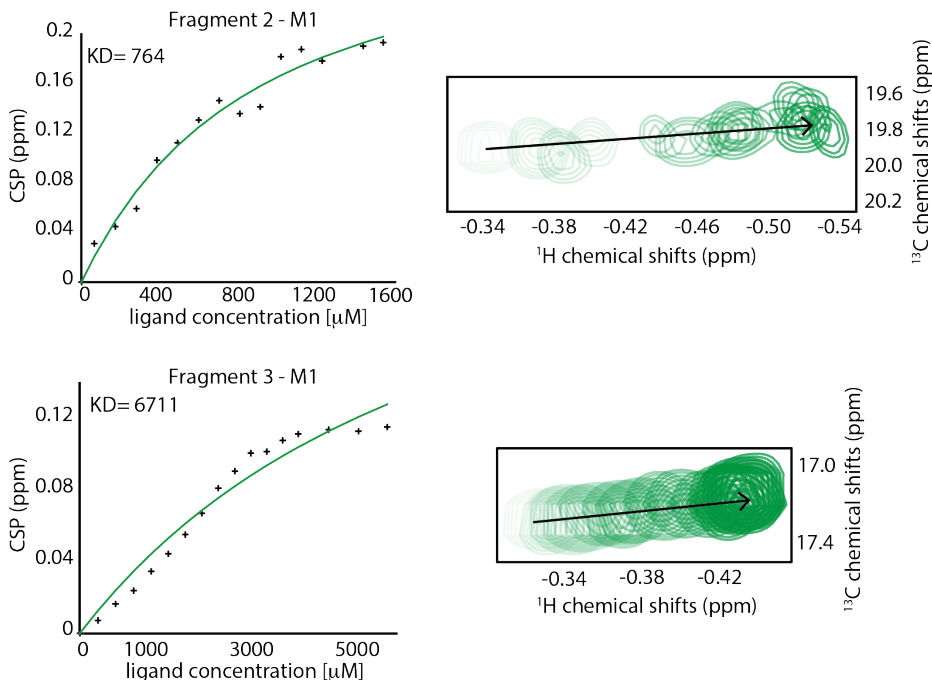


Figure 3.13: titration curves of the fragment 2 and 3 for the methyl M1. The CSP were obtained from the ^{13}C -HSQC spectra.

In the case of the fragment 1 and fragment 4, the $\tau_{c,\text{eff}}$ was directly derived from the sterically known distances from the fragments. In these cases, the distances that cannot be affected by conformational changes, e.g. the distance between the protons of the phenyl, or between the methyl and the phenyl protons in fragment 1, can be used to fit the $\tau_{c,\text{eff}}$ value by comparing calculated values to the expected value. All the protons of the benzindole ring are fixed as only one conformation is possible. The distances are measured on the energy minimized structures of the fragments. Finally, the distances are calculated from the NOE using eNORA for different values of $\tau_{c,\text{eff}}$. The difference between the calculated and the measured distances is plotted for the different $\tau_{c,\text{eff}}$, so that the $\tau_{c,\text{eff}}$ corresponding to the minimum difference is kept. It is important to note that only distances calculated with well fitted build-up curves can be considered in this approach, as mediocre quality build-ups can induce an offset in the fitted $\tau_{c,\text{eff}}$ value. Finally, the K_d of fragment 1 and 4 were back-calculated from the $\tau_{c,\text{eff}}$. This approach has the advantage to shortcut the titration experiments, which typically require a day of measurement. However, this could only perform well for fixed distances between the phenyl methyl and the aromatic protons.

Table 3.5: Affinity, K_D , and effective correlation time, τ_c , of the fragment 1, 2, 3, and 4

Fragment	Method	K_D [μ M]	$\tau_{c,eff}$ [ns]
1	Steric distances	260	4.8
2	Titration	760	2.4
3	Titration	6700	1.0
4	Steric distances	1450	2.0

Structure calculation

A series of F_1 -[^{15}N , ^{13}C]-Filtered [^1H , ^1H]-NOESY spectra were recorded on a 900 MHz spectrometer for each PIN1–fragment complex. The cross-relaxation rates derived from the NOE build-up curves were converted into distances using the effective correlation time of the complexes. A total of 17, 14, 18, and 15 intermolecular NOEs, could be measured for the fragments 1, 2, 3, and 4 respectively. Build-up curves of poor quality or showing quadratic behaviors characteristic of spin diffusion, were discarded. A set of 14, 11, 12, and 13 build-up curves for the fragment 1, 2, 3, and 4 respectively, were retained for the NMR² structure calculations.

The NMR² calculations were performed at the known catalytic site of PIN1 containing 6 different methyls (L61, L122, M130, and T152), using the experimentally derived distances restraints, the derived ligand structures and a starting structure of the protein arbitrarily taken from a previously determined crystal structure (PDB: 2xp6). Each side chains dihedral angle of the residues present in the binding site of PIN1 could rotate by 20° providing a large degree of flexibility to accommodate the ligand. However, the calculations output for each of the fragments provided 10 different poses that could not be discriminated by their TF. This situation is corresponding to what was observed in the benchmark, as a consequence of the small molecular size of the fragment and of the important experimental error (20%), the restraints are poorly violated in the case of a wrong pose. Therefore, the right pose cannot be differentiated from the wrong ones on the basis of the TF.

According to the ligand topology parameters (Table 3.4), the fragment 1 has the most favorable topology as it has the highest r_l and r_w values. Therefore, the effort on obtaining the PA was prioritized on the fragment 1, for which ^{13}C -ctHSQC and HCCH-TOCSY were recorded. The PA could be easily determined. The M130 methyl signal was readily derived from the ^{13}C -ctHSQC, taking advantage of the negative signal of the methionine resonance peaks. Furthermore, only one methionine, M130, is present in the binding site. The HCCH-TOCSY of the PIN1-1 complex could connect the methyl signals M1/M2 and M3/M4 as belonging to the same residue,

respectively. The NMR² calculations were performed once more with the obtained PA as additional input.

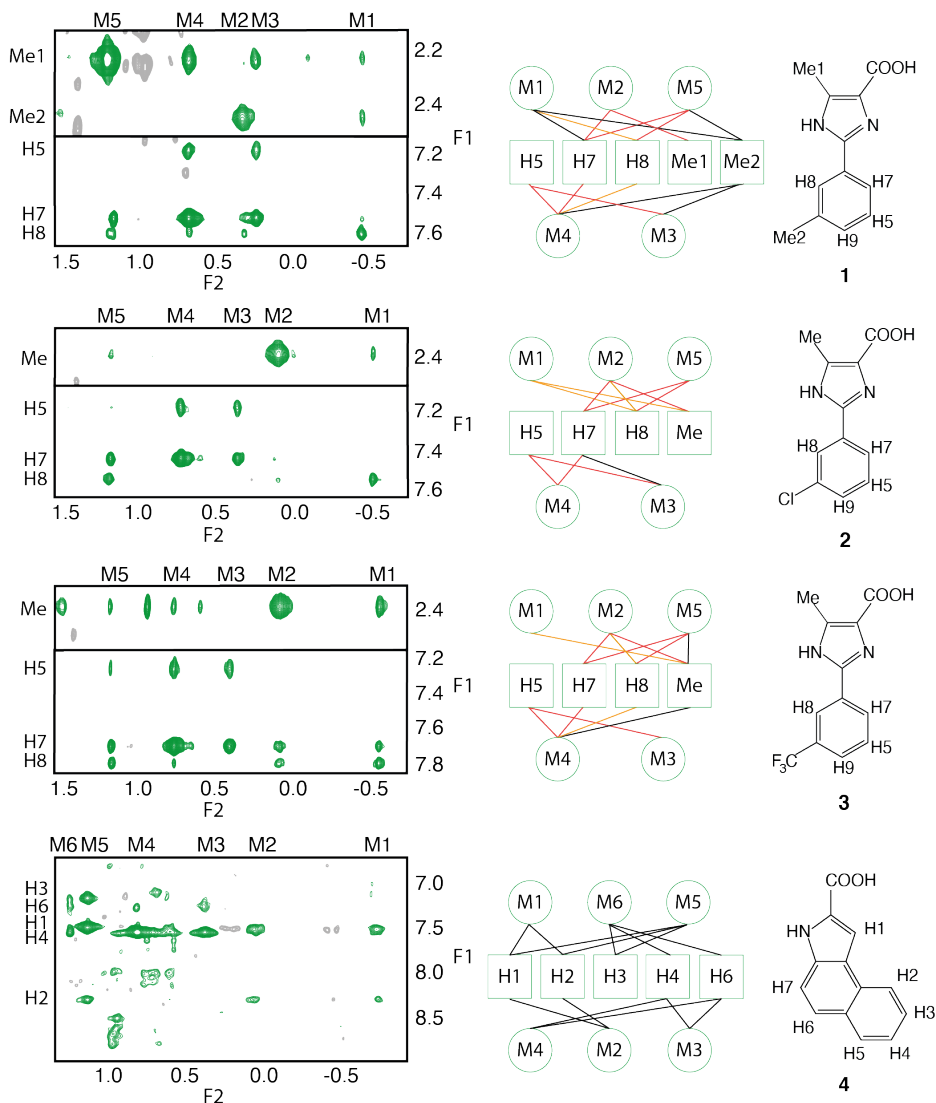


Figure 3.14: F_1 - $[^{13}\text{C}, ^{15}\text{N}]$ -filtered- $[^1\text{H}, ^1\text{H}]$ -NOESY of the different PIN1-fragment complexes. The methyls numbering is consistent from a complex to another. F_1 and F_2 are the protons frequency dimension. The distance restraints networks are pictured with red distances for the restraints present in the three fragment isosteres 1, 2, and 3; the orange restraints correspond to restraints present in at least two of the three fragment isosteres.

The structure calculation of the PIN1-1 complex converged, and the output structure with the lowest TF could be accepted as its TFC with the first wrong pose was equal to 0.44. Later, the alignment of the F₁-[¹⁵N,¹³C]-Filtered [¹H,¹H]-NOESY spectra with the ¹³C-ctHSQC spectrum of the complexes PIN1-2 and PIN1-3 enabled the identification of the NOE restraints involving the M130 methyl group. The NMR² structure determination for the complexes involving compounds 2 and 3, failed to converge, even when defining the methyls M1/M2 and M3/M4 as belonging to the same residues, respectively. The introduction of PA to the calculation of the PIN1-4 complex structure failed as well. Despite the absence of obvious correlation between the number of the restraints and the algorithm convergence, the successful structure calculation of PIN1-1 in comparison to the failed PIN1-2, 3, and 4 complex structure calculations suggests that the poor convergence of the algorithm was due to the paucity of restraints. Indeed 11, 12, and 13 intermolecular distance restraints for the complex involving fragment 2, 3, and 4, respectively, while 14 were measured for the complex involving 1. Moreover, in the light of the fragment topologies, the fragment 1 has higher r₁ and r_w as compared to 2, 3, and 4 (Table 3.4). According to the scattered plots of the Figure 3.12, it is also fragment 1 which has the best chance to be successfully calculated. This observation confirms the validity of the topology parameters for the prioritization of the fragments.

In order to calculate the complex structure impervious to NMR² structure calculation, we found that it was possible to determine structures of the two remaining isostere complexes which did not converge using the methyl assignments from the successfully completed PIN1-1 NMR² calculation. The alignment of the NOE patterns from the F₁-[¹⁵N,¹³C]-Filtered [¹H,¹H]-NOESY spectra enables the transfer of protein methyl assignments from the NMR² derived PIN1-1 complex structure to the distance restraints of the other fragment isosteres (2 and 3) whose complex structures could not be derived. The assignment transfer was in the first place only considered for the fragment 2 and 3 given their close structural proximity with the fragment 1. The transfer of assignments is greatly facilitated by following the chemical shift perturbations of the PIN1 methyl groups during the fragment titration, previously recorded to determine their binding affinities. The newly calculated and converged structures are overlaid with the X-ray crystal structure of PIN1-2 (2xp6) on Figure 3.15, and exhibit similar poses to the crystallographic reference with RMSDs of 1.1 Å, 2.5 Å, and 1.4 Å for the fragments 1, 2, and 3 respectively. The relatively high RMSD for 2 is due to a slight translation of the fragment toward the outside of the binding pocket. The high similarity in the fragment binding modes to PIN1 could be anticipated since a common NOE pattern can be identified for the three fragments suggesting beforehand a similar binding pose (Figure 3.14). Furthermore, the fragments 1, 2 and 3 are structurally very similar and only differ by an isostere substitution, therefore the binding mode is expected to remain similar. In the context of a FBDD campaign, the structure of the fragments binding to the target protein is expected to show diversity. However, the cooperative assignment strategy performs regardless of the fragment scaffolds as the methyl identity can be obtained from the alignment of the NOESY spectra. The structure calculation of the PIN1-fragment 4 complex showcases the versatility of the cooperative assignment strategy. Despite a different molecular scaffold, the same NOE cross-peaks pattern can be observed when comparing the F₁-[¹⁵N,¹³C]-Filtered [¹H,¹H]-NOESY spectra from 4 to the spectra of 1, 2, and 3. The methyl chemical shift perturbations in the ¹³C-ctHSQC spectra are used to confirm the identities of the protein methyls, M1-5. The transfer of the protein methyl assignments to the NMR² calculation of the PIN1-4 complex could solve the ligand pose. It is worth noticing, that in the case the NOESY patterns do not allow simple assignment transfer, the follow up of the CSP in the ¹³C-HSQC spectra of the compared complex, provide a straightforward way to recognize the methyls.

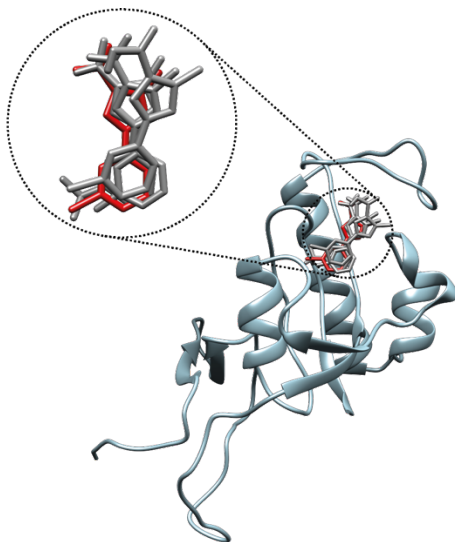


Figure 3.15: Overlay of the PIN1-fragment poses calculated with *NMR2* (grey) onto the X-ray crystal structure of PIN1-2 (red, pdb code: 2xp6).

The understanding of the SAR is a cornerstone of the FBDD. Once the key intermolecular interactions defined, and the structure of the complex is known, the further growing of the fragment, or linking of different fragments binding at different sites, can be rationalized. The SAR and evolution of the phenylimidazole fragments have been already described in the literature. However, the comparison with the structures obtained by *NMR2* is an additional validation of the method. The substituted phenyl group makes hydrophobic interactions with L122, M130, F134, T152, while the embedded histidines, H59 and H157, are candidates for possible π - π or cation- π interactions (not shown). The carboxylic acid is involved in salt bridges with the cationic pocket formed by R68, R69 and K63. The imidazole makes hydrogen bonds to S154 and C113 (not shown), and the trifluoromethyl groups may engage hydrogen bonds to Q131 and T152 (Figure 3.16A, B and C).²¹ In addition, the affinities of the fragment anticorrelate with the size of the substituents, with $K_{D,1}$ (Me) = 260 μ M, $K_{D,2}$ (Cl) = 670 μ M, and $K_{D,3}$ (CF₃) = 6700 μ M, suggesting that steric hindrance may be the primary reason for the decrease of affinity. Overall the three *NMR2* structures exhibit similar binding modes and are consistent with the interactions described previously from the structure derived by X-ray crystallography. The accepted structure of the PIN1-4 complex exhibits common patterns with the poses of the phenylimidazole fragments. The carboxyl group builds salt-bridge interaction with the residues R68 and K63 (Figure 3.16D). The benzindole N-H makes hydrogen bond to S154. The hydrophobic benzindole aromatic ring is inserting deep into the hydrophobic pocket formed by L61, M130, F134 (Figure 3.16D), and the histidines, H59 and H157, which could also make π - π or cation- π interactions (not shown). The affinity of the fragment 4 was found to be $K_{D,4}$ = 1500 μ M by back calculating from the effective correlation time ($\tau_{c,eff}$ = 2ns). The affinity is almost one order of magnitude lower than for the best of the phenylimidazole fragments, 1.

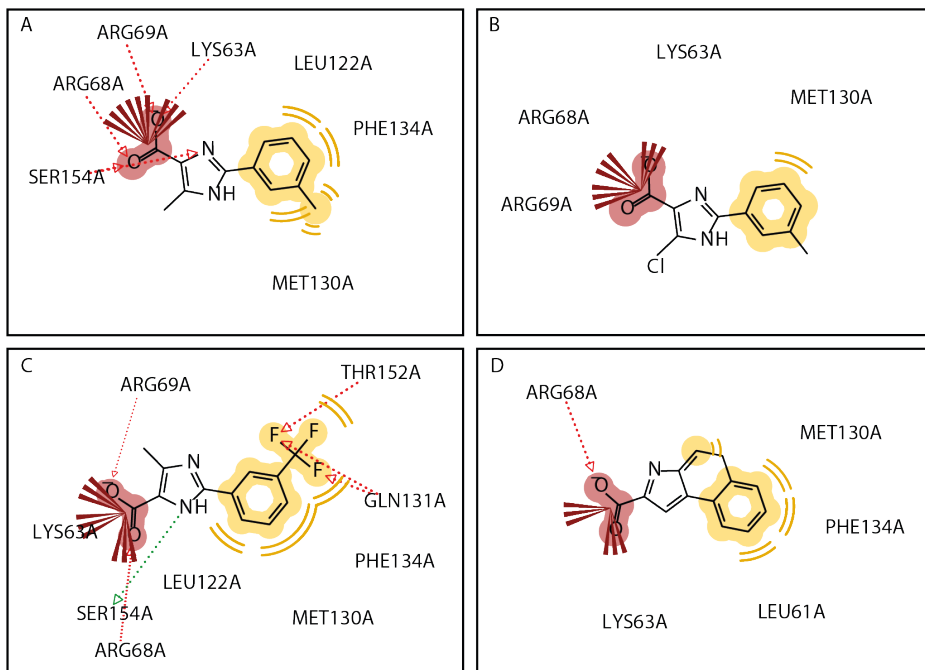


Figure 3.16: Pharmacophores of the different PIN1-fragment complexes, 1 (A), 2 (B), 3 (C), 4 (D). The yellow and red patches correspond to the hydrophobic and ionic interactions, respectively. The red and green arrows are for the acceptor and donor H-bonds, respectively.

Time-to-structure and outlook

In total, the measurement of the NOESY spectra required for each fragment 4 to 6 days of measurement on a 900 MHz spectrometer, the analysis of the data was relatively fast due to the simplicity of the fragment signal to assign and require no more than a couple of hours. The calculation of the PIN1-1 complex structure with the PA took only 3 minutes and the calculation of the PIN1-2,3, and 4 complexes with the assignments obtained by cooperativity were performed in a few seconds. Altogether, the 4 protein-fragment complex structures could be successfully measured and calculated within a month instead of 6-12 months if it would have been done with the classical NMR protocol. This turnaround time is closer to the medicinal chemistry timeline. However, in the context of a FBDD campaign, the number of protein-fragment complex structures to be calculated is expected to be significantly higher than four, i.e. several dozens of compounds. Moreover, the number of scans needed to achieve sufficient signal-to-noise can be doubled as compared to strong binders, depending on the affinity of the fragment and on its solubility, and can reach up to 48h for one mixing time point. Since the analysis time issue is efficiently tackled by the NMR^2 method, the time-to-structure is mostly governed by the measurement time. A first solution would be to use only one mixing time point to reduce the measurement time to maximum 2 days. This solution has been developed in the group, so that even with single mixing time NOE cross-peaks, it can be normalized to obtain high quality distance restraints (unpublished). The other solution would be to measure at a higher field. Indeed,

the same measurement taking 2 days on a 900 MHz spectrometer, would require only half-a-day of measurement on a 1.2 GHz spectrometer. Nevertheless, 1.2 GHz spectrometers at this time only fit 3 mm diameter NMR tube, corresponding to a volume of 160 μL versus 350 μL for the Shigemi tubes used in the present work.

Conclusion

Fragment-based drug discovery is a widely used approach in both pharmaceutical companies and academic laboratories. Fragment-based methods need significantly fewer compounds to be screened, synthesized and the fragment hits show usually high ligand efficiencies. Yet fragments contain usually only few protons, rendering the structure elucidation by NMR more difficult. The presented NMR^2 benchmark demonstrates the possibility of calculating NMR based protein-fragment complex structures. Despite of the quality of the data provided by NMR, it is often considered as a secondary technique in drug discovery due to the long analysis required prior to structure calculation. However, NMR^2 offers a promising alternative thanks to its reduced analysis time, especially in the context of FBDD campaign where several protein-fragment structures need to be defined. Furthermore, because of the analysis is automated and straightforward, the level of expertise required to obtain such a structure drastically reduced. NMR^2 is able to provide a structure in a few days, this includes the filtered NOESY measurement and the structure calculation. The data presented above enable the user to draw a clear understanding of the critical parameters yielding to successful calculations. The analysis of the data helps to establish a clear strategy that should systematically applied for the NMR^2 calculation of protein-fragment structures in the context of a FBDD campaign. The first important parameter is the binding site topology: proteins with less than three residues carrying a methyl group in the binding or proteins with a non-isotropic repartition of these residues in the binding are particularly challenging. A possible way of addressing these systems is to use mutated protein with improved methyl density in the binding site. Furthermore, the error to be applied to experimental NOE is considerable (+/- 20%) and significantly degrades the convergence of the calculations. We want to emphasize on the importance of using pulse sequences such as F_1 - $[^{13}\text{C}, ^{15}\text{N}]$ -Filtered- $[^1\text{H}, ^1\text{H}]$ -NOESY, providing NOE cross-peaks of high quality.⁴² The use of normalized NOE provides high-quality distance restraints⁸⁵ that were successfully used in prior studies. The attention of the reader is brought to the fact that an effort on the NOESY pulse sequence, or on the data analysis in order to reduce the error with a factor of two would yield to significant improvement in the field of NMR FBDD. Moreover, the use of PA significantly improved the calculation convergence, which is expressed by the increasing TFC. These increased TFC values are explained by the restriction of the degrees of freedom within a given fragment – binding pocket topology. The introduction of PA only requires at maximum a day of measurement time. The PA approach was demonstrated by solving the PIN1-1 complex structure. Moreover, the structures of the different complexes PIN1-2, PIN1-3, and PIN1-4 were solved taking advantage of the cooperativity between several NMR^2 datasets. It was demonstrated that the cooperative assignment strategy can overcome the experimental ambiguity due the lack of distance restraints that would otherwise prevent structure elucidation and consequently the corresponding structure-based drug design. Finally, the use of the protein aromatic signals is an additional way to improve the TFC and to discriminate between the right or wrong poses. The NMR^2 structure calculations only took a couple of minutes of computer time and only a few hours were needed to process and analyze the NMR data so that the complete process perfectly matches the timeline required for medicinal chemistry.

In conclusion, the benchmark and the application to PIN1-fragments provide a clear understanding on the route to be followed for successful protein-fragment complex structure calculation: 1) measurement and analysis of the best topology fragment; 2) introduction of PA, if needed; 3) calculation of this first complex structure; 4) introduction of the full assignments obtained from the first structure calculation to the further complex calculations, as it was already demonstrated in prior study. Moreover, it points out the weak points to be improved in the state-of-the-art, i.e. NOE error. Finally, the protein-fragment complex structures calculated from NMR data can be obtained in a few days or weeks and can be used by the medicinal chemist for fragment linking, fragment growing or pharmacophore building. The *NMR*² approach may open a new avenue in NMR drug design by enabling fast and robust fragment structure-based drug discovery. Having access to the structure of the binding site for each weak binder allows investigating chemical scaffolds that would otherwise be discarded which broadens the chemical knowledge as well as enables the druggability of the protein targets.

Chapter 4: NMR-based drug discovery for SARS-CoV-2

Part of this work will be used for publication after the thesis release.

Contributions of the author: F.T. performed the NMR screening, cross-checking with ^{15}N -TROSY, and the data analysis.

Introduction

In 2019 emerged in Wuhan, China, a novel coronavirus strain held to cause the, so far, major pandemic of the 21st century.¹⁴⁷⁻¹⁴⁸ From March 2020 until the writing of this thesis, the everyday life has been disrupted by social distancing, home office closing, and other prophylactic restrictions. Despite these global efforts, 170 million people have been acknowledged to be contaminated and 3.5 million people deceased (source: <https://covid19.who.int>) from the complication of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). At the moment when the laboratory started putting effort in the research against SARS-CoV-2, the promise of a vaccine remained unclear and the hottest lead was the use of hydroxychloroquine.¹⁴⁹ Since then, hydroxychloroquine was invalidated¹⁵⁰ and vaccination represents the most realistic hope to exit the world crisis. Nevertheless, the apparition of several variants threatens the durability of vaccine induced immunity, and the global vaccination campaign would benefit from being supported by curative drugs.¹⁵¹ Therefore, the finding of potential drug applicable to SARS-CoV-2 treatment would be a critical asset in the fight against the disease. The virus penetrates in the cells through binding to the ACE2 receptor which is highly expressed in the lung type II alveolar cells, thanks to the receptor binding domain (RBD) of its spike protein, capping the surface of the virus.¹⁵²⁻¹⁵³ The virions replicate inside the cells and can yield to different state of severity, going from asymptomatic to death. The most common symptoms are moderate respiratory illness with fever, cough, and tiredness. In the risk population (i.e., elderly, diabetes or people with cardiovascular diseases), the virus can trigger exaggerated immune response known as the cytokine storm, which causes a severe form of the disease which requires the admission in intensive care units and can evolve into death.¹⁵²⁻¹⁵³

The severe acute respiratory syndrome coronavirus-2 and relevant targets for drug discovery

The SARS-CoV-2 is a positive-sense single-stranded RNA beta-coronavirus. Among the several open reading frames (ORFs), ORF1 codes for two distinct large polypeptides, pp1a and pp1ab. These polypeptides are cleaved by the virus proteases, the main protease (M^{pro}) and the papain-like protease, and encode for the non-structural protein (nsp), such as M^{pro}. The remaining ORFs encode the structural proteins of the virus, which are necessary for the virus particle synthesis. Different targets have been imagined to tackle the SARS-CoV-2 penetration, replication, and survival in the human body. The first one is the inhibition of the RBD-ACE2 interaction at the origin of the cell penetration by the virus.¹⁵⁴⁻¹⁵⁶ This can be addressed mainly with antibodies, as the RBD is highly glycosylated, and the protein-protein interaction interface formed with the ACE2 still represents a challenge in medicinal chemistry. Additionally, the RNA-dependent RNA polymerase (RdRp) is composed by the catalytic subunit nsp12 and the accessory subunits nsp7 and nsp8, and is the enzyme involved in the genome replication as well as in the transcription.¹⁵³ Remdesivir, initially developed against Ebola, is known to inhibit the RdRp, and was repurposed as a treatment of SARS-CoV-2, with FDA approval.¹⁵⁷⁻¹⁵⁸ However, the poor risk-benefit balance did not yield to generalized use in clinics.¹⁵⁹ The M^{pro} of SARS-CoV-2 is a non-structural protein (nsp5) which active site, constituted of C144 and H41, recognizes and cleaves the amino-acid sequences L-Q↓-(S, A, G). This sequence is present at 11 sites of the polyprotein 1ab, for which the cleavage by M^{pro} is a mandatory maturation step.¹⁵³ Thus, the inhibition of the M^{pro} is a promising approach to stop the in-cell viral replication. M^{pro} shows important similarity to its homologous protein in SARS-CoV-1, with which it differs by only 12 amino-acids. However, only one mutation, A46S, is close to the binding site. Nevertheless, the remaining mutations affect

the dynamics of the protein enough to hinder the repurposing of the drugs targeting SARS-CoV-1 M^{pro}.¹⁶⁰⁻¹⁶²

Drug repurposing investigates the potential of already marketed drugs to be retargeted to an indication different than their initial one.¹⁶³⁻¹⁶⁵ Because approved drugs have already gone through the clinical trial process, the attrition rate is expected to be lower as compared to new drug candidates. Furthermore, as the safety profile of the drug is already known, and in case the repurposed drug is used in similar concentrations and administration way (*per os, bolus, ...*), many steps of the clinical trials can be skipped so that the drug can reach the market faster. Drug repurposing does not necessarily only consider approved drugs, drugs withdrawn from market, or drugs that failed to pass the late stage of clinical trial can be also considered. In the context of Covid-19, the urge for finding an efficient and available treatment oriented the investigation towards approved drugs or drugs that failed in the phase III of clinical trials, in order to reduce at a maximum, the attrition rate and the time to first-in-human.¹⁶⁵ Based on the target analysis, we therefore opted for an NMR-based drug repurposing campaign directed to M^{pro}. Furthermore, fragment screening specially designed for epitope mapping (MiniFrag), has been performed.¹⁶⁶

Many studies have been conducted with the goal to repurpose drugs for the treatment of Covid-19, and targeting M^{pro} in particular. Nevertheless, the structural data available to rationalize the finding of such drugs was particularly scarce. The Diamond Light Source released the structure of ~70 M^{pro}-fragment complexes, obtained from X-ray crystallization (source: <https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html>). However, crystal soaking does not provide any information on the affinity of the fragments, and therefore additional data are needed such as from SPR or NMR, to be analyzed extensively. On the other hand, in order to contribute to the global effort, we decided to order 12 molecules known to bind to M^{pro} from the literature. We selected molecules validated in at least two assays (SPR, TSA, ITC, enzymatic assay), and ordered them to cross-check the binding to M^{pro} by performing ¹⁵N-TROSY experiments.^{160, 167-169} The final goal being that the protein-ligands complex structures are calculated for the systems showing strong binding.

The Minifrag library contains 80 different low molecular weight fragments. The median is MW = 99 Da and the biggest fragment is only MW = 264 Da. The screening was performed using STD NMR, and cross-checked with ¹⁵N-TROSY experiments. This small molecular size is aimed to probe single pharmacophore interactions.¹⁶⁶

The FDA approved library is composed of 686 molecules used as drug treatments in human. Since, these molecules have already been through clinical trials prior to obtain FDA approval, they have the potential to be quickly repurposed for another pathology.

The NMR screening was performed by STD, in case a molecule interacts with M^{pro}, steady NOE transfers magnetization to the molecule's nuclear spins, resulting into a signal increase of the molecule (see Chapter 1).³⁶ In practice, only an important number of collisions during the saturation time will yield to a significant signal increase. For this reason, the molecules are present in excess (10-1000-fold) as compared to the target, and only transient interactions, i.e. weak binders, will be detected. Complementarily, a FRET assay evaluating the enzymatic activity was performed by the Scapozza Group in the university of Geneva on the FDA approved library.

The HPLC assay relies on the immobilization of M^{pro} onto Ni beads through the His-tag at the C-terminus of the protein. The protein is incubated with the FDA drug and pooled down by centrifugation, so that the molecules binding to M^{pro} are absent from the supernatant. The supernatant is filtered and analyzed by HPLC. The difference between the chromatograms of the

reference (ligand processed with beads and no M^{pro}) and the screening reveal the presence of a binding event. As the concentration of M^{pro} and of the ligands are of 50 μM only hits with dissociation constant up to ca 50 μM are expected to be observable.

All the hits are cross-checked by performing ¹⁵N-TROSY experiments (See Chapter 1 *Introduction*).³²⁻³³ As of the size of the M^{pro} dimer (74 kDa) the TROSY experiments are preferred over HSQC. In addition, the backbone deuteration of M^{pro}, by expression in 80% D₂O M9 medium, improved significantly the TROSY spectra, as the spin diffusion was reduced. The deuterated M^{pro} could be measured at concentrations of 0.3 mM within an hour, while the non-deuterated M^{pro} at the same concentration required 12 hours of measurement. Therefore, the deuterated M^{pro} was preferred to perform the TROSY hit cross-checking. The workflow is as follows: a reference sample is recorded for the apo-M^{pro}, and the samples to check contain M^{pro} and the eventual binders. In case chemical shift perturbations (CSP) are observed upon the addition of the molecule, this is the signature of an interaction of the ligand with the target.¹⁶⁻¹⁸

The initial goal was to determine the structures of M^{pro}-ligand complexes identified from the different screening using the NMR² protocol which has been demonstrated to be fast and reliable for the production of structure activity relationship data. The structures can be used as input for the design and refinement of a pharmacophore which is screened against bigger libraries of molecules to find potential strong binders. Alternatively, the complex structures can be used by medicinal chemists to rationally design M^{pro} inhibitors.

On the NMR study of the M^{pro} protein

The full-length M^{pro} was expressed, with a His tag at the C-terminus, and a glutathione S-transferase (GST) expression tag that is later cleaved by the tobacco etch virus (TEV) protease. M^{pro} has the advantage to be expressed in high yield (20-25 mg/L of culture), and it forms a dimer which ensures its enzymatic activity, and which is therefore relevant for the study of inhibitors. The protocol relative to the production of recombinant M^{pro} has been published elsewhere.¹⁷⁰ The construct of M^{pro} (GST-TEV-M^{pro}-HIS) is provided below:

MSPILGYWKIKGLVQPTRLLLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLQTQSMAIIRYIADKHNMLGGCPKERAEISMLGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIQIDKYLKSSKYIAWPLQGWQATFGGGDHPKSDLEVLFGQPLGSAVLCSENLYFQGS

SSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCP
RHVICTSEDMLNPYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCV
LKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPN
FTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNF
YGPVDRQTAQAAGTDTTITVNVLAWLAAVINGDRWFLNRFTTTLN
DFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNG
MNGRTILGSALLEDEFTPFDVVRQC

SGVTFQGP

HHHHHH

The color coding of the above-mentioned construct is as follows: in blue the GST tag sequence, and in green the his-tag used to bind to nickel containing column for purification; in bold, the sequence inserted containing the TEV cleavage site (in red); the part underlined is the

autocatalytic site of the M^{pro} (LQ↓S)¹⁵³ which was accidentally destroyed by inserting the TEV cleavage site. Yet this seems to have no deleterious effect on the protein.

The high molecular weight of the M^{pro} (MW = 74 kDa) is responsible for linewidth broadening, a consequence of the fast transverse relaxation of the system. The transverse relaxation increases with the rotational correlation time of the protein, and depends heavily on spin-spin interactions. To address this issue, M^{pro} was expressed in M9 culture media with 75% D_2O , to ensure the deuteration of the backbone and sidechains of the protein. As discussed in the Chapter 2, *NMR² method development and application to the bromodomain BRD4-BD2*, for perdeuteration, partial deuteration reduces the relaxation, narrows the linewidth and improves the signal-to-noise.^{34, 171-172} The partial deuteration of the M^{pro} improved significantly the SN and resolution of the ^{15}N -TROSY shown in Figure 4.1. In consequence, the recombinant expression of M^{pro} aimed to cross-check the hits from the screening with ^{15}N -TROSY was always carried out with 75% D_2O M9 minimal media.

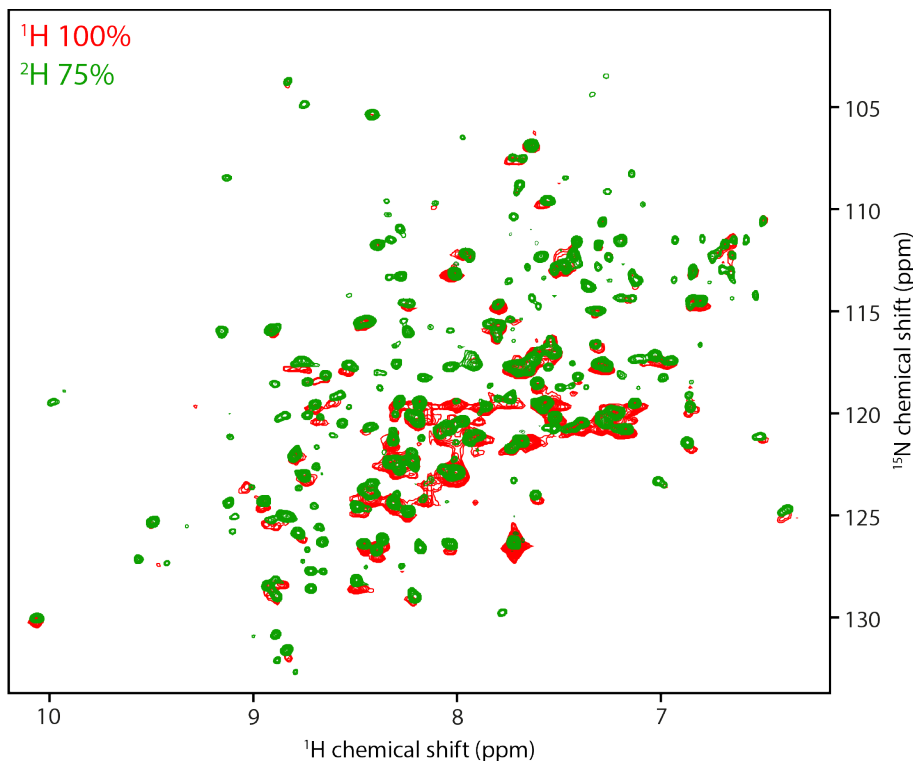


Figure 4.1: ^{15}N -TROSY spectra of M^{pro} expressed with normal water in red, and 75% D_2O in green. In both cases the sample was concentrated at 400 μM and measured at 303 K.

Despite the partial deuteration, the measurement of M^{pro} at room temperature (298 K) yields to poor signal-to-noise. Considering the number of samples that needed to be eventually cross-checked, it was important to ensure that the spectrum measurement time as short as possible with

the best SN as possible. The temperature of measurement was thus increased to accelerate the rotational correlation time, and slow down the relaxation rate.²⁵ The measurement at 303 K was found to be a good compromise to achieve satisfying SN (Figure 4.2) and ensure M^{pro} stability for several hours, enough to record good quality ¹⁵N-TROSY spectra.

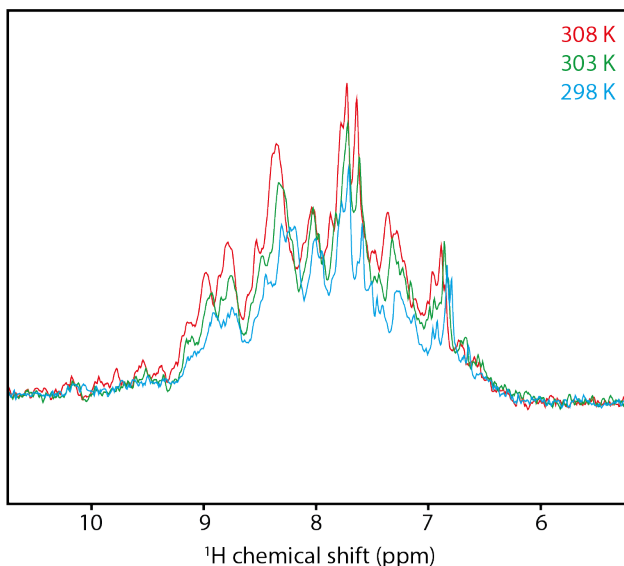


Figure 4.2: first 1D time point experiment from ¹⁵N-TROSY of M^{pro}, at different temperatures. The concentration is 400 μ M, the acquisition was carried for 1024 points with an interscan delay of 1 s, and 32 scans.

In summary, the full length M^{pro} requires partial deuteration and measurement at 303 K to provide satisfying ¹⁵N-TROSY spectra. Moreover, the stability of the protein is problematic, storage by freezing is not possible, and the protein precipitates after a couple of days at 4 °C. Besides, the concentration of the protein is limited and could not be pushed above 400 μ M, rendering the measurements particularly challenging. The production of a more stable construct, would be a critical improvement for the investigation of M^{pro} inhibitors.

Screening

Literature cross-checking

In the first months of the Covid-19 pandemic, considerable effort has been made all around the world to identify commercially available drug molecules that could be repurposed to treat the SARS-CoV-2. The sickness can be addressed through different leverage points such as the cytokine storm, the virus cell-penetrance, or the inhibition of key virus proteins. Among these lasts, M^{pro} is a cornerstone of the virus infectivity as it is necessary for the maturation of pp1a and pp1ab into the nsps constituting the virus machinery.¹⁵³ The inhibition of M^{pro} would then hinder the virus proliferation, and several studies on the M^{pro} inhibition with FDA approved, or

analogous protease inhibitors, were already published at this time point.¹⁶⁰⁻¹⁶¹ Some of the most promising molecules identified in the literature were ordered to cross-check the binding to M^{pro} by performing ¹⁵N-TROSY experiments, with the goal to observe or not the presence of CSPs. Hence, chloroquine¹⁶⁹, atazanavir¹⁶⁸, dipyrindamole¹⁶⁹, boceprevir, MG132, calpain inhibitor II and XII were ordered.¹⁶⁷ However, a quick solubility test in the presence or in the absence of M^{pro} revealed that only boceprevir, dipyrindamole and calpain inhibitor XII were soluble at concentrations above 100 μ M, necessary for this test. The ¹⁵N-TROSY spectra of the dipyrindamole-M^{pro} and calpain inhibitor XII-M^{pro} samples showed no CSPs in comparison to the apo-M^{pro} spectrum. Nevertheless, strong CSPs were observed upon the addition of boceprevir to M^{pro} (Figure 4.3), and the binding kinetics were found to be in the slow exchange regime. Boceprevir is also the only of the abovementioned molecules for which a structure of the complex formed with M^{pro} has been solved by X-ray crystallography (pdb: 6wnp, 7mco).¹⁷³ Boceprevir is a Michael acceptor thanks to its alpha-ketoamide moiety, and is therefore a reversible covalent binder of the M^{pro}. Literature reports a K_d of 1.18 μ M and a 50% inhibitory concentration (IC₅₀) of 4.13 μ M.

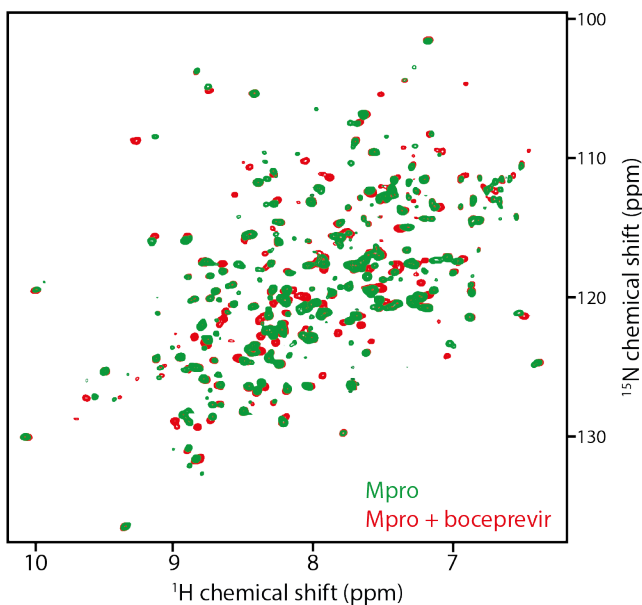


Figure 4.3: ¹⁵N-TROSY spectra of apo-M^{pro} (green) and M^{pro} in complex with boceprevir. The protein and the ligand are in ratio 1:1 at a concentration of 400 μ M. both spectra were measured at 303 K.

While the X-ray crystallography M^{pro}-boceprevir complex structure is known, solving the system with NMR² would be a good validation and may reveal some slightly different ligand pose. With the hope of obtaining the complex structure, doubly labeled M^{pro} was expressed to measure F₁-¹⁵N,¹³C-filtered-¹H,¹H-NOESY⁴² and calculate the M^{pro}-boceprevir complex structure using NMR².⁴³ Unfortunately, no exploitable intermolecular cross-peak could be observed in the filtered NOESY spectrum, which was explained by the size of the M^{pro} dimer (MW =74 kDa). The

absence of signal results from the excessively fast transverse relaxation at play. The use of partially deuterated double labeled M^{Pro} was intended to increase the signal-to-noise, as well as the measurement at higher temperature, with no success. To verify our hypothesis that the M^{Pro} is relaxing too fast, we recorded ¹³C-ctHSQC and observed at the FID signal amplitude for different indirect dimension time.¹⁷⁴ The indirect dimension time signal acquisition is constant and defined by the user, and in the first time point the signal is refocused so that the only modulations happening are the transverse relaxation, with the exchange contributions.²⁵ While a classic ¹³C-ctHSQC indirect time dimension is recorded with either 13.3 ms or 26.6 ms, the FID showed signal only for times below 2 ms, confirming the extremely fast relaxation. This result sadly invalidates the possibility of recording filtered NOESY of sufficient quality to calculate NMR² structures for M^{Pro}. Moreover, the protein showed poor stability at room temperature, and forms aggregates after a couple of days which is incompatible with the recording of reliable NOEs. The eventuality of using another construct containing only the N-terminus, more stable and remaining as a monomer has been investigated. Nonetheless, this construct is known for not being enzymatically active, and the ¹⁵N-TROSY in the presence of boceprevir showed only minor CSPs. Finally, the eligibility of boceprevir as a repurposed drug against SARS-CoV-2 is very unlikely. Indeed, boceprevir is quickly eliminated by the liver where it accumulates, which is not a problem for its initial purpose, treating hepatitis C virus, but a major hurdle in the case of Covid-19 treatment.

MiniFrag library screening

The screening of the MiniFrag library was performed for M^{Pro} in 2 days of successive measurement. As of the rather short measurement time and the high amount of molecule available (10 mg/each), the reference spectrum (ligand alone) was recorded and subtracted from the spectrum of the ligand in the presence of M^{Pro} spectrum. The subtraction of the reference spectra is the gold standard to discard false positives. The spectra for which a signal is observed witness from the transient interaction between the ligand and the magnetically saturated protein. In total, four hits could be identified from the difference spectra: phenol, piperazine, 4-bromo-pyrazole, and isopropylurea (Figure 4.4). The hit compounds from the STD screening were measured in ¹⁵N-TROSY experiments to cross-check the interactions. However, no CSP were observed for any of the fragment, even after increasing the fragments concentration up to 5 mM. The non-observation of N-H CSP is a recurrent phenomenon for very weak binders such as the very small molecules composing the MiniFrag library.¹⁶⁶ The use of ¹³C-HSQC could address this issue as sidechain CSP are more prone to be observed for very transient interaction. Since the M^{Pro} samples are not stable enough, ¹³C labeling was not considered as it is too costly. The MiniFrag library has been sent for FRET enzymatic assay, and result may infirm or confirm the observation in STD NMR.

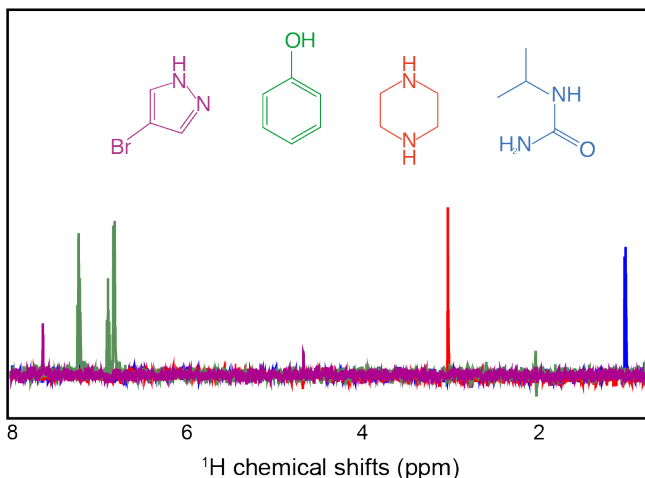


Figure 4.4: STD difference spectra between the sample spectrum (M^{pro} + fragment) and the reference (fragment), for the four hits of the MiniFrag library. The color code of the spectra is applied respectively to the color of the fragment depicted.

FDA library screening

The screening of the FDA approved library with M^{pro} took 14 days of consecutive measurement without recording reference experiments. *In fine*, 254 hits could be detected from the STD screening, which correspond to a hit rate of 37 %. This hit rate for drug sized molecules is far higher than expected. Indeed, as the molecular complexity increases, the probability for the drug to fit into the binding pocket and build the necessary interactions to form a stable interface gets lower. Furthermore, STD NMR detects transient protein-ligand interactions as it is the case for weak binders. Such transient interactions could also take place in the case the ligand exchanges between a free form and a bound form to supramolecular arrangement, e.g. micelles or aggregates. The data fetched from the databases such as PubChem, or DrugBank showed that only 35 hits are showing good solubility at the ligand concentration used during the screening (Table 4.1). Among these molecules only seven showed strong STD signal enhancement, but no CSP was observed in ^{15}N -TROSY experiments. The good solubility of a molecule, or a low $\log(P)$ does not guarantee the absence of transient supramolecular arrangements such as liquid droplets, or micelles. It can be concluded, in a first instance that a concentration of 200 μM is too high to perform drug like molecule STD NMR screening. One solution could have been to reduce the concentration of the molecules, and of the M^{pro} to maintain the excess ratio, but then the measurement time increases quadratically with the dilution. Moreover, STD is designed to find weak binders ($K_d \geq 50 \mu\text{M}$) so the drug repurposing potential of the STD active molecules is limited. Furthermore, the eventuality of fueling medicinal chemistry with drugs that can be modified to gain several orders of magnitude in affinity is rather low when starting from molecules with such a mature molecular size. Keeping to the goal of identifying drugs that could be potentially repurposed to the treatment of SARS-CoV-2, orthogonal experiments that are more suited to detect the strong binding of a molecule to a target have been conducted.

Table 4.1: soluble molecules showing positive STD enhancement

<i>Molecule</i>	<i>STD enhancement</i>
<i>Theophylline</i>	Strong
<i>Terbutaline</i>	Strong
<i>Niacin</i>	Strong
<i>Gallamine</i>	Strong
<i>tenoxicam</i>	Strong
<i>tropium</i>	Strong
<i>Dapsone</i>	Strong
<i>Antipyrine</i>	Medium
<i>methylaminolevulinate</i>	Medium
<i>penciclovir</i>	Medium
<i>acyclovir</i>	Medium
<i>etidronic acid</i>	Medium
<i>topotecan</i>	Medium
<i>topiramate</i>	Medium
<i>methocarbamol</i>	Medium
<i>tranexamic acid</i>	Medium
<i>hydroxyurea</i>	Medium
<i>aztreonam</i>	Medium
<i>cetirizine</i>	Medium
<i>Nicotine</i>	Medium

<i>Phenelzine</i>	Medium
<i>hydrochlorothiazide</i>	Medium
<i>doxycycline</i>	Medium
<i>ethosuximide</i>	Medium
<i>sumatriptan</i>	Medium
<i>sulpiride</i>	Medium
<i>oxyphenonium</i>	Medium
<i>raltegravir</i>	Medium
<i>salicylamide</i>	Medium
<i>metoprolol</i>	Medium
<i>fomepizole</i>	Medium
<i>acetaminophen</i>	Medium
<i>benzocain</i>	Medium
<i>trapidil</i>	Medium
<i>chlormezanone</i>	Medium

The HPLC screening was performed at concentrations comparable to the NMR screen (0.05 mM ligand and 0.05 mM M^{Pro}). The samples were prepared manually and disposed in nine 96 well-plates prior to analysis by HPLC. The plates 1 and 2 could be analyzed but the system was clogged after a while. The reason for this remains uncertain, as the samples were filtered before the injection in the HPLC. Nevertheless, the possibility of the delayed M^{Pro} aggregation in the HPLC system, forced to abandon this approach. From the subset of molecules that could be screened, nifedipine, liothyronine, docetaxel, butamben, perphenazin, pyrvinium, ibrutinib, and tazarotene were identified as hits. The hits were cross-validated with ¹⁵N-TROSY and yielded to the identification of ibrutinib (Figure 4.5) as a potentially promising hit. Ibrutinib is a tyrosine kinase inhibitor and shares structure proximity with the pelitinib which has been previously reported to be an allosteric binder of Mpro.¹⁷⁵ Given the lack of assignments, the CSP pattern for the M^{Pro}-ibrutinib complex could not determine whether the binding site was the enzymatic site or allosteric. Up until now, 42 tyrosine kinase inhibitors have been reaching the market, and many others were stopped during the pre-clinical or clinical trials.¹⁷⁶ Further investigation of the

tyrosine kinase inhibitors family could be of interest to find a stronger binder. However, tyrosine kinase inhibitors are potent drugs involved in key regulation pathway, severe side effects have been reported such as hepatotoxicity and cardiotoxicity. Therefore, if the contribution for fueling medicinal chemists with refined allosteric pharmacophore might be promising, the potential of tyrosine kinase inhibitors as a frontline treatment is unlikely.

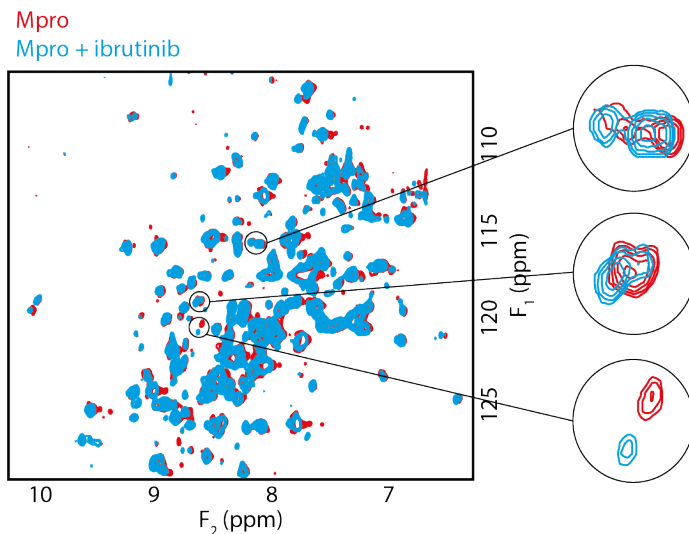


Figure 4.5: Overlaid ^{15}N -TROSY spectra of M^{pro} (red) and $M^{\text{pro}} + \text{ibrutinib}$ (blue) at $300 \mu\text{M}$ 1:1 ratio. F_1 is the nitrogen chemical shifts in ppm and F_2 is the proton chemical shifts in ppm.

From the FRET enzymatic assay, 77 hits were identified by the collaborators in Geneva, which makes a rate of 12%. Such rate is again unexpectedly high as of the molecular complexity of the molecules. The comparison with the STD data revealed that 36 of them are also positive in the STD NMR assay (Figure 4.6 and 4.7). The STD NMR hits are weak binders and considering the high hit rates of both assays these could be essentially an overlap of false positives. Nevertheless, weak binders giving rise to STD signal can perfectly interfere with the enzymatic activity and should be considered. For this reason, these compounds were prioritized for ^{15}N -TROSY cross-checking. The molecules showing the strongest STD enhancements (Figure 4.6 and 4.7, Table 4.2) were then cross-checked for binding to M^{pro} . However, no CSP could evidence any binding event, even if not all the molecules could be screened at this point due to limited amount of ^{15}N -labeled protein.

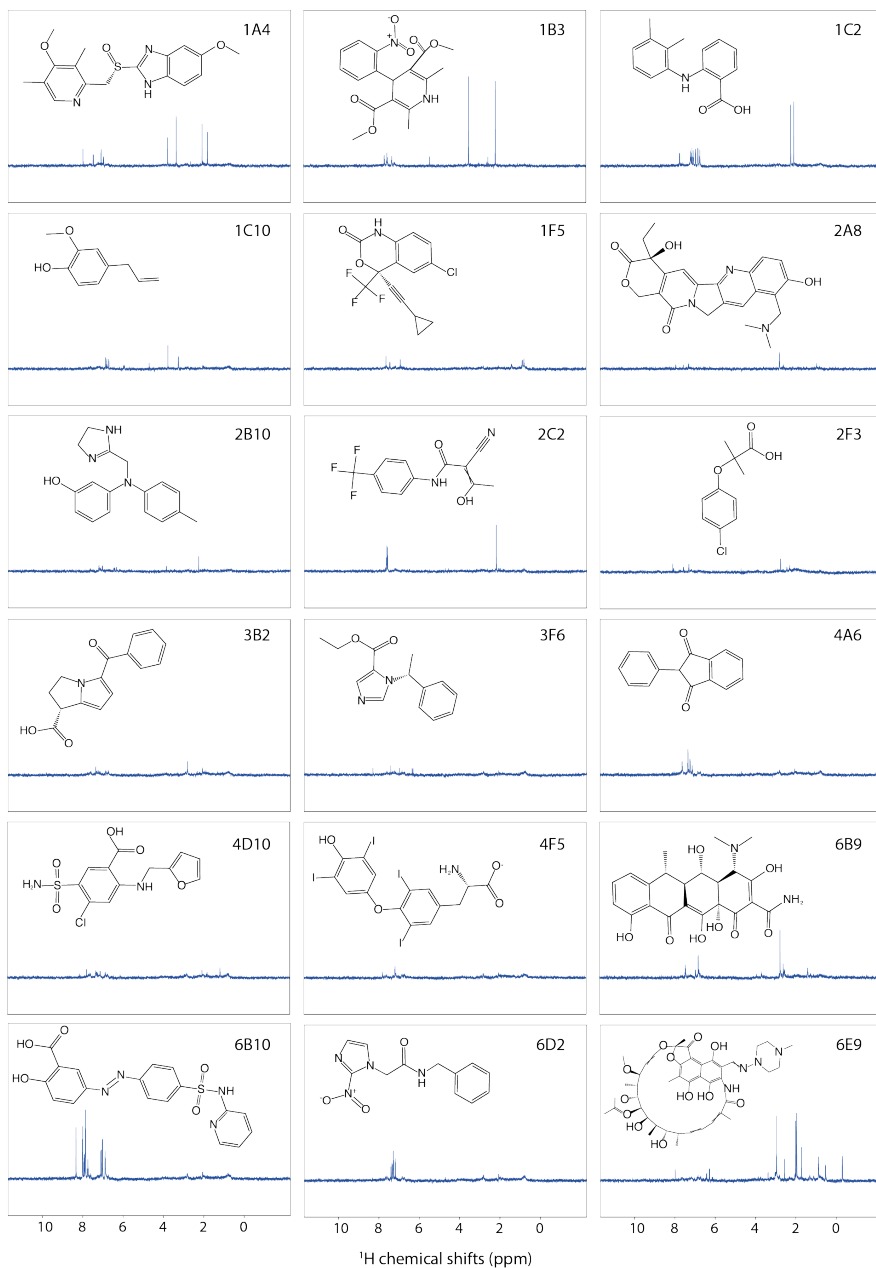


Figure 4.6: STD spectra of the hit molecules showing inhibitory properties in the FRET enzymatic assay.

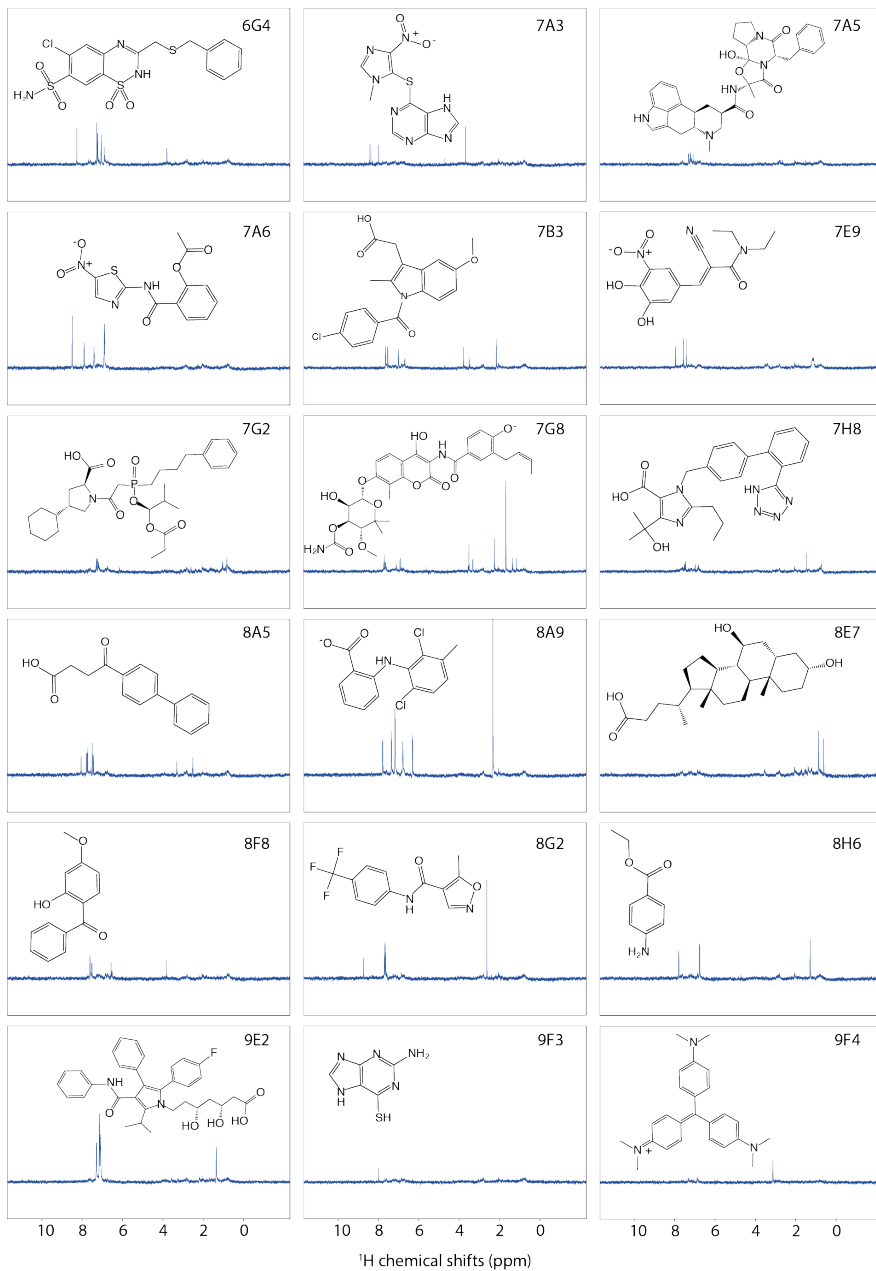


Figure 4.7: STD spectra of the hit molecules showing inhibitory properties in the FRET enzymatic assay.

Table 4.2: summary of the cross-validated molecules for

<i>Reference ID</i>	<i>Molecule</i>	<i>STD enhancement</i>
1A4	Esomeprazole	Strong
1B3	Nifedipine	Strong
1C2	Mefenamic acid	Strong
1C10	Eugenol	medium
1F5	Efavirenz	medium
2A8	Topotecan	weak
2B10	Phentolamine	weak
2C2	Teriflunomide	strong
2F3	Clofibrate	medium
3B2	ketorolac	weak
3F6	Oxybutynin	weak
4A6	Phenindione	medium
4D10	Furosemide	weak
4F5	Levothyroxine	weak
6B9	Doxycycline	strong
6B10	Sulfasalazine	strong
6D2	Benznidazole	medium
6E9	Rifampicin	strong
6G4	Benzthiazide	strong
7A3	Azathioprine	strong

7A5	Dihydroergotamine	weak
7A6	Nitazoxanide	strong
7B3	Indomethacin	strong
7E9	Entacapone	strong
7G2	Fosinopril	medium
7G8	Novobiocin	strong
7H8	Olmесartan	medium
8A5	Fenbufen	strong
8A9	Meclofenamic acid	strong
8E7	Ursodeoxycholic acid	medium
8F8	Oxybenzone	medium
8G2	Leflunomide	strong
8H6	Benzocaine	strong
9E2	Atorvastatin	strong
9F3	Thioguanine	weak
9F4	Gentian violet	weak

Yet, the most promising M^{pro} inhibitors in the enzymatic assay are most likely strong binders which would not be identified as hits in the STD screening. Indeed, four of the six compounds showing the strongest inhibition in the FRET enzymatic assay were negative in STD (etacrynic acid, benzethonium, dantrolene, and niclosamide). The two remaining are in the STD screening (sulfasalazine, nitazoxanide), suggesting rather low affinities. These six compounds were cross-checked in ^{15}N -TROSY experiments. The strongest observed CSP were observed for the etacrynic acid which is possibly a reversibly covalent binder, in the same manner as boceprevir (Compare Figure 4.8 and Figure 4.3). Moreover, CSP were also observed in the M^{pro} -sulfasalazine, M^{pro} -nitazoxanide and M^{pro} -niclosamide spectra (Figure 4.8). Nonetheless, in these latter spectra only minor CSP suggesting very weak binders could be observed. The ^{15}N -TROSY cross-checking was extended to other hit molecules which were not necessarily showing STD signal, but showing a relatively important enzymatic activity inhibition (pyrvinium, doxycycline, carmustine,

levothyroxine, ropivacaine). While a few minor intensity shifts could be observed, no significant binding was observed.

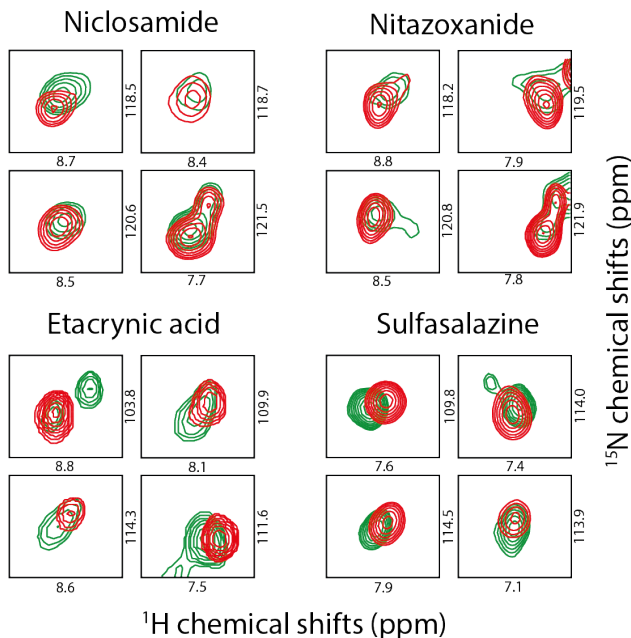


Figure 4.8: Chemical shift of apo-M^{Pro} (red) versus holo-M^{Pro} (green) for different FDA approved drugs, at 300 μM and ratio 1:1.

Finally, despite the orthogonal screening techniques involved, no high potential FDA approved molecule could be identified. Nevertheless, a few weak binders for which the critical chemical features are under evaluation may provide information on the M^{Pro} pharmacophore such as ibrutinib and etacrynic acid.

Outlook on the M^{Pro}-ligand complex structure resolution by NMR

The NMR based approach to identify FDA approved drugs that could be potentially repurposed to the treatment of Covid-19 has shown severe limitations. The first reason for that is that most of the NMR screening 1D experiments rely on the transient interaction of a ligand with a bigger entity (STD, T_{1ρ}, T₂).¹⁵ The propensity of drug like molecules to form aggregates at concentrations above 100 μM or to form micelles, give rise to a high rate of false positives, which is extremely complicated to narrow down. Moreover, in the context of drug repurposing, the hits should be nanomolar binders and would therefore not be identified with techniques probing transient interactions. In case one would decide to investigate potential repurposable drugs onto a target, the experiment of choice would be ¹⁵N-HSQC/TROSY, which would require large amount of labeled protein. Nevertheless, it would be possible to imagine a strategy where only low concentration sample (50 μM) of protein-ligand are measured with long acquisition time, the

time could be reduced by pooling together several drug candidates. Such a technique would still require considerable measurement time, especially for a protein providing weak signal-to-noise such as M^{pro}. In addition, the poor relaxation properties of M^{pro} impedes the structural resolution of its complexes with eventual binders by NMR², limiting structural investigation to X-ray crystallography. The methyl specific labeling described in the Chapter 2 on this thesis would eventually improve enough the signal-to-noise to obtain valuable NOE -based distance restraints, but this is an expensive approach and the stability issue of the protein would need to be solved first.

Conclusion

No FDA approved drug was demonstrated to have an anti-viral effect so far, and none of the many different target-based approaches tried by us and others have yielded to significant advances in the treatment of SARS-CoV-2. Therefore, it would be unfair to conclude that NMR performed poorly in comparison to other biophysics. Nevertheless, as of its net advantage at characterizing weakly interacting systems, NMR should be prioritized for fragment-based strategies. It is important to notice, that the most used repurposed treatment is dexamethasone, which was rationalized to fight the cytokine storm as its anti-inflammatory properties.¹⁷⁷ Such an approach, is phenotypic and not target-based.

Chapter 5: Exploration of the photo-CIDNP performances investigating new photosensitizers and chemical modifications of tryptophan

This chapter is an adaptation of two articles: Sobol A., [Torres F.](#), Aicher A., Renn A., and Riek R., *J Chem Phys*, 2019 & [Torres F.](#), Sobol A., Greenwald J., Renn A., Morozova O., Yurkovskaya A., and Riek R., *Phys Chem Chem Phys*, 2021.

Author contribution: F.T. performed the experiments and analysis. The initial dye screening was performed by A.S., and the TR-photo-CIDNP experiments and their analysis were performed by O.M. and A.Y..

Introduction

The intrinsic insensitivity of NMR remains over the years, the biggest hurdle for the technique. This is especially true in the context of biological applications, where the concentrations of biomolecules are in the low micromolar range. NMR sensitivity depends on the spin population polarization which obeys to the Boltzmann population distribution law (equation 5.1).⁶⁰ At biologically relevant temperatures, the energetic term ΔE is much lower than the thermic term kT , and the polarization is thus in the order of ppm, yielding a poor SN. The sensitivity of NMR improved significantly with the introduction of high field spectrometers, that brought the nuclear spin polarization up by improving the energy gap between the different spin states (equation 5.2).²⁵

$$\frac{N_{\alpha}}{N_{\beta}} = e^{(\Delta E)/kT} \quad (5.1)$$

And:

$$\Delta E = E_{\beta} - E_{\alpha} = \hbar\gamma B_0 \quad (5.2)$$

The SN of a given spectrometer is better appreciated using equation 5.3,²⁵ which includes most of the relevant parameters. It is important to notice that the number of spins, N , is a major parameter, imposing a concentration of samples far from physiological, typically for bio-NMR: 0.1-1 mM.

$$S/N \propto \frac{N\gamma^{5/2}B_0^{3/2}K}{\sqrt{\Delta f(T_c R_c + T_a[R_c + R_s] + T_s R_s)}} \quad (5.3)$$

Where N is the number of spins in the sample, γ is the gyromagnetic ratio, B_0 is the magnetic field of the spectrometer, and K the coil design dependent scaling factor. Δf is the receiver bandwidth, T_a is the amplifier temperature, R_c and T_c are the resistance and temperature of the coil, respectively. Finally, T_s and R_s are the temperature and the resistance induced by the sample in the coil, respectively. From equation 5.3, different technological improvements were derived that improved SN, such as an increased magnetic field, B_0 , as mentioned above. While the highest magnetic field commercially available, 1.2 GHz, improves the SN by a factor 2.82 in comparison to a 600 Mz, the former costs around 20M CHF and the latter less than 1M CHF. In parallel, cryoprobes have been developed to cool down the electronics of the amplifier (T_a) and the coil (T_c), resulting in lower resistance (R_c). Hence, the use of cryoprobes reduces the noise by a factor ~ 3 . In addition, reduced salt concentration or low mobility ions for the sample buffer salts, is important to decrease the resistance induced by the sample (R_s). Altogether, these improvements opened the way to NMR-based structural biology, and the study of biological systems in general. Yet, such studies are done at relatively high concentrations (0.1-1 mM) in comparison to the physiological concentration of most biomolecules (0.1-1 μ M). Furthermore, the timespan of bioNMR measurements is relatively long as compared to other biophysics analytics, going from hours to days of signal acquisition. The sole technological efforts to reduce noise and increase the sample polarization could never overcome the polarization barrier and showcase poor cost-effectiveness ratio. To overcome this hurdle, several hyperpolarization techniques have been proposed during the last 60 years such as Dynamic Nuclear Polarization (DNP),^{60, 178} Para-Hydrogen Induced Polarization (PHIP),⁶² Signal Amplification By Reversible Exchange (SABRE),⁶³⁻⁶⁴ and Chemically Induced Dynamic Nuclear Polarization (CIDNP), the last was independently discovered by Bargon et al.¹⁷⁹ and Ward et al.¹⁸⁰ in 1967. The CIDNP theory is

well established and relies on the radical pair mechanism proposed by Closs, Kaptein and Oosterhoff in 1969.^{71, 181} The extension by Adrian¹⁸² which uses the Noyes¹⁸³ diffusion theory constitutes an approximate theoretical framework used in the present work. The radical pair can originate from a flash photolysis, heat, or from the excitation of a photosensitizer that reacts with a molecule of interest. This technique is known as photo-CIDNP and presents the advantage to be easily triggered with a low energy light source, and to operate in solution at room temperature. The convenience of photo-CIDNP and the easiness of its setup (a light source plus an optic guide) makes it a promising technique to hyperpolarize biomolecules in a close to physiological environment.

Photo-CIDNP polarization theory

The photo-CIDNP effect involves a sequence of interactions between an optically excited photosensitizer and the molecule of interest in a magnetic field, known as the radical pair mechanism.⁷¹⁻⁷² The following section provides an overview on the different steps and the challenges associated with photo-CIDNP SN enhancement (SNE).

The first step is the optical excitation of the photosensitizer. At its wavelength of absorption, the photosensitizer electron can absorb a photon and jump to an orbital of higher energy. The excited photosensitizer can relax to its ground state by fluorescence or undergo intersystem crossing towards a triplet state (Figure 5.1). In the triplet state, the electron spins are parallel forbidding the relaxation to the ground state. Hence, the triplet state is long lived and can encounter and interact with other molecules. Due to the very short fluorescence life time of the excited S_1 state ($k_{\text{fluo}} \sim 10^8 \text{ s}^{-1}$) only a small fraction of molecules is in the excited state, S_1 , at moderate laser intensity (10 W/cm^2). For example, with a typical value of 1 W illuminated into a 3 mm NMR tube, 15 W/cm^2 corresponding to $1.5 \cdot 10^{20} \text{ phot/s}$ are obtained. With the absorption rate $k_{\text{ex}} = N_{\text{phot}}/\text{cm}^2 \cdot \sigma(\lambda)$ and assuming a typical absorption cross section of $\sigma(\lambda) = 10^{-16} \text{ cm}^2$, an excitation rate of $k_{\text{ex}} \sim 10^4$ is reached yielding to a S_1 population N_{S_1} of only 0.01% . Nevertheless, a considerable population N_T in the triplet state can be built up thanks to a high intersystem crossing rate, k_{ISC} , and a long lifetime of the triplet state:

$$N_{S_1} = N_{S_0} \frac{k_{\text{ex}}}{k_{\text{fluo}} + k_{\text{ISC}}} \quad (5.4)$$

$$N_T = N_{S_1} \frac{k_{\text{ISC}}}{k_T} = N_{S_0} \frac{k_{\text{ex}}}{k_{\text{fluo}} + k_{\text{ISC}}} \frac{k_{\text{ISC}}}{k_T} \quad (5.5)$$

With k_{fluo} describing the fluorescence relaxation of the singlet state and k_T is the triplet relaxation constant. In the case where the photo-CIDNP is performed at high magnetic field (i.e. $> 1 \text{ T}$), the degeneracy of the photosensitizer triplet state is removed, and three non-energetically equivalent states are present: T_+ , T_- , and T_0 .¹⁸⁴ As a consequence of its electronic structure, the triplet state dye is long-lived and highly redox-active. Among others, amino acids like tryptophan and tyrosine were found to be potent triplet quenchers by diffusion-controlled electron transfer. The rate depends on the reduction and oxidation potentials of the involved species and can reach as much as $2 \cdot 10^{10} \text{ Mol}^{-1}$ (collision limit).¹⁸⁵ Values of $3 \cdot 10^9$ and $2 \cdot 10^9 \text{ s}^{-1} \text{ Mol}^{-1}$ have been observed for tryptophan and tyrosine transferring an electron to Flavin, respectively.¹⁸⁶ With millimolar concentrations, rates are in the order of 10^6 s^{-1} . Electron transfer takes place from the highest occupied molecular orbital (HOMO) of an electron donor to one of the two singly occupied molecular orbitals (SOMO) of the photosensitizer, thereby forming a radical pair in the triplet state due to spin conservation (Figure 5.1). Competing processes are triplet to singlet relaxation (10^{-3} - 10^3 s^{-1}), quenching by oxygen ($10^9 \text{ M}^{-1} \text{ s}^{-1}$) vs. electron transfer ($10^9 \text{ M}^{-1} \text{ s}^{-1}$), which becomes

comparable at donor concentrations of 100 μM . Therefore, the sample is degassed by bubbling argon through or using enzymatic systems as described in the “Material and methods”.

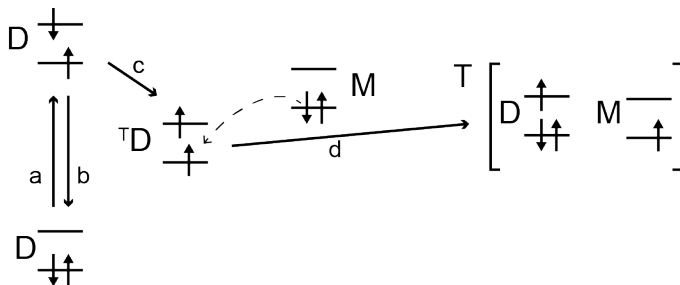


Figure 5.1: photosensitizer (D) excitation through photon absorption (a), fluorescence (b) or intersystem crossing (c) towards a triplet state. The triplet state photosensitizer can form a radical pair with a molecule (M) by electron transfer (d).

Since the radical pair is formed in the triplet states (T_+ , T_0 , and T_-) the energetically favorable immediate back transfer of the electron from the photosensitizer to the molecule is forbidden by Pauli principle. The recombination of the radical pair is only possible when the pair has evolved into a singlet state, and the electron-exchange coupling impedes the triplet singlet mixing of the radical pair. However, when the two radicals diffuse away from each other, the distance dependent electron-exchange coupling vanishes and the T_0 spin state oscillates between the triplet and singlet state with frequency-dependency on the Zeeman interaction of the radicals and the HFCC (Figure 5.2), according to the equation 5.6:

$$\omega_{ISC} = (g_M - g_D)\mu_B B_0 + \sum A_{M,a}m_{M,a} - \sum A_{D,b}m_{D,b} \quad (5.6)$$

With $g_{M,D}$ being the g-factors of the two radicals, μ_B the Bohr magneton, and A the HFCC which are summed over the nuclei a and b of the molecules M and the dye D , respectively. The equation yields to two different frequencies for the different nuclei spin states ($\pm 1/2$) denoted as m . Therefore, the recombination rate depends on the nuclear spin level. It is important to note, that only T_0 can evolve into a singlet state, which represent ca 1/3 of the total triplet population. According to Noyes diffusion theory,¹⁸³ the re-encounter probability of the spin correlated radicals from the pair is increased with respect to random collision with a radical from another pair, and exhibits a dependency of the order $t^{-3/2}$. When the radical pair is reformed after a time τ , the probability it is in a singlet state is given by equation 5.7:

$$|\langle S | \psi(\tau) \rangle|^2 = \sin^2(\omega_{ISC}\tau) \quad (5.7)$$

The radical reforming pairs in the singlet state recombine to regenerate the ground state molecules. As a consequence of equation 5.3, the molecules with different nuclear spin levels have different probability of being in the singlet state at the moment of the re-collision, this difference results in nuclear spin out-of-Boltzmann polarization. The resulting geminate polarization (P^G) is the difference in α and β spin populations originating from the probabilities of recombining for a spin α and β , respectively (equation 5.5).¹⁸⁷

$$P^G = |\omega_{ISC,\alpha}|^{1/2} - |\omega_{ISC,\beta}|^{1/2} \quad (5.8)$$

Where $|\omega_{ISC,\alpha}|^{1/2}$ and $|\omega_{ISC,\beta}|^{1/2}$ are the populations of α and β spins obtained by integration of equation (5.6) over the time dependent recollisional probability.

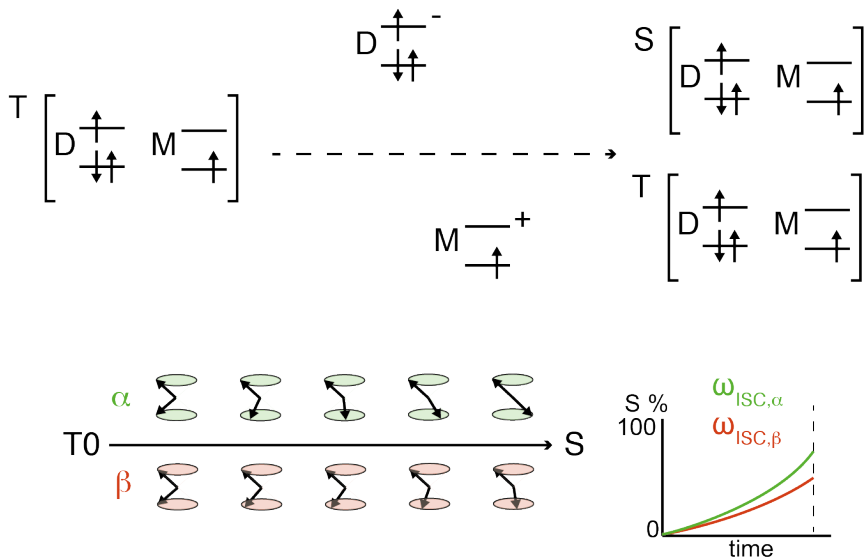


Figure 5.2: Evolution of the triplet state radical pair by splitting and recollision. During the time where the radicals are split apart, the electron state of the pair can evolve towards a singlet state under the influence of the magnetic parameters.

The radical pairs that recollide in any of the T spin states cannot recombine and therefore diffuse apart in the bulk solution, forming free radicals. Moreover, the nuclear spins in the radicals suffer paramagnetic enhanced relaxation compared to the diamagnetic species. Hence, the bulk paramagnetic polarization, initially $P^{\text{bulk}} = -P^G$, relaxes to the equilibrium polarization:

$$p^{\text{bulk}} = -P^G e^{-Kt} \quad (5.9)$$

Where K is the paramagnetic relaxation rate, and t is the time before random encountering with another free radical. The random collision of a free radical with a new partner forms F-pairs that can be in any electronic state. The singlet state F-pairs recombine directly and have no effect on the final polarization. The T_0 F-pair can generate polarization by the same nuclear spin dependent triplet-singlet electron spin mixing as it is generated in geminate polarization. The polarization arising from the repeated cycles of free M and D radical collisions is called F-pair polarization¹⁸⁸⁻¹⁹⁰ (P^F), and is added to the P^G to yield to the final polarization (P). The summary of the photo-CIDNP reaction mechanism is schematized in Figure 5.3.

The photo-CIDNP experiments were all conducted under continuous-wave (CW) laser irradiation on the order of seconds, except explicitly mentioned. Hence, the polarization present after a CW irradiation in the second timescale is a mix result from the P^G and the P^F . However, the fine analytics of the radical pair mechanism and the magnetic parameters governing the spin polarization can only be assessed using ns light pulses and deconvoluting the contribution of the F-pair polarization. This is done using time resolved (TR) photo-CIDNP. Several good and recent

reviews should be consulted for a more extensive description of this phenomenon, such as Okuno et al., and Morozova et al.^{184, 191}

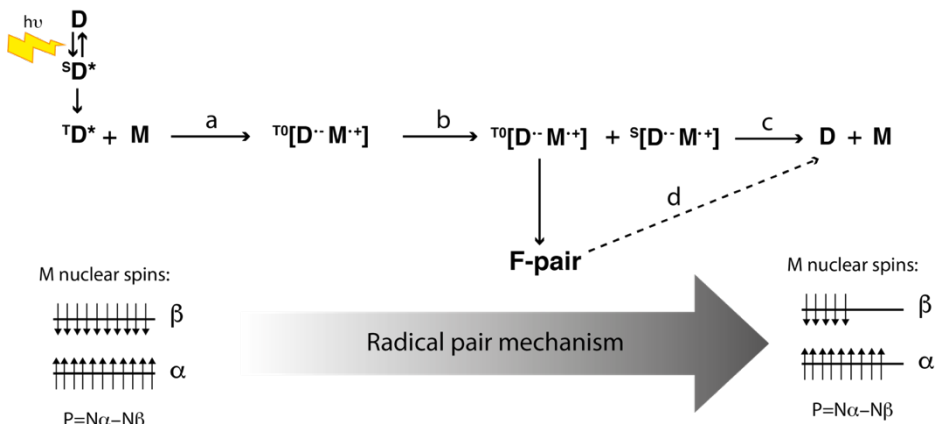


Figure 5.3: summary reaction mechanism of the photo-CIDNP mechanism. The triplet state dye forms a radical pair by electron transfer from the molecule of interest (a). The triplet state radical pair can split and then the intersystem crossing towards singlet state radical pair is possible before recollision (b). The singlet state radical pair recombines into ground state (c) and the nuclear spin populations are out of Boltzmann equilibrium due to the spin sorting in step b. The triplet state radical pair can escape and form free radical that can form F-pairs that undergo again through steps b and c to add up more polarization.

Dye investigation

As demonstrated from the photo-CIDNP theory, the photosensitizer plays a key role in the SNE. Since the hyperpolarization is an expression of the radical pair cyclic reaction yield, it is important to have important amount of triplet state photosensitizer that can react. To achieve such concentration of triplet state photosensitizer, it needs to absorb strongly at the wavelength used for irradiation, and to have a substantial intersystem crossing (ISC) rate, the expression of the rate at which the excited electronic singlet state converts to the electronic triplet state. Moreover, as of the optical density resulting from the strongly absorbing photosensitizers, the concentration must be kept in a range of a few dozens of micromolar. Hence, the triplet state must have a life time long enough to have the opportunity to interact with the molecule of interest and form the radical pair. Furthermore, the high ISC rate with a long triplet life time, promise a high quantum yield for the triplet state. Finally, the triplet state photosensitizer must be more prone to react through a photoreaction type I, which forms a radical pair, than II, which leads to the production of singlet state oxygen. Because of the necessity of strong light absorption, most of the photosensitizers in use are dyes. The classical dyes used in photo-CIDNP absorb in the ultraviolet (UV)-blue region: 3,3',4,4'-tetracarboxy-benzophenone (TCBP, 308 nm), bipyridyl (308 nm), and flavin mononucleotide (FMN, 450 nm). Nevertheless, UV irradiation threatens the molecular integrity of the dye as well as the molecule of interest. Moreover, the redshifting of the dye absorption properties is promising in case photo-CIDNP active dyes absorb in the near infrared,

this would match the optical window of human body and other mammals, possibly enabling in-vivo photo-CIDNP. Recently, the use of fluorescein monitored photo-CIDNP of tryptophan revealed promising results, with the detection of low and even sub micromolar concentrations of tryptophan. Interestingly, the performances of fluorescein with tryptophan were unmet with tyrosine. This thesis reports the CW-photo-CIDNP performances of Atto thio 12 (AT12) as a promising dye for the polarization of tyrosine and other “high g-factor” molecules. The AT12 dye also represent a considerable red-shift absorption as compared to the traditional photosensitizers, allowing irradiation at 532 nm. The different dyes used in photo-CIDNP and their respective wavelength are depicted in Figure 5.4. Although not absorbing in the red-infrared range, the use of green lasers represents a significant improvement in terms of sample stability, given the lower amount of energy delivered to the sample during irradiation.

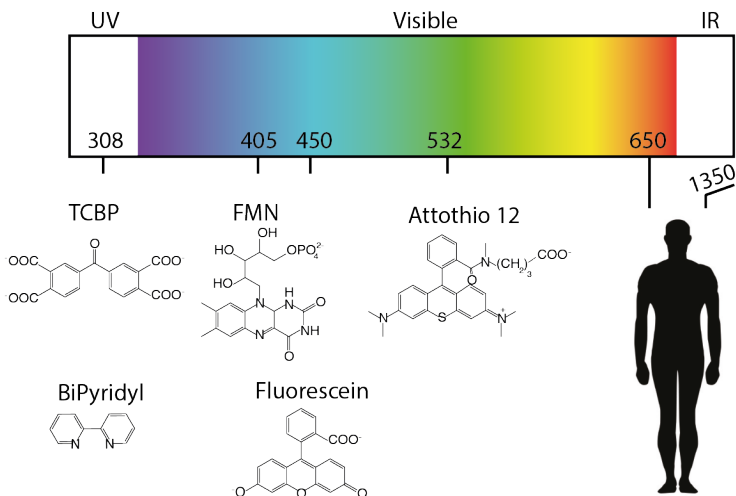


Figure 5.4: Summary of the dyes currently used in the photo-CIDNP, aligned with the light source wave length used for their excitation, in nm. The optical window of human body is also represented.

Chemical tuning of photo-CIDNP performances

If the parameters of the dyes used to monitor the photo-CIDNP effect can be optimized, the molecule of interest parameters must be adapted to the dye to enhance the polarization properties of the radical pair. This complex relationship is summarized in the equation 5.10:¹⁸⁴

$$P \propto \frac{HFCC}{12\sqrt{2}\Delta g\mu_B B_0/\hbar} \quad (5.10)$$

The equation 5.10 is a simplification of the result from geminate polarization but it can be also applied to steady-state photo-CIDNP, as it relates to the efficiency of the spin sorting. Of course, in the case of photo-CIDNP the F-pair polarization and parameters external to the magnetic parameters are expected to be of importance, but their relative influence is not defined well to predict from a chemical structure, its photo-CIDNP efficiency. Nevertheless, the dye-molecule complementarity appears under the form of the equation 5.10, where the relationship between the magnetic parameters HFCC and Δg is defined, and in the high field limit, high HFCC combined

to small Δg is in favor of high photo-CIDNP polarization. Few is known on the chemical features necessary to tune such conditions to an optimum and a serendipitous observation of a highly polarized impurity appearing after the repeated irradiation of tryptophan started our journey towards the understanding of such features (Figure 5.5). This impurity was characterized, synthesized and measured. The magnetic parameters were calculated and validated experimentally, and provided more insight into the dye-molecule couple dependency. Furthermore, the impurity was produced as a diastereoisomer mixture which could be isolated and exhibited significantly different polarization properties, demonstrating the critical importance of parameters else than magnetic to the obtention of high polarization yields.

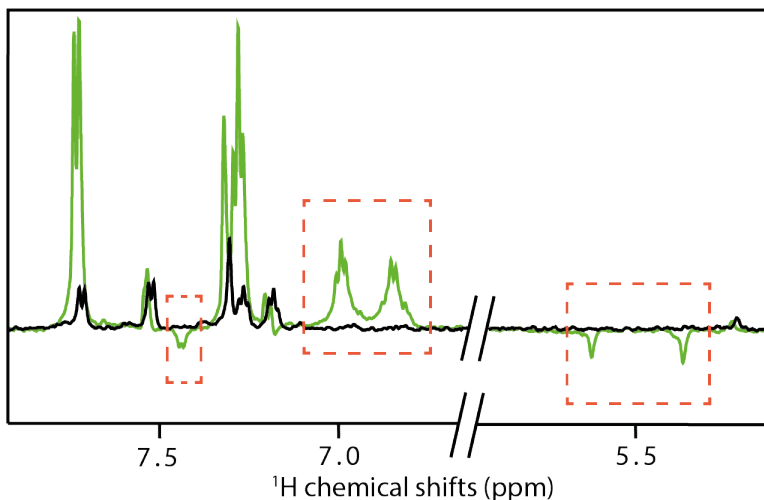


Figure 5.5: Unexpected polarisation of a trace degradation product. Photo-CIDNP of 1-day old 1 mM tryptophan sample measured by ^1H 1D NMR at 14.1 T (i.e. 600 MHz ^1H NMR frequency), irradiated in presence of 25 μM AT12 with a 1 W 532 nm laser. The dark spectrum is in black and the irradiated spectrum is in green. The red dashed boxes emphasize the highly polarized signals from an unknown molecule.

Dye investigation

While fluorescence spectroscopy reports many different dyes absorbing within a broad range of wavelengths, photosensitizers used for photo-CIDNP are relatively few and absorb in the UV-blue range: bipyridyl (308 nm),¹⁹² TCBP (308 nm),⁷⁶ FMN (450 nm).¹⁹³ The finding of performant photo-CIDNP active dyes that can be excited for different wavelength compatible with cheap commercial laser would have an impact on the field. Furthermore, UV lasers are unsuitable for CW-photo-CIDNP as they damage the optic fiber, and many biomolecules absorb in this range, contributing to the optical density and their eventual degradation. The use of redshifted wavelengths allows prolonged sample irradiation, thus maximizing the overall polarization. Moreover, dyes absorbing in the near-infrared region are biocompatible as they match the optical windows allowing potential *in vivo* applications. Whenever, the molecular basis correlating for

red-shifting and photo-CIDNP would be characterized this will help to discover near-infrared photo-CIDNP photosensitizers.

It was demonstrated in the introduction that the properties of the dyes, such as absorbance, quantum yield and triplet state lifetime were critical to the polarization efficiency.¹⁹⁴ Furthermore, it is important to adjust the photosensitizer according to the molecule to be polarized. Borrowing from the intense and highly successful science in single molecule fluorescence spectroscopy we investigated and compared the enhancement effect for a number of dyes. The performances of classic dyes such as bipyridyl and FMN was compared to xanthene dyes such as fluorescein, eosin Y, eosin B, erythrosine, and rose Bengal. These last dyes are all sharing a common aromatic system substituted with different groups. These substitutions are of interest for photo-CIDNP as the introduction of halogens yield to the internal heavy atom effect which increases the quantum yield of the triplet state (Table 5.1).¹⁹⁵ Also, these substitutions potentially to have an impact on the *g*-factor and could improve the performances. Furthermore, the introduction of a dye related to the xanthene with the oxygen in the central ring replaced by a sulfur atom, AT12, exhibited improved photo-CIDNP SNE, while moderate quantum yield improvement as compared to the halogen substituted xanthene dyes. The chemical modification of the xanthene aromatic moiety leads to a higher ISC rate and a slightly red shifted absorption spectrum $\lambda_{\text{max}} = 580$ nm. The redox potentials of the small molecules are also expected to be strongly affected by the chemical modifications. Assessment of the photosensitizer performances was done with tyrosine as a reference system a small screening of dyes with different ISC rates was conducted and we found AT12 to be a very promising sensitizer. In Figure 5.6 the absorption spectrum is shown, with the laser wavelengths used during this study.

With its maximum at 581 nm it fits well to a number of currently available laser diodes (LD) and the 532 nm diode pumped solid state laser (DPSSL), respectively. AT12 monitored Photo-CIDNP provided high tyrosine SNE, and relatively low tryptophan SNE (Figure 5.7). This difference has been interpreted as the expression of the value of *g*-factor of AT12. Also, the effect of the importance of the oxygen scavenging have been observed once more. The photo-CIDNP 1D ¹H NMR spectrum achieved with the dye AT12 (20 μ M) and tyrosine (0.1 mM) using a green laser (532 nm, 1 W during 4 s) at 298 K in a 100 mM phosphate buffer containing the enzymatic cocktail of glucose oxidase, catalase, and glucose in order to scavenge oxygen (see Material and Methods for more details). Typical for tyrosine the ¹H_{2,6} protons show a moderate enhancement, while the signals from the protons ¹H_{3,5} are strongly emissive with an up to 37-fold increase in peak intensity (Figure 5.7). Also, the aliphatic protons show a considerable enhancement.

In Table 5.1 the enhancement factors of the cross peaks of the ¹H_{3,5} tyrosine resonances for different sensitizers are listed including the well-known bipyridyl, FMN, and fluorescein. The dye AT12 shows an almost four and two-fold improvement with respect to the present “working horses” FMN and fluorescein, respectively. The other dyes tested rose Bengal and erythrosine show rather little enhancement albeit a high ISC, while also eosin with an ISC comparable to AT12 shows a signal enhancement in the order of 10. In addition, we conducted photo-CIDNP experiments with tryptophan. The photo-CIDNP performances of tryptophan with AT12 are extensively described in Table 5.2. The effect is very strong for the aromatic protons with high spin-density such as H₄ and H₆, and also a strong emissive effect is observed for the aliphatic protons H _{β} .

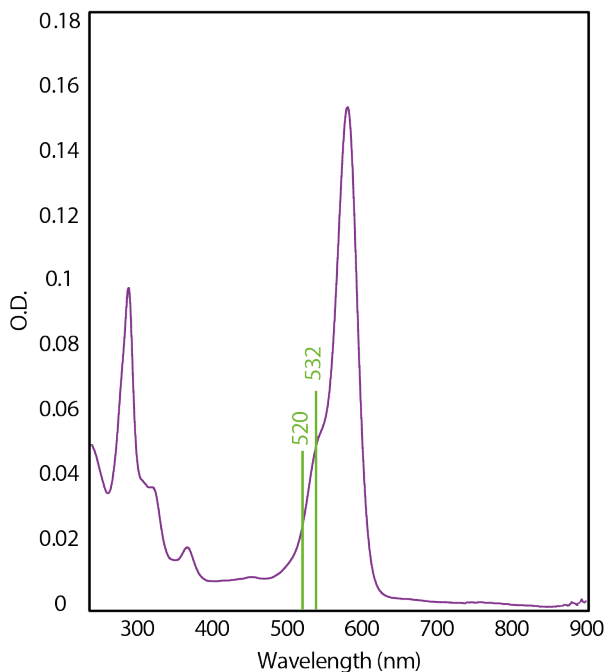


Figure 5.6: The absorption spectrum of AttoThio12 (AT12) along with the wavelength positions of currently available laser diodes (LD) and the green diode pumped solid state laser (DPSSL) wavelength (520 and 532 nm).

We concentrated on this first step by testing dye molecules with different ISC rates elucidated from single molecule fluorescence spectroscopy¹⁹⁶ and investigated and compared their CIDNP enhancement effects on tyrosine. We found AT12 to be a very promising sensitizer because it is at least twice as strong as the mostly used dyes bipyridyl, FMN, and fluorescein. However, there appears to be no correlation between a high ISC rate and CIDNP enhancement. Photo-CIDNP works for a variety of ISC rates/probability from 1 for the known systems like bipyridyl and FMN to 0.01 for the recently investigated fluorescein. While other dyes like rose Bengal and erythrosine also have a high ISC rate with almost no photo-CIDNP activity. Sensitizers with moderate ISC rate/probability of 0.2 like eosin and AT12 and even fluorescein with a low ISC probability of 0.01 are photo-CIDNP active with outstanding properties of AT12. Interestingly, it was recently demonstrated that the internal heavy atom effect was not showing improved photo-CIDNP performances, while external heavy atom effect using 3-iodo-1-propanol showed improved photo-CIDNP polarization of tryptophan with fluorescein.¹⁹⁵ Altogether, this suggests that the introduction of halogens to the dye are deleterious for photo-CIDNP. Obviously, the subsequent steps in the photo-CIDNP process including the redox potentials-involved electron transfer reaction¹⁹⁷ as well as the spin selecting recombination processes involving the g-factors, hyperfine couplings (HFCCs), and the magnetic field B_0 determine the photo-CIDNP activity most prominently. When compared within the example of tyrosine neither the magnetic field nor the HFCC of tyrosine are variables and thus only the g-factors and the electron transfer reaction rate could be a source of the outstanding performance of AT12. Given that the AT12 quantum yield

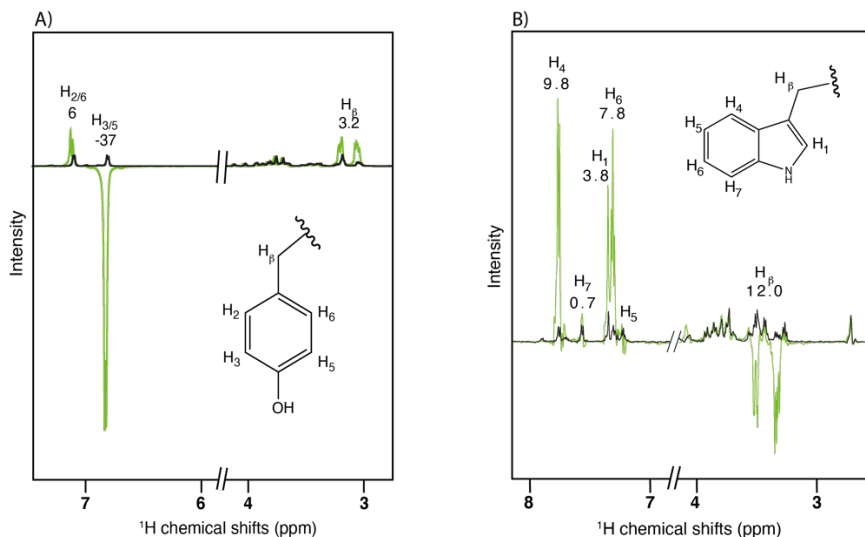


Figure 5.7: Photo-CIDNP ¹H-NMR spectra of 100 μM tyrosine (left) and 100 μM tryptophan (right) in the presence of 20 μM of AT12. Dark spectra are in black color and irradiated spectra are in green color, respectively. Irradiation was performed with a wavelength of 532 nm, during 4 s at power 1 W.

is not higher than FMN or bipyridyl and the lifetime of the triplet state is relatively short, this supports the hypothesis of a high photo-CIDNP efficiency due to an effective electron transfer. Alternatively, or in addition the Δg of the radical pair tyrosine-AT12 might be more optimized (smaller) than for the radical pair tryptophan-AT12 since the enhancement for tyrosine is ~3-fold higher (i.e. 38 over 18) when compared with tryptophan while fluorescein shows a 50-fold enhancement for tryptophan versus 20 for tyrosine. Moreover, the g-factor of the AT12 is expected to be higher as of the introduction of the sulfur atom to the aromatic ring which brings a higher spin orbit coupling, which is consistent with preferred polarization of the tyrosine. In summary, AT12 is a new efficient sensitizer for the photo-CIDNP of tyrosine and tryptophan giving a hint that dye-molecule pairs for increasing NMR polarization by several orders of magnitude may be out there.

Chemical tuning of photo-CIDNP performances

The serendipitous observation of an unexpected polarization while measuring the AT12 monitored CW-photo-CIDNP on the amino acid tryptophan⁷⁴ put into motion a study towards the characterization of this impurity. After a series of photo-CIDNP measurements on the same tryptophan-AT12 sample, a number of new peaks appeared in the light irradiated spectra that remained undetectable in the dark spectra, evidence of an exceptional polarization.

Table 5.1: Dye-dependent enhancement of the $^1\text{H}_{3,5}$ NMR cross peaks of tyrosine.

Dye	Oxygen	Oxygen Scavenged	Laser/Power	ISC (s^{-1})
<i>Bipyridyl</i>	-2.0	N/A	Excimer 308 nm / 1W	$1.5 \cdot 10^{10}$ ¹⁹²
<i>FMN</i>	-4.0	-11.0	LD 405 nm / 1W	$130 \cdot 10^6$ ¹⁹⁴
<i>Fluorescein</i>	0	-20.0	LD 450 nm / 1W	$3.7 \cdot 10^6$ ¹⁹⁴
<i>Eosin Y</i>	-5.0	-8.6	DPSSL 532 nm / 1W	$420 \cdot 10^6$ ¹⁹⁸
<i>Eosin B</i>	0	-0.4	DPSSL 532 nm / 1W	$100 \cdot 10^6$ ¹⁹⁸
<i>Erythrosin</i>	0	-1.8	DPSSL 532 nm / 1W	$1100 \cdot 10^6$ ¹⁹⁸
<i>Rose Bengal</i>	0	-1.0	DPSSL 532 nm / 1W	$800 \cdot 10^6$ ¹⁹⁸
<i>AT12</i>	-1.5	-38.6	DPSSL 532 nm / 1W	$118 \cdot 10^6$ ¹⁹⁶

Interestingly, traces of this photo-CIDNP-active product can also be found in the literature, notably in Okuno et al.,¹⁹⁵ where a photo-CIDNP active impurity is visible for the spectrum of irradiated tryptophan in the presence of fluorescein and 3-iodo-1-propanol. The highly polarized impurity was obviously suspected to be a photoinduced degradation product of tryptophan. Indeed, we identified it as an oxidocyclization product of tryptophan that results from the generation of $^1\text{O}_2$ by the triplet state of AT12. Subsequently, the resulting 3α -hydroxypyrrroloindole (HOPI) was successfully synthesized, characterized and its photo-CIDNP activity measured. Moreover, the combination of CW-photo-CIDNP, TR-photo-CIDNP experiments and density functional theory (DFT) simulations provided valuable information about the reasons for such an improvement of the signal enhancement in the 3α -hydroxypyrrroloindole as compared to tryptophan. The relevant parameters for CIDNP such as the Δg and the HFCCs were related to the observed anomalous lines in TR-photo-CIDNP experiments. Furthermore, insights into the reaction kinetics of the photo-CIDNP reaction emphasized the possible importance of conformation, and the molecular geometry for photo-CIDNP activity. All these mechanistic studies were done in order to evaluate on the potential of using CW-photo-CIDNP as a signal enhancement tool in biologically and chemically focused NMR.

Characterization of HOPI and synthesis

In order to identify the molecule that showed the large polarization enhancement depicted in Figure 5.5, the irradiated samples in which the highly polarized impurities could be observed were pooled together and purified by reversed-phases HPLC. Several peaks from the chromatogram were collected separately and screened by NMR with the photo-CIDNP experiment.

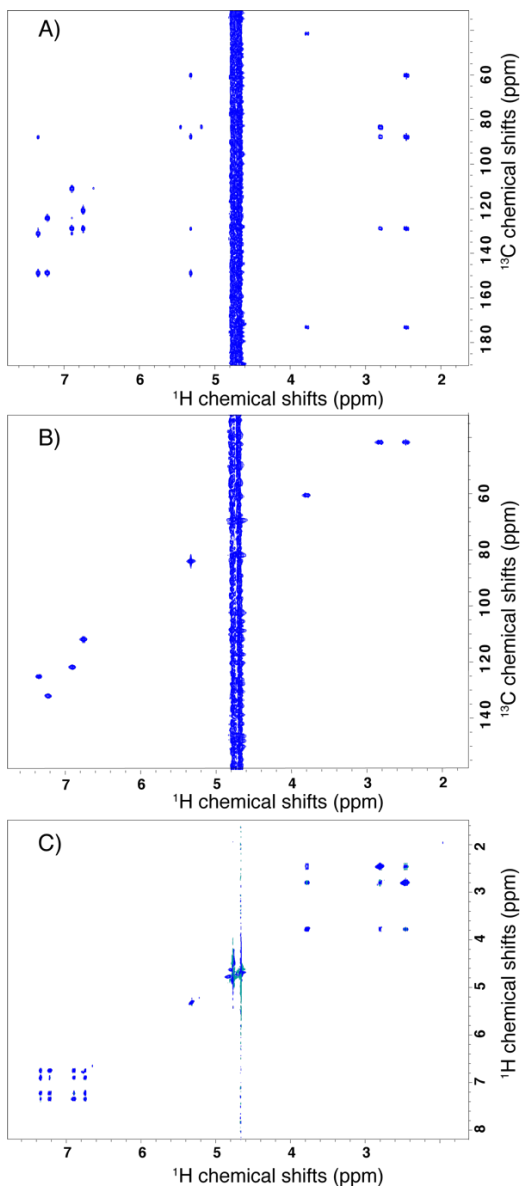


Figure 5.8: The HPLC purified fractions were analyzed with a set of state-of-the-art NMR experiments that enables one to establish the C-C and C-H connections and proximity and for the identification of the protons that belong to a scalar-coupled spin system A) Heteronuclear Multiple Bond Correlation (HMBC) B) Heteronuclear Multiple Quantum Correlation (HMQC) C) $1\text{H}, 1\text{H}$ -Total Correlated Spectroscopy (TOCSY).

More than one HPLC peak had significant photo-CIDNP signal enhancement and after identifying the peak with the highest photo-CIDNP signal enhancement, state-of-the-art NMR experiments were used to characterize its chemical structure, namely HMBC, HMQC, and TOCSY (Figure 5.8). In addition, the product was characterized by ESI-TOF mass spectrometry yielding a positive m/z of 221 (not shown). It is well known that triplet state dyes can relax to their ground state through the reaction with oxygen forming the very oxidative singlet oxygen. As the literature provided many characterized tryptophan oxidation products as potential candidates for the unknown molecules,¹⁹⁹⁻²⁰¹ those species were investigated further. The m/z ratio suggests the addition of an oxygen atom ($m/z(\text{tryptophan}) + 16$), and the 1D $^1\text{H-NMR}$ spectrum shows that all the hydrogens on the benzo group remain, eliminating all the hydroxy-tryptophans as candidates (Figure 5.9). The HOPI depicted in Figure 5.10 is consistent with the NMR data and the m/z ratio from MS. The reaction mechanism leading to HOPI involves an endoperoxide intermediate that can form upon singlet oxygen addition on both sides of the double bond of the pyrrole moiety of tryptophan. The direct consequence of this, is the presence of two diastereoisomers when the nitrogen of the amine group opens the endoperoxide following an $\text{S}_{\text{N}}2$ mechanism. This last step leads to cyclization. The reaction mechanism is well studied in the literature and is represented in the figure 5.11A.

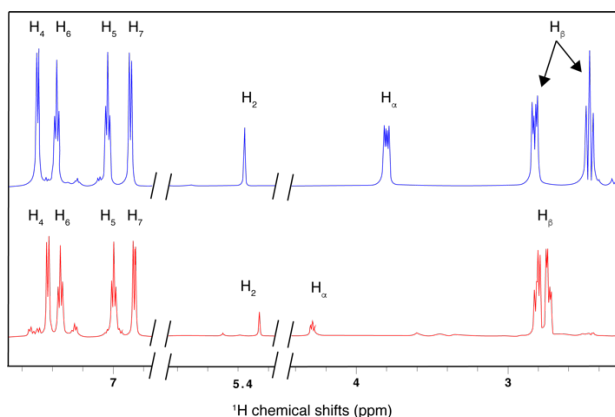


Figure 5.9: 1D ^1H NMR spectra of both diastereoisomers of HOPI with assignments. Upper spectrum is *cis*-HOPI, and lower is *trans*-HOPI. Both molecules were at 1 mM in 100 mM KPO_4 buffer, $\text{pH} = 7.1$. Spectra were recorded with 128 scans, 16k points acquisition and 3 s recovery delay.

The two diastereoisomers have been previously characterized.²⁰² The *trans* isomer is more polar than the *cis*, based on the C18 HPLC retention times (Figure 5.11C). Moreover, the chemical shifts associated to the different diastereoisomers were consistent with the ones previously described in the literature²⁰²; especially for the H_β and the H_α . Next, the HOPI diastereoisomer pair was synthesized according to the formerly described procedure²⁰³. The HOPI synthesis has been already described in the literature. In our case we used a LASER (532nm, 0.5W) as the irradiation source to generate $^1\text{O}_2$ from rose Bengal (RB), the solution was saturated in O_2 by bubbling oxygen through the solution. The reaction mixture was kept under stirring in a cold room (4°C) for 4 hours.

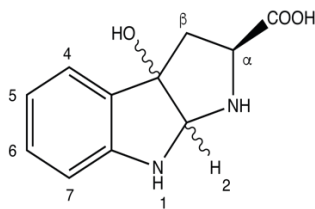


Figure 5.10: Structure of the *Cis/trans*-3 α -hydroxypyrroloindole

Then hydroperoxide is reduced by adding dimethyl sulfide (Me_2S) and stirring at room temperature for 30 min. Finally, the RB was removed from the reaction mixture via C18 solid phase extraction (SPE) in place of a liquid-liquid extraction. The sample was not acidified for SPE so that the RB remained visible and it could be monitored visually on the SPE column. The final purification was done on a semi-preparative C18 column (Figure 5.11 C). The synthesized products exhibited the same NMR spectra and polarization properties during photo-CIDNP experiments as observed for the degradation products of tryptophan (compare Figure 5.5 with Figure 5.12A and B).

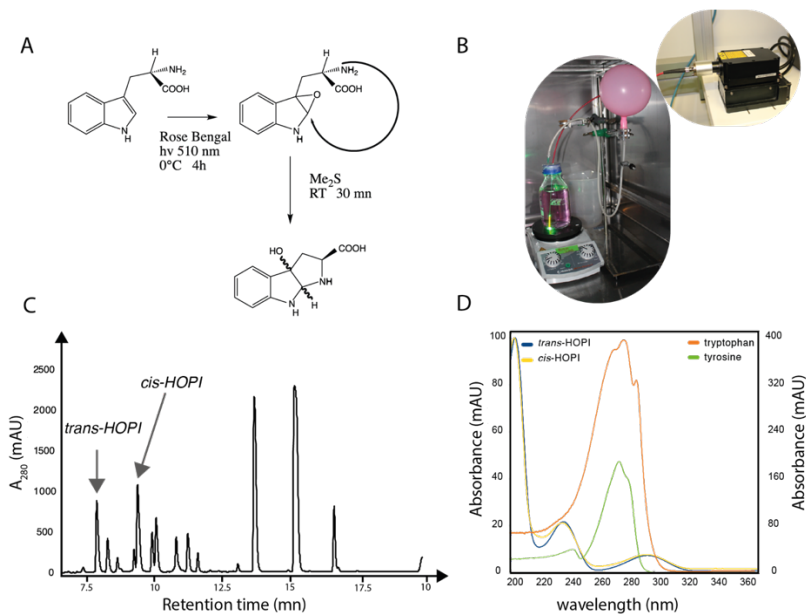


Figure 5.11: A) Reaction mechanism for the formation of the *cis/trans*-HOPI. B) Set up for the photo-oxidation of tryptophan. C) HPLC purification profile of HOPI from the tryptophan photo-oxidation reaction. D) UV absorption spectra for both diastereoisomers of HOPI compared to tryptophan and tyrosine. The maximal intensities were normalized to 100 mAu for visibility reasons. However, the tryptophan and tyrosine concentrations were 25 μM and the *cis/trans*-HOPI concentrations were of 50 μM . The left y-axis is referring to the *cis*- and *trans*-HOPI, and the right axis to the tryptophan and tyrosine.

Photo-CIDNP performances in the light of magnetic parameters

The oxidocyclization of tryptophan into HOPI greatly affects its CW-photo-CIDNP performances. Of particular interest is the strong AT12-based polarization found for ^1H with a factor of 73 for *trans*-HOPI and 68 for *cis*-HOPI at 100 μM HOPI concentration in a 14.1 T magnetic field (i.e. 600 MHz ^1H NMR frequency). Similarly, polarization in the presence of fluorescein yields a factor of -70 for *trans*-HOPI and -47 for *cis*-HOPI (Figure 5.12A and B, Table 5.2). Interestingly, both HOPIs show a strong concentration dependence of the signal enhancement but display contrary behaviors for the two diastereoisomers. At the lower concentration of 20 μM , *trans*-HOPI shows an enhancement of 120 for the H_5 resonance whereas for *cis*-HOPI it is only ca. 40 (section kinetics Figure 5.17A), while at concentration of ca 0.5 mM both show a similar enhancement of ca 60. At 4.7 T (200 MHz) the *trans*-HOPI is enhanced by 380-fold in the presence of AT12, and 160-fold in the presence of fluorescein (Table 5.2). The signal enhancement for the *trans*-HOPI at 4.7 T is illustrated in Figure 5.12C. In contrast to the single-scan photo-CIDNP 1D ^1H NMR experiment the reference dark spectrum with a factor of about three-fold less SN took 16 h to acquire with 11520 scans, a time difference of five orders of magnitude.

Table 5.2: Photo-CIDNP ^1H signal enhancement with different dyes and magnetic fields

Molecule	AT12	Fluorescein	AT12	Fluorescein
	at 14.1 T (600 MHz)	at 14.1 T (600 MHz)	at 4.7 T (200 MHz)	at 4.7 T (200 MHz)
<i>trans</i> -HOPI	73	-70	380	-160
<i>cis</i> -HOPI	68	-47	N.D.	N.D.
Tryptophan	18	53	90	225

The signal-to-noise enhancement factors correspond to the strongest aromatic resonance of HOPI (H_5) and tryptophan (H_4), both at 100 μM .

While tryptophan is typically better polarized with fluorescein than AT12, the *trans*-HOPI and to a lesser extent the *cis*-HOPI are equally hyperpolarized after irradiation in the presence of AT12 or fluorescein dyes (Table 5.2). Furthermore, the polarization sign for HOPI is inverted with fluorescein compared to AT12 (Figure 5.12A and B). Moreover, the pattern of the signal enhancement is perturbed. The photo-CIDNP spectrum of tryptophan with AT12 shows only absorptive lines for the aromatic protons and emissive lines for the H_β (Figure 5.7B), while the HOPI spectrum presents both absorptive and emissive for the aromatics and is emissive for the newly aliphatic H_2 (Figure 5.12A and B). The finding that dyes react preferentially with some molecules over others, yielding to preferential polarization, has already been reported. This is in particular the case for fluorescein and tryptophan which are a better pair than fluorescein and tyrosine¹⁹⁴ and conversely for tyrosine which is more efficiently hyperpolarized with AT12 than tryptophan.⁷⁴ In the high-field limit for fast tumbling molecules this difference is attributed to the offset in the Zeeman splitting (Δg) of the two single electrons of the respective dye and molecule

(equation 5.6), while the changes in sign and relative intensities are explained by the changes in the HFCC values upon chemical structure modification.

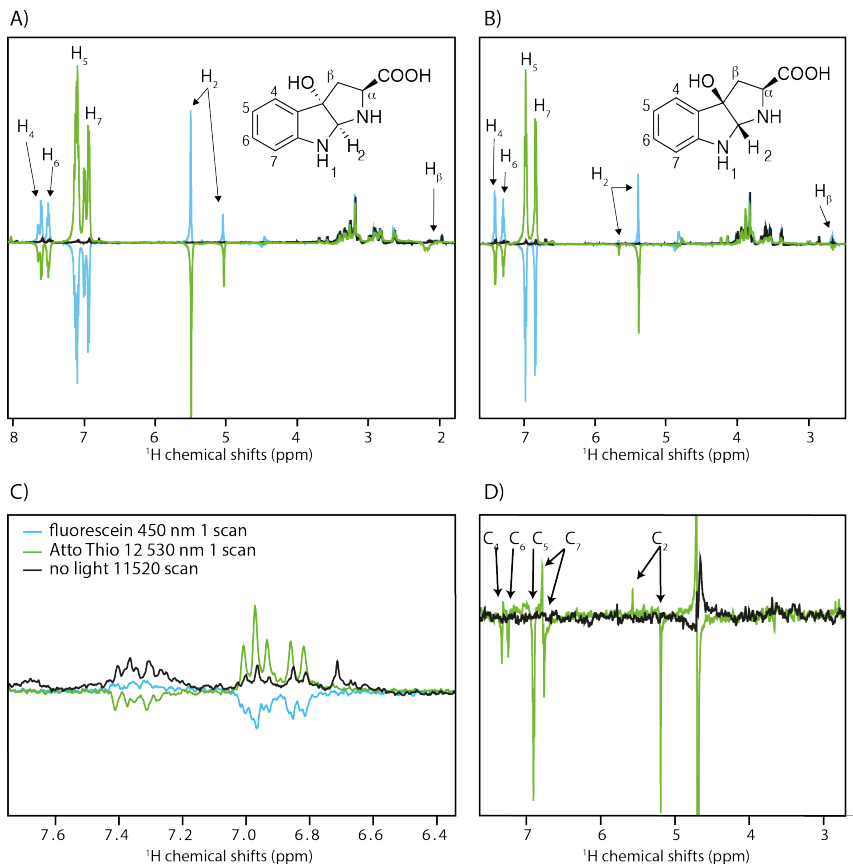


Figure 5.12: Photo-CIDNP ^1H 1D NMR spectra of HOPI, 100 μM , in the presence of AT12, 20 μM , (green lines) or fluorescein, 25 μM (blue lines). In black are the dark ^1H 1D NMR spectra. The irradiation was performed for 4 s at 1 W power level, and wave lengths of 450 nm for fluorescein and 532 nm for AT12. A) trans-HOPI at 600 MHz and B) cis-HOPI at 600 MHz. Note that trans-HOPI displays two H₂ and H₇ signals due to different conformations. C) Trans-HOPI at 200 MHz. The dark spectrum was accumulated over 11520 scans. For representation purposes the spectra were normalized to the noise level, hence the dark spectra are scaled by a factor of 0.0093 which correspond to $1/\sqrt{11520}$. D) Photo-CIDNP of ^{13}C detected on the attached ^1H of natural abundance trans-HOPI at 600 MHz. The dark spectrum (no light) was accumulated over 2560 scans. For representation purposes the spectra were normalized to the noise level, hence the dark spectrum is scaled by 0.02 which corresponds to $1/\sqrt{2560}$. The (very weak) signal at 3.8 ppm of the dark spectrum is due to the sugars, which are at a concentration of 2.5 mM (Supplementary), and were used for the cross-calculation of the signal enhancement.

Photo-CIDNP of ^{13}C resonances of HOPI

It is important to note that the here documented polarization enhancements are for ^1H while it is common that the polarization enhancements are reported for ^{13}C or other nuclei, which have either a lower gyromagnetic ratio or larger HFCC when compared with ^1H . The anisotropy of the ^{13}C HFCC needs to be considered for a proper analysis since their anisotropic character is much more pronounced than for hydrogen. For the carbons C5, C7, and C2 we could calculate HFCC values for the z-axis (in the molecular frame) of 2.51, 2.23, and 2.79 mT, respectively. Thus, for the following three reasons it is expected that HOPI will display an even stronger enhancement for ^{13}C : 1) its gyromagnetic ratio is 3.98-fold lower than that of ^1H , 2) the ^{13}C HFCCs are at least a factor of 2 higher than for the protons (Table 5.4), 3) the longitudinal relaxation of ^{13}C is slower than ^1H yielding less signal decay during the CW irradiation. Hence, at least a factor of 8 more is expected. Indeed, a ^{13}C signal enhancement of 1217-fold is estimated in a 1D single scan photo-CIDNP experiment on *trans*-HOPI and AT12 performed on a 600 MHz spectrometer starting with ^{13}C (steady state) magnetization that is detected on its covalently bounded ^1H via a refocusing INEPT²⁰⁴ (Figure 5.12D). In this case, only an estimation was possible since the *trans*-HOPI sample contained natural abundance ^{13}C , yielding a dark reference spectrum of extremely low signal intensity. It is noted that the estimated signal enhancement equals a 10^6 -fold reduction in measurement time. The irradiated spectrum measured in 1 scan (5 seconds) would have required $1'440'000$ scans for the dark spectrum which corresponds to 2000 hours of measurement ca 83 days.

g-factors differences (Δg)

In the following, the Zeeman splitting, the HFCC values as well as reaction kinetics of both HOPI diastereoisomers are investigated in some detail starting with Δg . As shown above in the photo-CIDNP theory chapter, in the high-field limit for fast tumbling molecules (i.e. < 1 ns), the singlet-triplet state mixing is mainly driven by the Δg . This parameter is resulting from the spin-orbital coupling of the unpaired electron with the SOMO. The oxidocyclization of the HOPI transforms the aromatic system from an indole ring to an aniline-like aromatic ring after the conversion of the pyrrole part of the indole group into a pyrrolidine. This modification of the shape and expansion of the SOMO is suggested to be the reason for the different *g*-factor values of tryptophan and HOPI. The *g*-factor calculations were obtained using density functional theory (DFT) calculations performed with Gaussian®. The geometries were optimized up to the 6-311+G(2d,2p) basis set. Implicit water solvent was used as well (PCM). Then computed with the *nmr = giao* keyword. This approach has been published elsewhere, showing good performances for the computation of *g*-factors of small molecules²⁰⁵. The calculated values are given in the Table 5.3. However, it is important to keep in mind that *g*-factor calculations are typically rather inaccurate. We could obtain a reasonably close value to the literature value for tryptophan (2.0027)²⁰⁶ with our method, but these values should be looked at with a qualitative rather than strictly quantitative eye. Nevertheless, the computed values of *g*-factor are consistent with the sign alternation observed when the dye is changed from fluorescein to AT12. However, since TCPB with a *g*-factor of 2.0035 behaves like AT12, HOPI's *g* factors are rather in the vicinity of a little bit lower than 2.0035 (but larger than 2.0034, the *g*-factor of fluorescein). As it is demonstrated later in this chapter, HOPI forms a radical pair by proton coupled electron transfer (PCET), and therefore the HOPI radicals used for the calculation are deprotonated on their aromatic nitrogen.

Table 5.3: *G*-factors (unitless) were calculated with unrestricted B3LYP, 6-311+G(2d,2p) or aug-cc-pVTZ basis set and implicit water solvation (PCM) for the *cis/trans*-HOPI, and cationic tryptophan (TRP) radicals.

Molecule	6-311+G(2d,2p)	Aug-cc-pVTZ
<i>trans</i> -HOPI	2.0035527	2.0035566
<i>cis</i> -HOPI	2.0035831	2.0035872
TRP ⁺	2.0028128	2.0028144

An experimental support of the *g*-factor modification through the chemical reaction is the sign alteration of the HOPI photo-CIDNP when the dye is changed from fluorescein to AT12 (Figure 5.12A and B). According to the Kaptein rules,²⁰⁷ this polarization sign alteration is explained by the dye-dependent sign alteration of the radical pair Δg . Indeed, no other conditions are changed between the two experiments except the dye. Furthermore, the observed magnetic field-dependence of the enhancements support the important role of the Δg in the photo-CIDNP signal enhancement. At 4.7 T (i.e., 200 MHz), the H₅ signal of *trans*-HOPI is enhanced 380-fold in the presence of excited AT12 (triplet at 6.98 ppm), and -160-fold in the presence of excited fluorescein (Table 5.2). The tryptophan signal is enhanced by 225 and 90-fold in the presence of fluorescein and AT12, respectively. The field dependent normalized photo-CIDNP polarizations have been calculated according to equation 5.8, and plotted into the Figure 5.13. The experimental signal enhancement values were normalized by the relative sensitivity of the 600 MHz (14.1 T) and 200 MHz (4.7 T) NMR spectrometers to obtain the normalized polarizations, and were fitted to the calculated curves. The photo-CIDNP polarizations monitored AT12 dye (Figure 5.13A) had to be calculated for the different possible *g*-values estimated from the photo-CIDNP anomalous line sign alternation of HOPI, and the *g*-values of tyrosine and tryptophan⁷⁴. The fitting of the experimental polarization to the calculated curves provided satisfactory results for all the possible *g*-values of AT12 except 2.0036, thus refining the range of *g*-values for AT12 from 2.0037 to 2.0040. Moreover, the optimal field for HOPI is strongly shifted to high field as compared to the tryptophan, which is explained by the cooperativity of the smaller Δg and higher HFCCs as compared to the tryptophan. The slightly higher field dependence of the AT12 monitored photo-CIDNP of HOPI as compared to tryptophan can be explained by this shift. The fluorescein monitored photo-CIDNP field dependence polarizations of HOPI is shifted towards higher field values, due to the particularly low difference in *g*-values (Figure 5.13B). Therefore, the optimal field for the HOPI-fluorescein photo-CIDNP polarization is 9-10 T, which would correspond to a 400 MHz spectrometer. This surprising finding is corroborated by the relatively lower polarization of HOPI at 200 MHz as compared to 600 MHz. In comparison, the tryptophan-fluorescein photo-CIDNP couple has a higher Δg and presents a slight increase of the polarization at 200 MHz as compared to 600 MHz. All these findings together show a stronger field-dependence of the AT12-based enhancement compared that of fluorescein. This dye-specific field-dependence is in great part explained by the Δg , the difference of the *g*-factor of the dye to the *g*-factor of the radicals. In principle, at lower fields, the HFCCs have a greater influence and the polarization resulting from each spin sorting step is higher and reaches a maximum when the

condition $\mu_B \Delta g B_0 = \text{HFCC}/2$ is fulfilled. However, as it is appearing with the HOPI-fluorescein couple, this greater influence of the HFCC can be observed at higher field if the Δg is small enough (equation 5.10). Fluorescein has a smaller Δg with tryptophan than AT12 as demonstrated elsewhere.¹² Furthermore, based on our calculation, the Δg of the fluorescein-HOPI pair is expected to be very low and on the order of 0.0001 and thus its effect on the polarization is not very magnetic field sensitive, while the Δg of the AT12-HOPI pair is several times larger yielding a stronger magnetic field dependence of the polarization.

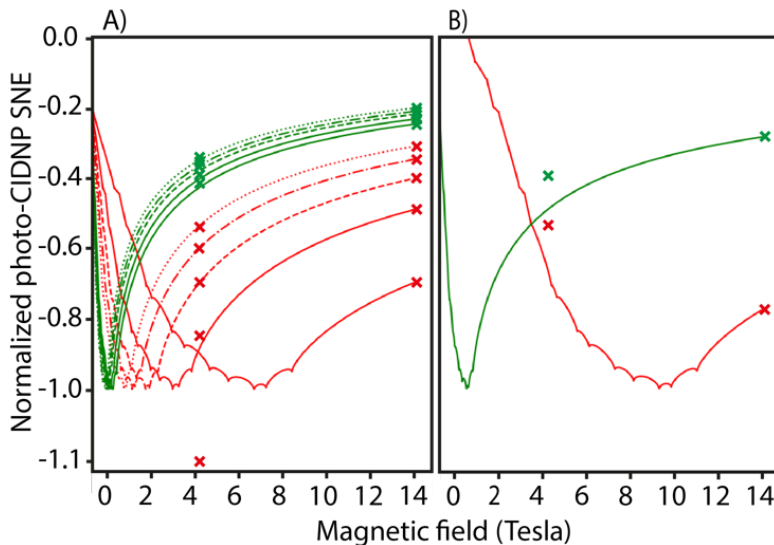


Figure 5.13: Field dependent normalized polarization of tryptophan (green) and HOPI (red) A) Monitored with AT12, dotted ($g_{\text{AT12}}=2.0040$), dash-dot ($g_{\text{AT12}}=2.0039$), dashed ($g_{\text{AT12}}=2.0038$), solid downfield ($g_{\text{AT12}}=2.0037$), solid upfield ($g_{\text{AT12}}=2.0036$). B) fluorescein

Whether the higher quantum yield of AT12 compared to fluorescein is also an important factor in the magnetic field dependence of polarization is open for discussion.^{194, 196} Altogether, this confirms the calculated g-factors for the HOPI radicals in the vicinity of 2.0035. Considering the sign of the anomalous lines, the g-factor of the HOPI radicals have been estimated above fluorescein's g-factor (2.0034) and below AT12's (between 2.0034 and 2.0041). However, the sign of the polarization when the experiments are conducted with TCBP dye with a g-factor = 2.0035 (Figure 5.14) enables to place the g-factor value of HOPIs in between 2.0034-2.0035. Furthermore, the polarization of HOPI at 200 MHz in the presence of TCBP is 3-fold lower as compared to tryptophan, which is expected since the Δg of HOPI with TCBP is estimated smaller than the Δg of HOPI with fluorescein, and shifts the optimal magnetic field to higher values.

Finally, since the Δg between HOPI and fluorescein is lower than the Δg between HOPI and AT12, one would expect the HOPI fluorescein couple yield to higher signal enhancements while considering only the magnetic parameters. However, it is the AT12 which provides the best performances as a photo-CIDNP sensitizer for HOPI, suggesting that other parameters than magnetics are to be considered to address the polarization quantitation.

Hyperfine couplings

Despite the fact that triplet-singlet state mixing of the radical pair electron spins is mainly driven by the Δg in fields above 1 Tesla (high-field), the spin sorting, and thus the hyperpolarization is the result of the HFCCs. These coupling constants are not field-dependent and are specific for each nucleus within the radical. The HFCC values depend on the Fermi contact interaction between the nucleus and the unpaired electron, and on the spin density.²⁰⁸ As a consequence of the differences in the conjugated systems of tryptophan and HOPI, their spin densities are different. Unlike the absorptive signals of the aromatic protons of tryptophan, the aromatic anomalous signals are emissive and absorptive for HOPI. Moreover, the H₂ proton which was aromatic in the tryptophan is now in the α -position of the conjugated system and has an intense emissive signal after irradiation in presence of AT12. To get further insights into the HFCC constants of HOPI with an emphasis on understanding which of the chemical features are important to generate a highly polarizable molecule, HFCCs were computed from DFT simulations run in Gaussian (Table 5.4). The geometries were first optimized with a quick molecular dynamics refinement using the UFF, and then with an unrestricted B3LYP functional and a 6-31+G(d,p) basis set in implicit water solvent (Polarizable Continuum Model, PCM). The HFCCs were calculated with the EPR-III basis set.

Table 5.4: Hyperfine coupling constants (mT) calculated for the radicals of *trans*-HOPI, *cis*-HOPI and cationic tryptophan (TRP)

Molecule	H2	H H5 4	H6	H7	Hb	Ha	
<i>trans</i> -HOPI	1.047	0.229	-0.850	0.156	-0.498	0.141 -0.022	0.067
<i>cis</i> -HOPI	1.273	0.231	-0.853	0.138	-0.480	-0.021 0.035	0.014
TRP ⁺	-0.177	-0.168	0.028	-0.105	-0.061	0.364 0.016	-0.083

In order to determine experimentally the HFCC and mechanistic insights (see below) TR-photo-CIDNP is necessary because in a CW-photo-CIDNP experiment the predominance of F-Pair over geminate polarization (slightly) deviates the total polarization and only the geminate polarization is purely related to the HFCC.²⁰⁶ In contrast to CW-photo-CIDNP studies TR-photo-CIDNP allows to determine the reaction mechanism and the hyperfine coupling constants in transient radicals. The NMR experiment was performed as following: a pre-saturation (WALTZ16) sequence prior to light irradiation by a 308 nm excimer laser and the CIDNP spectra is detected without any delay after the single short laser pulse (with duration of a few nanoseconds) using a short radio frequency pulse (μ s). As shown previously,²⁰⁹ the relative signal intensities in the geminate spectrum are proportional to the HFCC in the radicals constituting the geminate radical pair. TCBP has been used as the photosensitizer for the following reasons: For TR-photo-CIDNP

the experimental conditions were changed using TCBP as the dye because the hyperfine couplings of its radicals have been determined from EPR²¹⁰ and CIDNP^{209, 211} because its radicals show clear CIDNP patterns, that are distinctly different for electron transfer (ET) and PCET reactions. The experiments were performed at 4 mM concentration of HOPI and at a pH 7.5 (Figure 5.14). Unfortunately, in the geminate CIDNP spectra of both compounds (Figure 5.14), *cis* and *trans* HOPI, an admixture of tryptophan was clearly visible. For elimination of the unwanted contribution from the reaction of TCBP with tryptophan to CIDNP reference geminate spectra from the reaction of TCBP with tryptophan and N-acetyl tryptophan have been used (Figure 5.14).

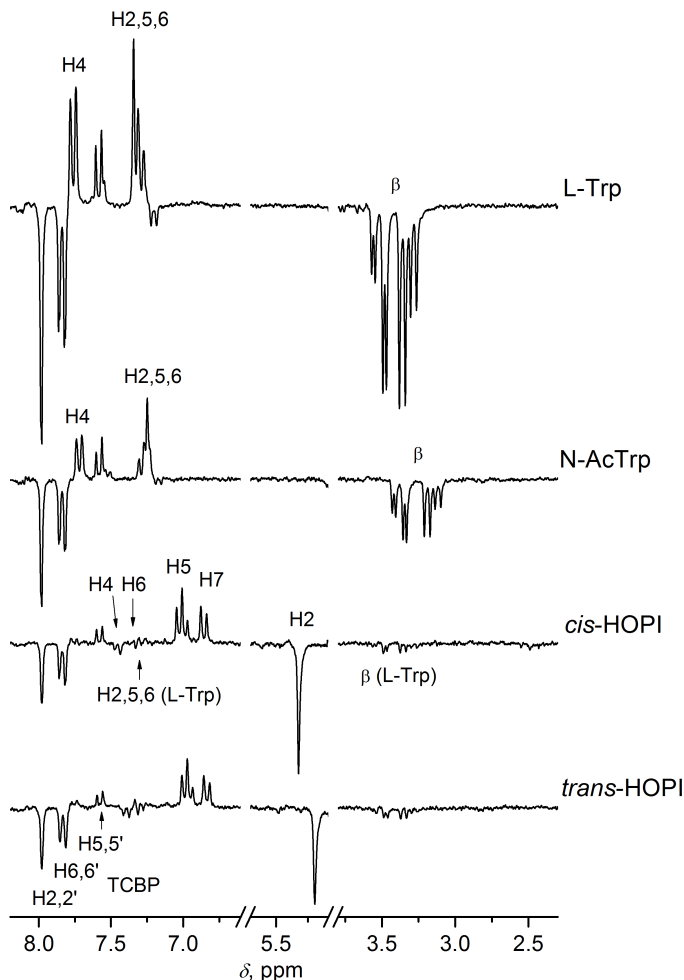


Figure 5.14: 200 MHz ¹H CIDNP spectra, obtained in the photoreaction between 4 mM TCBP and 4 mM tryptophan, 2 mM NAc-tryptophan, ~4 mM *cis*-HOPI and ~4 mM *trans*-HOPI at neutral pH. Spectra were recorded with no delay after the laser pulse

using a detecting RF pulse of 1 μ s, thus they contain only signals of the geminate reaction products.

Integration of the individual line intensities in the reference spectra allowed us to determine the relative contribution to CIDNP from tryptophan as 1:20 and to subtract it from the geminate CIDNP spectra of both HOPIs. The experimental HFCCs determined by the TR-photo-CIDNP anomalous corrected lines intensities (Figure 5.14) were then correlated with the calculated ones shown in Table 5.4 showing determination coefficients of 0.97 and 0.98 for *trans*-HOPI and *cis*-HOPI, respectively (Figure 5.15A and B). The good correlation between the experimental and the calculated HFCC, validates the values obtained with DFT calculations. It is thereby interesting to note that the H₂ HFCC was dramatically increased (i.e., by a factor of 5-6) by the aromaticity loss of the five membered-ring. The high HFCC of beta protons in tyrosine and tryptophan have already been predicted. However, because of the angular dependence of the HFCC,²¹² the rotation around the σ molecular bond was averaging out the expected hyperpolarization effect. From this observation, we conclude that in general, aliphatic protons in the α position to an aromatic system with hindered rotational mobility (i.e., in a cyclic aliphatic ring that) yield to high HFCC.

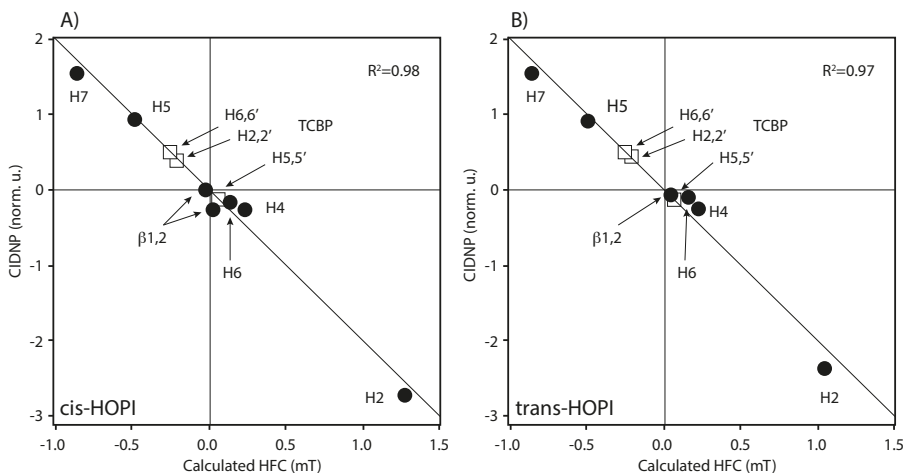


Figure 5.15: Correlation between the ¹H CIDNP intensities of the geminate products detected in the reversible photoreaction of TCBP and the calculated ¹H HFCCs of the *cis*-HOPI (A) and *trans*-HOPI (B) radicals (circles) and the combination of TCBP radicals I, III, IV (squares). Solid line: best fit by the function $P1i = -CA1i$ (fitting to circles), $P2j = -CA2j$ (fitting to squares). Concentration-dependent H5 Photo-CIDNP enhancement for *trans*- and *cis*-HOPI.

Since the HFCC of the TCBP dye are depending on its protonation state, the electron transfer mode can be addressed by analyzing the TR-photo-CIDNP anomalous line intensities.²⁰⁹ In the case of tryptophan or N-acetyl tryptophan the TCBP anomalous line intensities match with the unprotonated TCBP's HFCCs and thus a simple ET. However, the time resolved CIDNP spectra for *cis/trans*-HOPI yield a mix of protonated TCBP radicals as it can be observed from the correlation plot between the anomalous line intensities and the HFCCs of TCBP/H⁺ mix (Figures 5.15A and B, 5.16). This last observation suggests a PCET reaction mechanism for the HOPI radicals. The degenerate electron exchange takes place when the radical and the diamagnetic

molecule only differ by one electron. It is worth noticing that PCET is also the formation mechanism of tyrosyl radicals,²⁰⁹ which also shows improved signal enhancement when photo-CIDNP experiments are performed in the presence of AT12.⁷⁴ For TCBP, the experimentally determined values -0.252 mT (H6,6'), -0.209 mT (H2,2'), 0.067 mT (H5,5') were taken. These values correspond to the combination of radical T-I, T-III, and T-IV, which is indicative of PCET as the reaction mechanism (by contrast to the case of tryptophan, when ET results in correlation with the T-I radical alone).²¹³ T-I is the deprotonated TCBP, and T-III and T-IV are protonated form with one carboxy group protonated on position 3/3' and 4/4', respectively. The correlation plots are given in Figure 5.16. The radical species associated to ET or PCET and their corresponding apparent HFCC enable to evaluate the electron transfer mechanism, and are well documented.²¹¹

It is finally noted here that the impressive performances of HOPI in the CW-photo-CIDNP experiment using AT12 (Table 5.2) appears to be at odds to the TR-photo-CIDNP experiments in the presence of TCBP where the HOPI compounds are less enhanced as compared to tryptophan or N-acetyl-tryptophan. These apparent disappointing performances seems to be related to the different dyes and conditions used. About the latter, the TR-photo-CIDNP experiments are performed at mM concentration for which trans-HOPI shows less enhancement also with AT12 (Figure 5.20) and solely the geminate polarization is present as it is necessary for the fine analysis of the radical parameters. Moreover, the time resolved experiments were performed at pH = 7.5 resulting into 62% HOPI/H⁺ while the CW experiments were performed at pH = 7.1 resulting into 80% HOPI/H⁺. The photo-CIDNP effect has been observed to be 2.5 times higher when performed in more acidic conditions. Finally, it is expected that the optimal field for the TCBP-HOPI couple is way higher than the 4.7 T used for the TR-photo-CIDNP spectra. Therefore, the photo-CIDNP performances would be in the same regime as for the fluorescein-HOPI couple at 200 MHz (Figure 5.13B) and may be even lower with the assumption that the g-factor difference is smaller.

Diastereomeric-dependent reaction kinetics

At pH = 7.2 and at concentrations in the range of 0.1 mM to 0.5 mM, the photo-CIDNP enhancement of the H₅ signal is similar for *trans*-HOPI and *cis*-HOPI (73 versus 68, Table 5.2). However, the signal enhancement of the H₂ proton is different for the two diastereoisomers, and each exhibit two different chemical shifts that correspond to two slowly interconverting conformers. The larger enhancement factor for *trans*-HOPI H₂ is -82-fold for its upfield conformer and for *cis*-HOPI H₂ the larger enhancement is -47-fold for its downfield conformer. Moreover, the HOPI concentration dependence of the signal enhancement displays completely different behaviors for the two diastereoisomers. At the lower concentration of 20 μM, *trans*-HOPI shows an enhancement of 120 for the H₅ resonance whereas for *cis*-HOPI it is only ca. 40 (Figure 5.17A). Towards an understanding of this difference, it is noted that the radical pair formation obeys bimolecular steady state kinetics. Considering in a first approximation, under laser irradiation, the triplet state dye (D) is formed in a steady state described by the following:

$$[D] = K[D_0] = \frac{k_{ex}k_{ISC}}{k_T(k_{fluor}+k_{ISC})}[D_0] \quad (5.11)$$

With $[D_0]$ the total dye concentration, the rate constants, k_{ex} for the formation of an excited singlet state, k_{ISC} for the intersystem crossing between the excited singlet state to the triplet state, k_{fluor} for the fluorescence, and k_T for the triplet-singlet relaxation.

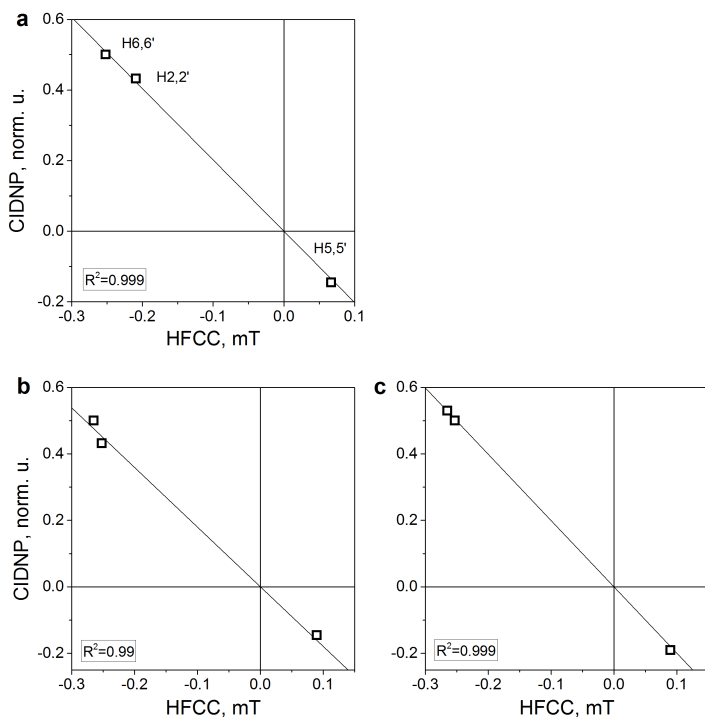
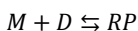


Figure 5.16: A) Correlation between ^1H CIDNP intensities of TCBP detected in the photoreaction of TCBP and trans-HOPI, and the HFCCs of the protons of TCBP radicals T-I:T-III:T-IV=0.56:0.18:0.25. B) correlation between ^1H CIDNP intensities of TCBP detected in the photoreaction of TCBP and trans-HOPI, and the HFCCs of the protons of TCBP anion T-I. C) correlation between ^1H CIDNP intensities of TCBP detected in the photoreaction of TCBP and N-AcTryptophan, and the HFCCs of the protons of TCBP anion T-I. Solid line: best fit by the function $P_{2j} = -C_2 A_{2j}$.

Since it is a cyclic process, the photo-CIDNP can be simplified to a reversible reaction:



Where M is the molecule to be hyperpolarized, and RP is the radical pair. As stated above, due to the long irradiation time (seconds) practiced in the present study, the system is assumed to be in a steady-state, i.e.:

$$\frac{d[RP]}{dt} = 0 \quad (5.12)$$

Hence, the solution to the equation of the bimolecular reaction:

$$\frac{d[RP]}{dt} = k_1[M][D] - k_{-1}[RP] \quad (5.13)$$

Is:

$$[RP] = \frac{k_1}{k_{-1}} [M][D] \quad (5.14)$$

Equation 5.14 describes the steady-state concentration in a reversible bimolecular reaction. The amount of RP is dependent on the concentration of both the D and M . This demonstrates that the slope of the concentration-dependent photo-CIDNP spin polarization shown in Figure 5.17B provides the time-independent rate constant ratio k_1/k_{-1} under the assumption that both the HOPIs are in surplus, where k_1 is the radical pair formation rate constant and k_{-1} the overall radical pair recombination rate constant. Based on the fits of the experimental data from Figure 5.17B, this ratio is higher for *trans*-HOPI (54 M^{-1}) than *cis*-HOPI (39 M^{-1}).

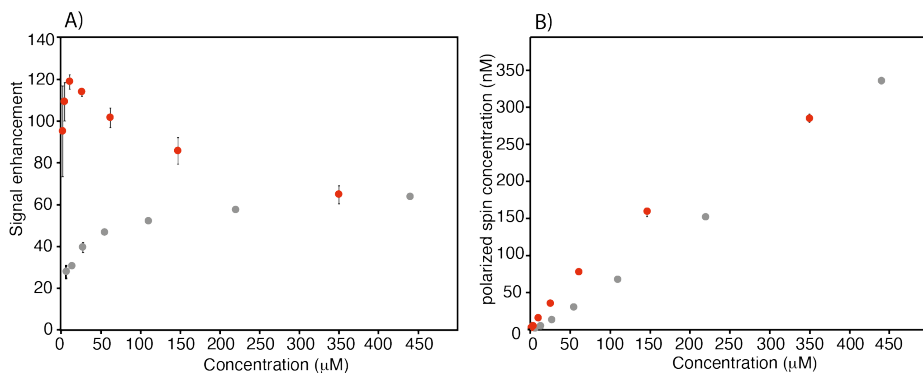


Figure 5.17: A) Concentration dependence of the enhancement of *trans*-HOPI shown as red circles and *cis*-HOPI data shown in grey circles. B) Concentration dependence of the polarized spin concentration. For the calculation it is assumed that the polarization at 14.1 T was 100 ppm. The data points shown in red are: *trans*-HOPI (350, 175, 88, 44, 22, 11, 5.5 μM) and in grey: *cis*-HOPI (440, 220, 110, 55, 27.5, 13.8, 6.9 μM).

Similar mechanistic insights are obtained from competition experiments with photo-CIDNP of tryptophan with either *cis*-HOPI or *trans*-HOPI with the latter showing a stronger interference on the photo-CIDNP of tryptophan (Figure 5.18). In principle, the conformational dependence of the g -factors and the HFCCs could modulate the recombination rate k_{-1} . However, neither the HFCC nor the g -factor calculations showed a significant difference between the two diastereoisomers (Tables 3 & 4) suggesting that the reason for different polarization properties is solely the reaction yield of the radical pair formation. Moreover, the photo-CIDNP enhancement of tryptophan in the presence of HOPI (both *trans* or *cis*) has been studied (Figure 5.18). For increasing concentrations of HOPI, the enhancement of tryptophan decreases, evidence of a competition in the formation of the radical pairs. Furthermore, the quenching of the tryptophan photo-CIDNP is more pronounced with the more photo-CIDNP active *trans*-HOPI diastereoisomer. The competing reactivities of the photo-CIDNP active species underscore the importance of kinetics for the spin sorting process.

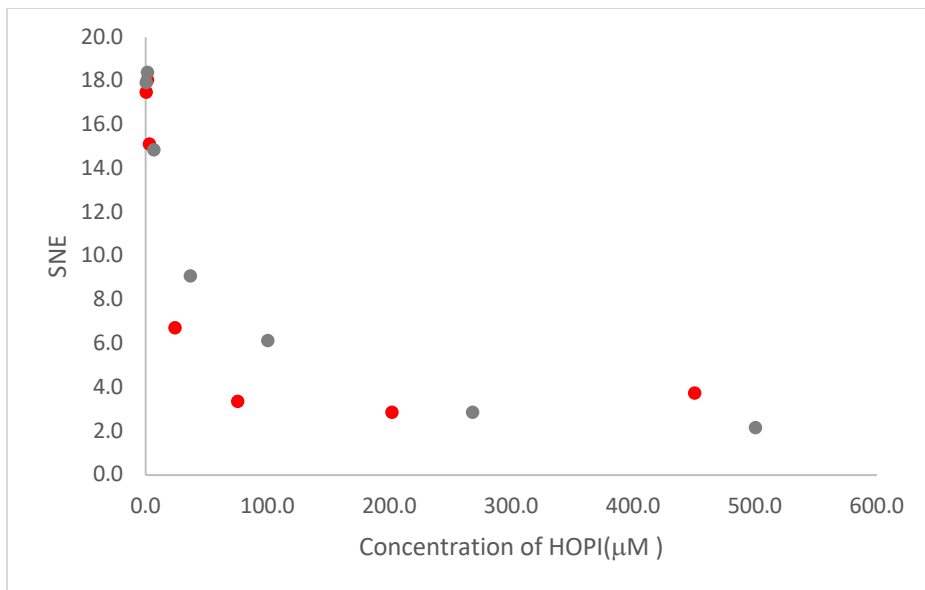


Figure 5.18: Quenching of the tryptophan ($100\mu\text{M}$) photo-CIDNP at 298 K with increasing concentration μM of *trans*-HOPI (red) and *cis*-HOPI (grey).²⁰⁸

This suggests other parameters such as the conformational flexibility, redox potential,²¹⁴ hydrodynamic radius or the dye-HOPI interaction potential to be key for the different polarizations observed. We propose here a straightforward explanation referring to the Marcus theory for electron-transfer, or in this case proton coupled electron transfer.²¹⁵⁻²¹⁶ Shortly, according to the Franck-Condon principle, the electron-transfer timescale being orders of magnitude faster than the nuclei motions, the radicals are generated in the wrong solvation environment, and in the wrong vibrational state. Marcus revised the well-known Franck-Condon principle to satisfy energy conservation. Hence, the vibrational states of the diamagnetic reactant must match the vibrational states of its radical product to enable the electron-transfer. The quantum mechanics simulations predict that the *cis*-HOPI has an internal H-bond between the hydroxy and the carboxy groups, while the *trans*-HOPI not. However, T_1 relaxation experiments revealed no decrease of the relaxation time T_1 for *cis*-HOPI as compared to *trans*-HOPI with T_1 values of 3.88 s and 3.26 s, respectively (Figure 5.19). This longer T_1 value invalidates the H-bond hypothesis.

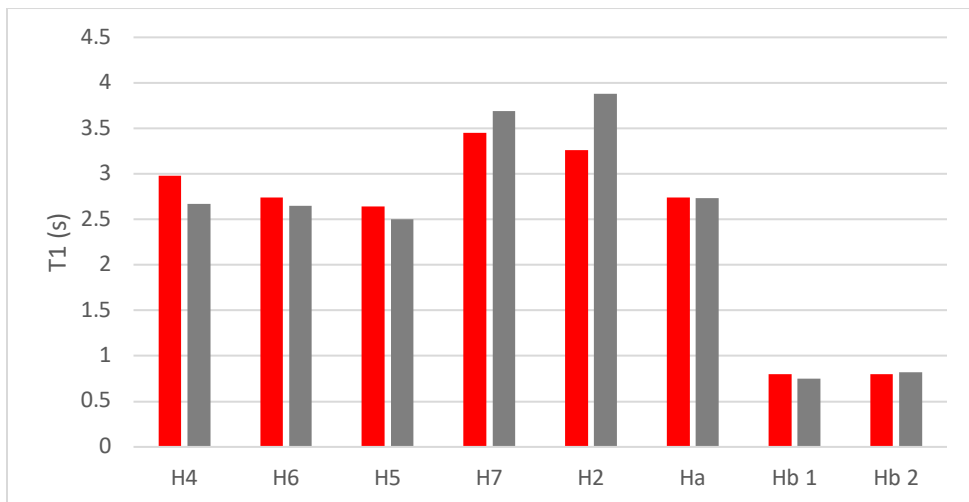


Figure 5.19: ¹H T1 (s) relaxation value of the different protons of the *trans/cis*-HOPI, in red and grey, respectively. The relaxation delay used were: 0.05, 0.10, 0.30, 0.50, 1.00, 1.50, 2.00, 3.00, 5.00, 8.00 seconds.

Nevertheless, the HPLC retention times suggest that the *trans*-diastereoisomer may be less hydrophobic than the *cis*. This difference in hydrophobicity is very likely related with a different solvation profile and may be the result of the different conformations. This particular conformation of the *cis*-HOPI interacts differently with its environment and is responsible for a higher energy penalty while jumping to the higher vibrational states as compared to the *trans*-HOPI. The result of this is a higher free energy barrier for the generation of the radical in the case of the *cis*-HOPI and then a slower reaction rate according to the equation 5.15:

$$k_{et} = Ae^{\frac{-\Delta G^*}{k_b T}} \quad (5.15)$$

The differences in the reactivities of the two diastereoisomers and their impact on the hyperpolarization efficacy suggest that diastereoisomeric compounds have to be tested separately regarding their photo-CIDNP performances. Because of higher energy and much shorter life time of the TCBP triplet state, these high performances were not observed during the TR-photo-CIDNP experiments at mM concentration in the presence of TCBP. In these conditions the two HOPI compounds were showing lower rate constant for quenching of the dye and lower enhancements as compared to tryptophan and N-Acetyl-tryptophan (Figure 5.14). Furthermore, no significant difference in the signal intensities could be observed. This emphasizes the fact that for CW-photo-CIDNP experiments at μ M concentration the polarization performances of a dye-molecule couple.

Bleaching of the CIDNP effect with cis/trans-HOPI compounds

A bleaching of the photo-CIDNP signal-to-noise enhancement (SNE) was observed by repeating the irradiation with a 1 min interval (Figure 5.20). There is a tremendous difference in the bleaching dynamics of the *cis*-HOPI which is fairly slow and the *trans*-HOPI. The bleaching can be also observed on the dye side, after the experiment the sample is transparent instead of purple

as prior to irradiation. Furthermore, the 1D spectra before and after the series of irradiation show no difference in the amount of *cis/trans*-HOPI, suggesting that the bleaching is due to the dye degradation. This finding implies two conclusions: first, the *cis*-HOPI is less reacting than the *trans*-HOPI which correlates with its relatively lower photo-CIDNP activity; second, the dye bleaching is the limiting factor of our photo-CIDNP system and is resulting from the interaction with the to-be-polarized molecule.

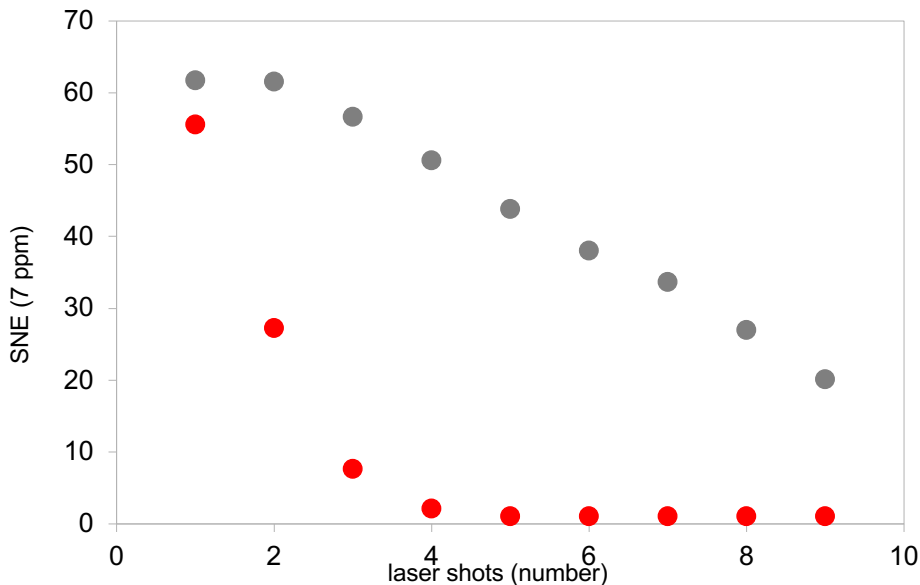


Figure 5.20: photo-CIDNP bleaching of *cis/trans*-HOPI (1 mM, grey and red, respectively) compounds in the presence of AT12 monitored by the observation of H7. The laser pulses were repeated every 1min with an intensity of 1 W and a duration of 4 s.

On the AT12-HOPI pair – an Outlook

The present study starting from a serendipitous observation identified a powerful CIDNP dye/ligand pair-AT12 and HOPI - yielding at relatively low concentrations signal enhancement in the order of two to three orders of magnitude explained by the symbiotic contributions of several factors such as hyperfine couplings, g-factor difference between dye and ligand and reaction kinetics. This pair opens on the one side an avenue towards high throughput screening or NMR experiments at very low concentration of ligand protein interactions studies relevant in biological sciences and drug research if the HOPI can be covalently linked as a CIDNP-NMR probe to the molecules of interest as usually required in fluorescence-based experiment/assays. The HOPI is thereby a rather small molecule when compared to the usual fluorescence dyes. Moreover, HOPI photo-CIDNP can be monitored with AT12 which is excited at 532 nm, an important redshift when compared to fluorescein (450 nm), FMN (450 nm), bipyridyl (308 nm), or TCBP (308 nm). On the other hand, the HOPI studies indicate that many small molecules of chemical and biological interest are potentially CIDNP-active as they comprise more often than

rare tryptophan-like or HOPI-like entities showing that there is a significant chemical space of CIDNP-active molecule to be investigated and potentially used.

Conclusion

Understanding the chemical features that make photo-CIDNP highly efficient are of great value in designing highly polarizable radical pairs. The previously reported maximum ^1H signal enhancement was for tyrosine⁷⁴ which could be enhanced up to 60 fold with AT12, while *trans*-HOPI could be polarized up to 120 fold at a magnetic field of 14.1 T. At a magnetic field of 4.7 T the ^1H signal enhancement observed is up to 380-fold and the ^{13}C enhancement at 14.1 T is estimated to be on the order of 1000-fold. The optimal field for the fluorescein-HOPI pair was estimated to be 400 MHz and improved photo-CIDNP performances are expected at this field. In addition, the optimal field for the HOPI-TCBP is potentially higher than 400 MHz. The optimization of photo-CIDNP pairs for high field could yield to unprecedented SN performances. The reasons for such enhancements are the larger HFCCs, reduced Δg and improved yields of the reactive species. It was already observed that aliphatic protons in the alpha position relative to an aromatic cycle presented the highest HFCC. Unfortunately, the dihedral dependence of the HFCC combined with the rotation of the aliphatic bonds were averaging these HFCCs to lower values.^{208, 217-219} The constraining of the aliphatic proton H_2 in a 5-membered ring impedes the rotation and provide a high HFCC that is responsible for a high photo-CIDNP signal enhancement. Moreover, smaller conjugated systems generate higher spin density and higher HFCCs, as well as modify the SOMO shape and the resulting HFCCs and g -factor. Furthermore, the difference in signal enhancement for the two diastereoisomers underlines the importance of the reactivity of the molecule in relation to the dye. It is an illustration of the Marcus theory and suggests to investigate diastereoisomers separately. Furthermore, given the intrinsic complexity of the CIDNP phenomenon the selection of a dye-molecule couple for CW-photo-CIDNP application cannot be simply derived from the quenching rate and magnetic parameters obtained with TR-photo-CIDNP. The latter is necessary to obtain the information about the electron transfer reaction and the HFCC. It gives also a clear idea of the quenching rates in play. Therefore, it is concluded that the selection of an optimal dye-polarizable aromatic molecule pair may yield unprecedented signal enhancement on the order of several orders of magnitude

Chapter 6: Basis for the empirical exploration of the photo-CIDNP performances

This chapter is an adaptation from the article: [Torres F., Renn A., and Riek R., MagRes, 2021.](#)

Author's contribution: F.T. performed the experiments and analysis.

Introduction

Photo-CIDNP is a versatile technique that can be used to study transient radicals that are too short lived for EPR,²²⁰⁻²²³ to study protein structure²²⁴ and folding,²²⁵⁻²²⁷ or to study electron-transfer mechanism.^{211, 221, 223, 228} The radical pair mechanism is described in the previous chapter and in different papers, that we recommend to the reader for a deeper understanding.^{184, 191, 229} It is our long-term goal to bring CIDNP towards a versatile and straightforward applicable tool in biomedical and biochemical research. In the previous chapter, our efforts have been to push the dye absorption towards more biocompatible wavelength such as near infrared (650-1350 nm). In parallel, the performing of photo-CIDNP with readily handled light source is a contemporary goal.²³⁰ Our approach is to use affordable light sources such as CW solid state lasers that shine light in the second range to maximize the polarization of low concentration. These efforts yielded to the discovery of the AT12 dye which monitored CW-photo-CIDNP experiments with a promising SNE after laser irradiation at 532 nm.⁷⁴ Furthermore, the light source could be setup within a few minutes on different Bruker spectrometers, in our case: 200 MHz Avance, 600 MHz Avance III and 700 MHz Avance Neo. The importance on the irradiation time is indicated in Figure 6.1 yielding for both the compounds tested (i.e., HOPI and tryptophan) an enhanced signal by a factor of 1.5 to more than 2 between one second and four second irradiation time. On this journey to establish a CW-photo-CIDNP in biomedicine, a strong dependency of the photo-CIDNP SNE on the dye-molecule couple was observed. For instance, tryptophan is poorly polarized in the presence of AT12 but highly polarized in the presence of fluorescein, and tyrosine is highly polarized in the presence of AT12 and less well polarized in the presence of fluorescein.^{74, 194} These changing performances were attributed in an initial approximation to the chemical structures of the aromatic ring and the atoms in their close vicinity yielding to different magnetic parameters, g-value and HFCC. This assumption was corroborated by the observation of anomalous line sign alternation for an oxidocyclization product of tryptophan (HOPI) while the dye monitoring the reaction is changed from AT12 to fluorescein.²³¹ Only a scarce fraction of the chemical space has been investigated so far (Table 6.1). Therefore, while the specific radical pair mechanism of these few molecules is well described, the general features ensuring a strong photo-CIDNP polarization and governing the dye-molecule couple optimization are not yet fully understood. The purpose of this chapter is to propose alternative ways simplifying the chemical space exploration to describe these general features. The molecules investigated are biologically relevant and are bioactive in concentrations ranging from nano to micromolar. The extensive study of the chemical space will help to develop predictive model for the CW-photo-CIDNP study of biomolecules at physiological concentrations.

Alternative to the calculation of magnetic parameters

The photo-CIDNP polarization is maximum when the equation 6.1 is satisfied:¹⁹¹

$$\Delta g \mu_B B_0 = HFCC/2 \quad (6.1)$$

In the high field limit, this condition is satisfied only if the g-factors of the two radicals are close to each other. Therefore, the dye-molecule matching relies heavily on monitoring the photo-CIDNP of a molecule of interest with a dye having a g-factor close to it.

Table 6.1: Principal photo-CIDNP active known molecules. FLUO stands for fluorescein.

<i>Molecule</i>	<i>Dye</i>
<i>Tryptophan</i>	FMN ^a , BIPY, FLUO ^f , AT12 ^g , TCBP ^e
<i>NAc-tryptophan</i>	FMN ^a , TCBP ^e
<i>1-methyl-tryptophan</i>	FMN ^a
<i>NAc-serotonin</i>	FMN ^a
<i>Methoxy-tryptamine</i>	FMN ^a
<i>Tyrosine</i>	FMN ^a , BIPY, FLUO ^f , AT12 ^g , TCBP ^e
<i>3-NO₂-tyrosine</i>	FMN ^a
<i>3-F-tyrosine</i>	FMN ^a
<i>3-amino-tyrosine</i>	FMN ^a
<i>NAc-tyrosine</i>	FMN ^a , TCBP ^e
<i>Histidine</i>	FMN ^c , TCBP ^e
<i>NAc-histidine</i>	FMN ^a , BIPY ^d , TCBP ^e
<i>1-methyl-histidine</i>	FMN ^a
<i>Methionine</i>	FMN ^a
<i>Adenine</i>	FMN ^{b,c}
<i>Guanine</i>	FMN ^{b,c}
<i>3/5-methyl-cytosine</i>	FMN ^b
<i>Thymine</i>	FMN ^{b,c}
<i>Porphyrin</i>	1,4 benzoquinone ^c
<i>polyphenol</i>	FMN ^c

a) ¹⁹³; b) ²³² c) ²³³ d) ²³⁴ e) ²³⁵ f) ¹⁹⁴ g) ⁷⁴

The g-factor of a radical can be obtained from DFT calculations, which is still inaccurate, and computationally expensive. Experimentally, this can be done measuring the field dependency of the photo-CIDNP effect, if the hyperfine couplings are known. This requires to use TR-photo-CIDNP and it is a tedious process, in the case many molecules have to be screened. The last solution is the measurement of the g-factor of the radical using electron paramagnetic resonance (EPR), nevertheless because the radicals are transient and the dye bleaches over long irradiation time, this solution usually involves flow through EPR, and is expensive in terms of material.²³⁶

In addition to the theoretical frame detailed before, Kaptein's rule captures the theory of CIDNP into a simple equation in order to qualitatively analyze the sign of a CIDNP line.²³⁷ According to the Kaptein's rules, considering a radical pair composed of molecules a and b the sign of the polarization on a nucleus i belonging to a is predicted by the following equation:

$$\Gamma_{ne} = \mu \varepsilon \Delta g H F C C_i \quad (6.2)$$

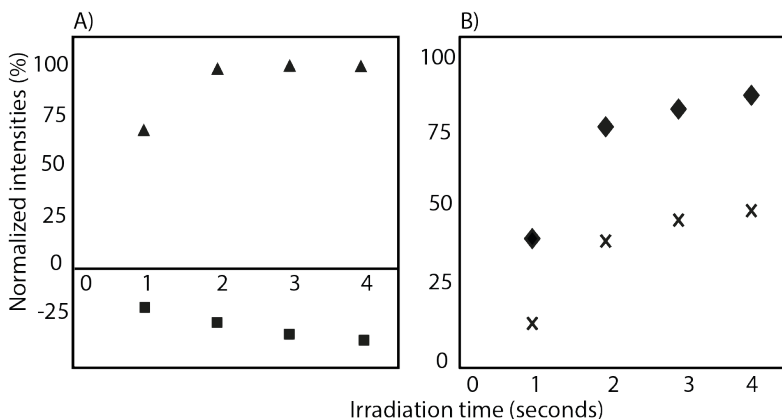


Figure 6.1: Intensity build-up curves according to the irradiation time. A) HOPI where the triangles and squares represent AT12 and fluorescein, respectively B) tryptophan where the rhombs and crosses represent AT12 and fluorescein, respectively.

Γ_{ne} is the net polarization sign of the radical a, μ and ε are Boolean values. μ is positive when the radical is formed from a triplet precursor, and negative otherwise. ε is positive for recombination products and negative for the radical escaped or the transfer reaction products. Δg is the sign of the g-factors difference between the two radicals, i.e., $g_a - g_b$, and $H F C C_i$ is the hyperfine coupling constant sign of the considered nucleus i in the radical a which makes the reaction nuclear spin selective. The Kaptein's rule equation predicts sign of polarization and reflects the complex nature of the reaction path that yields out-of-Boltzmann spin polarization. Therefore, it can be used for a qualitative analysis of the photo-CIDNP polarized products. This observation is related to the g-factor difference between the two molecules as we shall see, and is thus a simple manner to approximate the g-factor of a molecule. Indeed, we will demonstrate later in this chapter how Kaptein's rule can be used to rank the molecules on a g-factor scale by simply the photo-CIDNP sign alternation when monitored with different dyes, and how we use it to validate the prediction of magnetic parameters associated to chemical features.

In this chapter we also discuss the importance of side-chains in the intensity of the hyperpolarized lines in CW-photo-CIDNP experiments. Although the effect of the chemical modifications on the aromatic part of the molecules are well described by theory and confirmed experimentally,^{217, 219} the effect of the non-aromatic moieties are considered to be conditioning the triplet state dye quenching kinetics, as observed in TR-photo-CIDNP.²³⁵ While the exact mechanism of polarization can be only evaluated by TR-photo-CIDNP, in the context of biomedical NMR application aiming for the highest polarization at low micromolar molecule concentration, CW-photo-CIDNP appears to be the method of choice suggesting the exploration of a CW-photo-CIDNP-based empirical approach indicated. This work reports on this approach by screening 10 compounds with two different dyes. Moreover, the optimization of the photo-CIDNP system enables the detection of micromolar concentration of biologically relevant molecules.

Chemical tuning of g-factor

1.1 On the Kaptein rule of the oxidocyclization product of tryptophan α -hydroxypyrrroloindole

The differential effect of the dyes to trigger radical pair mechanism yielded to the serendipitous observation of HOPI an oxidocyclization product of tryptophan which is highly polarized after irradiation in the presence of AT12 (Figure 6.2). The study of HOPI revealed surprising features such as different polarization yields between the *cis* and *trans* diastereoisomers, and the sign alternation of the hyperpolarized intensities depending on the dye used to form the radical pair, fluorescein or AT12.²³¹ This sign alternation is here now assessed in the light of the Kaptein rules (equation 6.1). In the case of the photo-CIDNP reaction that is performed for all the experiments of this work, μ is positive since the radical pair is formed in a triplet-state. Moreover, the polarized species are the recombination products of the radical pairs and thus the parameter ϵ is positive. Hence, the variable parameters are the hyperfine coupling when the molecule changes, e.g., from tryptophan to HOPI, or/and the Δg . The Δg can also alter when the dye used is changed. Hence, for the same molecule a sign change of the NMR signal upon switch from a dye to another one is necessarily caused by an alternation of the sign of the Δg in the Kaptein rule equation (equation 6.2). As shown in Figure 6.2, the sign switch is observed in the case of HOPI (evidently for all the resonances) when the dye is altered from fluorescein to AT12. This finding can then be used to set the unknown g-factor of HOPI radicals between the g-factors of the dyes and in respect to the known values of tryptophan and tyrosine radicals as shown in Figure 6.3. Moreover, the observation of the anomalous lines signs in photo-CIDNP experiments monitored with TCBP in previous study, enabled to rank the HOPI compounds with a g-factor between 2.0034 (fluorescein) and 2.0035 (TCBP)²³¹.

This is not only an elegant illustration of the Kaptein's rules, but also a witness of a g-factor evolution upon chemical modification from tryptophan to its oxidocyclization product, HOPI (Figure 6.2). Since the g-factor originates from spin-orbit coupling, the aromatic system type has an influence on the g-factor value. This hypothesis is consistent with the results previously obtained for tyrosine which has a high g-factor (2.0041) and is preferentially polarized in the presence of AT12⁷⁴. Therefore, the photo-CIDNP spectrum of 2-3 dihydro-tryptophan (dH-TRP) which has the same aromatic system as HOPI (Figure 6.2), was recorded for both the dyes, AT12 and fluorescein (Figure 6.2). As expected, the polarization sign switch could be observed again upon dye change, confirming the idea that similar aromatic systems should/may yield close g-factors (as pinpointed to in Figure 6.2 and 6.3).

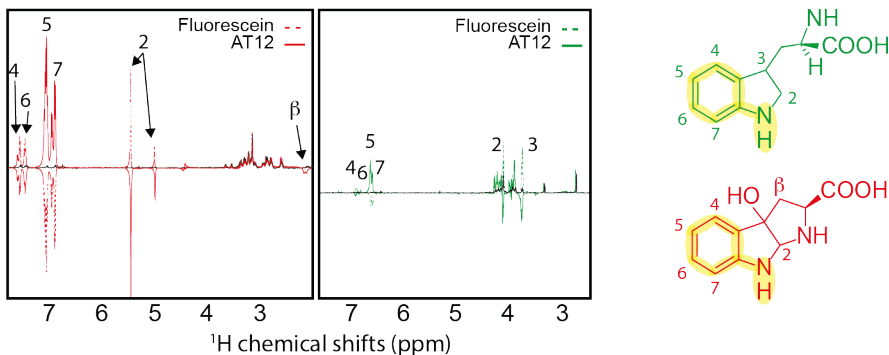


Figure 6.2: HOPI (red) and dH-TRP (green) photo-CIDNP spectra with fluorescein (dashed line). The aromatic system is emphasized with yellow background.

However, the good photo-CIDNP performance of the HOPI compound is not observed for dH-TRP as the polarization enhancement in the presence of AT12 was only 10-fold and 17-fold in the presence of fluorescein (Table 6.2). This difference cannot be attributed to a slight difference in the g-factor (towards the g-factor of fluorescein) because the SNE in the presence of fluorescein did not compensate the loss in SNE in the presence of AT12 as it would be expected if the enhancement would solely rely on a g-factor value change (Figure 6.2). Furthermore, the HFCC calculated with Gaussian® are comparable to the HFCC obtained for the HOPI (not shown). The lower performances of dH-TRP in comparison to HOPI, despite comparable magnetic parameters evidences the influence of other physicochemical parameters. The combination of the anomalous intensity sign alternation upon dye change and the analogy between aromatic systems enables the ranking of the molecules on a g-factor scale as it is shown in the Figure 6.3. Such a ranking is important to approach the condition described in equation (6.1) by ensuring a small Δg and thus optimizing the dye-molecule match.

Screening of the tryptophan and tyrosine chemical space

The distinct photo-CIDNP performances of the different dye-molecule couples was previously discussed in the literature.^{74, 194} For interest here, the tryptophan presented a higher SNE when polarized upon fluorescein irradiation when compared with the dye AT12 as shown in Figure 6.4 and listed in Table 6.2 whereas tyrosine was better polarized in the presence of AT12 (Table 6.2).⁷⁴ The lower performances of the dH-TRP in comparison to HOPI's despite similar g-factors turned the focus to the potential involvement of the side chain. Prior work has been done by Saprygina et al. (2014), studying the influence of N-acetylation on the quenching rate of TCBP. The replacement of the α -amine by a N-acetyl, resulted in the vanishing of the positive charge and lower quenching rates of the triplet-state photosensitizer accompanied by lower time-resolved photo-CIDNP enhancements and interpreted as causative of the N-acetylation. Similarly, to this approach, side chains modifications of the same molecular species were studied in the context of CW-photo-CIDNP.

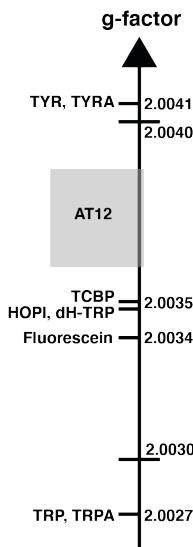


Figure 6.3: The sign alternation enables to locate the molecules on the g -values scale. The ranking of HOPI and dH-TRP provides fine information about their radicals g -values. For the non-elsewhere defined abbreviations: TRP stands for tryptophan, TRPA stands for tryptamine, TYR is tyrosine, TYRA is tyramine. (Figure adapted from Sobol et al.⁷⁴)

Here, first insights into the potential role of the side chain was gathered by a comparison of the tryptophan-derivative tryptamine (chemical structure see Figure 6.4) with tryptophan. Tryptamine differs from tryptophan by the absence of the carboxylic acid on its side chain. Indeed, improved CW-photo-CIDNP SNE is observed for tryptamine when compared with tryptophan, especially after irradiation in the presence of AT12 for which a further signal enhancement of a factor of 2-fold is documented (Figure 6.4, Table 6.2). Since both molecules have the same aromatic system and thus similar magnetic parameters,²³⁶ and a similar reaction mechanism is expected (i.e. ET), the improved polarization of tryptamine might be due to the charge of the molecule that differs from tryptophan by the absence of the carboxylic acid on its side chain causing potentially a change in the quenching kinetics: Fluorescein's contains a benzocarboxylate moiety of typical pKa 2-2.5 and a xanthenol of pKa 6.4,²³⁸ and is, in the buffer of interest, twice negatively charged. AT12 is neutral in the experimental conditions (pH = 7.1), however the aromatic system is globally carrying a positive charge (Figure 6.4). Due to its overall positive charge, it is expected that the quenching of fluorescein by tryptamine is faster than by tryptophan, which is globally neutral in the experimental conditions. However, this is contradictory with the fact that the strongest improvement in terms of CW-photo-CIDNP performances is for AT12 monitored experiments, despite the rather repulsive charges in play.

Table 6.2: Signal enhancement of tryptophan and tyrosine derivatives (100 μM) after irradiation in the presence of AT12 or fluorescein (20 μM) at 1 W for 4 seconds. The reported charges are of the diamagnetic molecules. The values for the tyrosine enhancement are taken from ref 20 and for the tryptophan and HOPI enhancement from ref 22. Nevertheless, sample conditions were identical. The $\log(P)$ is the logarithm of the partition coefficient P , and P is the ratio of the concentration of the compounds in a mixture of the two immiscible solvents octanol and water ($P = [\text{molecule}]_{\text{octanol}} / [\text{molecule}]_{\text{water}}$).

<i>molecule</i>	<i>AT12 SNE</i>	<i>Fluorescein SNE</i>	<i>Charge</i>	<i>Log(P)</i>
<i>tryptophan</i>	18 ^a	54 ^a	0	-1.1
<i>tryptamine</i>	45	70	+1	1.6
<i>tyrosine</i>	-38 ^b	-20 ^b	0	-2.3
<i>tyramine</i>	-60	-23	+1	1.1
<i>HOPI</i>	90 ^a	-70 ^a	0	-2.1
<i>2,3-dihydro-tryptophan</i>	10	-17	0	-1.5
<i>3-(2-(piperazin)ethyl)-indole</i>	44	103	+1	1.9
<i>indole propionic acid</i>	45	70	-1	1.8
<i>indole acetic acid</i>	27	54	-1	1.4

a 231, b74

In order to investigate further the impact of the side-chain on the SNE of AT12 monitored CW-photo-CIDNP experiments, such experiments were conducted on tyramine (Table 6.2), a derivative of tyrosine where the α -carboxylate moiety is absent. Tyrosine and tyramine are preferentially polarized upon irradiation in the presence of AT12 oppositely to tryptophan/tryptamine, due their higher g-factor, as shown in Figure 6.3. The CW-photo-CIDNP SNE in the presence of AT12 is significantly higher for tyramine when compared with tyrosine (Figure 6.4, Table 6.2). The minor SNE enhancement for tyramine versus tyrosine photo-CIDNP experiments monitored by fluorescein could be explained by the different Δg . This experiment supports the finding that the chemical modification of side-chains can significantly improve the SNE for CW-photo-CIDNP in the presence of AT12. Next, the CW-photo-CIDNP spectra of indole propionic acid (IPA) and indole acetic acid (IAA) have been recorded. IPA is the negatively charged analogue of tryptophan (Table 6.2) where the α -amine is lacking. IAA is similar to IPA but the carboxylate group is closer to the aromatic ring, since it is in the β -position.

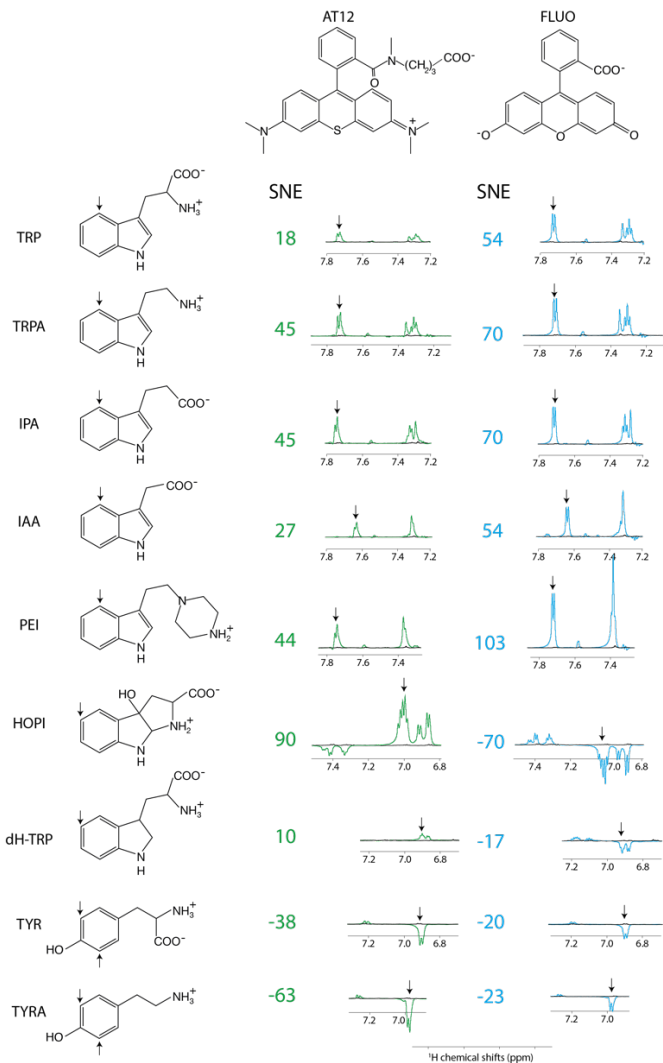


Figure 6.4: Photo-CIDNP spectra of tryptophan or tyrosine derivatives. The best aromatic proton's signal-to-noise enhancement (SNE) is provided and the corresponding signal is pinpointed with an arrow. All spectra were recorded after 4 sec of irradiation at 1 Watt, in the presence of either AT12 (20 μM, irradiation wavelength 532 nm) or fluorescein (20 μM, irradiation wavelength 450 nm), the to-be-polarized molecule is present in concentration of 100 μM. Color code: green are the AT12 monitored SNE values and photo-CIDNP spectra, blue are the fluorescein monitored SNE values and photo-CIDNP spectra, dark are the non-irradiated reference spectra. The structures are provided in their ionization state at experimental pH 7.1.

Unexpectedly, the IPA yielded to the same SNE as tryptamine (Table 6.2) whereas an interpretation of the SNE solely based on the charged, overall negative for IPA, predicted an opposite effect on the performance as compared to the positively charged tryptamine. Despite identical charge as IPA, IAA (Table 6.2) exhibits comparable performances as compared to tryptophan. Moreover, the 3-(2-(piperazin)ethyl)-indole (PEI) is an analogue of tryptamine where the α -amine is replaced by a piperazin moiety. In PEI, the overall charge is similar to tryptamine since the pKa of the tertiary amine is close to 4 and only the secondary amine is positively charged, due to its pKa around 9. PEI yielded similar polarization performances as tryptamine upon irradiation in the presence of AT12, and showed even higher SNE for fluorescein monitored photo-CIDNP experiments. An interpretation of these results solely based on the respective overall charges therefore fails.

Alternatively, it could be hypothesized that a different side-chain dynamics may play a role in the SNE of CW-photo-CIDNP since with the side chain alterations not only the charge of the side chain changed but also the dynamics with the tryptamine, IPA, IAA, PEI, and tyramine comprise faster side chain motion than tryptophan and tyrosine. This change in dynamics is indicated by the observation that the H_{β} resonances are split for tryptophan and not for the tryptamine, IPA, IAA and PEI (Figure 6.5). The same degeneracy of the H_{β} chemical shifts is observed when the amine group is removed from the tyrosine to become the tyramine (Figure 6.5). However, this hypothesis cannot be generalized since the dihydro-tryptophan which has a higher side-chain mobility as compared to the HOPI shows less polarization than HOPI.

The only summary of this first attempt to interpret the chemical space exploration is that the simultaneous presence of the α -carboxylate and the α -amine is suboptimal for CW-photo-CIDNP SNE when monitored with fluorescein or AT12 as supported by the less good polarization properties of tryptophan, tyrosine, and diH-TRP when compared with their analogues. A corollary of the presence of these α -carboxylate and α -amine is the water solubility of the small molecule and their solvation shell. This idea brings us to another difference between the different side-chains properties, which is hydrophobicity. This can be assessed with the logarithm of the calculated partition coefficient between octanol and water, $\log(P)$.

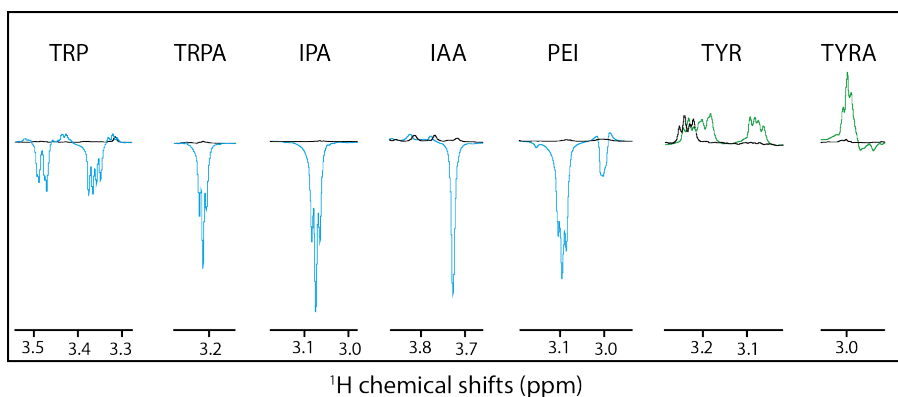


Figure 6.5: H_{β} anomalous signals for the different molecules. Dark lines are the reference spectra and blue lines are the irradiated spectra in the presence of fluorescein, and the green lines are the irradiated spectra in the presence of AT12. Scales between the different molecules are not respected for clarity purposes.

The evolution of the hydrophobicity within the different families of compounds and its influence on the CW-photo-CIDNP performances was therefore investigated. Within the tryptophan derivative group, i.e. tryptophan, dihydro-tryptophan, tryptamine, IAA, IPA, PEI, the increasing hydrophobicity is beneficial to the CW-photo-CIDNP performances when monitored by both fluorescein and AT12 dyes (Figure 6.6A and B). The same trend is suggested for the tyrosine derivatives, tyrosine and tyramine: Tyramine which is more hydrophobic is better polarized, especially in the presence of AT12, than tyrosine (Table 6.2). The HOPI was not included in this analysis since it is rather far away from the chemical space of the two series of interest. The positive influence of hydrophobicity on the SNE may be explained by two distinct mechanisms that also may work in concert. First, the aromatic nature of the dye-molecule interaction is favored for more hydrophobic molecules. Second, the water shell surrounding the molecule is perturbed by the different hydrophobicity of the side-chains as it can be observed from the H_{β} dynamics (Figure 6.5). Hence, the π - π stacking between the excited dye and the molecule, and therefore the orbital overlap, could be altered in a positive manner by increasing the hydrophobicity of the molecule. In summary, the hydrophobicity variation upon side-chain modification appears to have a qualitative impact on the CW-photo-CIDNP SNE unlike the charge and dynamic variation. With other words, the observed trends suggest a positive impact on the SNE for higher hydrophobicity of the molecules sharing a common aromatic moiety. While noting these findings, it must be stated (as above) that the exact nature of the polarization can only be determined by TR-photo-CIDNP. However, the presented empirical approach is regarded informative for CW-photo-CIDNP applications as the positive aspects of CW-photo-CIDNP with several seconds of light irradiation in terms of SN and easy and cheap set up is apparent.

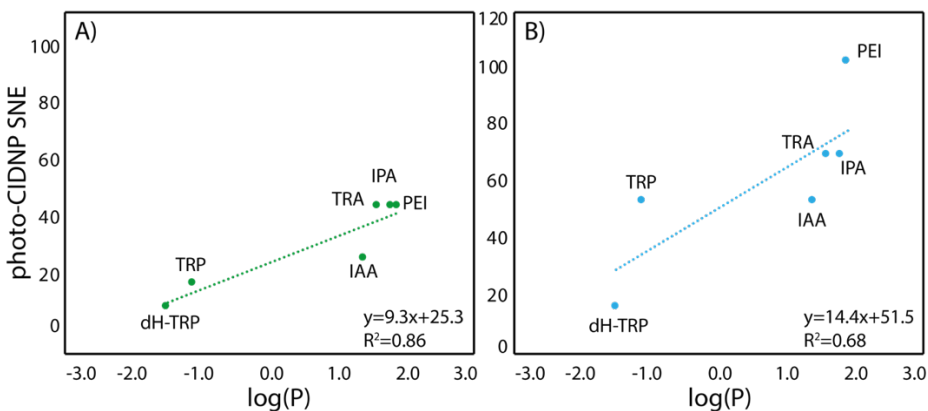


Figure 6.6: Correlation between the molecule hydrophobicity and the photo-CIDNP performances for the tryptophan derivatives. A) In the case where CW-photo-CIDNP is monitored by AT12. B) In the case where CW-photo-CIDNP is monitored by fluorescein. Log(P) is the logarithm of the partition coefficient, P, between octanol and water and were calculated with DataWarrior®. A statistical analysis of the trendlines using Pearson's R coefficient and the Student t-test for the hypothesis shows that for AT12 (A) a t of 4.87 and a p-value of 0.008 and for fluorescein (B) a t of 2.6 and a p-value of 0.060 is obtained. Hence, in both cases, the hypothesis H_0 (absence of correlation) is rejected, and the alternative hypothesis (correlation is non-null) is retained.

Another well described parameter is the pKa of the radical aromatic exchangeable protons. This results into a pH dependence of the photo-CIDNP effect, which has been extensively described for the tryptamine derivatives.¹⁹³ The pH dependency and therefore the pKa of the indole aromatics varies accordingly to the sidechains (Figure 6.7). This pH dependence is particularly important for the negatively charged IPA and IAA, when monitored with fluorescein (Figure 6.7A). Moreover, it appears that at pH > 8 the indole derivative compounds with negatively charged sidechains are the best enhanced.

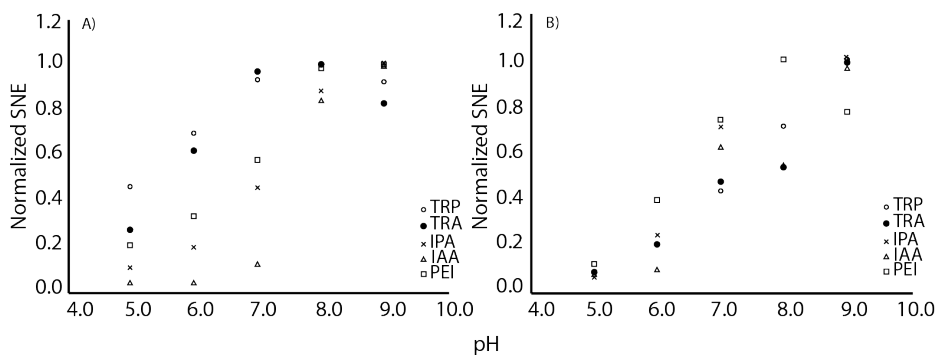


Figure 6.7: pH dependency of the SNE for fluorescein (A) and AT12 (B).

Encouraged by the prior observation, we decided to select the most hydrophobic indole derivative compound available in the lab, and we tried to detect it at micromolar concentration. Thus, the 5-methyl-3-carboxy-indole (MCI, $\log(P) = 2.4$) was measured at 10 and 1 μM , as shown in the Figure 6.8A and B, respectively. At such concentrations the dark spectra suffer low SN hindering any sort of analysis. The irradiated spectra show significant improvement already with a single scan, especially at 10 μM , where the SN is sufficient for the analysis, with a SNE of 50-fold for the proton 4 (Figure 6.8A).

However, at 1 μM the SN is not sufficient to identify all the aromatic peaks. Therefore, the signal was accumulated for 8 scans, yielding satisfying SN, with a SNE of 153-fold for the proton 4 (Figure 6.8B). In addition to the possibility of measuring low micromolar samples, the use of photo-CIDNP reduces the time measurement in a tremendous extent: the single experiment of the 10 μM sample was recorded in 5 seconds and would have needed 3 hours of signal accumulation to reach the same SN without light irradiation. More impressive even, the 8-scans photo-CIDNP measurement of 1 μM MCI, which was measured within a minute, and would have required 260 hours (10.8 days) of measurement without hyperpolarization. MCI is a known low affinity inhibitor of PIN1¹³¹; this opens the way for studying drug like molecules, or metabolites at the physiologic concentration. Many natural molecules such as metabolites or plant alkaloids, and approved drugs share an indole/indoline moiety. Therefore, a significant amount could be studied using photo-CIDNP polarization. As a validation of this hypothesis, we could observe photo-CIDNP hyperpolarization for several indole containing drugs such as tadalafil, etodolac, melatonin or sumatriptan (Table 6.3).

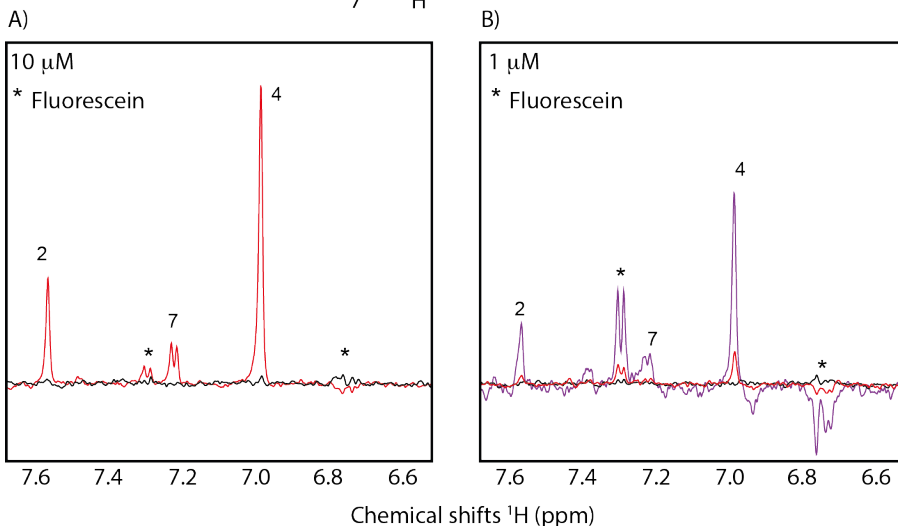
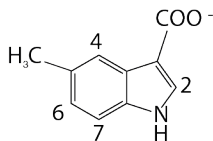


Figure 6.8: photo-CIDNP spectra of MCI at 10 μM (A) and 1 μM (B), in black the non-irradiated spectra, in red the irradiated single scan-spectra, in purple the irradiated 8-scans spectrum. The spectra were acquired on a 600 MHz NMR spectrometer in a 5 mm tube after 4 seconds irradiation at 532 nm.

Conclusion

The HOPI compound offers a nice illustration of the interplay between the choice of the dye and sign of the polarization according to the Kaptein rules. The modification of the chemical structure leads to a difference in the g-factor of the molecule and a relationship between this g-factor evolution and the aromatic system shape could be built without using computational methods. The compilation of CW-photo-CIDNP data for different dyes of known g-values will help to refine the g-factor scale presented in the present work. This easy ranking will help to rationalize CW-photo-CIDNP performances for new molecules of interest. Moreover, the effect of non-aromatic modifications could be observed. The side chains of the different molecules play a key role in the CW-photo-CIDNP SNE, which could not be predicted considering solely the ionic interaction monitoring the dye quenching or the radical pair stabilization. Indeed, the hydrophobicity of the molecules revealed to have an influence where the polarization performances improved gradually within the two classes of derivatives, respectively. The importance of the hydrophobicity opens the way for simple chemical space exploration, since it can be simply assessed by its $\log(P)$. Moreover, the pH dependency of the indole derivatives has been once more confirmed, and revealed the preferential enhancement of the carboxylated sidechains, in particular in the presence of fluorescein. This finding could not be predicted from

the previous data that is considering only the quenching rates of the triplet-state dye as a non-magnetic factor governing the photo-CIDNP efficiency. Once again, the presented results demonstrate on the complex interplay between molecular properties that lead to a CW-photo-CIDNP SN enhancement. Of great interest to us, is the possibility of empirically screening the chemical space for potential highly CW-photo-CIDNP-active small molecules. From this initial study, emerges the potential of making CW-photo-CIDNP a much broader method for signal enhancement in the biomedical NMR field, and the request for the exploration of the non-aromatic chemical-space within photo-CIDNP-active molecules. Finally, these findings, high hydrophobicity and carboxylated sidechain, were successfully applied to the detection of compounds at low micromolar concentration. In conclusion, the screening of more compounds will open the way to fine tuning of the photo-CIDNP conditions in order, to analyze compounds at micromolar concentration.

Table 6.3: photo-CIDNP SNE for different indole containing FDA approved drugs (100 μ M) monitored with fluorescein (25 μ M).

<i>Molecule</i>	<i>SNE (fluorescein)</i>	<i>SNE (AT12)</i>
<i>tadalafil</i>	35	13
<i>etodolac</i>	62	35
<i>Melatonin</i>	10	20
<i>Sumatriptan</i>	45	26

Conclusion and outlook

NMR is versatile technique which can be used at any stage of the drug discovery journey. In difference to X-ray crystallography, the sample conditions are rather physiological as NMR can be performed in solution, at room temperature and recently, also in-cell NMR shows promising results for the study of biomolecules in their native environment.²³⁹ Another advantage of NMR is its capacity to characterize transiently interacting systems such as fragments interacting with their target, often a protein, with dissociation constant in the millimolar range.^{15, 46, 121} In particular for the obtention of SAR at atomic resolution, NMR is the main alternative to X-ray crystallography. Nevertheless, NMR suffers from its own intrinsic hurdles, namely the complexity and tediousness of its data analysis combined to the insensitivity of the technique resulting from the low nuclear spin state polarization.

The time required for the data analysis is dominated by the time for the signal assignment.⁴⁶ Despite the development of many automation tools, the assignment step remains laborious and requires the full-time involvement of an experienced spectroscopist for several weeks, if not months. In comparison, X-ray crystallography diffracts a crystal in a few minutes and most of the analysis is automated and the structure can be established and validated within a few days.^{13, 22-23} Therefore, while X-ray crystallography is well aligned with the medicinal chemistry timeline, NMR is far off. Since the beginning of SBDD, different methods have been imagined to shortcut the assignment step and obtain the protein-ligand complex. Taking advantage of the preexisting protein structures deposited in the PDB, those techniques use ambiguous or non-ambiguous restraints to guide docking or rank the docking outputted ligand poses. Such methods are biased by the force-field of the docking software, which resolution is often not enough to provide high quality ligand poses. The only method, to our knowledge, using sparse data for the structure calculation of protein-ligand complex structures is NMR molecular replacement (*NMR*²), which addresses the protein signal assignment issue by handling semi-ambiguous restraints by a simulated annealing algorithm.^{43-44, 54} The *NMR*² method has been demonstrated to succeed previously for a number of targets, strong and weak binders. In this thesis the introduction of new features to the *NMR*² method is described, such as anti-NOE distance restraints, specific methyl labeling, interleaved filtered NOESY, PA, and cooperative assignment strategy. Altogether, these improvements fasten the algorithm convergence, improve the quality and the number of distance restraints, and enable to solve the structure of challenging systems such as protein-fragment complexes. The specific methyl labeling reduces the internal protein spin diffusion and suppresses the background signal in the 2D filtered NOESY spectra; using specific methyl labeling, one can tackle system of higher molecular weight, expectedly up to 100 kDa.³⁴ Such labeling scheme also opens the way to use interleaved filtered-NOESY experiments. When recorded with non-specifically labeled samples, the pulse sequence gives rise to artifacts hindering the analysis. The interleaved experiment, is interesting as it provides bidirectional NOEs leading to high quality distance restraints; it also provides the intra-protein methyl-methyl and the intra-ligand cross-peaks in separate spectra. These additional data have been used for the complex BRD4-BD2-iBET762 to derive the PA of the protein methyl, and the assignment of the ligand. Furthermore, the ligand bound conformation is constrained during the *NMR*² calculation using the distance restraints obtained from the intra-ligand spectrum. Nevertheless, the intra-ligand and the unidirectional NOEs are present in the 2D F_1 -[¹³C,¹⁵N]-filtered-[¹H,¹H]-NOESY, and are in most cases enough to derive the protein-ligand-complex structure. A simple filtered-NOESY is recorded four times faster than its interleaved version and contains half of the data. Therefore, the use of interleaved NOESY should be set aside for the most challenging systems, especially since it requires to specifically label the protein, a costly approach. The last advantage of specific

methyl labeling, is the use of anti-NOE restraints which densify significantly the distance restraints network, which may in principle help the algorithm convergence in challenging cases, even when applied conservatively as seen in the Chapter 2 for the BRD4-BD2 complex. However, it is not excluded that, with cautiousness, anti-NOE distance restraints could be introduced to the structure calculation of non-specifically labeled samples. In summary these features enabled the calculation of BRD4-BD2 in complex with its ligand iBET762, a system so far impervious to crystallization. The timeline for such a calculation is still within a week and aligns well with the medicinal chemistry turnaround table time. In Chapter 3 the applicability of NMR^2 to protein-fragments has been assessed and clear guidelines to guarantee success in most cases are emitted:

- 1) Record the NOESY spectra.
- 2) Calculate the topology parameters for each fragment.
- 3) Select the best ranked fragment based on the topology parameters and measure the HCCCH-TOCSY and the ^{13}C -ctHSQC spectra.
- 4) Include the PA obtained from complementary spectra in the NMR^2 calculation input, and calculate the structure of the best fragment from (3).
- 5) Identify the correct structure based on its target function and use the assignment combination to calculate the rest of the protein-fragment complex structures.

These guidelines introduce the use of PA obtained from two state-of-the-art experiments, HCCCH-TOCSY and ^{13}C -ctHSQC, to ease the algorithm convergence; it introduces as well the concept of cooperative assignment strategy. This latter feature of NMR^2 is particularly relevant in the context of FBDD campaigns where many different complexes need to be calculated to establish the SAR. The use of PA within the cooperative assignment strategy was successfully applied to the structure calculation of four PIN1-fragment complexes, each of which was calculated within a week. In comparison, the classic structure calculation of such fragment would have taken 2-3 months of work per fragment. In conclusion, NMR^2 is a promising alternative to X-ray crystallography, and it makes NMR structure calculation suitable for the routine in SBDD or FBDD pipelines. NMR^2 is particularly interesting to complete X-ray crystallography for FBDD as NMR data are known to be of better quality in the case of weakly interacting systems. If NMR^2 is now an established tool covering the application scope of NMR structure calculation, the screening step still suffers many hurdles. As it could be seen in the Chapter 3, NMR performed poorly for the screening of drug like molecules due to the concentrations at play (0.2 mM). The next challenge in the field is to develop proton 1D experiments operating at low concentration to reduce the false positive rate, and avoiding the use of labeled protein. Furthermore, the timespan of the screening is far too long as compared to other methods (796 compounds in 14 days). Such long measurement time are explained by the poor sensitivity of NMR due to the low polarization of its nuclear spin state populations. In comparison, in less than a year, X-ray crystallography soaking-based campaign from the diamond light screen provided screened several thousands of molecules and provided 197 complex structures.²⁴⁰ Nevertheless, most the molecules screened are fragments and do not necessarily directly impact the drug repurposing effort. An NMR method able to detect strong binding at a high throughput rate would therefore be a major contribution to the drug repurposing field.

This other hurdle of NMR, sensitivity, was addressed using photo-CIDNP hyperpolarization.¹⁹¹ With the ultimate goal to polarize low micromolar concentrated samples, CW-photo-CIDNP was the preferred technique, and preferred over other hyperpolarization technique for its simplicity of implementation and the compatibility of the method with close to physiological conditions.⁶⁶ However, while the theoretical framework of the radical pair mechanism,⁷⁰⁻⁷² governing the SNE of photo-CIDNP, is well known, the optimization of the technique for its application to pure

hyperpolarization purpose is scarcely described. Moreover, only a small subset of molecules has been polarized using photo-CIDNP, hindering the identification of molecular features favorable or not to the hyperpolarization. In the present thesis, the known chemical space of photo-CIDNP polarizable molecules is extended. Indeed, the polarization properties of tryptophan derivatives were described in the light of their magnetic parameters. Furthermore, the correlation between the magnetic parameters and the chemical structures is demonstrated, suggesting that chemical modification could be used to tune the photo-CIDNP properties of molecule to be polarized. Importantly, the influence of non-magnetic parameters is shown to be critical for the hyperpolarization with photo-CIDNP. Indeed, the two HOPI diastereoisomers have different SNE due to their different reaction rates.⁷⁵ This difference in reactivity is also at the origin of the sidechain effect, which is not related to the magnetic parameters. The sidechain effect has been previously studied in light of the electrostatic charge-charge interaction with the dye.⁷⁶ In TR-photo-CIDNP experiments in the millimolar concentration range, the influence of the sidechain charge on the triplet state dye quenching rate to form a radical pair was shown to be the main non-magnetic factor governing the photo-CIDNP performances. Nonetheless, such an impact of the sidechain charge was not observed with CW-photo-CIDNP at low concentration. Instead, the hydrophobicity was shown to have a great influence on the SNE upon photo-CIDNP experiments, and the presence of carboxylate groups was shown to be important regardless of the dye charge, and thus of the electrostatic interaction itself.²⁴¹ These preliminary results evidence the importance of chemical structure on the radical pair reactivity and the subsequent great influence on the photo-CIDNP SNE. Further screening of aromatic and sidechains combinations will provide a better understanding of the underlying parameters relevant to CW-photo-CIDNP. On the other side of the radical pair, the dye chemical space was explored as well to find redshifted dyes as compared to the traditional ones. In accordance with recently published study,¹⁹⁵ it was found that the introduction of halogens to generate internal heavy atom effect was of no benefit, and even deleterious, for the purpose of photo-CIDNP SN enhancement. However, the modification of the aromatic system of fluorescein, the xanthene moiety, by introducing sulfur and nitrogen atoms as in AT12 led to significant improvements in the hyperpolarization performance for high *g*-values molecules, such as tyrosine.⁷⁴ The fluorescein and the AT12 dyes show improved properties as compared to other dyes, while the exact reasons for that remain unclear, some features appear to be of importance for the research of further photo-CIDNP active dyes. In particular, the introduction of halogens to the dye structure with the purview to induce internal heavy atom effect¹⁹⁵ revealed to be deleterious for the photo-CIDNP performances. The exploration of further xanthene or phenyl-xanthene like aromatic systems, incorporating sulfur, nitrogen, and oxygen will help to define with more accuracy the photosensitizer chemical space. In conclusion, the present work opens the way for the exploration of the chemical space of photo-CIDNP on both sides of the radical pair, namely the dye and the molecule to be polarized. This exploration is to be done through the aromatic system chemical space as well as through the sidechain chemical space. The aromatic system of the dyes can be modified to modulate *g*-factor of the dye. Finally, this chemical space exploration will help matching the optimal condition for molecule hyperpolarization by selecting the appropriate dye to match $\Delta g = \pm HFCC/2$. In theory, photo-CIDNP is fully compatible with the study of biomolecules. Nevertheless, the barrier of photosensitizer's bleaching would need to be overcome to enable 2D or 3D NMR experiments. The bleaching is the result of both the photosensitizer and the small molecule degradation. Operating with photosensitizers absorbing at higher wave length, avoids the use of ultraviolet lasers which degrades the small molecules, and reduces also the sample heating due to the less energetic photons at play. The photosensitizer degradation is usually resulting from the transformation of the dye into a leucodye, the reduced form of the photosensitizer. The

optimization of the sample conditions to maximize the efficiency of the photo-CIDNP process will reduce the side reactions yielding the production of leucodye, and therefore minimize the dye bleaching.

In conclusion, the handling of semiambiguous distance restraints with simulated algorithm enables the fast structure calculation of protein-ligand complexes, reducing the analysis time from a couple of months to a few hours. The limiting factor is then measurement time, which is dictated by the low sensitivity of NMR. The presented results extend the applicability of photo-CIDNP to further systems and paves the way for the exploration of the critical parameters facilitating the SN enhancement. Our hope, is that in a near future, the sensitivity hurdle will be easily overcome and eventually combined to accelerated analysis methods such as NMR^2 , presented in this thesis, so that NMR could be applied in routine by the medicinal chemist, or other pharmaceutical scientists.

Material and methods

Chapter 2: NMR² method development and application to the bromodomain BRD4-BD2

Compound 1 was synthesized following a reported protocol,²⁴² and the discovery of compound 2 is described in detail.¹¹⁰ Purity of both compounds were >95% as judged by ¹H NMR.

The uniformly ¹³C,¹⁵N-labeled PHD-bromodomain construct of TRIM24 was expressed and purified as described previously.²⁴³

The second bromodomain of human BRD4 (Uniprot ID O60885; residues H341 to Glu460) was cloned into a pET28 vector (Novagen). The protein was expressed in *E. coli* BL21 Gold (DE3) cells in a D₂O based M9 medium containing 1 g/L ¹⁵NH₄Cl and 2 g/L ¹³C-Glucose-d₇. For selective methyl protonation at Ile- δ 1, Leu- δ 1/2 and Val- γ 1/2 the growth medium was supplemented 70mg/L 2-ketobutyric acid-4-¹³C sodium salt hydrate (Isotec) and 120 mg/L 2-keto-3-(methyl-¹³C)-butyric-4-¹³C acid sodium salt (Isotec) 30 min prior to induction. The expression medium contained 50 μ g/mL kanamycin for resistance marker selection and overexpression of protein was induced by the addition of 1mM IPTG at an OD₆₀₀ of 0.6. Cultures were grown over night at 18 °C before harvesting. Cells were resuspended in 50 mM TRIS, pH 8.0, 300 mM NaCl, 1 mM β -mercaptoethanol, 10 mM imidazole containing Complete Protease Inhibitor tablets (Roche) and benzonase nuclease (2.5 u/ml). Resuspended cells were lysed using a Constant Systems cell disruptor at 25 k.s.i. and clarified by centrifugation at 35,000g for 60 min at 4 °C. BRD4 BD2 was purified from the supernatant by nickel-affinity chromatography followed by treatment with TEV protease. Protein was further purified by size-exclusion chromatography using a Superdex 75 10/300 GL column in 20 mM HEPES, pH 7.4, 100 mM NaCl, 1 mM Tris(2-carboxyethyl) phosphine hydrochloride and subsequently concentrated to the required concentrations.

The identity of both proteins used in this work was confirmed by ESI-MS and purity was >95% as judged by SDS-PAGE.

Samples for TRIM24 contained 1.1 mM uniformly ¹⁵N- and ¹³C-labeled TRIM24-PHD-Bromo in 95% H₂O/5% D₂O at pH 7.4 in 50 mM HEPES, 100 mM NaCl, 1 mM TCEP. Compound 1 was added in equimolar amounts from a 20mM stock in DMSO-d₆. Resonance assignments of compound 1 were obtained on a 1mM sample in the same buffer as above.

NMR samples for BRD4 BD2 contained 350 μ M ILV- and otherwise uniformly ¹²C,²H- labelled BRD4 BD2 in 100% D₂O at pH 6.7 in 50 mM Na₂HPO₄, 0.5 mM d-TCEP. Compound 2 was added in 1.2 molar excess with respect to the protein. Excess unbound compound was subsequently removed by passing the sample over a PD-10 desalting column.

All NMR experiments involving protein samples were recorded at 298 K for BRD4 BD2, on a 800 MHz Bruker Avance III spectrometer equipped with a 5 mm TCI cryoprobe with z-axis gradients running Topspin 3.2 in 3 mm NMR tubes. NMR experiments for resonance assignment of compound 2 were carried out at 303 K on a 600 MHz Bruker Avance III spectrometer equipped with a 5 mm TCI cryoprobe with z-axis gradients running Topspin 3.5 in 5 mm NMR tubes.

For BRD4 BD2 bound to compound 2 a series of 4 double half-filtered NOESY experiments deploying an IPAP element to acquire all four combinations of $^{12}\text{C}/^{13}\text{C}$ filtering, with mixing times of 50, 100, 150 and 200 ms were recorded. Each experiment was acquired with 2048x512 complex points in the direct and indirect dimensions respectively for an experimental time of ca. 35 hours per spectrum. The spectra covered 11 kHz and 9.5 kHz in the direct and indirect dimension, respectively. Resonances of compound 2 were assigned with the help of double-half filtered DQF-COSY acquired with 2048x300 complex points, 16 scans covering 12.5 and 9.5 kHz in the direct and indirect dimension respectively. The previously described double half-filtered NOESYs were also used to aide compound resonance assignment. All spectra were processed with Topspin 3.1 (Bruker) and evaluated with ccpNMR analysis 2.4.²⁴⁴

τ_c the rotational correlation time of the protein, 10.5 ns, derived from the $^{15}\text{N}-T_1$, $^{15}\text{N}-T_{1\rho}$ relaxation rates using the software TENSOR2 and assumed to be the same for both bromodomain complexes.²⁴⁵ The NMR^2 method does not use a force field but instead employs a hard sphere repulsion model for the atoms as described in the program CYANA.²⁴⁶

Chapter 3: NMR^2 for fragment-based drug design

The NMR^2 calculations were performed on the OS Ubuntu 20.0 and uses slurm workload manager to utilize a cluster of 20 desktop with 4 CPU's each (Intel Xeon E3-1245 v5 @ 3.50GHz). The calculations were run restricting the number of CPU to 10 in order to be able to perform several structure calculations in parallel. The NMR^2 simulated algorithm calls the CYANA software to perform the structure calculation. NMR^2 does not set up the seed that initiate the trajectories' speed for the structure calculation. Therefore, the default seed is always used, yielding to similar structure calculation results. However, the error has an impact on the TF calculation by reducing the number of violations. Indeed, when error is introduced, the distance of the resulting calculation does not need to match exact distance restraints, but to fit in a range defined between the lower limits (lol) and the upper limits (upl). The NMR^2 calculations were performed with a side chain mobility of 20 degrees for the residues present in the binding site. The maximum number of files per structure was set up as 1000. Binding site was specified. The intermolecular distance restraints were measure between the protons of the ligands and the methyl's carbons from the proteins. Therefore, the anchor mode for the protein side was defined as carbon.

The introduction of white noise in order to simulated the error on the distance restraints was done as follow. To the exact distances (dist) is added or subtracted either 10% or 20% of their value. The resulting distances are lol = αdist , and the upl = βdist . α is equal to 0.9 or 0.8 for errors of 10% and 20%, respectively. β is equal to 1.1 or 1.2 for errors of 10% and 20%, respectively. The calculations of the complex structures use instead of a single exact distance restraints list, two lists of restraints: the lol and the upl lists.

PIN1 was expressed as doubly [^{15}N , ^{13}C]-labelled protein in order to discriminate between the NMR signal of the protons from the protein and the protons from the fragments. Three NMR samples containing 1.3-1.5 mM PIN1 and 3 mM ligands were prepared for subsequent NMR measurements (Material and methods). The protein methyl group resonances were identified by collecting ^{13}C -ctHSQC spectra for each complex (see below). The fragment assignments were straightforward using 1D ^1H NMR spectra in the free form and in complex with the PIN1, as well as the F1- ^{15}N , ^{13}C]-Filtered [^1H , ^1H]-NOESY experiments.

Protein expression, compounds and NMR samples [^{15}N , ^{13}C]-labelled recombinant PIN1 was produced as previously reported. PIN1 was buffer exchanged to 100% D $_2$ O, phosphate 20 mM, sodium chloride 50 mM at a pH of 6.6 and concentrated to 1.3 mM for 1 and 3, and to 1.5 mM for 2. The ligand concentration in the samples used for the NOESY experiments were 2.5, 3, and 3.5 mM for 1, 2, and 3, respectively. Compounds 1, 2, and 3 were ordered from Enamine Ltd.

All NMR experiments were recorded at 298 K. F_1 - ^{13}C , ^{15}N]-filtered [^1H , ^1H]-NOESY spectra were measured with [^{13}C , ^{15}N]-labelled PIN1 and unlabelled ligands on a Bruker Avance 900 MHz spectrometer with cryoprobe, with 2048 complex points in the direct time dimension (t_2) and 241, 256, 225, and 200 complex points for 1, 2, 3, and 4 respectively. The interscan delay was set to 1.5 s, the scans per increment was 160, $t_{2,\text{max}} = 190.05$ ms, and $t_{1,\text{max}} = 22.44$ ms for 1, 23.75 ms for 2, and 20.87 ms for 3. The mixing times used for the NOE build-ups of 1 were 30, 70 ms, 2: 40, 60, 90, 120 ms, 3: 30, 60, 90 ms, and 4: 40, 70, 100 ms. Constant time ^{13}C -HSQC (^{13}C -ctHSQC) for titration, $T_{1\rho}$ and T_1 ^{15}N -HSQC for relaxation experiment and F_1 - ^{13}C , ^{15}N]-filtered [^1H , ^1H]-TOCSY for fragment proton assignments were recorded on 600 MHz Bruker Avance spectrometer. PIN1 ^{15}N - T_1 and ^{15}N - $T_{1\rho}$ experiments were recorded at 0, 20, 40, 60, 80, 100, 120, 140, 180, 200 ms and 5, 10, 20, 30, 40, 50, 70, 90, 100, 120 ms relaxations delays, respectively. 64 (t_1) * 2048 (t_2) complex points were used with an interscan delay of 3 s (2 s for T_1), 32 scans per increment, $t_{1,\text{max}} = 138.35$ ms, and $t_{2,\text{max}} = 121.65$ ms (243 ms for T_1). Spectra were processed with NMRPipe.²⁴ For the assignment of the fragments, filtered F_1 - ^{13}C , ^{15}N]-filtered [^1H , ^1H]-TOCSY spectra were recorded with 1024(t_1)*2048(t_2) complex points, an interscan delay of 1.2 s, 128 scans per increment, $t_{1,\text{max}} = 13.30$ ms, and $t_{2,\text{max}} = 121.65$ ms. For the titration of 2 and 3, ^{13}C -ctHSQC spectra were recorded with 256(t_1)*2048(t_2) complex points, an interscan delay of 1.2 s, 8 scans per increment, a $t_{1,\text{max}} = 71.08$ ms, and a $t_{2,\text{max}} = 141.99$ ms. The fragment concentrations used for the titration were 60, 156, 252, 348, 444, 540, 636, 732, 828, 924, 1020, 1116, 1308, 1444 μM for 2 and 296, 592, 988, 1183, 1479, 1775, 2071, 2367, 2663, 2959, 3255, 3551, 3847, 4143, 4439, 4735, 5024, 5320, 5616 μM for 3. The spectra were processed with topspin 3.1 (Bruker) and analysed with ccpNMR Analysis 2.3.

Distances were computed from NOE build-up curves, assuming an isolated two spin system, using the program eNORA. The theoretical frame of eNORA distance restraints calculation is discussed in Chapter 2.

The NMR2 structure calculations were performed according to the published protocol for the complex PIN1-1, and the complexes PIN1-2 and PIN1-3 were calculated knowing the protein methyl group assignments. The fragment geometries were kept fixed during the calculation and optimized in a two steps process: first a molecular mechanics energy minimization was run with Avogadro (UFF force field, steepest descent algorithm); then an ab-initio energy minimization was run with Gaussian (6-31G(d,p) basis set, DFT B3LYP method). Frequency calculations were run afterwards and the Hessians were all positives. The structure activity relationship was derived from the pharmacophores obtained with LigandScout[®] software.

Chapter 4: NMR-based drug discovery for SARS-CoV-2

The Mpro has been recombinantly expressed according to the protocol published elsewhere.¹⁷⁰

The NMR screening were performed on a 600 MHz Bruker Avance III spectrometer equipped with cryoprobe. The spectrometer was equipped with a SamplerJet autosampler, and an in-house script was used to run the samples by batch of 80 tubes.

The STD spectra were acquired with 256 scans, 16k complex points, and an interscan delay of 2 s. The protein saturation is conducted with a train of gaussian-shaped frequency selective pulses of duration 50 ms. The train of pulses consists into 40 repetitions to reach 2 s irradiation. The on-resonance saturation rf-pulse was set at -1 ppm (-600 Hz), and the off-resonance saturation rf-pulse was set at 40 ppm (24000 Hz). The protein signal was filtered using a spinlock for 20 ms at 10 kHz power level.

The TROSY spectra were acquired with 128 scans, 128 (t_1) * 2048 (t_2) complex points, and an interscan delay of 0.6 s.

The samples for the Minifrag library screening were prepared with 20 mM phosphate buffer pH = 7.2, NaCl 50 mM, 0.050 mM Mpro and 1 mM fragment.

The samples for the FDA approved library screening were prepared with 20 mM phosphate buffer pH = 7.2, NaCl 50 mM, 0.01 mM Mpro and 0.200 mM ligand.

The enzymatic assay conditions were 20 μ M substrate 30 μ M compound and 5 μ M enzyme, in buffer: Tris-HCl 50 mM, EDTA 1 mM pH=7.55.

Chapter 5: Exploration of the photo-CIDNP performances investigating new photosensitizers and chemical modifications of tryptophan

The NMR measurements were performed at 298 K either on a Bruker Avance III 600 MHz spectrometer equipped with a cryoprobe or on a Bruker Avance 200 MHz spectrometer equipped with a room temperature probe. The irradiation of AT12 samples was performed with a Coherent Verdi V10 diode pumped solid state laser emitting at a wavelength of 532 nm. The laser used for the fluorescein samples was a Thorlabs L450P1600MM, a diode laser emitting at 450 nm. The laser light was coupled (using appropriate coupling optics) into an optical fiber (Thorlabs, FG950UEC) of length 10 m and a diameter of 0.95 mm. The end of the fiber was inserted into the sample solution in a 3 mm NMR tube to a depth of about 5 mm above the NMR coil region. The TR-photo-CIDNP experiments were performed on a Bruker DPX 200 MHz NMR spectrometer, irradiation was performed with COMPLEX Lambda Physik XeCl excimer laser (wavelength 308 nm, pulse energy up to 150 mJ).

The HPLC analysis and initial purification was performed on an Agilent Technologies® 1200 LC system using a semi-preparative Kinetex® C18 column. The runs were performed with an acetonitrile (0.1% trifluoroacetic acid, TFA) gradient going from 3 to 30 % in water (1% TFA) in 15 min at a flow rate of 3 mL/min, at 40°C. The UV spectra were recorded with a Jasco® V-650 UV-vis spectrometer at a concentration of 0.01 mM. Tyrosine and tryptophan (Sigma) were prepared as stock solutions of 0.2 mg/ml and 0.18 mg/ml, respectively, in a 0.1 M sodium/potassium phosphate buffer (pH 7.1) with 5% D₂O. The stock solution of AttoThio 12 (AT12) was 1 mg/ml in H₂O. To prevent dye quenching the enzyme cocktail Glucose oxidase (Go, 120 kDa), catalase (Cat, 240 kDa) and D-Glucose (G, 180 Da) was used at a concentration of 14 nM for each enzyme and 2.5 mM of Glucose, as described elsewhere^{194, 247}. The stock solutions were 0.25 μ M for Go and 0.16 μ M for Cat, respectively. The glucose stock solution was 500 mM in D₂O with 0.02% NaN₃. All the samples were prepared in a 100 mM KPO₄ buffer at pH = 7.1 with either 20 μ M AT12 or 25 μ M fluorescein, and 100 μ M target molecule (i.e. tyrosine,

tryptophan or HOPI) unless otherwise specified. The sample for TR-photo-CIDNP was prepared using 4 mM of HOPI and 4 mM of TCBP dye at pH = 7.5.

The T1 relaxation of the different HOPI diastereoisomers has been measured with NMR. The data were recorded using the Topspin pulse sequence (t1ir) modified to introduce a w5 water suppression scheme and the relaxation time were fitted with the fitting tool from Topspin 3.2.

Chapter 6: Basis for the empirical exploration of the photo-CIDNP performances

The NMR measurements were performed at 298 K either on a Bruker Avance III 600 MHz spectrometer equipped with a cryoprobe. The irradiation of AT12 samples was performed with a Coherent Verdi V10 diode pumped solid state laser emitting at a wavelength of 532 nm. The laser used for the fluorescein samples was a Thorlabs L450P1600MM, a diode laser emitting at 450 nm. The laser light was coupled (using appropriate coupling optics) into an optical fiber (Thorlabs, FG950UEC) of length 10 m and a diameter of 0.95 mm. The end of the fiber was inserted into the sample solution in a 3 mm NMR tube to a depth of about 5 mm above the NMR coil region.

Tyrosine, tyramine, tryptophan, tryptamine, IPA, IAA were purchased from Sigma, dH-TRP was purchased from Akos Pharma, and PEI was purchased from ChemSpace LLC. HOPI was synthesized in-house according to the previously published protocol²³¹. were prepared as stock solutions of 0.2 mg/ml and 0.18 mg/ml, respectively, in a 0.1 M sodium/potassium phosphate buffer (pH 7.1) with 5% D₂O. The stock solution of AttoThio 12 (AT12) was 1 mg/ml in H₂O. To prevent dye quenching the enzyme cocktail Glucose oxidase (Go, 120 kDa), catalase (Cat, 240 kDa) and D-Glucose (G, 180 Da) was used at a concentration of 14 nM for each enzyme and 2.5 mM of Glucose, as described elsewhere^{194, 247}. The stock solutions were 0.25 μM for Go and 0.16 μM for Cat, respectively. The glucose stock solution was 500 mM in D₂O with 0.02% Na₃N. All the samples were prepared in a 100 mM KPO₄ buffer at pH = 7.1 with either 20 μM AT12 or 25 μM fluorescein, and 100 μM target molecule.

Appendix 1: Structure calculation of the proSpC-Brichos domain

Introduction

The brichos domain contained in C-terminal part of the surfactant protein C (proSP-C) was suggested to act as a chaperone monitoring SP-C folding and preventing its aggregation.²⁴⁸ The proSP-C protein is later cleaved to form SP-C.²⁴⁹ Other brichos domains are found in the genome for which mutations are correlated to aggregative diseases, and mutations of the proSP-C brichos were identified in the interstitial lung disease or respiratory distress syndrome.²⁴⁸ Furthermore, the brichos domain has been shown to prevent amyloid fibrillization both *in vitro*²⁵⁰ and *in vivo*.²⁵¹ Later, aggregation kinetics of $\alpha\beta$ -42 using time dependent thioflavin T (ThT) fluorescence, have shown that brichos inhibits the secondary nucleation of the amyloid fibrils.²⁵² While the structure of the ProSP-C brichos domain could be solved by X-ray crystallography, the very flexible loop region (67-96) is not characterized.²⁵³ Our hypothesis is that this flexible loop region has a critical role in the brichos domain dynamics and complementary structural information obtained by NMR would provide insights into the loop structure. Brichos domain organize themselves as trimers, as of their important hydrophobic surface. Moreover, it has been shown that monomer is the active form of brichos.²⁵⁴ The trimer-monomer equilibrium can be displaced using small molecules or with single point mutation. In this study we present the NMR structure of the brichos trimer including the 67-96 loop, and new monomeric mutants with improved stability. The activity of the mutants against aggregation is cross-checked to validate the best of them. The structure of the more stable and more efficient aggregation inhibitor is calculated and compared with the wild type (wt)-trimer structure. Such comparison enlightens the structural features for brichos chaperone activity and aggregation prevention. Finally, we demonstrate the inhibition of α -synuclein with brichos, which was not revealed so far.

Results and discussion

Structure calculation of the wt-brichos trimer

The wt-brichos was recombinantly produced in *E. coli*. BL21*, with a GST-TEV-BRICHOS construct resistant to Kanamycin. The bacteria were grown in M9 media in the presence of ¹⁵NH₄Cl and ¹³C-glucose until optical density reaches 0.6, and then production phase was induced by introducing 1mM IPTG to the media. The sample produced for the backbone assignment of the trimer was produced in 70% D₂O M9 media.

The first step of the structure calculation consists into the backbone resonance assignment. Given the size of the trimer (34 kDa), the sample was partially deuterated (see above) and the experiments were recorded in their TROSY version, for resolution and sensitivity purposes.³² Hence, five different experiments corresponding to different polarization pathways have been measured: HNCa, HN(CO)Ca, HNCaCb, HNCO, and HN(Ca)CO (Figure A1.1A).²⁵ Although different, these five experiments rely on the magnetization transfer through the scalar coupling between the atoms of the protein backbone. Therefore, it is possible for a residue in the position *i* to identify the H^N, N, C α , C β and CO resonances.

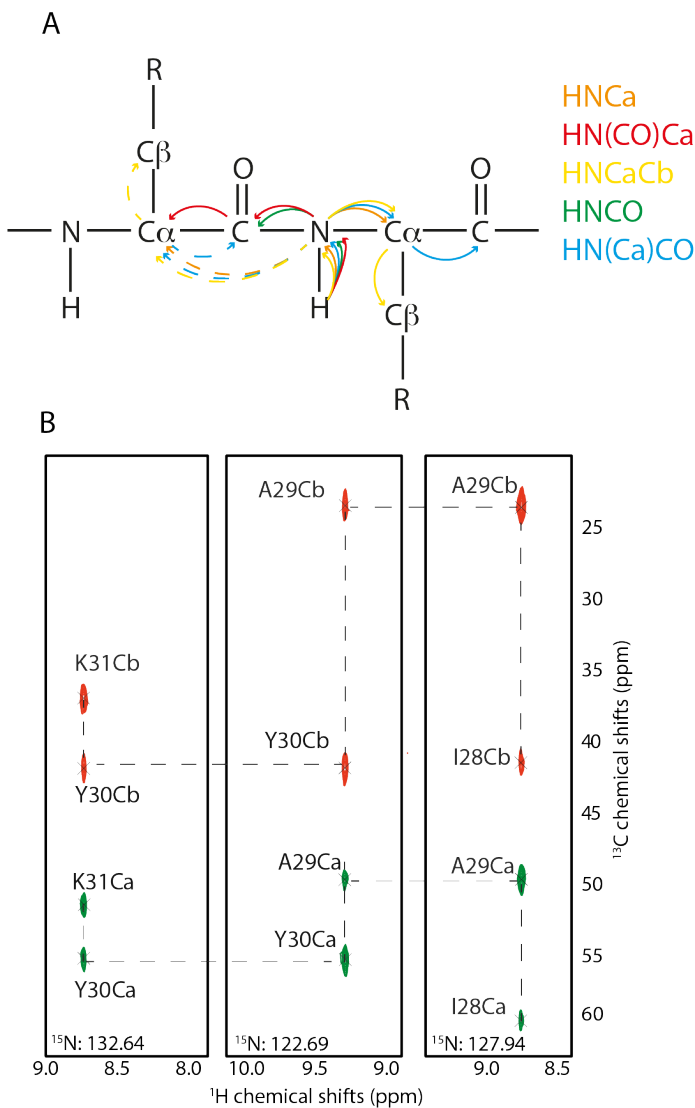


Figure A1.1:A) Magnetization pathway of the different 3D NMR experiment measured for backbone assignment, the nuclei into parenthesis are used as magnetization relays but are not marked with time evolution. B) Strip representation of different planes (^{15}N dimension) of the HNCaCb spectrum, the alignment of the signals enables to identify the residue in $i-1$ position. Furthermore, the characteristic chemical shift of the C_β of the alanine enables its identification.

Moreover, in the HNCa , HNCaCb , HN(Ca)CO , it is possible to observe the connection to the residue $i-1$ (Figure A1.1B) as of the very similar J_{NCa} coupling.⁵⁵ The only residues which are not

observable are the proline since they are missing the amide proton; yet they are observed as i-1 residues, which can be used as an information to evaluate the presence of a proline at this position. The characteristic upfield chemical shifts of the glycine C α (~40 ppm), and of the alanine C β (15-25 ppm) are useful to identify the residue type. Furthermore, the C β of the threonine and serine residues is downfield as compared to their C α in the HNCaCb experiments, and the C α and C β are identified thanks to their phase. Therefore, the residues Thr/Ser are easily identified. *In fine*, from this set of assignment it is possible to link the residues together thanks to the i, i-1, connectivity and to identify the residues Pro, Gly, Ala, Thr/Ser. Altogether with the protein sequence, the assignment is thus straightforward. 97% of the residue's amides could be assigned, and more than 90% of the C α , C β , and CO resonances. The backbone amide assignments are presented in the Figure A1.2. The chemical shifts of the backbone atoms can be used as input to determine the secondary structure of the sequence, using the Talos+ webserver.²⁵⁵ Talos+, provides an overview that can be aligned with the X-ray structure to have an estimate of the validity of the assignments and it provides angle restraints corresponding to the determined secondary structure, that will be later of use for the structure calculation.

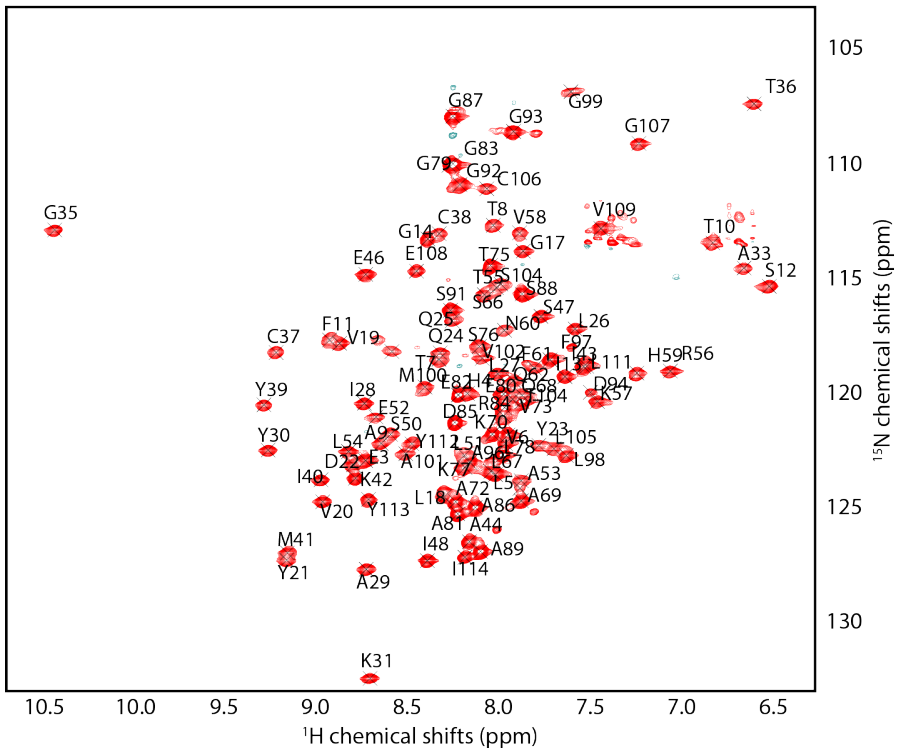


Figure A1.2: Backbone amide assignment represented on the ^{15}N -TROSY spectrum of wt-brichos.

The sidechain assignment is a compulsory step obtain a complete set of cross-relaxation NOE peaks providing the distance restraints to calculate the 3D structure of the brichos trimer. Moreover, it can be used to validate the backbone assignment; it is usually done by measuring

and assigning a HCCH-TOCSY spectrum in complementation to the NOESY spectra. The cross-relaxation peaks from which are derived the distance restraints are obtained from ^{13}C , ^{15}N -HMQC-NOESY spectra.²⁵⁶ Four HMQC-NOESY spectra with 25, 35, 45, and 55 ms mixing time were measured and the resonance assignments obtained from the other spectra have been transferred and the cross-peaks resulting from the cross-polarization between the different protons are assigned as well. Finally, 2538 distance restraints were obtained, comprising among them long range distance restraints, and distance restraints defining the interface. The structure of the trimer was calculated with CYANA,²⁵⁷ including the Talos+ angular restraints,²⁵⁵ the definition of the disulfide bridges. Furthermore, the recording of the ^{13}C -HSQC of a 10% ^{13}C -labeled sample provided the stereospecific assignment.²⁵⁸ Finally, the trimer structure was obtained with a backbone RMSD of 2.46 Å and a heavy atom RMSD of 1.66 Å (Figure A1.3). While the folded part of the structure overlays well with the reference X-ray structure (PDB: 2yad), the core seems more tightly packed than in the X-ray, which is explained as the CYANA structures are often a bit more packed than the X-ray crystallography structures. A solution to this problem could be a refinement in explicit water with an adapted forcefield, such as CNS (see Appendix 2). Interestingly, despite its supposed high mobility, the loop which was absent in the X-ray structure is rather well defined in the NMR-based structure.

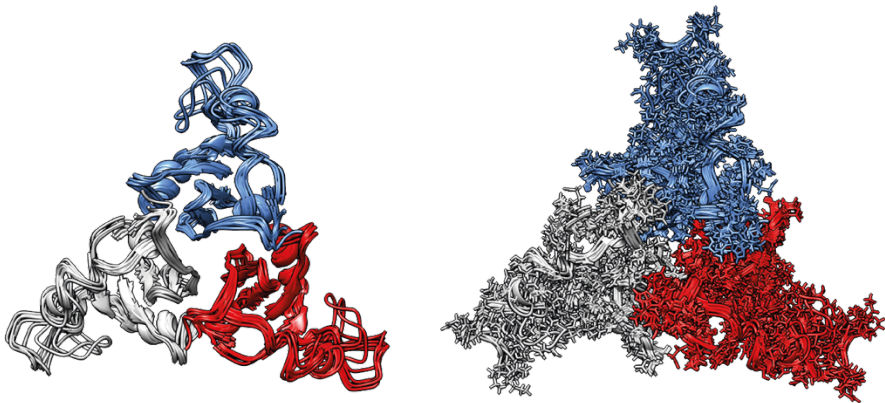


Figure A1.3: NMR structure of the trimer brichos. While the backbone structures overlay well, the sidechains spatial distribution is less well defined.

Nonetheless, the TF of the structure was high, with a value of around 250 for the 10 structures constituting the bundle. The apparent reason for that is the presence of many van der Waals violations. The van der Waals violations are resulting from the very diverse mobility in the protein, and the resulting different auto/cross-relaxation rates. Therefore, the NOE of the loop are extremely intense and the core NOE are much weaker, leading to biased distance restraints values, which could not be corrected using normalized NOE. Finally, the structure provides a model aligning well, visually, with the reference x-ray structure (pdb: 2yad), but of a quality not sufficient for deposition. Moreover, the structure of the loop is now known and its relative importance in the chaperone activity of brichos can be studied at an atomic level.

Design and selection of the brichos mutant

Evidences that brichos monomer is the active form of the protein accumulate in the literature. Indeed, the addition to the brichos trimer of Dianilino-1,1'-Binaphthyl-5,5'-Disulfonic 4,4'- Acid (bis-ANS) is known to break the non-covalent intermolecular interactions stabilizing the trimer, and the addition of bis-ANS-brichos to $\alpha\beta$ -amyloids shows a stronger inhibition of the secondary nucleation as compared to the addition of brichos alone. This suggests that the monomeric form is the active form of brichos.²⁵⁴ The interaction of bis-ANS with brichos has been characterized recorded the CSP upon addition of bis-ANS to a brichos sample (Figure A1.4A). It appears that the bis-ANS interacts with the trimer interface at the α -helix part, and with the hydrophobic core of the protein with Tyr, Ile, and Leu residues in the β -sheets (Figure A1.4B).

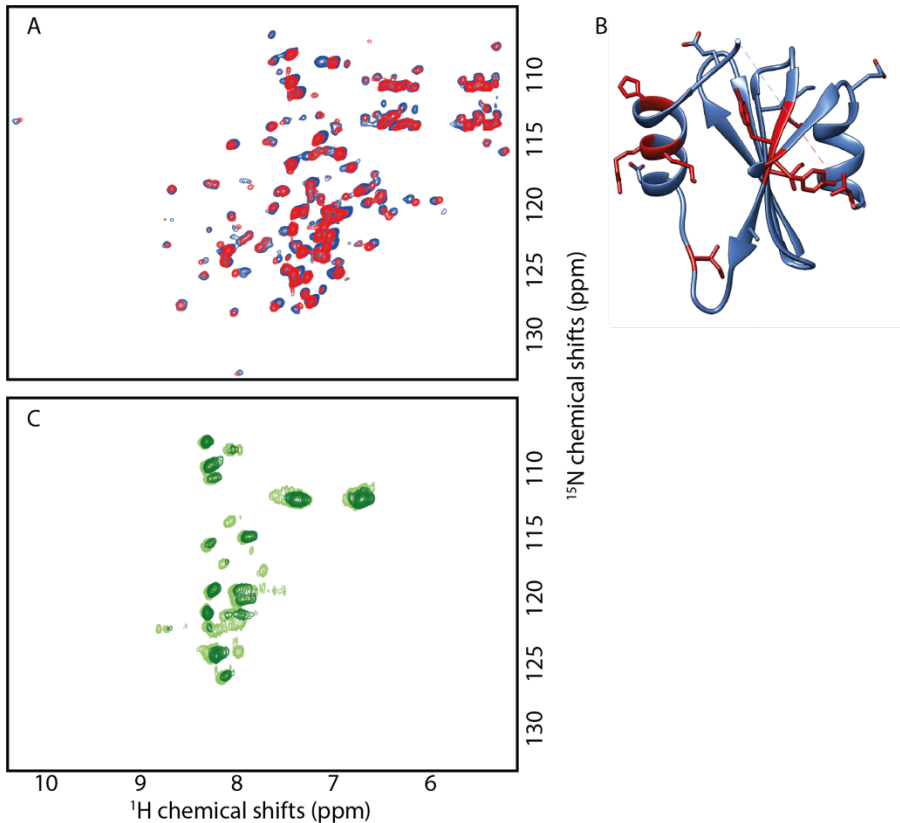


Figure A1.4: A) ¹⁵N-TROSY spectra of brichos (red) and brichos + bis-ANS (blue). B) Brichos structure (pdb: 2yad) with in red the residues for which a CSP is observed upon bis-ANS addition C) brichos ¹⁵N-TROSY spectra at higher concentration of bis-ANS, green 0.5 mM and dark green 1.5 mM.

Nevertheless, the structure characterization of the brichos-bis-ANS complex, in order to explain its improved anti-aggregation properties, requires the use of bis-ANS concentration between 0.5

to 1.5 mM, which provokes the unfolding of brichos (Figure A1.4C). This latest data hinders the structure calculation of a brichos-bis-ANS complex. On the other hand, the mutant S12R yields to the introduction of three positive charges at the trimer interface, the resulting repulsive interaction destabilizes the trimer towards a monomer. Once more, the inhibitory activity of the S12R mutant on the secondary nucleation has been shown to be improved as compared to the wild type brichos.²⁵⁴ The sequence of the S12R mutant is provided below:

GSEHLVTTATF**R**IGSTGLVVYDYQQLLIA YKPA PGTC CYIMK IAPESIPSLEALTRK VHN
FQMECSLQAKPAVPTSKL GQA EGRDAGSAPSGGDP AFLGMAVSTLCGEVPLYYI

The ¹⁵N-TROSY spectrum of the S12R mutant has been recorded with the view to calculate the structure (Figure A1.5). Nevertheless, the spectrum shows poor signal to noise and misses many signals in the beta-sheets region (>8.5 ppm). Such a spectrum suggests that the brichos is partly unfolded, which could be explained by the introduction of cationic residue in the hydrophobic interface of the protein. Furthermore, the introduction of a charged residue changes the isoelectric point of the protein and affects its stability. Hence, the S12R mutant was discarded as a possible system that could be studied by NMR, and mutations favorizing monomeric brichos, without affecting its stability were considered.

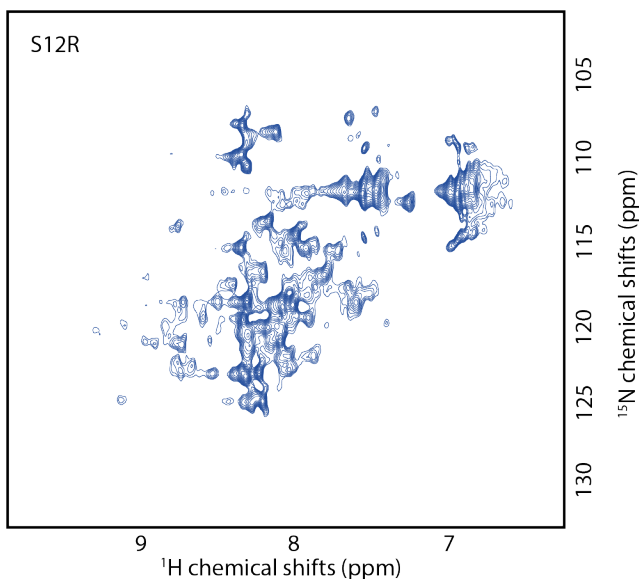


Figure A1.5: ¹⁵N-TROSY spectrum of the S12R brichos mutant.

To design these mutations, we focused on the two intermolecular salt bridges at the interface of the trimer, K31-E52 and E108-R56 (Figure A1.6). The working hypothesis was that these salt bridges were the main interaction stabilizing the trimer, as the typical energy contribution of such interactions is around 40 kJ/mol each. Two strategies have been imagined, with the first consisting in neutralizing the salt bridge electrostatic by replacing the lysine 31 by an isoleucine and the arginine 56 by a glutamine (K31I/R56Q). The second option is to create repulsive electrostatic by mutating the glutamate 52 by an arginine and the arginine 56 by a glutamate (E52R/R56E), the double switch is operated so that it preserves the isoelectric point of the protein.

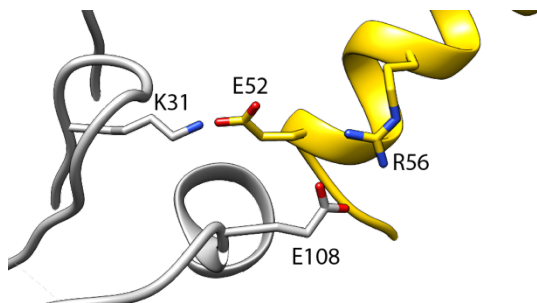


Figure A1.6: wt-brichos trimer interface. The grey and yellow colors correspond to different monomers, respectively. The residue identifiers are displayed as uniletter code and sequence position.

The two double mutant sequences are provided below.

K31I/R56Q:

GSEHLVTTATFSIGSTGLVVYDYQQLLIAYIPAPGTCCYIMKIAPEIPSLEALTQK**V**HNF
 QMECSLQAKPAVPTSKLQAEGRDAGSAPSGGDP AFLGMAVSTLCGEVPLY**Y**I

And E52R/R56E:

GSEHLVTTATFSIGSTGLVVYDYQQLLIAYKPAPGTCCYIMKIAPEIPSL**R**ALTE**K**VH**N**
 FQMECSLQAKPAVPTSKLQAEGRDAGSAPSGGDP AFLGMAVSTLCGEVPLY**Y**I

The mutants were produced following the same protocol as for the wt-brichos. The integrity of the mutants was assessed by performing the ^{15}N -TROSY so that the downfield CS witnessing from the secondary structure folding could be observed or not (Figure A1.7). The spectrum from the K31I/R56Q shows a more important number of peaks in the β -sheets region and the peak intensities and line width are more homogeneous in the spectrum suggesting a folded monomeric protein. Moreover, the multi angle light scattering experiment for K31I/R56Q revealed the presence of monomer in large excess as compared to the trimer (not shown). The mutant E52R/R56E was showing several peaks in the folded region of the ^{15}N -TROSY spectra, yet fewer than for the K31I/R56Q mutant. Moreover, the peak intensities and linewidth were varying significantly suggesting either partial unfolding or monomer-trimer equilibrium. Despite the use of an arginine buffer at low temperature to favorize the monomer, the β -sheets region was not significantly improved (Figure A1.7). The mutant K31I/R56Q was therefore selected for further structural investigation. The low overlap of the ^{15}N -TROSY signals suggests important conformational change operating between the wt-trimer and the K31I/R56Q mutant monomer.

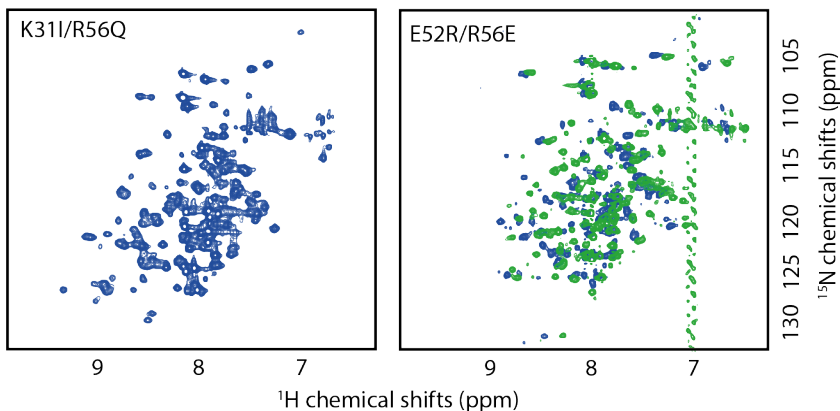


Figure A1.7: ^{15}N -TROSY spectrum of the K311/R56Q and the E52R/R56E double mutants. The green spectra of E52R/R56E is after addition of arginine buffer 50 mM and decrease of the temperature to 277 K for stabilization.

In parallel, the effect of brichos and its mutants on the aggregation kinetics of α -synuclein is determinant in the choice of the mutant to structurally investigate. Therefore, ThT fluorescence experiments where the fluorescence increases in stoichiometrically with the number of aggregated monomers²⁵⁹ have been conducted by collaborators in Prof. Tuomas Knowles group (Figure A1.8A). The introduction of wt-brichos reduces the aggregation by decreasing the secondary nucleation rate, and it is even reduced further upon the addition of the K311/R56Q (Figure A1.8B). However, the secondary nucleation is less efficiently inhibited using the E52R/R56E (Figure A1.8B) suggesting that the monomer stabilization using the proper mutations (K311/R56Q) was critical for the anti-aggregation properties of brichos.

The characterization of the secondary or tertiary structure modifications will provide an insight into the key features for the anti-amyloid brichos activity. Hence, double labeled ^{13}C , ^{15}N -K311/R56Q -brichos was expressed and the experimental set of experiment necessary for resonance assignment and structure calculation was measured: ^{13}C -HSQC, trHNca, trHN(CO)Ca, trHNCO, trHN(Ca)CO, trHNcaCb, HCCH-TOCSY, ^{15}N -TOCSY, [^{13}C , ^{15}N]-HMQC-NOESY, and HNHa.²⁵ The stereo-assignments of methyls were obtained from the ^{13}C -HSQC recorded for the 10% ^{13}C -labeled sample.²⁵⁸ The quality of the spectra was significantly better than for the wt-brichos, and promising preliminary structure calculation could be obtained by another Ph.D. student of the group, Dzmitry Ashkinadze, with a RMSD of 1.6 Å. The inclusion of $^3\text{J}(\text{H}_\text{N}\text{H}_\alpha)$ scalar coupling obtained from the HNHa spectrum, and eventually of residual dipolar couplings will improve the quality of the distance restraints in order to obtain multistate structure of the K311/R56Q-brichos monomer. Meanwhile, major rearrangement of the ternary structure is observed for the K311/R56Q in comparison to the wildtype (Figure A1.9A). This conformation change consists in the rearrangement of the 67-96 loop to the other face of the protein monomer (depicted in green in Figure A1.9A), and the bending of the core β -sheet, resulting in the exposure of so-far embedded residues (in blue in Figure A1.9A). Moreover, the position of the 67-96 loop seems to pull the α -helix in orange (Figure A1.9A), which unfolds and goes to the other face of the protein, uncapping the β -sheet residues in the back.

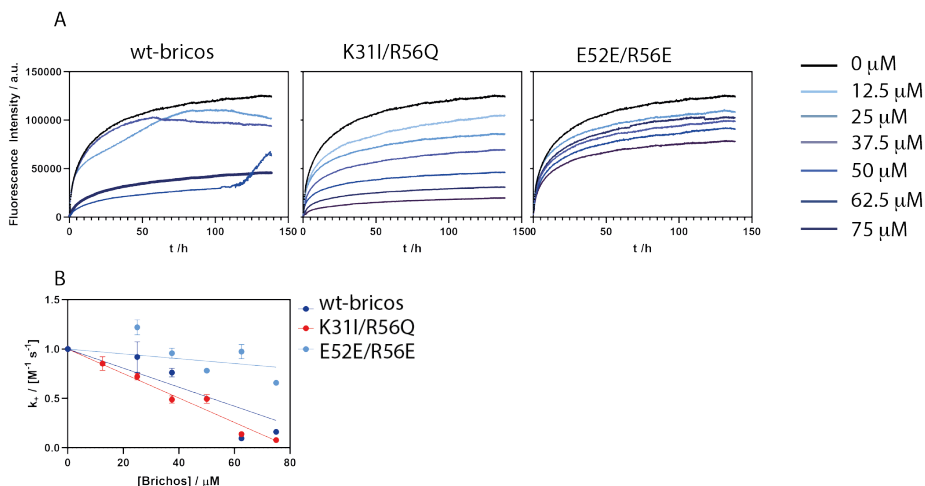


Figure A1.8: A) aggregation kinetics followed by ThT fluorescence increase for the

The revealed ternary structure of K311/R56Q is correlated to improved anti-aggregation properties. It is known that α -synuclein fibrils propagate thanks to the hydrophobic interaction with soluble monomers of α -synuclein, and the inhibitory mechanism of action of brichos is assumed to be the competitive binding to the hydrophobic pro-aggregating surface of α -synuclein fibrils. The coulombic surface of the wt-brichos and of the K311/R56Q-brichos is displayed in the Figure A1.9B and C, respectively. Interestingly, the coulombic surfaces reveal that the hydrophobic residues are solvent exposed in the mutant monomer while embedded in the wild type (compare Figure A1.9B and C). This is consistent with the observation in Figure A1.9A, as most of the hydrophobic residues are in the β -sheet, which was capped by the orange α -helix (residues 47-62) and embedded in the trimer interface. The different structures of the trimer and the monomer suggest that the displacement of the equilibrium from a trimer to a monomer induces an important rearrangement of the ternary structure which stabilizes an exposed hydrophobic surface which can competitively interact with the pro-aggregating α -synuclein surface.

Conclusion and outlook

Monomeric brichos is known to be the active form of the chaperone. Furthermore, the brichos monomer mutants show improved anti-aggregation properties as compared to the trimeric wt-brichos. We developed a stable monomer mutant (K311/R56Q) which is also a potent inhibitor of the α -synuclein aggregation. Moreover, these improved performances are shown to be associated to major structural changes. These changes involve the greater exposure of the hydrophobic surface which we hypothesize to be the reason for better affinity to the pro-aggregating hydrophobic patches of the α -synuclein fibrils.

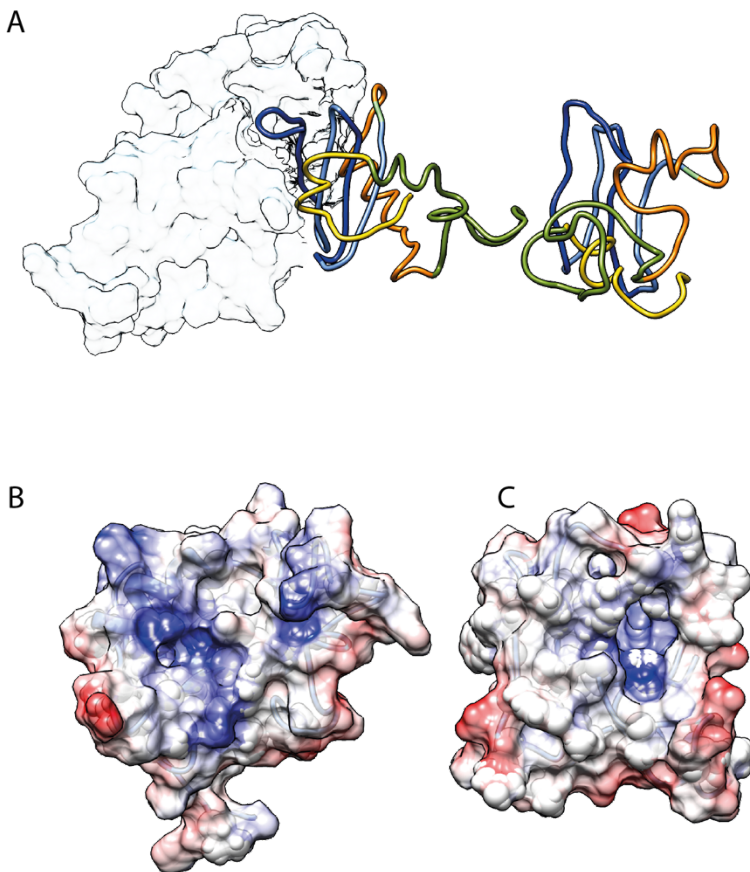


Figure A1.9:A) structural rearrangement of the K311/R56Q-brichos mutant . B) wt-brichos coulombic surface. C) K311/R56Q-brichos coulombic surface. The residues represented as spheres in the transparency correspond to the hydrophobic residues.

In order to confirm this hypothesis, relaxation measurement to probe the dynamics of the mutant should prove the mobility loss of the 67-96 loop thus validating the monomer as a stabilized structure. Moreover, the recording of 2D STD experiments will be used to mark the interaction surface of the K311/R56Q-brichos with α -synuclein fibrils and hopefully confirm the hydrophobic surface as the interacting site. Finally, obtaining the multistate structure will provide further insight into the structural features enabling the stability of an important hydrophobic surface in aqueous solution.

Appendix 2: NMR^2 structure deposition onto the PDB

The structures calculated using NMR^2 relies on experimental distance restraints and can thus be accepted as a structure.⁴³ Therefore, this opens the possibility to deposit NMR^2 structures onto the protein data bank (PDB), so that scientists from other groups can retrieve the structural information regarding the protein-ligand complex structures and use them for structure-based drug design. Nevertheless, the data submitted to the PDB are quite sparse because the protein intramolecular distance restraints are missing. On another hand, the structures are calculated using the CYANA engine which force-field is different from the force-field used in PDB, and the structure quality check may yield to poor results as of the difference in the van der Waals violations definition.²⁵⁷ Also, the electrostatics are often not present in the CYANA calculated output. To improve the quality of the structure towards PDB quality standards, we perform a water refinement in crystallography and NMR system (CNS) force field.²⁶⁰⁻²⁶¹ The description of the steps to perform the water refinement and later the PDB submission are described in this appendix chapter.

PDB submission and CNS water refinement

The protein data bank (PDB) was established in 1971 with the aim to archive the structures of biomolecules such as DNA, RNA, and proteins.²⁶² Since then PDB deposition ensure the public availability of these data which can be used for different purposes: molecular dynamic simulations, docking, homology modelling and recently NMR^2 . The worldwilde PDB organization operates as a single repository. The influence of structure deposition on structural biology and new medicine development makes it a necessity. Unlikely to the other accelerated structure determination techniques which are providing models,⁴⁶ NMR^2 output can be interpreted as a structure and is therefore eligible to be deposited in PDB. Nevertheless, the complex structure is obtained partially from an originally deposited structure, and the restraints traditionally required for PDB validation task force are not fulfilled here. The only restraints available are the ligand intramolecular distance restraints and the ligand-protein intermolecular distance restraints.

Moreover, the NMR^2 structures are calculated using the CYANA engine which force-field is adapted to the torsional angle structure calculation. As a result of this, the structures calculated with CYANA are often too packed yielding to van der Waals violations when validated within the PDB. In order to reduce the number of violations, the structures can be relaxed in explicit water in the CNS force field. The thus refined structures suffer less violations. One could perform the water refinement using the ARIA software suit. However, because the file translation is tedious in this precise case, the scripts from the RECOORD repository where used.²⁶³

Water refinement in CNS

Ligand parametrization

The ligand's topologies and parameters are not included in the CNS topology and parameter files. Therefore, they need to be calculated, and this is done with the server LigParGen: <http://zarbi.chem.yale.edu/ligpargen/>.²⁶⁴

The structure file (*pdb*) of the ligand can be uploaded and the charge needs to be carefully checked. The server will provide the *.pdb*, *.top* and *.param* files. The *.top* and *.param* have to be copied in the *topallhfg5.4.pro* and *parallhfg5.4.pro* files of the RECOORD folder, respectively. The dihedrals from the parameter file should not be copied to the *parallhfg5.4.pro*, and the improper angles in the *parallhfg5.4.pro* and *topallhfg5.4.pro* should be consistent, meaning having the same atom order.

CNS compatible complex structure preparation

The preparation into a CNS compatible format is done with the Cyana software. Cyana reads the *pdb* file and translates the atom naming into the Xplor format (CNS compatible).

```
>> read pdb xxx.pdb
```

```
>> translate xplor
```

```
>> write pdb xxx_trans.pdb
```

The restraints are transformed the same way. However, the names of the files are exported with: *.upl*, *.lol* or *.aco* extensions, while CNS accepts only *.tbl* extensions. Hence, the names are change to *dihedrals.tbl* for the *.aco* files, and *unambig.tbl* for the distance restraints, which contains the upper and lower limits.

The ligand structure file is copied at the end of the *pdb* file. The *ligand.pdb* is the one obtained from LigParGen so that the naming is consistent. Make sure that the protein structure ends with a line “TER”, and the file ends up with a “END” command.

Preparation of the water refinement

The folder containing all the scripts from RECOORD is the working directory. The first step is to run the script *changeScriptDir.sh* with the address of the current directory. Then the script *generate.sh* is run to generate the file needed by CNS.

```
>> ./generate.sh xxx_trans.pdb
```

A directory name *run1* is created, and the *xxx_trans_cns.pdb* and *xxx_trans_cns.mtf* are copied in the folder, but the *xxx* is replaced by *run1*. The file *xxx_trans_cns.pdb* is copied in the same folder as *run1_trans_cns_extended.pdb*. The *.tbl* restraints files are copied in the folder *run1* as well, it is important that they are named as mentioned previously.

Running the water refinement scripts

The script *annealing.sh* has to be run even though it generates only an error.

```
>> annealing.sh run1
```

Do not pay attention to that, the cause is not known but it does not affect the process. Once this script ran, a folder named *str* has been created in the *run1* folder, and the file *run1_cns.pdb* is to be copied in the *str* folder as *run1_cns_1.pdb*.

Finally, the script for the water refinement is run:

```
>> re_h2o.sh run1
```

The refined structures are in the folder *run1/str/wt/*.

PDB submission

Preparation of the ligand

In case the ligand is already part of the PDB, the easier is to search using the SMILES string of the ligand. Go to the PDB advanced search query builder <https://www.rcsb.org/search/advanced/chemical> to search for the particular ligand., and enter the sum formula or the SMILES code into the “Chemical” field.

E.g. C₂₂H₂₂CIN₅O₂ for the iBET-762 ligand:

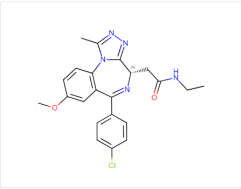

Figure A2.1: Advanced search query interface to mine deposited ligand.

or for the SMILES string:

(C1C1=CC=C(C2=N[C@@H](CC(NCC)=O)C3=NN=C(C)N3C4=CC=C(OC)C=C24)C=C1).

In case it is not part of PDB: the advantage is that the atoms in the ligand and the ligand itself can be named at convenience, even though it is probably best to avoid ligand names that are already taken. If the ligand is part of the PDB: make sure to use the ligand naming and atom nomenclature that all the other PDB entries with that ligand use. The below steps exemplify this on the iBET-762 ligand, which is called ‘EAM’ in the PDB.

So, you have found your ligand is in the PDB already. As an example we take the iBET-762 (C₂₂H₂₂CIN₅O₂, (C1C1=CC=C(C2=N[C@@H](CC(NCC)=O)C3=NN=C(C)N3C4=CC=C(OC)C=C24)C=C1) ligand, which is called ‘EAM’ in the PDB. If you look for this ligand, you should find (among others) the PDB entry ‘3P5O’. When you open that entry in your browser and scroll down you should find a link to ‘EAM’ in the ligand database which will take you to <https://www.rcsb.org/ligand/EAM> which is a summary page with much useful structural information for that ligand:

EAM View / Download Files ▾

2-[(4S)-6-(4-chlorophenyl)-8-methoxy-1-methyl-4H-[1,2,4]triazolo[4,3-a][1,4]benzodiazepin-4-yl]-N-ethylacetamide

Find entries where: EAM
 is present as a standalone ligand in 3 entries

Find related ligands:
 Similar Ligands (Stereo Specific)
 Similar Ligands (Incl/ Stereoisomers)
 Similar Ligands
 Similar Ligands (Substructure Stereo Specific)
 Similar Ligands (Substructure Incl/ Stereoisomers)

[View summary at Ligand Expo](#)

Chemical Component Summary		Chemical Details	
Name	2-[(4S)-6-(4-chlorophenyl)-8-methoxy-1-methyl-4H-[1,2,4]triazolo[4,3-a][1,4]benzodiazepin-4-yl]-N-ethylacetamide	Formal Charge	0
Identifiers	2-[(4S)-6-(4-chlorophenyl)-8-methoxy-1-methyl-4H-	Atom Count	52
		Chiral Atom Count	1

Figure A2.2: Summary of the ligand deposition in PDB.

The picture files (svg, png) and naming dictionaries for the ligand can be found by clicking on the “View summary at Ligand Expo” button/link above, which lead to the ligand index (<http://ligand-expo.rcsb.org/reports/E/EAM/index.html>). In the Figure A2.3 the atom naming for iBET762 is shown.

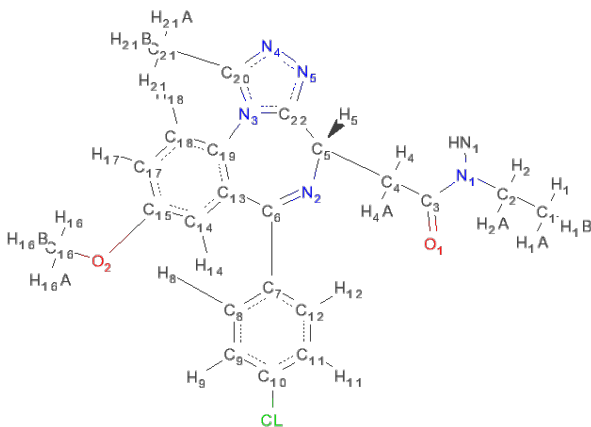


Figure A2.3: iBET762 picture and atom numbering according to the PDB nomenclature.

Getting prepared for the deposition

The deposition files consist in: 1) pdb, 2) NMR restraints (e.g. upl file), 3) NMR-Star file containing the assignments, 4) optional: NOESY peaklists. As of lately the PDB does actually no longer accept structure submissions in the pdb format. Only the mmCIF format is supported going forward. The pdb file can be converted into an mmCIF using the [pdb-extract](https://pdb-extract.wwpdb.org/) webserver: <https://pdb-extract.wwpdb.org/>. Note the amino acid sequence of the protein in 1-letter code is needed. Also, all the CYANA leftovers such as linker atoms, pseudoatoms etc. from the pdb file before conversion. Save the resulting mmCIF file to disk (output.cif) and rename ligand name and atom names as detailed below. Compare the naming scheme of the EAM ligand from the PDB to the naming scheme in the deposition files. Adapt the naming in the deposition files to the naming in the EAM ligand as it is found in the PDB. This can be done readily with shell scripts, and the restraints file(s) and NMR-Star file should be renamed accordingly. Rename the ligand moiety to 'EAM' (in the original CYANA file it was called 'BET').

Deposition

The inclusion of the optional NOESY peaklist is in principle not necessary, and could be skipped in case this generates problems. Nevertheless, the peaklist should be made available at least in the supplementary of the related article or deposited, when possible, in the BMRB database.

A word of warning also concerning the NMR-Star file. This was by far the reason for the most errors. A “minimal” NMR-Star file only is recommended, i.e.: only include the necessary assignment information. Unfortunately, every NMR program creates different NMR-Star files. The NMR-Star file modification is mostly a try and fail process. When an upload to the PDB fails, it usually gives some clues as to why it fails. It is then a matter of iteratively changing/deleting things in the NMR-Star file until the data upload succeeds. Once the data upload has succeeded the PDB offers a quite straight-forward, linear flow of points that must be addressed before deposition is complete.

Literature

1. Wlodawer, A.; Vondrasek, J., Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 249-84.
2. Blobaum, A. L.; Marnett, L. J., Structural and functional basis of cyclooxygenase inhibition. *J Med Chem* **2007**, *50* (7), 1425-41.
3. Ferreira, S. H.; Moncada, S.; Vane, J. R., Indomethacin and aspirin abolish prostaglandin release from the spleen. *Nat New Biol* **1971**, *231* (25), 237-9.
4. Vane, J. R., Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nat New Biol* **1971**, *231* (25), 232-5.
5. Smith, J. B.; Willis, A. L., Aspirin selectively inhibits prostaglandin production in human platelets. *Nat New Biol* **1971**, *231* (25), 235-7.
6. Vanderoudera, F. J.; Buytenhek, M.; Nugteren, D. H.; Vandorp, D. A., Acetylation of Prostaglandin Endoperoxide Synthetase with Acetylsalicylic-Acid. *Eur J Biochem* **1980**, *109* (1), 1-8.
7. Loll, P. J.; Picot, D.; Garavito, R. M., The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H2 synthase. *Nat Struct Biol* **1995**, *2* (8), 637-43.
8. Mashalidis, E. H.; Sledz, P.; Lang, S.; Abell, C., A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc* **2013**, *8* (11), 2309-24.
9. Batool, M.; Ahmad, B.; Choi, S., A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* **2019**, *20* (11).
10. Hall, M. D.; Yasgar, A.; Peryea, T.; Braisted, J. C.; Jadhav, A.; Simeonov, A.; Coussens, N. P., Fluorescence polarization assays in high-throughput screening and drug discovery: a review. *Methods Appl Fluoresc* **2016**, *4* (2), 022001.
11. Fang, X. N.; Zheng, Y. Z.; Duan, Y. K.; Liu, Y.; Zhong, W. W., Recent Advances in Design of Fluorescence-Based Assays for High-Throughput Screening. *Anal Chem* **2019**, *91* (1), 482-504.
12. Neumann, T.; Junker, H. D.; Schmidt, K.; Sekul, R., SPR-based fragment screening: Advantages and applications. *Curr Top Med Chem* **2007**, *7* (16), 1630-1642.
13. Blundell, T. L.; Jhoti, H.; Abell, C., High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* **2002**, *1* (1), 45-54.
14. Hajduk, P. J.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **2007**, *6* (3), 211-9.

15. Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W., One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *Journal of the American Chemical Society* **1997**, *119* (50), 12257-12261.
16. Gossert, A. D.; Jahnke, W., NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog Nucl Magn Reson Spectrosc* **2016**, *97*, 82-125.
17. Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **2013**, *73*, 1-16.
18. Orts, J.; Gossert, A. D., Structure determination of protein-ligand complexes by NMR in solution. *Methods* **2018**, *138-139*, 3-25.
19. Leavitt, S.; Freire, E., Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struc Biol* **2001**, *11* (5), 560-566.
20. Rossi, A. M.; Taylor, C. W., Analysis of protein-ligand interactions by fluorescence polarization. *Nature Protocols* **2011**, *6* (3).
21. Thomas, S. E.; Collins, P.; James, R. H.; Mendes, V.; Charoensuththivarakul, S.; Radoux, C.; Abell, C.; Coyne, A. G.; Floto, R. A.; von Delft, F.; Blundell, T. L., Structure-guided fragment-based drug discovery at the synchrotron: screening binding sites and correlations with hotspot mapping. *Philos Trans A Math Phys Eng Sci* **2019**, *377* (2147), 20180422.
22. Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H., Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **2005**, *48* (2), 403-13.
23. Patel, D.; Bauman, J. D.; Arnold, E., Advantages of crystallographic fragment screening: functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog Biophys Mol Biol* **2014**, *116* (2-3), 92-100.
24. Levitt, M. H., *Spin dynamics : basics of nuclear magnetic resonance*. John Wiley & Sons: Chichester ; New York, 2001; p xxiv, 686 p.
25. Cavanagh, J., *Protein NMR spectroscopy : principles and practice*. 2nd ed.; Academic Press: Amsterdam ; Boston, 2007; p xxv, 885 p.
26. Keeler, J., *Understanding NMR spectroscopy*. 2nd ed.; John Wiley and Sons: Chichester, U.K., 2010; p xiii, 511 p.
27. McIntosh, L. P.; Dahlquist, F. W.; Redfield, A. G., Proton NMR and NOE structural and dynamic studies of larger proteins and nucleic acids aided by isotope labels: T4 lysozyme. *J Biomol Struct Dyn* **1987**, *5* (1), 21-34.
28. McIntosh, L. P.; Griffey, R. H.; Muchmore, D. C.; Nielson, C. P.; Redfield, A. G.; Dahlquist, F. W., Proton NMR measurements of bacteriophage T4 lysozyme aided

by ¹⁵N isotopic labeling: structural and dynamic studies of larger proteins. *Proc Natl Acad Sci U S A* **1987**, *84* (5), 1244-8.

29. Fesik, S. W.; Zuiderweg, E. R., Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules. *Q Rev Biophys* **1990**, *23* (2), 97-131.

30. Clore, G. M.; Gronenborn, A. M., Applications of 3-Dimensional and 4-Dimensional Heteronuclear Nmr-Spectroscopy to Protein-Structure Determination. *Prog Nucl Mag Res Sp* **1991**, *23*, 43-92.

31. Bax, A.; Grzesiek, S., Methodological Advances in Protein Nmr. *Accounts Chem Res* **1993**, *26* (4), 131-138.

32. Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K., Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in C-13-labeled proteins. *Journal of the American Chemical Society* **1998**, *120* (25), 6394-6400.

33. Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K., Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *P Natl Acad Sci USA* **1997**, *94* (23), 12366-12371.

34. Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E., Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* **2002**, *124* (34), 10025-35.

35. Bibow, S.; Polyhach, Y.; Eichmann, C.; Chi, C. N.; Kowal, J.; Albiez, S.; McLeod, R. A.; Stahlberg, H.; Jeschke, G.; Guntert, P.; Riek, R., Solution structure of discoidal high-density lipoprotein particles with a shortened apolipoprotein A-I. *Nat Struct Mol Biol* **2017**, *24* (2), 187-+.

36. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem Int Ed Engl* **1999**, *38* (12), 1784-1788.

37. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B., WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J Biomol NMR* **2001**, *21* (4), 349-59.

38. Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M., Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J Biomol NMR* **2000**, *18* (1), 65-8.

39. Bodenhausen, G.; Ruben, D. J., Natural Abundance N-15 Nmr by Enhanced Heteronuclear Spectroscopy. *Chem Phys Lett* **1980**, *69* (1), 185-189.

40. Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M., Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. *Chem-Eur J* **2010**, *16* (26), 7803-7812.
41. Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J., Two-Dimensional NMR Lineshape Analysis. *Sci Rep-Uk* **2016**, *6*.
42. Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E., Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *Journal of the American Chemical Society* **1997**, *119* (29), 6711-6721.
43. Orts, J.; Walti, M. A.; Marsh, M.; Vera, L.; Gossert, A. D.; Guntert, P.; Riek, R., NMR-Based Determination of the 3D Structure of the Ligand-Protein Interaction Site without Protein Resonance Assignment. *J Am Chem Soc* **2016**, *138* (13), 4393-400.
44. Walti, M. A.; Riek, R.; Orts, J., Fast NMR-Based Determination of the 3D Structure of the Binding Site of Protein-Ligand Complexes with Weak Affinity Binders. *Angew Chem Int Ed Engl* **2017**, *56* (19), 5208-5211.
45. Fesik, S. W.; Zuiderweg, E. R.; Olejniczak, E. T.; Gampe, R. T., Jr., NMR methods for determining the structures of enzyme/inhibitor complexes as an aid in drug design. *Biochem Pharmacol* **1990**, *40* (1), 161-7.
46. Torres, F.; Orts, J., Nuclear magnetic resonance structure-based drug design. *Future Med Chem* **2018**, *10* (20), 2373-2376.
47. McCoy, M. A.; Wyss, D. F., Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. *J Biomol NMR* **2000**, *18* (3), 189-98.
48. Cioffi, M.; Hunter, C. A.; Packer, M. J.; Pandya, M. J.; Williamson, M. P., Use of quantitative (¹H) NMR chemical shift changes for ligand docking into barnase. *J Biomol NMR* **2009**, *43* (1), 11-9.
49. Sanchez-Pedregal, V. M.; Reese, M.; Meiler, J.; Blommers, M. J.; Griesinger, C.; Carlomagno, T., The INPHARMA method: protein-mediated interligand NOEs for pharmacophore mapping. *Angew Chem Int Ed Engl* **2005**, *44* (27), 4172-5.
50. Hajduk, P. J.; Mack, J. C.; Olejniczak, E. T.; Park, C.; Dandliker, P. J.; Beutel, B. A., SOS-NMR: a saturation transfer NMR-based method for determining the structures of protein-ligand complexes. *J Am Chem Soc* **2004**, *126* (8), 2390-8.
51. Otting, G., Prospects for lanthanides in structural biology by NMR. *J Biomol NMR* **2008**, *42* (1), 1-9.
52. Stratmann, D.; Boelens, R.; Bonvin, A. M., Quantitative use of chemical shifts for the modeling of protein complexes. *Proteins* **2011**, *79* (9), 2662-70.

53. Constantine, K. L.; Davis, M. E.; Metzler, W. J.; Mueller, L.; Claus, B. L., Protein-ligand NOE matching: a high-throughput method for binding pose evaluation that does not require protein NMR resonance assignments. *J Am Chem Soc* **2006**, *128* (22), 7252-63.
54. Walti, M. A.; Orts, J., The NMR2 Method to Determine Rapidly the Structure of the Binding Pocket of a Protein-Ligand Complex with High Accuracy. *Magnetochemistry* **2018**, *4* (1).
55. Ikura, M.; Kay, L. E.; Bax, A., A novel approach for sequential assignment of ¹H, ¹³C, and ¹⁵N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **1990**, *29* (19), 4659-67.
56. Ikura, M.; Marion, D.; Kay, L. E.; Shih, H.; Krinks, M.; Klee, C. B.; Bax, A., Heteronuclear 3D NMR and isotopic labeling of calmodulin. Towards the complete assignment of the ¹H NMR spectrum. *Biochem Pharmacol* **1990**, *40* (1), 153-60.
57. Faivre, E. J.; McDaniel, K. F.; Albert, D. H.; Mantena, S. R.; Plotnik, J. P.; Wilcox, D.; Zhang, L.; Bui, M. H.; Sheppard, G. S.; Wang, L.; Sehgal, V.; Lin, X.; Huang, X.; Lu, X.; Uziel, T.; Hessler, P.; Lam, L. T.; Bellin, R. J.; Mehta, G.; Fidanze, S.; Pratt, J. K.; Liu, D.; Hasvold, L. A.; Sun, C.; Panchal, S. C.; Nicolette, J. J.; Fossey, S. L.; Park, C. H.; Longenecker, K.; Bigelow, L.; Torrent, M.; Rosenberg, S. H.; Kati, W. M.; Shen, Y., Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **2020**, *578* (7794), 306-310.
58. Cheung, K.; Lu, G.; Sharma, R.; Vincek, A.; Zhang, R.; Plotnikov, A. N.; Zhang, F.; Zhang, Q.; Ju, Y.; Hu, Y.; Zhao, L.; Han, X.; Meslamani, J.; Xu, F.; Jaganathan, A.; Shen, T.; Zhu, H.; Rusinova, E.; Zeng, L.; Zhou, J.; Yang, J.; Peng, L.; Ohlmeyer, M.; Walsh, M. J.; Zhang, D. Y.; Xiong, H.; Zhou, M. M., BET N-terminal bromodomain inhibition selectively blocks Th17 cell differentiation and ameliorates colitis in mice. *Proc Natl Acad Sci U S A* **2017**, *114* (11), 2952-2957.
59. Zhao, Y.; Yang, C. Y.; Wang, S., The making of I-BET762, a BET bromodomain inhibitor now in clinical development. *J Med Chem* **2013**, *56* (19), 7498-500.
60. Abragam, A., *The principles of nuclear magnetism*. Clarendon Press: Oxford,, 1961; p 599 p.
61. Ardenkjaer-Larsen, J. H.; Fridlund, B.; Gram, A.; Hansson, G.; Hansson, L.; Lerche, M. H.; Servin, R.; Thaning, M.; Golman, K., Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. *P Natl Acad Sci USA* **2003**, *100* (18), 10158-10163.
62. Bowers, C. R.; Weitekamp, D. P., Transformation of Symmetrization Order to Nuclear-Spin Magnetization by Chemical-Reaction and Nuclear-Magnetic-Resonance. *Phys Rev Lett* **1986**, *57* (21), 2645-2648.

63. Adams, R. W.; Aguilar, J. A.; Atkinson, K. D.; Cowley, M. J.; Elliott, P. I.; Duckett, S. B.; Green, G. G.; Khazal, I. G.; Lopez-Serrano, J.; Williamson, D. C., Reversible interactions with para-hydrogen enhance NMR sensitivity by polarization transfer. *Science* **2009**, *323* (5922), 1708-11.
64. Adams, R. W.; Duckett, S. B.; Green, R. A.; Williamson, D. C.; Green, G. G., A theoretical basis for spontaneous polarization transfer in non-hydrogenative parahydrogen-induced polarization. *J Chem Phys* **2009**, *131* (19), 194505.
65. Adams, R. W.; Duckett, S. B.; Green, R. A.; Williamson, D. C.; Green, G. G. R., A theoretical basis for spontaneous polarization transfer in non-hydrogenative parahydrogen-induced polarization. *Journal of Chemical Physics* **2009**, *131* (19).
66. Kovtunov, K. V.; Pokochueva, E. V.; Salnikov, O. G.; Cousin, S. F.; Kurzbach, D.; Vuichoud, B.; Jannin, S.; Chekmenev, E. Y.; Goodson, B. M.; Barskiy, D. A.; Koptyug, I. V., Hyperpolarized NMR Spectroscopy: d-DNP, PHIP, and SABRE Techniques. *Chem-Asian J* **2018**, *13* (15), 1857-1871.
67. Bargon, J.; Fischer, H.; Johnsen, U., Kernresonanz-Emissionslinien Während Rascher Radikalreaktionen .I. Aufnahmeverfahren Und Beispiele. *Z Naturforsch Pt A* **1967**, *A 22* (10), 1551-&.
68. Lawler, R. G., Chemically Induced Dynamic Nuclear Polarization. *Journal of the American Chemical Society* **1967**, *89* (21), 5519-&.
69. Ward, H. R. L. R. G., Nuclear Magnetic Resonance Emission and Enhanced Absorption in Rapid Organometallic Reactions. *J Am Chem Soc* **1967**, *89* (21), 5518–5519.
70. Closs, G. L., A Mechanism Explaining Nuclear Spin Polarizations in Radical Combination Reactions. *Journal of the American Chemical Society* **1969**, *91* (16), 4552-&.
71. Kaptein, R.; Oosterhoff, J. L., Chemically induced dynamic nuclear polarization II - (Relation with anomalous ESR spectra). *Chem Phys Lett* **1969**, *4* (4), 195-197.
72. Kaptein, R.; Oosterhoff, L. J., Chemically induced dynamic nuclear polarization III (anomalous multiplets of radical coupling and disproportionation products). *Chem Phys Lett* **1969**, *4* (4), 214-216.
73. Cocivera, M., Optically Induced Overhauser Effect in Solution . Nuclear Magnetic Resonance Emission. *Journal of the American Chemical Society* **1968**, *90* (12), 3261-&.
74. Sobol, A.; Torres, F.; Aicher, A.; Renn, A.; Riek, R., Atto Thio 12 as a promising dye for photo-CIDNP. *J Chem Phys* **2019**, *151* (23), 234201.

75. Torres, F.; Sobol, A.; Greenwald, J.; Renn, A.; Morozova, O.; Yurkovskaya, A.; Riek, R., Molecular features toward high photo-CIDNP hyperpolarization explored through the oxidocyclization of tryptophan. *Phys Chem Chem Phys* **2021**, *23* (11), 6641-6650.
76. Saprygina, N. N.; Morozova, O. B.; Grampp, G.; Yurkovskaya, A. V., Effect of Amino Group Charge on the Photooxidation Kinetics of Aromatic Amino Acids. *J Phys Chem A* **2014**, *118* (2), 339-349.
77. Otting, G.; Wuthrich, K., Efficient Purging Scheme for Proton-Detected Heteronuclear Two-Dimensional Nmr. *J Magn Reson* **1988**, *76* (3), 569-574.
78. Otting, G.; Wuthrich, K., Extended Heteronuclear Editing of 2d H-1-Nmr Spectra of Isotope-Labeled Proteins, Using the X(Omega-1, Omega-2) Double Half Filter. *J Magn Reson* **1989**, *85* (3), 586-594.
79. Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C., Fragment-based lead discovery: leads by design. *Drug Discov Today* **2005**, *10* (14), 987-92.
80. Carvalho, A. L.; Trincao, J.; Romao, M. J., X-Ray Crystallography in Drug Discovery. *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols* **2010**, *572*, 31-56.
81. Orts, J.; Walti, M. A.; Marsh, M.; Vera, L.; Gossert, A. D.; Guntert, P.; Riek, R., NMR-Based Determination of the 3D Structure of the Ligand-Protein Interaction Site without Protein Resonance Assignment. *Journal of the American Chemical Society* **2016**, *138* (13), 4393-4400.
82. Walti, M. A.; Riek, R.; Orts, J., Fast NMR-Based Determination of the 3D Structure of the Binding Site of Protein-Ligand Complexes with Weak Affinity Binders. *Angew Chem Int Edit* **2017**, *56* (19), 5208-5211.
83. Wälti, M.; Orts, J., The NMR2 Method to Determine Rapidly the Structure of the Binding Pocket of a Protein-Ligand Complex with High Accuracy. *Magnetochemistry* **2018**, *4* (1), 12.
84. Torres, F.; Orts, J., Nuclear magnetic resonance structure-based drug design. *Future Med Chem* **2018**, *10* (20), 2373-2376.
85. Strotz, D.; Orts, J.; Chi, C. N.; Riek, R.; Vogeli, B., eNORA2 Exact NOE Analysis Program. *J Chem Theory Comput* **2017**, *13* (9), 4336-4346.
86. Orts, J.; Vogeli, B.; Riek, R., Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs. *Journal of Chemical Theory and Computation* **2012**, *8* (10), 3483-3492.
87. Orts, J.; Walti, M. A.; Marsh, M.; Vera, L.; Gossert, A. D.; Guntert, P.; Riek, R., NMR-Based Determination of the 3D Structure of the Ligand-Protein Interaction Site without Protein Resonance Assignment. *J Am Chem Soc* **2016**.

88. Ni, F., Recent Developments in Transferred Noe Methods. *Prog Nucl Mag Res Sp* **1994**, *26*, 517-606.
89. Vogeli, B.; Segawa, T. F.; Leitz, D.; Sobol, A.; Choutko, A.; Trzesniak, D.; van Gunsteren, W.; Riek, R., Exact Distances and Internal Dynamics of Perdeuterated Ubiquitin from NOE Buildups. *Journal of the American Chemical Society* **2009**, *131* (47), 17215-17225.
90. Verdone, L.; Caserta, M.; Di Mauro, E., Role of histone acetylation in the control of gene expression. *Biochem Cell Biol* **2005**, *83* (3), 344-53.
91. Workman, J. L.; Kingston, R. E., Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annual Review of Biochemistry* **1998**, *67*, 545-579.
92. Allfrey, V. G.; Faulkner, R.; Mirsky, A. E., Acetylation + Methylation of Histones + Their Possible Role in Regulation of Rna Synthesis. *P Natl Acad Sci USA* **1964**, *51* (5), 786-+.
93. Jenuwein, T.; Allis, C. D., Translating the histone code. *Science* **2001**, *293* (5532), 1074-80.
94. Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M. M., Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **1999**, *399* (6735), 491-496.
95. Lane, A. A.; Chabner, B. A., Histone deacetylase inhibitors in cancer therapy. *J Clin Oncol* **2009**, *27* (32), 5459-68.
96. Dawson, M. A.; Prinjha, R. K.; Dittmann, A.; Giotopoulos, G.; Bantscheff, M.; Chan, W. I.; Robson, S. C.; Chung, C. W.; Hopf, C.; Savitski, M. M.; Huthmacher, C.; Gudgin, E.; Lugo, D.; Beinke, S.; Chapman, T. D.; Roberts, E. J.; Soden, P. E.; Auger, K. R.; Mirguet, O.; Doehner, K.; Delwel, R.; Burnett, A. K.; Jeffrey, P.; Drewes, G.; Lee, K.; Huntly, B. J.; Kouzarides, T., Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **2011**, *478* (7370), 529-33.
97. Delmore, J. E.; Issa, G. C.; Lemieux, M. E.; Rahl, P. B.; Shi, J.; Jacobs, H. M.; Kastiris, E.; Gilpatrick, T.; Paranal, R. M.; Qi, J.; Chesi, M.; Schinzel, A. C.; McKeown, M. R.; Heffernan, T. P.; Vakoc, C. R.; Bergsagel, P. L.; Ghobrial, I. M.; Richardson, P. G.; Young, R. A.; Hahn, W. C.; Anderson, K. C.; Kung, A. L.; Bradner, J. E.; Mitsiades, C. S., BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **2011**, *146* (6), 904-17.
98. Mertz, J. A.; Conery, A. R.; Bryant, B. M.; Sandy, P.; Balasubramanian, S.; Mele, D. A.; Bergeron, L.; Sims, R. J., 3rd, Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc Natl Acad Sci U S A* **2011**, *108* (40), 16669-74.

99. Unzue, A.; Xu, M.; Dong, J.; Wiedmer, L.; Spiliotopoulos, D.; Caflisch, A.; Nevado, C., Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain. *J Med Chem* **2016**, *59* (4), 1350-6.
100. Clark, P. G. K.; Vieira, L. C. C.; Tallant, C.; Fedorov, O.; Singleton, D. C.; Rogers, C. M.; Monteiro, O. P.; Bennett, J. M.; Baronio, R.; Muller, S.; Daniels, D. L.; Mendez, J.; Knapp, S.; Brennan, P. E.; Dixon, D. J., LP99: Discovery and Synthesis of the First Selective BRD7/9 Bromodomain Inhibitor. *Angewandte Chemie-International Edition* **2015**, *54* (21), 6217-6221.
101. Demont, E. H.; Bamborough, P.; Chung, C. W.; Craggs, P. D.; Fallon, D.; Gordon, L. J.; Grandi, P.; Hobbs, C. I.; Hussain, J.; Jones, E. J.; Le Gall, A.; Michon, A. M.; Mitchell, D. J.; Prinjha, R. K.; Roberts, A. D.; Sheppard, R. J.; Watson, R. J., 1,3-Dimethyl Benzimidazolones Are Potent, Selective Inhibitors of the BRPF1 Bromodomain. *ACS Med Chem Lett* **2014**, *5* (11), 1190-5.
102. Bennett, J.; Fedorov, O.; Tallant, C.; Monteiro, O.; Meier, J.; Gamble, V.; Savitsky, P.; Nunez-Alonso, G. A.; Haendler, B.; Rogers, C.; Brennan, P. E.; Muller, S.; Knapp, S., Discovery of a Chemical Tool Inhibitor Targeting the Bromodomains of TRIM24 and BRPF. *J Med Chem* **2016**, *59* (4), 1642-7.
103. Hewings, D. S.; Rooney, T. P.; Jennings, L. E.; Hay, D. A.; Schofield, C. J.; Brennan, P. E.; Knapp, S.; Conway, S. J., Progress in the development and application of small molecule inhibitors of bromodomain-acetyl-lysine interactions. *J Med Chem* **2012**, *55* (22), 9393-413.
104. Yu, L.; Wang, Z.; Zhang, Z.; Ren, X.; Lu, X.; Ding, K., Small-molecule BET inhibitors in clinical and preclinical development and their therapeutic potential. *Curr Top Med Chem* **2015**, *15* (8), 776-94.
105. Filippakopoulos, P.; Knapp, S., Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev Drug Discov* **2014**, *13* (5), 337-56.
106. Gallenkamp, D.; Gelato, K. A.; Haendler, B.; Weinmann, H., Bromodomains and Their Pharmacological Inhibitors. *Chemmedchem* **2014**, *9* (3), 438-464.
107. Chung, C. W.; Coste, H.; White, J. H.; Mirguet, O.; Wilde, J.; Gosmini, R. L.; Delves, C.; Magny, S. M.; Woodward, R.; Hughes, S. A.; Boursier, E. V.; Flynn, H.; Bouillot, A. M.; Bamborough, P.; Brusq, J. M.; Gellibert, F. J.; Jones, E. J.; Riou, A. M.; Homes, P.; Martin, S. L.; Uings, I. J.; Toum, J.; Clement, C. A.; Boullay, A. B.; Grimley, R. L.; Blandel, F. M.; Prinjha, R. K.; Lee, K.; Kirilovsky, J.; Nicodeme, E., Discovery and characterization of small molecule inhibitors of the BET family bromodomains. *J Med Chem* **2011**, *54* (11), 3827-38.
108. Basheer, F.; Huntly, B. J., BET bromodomain inhibitors in leukemia. *Exp Hematol* **2015**, *43* (8), 718-31.

109. Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E., Selective inhibition of BET bromodomains. *Nature* **2010**, *468* (7327), 1067-73.
110. Mirguet, O.; Gosmini, R.; Toum, J.; Clement, C. A.; Barnathan, M.; Brusq, J. M.; Mordaunt, J. E.; Grimes, R. M.; Crowe, M.; Pineau, O.; Ajakane, M.; Daugan, A.; Jeffrey, P.; Cutler, L.; Haynes, A. C.; Smithers, N. N.; Chung, C. W.; Bamborough, P.; Uings, I. J.; Lewis, A.; Witherington, J.; Parr, N.; Prinjha, R. K.; Nicodeme, E., Discovery of epigenetic regulator I-BET762: lead optimization to afford a clinical candidate inhibitor of the BET bromodomains. *J Med Chem* **2013**, *56* (19), 7501-15.
111. Lane, A., The influence of spin diffusion and internal motions on NOE intensities in proteins. *J. Magn. Reson.* **1988**, *78*, 425.
112. Linge, J. P.; Habeck, M.; Rieping, W.; Nilges, M., Correction of spin diffusion during iterative automated NOE assignment. *J. Magn. Reson.* **2004**, *167*, 334.
113. Orts, J.; Vögeli, B.; Riek, R., Relaxation matrix analysis of spin diffusion for the NMR structure calculation with eNOEs. *J. Chem. Theory Comput.* **2012**, *8*, 3483.
114. Sprangers, R.; Velyvis, A.; Kay, L. E., Solution NMR of supramolecular complexes: providing new insights into function. *Nat Methods* **2007**, *4* (9), 697-703.
115. Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E., Selective methyl group protonation of perdeuterated proteins. *J Mol Biol* **1996**, *263* (5), 627-36.
116. Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J., Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 113.
117. Wüthrich, K., *NMR of proteins and nucleic acids*. Wiley: New York, 1986; p xv, 292.
118. Devlieg, J.; Boelens, R.; Scheek, R. M.; Kaptein, R.; Vangunsteren, W. F., Restrained Molecular-Dynamics Procedure for Protein Tertiary Structure Determination from Nmr Data - a Lac Repressor Headpiece Structure Based on Information on J-Coupling and from Presence and Absence of Noe. *Isr J Chem* **1986**, *27* (2), 181-188.
119. Macek, P.; Kerfah, R.; Erba, E. B.; Crublet, E.; Moriscot, C.; Schoehn, G.; Amero, C.; Boisbouvier, J., Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Sci Adv* **2017**, *3* (4).
120. Waring, M. J.; Chen, H. W.; Rabow, A. A.; Walker, G.; Bobby, R.; Boiko, S.; Bradbury, R. H.; Callis, R.; Clark, E.; Dale, I.; Daniels, D. L.; Dulak, A.; Flavell, L.;

Holdgate, G.; Jowitt, T. A.; Kikhney, A.; McAlister, M.; Mendez, J.; Ogg, D.; Patel, J.; Petteruti, P.; Robb, G. R.; Robers, M. B.; Saif, S.; Stratton, N.; Svergun, D. I.; Wang, W. X.; Whittaker, D.; Wilson, D. M.; Yao, Y., Potent and selective bivalent inhibitors of BET bromodomains. *Nat Chem Biol* **2016**, *12* (12), 1097-+.

121. Torres, F.; Ghosh, D.; Strotz, D.; Chi, C. N.; Davis, B.; Orts, J., Protein-fragment complex structures derived by NMR molecular replacement. *RSC Med Chem* **2020**, *11* (5), 591-596.

122. Doak, B. C.; Norton, R. S.; Scanlon, M. J., The ways and means of fragment-based drug design. *Pharmacol Ther* **2016**, *167*, 28-37.

123. Gorse, A. D., Diversity in medicinal chemistry space. *Curr. Top. Med. Chem.* **2006**, *6* (1), 3-18.

124. Reymond, J. L., The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48* (3), 722-730.

125. Fink, T.; Bruggesser, H.; Reymond, J. L., Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew. Chem. Int. Ed. Engl.* **2005**, *44* (10), 1504-8.

126. Mattos, C.; Ringe, D., Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* **1996**, *14* (5), 595-9.

127. Geschwindner, S.; Olsson, L. L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; de Beer, T.; Folmer, R. H., Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J. Med. Chem.* **2007**, *50* (24), 5903-11.

128. Ahn, S.; Kahsai, A. W.; Pani, B.; Wang, Q. T.; Zhao, S.; Wall, A. L.; Strachan, R. T.; Staus, D. P.; Wingler, L. M.; Sun, L. D.; Sinnavee, J.; Choi, M.; Cho, T.; Xu, T. T.; Hansen, G. M.; Burnett, M. B.; Lamerdin, J. E.; Bassoni, D. L.; Gavino, B. J.; Husemoen, G.; Olsen, E. K.; Franch, T.; Costanzi, S.; Chen, X.; Lefkowitz, R. J., Allosteric "beta-blocker" isolated from a DNA-encoded small molecule library. *Proc Natl Acad Sci U S A* **2017**, *114* (7), 1708-1713.

129. Murray, C. W.; Verdonk, M. L.; Rees, D. C., Experiences in fragment-based drug discovery. *Trends in Pharmacological Sciences* **2012**, *33* (5), 224-232.

130. Schuller, M.; al., e., Fragment Binding to the Nsp3 Macrodomein of SARS-CoV-2 Identified Through Crystallographic Screening and Computational Docking. *BioRxiv* **2020**.

131. Potter, A.; Oldfield, V.; Nunns, C.; Fromont, C.; Ray, S.; Northfield, C. J.; Bryant, C. J.; Scrace, S. F.; Robinson, D.; Matossova, N.; Baker, L.; Dokurno, P.; Surgenor, A. E.; Davis, B.; Richardson, C. M.; Murray, J. B.; Moore, J. D., Discovery of cell-active phenyl-imidazole Pin1 inhibitors by structure-guided fragment evolution. *Bioorg Med Chem Lett* **2010**, *20* (22), 6483-8.

132. Favia, A. D.; Bottegoni, G.; Nobeli, I.; Bisignano, P.; Cavalli, A., SERAPhiC: a benchmark for in silico fragment-based drug design. *J Chem Inf Model* **2011**, *51* (11), 2882-96.
133. Tanimoto, T., *An Elementary Mathematical Theory of Classification and Prediction*. International Business Machines Corporation: 1958.
134. Bajusz, D.; Racz, A.; Heberger, K., Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* **2015**, *7*, 20.
135. Chen, Y.; Wu, Y. R.; Yang, H. Y.; Li, X. Z.; Jie, M. M.; Hu, C. J.; Wu, Y. Y.; Yang, S. M.; Yang, Y. B., Prolyl isomerase Pin1: a promoter of cancer and a target for therapy. *Cell Death Dis* **2018**, *9* (9), 883.
136. Pu, W.; Zheng, Y.; Peng, Y., Prolyl Isomerase Pin1 in Human Cancer: Function, Mechanism, and Significance. *Front Cell Dev Biol* **2020**, *8*, 168.
137. Chao, S. H.; Greenleaf, A. L.; Price, D. H., Juglone, an inhibitor of the peptidyl-prolyl isomerase Pin1, also directly blocks transcription. *Nucleic Acids Res* **2001**, *29* (3), 767-73.
138. Potter, A. J.; Ray, S.; Gueritz, L.; Nunns, C. L.; Bryant, C. J.; Scrace, S. F.; Matassova, N.; Baker, L.; Dokurno, P.; Robinson, D. A.; Surgenor, A. E.; Davis, B.; Murray, J. B.; Richardson, C. M.; Moore, J. D., Structure-guided design of alpha-amino acid-derived Pin1 inhibitors. *Bioorg Med Chem Lett* **2010**, *20* (2), 586-90.
139. Ren, X. D.; Shi, Y. S.; Zhang, Y.; Liu, B.; Zhang, L. H.; Peng, Y. B.; Zeng, R., Novel Consensus Docking Strategy to Improve Ligand Pose Prediction. *Journal of Chemical Information and Modeling* **2018**, *58* (8), 1662-1668.
140. Morrone, J. A.; Weber, J. K.; Huynh, T.; Luo, H.; Cornell, W. D., Combining Docking Pose Rank and Structure with Deep Learning Improves Protein-Ligand Binding Mode Prediction over a Baseline Docking Approach. *Journal of Chemical Information and Modeling* **2020**, *60* (9), 4170-4179.
141. Bax, A.; Clore, G. M.; Gronenborn, A. M., H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins. *Journal of Magnetic Resonance* **1990**, *88* (2), 425-431.
142. Vuister, G. W.; Bax, A., Resolution Enhancement and Spectral Editing of Uniformly C-13-Enriched Proteins by Homonuclear Broad-Band C-13 Decoupling. *Journal of Magnetic Resonance* **1992**, *98* (2), 428-435.
143. Torres, F.; Ghosh, D.; Strotz, D.; Chi, C. N.; Davis, B.; Orts, J., Protein-fragment complex structures derived by NMR molecular replacement. *Rsc Medicinal Chemistry* **2020**, *11* (5), 591-596.

144. Kay, L. E.; Torchia, D. A.; Bax, A., Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **1989**, *28* (23), 8972-9.
145. Dosset, P.; Hus, J. C.; Blackledge, M.; Marion, D., Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* **2000**, *16* (1), 23-8.
146. Lakomek, N. A.; Ying, J.; Bax, A., Measurement of (1)(5)N relaxation rates in perdeuterated proteins by TROSY-based methods. *J Biomol NMR* **2012**, *53* (3), 209-21.
147. Hui, D. S.; E, I. A.; Madani, T. A.; Ntoumi, F.; Kock, R.; Dar, O.; Ippolito, G.; McHugh, T. D.; Memish, Z. A.; Drosten, C.; Zumla, A.; Petersen, E., The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* **2020**, *91*, 264-266.
148. Wu, J. T.; Leung, K.; Leung, G. M., Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **2020**, *395* (10225), 689-697.
149. Yao, X.; Ye, F.; Zhang, M.; Cui, C.; Huang, B.; Niu, P.; Liu, X.; Zhao, L.; Dong, E.; Song, C.; Zhan, S.; Lu, R.; Li, H.; Tan, W.; Liu, D., In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Clin Infect Dis* **2020**, *71* (15), 732-739.
150. Skipper, C. P.; Pastick, K. A.; Engen, N. W.; Bangdiwala, A. S.; Abassi, M.; Lofgren, S. M.; Williams, D. A.; Okafor, E. C.; Pullen, M. F.; Nicol, M. R.; Nascene, A. A.; Hullsiek, K. H.; Cheng, M. P.; Luke, D.; Lothar, S. A.; MacKenzie, L. J.; Drobot, G.; Kelly, L. E.; Schwartz, I. S.; Zarychanski, R.; McDonald, E. G.; Lee, T. C.; Rajasingham, R.; Boulware, D. R., Hydroxychloroquine in Nonhospitalized Adults With Early COVID-19 : A Randomized Trial. *Ann Intern Med* **2020**, *173* (8), 623-631.
151. Abdool Karim, S. S.; de Oliveira, T., New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. *N Engl J Med* **2021**, *384* (19), 1866-1868.
152. Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L., Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* **2021**, *19* (3), 141-154.
153. V'Kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V., Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **2021**, *19* (3), 155-170.
154. Letko, M.; Marzi, A.; Munster, V., Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **2020**, *5* (4), 562-569.
155. Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L., Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus:

implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* **2020**, *17* (6), 613-620.

156. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367* (6483), 1260-1263.

157. Yin, W.; Mao, C.; Luan, X.; Shen, D. D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y. C.; Tian, G.; Jiang, H. W.; Tao, S. C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Zhang, Y.; Xu, H. E., Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020**, *368* (6498), 1499-1504.

158. Gordon, C. J.; Tchesnokov, E. P.; Woolner, E.; Perry, J. K.; Feng, J. Y.; Porter, D. P.; Gotte, M., Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem* **2020**, *295* (20), 6785-6797.

159. Young, B.; Tan, T. T.; Leo, Y. S., The place for remdesivir in COVID-19 treatment. *Lancet Infect Dis* **2021**, *21* (1), 20-21.

160. Li, J.; Zhou, X.; Zhang, Y.; Zhong, F.; Lin, C.; McCormick, P. J.; Jiang, F.; Luo, J.; Zhou, H.; Wang, Q.; Fu, Y.; Duan, J.; Zhang, J., Crystal structure of SARS-CoV-2 main protease in complex with the natural product inhibitor shikonin illuminates a unique binding mode. *Sci Bull (Beijing)* **2021**, *66* (7), 661-663.

161. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R., Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **2020**, *368* (6489), 409-412.

162. Zhang, S.; Krumberger, M.; Morris, M. A.; Parrocha, C. M. T.; Griffin, J. H.; Kreutzer, A. G.; Nowick, J. S., Structure-Based Drug Design of an Inhibitor of the SARS-CoV-2 (COVID-19) Main Protease Using Free Software: A Tutorial for Students and Scientists. *ChemRxiv* **2020**.

163. Ashburn, T. T.; Thor, K. B., Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* **2004**, *3* (8), 673-83.

164. Pushpakom, S.; Iorio, F.; Eyers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; Doig, A.; Williams, T.; Latimer, J.; McNamee, C.; Norris, A.; Sanseau, P.; Cavalla, D.; Pirmohamed, M., Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* **2019**, *18* (1), 41-58.

165. Dotolo, S.; Marabotti, A.; Facchiano, A.; Tagliaferri, R., A review on drug repurposing applicable to COVID-19. *Brief Bioinform* **2021**, *22* (2), 726-741.

166. O'Reilly, M.; Cleasby, A.; Davies, T. G.; Hall, R. J.; Ludlow, R. F.; Murray, C. W.; Tisi, D.; Jhoti, H., Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discov Today* **2019**, *24* (5), 1081-1086.

167. Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J., Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res* **2020**, *30* (8), 678-692.

168. Fintelman-Rodrigues, N.; Sacramento, C. Q.; Ribeiro Lima, C.; Souza da Silva, F.; Ferreira, A. C.; Mattos, M.; de Freitas, C. S.; Cardoso Soares, V.; da Silva Gomes Dias, S.; Temerozo, J. R.; Miranda, M. D.; Matos, A. R.; Bozza, F. A.; Carels, N.; Alves, C. R.; Siqueira, M. M.; Bozza, P. T.; Souza, T. M. L., Atazanavir, Alone or in Combination with Ritonavir, Inhibits SARS-CoV-2 Replication and Proinflammatory Cytokine Production. *Antimicrob Agents Chemother* **2020**, *64* (10).

169. Xin Wang; Hai-Bin Luo, Z. L. X. L. Y.-Y. H. Y. W. L. Z. R. L. D. W. L. Z. H. L. X. X. Y. Z. J. C., FEP-based screening prompts drug repositioning against COVID-19

. *bioRxiv* **2020**.

170. Altincekic, N. K., S. M.; Qureshi, N. S.; Dujardin, M.; Ninot-Pedrosa, M. ; ...; Hengesbach, M.; Schlundt, A., Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Frontiers in Molecular Biosciences* **2021**, *8*.

171. Gardner, K. H.; Kay, L. E., The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 357-406.

172. Yamazaki, T. L., W.; Revington, M.; Mattiello, D. L.; Dahlquist, F. W.; Arrowsmith, C. H. K., L. E., An HNCA Pulse Scheme for the Backbone Assignment of 15N,13C,2H-Labeled Proteins: Application to a 37-kDa Trp Repressor-DNA Complex. *J Am Chem Soc* **1994**, *116*, 6464-6465.

173. Qiao, J.; Li, Y. S.; Zeng, R.; Liu, F. L.; Luo, R. H.; Huang, C.; Wang, Y. F.; Zhang, J.; Quan, B.; Shen, C.; Mao, X.; Liu, X.; Sun, W.; Yang, W.; Ni, X.; Wang, K.; Xu, L.; Duan, Z. L.; Zou, Q. C.; Zhang, H. L.; Qu, W.; Long, Y. H.; Li, M. H.; Yang, R. C.; Liu, X.; You, J.; Zhou, Y.; Yao, R.; Li, W. P.; Liu, J. M.; Chen, P.; Liu, Y.; Lin, G. F.; Yang, X.; Zou, J.; Li, L.; Hu, Y.; Lu, G. W.; Li, W. M.; Wei, Y. Q.; Zheng, Y. T.; Lei, J.; Yang, S., SARS-CoV-2 M(pro) inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021**, *371* (6536), 1374-1378.

174. Vuister, G. W. B., A., Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling. *J Magn Reson* **1992**, *98*, 428-435.

175. Gunther, S.; Reinke, P. Y. A.; Fernandez-Garcia, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.; Lorenzen, K.; Krichel, B.; Kopicki, J. D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Perez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hanle, A.; Groessler, M.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schlunzen, F.; Lourdu Xavier, P.; Werner, N.; Andaleeb, H.; Ullah, N.; Falke, S.; Srinivasan, V.; Franca, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitseva-Doyle, J. J.; Knoska, J.; Pena-Murillo, G. E.; Mashhour, A. R.; Hennicke, V.; Fischer, P.; Hakanpaa, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; von Stetten, D.; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia-Alai, M. M.; Niebling, S.; Gunther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E. C.; Mehrabi, P.; Karnicar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Uetrecht, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Gunther, S.; Turk, D.; Hinrichs, W.; Chapman, H. N.; Pearson, A. R.; Betzel, C.; Meents, A., X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372* (6542), 642-646.

176. Huang, L.; Jiang, S.; Shi, Y., Tyrosine kinase inhibitors for solid tumors in the past 20 years (2001-2020). *J Hematol Oncol* **2020**, *13* (1), 143.

177. Group, R. C.; Horby, P.; Lim, W. S.; Emberson, J. R.; Mafham, M.; Bell, J. L.; Linsell, L.; Staplin, N.; Brightling, C.; Ustianowski, A.; Elmahi, E.; Prudon, B.; Green, C.; Felton, T.; Chadwick, D.; Rege, K.; Fegan, C.; Chappell, L. C.; Faust, S. N.; Jaki, T.; Jeffery, K.; Montgomery, A.; Rowan, K.; Juszczak, E.; Baillie, J. K.; Haynes, R.; Landray, M. J., Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* **2021**, *384* (8), 693-704.

178. Ardenkjaer-Larsen, J. H.; Golman, K.; Gram, A.; Lerche, M. H.; Servin, R.; Thaning, M.; Wolber, J., Increase of signal-to-noise of more than 10,000 times in liquid state NMR. *Discov Med* **2003**, *3* (19), 37-9.

179. Bargon, J.; Fischer, H.; Johnsen, U., Kernresonanz-Emissionslinien Wahrend Rascher Radikalreaktionen .I. Aufnahmeverfahren Und Beispiele. *Z Naturforsch Pt A* **1967**, *A 22* (10), 1551-1555.

180. Ward, H. R.; Lawler, R. G., Nuclear Magnetic Resonance Emission and Enhanced Absorption in Rapid Organometallic Reactions. *J Am Chem Soc* **1967**, *89* (21), 5518-5519.

181. Closs, G. L., Mechanism explaining nuclear spin polarizations in radical combination reactions. *J. Am. Chem. Soc.* **1969**, *91* (16), 4552-4554.

182. Adrian, F. J., Role of Diffusion-Controlled Reaction in Chemically Induced Nuclear-Spin Polarization .2. General Theory and Comparison with Experiment. *J Chem Phys* **1971**, *54* (9), 3912-3917.
183. Noyes, R. M., A Treatment of Chemical Kinetics with Special Applicability to Diffusion Controlled Reactions. *J Chem Phys* **1954**, *22* (8), 1349-1359.
184. Okuno, Y.; Cavagnero, S., Photochemically Induced Dynamic Nuclear Polarization: Basic Principles and Applications. *Emagres* **2017**, *6* (2), 283-313.
185. Rehm, D.; Weller, A., Kinetics of Fluorescence Quenching by Electron and H-Atom Transfer. *Israel J Chem* **1970**, *8* (2), 259-&.
186. Tsentalovich, Y. P.; Lopez, J. J.; Hore, P. J.; Sagdeev, R. Z., Mechanisms of reactions of flavin mononucleotide triplet with aromatic amino acids. *Spectrochim Acta A* **2002**, *58* (9), 2043-2050.
187. Closs, G. L., Chemically Induced Dynamic Nuclear Polarization. In *Advances in Magnetic and Optical Resonance*, (Press), J. S. W. S. D. A., Ed. 1974; Vol. 7, pp 157-229.
188. Vollenweider, J. K.; Fischer, H.; Hennig, J.; Leuschner, R., Time-Resolved Cidnp in Laser Flash-Photolysis of Aliphatic-Ketones - a Quantitative-Analysis. *Chem Phys* **1985**, *97* (2-3), 217-234.
189. Vollenweider, J. K.; Fischer, H., Time-Resolved Cidnp in Laser Flash-Photolysis of Di-Tert-Butyl Ketone - Multiplet Versus Net Effects. *Chem Phys* **1986**, *108* (3), 365-372.
190. Vollenweider, J. K.; Fischer, H., Absolute Chemically-Induced Nuclear Polarizations and Yields from Geminate Radical Pair Reactions - a Test of High-Field Radical Pair Theories. *Chem Phys* **1988**, *124* (3), 333-345.
191. Morozova, O. B.; Ivanov, K. L., Time-Resolved Chemically Induced Dynamic Nuclear Polarization of Biologically Important Molecules. *Chemphyschem* **2019**, *20* (2), 197-215.
192. Tsentalovich, Y. P.; Morozova, O. B.; Yurkovskaya, A. V.; Hore, P. J., Kinetics and mechanism of the photochemical reaction of 2,2 '-dipyridyl with tryptophan in water: Time-resolved CIDNP and laser flash photolysis study. *Journal of Physical Chemistry A* **1999**, *103* (27), 5362-5368.
193. Stob, S.; Kaptein, R., Photo-Cidnp of the Amino-Acids. *Photochem Photobiol* **1989**, *49* (5), 565-577.
194. Okuno, Y.; Cavagnero, S., Fluorescein: A Photo-CIDNP Sensitizer Enabling Hypersensitive NMR Data Collection in Liquids at Low Micromolar Concentration. *Journal of Physical Chemistry B* **2016**, *120* (4), 715-723.

195. Okuno, Y.; Cavagnero, S., Effect of heavy atoms on photochemically induced dynamic nuclear polarization in liquids. *Journal of Magnetic Resonance* **2018**, *286*, 172-187.
196. Chmyrov, A.; Arden-Jacob, J.; Zilles, A.; Drexhage, K. H.; Widengren, J., Characterization of new fluorescent labels for ultra-high resolution microscopy. *Photoch Photobio Sci* **2008**, *7* (11), 1378-1385.
197. Mompean, M.; Sanchez-Donoso, R. M.; de la Hoz, A.; Saggiomo, V.; Velders, A. H.; Gomez, M. V., Pushing nuclear magnetic resonance sensitivity limits with microfluidics and photochemically induced dynamic nuclear polarization. *Nat Commun* **2018**, *9*.
198. Reindl, S.; Penzkofer, A., Triplet quantum yield determination by picosecond laser double-pulse fluorescence excitation. *Chem Phys* **1996**, *213* (1-3), 429-438.
199. Ronsein, G. E.; de Oliveira, M. C.; de Medeiros, M. H.; Di Mascio, P., Characterization of O(2) ((1)delta(g))-derived oxidation products of tryptophan: a combination of tandem mass spectrometry analyses and isotopic labeling studies. *J Am Soc Mass Spectrom* **2009**, *20* (2), 188-97.
200. Gracanic, M.; Hawkins, C. L.; Pattison, D. I.; Davies, M. J., Singlet-oxygen-mediated amino acid and protein oxidation: formation of tryptophan peroxides and decomposition products. *Free Radic Biol Med* **2009**, *47* (1), 92-102.
201. Savige, W. E., New Oxidation-Products of Tryptophan. *Aust J Chem* **1975**, *28* (10), 2275-2287.
202. Ronsein, G. E.; Oliveira, M. C. B.; Miyamoto, S.; Medeiros, M. H. G.; Di Mascio, P., Tryptophan oxidation by singlet molecular oxygen [O-2 ((1)Delta(g))]: Mechanistic studies using O-18-labeled hydroperoxides, mass spectrometry, and light emission measurements. *Chemical Research in Toxicology* **2008**, *21* (6), 1271-1283.
203. Nakagawa, M.; Kato, S.; Kataoka, S.; Kodato, S.; Watanabe, H.; Okajima, H.; Hino, T.; Witkop, B., Dye-Sensitized Photo-Oxygenation of Tryptophan - 3a-Hydroperoxytryptophan as a Labile Precursor of Formylkynurenine. *Chem Pharm Bull* **1981**, *29* (4), 1013-1026.
204. Lee, J. H.; Sekhar, A.; Cavagnero, S., H-1-Detected C-13 Photo-CIDNP as a Sensitivity Enhancement Tool in Solution NMR. *Journal of the American Chemical Society* **2011**, *133* (21), 8062-8065.
205. Wilson, D. J. D.; Mohn, C. E.; Helgaker, T., The rotational g tensor as a benchmark for density-functional theory calculations of molecular magnetic properties. *J Chem Theory Comput* **2005**, *1* (5), 877-888.
206. Kiryutin, A. S.; Morozova, O. B.; Kuhn, L. T.; Yurkovskaya, A. V.; Hore, P. J., H-1 and C-13 hyperfine coupling constants of the Tryptophanyl cation radical in

- aqueous solution from microsecond time-resolved CIDNP. *Journal of Physical Chemistry B* **2007**, *111* (38), 11221-11227.
207. Kaptein, R., Simple Rules for Chemically Induced Dynamic Nuclear Polarization. *J Chem Soc Chem Comm* **1971**, (14), 732-&.
208. Improta, R.; Barone, V., Interplay of electronic, environmental, and vibrational effects in determining the hyperfine coupling constants of organic free radicals. *Chem Rev* **2004**, *104* (3), 1231-1253.
209. Morozova, O. B.; Ivanov, K. L.; Kiryutin, A. S.; Sagdeev, R. Z.; Kochling, T.; Vieth, H. M.; Yurkovskaya, A. V., Time-resolved CIDNP: an NMR way to determine the EPR parameters of elusive radicals. *Phys Chem Chem Phys* **2011**, *13* (14), 6619-27.
210. Säuberlich; J. Brede, O.; Beckert, D., Photoionization of Benzophenone Carboxylic Acids in Aqueous Solution. A FT EPR and Optical Spectroscopy Study of Radical Cation Decay. *J Phys Chem* **1996**, *100* (46), 18101-18107.
211. Morozova, O. B.; Panov, M. S.; Fishman, N. N.; Yurkovskaya, A. V., Electron transfer vs. proton-coupled electron transfer as the mechanism of reaction between amino acids and triplet-excited benzophenones revealed by time-resolved CIDNP. *Phys Chem Chem Phys* **2018**, *20* (32), 21127-21135.
212. Torolabbe, A.; Maruani, J., The Conformational Dependence of the Hyperfine Couplings of Gamma-Protons in Pi-Radicals. *J Magn Reson* **1985**, *61* (2), 254-261.
213. Morozova, O. B.; Panov, M. S.; Fishman, N. N.; Yurkovskaya, A. V., Electron transfer vs proton-coupled electron transfer as the mechanism of reaction between amino acids and triplet-excited benzophenones revealed by time-resolved CIDNP. *Phys. Chem. Chem. Phys.* **2018**, *20* (32), 21127-21135.
214. Rehm, D.; Weller, A., Kinetics of Fluorescence Quenching by Electron and H-Atom Transfer. *Israel Journal of Chemistry* **1970**, *8* (2), 259-271.
215. Marcus, R. A., Electrostatic Free Energy and Other Properties of States Having Nonequilibrium Polarization .1. *Journal of Chemical Physics* **1956**, *24* (5), 979-989.
216. Marcus, R. A., On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer .1. *Journal of Chemical Physics* **1956**, *24* (5), 966-978.
217. Kuprov, I.; Craggs, T. D.; Jackson, S. E.; Hore, P. J., Spin relaxation effects in photochemically induced dynamic nuclear polarization spectroscopy of nuclei with strongly anisotropic hyperfine couplings. *J Am Chem Soc* **2007**, *129* (29), 9004-13.
218. Khan, F.; Kuprov, I.; Craggs, T. D.; Hore, P. J.; Jackson, S. E., 19F NMR studies of the native and denatured states of green fluorescent protein. *J Am Chem Soc* **2006**, *128* (33), 10729-37.
219. Kuprov, I.; Hore, P. J., Chemically amplified 19F-1H nuclear Overhauser effects. *J Magn Reson* **2004**, *168* (1), 1-7.

220. Closs, G. L.; Trifunac, A. D., Chemically Induced Nuclear Spin Polarization as a Tool for Determination of Spin Multiplicities of Radical-Pair Precursors. *J Am Chem Soc* **1969**, *91* (16), 4554-4555.
221. Morozova, O. B.; Kiryutin, A. S.; Yurkovskaya, A. V., Electron transfer between guanosine radicals and amino acids in aqueous solution. II. Reduction of guanosine radicals by tryptophan. *Journal of Physical Chemistry B* **2008**, *112* (9), 2747-2754.
222. Morozova, O. B.; Kiryutin, A. S.; Sagdeev, R. Z.; Yurkovskaya, A. V., Electron transfer between guanosine radical and amino acids in aqueous solution. I. Reduction of guanosine radical by tyrosine. *Journal of Physical Chemistry B* **2007**, *111* (25), 7439-7448.
223. Morozova, O. B.; Hore, P. J.; Sagdeev, R. Z.; Yurkovskaya, A. V., Intramolecular electron transfer in lysozyme studied by time-resolved chemically induced dynamic nuclear polarization. *Journal of Physical Chemistry B* **2005**, *109* (46), 21971-21978.
224. Kaptein, R.; Dijkstra, K.; Nicolay, K., Laser Photo-Cidnp as a Surface Probe for Proteins in Solution. *Nature* **1978**, *274* (5668), 293-294.
225. Hore, P. J.; Winder, S. L.; Roberts, C. H.; Dobson, C. M., Stopped-Flow Photo-CIDNP Observation of Protein Folding. *J Am Chem Soc* **1997**, *119* (21), 5049-5050.
226. Mok, K. H.; Nagashima, T.; Day, I. J.; Jones, J. A.; Jones, C. J.; Dobson, C. M.; Hore, P. J., Rapid sample-mixing technique for transient NMR and photo-CIDNP spectroscopy: applications to real-time protein folding. *J Am Chem Soc* **2003**, *125* (41), 12484-92.
227. Mok, K. H.; Hore, P. J., Photo-CIDNP NMR methods for studying protein folding. *Methods* **2004**, *34* (1), 75-87.
228. Morozova, O. B.; Yurkovskaya, A. V.; Vieth, H. M.; Sagdeev, R. Z., Intramolecular electron transfer in tryptophan-tyrosine peptide in photoinduced reaction in aqueous solution. *Journal of Physical Chemistry B* **2003**, *107* (4), 1088-1096.
229. Goetz, M., *An Introduction to Chemically Induced Dynamic Nuclear Polarization*. 1995 ed.; 1995; Vol. 7, p 69-86.
230. Bernarding, J.; Euchner, F.; Bruns, C.; Ringleb, R.; Mueller, D.; Trantzschel, T.; Bargon, J.; Bommerich, U.; Plaumann, M., Low-cost LED-based Photo-CIDNP Enables Biocompatible Hyperpolarization of F-19 for NMR and MRI at 7T and 4.7T. *Chemphyschem* **2018**, *19* (19), 2453-2456.
231. Torres, F.; Sobol, A.; Greenwald, J.; Renn, A.; Riek, R., Molecular features toward high photo-CIDNP hyperpolarization explored through the oxidocyclization of tryptophan. *Phys Chem Chem Phys* **2021**.

232. Kaptein, R.; Nicolay, K.; Dijkstra, K., Photo-Cidnp in Nucleic-Acid Bases and Nucleotides. *Journal of the Chemical Society-Chemical Communications* **1979**, (23), 1092-1094.
233. Hore, P. J.; Broadhurst, R. W., Photo-Cidnp of Biopolymers. *Prog Nucl Mag Res Sp* **1993**, *25*, 345-402.
234. Tsentalovich, Y. P.; Morozova, O. B.; Yurkovskaya, A. V.; Hore, P. J.; Sagdeev, R. Z., Time-resolved CIDNP and laser flash photolysis study of the photoreactions of N-acetyl histidine with 2,2'-dipyridyl in aqueous solution. *Journal of Physical Chemistry A* **2000**, *104* (30), 6912-6916.
235. Saprygina, N. N.; Morozova, O. B.; Grampp, G.; Yurkovskaya, A. V., Effect of amino group charge on the photooxidation kinetics of aromatic amino acids. *J Phys Chem A* **2014**, *118* (2), 339-49.
236. Connor, H. D.; Sturgeon, B. E.; Mottley, C.; Sipe, H. J.; Mason, R. P., L-tryptophan radical cation electron spin resonance studies: Connecting solution-derived hyperfine coupling constants with protein spectral interpretations. *J Am Chem Soc* **2008**, *130* (20), 6381-6387.
237. Kaptein, R., SIMPLE RULES FOR CHEMICALLY INDUCED DYNAMIC NUCLEAR POLARIZATION. *Journal of the Chemical Society D-Chemical Communications* **1971**, (14), 732-733.
238. Lavis, L. D.; Rutkoski, T. J.; Raines, R. T., Tuning the pK(a) of fluorescein to optimize binding assays. *Anal Chem* **2007**, *79* (17), 6775-82.
239. Luchinat, E.; Banci, L., In-cell NMR: a topical review. *IUCrJ* **2017**, *4* (Pt 2), 108-118.
240. Walsh, M. A.; Grimes, J. M.; Stuart, D. I., Diamond Light Source: contributions to SARS-CoV-2 biology and therapeutics. *Biochem Biophys Res Commun* **2021**, *538*, 40-46.
241. Torres, F.; Renn, A.; Riek, R., Exploration of the close chemical space of tryptophan and tyrosine reveals importance of hydrophobicity in CW-photo-CIDNP performances. *Magnetic Resonance* **2021**, *2*, 321-329.
242. Palmer, W. S.; Poncet-Montange, G.; Liu, G.; Petrocchi, A.; Reyna, N.; Subramanian, G.; Theroff, J.; Yau, A.; Kost-Alimova, M.; Bardenhagen, J. P.; Leo, E.; Shepard, H. E.; Tieu, T. N.; Shi, X.; Zhan, Y.; Zhao, S.; Barton, M. C.; Draetta, G.; Toniatti, C.; Jones, P.; Geck Do, M.; Andersen, J. N., Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *J Med Chem* **2016**, *59* (4), 1440-54.
243. Walser, R.; Renshaw, J.; Milbradt, A. G., Backbone resonance assignments for the PHD-Bromo dual-domain of the human chromatin reader TRIM24. *Biomol NMR Assign* **2016**, *10* (1), 207-11.

244. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D., The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **2005**, *59* (4), 687-96.
245. Blackledge, M.; Cordier, F.; Dosset, P.; Marion, D., Precision and uncertainty in the characterization of anisotropic rotational diffusion by N-15 relaxation. *Journal of the American Chemical Society* **1998**, *120* (18), 4538-4539.
246. Guntert, P., Automated NMR structure calculation with CYANA. *Methods Mol Biol* **2004**, *278*, 353-78.
247. Lee, J. H.; Cavagnero, S., A Novel Tri-Enzyme System in Combination with Laser-Driven NMR Enables Efficient Nuclear Polarization of Biomolecules in Solution. *Journal of Physical Chemistry B* **2013**, *117* (20), 6069-6081.
248. Sanchez-Pulido, L.; Devos, D.; Valencia, A., BRICHOS: a conserved domain in proteins associated with dementia, respiratory distress and cancer. *Trends Biochem Sci* **2002**, *27* (7), 329-32.
249. Beers, M. F.; Kim, C. Y.; Dodia, C.; Fisher, A. B., Localization, Synthesis, and Processing of Surfactant Protein Sp-C in Rat Lung Analyzed by Epitope-Specific Antipeptide Antibodies. *Journal of Biological Chemistry* **1994**, *269* (32), 20318-20328.
250. Nerelius, C.; Gustafsson, M.; Nordling, K.; Larsson, A.; Johansson, J., Anti-Amyloid Activity of the C-Terminal Domain of proSP-C against Amyloid beta-Peptide and Medin. *Biochemistry* **2009**, *48* (17), 3778-3786.
251. Coomaraswamy, J.; Kilger, E.; Wolfing, H.; Schafer, C.; Kaeser, S. A.; Wegenast-Braun, B. M.; Hefendehl, J. K.; Wolburg, H.; Mazzella, M.; Ghiso, J.; Goedert, M.; Akiyama, H.; Garcia-Sierra, F.; Wolfner, D. P.; Mathews, P. M.; Jucker, M., Modeling familial Danish dementia in mice supports the concept of the amyloid hypothesis of Alzheimer's disease. *P Natl Acad Sci USA* **2010**, *107* (17), 7969-7974.
252. Cohen, S. I. A.; Arosio, P.; Presto, J.; Kurudenkandy, F. R.; Biverstal, H.; Dolfe, L.; Dunning, C.; Yang, X. T.; Frohm, B.; Vendruscolo, M.; Johansson, J.; Dobson, C. M.; Fisahn, A.; Knowles, T. P. J.; Linse, S., A molecular chaperone breaks the catalytic cycle that generates toxic A beta oligomers. *Nat Struct Mol Biol* **2015**, *22* (3), 207-213.
253. Willander, H.; Askarieh, G.; Landreh, M.; Westermark, P.; Nordling, K.; Keranen, H.; Hermansson, E.; Hamvas, A.; Noguee, L. M.; Bergman, T.; Saenz, A.; Casals, C.; Aqvist, J.; Jornvall, H.; Berglund, H.; Presto, J.; Knight, S. D.; Johansson, J., High-resolution structure of a BRICHOS domain and its implications for anti-amyloid chaperone activity on lung surfactant protein C. *P Natl Acad Sci USA* **2012**, *109* (7), 2325-2329.
254. Biverstal, H.; Dolfe, L.; Hermansson, E.; Leppert, A.; Reifenrath, M.; Winblad, B.; Presto, J.; Johansson, J., Dissociation of a BRICHOS trimer into

- monomers leads to increased inhibitory effect on A beta 42 fibril formation. *Bba-Proteins Proteom* **2015**, 1854 (8), 835-843.
255. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A., TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular Nmr* **2009**, 44 (4), 213-223.
256. Frueh, D. P.; Leed, A.; Arthanari, H.; Koglin, A.; Walsh, C. T.; Wagner, G., Time-shared HSQC-NOESY for accurate distance constraints measured at high-field in N-15-C-13-ILV methyl labeled proteins. *Journal of Biomolecular Nmr* **2009**, 45 (3), 311-318.
257. Guntert, P.; Mumenthaler, C.; Wuthrich, K., Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **1997**, 273 (1), 283-298.
258. Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wuthrich, K., Stereospecific Nuclear Magnetic-Resonance Assignments of the Methyl-Groups of Valine and Leucine in the DNA-Binding Domain of the 434-Repressor by Biosynthetically Directed Fractional C-13 Labeling. *Biochemistry* **1989**, 28 (19), 7510-7516.
259. Xue, C.; Lin, T. Y. W.; Chang, D.; Guo, Z. F., Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. *Roy Soc Open Sci* **2017**, 4 (1).
260. Brunger, A. T., Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2007**, 2 (11), 2728-2733.
261. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* **1998**, 54, 905-921.
262. Unknown, Crystallography: Protein Data Bank. *Nature New Biology* **1971**, 233, 223.
263. Nederveen, A. J.; Doreleijers, J. F.; Vranken, W.; Miller, Z.; Spronk, C. A. E. M.; Nabuurs, S. B.; Guntert, P.; Livny, M.; Markley, J. L.; Nilges, M.; Ulrich, E. L.; Kaptein, R.; Bonvin, A. M. J. J., RECOORD: A recalculated coordinate database of 500+proteins from the PDB using restraints from the BioMagResBank. *Proteins-Structure Function and Bioinformatics* **2005**, 59 (4), 662-672.
264. Dodda, L. S.; de Vaca, I. C.; Tirado-Rives, J.; Jorgensen, W. L., LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res* **2017**, 45 (W1), W331-W336.