# Protein-ligand interactions revealed by liquide state NMR

Julien Orts

UNIVERSITÉ DE GRENOBLE

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physique pour les Sciences du Vivant**

Arrêté ministériel : 7 août 2006

Présentée par

**Julien Orts**

Thèse dirigée par **Martin Blackledge** et
codirigée par **Teresa Carlomagno / Christian Griesinger**

préparée au sein des laboratoires de
**- L'Institut de Biologie Structurale**
**dans l'École Doctorale de Physique**

**- NMR based Structural Biology**
**at the Max-Planck Institut für Biophysikalische Chemie**

**- Structural and Computational Biology**
**at the European Molecular Biology Laboratory**

# Caractérisation des interactions entre ligands et protéines par RMN en solution

Thèse soutenue publiquement le **11 Mai 2010,**
devant le jury composé de :

**Dr. Teresa Carlomagno**
Directeur de recherche à l'EMBL Heidelberg, Membre du jury
**Prof. dr. Christian Griesinger**
Directeur du MPI Goettingen, Membre du jury
**Dr. Martin Blackledge**
Directeur de recherche à l'IBS, Membre du jury
**Dr. Bruno Kieffer**
Directeur de recherche à l'IGBMC Strasbourg, Président, Rapporteur
**Prof. dr. Roland Riek**
Directeur de recherche à l'ETH Zurich, Rapporteur

Un des buts de la recherche pharmaceutique est l'inhibition de protéines avec l'aide de petites molécules (ligands). L'une des phases clefs de ce procédé est la détermination du mode d'interaction entre un ligand et son récepteur. Cette tâche peut être entravée par l'absence de structure du complexe protéine-ligand. C'est pour répondre à ce besoin que nous présentons dans ce travail de thèse, une méthode capable de déterminer la structure de complexes protéine-ligands. Dans la méthode INPHARMA (Inter-ligands Nuclear Overhauser Effect for Pharmacophore Mapping), les inter-ligands NOEs (INPHARMA NOEs) sont utilisés pour déterminer l'orientation relative de deux ligands qui interagissent de manière compétitive avec un même récepteur. Cette nouvelle approche ouvre la voie à des applications pharmaceutiques, également au stade initial du développement, quand l'information structurale via la cristallographie par Rayons X est difficile d'accès.

Mots clefs: ligand, protéine, interaction, structure, pharmacophore, RMN

In the process of structure-based drug design, the provision of the binding mode of ligands to the cellular receptor of interest is essential. This can suffer from limited access to protein/ligand structures, especially for the low affinity ligands that are commonly obtained from high throughput screening or fragment based lead discovery. In a common scenario crystal structures are available for one or several ligands but not for all chemical series of actual interest. Here, we present a new, NMR-based approach that allows overcoming this limitation. In the INPHARMA method interligand NOEs (Nuclear Overhauser Enhancement) are utilized to determine relative orientations of different chemical fragments binding competitively to a common receptor site. This novel methodology opens the way to the application of structure-based drug design already in an early stage of drug development, when structural information via crystallography is of difficult access.

Keywords: ligand, protein, interaction, structure, pharmacophore, NMR

# Protein-ligand interactions revealed by liquide state NMR

**Julien Orts**

# Protein-ligand interactions revealed by liquide state NMR

Inaugural Dissertation

Submitted in March 2010 to
The Faculty of Physics of the University Joseph Fourier
Grenoble, France
and
The European Molecular Biology Laboratory
Heidelberg, Germany

for the joint degrees of

**Doctor of Natural Sciences**

**Docteur ès Sciences Physiques**

presented by

**Julien Orts**

born in Saint-Louis, France in 1983

**Board of examiners**

Promotors:

Dr. T. Carlomagno
*European Molecular Biology Laboratory*
Prof. dr. C. Griesinger
*Max-Planck Institut für Biophysikalische Chemie*
Dr. M. Blackledge
*Institut de Biologie Structurale*

Reviewers:

Dr. B. Kieffer
*Institut de Génétique et de Biologie Moléculaire et Cellulaire*
Prof. dr. R. Riek
*Die Eidgenössische Technische Hochschule*

*A mes parents, à ma soeur, à mon frère*

*A Anne-Christine*

# List of symbols

| | |
|---|---|
| $\|\alpha\alpha\rangle$, $\|\alpha\beta\rangle$, $\|\beta\alpha\rangle$, $\|\beta\beta\rangle$ | states of a spin-1/2 pair |
| $[A]$, $[TB]$, $[M]$, $[L_1]$... | concentrations |
| Å | ångström |
| $a_{11}$, $a_{22}$, $a_{12}$, $a_{21}$, $a_{MM}$, $a_{NN}$... | amplitudes of auto and cross peaks |
| B | Boltzmann factor |
| $B_0$ | static field |
| $b$ | dipole-dipole coupling constant |
| $\delta(\omega)$ | Kronecker delta, $\delta(0) = 1$ |
| $\mathbf{e_{jk}}$ | unit vector parallel to the vector between spins j and k |
| $\mathrm{FT}_1$, $\mathrm{FT}_2$ | Fourier transformations in the indirect and direct dimension |
| $\phi_{1,2,3}$, $\psi$ | phases of radio frequency pulses |
| $\phi_{rec}$ | radio frequency phase shift of receiver reference signal during signal detection |
| $\gamma$, $\gamma_H$, $\gamma_j$, $\gamma_k$ | gyromagnetic ratios of the spin j, k |
| $\hbar$ | Planck constant divided by $2\pi$ |
| $H_{jk}^{\mathrm{DD}}$ | full form of the homonuclear dipole-dipole spin Hamiltonian |
| $H_{jk}^{\mathrm{DD,secular}}$ | secular homonuclear dipole-dipole spin Hamiltonian |
| $\mathbf{I_j}$, $\mathbf{I_k}$ | vectors operators for the spin angular momentum of nucleus j, k |

| | |
|---|---|
| $I_{1x},\ I_{2y},\ I_{2z}...$ | angular momentum operators |
| $< I >$ | expectation value of the operator $I$ |
| $< I >_{eq}$ | expectation value of the operator $I$ in thermal equilibrium |
| $\mathbf{I}$ | identity matrix |
| J | J-coupling |
| $J(\omega)$ | spectral density |
| $k,\ k_N,\ k_{12},\ k_{on}...$ | kinetic rate constants |
| $k_b$ | Boltzmann constant |
| $\mathbf{K}$ | chemical exchange matrix |
| $\mathcal{L}_{T_2}(\omega_1, \omega_2)$ | Lorentzian function |
| $\lambda,\ \lambda_+,\ \lambda_-$ | eigenvalues |
| $\mu_0$ | vacuum permeability |
| $\mathbf{M(t)},\ \mathbf{M_{eq}}$ | magnetization vectors at time t and at equilibrium |
| NOE | Nuclear Overhauser Effect |
| NOESY | Nuclear Overhauser SpectroscopY |
| $\omega$ | frequency |
| $\omega_1$, $\omega_2$ | frequency coordinates of a 2D spectra |
| $\omega_1$ | in addition, spin-lock field strength |
| $\omega_0$ | Larmor frequency |
| $\Omega_1,\ \Omega_2$ | resonance offset frequencies |
| $(\pi/2)_\phi$ | pulse flip angle $\pi/2$ with phase $\phi$ |
| $\vec{q_1}, \vec{q_2}$ | axis of rotations |
| $Q$ | quality factor |
| RMSD | Root Mean Square Deviation |
| $\rho$ | auto-relaxation rate constant |
| $R_{1\rho},\ R_1,\ R_2$ | rotating-frame, longitudinal transverse relaxation rate constants |
| $R_{ex}$ | exchange contribution to $R_2$ |
| $R_{i,i},\ R_{auto}$ | auto-relaxation rate constants |
| $R_{i,j},\ R_{cross}$ | cross-relaxation rate constants |
| $R_X^X,\ R_{T,X}^{TX},\ R_{X,T}^{TX}$ | block-sub-matrices of the relaxation matrix |
| $\mathbf{R}$ | relaxation matrix |

| | |
|---|---|
| $r,\ r_{jk}$ | inter-nuclei distances |
| $R^2$ | Pearson product-moment correlation coefficient |
| $\rho_{\alpha\alpha},\ \rho_{\beta\alpha}\cdots$ | population states for an ensemble of spin-1/2 pairs |
| $\rho_{(1)},\ \rho_{(2)}\cdots$ | spin density operators |
| $\sigma$ | cross-relaxation rate constant |
| $s_{cos}(t_1,t_2),\ s_{sin}(t_1,t_2)$ | cosine and sine signal in the States procedure |
| $S(\omega_1,\omega_2)$ | 2D NMR spectrum |
| $S^2$ | order parameter |
| $t_1,\ t_2$ | indirect and direct time dimensions in a 2D experiment |
| $T_2,\ T_1$ | transverse and longitudinal relaxation time constants |
| $T$ | temperature |
| $\tau_m$ | mixing time of a NOESY spectra |
| $\tau_c,\ \tau_{cA},\ \tau_{cT}\cdots$ | rotational correlation times |
| $\theta_1,\theta_2$ | angles and axis of rotations |
| $\theta_{jk}$ | angle between the vector $\mathbf{e_{jk}}$ and the magnetic field |
| $W_{\alpha+},\ W_{\beta\alpha},\ W_{--},\ W_0,\ W_1\ldots$ | transition probabilities per unit time |

# Contents

# Summary

The goal of pharmaceutical research is to control the activity of targets in such manner that a therapeutically favorable response is achieved. In many cases, scientists aim to inhibit proteins with small molecules. The task is complicated and request tremendous amount of development before a drug arrives to clinical applications. In the process of structure-based drug design, the determination of the binding mode of ligands to the cellular receptor of interest is essential. This can be hampered by limited access to protein-ligand structures, especially for low affinity ligands that are commonly obtained from high throughput screening or fragment based lead discovery. In a common scenario crystal structures are available for one or several ligands but not for all chemical series of actual interest.

In chapter 2, we present a NMR-based approach that allows overcoming this limitation. In the INPHARMA method interligand NOEs (INPHARMA NOEs) are utilized to determine relative orientations of different chemical fragments binding competitively to a common receptor site. This novel methodology opens the way to the application of structure-based drug design already in an early stage of drug development, when structural information via crystallography is of difficult access.

INPHARMA NOEs measured between two ligands binding competitively to a common macromolecular receptor are interpreted quantitatively to derive the binding mode of the two ligands. This approach relies on computationally demanding full relaxation matrix calculations. In chapter 3, we demonstrate that the INPHARMA NOEs measured in presence of a selectively-protonated receptor can be interpreted in a semi-quantitative way to discriminate between binding poses, thus relieving for the need of demanding computations.

In order to successfully apply INPHARMA, the methodology needs adjustment of various parameters that depend on the physical constants of the binding event and on the receptor size. In chapter 4, we present a thorough theoretical analysis of the INPHARMA interligand NOE effect in dependence of experimental parameters and physical constants. This analysis helps the experimentalist to choose the correct experimental parameters and consequentially to achieve optimal performance of the methodology.

In chapter 5, we investigate the scenarios in which the structure of the apo-receptor is not well defined, as it would be the case for low-resolution models obtained by homology modeling or experimental methods such as electron microscopy or lowest resolution X-ray crystallography. The method exhibits high success rates (ranging from 30% to 100%), making INPHARMA a powerful method to resolve small molecule protein complex structures.

# Zusammenfassung

Das Ziel pharmazeutischer Forschung ist die Kontrolle der Aktivität bestimmter Zielmoleküle, um eine therapeutisch gewünschte Reaktion herbeizuführen. Oftmals wird versucht, Proteine durch kleine Moleküle zu inhibieren. Diese Aufgabe ist kompliziert und es bedarf eines hohen Arbeitsaufwandes, bevor ein Wirkstoff Marktreife erlangt.

Im Prozess des strukturbasierten Wirkstoffdesigns ist die Aufklärung des Bindemodus zwischen Liganden und zellulären Rezeptormolekülen von zentraler Bedeutung. Dies setzt die Verfügbarkeit von Protein-Liganden-Strukturen voraus, was oftmals schwierig ist, insbesondere für Liganden geringer Affinität, wie sie häufig in high throughput screening Kampagnen oder fragmentbasierter Wirkstoffentwicklung vorkommen. Häufig sind Kristallstrukturen für einen einzelnen oder einige wenige Liganden verfügbar, nicht jedoch für die komplette chemische Serie.

In Kapitel 2 wird eine NMR-Methode vorgestellt, die dabei helfen kann, diese Limitierung zu überwinden. In der INPHARMA-Methode werden Inter-Liganden-NOEs genutzt, um die relative Orientierung verschiedener chemischer Fragmente aufzuklären, die kompetitiv an einen gemeinsamen Rezeptor binden. Diese neuartige Methode eröffnet die Möglichkeit strukturbasierten Wirkstoffdesigns auch in frühen Phasen der Wirkstoffentwicklung, wenn strukturelle Information kristallographisch nur schwer zugänglich ist. INPHARMA-NOEs können zwischen zwei kompetitiven Liganden gemessen werden, die an denselben Rezeptor binden und quantitativ genutzt werden, um den Bindemodus der beiden Liganden aufzuklären. Dieser Ansatz nutzt die rechenintensive Auswertung der full relaxation matrix.

In Kapitel 3 wird demonstriert, dass INPHARMA-NOEs in Gegenwart eines selektiv protonierten Rezeptors semi-quantitativ interpretiert werden können und es ermöglichen, zwischen verschiedenen Bindeposen zu unterscheiden; dadurch wird die zeitaufwändige Berechnung vereinfacht. Um die INPHARMA-Methode erfolgreich anwenden zu können, müssen verschiedene Parameter angepasst werden, die von physikalischen Konstanten der molekularen Interaktion und der Rezeptorgröße abhängen. In Kapitel 4 wird daher detailliert theoretisch die Abhängigkeit der INPHARMA-NOEs von experimentellen Parametern und physikalischen Konstanten untersucht. Dies ermöglicht es dem Experimentator, die korrekten experimentellen Parameter zu wählen und dadurch die bestmögliche Leistung der Methode zu erreichen.

In Kapitel 5 wird die Auswirkung mangelhafter struktureller Information über den Rezeptor auf die Genauigkeit der INPHARMA-Methode untersucht, wie sie beispielsweise für Homologiemodelle oder experimentelle Strukturen aus Elektronenmikroskopie oder gering aufgelöster Röntgen-Kristallographie vorliegt. Die Methode erzielt hohe Erfolgsraten (30 – 100%) und ist daher ein wertvolles Werkzeug zur Untersuchung von Liganden-Protein-Komplexen.

# Résumé


## Introduction

Le but de la recherche pharmaceutique est de contrôler l'activité de l'organisme de manière à obtenir une réponse thérapeutique. Dans plusieurs cas, on recherche à inhiber l'action de protéines avec l'aide de petites molécules (ligands). Elles sont supposées interagir spécifiquement avec la protéine d'intérêt pour contrôler son activité.
La découverte d'un tel inhibiteur est difficile et demande un développement de moyens conséquents avant qu'un médicament n'atteigne la phase clinique. L'une des phases clefs de ce procédé est la détermination du mode d'interaction entre un ligand et son récepteur. Cette tâche peut être entravée par l'absence de structure du complexe protéine-ligand.
En effet, les inhibiteurs de protéines sont souvent découverts à travers des tests d'activité à grande échelle ou l'on ne cherche pas à déterminer leurs représentations en trois dimensions mais plutôt à repérer leur potentiel d'inhibition. Cependant, même si la structure du complexe (protéine-ligand) n'est pas accessible pour chaque ligand, certaines peuvent être connues. De plus, la structure de la protéine seule (Apo structure) peut être répertoriée dans une banque de donnée telle que la ≪Protein Data Bank≫. C'est pour répondre à ce besoin que nous présentons dans ce travail de thèse, une méthode capable de déterminer la structure de complexes protéine-ligands.

## Résultats et Conclusion

Les chapitres 1 et 2 présentent les bases de la RMN qui permettront de surmonter le problème. Dans la méthode INPHARMA (Inter-ligands Nuclear Overhauser Effect for Pharmacophore Mapping), les inter-ligands NOEs (INPHARMA NOEs) sont utilisés pour déterminer l'orientation relative de deux ligands qui interagissent de manière compétitive avec un même récepteur. Les inter-ligands NOEs sont des mesures de transferts d'aimantations d'une molécule à une autre à travers l'espace. Etant donné que les ligands sont en compétition pour un même récepteur, ils se partagent le site d'interaction et par conséquent occupent le même espace au sein de la protéine à tour de rôle. Lors de cette cinétique d'échange des ligands, une partie de l'aimantation de chaque ligand va être transférée vers la protéine. Ce processus se fait par couplages dipolaires entre les protons de chaque molécule. L'aimantation qui est à présent sur la protéine va être conservée jusqu'à ce qu'un second ligand vienne à son tour dans la même poche d'interaction et récupère cette aimantation déposée par son compétiteur. Le même transfert s'opère entre la protéine et la petite molécule si bien qu'au finale tout apparait comme si les ligands s'étaient transmis l'aimantation de l'un directement à l'autre; Or c'est bien la protéine qui a relayé ce transfert. Les ligands quand à eux n'étaient jamais proche l'un de l'autre. Le transfert d'aimantation ne pouvait donc pas se faire sans la présence de la protéine. Les couplages dipolaires, sources de ces transferts, dépendent fortement des distances entre les protons de chaque molécule (ligands et protéine). C'est grâce à cette information que l'on peut déterminer la manière dont les ligands viennent interagir avec leur récepteur.

Cette nouvelle approche ouvre la voie à des applications pharmaceutiques, également au stade initial du développement, quand l'information structurale via la cristallographie par Rayons X est difficile d'accès. Les INPHARMA NOEs mesurés entre deux ligands en compétition pour un même récepteur macromoléculaire sont donc évalués quantitativement pour déterminer le mode d'interaction des deux ligands avec leur récepteur. Cette approche repose sur des calculs complexes qui font intervenir la matrice de relaxation par couplages dipolaires et échanges chimiques. Ces calculs sont très demandeur en termes de puissance et de quantité de processeurs. Dans le chapitre 3, nous démontrons que les INPHARMA NOEs, mesurés en présence de récepteurs sélectivement protonés, peuvent être interprétés de manière semi quantitative.

De cette manière, nous nous affranchissons du formalisme qui nécessite des moyens importants sans que la méthode ne perde sa capacité de prédiction de structures de complexe protéine-ligands.

Le succès de la mis en pratique de la méthode INPHARMA repose sur le choix de plusieurs paramètres physiques. Ces paramètres dépendent des caractéristiques du système étudié, tel que la taille des molécules, leurs poids qui définiront les temps de corrélations rotationnels, les affinités des ligands pour leur récepteur, le temps de mélange de l'expérience, etc. Le chapitre 4 présente une analyse théorique approfondie des INPHARMA NOEs dépendant des paramètres expérimentaux et des constantes physiques. Cette analyse permet de choisir les paramètres expérimentaux les mieux adaptés et par conséquent d'obtenir un résultat optimal.

Le chapitre 5 s'attache à étudier les performances de la méthode dans les situations où la résolution de la structure du récepteur est faible. C'est le cas des modèles basses résolutions obtenues par modélisations moléculaire, par homologie, ou par des méthodes expérimentales comme la microscopie électronique ou la cristallographie par Rayons X à basse résolution. Dans ces conditions nos prédictions sont largement correctes (en moyenne de 30 à 100 %) et bien au-delà des alternatives proposées par des méthodes purement *in silico*. La méthode INPHARMA apparaît ainsi comme performante et nous espérons qu'elle contribuera à résoudre des structures de complexes moléculaires protéine-ligand.

# Chapter 1

# Introduction

## 1.1 Nuclear Magnetic Resonance

### 1.1.1 Basic concepts

The nuclear Overhauser effect is certainly one the most prominent concepts in NMR spectroscopy. For structure determination, the NOE tool is exceptionally important. From the discovery and the first application to the most recent findings, the NOE has always played a key role in NMR spectroscopy. Indeed, the NOE is directly related to the inter-nucleus distance *(vide infra)*. This relationship is well known in the scientific community. But it appears that NOEs are richer in content and contain extremely complex information, such as reporting on motions occurring over a range of time-scales. In order to catch a glimpse of this tremendously abundant source of information, we need first to shed light on the origin of the NOE.

Running an NMR experiment consists of first exciting the molecular spins in an intelligent manner and then waiting for return to the equilibrium state. Independent of the equilibrium state, the process of reaching this state is termed relaxation. In other words, relaxation is the process by which, over time, the spins, and therefore the bulk magnetization, return to the equilibrium position. Describing the effect of relaxation is straightforward, while understanding its origin and its relationship to molecular motions is more difficult. It is recognized by the NMR community to be one of the most, if not the most, difficult part of the NMR theory. The goal here is not to derive rigorously all the key equations, but rather to illuminate the reader

intuitively about the key concepts. Our discussion will reveal the importance of the time-scale of molecular motions. Having explained the fundamental phenomenon of relaxation, we will proceed to describe the origin of NOE. Subsequently, the essential formalism will be developed.

For simplicity, and without losing generality, we consider henceforth that the equilibrium state is the thermal equilibrium state. This supposes that all coherences have vanished and the spin energy level populations are given by the Boltzmann distribution. Relaxation to this state may be divided into two processes. First, spin-spin relaxation concerns the decay of coherences. Second, spin-lattice relaxation refers to the relaxation of the spin state populations to the Boltzmann distribution values. They are also known as transverse relaxation and longitudinal relaxation, respectively.

For spin-$\frac{1}{2}$ nuclei (i.e. $^{1}$H , $^{13}$C and $^{15}$N ), relaxation is caused by fluctuation of the magnetic field at the site of the nuclear spins. This is due ultimately to molecular motions, both local and global, such as rotational tumbling in solution. There are many sources of fluctuations of the magnetic field. We can consider, as an example, the direct dipole-dipole coupling between two spins belonging to the same molecule. Each spin is associated with a small magnetic moment and therefore creates a small magnetic field around itself. As the molecule tumbles in solution, the direction and the magnitude of the magnetic field generated by the first spin at the site of the second spin changes. Another source of local field fluctuations is the so-called chemical shift anisotropy (CSA). The molecular electron current induced by an external magnetic field generates local fields. In a similar manner to the dipole-dipole interaction, molecular tumbling leads to changes in the direction and the magnitude of these local fields. The small fluctuations of the magnetic field are responsible for the relaxation of the spin system to its equilibrium state.

We can distinguish two types of field fluctuations. The first type consists of the transverse components of the local fields, which contain a component oscillating at the Larmor frequency (the resonance frequency of the spin system). These transverse fields represent the non-secular contribution to relaxation and are responsible for the longitudinal relaxation and part of the transverse relaxation. The second type of field fluctuation is that from the longitundinal components of the fluctuating local fields, which constitute the secular contribution and is partially responsible for transverse relaxation. In

quantum physics, a secular perturbation is one that changes the energy of the system but not its wavefunction, while a non-secular perturbation changes both. Above, we cited a few examples of sources of random field fluctuations. Different sources give rise to fluctuations of different amplitude. In liquid state NMR and for the majority of the time, the dipole-dipole interaction is dominant. If several mechanisms are active simultaneously, for example when the CSA becomes significant at high magnetic fields or when there are multiple dipole-dipole interactions, there is the possibility of observing cross-correlated relaxation. Random field relaxation theory derives the transition probabilities of a spin to change state, from an excited state, "spin up" to a low energy state, "spin down", and vice versa. For the sake of clarity and simplicity, we do not show the mathematical derivation of the theory. But we recall the main results for the case of the principal relaxation mechanism, the dipole-dipole interaction. This interaction is the origin of the NOE.

### 1.1.2 Dipole-dipole relaxation and the nuclear Overhauser effect

The complete form of the dipole-dipole interaction between spins $j$ and $k$ is represented in the spin Hamiltonian by $H_{jk}^{\mathrm{DD}}$, equation 1.1, where $\mathbf{e_{jk}}$ is the internuclear unit vector joining spin $j$ to spin $k$, and $\mathbf{I_j}$ and $\mathbf{I_k}$ are the nuclear spin angular momenta of spins $j$ and $k$, respectively. It is possible to simplify the spin Hamiltonian $H_{jk}^{\mathrm{DD}}$, by applying the so-called secular approximation. It arises because the spin interactions are dominated by the large interaction with the external magnetic field, which tends to hide some components of the internal spin interactions. Further simplification of the secular Hamiltonian is often possible as the result of rapid molecular motion. In isotropic liquids, the secular dipole-dipole interactions $H_{jk}^{\mathrm{DD,secular}}$ average to zero, $\langle 3\cos^2\theta_{jk} - 1\rangle = 0$. Consequently, this effect cannot be seen in a regular NMR spectrum. Nevertheless, both the non-secular part of the dipole-dipole coupling and the motion-averaged secular component can still cause relaxation.

$$\begin{cases} H_{jk}^{\mathrm{DD}} = -\dfrac{\mu_0}{4\pi}\dfrac{\hbar\gamma_j\gamma_k}{r_{jk}^3}(3(\mathbf{I_j}\cdot\mathbf{e_{jk}})(\mathbf{I_k}\cdot\mathbf{e_{jk}}) - \mathbf{I_j}\cdot\mathbf{I_k}) \\ H_{jk}^{\mathrm{DD,secular}} = -\dfrac{\mu_0}{4\pi}\dfrac{\hbar\gamma_j\gamma_k}{r_{jk}^3}\dfrac{1}{2}(3\cos^2\theta_{jk} - 1)(3I_{jz}I_{kz} - \mathbf{I_j}\cdot\mathbf{I_k}) \end{cases} \tag{1.1}$$

We consider a weakly coupled homonuclear spin system, e.g. two isolated protons with different chemical shifts. There are four eigenstates, $|\alpha\alpha\rangle$, $|\alpha\beta\rangle$, $|\beta\alpha\rangle$, $|\beta\beta\rangle$, and therefore twelve possible transitions between these states. The energy values are the eigenvalues of the interaction Hamiltonian in the canonical space of the spin system. There are eight single-quantum transitions. During such a transition, only one of the two spins is changing state, from "up" to "down" or vice versa (Fig. 1.1).

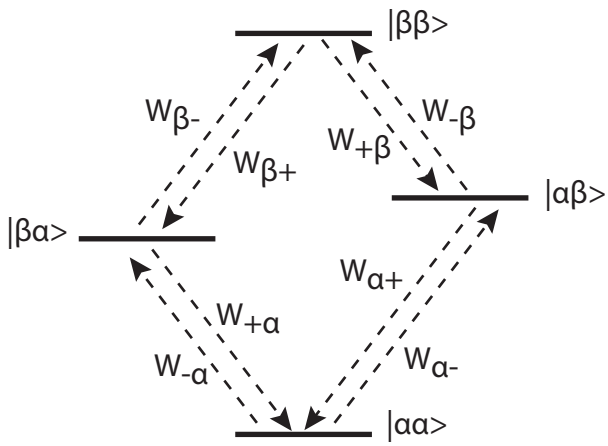



**Figure 1.1:** Single-quantum transitions in a homonuclear system

There are two double-quantum transitions and two zero-quantum transitions. In the first case, both spins invert their spin state in the same sense. In the second case, both spins also change state, but in the opposite sense (Fig. 1.2).

In the case of dipole-dipole interactions and using the results of random-field fluctuation theory we can determine the probability of the transitions.

$$\begin{cases} W_{\alpha+} = W_{+\alpha} = W_{\beta+} = W_{+\beta} = W_1(1 + \frac{1}{2}\mathrm{B}) \\ W_{\alpha-} = W_{-\alpha} = W_{\beta-} = W_{-\beta} = W_1(1 - \frac{1}{2}\mathrm{B}) \\ W_{++} = W_2(1 + \mathrm{B}) \\ W_{--} = W_2(1 - \mathrm{B}) \\ W_{+-} = W_{-+} = W_0 \end{cases} \tag{1.2}$$

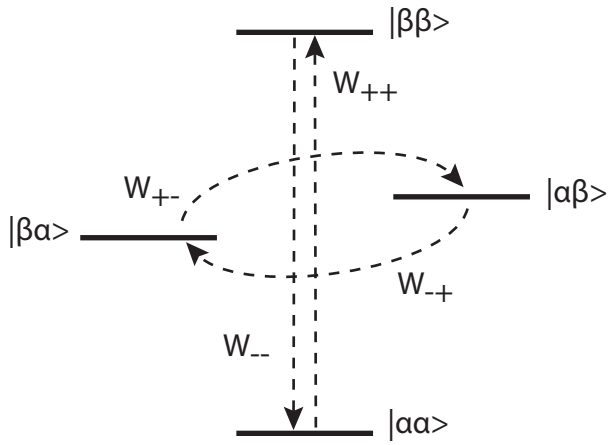



**Figure 1.2:** Double-quantum and zero-quantum transitions in a homonuclear system

$$W_0 = \frac{1}{10}b^2J(0), \quad W_1 = \frac{3}{20}b^2J(\omega_0), \quad W_2 = \frac{3}{5}b^2J(2\omega_0),$$
$$J(\omega) = \frac{\tau_c}{1+(\omega\tau_c)^2}, \quad b = -\frac{\mu_0}{4\pi}\frac{\hbar\gamma^2}{r^3} \tag{1.3}$$

$J(\omega)$ is the spectral density function; $\tau_c$ is the correlation time of the random field, which also corresponds to the rotational correlation time of the molecule. It is approximately equal to the time taken for a root-mean-square rotation of one radian. In general, small molecules have short correlation times while large molecules have long correlation times. The correlation time is also influenced by external parameters such as temperature and viscosity. In the first-order approximation of the Boltzmann distribution, B is a factor to thermally correct the transition probabilities, B = $\frac{\hbar\gamma B_0}{k_b T}$. Indeed, the equilibrium state of the spin system corresponds to a Boltzmann distribution of populations with the lowest energy state, $|\alpha\alpha\rangle$, more populated than the states $|\alpha\beta\rangle$, $|\beta\alpha\rangle$, and with $|\beta\beta\rangle$ being the least populated state. The terms B ensure that is the case.

The factor $b$ is a combination of fundamental physical constants and system parameters, incldуing the vacuum magnetic permeability, $\mu_0$, the gyromagnetic ratio, $\gamma$, and the distance between the spins, $r$. The magnetic permeability does not change appreciably and is constant within a homogenous sample.

The gyromagnetic ratio is dependent on the nucleus considered. If we excite carbons the transition probabilities and therefore the resulting relaxation effects will be weaker than if we excite the protons. Indeed, the gyromagnetic ratio of the proton is four times larger than that of carbon. Finally, the distances represent the link between the relaxation rates and molecular structure.

### 1.1.3 The Solomon equations

Equations describing the rate of change of population for each state can be derived by taking into account the probabilities of transitions to and from the other states, according to the diagrams in Fig. 1.1 and in Fig. 1.2. The kinetic equation for the population of the state $|\alpha\alpha\rangle$ is:

$$\frac{d}{dt}\rho_{\alpha\alpha} = -(W_{-\alpha}+W_{--}+W_{\alpha-})\ \rho_{\alpha\alpha}+W_{+\alpha}\ \rho_{\beta\alpha}+W_{++}\ \rho_{\beta\beta}+W_{\alpha+}\ \rho_{\alpha\beta} \quad (1.4)$$

Considering the relation between the spin-state populations and the longitudinal magnetization of the two spins:

$$\begin{aligned}\langle I_{1z}\rangle &= \frac{1}{2}\left(\rho_{\alpha\alpha}-\rho_{\beta\alpha}-\rho_{\beta\beta}+\rho_{\alpha\beta}\right)\\ \langle I_{2z}\rangle &= \frac{1}{2}\left(\rho_{\alpha\alpha}+\rho_{\beta\alpha}-\rho_{\beta\beta}-\rho_{\alpha\beta}\right)\end{aligned} \quad (1.5)$$

from which, after some algebra, we can derive the Solomon equations for the longitudinal relaxation of the two-spin system:

$$\frac{d}{dt}\begin{pmatrix}\langle I_{1z}\rangle\\ \langle I_{2z}\rangle\end{pmatrix} = \begin{pmatrix}-R_{auto} & R_{cross}\\ R_{cross} & -R_{auto}\end{pmatrix}\begin{pmatrix}\langle I_{1z}\rangle-\langle I_{1z}\rangle_{eq}\\ \langle I_{2z}\rangle-\langle I_{2z}\rangle_{eq}\end{pmatrix} \quad (1.6)$$

where $\langle I_{1z}\rangle_{eq}$ and $\langle I_{2z}\rangle_{eq}$ are the thermal equilibrium values of the angular momentum operators, given by $\langle I_{1z}\rangle_{eq} = \langle I_{2z}\rangle_{eq} = \frac{1}{4}\mathrm{B}$. $R_{auto}$ is the auto-relaxation rate constant, given by $R_{auto} = W_0 + 2W_1 + W_2$. The auto-relaxation rate increases with the rotational correlation time of the molecule, and therefore becomes faster as the molecule increases in size.
$R_{cross}$ is the cross-relaxation rate constant, given by $R_{cross} = W_0 - W_2$. The cross-relaxation rate can be positive or negative depending on the rotational correlation time and is null for $\tau_{c,null} = \frac{\sqrt{5}}{2\omega_0}$. Cross-relaxation leads to

the nuclear Overhauser effect, whereby the population difference across the transitions of one spin are affected by the population difference across the transitions of the other spin. It is therefore possible to cancel the NOE under certain conditions. The NOESY experiment, which exploits the NOE to transfer longitudinal magnetization between spatially proximal spins, is one of the most utilized for molecular structure determination.

### 1.1.4 NOESY

The NOESY pulse sequence is relatively short.

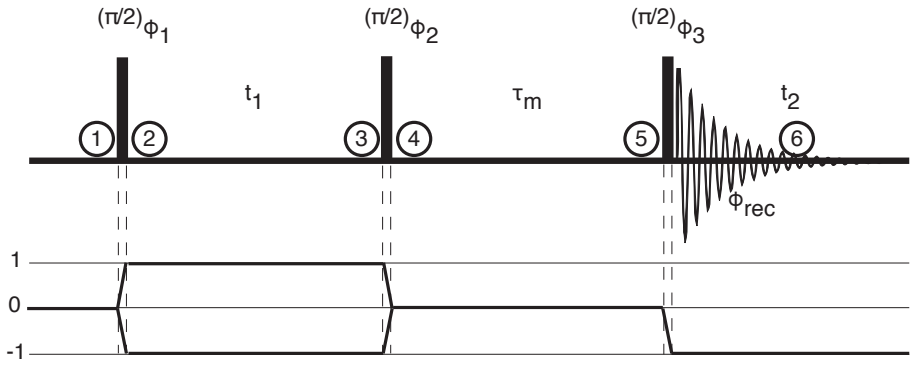



**Figure 1.3:** Diagram of the NOESY pulse sequence with the coherence transfer pathways

The Hamiltonian is dominated by the Zeeman interaction. We consider here that the two spins are sufficiently widely separated in the chemical structure to preclude any J-coupling (such coupling would only change the shape of the peak in the spectrum in any case). Furthermore, we saw above that the secular part of the dipolar interaction averages to zero with molecular tumbling in solution.

Without neglecting the chemical shift offsets of the two spins, we can follow the magnetization through the course of the pulse sequence with the product operator representation of the density matrix.

| Cycle | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_{rec}$ |
|---|---|---|---|---|
| 0 | $\psi$ | $\pi$ | 0 | 0 |
| 1 | $\psi+\pi$ | $\pi$ | 0 | $\pi$ |
| 2 | $\psi$ | $\pi$ | $\pi/2$ | $\pi/2$ |
| 3 | $\psi+\pi$ | $\pi$ | $\pi/2$ | $3\pi/2$ |
| 4 | $\psi$ | $\pi$ | $\pi$ | $\pi$ |
| 5 | $\psi+\pi$ | $\pi$ | $\pi$ | 0 |
| 6 | $\psi$ | $\pi$ | $3\pi/2$ | $3\pi/2$ |
| 7 | $\psi+\pi$ | $\pi$ | $3\pi/2$ | $\pi/2$ |

**Table 1.1:** Eight-step phase cycles for the NOESY experiment. The angles $\phi_{1,2,3}$ are the phases of the corresponding pulses in the NOESY pulse sequence and $\phi_{rec}$ is the receiver reference phase (Fig. 1.3). $\psi = (0; -\pi/2)$ changes according to the States procedure in order to generate pure absorption line-shapes.

$$\begin{aligned}
\rho_{(\mathbf{1})} &= I_{1z} + I_{2z} \\
\rho_{(\mathbf{2})} &= -I_{1y} - I_{2y} \\
\rho_{(\mathbf{3})} &= (-I_{1y}\cos(\Omega_1 t_1) + I_{1x}\sin(\Omega_1 t_1) - I_{2y}\cos(\Omega_2 t_1) + I_{2x}\sin(\Omega_2 t_1))\, e^{\frac{-t_1}{T_2}} \\
\rho_{(\mathbf{4})} &= (I_{1z}\cos(\Omega_1 t_1) + I_{1x}\sin(\Omega_1 t_1) + I_{2z}\cos(\Omega_2 t_1) + I_{2x}\sin(\Omega_2 t_1))\, e^{\frac{-t_1}{T_2}}
\end{aligned} \tag{1.7}$$

The phases of the pulses are taken from table 1.1. The transverse magnetization relaxes during the $t_1$ time with a relaxation time constant $T_2$. As can be seen from the coherence transfer pathway shown below the pulse sequence in Fig. 1.1, the desired magnetization between points 4 and 5 is longitudinal. Therefore, the phase cycle applied acts to suppress the transverse magnetization during this period. The remaining longitudinal relaxation evolves during the mixing time $\tau_m$ according to the Solomon equations (Eq. 1.6). After solving the equation *(vide infra)*, we observe that the magnetization is mixed as follows:

$$\begin{aligned}
\rho_{(\mathbf{5})} = &(I_{1z}a_{11}(\tau_m)\cos(\Omega_1 t_1) + I_{1z}a_{21}(\tau_m)\cos(\Omega_2 t_1))\, e^{\frac{-t_1}{T_2}} \\
&+ (I_{2z}a_{22}(\tau_m)\cos(\Omega_2 t_1) + I_{2z}a_{12}(\tau_m)\cos(\Omega_1 t_1))\, e^{\frac{-t_1}{T_2}}
\end{aligned} \tag{1.8}$$

where the amplitudes are given by:

$$a_{11}(\tau_m) = a_{22}(\tau_m) = \cosh(R_{cross}\tau_m)e^{-R_{auto}\tau_m} \tag{1.9}$$
$$a_{12}(\tau_m) = a_{21}(\tau_m) = \sinh(R_{cross}\tau_m)e^{-R_{auto}\tau_m} \tag{1.10}$$

Finally, the last pulse gives:

$$\begin{aligned}\rho_{(\mathbf{6})} = &(-I_{1y}a_{11}(\tau_m)\cos(\Omega_1 t_1) - I_{1y}a_{21}(\tau_m)\cos(\Omega_2 t_1))\, e^{\frac{-t_1}{T_2}} \\ &- (I_{2y}a_{22}(\tau_m)\cos(\Omega_2 t_1) - I_{2y}a_{12}(\tau_m)\cos(\Omega_1 t_1))\, e^{\frac{-t_1}{T_2}}\end{aligned} \tag{1.11}$$

The signal is therefore:

$$\begin{aligned}s_{cos}(t_1,t_2) = &a_{11}(\tau_m)\cos(\Omega_1 t_1)e^{i\Omega_1 t_2}e^{-\frac{t_1+t_2}{T_2}} \\ &+ a_{21}(\tau_m)\cos(\Omega_2 t_1)e^{i\Omega_1 t_2}e^{-\frac{t_1+t_2}{T_2}} \\ &+ a_{22}(\tau_m)\cos(\Omega_2 t_1)e^{i\Omega_2 t_2}e^{-\frac{t_1+t_2}{T_2}} \\ &+ a_{12}(\tau_m)\cos(\Omega_1 t_1)e^{i\Omega_2 t_2}e^{-\frac{t_1+t_2}{T_2}}\end{aligned} \tag{1.12}$$

The subscript *cos* denotes a signal with a cosine modulation in $t_1$. We need the corresponding sine-modulated signal in order to build the complete signal that will enable generation of pure absorption line-shapes after Fourier transform. All the product operator calculations were carried out with $\psi = 0$ as indicated in table 1.1. Following the "States" procedure, we should repeat them with $\psi = -\pi/2$. This provides the same signal with sine modulation:

$$\begin{aligned}s_{sin}(t_1,t_2) = &a_{11}(\tau_m)\sin(\Omega_1 t_1)e^{i\Omega_1 t_2}e^{-\frac{t_1+t_2}{T_2}} \\ &+ a_{21}(\tau_m)\sin(\Omega_2 t_1)e^{i\Omega_1 t_2}e^{-\frac{t_1+t_2}{T_2}} \\ &+ a_{22}(\tau_m)\sin(\Omega_2 t_1)e^{i\Omega_2 t_2}e^{-\frac{t_1+t_2}{T_2}} \\ &+ a_{12}(\tau_m)\sin(\Omega_1 t_1)e^{i\Omega_2 t_2}e^{-\frac{t_1+t_2}{T_2}}\end{aligned} \tag{1.13}$$

Once both modulations are obtained, they can be combined according to the "States" procedure in order to generate a signal with pure absorption lineshapes after Fourier transformation.

In the following expression, $FT_1$ and $FT_2$ represent Fourier transforms along the $t_1$ and $t_2$ dimensions, respectively:

$$\begin{aligned} S(\omega_1,\omega_2) =& \mathrm{Re}\Big\{ \mathrm{FT}_1 \left( \mathrm{Re}\Big\{ \mathrm{FT}_2 \left( s_{cos}(t_1,t_2) \right) \Big\} + i\mathrm{Re}\Big\{ \mathrm{FT}_2 \left( s_{sin}(t_1,t_2) \right) \Big\} \right) \Big\} \\ =& a_{11}\delta(\omega_1-\Omega_1) * \delta(\omega_2-\Omega_1) * \mathcal{L}_{T_2}(\omega_1,\omega_2) \\ &+ a_{21}\delta(\omega_1-\Omega_2) * \delta(\omega_2-\Omega_1) * \mathcal{L}_{T_2}(\omega_1,\omega_2) \\ &+ a_{22}\delta(\omega_1-\Omega_2) * \delta(\omega_2-\Omega_2) * \mathcal{L}_{T_2}(\omega_1,\omega_2) \\ &+ a_{12}\delta(\omega_1-\Omega_1) * \delta(\omega_2-\Omega_2) * \mathcal{L}_{T_2}(\omega_1,\omega_2) \end{aligned} \tag{1.14}$$

$$\mathcal{L}_{T_2}(\omega_1,\omega_2) = \frac{T_2^{-1}}{T_2^{-2}+\omega_1^2} \frac{T_2^{-1}}{T_2^{-2}+\omega_2^2} \tag{1.15}$$

The final signal in the resulting NOESY spectrum is composed of four peaks for the case of a two-spin system with no chemical exchange. Two of them are diagonal peaks ($\Omega_1$,$\Omega_1$), ($\Omega_2$, $\Omega_2$) and two are cross-peaks ($\Omega_1$,$\Omega_2$), ($\Omega_2$,$\Omega_1$). The sign of the cross-peaks relative to the diagonal peaks depends on the cross-relaxation rate. If the molecule is large it tumbles slowly, leading to a long rotational correlation time and positive cross-relaxation rate. This gives rise to positive cross-peaks (assuming that the diagonal peaks are also positive). In contrast, if the molecule is small, it will tumble rapidly and hence have a short correlation time. In this case, the cross-relaxation rate will be negative, leading to cross-peaks with negative intensity. In the particular case where the cross-relaxation rate is zero we do not see any cross-peaks in the NOESY spectrum. This would occur for a molecule having a rotational correlation time of 360 ps at a field of 500 MHz, for example. The Lorentzian function describes the peak shape, with the height and width being related to the transverse relaxation rate $T_2^{-1}$.

### 1.1.5 Chemical exchange

It is common that molecules adopt different conformations over time. If this is the case, then the spin (a proton, for example) will experience different electronic environments that may not be magnetically equivalent. We can describe such a situation using the formalism developed in the previous section (1.1.4). Now we have a single spin with two distinct magnetic states (with $I_{1z}$ and $I_{2z}$ characterizing two magnetic states of a single spin) instead of

two spins in fixed magnetic states (with $I_{1z}$ characterizing the magnetic state of one spin and $I_{2z}$ characterizing the magnetic state of the second spin). In this case, the cross-relaxation term can only be positive and is equal to the chemical exchange rate, and the auto-relaxation term is equal to the longitudinal relaxation rate. Therefore, the cross-peaks have the same sign as the diagonal peaks in the NOESY spectrum.

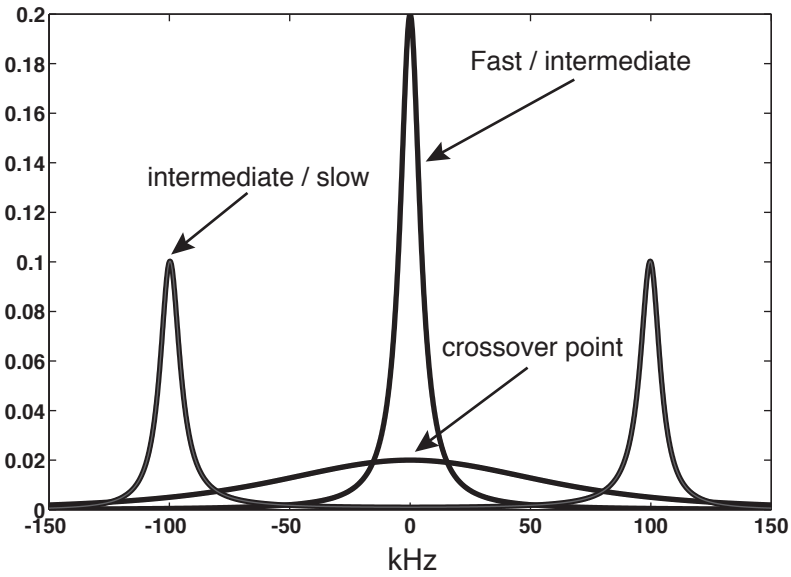



**Figure 1.4:** Peak line-shape under chemical exchange. The chemical exchange rate, $k$, ranges from fast ($k >> 100kHz$) to slow ($k << 100kHz$) exchange. The narrow line at 0 kHz corresponds to a fast chemical exchange rate, while the two lines at $\pm$ 100 kHz correspond to a slow chemical exchange rate, and reveal the two distinct electronic environments of the spin. The broad middle line corresponds to the crossover point where $k \approx 100kHz$.

While the NOE effect is well described by the previous formalism (Eq. 1.14), the signal line-shape will depend on the regime of the chemical exchange. Broadly speaking, we consider all motion that is not averaged by the tumbling of the molecule as chemical exchange. This translates to a time-scale ranging from sub-microsecond to seconds (or even longer for some special processes). This range of time-scales may be divided into three windows. The first, starting from sub-microsecond time-scales is characterized by fast and intermediate exchange. It ends at the so-called crossover point, where

the perturbation of the peak line-shape becomes strongly pronounced. The second window, at time-scales longer than the cross-over point, correspond to intermediate and slow exchange regimes. The final time window at very long time-scales corresponds to exchange-mediated longitudinal magnetization transfer observed during the mixing time (typically from ms to second) of the NOESY experiment.
In the slow exchange regime, the chemical shift (Zeeman) Hamiltonian is not averaged. Therefore, the two states are still visible as distinct peaks in the spectrum (Fig. 1.4).

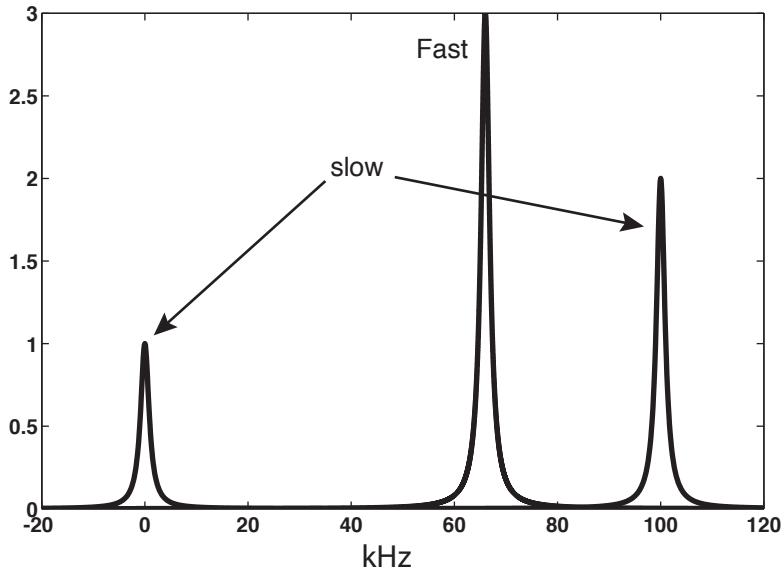



**Figure 1.5:** Peak lineshape under chemical exchange. The two magnetic sites are not equally populated. The fast exchange peak appears at the barycenter of the slow exchange peaks positions weighted by their population (2:1).

As the chemical exchange rate increases and reaches the cross-over point, the two lines merge and become very broad. This occurs at a chemical exchange rate approximately equal to half the difference in chemical shift (in Hz) between the two sites. This phenomenon can be seen as two consecutive processes. First, the two lines corresponding to the slow exchange regime, ± 100 kHz, start to converge towards the middle position, 0 kHz, and finally merge. Second, the rate of exchange becomes such that the net NMR signal

decays very rapidly, which, after Fourier transformation, gives a very broad signal in the frequency domain. At the cross-over point, the signal becomes so broad that it is usually lost in the noise (Fig. 1.4).
When the averaging between the environments becomes even faster, the single line begins to sharpen. Note that integral of the signal intensity is constant, and therefore as the peak gets sharper, the peak height increases. This is now the fast exchange regime, sometimes called motional narrowing (Fig. 1.4).

Thus far the populations of the two sites were implicitly assumed to be equal. If they are not equal, then motional average will be weighted according to the relative populations. Hence, the single peak in the fast exchange regime will not appear in the middle of the two slow-exchange peaks, but in the barycenter of the two positions weighted by the population of each site (Fig. 1.5).

### 1.1.6 The NOE

The time evolution of peak volumes during the mixing time of a NOESY experiment is described by the following equation, [1]:

$$\frac{\mathbf{dM(t)}}{\mathbf{dt}} = -(\mathbf{R} + \mathbf{K}) \cdot (\mathbf{M(t)} - \mathbf{M_{eq}}) \tag{1.16}$$

$\mathbf{M}$ is the matrix of volumes for the spins that we consider in our system, with $\mathbf{M_{eq}}$ representing the Boltzmann equilibrium magnetization. $\mathbf{R}$ is the relaxation matrix and $\mathbf{K}$ is the chemical exchange matrix. The solution of equation 1.16 provides all the peak volumes found in the NOESY spectrum acquired with a mixing time $\tau_m$.

$$\mathbf{\Delta M}(\tau_\mathbf{m}) = \exp\left[-(\mathbf{R} + \mathbf{K}) \cdot \tau_\mathbf{m}\right] \cdot \mathbf{\Delta M(0)} \tag{1.17}$$

#### Two-state system

In order to gain deeper insight into the underlying theory (called the full relaxation matrix approach) we develop here the calculations for one spin undergoing chemical exchange, illustrating aspects from the previous section about the NOE and chemical exchange. We call the two states experienced by the spin 'M' and 'N', and consider the following kinetic equation:

$$M \underset{k_N}{\overset{k_M}{\rightleftharpoons}} N \tag{1.18}$$

The chemical exchange matrix **K** is built from the following equations:

$$\frac{d[M]}{dt} = k_N[N] - k_M[M] \tag{1.19}$$

$$\frac{d[N]}{dt} = -k_N[N] + k_M[M] \tag{1.20}$$

Thus the **K** matrix is

$$\mathbf{K} = \begin{pmatrix} k_M & -k_N \\ -k_M & k_N \end{pmatrix} \tag{1.21}$$

The relaxation matrix **R** is:

$$\mathbf{R} = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix} \tag{1.22}$$

As the system we consider has only one spin, dipole-dipole cross-relaxation is not relevant. However, dipole-dipole auto-relaxation processes characterised by the rate constant $\rho$, are still possible due to interactions with the external environment.
Thus the complete equation system can be expressed as:

$$\frac{\mathbf{d}}{\mathbf{dt}} \begin{pmatrix} M \\ N \end{pmatrix} = -\begin{pmatrix} \rho + k_M & -k_N \\ -k_M & \rho + k_N \end{pmatrix} \begin{pmatrix} \Delta M \\ \Delta N \end{pmatrix} \tag{1.23}$$

Since the solution of the equation is an exponential we will transform the matrix into a diagonal matrix. We search now for the eigenvalues of this matrix.

$$det \begin{pmatrix} \rho + k_M - \lambda & -k_N \\ -k_M & \rho + k_N - \lambda \end{pmatrix} = 0 \tag{1.24}$$

which leads to

$$\lambda^2 - \lambda(2\rho + k_N + k_M) + \rho^2 + (k_N + k_M)\rho = 0 \tag{1.25}$$

with

$$\Delta = (k_N + k_M)^2 \tag{1.26}$$

thus

$$\begin{cases} \lambda_+ = \rho + k_N + k_M \\ \lambda_- = \rho \end{cases} \tag{1.27}$$

and finally

$$\begin{pmatrix} \Delta M \\ \Delta N \end{pmatrix} = \begin{pmatrix} a_{MM} & a_{MN} \\ a_{NM} & a_{NN} \end{pmatrix} \begin{pmatrix} \Delta M(0) \\ \Delta N(0) \end{pmatrix} \tag{1.28}$$

where $a_{MM}, a_{NN}$ are the volumes of the diagonal peaks and $a_{MN}, a_{NM}$ are the volumes of the cross-peaks.

$$\begin{cases} a_{MM} = \frac{1}{2}\left[\left(1 + \frac{k_M - k_N}{\lambda_+ - \lambda_-}\right) e^{-\lambda_+ t} + \left(1 - \frac{k_M - k_N}{\lambda_+ - \lambda_-}\right) e^{-\lambda_- t}\right] \\ a_{NN} = \frac{1}{2}\left[\left(1 - \frac{k_M - k_N}{\lambda_+ - \lambda_-}\right) e^{-\lambda_+ t} + \left(1 + \frac{k_M - k_N}{\lambda_+ - \lambda_-}\right) e^{-\lambda_- t}\right] \\ a_{MN} = \frac{k_N}{\lambda_+ - \lambda_-}\left(e^{-\lambda_- t} - e^{-\lambda_+ t}\right) \\ a_{NM} = \frac{k_M}{\lambda_+ - \lambda_-}\left(e^{-\lambda_- t} - e^{-\lambda_+ t}\right) \end{cases} \tag{1.29}$$

In the NOESY experiment, t is $\tau_m$, the mixing time. A Taylor expansion of the exponential terms for short mixing times should reveal that the cross-peak volumes depend directly on the cross relaxation rate, $R_{cross}$ (similar to Eq. 1.6 and Eq. 1.10). With $e^x \approx 1 + x$ as the first-order approximation, $a_{MN} \propto k_N \tau_m$ and $a_{NM} \propto k_M \tau_m$.

This result is similar to that we introduced previously (Eq. 1.10). The initial slope of the build-up curve is proportional to $R_{cross}\tau_m$, at short mixing times. Overall, the NOEs depend on the relaxation constants, the kinetic constants and the equilibrium populations of the spins involved.

We have described the exchange of longitudinal magnetization by cross-relaxation and by chemical exchange for a system of two spins or of a single spin in two magnetically inequivalent sites. What happens if we combine both aspects and consider a system comprised of many hundreds of spins, as would be the case for biomacromolecules (proteins, DNAs and RNAs)?

### 1.1.7 Multi-spin systems and the NOE

Without losing generality, we now consider a system composed of one ligand (protons C and D) and a receptor (proton T). The ligand is in exchange between a bound state, $C_b$ and $D_b$, to the receptor target (T) and a free state, $C_f$ and $D_f$. The transfer of magnetization between C and D when the ligand is unbound is presumed to be negligible (large inter-proton distances

and/or small size of the ligand). The nuclear magnetization can exchange between C, D and T via the following mechanisms:

$$C_f \overset{k}{\longleftrightarrow} C_b \overset{\sigma}{\longleftrightarrow} D_b \tag{1.30}$$

$$C_f \overset{k}{\longleftrightarrow} C_b \overset{\sigma}{\longleftrightarrow} T \tag{1.31}$$

$$C_f \overset{k}{\longleftrightarrow} C_b \overset{\sigma}{\longleftrightarrow} D_b \overset{k}{\longleftrightarrow} D_f \tag{1.32}$$

$$C_f \overset{k}{\longleftrightarrow} C_b \overset{\sigma}{\longleftrightarrow} T \overset{\sigma}{\longleftrightarrow} D_b \overset{k}{\longleftrightarrow} D_f \tag{1.33}$$

Where $k$ and $\sigma$ are the rate constants for magnetization transfer through the chemical exchange and cross relaxation, respectively. Thus, the spatial information from the bound state is transferred to the free state protons, which are more easily monitored. According to the schemes shown above, the observable cross peaks in the NOESY spectrum include the exchange peaks between the free and bound ligand resonances, the exchange-relayed NOE between a free ligand proton and a bound ligand proton, Eq. (1.30), or between a free ligand proton and protein proton(s) T (protein-ligand transferred NOEs, Eq. (1.31)), the exchange-relayed NOE between free ligand protons (intra-ligand transferred NOEs, Eq. (1.32)), and the exchange-relayed NOE between C and D mediated by protein proton(s) T (protein-ligand spin-diffusion, Eq. (1.33)). Protein-ligand transferred NOEs carry information about the contact sites between the ligand and the protein upon complex formation, while the intra-ligand transferred NOEs can be used to derive the bound conformation of the ligand. Hence it is possible to probe the bio-active protein-bound ligand conformation and the protein-ligand contacts.

In addition to the above, we propose here to investigate another type of NOE effect, based on the observation of inter-ligand tr-NOE. We now take into consideration two ligands, A and B. The ligands are in fast exchange with a common receptor, T. During the mixing time of the NOESY experiment, ligand A binds to the target, T, and transfers some magnetization to the receptor proton, i.e from $H_A$ to $H_T$. Then, during the same mixing time period, ligand A dissociates from the protein and is replaced by ligand B. The magnetization that was transferred from $H_A$ to $H_T$ is now transferred

from $H_T$ to $H_B$, leading to a NOESY cross-peak between $H_A$ and $H_B$. This special cross-peak is called an "INPHARMA NOE" *(vide infra)* (Fig. 1.6).

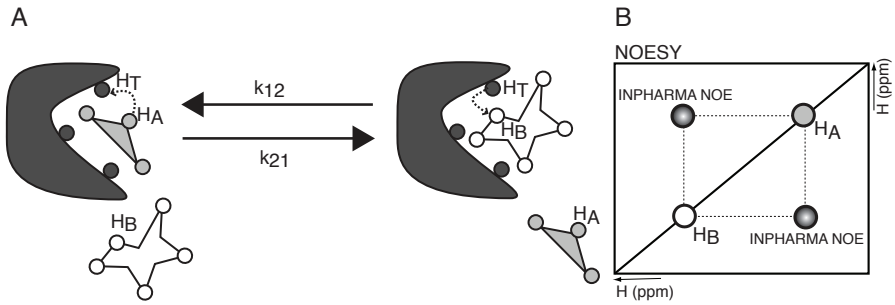



**Figure 1.6:** Schematic illustrating the principle of INPHARMA NOEs

This effect should not be confused with inter-ligand transferred-NOE peaks that appear when two ligands bind to the protein simultaneously. In the case of INPHARMA NOEs, the two ligands are never close to each other as they bind competitively. Observation of this phenomenon relies on the presence of spin diffusion. In fact, the ligand magnetization is spread throughout the protein during the mixing time. While direct NOEs can be observed up to internuclear distances of 5 Å, spin-diffusion-mediated NOEs can be observed between protons that are as far as 8 Å apart or more. If we take this information into account it becomes difficult to interpret the volume of the inter-ligand tr-NOE as direct distance information such as can be achieved with simple NOEs. We need to interpret the data via theoretical interpretation and calculations using the full relaxation matrix approach, [57, 62, 63]. Based on the experimental data, we seek to derive the relative binding mode of the two ligands in the binding pocket. This approach is called INPHARMA (Inter-ligand NOE for Pharmacophore Mapping). We explain below in more detail this method and the most salient points for understanding the form of the spectrum and the calculations.

### 1.1.8 Spin diffusion and intermolecular NOEs

We recall the equation 1.16, which describes evolution of peak volumes during the mixing time of a NOESY experiment, [1]:

$$\frac{\mathbf{dM(t)}}{\mathbf{dx}} = -(\mathbf{R} + \mathbf{K}) \cdot (\mathbf{M(t)} - \mathbf{M_{eq}}) \tag{1.34}$$

and its solution:

$$\mathbf{\Delta M}(\tau_\mathbf{m}) = \exp\left[-(\mathbf{R} + \mathbf{K}) \cdot \tau_\mathbf{m}\right] \cdot \mathbf{\Delta M(0)} \tag{1.35}$$

The kinetic model for the protein and the two ligands is assumed to be a two-state model. We consider four chemical entities, ligand A, ligand B, and the the ligand-protein complexes TA and TB. The ligands are assumed to present in a large molar excess, such that the concentration of free protein is negligible.

$$TA + B \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} TB + A \tag{1.36}$$

The relaxation matrix $\mathbf{R}$ contains the auto- and cross-relaxation terms for protons that belong to the different species, A, B, TA and TB.

$$\mathbf{R} = \begin{pmatrix} R_A^A & & & & & \\ & R_B^B & & & & \\ & & R_A^{TA} & R_{A,T}^{TA} & & \\ & & R_{T,A}^{TA} & R_T^{TA} & & \\ & & & & R_B^{TB} & R_{B,T}^{TB} \\ & & & & R_{T,B}^{TB} & R_T^{TB} \end{pmatrix} \tag{1.37}$$

The $\mathbf{R}$ matrix is written with the following conventions: $R_X^X$ contains the auto-relaxation and cross-relaxation terms between protons of ligand X (A or B) in the free form; $R_X^{TX}$ contains the auto-relaxation and cross-relaxation terms between protons of ligand X in the bound form TX; $R_{T,X}^{TX}$, $R_{X,T}^{TX}$ contain the cross-relaxation terms between the protons of ligand X and the protons of T in the complex TX; $R_T^{TX}$ contains the auto-relaxation and cross-relaxation terms between protons of the target. The $\mathbf{K}$ matrix represents the kinetic processes of the system at equilibrium. It necessarily has the same dimensions

as the **R** matrix; [X] is the concentration of the chemical entities, $k_i$ are the chemical exchange constants and **I** is the identity matrix. Each part of the matrix **K** represents one species of the system and therefore each term is a diagonal block matrix which has the same size as the corresponding block in the relaxation matrix **R**.

$$\mathbf{K} = \begin{pmatrix} k_{21}[TB]\mathbf{I} & & -k_{12}[B]\mathbf{I} & & & \\ & k_{12}[TA]\mathbf{I} & & & -k_{21}[A]\mathbf{I} & \\ -k_{21}[TB]\mathbf{I} & & k_{12}[B]\mathbf{I} & & & \\ & & & k_{12}[B]\mathbf{I} & & -k_{21}[A]\mathbf{I} \\ & -k_{12}[TA]\mathbf{I} & & & k_{21}[A]\mathbf{I} & \\ & & & -k_{12}[B]\mathbf{I} & & k_{21}[A]\mathbf{I} \end{pmatrix} \tag{1.38}$$

### Spin diffusion

All the transfers of magnetization by chemical exchange and dipole-dipole relaxation mechanisms are described by the matrices **K** and **R**. Therefore, if in the NOESY spectrum we observe cross-peaks between protons of ligand A and protons of ligand B, we must also see a connection between them in the matrix $\mathbf{K} + \mathbf{R}$:

$$\begin{pmatrix} k_{21}[TB]\mathbf{I}+R_A^A & & -k_{12}[B]\mathbf{I} & & & \\ & k_{12}[TA]\mathbf{I}+R_B^B & & & -k_{21}[A]\mathbf{I} & \\ -k_{21}[TB]\mathbf{I} & & k_{12}[B]\mathbf{I}+R_A^{TA} & R_{A,T}^{TA} & & \\ & & R_{T,A}^{TA} & k_{12}[B]\mathbf{I}+R_T^{TA} & & -k_{21}[A]\mathbf{I} \\ & -k_{12}[TA]\mathbf{I} & & & k_{21}[A]+R_B^{TB}\mathbf{I} & R_{B,T}^{TB} \\ & & & -k_{12}[B]\mathbf{I} & R_{T,B}^{TB} & k_{21}[A]\mathbf{I}+R_T^{TB} \end{pmatrix}$$

There are no terms in the matrix connecting the two block-sub-matrices, $k_{21}[TB]\mathbf{I}+R_A^A$ and $k_{12}[TA]\mathbf{I}+R_B^B$, which correspond to the protons of ligands A and B in their unbound states. Nevertheless, NOESY cross-peaks can arise between protons A and B in the free state. This is due to the fact that the magnetization is transferred from unbound A to unbound B via a multi-step process. We see that diagonal block-sub-matrices are 'linked' by chemical exchange constants. Hence, magnetization initially present on unbound A is transferred by chemical exchange to bound A (TA), followed by dipole-dipole mediated-transfer to the protons of the target (T). Another chemical exchange constant links the complex TA to the complex TB (Fig. 1.7). The

magnetization on TB is then transferred by dipole-dipole mechanisms to the protons of bound ligand B (TB). Finally, the last chemical exchange link allows transfer from bound ligand B to the unbound state (B). The ligand B then possesses the magnetization acquired while it was in the complex (TB). Initially, this magnetization arose from ligand A. This explains why cross-peaks are can appear between the two ligand A and B in their free states while there is never direct magnetization transfer between them.

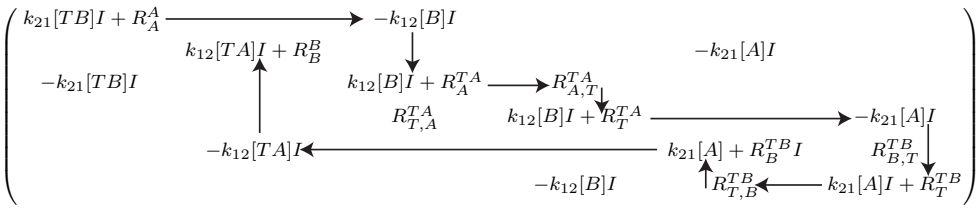



**Figure 1.7:** Illustration of the multi-step transfer of magnetization leading to INPHARMA NOEs.

We have seen that the multi-step transfer of magnetization involves the free states, the bound states and the intervening chemical exchange processes. While in the free state the relaxation rates of the ligands are quite slow because of their short rotational correlation times. In the bound state, however, the relaxation rates are fast because the ligands then tumble at the same slow rate as the protein. Thus, it is the bound-form relaxation rates that dominate the dipole-dipole-mediated transfer of magnetization. This transfer of magnetization not only reflects the bound conformation of the ligands but also their orientation in the binding pocket. Depending on the conformation and the orientation of the two ligands in their bound states, we will observe cross-peaks between different protons in the ligands (Fig. 1.8). Because the protons of the protein are crucial to the transfer of magnetization between the ligands, we must include them in the full relaxation matrix formalism. Therefore, we use the complete formalism as described by equations 1.34, 1.37 and 1.38 . We can already anticipate that a structural model for the receptor is required and specific conditions for the chemical and dipole-dipole relaxation rates are necessary for the success of the method. We will explore the detail

of the method in a following dedicated chapter.

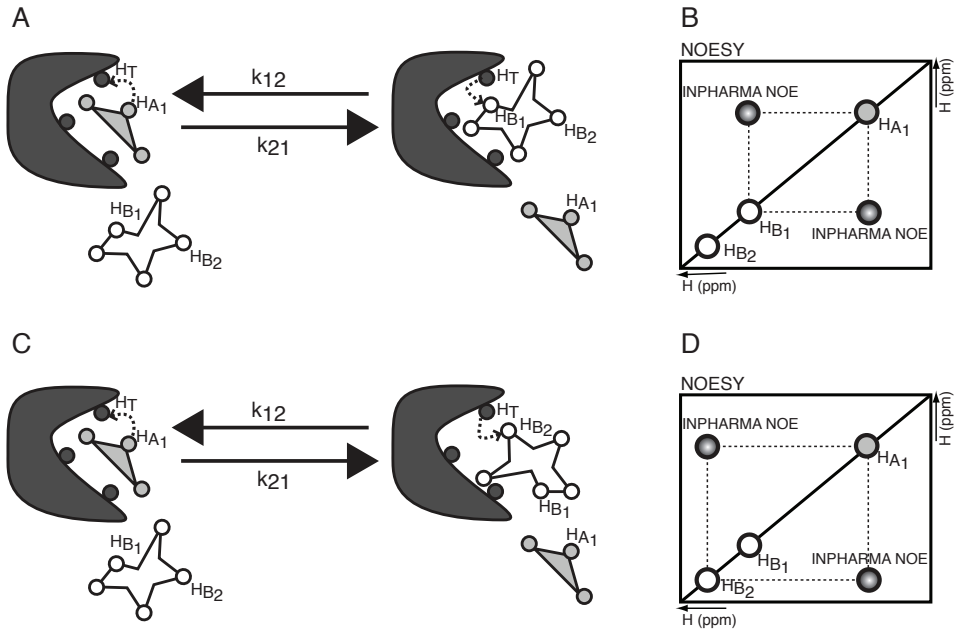



**Figure 1.8:** Qualitative representation showing distinction between two binding modes with the INPHARMA method. Panels A and C show two different binding modes for one of the two ligands. Panels B and D represent the expected NOESY spectra for the systems depicted in A and C, respectively.

## 1.2 Aim and scope of the thesis

### 1.2.1 The context

A considerable number of cellular machineries depend on highly specific interactions between protein receptors and their small-molecule ligands, such as cofactors, hormones, drugs or metabolites. Small molecules are also prime candidates for drug development. Knowledge of the interactions involved is critical for understanding the underlying natural processes and facilitating their manipulation for therapeutic purposes [21, 94, 109]. The structures of small-molecule-protein complexes are determined either by X-ray crystallography or by NMR spectroscopy [28, 36, 80]. But crystallizing the complex can be both very expensive and time-consuming. Almost thirty years ago, an alternative approach to the problem was initiated when the first *in silico* geometrical methods to determine structural information between ligand and protein were developed [50]. Today, dozens of docking and molecular dynamics programs are available and have been extensively tested [19, 22, 95]. Molecular modeling has thus become a tool of choice for the structural biologist and plays a key role in the drug discovery process. Nevertheless, the field is not yet sufficiently mature to guarantee acceptable success rates. Force-field imprecision, wide conformational spaces for both ligands and proteins, ignorance of the bioactive conformations, protein plasticity, protein dynamics, and lack of selectivity due to poor scoring functions are a few of the many problems that are currently hampering molecular modeling [52, 70, 97, 103]. Although significant breakthroughs are often made using a single technique (i.e NMR, X-ray crystallography or molecular modeling), we believe that only by a combination of methods may the quantitative basis of intermolecular interactions be understood in a rapid and reliable manner. Accordingly, we have developed a new approach, called INPHARMA, which allows the determination of the relative orientation of two competitive ligands in the receptor-binding pocket. The method is based on the observation, in a NOESY NMR experiment, of interligand, spin-diffusion-mediated, transferred NOEs between two ligands, which binding competitively and weakly to a common macromolecular receptor. My role in this project is to investigate the physical mechanisms that form the basis of the observed effec, both from theoretical and practical standpoints, and also to define the range of applicability of the method.

### 1.2.2 The studies

**Chapter 2**

During the first year of my PhD, I tested the efficacy of the method by applying it to the *de novo* derivation of the binding modes of two ligands to Protein Kinase A. The accuracy of the resulting complex structures was confirmed by comparison with the available crystal structures. The INPHARMA method was successful in determining the correct binding mode of the ligands. This chapter represents the proof-of-principle for the method.

**Chapter 3**

In the work described in this chapter, I investigated the potential of combining this new methodology with selective deuteration of the macromolecular target. Selective deuteration reduces the possible spin diffusion pathways, allowing extraction of more specific structural information. This approach has been applied, both theoretically and experimentally to the system of Protein Kinase A in complex with its two ligands. I formulated a new theoretical expression for the transferred magnetisation under reduced spin diffusion, which circumvents the computationally-intensive part of the INPHARMA method. The theoretical part was validated by the experimental results. In this case, we were also able to reproduce the correct binding modes for the ligands.

**Chapter 4**

A practical and theoretical guide for the application of the methodology in relation to both experimental and physicochemical parameters is presented. It aims to provide the information necessary for practising scientists to realise the full benefits of the method. Since the phenomenon is quite new and no detailed formalism has been released, this section sheds light on many previously unknown points.

**Chapter 5**

In this chapter, I describe my work to determine the influence of the starting protein model on the success of the method. First, I analysed the influence of the protein structure quality independently of the ligand orientation. Second, I combined both aspects, protein structure quality and ligand docking poses,

to derive the expected performance of the methodology in different 'real-case' scenarios. The method performs well, even when the receptor structure is not accurate and it is selective enough to identify the correct binding mode among several million trial modes.

All simulations were computed using a program written in C++, which is able to run in parallel on very large clusters ($\approx$ 2000 cpus). This program was written by a Mathematics/Informatics Masters student (Mael Bosson) under my guidance.

### 1.2.3 The ambition

The aim of the thesis is to provide a complete overview of the method. As mentioned above, traditionally, transfer of longitudinal magnetization between spins was restricted to spatially proximal spins within the same molecules. The novelty of the INPHARMA phenomenon is the exchange of magnetization between two competitive binders. The two ligands do not bind simultaneously to the receptor and therefore magnetization is not transferred in a "classical" manner, i.e. directly from proton to proton via spin diffusion. In reality, the magnetization is stored on the protein after being deposited by one ligand and marked by that ligand's characteristic chemical shifts. The fact that the magnetization remains on the protons of the protein, even though only for a short duration, is somewhat controversial in the NMR community. Initial opinions were that the magnetization would not "survive" long enough on the protein to allow its transfer to the second ligand. Chapter 4 unambiguously answers this concern, demonstrating that we can detect sufficient signal arising from this relayed transferred NOE under certain conditions. The signal measured is nonetheless rather weak and also described are recommendations to follow for maximizing the effect.

Having proven the hypothesis, we planned to show the fruitfulness of the method. Hence, we have chosen a system where the binding modes of the two ligands are known by a standard method, X-ray crystallography. Chapter 2 explains in detail how the method is able to retrieve the same complex structures as those determined by X-ray crystallography without needing to crystallize any complex. It represents a benchmark for the efficacy and utility of the method. Nevertheless, a structural model of the binding pocket is employed. The natural arising questions are: How precise is the method? What is required in terms of resolution for the receptor apo-structure?

Indeed, in many cases, scientists may be working on a system where the receptor structure is only poorly defined. How likely is it to be able to obtain a reliable result from the methodology under such conditions? They may want to know the accuracy of the outcome from the analysis. These questions are addressed in Chapter 5.

In some cases, scientists are able to produce protein with a particular isotopic labeling scheme. In this context, we proposed to use a specific labeling of the methyl groups of certain residues in a background of otherwise complete deutaration. This labeling scheme has the advantage of considerably reducing spin diffusion. The complete formalism developed in the chapter 4 is then no longer necessary but a new approach can be used, which considerably simplifies the interpretation and should allow further software development.

# Chapter 2

# Crystallography-Independent Determination of Ligand Binding Modes

**Abstract**

In the process of structure-based drug design, the provision of the binding mode of ligands to the cellular receptor of interest is essential. This can suffer from limited access to protein/ligand structures, especially for the low affinity ligands that are commonly obtained from high throughput screening or fragment based lead discovery. In a common scenario crystal structures are available for one or several ligands but not for all chemical series of actual interest. Here, we present a new, NMR-based approach that allows overcoming this limitation. In the INPHARMA method interligand NOEs (Nuclear Overhauser Enhancement) are utilized to determine relative orientations of different chemical fragments binding competitively to a common receptor site. This novel methodology opens the way to the application of structure-based drug design already in an early stage of drug development, when structural information via crystallography is of difficult access.

## 2.1 Introduction

Within the last decades structure based drug design (SBDD) has evolved to a powerful tool that has benefited the optimization of many low molecular weight lead structures to highly potent drugs targeting proteins of known three dimensional structure [86]. The principle of SBDD lies in the combination of different chemical moieties with the aim of obtaining a molecule that, while possessing the pharmacological properties necessary for a drug, is complementary in shape to the receptor binding pocket. This process requires knowledge of the exact structure of the protein/ligand complex. At present structural genomics initiatives provide protein structures of biomedically relevant targets at increasing rates [67] and recent structures of ion channels [3] and GPCRs [15, 83, 88] bring these protein classes within reach for SBDD. Despite these successes, the daily work of pharmaceutical discovery is often limited by the ability to obtain protein/ligand structures for the chemical series under investigation. Molecular docking is applicable for many proteins of known structure, but methods to select the correct binding pose amongst many docking models are often insufficient to deliver an unambiguous answer. High resolution crystal structures do provide an unequivocal solution for this problem but are often difficult to obtain for the lower affinity ligands (lead structures) that are commonly identified by high throughput screening or by fragment based lead discovery [86]. In view of the limited success rates of co-crystallization experiments in this stage of discovery, SBDD would benefit from methods providing the relative orientation of different chemical fragments binding competitively to a receptor site. Such an approach would provide protein/ligand structures of novel ligands or fragments in relation to the known co-crystal structure of a reference ligand.

Recently, we reported the observation by NMR of inter-ligand NOE (Nuclear Overhauser Enhancement) peaks occurring between two small ligands binding weakly and competitively to the same binding pocket of a common macromolecular receptor Fig. 2.1 [87, 89]. The measured mixture in solution contained two ligands A and B ($L_A$ and $L_B$) in a 10 to 50-fold excess to a target receptor T. As the ligands were competitive binders that never occupy the receptor binding pocket simultaneously, such NOEs did not originate from a direct transfer of magnetization between the two ligands [62]. We demonstrated that the interligand NOE peaks originate from a spin-diffusion process mediated by the protons of the receptor binding pocket. We proposed

that a number of such interligand NOEs can be used to define the relative orientation of the two ligands in the receptor binding pocket (relative binding mode) and we termed the novel effect INPHARMA (Internuclear Noes for PHARmacophore MApping) [89] (Fig. 2.1).

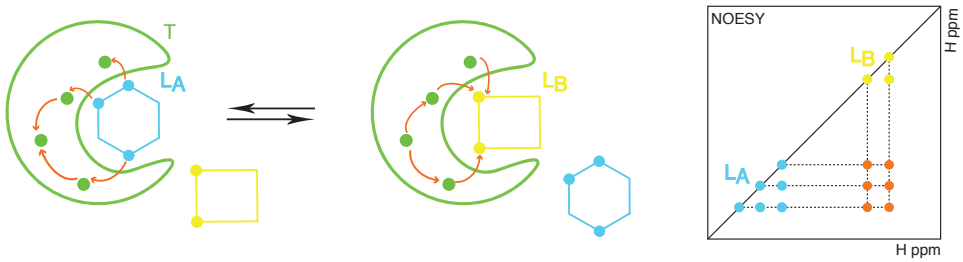



**Figure 2.1:** Schematic representation of the principle of the INPHARMA NOEs.

Here we demonstrate for the first time that INPHARMA can be applied in the context of SBDD, where it allows determining the relative, and in favorable cases even the absolute, binding mode of two low affinity ligands binding competitively to a common receptor site. In accordance with existing SBDD workflows, we employ the experimental information derived from the INPHARMA NOEs and measurable in a series of simple NOESY experiments, to select the correct binding mode among many possible binding orientations models [87].

As a test system in this study we selected the core hinge binding fragments of two known ATP competitive kinase inhibitors with affinities of 6 $\mu$M ($L_A$) and 16 $\mu$M ($L_B$; Fig 2.2) [96]. We performed the study with the catalytic subunit of the protein kinase A (PKA), an ubiquitous enzyme of known crystal structure [9, 43, 45], which regulates a large variety of cellular processes including ion flux, cell death and gene transcription.

## 2.2 Results

### 2.2.1 NMR experiments

NOESY spectra were acquired for the mixture of the catalytic subunit of the protein PKA (from now on referred to as PKA) and the two ligands $L_A$ (450$\mu$M) and $L_B$ (150 $\mu$M) at a 800 MHz spectrometer for mixing times $\tau_m$ = 150, 300, 450, 600 and 700 ms. The protein concentration was either 25 or 30 $\mu$M, as explained in the Methods section. For $\tau_m \geq$ 300 ms interligand INPHARMA NOEs were observed. These NOEs do not originate from aggregation of the ligands, as they are absent in NOESY experiments performed for the same ligands mixture in absence of PKA or in the presence of perdeuterated PKA.

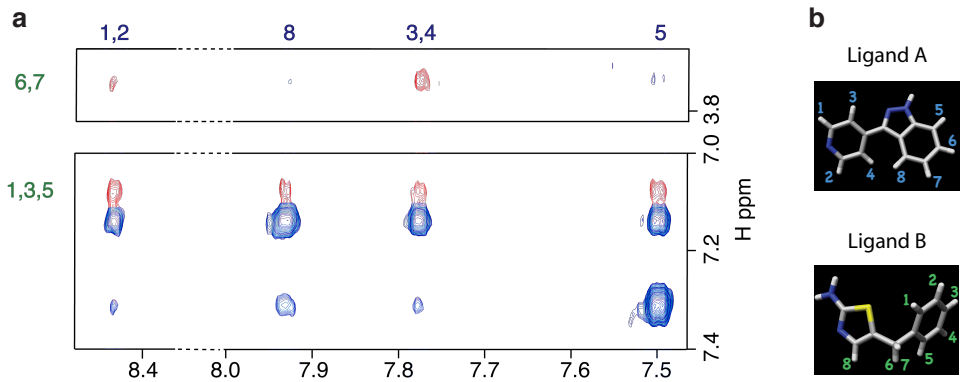



**Figure 2.2:** Experimental INPHARMA NOEs: (a) Region of the NOESY spectrum recorded for a mixture of the protein PKA (30 $\mu$M), $L_A$ (150 $\mu$M) and $L_B$ (450 $\mu$M) at 800 MHz for a $\tau_m$ = 600 ms. Intra-molecular transfer NOEs of $L_A$ are depicted in blue. Intermolecular "INPHARMA" NOEs between $L_A$ ($\omega_2$) and $L_B$ ($\omega_1$) are shown in red and are approximately two orders of magnitude less intense than intra-molecular transferred NOEs. (b) Chemical structures of the PKA inhibitors used in this study $L_A$ and $L_B$.

### 2.2.2 Docking models and calculation of theoretical INPHARMA NOEs

The scope of this work is to demonstrate that the INPHARMA NOEs allow selecting the correct orientation of a novel ligand of unkown crystal structure relative to a ligand of known crystal structure. As a basis for this study we

therefore determined crystal structures for the PKA/$L_A$ and PKA/$L_B$ complexes at 2.1 Å and 2.0 Å resolution, using described purification and crystallization protocols and standard crystallographic procedures (Fig. 2.3) [32]. The effectiveness of the INPHARMA method is demonstrated if the comparison of the experimental interligand NOEs with the theoretical interligand NOEs, expected for each pair of docking modes of the two ligands to PKA, allows the selection of docking modes that are similar to the crystal structures, either in the relative or in the absolute orientation of $L_A$ and $L_B$ in the PKA binding pocket.

For this analysis we generated "wrong" ligand orientations by rotation of $L_A$ and $L_B$ within the protein binding pocket around either the z or the y axis (Fig. 2.3), followed by energy minimization of each obtained complex structure (see Methods). The geometry of the binding site allowed for four different main orientations for $L_A$ and $L_B$ which were combined to give a total of 16 pairs of structures (Fig. 2.3), one of which reproduces the "correct" absolute binding mode of both ligands and is very similar to the crystal structures. In three pairs of structures the relative orientation of the two ligands is correct, while the absolute orientation of both ligands is wrong. In the remaining 12 pairs both the relative orientation of the two ligands and the absolute binding mode of at least one ligand are incorrect.

### 2.2.3 Calculation of theoretical INPHARMA NOEs

To investigate the effectiveness of the INPHARMA method we calculated interligand NOEs for each of the 16 pairs of ligand orientations. We then monitored whether the correlation of the calculated NOEs with the experimental NOEs allows for the selection of a preferred relative or absolute ligand orientation and whether this result reproduces the crystal structures. In the following sections, the "relative binding orientation" describes the position of the functional groups of the two ligands relative to one another, irrespectively of their orientation within the receptor. The "absolute binding orientation" indicates the position of the functional group of each ligand relative to the receptor binding pocket and thereby uniquely defines the intermolecular interactions in the complex.

The INPHARMA NOEs were calculated for the 16 pairs of structure using the full-relaxation matrix approach [8, 63, 108], as described in the Supplementary Methods, for a static complex structure. Subsequently, the theoret-

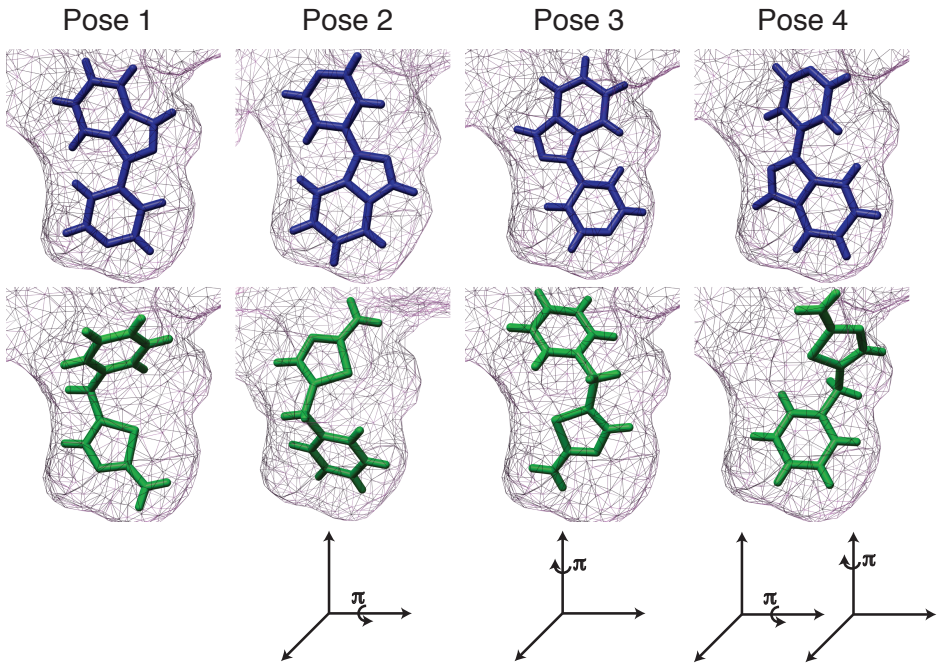



**Figure 2.3:** Binding orientations for ligands $L_A$ and $L_B$. (a) Starting the with the binding pose of the high resolution crystal structure (pose 1) other possible binding orientations were generated by (i) rotating the ligands by 180° around the y, the z axis or both, followed by energy minimization of the resulting pose (pose 2-4). Shown are the possible binding orientations for $L_A$ (blue model) and $L_B$ (green model) with the protein pocket shown as exclusion surface. Combination of the possible binding orientations of $L_A$ and $L_B$ results in 16 ligand pairs.

ical INPHARMA NOEs calculated for each models' pair are compared with the experimentally derived ones to select the best fitting models' pair.

Internal motions of the ligand in the binding pocket or of the protein side-chains are ignored in the calculation, due to the lack of a suitable theoretical model. The influence of internal motion on the measured inter-ligand NOEs results in a slope of the linear regression between experimental and calculated INPHARMA NOEs different from 1 at shorter mixing times, while at very long mixing times the presence of internal motions of variable amplitude at different sites in the complex are expected to deteriorate the overall quality of the fit.

### 2.2.4 Selection of docking modes

The selection of the correct binding mode of the two ligands is based on three criteria:

(i) linear correlation coefficient $R^2$ or quality factor $Q$. At this step all models' pairs showing a poor correlation between experimental and theoretical INPHARMA NOEs are discarded.

(ii) Systematic deviation of INPHARMA NOEs stemming from different structural moieties of the ligands. At this step models' pairs are excluded for which 180° rotations of one or both ligands around one (or more) pseudo-symmetry axis of the correct binding mode lead to binding poses with an acceptable overall correlation coefficient between experimental and theoretical INPHARMA NOEs but showing systematic deviations (under- or over-estimation) of groups of INPHARMA NOEs stemming from separate structural parts of the ligand(s).

(iii) Semi-quantitative use of weak INPHARMA NOEs, which are observable either in longer experiments at higher magnetic field, or at higher protein concentration and/or at longer mixing times. At this step, weak, additional INPHARMA NOEs are used as a discrimination criterion. However, the quantitative fitting of weak INPHARMA signals is often deteriorated by the effect of internal motions in multiple steps of spin diffusion; thus these NOEs are interpreted in a semi-quantitative manner.

Correlation of the calculated and measured INPHARMA NOEs for all 16 pairs of possible binding orientations are given in Table A.1. We found that the correlation coefficient for 4 out of 16 pairs differentiates clearly from the average correlation and therefore selected these 4 models' pairs with $R^2 > 0.9$ or $Q^2 < 0.15$ to pass the first selection criterion (i) (Fig. 2.4). In all models' pairs A-D the orientation of $L_B$ is uniquely defined (Fig. 2.1), while the orientation of $L_A$ is largely undefined. This finding likely results from the different three-dimensional shape of the two ligands. $L_B$ has a rather bent, asymmetric shape; thus, changes of the orientations of $L_B$ in the protein binding pocket (upon rotations around the y and/or z axis) result in dramatic changes in the environment of each proton, namely in the distribution of distances to the protons of the receptor (Supplementary 2.1). Being the INPHARMA NOEs mediated by the protons of the receptor, the distribution of distances between the ligand's and the receptor's protons determines the efficiency of the magnetization transfer. On the other hand, the more symmetric, flat

shape of $L_A$ allows for rotation of the ligand in the protein binding pocket with less dramatic changes in the intermolecular proton-proton distance distribution (Supplementary Fig. 2.1), which is reflected in the presence of all orientations of $L_A$ in the models' pairs with good correlation between the experimental and the theoretical INPHARMA NOEs.

To further differentiate amongst the remaining ligand orientations we analyzed the fit of the calculated and experimental INPHARMA NOEs of individual ring systems in the correlation graphs for the four models' pairs A-D (selection criterion (ii)). As seen in Fig. 2.4, pairs A and B perform significantly better than pairs C and D. For these last two pairs the INPHARMA NOEs stemming from the phenyl ring of $L_B$ to the pyridine ring of $L_A$ (color coded in yellow, Supplementary Fig. 2.2) and to the (4-Pyridyl)indazole ring of $L_A$ (color coded in red, Supplementary Fig. 2.2) are consistently over- and under-estimated, respectively, indicating that in model C/D the orientation of $L_A$ is wrong by a 180° rotation around the y axis. Thus, models pairs C and D can be safely excluded.

The degeneracy of the H1,2 and H3,4 protons of $L_A$ and the H1-5 protons of $L_B$ impedes the discriminations between the remaining pairs A and B, in which $L_A$ is rotated around the z axis, by both criterion (i) and (ii). According to selection criterion (iii), we recorded an additional NOESY spectrum at a 900 MHz spectrometer, with double number of scans and mixing time of 600 ms, with the aim of collecting more experimental data from the non-degenerated protons. This spectrum showed indeed additional interligand NOEs between protons H6 and H7 of $L_B$ and proton H5, H6 and H8 of $L_A$. The ratio of 1.4 between the H8($L_A$)-H6,7($L_B$) NOE and the H5($L_A$)-H6,7($L_B$) NOE is well reproduced in the back-calculated interligand NOEs for models' pair A (1.8), while is dramatically underestimated in models' pair B (0.06), thus leading to exclusion of the latter pair.

Application of the selection criteria (i)-(iii) allowed us to identify models' pair A as the one uniquely representing the experimental INPHARMA data. A comparison of the resulting orientation with the crystal structures of the PKA/$L_A$ and PKA/$L_B$ complexes furthermore shows that this model represents the correct relative orientation of the two ligands and the correct absolute orientation of both ligands with respect to the protein. Although the main scope of the methodology is to derive the relative binding mode of two low affinity fragment inhibitors, in the test case shown here the INPHARMA

approach exceeds the expectations and allows the determination of the absolute binding mode of both ligands. In general, we expect this to occur for those cases where the structure of the apo-protein closely resembles that of the protein in the complex.

Further docking models, described in Supplementary Figure A.4 & A.8 and providing a more complete statistics around some of the orientations of Fig. 2.3 were tested as well, leading to the selection of the same "crystal-structure-like" relative orientation of the two ligands in the PKA binding pocket as in models' pair A.

Finally, we verified that the INPHARMA NOEs calculated for the crystal structures of the PKA/$L_A$ and PKA/$L_B$ complexes reproduce the experimental INPHARMA NOEs in a comparable manner as for models' pair A (Supplementary Fig. A.4). The fitting of the calculated vs. experimental INPHARMA NOEs is excellent. The residual deviations can be accounted for invoking the presence of internal motions [46, 76], as proven by the slope of the linear regression of 0.3 [53, 91]. The good quality of the fit further verifies that the INPHARMA NOEs are suitable experimental data for the ranking of the docking poses.

### 2.2.5 Influence of the accuracy of the protein structure

In the case of PKA, the structure of the protein binding pocket in the various PKA/ligand complexes is largely unchanged and the PKA conformation used to calculate the different docking models is quite close to the "true" structure of the protein in the complexes (r.m.s.d of the binding pocket heavy atoms 0.3 Å). However, this might not be the case for other systems, where the binding pocket might be dynamic or ill-defined in the available structural models. In order to assess the influence of the quality of the protein structure on the outcome of the methodology, we used Molecular Dynamics simulations to generate a series of PKA structures with inaccurate conformations of the binding pocket. The models exhibit heavy atoms r.m.s deviations of 1.9, 1.7, 1.9, 1.6 and 1.9 Å in respect to the X-ray structure. Ligands $L_A$ and $L_B$ were docked in the four different orientations described above and sets of 16 model pairs were generated for each of the five PKA structures. (Supplementary Fig. A.5). The calculated INPHARMA NOEs for each of the 16x5 model pairs were correlated with the experimental interligand NOEs as described above (Supplementary Table A.1).

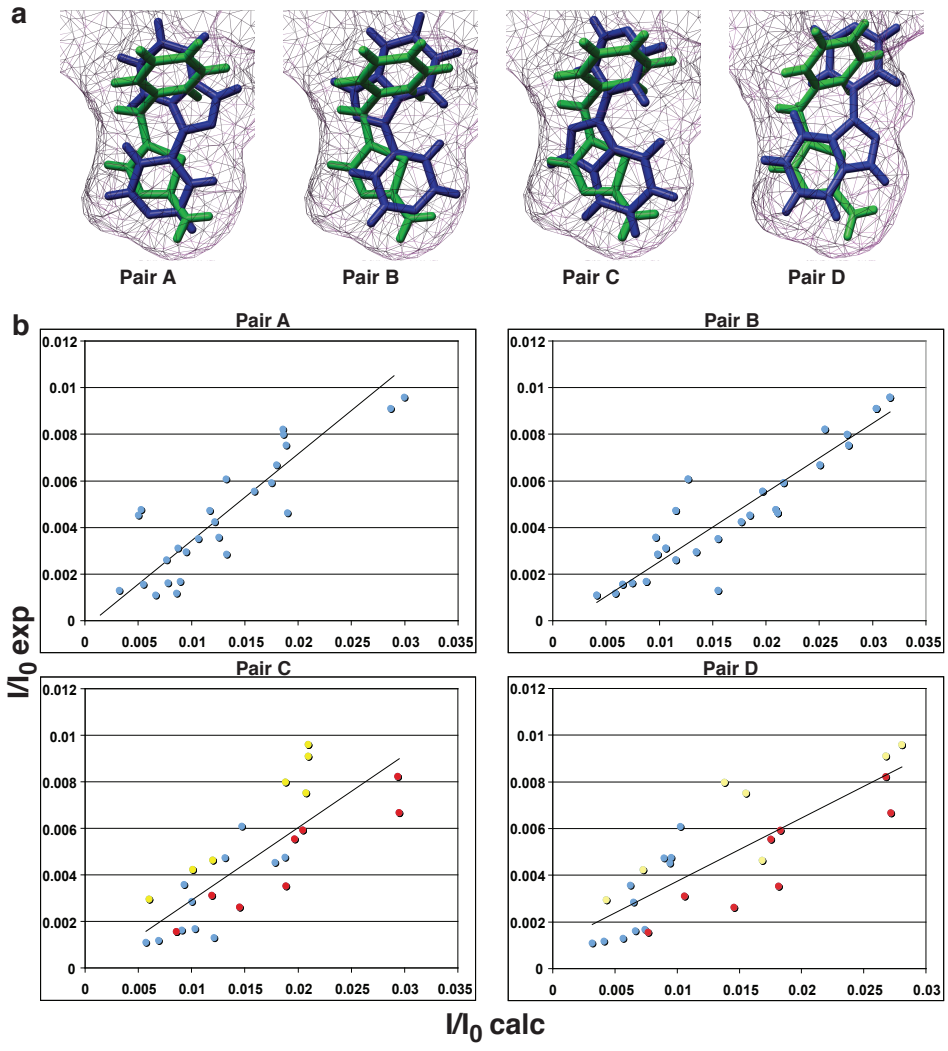



**Figure 2.4:** (a) Representation of the four models' pairs (A-D) that show a good correlation between the theoretical and the experimental INPHARMA NOEs (b) Correlations of the experimental INPHARMA NOEs, normalized with respect to the diagonal peak of the spin in $\omega_1$, with the back-calculated INPHARMA NOEs, for each models' pair of panel (a). The INPHARMA NOEs between the phenyl ring of $L_B$ to either the pyridine or the 3-(4-Pyridyl)indazole ring of $L_A$ of model C and D, color-coded in yellow and red, respectively (Supplementary Fig. A.2), are consistently under or over-estimated for the models' pairs C and D.

We found that the correct absolute orientation of the two ligands was selected by criteria (i)-(iii) for four of the five inaccurate protein models, whereas pose 4 (or 2) (Fig. 2.3b) was selected for both ligands in the fifth PKA model. Still, even for this model the correct relative orientation of $L_A$ and $L_B$ is predicted. This test showed that prediction of the relative binding mode of the two ligands is robust towards inaccuracy of the binding pocket structure, while the correct prediction of the absolute orientation of the ligands depends on the accuracy of the protein structural model used during docking.

## 2.3 Discussion

We have demonstrated for the first time that INPHARMA NOEs can be utilized to determine the relative orientation of low molecular weight, low affinity ligands which bind competitively to the same binding pocket of a common receptor. INPHARMA is based on the measurement of a series of NOESY spectra and is generally applicable to all receptor/ligand complexes with mid to low affinity. The bound ligand structure, easily obtainable from transferred NOE data [16, 73], and a structural model of the apo-protein, which are used to generate plausible docking modes, are the only requirements for the application of the method.

The INPHARMA method allows closing a gap in structure base drug discovery, where commonly crystal structures of some but not all interesting chemical lead series are accessible. Here INPHARMA provides an additional experimental approach to determine the complex structure of novel interesting lead series relative to a known protein ligand crystal structure.

Our data furthermore indicate that assessment of relative binding orientations is relatively stable against inaccuracies of the applied protein model, so that INPHARMA NOEs may allow aligning chemically different ligands as a basis for pharmacophore modeling in the absence of a protein crystal structure. In the specific case presented, it was possible to select the absolute orientation of the ligand pair in respect to the protein structure, solely based on interligand NOE data. Interestingly, this orientational signal in the first selection step was dominated by $L_B$, whereas $L_A$ exhibited no preferential orientations (Table A.1, Fig 2.3). The "orientational signal" of $L_B$ results from its bent, asymmetric shape.

The possibility of determining absolute protein-ligand orientations in the absence of a protein crystal structure suggest novel approaches for structure based design, where INPHARMA data of an appropriate selection of differently shaped low molecular weight ligands may provide a surprisingly detailed ligand based map of a receptor binding site.

## 2.4 Material and Methods

Protein expression and purification for the NMR experiments.The catalytic subunit of chinese hamster $C\alpha$ catalytic subunit of cAMP-dependent of protein kinase A (PKA) was expressed and purified according to the published procedure [51]. In specific, the expression plasmid pETPKA, containing the coding sequence of the cAMP dependent protein kinase, was used for transformation of E. coli strain BL21 (DE3) competent cells. The protein, containing a Histidine-tag, was purified on a Ni-NTA Fast-Flow column (Quiagen, Hilden Germany) following the manufactures recommendations. Cleavage of the His-tag was performed adding 80 $\mu$L of Tev-protease (1 mg mL-1) at room temperature overnight. Further purification was achieved by anion-exchange chromatography with a SourceQ column. Finally, samples (10-15 mL) were dialyzed against 1L NMR buffer (PBS-buffer, NaCl 150 mM).

### 2.4.1 NMR Experiments

Two sets of NOESY experiments were collected for a mixture of $L_A$ and $L_B$ and the protein PKA at 800 MHz and 700 MHz spectrometers using a cryogenetically cooled triple-resonance probe-head. The average measurement time was about 20 hours per NOESY spectrum. The first set of NOESY spectra was acquired at a 800 MHz spectrometer and comprised mixing times $\tau_m$ = 150, 300, 450, 600 and 750 ms. Data were recorded on two different samples: for $\tau_m$ = 300 and 600 ms the sample contained $[L_A]$ = 150 $\mu$M $[L_B]$ = 450 $\mu$M, [PKA] = 30 $\mu$M; for $\tau_m$ =150, 450 and 750 ms $[L_A]$ = 150 $\mu$M $[L_B]$ = 450 $\mu$M, [PKA] = 25 $\mu$M. The different protein concentration in the two samples was accounted for in the calculation of the theoretical INPHARMA NOEs. An additional NOESY spectrum was acquired at a 900 MHz spectrometer with $\tau_m$ = 600 ms. The sample contained: $[L_A]$ = 150 $\mu$M $[L_B]$ = 450 $\mu$M, [PKA] = 25 $\mu$M.

### 2.4.2 Crystal structures

Recombinant bovine C$\alpha$ catalytic subunit of cAMP-dependent protein kinase was crystallized according to Engh et al. [32]. Inhibitor complexes were obtained by soaking the crystals for 24-48 hours in the crystallization solution with 10 mM $L_A$ or $L_B$. The crystals were cryo-protected by the addition of 30% glycerol and frozen in liquid nitrogen. Diffraction data were collected at ESRF ID14-1. Datasets were processed, the structure solved and initially refined using APRV [49], XDS [41] and CNX [Accelrys, San Diego]. 7% of the reflections were set aside for cross-validation. Interactive model building and refinement was carried out using O [40], COOT [31], CNX and REFMAC5 [72], yielding a final R/Rfree of 21.5%/24.9% for the PKA:$L_A$ and 22.0%/26.3% for the PKA:$L_B$ complex (Supplementary Table A.2). Coordinates and structure factors are available at the protein data base (PDB) under entry codes 3DNE.pdb ($L_A$) and 3DND.pdb ($L_B$).

### 2.4.3 Docking models

Alternative binding poses were generated in order to evaluate whether INPHARMA NOEs can distinguish different ligand orientations in the receptor binding pocket. Starting with the crystal structure, the ligands were rotated by 180° either around the y or the z axis, or both (Fig. 2.3), followed by energy minimization with Xplor [92, 93], keeping the ligand rigid and using the repulsive part of the Van der Waals potential as driving force. Rotation of 90° around the y or z axis was prohibited by the shape of the binding pocket. Additional docking models were generated at Sanofi Aventis and represented smaller rotations of the ligands around the models of Fig. 2.3 (Supplementary Fig. A.3).

### 2.4.4 Calculation of the INPHARMA NOEs

The INPHARMA NOEs were calculated for all pairs of protein/ligand structures resulting from the obtained relative orientations, using the full relaxation matrix approach, as described in Supplementary Methods. All protons at a distance d $<$ 1 nm from any ligand proton were included in the calculations, whereas protons at a distance d $>$ 1 nm are not involved in the INPHARMA magnetization transfer process, as demonstrated previously [87]. The relative affinity of the two ligands was obtained by competition against

Y-27632 (ROCK Inhibitor), which delivered a $K_i$ of 6 $\mu$M for $L_A$ and 16 $\mu$M for $L_B$. Therefore a ratio of 1:3 was used for $k_{AB} : k_{BA}$ (Supplementary 4.4). The INPHARMA NOEs were calculated at each mixing time and divided by the diagonal peak with equal $\Omega_1$ frequency. This ensures that both the INPHARMA NOE and the reference diagonal peak originate from the same spin at the beginning of the NOESY experiment and therefore compensates for the incomplete longitudinal relaxation during the initial equilibration delay. Internal motions of both the protein and the ligands in the complex were neglected, as we lack a satisfactory model to describe them. The correlation time of the complex was optimized by fitting the experimental intraligand tr-NOEs to their theoretical values.

### 2.4.5 MD simulations

A Verlet Molecular Dynamics protocol at 298K for 15 ns in a continuum matter (dielectric constant = 78.4 Farad.m$^{-1}$) was used to generate PKA conformations other than the X-ray one. One structure was saved each 3 ns for a total of 5 different structures (Supplementary Fig. A.7).

# Chapter 3

# Specific methyl group protonation for the measurement of pharmacophore-specific inter-ligand NOEs

**Abstract**

INPHARMA NOEs can be measured between two ligands binding competitively to a common macromolecular receptor and can be interpreted quantitatively to derive the binding mode of the two ligands. This approach relies on computationally demanding full relaxation matrix calculations. Here we demonstrate that the INPHARMA NOEs measured in presence of a selectively-protonated receptor can be interpreted in a semi-quantitative way to discriminate between binding poses, thus relieving for the need of demanding computations.

## 3.1 Introduction

In the recent decades nuclear magnetic resonance (NMR) spectroscopy has emerged as a powerful tool in the drug-discovery field. Several NMR-based methods have proved beneficial for the optimization of low-molecular-weight lead structures. Among those, thechniques relying on the nuclear Overhauser effect (NOE) can provide information on pharmacophores at atomic resolution. The transferred-NOEs (tr-NOEs) [73] allow access to the bioactive conformation of ligands and can also be applied as a screening tool to identify the interaction of a small molecule with a macromolecular receptor. Recently, we have reported on the measurement of protein-mediated inter-ligand NOEs (INPHARMA), which are observed for a mixture of two ligands binding competitively and weakly to the same binding pocket of a common target [87, 89]. These inter-ligand NOEs do not result from a direct magnetization transfer between the protons of the two ligands, since the small compounds, being competitive binders, never occupy the protein binding pocket at the same time. Rather, a spin diffusion process, mediated by the protons of the receptor-binding pocket, leads to cross-peaks between the two ligands. We proposed and demonstrated that these inter-ligand NOEs can be used to determine the relative binding mode of two drug leads or, in favorable cases, even their absolute binding pose (INPHARMA approach) [87].

## 3.2 Principle

The principle underlying the methodology relies on the dependence of the INPHARMA NOEs on the protein environment and more specifically on the distances between the protons of each ligand and the protons of the receptor. Consequently, the size of the INPHARMA NOEs depends on the binding mode of each ligand to the receptor and a quantitative interpretation of such effects can be used to derive the binding modes. However, extensive spin diffusion among the receptor protons reduces the specificity of the INPHARMA NOE signals, and the pharmacophore signature has to be retrieved by theoretically simulating the effect of spin diffusion. Thus, the determination of the ligands' binding modes relies on computationally intensive calculations based on the full-relaxation matrix approach in presence of chemical exchange [62], including the protons of the ligands and all protons of the receptor within a distance of 8 Å from the binding pocket [73, 87].

In the daily work-flow of drug-discovery, it is desirable to extract information from the INPHARMA NOEs without need for the demanding full-relaxation matrix calculations. We reasoned that a clearer fingerprint of the binding modes of the two ligands to the receptor would be obtained from the values of the INPHARMA NOEs if spin-diffusion inside the protein could be either switched off or considerably reduced. An efficient way to attenuate intra-molecular spin-diffusion is to reduce the proton density in the receptor, for example by deuteration. In the past few years, a robust protocol has been developed for bacterial expression of proteins, in which methyl groups are selectively protonated in a highly deuterated background [99]. Here we employ this method to investigate the effects on protein-mediated inter-ligand NOEs of depleting the background protein proton density while maintaining protonated side-chains of specific amino-acids types [99]. Furthermore, we explore the attractive possibility of employing the INPHARMA method in combination with selectively protonated receptors to extract structural information on the relative binding mode of two competitive ligands without need for the time-consuming full-relaxation matrix calculations.

## 3.3 Methods

The system under investigation consists of the catalytic subunit of cAMP-dependent protein kinase A – a monomeric protein of 353 amino acids. The protein kinase A (PKA) is a ubiquitous enzyme of known crystal structure [9, 43, 45], which regulates a large variety of cellular processes including ion flux, cell death and gene transcription. The selected ligands were the core hinge binding fragments of two known ATP competitive kinase inhibitors $L_B$ and $L_A$, whose binding site has been previously identified (Fig. 3.1).

The catalytic unit of PKA (hereafter referred to as PKA) was expressed in three different forms: (i) fully protonated (FP), (ii) protonated at specific amino acid side chains (SP), or (iii) perdeuterated (PD). For the preparation of the specifically protonated protein, certain $\alpha$-ketoacids can serve as precursors for a number of methyl-bearing amino acids for proteins over-expressed in minimal media [99], In our case, among amino acids containing the aliphatic side-chain, only valine and leucine are part of the binding pocket (Fig. 3.3); therefore we chose $\alpha$-ketoisovaleric acid as the precursor molecule for the production of deuterated proteins with protonation restricted to the Leu$\delta$/Val$\gamma$

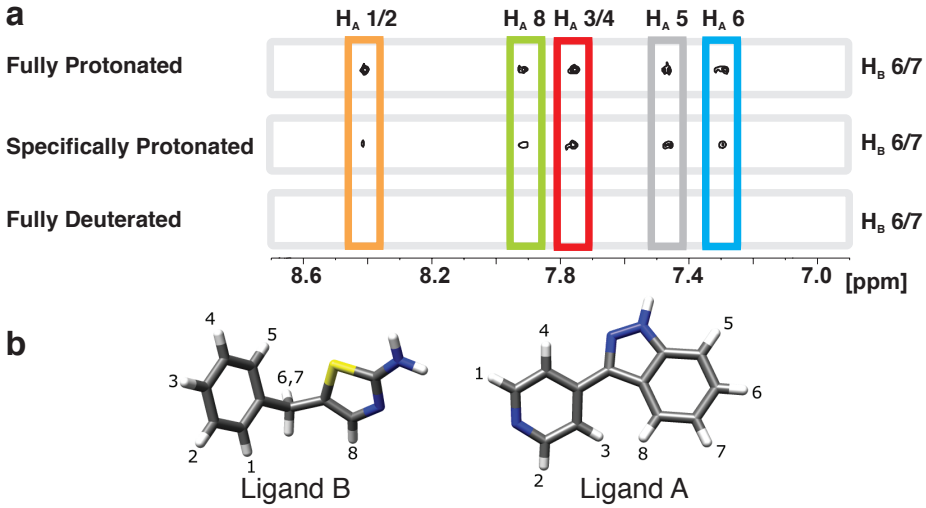



**Figure 3.1:** NOESY spectra for the measurement of inter-ligand NOEs. (a) Slices from 2D NOESY spectra at 3.72 ppm with a mixing time of 600 ms. Protein samples were expressed in a fully protonated form (upper slice); with the methyl group side chains of Leu and Val specifically protonated (middle slice); or fully deuterated (lower slice). (b) Chemical structures of the ligands $L_B$ and $L_A$ with proton numbering.

positions. Typically, precursors with the desired labeling patterns were added to $D_2O$-based growth medium approximately 1 hour prior to induction of protein over-expression, with expression times kept reasonably short (3-4h in our case) to maximize the incorporation levels [35]. Three samples were prepared containing 450 $\mu$M of $L_B$ and 150 $\mu$M of $L_A$. The 1:3 ratio in the concentration of the two ligands has been chosen to partially compensate for the different affinity of $L_B$ and $L_A$ for PKA ($K_{i,L_B}/K_{i,L_A} \approx 0.3$). The protein concentration was 25 $\mu$M for samples (ii, SP) and (iii, PD) (containing selectively protonated and perdeuterated protein, respectively), and 45 $\mu$M for sample (i, FP) (containing fully protonated protein). NOESY spectra were acquired on a 900 MHz spectrometer with a mixing time $\tau_m = 600$ ms (Fig. 3.1 a)

The protein concentrations were determined by fitting the intensity of intra-ligand tr-NOEs. The absence of NOEs in the sample containing the perdeuterated protein confirms that the inter-ligand NOEs do not originate

from aggregation of the ligands or from simultaneous binding to the protein, but are indeed mediated by the protein protons.

## 3.4 Results

Inter-ligand NOEs are observed for samples containing either fully or specifically protonated protein (Fig. 3.1), thus confirming that selective protonation of the Leu$\delta$/Val$\gamma$ positions in a perdeuterated background does not hamper the measurement of protein mediated inter-ligand NOEs. Figure 3.3 shows the binding pocket of PKA, in which $L_B$ and $L_A$ occupy alternately the same space. The short distances (d $<$ 5 Å) of the methyl groups of Leu49, Val57, Val104, Val123 and Leu173 from both ligands ensure an efficient transfer of magnetization between the two ligands in the Leu$\delta$/Val$\gamma$ selectively protonated sample. A comparison of the theoretical values of the INPHARMA NOEs expected for a sample containing $L_B$ (450 $\mu$M), $L_A$ (150 $\mu$M) and either fully or selectively protonated PKA (25 $\mu$M) shows that a much lower intensity is expected for the inter-ligand NOEs in the presence of selectively protonated protein, which is in agreement with the paucity of receptor protons available for the transfer (Fig. 3.3). The data set in the presence of the selectively protonated protein has been calculated assuming a typical protonation efficiency of 80%, which, for the five Val and Leu amino acids present in the binding pocket, results in 16 species with population p $>$ 0.5% (1 species with p = 32%, 10 species with p = 8.2% and 5 species with p =2%). All species with p $<$ 0.5% were neglected in the calculations. The two theoretical data sets correlate reasonably well, with the exception of two large outliers: the inter-ligand NOE between H8 of $L_B$ and H1,2 of $L_A$ and the NOE between H1,3,5 of $L_B$ and H5 of $L_A$ (Fig. 3.3). The ratio of the inter-ligand NOEs between H8-$L_B$ and H1,2-$L_A$ for the SP vs the FP sample is lower than average; this reflects the depletion of the protons of Phe327, which are the closest to both proton H8-$L_B$ and protons H1,2-$L_A$ and therefore are most responsible for the corresponding protein-mediated exchange of magnetization between them (Fig. B.1). On the other hand, the same ratio for the NOE between H1,3,5-$L_B$ and H5-$L_A$ is higher than average, reflecting the elimination of additional spin-diffusion pathways inside the protein. The principal mediators in the transfer of magnetization between the H1,3,5-$L_B$ protons and the H5-$L_A$ proton are the methyl groups of Val57, which are not depleted in the

selectively protonated protein. In general, it is expected that the ratio of the inter-ligand NOE peaks observed for the SP sample vs those observed for the FP sample is high for the ligand protons close to Leu$\delta$/Val$\gamma$ methyl groups, due to the absence of dissipating spin-diffusion pathways, while it is lower for the ligand protons that are distant from the protonated methyl groups. This should result in a specific fingerprint of the binding mode of the ligands.

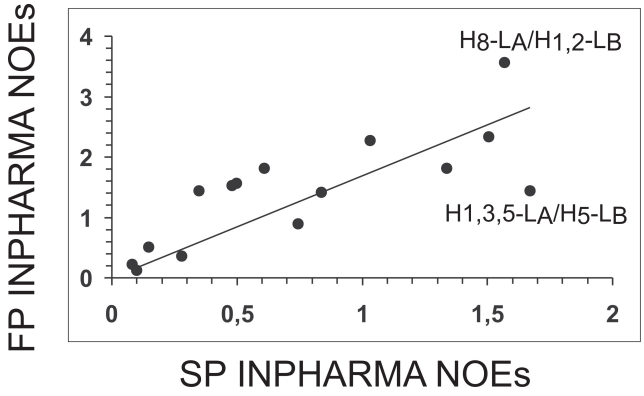



**Figure 3.2:** Correlation of the theoretical INPHARMA NOEs expected for a sample containing fully protonated PKA (FP) versus those expected using selective protonation (SP) of the Leu$\delta$/Val$\gamma$ positions. The NOEs are not normalized and are in arbitrary units.

### 3.4.1 Theoretical formulation

The higher specificity of the pharmacophore signature observable for the SP sample with respect to that of the FP sample prompted us to investigate the possibility of interpreting the INPHARMA NOEs in a semi-quantitative way, namely without considering the effect of spin diffusion inside the protein. For this purpose we defined the parameter **D**, as an indicator of the distances of each ligand proton to the PKA protons:

$$\mathbf{D} = \sum_{Val,Leu} \left( \sum_{i=1,2} d(H_A, Met_i)^{-6} \cdot \sum_{i=1,2} d(H_B, Met_i)^{-6} \right) \qquad (3.1)$$

where d($H_{A/B}, H_{Met}$) is the distance between proton $H_{A/B}$ of $L_{A/B}$ and the carbon of one of the two methyl groups of a particular valine or leucine amino acid. Whenever two or more ligand protons have a degenerate chemical

shift the sum extends over all degenerate protons and the resulting indicator $\mathbf{D}_{AB}$ is divided by the number of degenerate protons. Only distances d $< 5$ Å are considered in the calculations. Our goal is to demonstrate that the inter ligand-NOEs values correlate well with the indicator $\mathbf{D}_{AB}$ for each $H_A - H_B$ NOE and that the quality of the fitting can be used to distinguish between different binding modes of the two ligands.

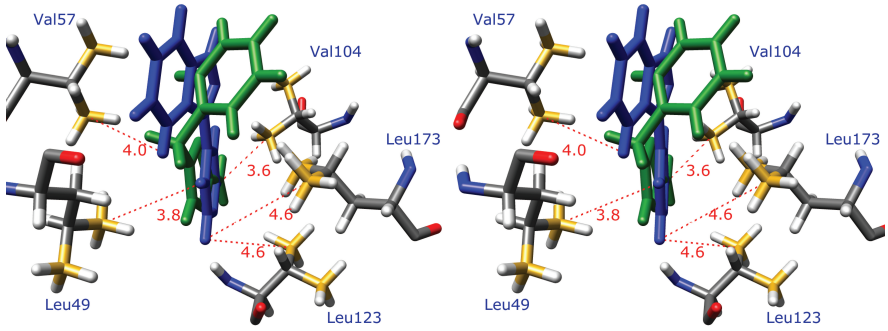



**Figure 3.3:** Overlap of the crystal structures of the PKA/$L_B$ and PKA/$L_A$ complexes (stereo view). The Val and Leu amino acids are represented with sticks and the. $L_B$ is in green and $L_A$ is in blue.

In Figure 3.4 we plot the experimental INPHARMA NOEs between $L_B$ and $L_A$ versus the indicator $D$ for five pairs of PKA/$L_B$ and PKA/$L_A$ complexes. The inter-ligand NOEs were measured for the mixture of $L_B$, $L_A$ and the selectively protonated PKA (SP) and were normalized with respect to the diagonal peak in $\omega_1$. The pair of docking modes in panel (a) corresponds to the crystal structures of both complexes and represents the "correct" docking poses. The pairs in panels (b-d) contain one wrong binding mode each and have been generated by rotating $L_B$ or $L_A$ by 180° around the z or y axis, as indicated in the figure. The difference in the quality of the correlation is striking, with the correct pair (a) exhibiting R = 0.83 for the linear correlation of the INPHARMA NOEs with the crude distance indicator $\mathbf{D}$, while the incorrect pairs show very poor correlations (R $< 0.5$). From this result we conclude that a semi-quantitative interpretation of the INPHARMA NOEs acquired in presence of selectively protonated protein might be sufficient to discriminate between docking modes. Such interpretation could replace the lengthy full-relaxation matrix interpretation necessary when the INPHARMA NOEs are acquired in the presence of fully protonated protein.

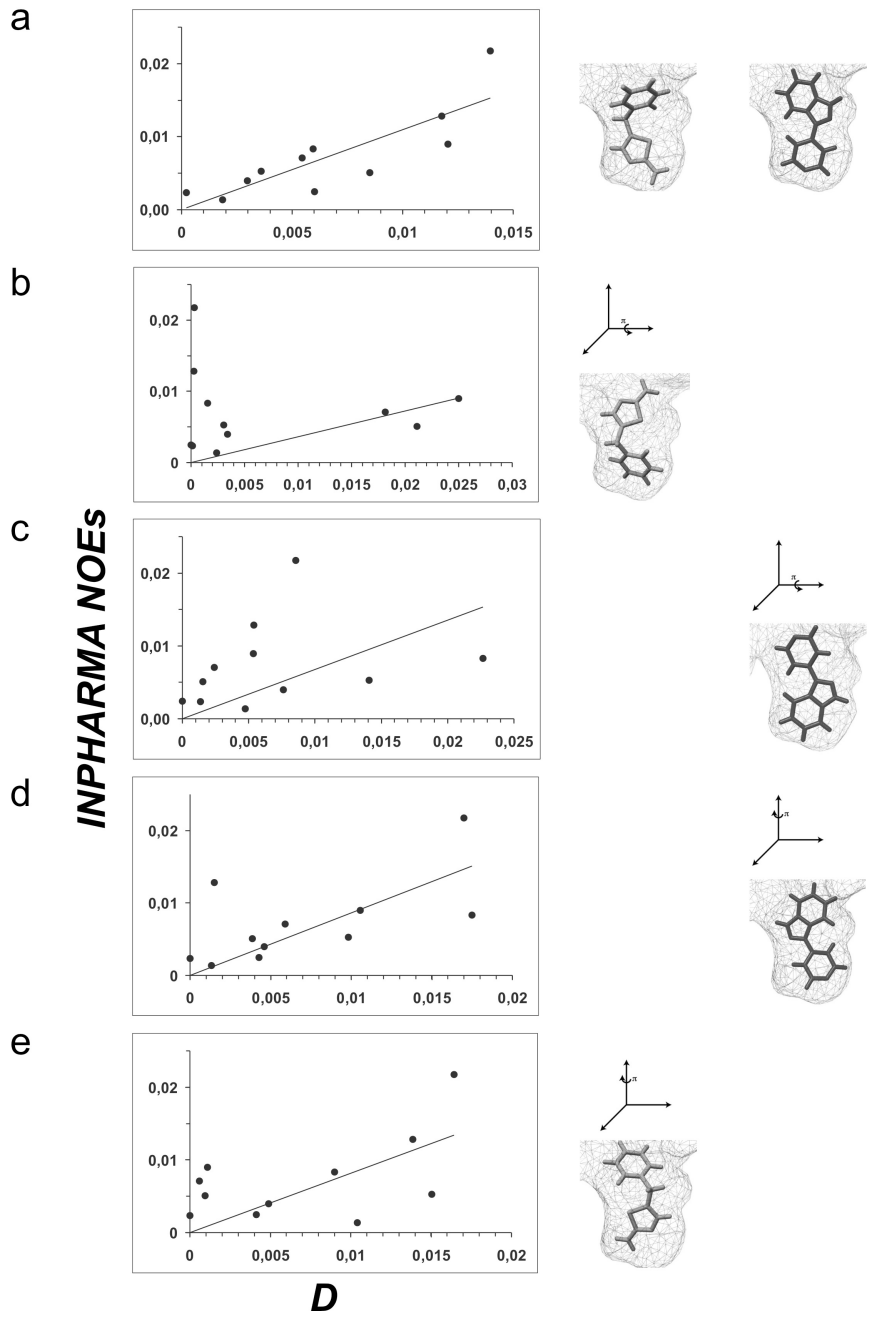



**Figure 3.4:** Correlation of the INPHARMA NOEs measured for a mixture of $L_B$, $L_A$ and selectively protonated PKA (Leu$\delta$/Val$\gamma$) with the distance indicator **D** calculated for five model-pairs of the PKA/$L_B$ and PKA/$L_A$ complexes. The first model-pair (panel (a)) corresponds to the crystal structures of the PKA/$L_B$ and PKA/$L_A$ complexes. Panels (b-d) present the correlations for model-pairs where one of the ligand has been rotated to an incorrect orientation in the binding pocket: (b) $L_B$ has been rotated by 180° around the y axis; (c) $L_A$ has been rotated by 180° around the y axis; (d) $L_A$ has been rotated by 180° around the z axis; (d) $L_B$ has been rotated by 180° around the z axis.

In addition, the specificity of the pharmacophore signature could be further improved by enlarging the SP INPHARMA NOE data-set. This can be achieved by means of multiple complementary schemes of selective protonation, targeting for example either the methyl groups or the aromatic side-chains [51].

The possibility of discriminating between binding poses by correlation of the INPHARMA NOEs with the distance indicator **D**, or a similar function, opens the way to easy implementation of the INPHARMA NOEs in structure calculation programs. While simulation of spin diffusion at each step of structure calculation and comparison of theoretical and experimental NOEs is a computationally demanding task, due to the large size of the matrices involved, the correlation between the INPHARMA NOEs and a distance indicator could be easily translated into an energy term. In this way, the INPHARMA NOEs could be used to actively drive the docking protocol towards the correct binding poses of the two ligands. Our laboratory is actively exploring this intriguing and promising approach to the determination of receptor-ligand structures.

## 3.5 Experimental Section

Protein expression and purification for the NMR experiments. The catalytic subunit of chinese hamster C$\alpha$ catalytic subunit of cAMP-dependent of protein kinase A (PKA) was expressed and purified according to the published procedure [51]. In detail, the expression plasmid pETPKA, containing the coding sequence of the cAMP dependent protein kinase, was used for transformation of E. coli strain BL21 (DE3) competent cells. To produce the deuterated and specifically protonated protein samples, cells were initially grown on normal M9 media and then transferred to 10 mL expression cultures with increasing percentages of $D_2O$. The main culture in 100% $D_2O$ was then inoculated with the adapted cells. Cells for the fully protonated protein sample were grown on $L_A$ medium. For the specifically protonated sample, the precursor with the desired labelling pattern (2-keto-3-methyl-butyrate) was added to the growth medium approximately one hour prior to induction. The protein, containing a Histidine-tag, was purified on a Ni-NTA Fast-Flow column (Qiagen, Hilden Germany) following the manufacturer's recommendations. Cleavage of the His-tag was performed by adding 80 $\mu$L

of Tev-protease (1 mg.mL$^{-1}$) and incubating overnight at room temperature. Further purification was achieved by anion-exchange chromatography with a SourceQ column. Finally, samples (10-15 mL) were dialyzed against 1L NMR buffer (PBS-buffer, NaCl 150 mM).

# Chapter 4

# The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method

**Abstract**

During the process of drug discovery, INPHARMA can be used to derive the structure of receptor/lead compound complexes binding to each other with a $K_d$ in the $\mu$M to mM range. To be successful, the methodology needs adjustment of various parameters that depend on the physical constants of the binding event and on the receptor size. Here we present a thorough theoretical analysis of the INPHARMA interligand NOE effect in dependence of experimental parameters and physical constants. This analysis helps the experimentalist to choose the correct experimental parameters and consequentially to achieve optimal performance of the methodology.

## 4.1 Introduction

In structure-based drug design a three-dimensional picture of the binding mode of known ligands to the target macromolecule is essential to the elaboration of a high affinity drug. When the complex cannot be crystallized, NMR can be used to obtain structural information in solution. However, the determination of a high resolution structure of the complex by NMR in solution is often limited by the availability and the physical properties of the target, which in many cases is either too large to be observed by NMR, too insoluble or not available with the necessary $^{13}$C /$^{15}$N /2D labeling from expression systems.

For ligands that bind weakly to the target ($K_d$ in the $\mu$M range) detailed structural information on the ligand bound conformation can be obtained by transferred-NOEs and transferred-CCR rates [12–14, 16, 59, 75]. This approach requires small quantities of unlabelled target (1-10 $\mu$M solution) and an excess of ligand, is applicable to any complex independently of the size of the target macromolecule and does not require any isotope labeling scheme. However, as the resonances of the target macromolecule are not observed, the methodology does not provide any structural information on the geometry of the intermolecular interactions in the complex. Models of the ligand binding mode can be obtained from the apo structure of the target macromolecule and the bound structure of the ligand by docking calculations. The docking models are ranked on the basis of the computed intermolecular interaction energy or with respect to experimental information, such as the ligand binding epitope, obtained by the STD or WATER-LOGSY approaches [2, 20, 68, 69]. However, this process mostly results in multiple models for the complex structure due to the lack of site specific structural information.

To overcome this problem, we developed the INPHARMA methodology (Interligand Noes for PHARmacophore Mapping) that allows mapping the structure of the binding pocket of a macromolecule on the NMR resonances of two competitively binding ligands [78, 79, 87, 89]. The method is based on the observation of interligand, spin diffusion mediated, transferred-NOE data, between two ligands $L_1$ and $L_2$, binding competitively and weakly to a macromolecular receptor T (Fig. 4.1). During the mixing time of the NOESY experiment, $L_1$ binds to the receptor and its protons (H$L_1$s) transfer their magnetization to the receptor protons (HTs). During the same mixing time of the NOESY experiment $L_1$ dissociates from the receptor and $L_2$ binds. The

magnetization that was transferred from H$L_1$ to HT can now be transferred from HT to H$L_2$. This leads to an intermolecular peak between H$L_1$ and H$L_2$ although $L_1$ and $L_2$ have never been close in space at any time during the NMR experiments. The NOE peak between H$L_1$ and H$L_2$ is a spin-diffusion mediated effect via the receptor proton HT. Clearly this effect can only occur if H$L_1$ and H$L_2$ are both close to the proton HT in the two complexes T$L_1$ and T$L_2$, respectively. A number of such interligand NOEs define the relative orientation of the two ligands in the receptor binding pocket.

The INPHARMA methodology belongs to the class of ligand-detected approaches (where the resonances of the target macromolecule are not seen in the spectrum), like transferred-NOEs, transferred CCR rates, STD and others [2, 6, 13, 14, 59]. The binding pocket of the target macromolecule is here indirectly mapped on the resonances of the two ligands. Due to the nature of the magnetization transfer, the information gained in the INPHARMA spectrum is highly site-specific, in contrast to STD or WATER-LOGSY experiments [20], which can only identify the face of the ligand that is solvent exposed with respect to that in contact with the receptor. The INPHARMA NOEs are used to rank and select binding modes of $L_1$ and $L_2$ obtained by docking the bound conformation of the ligands to a structural model of the apo-receptor. The direct employment of the INPHARMA NOEs in structure calculation programs (X-PLOR) is currently under development.

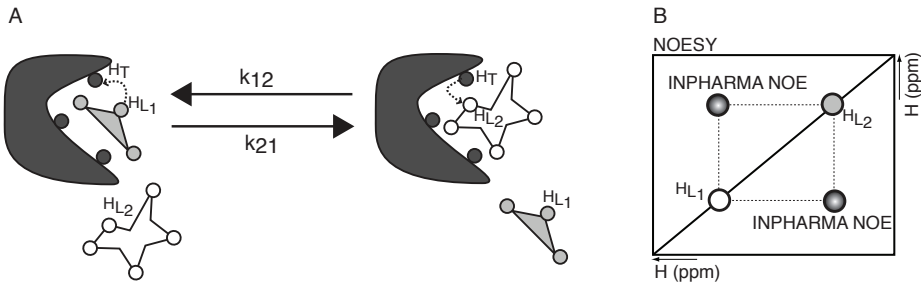



**Figure 4.1:** A) Schematic representation of the principle of the INPHARMA NOEs. At the beginning of the NOESY mixing time $L_1$ binds to the receptor and its proton H$L_1$ transfers magnetization to the proton of the receptor HT. Being $L_1$ a weak binder, it dissociates from the receptor during the NOESY mixing time and leaves the binding pocket free for $L_2$ to occupy it. At this point the magnetization deposited on HT from the H$L_1$ proton can be transferred to the H$L_2$ proton of $L_2$. This process results in a spin-diffusion mediated NOE peak between H$L_1$ and H$L_2$, depicted in B).

The method was originally developed to derive the binding mode of $L_2$ when the binding mode of $L_1$ is known. However, we have recently demonstrated that the INPHARMA approach is much more powerful and allows in favourable cases the de novo description of the binding mode of both $L_1$ and $L_2$ [78, 79]. In this work we have shown that the methodology is precise enough to unambiguously select one docking mode per ligand and accurate enough to select the correct docking mode, as compared to the crystal structures available for the investigated test cases [79].

Here we present a thorough theoretical analysis of the INPHARMA interligand NOE effects. We use simple model systems to describe the effect of various parameters, such as receptor proton density, NOESY mixing time, ligand concentrations, etc. on the size of the observed INPHARMA NOEs. Furthermore, we present a comprehensive description of the applicability of the INPHARMA approach. We provide all tools necessary to predict the size of the INPHARMA NOEs expected for complexes with different $K_d$ and $k_{off}$ values and to choose the best conditions, in terms of sample composition and experimental NMR parameters, for the observation of INPHARMA NOEs of respectable size. The theoretical data presented here are essential to ensure that the powerful INPHARMA method can be widely applied by non-NMR experts in the process of drug development.

## 4.2 Material and Methods

### 4.2.1 The model systems

The system that can be investigated with the INPHARMA method consists of a macromolecular receptor, usually a protein, and two competitively binding ligands. In this theoretical investigation we use simplified model systems to describe the influence of various structural parameters, such as intermolecular proton distances, proton densities in the receptor, depth of the binding pocket, etc., on the expected INPHARMA NOEs. Two simple geometries are used:

i) linear model: the ligands and the receptor consist of protons disposed in a linear arrangement with a fixed intra-molecular inter-proton distance d. The ligands contain from one to three protons (Fig. 4.2A);

ii) cubic model: the ligands consist of two or three protons linearly arranged with an inter-proton distance d; the receptor is a cube of different size (23 - 73).

In the receptor/ligand complex the protons are located at the nodes of a three-dimensional lattice with unit length d in all three dimensions (Fig. 4.2B). In the free receptor, the nodes of the binding pocket are empty. The ligands insert perpendicular to one face, in the middle of it, so that the last atom is part of this face.

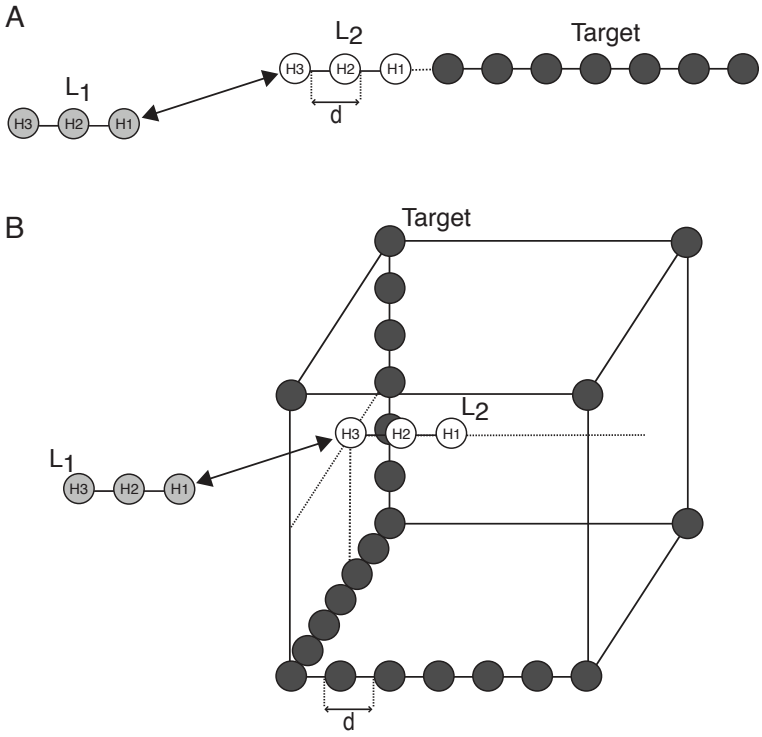



**Figure 4.2:** Simplified models of the system consisting of one receptor (in black) and two competitively binding ligands (in gray and white). A) Linear arrangement of the protons of both the receptor and the ligands; B) Linear arrangement for the protons of the ligands but cubic arrangement for the protons of the receptor. The ligands bind perpendicular to one face of the cube such that H3 stands in the middle of this face.

### 4.2.2 Theory

Intensities of NOE cross-peaks of a receptor/ligand complex in exchange with the free forms of both the ligand and the receptor are described by the following equation:

$$\frac{\mathbf{dM(t)}}{\mathbf{dx}} = -(\mathbf{R} + \mathbf{K}) \cdot (\mathbf{M(t)} - \mathbf{M_{eq}}) \tag{4.1}$$

with a solution of the form:

$$\mathbf{\Delta M}(\tau_\mathbf{m}) = \exp\left[-(\mathbf{R} + \mathbf{K}) \cdot \tau_\mathbf{m}\right] \cdot \mathbf{\Delta M(0)} \tag{4.2}$$

where **K** and **R** are the kinetic and relaxation matrices, M(0) the initial magnetization, $M_{eq}$ the equilibrium magnetization and the mixing time. Two different models have been considered for the chemical exchange. The first is a three-step model, including the state where both ligands and the receptor are found also in the free form, as described by:

$$TL_1 + L_2 \underset{k_{1on}}{\overset{k_{1off}}{\rightleftharpoons}} L_1 + T + L_2 \underset{k_{2off}}{\overset{k_{2on}}{\rightleftharpoons}} L_1 + TL_2 \tag{4.3}$$

In the second model, we assume that the receptor is never in the unbound state, due to presence of the two ligands in large excess. This model is described by:

$$TL_1 + L_2 \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} L_1 + TL_2 \tag{4.4}$$

#### Model 1

The relaxation matrix is a diagonal block matrix, where each sub-matrix describes the proton-proton relaxation pathway of one species; the species present in solution are the two ligands in the free state $L_1$, $L_2$, the two complexes $TL_1$ and $TL_2$ of the target macromolecule and $L_1$ or $L_2$, respectively [74], and the free target macromolecule T. The superindex of each $\mathbf{R}_s^x$ indicates the species (either the free ligands or the complexes $TL_1$ and $TL_2$) and the subindex indicates which protons of the species contribute to that sub-matrix. Thus, $\mathbf{R}_{L_1}^{L_1}$ describes the relaxation of the protons of ligand $L_1$ in the free state and $\mathbf{R}_{L_1}^{TL_1}$ the relaxation of the protons of ligand $L_1$ when bound in the complex $TL_1$. Analogously, $\mathbf{R}_{T}^{TL_1}$ describes the relaxation of the

protein protons in the complex T$L_1$ and $\mathbf{R}^{TL_1}_{L_1,T}$ contains the cross-relaxation terms between the protons of ligand $L_1$ and the protons of protein T in the complex T$L_1$.

$$\mathbf{R} = \begin{pmatrix} R^{L_1}_{L_1} & & & & & & & \\ & R^{L_2}_{L_2} & & & & & & \\ & & R^{TL_1}_{L_1} & R^{TL_1}_{L_1,T} & & & & \\ & & R^{TL_1}_{T,L_1} & R^{TL_1}_{T} & & & & \\ & & & & R^{TL_2}_{L_2} & R^{TL_2}_{L_2,T} & & \\ & & & & R^{TL_2}_{T,L_2} & R^{TL_2}_{T} & & \\ & & & & & & R^{T}_{T} \end{pmatrix} \tag{4.5}$$

The kinetic matrix is built according to the equations that rule the chemical equilibrium:

$$\begin{aligned} \frac{d[L_1]}{dt} &= -k_{1on} \cdot [T][L_1] + k_{1off} \cdot [TL_1] \\ \frac{d[TL_1]}{dt} &= -\frac{d[L_1]}{dt} \\ \frac{d[L_2]}{dt} &= -k_{2on} \cdot [T][L_2] + k_{2off} \cdot [TL_2] \\ \frac{d[TL_2]}{dt} &= -\frac{d[L_2]}{dt} \\ \frac{d[T]}{dt} &= -\frac{d[TL_1]}{dt} - \frac{d[TL_2]}{dt} \end{aligned} \tag{4.6}$$

and has the form:

$$\mathbf{K} = \begin{pmatrix} k_{1on}[T]\mathbf{I} & & -k_{1off}\mathbf{I} & & & & \\ & k_{2on}[T]\mathbf{I} & & & -k_{2off}\mathbf{I} & & \\ -k_{1on}[T]\mathbf{I} & & k_{1off}\mathbf{I} & & & & \\ & & & k_{1off}\mathbf{I} & & & -k_{1on}[L_1]\mathbf{I} \\ & -k_{2on}[T]\mathbf{I} & & & k_{2off}\mathbf{I} & & \\ & & & & & k_{2off}\mathbf{I} & -k_{2on}[L_2]\mathbf{I} \\ & & & -k_{1off}\mathbf{I} & & -k_{2off}\mathbf{I} & (k_{1on}[L_1]+k_{2on}[L_2])\mathbf{I} \end{pmatrix} \tag{4.7}$$

where $\mathbf{I}$ is the identity matrix of the same size as the corresponding block matrix in $\mathbf{R}$.

### Model 2

Compared to the matrix for model 1, the relaxation matrix $\mathbf{R}$ for model 2 lacks the last block representing the unbound target macromolecule.

$$\mathbf{R} = \begin{pmatrix} R_{L_1}^{L_1} & & & & & \\ & R_{L_2}^{L_2} & & & & \\ & & R_{L_1}^{TL_1} & R_{L_1,T}^{TL_1} & & \\ & & R_{T,L_1}^{TL_1} & R_T^{TL_1} & & \\ & & & & R_{L_2}^{TL_2} & R_{L_2,T}^{TL_2} \\ & & & & R_{T,L_2}^{TL_2} & R_T^{TL_2} \end{pmatrix} \tag{4.8}$$

The kinetics matrix **K** is derived from the equations:

$$\begin{aligned}
\frac{d[L_1]}{dt} &= k_{12} \cdot [TL_1][L_2] - k_{21} \cdot [TL_2][L_1] \\
\frac{d[L_2]}{dt} &= -\frac{d[L_1]}{dt} \\
\frac{d[TL_2]}{dt} &= \frac{d[L_1]}{dt} \\
\frac{d[TL_1]}{dt} &= -\frac{d[L_1]}{dt}
\end{aligned} \tag{4.9}$$

$$\mathbf{K} = \begin{pmatrix}
k_{21}[TL_2]\mathbf{I} & & -k_{12}[L_2]\mathbf{I} & & & \\
& k_{12}[TL_1]\mathbf{I} & & & -k_{21}[L_1]\mathbf{I} & \\
-k_{21}[TL_2]\mathbf{I} & & k_{12}[L_2]\mathbf{I} & & & \\
& & & k_{12}[L_2]\mathbf{I} & & -k_{21}[L_1]\mathbf{I} \\
& -k_{12}[TL_1]\mathbf{I} & & & k_{21}[L_1]\mathbf{I} & \\
& & & -k_{12}[L_2]\mathbf{I} & & k_{21}[L_1]\mathbf{I}
\end{pmatrix} \tag{4.10}$$

where I is the identity matrix of the same size as the corresponding block matrix in **R**. The elements of the kinetics matrices of the two models are derived from the binding constants and the equilibrium concentrations. The last are calculated from the mass conservation laws.

Model 1:

$$\begin{aligned}
k_{2on}[T][L_2] &= k_{2off}[TL_2] \\
k_{1on}[T][L_1] &= k_{1off}[TL_1] \\
TL_1] + [TL_2] + [T] &= [T]_{tot} \\
[TL_1] + [L_1] &= [L_1]_{tot} \\
[TL_2] + [L_2] &= [L_2]_{tot}
\end{aligned} \tag{4.11}$$

Model 2:

$$\begin{aligned} k_{12}[TL_1][L_2] &= k_{21}[TL_2][L_1] \\ [TL_1] + [TL_2] + [T] &= [T]_{tot} \\ [TL_1] + [L_1] &= [L_1]_{tot} \\ [TL_2] + [L_2] &= [L_2]_{tot} \end{aligned} \tag{4.12}$$

Individual elements of the relaxation matrix have the following form:

$$\begin{aligned} R_{i,i} = \rho_i &= \sum_{\substack{H_j \in A \\ j \neq i}} \frac{b^2}{d_{ij}^6} \cdot (J(0) + 3J(\omega) + 6J(2\omega)) \\ R_{i,j} = \sigma_{ij} &= \frac{b^2}{d_{ij}^6} \cdot (6J(2\omega) - J(0)) \\ J(\omega) &= \frac{2}{5} \left( \frac{\tau_c}{1 + (\omega\tau_c)^2} \right) \\ b &= \frac{1}{2} \cdot \frac{\mu_0}{4\pi} \hbar \gamma_H^2 \end{aligned} \tag{4.13}$$

where $\rho_i$ is the longitudinal relaxation rate of proton $H_i$, $\sigma_i j$ is the cross-relaxation rate of protons $H_i$ and $H_j$, $d_{ij}$ is the distance between protons $H_i$ and $H_j$ in the same chemical species (ligand, target macromolecule or complex) and "A" stand for the chemical species. All protons $H_j$ are considered in the calculation of the relaxation rates $\rho_i$ and $\sigma_{ij}$, irrespectively of the value of $d_{ij}$.

### 4.2.3 Analytical solution for a three spins system

A description of the transferred–NOE effect for a two-spin system has been given in great detail [54, 59, 60, 65]. In these studies the receptor was not included in the calculation, namely it was assumed that it does not contribute considerably to the process of magnetization transfer between the ligand protons. However, it has been reported that the contribution of the receptor cannot be neglected for an accurate prediction of transferred-NOEs [4, 5, 63, 73, 76, 108].

Clearly, for the calculation of the INPHARMA NOEs, the receptor protons must be taken into account as they mediate the transfer of magnetization between the two ligands. An analytical solution that describes the INPHARMA NOE transfer cannot be systematically found for a three spin system, where the two ligands $L_1$ and $L_2$ and the receptor T consist of one proton each, as this would correspond to searching for an analytical solution of a polynomial equation of sixth order [30]. However an analytical description of the INPHARMA transfer can be found by making two crude approximations: 1. the concentrations of the ligands and the receptor are equal; 2. all four species $L_1$, $L_2$, T$L_1$ and T$L_2$ have the same auto-relaxation rates $\rho$ and the cross-correlated relaxation rate $\sigma$ between the $H_T$ and $H_L$ protons in T$L_1$ and T$L_2$ are equal. Furthermore, the affinity of the two ligands is also chosen to be equal.

The first two approximations are clearly not realistic. However, an "easy-to-read" analytical solution for the three spin-system, even under the two coarse approximations made above is still useful to describe at a glance the overall dependence of the INPHARMA magnetization transfer on internal dynamics and correlation time. In the next section we show results calculated numerically for realistic scenarios with real physical constants and for multi-spin systems. Under the approximations mentioned above, the relaxation matrix and the kinetics matrix simplify to the following form:

$$\mathbf{R}+\mathbf{K}=\begin{pmatrix} k+\rho & 0 & -k & 0 & 0 & 0 \\ 0 & k+\rho & 0 & 0 & -k & 0 \\ -k & 0 & k+\rho & \sigma & 0 & 0 \\ 0 & 0 & \sigma & k+\rho & 0 & -k \\ 0 & -k & 0 & 0 & k+\rho & \sigma \\ 0 & 0 & 0 & -k & \sigma & k+\rho \end{pmatrix} \tag{4.14}$$

where $k = k_{12} = k_{21}$.

Solving Eq. 4.1 implies finding six eigenvalues for:

$$det(\mathbf{R} + \mathbf{K} - \lambda\mathbf{I}) = 0 \tag{4.15}$$

The INPHARMA NOE S(t) between $L_1$ and $L_2$ is defined by:

$$S(t) = -k^3\sigma^2 \sum_{i=1}^{6} \alpha_i \cdot e^{-\lambda_i t} \tag{4.16}$$

where $\lambda_i$ are coefficients that fulfilled the following equation system:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 \\ {\lambda_1}^2 & {\lambda_2}^2 & {\lambda_3}^2 & {\lambda_4}^2 & {\lambda_5}^2 & {\lambda_6}^2 \\ {\lambda_1}^3 & {\lambda_2}^3 & {\lambda_3}^3 & {\lambda_4}^3 & {\lambda_5}^3 & {\lambda_6}^3 \\ {\lambda_1}^4 & {\lambda_2}^4 & {\lambda_3}^4 & {\lambda_4}^4 & {\lambda_5}^4 & {\lambda_6}^4 \\ {\lambda_1}^5 & {\lambda_2}^5 & {\lambda_3}^5 & {\lambda_4}^5 & {\lambda_5}^5 & {\lambda_6}^5 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{4.17}$$

as described in [55]. The solution can be calculated from the known determinant of the Vandermonde matrix of Eq. 4.17:

$$\begin{aligned} S(t) &= -k^3\sigma^2 \sum_{i=1}^{6} \Pi(\lambda_i) \cdot e^{-\lambda_i t}, \\ \Pi(\lambda_i) &= \prod_{\substack{j=1 \\ j\neq i}}^{6} (\lambda_i - \lambda_j)^{-1} = \alpha_i \\ \lambda_{i=1,3,5} &= \frac{1}{3}(4k + 3\rho) - \frac{2}{3}\sqrt{4k^2 + 3\sigma^2}\cos\theta_i \\ \lambda_{i=2,4,6} &= \frac{1}{3}(2k + 3\rho) - \frac{2}{3}\sqrt{4k^2 + 3\sigma^2}\cos\theta_i \\ \theta_i &= \frac{1}{3}\left\{ \tan^{-1}\left( \frac{3\sqrt{3(32k^4\sigma^2 + 19k^2\sigma^4 + 4\sigma^6)}}{-16k^3 + 9k\sigma^2} \right) + \frac{\pi}{2}(1 - Sign(-16k^3 + 9k\sigma^2)) - i\pi \right\} \end{aligned} \tag{4.18}$$

Even with the drastic approximations made for this model system, the solution turns out to be quite complicated. However, a straightforward for-

mula is obtained for the asymptotic development of the variable k towards very high values (Fig. 4.3) and lead to the canonical expression of the INPHARMA NOE:

$$S(t) \underset{k\to\infty}{\sim} \frac{1}{4} \cdot e^{-\rho t} \left( \cosh\left(\frac{\sigma t}{\sqrt{2}}\right) - 1 \right) \tag{4.19}$$

This expression is reminiscent of the well known dependence of transferred-NOEs intensities on the mixing time t of the NOESY experiment:

$$S(t) \underset{k\to\infty}{\sim} \frac{1}{4} \cdot e^{-\rho t} \cdot \sinh\left(\frac{\sigma t}{\sqrt{2}}\right) \tag{4.20}$$

However, while the transferred-NOEs intensities depend on the time t according to the hyperbolic sine function, which can be approximated to a linear function at small mixing times, the INPHARMA NOEs show a hyperbolic cosine dependence on time, which can be approximated by a parabolic curve at short mixing times. This behavior reveals the spin-diffusion nature of the INPHARMA NOEs, which depend quadratically on $\sigma t$. The formula in Eq. **??** can be used to visualize the influence of the correlation time and internal motions on the build-up of the INPHARMA-NOE. The effect of internal motions is estimated using the Lipari Szabo´s model free approach in the limit of a very fast internal correlation time.

$$S(t) \underset{k\to\infty}{\sim} \frac{1}{4} \cdot e^{-S^2\rho t} \left( \cosh\left(\frac{S^2\sigma t}{\sqrt{2}}\right) - 1 \right) \tag{4.21}$$

where S is the order parameter varying from 0 to 1 and assuming equal order parameters for the auto-relaxation rate $\rho$ and the cross-relaxation rate $\sigma$.
In Figure 4.3, we show the dependence of the INPHARMA NOEs on the mixing time t, according to Eq. 4.21, for different correlation times of the molecule $\tau_c$ and different order parameters S. At short mixing times the curves have a parabolic shape, according to:

$$S(t) \underset{\substack{k\to\infty \\ t\to 0}}{\sim} \frac{1}{4} \cdot \left(\frac{S^2\sigma t}{\sqrt{2}}\right)^2 \tag{4.22}$$

For $S^2 < 1$, the cross-relaxation rate $\sigma$ is slower resulting in a delayed build-up of the INPHARMA NOEs. As for the transferred-NOEs, the presence of an order parameter $S \neq 1$ has a strong effect on the INPHARMA NOE at various mixing times. At 1s mixing time the difference is negligible while

at 400 ms, there is a factor of four between the INPHARMA NOE assuming $S^2 = 1$ and $S^2 = 0.5$ (full line and dashed-dotted line in Fig. 4.3).

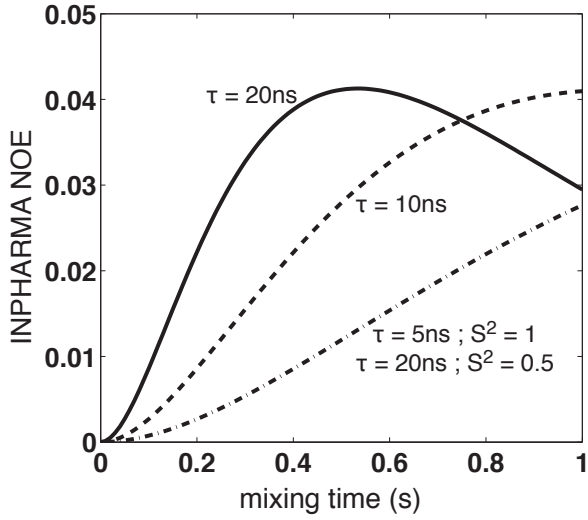



**Figure 4.3:** Theoretical dependence of the INPHARMA NOEs on the mixing time $\tau_m$ under the approximations used to derive the analytical expression of Eqs. 4.19 and 4.21. The inter proton distances d are equal to 2.5 Å. The full line curve was calculated assuming $\tau = 20$ ns and $S^2 = 1$. The dashed line curve was calculated assuming $\tau = 10$ ns and $S^2 = 1$. The dotted-dashed line curve was calculated assuming $\tau = 5$ ns and $S^2 = 1$. The last curve was calculated assuming $\tau = 20$ ns and $S^2 = 0.5$ (overlapped with the dotted-dashed curve). Other parameters are: field strength = 800 MHz, $[T]_{tot} = 50$ $\mu$M, $[L1]_{tot} = [L2]_{tot} = 500$ $\mu$M, $k_{off} = 1$ kHz.

This difference has to be taken in account in the interpretation of the INPHARMA NOE peaks that are transferred by highly dynamic parts of the binding pocket. Efforts to include order parameters in the interpretation of the data are on-going in our laboratories. Numerical solution for a multi-spin system. For more than three spins and for systems with realistic parameters, the equations system becomes too complex to search for an analytical solution. Therefore a numerical solution is calculated in Matlab for the linear and cubic model systems assuming both model 1 and model 2 for the chemical exchange and in dependence of physical parameters such as $\tau_c$, $\tau_m$, $k_{off}$, species concentrations and internal dynamics. The results of such simulations, which are essential to understand the influence of the physical and experimental

parameters on the size of the observed INPHARMA NOEs, are presented in the following section.

## 4.3 Results and Discussion

### 4.3.1 Choice of the kinetic model

One of the challenges of the INPHARMA method is to handle a large number of docking models for each of the two ligands, which need to be ranked with respect to the measured interligand NOEs. Thus, the calculation time needed to predict the interligand NOEs for one pair of models should be kept to a minimum. A medium size protein (20 kDa) contains about thousand protons, which translates into a matrix **R**+**K** of ca. nine million elements. The size of the **R**+**K** matrix can be reduced to four millions by choosing the kinetic model 2, where the presence of the free receptor in solution is neglected. In order to evaluate the error introduced on the calculated INPHARMA NOEs by neglecting the presence of the free receptor in solution, we simulated the interligand NOEs between the two protons H1 of $L_1$ and H1 of $L_2$ in presence of a common receptor consisting of a cube of 73 protons. In all simulations shown in this and the following paragraphs, the magnetization originates from $L_1$ and is transferred to $L_2$; the INPHARMA NOE shown on the y axis is normalized to the diagnol peak of the proton H1 of $L_1$ in a NOESY experiment with mixing time $\tau_m = 0$; a diffusion limited $k_{on}$ of $10^8$ $M^{-1}$ $s^{-1}$ is assumed; for the cubic model system (Fig. 4.2B) the ligand is placed deep in the receptor cube with the ligand proton H1 at coordinates (4,4,3), unless stated otherwise. Clearly the error introduced by using the approximated model 2 depends on the amount of receptor free in solution at equilibrium. For a system consisting of $[L_1]_{tot} = [L_2]_{tot} = 500$ $\mu$M, $[T]_{tot} = 50$ $\mu$M and with $K_d = 100$ $\mu$M for both ligands, the error made by neglecting the presence of the free receptor (4 $\mu$M) varies with the mixing time of the NOESY experiment and reaches ca. 10% or 5% at $\tau_m = 0.5$ s or 1 s respectively (Fig. 4.4A). However, for tighter binding ligands with $K_d = 10$ $\mu$M or lower the presence of the free receptor (0.5 $\mu$M) can be safely neglected, as the error on the interligand NOE is less than 3% to 5% over the complete range of mixing time from 0 s to 1 s (Fig. 4.4B). In our experience errors smaller than 5% - 10% are not expected to influence the outcome of the analysis.

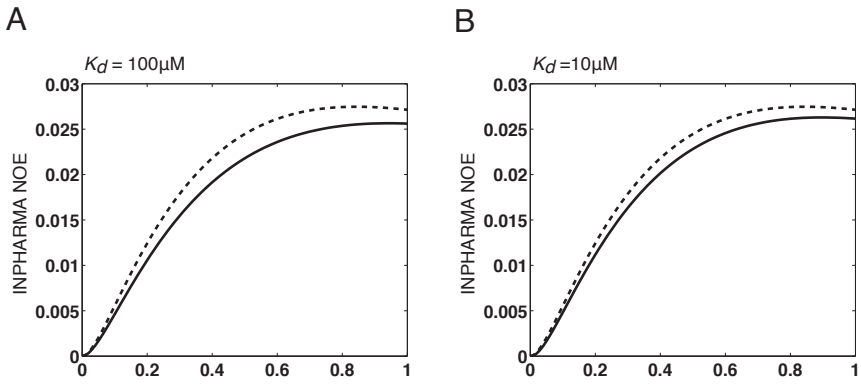



**Figure 4.4:** Dependence on the mixing time $\tau_m$of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a cubic receptor with 7 protons per dimension, as shown in Fig. 4.2B. The full line curve was calculated assuming the kinetic model 1, where the presence of the free receptor was explicitly taken into account as in Eq. 4.3; the dashed curve was calculated assuming the kinetic model 2, where the presence of the free receptors is neglected, as in Eq. 4.4. Other parameters are: $\tau_{cT} = 20$ ns, $\tau_{cA} = \tau_{cB} = 0.1$ ns, field strength = 800 MHz, d = 2.5 Å, $[T]_{tot} = 50$ $\mu$M, $[L1]_{tot} = [L2]_{tot} = 500$ $\mu$M. $k_{off} = 10$ kHz in panel A and $k_{off} = 1$ kHz in panel B.

### 4.3.2 Dependence on the $k_{off}$ of the two ligands

In order to observe the INPHARMA NOEs, the two ligands should bind weakly to the receptor, namely they should exchange a few times between their free and bound states during the mixing time of a NOESY experiment. Figure 4.5 shows the dependence on the $k_{off}$ of the INPHARMA NOE occurring between the two protons H1 of $L_1$ and H1 of $L_2$ in presence of a common receptor consisting of a cube of 73 protons, assuming model 1 for the exchange and equal affinity for the two ligands. For a mixing time of 500 ms and a $k_{off}$ of 1 kHz or higher (corresponding to a $K_d$ of 10 $\mu$M if one assumes a diffusion limited $k_{on}$ of $10^8$ $M^{-1}$ $s^{-1}$) the INPHARMA NOE has reached its maximum value, while for a $k_{off}$ of 100 Hz the INPHARMA NOE has an observable size equal to about two third of its maximum value. These results indicate that the transfer of magnetization between the two ligands improves with the number of exchange events in the binding pocket of the receptor. The condition on the $k_{off}$ is not too stringent and the INPHARMA NOE could be observed for most weakly binding ligands, typically available in an early phase of drug development, which we tested.

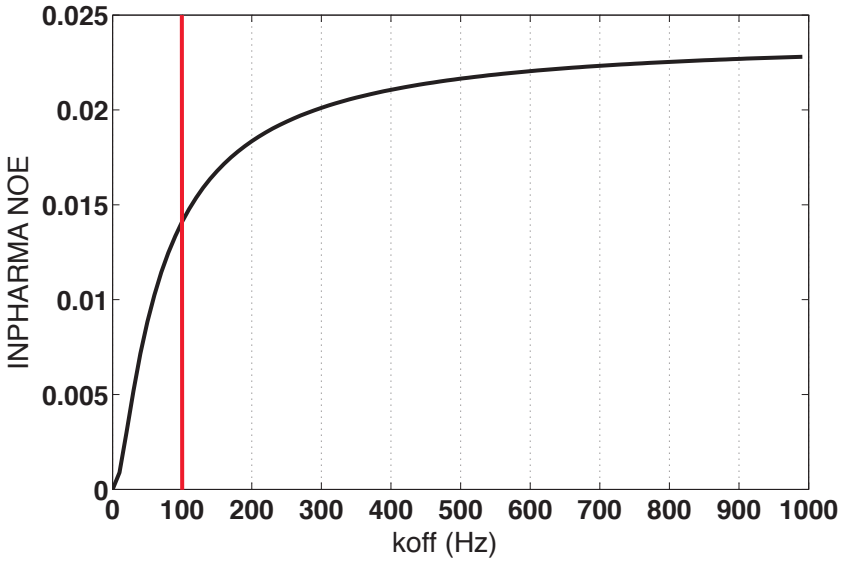



**Figure 4.5:** Dependence on the $k_{off} = k_{1off} = k_{2off}$ of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a cubic receptor with 7 protons per dimension, as shown in Fig. 4.2B. The curve was calculated assuming the kinetic model 1; Other parameters are $\tau_m = 500$ ms, $\tau_{cT} = 20$ ns, $\tau_{cA} = \tau_{cB} = 0.1$ ns, field strength = 800 MHz, d = 2.5 Å, $[T]_{tot} = 50$ $\mu$M, $[L1]_{tot} = [L2]_{tot} = 500$ $\mu$M.

More critical is the relative size of the dissociation rates for $L_1$ and $L_2$, $k_{1off}$ and $k_{2off}$. Clearly, the optimal situation for the receptor mediated transfer of magnetization between the two ligands is when $k_{1off} = k_{2off}$ and the two ligands spend an equal amount of time in the receptor bound state. Figure 4.6 shows the efficiency of the INPHARMA NOE transfer when $k_{1off}/k_{2off} \neq 1$. A considerable amount of magnetization transfer through the INPHARMA NOE is obtained for a $k_{1off}/k_{2off}$ value of up to 8, while for $k_{1off}/k_{2off} = 16$ the INPHARMA NOE is less than 1% at a mixing time of 1s and has decreased by a factor of 5 at 500 ms with respect to $k_{1off} = k_{2off}$. Thus, the method requires that the affinity of the two ligands is of comparable size.

An obvious way to compensate for the different affinity of the two ligands, and consequently for the different equilibrium concentrations of the $TL_1$ and $TL_2$ complexes in solution, might seem that of adjusting the concentration of the two ligands, as for example using $[L_1]_{tot} = 10[L_2]_{tot}$ if $k_{1off} = 10k_{2off}$.

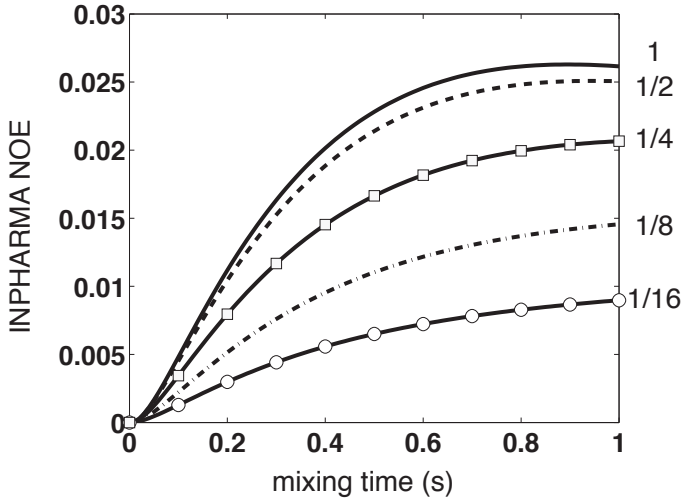



**Figure 4.6:** Dependence on the mixing time $\tau_m$ of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a cubic receptor with 7 protons per dimension, for different $k_{1off}/k_{2off}$ ratio. The curves were calculated assuming the kinetic model 1 and $k_{2off} = 1$ kHz. Other parameters are as in Fig. 4.5.

At small mixing times, for which the two-step transfer dominates the INPHARMA signal, a concentration ratio $[L_1]_{tot} = 10[L_2]_{tot}$ is optimal (Fig. 4.7). However, at longer mixing times the optimal concentration changes due to the contribution of additional pathways of magnetization transfer and spin diffusion inside the receptor. The use of longer mixing times renders the analysis of the INPHARMA NOE intensity more difficult, due to the contribution of many spin-diffusion pathways; on the other hand long mixing times are necessary to increase the intensity of the INPHARMA signal. For a complex with $\tau_c = 20$ ns the optimal value is found around $[L_2]_{tot}/[L_1]_{tot} = 0.33$ (Fig. 4.7B). On the other hand for a complex with $\tau_c = 200$ ns (Fig. 4.7C) the optimal concentration ratio of the two ligands for long mixing times is found for similar concentrations of $L_1$ and $L_2$ ($[L_2]_{tot}/[L_1]_{tot} = 0.81$) despite the differences in the $k_{off}$. In this case it is more important to achieve fast transfer of magnetization between the two ligands away from the fast relaxing receptor, in order to avoid the diffusion of the ligands magnetization in the receptor, rather than to have similar population of the $TL_1$ and $TL_2$ complexes. Figure 4.7A summarizes these results.

Here the INPHARMA NOE in dependence of the mixing time for ligands with equal affinity and equal concentration (full line), as compared to ligands with $k_{1off} = 10k_{2off}$ (dashed line) is shown. Not unexpectedly, the optimal INPHARMA integrals are observed when the affinities and the concentrations of the two ligands are equal. Changing the ratio of $[L_2]_{tot}/[L_1]_{tot}$ from 1 to 0.1 (circles) while keeping the overall concentration of the ligands constant ($[L_2]_{tot} + [L_1]_{tot}$= 1 mM) cannot compensate the decrease of the transfer efficiency due to unequal affinities. This result can be understood considering that while the $k_{off}$ represents the inverse of the lifetime of the complex, the concentration of the ligand influences the probability that the complex is formed. Thus, owing to the spin-diffusion loss between two consecutive complex formations, binding of the ligand to the receptor for twice as long does not have the same effect as binding twice as often. At shorter mixing times the INPHARMA NOE for $k_{1off} = 10k_{2off}$ and $[L_2]_{tot}/[L_1]_{tot} = 0.1$ (dashed) is larger than for $k_{1off} = 10k_{2off}$ and $[L_2]_{tot}/[L_1]_{tot} = 1$ (circles), while at longer mixing times the INPHARMA NOE shows the opposite behavior. The critical point where the INPHARMA NOE for $[L_2]_{tot}/[L_1]_{tot} = 0.1$ becomes worse than for$[L_2]_{tot}/[L_1]_{tot} = 1$ depends on the correlation time of the complex and shifts to shorter mixing times for larger $\tau_c$.

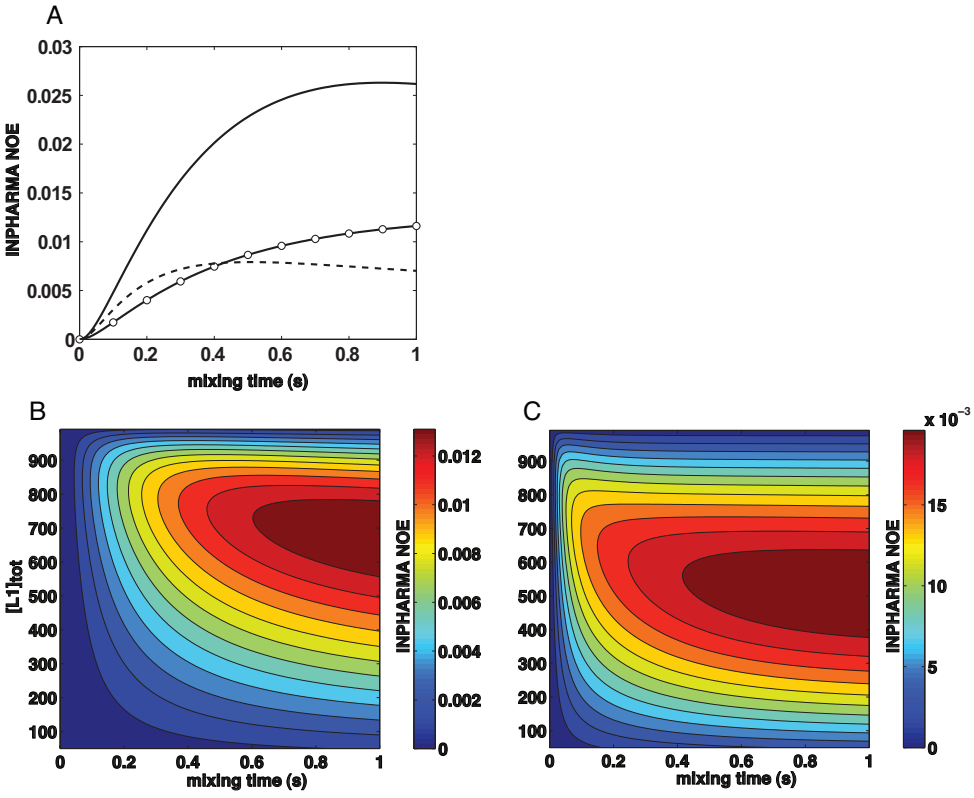



**Figure 4.7:** A) Dependence on the mixing time $\tau_m$ of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a cubic receptor with 7 protons per dimension, for $k_{1off}/k_{2off} = 1$ (full line) and $k_{1off}/k_{2off} = 10$ (circles). The curves were calculated assuming the kinetic model 1 and $k_{2off} = 1$ kHz. The concentrations of the species are: $[T]_{tot} = 50$ $\mu$M, $[L_1]_{tot} = [L_2]_{tot} = 500\mu$M (full); $[T]_{tot} = 50\mu$M, $[L_1]_{tot} = [L_2]_{tot} = 500\mu$M (circles); $[T]_{tot} = 50\mu$M, $[L_1]_{tot} = 910\mu$M,$[L_2]_{tot} = 90\mu$M (dashed). The dashed curve is multiplied by $[L_1]_{910}/[L_1]_{500}$ with $[L_1]_{910}$ and $[L_1]_{500}$ being the equilibrium concentrations of $[L_1]$ for $[L_1]_{tot} = 910\mu$M or 500 $\mu$M, respectively. This is necessary to compensate for the higher initial magnetization of $L_1$ by which the INPHARMA NOE intensity is normalized. Other parameters are as in Fig. 4.5. This panel shows that for short mixing times, the concentrations that compensate differences in the off rates give the largest INPHARMA signals. For longer mixing times, in which spin diffusion becomes more important, other concentrations are optimal. B-C) Contour plots of the INPHARMA NOE intensity in dependence of the $\tau_m$ (x axis) and of the $[L_1]$ tot concentration. (y axis) for and $k_{1off}/k_{2off}$ = 10. The total concentration of the two ligands $[L_1]_{tot} + [L_2]_{tot}$ is constant (1 mM). The INPHARMA NOE intensities are normalized to the initial magnetization of $L_1$ when $[L_1]_{tot}$ = 500 $\mu$M for comparison. B) transfer from proton H1 of $L_1$ and proton H1 of $L_2$ ; $\tau_{cT} = 20$ ns; C) transfer from proton H1 of $L_1$ and proton H1 of $L_2$ ; $\tau_{cT} = 200$ ns; Other parameters are: $\tau_{cA} = \tau_{cB} = 0.1$ ns, field strength = 800 MHz, d = 2.5 Å, $[T]_{tot} = 50\mu$M.

### 4.3.3 Influence of proton density

An important question is how the number of protons of both the ligands and the receptor influences the INPHARMA NOEs. Clearly, the most efficient transfer is obtained when the magnetization is not allowed to diffuse away through undesired spin-diffusion pathways and only the desired spin-diffusion transfer $L_1 \rightarrow receptor \rightarrow L_2$ occurs. In a real system, however, this is never the case, due to the presence of protons in the ligands and in the receptor that do not belong to the binding epitope or to the binding pocket.

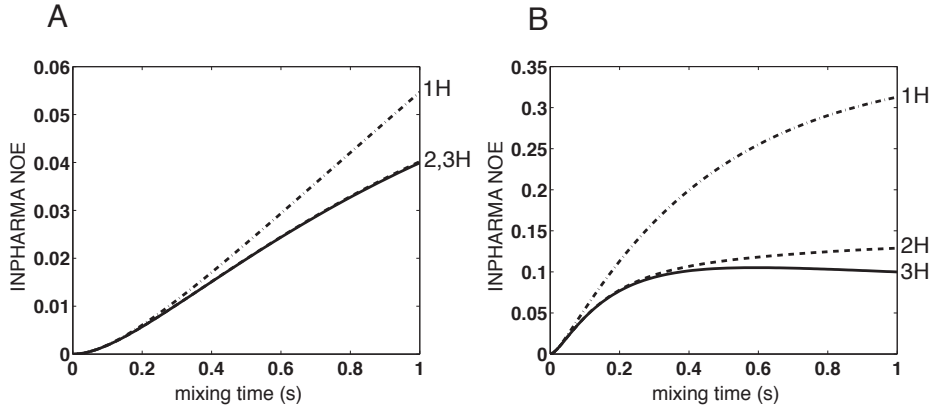



**Figure 4.8:** Dependence on the mixing time $\tau_m$ of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a linear receptor consisting of 7 protons as shown in Fig. 4.2A. The size of the ligands is variable, with the dotted-dashed curve showing the transfer for ligands containing only proton H1, the dashed curve for ligands consisting of H1 and H2 and the full curve for ligands consisting of H1, H2 and H3. The curves were calculated assuming the kinetic model 1 and $k_{1off} = k_{2off} = 1$ kHz. The concentrations of the species are: $[T]_{tot} = 50\mu$M, $[L1]_{tot} = [L_2]_{tot} = 500\mu$M. Curves in panel A and B were calculated assuming $\tau_{cT} = 20$ ns and $\tau_{cT} = 200$ ns, respectively.

We first investigate the effect of the ligand proton density by comparing the INPHARMA NOE in a linear system, as in Fig. 4.2A, for the case when the two ligands consist of only one, two or three protons (Fig. 4.8). The presence of protons H2 of $L_1$ and $L_2$ diminishes the efficiency of the transfer between H1 of $L_1$ and H1 of $L_2$ by a factor of three (Fig. 4.8B), due to the fact that the ligand protons H2 compete with the receptor protons for the H1 magnetization. However, the presence of additional ligand protons at $d > 5$ Å (H3 of $L_1$ and $L_2$) does not further affect the efficiency of the INPHARMA

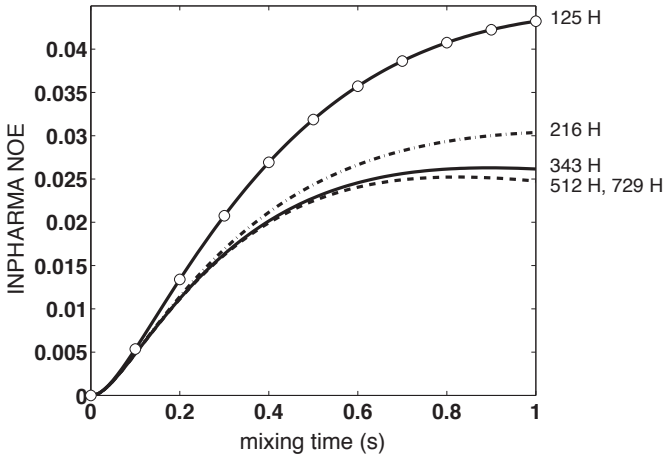



**Figure 4.9:** Dependence on the mixing time $\tau_m$ of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a cubic receptor of variable size: 53 (circles), 63 (doted-dashed), 73 (full), 83 and 93 (dashed, overlapped). The curves were calculated assuming the kinetic model 1; other parameters are $\tau_{cT} = 20$ ns, $\tau_{cA} = \tau_{cB} =$ 0.1 ns, field strength = 800 MHz, d = 2.5 Å, $[T]_{tot} = 50\mu$M, $[L_1]_{tot} = [L_2]_{tot} = 500\mu$M.

NOE.

Due to the dependence of spin diffusion on the size of the molecule, proton density on the large receptor is likely to have a stronger effect on the intensity of the INPHARMA NOE.

To simulate the dependence of the INPHARMA NOE on the number of protons of the receptor, while keeping the $\tau_{cT}$ constant, we used the artificial cubic system of Fig. 4.2B. Figure 4.9 shows the results of such simulations for a cube of size ranging from 33 to 93. As expected, the efficiency of magnetization transfer diminishes upon increasing the number of receptor protons, due to the undesired diffusion of the ligand(s) magnetization from the binding pocket to the receptor body. Such a loss reaches a maximum for a sphere of about 8 Å around the ligand binding pocket (size of the cube = 73); spin diffusion processes beyond this sphere do not influence the INPHARMA NOE to a measurable extent for molecules with a $\tau_c$ similar to that used in the simulation (20 ns). Clearly, increasing the size of the receptor directly translates into a higher velocity of the spin diffusion and to a larger radius of the sphere within which protons contribute to the magnetization transfer process.

However, also for very large receptors the mixing time of the NOESY experiment can be optimized to obtain the maximum INPHARMA NOE transfer, while minimizing dilution of the signal throughout the receptor (see next paragraph). For a receptor with $\tau_c = 20$ ns, we can conclude that spin-diffusion inside the receptor competes with the INPHARMA magnetization transfer up to a distance of 8 Å from the binding pocket and can therefore be considered as a "local" process.

### 4.3.4 Optimal choice of $\tau_m$: dependence on $\tau_m$

The mixing time that maximizes the intensity of the INPHARMA NOEs depends on the correlation time of the complex. Figure 4.10 shows the dependence of the INPHARMA NOE on the mixing time $\tau_m$ for receptors with increasing $\tau_c$. At short mixing times the INPHARMA NOE shows a parabolic dependence on $\tau_m$, which is typical of a second order effect (spin-diffusion) (see also Eq. 4.21). Increasing the correlation time increases the steepness of the initial part of the curve.

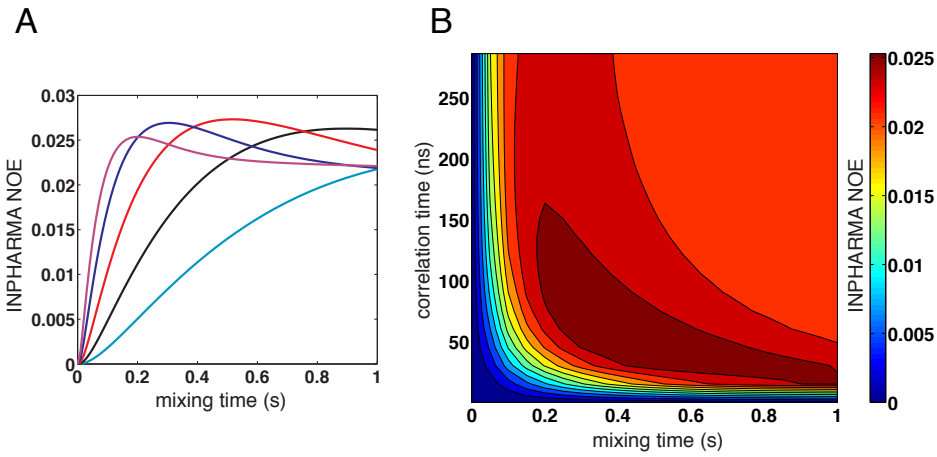



**Figure 4.10:** A.) Dependence on the mixing time $\tau_m$ of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a cubic receptor of size 73 and variable $\tau_c$: 10 ns (sky blue), 20 ns (black), 40 ns (red), 80 ns (dark blue), 160 ns (purple). The curves were calculated assuming the kinetic model 1; B) Contour plot of the INPHARMA NOE intensity in dependence of the $\tau_m$ (x axis) and of the $\tau_c$ of the receptor (y axis). Other parameters are: $\tau_{cA} = \tau_{cB} = 0.1$ ns, field strength = 800 MHz, d = 2.5 Å, $[T]_{tot} = 50\mu$M, $[L_1]_{tot} = [L_2]_{tot} = 500\mu$M.

As it is seen in Fig. 4.10B, for a receptor of 400 kDa or larger, the maximum INPHARMA NOE is obtained for $\tau_m < 100$ ms, while for a receptor of ca. 20 kDa, mixing times of the order of 500 ms or higher can be used. In our experience the best results, in terms of selection of the correct docking mode by comparison of the theoretical and experimental INPHARMA NOEs, are obtained when the experimental data are acquired at mixing times for which the INPHARMA NOE has not yet reached its maximum value. This is due to the fact that in this regime only „short" spin diffusion pathways contribute to the signal, thus minimizing the errors made by neglecting internal motions in the calculation of the theoretical NOEs *(vide infra)*. In fact, the more receptor protons contribute to the spin-diffusion transferred NOEs, the largest is the error made by neglecting their internal dynamics.

### 4.3.5 Dependence on receptor concentration

Clearly the size of the INPHARMA NOE depends not only on the $\tau_c$ of the receptor, the affinity of the ligands to the receptor and the kinetic constants but also on the concentration of the species in solution.

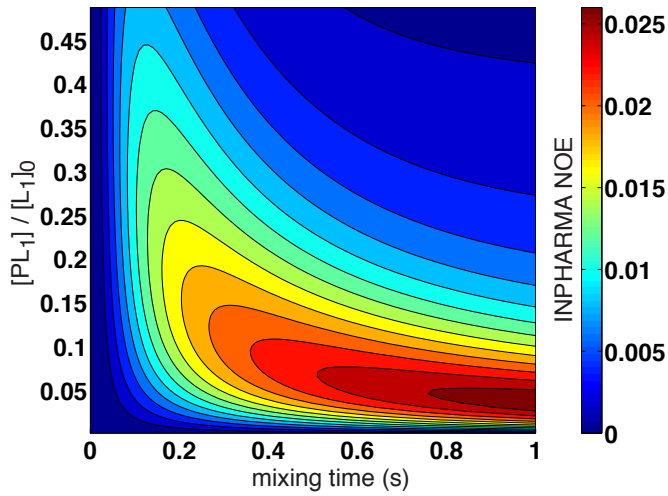



**Figure 4.11:** Contour plot of the dependence of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a cubic receptor of size 73 on the $\tau_m$ (x axis) and on the fraction of bound ligand $[TL_1]/[L_1]_0$ (y axis). The curves were calculated assuming the kinetic model 1. Other parameters are $\tau_{cT} = 20$ ns, $\tau_{cA} = \tau_{cB} = 0.1$ ns, field strength = 800 MHz, d = 2.5 Å, $[L_1]_{tot} = [L_2]_{tot} = 500\mu$M.

In Figure 4.11 we show a contour plot of the intensity of the INPHARMA NOE in dependence of the mixing time (x axis) and the fraction of bound ligand (y axis), defined as $[TL_1]/[L_1]_{tot} = [TL_2]/[L_2]_{tot}$. The ligands total concentration is 500 $\mu$M each. At short mixing times the INPHARMA NOE intensity linearly increases with the fraction of bound ligand. On the other hand, at long mixing times the intensity of the INPHARMA NOE steeply depends on the fraction of bound ligand. For $\tau_m > 600$ ms the optimal value of the INPHARMA NOE is achieved for a fraction of bound ligand ranging from 0.02 to 0.06, whereas at 150 ms bound ligand fractions ranging from 0.1 to 0.35 are required for optimal signal. However, the best results can be obtained at long mixing times using a low fraction of bound ligand (Fig. 4.11).

### 4.3.6 Effect of internal dynamics

Dynamics plays an important role in the NOE transfer mechanisms. Campbell and Sykes evaluated the influence of internal motions of the bound and free ligand on the transferred-NOE effect neglecting the receptor protons(27). Here, we analyse the effect of internal motions on the inter-ligand INPHARMA NOE effect using the model-free approach of Lipari Szabo(28). Clearly in this analysis we need to consider the receptor protons as well, as they mediate the transfer of magnetization between the two ligands. In order to quantitatively evaluate the error made by neglecting internal dynamics in the back-calculation of the INPHARMA NOE from a certain docking model, we performed theoretical simulations of the INPHARMA NOE, using the cubic system of Fig. 4.2B, in presence of internal dynamics of variable amplitude. First we assumed the same order parameter for all protons in all species, both free and bound. Figure 4.12 shows that the efficiency of the INPHARMA transfer decreases in presence of internal dynamics. At low mixing times the decrease of the magnetization transfer efficiency is proportional to $S^4$, as it was found in the analytical solution of Eq. 4.21. Thus the effect of internal motions at low mixing times corresponds to decreasing the correlation time of the species by a factor $S^4$ (Fig. 4.12B). The best choice of the mixing time depends not only on the $\tau_c$ of the complex but also on the presence of internal motions. The optimal compromise between optimizing the intensities of the INPHARMA NOEs and minimizing the error generating from neglecting internal motions is achieved, in our experience, by acquiring the experimental

data at mixing times for which the INPHARMA NOEs have not reached their maximum value, yet.

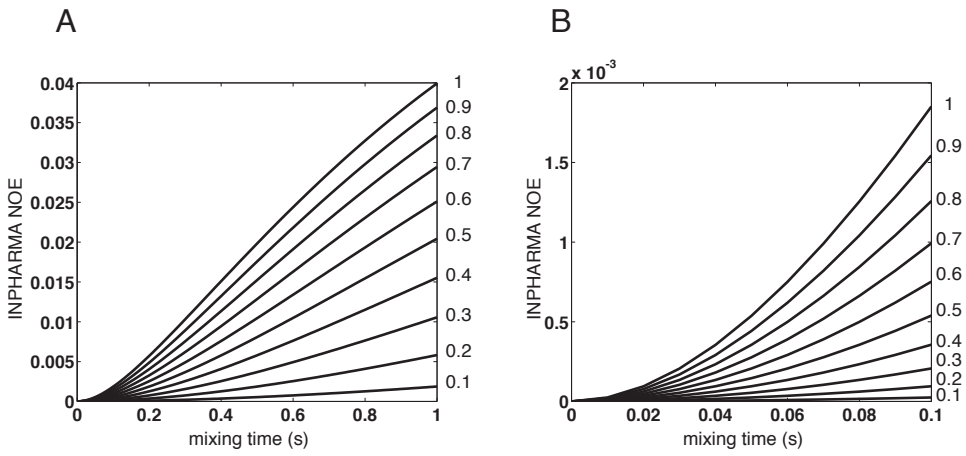



**Figure 4.12:** A) Dependence on the mixing time $\tau_m$ of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a linear receptor consisting of 7 protons as shown in Fig. 4.2A. The dynamic of the protons is variable with an $S^2$ ranging from 0.1 to 1. The curves were calculated assuming the kinetic model 1 with $k_{1off} = k_{2off} =$ 1 kHz. The concentrations of the species are: $[T]_{tot} = 50\mu$M, $[L_1]_{tot} = [L_2]_{tot} = 500\mu$M, $\tau_{cT} = 20$ ns, $\tau_{cA} = \tau_{cB} = 0.1$ ns, field strength = 800 MHz, d = 2.5 Å. B) Zoom of the panel A) for the mixing time between 0s to 0.1s.

In a more realistic system, different internal dynamics occurs for each species. Therefore, we simulated the efficiency of the INPHARMA magnetization transfer using three different order parameters $S^2_L, S^2_T$ and $S^2_{TL}$, for the free ligands protons, the free receptor protons and the receptor protons or ligand protons in the complexes, respectively (Fig. 4.13). Our simulations show that the INPHARMA magnetization transfer is not considerably affected either by the internal dynamics of the free ligands or by that of the free receptor (provided that the concentration of the free receptor is not too high). On the other hand, internal dynamics in the complexes measurably affects the efficiency of the INPHARMA magnetization transfer. For a medium size receptor, the efficiency of magnetization transfer at constant $\tau_m$ decreases upon increasing the amplitude of internal dynamics either for the receptor or the ligands protons in the complex, whereby the effect of receptor protons dynamics is smaller. The reason for this lies in the two counteracting effects of the internal dynamics of the receptor protons: on the one hand,

receptor protons dynamics reduces the spin-diffusion rate necessary for the INPHARMA transfer, thus resulting in smaller INPHARMA NOEs; on the other hand it reduces the rate of diffusion of the magnetization away from the binding pocket into the receptor body, thus improving the efficiency of the INPHARMA magnetization transfer.

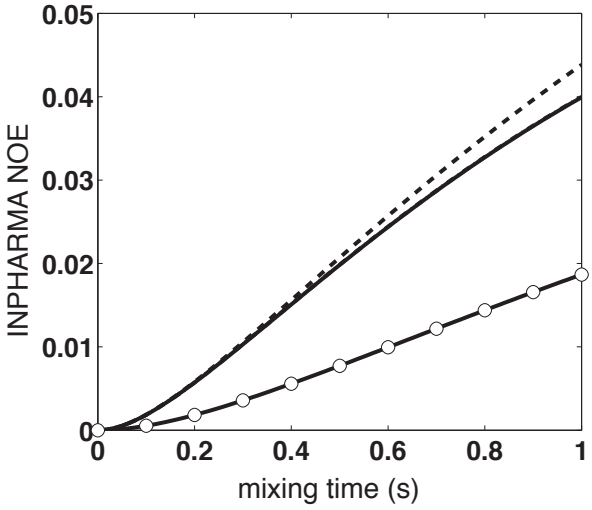



**Figure 4.13:** Dependence on the mixing time $\tau_m$ of the intensity of the INPHARMA NOE between proton H1 of $L_1$ and proton H1 of $L_2$ binding a linear receptor consisting of 7 protons as shown in Fig. 4.2A. Different amount of dynamics are taken into account: 1. No dynamics is present neither for the ligands nor for the receptor: $S^2 = 1$ (full line); 2. The free ligand is dynamic with an $S^2_L = 0.5$ (dashed line); 3. The free receptor is dynamic with an $S^2_T = 0.5$ (overlapped with the full line); 4. Both the ligands and the receptor in the complex show dynamics with $S^2_{TL} = 0.5$ (full line with circles). The curves were calculated assuming the kinetic model 1 with $k_{1off} = k_{2off} = 1$kHz. The concentrations of the species are: $[T]_{tot} = 50\mu$M, $[L_1]_{tot} = [L_2]_{tot} = 500\mu$M, $\tau_{cT} = 20$ ns, $\tau_{cA} = \tau_{cB} = 0.1$ ns, field strength = 800 MHz, d = 2.5 Å.

## 4.4 Conclusions

The INPHARMA method is a flexible, rapid and powerful technique to determine the relative, and in some favorable cases even the absolute, orientations of two ligands binding weakly and competitively to the same binding pocket of a common receptor. This methodology is particularly useful in the

process of optimizing drug leads as it provides a fast method to determine the relative orientation of different chemical skeletons in the receptor binding pocket. Here we provide a theoretical description of the INPHARMA effect and we describe its dependence on both kinetic and experimental variables. The equations and the graphs shown here provide an arsenal of tools to determine the applicability of the method to specific cases and to correctly choose the experimental parameters. We show that the choice of the experimental parameters, as for example the $\tau_m$ of the NOESY experiments, is critical and strictly depends on the system under investigation. Furthermore, we provide an estimation of the errors committed by using approximations, as for example neglecting the effect of internal motions. The analysis presented here helps to optimize the results from INPHARMA measurements.

# Chapter 5

# INPHARMA: Performance of the method

**Abstract**

INPHARMA proved its capability to discriminate between binding poses when the apo-protein structure is available. Here we investigate the scenarios in which the structure of the apo-receptor is not well defined, as it would be the case for low resolution models obtained by homology modeling or experimental methods such as electron microscopy or X-ray crystallography (when only low resolution data are available). The method exhibits high success rates (ranging from 30% to 100%), making INPHARMA a powerful method to resolve small molecule protein complex structure.

## 5.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy has emerged as a valuable tool in drug discovery over the past decades [38, 104]. NMR-based techniques have been developed to identify and subsequently optimize novel binders for a dedicated protein target. Among the different NMR approaches, those relying on the transferred Nuclear Overhauser Effect (tr-NOE) are particularly useful in providing structural information on the bioactive conformation of the ligand [56, 102]. In addition, we recently showed that protein mediated transferred-NOEs between two competitively binding ligands could be used to derive the binding mode of the ligands to their receptor, when the receptor structures are available [79].

However, it might be hard in the near future to gather sufficient experimental information in order to build accurate 3D models for the complete set of possible targets. Indeed, the gap between numbers of known protein sequences and structures has increased. As a comparison, Release of July, 2009 of UniProtKB/TrEMBL [100] contained 8,926,016 nonredundant sequences, while there were 59,790 structures (representing 34,480 nonredundant sequences) in Protein Data Bank (PDB) [7] as of August, 2009, two orders of magnitude fewer. Even if considerable progress has been made in the protein structure prediction, actual computational methods are still far from providing accurate models within an acceptable computing time. In general, for proteins sharing more than 30% sequence identity to their homologous templates, models are typically comparable to low-resolution experimental structures. Furthermore, when the sequence identity drops below 30%, the model accuracy decreases due to alignment errors [42,47,82]. Additionally, the protein apo-structure would not be a good representative of the bound conformation if a conformational rearrangement occurs upon ligand binding. In order to capture these situations, we evaluate the performance of our method on a system where accuracy of the protein target is deliberately degraded.

The system we investigate is the catalytic subunit of cAMP-dependent protein kinase A (PKA) in complex with two inhibitors. The two ligands target the ATP binding site of the PKA and their complex structure can be found in the protein data bank (pdb code: 3DNE and 3DND). All experimental protocols and measurements are described in the Chapter 2.

## 5.2 Protein structure exploration

The INPHARMA NOEs are built within a double exchange of magnetization between ligands and their receptor. The transfer process is mediated through the receptor. Initially, the magnetization is transfered from one ligand to the receptor, then the second transfer from the receptor to the other ligand. The bound-conformation of the target is typically unknown. We evaluate here the influence of the protein receptor conformation on the outcome of the analysis using molecular dynamics and homology modeling.

### 5.2.1 Molecular dynamics

We aim to verify that we obtain a good INPHARMA score even if the protein conformation around the ligands is not well defined. Hence, we sample the conformation of the protein while the ligand is kept fixed in the binding pocket. Molecular dynamics (MD) simulation was performed on each protein-ligand complex starting from the X-ray crystal structure with the software NAMD and the CHARMM force field [10, 11, 64, 81]. The protein was solvated in a 5 Å layer of water molecules, then minimized (Fig. 5.1). The ligands remained rigid during the simulation and fixed as in the X-ray structures. Langevin dynamics were performed for each complex with a 2 fs time step without coupling the hydrogen to the thermal bath (use of the SHAKE algorithm) and with a damping coefficient, $\gamma$, of 5 per picosecond. The water sphere was maintained with a spherical harmonic potential.

After minimization we gradually increased the temperature by increments of 50 K from 300K to 600K. A trajectory with ∼800 structures for each complex was generated. The models are sorted by the RMSD from the X-ray structure for the heavy atoms in the binding pocket and spanned the range of 0 to 5 Å RMSD. Note that the RMSD value of the whole complex is likely to be higher as the side chains on the surface are more affected compared to the buried ones present in the interaction site.

### 5.2.2 Homology modeling

Homology models were generated from 5 different starting structures selected as template from the MD trajectory. These structures have at minimum 1.4 Å heavy atom RMSD with respect to the X-ray structure. All side chains were mutated to alanine in a first place. The 3D models were reconstructed

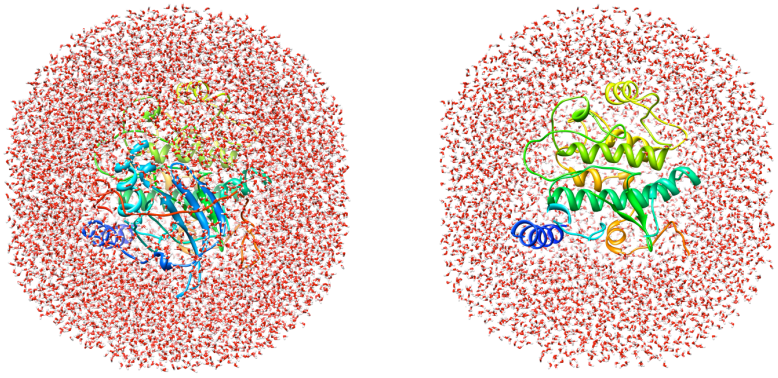


**Figure 5.1:** PKA solvated with a layer of water molecules maintained by a harmonic potential. On the right side of the panel a section of the sphere is shown.

while the ligand was kept in place. In total, 125 models for each complex were provided by Stefan Bartoschek and Peter Monecke, our collaborators at Sanofi-Aventis, in the NMR and molecular modeling department, respectively, in Frankfurt am Main (Fig. C.1). The models are sorted by RMSD values as, in the previous section (Fig. C.2A and Fig. C.2D).

## 5.3 Docking

Using the same approach described in the section "Molecular dynamics", we performed molecular dynamics simulations on just the protein. Whereas the ligands were included in the previous simulations we removed them for this simulation. In order to sample a larger conformational space of the protein, we increased the final temperature to 1200 K starting from 0 K. In Fig. 5.2 we follow the RMSD value of the receptor binding pocket with respect to the starting structure as a function of time dunring the MD simulation. We clearly see that in the last part of the trajectory the RMSD sharply increases. We discarded the last $\sim$ 80 structures, as they were seen to be unfolded. The first 700 structures were retained, in the set for docking analysis. All structures were minimized in water. The two ligands were docked rigidly to each model with the program Surflex [39].

Surflex utilizes an idealized active site ligand (a protomol) as a target to

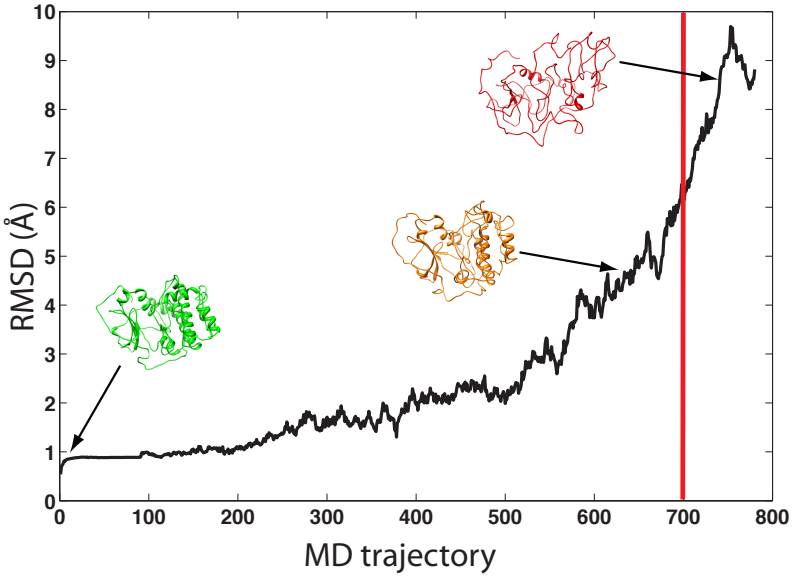



**Figure 5.2:** Binding pocket RMSD values with respect to the X-ray structure along the MD trajectory. The x-axis represents the MD trajectory and each unit corresponds to one model snapshot. The y-axis is the RMSD value in Å. The vertical red line shows the protein models retained in the set for docking ($\sim$ 700). Every structure above this line is discarded because of unfolding. Three structures are depicted on the graph to depict the evolution of the protein folds along the trajectory.

generate conjectural poses of molecules (Fig. C.3A). These hypothetical poses are scored using the Hammerhead scoring function. The following describes the overall procedure. There are two phases in the process.

1. Protomol Generation. An idealized binding site ligand is generated from the protein structure.

1.1. Input: (a) protein structure including hydrogens, (b) list of residues to identify the ligand binding site, used solely to identify residues proximal to the binding site.

1.2. Output: a protomol file that serves as a target to which putative ligands or ligand fragments are aligned on the basis of molecular similarity.

1.3. Procedure: three different types of molecular fragments are placed into the ligand binding site in multiple positions and are optimized for interaction to the protein. High-scoring nonredundant fragments collectively form the

protomol.
2. Docking. Ligands are docked into the protein to optimize the value of the scoring function.
2.1. Input: (a) protein structure, (b) protomol, (c) ligands.
2.2. Output: the optimized poses of docked ligands along with corresponding scores.
2.3. Procedure (for each hypothetical ligand) is the following.
2.3.1. Input ligand is fragmented, resulting in 1 to 10 molecular fragments, each of which may have some rotatable bonds.
2.3.2. Each fragment is aligned to the protomol to yield poses that maximize molecular similarity to the protomol.
2.3.3. The aligned fragments are scored and pruned on the basis of the scoring function and the degree of protein interpenetration.
2.3.4. Procedure used to construct full molecules from the aligned fragments: An incremental construction approach as in Hammerhead.
2.3.5. The best scoring poses are subjected to gradient-based optimization alignment, and the top scoring poses are returned along with their scores.

The 10 best scoring poses were kept for each docking trial, resulting in a total of 7000 poses (700 models · 10 best poses) for each of the two ligands docked to the 700 receptors models from the MD trajectory (Fig. C.3B). A filter based on similarity was applied to exclude similar binding mode found for a single protein model. We assumed that ligands with high similarity (less than 0.5 Å RMSD) would not add significant information to the analysis. Therefore, only one representative ligand was selected (Fig. C.4). Note that similar binding poses in two different protein models are kept as the protein conformation may play a role. The final sets of complexes consist of $\sim$ 4600 and $\sim$ 4700 poses for the two ligands, resulting in more than 21 millions possible combinations.

### 5.3.1 Descriptors

The putative binding poses with respect to the X-ray structures is characterized by descriptors based on quaternions. Unit quaternions provide a convenient mathematical tool for representing orientations and rotations of objects in 3D. Compared to Euler angles they are simpler to use and bypass the problem of gimbal lock. Compared to rotation matrices they are more numerically stable and more efficient in computation.

For each transformation of a rigid body in a 3D space, there exists a unique corresponding quaternion. Briefly, quaternion can be thought of a vector, an angle and a shift. The vector represents the axis of rotation, defines the positive sense of rotation and the direction for the shift. The descriptors are built with these parameters, see appendix C and Eq. C.1. Quaternions allow us to describe unambiguously two types of ligand orientations in the binding pocket: absolute orientation (absolute binding mode) and relative orientation (relative binding mode). The absolute binding mode characterizes the pose with respect to the one found in the crystal structures. On the other hand, the relative binding mode characterizes the accuracy of the orientations of the ligands with respect to each other in comparison to the X-ray structure. Hence if the two ligands are turned by 42° around the z-axis (in the complex reference frame) the absolute binding mode will not be correct, but the relative binding mode will be correct. In other words, if one pose of $L_A^{i\in[0;4600]}$ is characterized by the same quaternion as one pose for the ligand $L_B^{i\in[0;4700]}$, then the pair formed by these two poses will have the correct relative orientation.

## 5.4 Results and discussion

### 5.4.1 Receptor definition : MD and homology modeling

We sampled the protein conformation with molecular dynamics simulations producing two sets of receptor models, one for each ligand. A histogram of the binding pocket RMSD for the ensemble of structures is shown in figure 5.3A and 5.3B. We combined each structures from the two sets to form 64000 pairs ( $800 \cdot 800 = 640000$ pairs) and evaluated the INPHARMA NOEs. The pairs that exhibit a correlation coefficient above 0.85 ($R > 0.85$) between the experimental data and the calculated values, are considered to be positive and shown in the 2D-histogram in Fig. 5.3C . We observe positive results for protein models up to RMSDs of 3-3.5 Å.

Hence, the dependence of the outcome of the analysis appears to be quite insensitive to the quality of the receptor model within 3 Å RMSD. This result is quite encouraging, because it opens the horizon to application of the method to a multitude of poorly defined systems but potentially interesting for drug discovery. This conclusion is further supported by repeating the same analysis on the data sets produced by homology modeling (Fig. C.2).

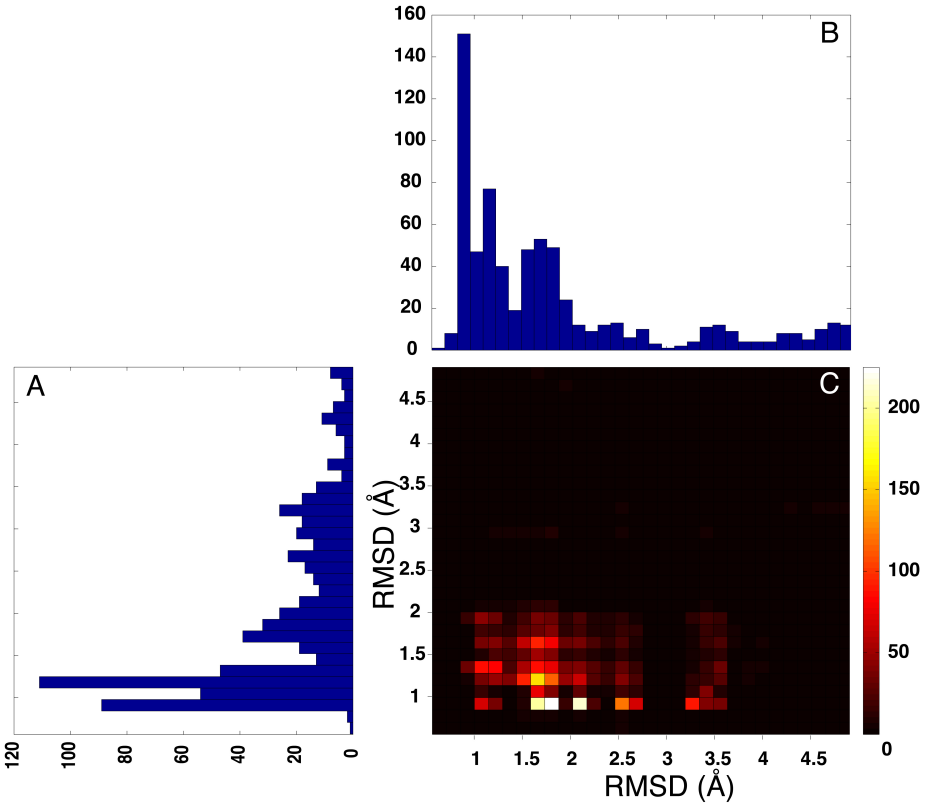



**Figure 5.3:** A), resp. B) Histogram of the binding pocket RMSD (models versus X-ray structure) distribution of the ligand A,resp. B for the models derived by molecular dynamic. The x axis represents the RMSD values in Å and the y axis is the numbers of structures having similar RMSD values. C) 2D-histogram of all possible pairs of models that correlated with $R > 85\%$ to the experimental values (INPHARMA NOEs).

### 5.4.2 Solving a complex structure in a general context

In a real case scenario, the binding mode of the ligands may not be available. Accordingly, we generated all possible docking modes with poorly defined receptors structures and completely unknown binding interactions, except for the binding site. In figure 5.4.2A and 5.4.2B we present the quality of the docking models generated by Surflex for both complexes. We observe that the putative docking models are not exceptionally good, especially for ligand A where almost no native poses are found by the program. Nevertheless, relative binding modes can be correct in many occurrences. We calculated

the INPHARMA NOEs for the ≈ 21 000 000 pairs and examined the accuracy of the method.

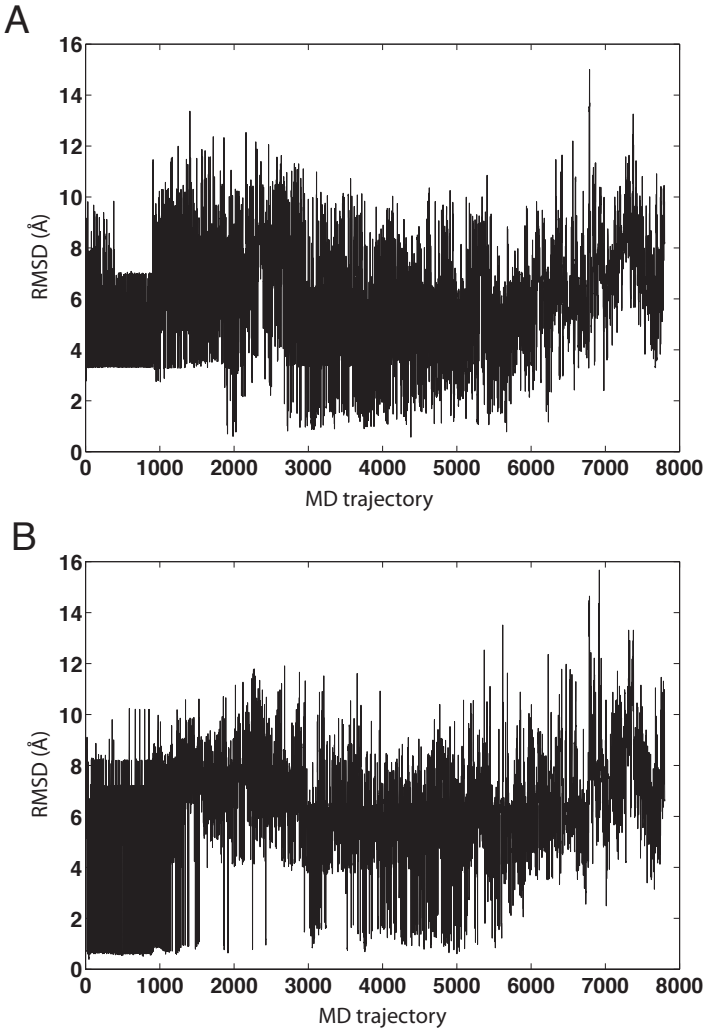



**Figure 5.4:** Representation of the quality of the docking poses measured by the RMSD to the native binding mode found in the crystal structures for ligand A (A) and ligand B (B). The x-axis represents the docked binding poses created by Surflex on each receptor model generated by MD simulation. 10 poses per protein model were kept resulting in a total of 7800 complex structures for each ligand. Note that only the first 7000 complex structures were retained as only the first 700 proteins models were retained in the MD simulation.

We defined the accuracy as the ratio between true positive binding modes over the selected binding modes (selected by the method). The true positives binding modes have a good correlation to the experimental data ($R > 0.85$) and fulfill the weak NOEs conditions (chapter 2), they are considered to represent a correct relative binding mode (Eq. C.1). The selected binding modes are the one passing the experimental conditions only. Therefore a pair of complex structures that pass the experimental condition but do not correspond to a good relative binding mode account for false positive and decreases the accuracy. In other words, if the predictions of the method match completely with relevant binding modes (absolute or relative binding modes) the accuracy is 100%.

Moreover, we present the accuracy of the method as a function of the quality of the receptor model (Fig. 5.5). As expected, the accuracy is very good when the receptor structure is reasonably well defined (RMSD $\leq$ 1 Å) and score about 90%. For less resolved protein structures, the accuracy drops to 30% and stays approximately constant over the rest of the RMSD range.

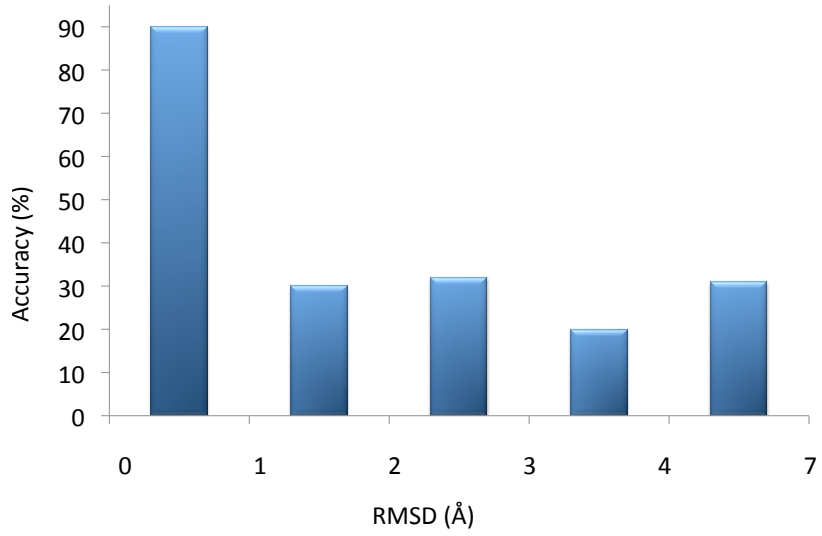



**Figure 5.5:** Accuracy of the INPHARMA predictions as a function of the receptor structure quality. The x axis represents the binding pocket RMSD (receptors models versus X-ray structure) in Å. The y axis is the Accuracy in % of the INPHARMA method calculated for the sets of complex structures within a certain receptor RMSD range defined on the x axis.

This result is very encouraging and has to be compared to the accuracy of purely *in silico* method that reaches $\approx 5\%$ in favorable cases. Moreover, buried binding pockets like the ATP binding pocket are unfavorable cases for *in silico* methods [26, 27]. This might explain why the docking software we used, had difficulties in predicting correct binding poses (Fig. 5.4.2). Taken altogether, the scores of the method are motivating, being substantially superior.

The search for protein-binding ligands is a crucial step in the inhibitor design process. Fragment screening represents an interesting method to rapidly find lead molecules, as it enables the exploration of a larger portion of the chemical space with a smaller number of compounds as compared to screening based on drug-sized molecules.

It is common in drug design that only one ligand interaction mode is unknown. This could be the case for lead optimization or fragment screening. In this particular context, we need to determine only one binding mode and the difficulty is greatly reduced. Due to the principle of the method, we still have to measure NOESY experiments with two ligands in presence, but one of the two ligand-protein complex structure should be known. In this case, that any of the two ligand binding mode is known, the predictions of the method are 100% accurate for a well defined receptor, and highly accurate, (66% and 100%), for a less resolved target structure (Fig.C.5A and Fig.C.5B). The accuracy of the method in this situation is excellent and ensure the reliability of the outcomes of the method.

We showed that the success rate of the methodology exceeds the one proposed by computational solutions with the same problematic. In the context of structure based drug design, INPHARMA appears to be a method of choice. Its ability to predict the correct binding mode of small molecules to target receptor with high accuracy and without specific labeling scheme in a short time, prompt the method among the fast and reliable NMR tools.

# Appendices

# Appendix A

The quality of the fitting between the experimental and the back-calculated interligand NOEs for different models' pairs was judged by the linear regression coefficient $R^2$ and by the quality factor $Q^2$:

$$Q^2 = \sum \frac{(I_{exp} - a \cdot I_{calc})^2}{I_{exp}^2} \tag{A.1}$$

where a is the slope of the linear regression $y = ax$. In the case of PKA, the first selection criteria chooses those pairs with $R^2 > 0.9 = R^2_{min}$ or $Q^2 < 0.15 = Q^2_{min}$ out the 16 docking models' pairs tested in Table A.1. The values of $R^2_{min}$ and $Q^2_{min}$ should be adjusted for each experimental dataset. In general, we consider $R^2_{min} = 0.9$ as a reasonable choice. However, if the experimental data are recorded using too long mixing times (large effect of internal motions) or the structure of the apoprotein binding pocket is far different from the one in the complex, the linear correlation coefficient can drop below 0.9 also for the correct binding poses. In those cases a lower threshold for $R^2_{min}$ should be used. A selection of the best model on the basis of small differences ( $\approx 0.05$) in the overall correlation coefficient between experimental and calculated INPHARMA NOEs is not recommended, as it could lead to wrong results. Internal motions, which cannot be taken into account theoretically due to the lack of a suitable model both for the protein side-chains and for the ligands in the binding pocket, hamper the quality of the overall fit, thus making it unsafe to rely on fine differences in the correlation coefficient ($\Delta R^2 \approx 0.05$).

**Effect of internal motions on the INPHARMA NOEs**

Due to the high non-linearity of the system (Eq. **??**), it is difficult to predict how the interligand NOEs are affected by different kinetic, thermodynamic and structural parameters. Internal dynamics at specific sites is an important parameter affecting magnetization transfer. In the case of spindiffusion-mediated interligand NOEs, the efficiency of magnetization transfer is influenced by the dynamics of both the ligand(s) and the protein side-chains. If the same order parameter $S$ is assumed for both the protein side-chains and the ligands protons, the interligand NOEs scale with $S^4$ at very short mixing times ($\tau_m \approx 1 - 10$ ms), while the intra-ligand NOEs scale with $S^2$. Thus the effect of internal dynamics is larger on the interligand than on the intraligand NOEs. At larger mixing times the value of the interligand NOEs can be predicted only by solving Eq. 4.1 under consideration of the internal dynamics at each proton site. Due to the lack of a suitable model to describe internal motions at each proton site in the complex, the effect of internal dynamics is neglected in the simulations. As a result of this, the slope of the linear regression between the experimental and the back-calculated interligand NOEs deviates from 1. In addition to the large divergence of the slope of the linear correlation from 1, the presence of internal motions can result in a deviation of the experimental data from the theoretical values ($R^2 < 1$), due to different amount of internal motions at the various protein and ligand sites in the complexes. At long mixing times and in presence of large internal motions, the regression curve will considerably deviate from linearity, even for the correct binding models. Therefore, it is advisable to record several NOESY spectra at different mixing times and to use in the analysis the experimental data stemming from the shortest mixing times for which a reasonable number of interligand NOEs are observed. The presence of internal motions in the PKA/ligand complexes is confirmed by the value of the slope of the linear correlation between calculated and experimental INPHARMA NOEs. In absence of internal motions the slope should be close to 1; however, we observe a value of 0.3 (Table A.1 and Fig. A.5), indicating a considerable amount of internal fluctuations either of the protein sidechains or of the ligands in the binding pocket, or of both. The influence of internal motions on the slope of the linear correlation curve between experimental and back-calculated INPHARMA NOEs depends on the size of the complex, on the nature of the motions and on the mixing times of the NOESY experiments,

as indicated by theoretical simulations (Fig. A.6). In Fig. A.6 we show the value of $I_{rel}$ in dependence of the order parameter $S$ for two mixing times and two protein sizes. The simulations are run on an artificial system: the ligand is represented by 3 protons in a line at a distance of 2.5 Å from one another, while the protein is represented by a cube of dimension 7x7x7 equidistant protons (2.5 Å). In the complex the ligand is placed in the middle of the cube (Fig. A.6 a). Interligand NOEs are calculated between the inner proton of $L_A$ (lattice position 3,3,3) and the middle proton of $L_B$ (lattice position 2,3,3 ). A homogeneous order parameter $S$ is assumed at each location on the protein and the ligand. $I_{rel}$ is defined as:

$$I_{rel}(S) = \frac{I_{inter-NOE}(S) \cdot I_{intra-NOE}(S=1)}{I_{inter-NOE}(S=1) \cdot I_{intra-NOE}(S)} \tag{A.2}$$

and is a measure of the ratio between the interligand NOE, normalized to the intraligand NOE value, for an order parameter $S \neq 1$ and the interligand NOE expected for $S = 1$, namely in absence of internal motions. The normalization to the intraligand NOE is necessary as in the process of correlating theoretical and experimental data we tune the simulation parameters, such as the protein $\tau_c$ or the protein concentration, to obtain a slope of 1 for the intraligand NOEs. If $I_{rel} = 1$, the NOE transfer is as efficient as in absence of motions. In this case the slope of the correlation line between the theoretical NOEs, calculated assuming no internal motions, and the experimental ones, is equal to 1. On the other hand, if $I_{rel}(S) = \mathrm{x}$, the efficiency of the NOE transfer for that specific $S$ value is x times that for $S$=1 and the slope of the correlation line between the theoretical NOEs, calculated assuming no internal motions, and the experimental ones, is equal to x. For a protein of the size of PKA (Fig. A.6b, squared symbols), the $I_{rel}$ value deviates from 1 for all order parameters $< 0.95$ at both 200 (green curve) and 600 (blue curve) ms. At 600 ms and for $S = 0.6$, $I_{rel}$ is equal to 0.3. Thus, the slope of 0.3 observed in the case of PKA (Fig. A.5 and Table A.1) can be explained assuming an order parameter $S$=0.6. This value should be considered as indicative but not quantitatively correct, as the exact value depends on the geometry of the binding pocket. Moreover, in this simplified system we assume the same order parameter for the inter- and intra ligand vectors, while in reality the two order parameters are likely to be different. For very large proteins (Fig. A.6b, round symbols ), $I_{rel}$ does not deviate from 1 in a range $0.6 <$

$S < 1$ at both 200 (green curve) and 600 (blue curve) ms. This reflects the fact that in extreme spin-diffusion conditions, NOEs reach the same plateau value, independently of the correlation time or of the internal motions. Such extreme spin-diffusion conditions can be reached for large systems before the magnetization relaxes back to equilibrium. This is in accord with the slope of 1 observed for the linear correlation between experimental and theoretical NOEs for the tubulin/epothilone and tubulin/baccatin complexes studied in ref. 7. An order parameter $S$ = 0.6-0.8 ($S^2$ = 0.36-0.64) is not surprising and falls well in the range of the order parameters for proton-proton vectors determined before [46, 90]. Low order parameters can be expected for the INPHARMA NOEs, as the magnetization transfer involves mainly protons from the protein side-chains, and in particular from the methyl groups, which are the closest neighbors to the ligand. Schleicher and Wijmenga [90] reported order parameters $S^2 < 0.6$ for long range NOEs including flexible side-chains. The order parameter $S^2$ for the NOE between methyl group protons and a proton of the environment can be as low as 0.6 [46], assuming fast rotation of the methyl groups. This order parameter can be further decreased by additional motion of the side-chain bearing the methyl group or of the ligand proton [91].

| $L_A$ | $L_B$ | $I_{NOE}$ | | | |
|---|---|---|---|---|---|
| | | $\tau_m$=300 ms | 450 ms | 600 ms | 750 ms |
| H8 | H1,H3,H5 | - | 0,0046 | 0,0096 | 0,0091 |
| H3,H4 | H1,H3,H5 | 0,0035 | 0,0026 | 0,0082 | 0,0067 |
| H1,H2 | H1,H3,H5 | 0,0031 | 0,0016 | 0,0059 | 0,0055 |
| H5 | H1,H3,H5 | 0,0029 | 0,0042 | 0,0080 | 0,0075 |
| H3,H4 | H6,H7 | 0,0016 | 0,0017 | 0,0061 | 0,0047 |
| H1,H2 | H6,H7 | 0,0011 | 0,0012 | 0,0028 | 0,0036 |
| H5 | H6,H7 | - | 0,0013 | 0,0048 | 0,0045 |

**Table A.1:** Normalized INPHARMA NOEs used to calculate the correlation graphs of Fig. 2.4. The INPHARMA NOE intensities INOE were normalized by the diagonal peak with equal $\omega_2$ frequency at the lowest mixing time. The experimental conditions are described in the Experimental Section.

| $L_A$ | $L_B$ | $R^2$ | $Q^2$ | $a$ |
|---|---|---|---|---|
| Pose 1 | Pose 1 | 0.93 | 0.06 | 0.32 |
| Pose 2 | Pose 1 | 0.91 | 0.11 | 0.27 |
| Pose 3 | Pose 1 | 0.96 | 0.04 | 0.29 |
| Pose 4 | Pose 1 | 0.91 | 0.09 | 0.31 |
| Pose 1 | Pose 2 | 0.64 | 0.65 | 0.07 |
| Pose 2 | Pose 2 | 0.73 | 0.43 | 0.14 |
| Pose 3 | Pose 2 | 0.58 | 0.86 | 0.02 |
| Pose 4 | Pose 2 | 0.70 | 0.48 | 0.08 |
| Pose 1 | Pose 3 | 0.82 | 0.29 | 0.14 |
| Pose 2 | Pose 3 | 0.59 | 0.70 | 0.05 |
| Pose 3 | Pose 3 | 0.26 | 0.88 | 0.00 |
| Pose 4 | Pose 3 | 0.34 | 0.98 | 0.00 |
| Pose 1 | Pose 4 | 0.72 | 0.52 | 0.08 |
| Pose 2 | Pose 4 | 0.71 | 0.51 | 0.10 |
| Pose 3 | Pose 4 | 0.59 | 1.16 | -0.02 |
| Pose 4 | Pose 4 | 0.70 | 0.51 | 0.07 |

**Table A.2:** Evaluation of the back-calculated INPHARMA NOEs for the sixteen combinations of the docking poses of $L_A$ and $L_B$, as in Fig. 2.3, with respect to the experimental data. $R^2$ is the Pearson correlation coefficient calculated for a linear regression to the linear function $y = ax$. A value close to 1 represents a good quality fit. $Q^2$ is the quality factor. It measures the deviation of the data to the best-fit line. A value close to zero reflects a good quality of the fit. $a$ is the slope of the linear regression between experimental and back-calculated INPHARMA NOEs.

| $L_A$ | $L_B$ | $R^2$ | $L_A$ | $L_B$ | $R^2$ | $L_A$ | $L_B$ | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| | MD1 | | | MD3 | | | MD5 | |
| Pose 1 | Pose 1 | 0.95 | Pose 1 | Pose 1 | 0.83 | Pose 2 | Pose 2 | 0.84 |
| Pose 4 | Pose 4 | 0.81 | Pose 4 | Pose 2 | 0.81 | Pose 4 | Pose 4 | 0.80 |
| Pose 1 | Pose 3 | 0.73 | Pose 1 | Pose 4 | 0.77 | Pose 2 | Pose 4 | 0.80 |
| Pose 4 | Pose 2 | 0.67 | Pose 2 | Pose 2 | 0.66 | Pose 1 | Pose 4 | 0.80 |
| Pose 3 | Pose 1 | 0.66 | Pose 1 | Pose 3 | 0.65 | Pose 4 | Pose 2 | 0.79 |
| Pose 2 | Pose 4 | 0.64 | Pose 2 | Pose 4 | 0.57 | Pose 1 | Pose 1 | 0.77 |
| Pose 3 | Pose 4 | 0.57 | Pose 1 | Pose 2 | 0.56 | Pose 3 | Pose 4 | 0.73 |
| Pose 4 | Pose 1 | 0.56 | Pose 4 | Pose 4 | 0.54 | Pose 2 | Pose 1 | 0.66 |
| Pose 1 | Pose 4 | 0.55 | Pose 3 | Pose 2 | 0.52 | Pose 1 | Pose 2 | 0.64 |
| Pose 2 | Pose 2 | 0.52 | Pose 3 | Pose 1 | 0.50 | Pose 3 | Pose 2 | 0.61 |
| Pose 3 | Pose 2 | 0.42 | Pose 3 | Pose 4 | 0.49 | Pose 1 | Pose 3 | 0.60 |
| Pose 1 | Pose 2 | 0.39 | Pose 2 | Pose 1 | 0.43 | Pose 3 | Pose 1 | 0.59 |
| Pose 2 | Pose 1 | 0.39 | Pose 4 | Pose 1 | 0.30 | Pose 4 | Pose 1 | 0.52 |
| Pose 2 | Pose 3 | 0.31 | Pose 2 | Pose 3 | 0.26 | Pose 2 | Pose 3 | 0.20 |
| Pose 3 | Pose 3 | 0.21 | Pose 3 | Pose 3 | 0.18 | Pose 3 | Pose 3 | 0.16 |
| Pose 4 | Pose 3 | 0.20 | Pose 4 | Pose 3 | 0.16 | Pose 4 | Pose 3 | 0.15 |
| | MD2 | | | MD4 | | | | |
| Pose 1 | Pose 1 | 0.90 | Pose 1 | Pose 1 | 0.79 | | | |
| Pose 4 | Pose 2 | 0.72 | Pose 4 | Pose 2 | 0.79 | | | |
| Pose 3 | Pose 1 | 0.67 | Pose 1 | Pose 4 | 0.74 | | | |
| Pose 1 | Pose 4 | 0.67 | Pose 2 | Pose 2 | 0.72 | | | |
| Pose 4 | Pose 1 | 0.63 | Pose 1 | Pose 3 | 0.69 | | | |
| Pose 4 | Pose 4 | 0.60 | Pose 2 | Pose 4 | 0.62 | | | |
| Pose 1 | Pose 3 | 0.59 | Pose 4 | Pose 4 | 0.62 | | | |
| Pose 2 | Pose 1 | 0.58 | Pose 3 | Pose 4 | 0.60 | | | |
| Pose 2 | Pose 2 | 0.52 | Pose 3 | Pose 1 | 0.59 | | | |
| Pose 3 | Pose 4 | 0.45 | Pose 3 | Pose 2 | 0.59 | | | |
| Pose 1 | Pose 2 | 0.43 | Pose 2 | Pose 1 | 0.58 | | | |
| Pose 2 | Pose 4 | 0.42 | Pose 1 | Pose 2 | 0.58 | | | |
| Pose 3 | Pose 2 | 0.40 | Pose 4 | Pose 1 | 0.42 | | | |
| Pose 2 | Pose 3 | 0.38 | Pose 2 | Pose 3 | 0.38 | | | |
| Pose 3 | Pose 3 | 0.19 | Pose 3 | Pose 3 | 0.22 | | | |
| Pose 4 | Pose 3 | 0.18 | Pose 4 | Pose 3 | 0.21 | | | |

**Table A.3:** Evaluation of the back-calculated INPHARMA NOEs with respect to the experimental data for the sixteen combinations of the binding poses of $L_A$ and $L_B$ docked to the five different models of the PKA structure (MD1-MD5) generated by molecular dynamics (Fig. A.7).

| | PKA/$L_A$ | PKA/$L_B$ |
|---|---|---|
| Data collection | | |
| Space group | P212121 | P212121 |
| Cell dimensions | | |
| a, b, c (Å) | 72.6, 75,9, 79.9 | 73.0, 77.6, 80.1 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 2.01 | 2.26 |
| Rsym | 5.3(33.1) | 9.1(40.9) |
| I/s I | 21.6(5.37) | 13.1(4.2) |
| Completeness (%) | 99.4(98.9) | 99.9(100) |
| Redundancy | 5.4(5.3) | 5.4(5.5) |
| Refinement | | |
| Resolution | 2.01 | 2.26 |
| No. reflections (work/free) | 27786/2100 | 202256/1521 |
| Rwork/Rfree | 0.215/0.249 | 0.189/0.253 |
| No. atoms | | 3325 |
| Protein | 2793 | 2793 |
| Peptide | 157 | 157 |
| Water | 314 | 264 |
| Ligand | 15 | 13 |
| B-factors | | |
| Protein | 29.6 | 30.6 |
| Peptide | 29.8 | 27.3 |
| Water | 45.8 | 36.6 |
| Ligand | 25.5 | 49.4 |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.006 | 0.021 |
| Bond angles (°) | 1.21 | 1.85 |
| Ramachandran analysis (%) | | |
| favoured regions | 96.3 | 94.5 |
| allowed regions | 3.1 | 4.9 |
| outlier | 0.6 | 0.6 |

**Table A.4:** Data collection and refinement statistics for the crystallographic structures. Highest resolution shell is shown in parenthesis.

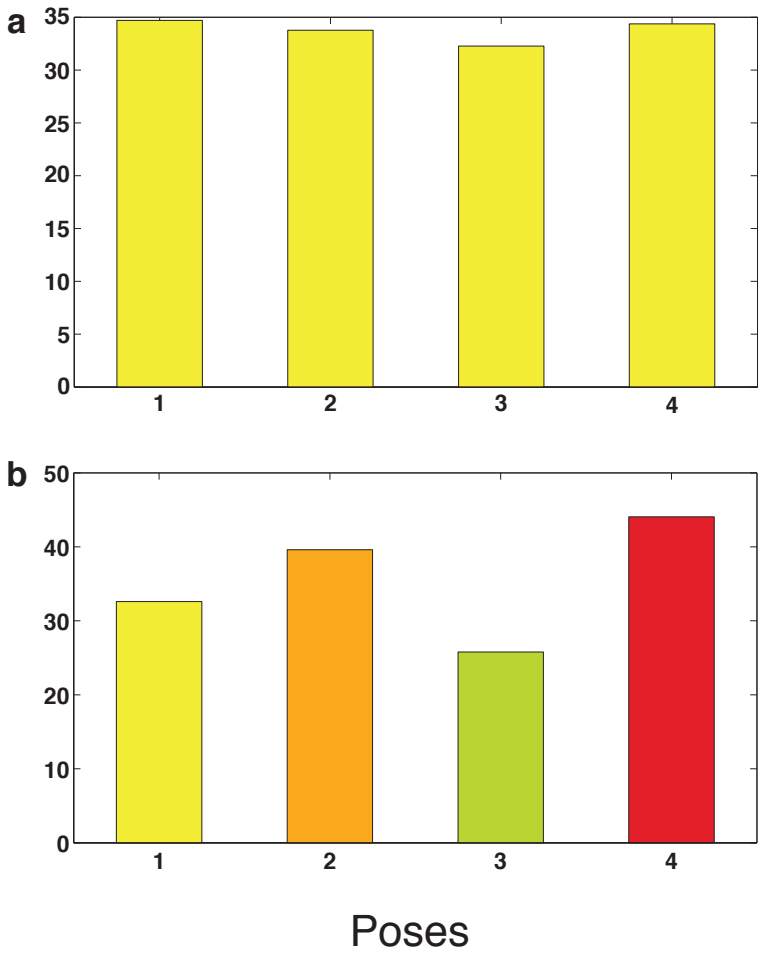



**Figure A.1:** Schematic representation of the proton environment of docked binding modes of the ligand $L_A$ (panel (a) and $L_B$ (panel (b). The binding poses follow the numbering of Fig. 2.3. The proton environment D is evaluated by:

$$D = \sum_{H_i \in ligand} \left( \sum_{\substack{H_j \in protein \\ d(H_i,H_j) < 5\mathring{\mathrm{A}}}} d_{i,j}^{-1} \right) \tag{A.3}$$

While the coefficient D conspicuously differs in the four docking modes of

$L_B$, it remains almost constant throughout the four docking models of $L_A$, indicating that the average proton environment is more independent of the docking pose for the flat $L_A$ than for the bent $L_B$.

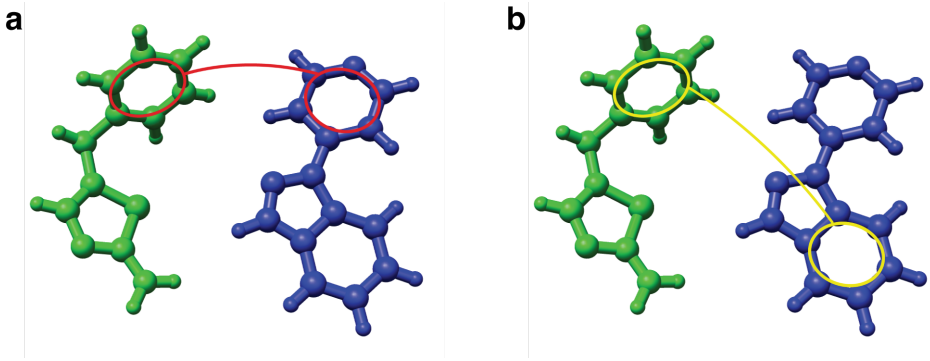



**Figure A.2:** Schematic representation of the color-coding for the INPHARMA NOEs of panel (c) and (d) of Fig. 2.4. (a) NOEs from the phenyl ring of $L_B$ to the pyridine ring of $L_A$ are color-coded in yellow; (b) NOEs from the phenyl ring of $L_B$ to the 3-(4-Pyridyl)indazole ring of $L_A$ are colorcoded in red.

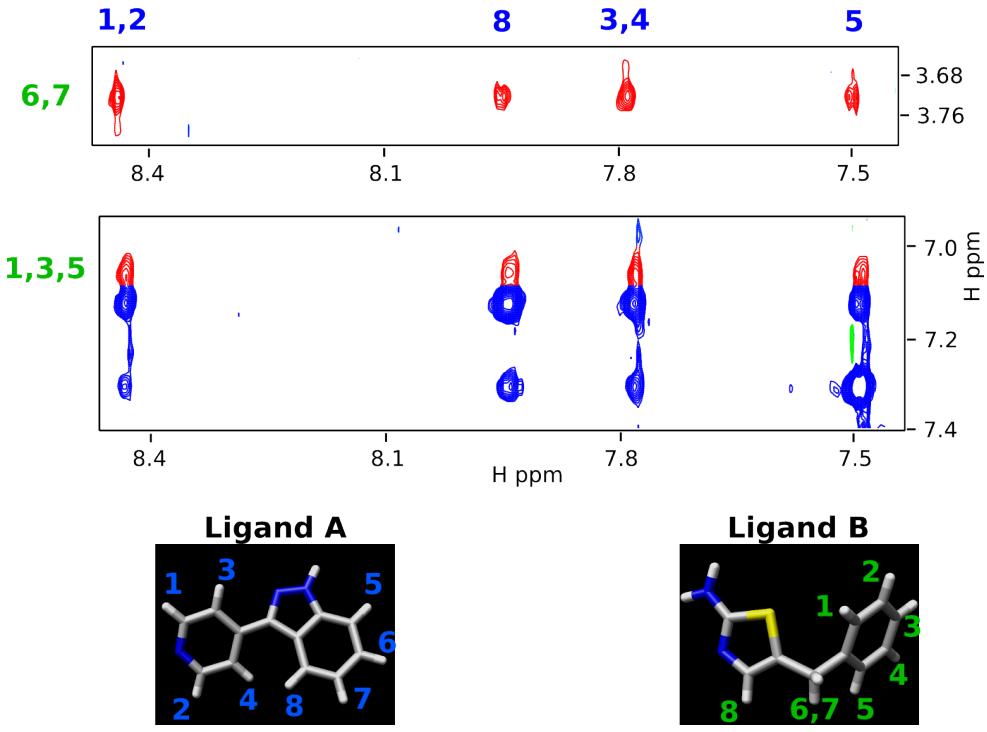



**Figure A.3:** Regions of a NOESY spectrum, acquired at a 900 MHz spectrometer with $\tau_m$ = 600 ms. The sample contained: $[L_A]$ = 150 mM $[L_B]$ = 450 mM, [PKA] = 25 mM. NOEs between H6 and H7 of $L_B$ and H5 and H8 of $L_A$ are visible in this spectrum.

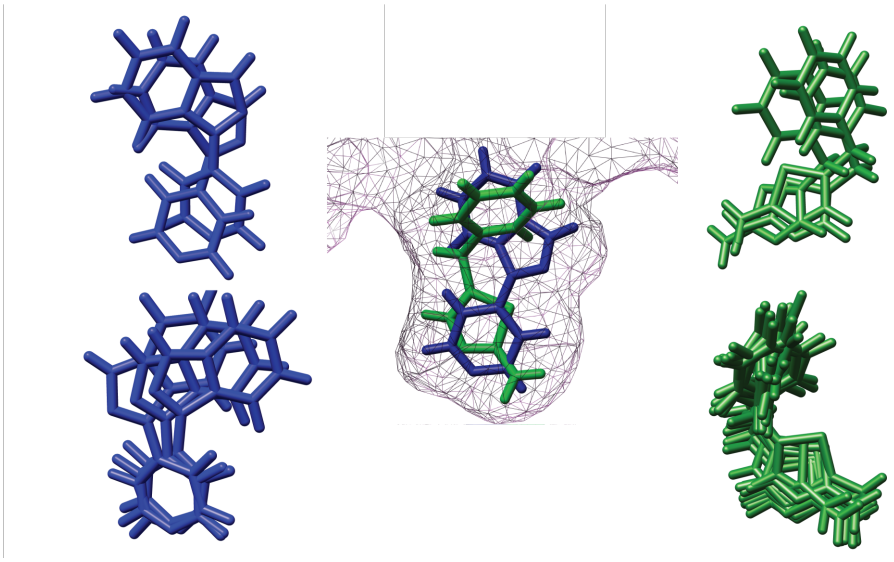


**Figure A.4:** Representation of the additional docking modes, generated at Sanofi Aventis, for which the INPHARMA NOEs were back-calculated and fitted to the experimental ones. In the central panel the binding modes of $L_A$ and $L_B$ observed in the crystal structures are represented. The lower and upper left panels represent the docked orientation of $L_A$, with the upper ones being close to the crystal structure and the lower ones differing from the crystal structure by a rotation $< 180^\circ$ around the z axis. The conformation of $L_A$ differs from the bioactive one in both panels by a rotation of the 3-(4-Pyridyl)indazole ring with respect to the pyridine ring $< 50^\circ$. The two upper and lower right panels represent the docked orientation of $L_B$, with the upper ones differing from the crystal structure by a rotation $< 180^\circ$ around the z axis and the lower ones being close to the crystal structure. The conformation of $L_B$ differs from the bioactive one in both panels by a rotation of the phenyl ring with respect to the thiazole ring $< 50^\circ$.

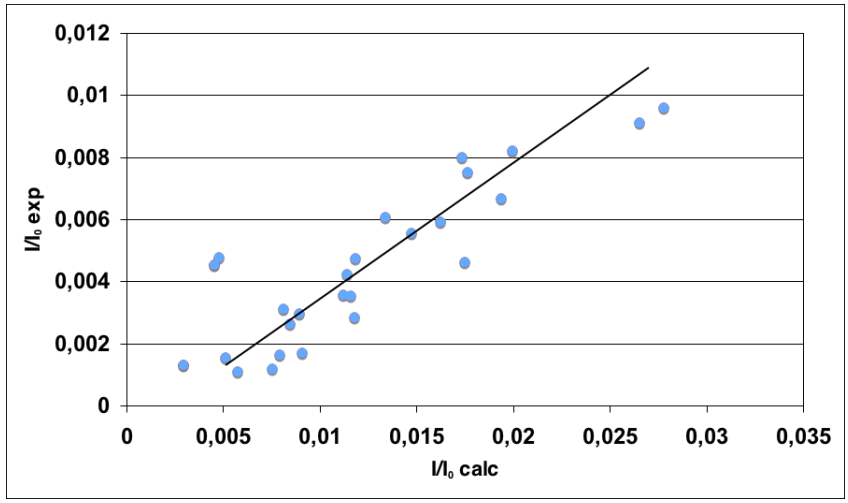



**Figure A.5:** Correlation between the experimental INPHARMA NOEs, normalized to the diagonal of the spin in $\omega_2$, and the normalized INPHARMA NOEs calculated for the two crystal structures of the PKA/$L_A$ and PKA/$L_B$ complexes.

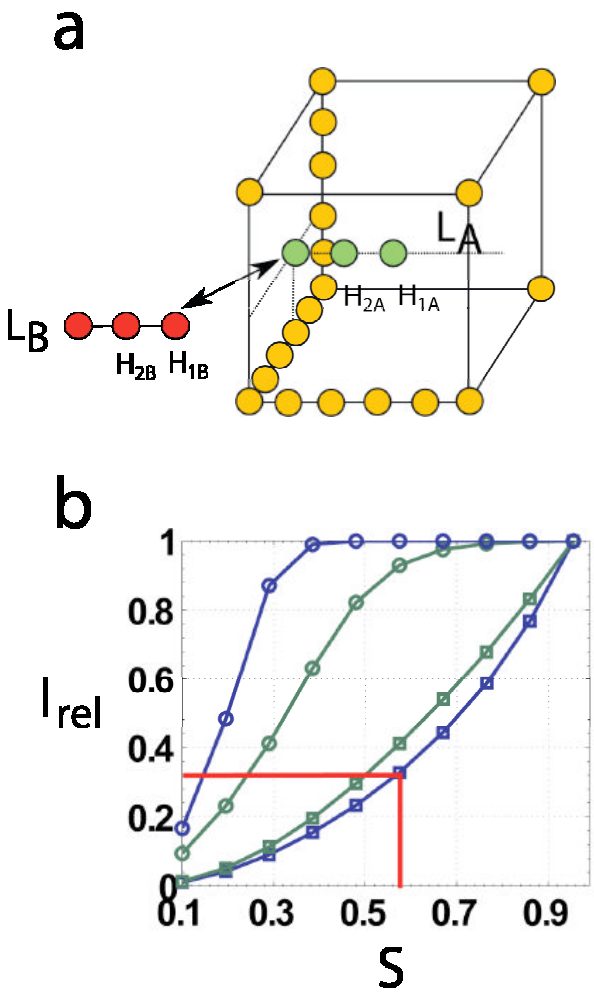
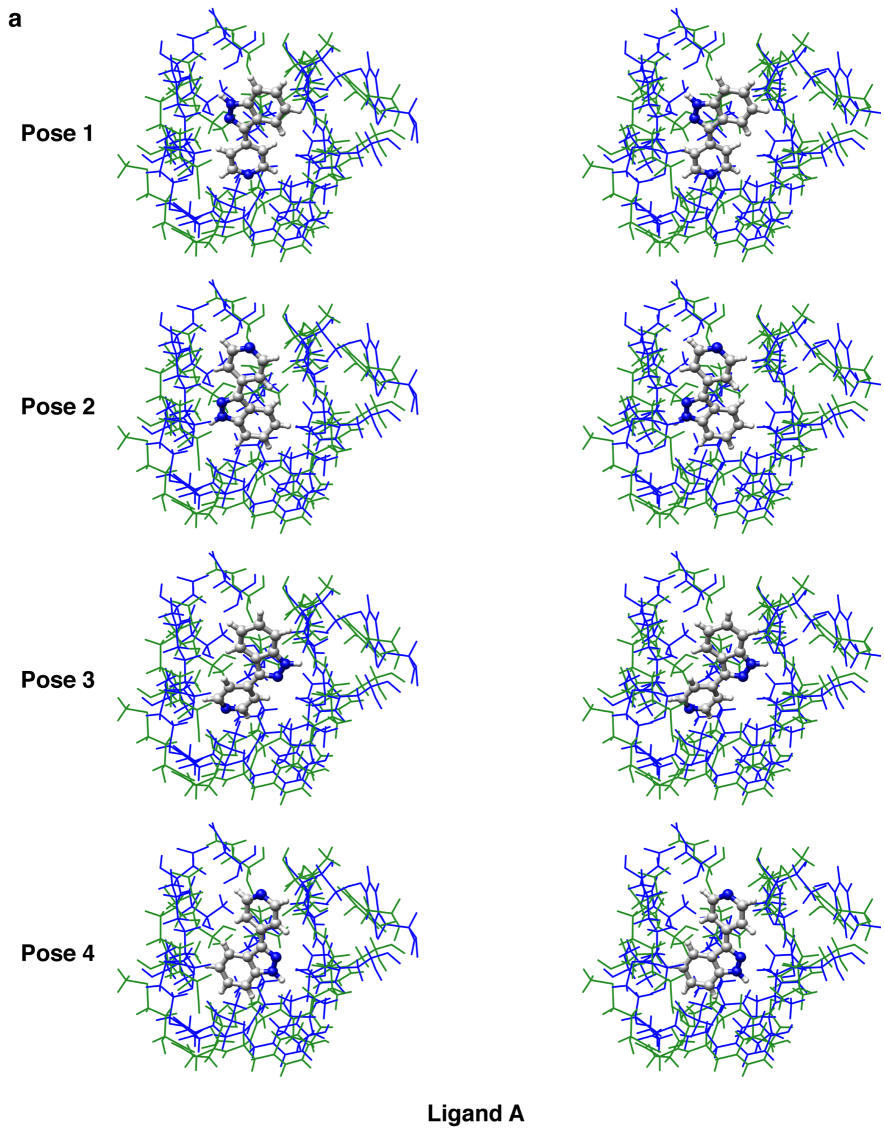
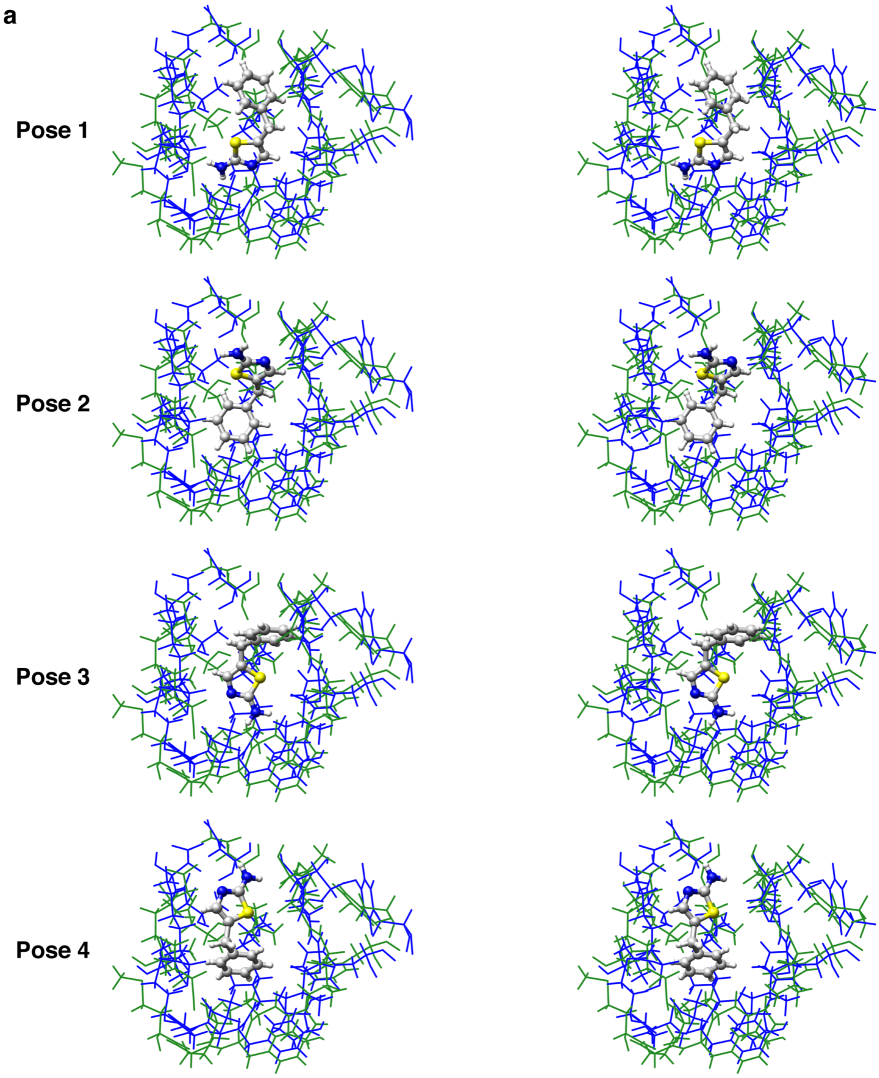
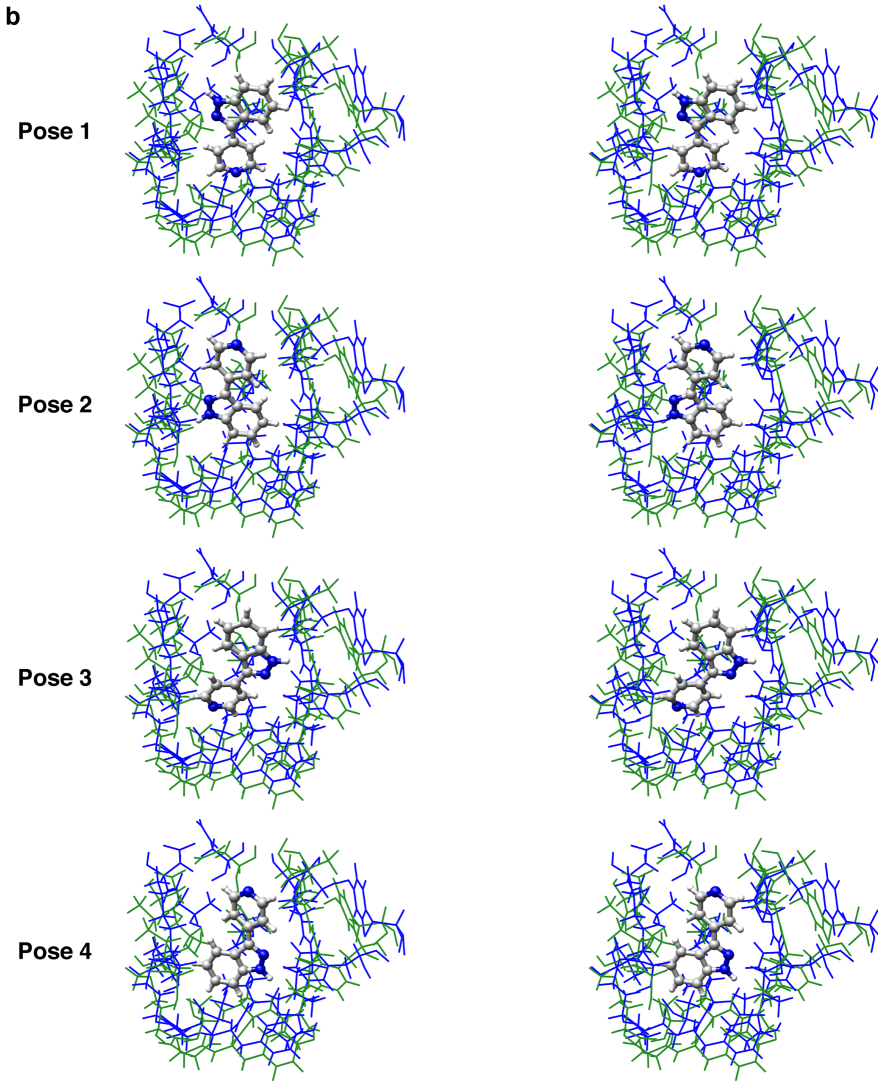


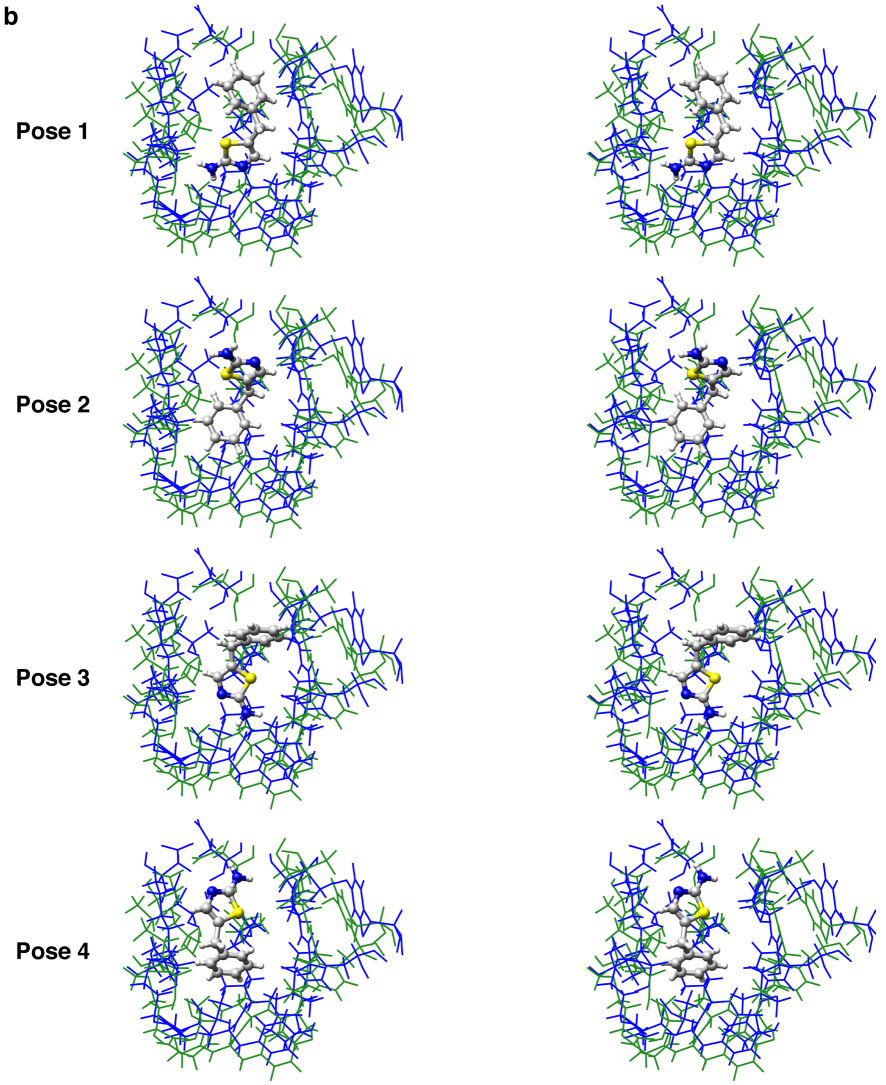

**Figure A.6:** (a) Schematic representation of the model system used to investigate the effect of internal motions on the INPHARMA NOEs. The protein is represented by a cube of dimensions 7x7x7 while the ligands $L_A$ and $L_B$ are linear and contain 3 protons each. All protons of the protein and of the ligand are at a distance of 2.5 Å from each other. When $L_A$ binds to the protein, H1A is in the middle of the protein cube. (b) $I_{rel}$ for the INPHARMA NOE between H1A and H2B normalized to the intraligand NOE between H1A and H2A.Green, $\tau_m$ = 200ms; blue, $\tau_m$ = 600 ms; the squared symbol are for a model system with parameters equal to the PKA system ($[L_A]$ = 150 mM $[L_B]$ = 450 mM, [PKA] = 30 mM, $\tau_c$(PKA) = 20 ns ), while the round symbols are for a model system with parameters equal to the tubulin system ($[L_A]$ = 500 mM $[L_B]$ = 500 mM, [PKA] = 50 mM, $\tau_c$(Tub) = 1300 ns). The red lines indicate a slope of the correlation between theoretical and experimental INPHARMA NOEs equal to 0.3 for mixing times around 600 ms.

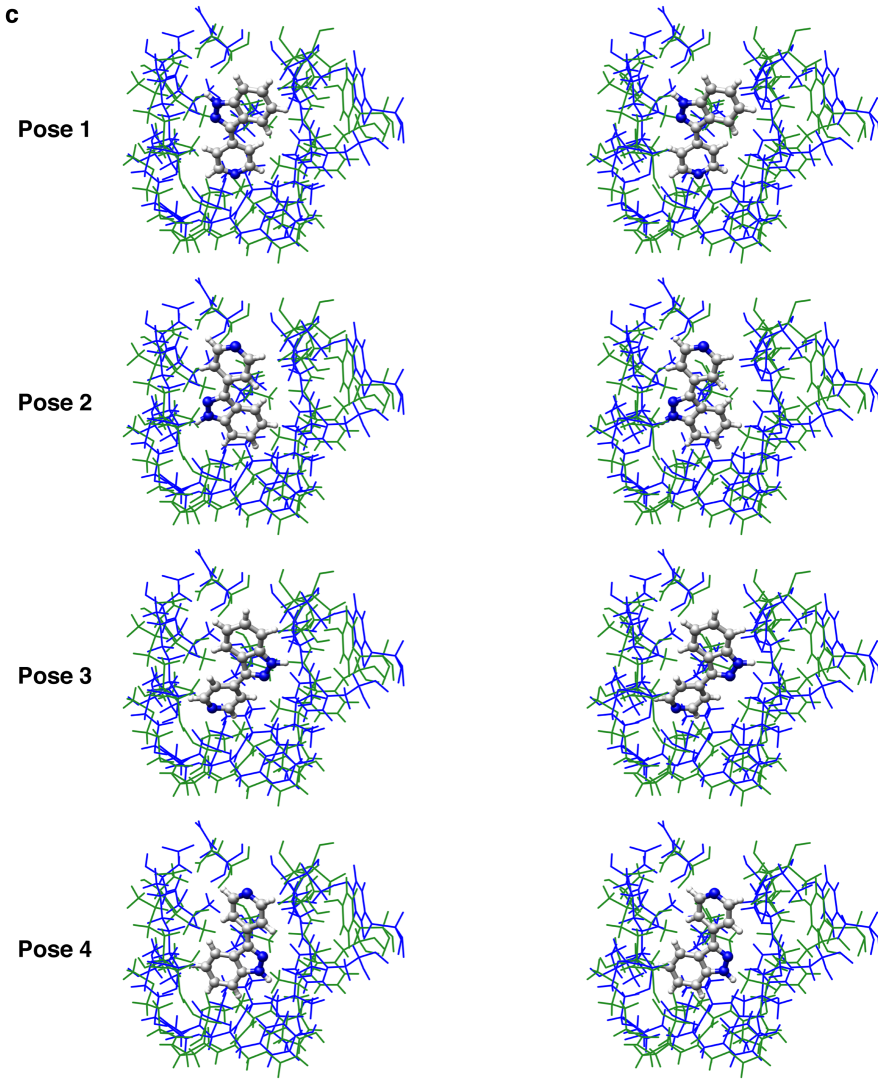
**Figure A.7:** In each panel, one snapshot of a Molecular Dynamics simulation of PKA is shown in green overlapped with the protein X-ray structure in blue. To each of the five snapshot structures the two ligands were docked in the four orientations of Fig. 2.3, to generate a total of 16 models pairs per snapshot structure. The r.m.s.d of the binding pocket heavy atoms to the X-ray structure is 1.9 Å for panel (a), 1.7 Å for panel (b), 1.9 Å for panel (c), 1.6 Å for panel (d) and 1.9 Å for panel (e). The ligand orientations in the first row correspond to the X-ray like orientations. Comparison of the INPHARMA NOEs calculated for all 16 models' pairs for each snapshot structure with the experimental INPHARMA NOE allowed us selecting the following ligand orientations: panel (a)- (d): $L_A$, pose 1; $L_B$, pose 1 (X-ray like); panel (e): $L_A$, pose 2; $L_B$, pose 2 (both ligands are turned 180° around the y axis). The conformational inaccuracy of the snapshot structure in panel (e) allows the determination of the relative orientation of the two ligands by INPHARMA, while the absolute binding modes are not correctly predicted.

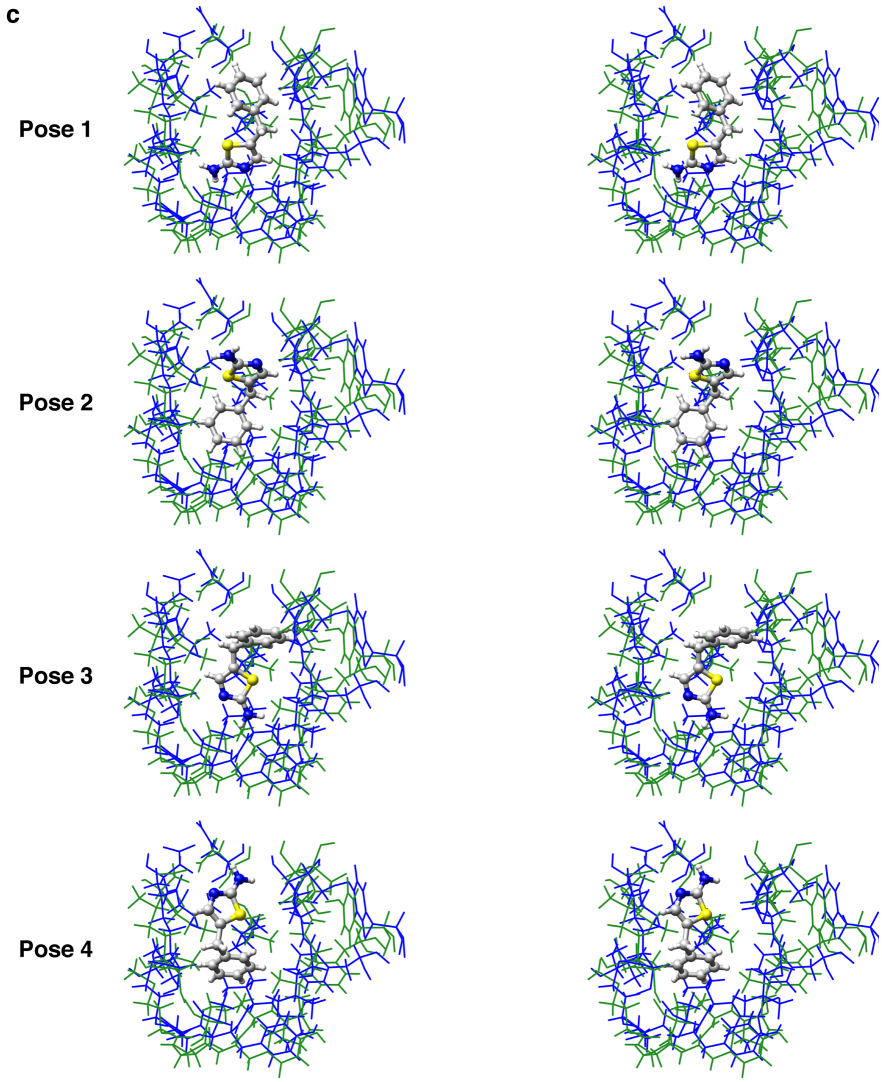


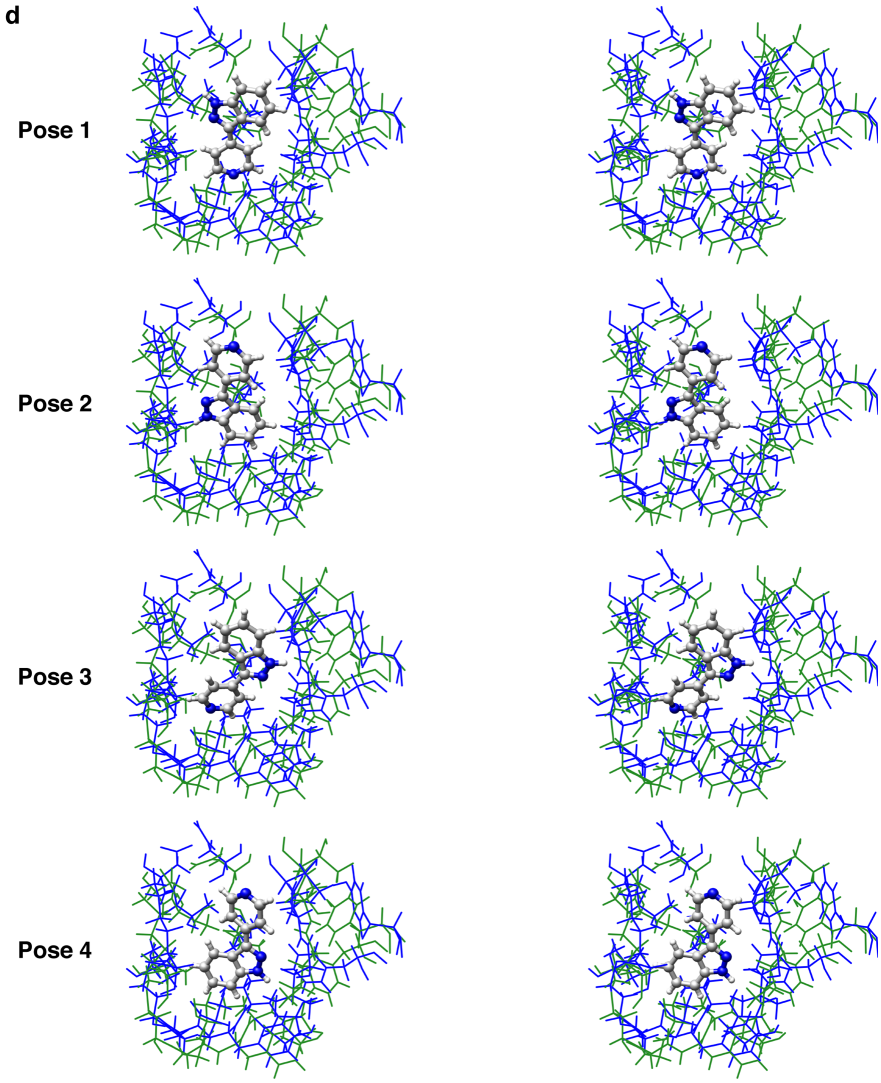

a
Pose 1
Pose 2
Pose 3
Pose 4
Ligand B

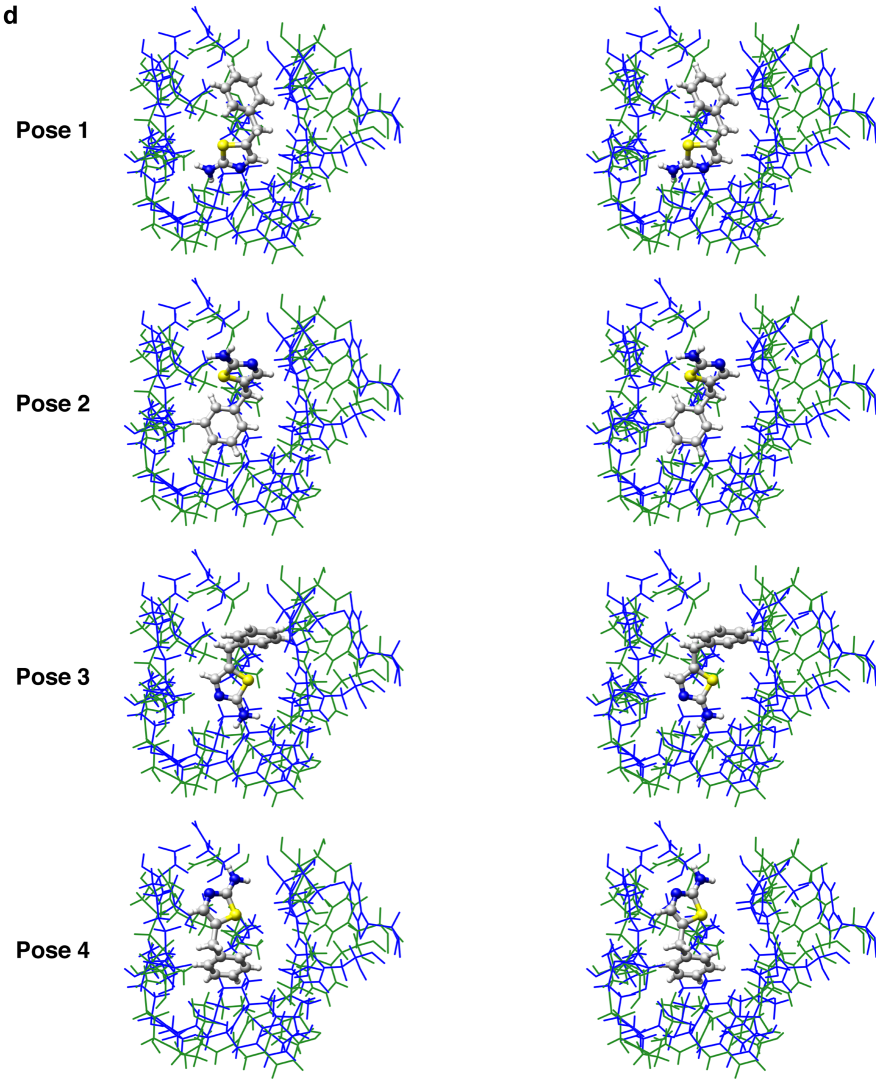

b
Pose 1
Pose 2
Pose 3
Pose 4
Ligand A

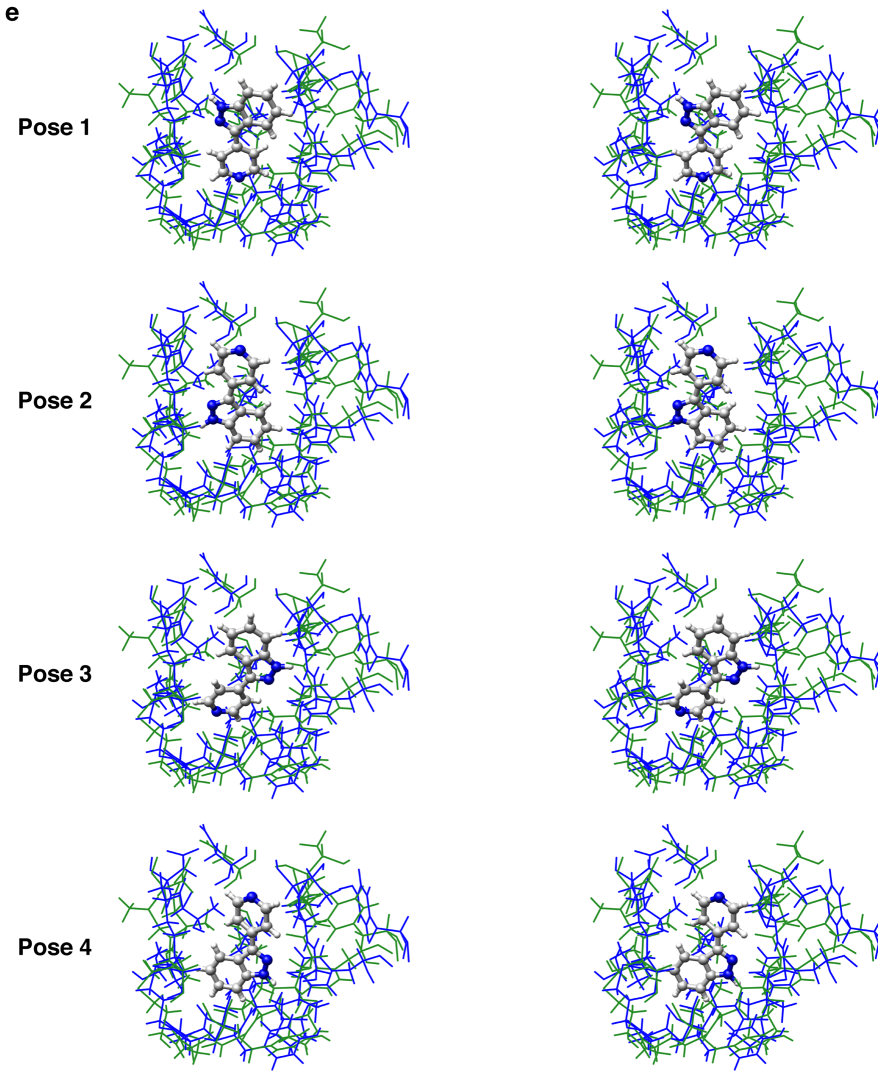

b
Pose 1
Pose 2
Pose 3
Pose 4
Ligand B

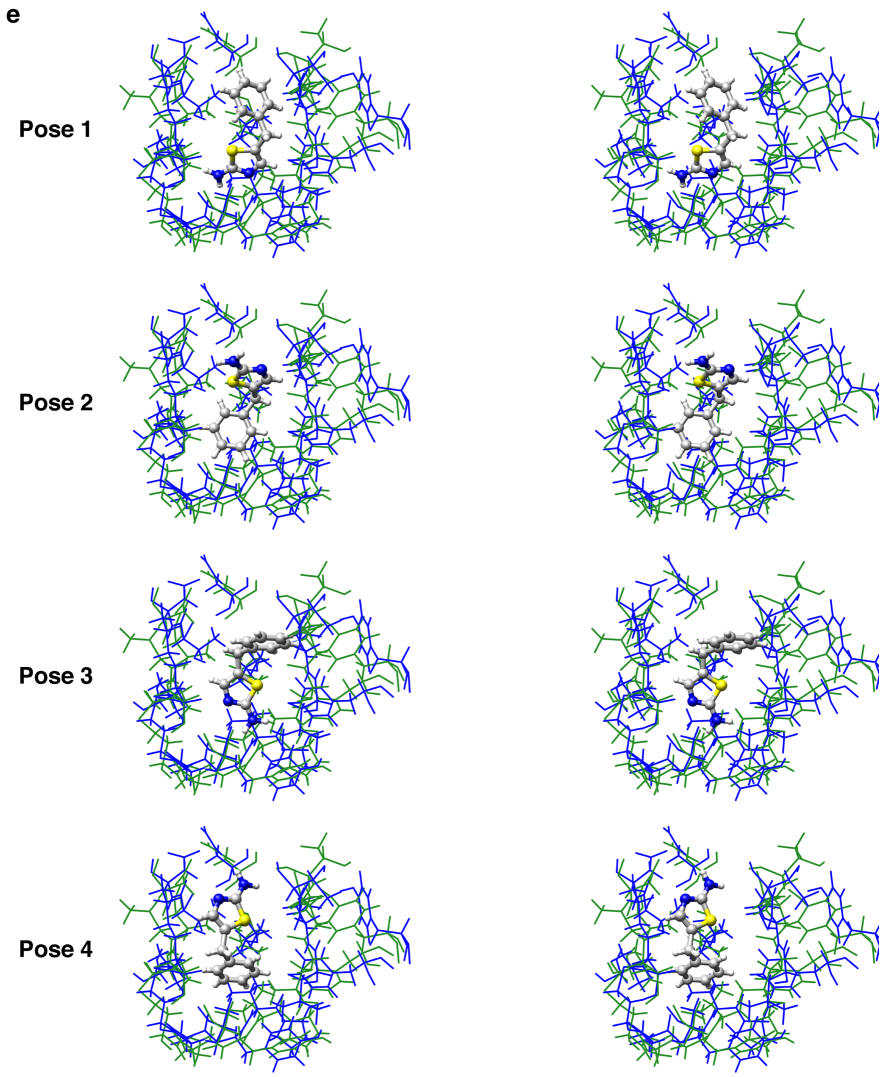

c
Pose 1
Pose 2
Pose 3
Pose 4
Ligand A


c
Pose 1
Pose 2
Pose 3
Pose 4
Ligand B


d
Pose 1
Pose 2
Pose 3
Pose 4
Ligand A


d
Pose 1
Pose 2
Pose 3
Pose 4
Ligand B


e
Pose 1
Pose 2
Pose 3
Pose 4
Ligand A


e
Pose 1
Pose 2
Pose 3
Pose 4
Ligand B

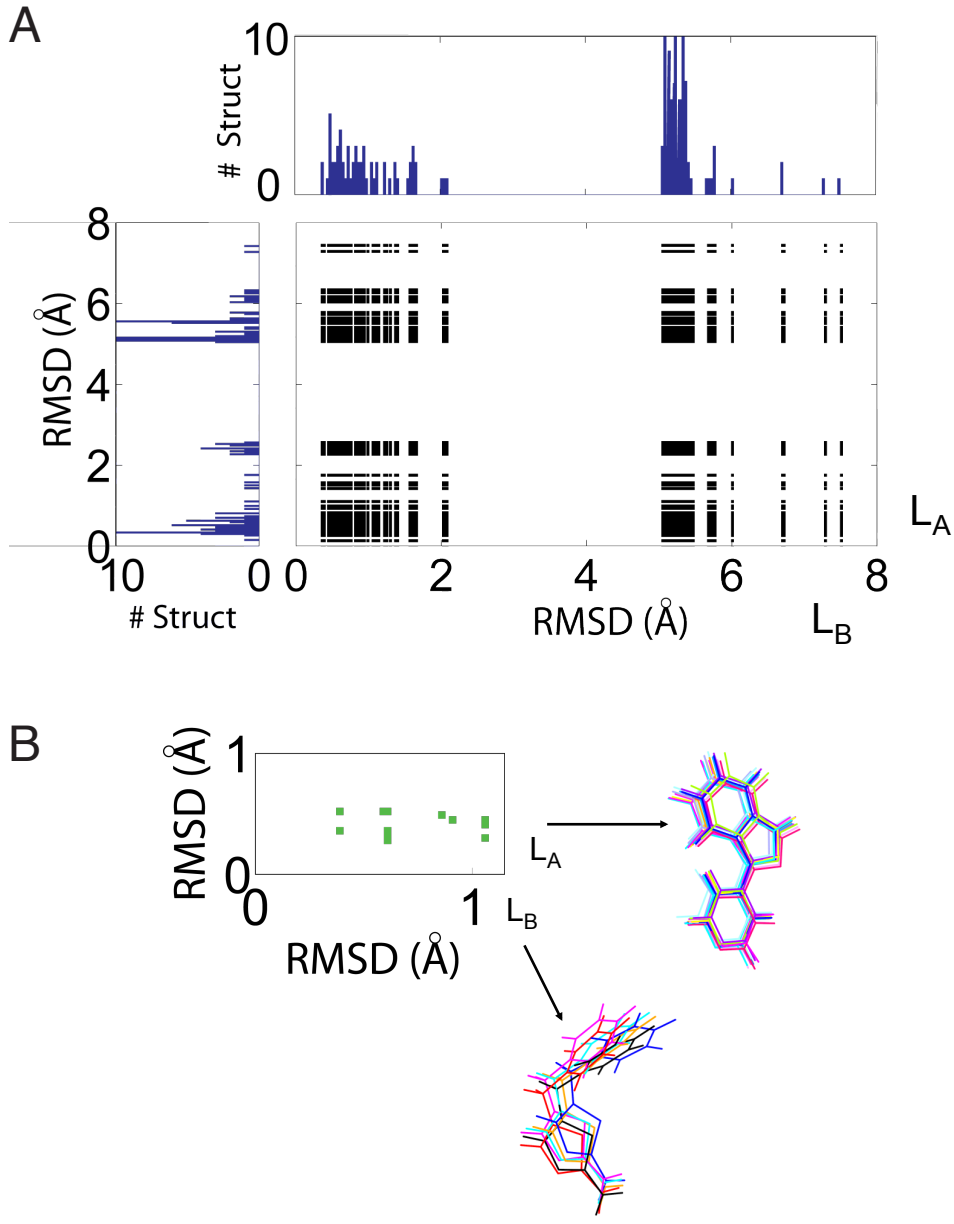



**Figure A.8:** A) Two Histograms of the docking poses RMSD (models versus Xray structure) distribution of the ligand A (represented vertically) and B (represented horizontally) for 200 models. Ligand A and B were docked in the ATP binding site of PKA while the protein structure was kept rigid. The histograms show the sets of RMSD values (in Å) between the docking poses (of ligand A and ligand B) and the corresponding native pose found in the Xray structure. 2D grid of all possible combinations pairs between the two sets of models (X axis corresponds to the ligand B and Y axis corresponds to the ligand A). B) Pairs of models from the grid in A) that pass the selection criteria described in chapter 2.

# Appendix B

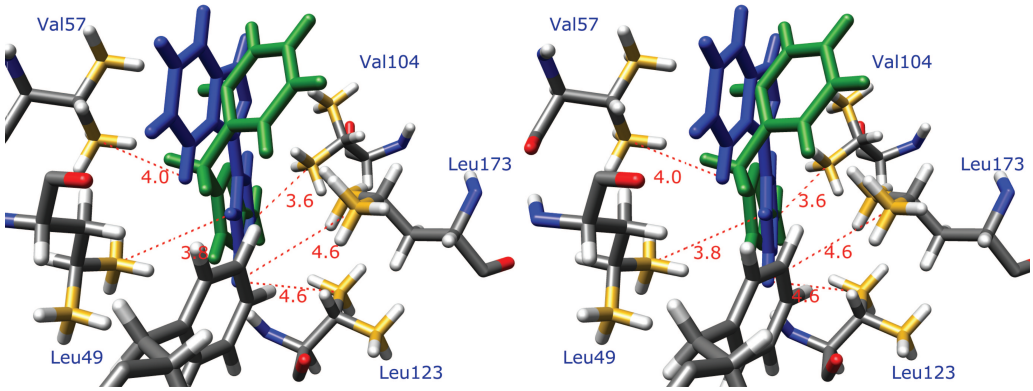



**Figure B.1:** Overlap of the crystal structures of the PKA/$L_A$ and PKA/$L_B$ complexes (stereoview). The Phe327, which is close to both the H8-$L_A$ and protons H1,2-$L_B$, is shown as well.,

# Appendix C

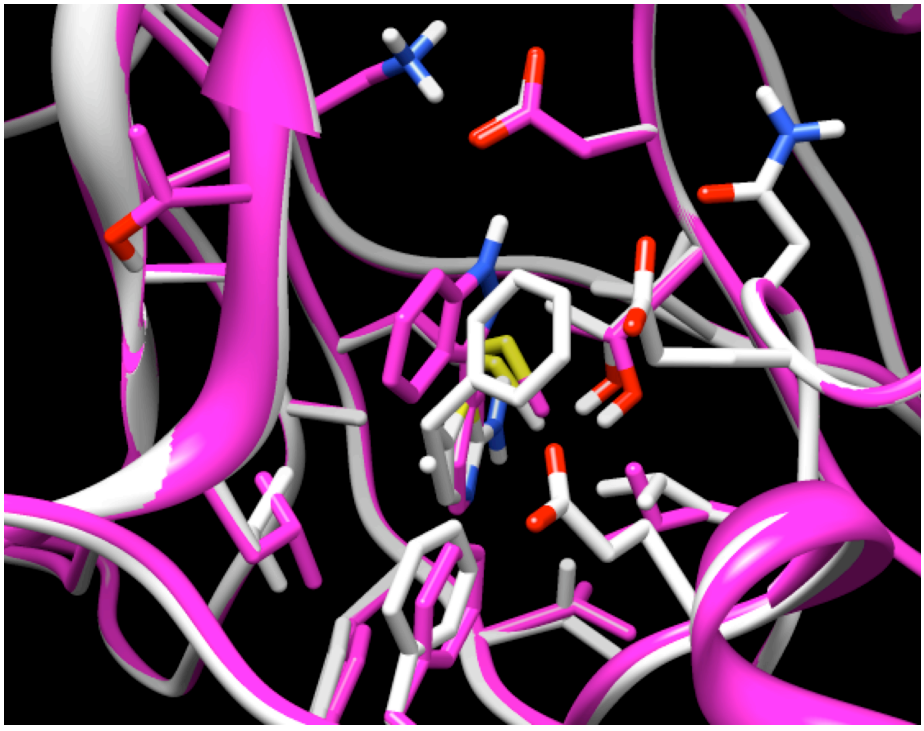


**Figure C.1:** Overlap of two homology models created for the complex PKA/ligand A (pink) and PKA/ligand B (white).

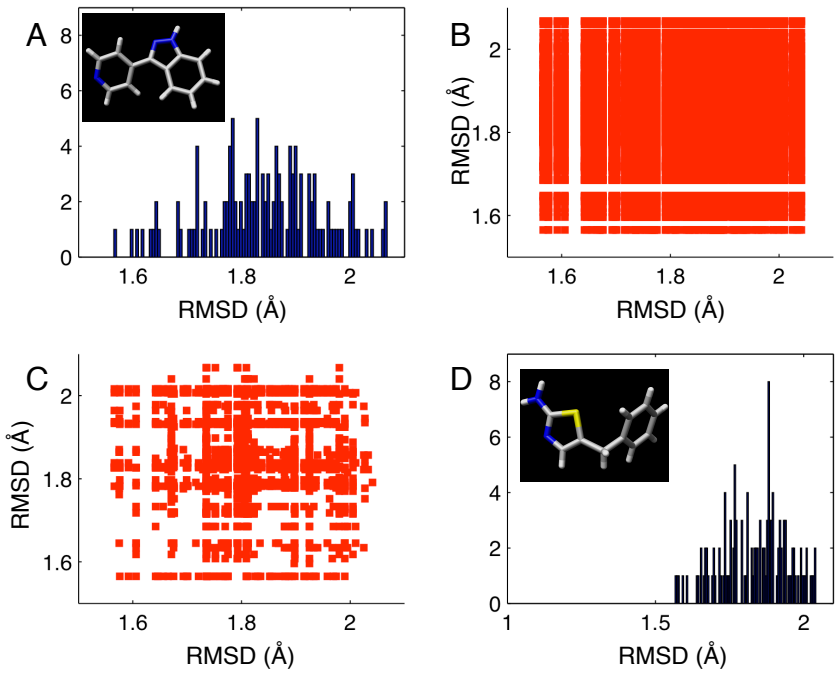



**Figure C.2:** A), resp. D) Histogram of the binding pocket RMSD (models versus X-ray structure) distribution of the ligand A, resp. B for the 125 homolgy models. The x-axis represents the RMSD values in Å and the y-axis is the numbers of structures having similar RMSD values. B) 2D grid of all possible combinations pairs between the two sets of models. C) Pairs of models from the grid in B) that correlated at minimum $R > 85\%$ to the experimental values (INPHARMA NOEs).

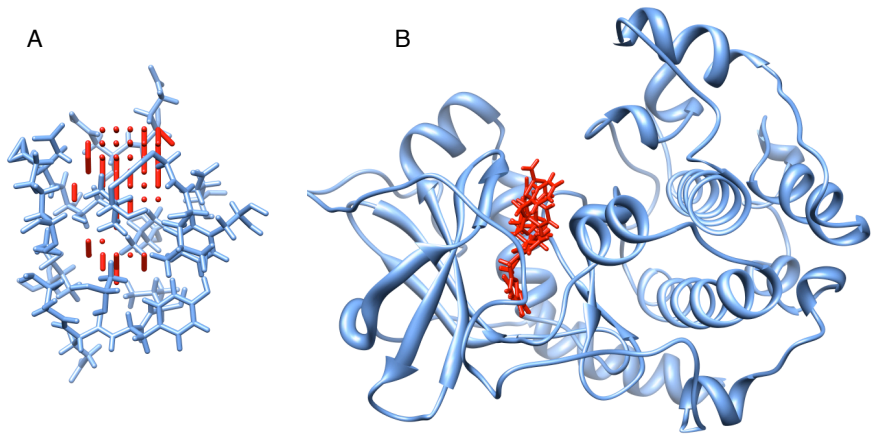



**Figure C.3:** a) Protomol generated by Surflex for the ATP binding site of PKA. B) Outcome of the Surflex docking protocol: 10 best poses for the ligand B dock to the ATP intereaction site of a MD model of PKA

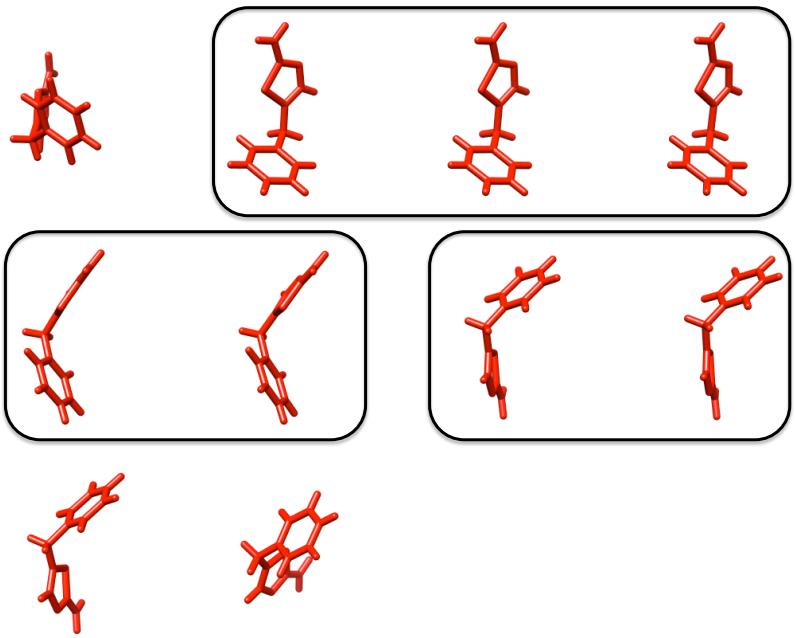


**Figure C.4:** 10 best poses for the ligand B docked to the ATP intereaction site of one PKA model derived from the MD simulation. Three black boxes surround similar docking poses. Only the pose with the highest score among the redundant ones is kept. In this particular case six docking models will pass the filter.

We build two criteria to select the correct relative orientation of the ligand based on five parameters related to quaternions $(\theta_1, \theta_2, \vec{q_1}, \vec{q_2}, shift)$, where $\vec{q_i}$ represents the axis of rotation, $\theta_i$ the angle of rotation and shift is the distance between the centers of the ligands (center of mass). The two selection functions are:

$$\begin{cases} f_1 = (\theta_1 - sign(\vec{q_2} \cdot \vec{q_2})\ \theta_2) \cdot (1 - 0.7 \cdot \cos(\vec{q_1} \cdot \vec{q_2})^2)\ mod(360^\circ) \\ f_2 = (\theta_{i=1,2} \cdot (1 - \cos(\vec{q_1} \cdot \vec{q_2})^2) \end{cases} \tag{C.1}$$

The function $f_1$ allows different degree of freedom of the ligands depending on the relative position of the quaternion vectors axis. We tolerate a higher degree of rotation if the two quaternion vectors axis are parallel compared to the case where the vector axis are perpendicular. The function $f_2$ ensures that when the axis are perpendicular, $\cos(\vec{q_2} \cdot \vec{q_2}) \approx 0$ and $(\theta_1 - \theta_2) \approx 0$, the rotation angles, $\theta_{1,2}$, are restricted to reasonable values. For the calculations we considered pairs as positive when $f_1 \leq 17$, $f_2 \leq 29$ and $shift \leq 1$ Å (to ensure that the two ligands do not bind too far on the protein as they are competitive binders).

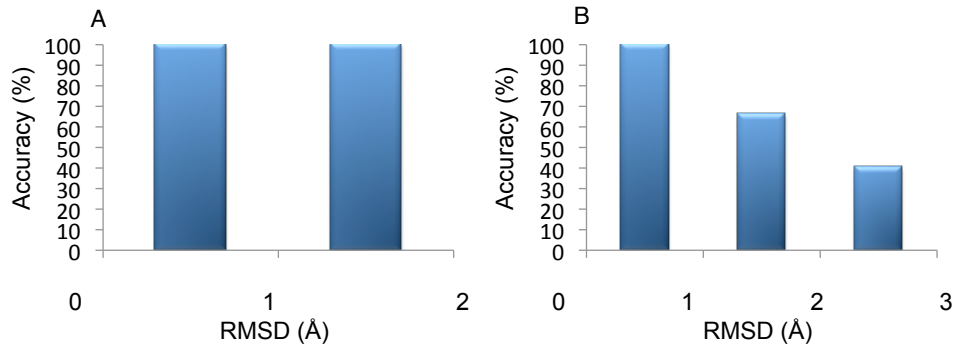



**Figure C.5:** A), resp. B) Accuracy of the INPHARMA predictions in dependence of the receptor structure quality when the bioactive binding mode of the ligand A, resp. B is known. The x-axis represents the binding pocket RMSD (receptors models versus X-ray structure) in Å. The y-axis is the Accuracy in % of the INPHARMA method calculated for the set of complexe structures within a certain receptor RMSD range defined on the x-axis.

# Appendix D

# RNA dynamics

### Introduction

The past years have seen a growing interest in the dynamic behavior of proteins and their correlation with function. Protein dynamics on several timescales are studied with state-of-the-art methodology, such as relaxation and dipolar couplings. However, dynamic investigations of RNA molecules are still sparse, with the exception of a few seminal studies dealing both with fast internal motions and slower inter-domain motions [29,37,98]. Much work has still to be done to understand RNA inter and intra-domain dynamics.

Clearly, understanding how an RNA sequence codes for a particular function at a molecular level requires both the structural description of its conformation(s) and characterization of its dynamics. In addition RNA molecules have proven to be much more dynamic than proteins and their conformation(s) can be subtly influenced by the environmental conditions. Thus, developing techniques to depict the dynamics process underlining activity and functional regulation of the RNA is one of the biggest challenges in structural biology. Between several techniques such as X-ray crystallography, fluorescence spectroscopy, EPR spectroscopy, and molecular dynamics simulations, NMR spectroscopy rank among the most promising methods since it provides atomic resolution for both structure and dynamics over large and relevant time scale (ps-ms) for biological processes. Part of my Ph.D. work focuses on characterizing RNA dynamics, first on the ps-ns time scale and later until the ms time scale. The RNA molecule I am working on is the U4 5' stem-loop RNA [77, 101], which intervenes during the splicing cycle and is known to

assume a kink-turn fold when bound to 15.5K (Fig. D.1) [58, 61]. However, MD [17, 18, 84, 85], fluorescence and biochemical studies have hypothesized large inter-domain motions and even the unfolding of stem II for the RNA free in solution [44].

Here we present the part of the study on the fast dynamics (ps-ns) of the U4 RNA molecule. We first address the question of the stability of the fold of stem II. Subsequently, we address the question of decoupled motions between the stem I and stem II of the U4 RNA.

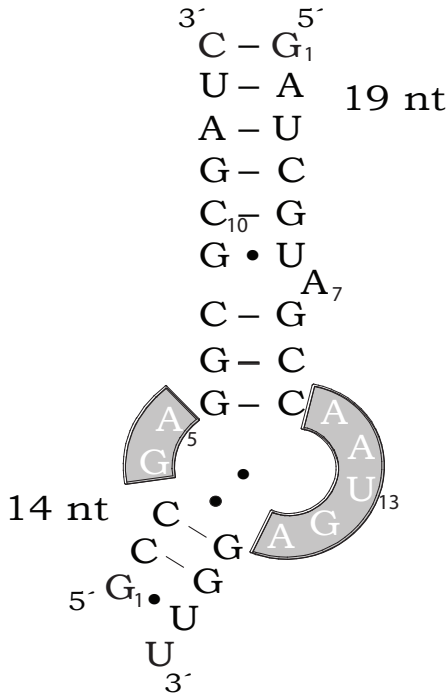



**Figure D.1:** Seondary structure of the U4 RNA construct used in this work. Strand I is composed of 14 nucleotides and strand II of 19 nucleotides. The region where the protein 15.5K binds the RNA is highlighted in grey and connect stem I, the shorter stem, to stem II. It countains the three nucleotides that form the kink-turn, A11 A12 and U13. Non canonical base pairs are represented by a dot, while the canonical ones are depicted by a line.

**U4 RNA dynamics: local motions faster than the correlation time**

We aim at determining the local dynamics of all bases and riboses of the U4 RNA. It would answer the question about the stability of the stem II and also set the first step towards a better understanding of the recognition of 15.5K by U4. Motions on a time-scale faster than the correlation time can be detected through the relaxation rates of the individual spin in each nucleotide. Therefore, we measured relaxation rates constants, $R_1$, $R_{1\rho}$ and the $^{13}$C -($^1$H ) steady-state NOE for the carbons C6, C8 on the nucleobases and C1' on the ribose moieties for the 33 nucleotides at 298 K. In order to reduce peaks overlapping in the spectra only one strand of the U4 construct was $^{13}$C ,$^{15}$N labeled at a time. Hence, we had to repeat twice the experiments to evaluate the complete set of parameters ($R_1$, $R_{1\rho}$ and the $^{13}$C -($^1$H ) steadystate NOE). The samples were prepared by Claudia Schwiegk and Melanie Falb (Ph.D. student who set up the expression and purification protocol). The assignment of the $^1$H , $^{13}$C and $^{15}$N resonances of the U4 RNA was established by Melanie Falb.

NMR experiments were carried out on 600 MHz, 700 MHz and 800 MHz Bruker spectrometers. $^{13}$C $R_1$ and $R_{1\rho}$ and the $^{13}$C -($^1$H ) steady-state NOE data were obtained from the pulse sequences hsqct1etf3gpsi3d, hsqctretf3gpsi3d.2 and invnoef3gpsi included in the Bruker pulse sequence library [33, 48, 71] and adapted to measure carbon relaxation. The carrier frequency was set to 138 ppm for the aromatic carbons C6, C8 and 89 ppm for C1'. The spectral width was 10 ppm for the bases and 6 ppm for the sugars. About 160 complex points were acquired in the indirect dimension. Off-resonant carbon IBURP1 pulses were applied during carbon evolution in order to suppress the J(C5, C6), resp. J(C1', C2' ) coupling. Interscans relaxation delays of 2 s were used for the $R_1$ and $R_{1\rho}$ measurements, 5 s were used for the NOE. $R_1$, $R_{1\rho}$ and the NOE data were obtained with 160 scans for each t1-increment. $R_1$ and $R_{1\rho}$ subspectra with different relaxation delays were acquired in one scan-interleaved pseudo-3D experiment. The NOE experiments were also recorded scan-interleaved, with alternating proton-presaturated and non-presaturated spectra. While in the latter case, a relaxation delay of 5 s was used, proton-presaturation was applied for 3 s subsequent to a 2 s relaxation delay in the presaturated spectra. The interleaved spectra were separated by a modified Bruker standard macro. For the acquisition of $R_1$ relaxation rates, the relaxation delay $\tau$ was set to 10 ms, 50 ms, 100 ms, 200 ms, 400

ms, 700 ms, 1 s and 1.5 s. Spectra with $\tau$=50 ms and 400 ms were measured twice for error determination. $R_{1\rho}$ rates were acquired as described [48, 71]. Random length proton decoupling pulses were used during the carbon spin-lock period. Adiabatic Mulder pulses [71] were applied to rotate the carbon magnetization into the transverse plane. A spin-lock field of 2 kHz with an offset of 2000 Hz was applied for the variable relaxation delay ($\tau$=12 ms, 24 ms, 36 ms, 48 ms, 64 ms, 80 ms, 104 ms, 128 ms). Duplicate measurements were recorded for $\tau$=24 ms and 80 ms. Figure D.2 (spectra green and yellow) shows plans of the $R_1$ pseudo-3D experiment measured with labeled U4 samples at the 600 Mhz spectrometer.

The data were processed using Topspin and analyzed in Sparky. $R_1$ and $R_{1\rho}$ relaxation rates were fitted from peak heights to monoexponential functions with two unknown parameters $Ae^{-Rt}$ (Fig. D.3). $R_{1\rho}$ contains the effect of conformational exchange, the spin-lock field strength and spin-lock offset:

$$R_{1\rho} = R_1 \cos(\theta)^2 + R_2 \sin(\theta)^2 + R_{ex} \sin(\theta)^2 \tag{D.1}$$

in which $\theta = \arctan(\omega_1/\Delta\omega)$ is the inclinaton angle between the static field, $\Delta\omega = \omega - \omega_0$, and the effective spin-lock field, $\omega_e = (\omega_1^2 + \Delta\omega^2)^{1/2}$, in the rotating frame; $\omega$ is the spin-lock rf frequency; $\omega_1$ is the spin-lock field strength in units of rad.s$^{-1}$; $\omega_0$ is the population-averaged chemical shift; $R_1$ and $R_2$ are the longitudinal and transverse relaxation rate constants. $R_{ex}$ is the contribution to the transverse relaxation rate from exchange processes.

Separation of the exchange contribution to the transverse relaxation will be carried out with the model-free analysis with either Modelfree of A. Palmer and co-workers [66] or with Relax of E. d'Auvergne [23–25]. Diffusion parameters will be obtained from the program Modelfree and Relax and compare to hydrodynamic calculations performed with the program hydronmr [34].

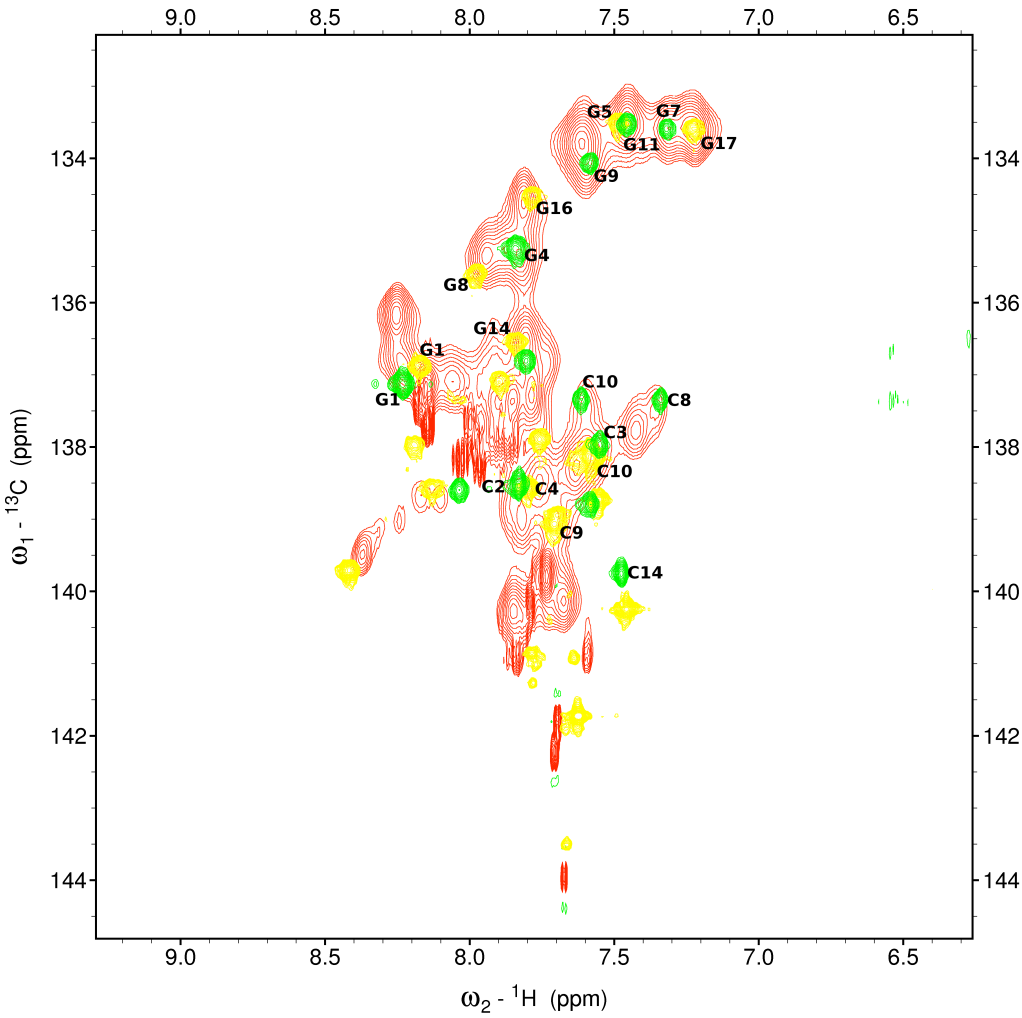



**Figure D.2:** Overlap of the first 2D-plan of three $R_1$ 3D-scan-interleaved experiments on C6-H6 and C8-H8 measured at 600 MHz (three spectra, green, yellow and red). The green and yellow spectra correspond respectively to U4 samples with the stem I labelled and stem II labelled only. The red spectrum corresponds to the A-U elongated and C,G labelled sample (Fig. D.4).

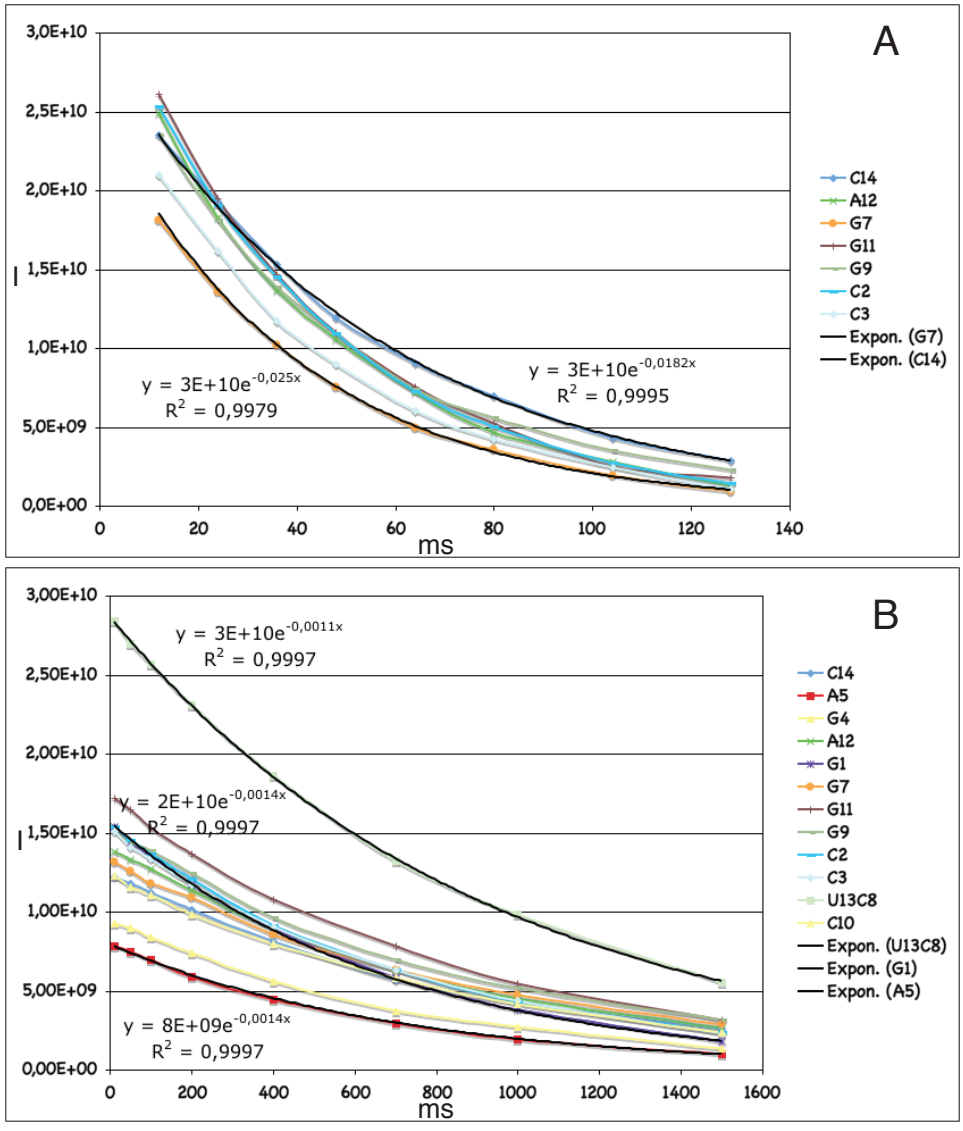



**Figure D.3:** Peaks intensities are plotted against the relaxation time delay $\tau$ (ms). In panel A, resp. B, $R_{1\rho}$, resp. $R_1$ of the C1' are plotted for different riboses.

**Motions coupling between stem I and stem II of the U4 RNA**

We wish to understand the inter-domain dynamics, specifically if one stem moves with respect to the second stem hinging on the kink-turn motive. Here we will make use of recently developed NMR techniques by Al-Hashimi [37, 105, 106]. Domain elongation strategy allows to resolve domain motions by nuclear magnetic resonance spectroscopy. It requires an elongated RNA with a hydrodynamic shape (rotational diffusion tensor) that is less sensitive to domain motions. As in the U4 no stem is much longer than the other one, we designed new constructs, where stem I or stem II is elongated (Fig. D.4). If the diffusion tensor of the RNA molecule is dominated by a single long helix, the total diffusion tensor can be calculated and the dynamics of the second stem relative to the long helix can be determined [105–107]. With these constructs we will study the dynamics of stem I with respect to stem II and vice versa. As the length of the elongated stem is greatly superior to the other one, the correlation time of the shorter stem will be decoupled from the correlation time of the complete RNA and appears as internal motions. Again, to reduce spectral overlap upon elongation of one stem, special labeling scheme are employed. When elongation is done with AU base pairs, C and G residues are $^{13}$C ,$^{15}$N labeled and therefore observed in the spectra. The same procedure is applied for labelled A, U residues where GC base pairs are used for elongation. In this way, number of the peaks overlapping does not increase. Note that we need few CG base pairs for stabilization of the elongated stem when the elongation is done with A-U base pairs and a UUCG tetra loop is added to ensure proper folding (Fig. D.4). With the same protocol described in the previous section, we measured $R_1$ and $R_{1\rho}$ on the elongated U4 RNAs (Fig. D.3), (complete data set not shown). Comparison of the dynamics parameters obtained after the modelfree anaylsis for the short and elongated constructs will reveal additional dynamics of the stems that is otherwise masked by the overall reorientation of the molecule. Indeed, by comparing the dynamics parameters determined for the short U4 RNA with respect to the corresponding ones calculated for the elongated U4 RNA (on the stem which was not elongated), we could conclude that the two stems are either rigid with respect to each other, or move completely independent of each other.

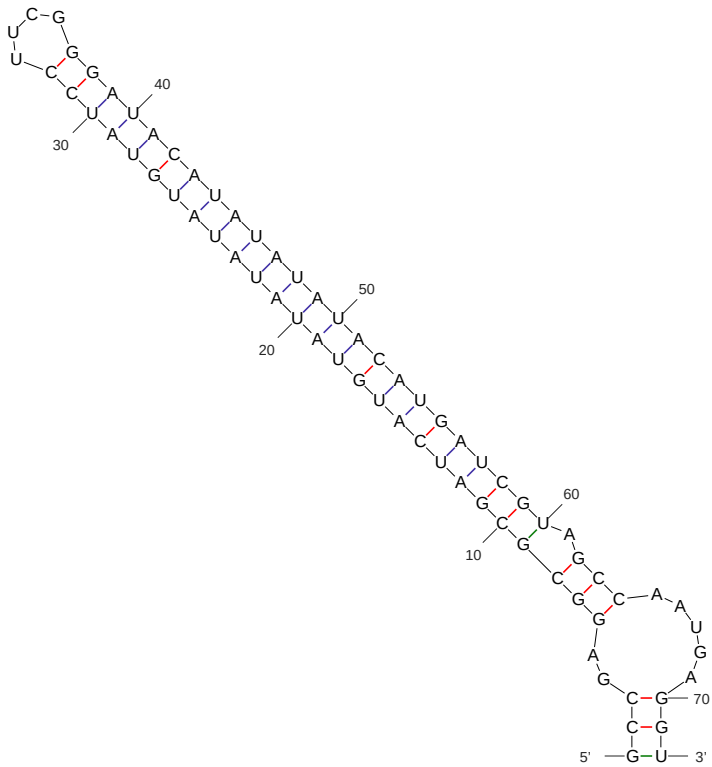



**Figure D.4:** Seondary structure of the A-U elongated C,G labelled U4 RNA construct used in this work

# Publications

Orts J, Griesinger C, Carlomagno T. The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method.
*J. Magn. Reson.* , 200(1):64-73, **2009**.

Orts J, Tuma J, Reese M, Grimm SK, Monecke P, Bartoschek S, Schiffer A, Wendt KU, Griesinger C, Carlomagno T. Crystallography-independent determination of ligand binding modes.
*Angew. Chem. Int. Ed. Engl.*, 2008;47(40):7736-40**2008**.

Orts J, Grimm SK, Griesinger C, Wendt KU, Bartoschek S, Carlomagno T. Specific methyl group protonation for the measurement of pharmacophore-specific interligand NOE interactions.
*Chemistry* , 14(25):7517-20, **2008**.

# Acknowledgement

A Ph.D. thesis is rarely the accomplishment of a single person. During the last three years I crossed the way of many persons. All of them contributed to make my research quest enjoyable and instructive.

I would first thank my guides. Teresa Carlomagno, who always supported me and positively criticized my work. Our discussions were always very instructive both from a scientific and human point of view. Thank you for giving me independence and putting trust in my capabilities. I enjoyed working on theses very nice projects and I appreciated the constant interest you had for my research. I did not have the time to learn Italian, but I learned so many other things, grazie!
Christian Griesinger who initially welcomed me in his department at the Max Planck Institute. His energy and inexhaustible interest for my work were motivating and inspiring. I will not forget our fruitful discussions in the middle of the nights. I was already convinced that everything is a matter of willing, now I am sure about it!
Finally Martin Blackledge for following my research and accepting to be my promoter at "home" in Grenoble.

My thoughts go to the Aventis team with whom I had a great collaboration. Thank you Stefan, Peter and Elke.
Thank you to Toby Gibson and Anne Ephrussi for accepting to take part of my Thesis Advisory committee.
It is a honour that Bruno Kieffer and Roland Riek accepted to review this work. I am grateful to John, Mayako, Benjamin and Teresa for giving very good comments on the content, the style of the thesis and the german sum-

mary: "Vielen Dank".
Special thanks to Wolfgang Jahnke, the very first actor of this story.

I will surely always remember my first team mate, Karel the chocolate eater, who initiated me with the NMR spectrometers and taught me good basic practices in the "wet lab". Many good souvenirs will stay in memory, as the "MRI water calibration or the console rebooting..." Of course the team would not be complete if I do not mention Christophe & Peter the twins fighter, Nils the wise-man, Edward d'Auvergne actually from Australia, Pierre and Nicolas the french "couple", Min-Kyu and Hai-Young our real korean couple, Vinesh and co, Sigrun, Monika & Laura the lively office mates in Goettingen, Steffen Grimm our storyteller, Ping and Volker the artists, Joerg and Melanie working on the "monster", Fernando the quiet, Irene the keyboard killer, Marcel the gnu-plotter, Jenny Dirk and JJ in the cold room, Lukasz and Donghan for the south park parody, Korvin more dangerous than a gremlins but as funny, Claudia the wet lab master, Benedict the adventurer, Sergey, Mate, Fabian, Edith the sporty and funny chemists (lab did not burst), the solid state crew and all the others that I do not list but also made my time in Goettingen so memorable.

Over the Ph.D., I changed town and laboratory. I went from the excellent MPI in the rainy and cute Goettingen to the excellent EMBL in the cute and rainy Heidelberg.
I had the great honour to share the office with John the british gentleman and Tea from Bari that can speak louder than Bernd the spectrometer master.
Today I share the office with Ben the poker player, Domenico, Mariarosaria the south Italians team, Irene still killing keyboard, Carmen that knows more people you would in 10 lives, Thierry the last real "Alsacien", Audri and Mayako the new comers. We are supported by Magda in the wet-lab with her marvellous cakes and Bernd, who is looking after the spectrometers with a lot of humour.

It was a real pleasure to meet you!

Le tableau ne serait pas complet sans les personnes qui comptent affectivement. Les encouragements et le soutient de mes parents m'ont toujours donné confiance en moi. La gentillesse de mon frère, Guillaume et de ma sœur, Nathalie, fait partie intégrante des moments en famille si important pour garder un certain équilibre. Ces instants si précieux, toute la famille y a contribué et j'espère qu'ils se feront toujours au sein d'une famille encore plus nombreuse. J'ai une pensée pour mon grand-père parti qui doit être fier de moi. Je porte toute mon affection aux trois autres grands-parents et à ma grande tante. Sans oublier mes oncles, tantes, cousins, belles-sœurs, beaux-parents et autres proches. C'est enfin vers mon autre moitié, Anne-Christine, que je me tourne pour témoigner toute ma reconnaissance envers ta patience, ta bienveillance et ton affection, merci.

Julien

# Bibliography


[1] A. Abraham. *Principles of nuclear magnetism*. Oxford University Press, Oxford, United Kingdom, **1961**.

[2] J. Anglister, F. Naider, J. Anglister, T. Scherf, B. Zilber, R. Levy, A. Zvi, R. Hiller, and D. Feigelson. 2-dimensional nmr investigations of the interactions of antibodies with peptide antigens. *Methods in Enzymology*, 7(12):1154–1162, 1993.

[3] F. M. Ashcroft. From molecule to malady. *Nature*, 440(7083):440–7, 2006.

[4] P. Balaram, Bothnerb.Aa, and E. Breslow. Localization of tyrosine at binding-site of neurophysin ii by negative nuclear overhauser effects. *Journal of the American Chemical Society*, 94(11):4017–&, 1972.

[5] P. Balaram, Bothnerb.Aa, and J. Dadok. Negative nuclear overhauser effects as probes of macromolecular structure. *Journal of the American Chemical Society*, 94(11):4015–&, 1972.

[6] R. W. Behling, T. Yamane, G. Navon, and L. W. Jelinski. Conformation of acetylcholine bound to the nicotinic acetylcholine-receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 85(18):6721–6725, 1988.

[7] H M Berman, J D Westbrook, M J Gabanyi, W Tao, R Shah, A Kouranov, T Schwede, K Arnold, F Kiefer, L Bordoli, J Kopp, M Podvinec, P D Adams, L G Carter, W Minor, R Nair, and J La Baer. The protein structure initiative structural genomics knowledgebase. *Nucleic Acids Res*, 37(Database issue):365–368, Jan 2009.

[8] B. A. Borgias, M. Gochin, D. J. Kerwood, and T. L. James. Relaxation matrix analysis of 2d-nmr data. *Prog. Nucl. Magn. Res. Spec.*, 22:83–100, 1990.

[9] D. Bossemeyer, R. A. Engh, V. Kinzel, H. Ponstingl, and R. Huber. Phosphotransferase and substrate binding mechanism of the camp-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 a structure

of the complex with mn2+ adenylyl imidodiphosphate and inhibitor peptide pki(5-24). *Embo J*, 12(3):849–59, 1993.

[10] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 30(10, Sp. Iss. SI):1545–1614, JUL 30 2009.

[11] B.R. Brooks, R.E. Bruccoleri, D.J. Olafson, D.J. States, S. Swaminathan, and M. Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:187–217, 1983.

[12] T. Carlomagno, W. Bermel, and C. Griesinger. Measuring the chi 1 torsion angle in protein by ch-ch cross-correlated relaxation: a new resolution-optimised experiment. *J Biomol NMR*, 27(2):151–7, 2003. Journal of biomolecular NMR.

[13] T. Carlomagno, M. J. Blommers, J. Meiler, B. Cuenoud, and C. Griesinger. Determination of aliphatic side-chain conformation using cross-correlated relaxation: application to an extraordinarily stable 2'-aminoethoxy-modified oligonucleotide triplex. *J Am Chem Soc*, 123(30):7364–70, 2001. Journal of the American Chemical Society.

[14] T. Carlomagno, V. M. Sanchez, M. J. Blommers, and C. Griesinger. Derivation of dihedral angles from ch-ch dipolar-dipolar cross-correlated relaxation rates: a c-c torsion involving a quaternary carbon atom in epothilone a bound to tubulin. *Angew Chem Int Ed Engl*, 42(22):2515–7, 2003. Angewandte Chemie (International ed.

[15] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, and R. C. Stevens. High-resolution crystal structure of an engineered human beta2-adrenergic g protein-coupled receptor. *Science*, 318(5854):1258–65, 2007.

[16] G. M. Clore and A. M. Gronenborn. Theory of the time-dependent transferred nuclear overhauser effect - applications to structural-analysis of ligand protein complexes in solution. *Journal of Magnetic Resonance*, 53(3):423–442, 1983.

[17] V Cojocaru, R Klement, and T M Jovin. Loss of g-a base pairs is insufficient for achieving a large opening of u4 snrna k-turn motif. *Nucleic Acids Res*, 33(10):3435–3446, 2005.

[18] V Cojocaru, S Nottrott, R Klement, and T M Jovin. The snrnp 15.5k protein folds its cognate k-turn rna: a combined theoretical and biochemical study. *RNA*, 11(2):197–209, Feb 2005.

[19] P. Cozzini, G. E. Kellogg, F. Spyrakis, D. J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L. A. Kuhn, G. M. Morris, M. Orozco, T. A. Pertinhez, M. Rizzi, and C. A. Sotriffer. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem*, 51(20):6237–55, 2008. eng.

[20] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, and B. Stockman. Waterlogsy as a method for primary nmr screening: practical aspects and range of applicability. *J Biomol NMR*, 21(4):349–59, 2001. Journal of biomolecular NMR.

[21] D. E. Danley. Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 6):569–75, 2006. eng.

[22] R. Das and D. Baker. Macromolecular modeling with rosetta. *Annu Rev Biochem*, 77:363–82, 2008. eng.

[23] E J d'Auvergne and P R Gooley. Model-free model elimination: a new step in the model-free dynamic analysis of nmr relaxation data. *J Biomol NMR*, 35(2):117–135, Jun 2006.

[24] E J d'Auvergne and P R Gooley. Optimisation of nmr dynamic models i. minimisation algorithms and their performance within the model-free and brownian rotational diffusion spaces. *J Biomol NMR*, 40(2):107–119, Feb 2008.

[25] E J d'Auvergne and P R Gooley. Optimisation of nmr dynamic models ii. a new methodology for the dual optimisation of the model-free parameters and the brownian rotational diffusion tensor. *J Biomol NMR*, 40(2):121–133, Feb 2008.

[26] I W Davis, K Raha, M S Head, and D Baker. Blind docking of pharmaceutically relevant compounds using rosettaligand. *Protein Sci*, 18(9):1998–2002, Sep 2009.

[27] Ian W. Davis and David Baker. Rosettaligand docking with full ligand and receptor flexibility. *Journal of Molecular Biology*, 385(2):381–392, 2009. http://www.sciencedirect.com/science/article/B6WK7-4TY9MY4-2/2/2fffff31fbfc1ba0d9301da014f5e7de.

[28] L. J. DeLucas. Protein crystallization - is it rocket science? *Drug Discov Today*, 6(14):734–744, 2001. Eng.

[29] E Duchardt and H Schwalbe. Residue specific ribose and nucleobase dynamics of the cuucgg rna tetraloop motif by mnmr 13c relaxation. *J Biomol NMR*, 32(4):295–308, Aug 2005.

[30] H. M. Edwards. Galois theory. *Springer-Verlag*, ISBN 0-387-909080-X, 1984.

[31] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr.*, 60(Pt 12 Pt 1):2126–32, 2004.

[32] R. A. Engh, A. Girod, V. Kinzel, R. Huber, and D. Bossemeyer. Crystal structures of catalytic subunit of camp-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors h7, h8, and h89. structural implications for selectivity. *J. Biol. Chem.*, 271(42):26157–64, 1996.

[33] N A Farrow, O Zhang, J D Forman-Kay, and L E Kay. A heteronuclear correlation experiment for simultaneous determination of 15n longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR*, 4(5):727–734, Sep 1994.

[34] J García De La Torre, M L Huertas, and B Carrasco. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J*, 78(2):719–730, Feb 2000.

[35] N K Goto, K H Gardner, G A Mueller, R C Willis, and L E Kay. A robust and cost-effective method for the production of val, leu, ile (delta 1) methyl-protonated 15n-, 13c-, 2h-labeled proteins. *J Biomol NMR*, 13(4):369–374, Apr 1999.

[36] P. J. Hajduk, J. C. Mack, E. T. Olejniczak, C. Park, P. J. Dandliker, and B. A. Beutel. Sos-nmr: a saturation transfer nmr-based method for determining the structures of protein-ligand complexes. *J Am Chem Soc*, 126(8):2390–8, 2004. eng.

[37] A L Hansen and H M Al-Hashimi. Dynamics of large elongated rna by nmr carbon relaxation. *J Am Chem Soc*, 129(51):16072–16082, Dec 2007.

[38] H Hanzawa and T Takizawa. [nmr screening in the fragment-based drug discovery]. *Yakugaku Zasshi*, 130(3):325–333, Mar 2010.

[39] A. N. Jain. Surflex-dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *Journal of computer-aided molecular design*, 21(5):281–306, May 2007.

[40] T. A. Jones, J. Y. Zou, S. W. Cowan, and M. Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta. Crystallogr. A*, 47 ( Pt 2):110–9, 1991.

[41] W. Kabsch. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, 26:795–800, 1993.

[42] V Katritch, M Rueda, P C Lam, M Yeager, and R Abagyan. Gpcr 3d homology models for ligand screening: lessons learned from blind predictions of adenosine a2a receptor complex. *Proteins*, 78(1):197–211, Jan 2010.

[43] C. Kim, N. H. Xuong, and S. S. Taylor. Crystal structure of a complex between the catalytic and regulatory (rialpha) subunits of pka. *Science*, 307(5710):690–6, 2005.

[44] D J Klein, T M Schmeing, P B Moore, and T A Steitz. The kink-turn: a new rna secondary structure motif. *EMBO J*, 20(15):4214–4221, Aug 2001.

[45] D. R. Knighton, J. H. Zheng, L. F. Ten Eyck, V. A. Ashford, N. H. Xuong, S. S. Taylor, and J. M. Sowadski. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, 253(5018):407–14, 1991.

[46] T. M. G. Koning, R. Boelens, and R. Kaptein. Calculation of the nuclear overhauser effect and the determination of proton proton distances in the presence of internal motions. *Journal of Magnetic Resonance*, 90(1):111–123, 1990.

[47] S Kortagere and S Ekins. Troubleshooting computational methods in drug discovery. *J Pharmacol Toxicol Methods*, Feb 2010.

[48] D M Korzhnev, N R Skrynnikov, O Millet, D A Torchia, and L E Kay. An nmr experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J Am Chem Soc*, 124(36):10743–10753, Sep 2002.

[49] M. Kroemer, M. K. Dreyer, and K. U. Wendt. Aprv - a program for automated data processing, refinement and visualization. *Acta. Crystallogr. D Biol. Crystallogr.*, 60(Pt 9):1679–82, 2004.

[50] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. A geometric approach to macromolecule-ligand interactions. *J Mol Biol*, 161(2):269–88, 1982. eng.

[51] T Langer, M Vogtherr, B Elshorst, M Betz, U Schieborr, K Saxena, and H Schwalbe. Nmr backbone assignment of a protein kinase catalytic domain by a combination of several approaches: application to the catalytic subunit of camp-dependent protein kinase. *Chembiochem*, 5(11):1508–1516, Nov 2004.

[52] A. R. Leach, B. K. Shoichet, and C. E. Peishoff. Prediction of protein-ligand interactions. docking and scoring: successes and gaps. *J Med Chem*, 49(20):5851–5, 2006. eng.

[53] A. L. Lee, P. F. Flynn, and A. J. Wand. Comparison of h-2 and c-13 nmr relaxation techniques for the study of protein methyl group dynamics in solution. *J. Am. Chem. Soc.*, 121(12):2891–2902, 1999. J Am Chem Soc J Am Chem Soc.

[54] W. LEE and N. R. Krishna. Influence of conformational exchange on the 2d noesy spectra of biomolecules existing in multiple conformations. *Journal of Magnetic Resonance*, 98(1):36–48, 1992.

[55] IE Leonard. The matrix exponential. *SIAM Review*, 38:507–512, 1996.

[56] M Leone, H H Freeze, C S Chan, and M Pellecchia. The nuclear overhauser effect in the lead identification process. *Curr Drug Discov Technol*, 3(2):91–100, Jun 2006.

[57] D. Li, E. F. DeRose, and R. E. London. The inter-ligand overhauser effect: a powerful new nmr approach for mapping structural relationships of macromolecular ligands. *J Biomol NMR*, 15(1):71–6, 1999. eng.

[58] P Li, J Kirkpatrick, and T Carlomagno. An efficient strategy for the determination of the three-dimensional architecture of ribonucleoprotein complexes by the combination of a few easily accessible nmr and biochemical data: intermolecular recognition in a u4 spliceosomal complex. *J Mol Biol*, 388(2):283–298, May 2009.

[59] G. Lippens, K. Hallenga, D. Vanbelle, S. J. Wodak, N. R. Nirmala, P. Hill, K. C. Russell, D. D. Smith, and V. J. Hruby. Transfer nuclear overhauser effect study of the conformation of oxytocin bound to bovine neurophysin-i. *Biochemistry*, 32(36):9423–9434, 1993. Biochemistry-Us Biochemistry-Us.

[60] G. M. Lippens, C. Cerf, and K. Hallenga. Theory and experimental results of transfer-noe experiments .1. the influence of the off rate versus cross-relaxation rates. *Journal of Magnetic Resonance*, 99(2):268–281, 1992.

[61] S Liu, P Li, O Dybkov, S Nottrott, K Hartmuth, R Lührmann, T Carlomagno, and M C Wahl. Binding of the human prp31 nop domain to a composite rna-protein platform in u4 snrnp. *Science*, 316(5821):115–120, Apr 2007.

[62] R. E. London. Theoretical analysis of the inter-ligand overhauser effect: a new approach for mapping structural relationships of macromolecular ligands. *J Magn Reson*, 141(2):301–11, 1999. eng.

[63] R. E. London, M. E. Perlman, and D. G. Davis. Relaxation-matrix analysis of the transferred nuclear overhauser effect for finite exchange-rates. *Journal of Magnetic Resonance*, 97(1):79–98, 1992.

[64] A.D. MacKerel Jr., C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, volume 1 of *The Encyclopedia of Computational Chemistry*, pages 271–277. John Wiley & Sons: Chichester, 1998.

[65] S. Macura and R. R. Ernst. Elucidation of cross relaxation in liquids by two-dimensional n.m.r. spectroscopy (reprinted from molecular physics, vol 41, pg 95-117, 1980). *Molecular Physics*, 100(1):135–147, 2002.

[66] A M Mandel, M Akke, and A G Palmer. Backbone dynamics of escherichia coli ribonuclease hi: correlations with structure and function in an active enzyme. *J Mol Biol*, 246(1):144–163, Feb 1995.

[67] B. W. Matthews. Protein structure initiative: getting into gear. *Nat. Struct. Mol. Biol.*, 14(6):459–60, 2007.

[68] M. Mayer and B. Meyer. Mapping the active site of angiotensin-converting enzyme by transferred noe spectroscopy. *J Med Chem*, 43(11):2093–9, 2000. Journal of medicinal chemistry.

[69] M. Mayer and B. Meyer. Group epitope mapping by saturation transfer difference nmr to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc*, 123(25):6108–17, 2001. Journal of the American Chemical Society.

[70] S. L. McGovern and B. K. Shoichet. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J Med Chem*, 46(14):2895–907, 2003. eng.

[71] F A A Mulder, de Graaf RA, R Kaptein, and R Boelens. An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations. *J Magn Reson*, 131(2):351–357, Apr 1998.

[72] G. N. Murshudov, A. A. Vagin, and E. J. Dodson. Refinement of macromolecular structures by the maximum-likelihood method. *Acta. Crystallogr. D Biol. Crystallogr.*, 53(Pt 3):240–55, 1997.

[73] F. NI and H. A. Scheraga. Use of the transferred nuclear overhauser effect to determine the conformations of ligands bound to proteins. *Accounts Chem. Res.*, 27(9):257–264, 1994. Accounts Chem Res Accounts Chem Res.

[74] F. NI and Y. ZHU. Accounting for ligand-protein interactions in the relaxation-matrix analysis of transferred nuclear overhauser effects. *Journal of Magnetic Resonance Series B*, 103(2):180–184, 1994.

[75] Q Ning, D R Ripoll, Z Szewczuk, Y Konishi, and F Ni. Thrombin-bound conformation of a cyclic anticoagulant peptide using transferred nuclear overhauser effect (noe), distance geometry, and noe simulations. *Biopolymers*, 34(9):1125–1137, Sep 1994.

[76] N. R. Nirmala, G. M. Lippens, and K. Hallenga. Theory and experimental results of transfer noe experiments .2. the influence of residual mobility and relaxation centers inside the protein on the size of transfer noes. *Journal of Magnetic Resonance*, 100(1):25–42, 1992.

[77] S Nottrott, K Hartmuth, P Fabrizio, H Urlaub, I Vidovic, R Ficner, and R Lührmann. Functional interaction of a novel 15.5kd [u4/u6.u5] tri-snrnp protein with the 5' stem-loop of u4 snrna. *EMBO J*, 18(21):6119–6133, Nov 1999.

[78] J. Orts, S. K. Grimm, C. Griesinger, K. U. Wendt, S. Bartoschek, and T. Carlomagno. Specific methyl group protonation for the measurement of pharmacophore-specific interligand noe interactions. *Chemistry*, 14(25):7517–20, 2008. eng.

[79] J. Orts, J. Tuma, M. Reese, S. K. Grimm, P. Monecke, S. Bartoschek, A. Schiffer, K. U. Wendt, C. Griesinger, and T. Carlomagno. Crystallography-independent determination of ligand binding modes. *Angew Chem Int Ed Engl*, 47(40):7736–40, 2008. Angewandte Chemie (International ed.

[80] M. Pellecchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T. L. James, S. W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe, and G. Siegal. Perspectives on nmr in drug discovery: a technique comes of age. *Nat Rev Drug Discov*, 7(9):738–45, 2008. eng.

[81] J C Phillips, R Braun, W Wang, J Gumbart, E Tajkhorshid, E Villa, C Chipot, R D Skeel, L Kalé, and K Schulten. Scalable molecular dynamics with namd. *J Comput Chem*, 26(16):1781–1802, Dec 2005.

[82] B K Rai, G J Tawa, A H Katz, and C Humblet. Modeling g protein-coupled receptors for structure-based drug discovery using low-frequency normal modes for refinement of homology models: application to h3 antagonists. *Proteins*, 78(2):457–473, Feb 2010.

[83] S. G. Rasmussen, H. J. Choi, D. M. Rosenbaum, T. S. Kobilka, F. S. Thian, P. C. Edwards, M. Burghammer, V. R. Ratnala, R. Sanishvili, R. F. Fischetti, G. F. Schertler, W. I. Weis, and B. K. Kobilka. Crystal structure of the human beta2 adrenergic g-protein-coupled receptor. *Nature*, 450(7168):383–7, 2007.

[84] F Rázga, N Spackova, K Réblova, J Koca, N B Leontis, and J Sponer. Ribosomal rna kink-turn motif–a flexible molecular hinge. *J Biomol Struct Dyn*, 22(2):183–194, Oct 2004.

[85] F Rázga, M Zacharias, K Réblová, J Koca, and J Sponer. Rna kink-turns as molecular elbows: hydration, cation binding, and large-scale dynamics. *Structure*, 14(5):825–835, May 2006.

[86] D. C. Rees, M. Congreve, C. W. Murray, and R. Carr. Fragment-based lead discovery. *Nat. Rev. Drug. Discov.*, 3(8):660–672, 2004. Nat Rev Drug Discov Nat Rev Drug Discov.

[87] M. Reese, V. M. Sanchez-Pedregal, K. Kubicek, J. Meiler, M. J. J. Blommers, C. Griesinger, and T. Carlomagno. Structural basis of the activity of the microtubule-stabilizing agent epothilone a studied by nmr spectroscopy in solution. *Angew. Chem. Int. Edit.*, 46(11):1864–1868, 2007. Angew Chem Int Edit Angew Chem Int Edit.

[88] D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, X. J. Yao, W. I. Weis, R. C. Stevens, and B. K. Kobilka. Gpcr engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science*, 318(5854):1266–73, 2007.

[89] V. M. Sanchez-Pedregal, M. Reese, J. Meiler, M. J. Blommers, C. Griesinger, and T. Carlomagno. The inpharma method: protein-mediated interligand noes for pharmacophore mapping. *Angew. Chem. Int. Ed. Engl.*, 44(27):4172–5, 2005.

[90] J Schleucher and S S Wijmenga. How to detect internal motion by homonuclear nmr. *J Am Chem Soc*, 124(20):5881–5889, May 2002.

[91] D. M. Schneider, M. J. Dellwo, and A. J. Wand. Fast internal main-chain dynamics of human ubiquitin. *Biochemistry*, 31(14):3645–3652, 1992. Biochemistry-Us Biochemistry-Us.

[92] C. D. Schwieters, J. J. Kuszewski, and G. M. Clore. Using xplor-nih for nmr molecular structure determination. *Progr. Nucl. Magn. Res. Spec.*, 48:47–62, 2006.

[93] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore. The xplor-nih nmr molecular structure determination package. *J. Magn. Reson.*, 160(1):65–73, 2003.

[94] Suzanne B. Shuker, Philip J. Hajduk, Robert P. Meadows, and Stephen W. Fesik. Discovering high-affinity ligands for proteins: Sar by nmr. *Science*, 274(5292):1531–1534, 1996. http://www.sciencemag.org/cgi/content/abstract/274/5292/1531.

[95] S. F. Sousa, P. A. Fernandes, and M. J. Ramos. Protein-ligand docking: current status and future challenges. *Proteins*, 65(1):15–26, 2006. eng.

[96] M J Stocks, S Barber, R Ford, F Leroux, S St-Gallay, S Teague, and Y Xue. Structure-driven htl: design and synthesis of novel aminoindazole inhibitors of c-jun n-terminal kinase activity. *Bioorg Med Chem Lett*, 15(14):3459–3462, Jul 2005.

[97] S. J. Teague. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov*, 2(7):527–41, 2003. eng.

[98] L Trantírek, E Caha, P Kaderávek, and R Fiala. Nmr (13)c-relaxation study of base and sugar dynamics in gcaa rna hairpin tetraloop. *J Biomol Struct Dyn*, 25(3):243–252, Dec 2007.

[99] V Tugarinov, P M Hwang, and L E Kay. Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem*, 73:107–146, 2004.

[100] UniProt Consortium. The universal protein resource (uniprot). *Nucleic Acids Res*, 36(Database issue):190–195, Jan 2008.

[101] I Vidovic, S Nottrott, K Hartmuth, R Lührmann, and R Ficner. Crystal structure of the spliceosomal 15.5kd protein bound to a u4 snrna fragment. *Mol Cell*, 6(6):1331–1342, Dec 2000.

[102] M Vogtherr and K Fiebig. Nmr-based screening methods for lead discovery. *EXS*, (93):183–202, 2003.

[103] G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head. A critical assessment of docking programs and scoring functions. *J Med Chem*, 49(20):5912–31, 2006. eng.

[104] N Yanamala, A Dutta, B Beck, B van Fleet, K Hay, A Yazbak, R Ishima, A Doemling, and J Klein-Seetharaman. Nmr-based screening of membrane protein ligands. *Chem Biol Drug Des*, 75(3):237–256, Mar 2010.

[105] Q Zhang and H M Al-Hashimi. Domain-elongation nmr spectroscopy yields new insights into rna dynamics and adaptive recognition. *RNA*, 15(11):1941–1948, Nov 2009.

[106] Q Zhang, X Sun, E D Watt, and H M Al-Hashimi. Resolving the motional modes that code for rna adaptation. *Science*, 311(5761):653–656, Feb 2006.

[107] Q Zhang, R Throolin, S W Pitt, A Serganov, and H M Al-Hashimi. Probing motions between equivalent rna domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. *J Am Chem Soc*, 125(35):10530–10531, Sep 2003.

[108] J. Zheng and C. B. Post. Protein indirect relaxation effects in exchange-transferred noesy by a rate-matrix analysis. *Journal of Magnetic Resonance Series B*, 101(3):262–270, 1993.

[109] X. Zhu, J. L. Kim, J. R. Newcomb, P. E. Rose, D. R. Stover, L. M. Toledo, H. Zhao, and K. A. Morgenstern. Structural analysis of the lymphocyte-specific kinase lck in complex with non-selective and src family selective kinase inhibitors. *Structure*, 7(6):651–61, 1999. eng.