Development and application of NMR methods for challenges in drug discovery

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

im Promotionsprogramm Chemie
der Georg-August University School of Science (GAUSS)

vorgelegt von

Jens Pilger

aus Stuttgart

Göttingen, 2013

Betreuungsausschuss

- Prof. Dr. Christian Griesinger, NMR-based Structural Biology, MPI for Biophysical Chemistry
- Prof. Dr. Ulf Diederichsen, Organische und Biomolekulare Chemie, Georg-August Universität Göttingen

Mitglieder der Prüfungskommision

- Referent: Prof. Dr. Christian Griesinger, NMR-based Structural Biology, MPI for Biophysical Chemistry
- Korreferent: Prof. Dr. Ulf Diederichsen, Organische und Biomolekulare Chemie, Georg-August Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

- Prof. Dr. Henning Urlaub, Bioanalytical Mass Spectrometry, MPI for Biophysical Chemistry
- Prof. Dr. Bert de Groot, Computational Biomolecular Dynamics, MPI for Biophysical Chemistry
- Dr. Lars Kuhn, Biomolecular NMR Spectroscopy, European Neuroscience Institute Göttingen
- PD. Dr. Thomas Exner, Theoretische Pharmazeutische Chemie, Eberhard Karls Universität Tübingen

Tag der mündlichen Prüfung: 2.April 2013

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine ausser den angegebenen Literaturstellen und Hilfsmitteln verwendet habe.

Göttingen, 2013

Jens Pilger

Publications

The work presented in this thesis is published or will be published as follows:

1. **J. Pilger***, A. Mazur*, P. Monecke , H. Schrader T. Langer, A. Schiffer, M. Wegstroth, D. Lee, G. Hessler, K. U. Wendt, S. Becker and C. Griesinger:
   A Spin Diffusion Based NMR Method to Determine Protein-Ligand Complex Structures Demonstrated on Three Enzymes of Pharmazeutical Interest.

   -in final preparation-

2. A. Lange*, H. Sun*, **J. Pilger***, U. Reinscheid and H. Gross:
   Predicting the Structure of Cyclic Lipopeptides by Bioinformatics: Structure Revision of Arthrofactin.

   ChemBioChem, 2012, 13: 2671-2675

3. J.Wagner*, S. Ryazanov*, A. Leonov*, J. Levin*, S. Shi, F. Schmidt, C. Prix, F. Pan-Montojo, U. Bertsch, G. Mitteregger-Kretzschmar, M. Geissen, M. Eiden, F. Leidel, T. Hirschberger, P. Tavan, **J. Pilger**, M. Zweckstetter, T. Frank, M. Bähr, J. Weishaupt, M. Uhr, H. Urlaub, U. Teichmann, M. Samwer, K. Bötzel, M. Groschup, H. Kretzschmar, C. Griesinger and A. Giese:
   Anle138b: A Novel Oligomer Modulator for the Treatment of Prion and Parkinson Disease.

   Acta Neuropathologica, 2013, online ahead of print

*equal contribution

Abbreviations

AD	Alzheimer‘s disease
ADMET	Absorption, distribution, metabolism, excretion, toxicity properties
ATP/AMP	Adenosine triphosphate/monophosphate
Bac	Baccatin
BSE	Bovine spongiform encephalopathy
CDK	Cyclin-dependent kinase
CHO	Chinese hamster ovary
CJD	Creutzfeld-Jakob disease
CLP	Cyclic lipopeptide
COMT	Catechol-methyl-transferase
COSY	Correlation spectroscopy
CYP450	Cytochrome P450
DFG	Aspartic acid-phenylalanine, glycine
DMSO	Dimethylsulfoxid
DTT	Dithiothreitol
EC	Electron crystallography
EC50	Effective concentration 50
ECL	Extra-cellular loop
EDTA	Ethylene diamine tetraacetic acid
e.g.	Exempli gratia
EGTA	Ethylene glycol tetraacetic acid
EM	Energy minimization
Epo	Epothilone
ESI	Electron spray ionization
FA	Fluorescence anisotropy
FBDD	Fragment-based drug design
FFAR	Free fatty acid receptor
FL	Fluorescence lifetime
FT	Fourier transformation
GAFF	General amber force field
GBSA	Generalized Born surface area
GPCR	G-protein coupled receptor
HDX	Hydrogen deuterium exchange
HMBC	Heteronuclear multiple bond correlation

HPLC	High-performance liquid chromatography
HSA	Human serum albumin
HSQC	Heteronuclear single quantum coherence
i.e.	Id est
INPHARMA	Inter-ligand NOEs for pharmacophore mapping
ITC	Isothermal titration calorimetry
LC	Liquid chromatography
MAD	Multiple wavelength anomalous dispersion
MD	Molecular dynamics
MDMA	3, 4-methylenedioxymethamphetamine
MPTP	1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridin
MS	Mass spectrometry
MSM	Markov state model
MT	Microtubule
NS	Number of scans
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NRPS	Non-ribosomal peptide synthetase
PD	Parkinson‘s disease
PDB	Protein data base
PIPES	Piperazine-N,N-bis(2-ethanesulfonic acid)
PKA	Protein kinase A
PMF	Potential of mean force
PrP	Prion protein
RF	Radio frequency
RMSD	Root-mean-square-deviation
RP	Reversed phase
SAR	Structure-activity relationship
SIFT	Scanning for intensely fluorescent targets
SPE	Solid phase extraction
SPR	Surface plasmon resonance
STD	Saturation transfer difference
STI	STD, trNOE and INPHARMA combination
SULT	Sulfotransferase
TCEP	Tris(2-carboxyethyl)phosphine
TD	Time domaine
TE	Thioesterase
TEMPO	2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
TM	Transmembrane helix
TOCSY	Total correlation spectroscopy
trNOE	Transferred NOE

Summary

NMR spectroscopy plays an important role in all stages of the drug discovery process. Namely in the structure elucidation of natural products, synthetic ligands and metabolites, as well as a high-throughput screening technique. Yet, the application of NMR in structure-based drug design, e.g. in finding the binding mode of a small molecule drug to a macromolecular protein receptor is far from making the most from the opportunities available to it. Structure-based drug design is a powerful and widely used tool for the optimization of low molecular weight compounds that should be turned into highly efficient drugs. The method mainly relies on high-resolution crystal structures of the receptor-ligand complex to obtain the required information for optimizing target binding of small molecules. However, obtaining crystals and structures of sufficient quality cannot be achieved for nearly the half of pharmaceutically relevant protein targets. For those target proteins that cannot be crystallized, NMR spectroscopy is an alternative and structures of protein-ligand complexes can be determined, provided the protein can be labelled with stable isotopes such as ^{13}C or ^{15}N. However, pharmaceutically relevant non-crystallizable target proteins are often non-tractable by NMR, because they are too large and result in overcrowded spectra or they cannot be expressed in bacteria and therefore cannot be labelled with stable isotopes enabling heteronuclear NMR. In such cases one can employ INPHARMA (Inter-Ligand NOEs for PHARmacophore MApping). It utilizes two ligands that bind competitively to the same binding pocket of a protein. INPHARMA peaks in a NOESY spectrum emerge from the magnetization transfer from the protons of one ligand to the protons of the other ligand via the protein protons, provided the ligands dissociate from the protein several times during the NOESY mixing time. The method is further developed and it is investigated whether the methodology can be improved by inclusion of Saturation Transfer Difference (STD) restraints and transferred NOE (trNOE) restraints in addition to the INPHARMA restraints. STD is a frequently used technique in NMR spectroscopy and NMR-based screening for protein binders. The technique is developed and tested on protein kinase

A, where crystal structures of the protein/ligand complexes are known. The results show that the combination of the NMR methods INPHARMA, tr-NOE and STD results in a precise scoring function for docking modes and therefore the determination of ligand binding modes. It is demonstrated that the method is superior to docking scoring functions alone and can lead to the correct result by using a molecular dynamics simulation driven refinement, even if the initial conformation of the protein side chains is not correct. Multiplexing of several ligands improves the reliability of the scoring function further. Then the technique is extended the G-protein coupled receptor GPR40, a membrane protein, for which only homology models exist and which is an interesting drug target in on-going research. For this system, the ligand binding mode found is supported by SAR data. The binding mode of epothilone to tubulin, an important interaction for cancer therapy is reinvestigated using STD data. The binding mode found by INPHARMA is confirmed and further optimized, while the electron crystallography derived structure does not fit to the experimental NMR data. The NMR-based ligand binding mode determination method is presented to derive binding modes of ligands based on simple NMR experiments (NOESY and STD). It is demonstrated on the examples of PKA, GPR40 and the tubulin-epothilone complex, that based on a crystal structure or homology model of the protein, binding modes can be determined that can be used for pharmacophore mapping and drug optimization.

In the second part the drug metabolism of anle138b, a modulator of toxic protein oligomers in prion and Parkinson's disease is investigated. A methodology is developed to extract the drug from organs and to determine its concentration in the brain. It was confirmed that anle138b is the only active compound in the brain, while metabolites are only formed in liver and kidney. With combined HPLC, mass spectrometry and NMR techniques, the structures of the metabolites were determined and the drug metabolism of anle138b in the mice and rat model was revealed.

In the last part NMR spectroscopy is applied to reinvestigate the structural and stereochemical features of arthrofactin, a potentially antibiotic natural product. Arthrofactin was initially reported in 1993 as a bioactive cyclic lipopeptide from the bacterium *Pseudomonas sp.* The structure of arthrofactin and its derivatives was reassigned on the basis of extensive NMR experiments and chiral HPLC analysis. A new approach of phylogenetic structure prediction is tested and was successfully approved with NMR data.

In conclusion, NMR spectroscopy is applied and further developed in this thesis to several challenges of the drug discovery process.

Zusammenfassung

NMR Spektroskopie spielt eine wichtige Rolle in allen Arbeitsschritten der pharmazeutischen Wirkstoffentwicklung. Genauer gesagt in der Strukturbestimmung von Naturstoffen, synthetischen Liganden und Metaboliten, ebenso wie als Hochdurchsatz Screening Technik. Im Moment jedoch liegt die NMR Anwendung im Bereich der strukturbasierten Wirkstoffentwicklung, also im Bestimmen des Bindungsmodus eines kleinen Wirkstoffmoleküls an einen makromolekularen Rezeptor, weit hinter ihren Möglichkeiten. Die strukturbasierte Wirkstoffentwicklung ist ein mächtiges und weit verbreitetes Werkzeug für die Optimierung niedermolekularer Stoffe, welche zu effizienten Wirkstoffen verbessert werden sollen. Die Methode basiert hauptsächlich auf hochauflösenden Kristallstrukturen des Rezeptor/Ligand Komplexes um die Bindung des kleines Moleküles an seinen Ziel-Rezeptor zu optimieren. Dennoch können Kristalle und dementsprechend Kristallstrukturen für fast die Hälfte aller pharmazeutisch interessanten Zielproteine nicht realisiert werden. Für diejenigen Zielproteine die nicht kristallisiert werden können stellt NMR Spektroskopie eine Alternative dar um die Protein/Ligand Komplexstruktur zu erhalten, insofern das Protein mit stabilen Isotopen wie ^{13}C oder ^{15}N markiert werden kann. Dennoch sind pharmazeutisch relevante, nicht kristallisierbare Zielproteine oft nicht mit NMR messbar, da sie zu gross sind und zu überfüllten Spektren führen würden. Oder sie können nicht in Bakterien exprimiert werden, und damit nicht mit den für heteronukleare NMR notwendigen stabilen Isotopen markiert werden. In solchen Fällen empfiehlt sich die Verwendung der INPHARMA (Inter-Ligand NOEs für PHARmacophore MApping) Methode, welche zwei Liganden benötigt die um dieselbe Bindungsstelle eines Proteins konkurrieren. INPHARMA Signale innerhalb eines NOESY Spektrums entstehen durch den Transfer von Magnetisierung von den Protonen des Liganden über die Proteinprotonen zu den Protonen des anderen Liganden, wenn gegeben ist, dass die Liganden während der NOESY Mischzeit mehrmals die Bindungstelle verlassen. Die Methodik wird nun weiterentwickelt und es wird untersucht ob sich die Technik verbessern lässt, wenn zusätzlich Sättigungstransfer Differenz

(STD) Beschränkungen und transferierter NOE (trNOE) Beschränkungen zu den INPHARMA Beschränkungen verwendet werden. STD ist eine häufig verwendete Technik in der NMR Spektroskopie, speziell wenn es um NMR basiertes Screening von Proteinbindern geht. Die Methode wird entwickelt und getestet an Protein Kinase A; hier sind Kristallstrukturen der Protein/Ligand Komplexe bekannt. Die Ergebnisse zeigen, dass die Kombination der NMR Methoden INPHARMA, trNOE und STD in einer präzisen Bewertungsfunktion für Dockingmodelle resultiert und daher die Bestimmung des Liganden Bindungsmodus ermöglichen. Es wird gezeigt, dass die Methodik den rein theoretischen Docking-Bewertungsfunktionen überlegen ist. Weiterhin kann das korrekte Ergebnis auch durch eine Optimierung mittels molekulardynamischen Simulationen erhalten werden, selbst wenn die initiale Konformation der Aminosäure Reste nicht richtig ist. Eine weitere Verbesserung entsteht durch die Anwendung mehrer verschiedener Ligandkombinationen, genannt Multiplexing. Danach wird die Methode auf den G-Protein gekoppelten Rezeptor GPR40, einem Membranprotein, übertragen, von welchem nur Homologie Modelle existieren und welches in der aktuellen pharmazeutischen Forschung eine wichtige Rolle spielt. Für dieses System wird ein Bindungsmodus gefunden, der den SAR Daten der Literatur entspricht. Der Bindungsmodus von Epothilon an Tubulin, einer wichtigen Interaktion in der Krebsforschung, wird mittels STD Daten neu untersucht. Der Bindungsmodus der einst mit INPHARMA gefunden wurde wird bestätigt und weiter optimiert, während der Bindungsmodus der mittels Elektronen-Kristallographie gefunden wurde sich nicht mit den NMR Daten vereinigen lässt. Die Methodik zur NMR-basierten Wirkstoffbindungsmodus Bestimmung wird vorgestellt, um Bindungs-modi von Liganden mittels einfacher NMR Experimente (NOESY und STD) zu erhalten. An den Beispielen PKA, GPR40 und dem Tubulin-Epothilon Komplex wird gezeigt, dass basierend auf einer Kristallstruktur oder einem Homologie Modell des Proteins Bindungsmodi bestimmt werden können, welche dann zur Wirkstoffoptimierung genutzt werden können.

Im zweiten Teil wird der Metabolismus des Wirkstoffs Anle138b untersucht, einem Modulator von toxischen Protein Oligomeren der Prionen und Parkinson Krankheit. Eine Methode wurde entwickelt um Anle138b aus Organen zu extrahieren und seine Konzentration im Gehirn zu bestimmen. Es wird bestätigt, dass Anle138b der einzig aktive Wirkstoff im Gehirn ist, während Metabolite in den Organen Leber und Niere gefunden wurden. Mit einer Kombination von HPLC, Massenspektrometrie und NMR gelingt die Strukturbestimmung der Metaboliten und der Metabolismus von Anle138b im Maus und Ratten Modell wird erklärt.

Im letzten Teil wird NMR Spektroskopie angewandt, um die strukturellen

und stereochemischen Eigenschaften von Arthrofactin, einem potenziell antibiotisch wirksamen Naturstoff neu zu untersuchen. Arthrofactin wurde zuerst 1993 als bioaktives, cyclisches Lipopeptid des Bakteriums *Pseudomonas sp.* beschrieben. Die Struktur des Arthrofactins und seiner Abkömmlinge wurde neu bestimmt auf der Basis ausführlicher NMR und chiraler HPLC Analytik. Ein neuer Ansatz der phylogenetischen Strukturbestimmung wurde getestet und mittels NMR Daten verifiziert.

Zusammengefasst wird NMR Spektroskopie in dieser Dissertation auf einige aktuelle Herausforderungen der pharmazeutischen Wirkstoffentwicklung angewandt und weiterentwickelt.

Acknowledgement

I would like to thank:

- Prof. Dr. Christian Griesinger for giving me the oppertunity to work in these interesting and challenging projects, as well as for the support and discussions.
- Prof. Dr. Ulf Diederichsen for supervision from the university and second correction.
- Dr. Adam Mazur, Phillip Lottman, Dr. Karel Kubicek, Harindranath Kadavath, Dr. Stefan Becker, Melanie Wegstroth, Dr. Donghan Lee, Dr. Korvin Walter, Dr. Marcel Reese, Dr. Jennifer Tuma, PD. Dr. Thomas Exner, Dr. Tim ten Brink, Jan Will and Dr. Teresa Carlomagno for help, samples, data and discussions.
- Dr. Peter Monecke, Dr. Stefan Bartoschek, Dr. Alexander Schiffer, Dr. Bettina Elshorst, Dr. Herman Schrader, Dr. Thomas Langer, Dr. Gerhard Hessler and Dr. K. Ulrich Wendt from Sanofi for supplies and discussions and the fruitful collaboration.
- Prof. Dr. Uwe Reinscheid, Dr. Han Sun and Prof. Dr. Harald Gross for help, samples and discussions within the arthrofactin project.
- Sergey Ryazanov, Dr. Andrei Leonov, Prof. Dr. Henning Urlaub, Gerhard Wolf, Dr. Jens Weishaupt, Dr. Sebastian Kügler, Prof. Dr. Armin Giese and Jens Wagner for help, samples and discussions within the anle138b project.
- All of my NMR2 friends, especially Han, Rakhi and Guo-Wei, for the pleasant time.

I heartily thank my wife Angelika, my daughters Alice and Jana and my parents Susann and Peter for all the joy and support.

Contents

1 Introduction 1
- 1.1 Structure-based drug design 1
  - 1.1.1 The drug discovery process 1
  - 1.1.2 3-D protein structures 3
  - 1.1.3 Molecular docking 6
  - 1.1.4 Pharmacophore mapping 7
  - 1.1.5 Molecular dynamics simulations 9
  - 1.1.6 Drug metabolism 11
- 1.2 Nuclear Magnetic Resonance 14
  - 1.2.1 NMR applied to drug discovery 14
  - 1.2.2 The Nuclear Overhauser Effect (NOE) 17
  - 1.2.3 Transferred NOE (trNOE) 25
  - 1.2.4 INPHARMA 28
  - 1.2.5 Saturation transfer difference (STD) 31

2 Protein kinase A 35
- 2.1 Introduction 35
  - 2.1.1 Protein kinases as drug targets 35
  - 2.1.2 ATP binding site 36
- 2.2 Material & Methods 39
  - 2.2.1 Protein expression and sample preparation 39
  - 2.2.2 NMR spectroscopy 39
  - 2.2.3 Docking and molecular dynamics simulations 40
  - 2.2.4 Back-calculation of peak volumes 40
- 2.3 Results I - Method development 40
  - 2.3.1 Molecular docking 41
  - 2.3.2 INPHARMA 42
  - 2.3.3 STD 47
  - 2.3.4 INPHARMA-STD 50
  - 2.3.5 Multiple ligands (Multiplexing) 53
  - 2.3.6 Molecular dynamics simulations and refinement 56

2.4 Results II - Application and testing 60
 2.4.1 Ligands and conformational changes 60
 2.4.2 Advanced binding mode I - Biased crystal structure . . 60
 2.4.3 Advanced binding mode II - Backpocket penetration . 62
 2.4.4 Advanced binding mode III - Docking failure 67
2.5 Discussion & Conclusion 73

3 G-protein coupled receptor 40 77
3.1 Introduction 77
 3.1.1 G-protein coupled receptors 77
 3.1.2 GPR40: Homology models and proposed ligand binding modes 81
3.2 Materials & Methods 84
 3.2.1 Sample preparation 84
 3.2.2 NMR spectroscopy 85
 3.2.3 Docking and alignment 85
 3.2.4 Energy minimization and molecular dynamic simulations 85
 3.2.5 Back-calculation of peak volumes 85
3.3 Results 87
 3.3.1 Docking, STD and INPHARMA Ligands 5 and 6 . . . 87
 3.3.2 Modeling and molecular dynamics simulations 90
 3.3.3 STD Ligand 7 93
3.4 Discussion & Conclusion 95

4 Tubulin-epothilone complex 103
4.1 Introduction 103
 4.1.1 Tubulin and antimitotic drugs 103
 4.1.2 The binding mode of epothilone 105
4.2 Material & Methods 108
 4.2.1 STD experiments 108
 4.2.2 Back-calculation of peak volumes 109
 4.2.3 Molecular dynamics simulations 109
4.3 Results 110
 4.3.1 STD experiments 110
 4.3.2 Molecular dynamics simulations 113
4.4 Discussion & Conclusion 121

5 Drug metabolism of anle138b 123
5.1 Introduction 123
 5.1.1 Toxic protein aggregates in neurodegenerative diseases 123
 5.1.2 Anle138b - a novel aggregation modulator 124

5.2 Material & Methods 126
5.2.1 HPLC analysis 126
5.2.2 Mass spectrometric analysis 126
5.2.3 NMR analysis 126
5.3 Results 127
5.3.1 Anle138b analysis 127
5.3.2 Metabolite analysis 130
5.3.3 Pharmakokinetic studies 133
5.4 Discussion & Conclusion 135

6 Structure revision of arthrofactin 139
6.1 Introduction 139
6.2 Materials & Methods 141
6.2.1 Sample production and bioinformatics 141
6.2.2 NMR spectroscopy 141
6.3 Results 143
6.3.1 Bioinformatic studies 143
6.3.2 NMR assignment 144
6.4 Discussion & Conclusion 147

Bibliography 149

A Protein kinase A 171
A.1 Rescoring of docking poses with GOLD 171
A.2 NOESY spectra of ligand combinations 173

B G-protein coupled receptor 40 179
B.1 NMR spectra of ligand combination 179

C Anle138b 183
C.1 Mass spectrometric and NMR spectroscopic data 183
C.2 NMR spectra of anle138b and its metabolites 184
C.3 Fragmentation patterns of the metabolites 188

D Arthrofactin 191
D.1 Assignment of arthrofactin 191
D.2 NMR spectra of arthrofactin 193

Chapter 1

Introduction

1.1 Structure-based drug design

1.1.1 The drug discovery process

Health is one of the most important aspects of life. Therefore a main aspect of science is to provide the knowledge of medicine, therapies and drugs to cure sick persons and restore their health and quality of life. Natural products that can be used as drugs are known during the whole history of mankind. Still, it is only a century ago, that is was possible to develop drugs in a rational way. This methodology is called drug discovery and aims to identify and modify molecules chemically to fit best the designation of a drug, being able to cure a disease. The first thoughts about a rational drug development process came up at the end of the 19th century, when Paul Ehrlich postulated the existence of chemoreceptors [Ehrlich, 1900]. The structure of suitable molecules have to be modified and optimized to best fit these receptors. This was expressed in the phrase: 'we have to learn to aim chemically'. Indeed, in his laboratories the first screening was done by using hundreds of new synthesized organic arsenic compounds to discover arsphenamine to cure syphilis [Ehrlich, 1910]. This ground-breaking study influenced the following generations and led to the discovery of many drugs, e.g. penicillin [Fleming, 1929] and the establishment of chemotherapy [Strebhardt and Ullrich, 2008]. In 1905 the concept of receptors was refined by describing them as switches that can be activated (agonists) and blocked (antagonists) [Langley, 1905]. But it took until the early fifties that this finding could be exploited, when the different forms of adrenergic receptors were described [Ahlquist, 1948] and drugs like adrenalin (β-adrenoreceptor agonist), β-blockers or benzodiazepines were discovered and further developed [Drews, 2000].

The driving force of drug discovery is still high-throughput screening

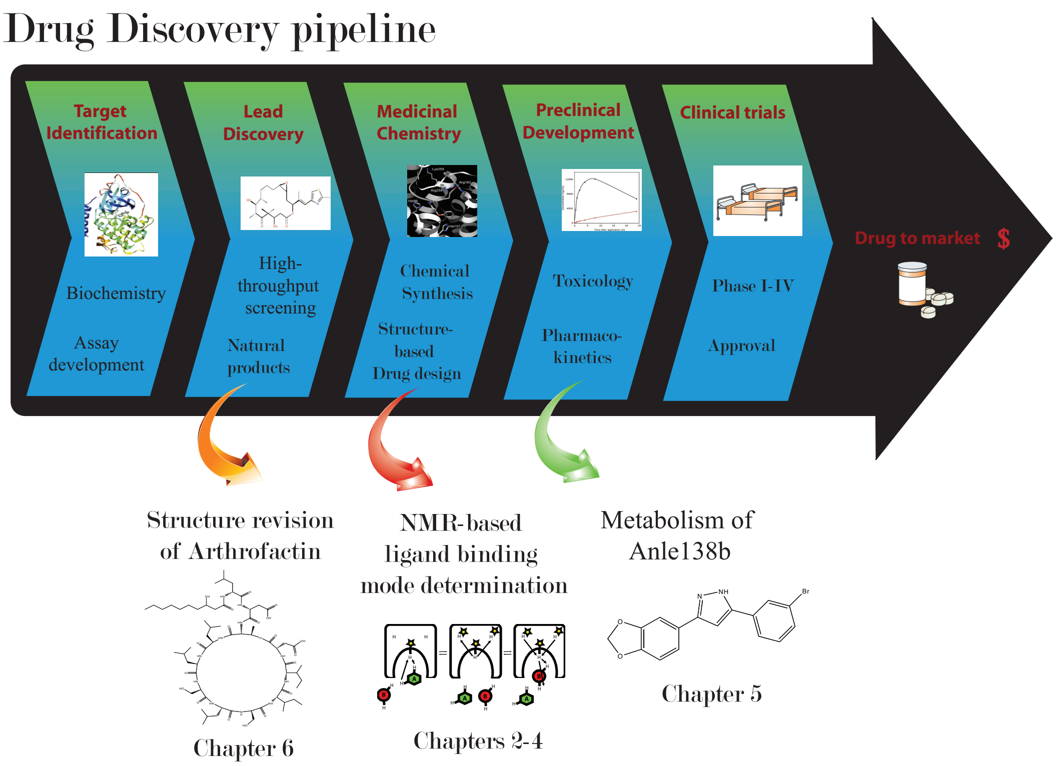

Drug Discovery pipeline
Target Identification
Biochemistry
Assay development
Lead Discovery
High-throughput screening
Natural products
Medicinal Chemistry
Chemical Synthesis
Structure-based Drug design
Preclinical Development
Toxicology
Pharmaco-kinetics
Clinical trials
Phase I-IV
Approval
Drug to market $
Structure revision of Arthrofactin
Chapter 6
NMR-based ligand binding mode determination
Chapters 2-4
Metabolism of Anle138b
Chapter 5

Figure 1.1: Overview of the drug discovery process from target identification to approval for market. Indicated are the topics investigated in this thesis.

[Macarron et al., 2011, Lahana, 1999] which is nowadays fully automated and provides the 'hits', which are binding scaffolds that are then optimized to lead structures.

Hand in hand with the improvements in structural biology, more 3-D structures of proteins became available and the medicinal chemist can now optimize compounds in a more rational and faster manner. These structural informations also provide the opportunity to screen molecules in a virtual manner, which is obviously less resource-expensive [Kitchen et al., 2004] and still provides a good selection [Clark, 2006]. Additionally the properties of drugs have been understood more profoundly today. Especially with the appliance of the Lipinski rules of five, which states that: a drug molecule should have a i) lipophilicity of logP below 5, ii) weight less than 500 Da and has iii) less than 5 hydrogen bond donors and 10 acceptors [Lipinski et al., 2001]. Implementation of such rules helps to identify drugs at an early stage, that would fail in later stages due to toxicity or too low bioavailability [Kubinyi, 2003]. Besides the developments in screening, the application and optimizations by the addition of functional groups to natural products as drugs has again come into the focus [Koehn and Carter, 2005].

The overall drug discovery process (Fig. 1.1) is very long and very cost intensive [Lombardino and Lowe, 2004]. The estimations of time and money that are spent differ a lot, but all are in an average range of 10 years and 1 billion dollar [Adams and Brantner, 2006]. To save resources in the development phase it is even common to use known drugs on new targets [Haupt and Schroeder, 2011].

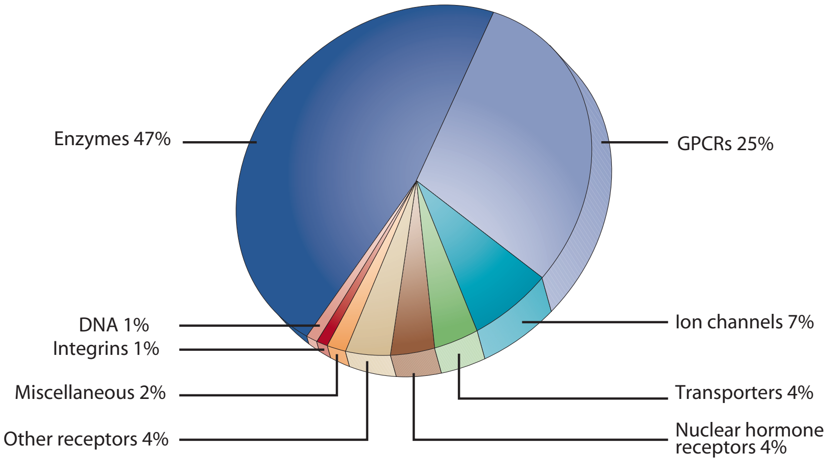


Figure 1.2: Marketed small-molecules drug targets by biochemical classes [Hopkins and Groom, 2002].

An interesting questions arising from the drug discovery process is: how many drug targets are there in the end? The human genome comprises around 30.000 genes which encode for a much higher number of proteins, if alternative splicing, post-translational modifications and protein complex formations are also considered. Still, predicting a number of how many of these proteins can be targeted by drugs is not possible. The number of approved drug targets is approximately 324 [Overington et al., 2006] and an estimation of 600-1500 possible drug targets is proposed in the literature [Hopkins and Groom, 2002]. As seen in Fig. 1.2 most of the approved drugs are enzymes like protein kinases, but more than 40% are membrane proteins like G-protein coupled receptors, ion channels and transporters. Interestingly the membrane proteins are very difficult to crystallize and therefore for most of these important targets no 3-D crystal structures are available.

1.1.2 3-D protein structures

The starting point of structure-based drug design is a 3-dimensional model of the macromolecular receptor with atomic resolution. Such a model can be provided by the free apo-form of the protein or the holo-form, when a ligand is bound. The latter provides even more information, especially when dealing with an induced-fit situation. But the holo-form also comprises the side chain conformations of the bound ligand, which might change when using another binder. 3-D protein structures are open to the public, due to the RCSB Protein Data Bank (PDB, www.rcsb.org/pdb/) [Berman et al., 2000], where these data can be deposited. In 2012 the PDB contained around 80.000 structures, many of them with a bound ligand.

The standard technique to obtain a 3-D structure of a molecule is X-ray crystallography, which was applied for 90% of the structures found in the PDB. The technique is well established and documented with its benefits and limitations [Davis et al., 2003]. Initially X-ray crystallography was used only for the structure elucidation of small molecules. This is done since a century and the world repository, the Cambridge Structural Database (CSD, www.ccdc.cam.ac.uk), comprises around half a million deposited structures. The scattering pattern of protein crystals was obtained already for pepsin in 1934 [Bernal and Crowfoot, 1934]. Structure calculation was possible at that time for small molecules but due to the lack of highly efficient computational resources, protein structures were not solved. The landmark event of protein X-ray crystallography happened in 1958, when the structure of myoglobin [Kendrew et al., 1958] was solved. The technique is based on the observation that an X-ray photon can be scattered by an electron by producing secondary, spherical waves that can be detected. In practice, X-ray crystallography relies on a crystal of the protein, which is often only obtained in a time consuming process. Especially in the case of membrane proteins it is often not possible. Given that a crystal exists, it is exposed to an X-ray beam and investigated from all orientations, by rotating the crystal. Hereby the crystal must be of sufficient quality to remain stable. During this process, the X-ray diffraction is detected and results in a distinct diffraction pattern. Hereby every spot of the diffraction pattern represents one lattice plane and depends on i) the unit cell, ii) wavelength and iii) crystal orientation in the beam. The unit cell is the imaginary smallest unit inside the crystal, which is normally around 50 $Å^3$ in size. A second challenging task, after obtaining the crystal, is to solve the phase problem. Only the amplitudes can be directly derived from the measured intensities, but not the phases. The technique most frequently used to solve this problem nowadays is multiple wavelength anomalous dispersion (MAD). Therefore anomalous diffraction is recorded at different wavelength, created by a special atom like selenium which can be brought into the protein during the expression by the modified amino acid selenomethionine. Another method is isomorphous displacement, which relies on the soaking of heavy atoms into the crystal structure. This has the disadvantage that several crystals with different heavy atoms have to be grown, largely increasing the amount of necessary protein. When the phase problem is solved, which has to be done only once for a specific protein, an electron density map is obtained and the crystallographer can fit the protein coordinates into the electron density. This task becomes easier with better resolution of the density map, which is normally around 2 Å. The structure obtained by X-ray crystallography is not a photograph in atomic detail, but is a model supported by strong experimental data [Podjarny et al., 2011]. Once the structure of a protein is solved, complexes with many different ligands can be obtained very fast. For the approach of fragment based drug design, it is even common to do a high-throughput crystallography [Blundell et al., 2002].

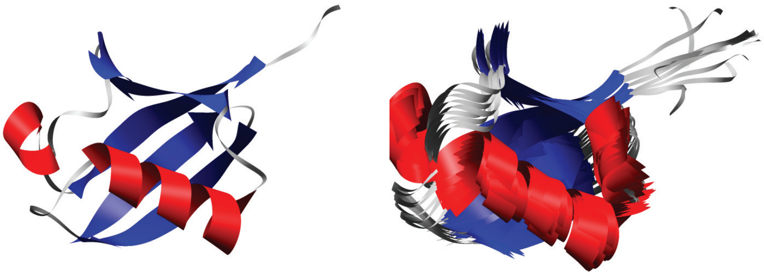


Figure 1.3: 3-D structure of ubiquitin, derived as a single model by X-ray crystallography (left) and as a structure ensemble by NMR spectroscopy (right).

The second important method to reveal 3-D structures of proteins is NMR spectroscopy and it accounts for approximately 10% of the structures in the PDB. The advantage in comparison to X-ray crystallography is, that no crystal is needed and the sample can be measured in solution, which also excludes the problem of crystal contact sides. The limitation on NMR is first the size of the protein, being typically less than 40 kDa, even though much larger structures have also been solved or were investigated, e.g. the 82 kDa protein malate synthase [Grishaev et al., 2008]. The second limitation of NMR is the possibility to express the protein in bacteria and label it with heteronuclear NMR-enabling stable isotopes like ^{13}C or ^{15}N. Protein structures can be calculated based on the experimental NMR data of i) NOESY spectra that yield distances and ii) chemical shifts, that provide chemical environments. Additional informations like residual dipolar couplings (RDC) or pseudo-contact shifts (PSD) are frequently used. The application of these experimental restraints requires an assignment of most of the protein signals in the NMR spectrum, which can be a very time consuming task, hampering the industrial workflow. The NMR methodology is described in detail in chapter 1.2.

1.1.3 Molecular docking

Given that a model with atomic resolution exists for a protein target, the binding of small molecules can be simulated with an approach termed molecular docking. This structure-based approach starts from a model of the apo-protein (uncomplexed conformation) or the holo-protein (complexed conformation). The ligand is now computationally placed into the binding pocket of the protein structure to propose a suitable binding mode. Molecular docking is done in two steps called posing and scoring. The quality of posing decides if the ligand can be docked correctly. Furthermore the quality of the scoring functions decides, if the correct binding mode is really identified. Posing refers to the actual docking, the placement of the ligand, and scoring to an energetic evaluation of the ligand binding pose. Posing can be done in a systematic or random manner. Systematic algorithms explore all degrees of freedom in the molecule and often grow the ligand incrementally in the binding site and is implemented in the programs FlexX, Dock or Glide. Random search approaches are Monte Carlo simulations like in AutoDock or genetic algorithms as in GOLD.

The quality of a docked ligand pose can be evaluated with a scoring function, based on i) a force field, ii) empirical or iii) knowledge-based manner [Kitchen et al., 2004]. Force fields like Amber or Tripos can be used to calculate the energy of the protein-ligand interaction, as well as the energy inside the ligand. This approach is accurate, but time consuming. Much faster is the empirical approach, that applies potentials obtained by experimental data sets to score the binding energies in a fast manner. As these potentials represent a simplification of the force fields they are not as accurate as the further. Knowledge-based scoring functions like Astex Statistical Potential (ASP) or DrugScore rely on empirical data more than on binding energies.

The state-of-the-art program GOLD [Jones et al., 1995] applies a genetic algorithm for posing and an empirical (ChemScore) or a force field based scoring function (GOLDscore). It was shown that the best scoring results can be obtained, when consensus scoring is done, which means that the docking structures are rescored by different scoring functions, e.g. ChemScore, GOLDScore and ASP [O'Boyle et al., 2009]. Comparable with GOLD is the new docking program PLANTS [Korb et al., 2006], which uses an ant colony algorithm for posing and the empirical scoring function ChemPLP [Korb et al., 2009a]. It is a consensus in the drug design community that a docked structure is considered correct, if the RMSD to the crystal structure is less than 2 Å, as can be seen in the comparisons and tests of docking programs, e.g. [Meiler and Baker, 2006, Moitessier et al., 2008, Cross et al., 2009, Cortés Cabrera et al., 2012]. In modern approaches the small molecule is kept flexible during the docking and the protein rigid, or partially flexible with specific rotatable amino acid residues. If a ligand binds to the surface of the protein, it is advisable to use algorithms for protein-protein docking (e.g. HADDOCK [Dominguez et al., 2003]), which are specifically developed for the purpose that interaction of proteins take place on their surfaces. Here the protocols start with rigid molecule docking, which is then refined with semi-flexible side chains, flexible backbone and refinement in a water solvent environment. Docking results can be improved by the inclusion of experimental NMR data in the programs HADDOCK and PLANTS [Korb et al., 2010b].

The docking approach relies on a protein model of high quality from X-ray crystallography or NMR spectroscopy. If no model is available, a homology model can be used, but results have to be treated very carefully [Hillisch et al., 2004]. Studies in the field of GPCR modeling have shown the difficulty of the approach. Failure is often due to wrong amino acid conformations [Michino et al., 2009, Kufareva et al., 2011], that block the correct orientation of the ligand.

1.1.4 Pharmacophore mapping

The pharmacophore concept is a very useful technique for the identification of hit molecules. Additionally it can be used for binding site mapping or virtual screening, if no structure of the receptor is available or the docking approach would be too time-consuming. The method is ligand-based and focusses on the assignment of chemical features of the ligand to find other molecules with similar properties by ligand alignment. The IUPAC definition of a pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger or to block its biological response [Wermuth et al., 1998].

In principle, the approach could be started with one binder that has to be superimposed to all molecules in a library, in order to find similar structure elements and identify new binders. This concept can be strongly improved by screening not for a ligand structure but for the pharmacophore, including the necessary structure-activity-relationship informations of donors, acceptors or aromatic moieties. The more known binding ligands are taken into account, the preciser the information of the pharmacophore will be. Such a pharmacophore alone gives already a map of the binding site, even if the structure of the binding site is not known. The chemical features of the pharmacophore are then screened through a compound library and often result in specific binding ligands [Langer and Hoffmann, 2006].

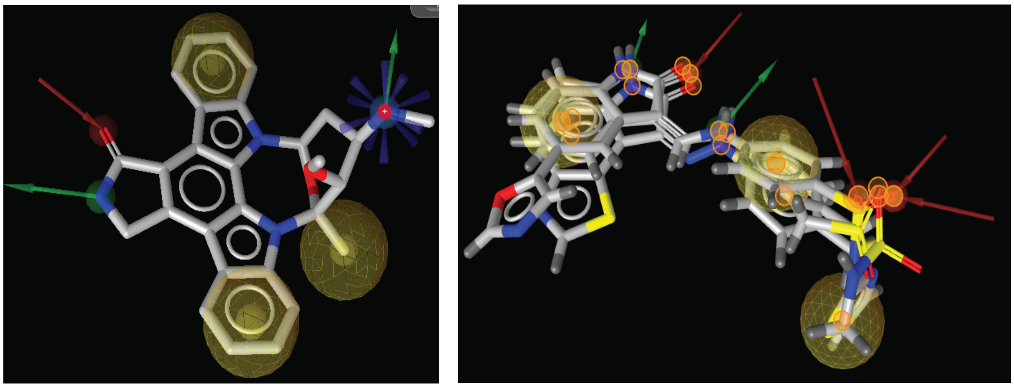


Figure 1.4: Pharmacophore of staurosporine (left) derived from the crystal structure of the complexes (pdb structures 1STU). Green arrows indicate hydrogen donors, red arrows indicate hydrogen acceptors and spheres illustrate the aromatic features and methyl groups. On the right a pharmacophore derived from the alignment of multiple ligands that bind CDK2 (image taken from http://macinchem.org/reviews/ligandscout-update.php). Hereby the orange circles indicate common pharmacophore features. The protein structures were omitted for clarity.

If several molecules are to be taken into account for deriving the pharmacophore, they have to be superimposed. This 3-D alignment is a challenge, as it incorporates the problem of conformational flexibility [Wolber et al., 2006]. Normally this problem is solved by the creation of many 3-D conformations of each molecule, which are then aligned pairwise. A scoring function chooses then the best fitting pair. Still, this process of alignment and assignment of the chemical features has to be inspected visually and is often driven by experience. The chemical features can be assigned in the most effective manner, if a 3-D structure of the ligand-protein complex is available and the donor/acceptor interactions are visible [Steindl et al., 2006]. This is exemplary shown in Fig. 1.4, where the pharmacophore of staurosporine bound to PKA was automatically derived. If the crystal structure of the complex is available, the pharmacophore can be assigned in the most efficient manner. Obviously the methodology can only work, if the ligand binding mode to the protein is correct. Besides the staurosporine figure, a figure is shown of multiple aligned crystal structures of ligands that bind CDK2. This multiple alignment gives a more robust pharmacophore and illustrates the necessity to

correctly align ligands. These can be very similar in shape as in the example or completely different.
When a 3-D structure of the protein is available, one can apply virtual screening using molecular docking, but the pharmacophore approach is orders of magnitudes faster. The state-of-the-art program for aligning molecules, assigning pharmacophores and perform virtual screening is Catalyst [Kurogi and Güner, 2001]. LigandScout [Wolber and Langer, 2005] displayed a comparable performance [Wolber et al., 2008]. Recently the program pharmACOphore [Korb et al., 2010a] for molecule alignment was implemented in the docking program PLANTS and follows a new ant colony optimization based algorithm.

1.1.5 Molecular dynamics simulations

3-D structures of a molecule represent a static view, which is of course not the reality. Molecules with rotatable bonds are often highly dynamic and this has to be taken into account, when they are investigated at the molecular level. The movement at an atomic resolution can be efficiently described with the molecular dynamics (MD) simulation approach. The basic equation for the dynamics of a quantum system and is evolution in time (wave function) is the Schrödinger equation:

$$i\hbar\partial t\psi(r,R) = H\psi(r,R) \tag{1.1}$$

Where i is the imaginary unit, ψ the wave function and H the Hamilton operator, indicating that the motion is based on energy. The calculation effort that has to be full-filled in order to run such a calculation for a whole protein is too high to be accomplished. Therefore the atoms are described by the classic Newton mechanics. The electrons are ignored and only the nuclei are described (Born-Oppenheimer approximation) by the classical Newton equation of motion:

$$m_i \frac{d^2}{dt^2}\vec{R}_i = -\vec{\nabla}_i E(\vec{R}_1) \tag{1.2}$$

Two different interactions between nuclei are distinguished: those nuclei i) connected by a bond and those ii) close in space. The interactions are described as physical springs that undergo attraction and repulsion. Bond parameters are bond-length, bond angle, dihedral bond angle and extra-planar bond angle. Nuclei close in space which are not connected by a bond are described by the Lennard-Jones potential in the case of Van-der-Waals interaction and with an electrostatic potential in the case of a charged interaction. Bond length, angles and charges of the specific atom types are described by the force field and have to be parametrized. These parameters result from empirical and quantum chemical investigations. Several different force fields are available, for example Amber, CHARMM, Gromos or Tripos. When the atoms of the molecule are parametrized, the integration of the Newtonian equation of motion is done numerically via the leapfrog scheme. Therefore, the equation is solved for all atoms in the system, followed by a step where the atoms move and then a second step of integration. Attraction and repulsion of the mechanically treated nuclei produce the movement of the whole biopolymer. A more detailed description of the system, if needed, is obtained when the environment is also simulated, i.e. the solvent, ions or the lipid bilayer in the case of a membrane protein.

The limitation of MD is the amount of computer time that is necessary. Today protein-water simulations in the ns range are standard, extendable to μs. MD simulations in the millisecond range were so far only achieved by the super-computer Anton [Shaw et al., 2010] or the distributed-computing project Foldingathome [Voelz et al., 2010]. MD simulations start from a crystal structure, a homology model, a de novo protein model or an NMR structure, whereas NMR data can also restrain an MD trajectory. MD simulations are a great help for the understanding of very different kinds of protein motions [Henzler-Wildman and Kern, 2007]. For example the function of the aquaporin water channel was explained by an MD simulation [de Groot and Grubmüller, 2001].

Recently it was possible for the first time to simulate the way of a charged ligand into its binding pocket during an MD simulation and to completely reconstruct the ligand binding process in terms of energy and kinetics [Buch et al., 2011]. Therefore 495 MD simulations of 100 ns each were obtained from the enzyme-inhibitor complex trypsin-benzamidine and showed the binding event with an RMSD less than 2 Å from the crystal structure in 187 cases. A Markov state model (MSM) was produced and the obtained potential of mean force (PMF) is show in Fig. 1.5 and reveals several metastable states. It is interesting to see how often the ligand resides at other binding sites and gives an idea about the unspecific binding events of a ligand. Simulating the binding event of a ligand to a protein is very interesting but often not possible, as the simulation time is not long enough for an occurrence of a binding event or too high energy barriers that have to be crossed. An alternative to the latter can be metadynamics, that have shown to reproduce correct binding modes by the combination of MD and a new methodology to cross the energy barriers [Gervasio et al., 2005]. Besides its functionality in describing protein mechanisms and ligand movements in the binding site, MD is very useful to sample a wide conformational space, especially in the case of small molecules. If described accurately by the force field, the conformational analysis can provide crucial information for structure determination [Sun et al., 2011].

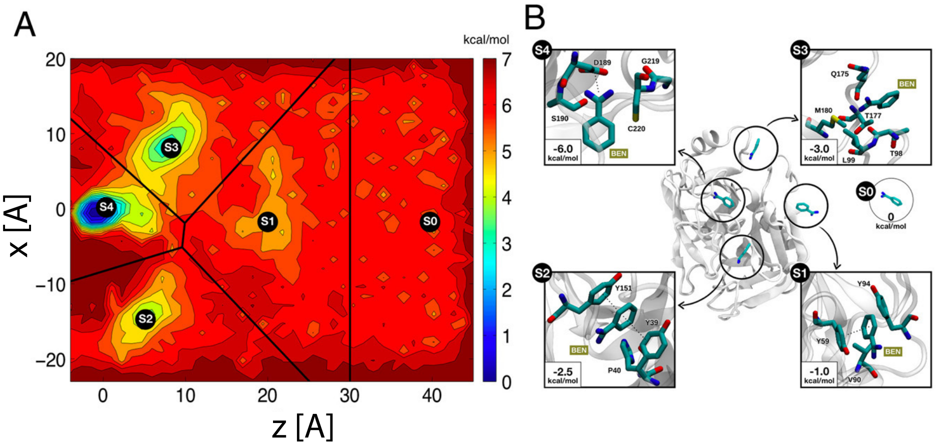


Figure 1.5: PMF in the xz plane identifies five different metastable states (S0-S4) of the ligand benzamidine binding to trypsin. The states are represented on the right site and illustrate specific and unspecific binding events of the ligand [Buch et al., 2011].

1.1.6 Drug metabolism

The discovery and optimization of a drug molecule to inhibit a specific macromolecule is just one side of the medal. It is essential to understand more than the action of the drug to the organism, which is termed pharmacodynamics. Equally important is pharmacokinetics, which is the action that the organism takes on the drug, also called drug metabolism. Five questions have to be answered to understand the pharmacokinetics: i) absorption (A), is the drug bioavailable? ii) distribution (D), does the drug reach the supposed location of action, e.g the brain? iii) metabolism (M), how is the drug biotransformed? iv) excretion (E), does the drug leave the organism after a certain time and how? and v) toxicity (T), is the drug or one of the metabolites toxic? These characteristics of the drug are summarized as the ADMET properties and need to be determined for every potential drug candidate. The reason for withdrawal of a drug, besides the lack of efficacy, is most often toxicity and side effects [Schuster et al., 2005]. Therefore ADMET properties are often the most difficult part of the drug discovery process. Given that no toxic metabolites are produced, every new drug is regularly tested for inhibition of metabolic enzymes or of vitally essential structures as the hERG heart channel.

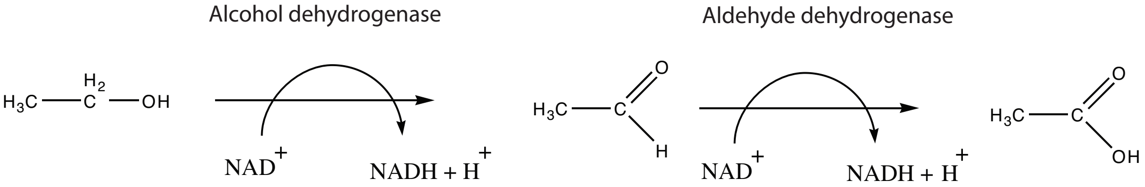


Figure 1.6: Schematic metabolism and elimination of ethanol.

A common example for drug metabolism and ADMET properties is ethanol (Fig. 1.6). Ethanol has a very good bioavailability and as it passes through all membranes it is distribution everywhere in the organism. Ethanol is oxidated by the enzyme alcohol-dehydrogenase to ethanal. This substance gets further oxidized by aldehyde-dehydrogenase to acetic acid, which can enter the citric acid cycle or be stored in fatty acid chains. Thus, ethanol can be completely eliminated during metabolism. If ethanol is not pure and for example contaminated with methanol, the drug metabolism is very dangerous. Methanol follows the same metabolic pathway, but the metabolic products methanal (formaldehyde) and formic acid are highly toxic.

Normally drugs are not completely metabolized as it is the case for ethanol. Elimination takes also place over the kidneys (urine) and the liver (bile). The biotransformation is normally done by the cytochrome P450 (CYP450) enzyme superfamily 1.7. All members of the CYP450 family are attached to membranes, most often in liver cells, and possess a heme group in the reactive center. Differences in the size of the binding pocket make them suitable to oxidize or reduce most xenobiotic substances that enter the organism. Examples for the reactions are demethylenation, demethylation, N-dealkylation and hydroxylation. The reaction mechanisms, which are often coupled with NADPH-P450-reductase, are rather complicated and documented in the literature [Zhang et al., 2008]. These oxidative and reductive reactions are termed the phase I reactions of the metabolism. Additionally to the CYP450 enzymes, the actions of alcohol-dehydrogenase, aldehyde-dehydrogenase, monoamine oxidase (MAO) or different esterases belong to the phase I group.

In contrast, phase II reactions comprise the addition of a functional group to the drug in the cytosol. These phase II reactions often follow directly a phase I reaction, for example the opening or addition of a hydroxy group. The most important functional group added is glucose (glykosylation) by the enzyme UDP-transferase. Very common is also a sulfate (sulfurylation) group, done

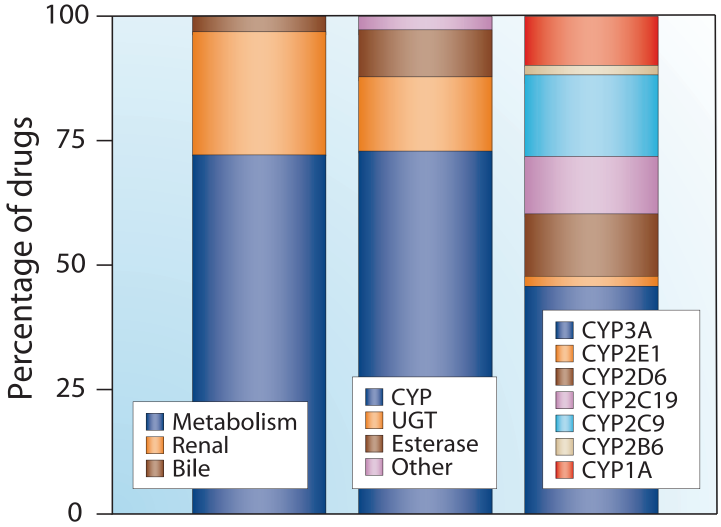


Figure 1.7: Routes of elimination of the top 200 most prescribed drugs in 2002. Most drugs are eliminated by biotransformation during metabolism. More than three quarters of the drugs are metabolized by the cytochrome P450 superfamily. The majority of reactions is carried out by CYP3A, CYP2C9 and CYP2D6 (from [Wienkers and Heath, 2005]).

by the SULT family, especially SULT1A3. Further functional groups comprise i) methyl-groups, which are transferred by catechol-O-methyltransferase (COMT), ii) gluthatione, transferred by gluthathione-S-transferase and iii) acetate, transferred by acetyl-CoA.

Investigation of drug metabolism in the pharmaceutical industry is an established procedure. Due to the very complicated issue of drug-drug interactions, it is often not possible to predict the metabolites of a drug correctly in advance. Urine or blood samples from humans or animals are taken at different time points to measure the concentration of the drug to understand the absorption and elimination behavior. In the case of animals, organ samples can be taken as well to investigate the distribution in the body and check for example if a compound passes the blood-brain-barrier. After extraction, the compound can by separated and quantified with HPLC. The standard procedure for the investigation of biotransformations and identifications of metabolites is LC/MS [Clarke et al., 2001]. There are many softwares, e.g. MetaSite [Cruciani et al., 2005] that predict metabolic sites in small molecules and can speed up the metabolite identification. For structure elucidation of the metabolites the LC/MS and MS/MS approach is often sufficient, as there is only a limited number of functional groups (see above) that can be attached to the compound during drug metabolism. Even though, it is often not pos-

sible to predict the exact metabolic site, e.g. in which position a hydroxy group is added to a phenyl ring. In this case the metabolite structure can be elucidated with NMR, given that the metabolite can be concentrated high enough to enable 2-dimensional NMR.
When the crystal structure of CYP2C9 was solved [Williams et al., 2003], it revealed the complexity of the drug transformation. In Fig. 1.8 the drug warfarin is bound to CYP2C9, but in a position with the closest distance 10 Å away from the catalytic heme site iron. Still, the metabolic site of warfarin [Kaminsky and Zhang, 1997] is closest to the catalytic site and it was shown, e.g. on the cytochrome bc1 complex [Bartoschek et al., 2001], that electron transfers can take place between distances like 10 Å. The binding pocket is large enough to accommodate a second ligand, which could be an explanation for complex drug-drug interactions.

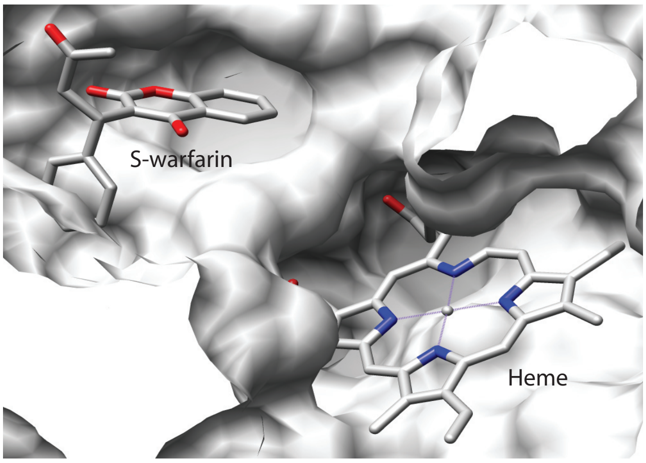


Figure 1.8: Crystal structure (PDB entry code: 1OG5) of CYP2C9 with bound ligand S-warfarin. Indicated is the heme group, which is the catalytic side.

1.2 Nuclear Magnetic Resonance

1.2.1 NMR applied to drug discovery

The whole drug discovery and development process is accompanied by NMR spectroscopy at certain stages. The most important NMR application is the analytical identification and structure elucidation of small molecules. These comprise newly discovered natural products, chemically synthesized compounds or extracted metabolites. In the first stage of the drug discovery process, which is basic research, NMR can be used to determine protein structures and investigate the dynamics and interactions with drug candidates and other molecules. While the protein structure determination of NMR is less and less done in pharmaceutical industries, the studies especially of weak interactions in the early phase of drug development is a growing field. In the final stage of drug development, the clinical trials, NMR (than termed MRI) can be used to monitor patients organs and even the distribution of labeled compounds. NMR gets a growing importance as a screening tool to investigate large compound libraries for their binding behavior to a target protein [Pellecchia et al., 2008], [Lepre et al., 2004] and [Meyer and Peters, 2003]. This is especially important in the screening for weakly binding fragments, where NMR can outperform screening assays that require a minimal affinity which may not be reached in the first screening. In [Dalvit, 2009] NMR methods in fragment screening were compared to the biophysical techniques surface plasmon resonance (SPR), fluorescence anisotropy (FA) and fluorescence lifetime (FL) and lead to the conclusion that NMR has the largest dynamic range. In Tab. 1.9, a comparison of the different NMR techniques SAR-by-NMR, spin labeling, STD, trNOE and INPHARMA is shown, which can be exploited to obtain informations upon ligand binding to a target protein.

All of these are frequently used in pharmaceutical research. Discussed in the following are the interesting methods i) SAR-by-NMR, which is based on chemical shift perturbations and ii) SLAPSTIC, that uses a spin labeled protein or ligand. These methods are especially interesting, because they are applied for second site ligand screening, a technique highly desired for the upcoming and often successful approach of fragment based drug discovery (FBDD) [Rees et al., 2004]. Hereby medium binding ligands are linked to create higly potent drugs. The NOE based methods STD, trNOE and INPHARMA, which are the most important technical basis for this thesis, will be discussed in detail after the introduction of the NOE concept.

Chemical shift perturbations and SAR-by-NMR

An important application of NMR is the mapping of ligand binding sites on the protein, including the search for a second-site binder by SAR-NMR [Shuker et al., 1996]. Therefore the protein target has to be labeled with a heteronuclear NMR enabling stable isotope like ^{15}N, or ^{13}C labeling of protein methyl groups. Then an HSCQ spectrum can be recorded to obtain a chemical shift map of the protein residues. Upon ligand binding, chemical shifts of residues in the close environment of the ligand will change. Given that an assignment of the protein was done, the binding site of the ligand can be determined by these shift perturbations. This methodology is also useful

	SAR by NMR	Spin labeling	STD	trNOE	INPHARMA
Large protein (> 30 kDa)	limited	yes	yes	yes	yes
small protein (< 10 kDa)	yes	yes	no	no	no
Isotope-labeled protein required	yes	yes	no	no	no
Binding epitope on protein	yes	no	no	no	no
Binding epitope on ligand	no	no	yes	yes	yes
Amount of protein [µM] at 500 MHz	~25	~1	~ 0.1	~5	~25
K_D tight binding	no limit	100 pM	100 pM	~10 µM	~ 10 µM
K_D weak binding	~1 mM	~10 mM	~10mM	~ 10 mM	~10 mM
Identification of ligand	no	yes	yes	yes	yes
Comments	robust method	sensitive method, but results ambiguous if lysine positions unknown	robust method	robust method to determine the bound ligand conformation	yields the relative binding mode of two ligands that target the same binding site

Figure 1.9: Comparison of NMR spectroscopy techniques for the identification and characterization of binding of ligands to proteins. The table is a modification of [Meyer and Peters, 2003].

as a screening tool for a wide range of binding affinities from weak mM to strong nM ligand binders. Especially interesting in this sense is the SAR-by-NMR technique, where ligands are screened for binding in the presence of a known binder. If additional shift perturbations occur for residues close to the binding site of the first ligand, it can be concluded that both ligands bind to a similar binding site. With the information of a second-site ligand both molecules can be linked as is often done in FBDD approaches, which can result in a binder that is orders of magnitudes stronger than the initial binders. The methodology can be even improved together with the NMR methodology ILOE [Becattini et al., 2006].

Spin-labeled protein/ligand and SLAPSTIC

The approach of spin labels attached to protein side chains as a tool to identify interacting compounds (SLAPSTIC) was in the first place used as a screening tool of ligands against a target protein [Jahnke, 2002]. Therefore a side chain of the protein is spin labeled with a paramagnetic organic compounds such as TEMPO. The paramagnetic label leads to drastically increased relaxation rates of neighboring protons, therefore a strong decrease or even quenching of a ligand proton signal indicates a binding event of the ligand to the labeled protein. This effect can be used as a screening tool for binding ligands. The methodology can also be applied vice versa with a spin labeled ligand, then used for second-site screening. When a second ligand binds to a binding site close to the first ligand, the proton signals of this ligand will relax very fast due to the proximity of the paramagnetic spin label of the first ligand. This effect can be measured in a 1D experiment or a $T1\rho$ relaxation experiment. When a second site ligand is found, it can be linked via chemical synthesis to the first ligand and the best linked ligand chosen for further development. There are several successful examples of this approach for FBDD [Jahnke et al., 2003,Poppe et al., 2007,Leone et al., 2010], which can also be combined with SAR-by-NMR or ILNOE. SLAPSTIC is very sensitive and robust against false positive detection. Yet, the limitation is the need of a spin-labeled first ligand, which is often limited by chemical resources.

1.2.2 The Nuclear Overhauser Effect (NOE)

Longitudinal and transverse relaxation

The term relaxation refers to the return of the spins to their equilibrium states. This comprises i) longitudinal relaxation, which is the spin-lattice relaxation along the z-axis and ii) transverse relaxation, which is the spin-spin relaxation within the xy-plane. Furthermore the cross-relaxation between two dipoles leads to a unique and very useful NMR phenomenon: the Nuclear Overhauser Effect (NOE). The NOE crossrelaxation rate contains a distance information that displays one of the most important properties of NMR-based structure calculations. Moreover, the NOE can be used to study dynamics as it also contains information on relaxation mechanisms. In the following section a detailed description of the NOE is given and how it can be back-calculated. This will be the theoretical basis for the largest part of this thesis.

As stated above, relaxation contains longitudinal and transverse relaxation. The first is the return of the spins to the equilibrium state along the z-axis and is also termed spin-lattice relaxation. The longitudinal relaxation time T_1 can be easily measured with an inversion recovery experiment. A 90° pulse places the magnetization in the xy-plane and a 180° pulse inverts it.

By variation of the time τ between the two pulses, a straight line of the relaxation time constant can be obtained. A quick estimate can be performed by measuring the time where τ equals zero (τ_0). Then T_1 can be derived:

$$T_1 = \frac{\tau_0}{ln2} \tag{1.3}$$

The transverse relaxation is the return of the spins to the equilibrium state in the xy-plane and is also termed spin-spin relaxation. The transverse relaxation T_2 can be easily estimated, as it is proportional to the linewidth ν:

$$\Delta\nu = \frac{\Delta\omega}{2\pi} = \frac{1}{(\pi T_2)} \tag{1.4}$$

This is just a rough estimate and the linewidth contains also other contributions such as chemical exchange or unresolved couplings. T_2 can be measured accurately with the CPMG (Carr-Purcell-Meiboom-Gill) experiment, also called spin echo. Therefore a 90° pulse is followed by several 180° pulses. The magnetization in the xy-plane is constantly reduced and the amplitude of the decay gives T_2.

T_2 is further dependent on the size of the molecule: during the brownian rotation diffusion in solution, the correlation time τ_c is the time it takes for the molecule to rotate by one radian and depends on the molecular size. τ_c can be estimated by applying the Stokes-Einstein relation [Görler and Kalbitzer, 1997]:

$$\tau_c = \frac{4\pi\eta d^3}{3kT} \tag{1.5}$$

Hereby η is the viscosity, d the radius of the molecule (it is assumed that the molecule is spherical), k the Boltzmann constant and T the temperature. As can be seen, the larger the molecule, the higher the viscosity and the lower the temperature, the larger is τ_c.

So far a description of the processes and motions occurring during dipole-dipole relaxation was obtained. The next step is the connection of these motions to the Larmor frequency. Therefore the amount of motion present at the observed Larmor frequency has to be found. This is described by a correlation function $G(t)$, which is frequently used in physics. For the present case it can be derived as follows [Keeler, 2010]:

$$G(t) = e^{\frac{-t}{\tau_c}} \tag{1.6}$$

Yet, the correlation function is a function of time. Fourier transformation of the correlation function results in the spectral density function $J(\omega)$:

$$J(\omega) = \frac{2}{5}(\frac{\tau_c}{1+\omega^2\tau_c^2}) \tag{1.7}$$

This equation assumes a protein system with rigid side chains. Dynamics of different amino acid types can be involved by using order parameters, as introduced by the model-free formalism of Lipari-Szabo [Lipari and Szabo, 1982]:

$$J(\omega) = \frac{2}{5}(\frac{S^2\tau_c}{1+\omega^2\tau_c^2}\frac{(1-S^2)\tau}{1+\omega^2\tau^2}) \tag{1.8}$$

Hereby the order parameter S^2 is an estimated or experimentally determined descriptor of the internal motion. The values range from 0 to 1, whereby 1 would represent a rigid system without internal dynamics.

The Solomon equations

The return of a two spin system to the equilibrium state can be described with the Solomon equations [Solomon, 1955]. The origin of the NMR signal are transitions of spins between the different energy levels of the α and β state. In the case of dipole-dipole relaxation the interactions of the spins can be understood, when taking a close look on the transitions. On the energy level diagram in Fig. 1.10 the transitions of the spin states are shown. Hereby the rate constants W of the transitions are given: i) $W_{1(\alpha\ or\ \beta)}$ indicates single-quantum transitions, ii) W_2 indicates double-quantum transitions and iii) W_0 zero-quantum transitions. There are eight single-quantum transitions, where only one spin changes the state and these lead to autorelaxation. Additionally there are two double- and two zero-quantum transitions, where both spins change the state and these lead to auto- and crossrelaxation. The energy levels of the $\alpha\beta$ and $\beta\alpha$ states have the same energy level in the diagram, because a homonuclear system is shown. In a heteronuclear system of e.g. ^{1}H and ^{13}C, these levels would have different heights. The diagram in Fig. 1.10 illustrates the probabilities of the transitions to occur:

$$W_0 = b^2 \frac{1}{r^6} J(0) \qquad W_2 = b^2 \frac{1}{r^6} 6J(2\omega_0) \qquad W_1 = b^2 \frac{1}{r^6} 3J(\omega_0) \quad (1.9)$$

$$b = \frac{\mu_0}{8\pi} \hbar \gamma_H^2 \quad (1.10)$$

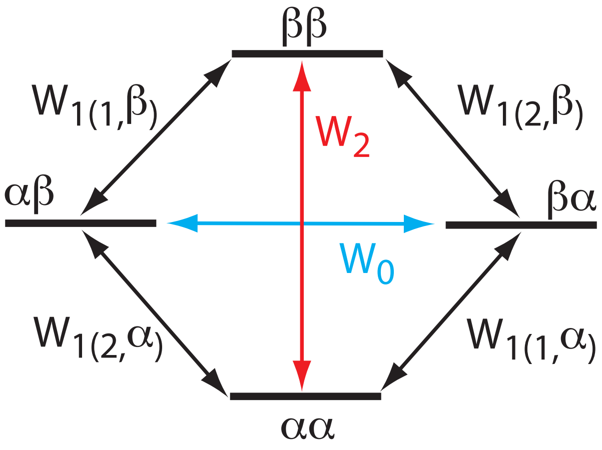


Figure 1.10: Energy level diagram of a homonuclear two spin system. Indicated are the 12 possible transitions between the spin states.

The rate constants for the transitions between the energy levels are termed W_0, W_1 and W_2. From these the autorelaxation ρ and crossrelaxation σ terms of the spins I and S can be derived:

$$\begin{aligned} \rho_I &= W_2 + 2W_1 + W_0 \\ \rho_S &= W_2 + 2W_1 + W_0 \\ \sigma_{IS} &= W_2 - W_0 \end{aligned} \quad (1.11)$$

These lead to the Solmon equations of a two spin system:

$$\frac{d\Delta I_z(t)}{dt} = -\rho_I \Delta I_z(t) - \sigma_{IS} \Delta S_z(t) \quad (1.12)$$

$$\frac{d\Delta S_z(t)}{dt} = -\rho_S \Delta S_z(t) - \sigma_{IS} \Delta I_z(t) \quad (1.13)$$

The motional regimes

The autorelaxation rate ρ of the dipole-dipole interaction is always a negative term as can be seen in equations shown above. In comparison to this, the crossrelaxation rate can be positive or negative, depending on the value of the correlation time τ_c. Hence, if the molecule is small and τ_c is short, the dominant transition will be double quantum and the rate constant is positive. While, if the molecule is large and τ_c is long, the dominant transition is zero-quantum and the rate constant is negative.

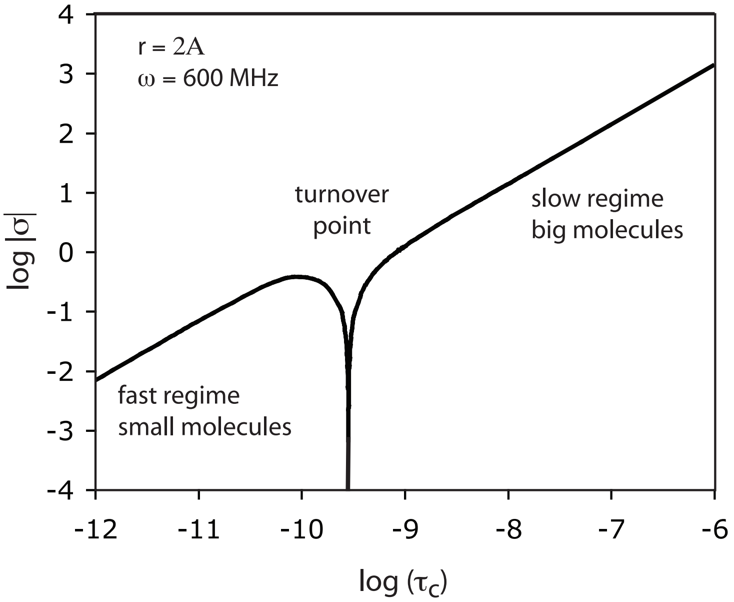


Figure 1.11: Simulation of the crossrelaxation rate σ between two protons (distance 2 Å) at a Larmor frequency of 600 MHz. While the correlation time τ_c of the molecule is increasing, the sign of the signal changes from positive to negative. The turnover point is at $\tau_c = \frac{\sqrt{5}}{2\omega_0}$, extending out of the graph.

The turnover point between these two motional regimes is at a correlation time of $\tau_c = \frac{\sqrt{5}}{2\omega_0}$. At this point no NOE signal is observed. The different motional regimes are a very useful phenomenon for drug discovery, as a small molecule that binds to a big molecule will adopt the correlation time of the latter.

The NOESY experiment

The phenomenon of cross-relaxation gives rise to a signal between spins close in space, if the distance is less than approximately 5 Å [Cavanagh et al., 2007].

The NOE effect can be measured in a 1D spectrum as a transient or steady-state NOE experiment (described below). In this case it is designed as a difference experiment, where two spectra are recorded (on- and off-resonance irradiation) and subtracted. The extension to a 2-dimensional NOESY experiment with diagonal peaks and cross-peaks can be easily achieved. Therefore an initial 90° pulse is followed by the evolution time t_1 and a second 90° pulse brings the magnetization to the z-axis. During the mixing time, the magnetization is kept longitudinal and all magnetization except I_z is suppressed by phase cycling or pulse field gradients. This is followed by a third 90° read-out pulse. The pulse sequence of the described experiment is shown in Fig. 1.12.

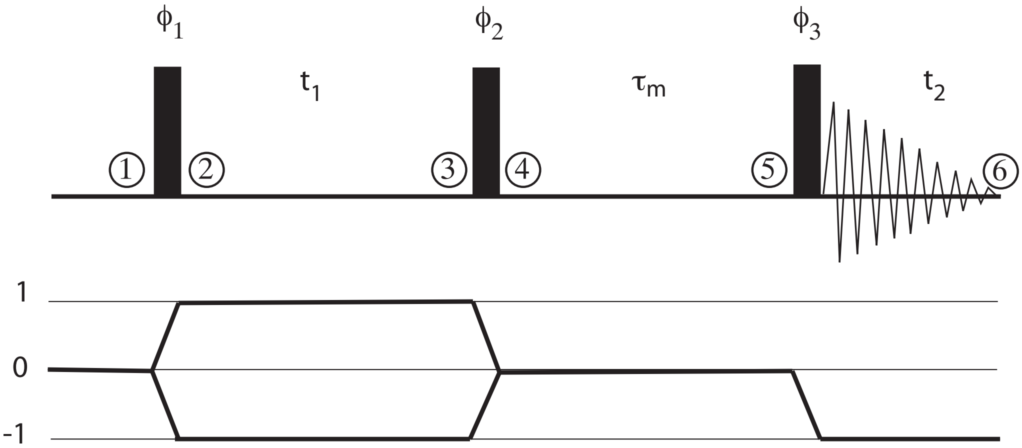


Figure 1.12: Pulse sequence of a standard NOESY experiment. Indicated are the pulses and the coherence transfer pathway.

In the following the magnetization at the indicated numbers of Fig. 1.12 during the pulse sequence are described by the product operator formalism. Product operators are a practical representation of the density matrix description of spin dynamics. The first two pulses are identical to the well known COSY sequence:

$$1) + I_z + S_z \tag{1.14}$$

$$2) - I_y - S_y \tag{1.15}$$

$$\begin{aligned} 3) &- I_y cos\Omega_I t_1 cos\pi I_{IS} t_1 + 2I_x S_z cos\Omega_I t_1 sin\pi J_{IS} t_1 \\ &+ I_x sin\Omega_I t_1 cos\pi J_{IS} t_1 + 2I_y S_z sin\Omega_I t_1 sin\pi J_{IS} t_1 \end{aligned} \tag{1.16}$$

$$\begin{aligned} 4) &- I_z cos\Omega_I t_1 cos\pi I_{IS} t_1 - 2I_x S_y cos\Omega_I t_1 sin\pi J_{IS} t_1 \\ &+ I_x sin\Omega_I t_1 cos\pi J_{IS} t_1 - 2I_z S_y sin\Omega_I t_1 sin\pi J_{IS} t_1 \end{aligned} \tag{1.17}$$

The interesting part starts here after the second 90° pulse. The NOESY has the same pulses as e.g. the DFQ-COSY, yet the difference is the coherence selection. During the mixing time τ_m only longitudinal magnetization and zero-quantum (ZQ) coherence is accepted and allowed to evolve based on the phase cycle:

$$\begin{aligned} 5) &- \mathbf{a} I_z cos\Omega_I t_1 cos\pi J_{IS} t_1 + ZQ_y cos\Omega_I t_1 sin\pi J_{IS} t_1 cos(\Omega_I - \Omega_S)\tau_m \\ &- \mathbf{b} S_z cos\Omega_I t_1 cos\pi J_{IS} t_1 - ZQ_x cos\Omega_I t_1 sin\pi J_{IS} t_1 sin(\Omega_I - \Omega_S)\tau_m \end{aligned} \tag{1.18}$$

The amplitudes **a** and **b** are matrices that describe the efficiency of the magnetization transfer during the mixing time and can be derived from the Solomon equations:

$$\mathbf{a}_{(\tau_m)} = e^{-\mathbf{R}\tau_m} \qquad \mathbf{R} = \left\{ \begin{array}{cc} \rho_I & \sigma_{IS} \\ \sigma_{SI} & \rho_S \end{array} \right\} \tag{1.19}$$

R is in place the relaxation matrix for a two spin system. In the case of a large molecule, **R** will increase tremendously and the multistep transfers will lead to spin diffusion. The chemical exchange also contributes to **a** and can be described with a kinetic matrix, as will be described later in the text. The relaxation matrix consists of the elements of the autorelaxation rate ρ and crossrelaxation rate σ, which are defined as:

$$Autorelaxation : \rho_I = b^2 \frac{1}{r_{IS}^6}[6J(2\omega) + 3J(\omega) + J(0)] \tag{1.20}$$

$$Crossrelaxation : \sigma_{IS} = b^2 \frac{1}{r_{IS}^6}[6J(2\omega) - J(0)] \tag{1.21}$$

Back to the NOESY experiment, the final 90° pulse leads to the signal and gives:

$$\begin{aligned}6) &+ \mathbf{a}I_y cos\Omega_I t_1 cos\pi J_{IS} t_1 \\ &+ \frac{1}{2}(2I_zS_x - 2I_xS_z) cos\Omega_I t_1 sin\pi J_{IS} t_1 cos(\Omega_I - \Omega_S)\tau_m \\ &+ \mathbf{b}S_y cos\Omega_I t_1 cos\pi J_{IS} t_1 \\ &- \frac{1}{2}(2I_xS_x + 2I_zS_z) cos\Omega_I t_1 sin\pi J_{IS} t_1 sin(\Omega_I - \Omega_S)\tau_m\end{aligned} \tag{1.22}$$

The terms $\mathbf{a}I_y$ and $\mathbf{b}S_y$ lead to the desired diagonal peaks and cross-peaks. They are in-phase and absorptive in both dimensions. The term $\frac{1}{2}(2I_zS_x - 2I_xS_z)$ is anti-phase and dispersive in both dimensions. It represents the unwanted zero-quantum contribution to diagonal peaks and cross-peaks. For large molecules it can be neglected, as the ZQ relaxation is fast and the anti-phase contribution small. In the case of small molecules it can be avoided, for example by using a ZQ-filter [Thrippleton and Keeler, 2003] or reduced with a 180° refocusing pulse during the mixing time. The last term $-\frac{1}{2}(2I_xS_x + 2I_zS_z)$ is double quantum coherence and not observable. The presented description of the NOESY experiment is based on [Hore et al., 2000, Cavanagh et al., 2007, Keeler, 2010].

The magnetization transfer can not only take place in the longitudinal, but also in the transverse plane. This is done in a ROESY (rotating-frame nuclear Overhauser spectroscopy) experiment, where the mixing time is substituted by a spin-lock field to keep the magnetization in the transverse plane. As seen in equation 1.24 the sign of the ROESY cross relaxation rate is always positive (in comparison to the NOE cross relaxation rate as shown above), as it is an addition and not a difference of two positive terms. Yet, ROE has the disadvantages of a less favorable spin diffusion than NOE and artifacts from J-coupling. Still, the ROE experiment is useful for molecules that display a NOE cross relaxation rate close to zero between the fast and slow motion regime.

$$\sigma_{NOE_{ij}} = b^2 \frac{1}{r_{ij}^6}[6J(2\omega) - J(0)] \tag{1.23}$$

$$\sigma_{ROE_{ij}} = b^2 \frac{1}{r_{ij}^6}[3J(2\omega) + 2J(0)] \tag{1.24}$$

Back-calculation of the NOE

The NOE peak volumes **V** that evolve during the mixing time of a NOESY experiment are described with a differential equation, derived from the Solomon equations:

$$\frac{d\mathbf{V}(\mathbf{t})}{dt} = -(\mathbf{R} + \mathbf{K}) \cdot (\mathbf{V}(t) - \mathbf{V}_{eq}) \tag{1.25}$$

V represents the volume matrix of all considered spins that evolve during time t. The Boltzmann equilibrium magnetization is represented by $\mathbf{V}_{eq}$. **R** is the relaxation matrix and **K** is the kinetic matrix of the chemical exchange. **K** is derived from the chemical equilibrium rate constants for example from a ligand to protein binding event. **R** and **K** can be united to the matrix **D**. If the mixing time τ_m is also considered, the differential equation 1.25 can be solved as follows:

$$\mathbf{V}(\tau_m) = e^{-\mathbf{D}\cdot\tau_m} \cdot \mathbf{V}(0) \tag{1.26}$$

The dynamic matrix $\mathbf{D} = \mathbf{R} + \mathbf{K}$ is asymmetric and therefore not practicable for computation. **D** has to be turned into a symmetric and diagonalized matrix, which can be efficiently computed. For this operation **D** has to be multiplied with a symmetrization matrix **S** and its inverse matrix $\mathbf{S}^{-1}$ to give the symmetrical $\mathbf{D_s}$ matrix. The result has then to be multiplied with a diagonalization matrix **X** and its transpose $\mathbf{X}^T$ to yield the diagonal, symmetrical matrix λ:

$$\begin{aligned}
\mathbf{V}(\tau_m) &= e^{-\mathbf{D}\cdot\tau_m} \cdot \mathbf{V}(0) && (1.27)\\
&= \mathbf{S} \cdot e^{-\mathbf{D_s}\cdot\tau_m} \cdot \mathbf{S}^{-1} \cdot \mathbf{V}(0) && (1.28)\\
&= \mathbf{S} \cdot \mathbf{X} \cdot e^{-\lambda\cdot\tau_m} \cdot \mathbf{X}^T \cdot \mathbf{S}^{-1} \cdot \mathbf{V}(0) && (1.29)
\end{aligned}$$

Using the final equation 1.29, the NOE peak volumes can be efficiently back-calculated for all relevant protein and ligand systems.

1.2.3 Transferred NOE (trNOE)

As was discussed in the section above, there are two motional regimes for the NOE that depend on the size of the molecule. The consequence is, that large molecules display much larger signals in the spectrum than small molecules due to their larger correlation time. If a small molecules binds to the large molecule, there will be a magnetization transfer and the signals of the small molecule will adopt the same strong, negative intensities as with the large correlation time. This happens because these large intensities of the bound form just overlay the small, positive signals of the free form. As a result one sees the bound conformation of the small molecule in the NOESY spectrum and not the free form. Especially if the molecule has many torsion angles, this is a very valuable information.

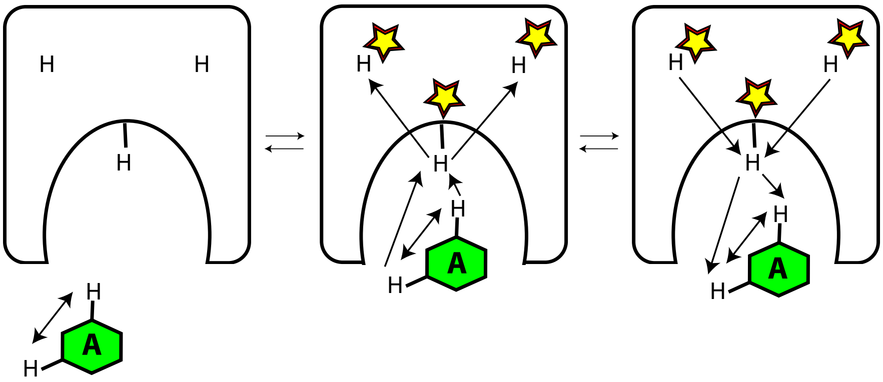


Figure 1.13: Principle of transferred NOE magnetization, observable in an NMR NOESY spectrum. Magnetization of the ligand protons spreads to the protein protons and back.

The magnetization transfer of trNOE is shown in Fig. 1.13. The case for long mixing times is shown, where the magnetization spreads by spin diffusion all over the protein. Therefore the NOE peaks also contain information of the protein surroundings. If one is just interested in the bound conformation, it is feasible to record at very short mixing times to avoid spin diffusion or even measure with a deuterated protein in order to see only the transfer between ligand protons. First applications of the trNOE were done in the 1970s [Balaram et al., 1972, Balaram et al., 1973].

The theoretical description and back-calculation is derived from the approach by [Clore and Gronenborn, 1982] and [Ni and Zhu, 1994]. In the kinetic model it is assumed that ligand A binds to a target protein T to form a complex TA. This reaction is reversible and described by the rate constants k_{12} and k_{21}. Due to the large excess of the ligand, it is assumed that the target is never in the unbound state:

$$A + T \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} TA \tag{1.30}$$

The relaxation matrix **R** can be set up as a block matrix. Diagonal elements represent the auto-relaxation terms and the others the cross-relaxation terms, whereby cross-relaxation takes only place in the bound states:

$$\mathbf{R} = \left\{ \begin{array}{ccc} \mathbf{R}_A^A & 0 & 0 \\ 0 & \mathbf{R}_A^{TA} & \mathbf{R}_{A,T}^{TA} \\ 0 & \mathbf{R}_{T,A}^{TA} & \mathbf{R}_T^{TA} \end{array} \right\} \tag{1.31}$$

Hereby the superindex of each **R** indicates the species and the subindex indicates the protons of the species for which the relaxation is described. So $\mathbf{R}_A^A$ represents the free ligand, $\mathbf{R}_A^{TA}$ is the relaxation of ligand protons in the bound form, $\mathbf{R}_T^{TA}$ is the relaxation of protein protons in the ligand bound protein form. Finally $\mathbf{R}_{T,A}^{TA}$ and $\mathbf{R}_{A,T}^{TA}$ represent the cross-relaxation between the protons of the bound ligand and the target protons, respectively. The kinetic matrix **K** represents the processes of the system at equilibrium. It includes the rate constants k_{12} and k_{21} and the concentrations of [TA] and [A] . **I** represents the identity matrix:

$$\mathbf{K} = \left\{ \begin{array}{ccc} k_{12}[TA]\mathbf{I} & -k_{21}[A]\mathbf{I} & 0 \\ -k_{12}[TA]\mathbf{I} & k_{21}[A]\mathbf{I} & 0 \\ 0 & 0 & k_{21}[A]\mathbf{I} \end{array} \right\} \tag{1.32}$$

The trNOE is a very useful technique to derive the bound conformation of a ligand to a protein, given that several off-events take place during the mixing time, i.e. the binding is not too strong. It can also be used to check if a ligand binds to a protein, but this application is now mostly done with the much faster STD method (see below).

Interligand NOE

In the case that the binding pocket of a protein is large enough that it can accommodate two ligands at the same time, an interligand NOE (ILOE) signal between these can be observed [Li et al., 1999]. Also it is possible that a ligand binds outside of the binding pocket and an NOE is observed to the ligand inside the binding pocket. As for the trNOE there is spin diffusion enhancing the signal, but the original magnetization transfer takes place between the two ligands. If the protein would be deuterated or the spectrum would be recorded at very short mixing times, minimizing extremely the effect of spin diffusion, there would still be the interligand NOE visible. This has been theoretically described in detail [London, 1999]. The technique is very useful for linking small fragments to a potent inhibitor and has already been applied successfully in drug discovery [Becattini et al., 2006,Sledz et al., 2010].

1.2.4 INPHARMA

The INPHARMA (Interligand NOE for PHARmacophore MApping) effect was first observed when epothilone A and baccatin III were measured together with tubulin in one NMR sample [Sánchez-Pedregal et al., 2005]. It was possible to observe a magnetization flow between the two ligands, which was transferred via the protein by spin diffusion. The result are NOE signals between the binding molecules, yielding information about their relative binding orientation. As the magnetization flow goes via the protein, the information of protein density around the ligands is also given. If the protein would be deuterated, or very short mixing times would be chosen, no spin diffusion would take place and the INPHARMA peaks would not be observable. This has been shown with deuterated PKA samples [Orts et al., 2008a]. Because the information of the protein environment is included in the INPHARMA peaks, even the absolute binding mode of the ligands in the binding pocket can potentially be derived. The magnetization transfer of INPHARMA during a NOESY experiment is shown in Fig. 1.14. One ligands binds to the protein and the magnetization spreads by spin diffusion to the protein protons. When the second ligand binds, the magnetization is transferred onto this ligand.

The magnetization transfer of INPHARMA can be described similar to the trNOE, but the relaxation and kinetic matrix have to be adjusted to the situation of two ligands. Such a description was done for the ILOE effect by [London, 1999] and proposes five different states: Ligands A and B in the free form, ligands A and B bound and both ligands bound. Transferring this to the INPHARMA situation we end up with four different chemical species: Ligands A and B in the free form and ligand A and B bound to the target (in the on- and in the off-state). Setting up these equations was described in detail [Reese et al., 2007,Orts et al., 2008b,Orts et al., 2009]. In this kinetic model it is assumed that the target is never in the unbound state due to the large excess of the two ligands and therefore the unbound target is negligible. The complexes then form as follows:

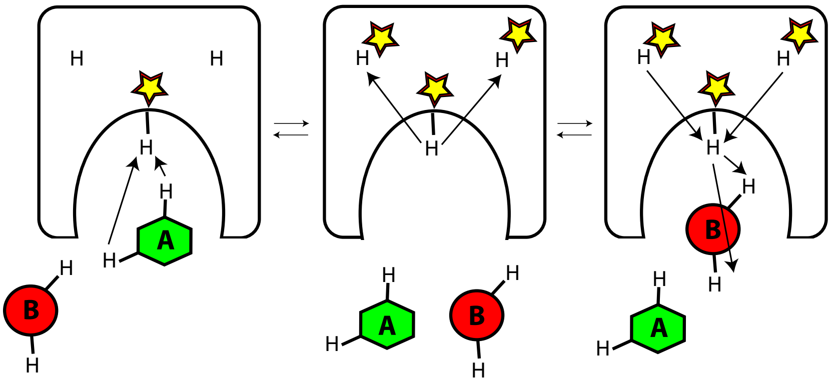


Figure 1.14: Principle of INPHARMA magnetization transfer, observable in an NMR NOESY spectrum. Magnetization of the first ligand spreads to the protein protons and is transferred to the second ligand upon binding. NOESY peaks between the two ligands will appear in the spectrum.

$$TA + B \quad \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \quad TB + A \tag{1.33}$$

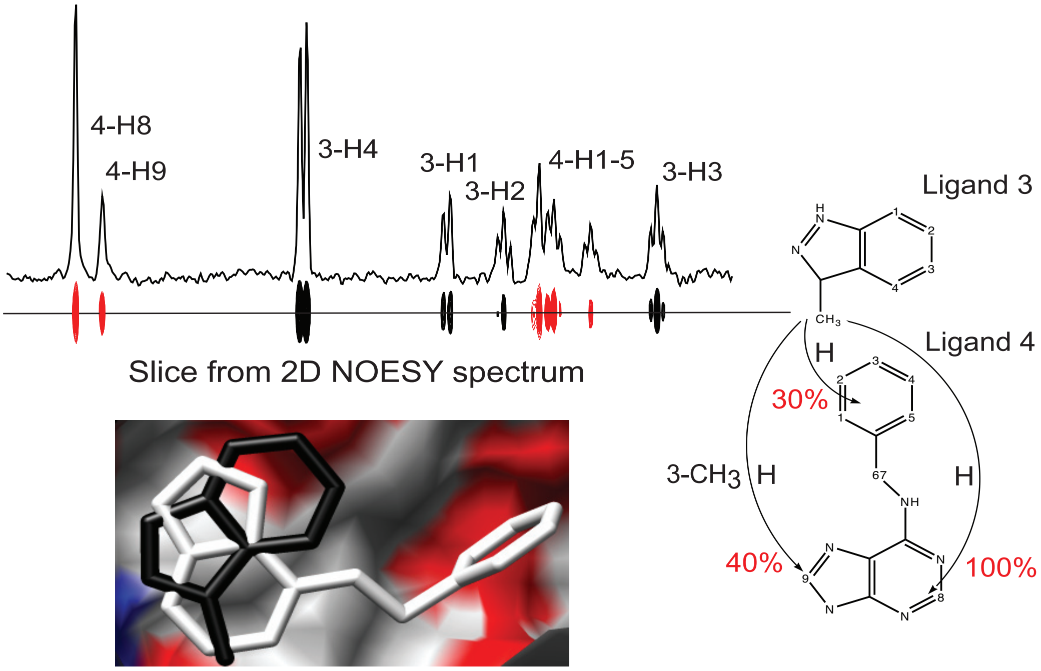
The relaxation matrix can be set up in the same way as for the trNOE, just the ligand B has to be added:

$$\mathbf{R} = \left\{ \begin{array}{cccccc} \mathbf{R}_A^A & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{R}_B^B & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{R}_A^{TA} & \mathbf{R}_{A,T}^{TA} & 0 & 0 \\ 0 & 0 & \mathbf{R}_{T,A}^{TA} & \mathbf{R}_T^{TA} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{R}_B^{TB} & \mathbf{R}_{B,T}^{TB} \\ 0 & 0 & 0 & 0 & \mathbf{R}_{T,B}^{TB} & \mathbf{R}_T^{TB} \end{array} \right\} \tag{1.34}$$

Hereby $\mathbf{R}_T^{TX}$ (X = ligand A or B) represents relaxation between protons of the target, while the ligand is in the bound state. $\mathbf{R}_X^{TX}$ represents the relaxation of ligand protons in the bound form. $\mathbf{R}_{T,X}^{TX}$ and $\mathbf{R}_{X,T}^{TX}$ represent the cross-relaxation between the protons of the bound ligand and the target protons. The kinetic matrix can be derived from the trNOE in the same way:

$$\mathbf{K} = \left\{ \begin{array}{cccccc} k_{21}[TB]\mathbf{I} & 0 & -k_{12}[B]\mathbf{I} & 0 & 0 & 0 \\ 0 & k_{12}[TA]\mathbf{I} & 0 & 0 & -k_{21}[A]\mathbf{I} & 0 \\ -k_{21}[TB]\mathbf{I} & 0 & k_{12}[B]\mathbf{I} & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{12}[B]\mathbf{I} & 0 & -k_{21}[A]\mathbf{I} \\ 0 & -k_{12}[TA]\mathbf{I} & 0 & 0 & k_{21}[A]\mathbf{I} & 0 \\ 0 & 0 & 0 & -k_{12}[B]\mathbf{I} & 0 & k_{21}[A]\mathbf{I} \end{array} \right\} \tag{1.35}$$

Summarizing the relaxation and the kinetic matrix according to equation 1.26 results in an accurate description of the multi-step transfer which ultimately results in INPHARMA NOESY spectra.



Slice from 2D NOESY spectrum

Superimposed crystal structures
of ligand 3 and ligand 4.

Figure 1.15: Example of the INPHARMA method, demonstrating the ability to give distance related superimpositions of bound ligands. These observations are proven by the superimposed crystal structures, shown below.

An example of the INPHARMA method is shown in Fig. 1.15 for the PKA binding ligands 3 and 4. The slice through the NOESY spectrum at the resonance of the methyl group of ligand 3 in ω_1 demonstrates that the magnetization is transferred via protein protons (H) to ligand 4 and suggests e.g. a larger spatial proximity of 3-CH_3 to 4-H8 than 4-H1-5. Below are the crystal structures of ligand 3 (black) and ligand 4 (white) superimposed to demonstrate the correct prediction of INPHARMA. The sample contained 0.045 mM PKA, 1.2 mM of ligand 3 and 0.9 mM of ligand 4 in PBS D_2O buffer.

1.2.5 Saturation transfer difference (STD)

The two-dimensional NOESY is very useful, but also requires long NMR measurement time. The NOE can be measured also in a one-dimensional fashion. For this purpose the signal of one specific protein proton frequency will be saturated that does not overlap with proton frequencies of the ligand. This is done by RF irradiation onto this frequency. The single-quantum transitions are significantly reduced and the same holds true for the signal.
In practice two different spectra are recorded. One spectrum with the RF irradiation far away from any molecule signal (off-resonance) and one spectrum with the RF irradiation on one specific molecule frequency (on-resonance). The difference spectrum shows a strongly reduced signal at the saturated frequency and also reduced signals of protons close in space to the saturated proton. This one-dimensional NOE can be used in a fast and easy manner to determine the constitution of small molecules.
The method also works in the case of chemical exchange, because the saturation can be transferred by spin diffusion. The consequence is that the saturation of one frequency of a big molecule will quickly spread over the whole protons. If a small molecule binds weakly, it will also get saturated. Given that there are several off-events during the saturation time, the binding event is visible by a change in the difference spectrum. The saturation is normally set to a frequency where no ligand signals are to be found: around 0.5 ppm in the methyl group region of the protein, or around 11 ppm beyond the aromatic region of the protein. The magnetization transfer is shown for irradiation on methyl groups in Fig. 1.16. This technique is termed saturation transfer difference (STD) and was first introduced by [Mayer and Meyer, 1999].

The back-calculation of STD peak volumes is done with the complete relaxation matrix approach similar to the NOESY spectra. Yet, the saturation by RF irradiation has to be taken into account. The following description of the back-calculation is based on the complete relaxation and conformational exchange matrix (CORCEMA) theory developed by [Jayalakshmi and Krishna, 2002] and [Jayalakshmi et al., 2004]. Here an ideal case is assumed with an infinite recovery delay and an instantaneous saturation by RF irradiation of the saturated spins, e.g. methyl groups. The equation for the

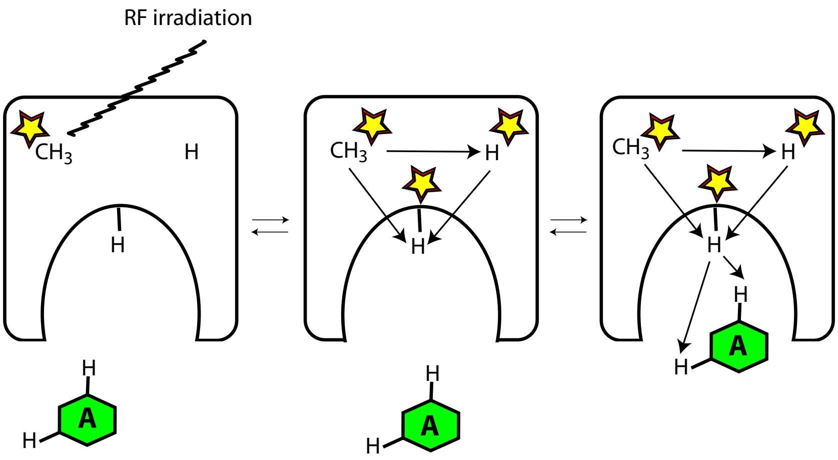


Figure 1.16: STD magnetization transfer. Protein protons of a specific frequency (here methyl groups) are saturated by RF irradiation. Magnetization spreads to all protein protons and the bound ligand.

back-calculation of peak volumes is set-up similar to equation 1.26, with an additional matrix $\mathbf{Q}$ to describe the saturated protons:

$$\mathbf{V}(t) = \mathbf{V}(0) + (\mathbf{1} - e^{-\mathbf{D}\cdot t}) \cdot \mathbf{D}^{-1} \cdot \mathbf{Q} \tag{1.36}$$

$\mathbf{Q}$ is a matrix that describes the cross-relaxation terms between the saturated protein protons (e.g. from methyl groups) and the rest of protein protons, as well as the transfer to the bound ligand. $\mathbf{Q}$ is build up like $\mathbf{R}$ of the trNOE and STD shown above, yet it contains a new species: the protein protons T^{RF} that receive saturation by RF directly. As the free ligand does not receive magnetization, the saturation its equal to 0.

Besides the detection of ligand binding and determination of the rate constants of the same, STD can be used to obtain information on the ligand binding mode. This was first demonstrated on oligosaccharides binding a protein by group epitope mapping [Mayer and Meyer, 2001]. Oligosaccharides are a feasible example: they are very large ligands and the parts that are buried in the protein and exposed to the solvent are well separated in space. Buried protons get stronger saturation from the protein, therefore they can be distinguished easily from those exposed that receive less saturation. The ability of STD to derive the binding epitope of even a small ligand to a protein is demonstrated in Fig. 1.17: ligand protons in close contact to the protein such as 4-H8 and 4-H9 suffer from stronger saturation via protein protons than solvent exposed protons such as 4-H1-5.

The STD sample contained 0.045 mM PKA and 0.9 mM of ligand 4. On

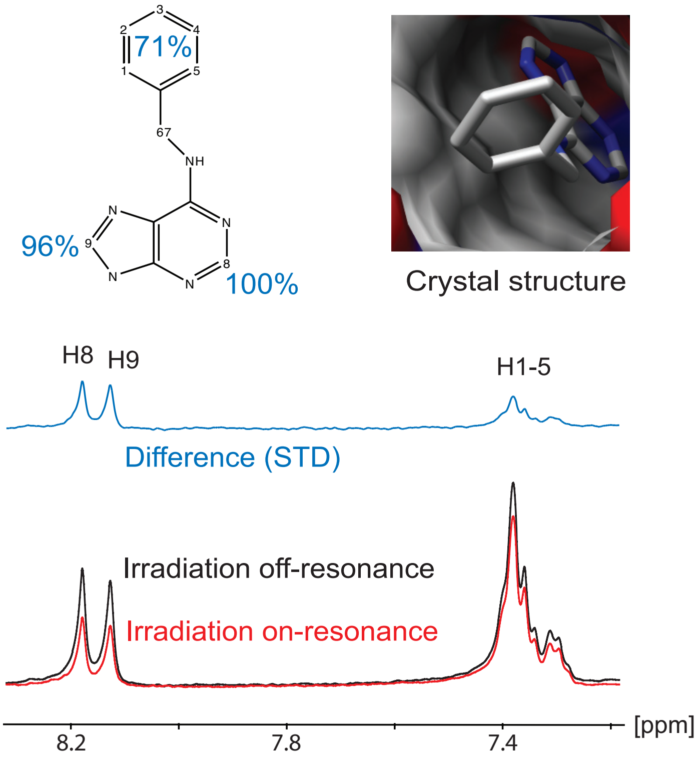


Figure 1.17: Example of the STD effect. The signal of the aromatic ring of the ligand is stronger than those of the single protons. After subtracting the 1D from an STD, it becomes clear that the single protons are deeper in the binding site and receive more magnetization from the protein. This observation is proven by the corresponding crystal structure.

the right side the crystal structure of ligand 4, demonstrating the exposition of the aromatic ring to the solvent and the close contact of protons 4-H8 and 4-H9 to the protein protons.

Taking together the widely used screening capabilities of STD with the speed of the experiment and the binding mode information obtainable, this technique is extremely useful for NMR and drug discovery. It is criticized that the method depends too much on T_1 and may give wrong results for some spins [Yan et al., 2003, Mizukoshi et al., 2012]. But this problem can be overcome with the accurate description delivered by the complete relaxation matrix approach.

WaterLOGSY

WaterLOGSY is an alternative to STD in NMR screening. The mechanism is opposite of STD: instead of saturating the protein, the bulk water is saturated. This is achieved by an 180° pulse on the water frequency, followed by a strong water suppression block [Dalvit et al., 2001]. While the STD represents the steady-state version of the one-dimensional NOE, the WaterLOGSY is recorded in a transient state manner. The magnetization will spread to protons exposed to the bulk water which affects the free and the bound form of the ligand. Yet, the magnetization that spreads over the whole protein and the bound ligand will change the sign of the NOE of the ligand protons, due to the long correlation time of the macromolecule. This will not happen to the free form of the ligand that will have a positive NOE. Therefore WaterLOGSY can be used to identify ligand binding. Ligand protons that are buried in a protein binding pocket do not get as strongly affected by the magnetization transfer as protons directly exposed to the bulk water. Therefore it is also possible to receive the binding epitope of a ligand. When information on the binding epitope is desired, the method can be used complementary to STD, as the difference spectrum should display the opposite information of the STD spectrum, i.e. the protons exposed to water like in the SALMON approach [Ludwig et al., 2008a,Ludwig et al., 2008b]. The WaterLOGSY experiment has the advantage that no protein proton frequency has to be saturated, which can be useful for ligands with many methyl groups that overlap with protein methyl groups in the spectrum. Still, the WaterLOGSY experiment is significantly longer than STD and is therefore not the first choice in screening.

Chapter 2

Protein kinase A

2.1 Introduction

2.1.1 Protein kinases as drug targets

Protein kinases are a family of enzymes that modify and activate other proteins by chemically adding a phosphate group. This phosphate group comes from an ATP molecule that binds to the very conserved binding pocket of the protein kinase. In the human genome 518 putative protein kinase genes were identified [Manning et al., 2002], which would account for 1.7 % of the human genes. Protein kinases can be divided in threonine/serine kinases (e.g. PKA, CDK) and tyrosine kinases (e.g. ABL), depending on which amino acid of the target protein they phosphorylate. Protein phosphorylation plays a key role in most aspects of the regulation in the cell, especially signal transduction. Abnormal phosphorylation, e.g. caused by mutations in the protein kinase can lead to unregulated cell growth, which is a crucial step in the development of cancer. Therefore protein kinases have become the second most important family of drug targets after G-protein coupled receptors, accounting for 20-30 % of the drug discovery programmes [Cohen, 2002]. The history of kinase drug design started in the early eighties, when it was found that tumour promoting phorbol esters are activators of protein kinase C [Castagna et al., 1982]. The first drugs developed for kinase inhibition were isoquinolinesulfonamides [Hidaka et al., 1984], which target the ATP binding site and showed good cell permeability. These substances led to the discovery of compounds like fasudil [Asano et al., 1998], which were approved in Japan for the market to treat cerebral vasospasm. Yet, these inhibitors showed a comparably low potency, which made it hard to compete with the ATP concentration inside the cell (2-10 mM). The situation changed when a natural, nanomolecular inhibitor was discovered [Tamaoki et al., 1986]: the antifungal agent staurosporine, produced by the bacterium *Streptomyces*, inhibits most kinases and was used a a lead structure by several companies (Fig. 2.2). But even with the discovery of a very strong inhibitor a main problem still remained: the low selectivity, meaning that many different kinases are targeted unspecifically. The first clinical success came from another natural product, rapamycin, that showed strong effects in immunosuppression by inhibiting the protein kinase mTOR specifically [Davies et al., 2000]. The greatest success of a specific protein kinase inhibitor and a landmark in structure-based drug design in general is imatinib. This compound was developed by Novartis under the treading name Gleevec and targets the Abelson tyrosine kinase (ABL). ABL fuses with the breakpoint cluster region (BCR) and forms an enhanced kinase that is observed in nearly all cases of chronic myelogenous leukemia (CML) [Druker et al., 1996]. Patients of CML can be cured successfully with imatinib and Gleevec was soon approved to the market and a huge financial success. Today an interesting drug target is epidermal growth factor receptor (EGFR) tyrosine kinase, which is overexpressed in lung and breast cancer. The Astra-Zeneca compound Iressa [Morin, 2000] was approved for the market. Another important kinase family are the cyclin-dependent kinases (CDKs) which play essential roles in cell proliferation. Hereby flavopiridol [Senderowicz, 2000] from Sanofi is an interesting candidate. At the time of this thesis only 9 small molecule drugs that target kinases are approved to the market, while nearly 30% of all industrial drug discovery programs focus on protein kinases [Backes et al., 2008].

2.1.2 ATP binding site

The first 3-dimensional structure of a protein kinase was obtained in 1991, when the crystal structure of the catalytic subunit of the cAMP-dependent protein kinase A (PKA) was solved [Knighton et al., 1991]. The mechanism of action of the natural binder ATP is well understood [Bossemeyer et al., 1993], as well the binding modes of many diverse inhibitors [Engh et al., 1996]. In the physiological environment PKA is a tetramer that consists of two catalytic and two regulatory units. Activation by four second messenger cAMP molecules releases the two catalytic subunits. Generally the catalytic subunit of a protein kinase is amenable for crystallization. For most important inhibitors and many fragment drugs crystal structures exist in the public PDB as well as in the proprietary databases of pharmaceutical companies. Protein kinase structures are not amenable to be solved by NMR, as the catalytic subunit is already too large (42 kDa in the case of PKA). Nevertheless a 70 % assignment of PKA was presented by the combination of

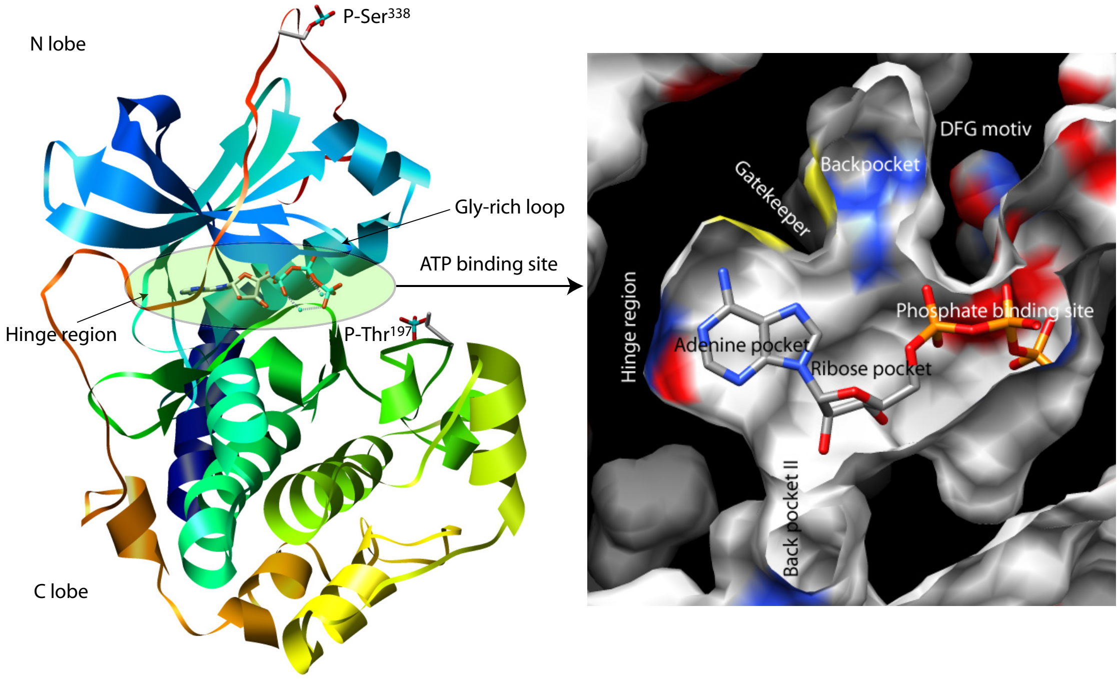

N lobe
P-Ser^{338}
Gly-rich loop
ATP binding site
Hinge region
P-Thr^{197}
C lobe
DFG motiv
Backpocket
Gatekeeper
Hinge region
Phosphate binding site
Adenine pocket
Ribose pocket
Back pocket II

Figure 2.1: Crystal structure of ATP bound to PKA (PDB entry code: 1ATP). Indicated are the N-lobe and C-lobe, the binding site at the hinge region, and the phosphorylated residues Thr^{197} and Ser^{338}. On the right side there are the different binding pockets indicated: the common binding motif at the hinge region and the DFG motif of the backpocket, which is guarded by the gatekeeper residue.

different methods, including chemical shift calculation, paramagnetic tagging and selective labeling. Investigations with TROSY revealed then conformational changes upon phosphorylation [Langer et al., 2004]. Phosphorylation activates the kinase and induces a conserved, major conformational change. In the case of protein kinase A, phosphorylation takes place on Thr^{197} and Ser^{338}. A new hydrogen bond between P-Thr^{197} and His^{87} is the key element of activation and induces a cascade of new hydrogen bonds, also stabilizing the conserved DFG motiv [Kornev et al., 2006]. This region of Asp-Phe-Gly (DFG) forms the backpocket next to the ATP-binding site. The entrance to this pocket is guarded by the gatekeeper, normally a bulky amino acid like phenylalanine. In the case of PKA it is Met^{120}. Some ligands have the ability to re-orient the gatekeeper and penetrate into the backpocket. This behavior is highly desirable in kinase drug design, as it opens the way for higher selectivity. ATP is attached with a hydrogen bond to the Val^{123}-NH and the Glu^{121}-CO at the hinge region and covers the adenine pocket, the ribose pocket and the phosphate binding region, but not the backpockets (Fig. 2.1).

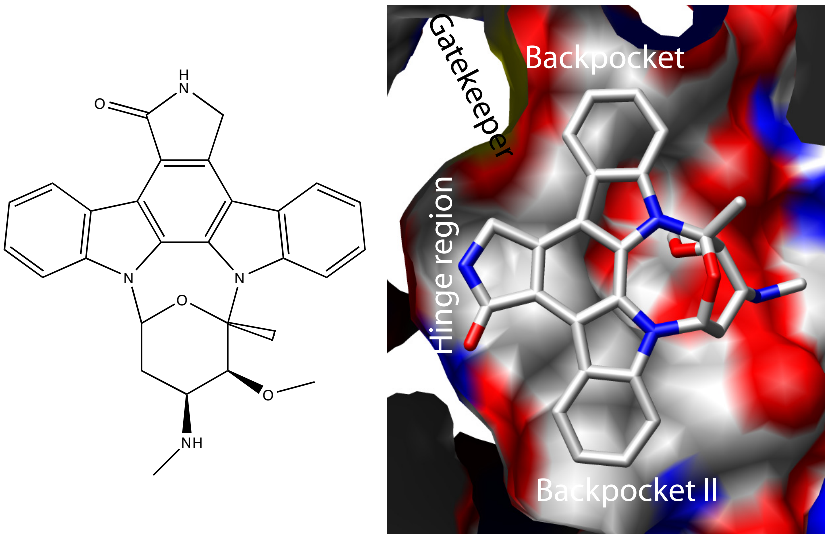


Figure 2.2: Constitution of the natural product staurosporine on the left side and on the right the crystal structure of staurosporine bound to PKA (PDB code: 1STC). Staurosporine forms hydrogen bonds to the hinge region (Val^{123} and Glu^{121} and Leu^{172}). It also moves aside the gatekeeper residue Met^{120} and Phe^{327} on the other side and penetrates into both backpockets.

In the case of p38 MAP kinase an additional hydrophobic groove is formed that allows the specific binding of the ligand SB203580, together with specific amino acids in the backpocket like Thr^{106}, [Tong et al., 1997]. NMR investigations of the complex using RDCs showed that the inhibitor and the protein are highly dynamic in comparison to the rigid crystal structure. So the interpretation of ligand binding modes considering conformational changes profits from being studied with NMR in addition to X-ray crystallography [Honndorf et al., 2008, Honndorf et al., 2012]. As mentioned above it is generally easy to obtain the crystal structure of a protein kinase with its bound ligand. Still, this is not the case for all ligands and especially flexible molecules are sometimes difficult or impossible to co-crystallize. Therefore the application of an NMR-based scoring function for docking and molecular dynamic simulations of ligand binding modes is highly desirable. It was shown in 2008, that the NMR method INPHARMA can be applied successfully to PKA [Orts et al., 2008b, Orts et al., 2008a], which is a representative for all members of the protein kinase family. Two ligand fragments targeting the ATP binding site were investigated and the crystal structures of these ligands were correctly identified out of several artificial docking modes, if some ad-hoc assumptions are taken. In the following part it will be demonstrated that the methodology can be substantially improved by taking more NMR data into account. Therefore no more assumptions have to be taken and the methodology is expanded from the original ligand pair to a diverse ligand set.

2.2 Material & Methods

2.2.1 Protein expression and sample preparation

The chinese hamster Cα catalytic subunit of cyclic adenosine monophosphate (cAMP) dependent protein kinase A (PKA) was expressed and purified according to the published procedure [Langer et al., 2004].
0.38 mL D_2O PBS-buffer with a PKA concentration of 0.045 mM and additional 150 mM NaCl were combined with 0.020 mL DMSO-d_6 with appropriate ligand concentration (L1: 0.3 mM, L2: 0.9 mM, L3: 1.2 mM, L4: 0.9 mM). 2 mM TCEP were added as an anti-aggregation agent.

2.2.2 NMR spectroscopy

INPHARMA spectra were recorded using the standard NOESY pulse program noesyphpr. Ligand combinations 1&2, 2&3 and 3&4 were recorded on a 700 MHz spectrometer and combination 1&3 on a 900 MHz spectrometer. Both spectrometers are equipped with cryogenically cooled probe heads (Bruker, Karlsruhe). The mixing times 400 ms, 600 ms, 800 ms and 1 s were recorded.
Combination B3&2 was recorded on an 700 MHz and B5&3 on an 800 MHz spectrometer with a mixing time of 800 ms. Concentrations and conditions are in the appendix.
STD spectra were recorded with a standard pulse sequence (stddiffesgp.2) with 8 scans on a 400 MHz spectrometer. Saturation was achieved by a train of shaped 90° pulses of 50 ms length. A number of 160 selective pulses was applied, leading to a total length of saturation of 8 s. 8 s were used, since the usually applied 2 s did not fully saturate the resonances. The on-resonance irradiation was performed at $+0.5$ ppm and off-resonance irradiation was set to $+30$ ppm.

2.2.3 Docking and molecular dynamics simulations

For ligands 1-4, 1000 binding modes per ligand in the crystal structure 3DNE were created using the docking software PLANTS [Korb et al., 2006]. These were clustered with 2 Å RMSD similarity filter and the best scoring structure from every cluster was used as a representative structure.
For ligands B3 and B5, 10 docking modes within the crystal structure 3DNE were created respectively. TrNOE data were used to restrain the docking of ligand B3. Three-dimensional structures of the ligands were created using CORINA [Sadowski and Gasteiger, 1994].
Energy minimizations and molecular dynamic simulations were done as implemented in Gromacs [Hess et al., 2008], using the amber99sb force field [Hornak et al., 2006] for the protein and the General Amber Force Field [Wang et al., 2004] for the ligand. EM was done with conjugate gradient integrator in 1000 steps. MD was done using steepest descent integrator with a stepsize of 2 fs. Simulations were done for 1 ns in implicit solvent, using the Generalized Born model augmented with the hydrophobic solvent accessible Surface Area (GBSA). From every trajectory 100 or 500 snapshots were saved, respectively.

2.2.4 Back-calculation of peak volumes

All INPHARMA, trNOE and STD peak volumes were back-calculated with the software SpINPHARMA, using the complete-relaxation matrix approach [London, 1999,Orts et al., 2009] as described in the introduction 1.2. Methyl-groups were described using a 3-site jump model [Tropp, 1980]. STD saturation was applied to all methyl groups. Protons within a distance of 8 Å from any ligand proton were considered. The INPHARMA peaks were normalized to the sum of all peaks in the direct dimensions [Koeck and Griesinger, 1994]. The correlation times were estimated to be 15 ns. Kd values were determined by STD competition experiments (L1: 0.03 mM, L2: 0.09 mM, L3: 0.09 mM, L4: 0.09 mM). On-rates (k_{on}) were assumed to be in the diffusion limit 10^8 $M^{-1}s^{-1}$. The multiplexing approach was realized with an in-house written Python program.

2.3 Results I - Method development

In the first part of the results it is described how the methodology was further developed, starting from the reproduction of the [Orts et al., 2008b] data. In the second part the application to several complicated ligand binding modes, including conformational changes of the protein, will be shown. In the following the test set of ligands is described and their behavior in molecular docking. Then the NMR spectroscopic data and their back-calculation is discussed which finally leads to ligand binding mode identification. In the end molecular dynamics simulations will be used to refine the structures.

2.3.1 Molecular docking

In the beginning the two ligands 1 and 2 from [Orts et al., 2008b] were used and then accompanied by the new ligands 3 and 4. These are shown in Fig. 2.3. The crystal structures of these ligands in complex with PKA are deposited in the PDB under the accession codes 3DNE, 3DND, 4EXF and 4EXD, respectively.

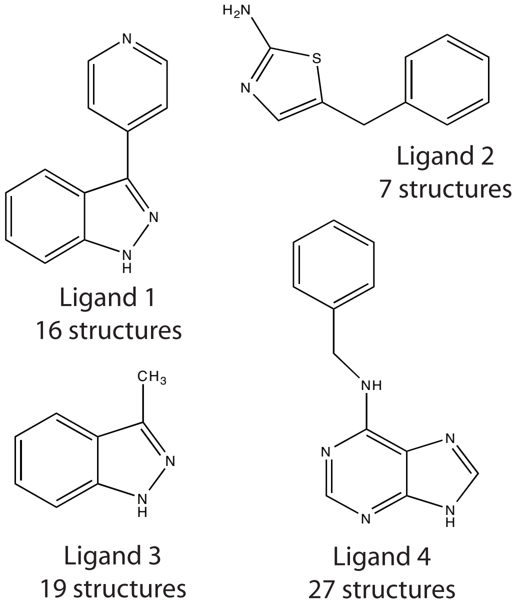


Figure 2.3: Four indazole-based PKA binding ligands. Indicated is the number of representative structures after docking.

The crystal structure of PKA in complex with ligand 1 (PDB code 3DNE) was applied to obtain the protein coordinates. This structure was chosen to make the results comparable to previous INPHARMA studies, where it was also used. 1000 docking structures were created for every ligand within the crystal structure 3DNE by the ant colony algorithm based docking program PLANTS [Korb et al., 2006]. This docking program was found to compare well with the established program GOLD [Korb et al., 2009b]. Docking models were clustered by a 2 Å RMSD similarity filter and the best scoring docking mode from every cluster was taken as a representative structure for this cluster. The docking results were scored by the PLANTS scoring function ChemPLP, and only for ligand 1 the correct binding mode received the best score. Alternative scoring of the PLANTS docking modes with ChemScore and ASP led to a comparable bad performance of docking scoring functions. Only GOLDscore was able to identify correctly the binding modes of ligand 1, 2 and 4, but not for ligand 3. Scores of the docking scoring function can be found in the appendix.

2.3.2 INPHARMA

In the first place, the experimental data from [Orts et al., 2008b] were reproduced. Sample conditions were improved by the addition of TCEP to stabilize the PKA sample against aggregation. Also the increase of protein and ligand concentrations led to better resolved spectra. The NOESY spectra of the ligand combinations 1&2 (Fig. 2.6) and 3&4 (Fig. 2.7) are shown in the following and 2&3 and 1&3 are shown in the appendix. INPHARMA signals were clearly observed for every combination. Several programs were tested for peak integration, including Sparky, TopSpin, Cara and NMRView. The box integration scheme of NMRView was the most effective tool to correctly integrate even very small and overlapping peaks, so it was used in all following investigations.

In [Orts et al., 2008b], it was demonstrated that the linear correlation coefficient $R_{INPHARMA}$ between the experimental and back-calculated INPHARMA peak volumes is better for the crystal structures, than for synthetic binding modes, as seen in Fig. 2.4.

In the present study, no synthetic binding modes were used, but real docking modes as proposed by the state-of-the-art software PLANTS. The peak volumes were back-calculated for all 70 possible double combinations of the docking modes of ligand pair 1&2 and correlated with the experimental, integrated peak volumes. The linear correlation coefficient $R_{INPHARMA}$ was used to score the quality of the fit between experimental and back-calculated peak volumes of every docking pair. The result can be seen in Fig. 2.5: $R_{INPHARMA}$ is plotted against the RMSD of the two PKA/ligand structures compared to the average crystal structures. The structure pair closest to the crystal structures is considered to be the correct answer. Yet, $R_{INPHARMA}$ places the correct answer amongst the 10% best structure pairs, but too many false positives prevent the identification of the correct structures. The identified binding mode includes the correct binding mode of ligand 2, but is far off for ligand 1.

The similar problem was found for all other double combinations (1&2,

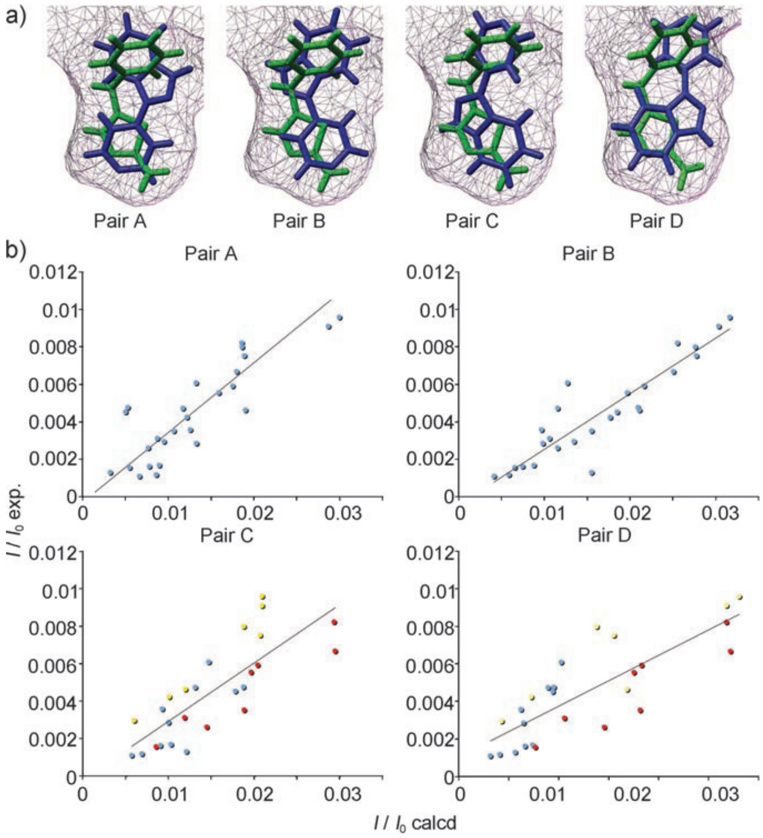


Figure 2.4: Result from [Orts et al., 2008b]: The crystal structures (pair A) show a better correlation between experimental and back-calculated INPHARMA data than synthetic rotations of these (pairs B-D). Some ad-hoc criteria defined in this publication also support this finding.

2&3, 1&3 and 3&4): The correct answer was scored approximately amongst the 10% best scoring structures, but the identified binding modes were far off the correct answer. An approach to use a higher number of docking modes (one thousand structures per ligand, resulting in one million double combinations), resulted just in higher numbers of false positives (data not shown). The final conclusion is, that not the number of structures should be increased, but the amount of data input, given that no additional NMR work is necessary. From the INPHARMA NOESY spectra the trNOE can be extracted, yielding information on the bound ligand conformation. Still, this information is not very useful for small, rigid molecules as in the present case.

Before the recording of an INPHARMA NOESY, the ligands are checked for binding to the protein by STD. Remembering the introduction part, STD

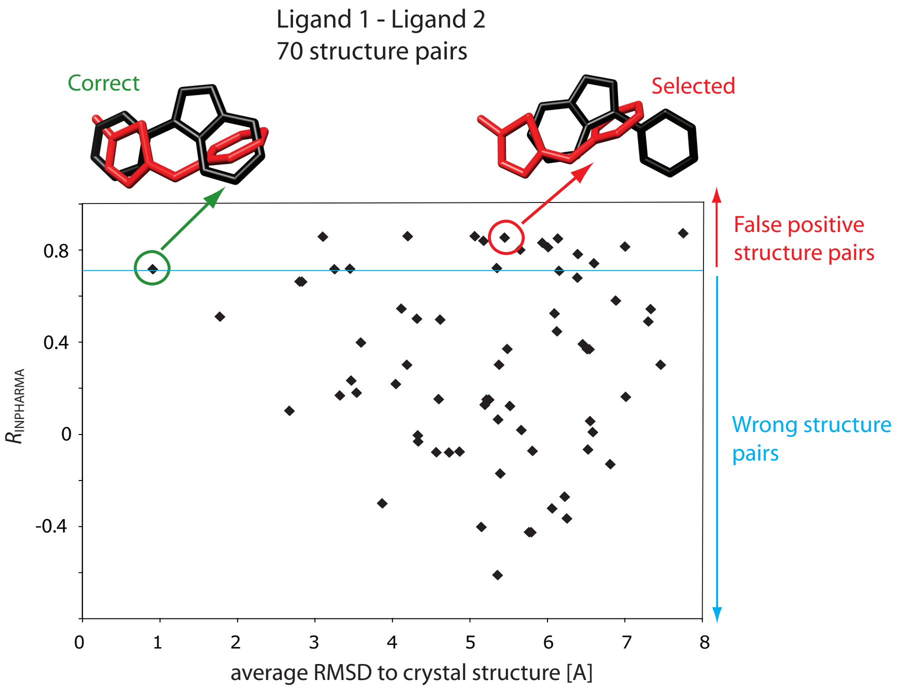


Figure 2.5: $R_{INPHARMA}$ plotted against the RMSD of the two PKA/ligand structures compared to the crystal structures, the best structure according to $R_{INPHARMA}$ has a red circle, the best answer has a green circle. $R_{INPHARMA}$ does not pick the correct structure as there are too many false positives.

not only gives information on the binding behavior of a ligand, but also on the ligand binding mode.

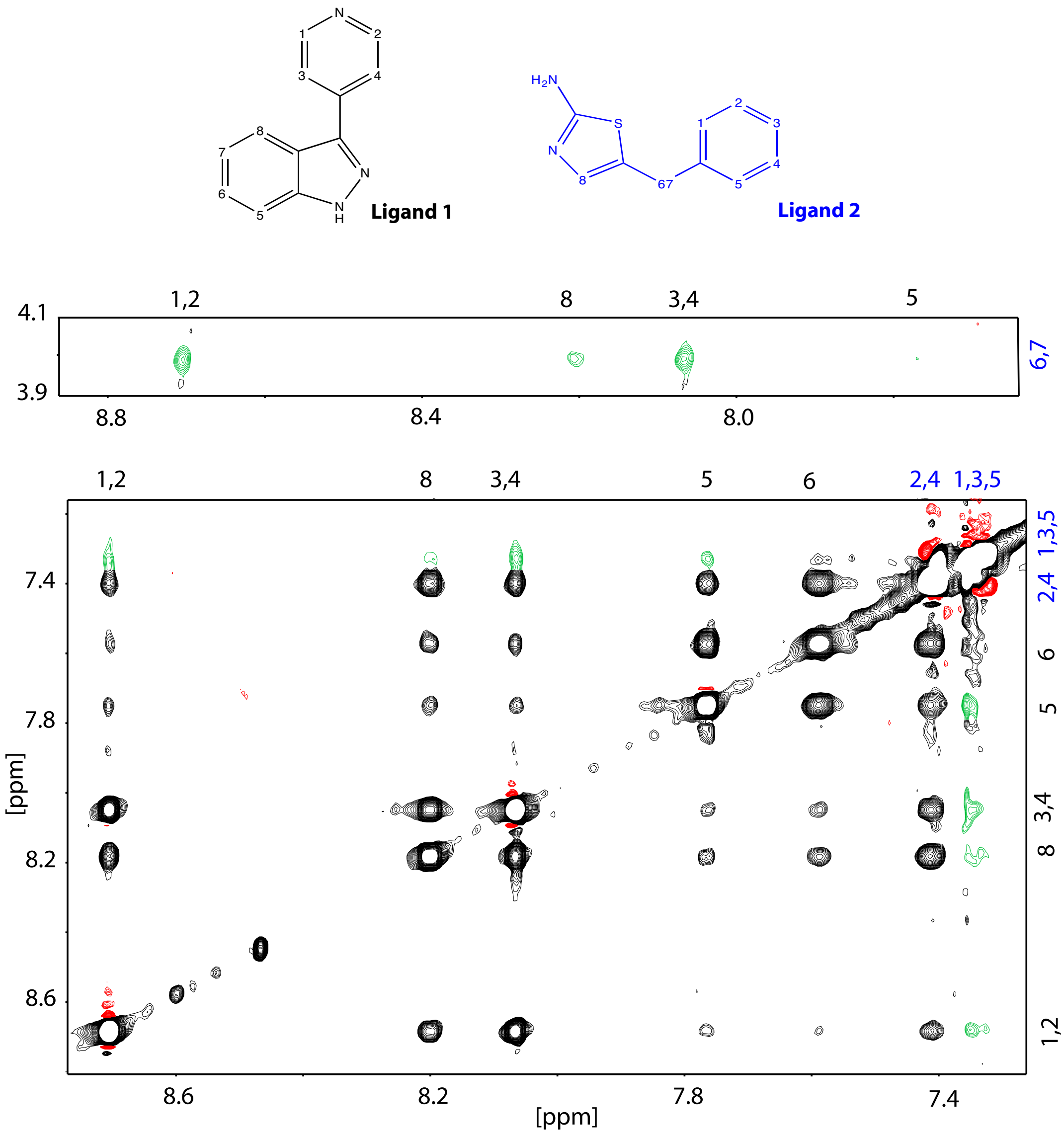


Figure 2.6: NOESY spectrum of ligand 1 (0.3 mM) and ligand 2 (0.9 mM) in the presence of PKA (0.045 mM). INPHARMA peaks chosen are marked in green. The mixing time was 800 ms on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 293 K with 64 scans, 2048 points in F2 and 580 points in F1.

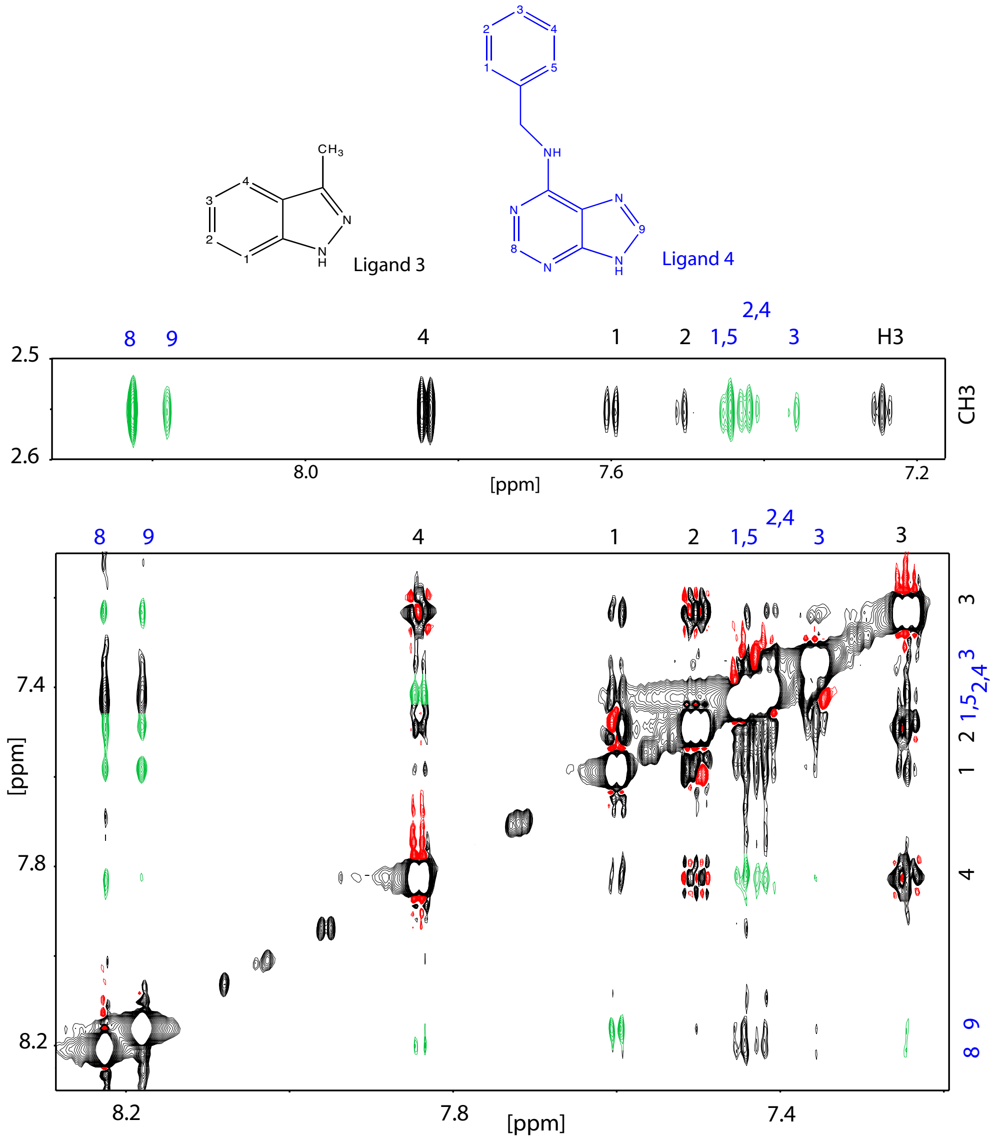


Figure 2.7: NOESY spectrum of ligand 3 (0.9 mM) and ligand 4 (0.9 mM) in the presence of PKA (0.045 mM). INPHARMA peaks chosen are marked in green. The mixing time was 800 ms on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 293 K with 64 scans, 2048 points in F2 and 580 points in F1.

2.3.3 STD

In principle one could simply use the STD spectra that were recorded prior to the INPHARMA NOESY measurements. In the present study, every ligand was measured with PKA on its own, to avoid signal overlaps. When STD is used to investigate if a ligand is binding to a protein, very small protein concentrations are used. Now, that the binding epitope should be derived, the protein was as high concentrated as for the INPHARMA measurements (0.045 mM). In this manner the binding epitope was found to be derived in the optimal way. Different saturation times (2 s, 4 s, 6 s and 8 s) were tested for ligand 1 and a build-up series of the ligand signals was recorded. As can be seen in Fig. 2.8, the signal increases up to 6 s and then stays at this intensity level. Finally, 8 s were chosen as the optimum saturation time, to ensure that all resonances are saturated and the optimal signal intensity is achieved.

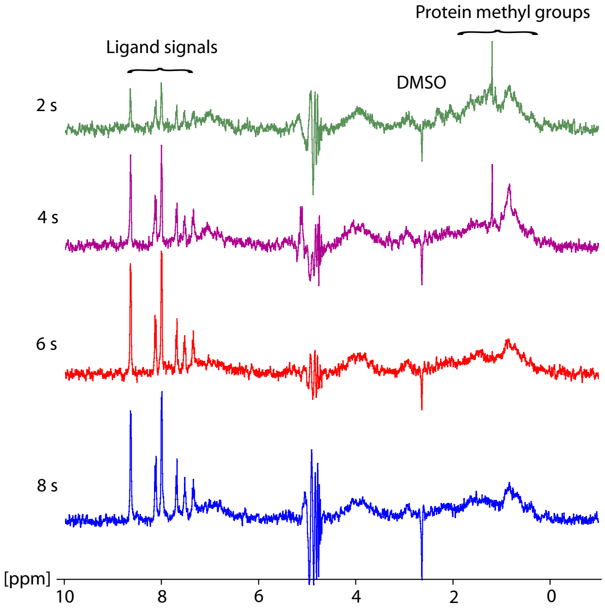


Figure 2.8: STD build-up curve of ligand 1 (0.3 mM) in solution with PKA (0.045 mM). Saturation was done on methyl groups. Signal intensities increase to a saturation time of 8 s.

One spectrum was recorded with on-resonance irradiation (0.5 ppm) and a second spectrum with off-resonance irradiation (30 ppm). The STD spectrum was obtained by subtracting the first from the latter manually. In this manner the 1D (off-resonance irradiation) and the STD spectrum are better comparable than, if only an interleaved difference spectrum is recorded, as it is normally done. For all four ligands, the binding epitope could be derived from the STD spectra. This is shown in Fig. 2.9.

For the back-calculation of STD, the experimental peak volumes (V) were used as defined in equation:

$$V_{STD} = \frac{V_{off-resonance} - V_{on-resonance}}{V_{off-resonance}} \tag{2.1}$$

Back-calculated peak volumes were treated in the same way. Both, experimental and back-calculated volumes are multiplied by 100 as a scaling factor for better representation.

For the graphical representation of STD in percent that was introduced by [Mayer and Meyer, 2001] and is shown in the figures, the largest STD peak volume $V_{Max.STD}$ was set to 100 % and the other peak volumes were normalized according to:

$$STD_{in\%} = \frac{V_{STD}}{V_{Max.STD}} * 100 \tag{2.2}$$

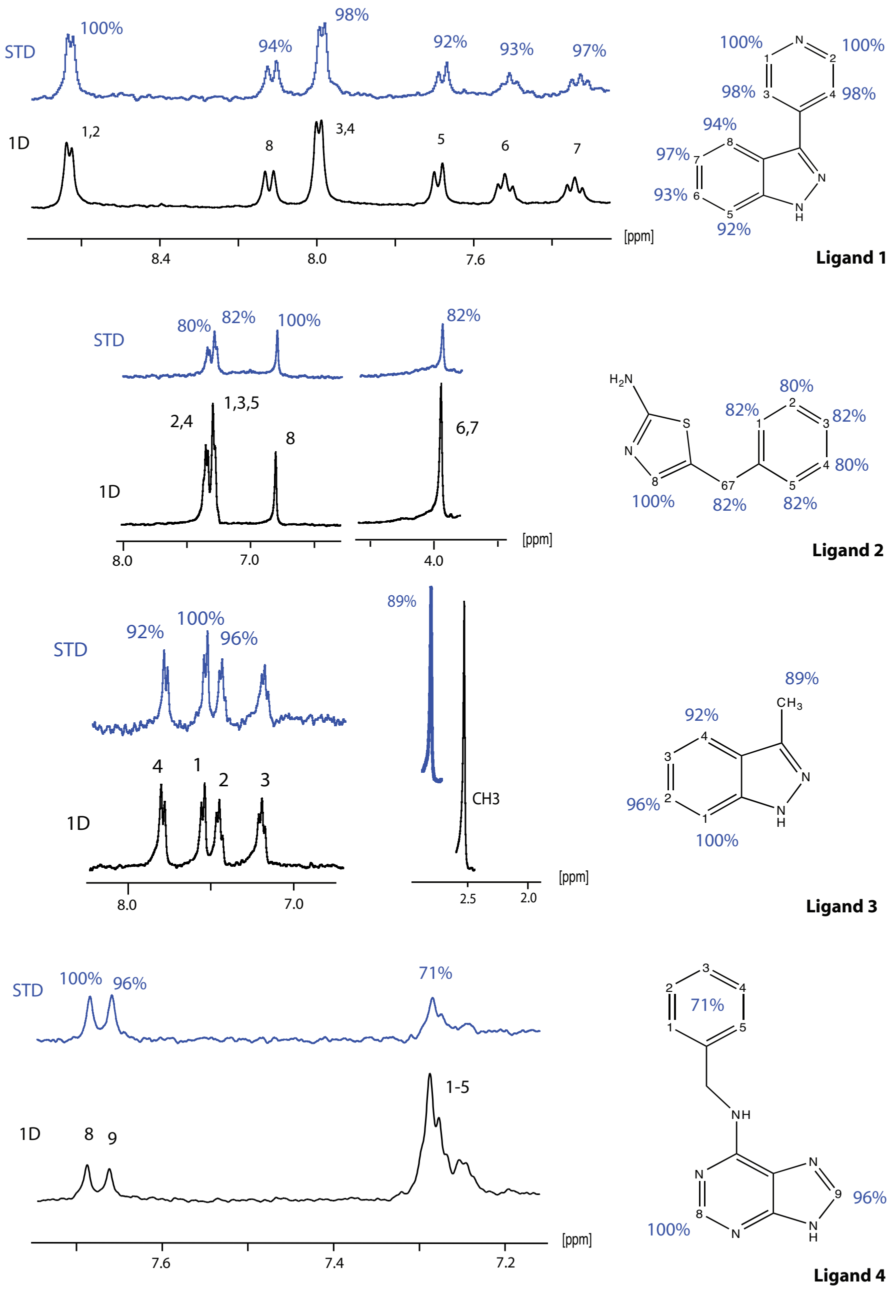


Figure 2.9: Experimental STD data of the four PKA binding ligands. The normal 1D in the presence of 0.045 mM PKA is shown in black and the STD spectrum in blue. Ligand concentrations were 1 (0.3 mM), 2 (0.9 mM), 3 (1.2 mM) and 4 (0.9 mM). Saturation was applied for 8 s on 0.5 ppm on a 400 MHz spectrometer (number of scans = 8; temperature = 298 K; time domain = 16384 points). Proton 3 of ligand 3 was dismissed due to broadening, caused by an overlap with broadened protein signals. The $V_{Max.STD}$ values are: L1: 13.75; L2: 28.6; L3: 16.91; L4: 32.66.

2.3.4 INPHARMA-STD

In the following a new approach was tested to combine INPHARMA with the STD data, also considering trNOE. The different data sets were combined using a simple average of the correlation coefficients R_{STD}, R_{trNOE} and $R_{INPHARMA}$ to R_{STI}:

$$R_{STI} = \frac{R_{STD} + R_{trNOE} + R_{INPHARMA}}{3} \tag{2.3}$$

Hereby STD and trNOE of the two ligands are combined as:

$$\begin{aligned} R_{STD} &= \frac{R_{STD(Ligand1)} + R_{STD(Ligand2)}}{2} \\ R_{trNOE} &= \frac{R_{trNOE(Ligand1)} + R_{trNOE(Ligand2)}}{2} \end{aligned} \tag{2.4}$$

In the present case the software SpINPHARMA describes already the combined trNOE for the ligand pair, so the last equation was not necessary. The results for the double combination are shown in the following figures. They are striking in three ways: i) the combination of INPHARMA and STD finds the correct (or close to correct) binding mode in every example, ii) this works even though INPHARMA and STD would fail respectively on their own and iii) the NMR approach performs better than the docking scoring functions ChemPLP and GOLDScore.

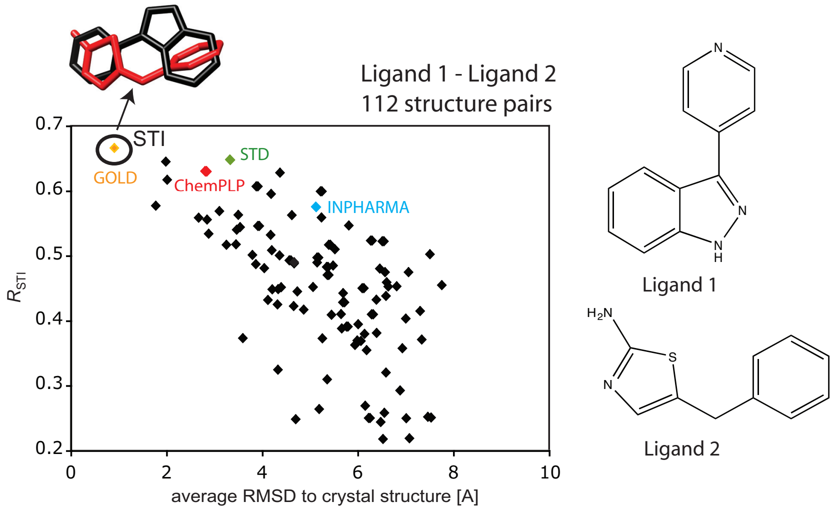


Figure 2.10: R_{STI} plotted against the RMSD of the two ligand/PKA structures compared to the crystal structures, the best structure according to R_{STI} has a black circle. The best structures selected by different scoring are represented in blue (INPHARMA), green (STD), red (ChemPLP) and yellow (GOLDScore). R_{STI} selects the correct docking mode, i.e. the mode with the smallest RMSD to the average crystal structure.

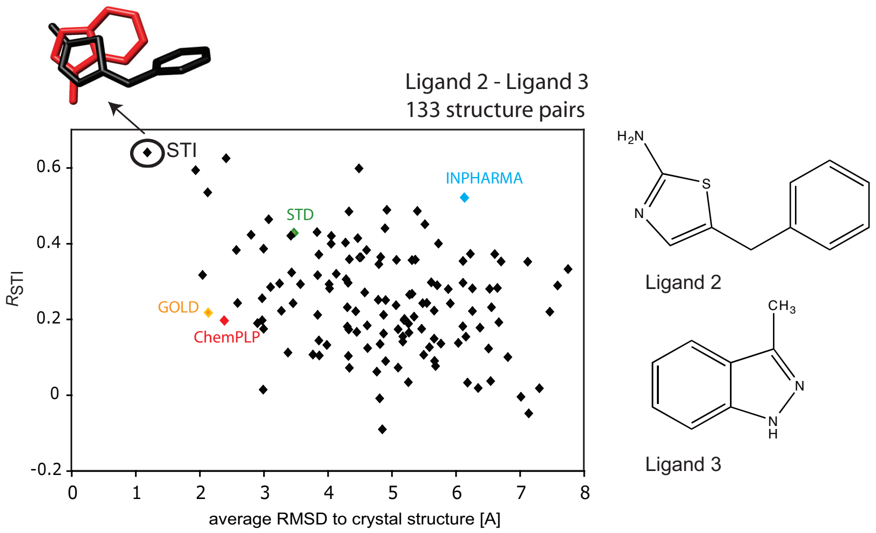


Figure 2.11: R_{STI} plotted against the RMSD of the two ligand/PKA structures compared to the crystal structures, the best structure according to R_{STI} has a black circle. The best structures selected by different scoring are represented in blue (INPHARMA), green (STD), red (ChemPLP) and yellow (GOLDScore). R_{STI} selects the correct docking mode, i.e. the mode with the smallest RMSD to the average crystal structure.

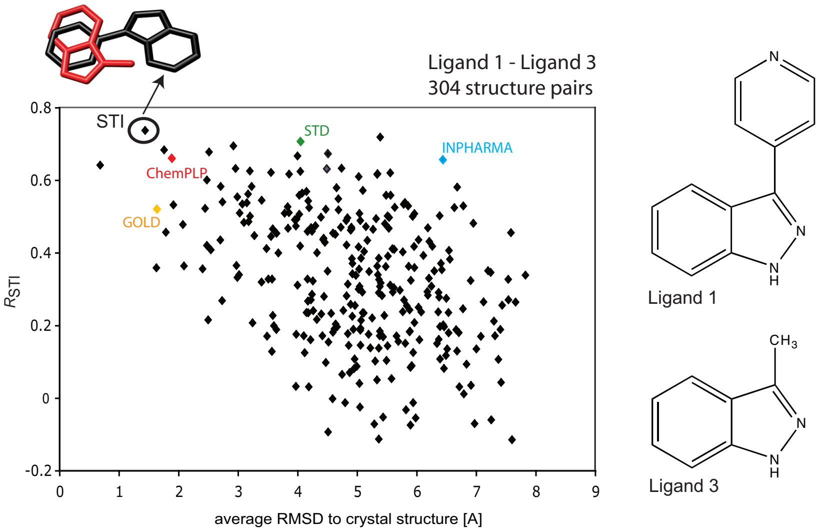


Figure 2.12: R_{STI} plotted against the RMSD of the two ligand/PKA structures compared to the crystal structures, the best structure according to R_{STI} has a black circle. The best structures selected by different scoring are represented in blue (INPHARMA), green (STD), red (ChemPLP) and yellow (GOLDScore). R_{STI} selects the second best docking mode, i.e. the mode with the smallest RMSD to the average crystal structure.

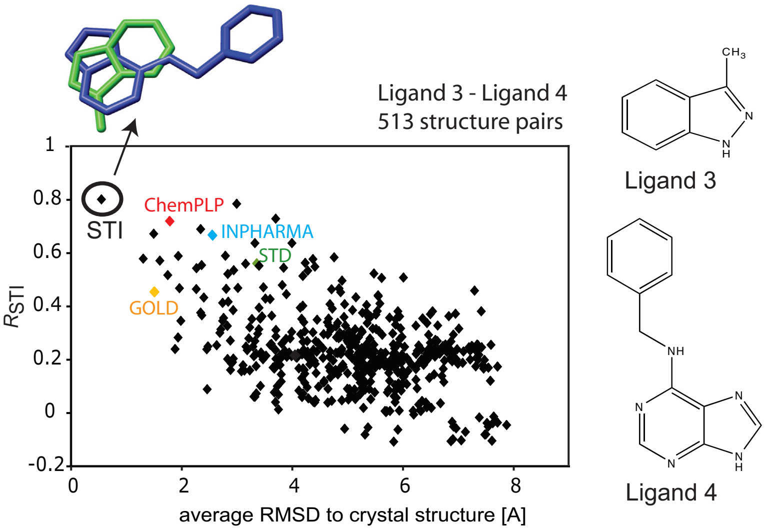


Figure 2.13: R_{STI} plotted against the RMSD of the two ligand/PKA structures compared to the crystal structures, the best structure according to R_{STI} has a black circle. The best structures selected by different scoring are represented in blue (INPHARMA), green (STD), red (ChemPLP) and yellow (GOLDScore). R_{STI} selects the correct docking mode, i.e. the mode with the smallest RMSD to the average crystal structure.

2.3.5 Multiple ligands (Multiplexing)

In order to evaluate the impact of having more ligands than two (multiplexing), the R_{STI} scores of all ligand pairs were combined to $R_{STI}^{Multiplexing}$, using the following equation:

$$R_{STI}^{Multiplexing} = \sqrt{\frac{R_{STI}^2 1\&2 + R_{STI}^2\ 1\&3 + R_{STI}^2\ 2\&3 + R_{STI}^2\ 3\&4}{4}} \tag{2.5}$$

The ligand combinations 1&2, 1&3, 2&3 and 3&4 were chosen, as they give well dispersed spectra, while the combinations 1&4 and 2&4 display many overlaps and therefore not enough quantifiable INPHARMA peak. As seen in Fig. 2.14, the differentiation between the correct structures and those with larger RMSD is even more pronounced and the correct binding mode can be identified with $R_{STI}^{Multiplexing}$ more easily. It is noteworthy that this multiplexing of more than two ligands substantially improves INPHARMA scoring, while STD, trNOE and the docking scoring functions do not improve. In the example presented, the multiplexed INPHARMA scoring found the correct binding mode, while it failed with only one pair. To minimize INPHARMA measurement time and optimize INPHARMA cross peak integrals each ligand should then be measured only twice in one pair with the ligand that binds slightly weaker and in the other pair with the ligand that binds slightly stronger: (n-1). Hereby n is the number of ligand pairs. This approach is done by using the equation:

$$R_{STI}^{Multiplexing} = \sqrt{\frac{R_{STI}^2\ 1\&2 + R_{STI}^2\ 2\&3 + R_{STI}^2\ 3\&4}{3}} \tag{2.6}$$

and is shown in Fig. 2.15. $R_{STI}^{Multiplexing}$ finds again the correct structures, but INPHARMA picks structures as the best scoring ones, which are far away in RMSD (more than 5 Å) from the correct structures. This observation underlines, that $R_{STI}^{Multiplexing}$ is a very robust scoring factor, while $R_{INPHARMA}$ can perform very well, but does not show the same level of robustness as $R_{STI}^{Multiplexing}$ and depends strongly on the number of structure pairs that are taken into account.

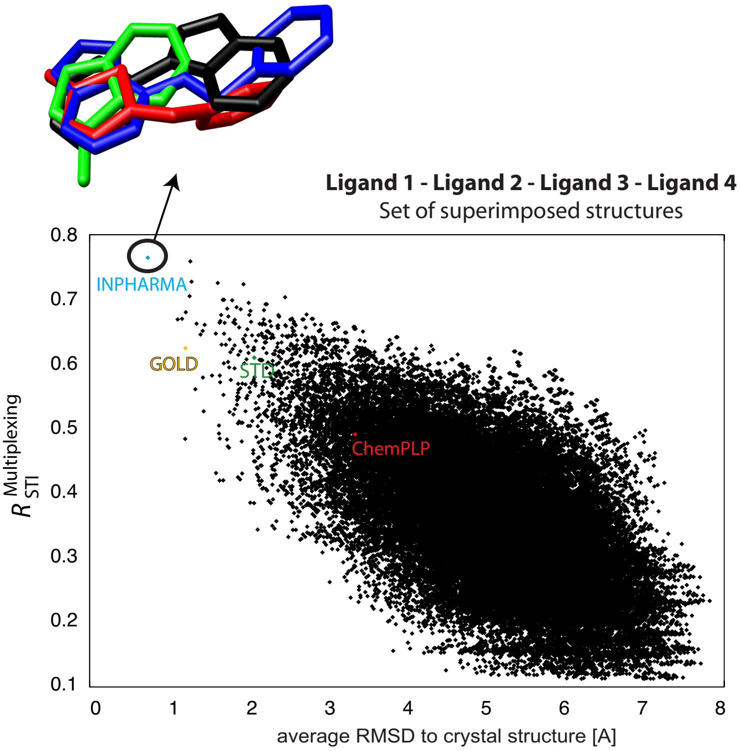


Figure 2.14: $R_{STI}^{Multiplexing}$ plotted against the RMSD of all four ligand/PKA structures compared to the crystal structures, the best structure according to $R_{STI}^{Multiplexing}$ has a black circle. The best structures selected by the different scoring are represented in blue (INPHARMA), green (STD), purple (trNOE) red (ChemPLP) and yellow (GOLDScore). $R_{STI}^{Multiplexing}$ and INPHARMA find the correct docking mode, i. e. the mode with the smallest RMSD to the average crystal structure.

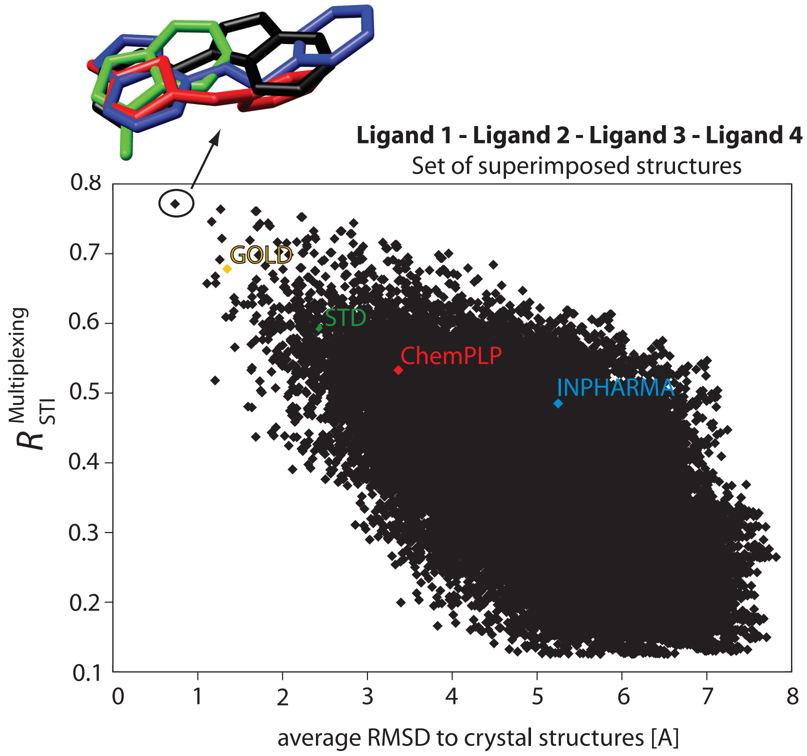


Figure 2.15: $R_{STI}^{Multiplexing}$ plotted against the RMSD of all four ligand/PKA structures compared to the crystal structures, the best structure according to $R_{STI}^{Multiplexing}$ has a black circle. The best structures selected by the different scoring are represented in blue (INPHARMA), green (STD), purple (trNOE) red (ChemPLP) and yellow (GOLDScore). $R_{STI}^{Multiplexing}$ finds the correct docking mode, i. e. the mode with the smallest RMSD to the average crystal structure.

2.3.6 Molecular dynamics simulations and refinement

In the previous studies a rigid system was assumed and the docking mode in the crystal structure is an average of what is seen in solution. This is a very simple approximation but useful to derive the correct binding orientation of the ligand without using too many structural models. The question is now arising, if the best docking mode is really the most faithful? In solution the protein is a dynamic system and the ligand will be subjected to strong forces, moving it in the binding pocket and modulating its conformation. Therefore investigations with molecular dynamics simulations using a given force field without experimental restraints were done, starting with the docking model of each ligand, respectively. MD studies were done on all ligands presented here. As an example ligand 1 is shown in Fig. 2.16.

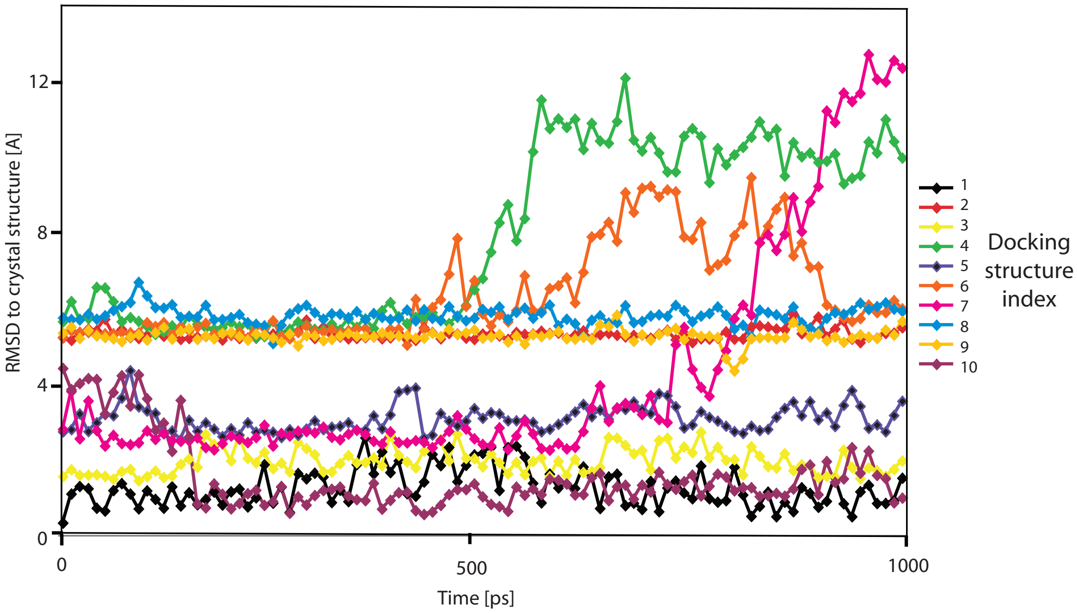


Figure 2.16: Ten docking modes of ligand 1 to PKA were sent to a molecular dynamics simulation of 1 ns. Snapshots were taken every 10 ps. Indicated is the change in RMSD to the crystal structure. While some binding modes remain stable in the binding pocket, others show strong dynamics or even move out of the binding site.

The Pearson correlation coefficient R was used to score the fitting between the experimental and back-calculated INPHARMA, trNOE and STD data. R was not squared to take into account, that small ligands only display a few proton peaks, and the differentiation of a positive or negative correlation is the valuable information, as it reflects directly the buried and exposed parts

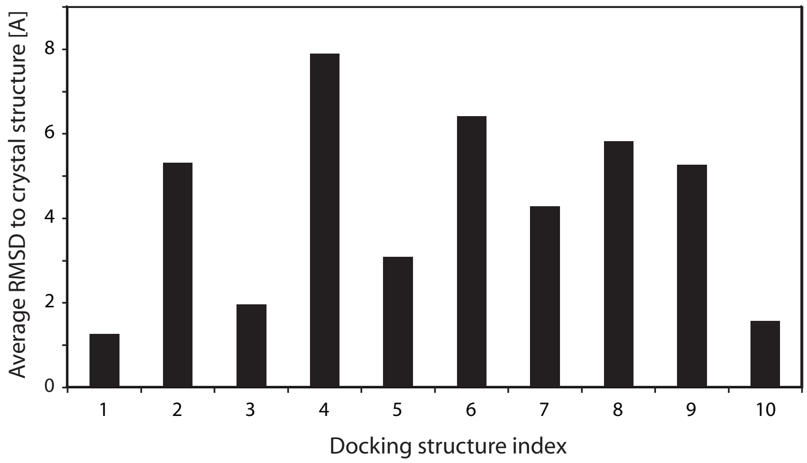


Figure 2.17: Averages of the trajectories shown in Fig. 2.16.

of the bound ligand in the case of STD. R can differentiate this very well, while R^2 can not. The same holds true for a quality factor defined here as:

$$Q^2 = \frac{\sum(V_{exp} - V_{calc})^2}{\sum(V_{exp})^2} \tag{2.7}$$

In order to cover a larger conformational space, docking structures with good R_{STI} values can be run in an unrestrained MD simulation. By taking snapshots of the resulting trajectories many new structures close to the starting structure are generated. These can then be cross-validated against the experimental data in order to find the best one, respectively. For this purpose every structure obtained from a snapshot was again scored by R_{STI}. In this case the application of a quality factor (equation 2.7) is appropriate, because small changes in the geometry are balanced by the high number of calculated structures. Hereby the Q-Factor is the most important selection criterion. In Fig. 2.18 and Fig. 2.19 the refinement is done against STD, trNOE and INPHARMA data independently. All refinement methods end up in a structure pair very close to the crystal structure reference and display very good correlations between experimental and back-calculated peak volumes.

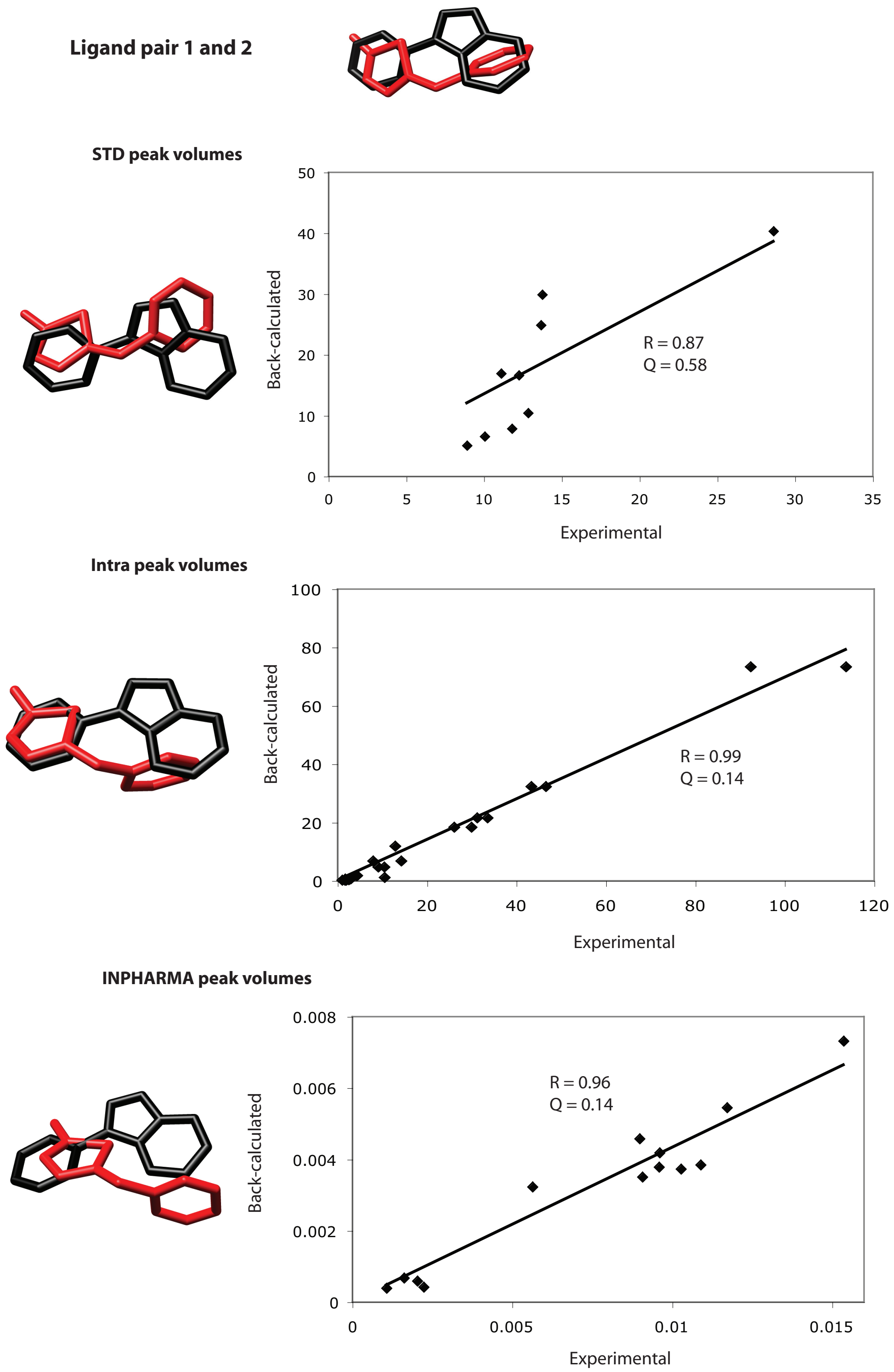


Figure 2.18: The ligand pair 1 and 2 with the best R_{STI} score was run in an unrestrained MD simulation for 1 ns and resulting snapshot structures (500 per ligand) were rescored against STD, trNOE and INPHARMA peak volumes, respectively. Correlation of the experimental and back-calculated peak volumes are shown.

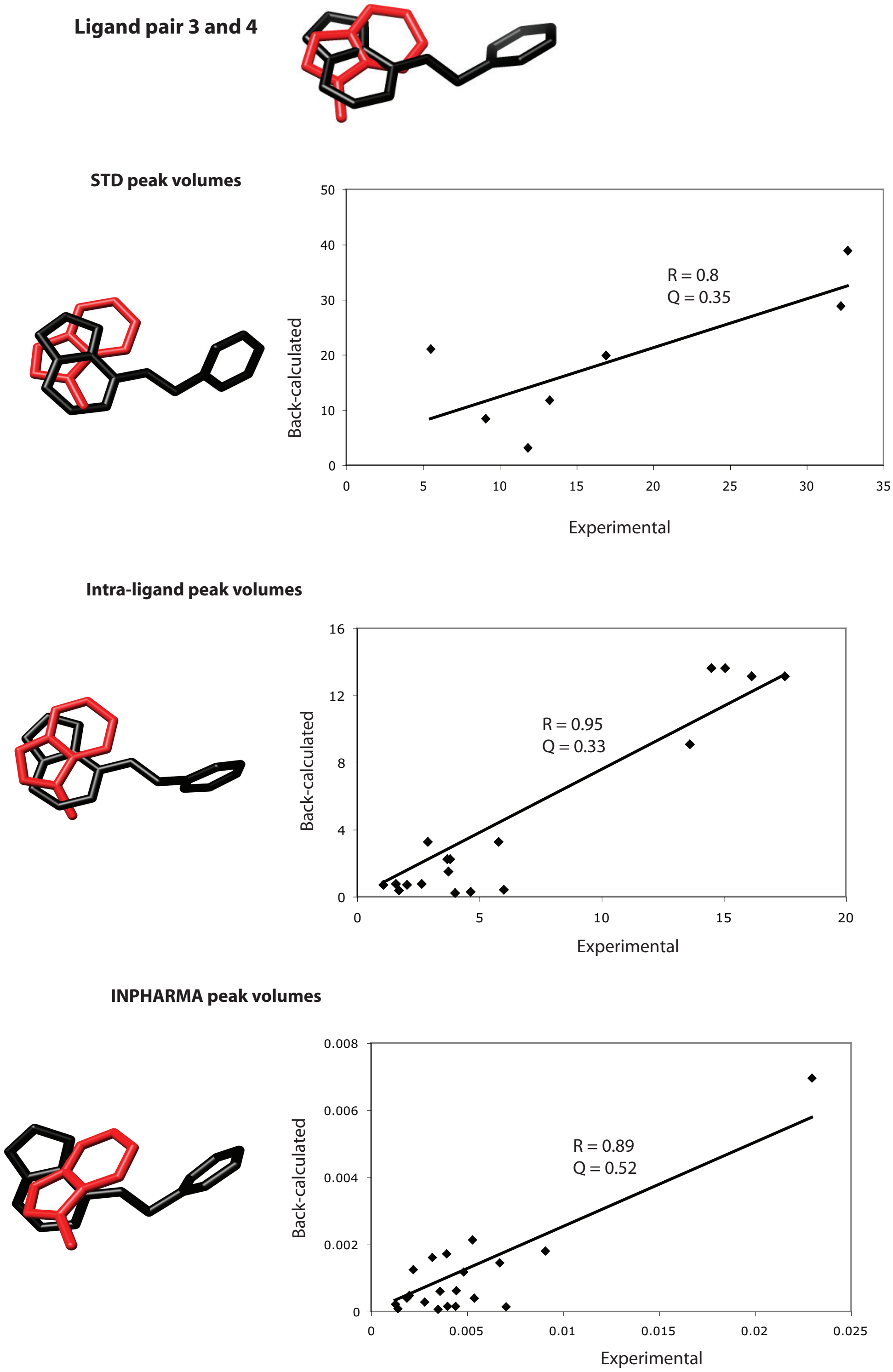


Figure 2.19: The ligand pair 3 and 4 with the best R_{STI} score was run in an unrestrained MD simulation for 1 ns and resulting snapshot structures (500 per ligand) were rescored against STD, trNOE and INPHARMA peak volumes, respectively. Correlation of the experimental and back-calculated peak volumes are shown.

2.4 Results II - Application and testing

2.4.1 Ligands and conformational changes

In the preceding chapter the methodology was developed and successfully yielded the correct binding mode of ligands to PKA. It has to be mentioned that the ligands could be docked correctly into the protein crystal structure by the docking program PLANTS, which is very favorable. It is very important for docking that the side chains of the protein crystal structure have an orientation that allows the ligand to adopt the correct binding mode during the docking process. In other words, the correct binding mode can obviously not be selected, if it is not found during docking. This represents one of the largest issues in molecular docking: to dock a ligand correctly into the protein binding site, the side chains must be in an adequate conformation. When the quality of docking programs is evaluated, the procedure is often as follows: The ligand is taken out of the protein-ligand complex crystal structure and then docked back into the empty binding pocket, e.g. done in [Meiler and Baker, 2006, Cross et al., 2009]. Around 80% of ligands can be docked correctly this way with standard docking programs like GOLD [Korb et al., 2009b]. Yet, this hit rate would be much lower, if the side chains would not be in the adequate conformation. Approaches of using flexible side chains or backbone motions via molecular dynamics simulations within the docking process are normally very time consuming and it is not guaranteed, that the correct solution will be found. In the following, the developed methodology of NMR-based protein-ligand complex structure determination is applied to three advanced binding modes, meaning that the correct binding mode is more difficult to be found, as for example a conformational change of a side chain is necessary.

2.4.2 Advanced binding mode I - Biased crystal structure

Ligand e (Fig. 2.21) is a small fragment termed chlor-indazole and binds PKA with a very low affinity (Kd = 1.2 mM) and the binding mode could not be derived from the crystal structure reliably. Even though the crystal structure was obtained, two possible binding modes were found and these could not be further distinguished. These two crystal structures are shown in Fig. 2.20. Both make a hydrogen bond with the NH of Val^{123}, but they are flipped by 180° relative to each other. Now that the methodology was established in the previous chapter, the ligand e was again investigated, using INPHARMA and STD data.

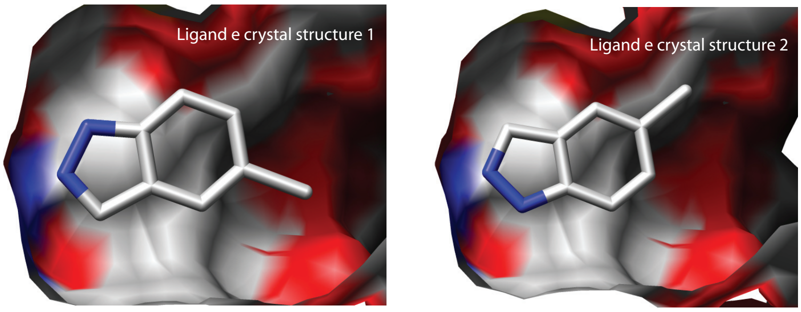


Figure 2.20: The two different crystal structures that were obtained for ligand e.

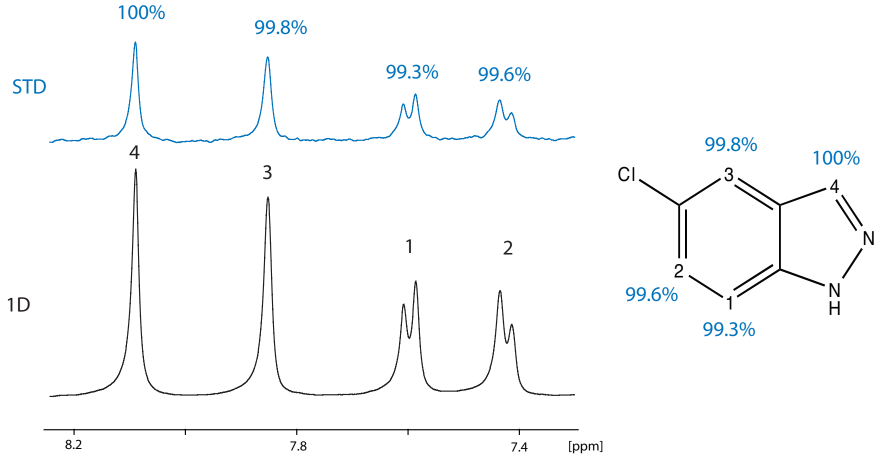


Figure 2.21: Experimental STD data of PKA binding ligand e. The normal 1D in the presence of 0.03 mM PKA is shown in black, and the STD spectrum in blue (enlarged by a factor of 4). Ligand concentration was 1.2 mM. Saturation was applied for 8 s on 0.5 ppm on a 400 MHz spectrometer (number of scans = 8; temperature = 298 K; time domain = 16384 points).

An STD of ligand e with PKA was recorded as shown in Fig. 2.21. Due to the low affinity, the difference spectrum only showed very low differences between the peak volumes, when the binding epitope was derived, but they were measurable.
The INPHARMA NOESY was recorded together with ligand 3, as this ligand pairs did not show overlaps in the aromatic region. R_{STI} was calculated for ten docking models of ligand e, including the two crystal structures and is shown in Fig. 2.22. One structure is clearly favored with an R_{STI} of 0.73 and represents crystal structure 1. Crystal structure 2 received an R_{STI} of only 0.42. Therefore it is concluded that crystal structure 1 is the favored binding mode of ligand e in solution.

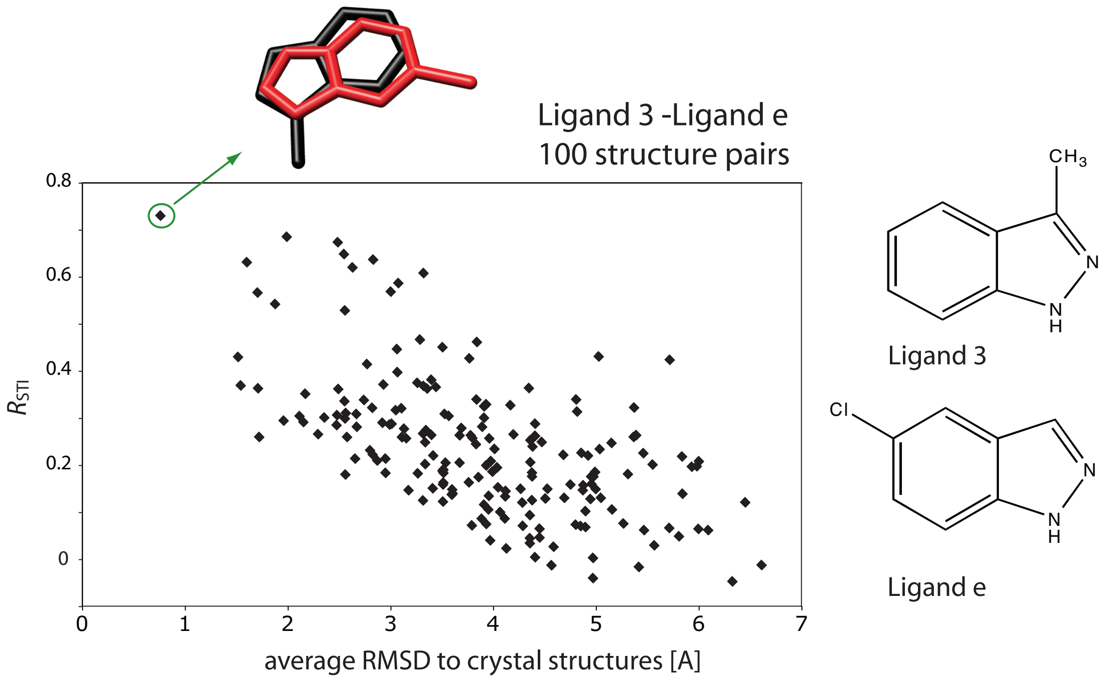


Figure 2.22: R_{STI} plotted against the RMSD of the two ligand/PKA structures compared to the crystal structures (in the case of ligand e crystal structure 1 is shown), the best structure according to R_{STI} has a black circle.

2.4.3 Advanced binding mode II - Backpocket penetration

It was described in section 2.1.2 that the ATP binding site of PKA hosts a backpocket, which is guarded by the gatekeeper residue Met^{120}. Most kinases display a similar motive, that a backpocket is blocked by a gatekeeper residue. The ATP binding site of kinases is not very specific and therefore it is crucial for the medicinal chemist to develop drugs, that are specific for the ATP-binding site of the targeted kinase. For this purpose the penetration of a ligand into the backpocket can highly improve selectivity. As was described before, the kinase inhibitor staurosporine occupies both backpockets of PKA. At this stage, the developed NMR method can only deal with ligands that bind to the PKA structure with a closed backpocket. To get the correct binding mode of a backpocket penetrating ligand, the gatekeeper residue has to change its conformation. How can this be realized for the NMR method? In the following, it will be shown that the methodology can actually already handle this issue. Therefore ligand B5 was chosen, where it is known from the crystal structure, that a penetration into the backpocket takes place, a process which is described in detail in section 2.1.2.

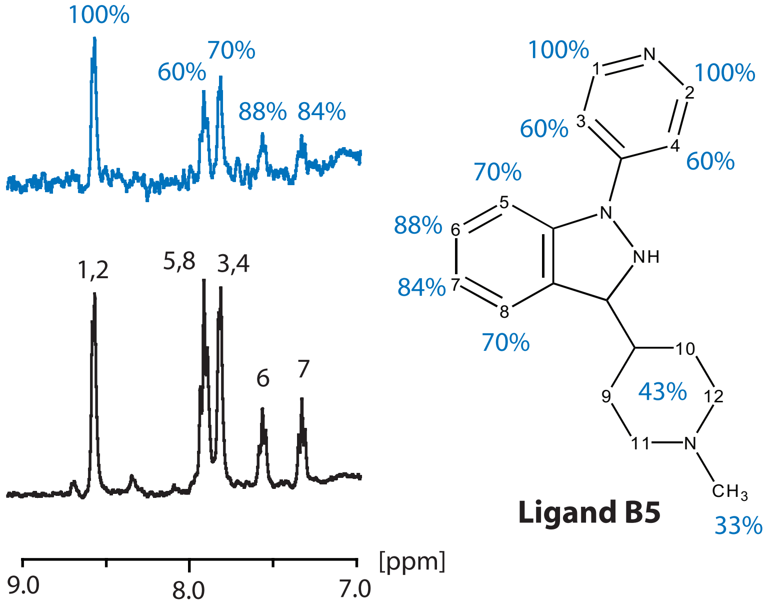


Figure 2.23: Experimental STD data of PKA binding ligand B5. The normal 1D in the presence of 0.03 mM PKA is shown in black, the difference spectrum between 1D and STD in blue. Ligand concentration was 0.5 mM. Saturation was applied for 8 s on 0.5 ppm on a 400 MHz spectrometer (number of scans = 8; temperature = 298 K; time domain = 16384 points). The $V_{Max.STD}$ value is 30.97.

In the first place, an STD spectrum was recorded, as seen in Fig. 2.23. Visual inspection already gives a clear proposal: Protons H1-H4 will be buried in the protein, while the sugar group is exposed to the solvent. The ligand recalls partially to ligand 1, yet the binding mode seems to be different, as the protons H6 and H7 receive unexpected strong magnetization in comparison to protons H5 and H8.

Second, an INPHARMA NOESY was recorded. Ligand 3 was chosen as the competitive ligand, as there were least overlaps in the 1D spectrum with this ligand in comparison to others. Visual inspection of the spectrum (shown in the appendix) confirms the STD observations: The protein buried methyl group of ligand 3 is very close to protons H1-H4, but far from the sugar group and the other protons of ligand B5. 10 docking modes of ligand B5 were created and STD, trNOE and INPHARMA were back-calculated for all docking modes. The result can be seen in Fig. 2.24: Two docking modes, very close in RMSD, score better than the other structure pairs. If it is assumed that the correct binding mode of ligand 3 is already known, the selection gets even

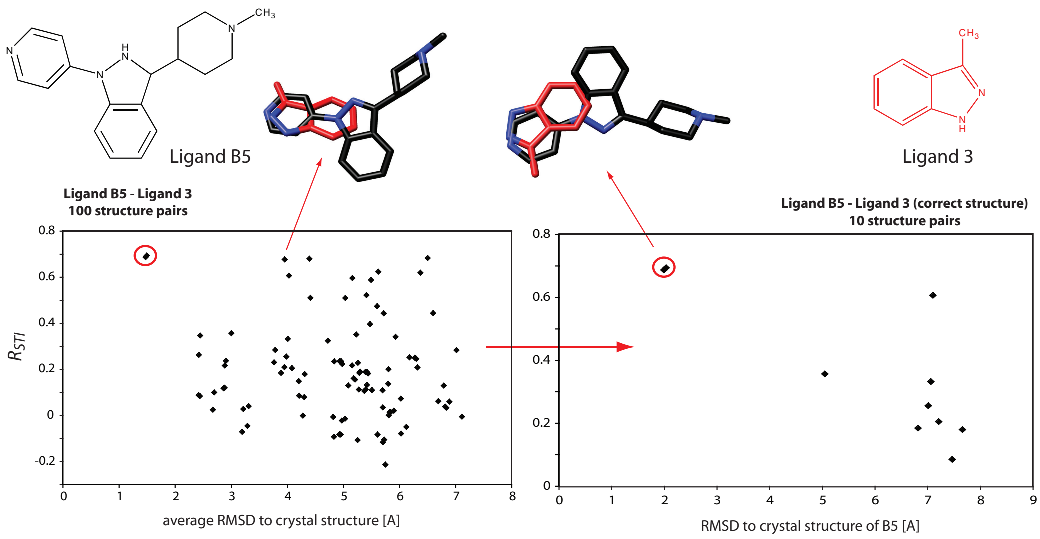


Figure 2.24: The 100 pairwise combinations of ligand B5 (black) and ligand 3 (red) scored against R_{STI} on the left side. On the right side only the pairs with the correct structure of ligand 3 are shown. Therefore only the RMSD of ligand B5 to the crystal structure is shown and not the average RMSD.

much clearer.

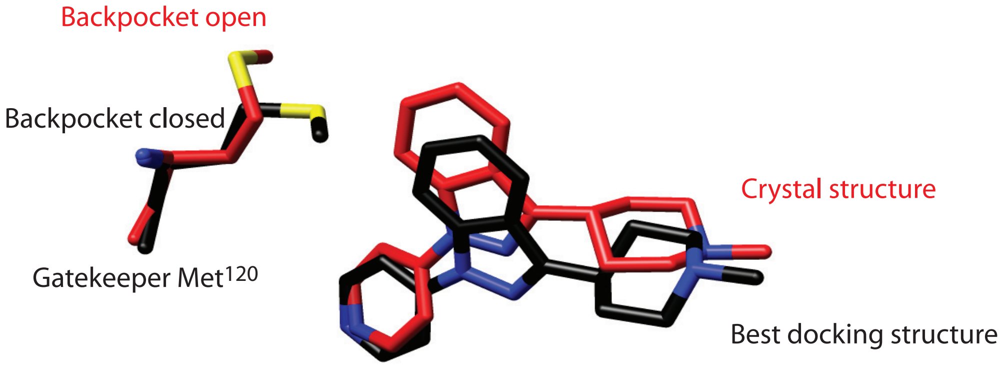

Backpocket open
Backpocket closed
Gatekeeper Met120
Crystal structure
Best docking structure

Figure 2.25: Best docking model of ligand B5 by R_{STI} in black and the crystal structure in red. The gatekeeper residue Met^{120} normally closes the backpocket in PKA. But for ligand B5 in the bound state, the backpocket is open as seen in the crystal structure.

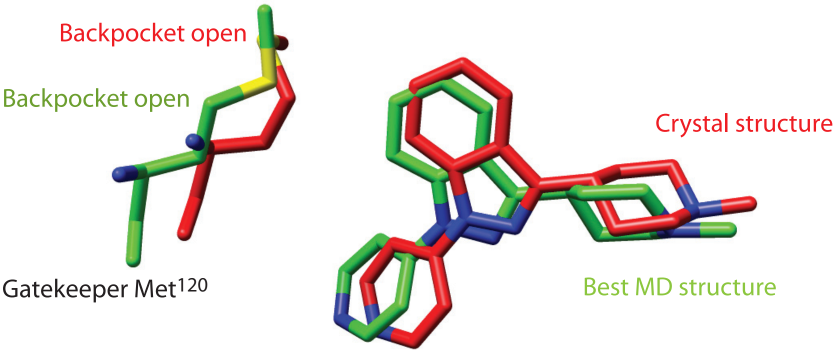


Figure 2.26: Best MD structure of ligand B5 by R_{STI} in green and the crystal structure in red. The gatekeeper residue Met^{120} is in the open position in both structures.

If the obtained binding mode is inspected and compared with the crystal structure (Fig. 2.25), it can be seen that they are very close in RMSD. Yet, the backpocket can not be entered, as the gatekeeper residue Met^{120} is in the closed and not in the open conformation. Therefore it would not be noticed, that this ligand penetrates into the backpocket and has the potential to be optimized to a highly specific binder.
Now the obtained ligand binding mode was run in an unrestrained MD simulation for 1 ns, as was done in the MD refinement described above. Indeed, during the MD simulation the gatekeeper immediately opens (Fig. 2.27) and ligand B5 penetrates into the backpocket. And most interestingly, these structures with the backpocket penetration score (best $R_{INPHARMA}$: 0.94) better than the initial docking model with the closed backpocket (best $R_{INPHARMA}$: 0.86. The structure with the best score is shown in Fig. 2.26. The RMSD is only slightly closer to the crystal structure in comparison to the best docking mode but the interesting observation, the backpocket penetration, is obtained.

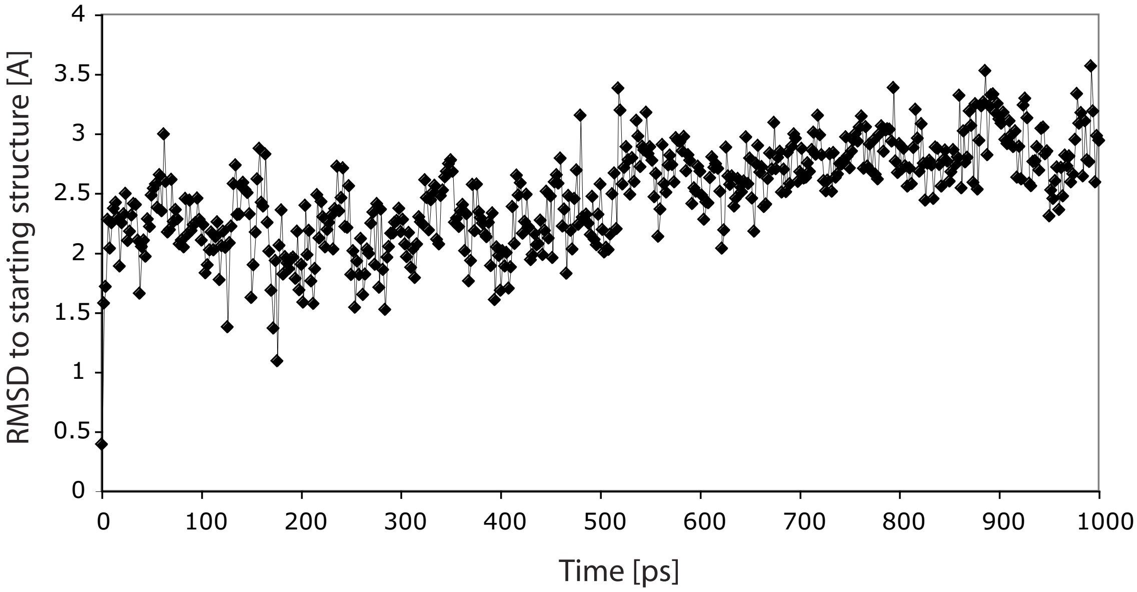


Figure 2.27: RMSD of the gatekeeper residue Met^{120} during an MD simulation of 1 ns with bound ligand B5. The gatekeeper immediately opens, as can be seen by the large RMSD change in the first step. Around 280 ps the gatekeeper closes, but most of the time it is forced to the open position by ligand B5. A general trend is observable, that the gatekeeper keeps on opening. Still, it can be assumed that the gatekeeper would not move away further in a simulation that last longer than 1 ns.

2.4.4 Advanced binding mode III - Docking failure

The third example presents ligand B3 and treats two problems that occur within the docking process of this ligand, when compared with the crystal structure of the complex PKA:B3: i) the ligand has two rotatable bonds which are not correctly rotated (wrong by 180°) and ii) a side chain (Asp^{184}) blocks the correct orientation of the solvent exposed part of the ligand. In the first place, an STD spectrum was recorded, as seen in Fig. 2.28. The binding epitope of the ligand could be clearly identified. Protons H1 and H2 are buried in the protein, while the methyl group is exposed to the solvent. Strong differences in intensities between protons H1 and H3, as well as H4 and H5, indicate that the binding mode is not balanced in the middle of the binding pocket, but protons H1 and H4 are closer to the protein side than the others.

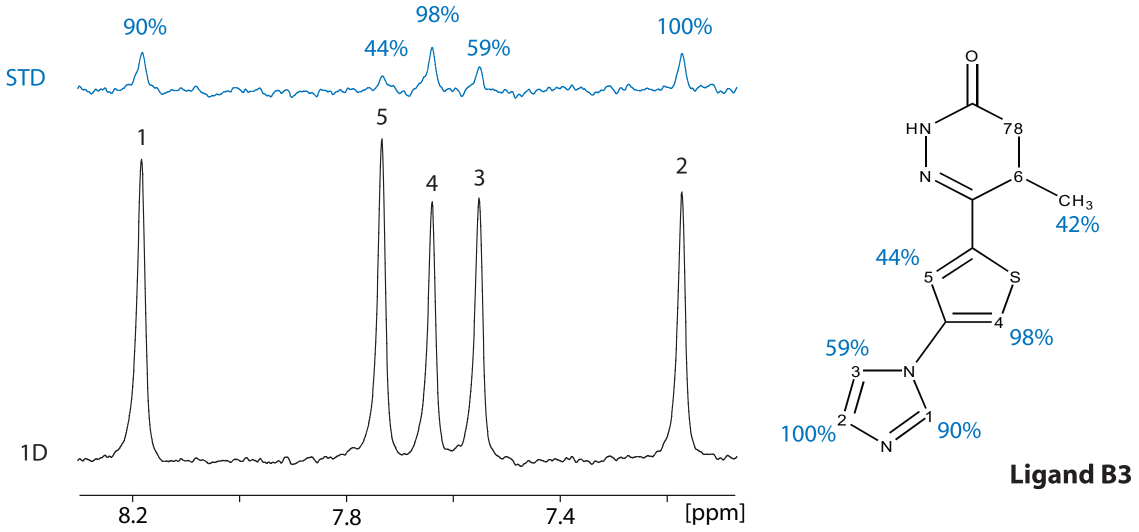


Figure 2.28: Experimental STD data of PKA binding ligand B3. The normal 1D in the presence of 0.045 mM PKA is shown in black and the STD spectrum in blue. Ligand concentration was 1 mM. Saturation was applied for 8 s on 0.5 ppm on a 400 MHz spectrometer (number of scans = 8; temperature = 298 K; time domain = 16384 points). The $V_{Max.STD}$ value is 15.71.

In comparison to the other PKA ligands, B3 has torsion angles from which the bound conformation can be characterized by trNOE. A trNOE spectrum was recorded as seen in Fig. 2.29. The spectrum clearly indicates that the protons i) H1 and H4, ii) H3 and H5, iii) H5 and the methyl group, are close in space in the bound conformation. A NOESY spectrum of ligand B3 in solution without PKA (not shown) confirms that the two bonds freely rotate in the unbound form.

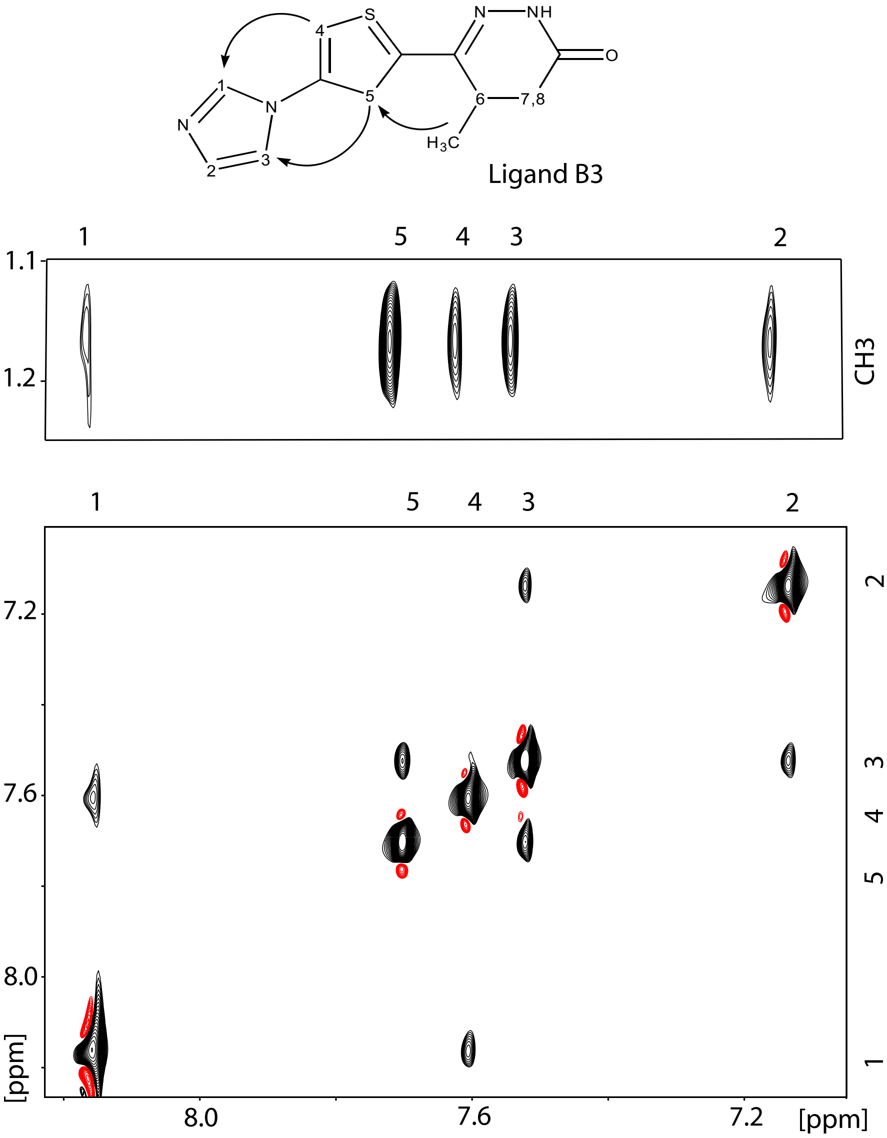


Figure 2.29: NOESY spectrum of ligand B3 (1 mM) in the presence of PKA (0.045 mM). The bound conformation of B3 can be derived from trNOE. The mixing time was 800 ms on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 293 K with 64 scans, 2048 points in F2 and 388 points in F1.

10 docking modes of B3 in PKA were created and visually inspected. Surprisingly, non of the docked structures displayed the bound conformation known from trNOE. This indicates that trNOE have to be applied whenever

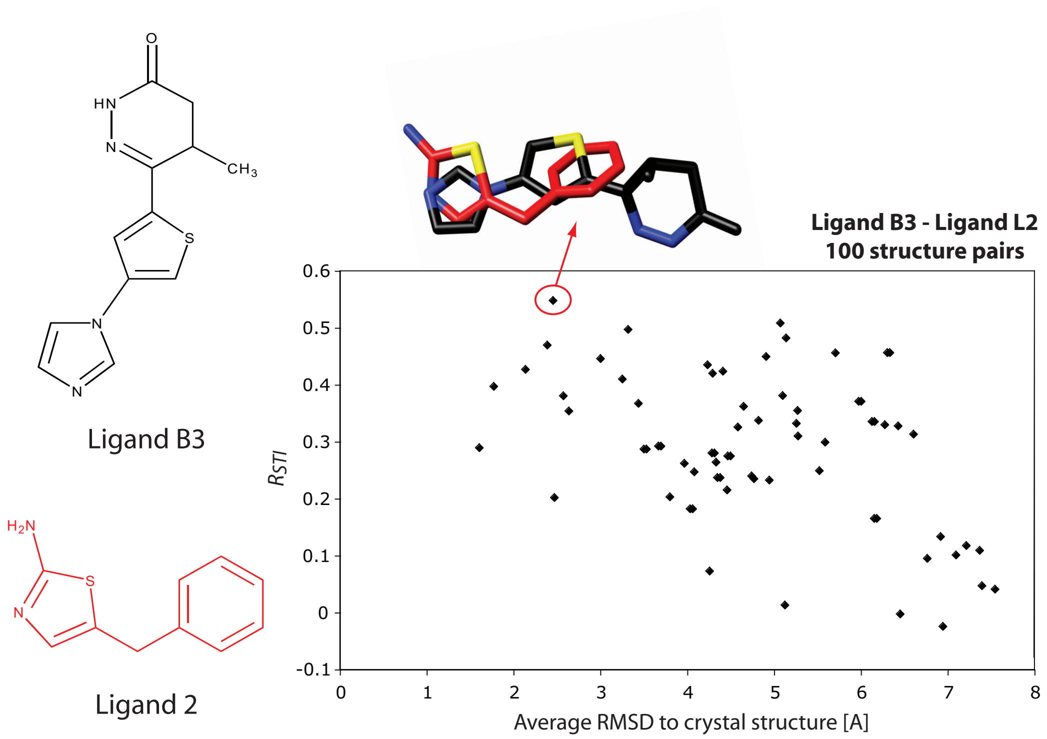


Figure 2.30: The 100 pairwise combinations of ligand B3 and ligand 2 scored against R_{STI}. The identified structure pair is close, but not the closest selection to the crystal structure. It should be noticed that none of the structure pairs correctly reproduced the binding orientation of the crystal structure.

possible. Now two solutions are on hand: i) using a rigid docking with the correct conformation of ligand B3 or ii) restrain the torsion angles during docking. The first approach is not the ideal solution, because the trNOE gives a good estimation, but not an exact description of the torsion angle. And a small distinction from the correct angle might have a negative influence on the docking, as the ligand might not fit with the restrained conformation. The second approach gives the docking program a range in which the distance between two protons has to be placed. The distances between the protons close in space, as interpreted based on the trNOE spectrum shown in Fig. 2.29, were restraint to the following ranges: H1-H4: 1-3 Å, H3-H5: 1-3 Å, and H5-CH3: 1-5 Å. These distances were chosen, as a rotation of the torsion angle around 180° from the expected conformation resulted in a distance larger than 3 Å or 5 Å, respectively. It has to be mentioned here, that PLANTS is the only docking program at the moment that allows such a distance restraint. This approach turned out to be successful, as most of the

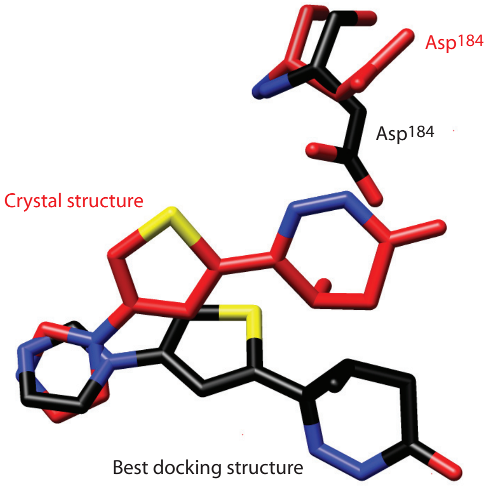


Figure 2.31: Best docking model of ligand B3 by R_{STI} in black and the crystal structure in red. Ligand orientation and torsion angles are correct, but the solvent exposed part of the ligand does not well reproduce the crystal structure. It can be seen that the residue Asp^{184} is rotated in the crystal structure and allows the ligand to adopt a binding mode which is restricted in the rigid protein structure that is used during the docking

docked structures now display the correct bound conformation.

Finally an INPHARMA NOESY was recorded with ligand 2 as the competing ligand. The spectrum is shown in the appendix and displays more than 12 INPHARMA peaks. These were back-calculated for all 10 docking models and the result of R_{STI} is shown in Fig. 2.30. One structure pair is favored over the others and shown in Fig. 2.31, together with the crystal structure of B3. It can be seen that the torsion angles and the orientation of the ligand are correct. Yet, the structure is 3.49 Å RMSD away from the crystal structure. The problem hereby is, that the methyl-group bearing ring of ligand B3 can not rotate into its correct position, because it is hindered by the conformation of side chain Asp^{184}. It can be seen in the crystal structure, that such a conformational change has to take place to allow the ligand to adopt the correct binding mode. A refinement of the docking structure with MD was applied, but the side chain did only slightly rotate. The RMSD to the crystal structure is now 3.02 Å for the best scoring MD structure, and the R_{STI} improved from 0.54 to 0.74.

An alternative approach is presented in the following, which exploits the information of trNOE first and will therefore be termed TSI instead of STI. Following the argument that this ligand displays very informative trNOE data, which were necessary to get the torsion angles right during the docking process, these data should be used as a first selection criterion. As can be seen in Fig. 2.32, the docking structure with index 8 clearly has the best trNOE correlation. It also scores very good for STD. For the ligand 2, it is already known from the previous chapter, that only one binding mode has a convincing STD correlation. When the INPHARMA correlation of structure 8 with the correct binding mode of ligand 2 is investigated, the score is only 0.19. On the other hand, structures with better INPHARMA scores display very bad STD or trNOE scores. When structure 8 was run in an unrestrained MD simulation for 1 ns, the residue Asp^{184} immediately changed its conformation and ligand B3 was allowed to adopt the correct orientation as can be seen in Fig. 2.33. The best scoring binding mode obtained an R_{STI} of 0.85 ($R_{INPHARMA}$: 0.77, R_{STD}: 0.84, R_{trNOE}: 0.93).

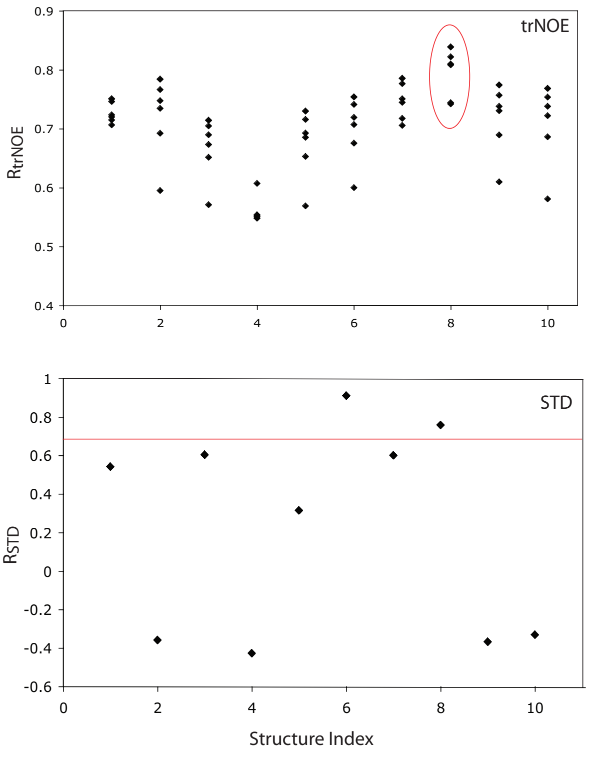


Figure 2.32: Selection of the binding mode starting by trNOE and STD (TSI guided): From the ten docking structures (1-10), structure 8 clearly has the best trNOE correlation (circled red) and the second best STD score as seen below.

This optimization is very impressive when one takes into account that the R_{STI} was only 0.29 in the starting structure. With the original procedure STI applied, a structure of 3.02 Å RMSD to the crystal structure would have

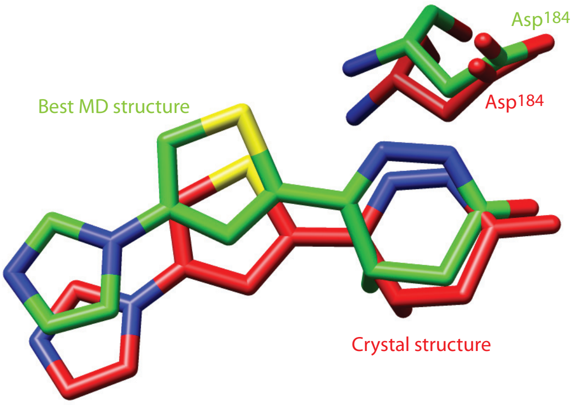


Figure 2.33: Best MD structure of ligand B3 by R_{TSI} in green and the crystal structure in red. The residue Asp^{184} opens during the MD simulations and gives way for the correct binding mode, which score significantly higher than the original docking structure.

been selected. But with the second approach TSI, the selected structure had an RMSD of only 1.12 Å after optimization. It has to be concluded, that the correct binding mode can not always be found in a linear manner, but sometimes different approaches have to be tested and compared. For the presented case, the trNOE data were essential. The docking program would have failed in the first step, and the NMR-based ligand binding mode determination method would have failed in the second step without trNOE data. This issue underlines again the importance of taking all available data into account and scale them in comparison to each other, if necessary. The conclusion is, that a ligand with torsion angles can be guided by trNOE during the docking and the trNOE should be scored first to take only structures into account, that fulfill the trNOE restraints. In the case of a ligand with several torsion angles, it is therefore advisable to apply the methodology STI developed in the previous chapters and the TSI method of the present chapter and then compare the scores. In the present case of B3, the structure obtained by the TSI guided method led to a better score (R_{STI}: 0.85; RMSD: 1.12 Å) than the STI guided method (R_{STI}: 0.74; RMSD: 3.02 Å).

2.5 Discussion & Conclusion

Summary

A methodology was developed to find the correct ligand binding mode out of many diverse docking models by the application of fast obtainable experimental NMR data. A flowchart of the methodology, termed NMR-based ligand binding mode determination, can be seen in Fig. 2.34.

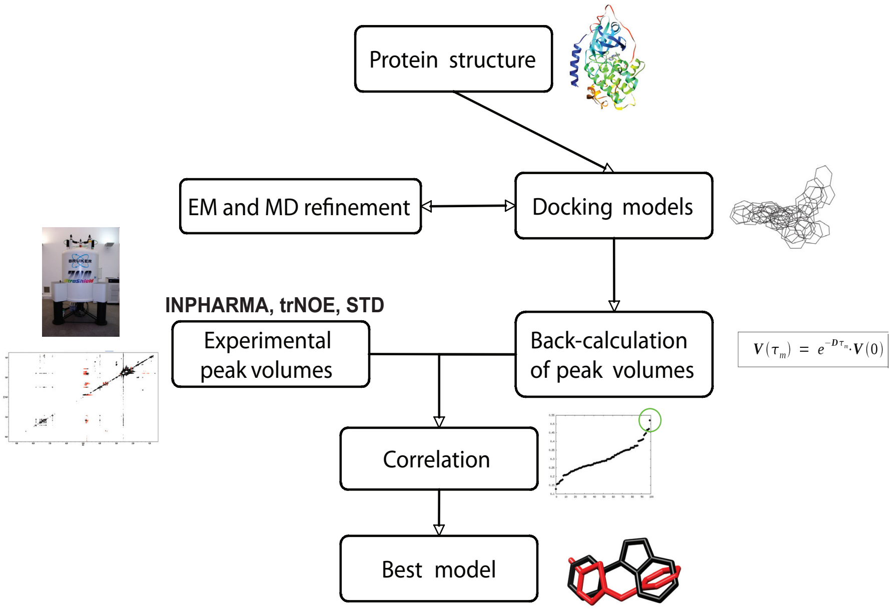

Protein structure
EM and MD refinement
Docking models
INPHARMA, trNOE, STD
Experimental peak volumes
Back-calculation of peak volumes
$V(\tau_m) = e^{-D\tau_m} \cdot V(0)$
Correlation
Best model

Figure 2.34: The flowchart gives an overview of the NMR-based binding mode determination methodology. Starting with a crystal structure or homology model docking models are created. INPHARMA, trNOE and STD peak volumes for every docking model are back-calculated. Then the back-calculated volumes are correlated with the experimental NMR data and the best scoring model is chosen. In the case of a ligand with many torsion angles, it is advisable to do the TSI methodology in parallel.

INPHARMA was shown to be very useful for the relative binding mode determination, yet it lacked efficiency in the determination of the absolute binding mode. TrNOE is very useful for the bound conformation, but performs not very well, when applied to find the absolute binding mode by including the protein environment. Finally STD finds very efficiently the protein buried and solvent exposed part of the ligand, but this information is often not sufficient to find the absolute binding mode. The main point of this thesis is, that the combination of all three methods results in a powerful tool to decipher the correct binding mode. Additionally, the multiplexing of many ligands strongly improves INPHARMA and therefore the overall correlation approach. Even for difficult docking cases, where conformations of amino acids have to change to enable the correct binding mode, the method can be successfully applied. Furthermore the refinement with molecular dynamics simulations was found to be an optimal tool. The amount of structures that have to be calculated for the MD approach is high, therefore the software SpINPHARMA is speed-optimized to fit this purpose. The Pearson correlation coefficient R is used in a first stage, where the correct binding mode has to be found out of diverse docking structures. In a second step a geometry optimization can be done using energy minimization and molecular dynamic simulations, which is then scored together by R and a quality factor Q, which is more sensitive to small changes in the geometry. In the following some issues are critically discussed, which may lead to difficulties or even wrong results. Afterwards an outlook is given.

Consideration of the experimental error

For the INPHARMA peaks, random areas of the background within the NOESY spectrum were integrated and an S/N ratio of 10 was found. Therefore the error is estimated to be between 10 % and 20 %. For the intraligand (trNOE) peaks 5% were estimated and for the diagonal peak volumes 2%. For the STD spectra the same was done and results in a similar error estimation between 10% and 20%. The intraligand peaks have the lowest error, but also contain the least information, especially with rigid small molecules. Therefore it is justified to apply a 1:1:1 weighting, when the three parameters are combined. If the spectra of either the INPHARMA NOESY or STD have a low S/N ratio (e.g. 5), it is probably advisable to use another weighting scheme and to put more weight on the parameter with the best S/N ratio. Yet, this approach was not applied within this thesis.

Hidden binding site for INPHARMA

R_{STD}, R_{trNOE} and $R_{INPHARMA}$ are combined to R_{STI} as described in the equation 2.3 with a simple average. Yet, STD has different requirements to off-rates than INPHARMA. Indeed, STD requires only one off-event during the saturation lasting several seconds, while INPHARMA requiring at least two off-events in the mixing time of a few 100 ms. When STD is recorded, the Kd values of the ligands can be estimated. Given that the Kd value of a ligand is lower than 10 μM and still shows INPHARMA signals, it is advisable to be careful that INPHARMA and STD really emerge from the same binding site. One can imagine that the STD signals show a strong ligand binding event and the INPHARMA signals a weaker binding event to another binding site, for example the protein glycogen phosphorylase has four different ligand binding sites. However, no such effect was observed so far. To make sure that an effect of a strong binding site for STD does not occur, we recommend to determine the Kd values of every ligand. This can be done with STD if a reference ligand exists. A reference ligand has a known Kd value and can be removed from the binding site by titration with another ligand. If no reference ligand is available, other biophysical methods like isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR) could be used. Given that there are ligands with a binding affinity below 10 μM, it can be assumed that they are strong binders and potentially invisible to INPHARMA. One option to handle ligands with binding affinities below 10 μM is to titrate them with weaker binding ligands (typically in the 100-1000 μM range) during an STD experiment. When the signal intensity of the ligand goes down, it binds the same binding pocket as the weaker binder and the INPHARMA signal can be trusted.

Unspecific binding

It is not uncommon that i) different binding modes of one ligand exist, ii) unspecific binding of the ligand to the surface or subpockets of the protein occurs or iii) long channels on the way to the binding pocket are present, where the ligand has to go through. All of these issues contribute to the NMR signal, additionally to the time that the ligand resides at the principal binding site. STD and trNOE signals are especially affected by the contribution of unspecific surface and subpocket binding or long channel pathways. The same holds true for INPHARMA, even tough one could imagine that hereby the unspecific contribution is lower, as both ligands would have to reside at the same unspecific binding site for a certain time and this would then rather be the case for subpockets and channels, than for diffusion controlled surface binding events. Such unspecific binding effects are yet not characterized well enough. An STD study was done on human serum albumin (HSA) with L-tryptophan as the ligand, while the principal binding site was blocked by the very strong binder naxopren. The result was a strong remaining STD signal for the ligand, caused by unspecific binding [Ji et al., 2008]. This approach is very interesting and could also be applied to INPHARMA. Therefore the binding pocket of PKA could be blocked by staurosporine, and two other PKA binding ligands could be measured for their unspecific INPHARMA signals. On the other hand, it is not clear if the remaining STD signals are really unspecific binding, or if the strong binder does indeed block all binding events to the specific binding pocket.

Aggregation of the protein

If one takes a closer look on the molecular weights of the systems used for INPHARMA so far: i) PKA (40 kDa plus dimer- and tetramerization), GPR40 (30 kDa plus membrane fragment), tubulin (50 kDa plus aggregate formation) and iv) glycogen phosphorylase (100 kDa plus dimerization), it is clear that they are very large molecules with a very efficient magnetization transfer. In the case of PKA, it was observed that the signal-to-noise ratio strongly increases, if no anti-aggregation agent is added to the sample. When the protein aggregates, the complexes become very huge and the magnetization transfer most efficient. The quality of the correlation between experimental data and back-calculated data will also strongly increase (Luca Codutti and Steffen Grimm, oral communication). Nevertheless, in the presented study, there was always TCEP as an anti-aggregation agent added. The logic behind this is, that the sample is stable for weeks instead of days and that possible changes in the binding pocket, that could putatively occur during the aggregation process, can be neglected. Also the aim of this study is the monomeric PKA and not an aggregated state.

Outlook

Following studies should be done on soluble proteins like other kinases or phosphatases, where the crystal structure is known, to undertake a proof-of-concept study and further develop the methodology and the MD refinement strategy. The methodology could also be used on whole cells to better simulate the physiological environment. It is also imaginable to use even more NMR parameters in a more elaborate strategy, namely chemical shift changes upon ligand binding. These could be back-calculated in the case of very strong binders and in the case of weak binders, it would be possible to back-calculate the chemical shifts of the protein signals, that change upon ligand binding in an HSQC spectrum. Such an approach is currently in the focus of a group in Lyon. But these approaches would then require labeling of the ligand in the strong binding scenario and labeling of the protein in the weak binding scenario, which is time consuming and not always possible. Still, these approaches are much faster than the classical NMR approach of a full assignment of the protein signals and the following structure calculation of the protein-ligand complex.

Chapter 3

G-protein coupled receptor 40

3.1 Introduction

3.1.1 G-protein coupled receptors

G-protein coupled receptors (GPCRs) are the largest family of membrane proteins. GPCRs play the most important role in hormone and neurotransmitter response, followed by ion channels. Therefore GPCRs are responsible for signal transduction inside the whole organism, control of nerves, muscles and senses like vision, olfaction and taste. Within the neurotransmitter response for example, up and down regulation of body functions is controlled by i) the adrenergic receptor controlled (via adrenaline) sympathtic and ii) the muscarinic receptor controlled (via acetylcholine) parasympathic system. Both receptor families comprise different GPCR subtypes.

The human genome contains around 800 GPCRs which can be further divided in five classes, with the rhodopsin-like family being by far the largest with 670 members. All GPCRs share the common motif of seven α-helices that wind through the membrane and face the extracellular space with a ligand binding site. The ligand binding site is buried deep in the GPCR's hydrophobic pocket and the intracellular facing site of the GPCR binds a specific heterotrimeric G-protein. Generally the mechanisms of a GPCR can be demonstrated with the β_2-adrenergic receptor: adrenaline (or noradrenaline) binds to the GPCR ligand binding site and activates the release of Gα-protein. This stimulates adenylate cyclase, turning then AMP into cAMP. Increase of cAMP will in turn activate cAMP-dependent protein kinase A (PKA) to phosphorylate proteins involved in cell-specific modes of action, e.g. muscle-cell contraction [Rosenbaum et al., 2009]. Yet the signaling behavior of GPCRs can be much more complex, as seen in Fig. 3.1, where the β_2-adrenergic receptor can also trigger another pathway, when the

G-protein binding site is phosphorylated. A molecule which binds and activates the GPCR is termed agonist, while a molecule that binds, but does not activate the GPCR is called antagonist. It is observed that GPCRs tend to have certain noise levels, where they are signaling without being activated. This behavior can be turned off by an inverse agonist, a molecule that binds and completely inhibits signaling.

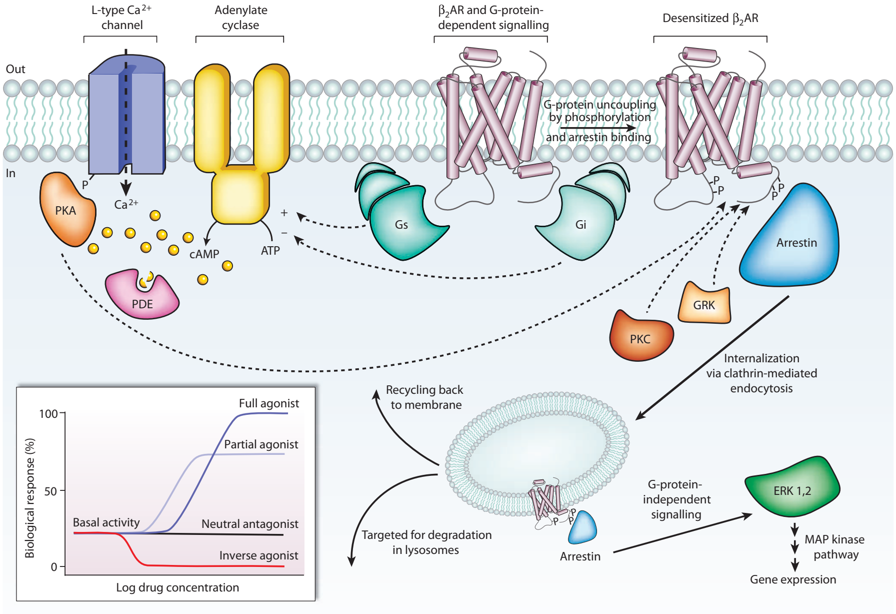

L-type Ca2+ channel
Adenylate cyclase
β2AR and G-protein-dependent signalling
Desensitized β2AR
Out
In
G-protein uncoupling by phosphorylation and arrestin binding
P
PKA
Ca2+
cAMP
ATP
+
–
Gs
Gi
P
P
P
P
Arrestin
PDE
GRK
PKC
Internalization via clathrin-mediated endocytosis
Recycling back to membrane
Full agonist
Partial agonist
100
50
0
Biological response (%)
Basal activity
Neutral antagonist
Inverse agonist
Log drug concentration
G-protein-independent signalling
ERK 1,2
MAP kinase pathway
Gene expression
Targeted for degradation in lysosomes
Arrestin
P
P

Figure 3.1: Schematic presentation of the complex signaling pathway of the β_2-adrenergic receptor GPCR [Rosenbaum et al., 2009].

Groundbreaking work in the field of GPCR research was done by Robert Lefkowitz and Brian Kobilka and was rewarded with the Nobel price in chemistry 2012. The first being responsible for major achievements in the understanding of the signaling pathways and the latter for the structure determination, especially of the β_2-adrenergic receptor. Detailed understanding of the GPCR signaling mechanism on an atomic level was obtained with the crystal structure of rhodopsin [Palczewski et al., 2000]. Retinal is bound to opsin and the absorption of light changes its configuration, activating a signal pathway that finally leads to vision. It could be understood that the configuration change activates a tryptophan conformation change, the so called rotamer toggle switch, leading to a global conformational change of the GPCR that results in the release of G-protein [Hofmann et al., 2009].

The crystal structure of the β_2-adrenergic receptor [Rasmussen et al., 2007] with the bound inverse agonist carazolol was of high interest, as the substrate of this receptor are neurotransmitters and therefore more relevant for drug discovery than rhodopsin. The conclusion is, that the binding of an agonist leads to the activation of the rotamer toggle switch as in rhodopsin and the separation of the transmembrane helices TM3 and TM6. The ionic lock between ECL2 and ECl3 will break during the movement of TM6 [Audet and Bouvier, 2008]. It seems that the ionic lock can break even in the absence of an agonist. This does not happen when the inverse agonist binds, so it is thought that the inverse agonist stabilizes TM6 and therefore the ionic lock. G-protein coupled receptors are membrane proteins and very difficult to crystallize due to the lack of exposed polar surfaces, a generally low expression level and poor thermal stability. In pharmaceutical industry they are often represented by homology models [Hillisch et al., 2004]. The seven transmembrane helices are a conserved motif and can easily be modeled, even though the sequence similarity between different GPCRs is often only around 10%. More problematic for example is the extracellular loop 2 (ECL2), which can occur in very different conformations (Fig. 3.2) and works like a lid over the binding pocket. It was even proposed that in some cases the results of GPCR homology modeling would be better, if the ECL2 loop is excluded from the structure [Costanzi, 2008]. For a long time the only available crystal structure and therefore a template for homology modeling was rhodopsin.

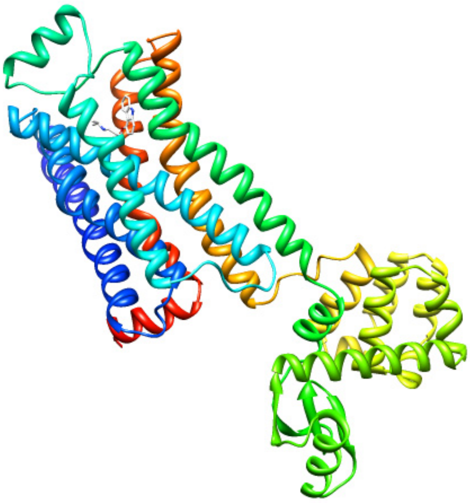


Figure 3.2: Homology model of the β-adrenergic receptor with bound ligand and coupled G-protein. It is noteworthy that the extracellular loop 2 has the form of a helix.

Nowadays the technique of GPCR crystallization improved substantially and more crystal structures become available, e.g. recently the M2 muscarinic receptor [Haga et al., 2012] bound to a ligand. Even the ligand free form of a GPCR was now found for the GPCR opsin, without bound retinal [Park et al., 2008]. Such models can be a very useful starting point for drug discovery with the virtual screening approach. Yet, in a virtual screening approach for the β_2-adrenergic receptor several strong binders were identified, but most of them turned out to be inverse agonists [Kolb et al., 2009]. This indicates that stabilizers of the inactive, ligand bound form of the GPCR were identified. The conclusion would be, that it is also very useful to obtain the active, ligand free form of a GPCR, in order to develop agonists instead of inverse agonists or antagonists. It depends on the desired result, if a GPCR should be up- or down-regulated, i.e. if an agonist or an antagonist has to be found. Next to the described orthosteric ligand that binds at the binding pocket, there are allosteric ligands that do not bind at the binding site, but still trigger a reaction on the GPCR. Recently it was possible to design a dualsteric ligand, that combined both properties [Mohr et al., 2013]. The application of other methods such as NMR spectroscopy to understand the complex dynamics of the highly flexible GPCR molecules and the binding of ligands to the same are very promising. Hereby, not the whole structure is in the focus, but smaller areas of specific interest are investigated. It was demonstrated that the dynamics of the ionic lock of the β_2-adrenergic receptor could be accurately described for an agonist, an antagonist and an inverse agonist using NMR spectroscopy [Bokoch et al., 2010] in conjunction with solved X-ray structures. The action of these different ligands on the ionic lock and therefore on the extracellular surface of the GPCR could not be seen in the corresponding crystal structure of the GPCR. After successful labeling, the dynamics of the rhodopsin C-terminus were investigated [Werner et al., 2008]. Furthermore it was possible with NMR to reveal the bound conformation of the peptides PACAP [Inooka et al., 2001] and neurotensin [Luca et al., 2003] in its specific GPCR. Drug design for GPCRs is a challenge that takes into account all available information of the few existing crystal structures, homology models, ligand-based pharmacophore search and libraries with privileged GPCR binding building blocks. Nearly 50% of recently launched drugs target GPCRs, making them the most important drug target and also the most beneficial one, with around 30 billion $ sales in 2001 [Klabunde and Hessler, 2002].

3.1.2 GPR40: Homology models and proposed ligand binding modes

The G-protein coupled receptor 40 (GPR40), also known as the free fatty acid receptor 1 (FFAR1), is an interesting drug target in diabetes research. For a long time it was an orphan receptor, which natural binder and function was unknown. This suddenly changed in 2003, when it was found that long-chain fatty acids, especially linoleic acid, bind to GPR40 and activate a signal cascade, taking part in the regulating of insulin secretion [Itoh et al., 2003]. GPR40 is preferentially expressed by the gene FFAR1 in the β-cells in the pancreas. Activation of GPR40 by fatty acids induces a signaling pathway which leads to a higher Ca^{2+} concentration in the cells. This ultimately stimulates the release of insulin from the β-cells, a mechanism which does not work properly in type 2 diabetes patients.

Today several pharmaceutical companies are developing drugs as GPR40 agonists. At the time this thesis is written, Takeda has a compound [Negoro et al., 2010] in phase III of clinical trials and other companies have presented interesting candidates, e.g. Amgen [Lin et al., 2011] or Merck [Zhou et al., 2010]. Yet, these compounds were identified during high-throughput screening and chemically optimized on a trial-and-error base, as there is no crystal structure of GPR40 available.

A homology model for GPR40 based on rhodopsin (Fig. 3.3) was made by the Costanzi group [Tikhonova et al., 2007]. The group is very experienced in the modeling of G-protein coupled receptors. In the GPCR dock contest 2008, which takes place every two years, the group submitted a homology model of the adenosine receptor A_{2A} with the bound ligand ZM241385 and reached the closest RMSD to the newly obtained crystal structure [Michino et al., 2009].

The binding of ligand GW9508 to GPR40 was investigated in detail by docking to the homology model and experimental SAR data by amino acid exchange [Sum et al., 2007]. In Fig. 3.3 the SAR supported binding mode of GW9508 to GPR40 is seen in detail. The carboxylic moiety is anchored between Arg^{183}, Arg^{258} and Asn^{244}. The hydrophobic tail is placed in the hydrophobic pocket, the linker stabilized by His^{86} and the last aromatic ring of GW9508 is stabilized by an NH-π interaction with His^{137}.

This homology model turned out to be very useful and was applied as a basis for docking studies [Lu et al., 2010], QM/MM studies [Lu et al., 2011] and to successful virtual screening [Tikhonova et al., 2008] for new binders.

Another homology model was made based on the β-adrenergic receptor, that displayed similar binding modes of ligand GW9508 and linoleic acid as for the previous homology model. Yet, the new model seemed to describe

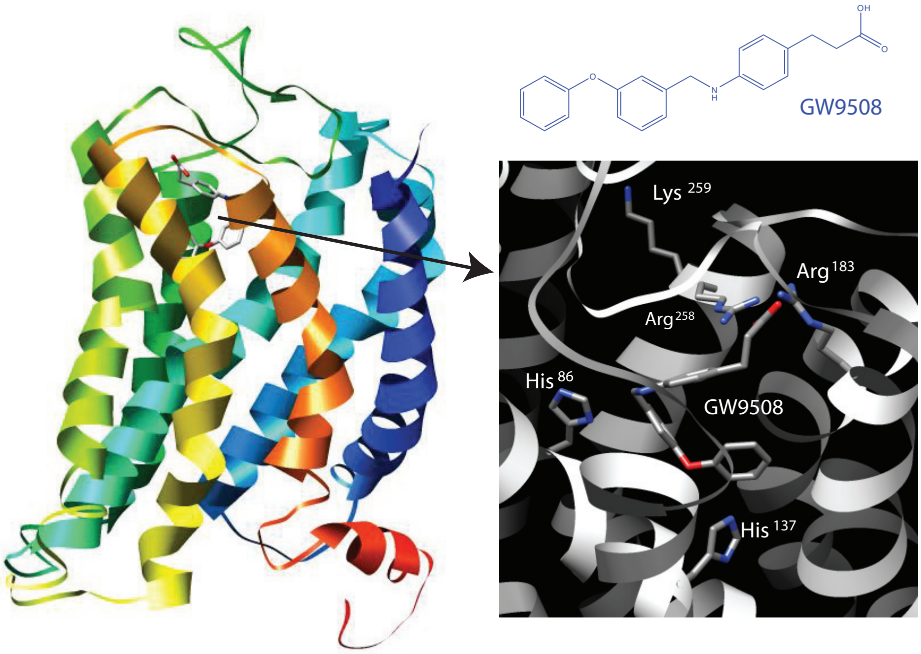


Figure 3.3: Homology model of GPR40 with bound ligand GW9508 in a detailed view. The binding mode reproduced best the site-directed mutagenesis data, a hydrogen bonding between the carboxylic group and the arginine residues, as well as a stabilizing contact to His^{137} [Tikhonova et al., 2007, Sum et al., 2009].

the structure of the ECL2 loop more accurately and a mechanism for the activation of GPR40 could be proposed [Sum et al., 2009]. An ionic lock between two arginine (Arg^{183} and Arg^{258}) and two glutamic acid (Glu^{145} and Glu^{172}) residues is proposed to be released by linoleic acid binding to the arginines with its carboxylic group, while the hydrophobic tail is stabilized in the hydrophobic pocket. This mechanism explains well the SAR data, that the two arginines are crucial for binding and every ligand that binds to GPR40 has a carboxylic moiety or at least a comparable chemical structure.

The most interesting GPR40 binding ligand at the moment is TAK-875 from Takeda, which reached phase III of clinical trials and binds with an EC_{50} of 50 nM. A binding mode was proposed by Takeda [Negoro et al., 2010] by docking of TAK-875 to an internal homology model. The anchoring of the carboxylic moiety of TAK-875 to the arginines is in common with the GW9508 model, but the rest of the ligand does not enter the hydrophobic pocket and the sulfate group forms a hydrogen bond to Lys^{259}. There is no amino acid exchange data to prove this interaction, and it has to be said that the homology model was done without the ECL2 loop, resulting in an

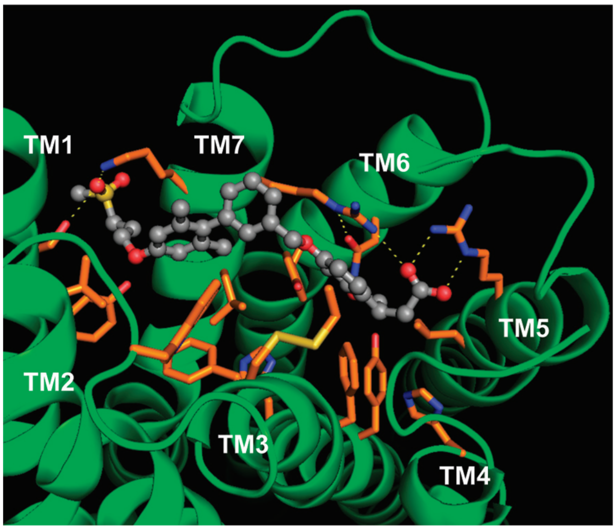

TM1
TM7
TM6
TM5
TM2
TM3
TM4

Figure 3.4: Homology model of GPR40 with ligand TAK-875 docked to the binding site. Proposed is a hydrogen bonding between the carboxylic group of TAK-875 to the arginine residues and between the sulfate group and Lys^{259} [Negoro et al., 2010].

unrealistic representation of GPR40.

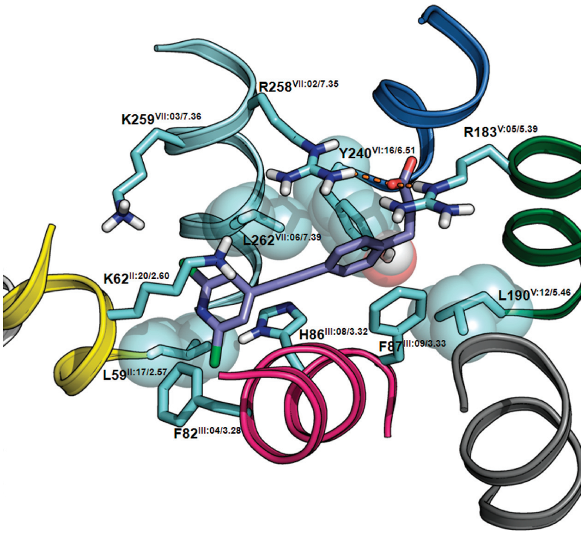

R258VII:02/7.35
K259VII:03/7.36
R183V:05/5.39
Y240VI:16/6.51
L262VII:06/7.39
K62II:20/2.60
L190V:12/5.46
H86III:08/3.32
F87III:09/3.33
L59II:17/2.57
F82III:04/3.28

Figure 3.5: Homology model of GPR40 with compound 37 manually docked to the binding site. Proposed is a hydrogen bond to the arginine residues and a stabilizing contact between the nitrogen of ligand 37 and His^{86} [Christiansen et al., 2011].

A third binding mode was proposed for compound 37 by [Christiansen et al., 2011] to an internal homology model (Fig. 3.5). This model again shares the anchoring of the head group to the arginines, but makes the stabilizing contact to His^{86}. The ligand was docked manually, as there is no homology model available with a binding site that allows a docking pose to this amino acid. This assumption is realistic, as the ligand is very rigid, shorter than linoleic acid, and might not be able to form the stabilizing NH-π interaction with His^{137}.

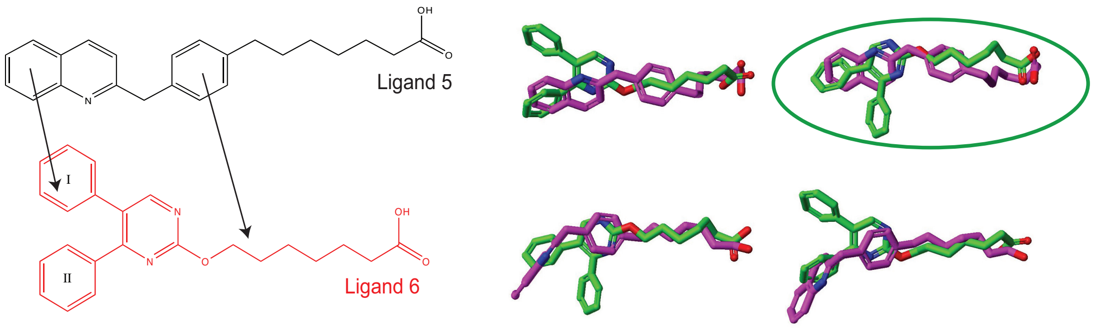


Figure 3.6: Alignment of ligands 5 and 6 based on INPHARMA signals (simplified for clarity) observed between the indicated regions. From the four proposals by Catalyst, only one fits the experimental data.

In 2010, it was shown that a simple GPR40 sample in solution can be produced by the lysation of cells that over-express GPR40. The buffer solution contains membranes including GPR40 and is amenable for NMR studies. It could be shown that EC$_{50}$ values could be accurately reproduced with STD experiments and that INPHARMA peaks between two competing ligands are observable. Using the NMR data it was possible to align the ligands in a realistic manner and choose between different proposals, obtained by the alignment program Catalyst (Fig 3.6) [Bartoschek et al., 2010].

3.2 Materials & Methods

3.2.1 Sample preparation

GPR40 was overexpressed by the pharmaceutical company Sanofi as described [Bartoschek et al., 2010]. Before sample preparation the membranes were thawed and dialyzed with D$_2$O buffer containing DSS as an internal reference, deuterated Tris buffer and deuterated EDTA. Sample preparation was done in Shigemi tubes with a total volume of 250 μM.

3.2.2 NMR spectroscopy

INPHARMA spectra were recorded using the standard NOESY pulse program noesyphpr. Ligand combinations 5&6 and 5&7 were recorded on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head (Bruker, Karlsruhe). STD spectra were recorded with a standard pulse sequence (std-diffesgp.2) with 16 scans on a 400 MHz spectrometer Saturation was achieved by a train of shaped 90° pulses of 50 ms length. A number of 160 selective pulses was applied, leading to a total length of saturation of 8 s. The on-resonance irradiation was performed at -2 ppm and off-resonance irradiation was set to $+30$ ppm.

3.2.3 Docking and alignment

10 binding modes per ligand were created in the homology model of GPR40 [Tikhonova et al., 2007] using the docking software PLANTS [Korb et al., 2006]. The ligands were kept flexible and the protein rigid. Three-dimensional structures of the ligands were created using CORINA [Sadowski and Gasteiger, 1994].

3.2.4 Energy minimization and molecular dynamic simulations

Energy minimization and molecular dynamic simulations were done as implemented in Gromacs [Hess et al., 2008], using the amber99sb force field [Hornak et al., 2006] for the protein and the General Amber Force Field [Wang et al., 2004] for the ligand. EM was done with conjugate gradient integrator in 1000 steps. MD was done using steepest descent integrator with a stepsize of 2 fs. Simulations were done for 1 ns for every docking mode respectively in implicit solvent, using the Generalized Born model augmented with the hydrophobic solvent accessible Surface Area (GBSA). From every trajectory 100 snapshots were saved and the RMSD to the starting structure was calculated.

3.2.5 Back-calculation of peak volumes

All INPHARMA, trNOE and STD peak volumes were back-calculated with the software SpINPHARMA, using the complete-relaxation matrix approach [London, 1999,Orts et al., 2009]. Methyl-groups were described using a 3-site jump model [Tropp, 1980]. STD saturation was applied to all methyl groups. Protons within a distance of 8 Å from any ligand proton were considered.

The correlation time was estimated to be 100 ns. Kd values were determined by STD competition experiments (L5: 340 nM, L6: 160 nM). On-rates (k_{on}) were assumed to be in the diffusion limit 10^8 $M^{-1}s^{-1}$.

3.3 Results

3.3.1 Docking, STD and INPHARMA Ligands 5 and 6

The ligands 5 and 6 (Fig. 3.7) were identified from high-throughput screening by Sanofi and an alignment based on experimental NMR data, i.e. the relative binding mode, is known. To get the absolute binding mode, docking models were created in the public, SAR data supported homology model of GPR40 [Tikhonova et al., 2007]. 10 binding modes per ligand were obtained and manually investigated. As seen in Figs. 3.8 and 3.9, these binding modes are diverse and display different orientations in the binding pocket. The docking modes obtain good ChemPLP scores, indicating that they fit well to the binding site. In Fig. 3.10, the docking modes of ligand 5 are shown together with the SAR-supported ligand GW9508, indicating that the ligand fits well the binding site and the best scoring docking mode is very close to the best known solution.

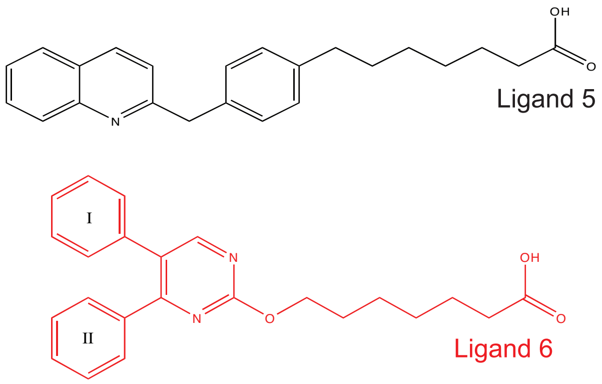
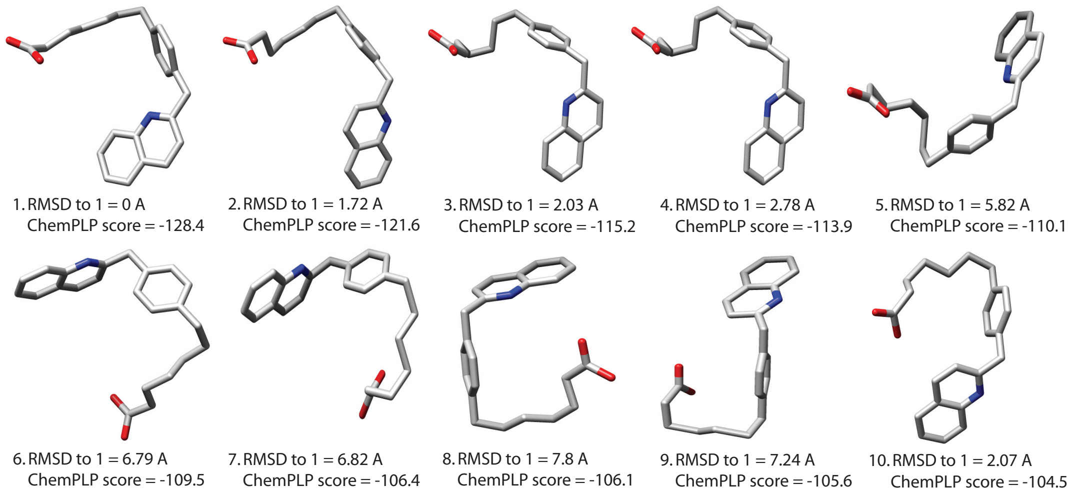


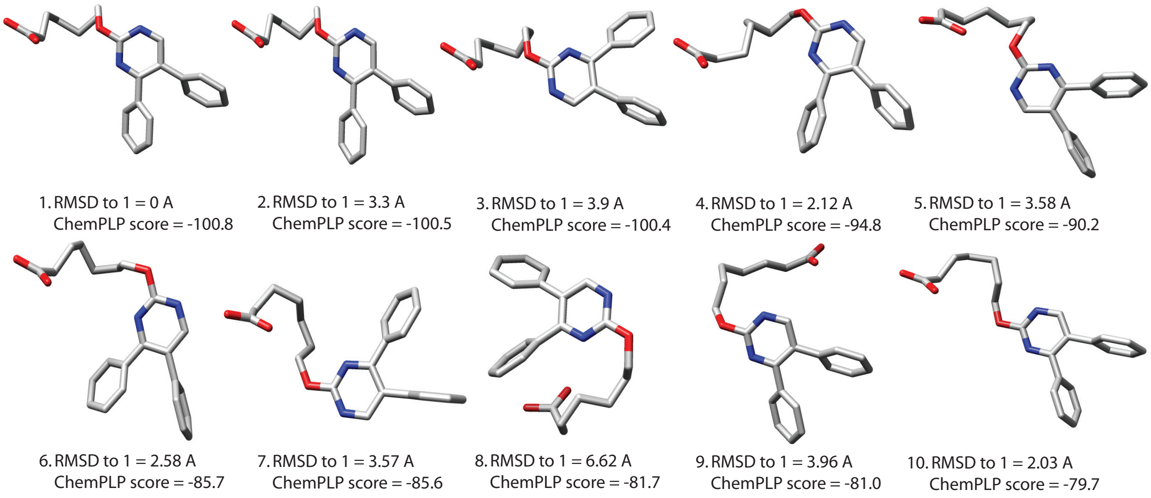
Figure 3.7: Ligands 5 ($EC_{50} = 340$ nM) and 6 ($EC_{50} = 280$ nM) identified during high-throughput screening by Sanofi.

Extensive INPHARMA NOESY measurements were done with the ligand pair 5 and 6 [Bartoschek et al., 2010] and used as a basis for the experimental data. For every docking model obtained the STD peak volumes were back-calculated, as well as for every pair of docking models the trNOE and INPHARMA peak volumes. The correlation coefficients R between experimental and back-calculated peak volumes of STD, trNOE and INPHARMA were combined to R_{STI}. The best scoring model is shown in Fig. 3.11 with important amino acids and hydrogen bonds indicated. The carboxylic group of both ligands is anchored between the arginines Arg^{183}, Arg^{258} and Asn^{244} and Ser^{247}. The rest of the ligands is placed in the hydrophobic pocket deeper buried in the GPR40 structure. In Fig. 3.12 all 100 combinations of the 10



1. RMSD to 1 = 0 A ChemPLP score = -128.4
2. RMSD to 1 = 1.72 A ChemPLP score = -121.6
3. RMSD to 1 = 2.03 A ChemPLP score = -115.2
4. RMSD to 1 = 2.78 A ChemPLP score = -113.9
5. RMSD to 1 = 5.82 A ChemPLP score = -110.1
6. RMSD to 1 = 6.79 A ChemPLP score = -109.5
7. RMSD to 1 = 6.82 A ChemPLP score = -106.4
8. RMSD to 1 = 7.8 A ChemPLP score = -106.1
9. RMSD to 1 = 7.24 A ChemPLP score = -105.6
10. RMSD to 1 = 2.07 A ChemPLP score = -104.5

Figure 3.8: Docking modes of ligand 5 in the homology model of GPR40. Indicated is the RMSD to the first docking model and the ranking by the docking scoring function ChemPLP.



1. RMSD to 1 = 0 A ChemPLP score = -100.8
2. RMSD to 1 = 3.3 A ChemPLP score = -100.5
3. RMSD to 1 = 3.9 A ChemPLP score = -100.4
4. RMSD to 1 = 2.12 A ChemPLP score = -94.8
5. RMSD to 1 = 3.58 A ChemPLP score = -90.2
6. RMSD to 1 = 2.58 A ChemPLP score = -85.7
7. RMSD to 1 = 3.57 A ChemPLP score = -85.6
8. RMSD to 1 = 6.62 A ChemPLP score = -81.7
9. RMSD to 1 = 3.96 A ChemPLP score = -81.0
10. RMSD to 1 = 2.03 A ChemPLP score = -79.7

Figure 3.9: Docking modes of ligand 6 in the homology model of GPR40. Indicated is the RMSD to the first docking model and the ranking by the docking scoring function ChemPLP.

docking models per ligand are shown, sorted by their R_{STI} values. On the right side of the diagram, the RMSD to the best scoring structure is indicated. For the calculation of the RMSD only the carboxylic moieties, the chinoline moiety of ligand 5 and ring I of ligand 6 were used, as the connector part of the ligand is very flexible and not the crucial part for the binding. A general trend can be observed: the closer the RMSD is to the best scoring

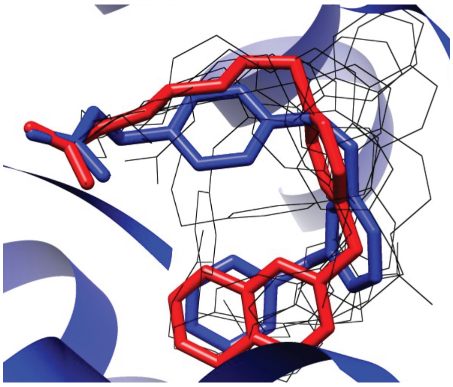


Figure 3.10: 10 docking modes of ligand 5 in the homology model of GPR40. Indicated is the best scoring structure in red and the SAR-supported ligand GW9508 [Tikhonova et al., 2007] in blue.

structure, the better is its score.

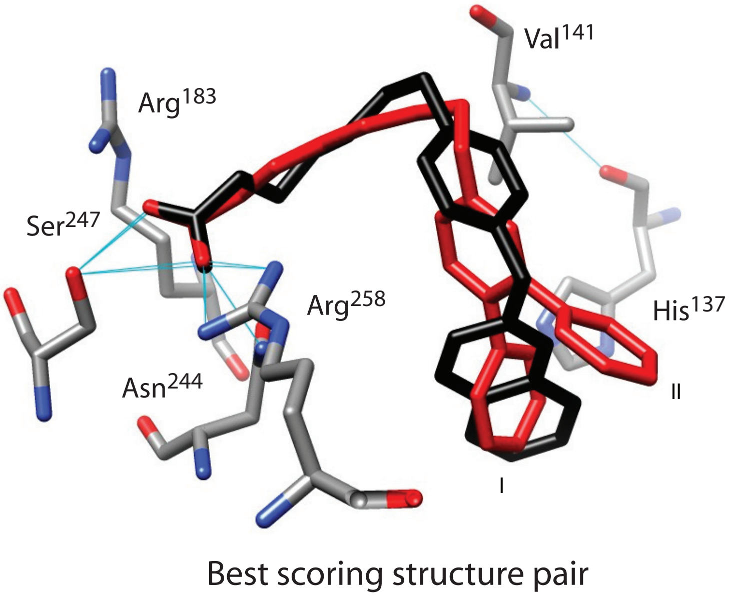

Val141
Arg183
Ser247
Arg258
His137
Asn244
II
I
Best scoring structure pair

Figure 3.11: Detail of the binding mode of the two ligands (5 black, 6 red) in the pocket. The best scoring pair is shown schematically superimposed with the important amino acids and hydrogen bonds at the binding site indicated.

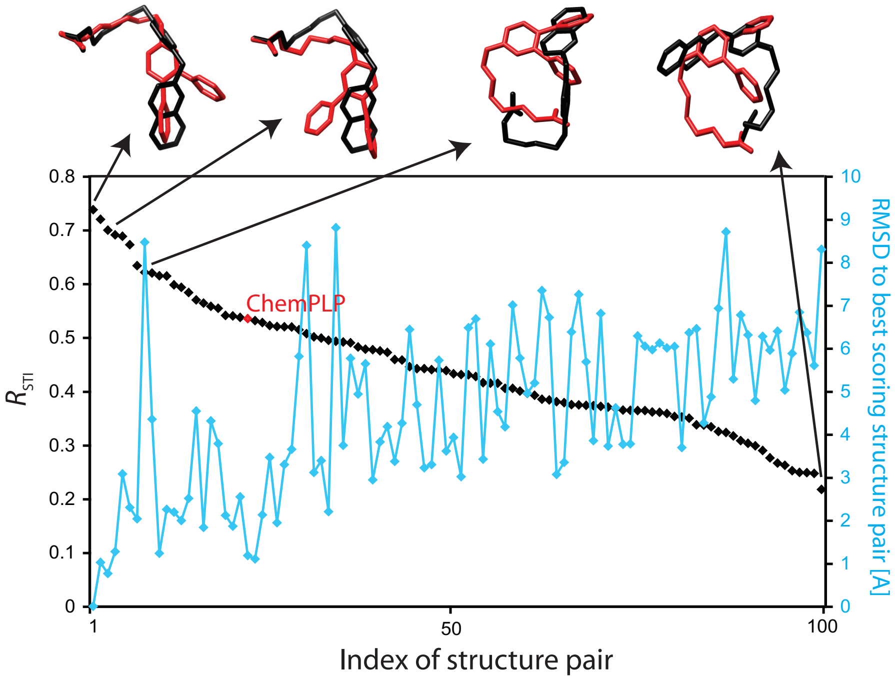


Figure 3.12: R_{STI} scores for the 100 pairwise combinations, including the RMSDs to the best scoring structure pair. Structure pairs (on top) with the best scores display the same motif of an overlap of the ligand 6 ring II and the chinoline moiety of ligand 5. Indicated in red the best scoring pair using the scoring function ChemPLP.

3.3.2 Modeling and molecular dynamics simulations

Molecular dynamics simulations of the ligands 5 and 6 within the binding pocket of the GPR40 homology model were performed to obtain a better understanding of the ligand dynamics in the binding site, as well as to refine the obtained binding modes from INPHARMA. An MD simulation was done for 1 ns in implicit solvent for every of the 10 docking models per ligand, resulting in 20 MD trajectories. In Figs. 3.13 and 3.14 the RMSD of the final structure to the starting structure (obtained from docking) of each trajectory is shown. It was visually inspected that around 5 Å RMSD from the starting structure, the ligand starts to leave the binding pocket with the hydrophilic tail.

It can be seen that some structures are stable at the binding site, while others are moving far off and even leave the binding pocket. Generally it can be said that the ligands are very dynamic at the binding site and also those that remain in a stable binding mode change in RMSD around 3 Å.

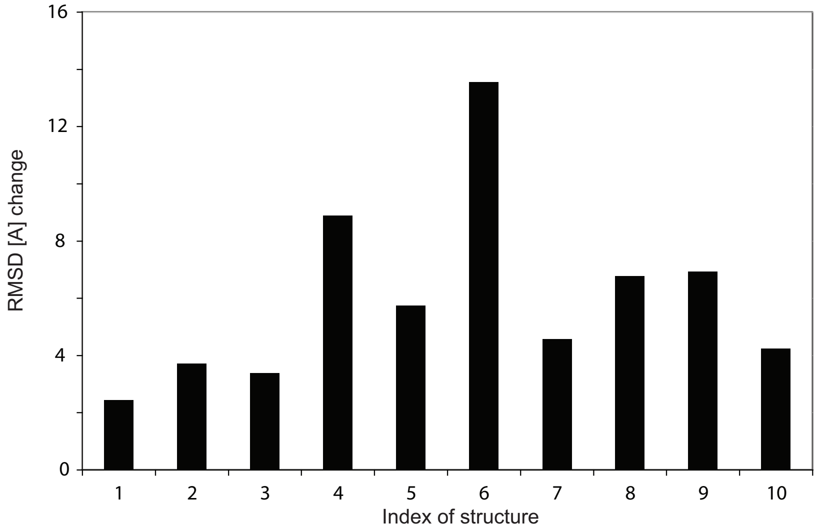


Figure 3.13: RMSD value between the final structure to the initial binding mode during an MD simulation of 1 ns of ligand 5. Not all ligand binding modes stay in a stable position in the binding pocket. The higher the RMSD value, the more unstable is the binding mode.

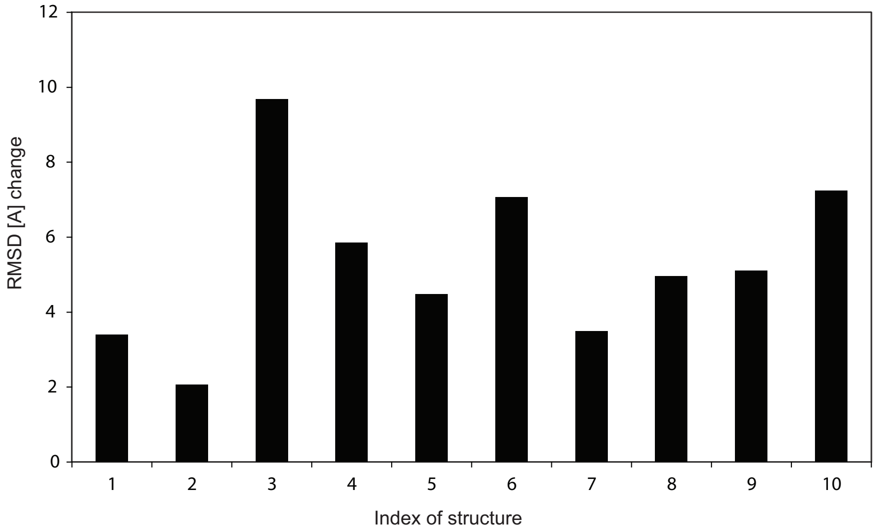


Figure 3.14: RMSD value between the final structure to the initial binding mode during an MD simulation of 1 ns of ligand 5. Not all ligand binding modes stay in a stable position in the binding pocket. The higher the RMSD value, the more unstable is the binding mode.

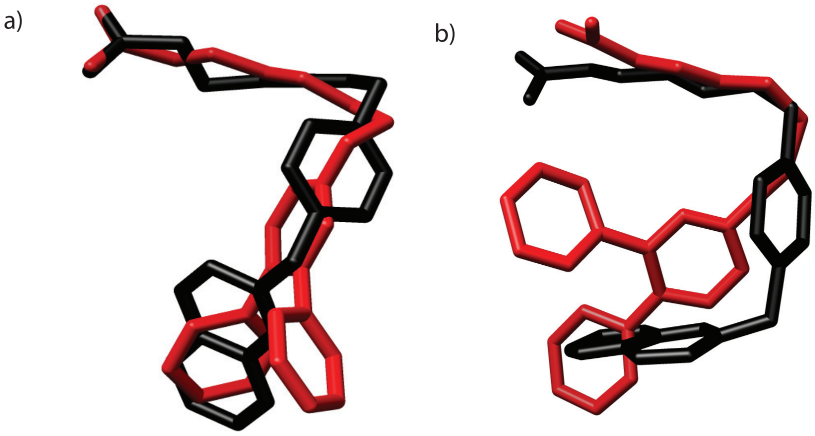
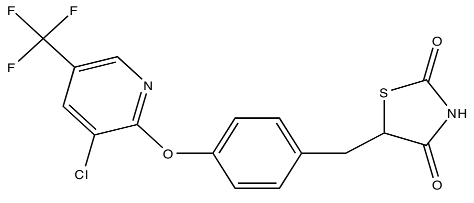


Figure 3.15: The superimposition of the PLANTS structures with the best score (a) and the superimposition of the MD structures with the best score (b) after refinement against R_{STI}. It is noteworthy that ring II of ligand 6 performs a flip during the MD simulation of 1 ns.

3.3.3 STD Ligand 7

Ligand 7 (Fig. 3.16), termed 9-43 in the literature, binds GPR40 with an EC_{50} of 69 nM [Zhou et al., 2010].



Ligand 7

Figure 3.16: Ligand 7, agonist of GPR40 with $EC_{50} = 69$ nM.

Ligand 7 is soluble up to 0.05 mM and displayed a strong STD signal in solution with GPR40 membranes. A binding group epitope could be extracted as seen in Fig. 3.17 and nearly no STD effect was observed in solution with empty membranes. Yet, the binding affinity is too strong to expect INPHARMA peaks.

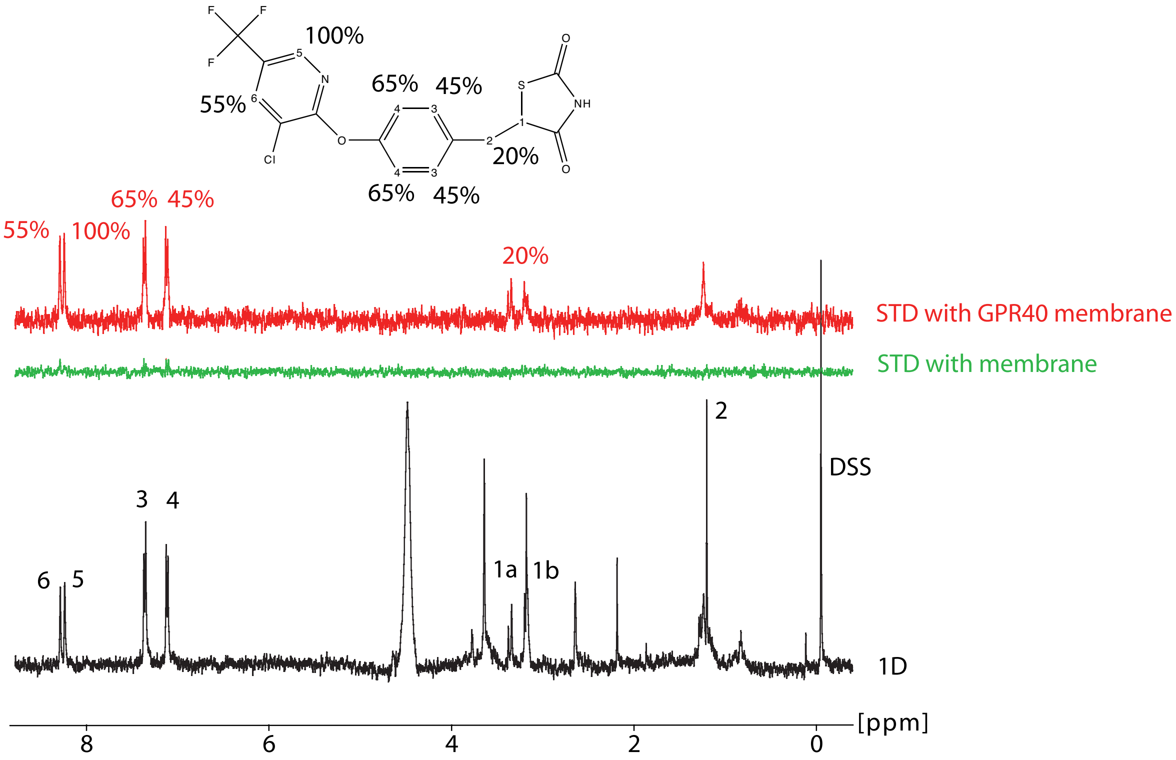


Figure 3.17: STD spectrum and binding epitope of ligand 7, binding to GPR40 membranes. Nearly no binding is observed to membranes without GPR40.

3.4 Discussion & Conclusion

Here it is shown that the absolute binding mode of ligands to the membrane protein and GPCR GPR40 can be determined based on a published homology model [Tikhonova et al., 2007]. Ten docking models for each of the two ligands 5 and 6 in the homology model were created using PLANTS and subsequently energy minimized using the Amber force field. The INPHARMA, trNOE and STD integrals necessary for R_{STI} scoring were back-calculated for every model.

Correlation of these back-calculated values with the experimental data results in a clear proposal for the optimal binding mode. It is also observed that the next R_{STI} scoring structure pairs have a small RMSD. Yet, there are some structure pairs (e.g. the 8^{th} structure), which have a large RMSD but still score very well ($\mathrm{R}_{STI} = 0.62$ compared to 0.75 of the best scoring pair). This structure, however, can be excluded on the basis of mutant analysis data (see below). The best scoring structure in terms of R_{STI} is in good agreement with the previously determined relative binding mode [Bartoschek et al., 2010], which was confirmed by the observation that the presence of the aromatic ring II of ligand 6 does not strongly affect the binding affinity, while the aromatic ring I is crucial for binding. Site directed mutagenesis experiments with the ligand GW9508 [Sum et al., 2007], which is very similar in structure to ligands 5 and 6, supports the concept that the carboxyl moiety is anchored between Arg^{183}, Arg^{258} and Asn^{244}, as the mutation of each of these residues to Ala almost abolishes binding. Additionally, the mutation of His^{137} to Ala lowered the binding affinity, which is compatible with a polarized NH-π-interaction of the 3-phenoxy moiety with His^{137} [Tikhonova et al., 2007]. These data are in excellent agreement with our model. We observe the carboxyl moieties anchored between Arg^{183}, Arg^{258}, Asn^{244} and Ser^{247}, while the chinoline group of ligand 5 as well as the aromatic ring I of ligand 6 are in close contact to His^{137}. This binding mode explains well why the aromatic ring I of ligand 6 is necessary for binding, while the aromatic ring II is not; ring II is in fact too far to contact His^{137}. The 8^{th} structure with a good R_{STI} but large RMSD displays an overlay of the rings as expected, but both ligands are oriented completely differently in the binding site and no contact with the crucial arginines is formed; this binding mode would not explain the mutation data, which indicate that the arginines are essential for ligand binding. A comparison between the GW9508 and the received binding modes of ligand 5 and ligand 6, together with the SAR data indicated can be seen in Fig. 3.18.

To get a better understanding of the obtained binding modes, each was taken as the initial structure for a molecular dynamics simulation of 1 ns.

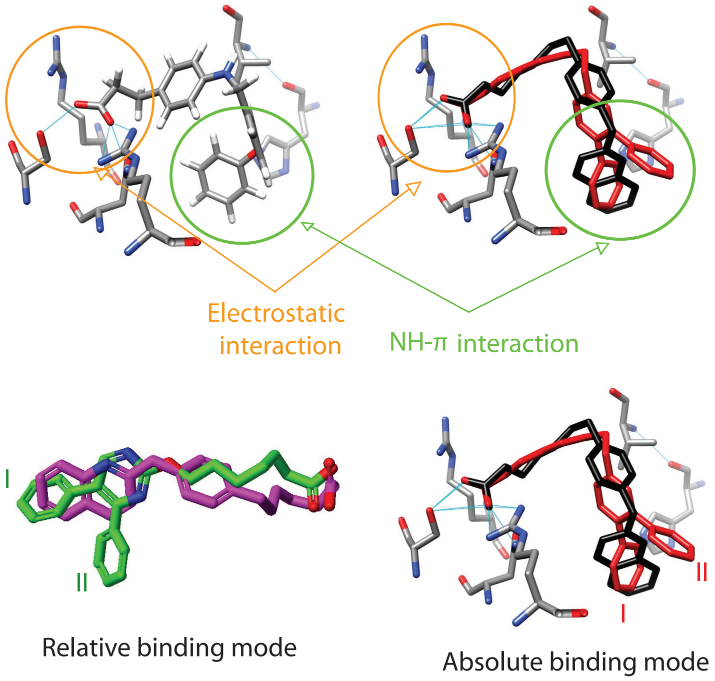


Figure 3.18: Comparison of the binding mode of ligand GW9508, obtained by docking and SAR data [Tikhonova et al., 2007] and the R_{STI} derived binding modes of ligands 5 and 6 superimposed. Both share an anchoring of the carboxylic moiety to charged amino acids and a close contact of an aromatic ring to His^{137}, proposing a NH- π stabilization. Below a comparison of the relative binding mode of the two ligands derived by INPHARMA [Bartoschek et al., 2010] and the absolute binding modes derived from the combined scoring. Both share an overlap of ring I of ligand 6 with the chinoline moiety on ligand 5.

It was shown that the chosen binding mode and those close in RMSD to it, remain at the initial binding side with a fluctuation less than 5 Å, while those with a larger RMSD display a less stable binding mode. Generally it can be said, that the ligand starts to penetrate deeper in the binding pocket and that the ligands behave very flexible with several flips occurring in the hydrophobic tail, the chinoline moiety in ligand 5 and the aromatic rings I and II in ligand 6. These simulations explain well how flexible the ligands bind. The hydrophobic part is able to flip, so it can be assumed that all docking structures close in RMSD to the chosen one, represent a correct conformation observed during the binding event. The new ligand 7 is not a fatty acid derivative as the ligands 5 and 6 and displays therefore a possibly different binding mode. The binding of ligand 7 to GPR40 membranes was confirmed and investigated with STD. The carboxylic group of ligand 7 seems to bind to the same site as the thiazolidinedione head group of ligands 5 and 6. The linker part also aligns, while the tails seem to have different binding sites. This is reasonable, as ligand 5 is longer and more flexible and is thought to interact with His^{137}. Ligand 7 instead is shorter and less flexible and is potentially stabilized by His^{86} as proposed for the binding mode of compound 37 [Christiansen et al., 2011]. Such a stabilization was also found in docking of the short and rigid ligand TUG-424 to the homology model [Lu et al., 2010].

The results are promising, especially in view of the fact that the determination of ligand binding modes in GPCRs remains a difficult task, as can be regularly seen in docking competitions [Michino et al., 2009, Kufareva et al., 2011].

The GPR40 membrane preparations in the light of new experimental data

In the preceding text the situation was described in the context that the experimental data from [Bartoschek et al., 2010] result from specific binding of the ligands to GPR40. This was supported in the paper by the removal of the ligand 5 STD signal by the titration with the natural binder linoleic acid. This is a robust argument, as the probability that linoleic acid blocks the membranes and therefore lowers the STD binding signal of ligand 5 is very unlikely. On the other hand, the paper did not present a NOESY experiment, of the ligands within a sample of normal membranes that do not over-express GPR40. Such experiments were recently done on mammalian CHO and human HEK cells that do not over-express GPR40. According to these experiments, there is no difference in the NMR data between the membranes that over-express GPR40, and those that do not. STD and INPHARMA peaks are in principle the same for the GPR40 over-expressing cells and the normal cells. This would imply that the STD, trNOE and INPHARMA peaks result from unspecific binding of the ligands in the membranes or to unidentified lipid binding membrane proteins.

Additionally it was observed that linoleic acid forms micelles at a certain concentration (0.5 mM) and it is very probable that the fatty acid derived ligands 5 and 6 do the same. Indeed, if the concentration of ligand 5 is increased in solution an STD signal rises in the absence of membranes. It could be assumed that a titration of ligand 6 with ligand 5 would lead to a decrease in the STD signal of the first. Instead, the signal of ligand 6 stays the same and strongly increases at a concentration of 200 μM ligand 5, as can be seen in Fig. 3.19. This supports the idea that ligand 5 forms micelles at high concentrations where other ligands can bind to.

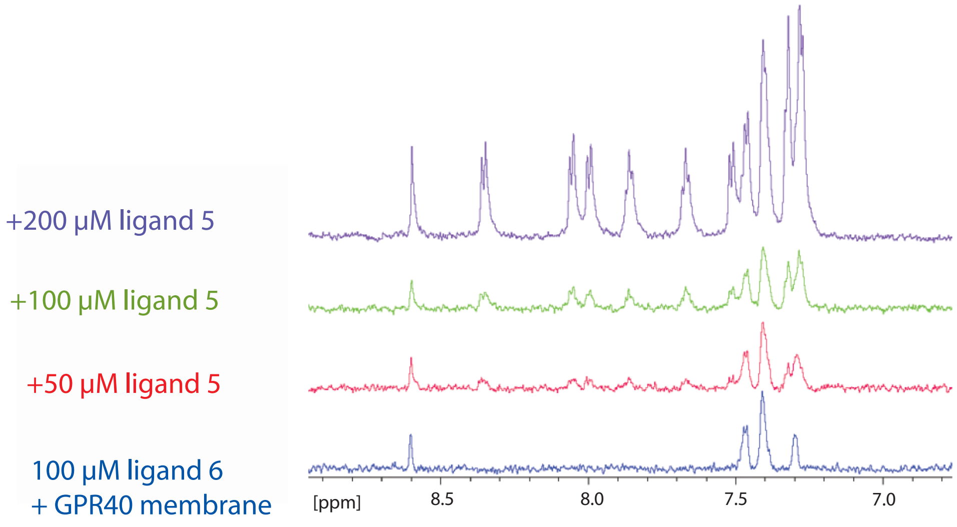


Figure 3.19: Titration of a sample containing GPR40 membranes (HEK) and 100 μM of ligand 6 with ligand 5. While the concentration of ligand 5 is increased, the STD signal of ligand 6 stays constant. At a concentration of 200 μM ligand 5, both STD signals strongly increase. The spectra were recorded by Bettina Elshorst at 298 K with on-resonance irradiation at -0.5 ppm.

The ligands applied are very lipophilic and tend to micelle formation and membrane binding. Probably the presence of the membranes even triggers the formation of micelles. With these new data given it seems possible that STD and INPHARMA can arise from i) micelle formation, ii) membrane binding or iii) unidentified membrane protein binding or a combination of these. It is therefore imaginable that the specific interaction with GPR40 is not the main source of the STD and INPHARMA signals. These observations also question the results and the modeled structures that were shown above and imply that they have to be treated with care.

When the HEK cells were investigated with GPR40-specific antibodies (4093 TEA, Fig. 3.20), it became clear that the over-expression level is not much higher than in the normal membrane. Also it is shown that GPR40 is present in the normal HEK membrane. This led to the question, if there is at all a difference between the membranes that over-express GPR40, and those that do not. This is in turn an argument for the specific interaction of the ligands with GPR40. In the publication [Bartoschek et al., 2010] there were most likely membranes from High Five insect cells applied. These gave, in comparison to HEK and CHO cells, a high over-expression level of GPR40 which is magnitudes higher than in normal membranes. There were no High

Five insect cells without GPR40 over-expression available, but Sf9 insect cells that also originate from a lepidopteran cell line. The experiment was repeated with these membranes, which definitely do not contain human GPR40, and binding signals from STD and INPHARMA were observed. Still, it could well be imagined that insect cells contain yet unidentified fatty acid binding membrane proteins.

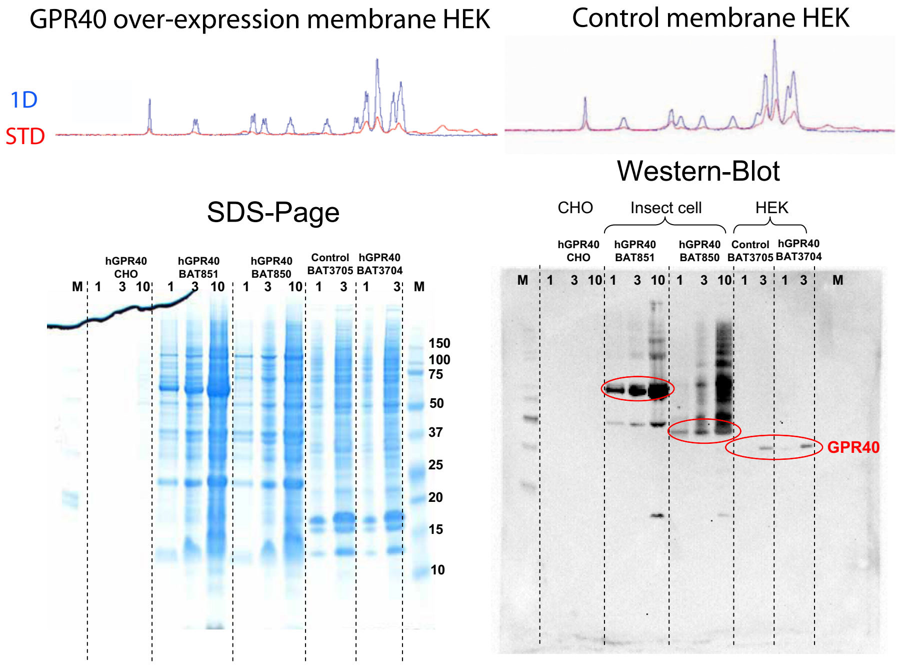


Figure 3.20: Comparison of GPR40 over-expressing membranes and control membranes. The STD signal of both HEK membranes is the same as seen above. This is confirmed by the SDS-Page and Western Plots shown below. GPR40 concentration is nearly equal. CHO cells give no signal at all, but High Five insect cells show a very high over-expression level. These experiments were done within Sanofi.

The question is further complicated by the nature of the ligands, which are very easy to be aligned with their carboxylic moieties and the hydrophobic aromatic tails. It is very probable that their relative binding mode is the same in i) GPR40, ii) the membranes, iii) other membrane proteins or iv) micelles. It can well be that the binding mode which was modeled above based on NMR data is present in GPR40 and at other binding sites. Therefore one would not see a difference whether the ligand is binding specific to GPR40 or not.

The ligands 5 and 6 were measured with liposomes to see if unspecific binding is taking place rather to the membrane or to membrane proteins. The 1D and STD spectra showed a very strong broadening of the signals (Fig. 3.21), which was not observed in the presence of membranes. This indicates that the ligands go to the liposomes and bind there. Yet, in the presence of cell membranes the signal line shapes look very different. The NOESY spectrum of the ligands/liposome sample showed trNOE, but also broadened and no INPHARMA. This is in agreement with the argumentation in [Bartoschek et al., 2010], that the ligands do not exchange enough in the vast space of the membranes to give rise to INPHARMA signals. It can be concluded that the membrane binding is contributing to the NMR signal of ligand/membrane samples, but is not the main contribution to the signal. Therefore in turn membrane protein binding can be assumed as the main source to the NMR signal.

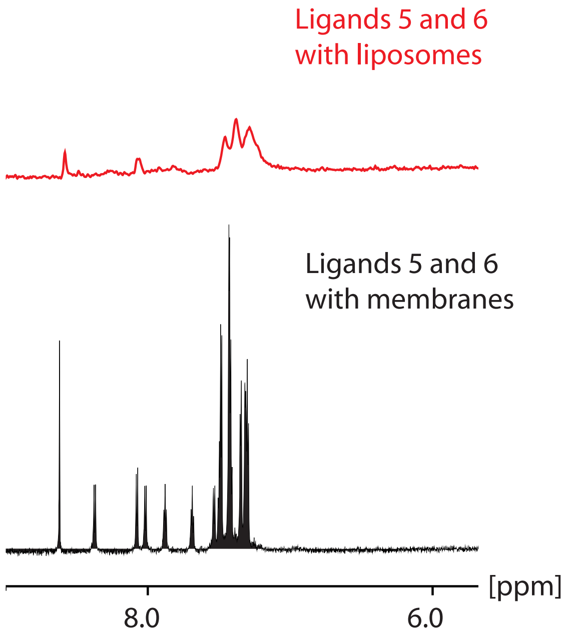


Figure 3.21: Comparison of the ligands 5 and 6 aromatic region together with liposomes (red) and with cell membranes (black). In the case of liposomes an enormous line broadening can be observed. Spectra were recorded on a 400 MHz spectrometer by Adam Mazur.

The putative conclusion is, that there are off-target binding events contributing to the signal and therefore the spectra of GPR40 over-expressing membranes and normal membranes look very similar. This would also be in agreement with the observation of [Bartoschek et al., 2010], that linoleic acid can remove a ligand from its binding site. The difference is just, that the ligand is not removed from GPR40 only, but also from other fatty acid binding membrane proteins. In turn, a titration with ligand 5 can not be successful, as this ligand starts to form micelles at concentrations approximately higher than 0.2 mM.

It is possible that these issues are specific for the ligands of the GPR40 system and do not turn up with other protein systems. Investigations of GPCR and ion channel membrane preparations are at the moment done within proprietary projects at the pharmaceutical company Sanofi to address these questions. It is very important to finally answer these issues to decide if STD, trNOE and INPHARMA are amenable for membrane proteins with the presented sample preparation. First results on the cannabinoid receptor 1 (CB1) showed similar issues to the GPR40 system, i.e. STD and INPHARMA signals in the presence of membranes that do not over-express CB1. On the other hand, STD investigations on whole cells with a CB1 binder were shown to work only with CB1 over-expressing cells and not with others [Pereira et al., 2009]. In this case the ligand was hydrophilic and a large steroid derivative, so there was most probably neither membrane binding, nor unspecific membrane protein binding, nor micelle formation taking place.

In conclusion, it was found that the NMR-based protein-ligand complex structure determination methodology should be applicable to membrane protein preparations if the following issues do not occur: i) micelle formation (can be regulated with ligand concentration), ii) membrane binding (can be regulated with ligand lipophilicity), iii) low target over-expression levels (can be tested with antibodies) and iv) specific or unspecific binding to other membrane proteins (can be tested with normal cell membrane preparations). The fourth point excludes many fragment-like ligands and also the ligands 5 and 6. The modeled binding mode of these ligands can therefore be seen as a general binding motif of these ligands to GPR40 and other fatty acid binding membrane proteins.

Chapter 4

Tubulin-epothilone complex

4.1 Introduction

4.1.1 Tubulin and antimitotic drugs

The protein tubulin is the building block for the microtubules, a cytoskeletal assembly which reaches out from the centrosome into all directions inside the cell. The functions of the microtubules include the stabilization of the cell, the transport of organelles and cell division. During the latter the mitotic spindle is formed by the microtubules, which is the driving force for the division of the cell.

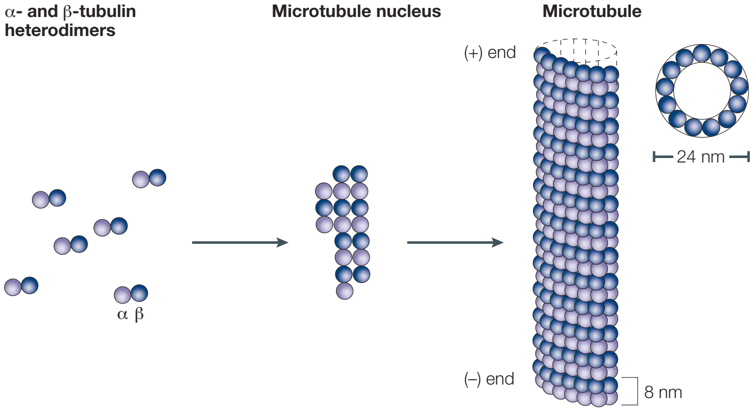


Figure 4.1: Polymerization mechanism of microtubules [Jordan and Wilson, 2004].

In the nucleation phase of the microtubule assembly there are heterodimers of the α- and β-subunits of tubulin formed. These heterodimers have a molecular weight of 100 kDa and range over a distance of 8 nm. Once these heterodimers accumulate with other tubulin heterodimers to a certain size, the elongation of the microtubules starts. The final microtubule assembly structure consists of closed ring made out of 13 protofilaments of head-to-tail assembled heterodimers with a diameter of 24 nm. At the β-tubulin exposing (+) end the microtubule is then further elongated. This procedure of microtubule assembly is illustrated in Fig. 4.1.

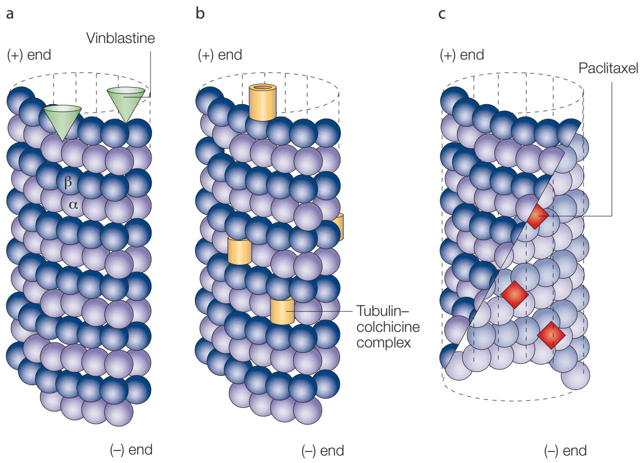


Figure 4.2: Binding sites of antimitotic drugs to microtubules [Jordan and Wilson, 2004]. a) vinblastine binds to the (+) end and suppresses MT assembly. b) colchicine binds to tubulin heterodimers fixing the conformation to suppress MT dynamics. c) paclitaxel binds to the interior surface and stabilizes the MT.

A modulation of the tubulin polymerization (and depolymerization) mechanism has a direct influence on the mitotic spindle that is responsible for cell division, and if mitosis does not function properly, the cell will die or undergo apoptosis. The consequence is, that antimitotic drugs aiming to destroy the mitotic spindle will especially harm cells that tend to divide very often. This holds true for cancer cells and therefore antimitotic drugs are of highest importance for chemotherapy as cytostatica. Three different mechanisms of tubulin modulators have been found within natural products. They differ by their binding site to tubulin or the microtubule: i) the first group comprises the vinca alkaloids from the periwinkle, e.g. vinblastine. They bind to the (+) end of the microtubule, between α and β tubulin, as seen in crystal structure 1Z2B and inhibit polymerization, ii) colchicine from the autumn crocus binds at a distinct site of the β subunit, stabilizing a conformation that inhibits straightening of the microtubule. It is used to cure gout and related inflammatory disorders and iii) paclitaxel, epothilones and related substances bind to the taxane binding site of β-tubulin, which is at the interior of the microtubule and stabilize the microtubule or affect the number of protofilaments. Hereby they inhibit depolymerization or produce a microtubule without proper function. These three different binding sites are illustrated in Fig. 4.2. For cyclostreptin, a covalent binding to Thr^{220} very close to the taxane binding site was observed and proposes a weak binding site en route to the taxane binding site [Buey et al., 2007]. These presented molecules display only a small selection of tubulin binding drugs and many more are under investigation [Dumontet and Jordan, 2010, Amos, 2011]. The most frequently applied antimitotic drug is paclitaxel (trading name: Taxol), originally extracted from the pacific yew tree, that binds microtubules with high affinity and stabilizes them. The binding mode of paclitaxel to Zn-stabilized tubulin sheets is known from the electron crystallography derived structure [Löwe et al., 2001] and it is assumed that the binding mode in tubulin monomers or heterodimers is the same. The binding is stabilized by two phenyl-rings of paclitaxel, that penetrate into grooves on the tubulin surface and stabilize the conformation of the highly flexible M-loop. Paclitaxel is poorly soluble in water and therefore is sometimes replaced in studies by its baccatin III (bacIII) core. First, bacIII was believed to be an inactive derivative of paclitaxel, but then bacIII was found to be a stabilizer of microtubules like paclitaxel with an activity around ten times lower than paclitaxel [Chatterjee et al., 2001, Andreu and Barasoain, 2001]. This indicates that bacIII also specifically targets the taxane site and stabilizes the M-loop in a similar manner to paclitaxel. Yet, there is no data on the binding mode of bacIII published, and it is unknown if the bacIII binding mode is similar or different to the one of paclitaxel.

4.1.2 The binding mode of epothilone

Extensive studies using NMR spectroscopy were done on the binding modes of drug molecules binding to the taxane-binding site of β-tubulin. The bound conformation of discodermolide [Sanchez-Pedregal et al., 2006], dictyostatin [Canales et al., 2008], docetaxel [Canales et al., 2011] and tubulysin [Kubicek et al., 2010] were derived using trNOE and INPHARMA data. The natural product epothilone A (epoA, Fig. 4.4) is also in the group of anticancer agents that target the taxane-binding site of β-tubulin. Epothilone showed activity in cancer cells that were resistant against Taxol and the epothilone derivative ixabepilone was released in the United States under the trading name Ixempra. An electron crystallography (EC) derived binding mode of epothilone to Zn-stabilized tubulin sheets is known [Nettles et al., 2004].

There is a severe disagreement between the bound conformation in solution found by the NMR methods trNOE and CCR [Carlomagno et al., 2003] and the EC-derived structure. An NMR model of the binding mode, based

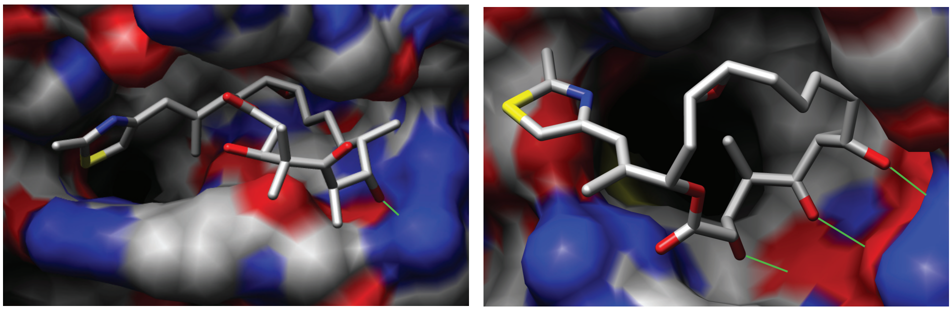


Figure 4.3: Binding mode of epothilone A to tubulin heterodimers as determined by NMR spectroscopy (left) and to Zn-stabilized tubulin sheets as determined by electron crystallography (right).

on molecular modeling with HADDOCK and scoring by INPHARMA was published [Reese et al., 2007] and shows a very different binding mode of epoA to β-tubulin compared to the EC-derived structure. An investigation by solid state NMR of epothilone B (patupilone) bound to microtubules revealed details about the pharmacophore. Still, it was not possible to decide if the EC or the NMR structure is the correct binding mode [Kumar et al., 2010].

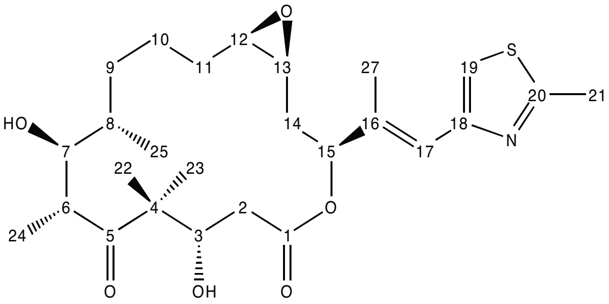


Figure 4.4: Constitution and configuration of epothilone A.

The binding site of cyclostreptin was discovered to be in the pore site, which is situated between the two β units of two tubulin dimers. Following this observation, some authors ([Canales et al., 2008], [Magnani et al., 2009], [Canales et al., 2011]) proposed that the pore site is also the binding site in solution of paclitaxel-like antimitotic drugs that bind the taxane site inside the microtubule. The logic behind this was, that STD data showed differences in the binding mode between the tubulin and the microtubule samples. Additionally it was observed that the binding affinities are not as strong as expected in the tubulin samples. Therefore these drugs would bind at the pore site between dimers that are formed in solution. Conse-

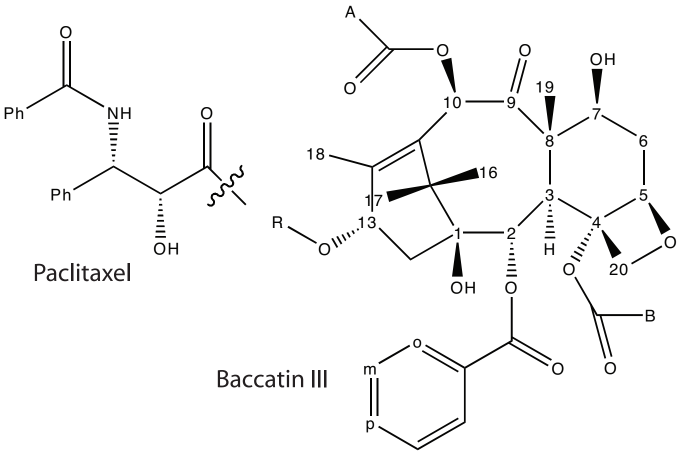


Figure 4.5: Constitution and configuration of the baccatin III core, derived from paclitaxel.

quently the drugs only bind to the taxane site, if a microtubule is either in the growing phase or already present. An exception could be the Zn-stabilized sheets, that were used to determine the crystal structures of paclitaxel and epothilone and which show that the drugs bind to the taxane site. As there are INPHARMA signals between the taxane binding drugs like baccatin, epothilone and discodermolide; epothilone would also be affected. Indeed, there were also INPHARMA peaks observed between epothilone and tubulysin. Tubulysin is a peptide with a structure very similar to soblidotin, a dolastatin derivative. Dolastatin is known from X-ray crystallography to bind close to the vinblastine and cyclostreptin (pore) binding site [Cormier et al., 2008]. Therefore it had to be concluded that the possibility of another binding site on tubulin in solution is probable [Kubicek et al., 2010].

On the other hand, experimental data obtained by H/D exchange (HDX) followed by mass spectrometry measurements give another picture. Based on these data, the taxane binding site is stabilized and therefore protected against deuteration, if the taxane binding site drugs are in solution [Khrapunovich-Baine et al., 2009]. Furthermore three new binding modes of epothilone A are proposed, which are supported by the H/D data [Khrapunovich-Baine et al., 2011]. These are shown in Fig. 4.6 and are closer to the NMR-, than to the EC-derived structure.

At the final stage of this thesis, a new crystal structure of epothilone bound to tubulin was published [Prota et al., 2013]. The structure is deposited in the PDB under the entry code 4I50 and shows a complex between $\alpha\beta$ tubulin, the stathmin-like protein RB3, tubulin tyrosine ligase and epoA. The structure is shown later in the text in Fig. 4.19 and epoA is contrary orientated at the binding site to all above mentioned structures. This new crystal structure shows epothilone bound to tubulin with four hydrogen bonds and fixes the M-loop of tubulin as a structured alpha-helix. The hydrogen bonds are induced by OH3, OH7, O1 and the N of the thiazole ring of epoA and are in good agreement with SAR data [Nicolaou et al., 1997, Regueiro-Ren et al., 2002, Buey et al., 2004]. The crystal structure also explains well, why an addition of a methyl group to C12 results in a much stronger binding affinity, as in this structure there is a cavity to fill by the methyl group.

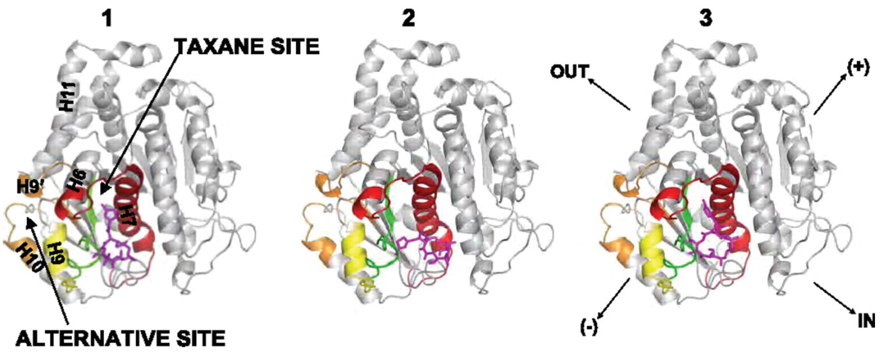


Figure 4.6: Binding mode proposals of epothilone A (shown in magenta) to tubulin, supported by HDX data [Khrapunovich-Baine et al., 2011].

To further investigate the tubulin/epoA system, new STD data were recorded and applied in the following section as a selection criterion to choose between the ambiguous binding modes. As the INPHARMA method depends on two different molecules that target competitively the same binding site, baccatin III (bacIII) was used complementary to epothilone A. BacIII is the core element of paclitaxel (Fig. 4.5, lacking the rest that bears two phenyl rings. The logic behind this is, that bacIII is much better soluble than paclitaxel. Interestingly, INPHARMA found for bacIII another binding mode than it would have been expected from the EC structure of paclitaxel. Thus, for bacIII there were also new STD data obtained.

4.2 Material & Methods

4.2.1 STD experiments

The epoA sample was prepared with 0.5 mM epothilone A and 0.012 mM tubulin (Cytoskeleton Inc., Denver (USA)) in D_2O containing phosphate (3 mM), calcium (1.5 mM) and sodium (0.7 mM). Saturation was applied for 1

s on -3.5 ppm on 298 Kelvin a 900 MHz spectrometer equipped with a cryogenically cooled probe head (ns: 16, TD: 8K). The epoA sample preparation and measurement was done by Karel Kubicek.
The bacIII sample was prepared with 2 mM baccatin III and 0.020 mM tubulin in D_2O PIPES buffer containing sodium phosphate (50 mM), EGTA (1mM), DTT (1mM) and calcium chloride (2mM). Saturation was applied for 1.5 s on -1.5 ppm at 278 Kelvin on a 900 MHz spectrometer equipped with a cryogenically cooled probe head (ns: 512, TD: 8K)

4.2.2 Back-calculation of peak volumes

All INPHARMA, trNOE and STD peak volumes were back-calculated with the software SpINPHARMA, using the complete-relaxation matrix approach [London, 1999, Orts et al., 2009] as described in the introduction part 1.2. Methyl-groups were described using a 3-site jump model [Tropp, 1980]. STD saturation was applied to all methyl groups. Protons within a distance of 8 Å from any ligand proton were considered. The INPHARMA and intraligand peaks were normalized to the sum of all peaks in the direct dimensions [Koeck and Griesinger, 1994]. The correlation time was estimated to be 100 ns. Kd values were estimated to be 0.001 mM or 0.01 mM for epoA and 0.05 mM for bacIII. On-rates (k_{on}) were assumed to be in the diffusion limit 10^8 $M^{-1}s^{-1}$.

4.2.3 Molecular dynamics simulations

The INPHARMA derived HADDOCK model of the epothilone-tubulin complex [Reese et al., 2007] and the EC derived structure [Nettles et al., 2004] were used as starting structures for energy minimization and molecular dynamics simulations. Additionally the same was done for bacIII. Therefore the INPHARMA derived HADDOCK model of the tubulin/bacIII complex [Reese et al., 2007] was used for the NMR model. For the EC structure of bacIII, the EC-derived and refined structure of paclitaxel bound to the tubulin dimer (PDB code 1JFF) was used [Löwe et al., 2001] and paclitaxel was computationally modified to bacIII, followed by energy minimization. For the crystal structure the epothilone binding monomer (chain D) of PDB structure 4I50 was used. Energy minimizations and molecular dynamic simulations were done as implemented in Gromacs [Hess et al., 2008], using the amber99sb force field [Hornak et al., 2006] for the protein and the General Amber Force Field [Wang et al., 2004] for the ligand. EM was done with conjugate gradient integrator in 1000 steps. MD was done using steepest descent integrator with a stepsize of 2 fs. Simulations were run for 1 ns at

300 K and 1 atm in implicit solvent, using the Generalized Born model augmented with the hydrophobic solvent accessible Surface Area (GBSA). From every trajectory 100 snapshots were saved, respectively. RMSD changes of the ligands were calculated within Gromacs using least square fit method to reference to the whole system and to obtain RMSD values, which are not influenced by spatial motions of the protein.

4.3 Results

4.3.1 STD experiments

STD spectra were recorded for epoA and bacIII as described in the methodology section. STD effects were clearly visible and the group binding epitopes could be derived from the spectra, as is shown in Figs. 4.7 and 4.8.

STD data epothilone A

For epoA, the protons H17, H19 and H15 get the strongest saturation. These protons are situated at the linker between the macrocycle and close to the thiazole ring. The methyl group H21, that is also situated at this ring gets the strongest saturation of the epothilone methyl groups. In comparison, the methyl groups H22, H23, H24 and H25, which are close in space, all get much lower saturated than H21. The protons of the hydrophobic stretch get comparably highly saturated, while the protons of the other side of the macrocycle with the three oxygen groups are weakly saturated. These values already result in a rough picture of the binding epitope: The thiazole ring and the linker are very close to the protein, as well as the hydrophobic stretch of the macrocycle. Farther away from the protein and more exposed to the solvent are the methyl groups H22-H25 and the oxygen bearing stretch. These observations are supported and complemented by the WaterLOGSY spectra (not shown). When one inspects the electron crystallography structure of epothilone (PDB code 1TVK), it becomes clear that this group epitope might not fit very well. As seen in the crystal structure, the binding mode is mainly supported here with the oxygen bearing stretch being close to the protein, forming hydrogen bonds with the same. To test this observation, the R_{STD} values for the EC-derived structure and the NMR-derived structure were back-calculated and compared with the experimental data. The result was an R_{STD} for the NMR structure with 0.55 and -0.025 for the EC structure, whereas the NMR structure is clearly favored. An energy minimization improved the NMR structure to 0.56 and the EC structure to 0.11. The crystal structure 4I50 obtained the highest R_{STD} of 0.81.

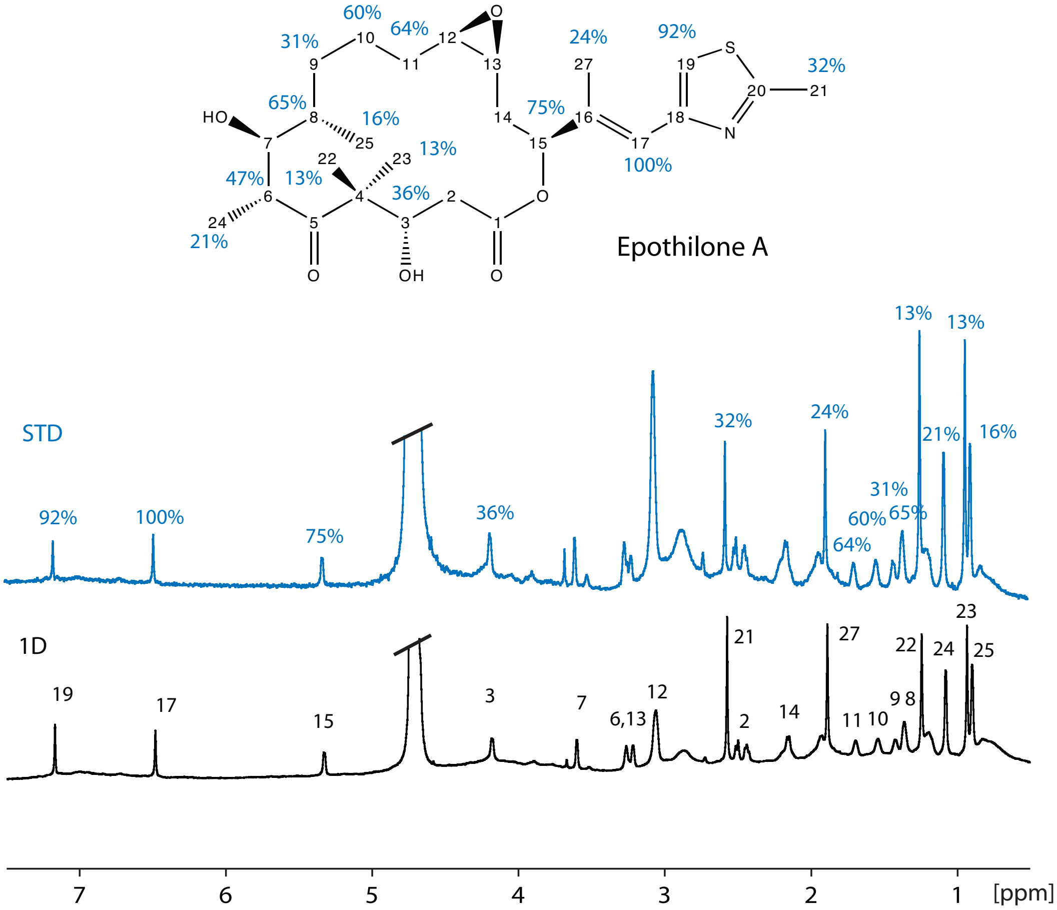


Figure 4.7: Comparison of the 1D and STD spectra of epothilone A in solution with tubulin heterodimers. The $V_{Max.STD}$ value is 96.18.

STD data baccatin III

A comparable study was done for bacIII. Hereby the NMR derived structure was used and for the EC-derived structure the binding mode from paclitaxel to tubulin (PDB code 1JFF) was applied. Therefore the structure of paclitaxel was computationally changed to bacIII and the resulting structure was energy minimized. Accordingly, the NMR structure was also energy minimized with the same (Amber) force field. The group epitope was not as clear as in the case of epothilone. Yet, it can be said that the phenyl ring must be buried in a protein cavity. Calculation of R_{STD} resulted in 0.65 for the NMR structure and 0.67 for the EC structure. The STD values were then combined with experimental trNOE and INPHARMA data of epoA and bacIII from the literature [Reese et al., 2007] to yield R_{STI} values. Back-calculation and

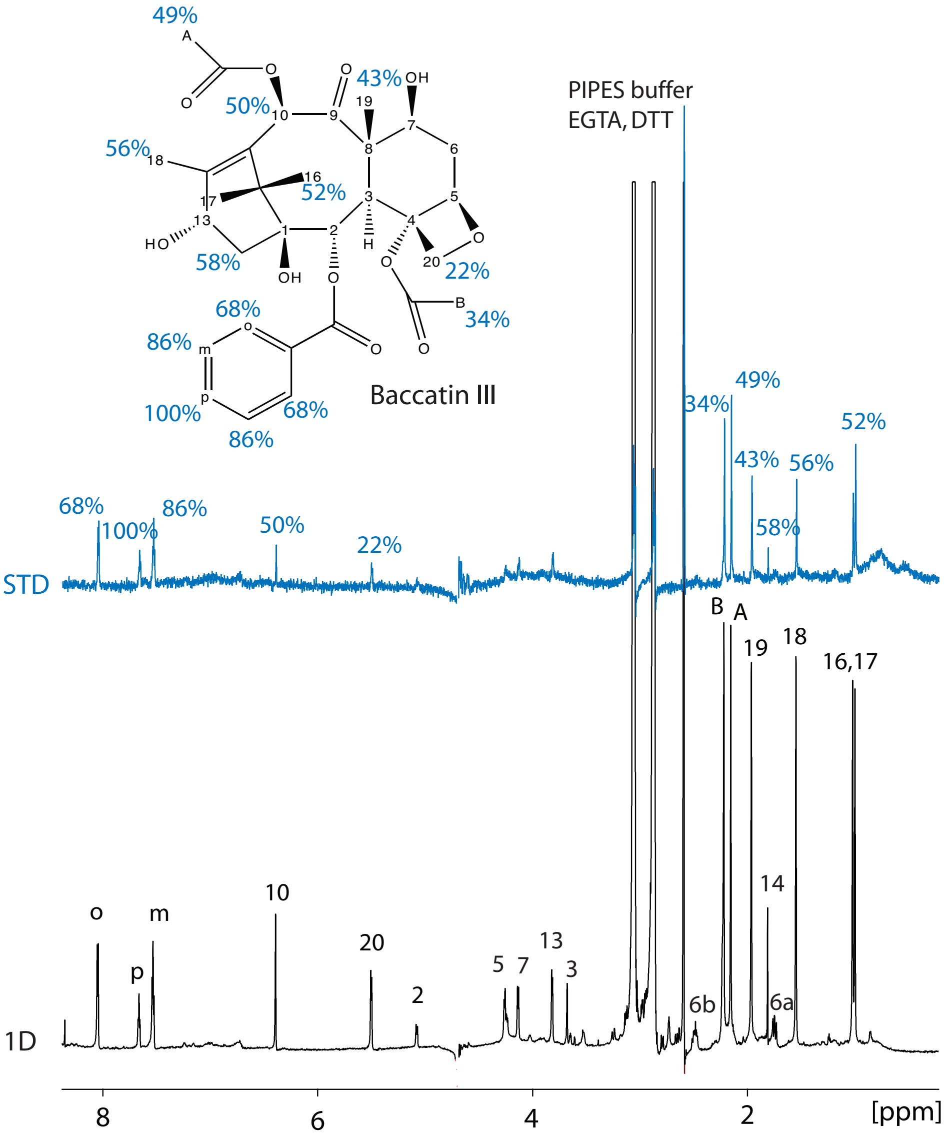


Figure 4.8: Comparison of the 1D and STD spectra of baccatin III in solution with tubulin heterodimers. The $V_{Max.STD}$ value is 14.68.

combination of the parameters was done as described before. For the energy minimized NMR structures of epoA and bacIII, the R_{STI} is 0.63 and for the energy minimized EC structures of epoA and bacIII, the R_{STI} is 0.44. This implies that the NMR structure of epoA is clearly favored, but the bacIII structures are both very similar in their scores. The crystal structure 4I50 obtained an R_{STI} of 0.69 with the NMR structure of bacIII and an R_{STI} of 0.59 with the EC structure of bacIII.

4.3.2 Molecular dynamics simulations

MD epothilone A

To investigate whether scoring against R_{STD} and finally against R_{STI} can be further optimized, an unrestrained molecular dynamics simulation was run for 1 ns in implicit solvent, starting from the EC and the NMR structures, respectively. In Fig. 4.9 and Fig. 4.10, it can be seen how much the binding modes of epoA changed from the starting structure during the MD simulation. It is remarkable, that the flexible M-loop of tubulin, which is oriented differently in the NMR structure than in the EC structure, moves closer to the latter during the simulation of the NMR structure. This can be seen later in Fig. 4.15. 100 snapshots from the MD trajectories were taken respectively and the peak volumes were back-calculated to yield the R_{STD} values. The best scoring NMR MD structure gained an R_{STD} of 0.75. The best EC MD structure gained an R_{STD} of 0.61. The final structure of the MD trajectory gave an R_{STD} of 0.65 for the NMR structure and an R_{STD} of 0.45 for the EC structure. Still, the NMR structure scores better, but the EC structure was significantly improved. As can be seen in the trajectory 4.10, the position of epoA in the EC structure changes immediately, which also supports that the starting structure is energetically not very favorable.
An additional MD trajectory lasting 10 ns was recorded for the epoA bound monomer of the crystal structure 4I50 and is shown in Fig. 4.11. The MD was recorded for 10 ns, as after 1 ns a sudden change in RMSD happened. In the end it turned out to be an outlier, caused by a sudden movement of the M-loop towards the solvent. EpoA in this structure stays stable and the final structure changed only 1.9 Å in RMSD from the starting structure. 100 snapshots from the MD trajectories were taken respectively and the final structure of the MD trajectory got an R_{STD} of 0.76 and best scoring structure an R_{STD} of 0.84.

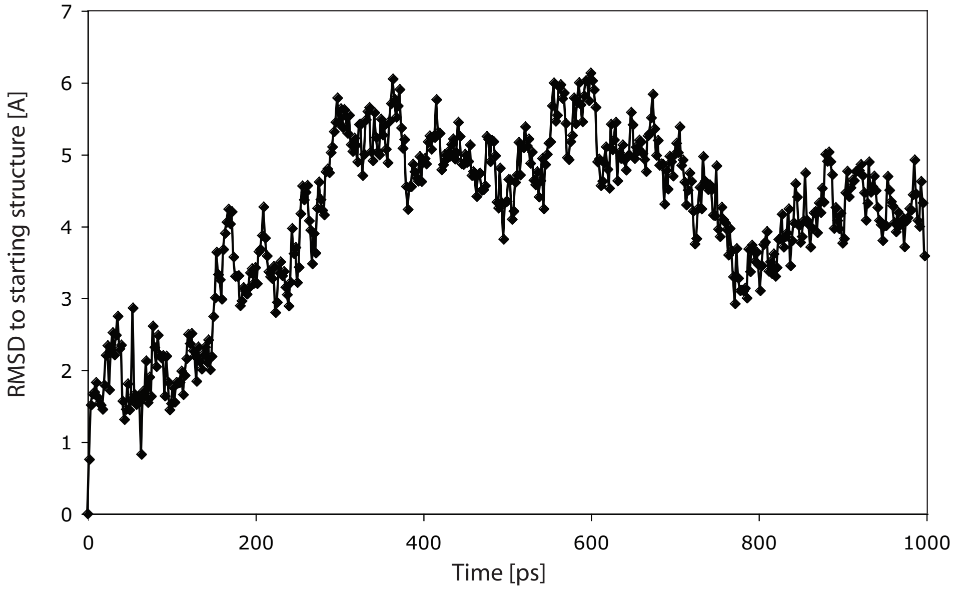


Figure 4.9: MD simulation starting from the NMR-derived tubulin/epoA complex structure for 1 ns. RMSD changes of epoA relative to the starting structure are indicated.

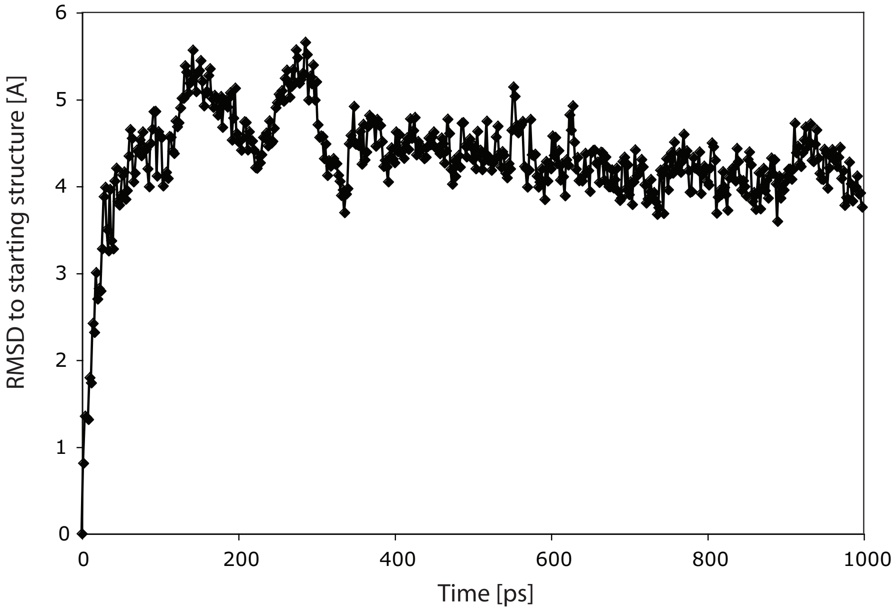


Figure 4.10: MD simulation starting from the EC-derived tubulin/epoA complex structure for 1 ns. RMSD changes of epoA relative to the starting structure are indicated.

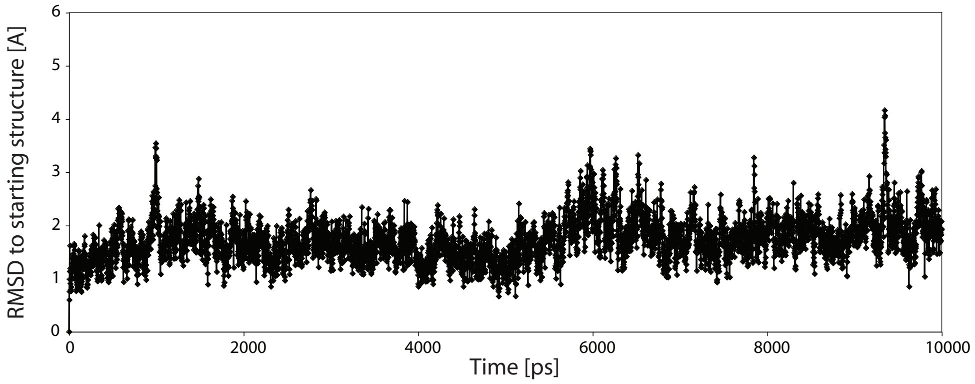


Figure 4.11: MD simulation starting from the crystal structure 4I50 of the tubulin/epoA complex structure for 10 ns. RMSD changes of epoA relative to the starting structure are indicated. Outliers are caused by sudden movements of the M-loop towards the solvent.

MD baccatin III

A similar MD simulation as for epoA was done for bacIII. The trajectories of the EC and NMR structure are shown in Fig. 4.12 and Fig. 4.13. The best scoring NMR MD structure by R_{STD} got 0.75 and the best EC MD structure gained a comparable R_{STD} of 0.83. The final structure of the MD trajectory gave an R_{STD} of 0.66 for the NMR structure and an R_{STD} of 0.43 for the EC structure. As can be seen in the RMSD values of the figures, is the NMR derived structure more stable, i.e. does change less from the original structure than the EC derived model. The latter changes around 5 Å RMSD compared to the starting structure, while the NMR structure changes only around 2.5 Å. To evaluate how representative this stability is, another unrestrained MD simulation of the original EC-derived structure of paclitaxel bound to tubulin was run. This structure is generally considered to be correct and reproducible. It can be seen in Fig. 4.14, that paclitaxel is more stable, changing only around 2.5 Å RMSD from the starting structure, compared to the derived baccatin III binding mode. These simulations prefer therefore the NMR derived binding mode as the more stable one.

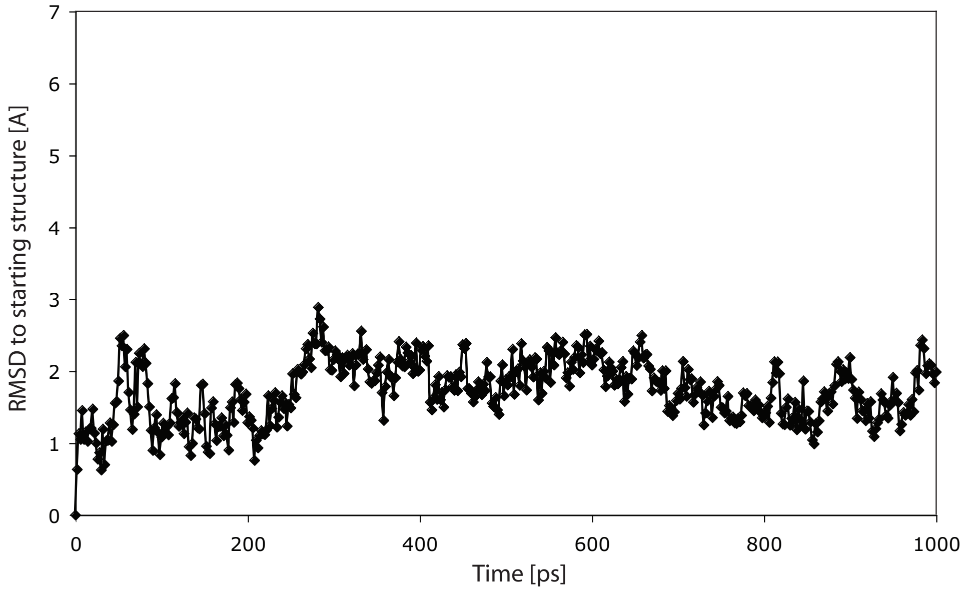


Figure 4.12: MD simulation starting from the NMR-derived baccatinIII/tubulin complex structure for 1 ns. RMSD changes of baccatinIII relative to the starting structure are indicated.

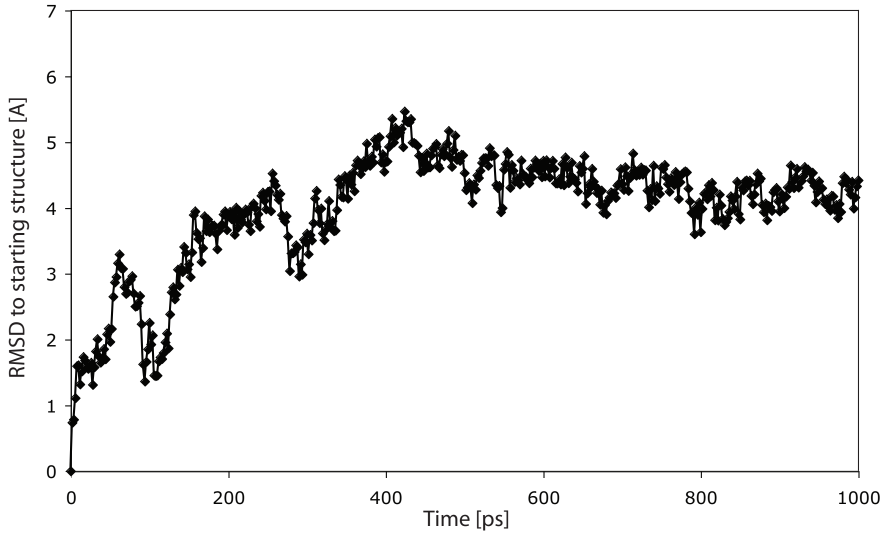


Figure 4.13: MD simulation starting from the EC-derived (derived from paclitaxel) tubulin/bacIII complex structure for 1 ns. RMSD changes of bacIII relative to the starting structure are indicated.

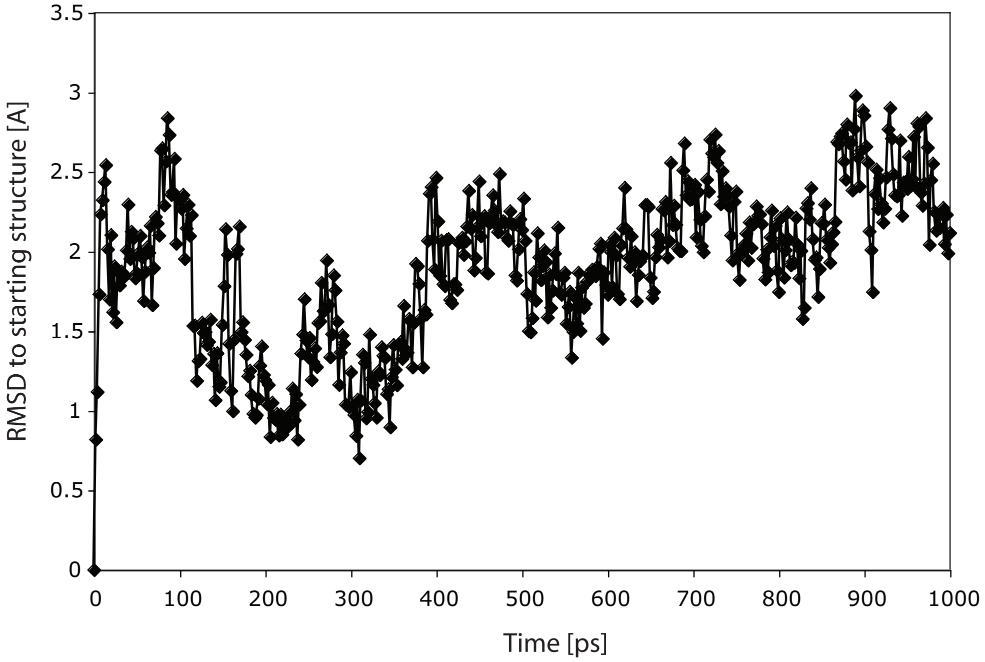


Figure 4.14: MD simulation starting from the EC-derived tubulin/paclitaxel complex structure for 1 ns. RMSD changes of paclitaxel relative to the starting structure are indicated.

Scoring by R_{STI}

The new MD structures for epoA and bacIII, can be combined pairwise to R_{STI} and scored accordingly. The best NMR structure pair of epoA and bacIII by R_{STI} is 0.76 and the best EC structure pair obtained an R_{STI} of 0.7. The NMR structures are seen in Figs. 4.15 and 4.17; the EC structures in Figs. 4.16 and 4.18 in comparison with the starting structures that obtained 0.63 for the NMR, and 0.44 for the EC structure pair. At the end of the simulation, the NMR pair obtained an R_{STI} of 0.5 and the EC pair 0.53. The crystal structure 4I50 of epoA in combination with the EC structure of bacIII got as best R_{STI} 0.69 and 0.55 for the final MD structure pair. Together with the NMR structure of bacIII, the crystal structure of epoA obtained a best R_{STI} value 0.79, and 0.73 for the final MD structure pair.

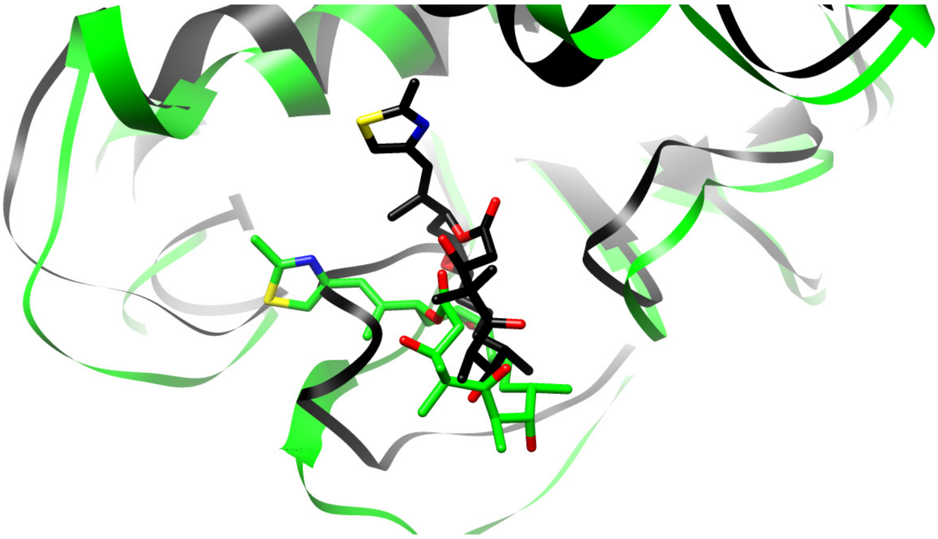


Figure 4.15: Tubulin/epothilone A complex structure derived from NMR [Reese et al., 2007] in black (R_{STI}: 0.63) and in green the structure that obtained the best R_{STI} value (0.76) during an MD simulation of 1 ns. It can be seen, that the thiazole ring moves to the left, as the flexible M-loop does the same.

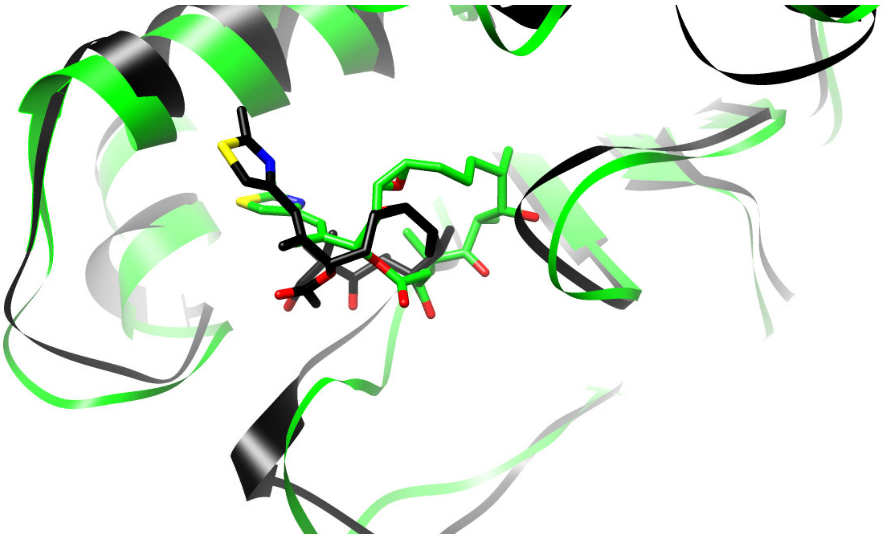


Figure 4.16: Tubulin/epothilone A complex structure derived from EC [Nettles et al., 2004] in black (R_{STI}: 0.44) and in green the structure that obtained the best R_{STI} value (0.7) during an MD simulation of 1 ns. The macrocycle moves into another position, while the orientation of the M-loop does not change.

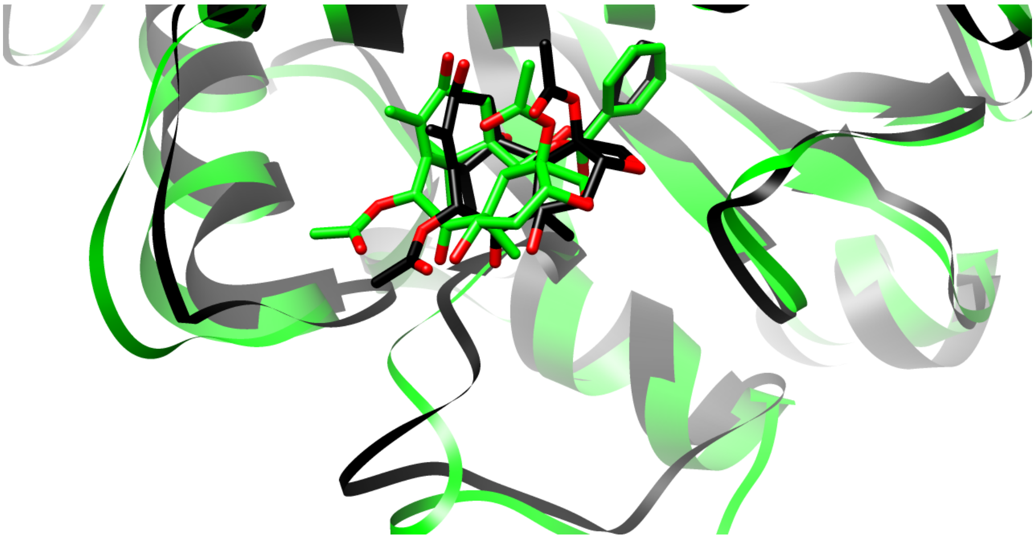


Figure 4.17: Tubulin/baccatin III complex structure derived from NMR [Reese et al., 2007] in black and in green the structure that obtained the best R_{STI} value during an MD simulation of 1 ns. The position of baccatin III does not change very much and the position of the phenyl ring does not change at all. Also the orientation of the M-loop does not change much.

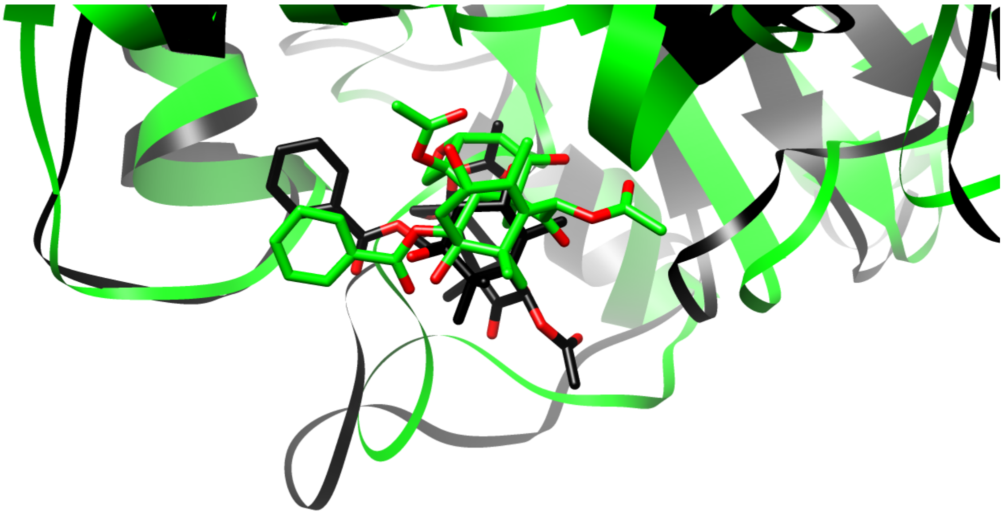
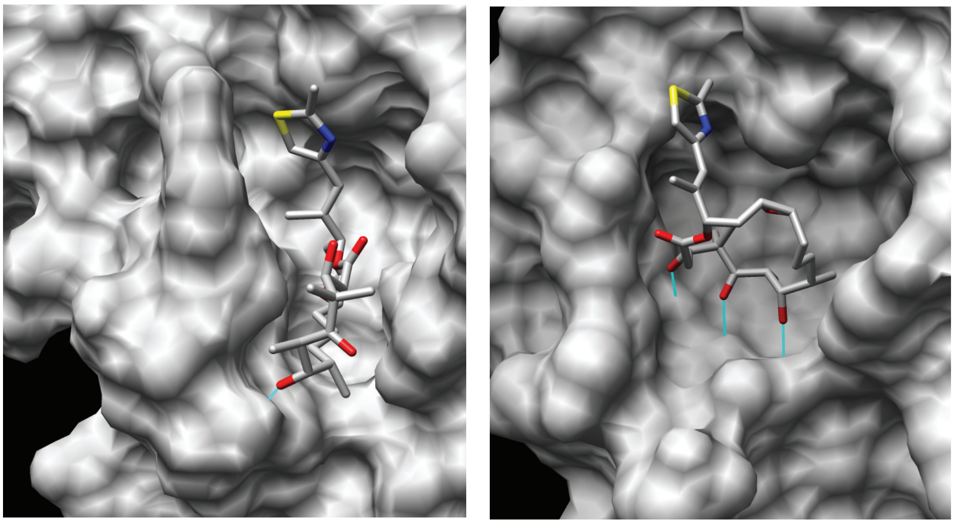


Figure 4.18: Tubulin/baccatin III complex structure derived from the EC structure of paclitaxel [Löwe et al., 2001] in black and in green the structure that obtained the best R_{STI} value during an MD simulation of 1 ns. Baccatin III begins to leave the binding site and the M-loop changes its orientation.

4.4 Discussion & Conclusion

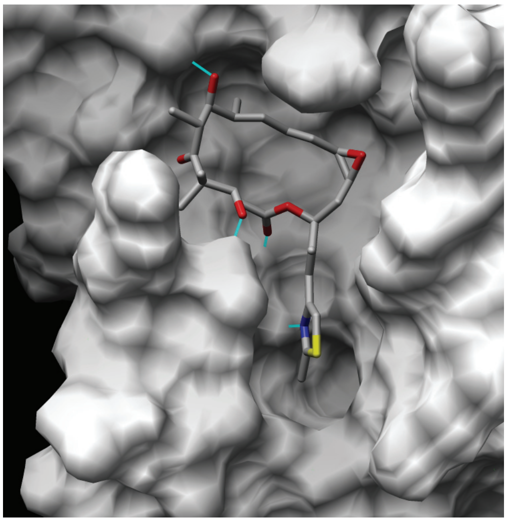
The ambiguous binding mode of epothilone to tubulin in solution was reinvestigated based on new data. Therefore the existing structures were run in an unrestrained MD simulations to create a large ensemble of conformations. These were then evaluated by new experimental STD data measured for epothilone A and baccatin III. Finally they were compared by their respective R_{STD} and R_{STI} values at the start and end of an MD simulation of 1 ns, as well as by the best scoring structure within this simulation. According to these data, the INPHARMA derived binding mode of epothilone is favored. It scores much better than the EC derived binding mode by R_{STD} at the start and at the end of the MD simulation. Also the R_{STI} values of the NMR structure are better. The new crystal structure 4I50 of epoA displays the best R_{STD} and R_{STI} values, whereas the combination with the NMR structure bacIII is favored over the EC structure of bacIII.
In the case of baccatin III, the situation is more difficult. The NMR structures display R_{STD} and R_{STI} values very close to those for the EC structure. Yet, the NMR structure is stable during the MD simulation, as the paclitaxel structure is, while the EC structure is changing much more in RMSD. It can not be said for sure in which protein cavity the phenyl ring penetrates and which of the structures is closer to the actual binding mode. On the other hand, it is possible that the phenyl ring of bacIII could penetrate into both cavities respectively and therefore bacIII possesses two distinct binding modes. The NMR signal would then include the information of both binding modes and both would obtain very similar scores, as it is observed.

So far trNOE, INPHARMA and STD data were applied from the NMR side to reveal the binding mode of epothilone to tubulin. Until now it was not possible to express tubulin in bacteria, meaning that no labeling with stable isotopes for heteronuclear NMR is possible. Therefore nothing more can be done from the NMR side. Yet, there are chemical shift data from the solid state investigations of epothilone and microtubules. These can be back-calculated and compared with different binding modes. Such an approach could yield new informations on the binding mode of epothilone in the future.



NMR structure
$R_{STD} = 0.56$
$R_{STI} = 0.63$ (with bacIII NMR)
RMSD change during MD of 1 ns: 3.9 A

EC structure (PDB code 1TVK)
$R_{STD} = 0.11$
$R_{STI} = 0.44$ (with bacIII EC)
RMSD change during MD of 1 ns: 4 A



Crystal structure (PDB code 4I50)
$R_{STD} = 0.81$
$R_{STI} = 0.69$ (with bacIII NMR)
$R_{STI} = 0.59$ (with bacIII EC)
RMSD change during MD of 10 ns: 1.9 A

Figure 4.19: Comparison of the three binding modes of epothilone to tubulin that were mentioned in the text from left to right: NMR structure, EC structure and crystal structure. The individual R_{STD} and R_{STI} values are indicated, as well as the change in RMSD change from the starting structure during an MD simulation.

Chapter 5

Drug metabolism of anle138b

5.1 Introduction

5.1.1 Toxic protein aggregates in neurodegenerative diseases

The hallmark feature of neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD) and prion diseases is the ability of particular proteins to fold into stable alternative conformations and to form neurotoxic aggregates that accumulate in the brain [Soto and Estrada, 2008]. These deposits share similar morphological, structural and staining characteristics, although the main protein component is disease-specific.
Alzheimer's disease is the most common form of dementia and is characterized by extracellular plaques of β-amyloid and intracellular tangles, consisting of paired helical filaments of the microtubule-associated protein tau. PD affects about 1% of the population beyond 65 years of age. Pathologically, PD is characterized by depositions of α-synuclein aggregates and degeneration of dopaminergic neurons in the substantia nigra, resulting in impaired motor functions [Forman et al., 2005]. Prion diseases are invariably fatal neurodegenerative diseases that include Creutzfeldt-Jakob disease (CJD) and bovine spongiform encephalopathy (BSE). They are caused by an unconventional infectious agent which consists primarily of the misfolded, aggregated and beta-sheet rich PrP^{Sc} isoform of the membrane glycoprotein PrP^{C}. The available evidence suggests that PrP^{Sc} acts both as a template for this conversion and as a neurotoxic agent causing neuronal dysfunction and cell death [Giese and Kretzschmar, 2001]. Recently it was shown that the drug bexarotene promotes expression level of ApoE, leading to the clearance of β-amyloid from the brain [Cramer et al., 2012]. Nevertheless it is

more straight forward and potentially more effective to develop drugs that destabilize the pathological protein aggregates in the case of β-amyloid, tau, α-synuclein and the prion protein. No such drugs exist at the moment on the market, but the need to find a cure for neurodegenerative diseases beyond the symptomatic level is essential in a society constantly growing in age.

5.1.2 Anle138b - a novel aggregation modulator

In 2000, the first method for the detection of pathological prion protein (PrP^{Sc}) in cerebrospinal fluid was developed. This was achieved by using a set-up for confocal dual-color fluorescence correlation spectroscopy. The methodology was especially useful to get a definitive diagnosis for the human form of the prion disease Creutzfeld-Jakob. As there is no drug available to cure this disease, the set-up was applied for a high-throughput screening of a chemical library against the aggregation of the prion protein. The results of the antiprion drug discovery were presented in 2005 and revealed a new class of substances as potential lead structure: N-benzylidene-benzohydrazide derivatives [Bertsch et al., 2005]. Indeed, these substances showed anti-aggregation effects on the prion protein in cell cultures, but interestingly also against the aggregation of α-synuclein (Parkinson's disease) and PolyQ (Huntington's disease). Based on these results, the stability of the molecular structure was improved by a ring closure, leading to di-phenyl-pyrazoles. A library of 150 compounds based on this lead structure were tested simultaneously against the prion protein and α-synuclein in SIFT assays, as well as *in vivo*, e.g. with prion-infected or MPTP mice. The best results for both targets (prion protein and α-synuclein), together with the best performance in penetrating the blood-brain-barrier led to the selection of anle138b as the drug candidate. An overview of this drug discovery process is given in Fig. 5.1. At this point the pharmacokinetics, including drug metabolism and toxicity of anle138b were unknown. The motivation of this chapter is to shed light on this very important part of the drug development process.

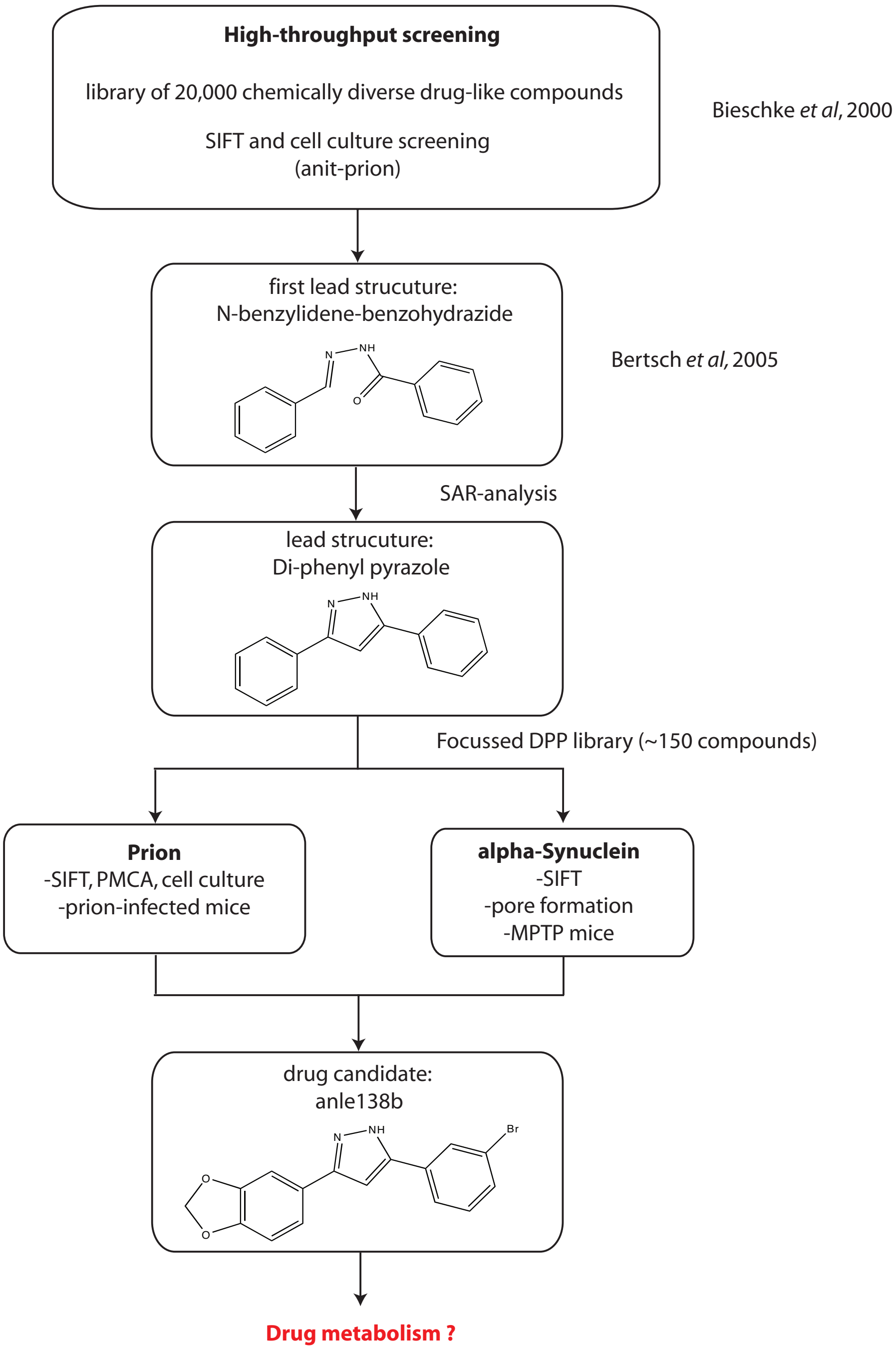

High-throughput screening
library of 20,000 chemically diverse drug-like compounds
SIFT and cell culture screening
(anit-prion)
Bieschke *et al*, 2000
first lead strucuture:
N-benzylidene-benzohydrazide
N
NH
O
Bertsch *et al*, 2005
SAR-analysis
lead strucuture:
Di-phenyl pyrazole
N
NH
Focussed DPP library (~150 compounds)
Prion
-SIFT, PMCA, cell culture
-prion-infected mice
alpha-Synuclein
-SIFT
-pore formation
-MPTP mice
drug candidate:
anle138b
N
NH
Br
O
O
Drug metabolism ?

Figure 5.1: Summary of the anle138b drug discovery process and the emerging question concerning the drug metabolism.

5.2 Material & Methods

5.2.1 HPLC analysis

Treated mice were sacrified, dissected, and samples were frozen at -80°C. The tissues were de-frozen at 4°C prior to use. Tissues were homogenized twice in 5 mL of acetonitrile at maximum speed for 3 minutes using a homogenizer (IKA ULTRA-TURRAX Tube drive workstation, Germany). The homogenate was ultrasonicated at 30°C for 5 minutes and centrifuged at 5000 g for 10 min. An aliquot (0.1 mL) of supernatant was injected into the HPLC system. Analytical high performance liquid chromatography (HPLC) was performed using a Waters HPLC system with a Waters 996 Photodiode Array Detector. All separations involved a mobile phase of 0.1% trifluoroacetic acid (TFA) (v/v) in water (solvent A) and 0.1 % TFA in acetonitrile (solvent B). HPLC was performed using reversed-phase (RP) column Eurospher RP 18, 100 Å, 5 μm, 250 x 4.6 mm at flow rates of 1 mL/min with a gradient of solvent B from 0% to 100% in 50 minutes. The effluent was monitored for UV absorption at 260 nm. Samples were quantified using the peak area ratio of the measured compound to synthesized anle138b as external standard.

5.2.2 Mass spectrometric analysis

Peaks were collected and injected into an LTQ Orbitrap FT-mass spectrometer. Measurement was done in ESI positive ion mode with and without collision-induced fragmentation. Due to the nearly equal natural abundances of the two bromine isotopes in anle138b, double peaks with the mass difference m/z 2 were manually selected and separated by the LTQ quadrupole for ion scan. These were then fragmentated by the collision induced decay (CID) at different energy levels and the specific fragments were measured in the FT-MS.

5.2.3 NMR analysis

Anle138b and anle138c were dissolved in DMSO-d_6 and were measured on a 400 MHz spectrometer. For the metabolites, peaks were collected from HPLC and the samples were frozen in liquid nitrogen. Then, the samples were lyophilized to yield a white powder, which was dissolved in DMSO-d_6. 0.2 mL of the sample solution were taken for NMR measurements. These were carried out on a 600 MHz or 900 MHz spectrometer, both equipped with a cyrogenically cooled probehead. Detailed sample conditions are in the appendix.

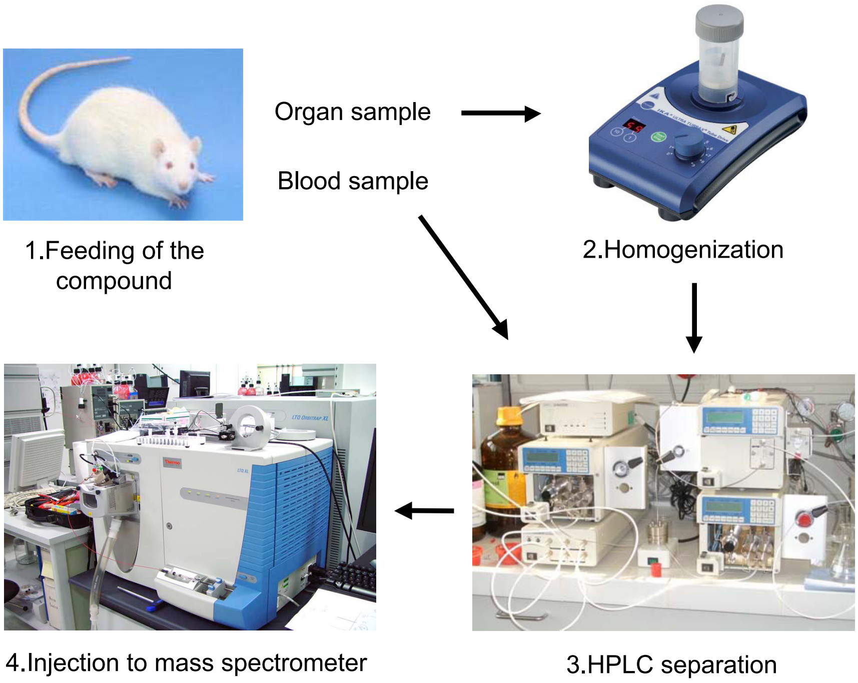


Figure 5.2: Investigation of drug metabolism: 1) the drug is fed to an animal, which is afterwards 2) sacrificed to deliver organ samples or blood samples are taken. 3) Homogenized organs or blood is run over an HPLC column to separate the organic substances and 4) these are then further investigated by mass spectrometry.

5.3 Results

5.3.1 Anle138b analysis

In the first place a chemical analysis of anle138b was done to develop and test the applicability of the method. In further studies the same methodology was used to determine concentrations of anle138b in blood and tissue samples. The molecule was clearly identified by mass spectrometry and displayed the bromine natural isotope twin peak as seen in Fig. 5.5. After anle138b was identified with mass spectrometry, it was measured using induced fragmentation within the orbitrap mass spectrometer to obtain the fragmentation pattern. During fragmentation, anle138b undergoes two major rearrangements of the methylenedioxy group or a release of the bromine atom. The pattern is interpreted in Fig. 5.3.

Anle138b was measured with NMR spectroscopy and was also clearly identified (see appendix). Here the tautomeric constitution is expressed as equally distributed twin peaks in the spectrum. The compound is soluble in

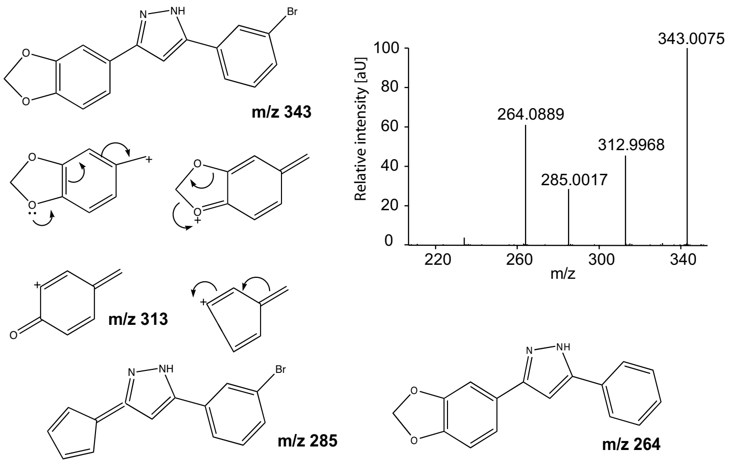

m/z 343
m/z 313
m/z 285
m/z 264
Relative intensity [aU]
100
80
60
40
20
0
220
260
300
340
m/z
343.0075
264.0889
312.9968
285.0017

Figure 5.3: The mass spectrum of anle138b, after selection of the *m/z* 343 isotope during induced fragmentation. In addition to the original mass, three masses are visible. *m/z* 285 and *m/z* 318 can be created during a rearrangement as described in [Yan et al., 2007]. *m/z* 264 originates from a deletion of the bromine.

acetonitrile and can be quantified with HPLC diode array, using a standard. In Fig. 5.4 the concentration of anle138b in rat plasma is visible at different time points.

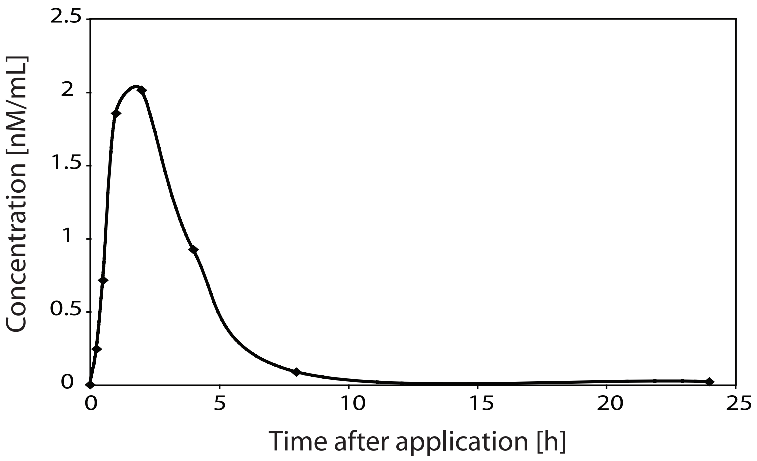


Figure 5.4: Concentration of anle138b in rat plasma at different timepoints. The maximum concentration is reached after 2-3 hours and then constantly declines until it is gone after 12 hours.

Different extraction protocols of anle138b from mouse tissues were tested, including methanol, ethanol, dichlormethane, trichlormethane and acetonitrile. Acetonitrile turned out to extract and dissolve anle138b the best, so it was used in further applications. As described in the methods section, the extraction was supported by ultrasonication. Cell debris was removed by centrifugation. After the methodology was validated, it was applied on mouse brains to search for metabolites in the brain. As can be seen in Fig. 5.5, the HPLC chromatogram of a mouse brain homogenate clearly displays anle138b (confirmed by mass spectrometry), but no additional metabolite. This observation was confirmed by other mouse brains (with and without MPTP) and rat brain samples. The reason for the background free detection of anle138b in the HPLC chromatogram is the following: i) the level of anle138b in the brain is high and anle138b has a high solubility in acetonitrile, ii) many other small organic molecules (e.g. amino acids, lipids, dopamine) that are present in the brain have a lower concentration and/or a lower solubility in acetonitrile and/or a low extinction coefficient at 260 nm. Therefore several small peaks observed in the chromatogram could reflect the presence of other UV-active endogenous compounds from the brain and iii) proteins and other macromolecules are insoluble in acetonitrile.

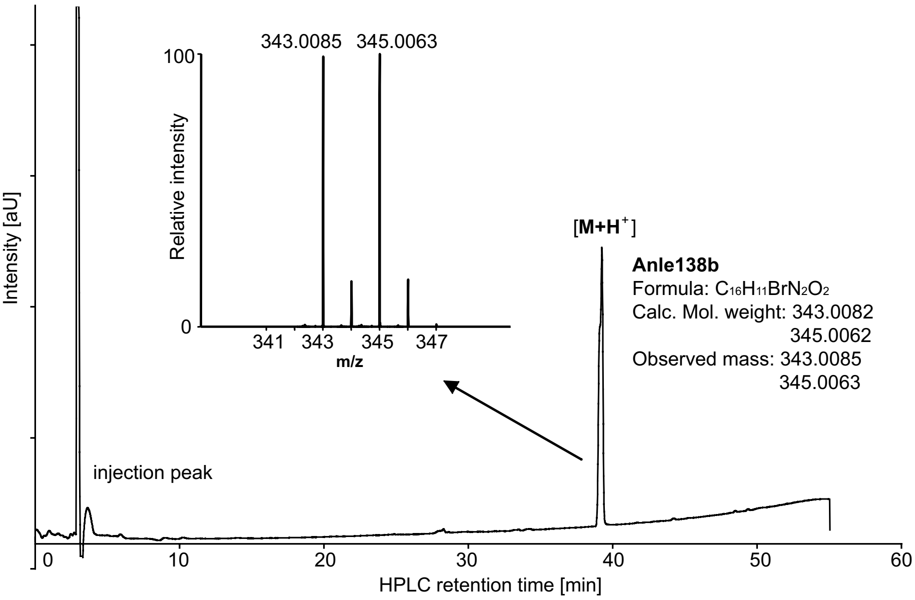


Figure 5.5: HPLC chromatogram of an MPTP mouse brain homogenate. The high resolution mass spectrum (ESI$^+$ mode) identifies the present compound clearly as anle138b from the mass and the two equally populated isotopes ^{79}Br and ^{81}Br and the two additional small peaks shifted by one mass unit originating from the ^{13}C isotope for 16 carbons in the molecule.

No metabolites were found in the brain. It has to be determined clearly if they were absent or just not extractable with the applied methodology. So following the brain analysis, organs that are heavily involved in metabolism like liver and kidney were also investigated. And indeed, in liver and kidney (not in other organs) more metabolites were detected. This can be seen in Fig. 5.6. These additional HPLC peaks were collected and investigated for their masses. As anle138b contains a bromine atom, all metabolites also display the same bromine double mass peak. It would be very unlikely that a bromine atom gets removed during metabolism, but the other peaks were also investigated and identified as impurities or relicts from the chemical synthesis.

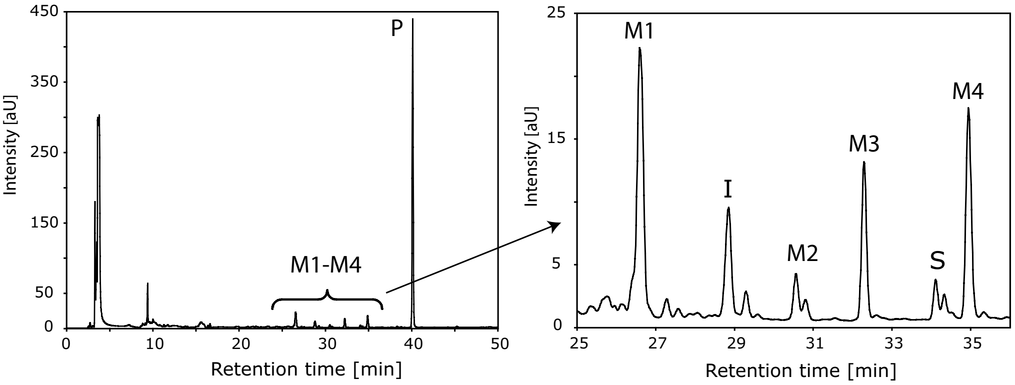


Figure 5.6: Representative HPLC chromatogram of a rat liver. The parent drug anle138b (P) is highly concentrated. Minor concentrated are the four metabolites (M). Also present is an unknown impurity (I) and a relict from the compound synthesis (S).

5.3.2 Metabolite analysis

The metabolites identified during HPLC analysis were collected after separation and intensively studied with high resolution mass spectrometry and NMR spectroscopy and are described in the following section. The mass spectra that follow induced fragmentation together with their interpretation and NMR spectra can be found in the appendix. As seen in Fig. 5.7, the structures of all four metabolites were elucidated.

M1 - Sulfurylated anle138c

In the first place the double peak at m/z 410/412 was identified. Induced fragmentation resulted in the mass peaks m/z 331/333. This product was clearly identified with further fragmentation as anle138c. So a functional group of m/z 79 is added to anle138c. This corresponds well with the mass

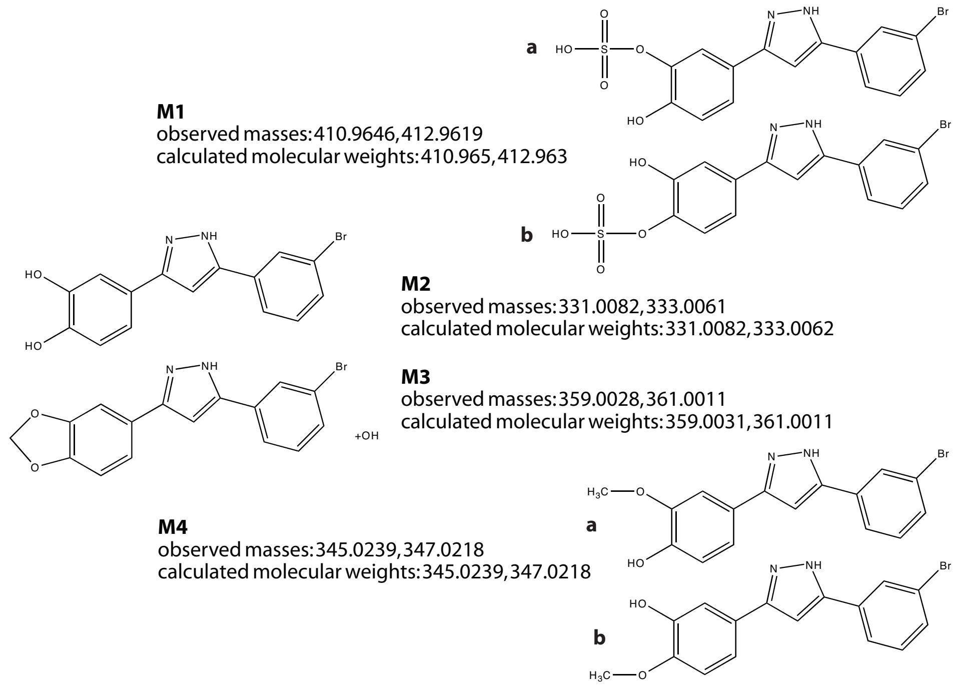

a
M1
observed masses:410.9646,412.9619
calculated molecular weights:410.965,412.963
b
M2
observed masses:331.0082,333.0061
calculated molecular weights:331.0082,333.0062
M3
observed masses:359.0028,361.0011
calculated molecular weights:359.0031,361.0011
+OH
M4
observed masses:345.0239,347.0218
calculated molecular weights:345.0239,347.0218
a
b

Figure 5.7: The observed masses were compared with probable metabolic structures of anle138b. Due to the highly accurate observed masses, the chemical formulas of the metabolites could be determined.

of a sulfate group. Sulfurylation is a typical phase II reaction, normally done by the enzyme SULT1A3. Further investigations with NMR confirmed the structure that yields two different molecules, depending which hydroxy group is sulfurylated. As both hydroxy groups of anle138c are possible to get sulfurylated, there are two metabolites: M1a and M1b.

M2 - Demethylenated anle138b (anle138c)

The double peak m/z 331/333 was identified, which corresponds to the compound anle138c. Anle138c was synthesized as part of the DPP compound library and can therefore easily be analyzed. It is identical with metabolite M2. From the metabolism of the amphetamine drug MDMA (ecstasy) it is known that a demethylenation of the methylenedioxy group to a catechol group occurs [Tucker et al., 1994]. A similar reaction could be expected for anle138b and the catechol derivative is then anle138c.

M3 - Hydroxylated anle138b

The double peak m/z 359/361 was identified. This mass corresponds to an addition of two methyl groups to anle138c. Yet, the observed mass m/z 359,0028 does not correspond well with the calculated mass 359.02 of this dimethylated molecule. Another interpretation is a demethylenation of anle138b, followed by two rearrangements, as observed for the anle138b fragmentation. This is supported by the two other mass fragments at m/z 301/303 and the anle138b fragment with the bromine atom released at m/z 280. The mass difference to anle138b for all fragments is 16, corresponding well to an additional hydroxy group (in exchange to a hydrogen). Some CYP450 enzymes are known to transfer hydroxy groups to drug molecules. Yet, the position of the hydroxy group at M3 can not be determined by mass spectrometry, as it does not go off during induced fragmentation.

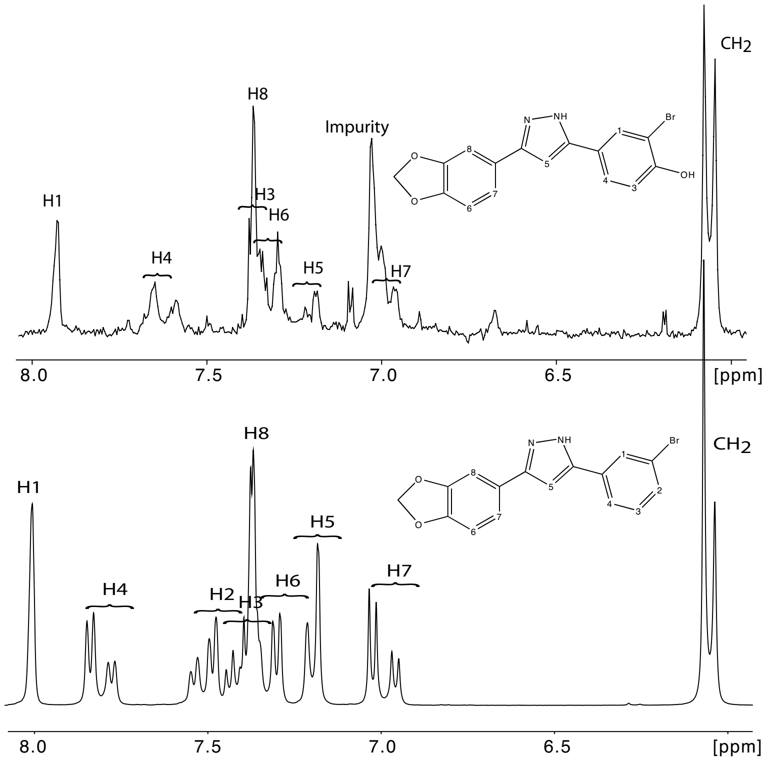


Figure 5.8: Comparison of the 1D spectra of M3 and beneath of anle138b. M3 lacks the signal of H2, so this is the most probable hydroxylation site.

Analysis with NMR was done and similar signals to anle138b were observed, but not proton H2, as seen in Fig. 5.8. hence, it is assumed that the hydroxy group is placed to the aromatic ring at the position of H2. This is supported by empirical chemical shift predictions, that would expect an upfield shift of the resonances, when a hydroxy group is added to the bromine bearing aromatic ring.

M4 - Methylated anle138c

The double peak m/z 345/347 was identified. Induced fragmentation resulted in the mass peak m/z 330/332 and further fragmentation into the mass m/z 313/315. The first change comprises a functional group of m/z 15, which corresponds well to a methyl group. This methyl group was also observed in an NMR spectrum and suggests that M4 is a methylated derivative of anle138c, whereas one hydroxy group gets methylated (done by the enzyme COMT), respectively. The second mass m/z 313/315 also confirms that M4 is a derivative of anle138c. As both hydroxy groups of anle138c are possible to get methylated, there are two metabolites: M4a and M4b.

5.3.3 Pharmakokinetic studies

In the following the distribution properties of anle138b are investigated. So far it is known that anle138b is efficiently absorbed and reaches the brain after a short time. In the following study done with rats, the concentration of anle138b within liver, kidney, spleen, thymus and fat tissue at the timepoints 4, 8 and 24 hours post application was investigated. It can be seen in Fig. 5.9 that anle138b reaches very similar concentration levels in brain, liver, kidney, thymus and spleen; then the concentration declines constantly. Only in liver, where the main metabolism takes place, the maximum is reached at 8 hours before it decreases. It is a striking observation how strong anle138b accumulates in the fat tissue. The concentration in the fat is very high and increases until it reaches a putative saturation that was not further investigated here.

To test the behavior of the metabolite anle138c (M2) over time and to compare its concentration in blood plasma with anle138b, time points during 24 hours were measured from mouse blood plasma. In comparison to the other metabolites, the concentration of anle138c in the liver is not very high. But if seen on a longer timescale of 24 hours, it becomes clear that the concentration of anle138c is constantly increasing while it is metabolized from anle138b. Anle138b is in turn decreasing in concentration. This can be seen in Fig. 5.10.

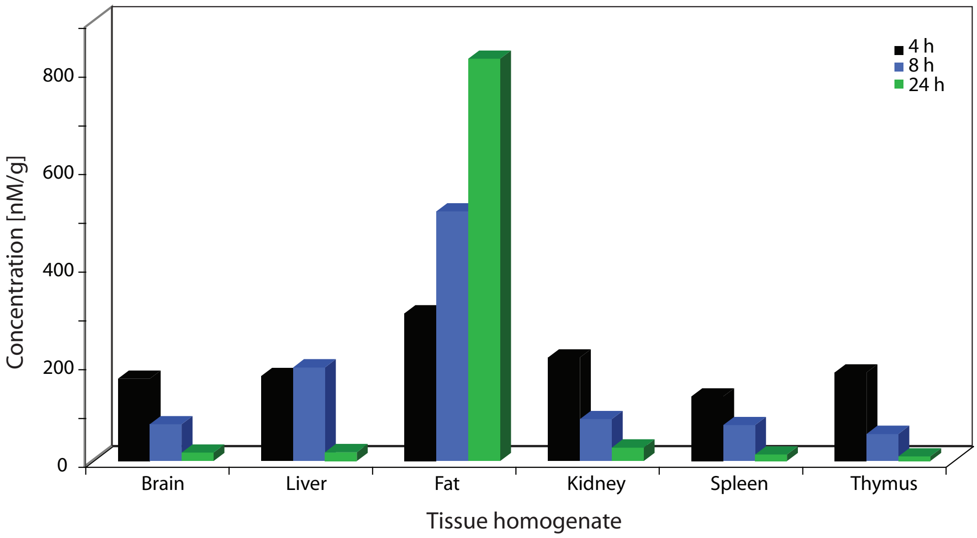


Figure 5.9: Concentration of anle138b in different rat tissue homogenates after 4, 8 and 24 hours. The compound leaves the organs after 4 hours, but accumulates in the fat tissue.

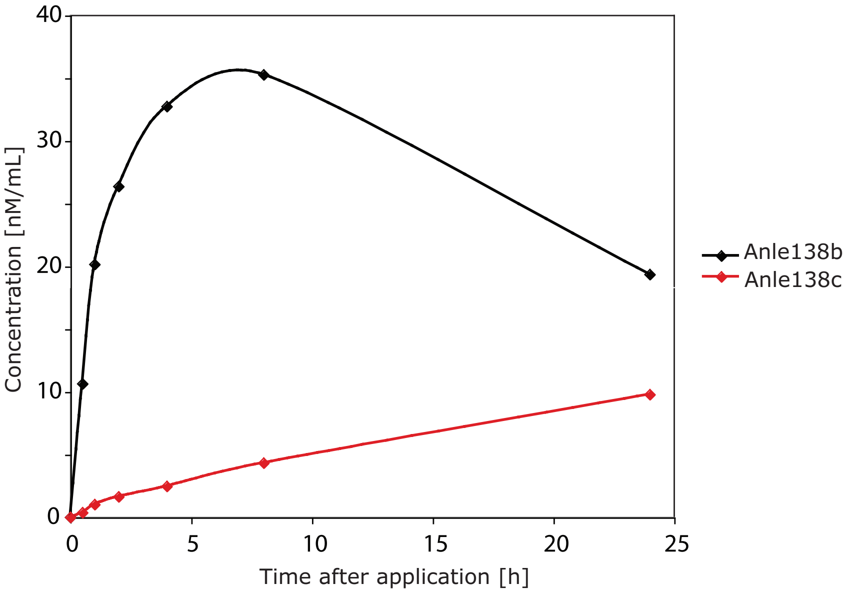


Figure 5.10: Concentration of anle138b and the metabolite anle138c in rat blood plasma at different timepoints during 24 hours. Parallel to the decrease of anle138b, the concentration of anle138c increases.

5.4 Discussion & Conclusion

The metabolism of the toxic protein aggregate modulator anle138b was intensively investigated. During this pharmacokinetic studies many informations on the ADME properties were determined. Anle138b was found to be absorbed efficiently and is distributed to all organs. It passes the blood-brain-barrier and is highly concentrated in the brain. Metabolism takes place in the liver and kidney, where four metabolites were identified and their chemical structures were elucidated. Anle138b gets demethylenated during or hydroxylated during phase I metabolism by CYP450 enzymes. During phase II metabolism the demethylenated metabolite, anle138c, gets sulfurylated or methylated. Anle138b is metabolized and excreted from the body and no anle138b is detected in blood and organs after 24 hours. An exception is the fat tissue, were anle138b reaches high concentrations (Fig. 5.9) and a putative saturation level, which was not further investigated here. A summary of the elucidated metabolism pathway in shown in Fig. 5.11.

There are several possibilities how the studies on anle138b pharmacokinetics could continue. First it should be investigated with CYP450 assays, which CYP450 is active in which transformation. Based on data for MDMA metabolism it is assumed that CYP2D6 is doing the demethylenation [Tucker et al., 1994, Kreth et al., 2000], but this should be further investigated. Metabolite M3 should be synthesized to compare the NMR spectra and HPLC chromatogram of the synthetic compound with the metabolite to make sure in which position the bromine ring gets hydroxylated. If all metabolites could be synthesized, it would be advantageous to test them in animal experiments for their toxicity. In the end, anle138b has to be tested on humans and one would have to see, if the human body produces the same metabolites as rats and mice or different ones which could be toxic. An interesting topic would be the combination of the metabolic investigations with the NMR-based ligand binding mode determination. It could be possible to decipher the binding mode of anle138b to its demethylenation enzyme, e.g. CYP2D6. A docking model of anle138b to the crystal structure of CYP2D6 (PDB code: 2F9Q) [Rowland et al., 2006] done with PLANTS is shown in Fig. 5.12. It proposes two distinct binding modes, that offer an explanation for the demethylenation process, as well as for the hydroxylation site.

Such a model could be refined with NMR data like STD. It would be necessary to investigate how the paramagnetic effects of the iron in the heme group influences the result. It should be mentioned here, that the binding mode of MPTP to its metabolizing enzyme (CYP2D6) was determined with NMR spectroscopy by using T_1 relaxation measurements as distance informations of protons to the heme iron [Modi et al., 1997].

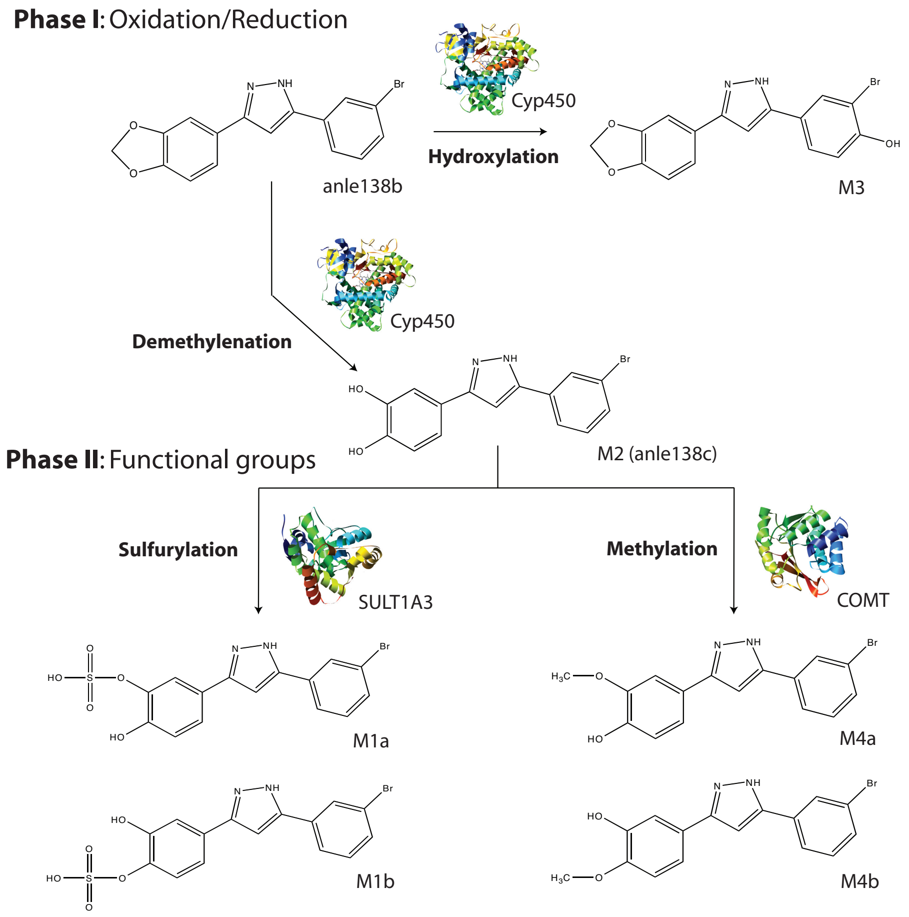

Phase I: Oxidation/Reduction
Cyp450
Hydroxylation
anle138b
M3
Cyp450
Demethylenation
M2 (anle138c)
Phase II: Functional groups
Sulfurylation
SULT1A3
Methylation
COMT
M1a
M4a
M1b
M4b

Figure 5.11: Metabolism and biotransformation pathway of anle138b, as revealed by the presented study. In phase I, anle138b is hydroxylized and demethylenated to anle138c by a CYP450 enzyme. Anle138c gets further sulfurylated by SULT1A3 and methylated by COMT.

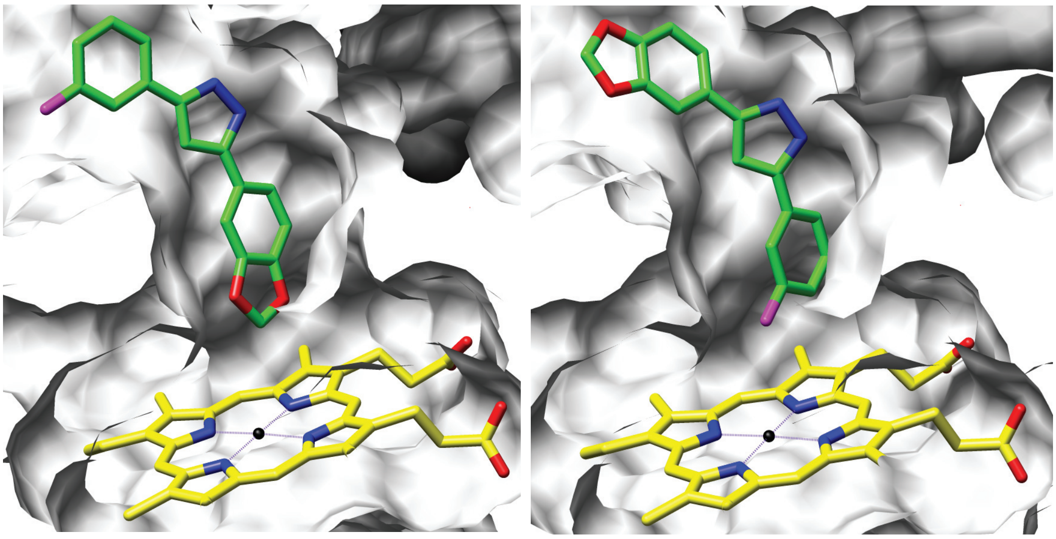


Figure 5.12: Best scoring docking models of anle138b into the crystal structure of CYP2D6. This metabolic enzyme possesses a tunnel-like pathway to the reaction center at the heme group. On the left side the methylene group of anle138b is closest to the reaction center at the heme group. On the right side, the aromatic ring of anle138b is closest to the reaction center with proton H2. This position is thought to become hydroxylated.

Chapter 6

Structure revision of arthrofactin

6.1 Introduction

Biosurfactants are surface active substances, derived from living organisms [Cooper and Zajic, 1980]. Besides numerous industrial applications, biosurfactants are also of high relevance for medical applications. In this respect, particularly the multiple roles of biosurfactants on biofilm formation are of high interest because of the emerging problem of biofilm formation in medical devices and instruments. Knowledge about the underlying mode of action, but also the exact molecular structure of the involved molecules are essential for deeper insights and progress in this research field.

In 1993 arthrofactin was isolated from the bacterium *Pseudomonas* sp. MIS38 (at that time identified as *Arthrobacter* sp.) [Morikawa et al., 1993]. This new cyclic lipopeptide (CLP) showed a capability to lower the surface tension of water from 72 to 24 $\frac{mN}{m}$, making it the most effective biosurfactant today, being 5-7 times more effective than surfactin. Arthrofactin belongs to the family of lipoundecapeptides, together with amphisin, lokisin, tensin and pholipeptin. The biosynthetic gene cluster of arthrofactin encodes the three nonribosomal peptide synthetase (NRPS) subunits ArfA, ArfB and ArfC, in which 11 modules and a terminal tandem thioesterase (TE) are observed [Roongsawang et al., 2003]. In a last step, the terminal thioesterases mediate the cyclization and release of the completed arthrofactin. Almost typical for Pseudomonas-CLP-biosynthesis gene cluster, at the C-terminal end of the NRPS a tandem of two distinct TE domains can be observed [Raaijmakers et al., 2006, Gross and Loper, 2009].

It is crucial to know the configuration and conformation of a potential drug molecule. Unfortunately there are three different configurations and ring closures of arthrofactin in the literature. All of them lack spectroscopic evidence and there was no crystallization of arthrofactin achieved. The first structure was presented together with the identification of arthrofactin and the ring closure was proposed to be between the 11-Asp and a CH atom of the fatty acid part, as in surfactin [Morikawa et al., 1993]. The structure of a very similar CLP, pholipeptin, was solved by NMR and the ring closure was determined to be between the residues Asp-11 and Thr-3, as a Asp-11 β peptide linkage [Ui et al., 1997]. In 2002, the crystal structure of lokisin was solved [Sorensen et al., 2002] and confirmed a ring closure between Asp-11 and Thr-3, but via a Asp-11 α peptide linkage to the *allo*-isomer of threonine. A similar structure was then proposed for arthrofactin during a study of the thioesterase of arthrofactin synthetase [Roongsawang et al., 2007, Washio et al., 2010]. Nevertheless in the meantime a study on the non-ribosomal peptide synthetase (NRPS) assembly line proposed a ring closure between Asp-11 α and the H3 hydroxy group of the decanoic fatty acid side chain, based on the involved enzymes [Balibar et al., 2005].

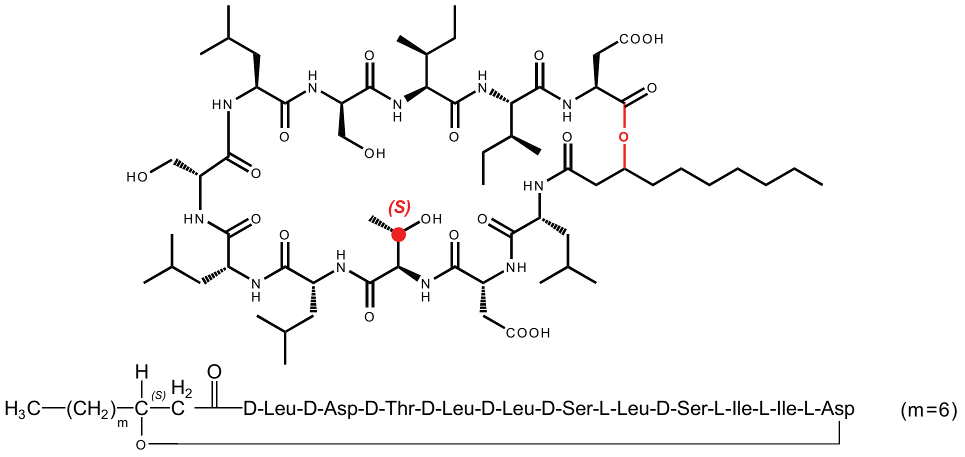

COOH
O
HN
OH
HO
(S)
NH
COOH
H₃C—(CH₂)ₘ—C(H)(S)—C(H₂)—C(=O)—D-Leu-D-Asp-D-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Ser-L-Ile-L-Ile-L-Asp
(m=6)

Figure 6.1: Constitution and configuration of arthrofactin proposed by [Balibar et al., 2005]

A structural revision of arthofactin is done with three independent approaches: i) an intensive NMR investigation, comprising the complete assignment and configurationa and ring closure determination, ii) a phylogenetic tree of peptide synthetase domains of bacteria was done, with a special focus on the thioesterase domain. Placing the *Pseudomonas* strain then into this tree should reveal the nature of the ring closure mechanism and iii) the nature of the amino acids was clarified with HPLC and mass spectrometry.

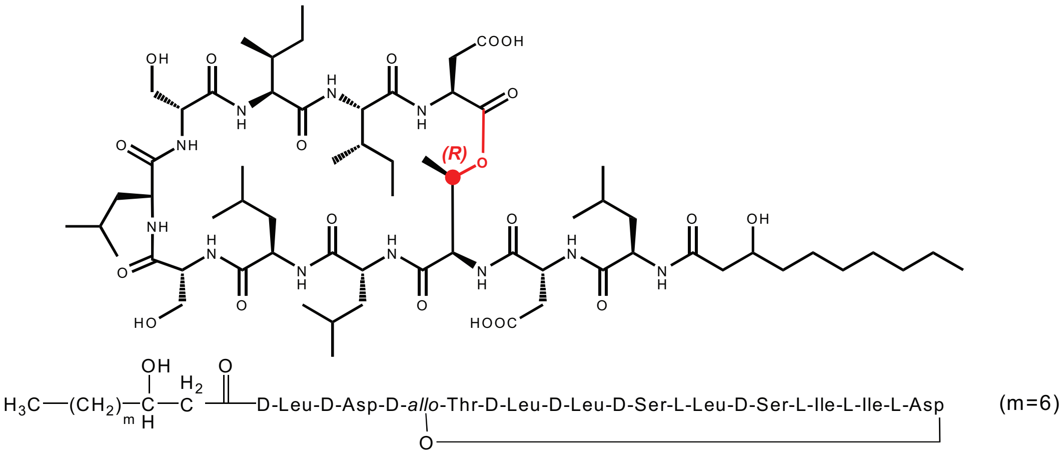

(R)
COOH
HOOC
OH
HO
H3C—(CH2)m—C(OH)H—C H2—C(=O)—D-Leu-D-Asp-D-allo-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Ser-L-Ile-L-Ile-L-Asp
(m=6)

Figure 6.2: Constitution and configuration of arthrofactin proposed by [Roongsawang et al., 2007]

6.2 Materials & Methods

6.2.1 Sample production and bioinformatics

Pseudomonas sp. MIS38 was cultivated in 7.5 L of Davis minimal broth. After acidification to pH = 2, the broth was repetitively extracted with EtOAc. The crude extract (1.1 g) was further fractionated via reversed-phase SPE chromatography and followed by reversed-phase HPLC to afford 11.5 mg of pure arthrofactin and minute amounts of its derivatives.

The amino acid sequences of TE proteins of all currently available cyclic lipopeptides were retrieved from publicly accessible databases (http://www.ncbi.nlm.nih.gov/entrez/). For phylogenetic analyses, alignments were made with ClustalW2, available at http://www.ebi.ac.uk/Tools/msa/clustalw2/. Trees were inferred by Neighbour Joining in ClustalW2.A phylogenetic tree was constructed with all currently published amino acid sequences of 35 TE domains identified in known bacterial and NRPS-derived CLP biosynthesis clusters. Sample production and bioinformatics were done by Anna Lange and Prof. Harald Gross.

6.2.2 NMR spectroscopy

10 mg of arthrofactin were dissolved in 0.5 mL of DMSO-d$_6$. All NMR spectra of arthrofactin were recorded at 25°C in DMSO-d_6 and referenced to the residual solvent signals with resonances at $\delta_{H/C}$ 2.51/40.2. The internal referencing of Bruker was used for ^{15}N chemical shifts.

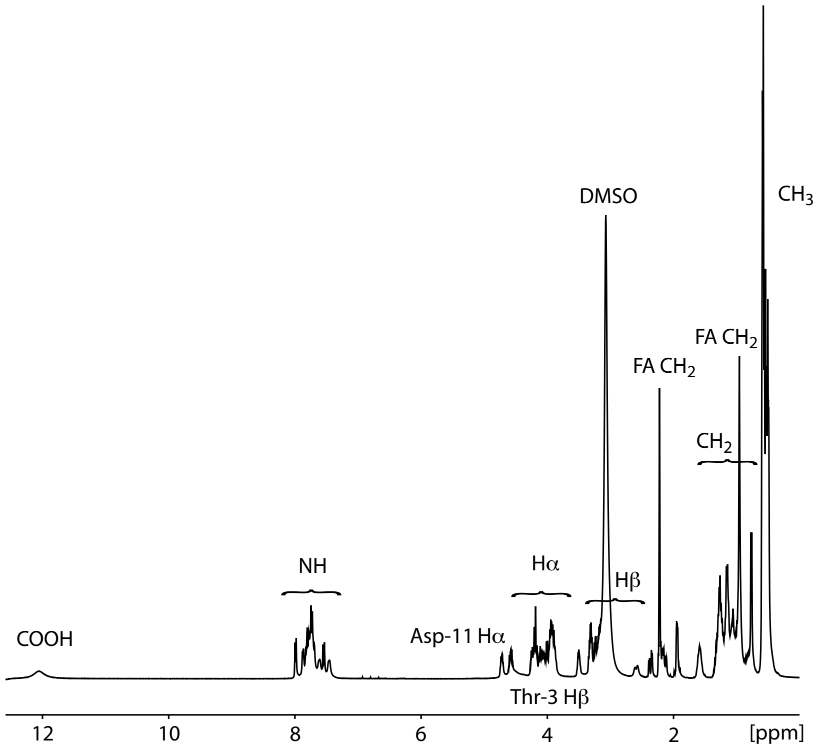


Figure 6.3: 1D spectrum of arthrofactin in DMSO. Indicated are the main functional groups from the peptide and the fatty acid (FA) part.

To allow for correct ordering chemical shifts are listed with ± 0.01 ppm for ^{13}C and ^{15}N and ± 0.001 ppm for ^{1}H, respectively. 1D ^{1}H, 2D ^{1}H-^{1}H-COSY, ^{1}H-^{1}H-NOESY, ^{1}H-^{1}H-TOCSY, ^{1}H-^{13}C-HSQC, ^{1}H-^{15}-HSQC and ^{1}H-^{13}C-HMBC have been recorded using a Bruker 700 MHz spectrometer, and ^{1}H-^{15}N-HMBC with a Bruker 600 MHz spectrometer. Both spectrometers are equipped with a cryogenically cooled triple-resonances probe head. The ^{1}H spectrum was acquired using 16K data points at a spectral width of 7 kHz, with 1 scan and a 2.0 s delay between transients. The data matrix was zero-filled to 32K points and processed with an exponential window function. All homonuclear 2D spectra were measured over 4K complex points in F$_2$, 1024 increments in F$_1$ for TOCSY, 800 for COSY and 600 for NOESY, collecting 8 (TOCSY) or 160 (COSY and NOESY) scans per increment with a relaxation delay of 2.0 s. The spectral widths in the F$_2$ and F$_1$ dimensions were both 6.3 kHz for all homonuclear spectra. The mixing time of TOCSY and NOESY experiments was set to 80 ms and 150 ms, respectively. The spectra width of the heteronuclear spectra in the F$_2$ and F$_1$ dimensions were 4.2 kHz and 3.6 kHz for ^{1}H-^{13}C-HSQC, 1.4 kHz and 1.6 kHz for ^{1}H-^{15}N-HSQC, 6.3 kHz and 35.2 kHz for ^{1}H-^{13}C-HMBC, 6.3 kHz and 1.6 kHz for ^{1}H-^{15}N-HMBC respectively. 4K x 1K data points were acquired with 16 scans per increment for the ^{1}H-^{13}C-HSQC experiment, 4K x 320 data points with 64 scans for the ^{1}H-^{15}N-HSQC experiment, 8K x 1K data points with 48 scans for the ^{1}H-^{13}C-HMBC experiment and 4K x 114 data points with 1000 scans for the ^{1}H-^{15}N-HMBC experiment. In the ^{1}H-^{13}C-HMBC and ^{1}H-^{15}N-HMBC sequence, the delay for the evolution of long-range couplings was set to 8.0 Hz and 6.0 Hz, respectively.

6.3 Results

6.3.1 Bioinformatic studies

The first TE domain encoded by the arthrofactin biosynthesis gene cluster clearly groups with subclass II (Fig. 6.4) as determined by the phylogenetic analysis. Therefore the structure of arthrofactin should be cyclized between Asp-11 and Thr-3, instead of lactone ring formation between Asp-11 and the H3 hydroxy group of the decanoic fatty acid side chain as originally suggested. This is the clear proposal from bioinformatics that will be tested in the following against experimental data.

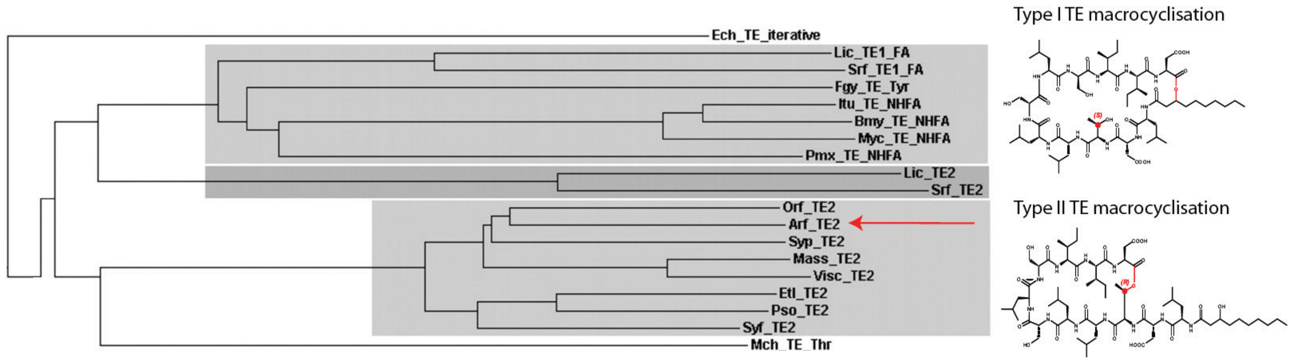

Type I TE macrocyclisation
Ech_TE_iterative
Lic_TE1_FA
Srf_TE1_FA
Fgy_TE_Tyr
Itu_TE_NHFA
Bmy_TE_NHFA
Myc_TE_NHFA
Pmx_TE_NHFA
Lic_TE2
Srf_TE2
Type II TE macrocyclisation
Orf_TE2
Arf_TE2
Syp_TE2
Mass_TE2
Visc_TE2
Etl_TE2
Pso_TE2
Syf_TE2
Mch_TE_Thr
(S)
(R)

Figure 6.4: Phylogenetic tree of the tandem thioesterase domain gene cluster of 35 amino acid sequences from literature. AF-TE2 that encodes for arthrofactin plots in the group TE2 with a macrocyclisation via amino acids, unlike TE1, where the macrocyclisation takes place via the fatty acid. Therefore a macrocyclisation via amino acids is proposed by bioinformatics. This tree was simplified for clarity.

6.3.2 NMR assignment

The assignment was started from the Hβ protons of serine, because they have very characteristic chemical shifts and appear in pairs in the ^{1}H-^{13}C-HSQC spectrum. In arthrofactin two serine residues are present, which are separated by one leucine. With the help of ^{1}H-^{1}H-COSY and ^{1}H-^{1}H-TOCSY experiments, the Hα and NH protons of each serine can be assigned with the knowledge of the chemical shifts of the Hβ protons. However, which serine belongs to which serine in the structure formula still remained unknown. We started at the Hα proton with the chemical shift of 4.536 ppm. In the ^{1}H-^{1}H-NOESY spectrum it shows two cross-peaks with the protons in the amide region. One cross-peak belongs to the intra-residue connection of the NH and Hα protons. Since the NH of serine was assigned, the other proton could be immediately assigned as the NH proton of the neighbored residue. Subsequently, we assigned the corresponding Hα from the COSY spectrum and this Hα proton shows a very strong peak with the NH proton of the second serine residue in the NOESY spectrum. With this procedure we identified the serine that we have started with: it is Ser-6 and the second one is Ser-8. The assignment of the backbone protons of the remaining residues was performed in the same way using a combination of TOCSY, NOESY and COSY spectra. The chemical shifts of the carbonyl- and nitrogen-atoms were assigned with HMBC experiments, and the chemical shifts of carbons were directly assigned from the HSQC spectrum (Fig. 6.5).

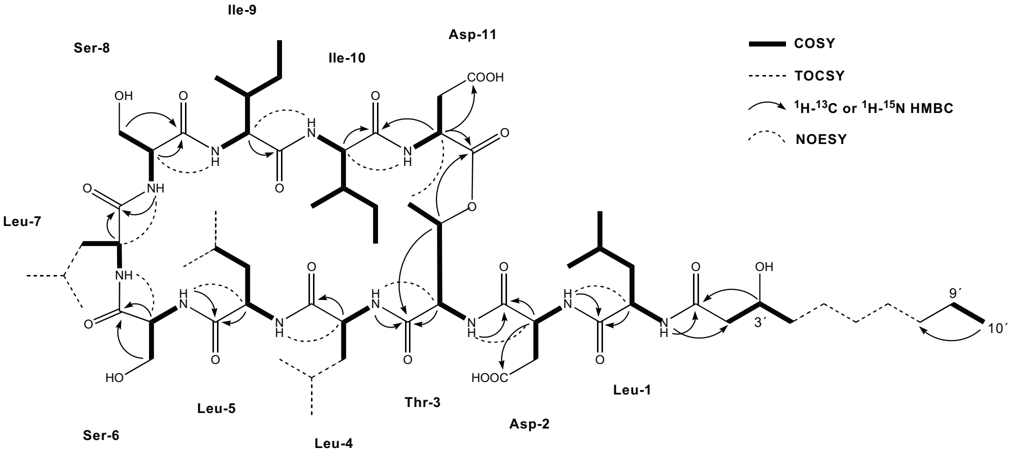


Figure 6.5: Key correlations used for the NMR assignment of arthrofactin.

The closure of the cyclic peptide between Thr-3 and Asp-11 was proven by the following two evidences: (i) the Asp-11 Hα proton shows a HMBC cross-peak with the carbonyl carbon of Asp-11 as well as the Thr-3 Hβ proton

(Fig. 6.6), ii) the Asp-11 Hα and Thr-3 Hβ show NOESY crosspeaks (Fig. 6.7) and (iii) the Thr-3 Hγ shows a NOESY peak with the Asp-11 Hα (Fig. 6.9). Therefore, the closure of the ring must be situated here, between the Asp-11 carbonyl group and the Thr-3 hydroxyl group. The sequence was further verified using ^{1}H-^{15}N HMBC derived connectivities. The closure of the lactone ring in the old formula was disproven by the absence of decisive cross peaks: H-3 of the fatty acid to Asp-11 CO (^{1}H-^{13}C HMBC) and H-3 to Asp-11 Hα (NOESY). The presence of a NOESY crosspeak of Asp-11 Hα to Thr-3 Hβ and the absence of a crosspeak of the Hβ methylene group protons of Asp-11 with Thr-3 Hβ indicated that no isoaspartyl linkage is present. Likewise, we could rule out the beta-linked isomer for Asp-2.

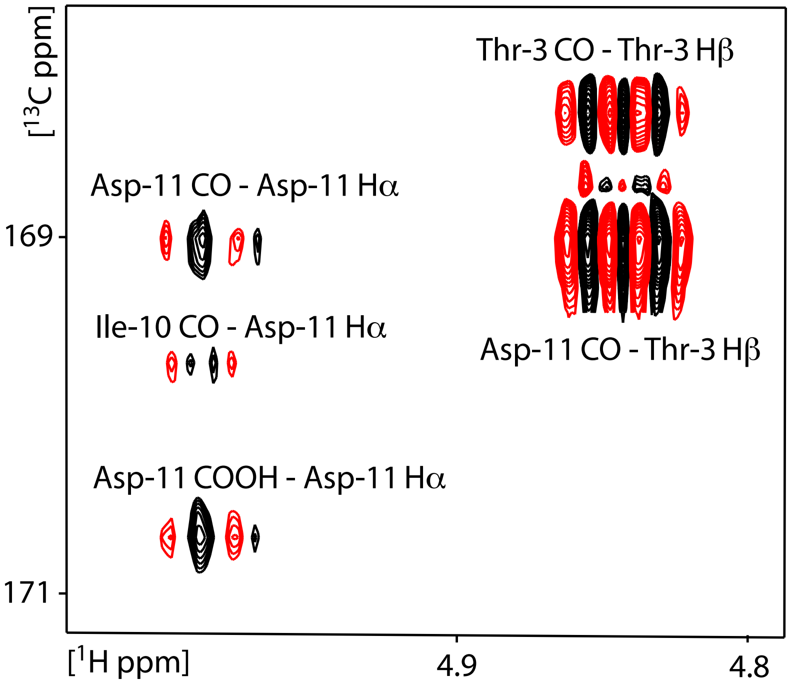


Figure 6.6: ^{1}H-^{13}C-HMBC spectrum of arthrofactin in DMSO. It can be seen that the Asp-11 CO carbon atom is coupled with the Asp-11 Hα, as well as Thr-3 Hβ, indicating a ring closure between Asp-11 and Thr-3.

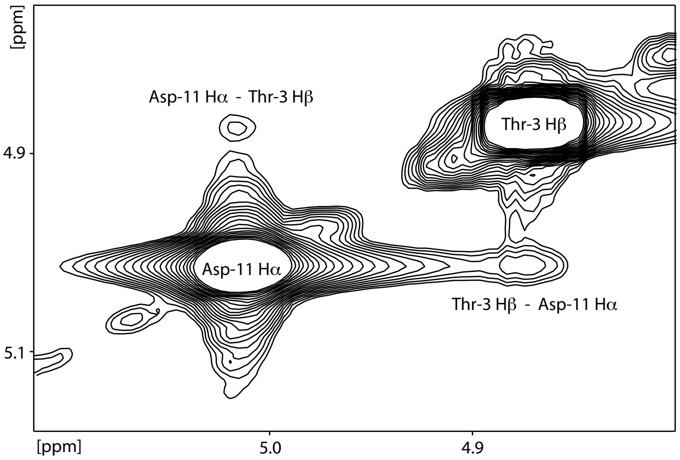


Figure 6.7: NOESY spectrum of arthrofactin in DMSO. The NOESY peaks between Thr-3 Hβ and Asp-11 Hα indicate a ring closure between Asp-11 and Thr-3.

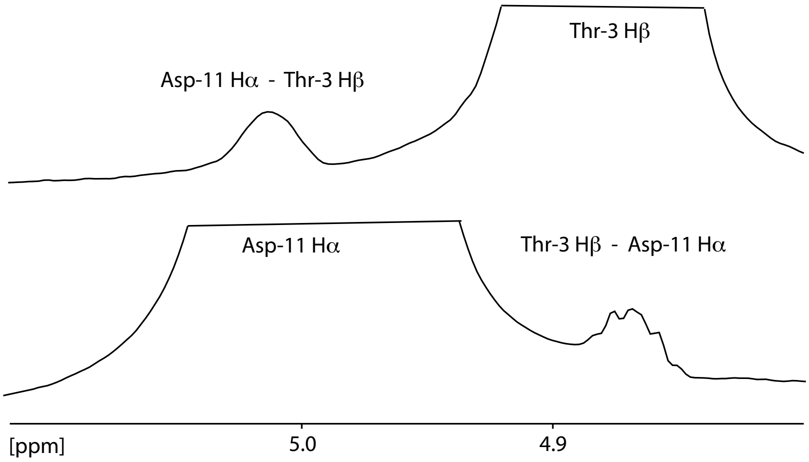


Figure 6.8: Slices from the NOESY spectrum shown above to illustrate that the cross peaks are not artifacts from T1 noise. Diagonal peaks are cut for clarity.

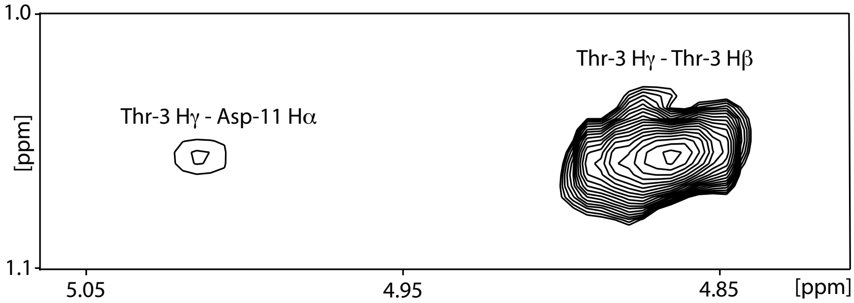


Figure 6.9: NOESY spectrum of arthrofactin in DMSO. The NOESY peak between Thr-3 Hγ and Asp-11 Hα indicates a ring closure between Asp-11 and Thr-3.

6.4 Discussion & Conclusion

In summary, using a bioinformatics approach in combination with spectral experiments, the revised structure of arthrofactin has been determined, as depicted in figure 6.2. Arthrofactin has been shown to possess a cyclization scheme between an D-*allo*-Thr and an Asp, as determined by the NMR experiments HMBC and NOESY. The presence of D-*allo*-Thr was confirmed by chiral HPLC and mass spectrometry. All TE domains of available CLP gene cluster were systematized and propose a phylogenetic model which allows the prediction of the regioselectivity regarding the macrocyclization process during CLP biosynthesis. The data presented here show also the progress of genetic sequence-based prediction of NRPS-derived natural products, which reached already a significant level of accuracy and reliability. This bioinformatic prediction is verified by intensive NMR studies, including the full assignment of arthrofactin and comprising clear evidence for the regioselectivity of the ring closure. To further optimize the methodology other natural product structures should be predicted by bioinformatics in the future. These have to be validated with NMR spectroscopy and mass spectrometry. Still, it is imaginable that the methodology can be optimized to a degree, where experimental support can be reduced to a minimum. This is comparable to the de novo determination of protein structures, just based on sequence alignment. The data and interpretation described in this chapter is published [Lange et al., 2012].

Bibliography

[Adams and Brantner, 2006] Adams, C. P. and Brantner, V. V. (2006). Estimating the cost of new drug development: is it really 802 million dollars? *Health Aff (Millwood)*, 25(2):420–428.

[Ahlquist, 1948] Ahlquist, R. P. (1948). A study of the adrenotropic receptors. *Am J Physiol*, 153(3):586–600.

[Amos, 2011] Amos, L. A. (2011). What tubulin drugs tell us about microtubule structure and dynamics. *Semin Cell Dev Biol*, 22(9):916–926.

[Andreu and Barasoain, 2001] Andreu, J. M. and Barasoain, I. (2001). The interaction of baccatin iii with the taxol binding site of microtubules determined by a homogeneous assay with fluorescent taxoid. *Biochemistry*, 40(40):11975–11984.

[Asano et al., 1998] Asano, T., Ikegaki, I., Satoh, S., Seto, M., and Sasaki, Y. (1998). A protein kinase inhibitor, fasudil (at-877): A novel approach to signal transduction therapy. *Cardiovascular Drug Reviews*, 16(1):76–87.

[Audet and Bouvier, 2008] Audet, M. and Bouvier, M. (2008). Insights into signaling from the beta2-adrenergic receptor structure. *Nat Chem Biol*, 4(7):397–403.

[Backes et al., 2008] Backes, A. C., Zech, B., Felber, B., Klebl, B., and Müller, G. (2008). Small molecule inhibitors binding to protein kinases. part i: exceptions from the traditional pharmacophore approach of type i inhibition. *Expert Opin. Drug Discov.*, 3(12):1409–1425.

[Balaram et al., 1973] Balaram, P., Bothner-By, A. A., and Breslow, E. (1973). Nuclear magnetic resonance studies of the interaction of peptides and hormones with bovine neurophysin. *Biochemistry*, 12(23):4695–4704.

[Balaram et al., 1972] Balaram, P., Bothner-By, A. A., and Dadok, J. (1972). Negative nuclear overhauser effects as probes of macromolecular structure. *J Am Chem Soc*, 94(11):4015–4017.

[Balibar et al., 2005] Balibar, C. J., Vaillancourt, F. H., and Walsh, C. T. (2005). Generation of d amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol*, 12(11):1189–1200.

[Bartoschek et al., 2001] Bartoschek, S., Johansson, M., Geierstanger, B. H., Okun, J. G., Lancaster, C. R., Humpfer, E., Yu, L., Yu, C. A., Griesinger, C., and Brandt, U. (2001). Three molecules of ubiquinone bind specifically to mitochondrial cytochrome bc1 complex. *J Biol Chem*, 276(38):35231–35234.

[Bartoschek et al., 2010] Bartoschek, S., Klabunde, T., Defossa, E., Dietrich, V., Stengelin, S., Griesinger, C., Carlomagno, T., Focken, I., and Wendt, K. U. (2010). Drug design for g-protein-coupled receptors by a ligand-based nmr method. *Angew Chem Int Ed Engl*, 49(8):1426–1429.

[Becattini et al., 2006] Becattini, B., Culmsee, C., Leone, M., Zhai, D., Zhang, X., Crowell, K. J., Rega, M. F., Landshamer, S., Reed, J. C., Plesnila, N., and Pellecchia, M. (2006). Structure-activity relationships by interligand noe-based design and synthesis of antiapoptotic compounds targeting bid. *Proc Natl Acad Sci U S A*, 103(33):12602–12606.

[Berman et al., 2000] Berman, H., Westbrok, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucl Acids Res*, 28(1):235–242.

[Bernal and Crowfoot, 1934] Bernal, J. D. and Crowfoot, D. (1934). X-ray photographs of crystalline pepsin. *Nature*, 133:794–795.

[Bertsch et al., 2005] Bertsch, U., Winklhofer, K. F., Hirschberger, T., Bieschke, J., Weber, P., Hartl, F. U., Tavan, P., Tatzelt, J., Kretzschmar, H. A., and Giese, A. (2005). Systematic identification of antiprion drugs by high-throughput screening based on scanning for intensely fluorescent targets. *J Virol*, 79(12):7785–7791.

[Blundell et al., 2002] Blundell, T. L., Jhoti, H., and Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov*, 315(4):677–685.

[Bokoch et al., 2010] Bokoch, M. P., Zou, Y., Rasmussen, S. G., Liu, C. W., Nygaard, R., Rosenbaum, D. M., Fung, J. J., Choi, H. J., Thian, F. S., Kobilka, T. S., Puglisi, J. D., Weis, W. I., Pardo, L., Prosser, R. S., Mueller, L., and Kobilka, B. K. (2010). Ligand-specific regulation of the extracellular surface of a g-protein-coupled receptor. *Nature*, 463(7277):108–112.

[Bossemeyer et al., 1993] Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., and Huber, R. (1993). 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):849–859.

[Buch et al., 2011] Buch, I., Giorgino, T., and De Fabritiis, G. (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci U S A*, 108(25):10184–10189.

[Buey et al., 2007] Buey, R. M., Calvo, E., Barasoain, I., Pineda, O., Edler, M. C., Matesanz, R., Cerezo, G., Vanderwal, C. D., Day, B. W., Sorensen, E. J., López, J. A., Andreu, J. M., Hamel, E., and Díaz, J. F. (2007). Cyclostreptin binds covalently to microtubule pores and lumenal taxoid binding sites. *Nat Chem Biol*, 3(2):117–125.

[Buey et al., 2004] Buey, R. M., Díaz, J. F., Andreu, J. M., O'Brate, A., Giannakakou, P., Nicolaou, K. C., Sasmal, P. K., Ritzén, A., and Namoto, K. (2004). Interaction of epothilone analogs with the paclitaxel binding site: relationship between binding affinity, microtubule stabilization, and cytotoxicity. *Chem Biol*, 11(2):225–236.

[Canales et al., 2008] Canales, A., Matesanz, R., Gardner, N. M., Andreu, J. M., Paterson, I., Díaz, J. F., and Jiménez-Barbero, J. (2008). The bound conformation of microtubule-stabilizing agents: Nmr insights into the bioactive 3d structure of discodermolide and dictyostatin. *Chemistry*, 14(25):7557–7569.

[Canales et al., 2011] Canales, A., Rodríguez-Salarichs, J., Trigili, C., Nieto, L., Coderch, C., Andreu, J. M., Paterson, I., Jiménez-Barbero, J., and Díaz, J. F. (2011). Insights into the interaction of discodermolide and docetaxel with tubulin. mapping the binding sites of microtubule-stabilizing agents by using an integrated nmr and computational approach. *ACS Chem Biol*, 6(8):789–799.

[Carlomagno et al., 2003] Carlomagno, T., Blommers, M. J., Meiler, J., Jahnke, W., Schupp, T., Petersen, F., Schinzer, D., Altmann, K. H., and Griesinger, C. (2003). The high-resolution solution structure of epothilone a bound to tubulin: an understanding of the structure-activity relationships for a powerful class of antitumor agents. *Angew Chem Int Ed Engl*, 42(22):2511–2515.

[Castagna et al., 1982] Castagna, M., Takai, Y., Kaibuchi, K., Sano, K., Kikkawa, U., and Nishizuka, Y. (1982). Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor-promoting phorbol esters. *J Biol Chem*, 257(13):7847–7851.

[Cavanagh et al., 2007] Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., and Skelton, N. J. (2007). Protein nmr spectroscopy. *Book*, Elsevier.

[Chatterjee et al., 2001] Chatterjee, S. K., Barron, D. M., Vos, S., and Bane, S. (2001). Baccatin iii induces assembly of purified tubulin into long microtubules. *Biochemistry*, 40(23):6964–6970.

[Christiansen et al., 2011] Christiansen, E., Urban, C., Grundmann, M., Due-Hansen, M. E., Hagesaether, E., Schmidt, J., Pardo, L., Ullrich, S., Kostenis, E., Kassack, M., and Ulven, T. (2011). Identification of a potent and selective free fatty acid receptor 1 (ffa1/gpr40) agonist with favorable physicochemical and in vitro adme properties. *J Med Chem*, 54(19):6691–6703.

[Clark, 2006] Clark, D. E. (2006). What has computer-aided molecular design ever done for drug discovery? *Exp Opin Drug Dis*, 1(2):103–110.

[Clarke et al., 2001] Clarke, N. J., Rindgen, D., Korfmacher, W. A., and Cox, K. A. (2001). Systematic lc/ms metabolite identification in drug discovery. *Anal Chem*, 73(15):–439.

[Clore and Gronenborn, 1982] Clore, G. M. and Gronenborn, A. M. (1982). Theory and applications of the transferred nuclear overhauser effect to the study of the conformations of small ligands bound to proteins. *J Magn Reson*, 48:402–417.

[Cohen, 2002] Cohen, P. (2002). Protein kinases–the major drug targets of the twenty-first century? *Nat Rev Drug Discov*, 1(4):309–315.

[Cooper and Zajic, 1980] Cooper, D. G. and Zajic, J. E. (1980). Surface-active compounds from microorganisms. *Adv Appl Microbiol*, 26:229–253.

[Cormier et al., 2008] Cormier, A., Marchand, M., Ravelli, R. B., Knossow, M., and Gigant, B. (2008). Structural insight into the inhibition of tubulin by vinca domain peptide ligands. *EMBO Rep*, 9(11):1101–1106.

[Cortés Cabrera et al., 2012] Cortés Cabrera, Á., Klett, J., Dos Santos, H. G., Perona, A., Gil-Redondo, R., Francis, S. M., Priego, E. M., Gago, F., and Morreale, A. (2012). Crdock: an ultrafast multipurpose protein-ligand docking tool. *J Chem Inf Model*, 52(8):2300–2309.

[Costanzi, 2008] Costanzi, S. (2008). On the applicability of gpcr homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. *J Med Chem*, 51(10):2907–2914.

[Cramer et al., 2012] Cramer, P. E., Cirrito, J. R., Wesson, D. W., Lee, C. Y., Karlo, J. C., Zinn, A. E., Casali, B. T., Restivo, J. L., Goebel, W. D., James, M. J., Brunden, K. R., Wilson, D. A., and Landreth, G. E. (2012). Apoe-directed therapeutics rapidly clear ?-amyloid and reverse deficits in ad mouse models. *Science*, 335(6075):1503–1506.

[Cross et al., 2009] Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., Fan, K. Y., Hu, Y., and Humblet, C. (2009). Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J Chem Inf Model*, 49(6):1455–1474.

[Cruciani et al., 2005] Cruciani, G., Carosati, E., De Boeck, B., Ethirajulu, K., Mackie, C., Howe, T., and Vianello, R. (2005). Metasite: understanding metabolism in human cytochromes from the perspective of the chemist. *J Med Chem*, 48(22):6970–6979.

[Dalvit, 2009] Dalvit, C. (2009). Nmr methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov Today*, 14(21-22):1051–1057.

[Dalvit et al., 2001] Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., and Stockman, B. (2001). Waterlogsy as a method for primary nmr screening: practical aspects and range of applicability. *J Biomol NMR*, 21(4):349–359.

[Davies et al., 2000] Davies, S. P., Reddy, H., Caivano, M., and Cohen, P. (2000). Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem J*, 351(Pt 1):95–105.

[Davis et al., 2003] Davis, A. M., Teague, S. J., and Kleywegt, G. J. (2003). Application and limitations of x-ray crystallographic data in structure-based ligand and drug design. *Angew Chem Int Ed Engl*, 42(24):2718–2736.

[de Groot and Grubmüller, 2001] de Groot, B. L. and Grubmüller, H. (2001). Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and glpf. *Science*, 294(5550):2353–2357.

[Dominguez et al., 2003] Dominguez, C., Boelens, R., and Bonvin, A. M. (2003). Haddock: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, 125(7):1731–1737.

[Drews, 2000] Drews, J. (2000). Drug discovery: A historical perspective. *Science*, 287:1960–1964.

[Druker et al., 1996] Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J., and Lydon, N. B. (1996). Effects of a selective inhibitor of the abl tyrosine kinase on the growth of bcr-abl positive cells. *Nat Med*, 2(5):561–566.

[Dumontet and Jordan, 2010] Dumontet, C. and Jordan, M. A. (2010). Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat Rev Drug Discov*, 9(10):790–803.

[Ehrlich, 1900] Ehrlich, P. (1900). Gesammelte arbeiten. *Springer-Verlag Berlin, 1957*.

[Ehrlich, 1910] Ehrlich, P. (1910). Die behandlung des syphilis mit dem ehrlichschen präparat 606. *Deutsche medizinische Wochenzeitschrift*, pages 1893–1896.

[Engh et al., 1996] Engh, R. A., Girod, A., Kinzel, V., Huber, R., and Bossemeyer, D. (1996). 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):26157–26164.

[Fleming, 1929] Fleming, A. (1929). On the antibacterial action of ciltures of a penicillium, with special reference to their use in the isolation of b. influenzae. *Br J Exp Pathol*, 10(31):226–236.

[Forman et al., 2005] Forman, M. S., Lee, V. M., and Trojanowski, J. Q. (2005). Nosology of parkinson's disease: looking for the way out of a quagmire. *Neuron*, 47(4):479–482.

[Gervasio et al., 2005] Gervasio, F. L., Laio, A., and Parrinello, M. (2005). Flexible docking in solution using metadynamics. *J Am Chem Soc*, 127:2600–2607.

[Giese and Kretzschmar, 2001] Giese, A. and Kretzschmar, H. A. (2001). Prion-induced neuronal damage–the mechanisms of neuronal destruction in the subacute spongiform encephalopathies. *Curr Top Microbiol Immunol*, 253:203–217.

[Görler and Kalbitzer, 1997] Görler, A. and Kalbitzer, H. R. (1997). Relax, a flexible program for the back calculation of noesy spectra based on complete-relaxation-matrix formalism. *J Magn Reson*, 124(1):177–188.

[Grishaev et al., 2008] Grishaev, A., Tugarinov, V., Kay, L. E., Trewhella, J., and Bax, A. (2008). Refined solution structure of the 82-kda enzyme malate synthase g from joint nmr and synchrotron saxs restraints. *J Biomol NMR*, 40(2):95–106.

[Gross and Loper, 2009] Gross, H. and Loper, J. E. (2009). Genomics of secondary metabolite production by pseudomonas spp. *Nat Prod Rep*, 26(11):1408–1446.

[Haga et al., 2012] Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T., and Kobayashi, T. (2012). Structure of the human m2 muscarinic acetylcholine receptor bound to an antagonist. *Nature*, 482(7386):547–551.

[Haupt and Schroeder, 2011] Haupt, V. J. and Schroeder, M. (2011). Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief Bioinform*, 12(4):312–326.

[Henzler-Wildman and Kern, 2007] Henzler-Wildman, K. and Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450(7172):964–972.

[Hess et al., 2008] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theor Comp*, 4:435–447.

[Hidaka et al., 1984] Hidaka, H., Inagaki, M., Kawamoto, S., and Sasaki, Y. (1984). Isoquinolinesulfonamides, novel and potent inhibitors of cyclic nucleotide dependent protein kinase and protein kinase c. *Biochemistry*, 23(21):5036–5041.

[Hillisch et al., 2004] Hillisch, A., Pineda, L. F., and Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug Discov Today*, 9(15):659–669.

[Hofmann et al., 2009] Hofmann, K. P., Scheerer, P., Hildebrand, P. W., Choe, H. W., Park, J. H., Heck, M., and Ernst, O. P. (2009). A g protein-coupled receptor at work: the rhodopsin model. *Trends Biochem Sci*, 34(11):540–552.

[Honndorf et al., 2008] Honndorf, V. S., Coudevylle, N., Laufer, S., Becker, S., and Griesinger, C. (2008). Dynamics in the p38alpha map kinase-sb203580 complex observed by liquid-state nmr spectroscopy. *Angew Chem Int Ed Engl*, 47(19):3548–3551.

[Honndorf et al., 2012] Honndorf, V. S., Coudevylle, N., Laufer, S., Becker, S., Griesinger, C., and Habeck, M. (2012). Inferential nmr/x-ray-based structure determination of a dibenzo[a,d]cycloheptenone inhibitor-p38? map kinase complex in solution. *Angew Chem Int Ed Engl*, 51(10):2359–2362.

[Hopkins and Groom, 2002] Hopkins, A. L. and Groom, C. R. (2002). The druggable genome. *Nat Rev Drug Discov*, 1(9):727–730.

[Hore et al., 2000] Hore, P. J., Jones, J. A., and Wimperis, S. (2000). Nmr: The toolkit. *Oxford Chemistry Press.*

[Hornak et al., 2006] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65:712–725.

[Inooka et al., 2001] Inooka, H., Ohtaki, T., Kitahara, O., Ikegami, T., Endo, S., Kitada, C., Ogi, K., Onda, H., Fujino, M., and Shirakawa, M. (2001). Conformation of a peptide ligand bound to its g-protein coupled receptor. *Nat Struct Biol*, 8(2):161–165.

[Itoh et al., 2003] Itoh, Y., Kawamata, Y., Harada, M., Kobayashi, M., Fujii, R., Fukusumi, S., Ogi, K., Hosoya, M., Tanaka, Y., Uejima, H., Tanaka, H., Maruyama, M., Satoh, R., Okubo, S., Kizawa, H., Komatsu, H., Matsumura, F., Noguchi, Y., Shinohara, T., Hinuma, S., Fujisawa, Y., and Fujino, M. (2003). Free fatty acids regulate insulin secretion from pancreatic beta cells through gpr40. *Nature*, 422(6928):173–176.

[Jahnke, 2002] Jahnke, W. (2002). Spin labels as a tool to identify and characterize protein-ligand interactions by nmr spectroscopy. *Chembiochem*, 3(2-3):167–173.

[Jahnke et al., 2003] Jahnke, W., Flörsheimer, A., Blommers, M. J., Paris, C. G., Heim, J., Nalin, C. M., and Perez, L. B. (2003). Second-site nmr screening and linker design. *Curr Top Med Chem*, 3(1):69–80.

[Jayalakshmi et al., 2004] Jayalakshmi, V., Biet, T., Peters, T., and Krishna, N. R. (2004). Refinement of the conformation of udp-galactose bound to galactosyltransferase using the std nmr intensity-restrained corcema optimization. *J Am Chem Soc*, 126(28):8610–8611.

[Jayalakshmi and Krishna, 2002] Jayalakshmi, V. and Krishna, N. R. (2002). Complete relaxation and conformational exchange matrix (corcema) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J Magn Reson*, 155(1):106–118.

[Ji et al., 2008] Ji, Z., Yao, Z., and Liu, M. (2008). Saturation transfer difference nuclear magnetic resonance study on the specific binding of ligand to protein. *Analytical Biochemistry*, 385:380–382.

[Jones et al., 1995] Jones, G., Willett, P., and Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol*, 245(1):43–53.

[Jordan and Wilson, 2004] Jordan, M. A. and Wilson, L. (2004). Microtubules as a target for anticancer drugs. *Nat Rev Cancer*, 4(4):253–265.

[Kaminsky and Zhang, 1997] Kaminsky, L. S. and Zhang, Z. Y. (1997). Human p450 metabolism of warfarin. *Pharmacol Ther*, 73(1):67–74.

[Keeler, 2010] Keeler, J. (2010). Understanding nmr spectroscopy. *Wiley*, 2nd edition.

[Kendrew et al., 1958] Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666.

[Khrapunovich-Baine et al., 2009] Khrapunovich-Baine, M., Menon, V., Verdier-Pinard, P., Smith, A. B., Angeletti, R. H., Fiser, A., Horwitz, S. B., and Xiao, H. (2009). Distinct pose of discodermolide in taxol binding pocket drives a complementary mode of microtubule stabilization. *Biochemistry*, 48(49):11664–11677.

[Khrapunovich-Baine et al., 2011] Khrapunovich-Baine, M., Menon, V., Yang, C. P., Northcote, P. T., Miller, J. H., Angeletti, R. H., Fiser, A., Horwitz, S. B., and Xiao, H. (2011). Hallmarks of molecular action of microtubule stabilizing agents: effects of epothilone b, ixabepilone, peloruside a, and laulimalide on microtubule conformation. *J Biol Chem*, 286(13):11765–11778.

[Kitchen et al., 2004] Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 3(11):935–949.

[Klabunde and Hessler, 2002] Klabunde, T. and Hessler, G. (2002). Drug design strategies for targeting g-protein-coupled receptors. *Chembiochem*, 3(10):928–944.

[Knighton et al., 1991] Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Xuong, N. H., Taylor, S. S., and Sowadski, J. M. (1991). Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, 253(5018):414–420.

[Koeck and Griesinger, 1994] Koeck, M. and Griesinger, C. (1994). Fast noesy experiments - an approach for fast structure determination. *Angew Chem Int Ed Engl*, 33:332–334.

[Koehn and Carter, 2005] Koehn, F. E. and Carter, G. T. (2005). The evolving role of natural products in drug discovery. *Nat Rev Drug Discov*, 4(3):206–220.

[Kolb et al., 2009] Kolb, P., Rosenbaum, D. M., Irwin, J. J., Fung, J. J., Kobilka, B. K., and Shoichet, B. K. (2009). Structure-based discovery of beta2-adrenergic receptor ligands. *Proc Natl Acad Sci U S A*, 106(16):6843–6848.

[Korb et al., 2010a] Korb, O., Monecke, P., Hessler, G., Stützle, T., and Exner, T. E. (2010a). pharmacophore: multiple flexible ligand alignment based on ant colony optimization. *J Chem Inf Model*, 50(9):1669–1681.

[Korb et al., 2010b] Korb, O., Mšller, H. M., and Exner, T. E. (2010b). Nmr-guided molecular docking of a protein-peptide complex based on ant colony optimization. *ChemMedChem*, 5(7):1001–1006.

[Korb et al., 2009a] Korb, O., Stützle, T., and Exner, T. E. (2009a). Empirical scoring functions for advanced protein-ligand docking with plants. *J Chem Inf Model*, 49(1):84–96.

[Korb et al., 2009b] Korb, O., Stützle, T., and Exner, T. E. (2009b). Empirical scoring functions for advanced protein-ligand docking with plants. *J Chem Inf Model*, 49(1):84–96.

[Korb et al., 2006] Korb, O., Stützle, T., and Exner, T. E. (2006). Plants: Application of ant colony optimization to structure-based drug design. *LNCS*, 4150:247–258.

[Kornev et al., 2006] Kornev, A. P., Haste, N. M., Taylor, S. S., and Eyck, L. F. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci U S A*, 103(47):17783–17788.

[Kreth et al., 2000] Kreth, K., Kovar, K., Schwab, M., and Zanger, U. M. (2000). Identification of the human cytochromes p450 involved in the oxidative metabolism of "ecstasy"-related designer drugs. *Biochem Pharmacol*, 59(12):1563–1571.

[Kubicek et al., 2010] Kubicek, K., Grimm, S. K., Orts, J., Sasse, F., and Carlomagno, T. (2010). The tubulin-bound structure of the antimitotic drug tubulysin. *Angew Chem Int Ed Engl*, 49(28):4809–4812.

[Kubinyi, 2003] Kubinyi, H. (2003). Drug research: myths, hype and reality. *Nat Rev Drug Discov*, 2(8):665–668.

[Kufareva et al., 2011] Kufareva, I., Rueda, M., Katritch, V., Stevens, R. C., Abagyan, R., and GPCR Dock 2010 participants (2011). Status of gpcr modeling and docking as reflected by community-wide gpcr dock 2010 assessment. *Structure*, 19(8):1108–1126.

[Kumar et al., 2010] Kumar, A., Heise, H., Blommers, M. J., Krastel, P., Schmitt, E., Petersen, F., Jeganathan, S., Mandelkow, E. M., Carlomagno, T., Griesinger, C., and Baldus, M. (2010). Interaction of epothilone b (patupilone) with microtubules as detected by two-dimensional solid-state nmr spectroscopy. *Angew Chem Int Ed Engl*, 49(41):7504–7507.

[Kurogi and Güner, 2001] Kurogi, Y. and Güner, O. F. (2001). Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr Med Chem*, 8(9):1035–1055.

[Lahana, 1999] Lahana, R. (1999). How many leads from hts? *Drug Discov Today*, 4(10):447–448.

[Lange et al., 2012] Lange, A., Sun, H., Pilger, J., Reinscheid, U. M., and Gross, H. (2012). Predicting the structure of cyclic lipopeptides by bioinformatics: structure revision of arthrofactin. *Chembiochem*, 13(18):2671–2675.

[Langer and Hoffmann, 2006] Langer, T. and Hoffmann, R. D. (2006). Pharmacophore modelling: applications in drug discovery. *Expert Opin Drug Discov*, 1(3):261–267.

[Langer et al., 2004] Langer, T., Vogtherr, M., Elshorst, B., Betz, M., Schieborr, U., Saxena, K., and Schwalbe, H. (2004). 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.

[Langley, 1905] Langley, J. N. (1905). On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. *J Physiol*, 33(4-5):374–413.

[Leone et al., 2010] Leone, M., Barile, E., Vazquez, J., Mei, A., Guiney, D., Dahl, R., and Pellecchia, M. (2010). Nmr-based design and evaluation of novel bidentate inhibitors of the protein tyrosine phosphatase yoph. *Chem Biol Drug Des*, 76(1):10–16.

[Lepre et al., 2004] Lepre, C. A., Moore, J. M., and Peng, J. W. (2004). Theory and applications of nmr-based screening in pharmaceutical research. *Chem Rev*, 104(8):3641–3676.

[Li et al., 1999] Li, D., DeRose, E. F., and London, R. E. (1999). The interligand overhauser effect: a powerful new nmr approach for mapping structural relationships of macromolecular ligands. *J Biomol NMR*, 15(1):71–76.

[Lin et al., 2011] Lin, D. C., Zhang, J., Zhuang, R., Li, F., Nguyen, K., Chen, M., Tran, T., Lopez, E., Lu, J. Y., Li, X. N., Tang, L., Tonn, G. R., Swaminath, G., Reagan, J. D., Chen, J. L., Tian, H., Lin, Y. J., Houze, J. B., and Luo, J. (2011). Amg 837: a novel gpr40/ffa1 agonist that enhances insulin secretion and lowers glucose levels in rodents. *PLoS One*, 6(11).

[Lipari and Szabo, 1982] Lipari, G. and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *J Am Chem Soc*, 104(17):4546–4559.

[Lipinski et al., 2001] Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, 46(1-3):3–26.

[Lombardino and Lowe, 2004] Lombardino, J. G. and Lowe, J. A. (2004). The role of the medicinal chemist in drug discovery–then and now. *Nat Rev Drug Discov*, 3(10):853–862.

[London, 1999] London, R. E. (1999). Theoretical analysis of the inter-ligand overhauser effect: a new approach for mapping structural relationships of macromolecular ligands. *J Magn Reson*, 141(2):301–311.

[Löwe et al., 2001] Löwe, J., Li, H., Downing, K. H., and Nogales, E. (2001). Refined structure of alpha beta-tubulin at 3.5 a resolution. *J Mol Biol*, 313(5):1045–1057.

[Lu et al., 2010] Lu, S. Y., Jiang, Y. J., Lv, J., Wu, T. X., Yu, Q. S., and Zhu, W. L. (2010). Molecular docking and molecular dynamics simulation studies of gpr40 receptor-agonist interactions. *J Mol Graph Model*, 28(8):766–774.

[Lu et al., 2011] Lu, S. Y., Jiang, Y. J., Zou, J. W., Luo, H. B., and Wu, T. X. (2011). Insight into analysis of interactions of gw9508 to wild-type and h86f and h137f gpr40: a combined qm/mm study and pharmacophore modeling. *J Mol Graph Model*, 29(6):818–825.

[Luca et al., 2003] Luca, S., White, J. F., Sohal, A. K., Filippov, D. V., van Boom, J. H., Grisshammer, R., and Baldus, M. (2003). The conformation of neurotensin bound to its g protein-coupled receptor. *Proc Natl Acad Sci U S A*, 100(19):10706–10711.

[Ludwig et al., 2008a] Ludwig, C., Michiels, P. J., Wu, X., Kavanagh, K. L., Pilka, E., Jansson, A., Oppermann, U., and Günther, U. L. (2008a). Salmon: solvent accessibility, ligand binding, and mapping of ligand orientation by nmr spectroscopy. *J Med Chem*, 51(1):1–3.

[Ludwig et al., 2008b] Ludwig, C., Michiels, P. J. A., Lodi, A., Ride, J., Bunce, C., and Günther, U. L. (2008b). Evaluation of solvent accessibility epitopes for different dehydrogenase inhibitors. *Chem Med Chem*, 3:1371–1376.

[Macarron et al., 2011] Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., Schopfer, U., and Sittampalam, G. S. (2011). Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov*, 10(3):188–195.

[Magnani et al., 2009] Magnani, M., Maccari, G., Andreu, J. M., Díaz, J. F., and Botta, M. (2009). Possible binding site for paclitaxel at microtubule pores. *FEBS J*, 276(10):2701–2712.

[Manning et al., 2002] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934.

[Mayer and Meyer, 1999] Mayer, M. and Meyer, B. (1999). Characterization of ligand binding by saturation transfer difference nmr spectroscopy. *Angew Chem Int Ed Engl*, 38:1784–1788.

[Mayer and Meyer, 2001] Mayer, M. and Meyer, B. (2001). 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*, 123:6108–6117.

[Meiler and Baker, 2006] Meiler, J. and Baker, D. (2006). Rosettaligand: protein-small molecule docking with full side-chain flexibility. *Proteins*, 65(3):538–548.

[Meyer and Peters, 2003] Meyer, B. and Peters, T. (2003). Nmr spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl*, 42(8):864–890.

[Michino et al., 2009] Michino, M., Abola, E., GPCR Dock 2008 participants, Brooks, C. L., Dixon, J. S., Moult, J., and Stevens, R. C. (2009). Community-wide assessment of gpcr structure modelling and ligand docking: Gpcr dock 2008. *Nat Rev Drug Discov*, 8(6):455–463.

[Mizukoshi et al., 2012] Mizukoshi, Y., Abe, A., Takizawa, T., Hanzawa, H., Fukunishi, Y., Shimada, I., and Takahashi, H. (2012). An accurate pharmacophore mapping method by nmr spectroscopy. *Angew Chem Int Ed Engl*, 51(6):1362–1365.

[Modi et al., 1997] Modi, S., Gilham, D. E., Sutcliffe, M. J., Lian, L. Y., Primrose, W. U., Wolf, C. R., and Roberts, G. C. (1997). 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine as a substrate of cytochrome p450 2d6: allosteric effects of nadph-cytochrome p450 reductase. *Biochemistry*, 36(15):4461–4470.

[Mohr et al., 2013] Mohr, K., Schmitz, J., Schrage, R., Tränkle, C., and Holzgrabe, U. (2013). Molecular alliance-from orthosteric and allosteric ligands to dualsteric/bitopic agonists at g?protein coupled receptors. *Angew Chem Int Ed Engl*, 52(2):508–516.

[Moitessier et al., 2008] Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., and Corbeil, C. R. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol*, 153 Suppl 1:7–26.

[Morikawa et al., 1993] Morikawa, M., Daido, H., Takao, T., Murata, S., Shimonishi, Y., and Imanaka, T. (1993). A new lipopeptide biosurfactant produced by arthrobacter sp. strain mis38. *J Bacteriol*, 175(20):6459–6466.

[Morin, 2000] Morin, M. J. (2000). From oncogene to drug: development of small molecule tyrosine kinase inhibitors as anti-tumor and anti-angiogenic agents. *Oncogene*, 19(56):6574–6583.

[Negoro et al., 2010] Negoro, N., Sasaki, S., Mikami, S., Ito, M., Suzuki, M., Tsujihata, Y., Ito, R., Harada, A., Takeuchi, K., Suzuki, N., Miyazaki, J., Santou, T., Odani, T., Kanzaki, N., Funami, M., Tanaka, T., Kogame, A., Matsunaga, S., Yasuma, T., and Momose, Y. (2010). Discovery of tak-875: A potent, selective, and orally bioavailable gpr40 agonist. *ACS Med Chem Lett*, 1(6):290–294.

[Nettles et al., 2004] Nettles, J. H., Li, H., Cornett, B., Krahn, J. M., Snyder, J. P., and Downing, K. H. (2004). The binding mode of epothilone a on alpha, beta-tubulin by electron crystallography. *Science*, 305(5685):866–869.

[Ni and Zhu, 1994] Ni, F. and Zhu, Y. (1994). Accounting for the ligand-protein interactions in the relaxation-matrix analysis of transferred nuclear overhauser effects. *J Magn Reson*, 102:180–184.

[Nicolaou et al., 1997] Nicolaou, K. C., Winssinger, N., Pastor, J., Ninkovic, S., Sarabia, F., He, Y., Vourloumis, D., Yang, Z., Li, T., Giannakakou, P., and Hamel, E. (1997). Synthesis of epothilones a and b in solid and solution phase. *Nature*, 387(6630):268–272.

[O'Boyle et al., 2009] O'Boyle, N. M., Liebeschuetz, J. W., and Cole, J. C. (2009). Testing assumptions and hypotheses for rescoring success in protein-ligand docking. *J Chem Inf Model*, 49(8):1871–1878.

[Orts et al., 2009] Orts, J., Griesinger, C., and Carlomagno, T. (2009). The inpharma technique for pharmacophore mapping: A theoretical guide to the method. *J Magn Reson*, 200(1):64–73.

[Orts et al., 2008a] Orts, J., Grimm, S. K., Griesinger, C., Wendt, K. U., Bartoschek, S., and Carlomagno, T. (2008a). Specific methyl group protonation for the measurement of pharmacophore-specific interligand noe interactions. *Chemistry - A European Journal*, 14(25):7517–7520.

[Orts et al., 2008b] Orts, J., Tuma, J., Reese, M., Grimm, S. K., Monecke, P., Bartoschek, S., Schiffer, A., Wendt, K. U., Griesinger, C., and Carlomagno, T. (2008b). Crystallography-independent determination of ligand binding modes. *Angew Chem Int Ed Engl*, 47(40):7736–7740.

[Overington et al., 2006] Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006). How many drug targets are there? *Nat Rev Drug Discov*, 5(12):993–996.

[Palczewski et al., 2000] Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000). Crystal structure of rhodopsin: A g protein-coupled receptor. *Science*, 289(5480):739–745.

[Park et al., 2008] Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W., and Ernst, O. P. (2008). Crystal structure of the ligand-free g-protein-coupled receptor opsin. *Nature*, 454(7201):183–187.

[Pellecchia et al., 2008] Pellecchia, M., Bertini, I., Cowburn, D., Dalvit, C., Giralt, E., Jahnke, W., James, T. L., Homans, S. W., Kessler, H., Luchinat, C., Meyer, B., Oschkinat, H., Peng, J., Schwalbe, H., and Siegal, G. (2008). Perspectives on nmr in drug discovery: a technique comes of age. *Nat Rev Drug Discov*, 7(9):738–745.

[Pereira et al., 2009] Pereira, A., Pfeifer, T. A., Grigliatti, T. A., and Andersen, R. J. (2009). Functional cell-based screening and saturation transfer double-difference nmr have identified haplosamate a as a cannabinoid receptor agonist. *ACS Chem Biol*, 4(2):139–144.

[Podjarny et al., 2011] Podjarny, A., Dejaegere, A. P., and Kieffer, B., editors (2011). *Biophysical approaches determining ligand binding to biomolecular targets: detection, measurement and modelling.* RSC Publishing.

[Poppe et al., 2007] Poppe, L., Harvey, T. S., Mohr, C., Zondlo, J., Tegley, C. M., Nuanmanee, O., and Cheetham, J. (2007). Discovery of ligands for nurr1 by combined use of nmr screening with different isotopic and spin-labeling strategies. *J Biomol Screen*, 12(3):301–311.

[Prota et al., 2013] Prota, A. E., Bargsten, K., Zurwerra, D., Field, J. J., Díaz, J. F., Altmann, K. H., and Steinmetz, M. O. (2013). Molecular mechanism of action of microtubule-stabilizing anticancer agents. *Science*, 339(6119):587–590.

[Raaijmakers et al., 2006] Raaijmakers, J. M., de Bruijn, I., and de Kock, M. J. (2006). Cyclic lipopeptide production by plant-associated pseudomonas spp.: diversity, activity, biosynthesis, and regulation. *Mol Plant Microbe Interact*, 19(7):699–710.

[Rasmussen et al., 2007] Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I., and Kobilka, B. K. (2007). Crystal structure of the human beta2 adrenergic g-protein-coupled receptor. *Nature*, 450(7168):383–387.

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

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

[Regueiro-Ren et al., 2002] Regueiro-Ren, A., Leavitt, K., Kim, S. H., Höfle, G., Kiffe, M., Gougoutas, J. Z., DiMarco, J. D., Lee, F. Y., Fairchild, C. R., Long, B. H., and Vite, G. D. (2002). Sar and ph stability of cyano-substituted epothilones. *Org Lett*, 4(22):3815–3818.

[Roongsawang et al., 2003] Roongsawang, N., Hase, K., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S. (2003). Cloning and characterization of the gene cluster encoding arthrofactin synthetase from pseudomonas sp. mis38. *Chem Biol*, 10(9):869–880.

[Roongsawang et al., 2007] Roongsawang, N., Washio, K., and Morikawa, M. (2007). In vivo characterization of tandem c-terminal thioesterase domains in arthrofactin synthetase. *Chembiochem*, 8(5):501–512.

[Rosenbaum et al., 2009] Rosenbaum, D. M., Rasmussen, S. G., and Kobilka, B. K. (2009). The structure and function of g-protein-coupled receptors. *Nature*, 459(7245):356–363.

[Rowland et al., 2006] Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Lewis, C. J., Tennant, M. G., Modi, S., Eggleston, D. S., Chenery, R. J., and Bridges, A. M. (2006). Crystal structure of human cytochrome p450 2d6. *J Biol Chem*, 281(11):7614–7622.

[Sadowski and Gasteiger, 1994] Sadowski, J. and Gasteiger, J. (1994). Comparison of automatic three-dimensional model builder using 639 x-ray structures. *J. Chem. Inf. Comput. Sci.*, 34:1000–1008.

[Sanchez-Pedregal et al., 2006] Sanchez-Pedregal, V. M., Kubicek, K., Meiler, J., Lyothier, I., Paterson, I., and Carlomagno, T. (2006). The tubulin-bound conformation of discodermolide derived by nmr studies in solution supports a common pharmacophore model for epothilone and discodermolide. *Angew Chem Int Ed Engl*, 45(44):7388–7394.

[Sánchez-Pedregal et al., 2005] Sánchez-Pedregal, V. M., Reese, M., Meiler, J., Blommers, M. J., Griesinger, C., and Carlomagno, T. (2005). The inpharma method: protein-mediated interligand noes for pharmacophore mapping. *Angew Chem Int Ed Engl*, 44(27):4172–4175.

[Schuster et al., 2005] Schuster, D., Laggner, C., and Langer, T. (2005). Why drugs fail–a study on side effects in new chemical entities. *Curr Pharm Des*, 11(27):3545–3559.

[Senderowicz, 2000] Senderowicz, A. M. (2000). Small molecule modulators of cyclin-dependent kinases for cancer therapy. *Oncogene*, 19(56):6600–6606.

[Shaw et al., 2010] Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346.

[Shuker et al., 1996] Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. (1996). Discovering high-affinity ligands for proteins: Sar by nmr. *Science*, 274(5292):1531–1534.

[Sledz et al., 2010] Sledz, P., Silvestre, H. L., Hung, A. W., Ciulli, A., Blundell, T. L., and Abell, C. (2010). Optimization of the interligand overhauser effect for fragment linking: application to inhibitor discovery against mycobacterium tuberculosis pantothenate synthetase. *J Am Chem Soc*, 132(13):4544–4545.

[Solomon, 1955] Solomon, I. (1955). Relaxation processes in a system of two spins. *Phys. Rev.*, 99(2):559–565.

[Sorensen et al., 2002] Sorensen, D., Nielsen, T. H., J, S., and C, C. (2002). Cyclic lipoundecapeptide lokisin from pseudomonas sp. strain dss41. *Tetrahedron letters*, 43(25):4421–4423.

[Soto and Estrada, 2008] Soto, C. and Estrada, L. D. (2008). Protein misfolding and neurodegeneration. *Arch Neurol*, 65(2):184–189.

[Steindl et al., 2006] Steindl, T. M., Schuster, D., Wolber, G., Laggner, C., and Langer, T. (2006). High-throughput structure-based pharmacophore modelling as a basis for successful parallel virtual screening. *J Comput Aided Mol Des*, 20(12):703–715.

[Strebhardt and Ullrich, 2008] Strebhardt, K. and Ullrich, A. (2008). Paul ehrlich's magic bullet concept: 100 years of progress. *Nat Rev Cancer*, 8(6):473–480.

[Sum et al., 2009] Sum, C. S., Tikhonova, I. G., Costanzi, S., and Gershengorn, M. C. (2009). Two arginine-glutamate ionic locks near the extracellular surface of ffar1 gate receptor activation. *J Biol Chem*, 284(6):3529–3536.

[Sum et al., 2007] Sum, C. S., Tikhonova, I. G., Neumann, S., Engel, S., Raaka, B. M., Costanzi, S., and Gershengorn, M. C. (2007). Identification of residues important for agonist recognition and activation in gpr40. *J Biol Chem*, 282(40):29248–29255.

[Sun et al., 2011] Sun, H., Reinscheid, U. M., Whitson, E. L., d'Auvergne, E. J., Ireland, C. M., Navarro-Vázquez, A., and Griesinger, C. (2011). Challenge of large-scale motion for residual dipolar coupling based analysis of configuration: the case of fibrosterol sulfate a. *J Am Chem Soc*, 133(37):14629–14636.

[Tamaoki et al., 1986] Tamaoki, T., Nomoto, H., Takahashi, I., Kato, Y., Morimoto, M., and Tomita, F. (1986). Staurosporine, a potent inhibitor of phospholipid/ca^{++} dependent protein kinase. *Biochem Biophys Res Commun*, 135(2):397–402.

[Thrippleton and Keeler, 2003] Thrippleton, M. J. and Keeler, J. (2003). Elimination of zero-quantum interference in two-dimensional nmr spectra. *Angew Chem Int Ed Engl*, 42(33):3938–3941.

[Tikhonova et al., 2008] Tikhonova, I. G., Sum, C. S., Neumann, S., Engel, S., Raaka, B. M., Costanzi, S., and Gershengorn, M. C. (2008). Discovery of novel agonists and antagonists of the free fatty acid receptor 1 (ffar1) using virtual screening. *J Med Chem*, 51(3):625–633.

[Tikhonova et al., 2007] Tikhonova, I. G., Sum, C. S., Neumann, S., Thomas, C. J., Raaka, B. M., Costanzi, S., and Gershengorn, M. C.

(2007). Bidirectional, iterative approach to the structural delineation of the functional 'chemoprint' in gpr40 for agonist recognition. *J Med Chem*, 50(13):2981–2989.

[Tong et al., 1997] Tong, L., Pav, S., White, D. M., Rogers, S., Crane, K. M., Cywin, C. L., Brown, M. L., and Pargellis, C. A. (1997). A highly specific inhibitor of human p38 map kinase binds in the atp pocket. *Nat Struct Biol*, 4(4):311–316.

[Tropp, 1980] Tropp, J. (1980). Dipolar relaxation and nuclear overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances. *J Chem Phys*, 72(11):6035–6043.

[Tucker et al., 1994] Tucker, G. T., Lennard, M. S., Ellis, S. W., Woods, H. F., Cho, A. K., Lin, L. Y., Hiratsuka, A., Schmitz, D. A., and Chu, T. Y. (1994). The demethylenation of methylenedioxymethamphetamine ("ecstasy") by debrisoquine hydroxylase (cyp2d6). *Biochem Pharmacol*, 47(7):1151–1156.

[Ui et al., 1997] Ui, H., Miyake, T., Iinuma, H., Imoto, M., Naganawa, H., Hattori, S., Hamada, M., Takeuchi, T., Umezawa, S., and Umezawa, K. (1997). Pholipeptin, a novel cyclic lipoundecapeptide from pseudomonas fluorescens. *J Org Chem*, 62(1):103–108.

[Voelz et al., 2010] Voelz, V. A., Bowman, G. R., Beauchamp, K., and Pande, V. S. (2010). Molecular simulation of ab initio protein folding for a millisecond folder ntl9(1-39). *J Am Chem Soc*, 132(5):1526–1528.

[Wang et al., 2004] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J Comput Chem*, 25(9):1157–1174.

[Washio et al., 2010] Washio, K., Lim, S. P., Roongsawang, N., and Morikawa, M. (2010). Identification and characterization of the genes responsible for the production of the cyclic lipopeptide arthrofactin by pseudomonas sp. mis38. *Biosci Biotechnol Biochem*, 74(5):992–999.

[Wermuth et al., 1998] Wermuth, C. G., Ganellin, C. R., Lindberg, P., and Mitscher, L. A. (1998). Glossary of terms used in medicinal chemistry (iupac recommendations 1997). *Annu Rep Med Chem*, 33:385–395.

[Werner et al., 2008] Werner, K., Richter, C., Klein-Seetharaman, J., and Schwalbe, H. (2008). Isotope labeling of mammalian gpcrs in hek293 cells and characterization of the c-terminus of bovine rhodopsin by high resolution liquid nmr spectroscopy. *J Biomol NMR*, 40(1):49–53.

[Wienkers and Heath, 2005] Wienkers, L. C. and Heath, T. G. (2005). Predicting in vivo drug interactions from in vitro drug discovery data. *Nat Rev Drug Discov*, 4(10):825–833.

[Williams et al., 2003] Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak VinkoviÄ‡, D., and Jhoti, H. (2003). Crystal structure of human cytochrome p450 2c9 with bound warfarin. *Nature*, 424(6947):464–468.

[Wolber et al., 2006] Wolber, G., Dornhofer, A. A., and Langer, T. (2006). Efficient overlay of small organic molecules using 3d pharmacophores. *J Comput Aided Mol Des*, 20(12):773–788.

[Wolber and Langer, 2005] Wolber, G. and Langer, T. (2005). Ligandscout: 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*, 45(1):160–169.

[Wolber et al., 2008] Wolber, G., Seidel, T., Bendix, F., and Langer, T. (2008). Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov Today*, 13(1-2):23–29.

[Yan et al., 2007] Yan, G., Li, Q., Tan, H., and Ge, T. (2007). Electrospray ionization ion-trap time-of-flight tandem mass spectrometry of two furofurans: sesamin and gmelinol. *Rapid Commun Mass Spectrom*, 21(22):3613–3620.

[Yan et al., 2003] Yan, J., Kline, A. D., Mo, H., Shapiro, M. J., and Zartler, E. R. (2003). The effect of relaxation on the epitope mapping by saturation transfer difference nmr. *J Magn Reson*, 163(2):270–276.

[Zhang et al., 2008] Zhang, D., Zhu, M., and Humphreys, W. G., editors (2008). *Durg metabolism in drug design and development*. John Wiley & sons.

[Zhou et al., 2010] Zhou, C., Tang, C., Chang, E., Ge, M., Lin, S., Cline, E., Tan, C. P., Feng, Y., Zhou, Y. P., Eiermann, G. J., Petrov, A., Salituro, G., Meinke, P., Mosley, R., Akiyama, T. E., Einstein, M., Kumar, S., Berger, J., Howard, A. D., Thornberry, N., Mills, S. G., and Yang, L. (2010). Discovery of 5-aryloxy-2,4-thiazolidinediones as potent gpr40 agonists. *Bioorg Med Chem Lett*, 20(3):1298–1301.

Appendix A

Protein kinase A

A.1 Rescoring of docking poses with GOLD

Name	RMSD to X-ray	ChemPLP	ChemScore	ASP	GOLDScore
Ligand 1					
1	0.41	-84.5	33.8	28.27	45.78
2	5.1	-73.1	26.89	26	32.09
3	5.53	-63.3	25.8	26.87	34.22
4	2.62	-62.9	25.26	22.97	24.5
5	5.5	-59.8	25.56	21.21	32.66
6	6.33	-58.6	20.48	20.03	29.99
7	7.83	-57.6	18.15	15.36	26.02
8	3.94	-54.6	18.97	20.49	25.24
9	7.34	-53.7	17.69	15.73	24.37
10	5.24	-52.7	19.62	14.65	24.44
11	2.56	-52.6	21	19.56	28.58
12	7.38	-54.5	19.63	20.66	21.07
13	5.59	-54.3	23.8	19.31	30.93
14	7.25	-54	20.26	16.15	17.66
15	6.18	-53.5	19.78	19.69	29.4
16	6.99	-52.4	17.01	15.37	21.84

Name	RMSD to X-ray	ChemPLP	ChemScore	ASP	GOLDScore
Ligand 2					
1	5.2	-70.9	27.91	26.24	36.22
2	5.26	-69.4	25.02	24.19	35.66
3	1.41	-66.3	24.53	26.06	39.99
4	7.68	-61.5	17.33	13.16	30.25
5	3.14	-58.3	16.96	16.57	32.47
6	5.8	-58	17.12	17.47	31.92
7	6.78	-55.8	15.42	14.04	24.9

Name	RMSD to X-ray	ChemPLP	ChemScore	ASP	GOLDScore
Ligand 3					
1	3.36	-60.5	25.53	20.42	31.26
2	3.78	-57.9	22.29	20.61	29.33
3	4.4	-57.5	23.8	17.02	29.45
4	2.86	-55.4	24.25	21.5	35.89
5	2.85	-51.6	22.28	16.08	30.81
6	4.59	-51.5	21.45	15.52	28.51
7	5.54	-49.3	18.55	15.61	25.17
8	6.57	-47.7	16.76	11.53	21.5
9	3.74	-47.7	18.16	16.33	25.66
10	0.96	-47	18.97	14.35	27.32
11	2.47	-44.7	18.21	16.11	23.25
12	7.26	-44.5	13.63	13.1	20.69
13	5.44	-44.5	15.82	13.1	18.01
14	5.74	-44.1	16.21	12.41	12.73
15	4.55	-43.4	16.99	14.55	19.32
16	7.83	-43.1	14.61	11.05	18.7
17	4.55	-43	18.01	14.36	-13.62
18	7.5	-43	14.36	12.56	20.65
19	3.42	-42.8	16.78	15.12	28.7

Name	RMSD to X-ray	ChemPLP	ChemScore	ASP	GOLDScore
Ligand 4					
1	4	-66.7	10.36	23.8	29.1
2	3.85	-66	11.97	23.66	36.58
3	0.15	-64.3	16.7	27.5	40.39
4	6.1	-63.8	17.09	25.5	34.44
5	7.6	-62.6	12.18	19.4	33.01
6	5.99	-62.5	17.25	24.3	29.35
7	7.22	-61.4	17.43	22.66	35.46
8	7.43	-60.6	11.61	20.45	33.37
9	7.4	-59.7	16.58	22.82	32.46
10	6.29	-59.1	17.76	25.94	37.76
11	2.92	-58.8	14.4	25.6	37
12	5.92	-58.7	18.32	22.29	33.03
13	6.94	-58.7	15.1	23.05	35.51
14	5.96	-58.3	12.66	21.58	33.41
15	2.26	-57.8	11.11	23.94	28.63
16	7.14	-57.4	16.67	24.11	29.15
17	3.02	-56.4	11.19	22.12	38.04
18	6.44	-55.9	15.94	26.62	35.21
19	7.46	-55.5	10.89	20.05	29.37
20	6.01	-54.7	12.51	21.35	28.46
21	7.93	-54.5	13.71	18.14	26.73
22	7.18	-54.5	18.55	20.38	31.15
23	2.81	-54.8	10.95	21.98	20.48
24	4.57	-58.5	10.68	21.59	34.51
25	7.26	-56.1	16.32	19.17	28.91
26	2.07	-56.1	11	24.43	30.65
27	7.34	-56	12.65	20.05	28.97

A.2 NOESY spectra of ligand combinations

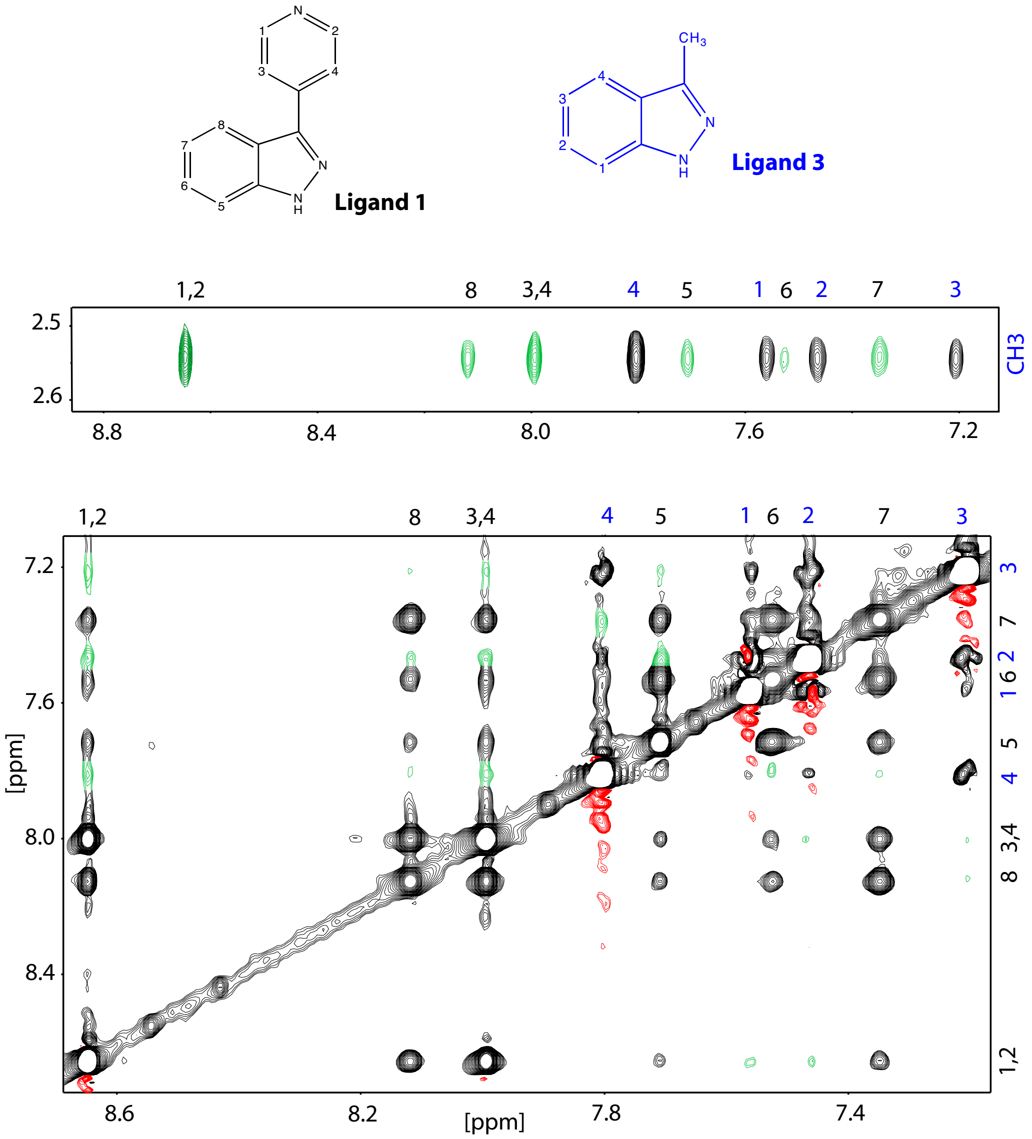


Figure A.1: NOESY spectrum of ligand 1 (0.3 mM) and ligand 3 (0.9 mM) in the presence of PKA (0.045 mM). INPHARMA peaks chosen are marked in green. The mixing time was 800 ms on a 900 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 293 K with 64 scans, 2048 points in F2 and 580 points in F1.

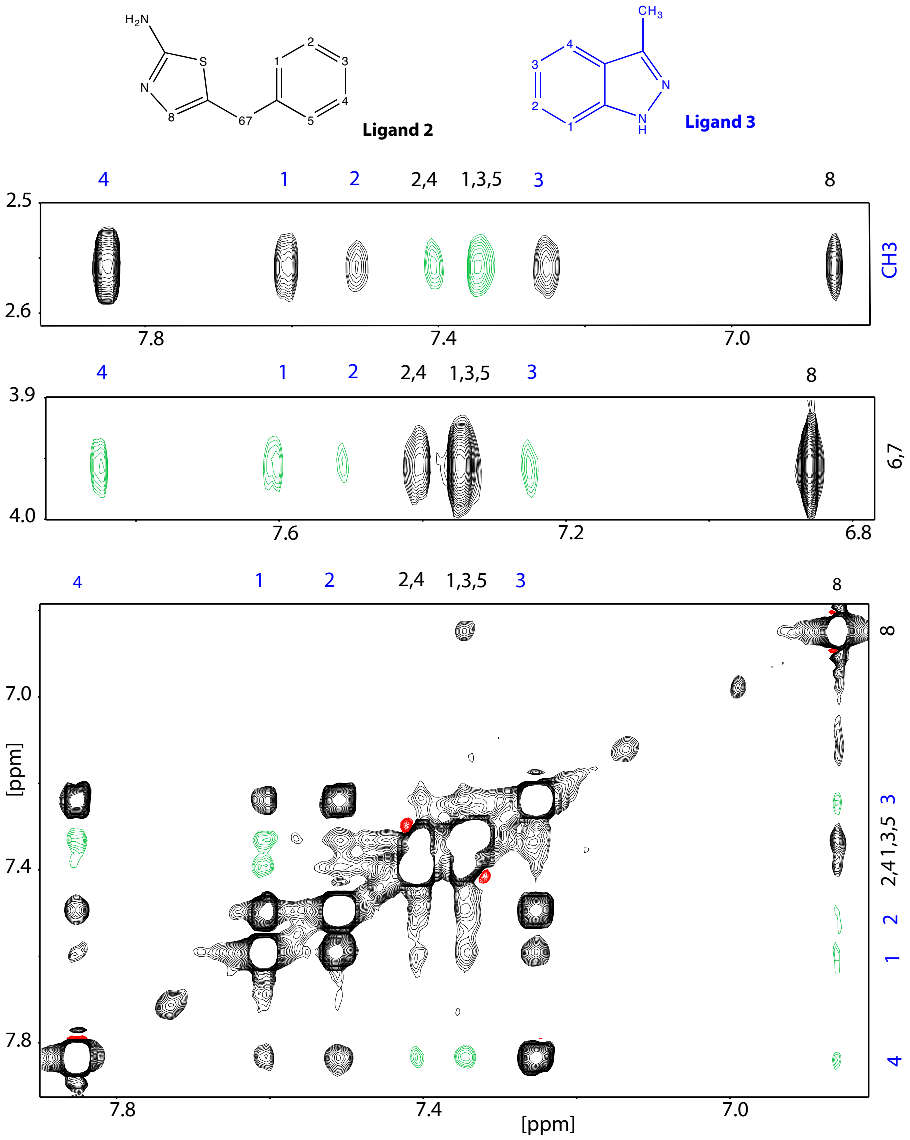


Figure A.2: NOESY spectrum of ligand 2 (0.9 mM) and ligand 3 (1.2 mM) in the presence of PKA (0.045 mM). INPHARMA peaks chosen are marked in green. The mixing time was 800 ms on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 293 K with 64 scans, 2048 points in F2 and 580 points in F1.

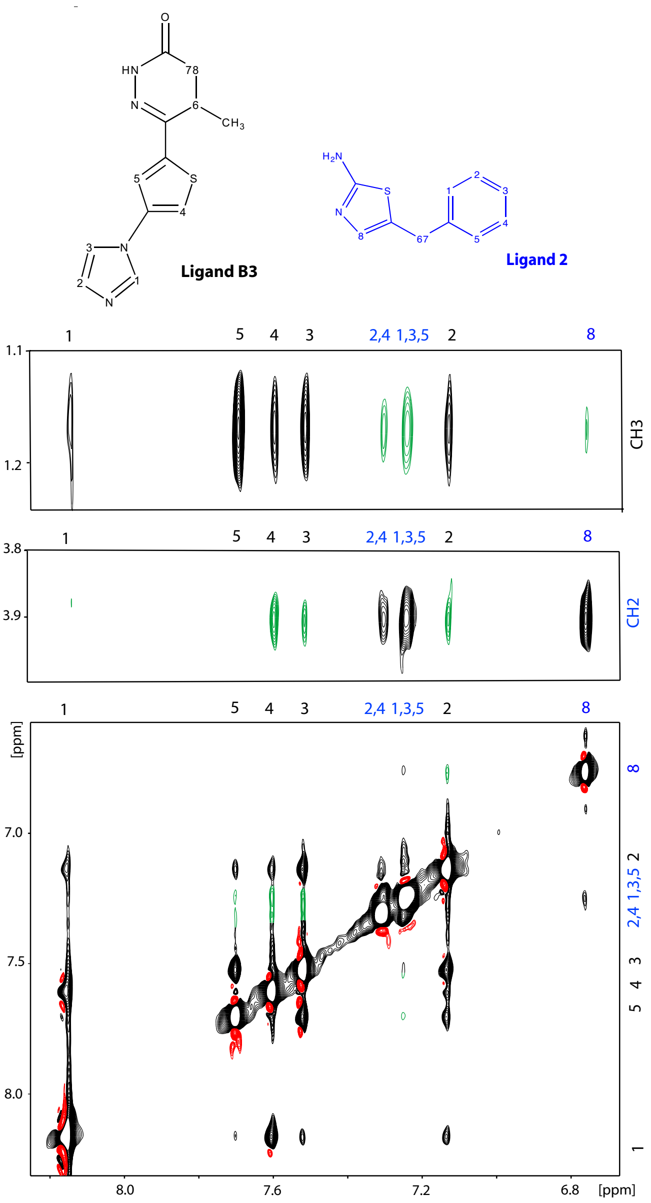


Figure A.3: NOESY spectrum of ligand B3 (1 mM) and ligand 2 (0.3 mM) in the presence of PKA (0.045 mM). INPHARMA peaks chosen are marked in green. The mixing time was 800 ms on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 293 K with 64 scans, 2048 points in F2 and 388 points in F1.

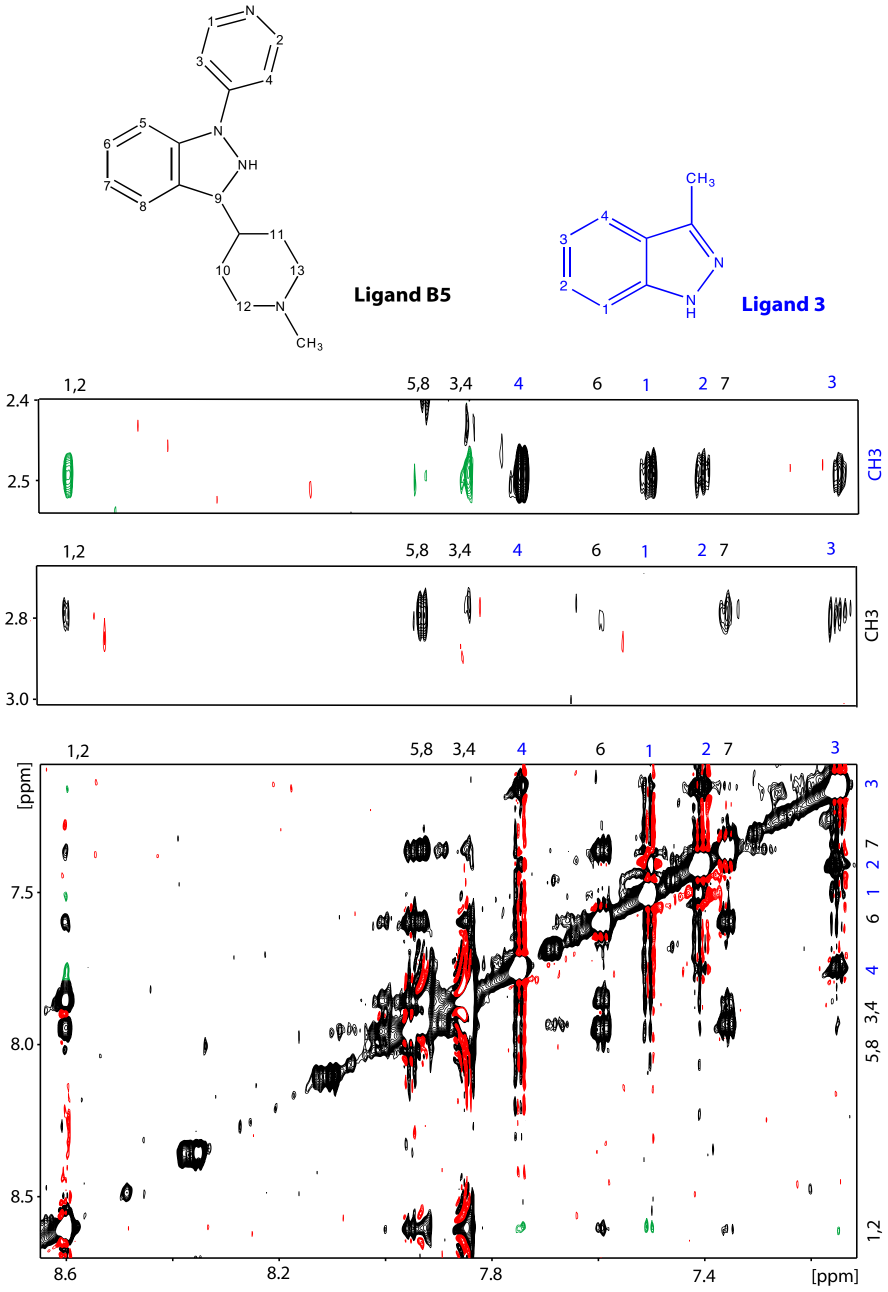


Figure A.4: NOESY spectrum of ligand B5 (0.5 mM) and ligand 3 (1 mM) in the presence of PKA (0.045 mM). INPHARMA peaks chosen are marked in green. The mixing time was 800 ms on a 800 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 293 K with 64 scans, 4096 points in F2 and 512 points in F1.

Appendix B

G-protein coupled receptor 40

B.1 NMR spectra of ligand combination

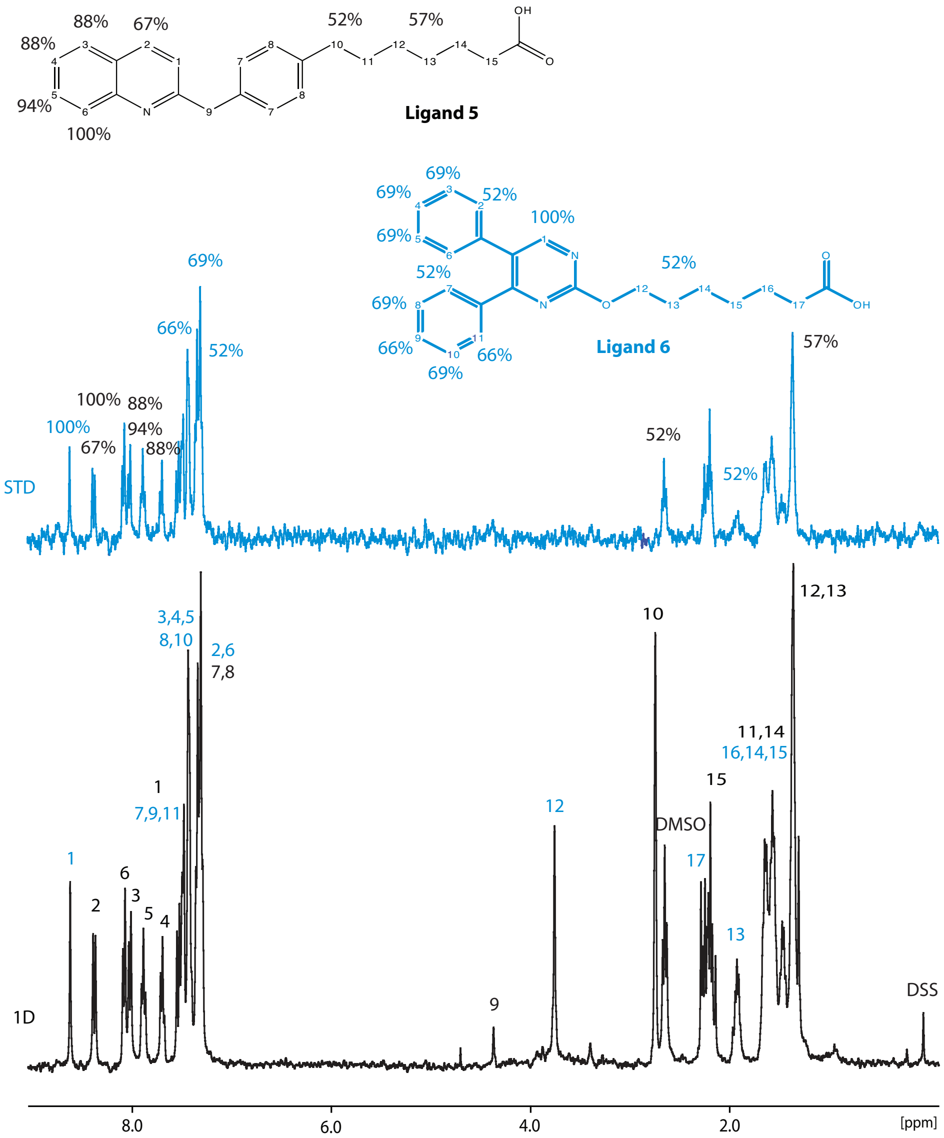


Figure B.1: Experimental STD data of the GPR40 binding ligands 5 and 6. The normal 1D spectrum in the presence of GPR40 membranes is shown in black, the STD spectrum in blue. For back-calculation only ligand 5 was considered, as ligand 6 had too many overlapping signals in the aromatic region. The sample contains 0.2 mM ligand 5, 0.1 mM ligand 6, 25% (v/v) GPR40 membranes, 10 mM perdeuterated Tris and 1 mM perdeuterated EDTA. Saturation was done for 8 s on -2 ppm on a 400 MHz spectrometer (number of scans = 16; temperature = 310 K; time domain = 32768 points).

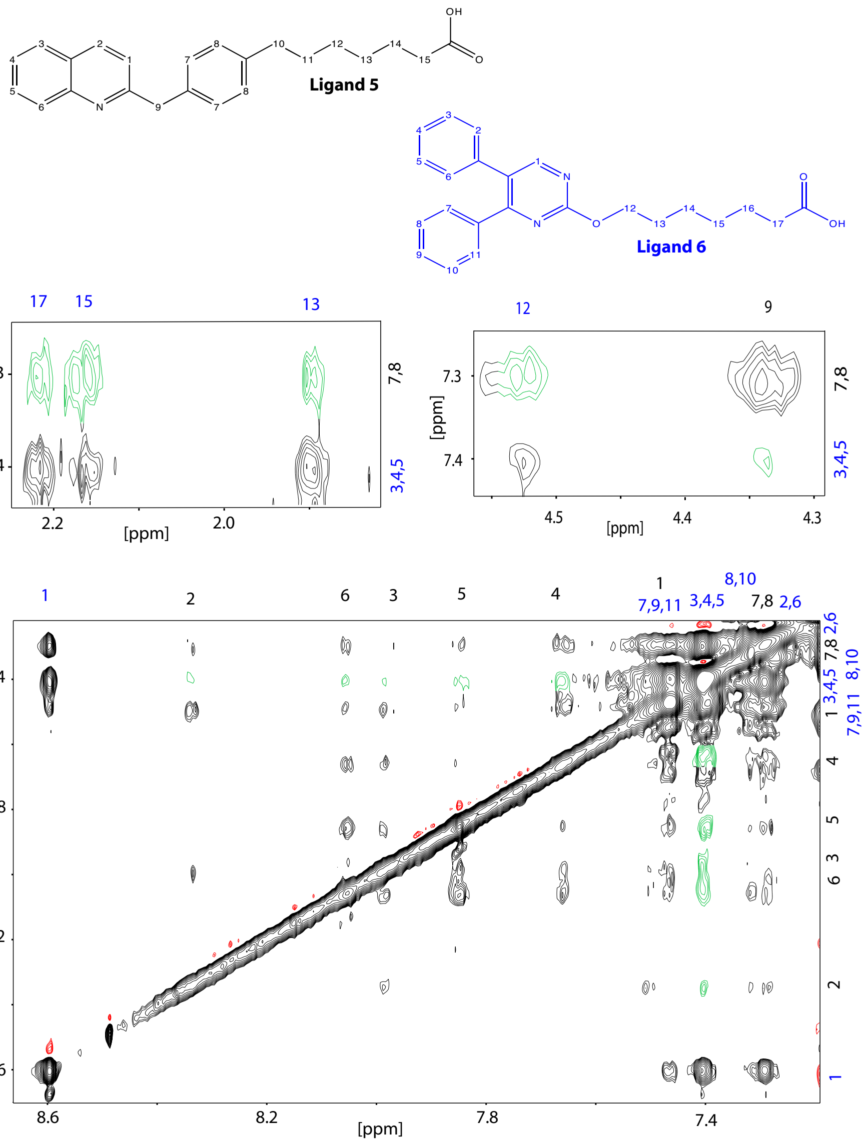


Figure B.2: NOESY spectrum of ligand 5 (0.2 mM) and ligand 6 (0.1 mM) in the presence of 25% (v/v) GPR40 membranes in D_2O (10 mM Tris, 1 mM EDTA). INPHARMA peaks chosen are marked in green. The mixing time was 300 ms on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 310 K with 128 scans and 8192 points in F2 and 256 points in F1.

Appendix C

Anle138b

C.1 Mass spectrometric and NMR spectroscopic data

Anle138b
3-(1,3-benzodioxol-5-yl)-5-(3-bromophenyl)-1H-pyrazole
ESI-MS m/z (acetonitrile, positive mode): $C_{16}H_{11}BrN_2O_2$ $[M + H^+]$: calculated molecular weights: 343.0082, 345.0062; observed masses: 343.0085, 345.0063.
^{1}**H NMR** (DMSO, 400 MHz): $\delta = 8.00$ (s, 1H), 7.81 (d, $J = 8.1$ Hz, 1H), 7.5 (d, $J = 8.0$ Hz, 1H), 7.4 (d, $J = 7.9$ Hz, 1H), 7.37 (d, $J = 1.5$ Hz, 1H), 7.32 (d, $J = 8.0$ Hz, 1H), 7.19 (s, 1H), 7.0 (d, $J = 8.0$ Hz, 1H), 6.06 (s, 2H).

M1 - sulfurylated anle138c
5-(5-(3-bromophenyl)-1H-pyrazol-3-yl)-2-hydroxyphenyl hydrogen sulfate
4-(5-(3-bromophenyl)-1H-pyrazol-3-yl)-2-hydroxyphenyl hydrogen sulfate
ESI-MS m/z (acetonitrile, positive mode): $C_{15}H_{11}BrN_2O_5S[M + H^+]$: calculated molecular weights: 410.965, 412.963; observed masses: 410.9646, 412.9619.
^{1}**H NMR** (DMSO, 600 MHz): $\delta = 8.04$ (s, 1H), 7.87 (d, $J = 7.8$ Hz), 7.59 (d, $J = 1.9$ Hz, 1H), 7.49 (d, $J = 7.8$ Hz, 1H), 7.42 (dd, $J = 8.2$ Hz, 1.9 Hz, 1H), 7.38 (t, $J = 7.8$, 1H), 7.12 (s, 1H), 6.93 (d, $J = 8.2$, 1H).
^{1}**H NMR** (DMSO, 600 MHz): $\delta = 8.04$ (s, 1H), 7.82 (d, $J = 7.8$ Hz), 7.59 (d,$J = 1.9$ Hz, 1H), 7.38 (d, $J = 7.8$ Hz, 1H), 7.46 (dd, $J = 8.2$ Hz, 1.9 Hz, 1H), 7.38 (t, $J = 7.8$, 1H), 7.12 (s, 1H), 6.87 (d, $J = 8.2$, 1H).

M2 - anle138c
3-(1,3-benzodioxol-5-yl)-5-(3-bromophenyl)-1H-pyrazole

ESI-MS m/z (acetonitrile, positive mode): $C_{15}H_{11}BrN_2O_2[M+H^+]$: calculated molecular weights: 331.0082, 345.0062; observed masses: 331.0082, 333.0061.
^{1}**H NMR** (DMSO, 400 MHz): $\delta = 8.02$ (s, 1H), 7.83 (d, $J = 8.1$ Hz, 1H), 7.5 (d, $J = 7.8$ Hz, 1H), 7.38 (t, $J = 7.8$ Hz, 1H), 7.19 (d, $J = 1.9$ Hz, 1H), 7.08 (dd, $J = 8.2$ Hz, 1.9 Hz, 1H), 7.02 (s, 1H), 6.8 (d, $J = 8.2$ Hz, 1H).

M3 - Hydroxylated anle138b

4-(3-(benzo[d][1,3]dioxol-6-yl)-1H-pyrazol-5-yl)-2-bromophenol
3-(3-(benzo[d][1,3]dioxol-6-yl)-1H-pyrazol-5-yl)-5-bromophenol
ESI-MS m/z (acetonitrile, positive mode): $C_{16}H_{11}BrN_2O_3$ $[M+H^+]$: calculated molecular weights: 359.0031, 361.0011; observed masses: 359.0028, 361.0011.
^{1}**H NMR** (DMSO, 900 MHz): $\delta = 7.92$ (s, 1H), 7.65 (d, $J = 8.1$ Hz, 1H), 7.4 (d, $J = 7.9$ Hz, 1H), 7.36 (m, 2H), 7.34 (d, $J = 8.0$ Hz, 1H), 7.21 (s, 1H), 7.04 (d, $J = 8.0$ Hz, 1H), 6.07 (s, 2H).
^{1}**H NMR** (DMSO, 900 MHz): $\delta = 7.92$ (s, 1H), 7.59 (d, $J = 8.1$ Hz, 1H), 7.4 (d, $J = 7.9$ Hz, 1H), 7.36 (m, 2H), 7.34 (d, $J = 8.0$ Hz, 1H), 7.18 (s, 1H), 6.96 (d, $J = 8.0$ Hz, 1H), 6.07 (s, 2H).

M4 - Methylated anle138c

5-(5-(3-bromophenyl)-1H-pyrazol-3-yl)-2-methoxyphenol
4-(5-(3-bromophenyl)-1H-pyrazol-3-yl)-2-methoxyphenol
ESI-MS m/z (acetonitrile, positive mode): $C_{16}H_{11}BrN_2O_3$ $[M+H^+]$: calculated molecular weights: 345.0239, 347.0218; observed masses: 345.0239, 347.0218.
^{1}**H NMR** (DMSO, 600 MHz): $\delta = 8.03$ (s, 1H), 7.86 (d, $J = 8.1$ Hz, 1H), 7.55 (d, $J = 7.8$ Hz, 1H), 7.39 (m, 1H), 7.23 (s, 1H), 7.22 (m, 1H), 7.03 (s, 1H), 6.86 (d, $J = 8.2$ Hz, 1H), 3.86 (s, 3H).
^{1}**H NMR** (DMSO, 600 MHz): $\delta = 8.03$ (s, 1H), 7.81 (d, $J = 8.1$ Hz, 1H), 7.5 (d, $J = 7.8$ Hz, 1H), 7.39 (m, 1H), 7.23 (s, 1H), 7.14 (m, 1H), 6.97 (s, 1H), 6.82 (d, $J = 8.2$ Hz, 1H), 3.86 (s, 3H).

C.2 NMR spectra of anle138b and its metabolites

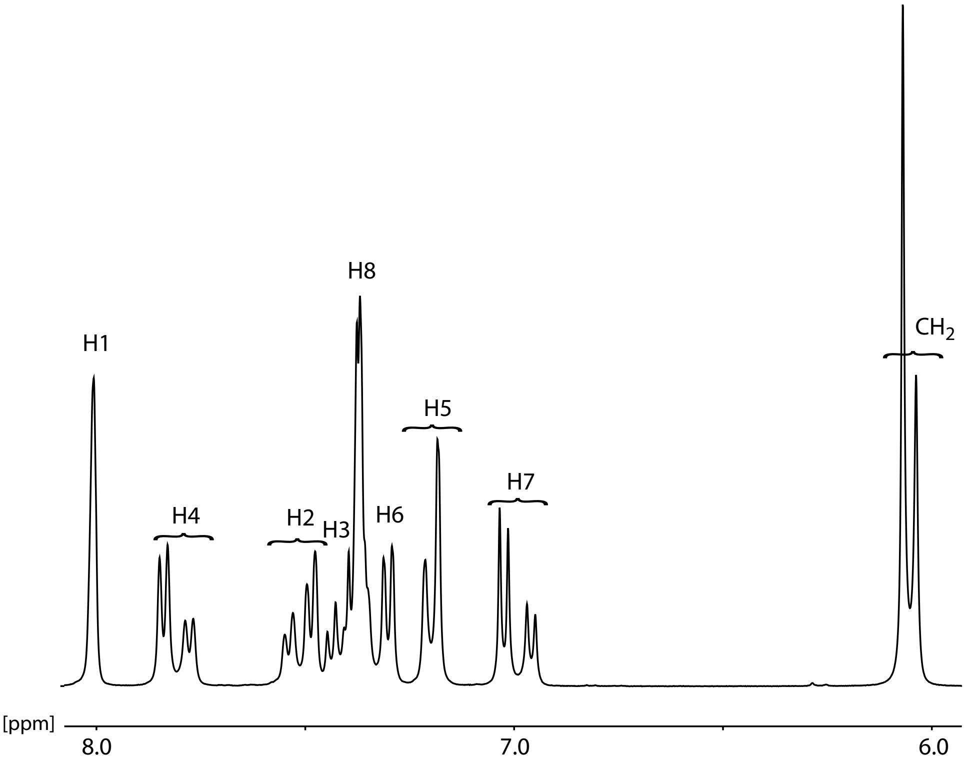


Figure C.1: 1D spectrum of anle138b in DMSO-d_6. Anle138b is present as a tautomer, as the proton switches between the two nitrogen atoms.

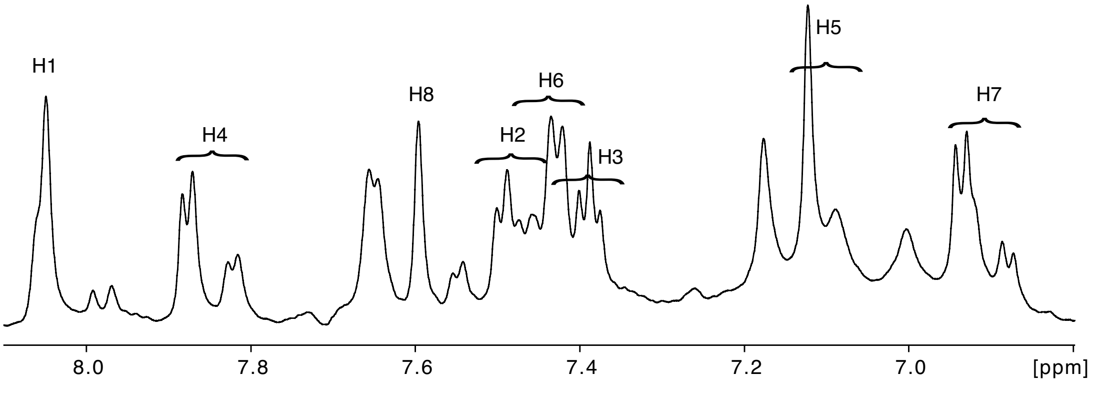


Figure C.2: 1D spectrum of metabolite M1 in DMSO-d_6.

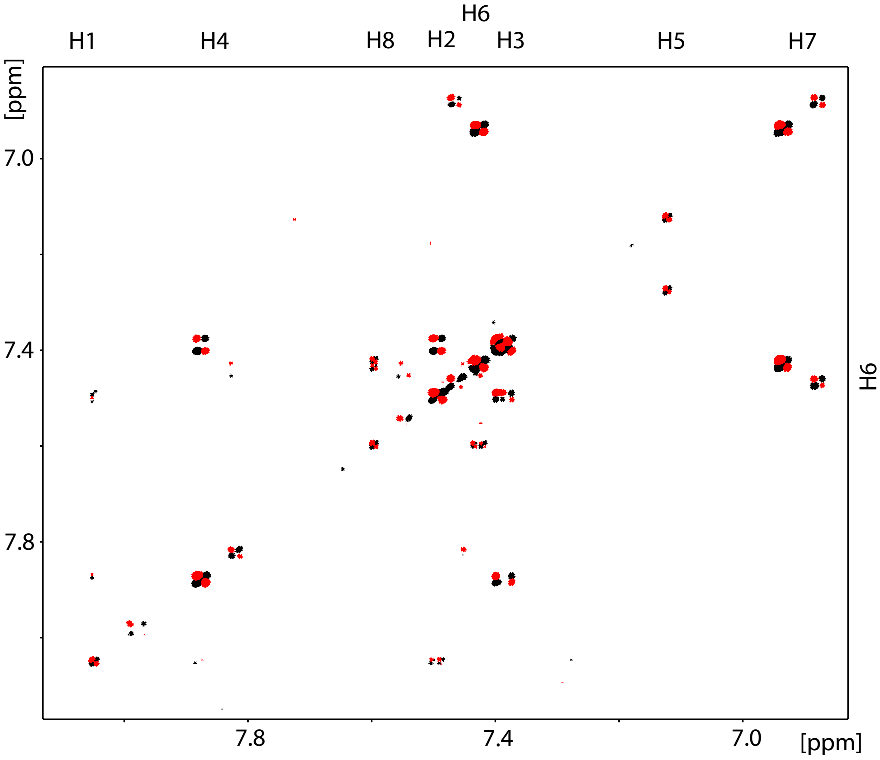


Figure C.3: DQF-COSY spectrum of metabolite M1 in DMSO-d_6.

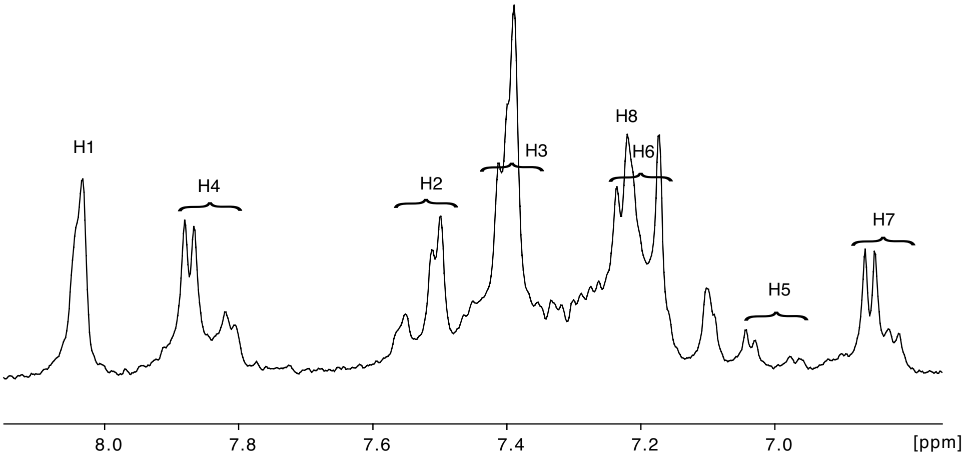


Figure C.4: 1D spectrum of metabolite M4 in DMSO-d_6, recorded on a 600 MHz spectrometer.

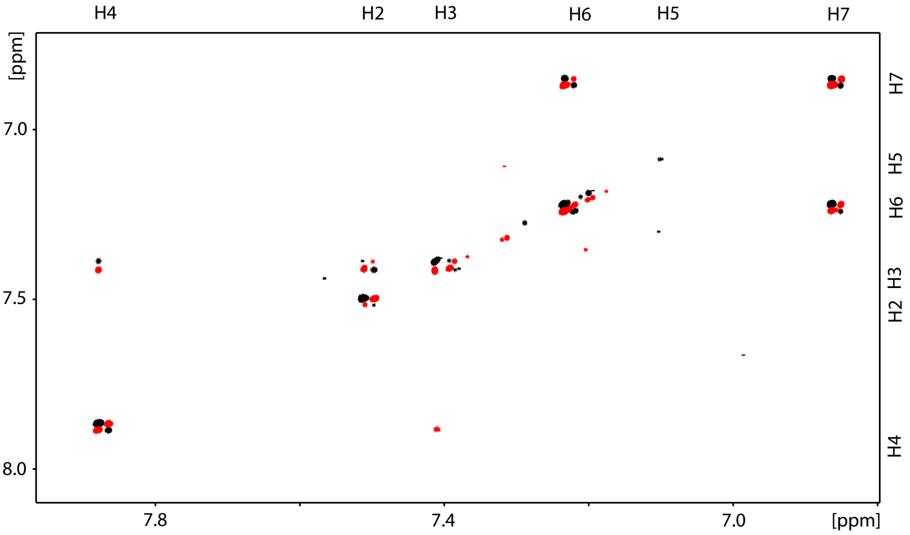


Figure C.5: DQF-COSY spectrum of metabolite M4 in DMSO-d_6, recorded on a 600 MHz spectrometer.

C.3 Fragmentation patterns of the metabolites

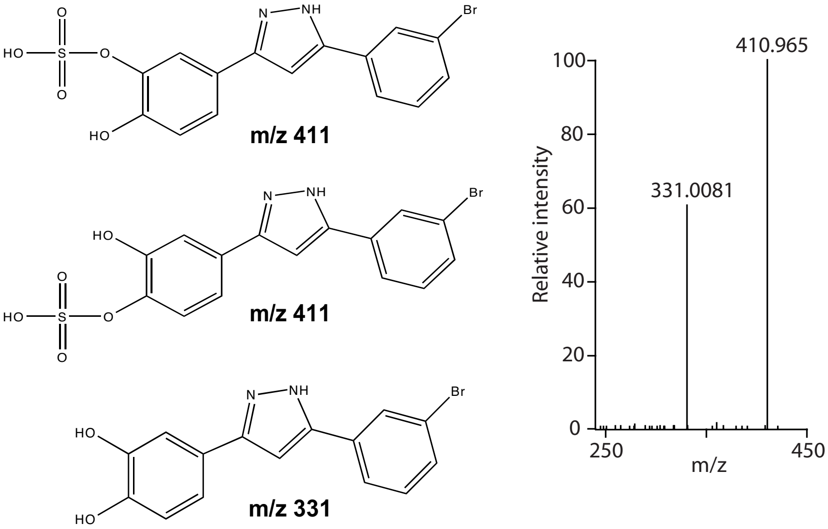

m/z 411
m/z 411
m/z 331
410.965
331.0081
Relative intensity
m/z
0
20
40
60
80
100
250
450

Figure C.6: Interpretation of the fragmentation pattern of M1, based on the mass spectrometric fragmentation data. The sulfate group is released and M1 is degraded to m/z 331/333, which equals M2 (anle138c).

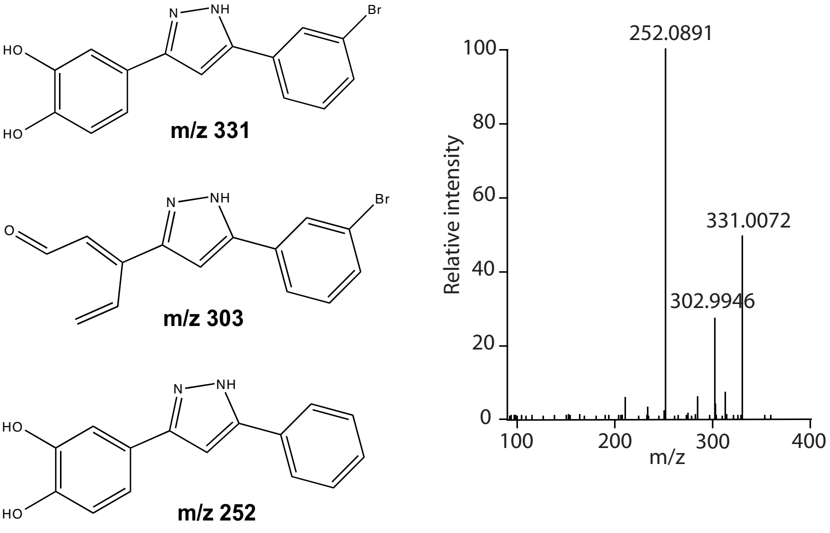

m/z 331
m/z 303
m/z 252
252.0891
331.0072
302.9946
Relative intensity
m/z
0
20
40
60
80
100
100
200
300
400

Figure C.7: Interpretation of the fragmentation pattern of M2, based on the mass spectrometric fragmentation data. Either the aromatic ring is broken, leading to the fragment m/z 303/305, or the bromine is released, resulting in the product m/z 252.

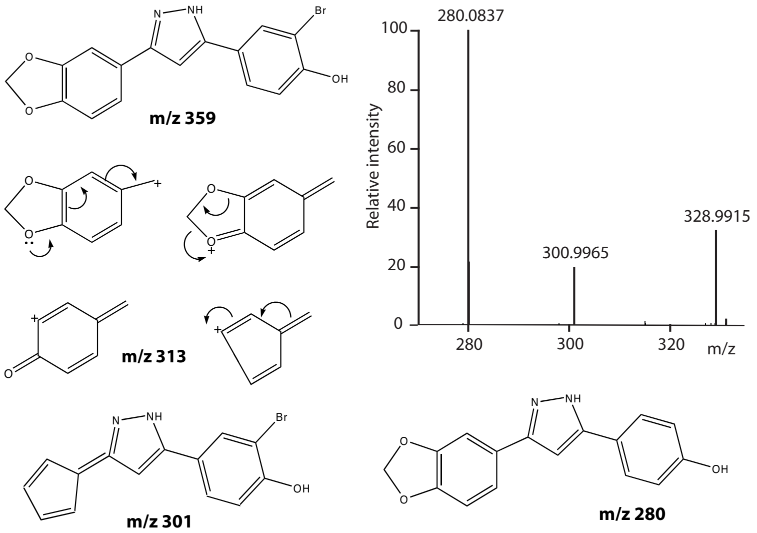

m/z 359
m/z 313
m/z 301
m/z 280
N—NH
Br
OH
280.0837
300.9965
328.9915
Relative intensity
100
80
60
40
20
0
280
300
320
m/z

Figure C.8: Interpretation of the fragmentation pattern of M3, based on the mass spectrometric fragmentation data. M3 behaves similar to anle138b with a rearrangement of the aromatic ring to *m/z* 313/315 or *m/z* 301/303. Release of the bromine leads to product *m/z* 280. The hydroxy group seems so be stable against fragmentation.

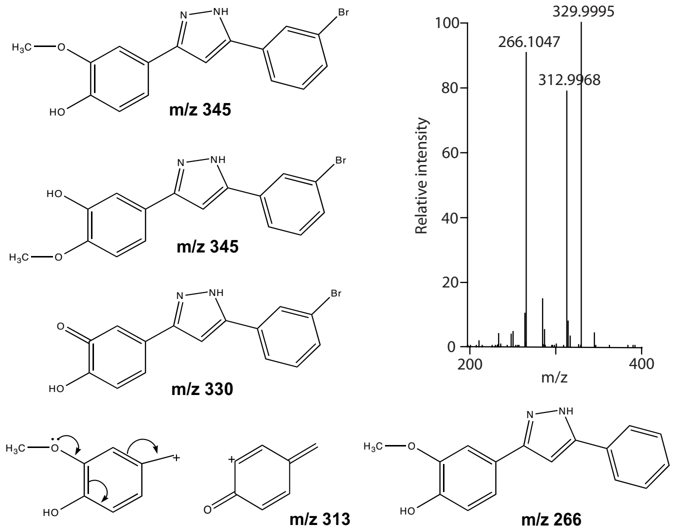

m/z 345
m/z 345
m/z 330
m/z 313
m/z 266
H3C—O
HO
N—NH
Br
329.9995
266.1047
312.9968
Relative intensity
100
80
60
40
20
0
200
400
m/z

Figure C.9: Interpretation of the fragmentation pattern of M4, based on the mass spectrometric fragmentation data. Demethylation results in the M2 (anle138c)-like product *m/z* 330/332. Rearrangement of the aromatic ring leads to the product *m/z* 313/315. Finally release of the bromine results in product *m/z* 266.

Appendix D

Arthrofactin

D.1 Assignment of arthrofactin

ESI-MS *m/z* (acetonitrile, positive mode): $C_{64}H_{111}N_{11}O_{20}Na$ $[M + Na^{+}]$: calculated molecular weights: 1376.79; observed masses: 1376.7907.
^{1}**H NMR** (DMSO, 700 MHz): δ = 8.28; 8.16; 8.11; 8.078; 8.05; 8.04; 8.03; 7.99; 7.9; 7.82; 7.75; 5.004; 4.86; 4.52; 4.48; 4.47; 4.397; 4.342; 4.29; 4.236; 4.23; 4.186; 4.171; 3.79; 3.61; 3.6; 3.52; 3.46; 2.87; 2.64; 2.49; 2.42; 2.23; 1.88; 1.86; 1.6; 1.579; 1.56; 1.549; 1.52; 1.465; 1.439; 1.43; 1.37; 1.34; 1.24; 1.13; 1.08; 1.06; 0.86; 0.79;
^{13}**C NMR** (DMSO, 700 MHz): δ = 172.8; 172.3; 172.3; 172.11; 171.9; 171.7; 171.4; 171.1; 170.6; 170.3; 170.28; 170; 168.8; 71; 68; 62.7; 61.15; 57.6; 56.59; 56.5; 55.66; 54.5; 51.91; 51.52; 51.07; 49.95; 48.38; 43.9 41; 40.82; 40.8; 40.16; 37; 36.9; 36.72; 35.71; 35.6; 25.2; 24.7; 24.51; 24.48; 24.4; 24.3; 23.52; 23.4; 23.03; 22.66; 21.9; 21.8; 16.81; 14.44; 14.37; 11.85; 11.7; 10.6;
^{15}**N NMR** (DMSO, 700 MHz): δ = 124.17; 123. 87; 121.02; 120.35; 119.62; 116.83; 116.53; 116.41; 115.49; 112.83;

Atom Name	^{1}H Shift	^{13}C Shift	^{15}N Shift
FA-1	0.86	14.37	
FA-6	1.24	25.2	
FA-7	1.34	36.9	
FA-8	3.79	68	
FA-9	2.23	43.9	
FA-O	-	171.4	
Leu1-NH	8.03	-	124.17
Leu1-a	4.29	51.07	
Leu1-b	1.43	41	
Leu1c	1.6	24.4	
Leu1-CH3	0.878 + 0.836	23.4 + 21.8	
Leu1-O	-	172.8	
Asp2-NH	8.28	-	116.83
Asp2-a	4.48	49.95	
Asp2-b	2.64 + 2.49	35.71	
Asp2-COOH	-	171.9	
Asp2-O	-	170.6	
Thr3-NH	7.82	-	112.83
Thr3-a	4.47	55.66	
Thr3-b	4.86	71	
Thr-CH3	1.06	16.81	
Thr3-O	-	168.8	
Leu4-NH	8.11	-	123.87
Leu4-a	4.23	56.5	
Leu4-b	1.465	40.8	
Leu4-c	1.579	24.48	
Leu4-CH3	0.867 + 0.815	23.52 + 21.9	
Leu4-O	-	172.3	
Leu5-NH	8.04	-	119.62
Leu5-a	4.342	51.52	
Leu5-b	1.439	40.82	
Leu5-c	1.549	24.51	
Leu5-CH3	0.868 + 0.783	23.03 + 22.66	

Leu5-O	-	171.7	
Ser6-NH	8.078	-	115.49
Ser6-a	4.52	54.5	
Ser6-b	3.6 + 3.46	62.7	
Ser6-O	-	170.6	
Leu7-NH	8.05	-	120.35
Leu7-a	4.171	51.91	
Leu7-b	1.56	40.16	
Leu7-CH3	0.866 + 0.8	14.44	
Leu7-O	-	172.11	
Ser8-NH	8.16	-	116.41
Ser8-a	4.236	56.59	
Ser8-b	3.61 + 3.52	61.15	
Ser8-O	-	170.3	
Ile9-NH	7.75	-	116.53
Ile9-a	4.186	57.6	
Ile9-b	1.86	36.72	
Ile9-c	1.37	24.3	
Ile9-d	1.08	11.85	
Ile9-e	0.79	11.7	
Ile9-O	-	170.28	
Ile10-NH	7.9	-	not observed
Ile10-a	4.397	56.5	
Ile10-b	1.88	35.6	
Ile10-c	1.52	24.7	
Ile10-d	1.13	10.6	
Ile10-e	0.79	-	
Ile10-O	-	171.1	
Asp11-NH	7.99	-	121.02
Asp11-a	5.004	48.38	
Asp11-b	2.87 + 2.42	37	
Asp11-COOH	-	172.3	
Asp11-O	-	170	

D.2 NMR spectra of arthrofactin

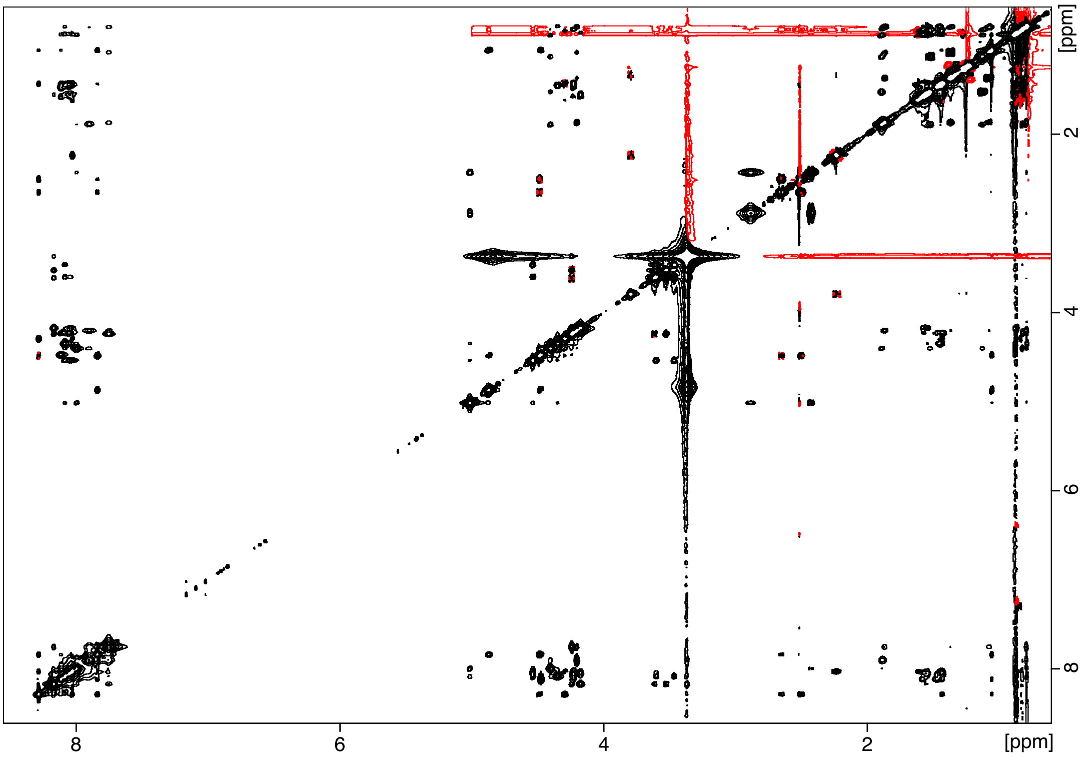


Figure D.1: NOESY spectrum of arthrofactin in DMSO. The mixing time was 150 ms on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. The spectrum was recorded at 298 K with 16 scans.

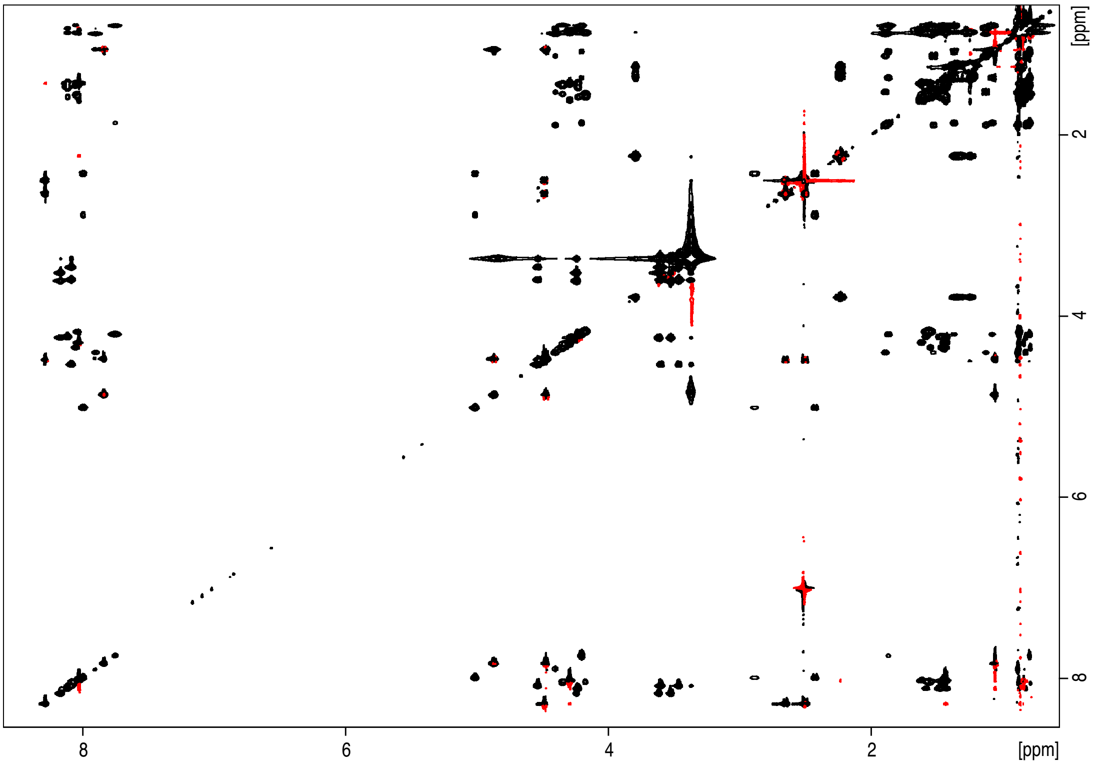


Figure D.2: TOCSY spectrum of arthrofactin in DMSO. The spectrum was recorded at 298 K with 8 scans, on a 700 MHz spectrometer, equipped with a cryogenically cooled probe head. Spin lock (DIPSI) was applied for 80 ms.

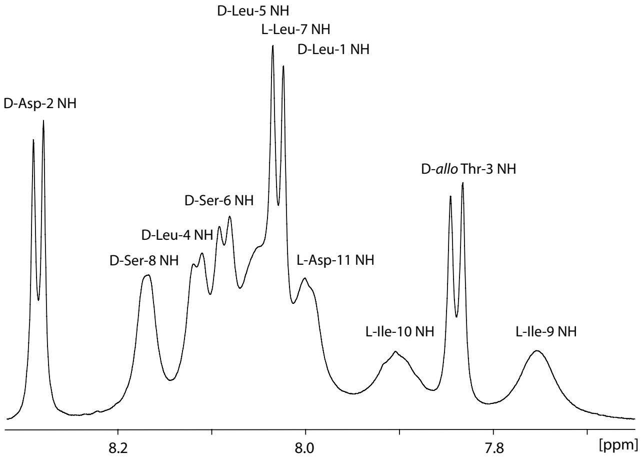


Figure D.3: Detail of the 1D spectrum of arthrofactin in DMSO. With ^{1}H-^{15}N-HMBC an assignment of all NH signals was possible.

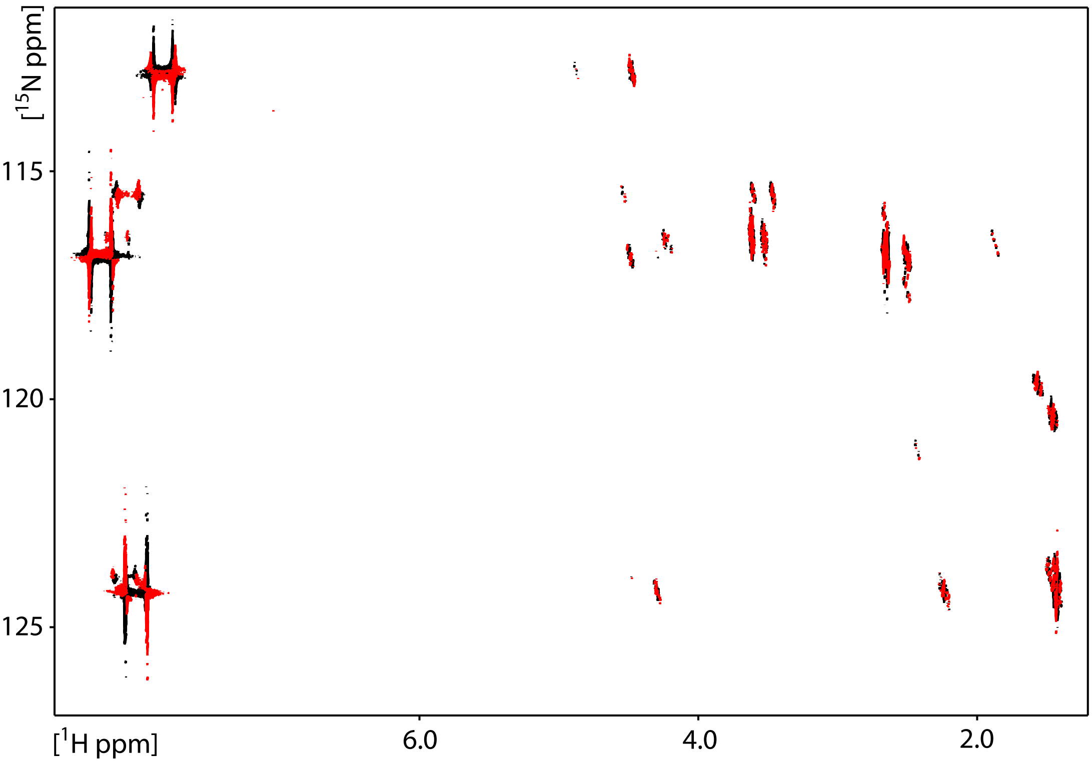


Figure D.4: ^{1}H-^{15}N-HMBC spectrum of arthrofactin in DMSO. The spectrum was recorded at 298 K with 1000 scans, on a 600 MHz spectrometer, equipped with a cryogenically cooled probe head.

Curriculum vitae

Name	**Jens Pilger**
Geburtsname	Jens-Peter Kurz
Geburtsdatum	8. August 1982
Geburtsort	Stuttgart
Staatsangehörigkeit	Deutsch
Familienstand	Verheiratet, 2 Kinder

Seit März 2009	Doktorand in der Abteilung NMR-basierte Strukturbiologie am Max-Planck-Institut für biophysikalische Chemie in Göttingen.
April 2008 - Januar 2009	Diplomarbeit in der Abteilung Geobiologie im geowissenschaftlichen Institut der Universität Göttingen bei Prof. Dr. Thiel: *Trace elements and biomarkers from the tunnel of Äspö, Sweden.*
Februar 2007-Mai 2007	Bachelorarbeit in der Abteilung Geochemie im geowissenschaftlichen Institut der Universität Göttingen bei Prof. Dr. Wörner: *Geochemisches Fingerprinting am Weinstein System mittels ICP-MS.*
Oktober 2006 - Januar 2009	Studium der Geowissenschaften an der Universität Göttingen.
August 2005 - Mai 2006	Studium der Biochemie, Anatomie und Paläotologie an der Montana State University.
Oktober 2003 - Juli 2005	Studium der Geowissenschaften an der Universität Tübingen und Vordiplom.
September 2002 - Mai 2003	Zivildienst bei Kliniken Schmieder (Neurorehabilitation) in Stuttgart.
1998 - 2002	Besuch des Geschwister-Scholl-Gymnasiums in Stuttgart und Abitur.
1993 - 1998	Besuch des Wirtemberg Gymnasiums in Stuttgart.