Universiteit Leiden
The Netherlands

NMR structural studies of protein-small molecule interactions

Shah, D.M.

Citation
Shah, D. M. (2014, June 17). *NMR structural studies of protein-small molecule interactions*. Retrieved from https://hdl.handle.net/1887/26922

Version: Not Applicable (or Unknown)
License: Leiden University Non-exclusive license
Downloaded from: https://hdl.handle.net/1887/26922

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The handle http://hdl.handle.net/1887/26922 holds various files of this Leiden University dissertation

Author: Shah, Dipen M.
Title: NMR structural studies of protein-small molecule interactions
Issue Date: 2014-06-17

NMR structural studies of protein-small molecule interactions

Dipen M. Shah

Cover Image: The NMR structure of the Hsp90-ligand complex (front cover) and the Rit1-GDP-ligand ternary complex (back cover)

Cover designed by Dipen Shah

Doctoral Thesis, Leiden University, 2014
2014, Dipen M. Shah

Printed by Ipskamp Printers, Enschede

ISBN: 978-94-6259-201-8

NMR structural studies of protein-small molecule interactions

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Prof. mr. C.J.J.M.Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 17 juni 2014
klokke 15:00 uur

door

Dipen M. Shah
Geboren te Mangalore, India in 1985

Promotiecommissie

Promotor:	Prof.Dr. M.Ubbink
Co-promotor:	Dr. G. Siegal
Overige leden:	Prof.Dr. J. Brouwer
	Prof.Dr. A. P. IJzerman
	Prof.Dr. H. S. Overkleeft
	Prof.Dr. P. C. Driscoll (National Institute of Medical Research, Medical Research Council, U.K.)

“Dream is not that which you see while sleeping, it is something that does not let you sleep”

Dr. A.P.J. Abdul Kalam, Wings of Fire

To all who made this possible

Table of Contents

	List of abbreviations	8
Chapter 1	General Introduction	9
Chapter 2	Discovery of small molecule fragments that inhibit the DNA binding of the TEL-ETS domain	53
Chapter 3	Inhibition of small GTPases by stabilization of the GDP complex, a novel approach applied to Rit1, a target for rheumatoid arthritis.	75
Chapter 4	Rapid protein-ligand costructures from sparse NOE data	135
Chapter 5	General Discussion	159
	Summary	171
	Samenvatting	177
	Publication list	183
	Curriculum Vitae	184

Abbreviations

ATP	adenosine triphosphate
ADMET	absorption,distribution,metabolism, excretion, toxicology
CPMG	Carr-Purcell-Meiboom-Gill
CRINEPT	cross-correlated relaxation enhanced polarisation
CSP	chemical shift perturbation
DTT	dithiothreitol
EDTA	ethylenedimaminetetraaceticacid
EPR	electron paramagnetic resonance
FBDD	fragment based drug discovery
GTP	guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	heteronuclear single quantum coherence
Hsp90	heat shock protein 90
HTS	high throughput screening
ILOE	inter-ligand overhauser effect
IPTG	isopropyl β-D-1-thiogalactopyranoside
MALDI-TOF	matrix-assited laser desorption/ionization time of light
MTS	[(1-Acetyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate]
MTSL	[(1-Oxyl-2,2,5,5-tetramethylpyrroline-3-yl)methyl]-methanethiosulfonate
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser effect spectroscopy
PDB	protein data bank
PRE	paramagnetic relaxation enhancement
PCS	pseudocontact shift
RF	radio frequency
RMSD	root mean square deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
STD	saturation transfer difference
TCEP	tris(2-carboxyethyl)phosphine
TINS	target immobilized NMR screening
TOCSY	total correlation spectroscopy
Tris-HCl	trisaminomethane hydrochloride
TROSY	transverse-relaxation optimized spectroscopy

Chapter 1

General Introduction

Drug Discovery

Discovery of new medicines has transitioned from serendipity to rationality over a period of time. A drug may be referred to as a substance that is either used in diagnosis, prevention or treatment of a disease and which carries out a physiological effect when ingested or otherwise introduced into the body. A drug carries out its action by binding to a therapeutic target.[1] The pharmaceutical industry today invests between 10-20% of annual sales revenue in research and development, far greater as compared to other research-based sectors. The discovery of a drug molecule takes about 10 years and these timelines have led to a rise in the financial expenditure, estimated to be more than $500 million. These higher costs are associated with a significant risk, since many drug candidates fail to reach the clinic. Newer strategies are needed at an early stage of the drug discovery process to reduce the risk of failure and successfully identify potential drug candidates. Two main broad types of screening strategies are typically employed to find optimal drug candidates at a preclinical stage - phenotypic screening and target-based screening.

Phenotypic v/s Target based Screens

Phenotypic screening looks at the effects, or phenotypes (a set of observable characteristics of a disease), induced by the compounds in cells, tissues or whole organisms whereas target based screens measures the effect of compounds on a target protein using *in vitro assays*. Phenotypic screening leads to the identification of a small molecule that either modifies or alters disease phenotype by acting on an unknown target or by acting simultaneously on one or more targets. However, the challenge with phenotypic screening is that the subsequent determination of a relevant target or targets that interact with the candidate molecule has proven slow and difficult. [2,3]

The strength of target based screening is that the small molecule screening strategies can be applied against a known target (mostly in high-throughput formats). One can also apply molecular knowledge to investigate specific mechanisms such as if a binding of a drug results in an inhibition or activation of the target protein. Recent advances in molecular biology and chemical genomics have led to the identification of novel drug targets that are implicated in a number of diseases. As a result, phenotypic screens are now largely replaced by target based screens. The initial stage of target-based drug discovery programs consists of many sequential and iterative steps as illustrated in Figure 1.[4,5] Most pharmaceutical companies carry out multiple target-based screens in a drug discovery pipeline to achieve desired success in the drug development. [2,3]

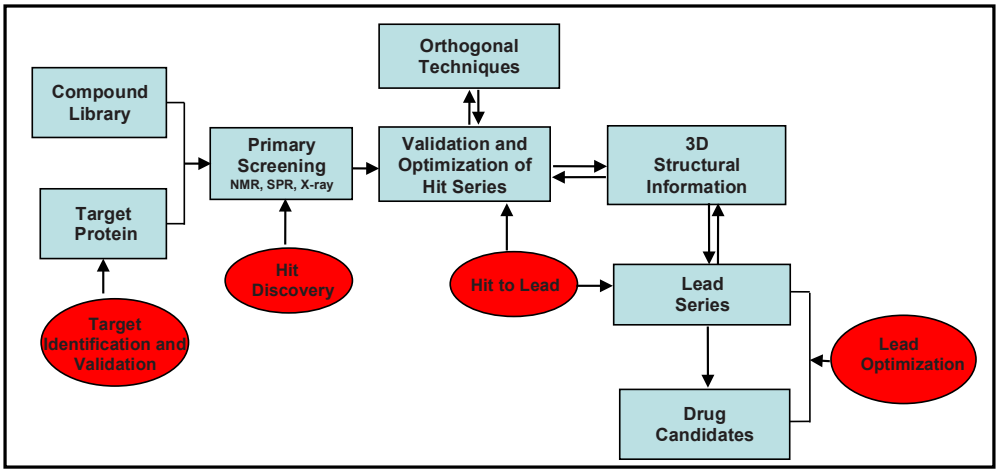


Figure 1. Outline of initial stages involved in a target based drug development.

A. Target Identification/Validation

One of the important steps in developing a new drug is the identification of a protein target and validating its role in a particular disease. A “target” is a protein

whose activity in the cell is associated with the onset or progression of a particular disease. A 'ligandable' target is one which is accessible to the putative drug molecule, be that a small molecule or larger biologicals. The understanding of fundamental processes and cellular networks associated with the target protein and cellular changes caused upon activation/inhibition of the target has been the underlying approach to identify suitable targets for drug intervention. Validation of new drug targets is the process of physiologically, pathologically and pharmacologically evaluating the effects of a molecule on a protein target implicated in disease. Target validation can be performed at a molecular, cellular or the whole animal level.[4,5]

B. Hit Discovery

The main goal of the hit discovery stage is to identify small molecules or hits that harbor the potential to modulate the functional activity of the target protein. A hit is defined as a compound which exhibits desired activity in a small molecule screen against a target protein and whose activity can be confirmed upon retesting. Many potential screening approaches exist to discover hits. Most commonly used approaches are high-throughput screening (HTS), fragment and/or knowledge based screening.[6,7]

High-throughput screening (HTS) is the process of testing a large number (at least 10's of thousand to a few million) of diverse 'drug-like' or 'lead-like' chemical structures (molecular weight between 250 - 600 Da) against disease targets to identify hits using binding, enzymatic or cellular assays.[8-10] The emphasis in HTS is to select those compounds that bind to the target protein with a higher potency/binding affinity (typically <1 μM). Therefore, desired compounds need to make a sufficient number of appropriate interactions (such as hydrogen bonding,

hydrophobic and ionic interactions etc.) within the active site of a protein target.[11,12,13] However, HTS has potential limitations. As HTS involves screening of large compound libraries, it becomes difficult to monitor the quality of compounds and manage the chemical diversity space of the HTS library (the chemical diversity space refers to the extent of variety in the atomic composition within a set of compounds). Inevitably, HTS screening decks may contain molecules that are not drug-like i.e. highly lipophilic and with poor aqueous solubility. As a result, hits that come from HTS consist of a large number of false positives (compounds that cause aggregation, are reactive molecules or redox active), false negatives and compounds with poor ADMET properties (absorption, distribution and metabolism). The inappropriate physico-chemical properties of compounds have led to high attrition rates during drug development (the attrition rate reflects the level of loss of new candidate drugs during the process from pre-clinical to clinical and through their clinical development).[14-16] The compounds with higher molecular weight and lipophilicity are the main drivers for attrition of molecules as they directly influence the ADMET properties. To reduce attrition rates, Lipinski and coworkers[17] have proposed the famous “Rule of Five” (defined below).[17] The rule provides the framework to develop drugs with better aqueous solubility and oral bioavailibility. The rule of 5 is the outcome from the analysis of physico-chemical properties of more than 2000 drugs and drug-like molecules in clinical trials. The rule concludes that a drug-like molecule is more likely to be membrane permeable and easily absorbed by the body if it matches the following criteria:

- Molecular weight < 500
- The compound's lipophilicity, expressed as a quantity known as Clog*P* (the calculated logarithm of the partition coefficient between water and 1-octanol),

< 5.

- The number of groups in the molecule that can donate hydrogen atoms to hydrogen bonds (usually the sum of hydroxyl and amine groups in a drug molecule) < 5.

The number of groups that can accept hydrogen atoms to form hydrogen bonds (estimated by the sum of oxygen and nitrogen atoms) < 10.

Another potential limitation of HTS is that large compound libraries represent only a tiny fraction of chemical diversity space. It is estimated that there are about 10^{60-200} possible drug-like compounds of HTS size (250-600 Da) while there are only approximately 10^{9} possible molecules with 11 or fewer heavy atoms (C, N, O and F). This suggests that screening of 1,000 low molecular weight molecules (< 16 heavy atoms per compound) might sample total chemical space more effectively than screening 1,000,000 more typical, higher molecular weight HTS compounds (< 36 heavy atoms per compound). This poor sampling by HTS libraries has limited the confidence in finding of good starting points for subsequent optimization and development. There are various examples in several drug discovery projects where HTS has failed to generate meaningful potential hits. There is a constant need for alternative approaches so as to overcome the problems posed by HTS.[18,19]

The fragment-based drug discovery (FBDD) approach is able to overcome the limitations posed by HTS and is an established method used within the pharmaceutical industry to develop drugs against a variety of diseases. FBDD involves the generation of very small molecular weight compounds (fragments) libraries that are screened at high concentrations. These fragments are then elaborated or grown into potent drug molecules. FBDD combines a stepwise medicinal chemistry approach and takes into account the structural aspect of the

biological targets to enable efficient hit to lead development (a lead molecule is a chemically optimized fragment hit with a better binding affinity to the protein target and is a more drug-like compound). FBDD is the primary approach in the work described in this thesis and will be described in detail in a later section.[20]

Focused or knowledge-based screening involves selecting from the chemical library smaller subsets of molecules that are likely to have activity against the target protein. This selection of molecules is based on the prior knowledge of the target protein from literature and from the chemical classes that are likely to have activity at the drug target.[5-7]

C. ***Hit to Lead and Lead optimization***

Once a set of hits is obtained from compound screening, the next step is to narrow down which compounds are the best to progress. The initial refinement or also termed as “hit validation” is to generate dose-response curves in a suitable assay for each hit.[4] A validated hit should ideally act reversibly with the drug target. It is important to initiate a drug discovery program with small simple molecules as the follow-up medicinal chemistry efforts tend to improve the potency at the expense of an increase in the molecular weight of a compound.[8,9] The goal of the lead optimization phase is to maintain favorable ADMET properties while improving the deficiencies (chemical groups that are not critical for binding affinity) of the lead structure. The success of drug discovery programs largely depends on the successful development of lead series as these are pursued as potential drug candidates for subsequent studies. The availability of 3D structural information on the target protein-small molecule complexes is crucial at this stage of drug development as its inclusion allows developing lead compounds with better potency and favorable physico-chemical properties.

Fragment Based Drug Discovery

Fragment based drug discovery (FBDD) is an established method. FBDD has significantly developed since last 10 years and a large number of pharmaceutical companies and academic groups are now actively involved. Despite being only few years in existence, FBDD has been able to deliver drug candidates in a timely fashion and there are approved drugs already on the market in addition to several clinical and pre-clinical drug candidates against a variety of protein targets. An extensive list of fragment-derived compounds has entered the clinical trials from various pharmaceutical companies. The current list contains 16 compounds in Phase I, 11 compounds in Phase II and 1 compound in Phase III clinical trials. One drug discovered from the fragment-screen has received FDA approval and is marketed under the name *Zelboraf*. The drug was discovered at Plexxikon Inc. and developed in partnership with Roche. The drug has shown dramatic clinical results and extends life of patients with a deadly form of skin cancer. More details on the clinical trial progression of fragment derived compounds is reviewed elsewhere.[20,23]

FBDD starts with screening of small molecules, called "*fragments*" i.e. the minimal recognition motifs or molecular anchors.[20] The origins for FBDD approach can be traced to a paper by Williams Jencks, in which it was proposed that weakly binding fragments can form high quality interactions (high binding energies per unit of molecular mass) with the target and later these fragments can be optimized to deliver highly potent lead-like molecules. Nakamura and Abeles applied FBDD approach and demonstrated that indeed it was possible to obtain potent HMG-CoA inhibitors when starting from weakly binding fragments.[21,22] The fragment based drug discovery process, in general consists of three stages:

1) Fragment library - in which a fragment collection is assembled

2) Fragment screening- in which the fragment library is screened on a purified protein target using an array of biophysical techniques that are able to detect weak, non-covalent binding to the target of interest and

3) Fragment elaboration- during which validated fragments are developed in lead compounds guided by structure based drug design (SBDD) and biochemical data.

Fragment based approaches offer a number of attractive features: (i) a significantly larger proportion of chemical space can be sampled within a fragment library (usually ~10^3 fragments) than with the ~10^5–10^6 larger molecules typical for an HTS campaign. As a result, less number of compounds (about a few hundred up to a few thousand) are typically screened against the target protein (ii) fragments are small and have a greater probability of correctly matching the binding site of the target protein by forming high quality interactions. As a large number of atoms in a fragment hit are involved in direct protein-binding interaction, fragments are considered to be highly ligand efficient binders {Ligand efficiency (LE) is a measurement of the binding energy per atom of a ligand to its target protein and is a valuable metric to small molecules with different sizes} (iii) the chemical optimization of fragment hits (parameters such as potency, target selectivity, ADMET properties and LEs > 0.3) can be achieved when the protein-ligand binding interaction is structurally validated and (iv) fragments in a library are chosen such that they exhibit good aqueous solubility and should lead to fewer false positives arising from aggregation, a common problem encountered in HTS programs.[18,19,23,24]

Fragment Library

Although Lipinski's rule of 5 provides useful guidelines to maximize an oral drug candidate success in the drug development, it may not be relevant to assess optimum properties of lead-like molecules. It has been reported that the libraries containing compounds with molecular weight 100-350 Da and clogP of 1-3 do result in hits which can be optimized into lead molecules with favorable drug-like properties. This suggests that smaller is better for efficient drug development. Fragment molecules in a library typically are compliant with the "Rule of Three" as proposed by Congreve and colleagues.[24,25] Rule of 3 is used as the selection criteria and include physico-chemical characteristics such as molecular weight < 300 Da (~150-300 Da), fewer number of heavy atoms and a limited number of hydrogen bond donors ≤ 3 and the number of hydrogen bond acceptor ≤ 3. Other criteria include the solubility, ClogP ≤ 3 and the number of rotatable bonds ≤ 3.[23,24]

Fragment Screening

Fragment hits are simple molecules and tend to bind weakly to the target protein. The typical binding affinities exhibited by fragment hits range between 0.01-10 mM. Hence, to detect weak fragment hits, sensitive biophysical techniques are required. NMR (Nuclear Magnetic Resonance) and X-ray crystallography are commonly used techniques as they are able to detect hits within a range of binding affinities. The application of X-ray crystallography as a screening tool depends on number of factors such as availability of a large amount of protein, and access to synchrotron and time involved to screen for crystallization conditions.[26,27] Other techniques like mass spectrometry, high concentration functional screening, calorimetry, surface plasmon resonance (SPR) and NMR based screening methods

may be easier to set-up for small molecule screening. [28-32] Significant importance should also be laid on choosing an appropriate biophysical technique for screening of fragments. For example, as noted by Jhoti and colleagues, only about 5% of the fragment population needs to interact with the protein to be detected as an NMR hit, whereas in X-ray crystallography experiments a fragment needs to have at least 70% occupancy of the binding site to be defined as a hit. Also, NMR can detect hits with solubilities lower than their potency for the target protein.[33] To avoid the loss of any potential hits that come from a particular screening method and to aid in reliable identification of hits, often two or more techniques are employed early in the FBDD process.

NMR as a screening tool in FBDD

The popularity of NMR as a screening technique in the drug discovery process is increasing due to its sensitivity for the detection of the low affinity compounds. This section will mainly focus on NMR methods that are capable of detecting binding of small molecules to a protein target by screening of a compound library. NMR based screening can be implemented as ligand- or protein-detected methods.[34] It is necessary to introduce some basic NMR concepts for a better understanding of the methods described in the later sections of this chapter.

Basic Concepts for Ligand detected NMR Methods:

Magnetization

The nuclear magnetic resonance (NMR) spectroscopy experiment involves using energy in the form of electromagnetic radiation to transit the excess alpha nuclei (low energy ground state) into the beta state (high energy excited state). The energy in the form of radio waves is appropriate for the low energy transition involved in NMR.

This energy is at a specific resonance frequency that depends on the strength of the applied external magnetic field and the magnetic properties of NMR active nuclei. The term 'magnetization' in NMR is simply the sum of all the individual nuclear magnetic moments possessed by respective nuclear spins in presence of the externally applied magnetic field. There is a very small energy difference between α (low energy ground state) and β (high energy excited state) energy states of a nuclear spin orientation in a magnetic field and this results in a very small excess population of nuclei in the ground state (Boltzmann distribution). For example, the population difference is only on the order of 1 in 10^5 for ^1H spins in an 11.7T magnetic field. It is this small difference in the population that is responsible for an NMR signal.[35.36]

Relaxation

Relaxation in NMR is an important process and an understanding is required for proper measurement of the NMR spectra. Any spin system that is not in the equilibrium state will relax back to its Boltzmann distribution. This happens via two mechanisms called *spin-lattice relaxation* and *spin-spin relaxation*. The spin-lattice relaxation is a process by which the spins exchange energy with their surrounding medium. This can be pictured as a movement of the bulk magnetization of spins back into the direction of the external magnetic field. It is therefore also called *longitudinal relaxation* or T_1*-relaxation*. The spin-spin relaxation is characterized by the loss of coherence among the spins. The spin-spin relaxation is also called *transverse relaxation* or T_2*-relaxation*.[35,36]

Chemical Exchange

As the binding of a small molecule ligand to the protein requires exchange between the free and the bound states, the resulting binding kinetics and exchange

rates affect the properties of the NMR spectra. The interconversion between the bound and free states is dependent on the Larmor frequency of the observed nucleus and field strength.If the exchange is slow i.e. in *slow exchange* as per NMR time scale (NMR time scale depends on the strength of the magnetic field and a particular experimental set-up), two separate resonances can be observed for the free and the bound state. In the intermediate exchange regime, target resonances that are sharp/intense at low ligand concentrations, broaden and sometimes disappear as the ligand concentration is increased. These broadened resonances reappear at high ligand concentrations and perhaps exhibit a small chemical shift. In “fast exchange” regime the spectral characteristics in the bound state are transferred to the free state of the ligand. The observed resonances are the population weighted averages of the signals of the free and the bound state and a single sharp signal can be detected.[35-37]

Intrinsic NMR characteristics of a protein and a small molecule ligand

The detection and characterization of protein-ligand interactions require different NMR techniques depending on the binding affinity, molecular weight and chemical exchange between free and bound state. A protein target exhibits characteristic properties such as slow diffusion, fast relaxation due to slower tumbling and fast spin-diffusion (spontaneous exchange of magnetization among nearby nuclear spins). A small molecule ligand possesses opposite properties such as fast diffusion, slower relaxation and negligible spin diffusion. In a protein-ligand complex, the properties of the bound ligand become similar to that of the protein. Acquisition of 1D ^{1}H NMR spectra of a ligand in the presence and absence of the protein may indicate binding via broadening of the resonance signal and/or loss of signal intensity of the ligand resonances. Most NMR assays exploit these differences in the properties of the ligand caused by its binding to the protein.[36,37]

Ligand detected NMR Methods

NMR is a sensitive technique for probing the binding interaction of target protein with a small molecule or fragment using ligand-based NMR detection methods. Ligand-observed NMR methods are routinely used to generate hit matter (a primary set of compounds that bind the target protein). There are a number of ligand detected methods that are based on the acquisition of simple 1D ^{1}H NMR experiments. There are two ways to detect ligand binding by ligand-detected NMR: 1) exploiting the difference in the tumbling of the ligand in the presence and absence of the protein targets (ligands that are bound to the protein will experience slower mobility and altered relaxation parameters). CPMG experiments and TINS (Target Immobilized NMR Screening) are good examples.[38,39,40] In CPMG (Carr-Purcell-Meiboom-Gill) experiment, a simple measure as the observation of a reduction in the intensity of the ligand proton resonance observed only in the presence of the protein target is considered as an indicator for ligand binding.[38,39] TINS (Target immobilized NMR screening) is one of the technologies used extensively in the work described in this thesis and will be described in detail below.[40] 2) the transfer of ^{1}H magnetization from protein to the bound ligand (only ligand molecules that bind to the protein will experience the magnetization transfer). STD and WaterLOGSY are routinely used techniques that exploit the magnetization transfer.[41,42] There are a number of approaches that are developed to detect ligand binding such as diffusion editing, ^{19}F fluorine screening and competition binding studies (provides information on the binding site) that can be implemented but will not be discussed here.[26,43]

TINS

TINS exploits the enhanced transverse relaxation rate (measure of how fast the spins exchange energy in transverse plane and this is responsible for a true

linewidth of an NMR signal) of the ligand caused by binding to a protein target immobilized on a solid support. The difference in the transverse relaxation rate between protein bound and free state of ligand is at least 2 orders of magnitude. Use of a reference protein along with the target protein eliminates those fragments that exhibit non-specific interactions to the protein surface. Target and reference proteins are immobilized on a commercially available resin. A flow-injection, dual-cell sample holder is placed into the magnet into which repeated cycles of mixtures of fragments about 3-6 fragments per mix are injected simultaneously on both the immobilized target and the reference cells. After injection, flow is stopped and NMR data are acquired followed by extensive washing prior to the next injection. In TINS, a small quantity of protein target is used and a fragment library of ~2000 compounds can be screened in less than 5 days. The change in the signal amplitude caused by the interaction of the fragment with either the target protein or the reference protein is termed as the TINS effect.[12,40,44,45] TINS can also be set-up in competition mode allowing one to rapidly characterize the ligand binding site using a known competitor. TINS NMR screening has been successfully applied to diverse classes of protein targets including kinases, viral RdRP's (RNA dependent RNA polymerase), GTPases and also to challenging membrane proteins. The technique is sensitive, robust and efficient for the detection of weak binders.

STD and WaterLOGSY

The saturation transfer difference (STD) experiment is widely adopted for screening purposes and was developed originally by Mayer and Meyer.[41] The STD experiment is the difference between two separate experiments. In a first experiment so called *"on resonance"*, the protein proton magnetization is saturated (saturation pulse equalizes the population difference between the ground and the excited state

and as a result no NMR signal is observed) by a train of selective rf (*radio frequency*) pulses. The saturation is placed on the isolated resonances of the protein, usually around the methyl region (~0.0 to -1.0 ppm) such that these do not overlap with the ligand resonances. The selective saturation of e.g. the methyl ^{1}H's is then transferred to nearby protons within the protein *via* intramolecular ^{1}H-^{1}H cross relaxation pathways, referred to as spin diffusion. When a ligand binds to the protein, the saturation is transferred via intermolecular ^{1}H-^{1}H cross relaxation at the binding interface. These saturated ligand molecules upon dissociation from the target protein are exchanged back into solution where their saturated state persists. As more ligand molecules are exchanged *on* and *off* the protein, the population of the saturated ligand builds up in solution. In another experiment so called *"off resonance",* the saturation is applied far away from protein resonances, e.g. at 100 ppm, such that no saturation of the protein is observed. The "*off resonance"* spectrum is used as a reference. The *"on resonance"* and *"off resonance"* experiments are acquired in an interleaved fashion and then subtracted. The resulting difference spectrum between the *"on resonance"* and "*off resonance"* experiments yields only those ligand resonances that experience saturation arising exclusively from the ligand binding to the target protein. STD experiments have several advantages: i) STD experiments can be carried out with less protein concentration (~1-5 μM) compared to other ligand detected NMR methods, (ii) compounds can be screened in mixtures (iii) ease of implementation and (iv) applicability to large molecular weight targets.[43] Mayer and co-workers[46] have also demonstrated that STD can be used for group epitope mapping in which only those specific protons of the ligand that are closer to the protein surface experience a higher degree of saturation. This type of information can be vital to medicinal chemists to guide ligand elaboration.[37,46]

There are potential pitfalls with STD technique that need to be considered before implementing to a protein-ligand system. The STD effect is limited by the exchange regime. To get maximum sensitivity of STD signal, ligand has to dissociate at a faster rate, this is caused by the difference in the relaxation rate of the ligand when free and when bound to the protein target. During STD experiment, one assumes that protein is 100% saturated. However, this may not be always true as the saturation is also lost to solvent protons and other saturated protein protons. Also, some protein targets are suboptimal for STD, in which another similar approach, WaterLOGSY (WaterLOGSY (Water-ligand Observed via Gradient SpectroscopY) may be more effective.[41,43] In WaterLOGSY experiment, the saturation pulse is applied at the resonance frequency of the bulk water. The transfer of saturation is achieved from water to the protein target and subsequently from protein target to the ligand. This transfer relies on the presence of bound water molecules within ligand binding site of the protein. It is to be noted that the ligand observed NMR techniques are dependent on the “fast exchange regime” between the ligand and the target protein. A ligand that is bound tightly is in “slow exchange” and is not suitable for detection by ligand observed NMR methods.[37,43]

Fragment Elaboration and Structure Based Drug Design

Typically, FBDD campaigns are combined with structure-based drug design (SBDD). SBDD has emerged as a new tool in medicinal chemistry. Identification of initial fragments from a direct binding assay is most useful if it is also supported by structural information such as the binding site of the fragment. The initial fragment hits can be considered as building blocks of a complex series of lead compounds. The evolution and growing of fragment hits to tighter-binding molecules can be

achieved by designing of small subset of compounds that make additional specific interactions within the binding site of the target protein. SBDD requires the three dimensional structure or closely related structure of a homolog or a NMR derived structure of the target protein, preferentially complexed with a ligand. The 3D structure of the protein-ligand complex reveals the binding mode and the conformation of a compound under investigation and indicates the essential molecular interactions determining its binding affinity. Small molecules could be modeled into a binding pocket of the drug target using various computational tools.[37-39,51] As noted by Hajduk and Greer, the inclusion of structural information derived from methods such as X-ray crystallography and NMR could dramatically influences the success of fragment based drug design.[24] The ability to increase the potency of inhibitors nearly triples with the aid of structure-based design. The development of selective inhibitors for cyclin-dependent kinase-2 (CDK-2), Src-kinase, matrix metalloproteinase 3 and Hsp90 are very good examples where SBDD have helped to achieve potent compounds.[52-55] Fragment-based screening coupled with structure-based drug design provides a powerful combination for maximizing the representation of chemical diversity space and generating novel, potent inhibitors for various protein targets.

The FBDD approach has been highly successful and to date at least 25 drugs derived from FBDD and have entered the clinical trials. One drug discovered from the fragment-screen has received FDA approval and is marketed under the name *Zelboraf*. The drug was discovered at Plexxikon Inc. and developed in partnership with Roche. The drug has shown dramatic clinical results and extends life of patients with a deadly form of skin cancer.[20,23,47] A particularly impressive example was the development of inhibitors of the protein–protein interaction between Bcl-X_L and its

partner proteins. The initial work was published in 2005 showing how very potent inhibitors (< 100 nM) were discovered starting from the identification and subsequent NMR structure determination of two weak fragments with 300 μM and 4 mM affinities, respectively.[49] The example shows how FBDD has the potential to deliver drugs even for protein-protein interaction targets. A more recent application was the identification of small molecule binding to a novel site on the protein survivin.[50] The structure of an NMR-derived protein–ligand complex was determined for one of the fragments obtained from a fragment screen and subsequent optimization allowed the identification of compounds with affinities of < 100 nM. These compounds are suitable probes for understanding the role of the novel binding site in cancer biology.[48,50] About 13 different institutions reported the development of more than 50 potent compounds (IC_{50} < 100 nM; IC_{50} is the concentration of an inhibitor required for 50% inhibition of an enzyme) against different protein targets starting from weakly binding fragments.[22] There are many successful examples of application of FBDD to obtain high affinity leads which are reviewed elsewhere.[20,21]

X-ray Crystallography as a Structural Tool

To date, X-ray crystallography has been the main driver for structure determination purposes. Soaking of small molecules into protein crystals is a successful approach to obtain high resolution 3D structural information of the protein-small molecule complexes. There have been many examples where the use of X-ray crystallography has aided the successful discovery of nanomolar potent inhibitors against protein targets such as p38 MAP kinase, β-secretase, urokinase, Src-kinase, cyclin dependent kinase-2 (CDK2) to name a few.[56] The X-ray structure of the fragment bound to the protein provides the final binding evidence and, in addition, it

delivers the relevant structural information for initiating lead optimization via medicinal and/or combinatorial chemistry attempts. However, in practice, it is not always possible to obtain protein-small molecule costructures because of different experimental causes. The set-up for X-ray based structure determination is not trivial and is both resource and time intensive. Common problem is faced with the interpretation of electron density maps determined from X-ray diffraction experiments (An electron density map is a three-dimensional description of the electron density in a crystal structure, determined from X-ray diffraction experiments). This can be ambiguous and even at resolution of 1.5 to 2.5 Å, there are uncertainties in the placement of amino acid residues like asparagine, glutamine and histidine because of their internal pseudo-symmetry. In the case of asparagine and glutamine, the side-chain N and O atoms will have similar electron densities, and in the case of histidine, the N and C atoms of the imidazole ring will usually be indistinguishable (and consequently the side-chains of these residues can typically be built in two orientations). Errors in the placement of ligands (including fragments) in macromolecular crystal structures can also arise from several causes. Non-covalently bound compounds may exhibit greater thermal motion or conformational disorder than the surrounding protein, leading to poor electron density.[57-60] Artifacts may be generated by the crystallization process itself. Potential blocking of the target site by crystal contacts (interchain or intermolecular contacts that occur solely as a result of crystallization) can result in false-negatives upon soaking ligands. Similarly, residues surrounding the site of interest could be held in an inappropriate conformation for ligand binding, or could be blocked by other ligands.[61] Additionally, crystals that are grown at extremes of pH may not yield ligand-binding modes observed at physiological pH due to protonation/deprotonation of susceptible side-

chains. Also, not all protein targets can be crystallized and may not be amenable for crystallography set-up. NMR spectroscopy has proven to be an alternative choice and in some cases an appropriate technique to obtain 3D structural information on the protein-ligand complexes.

NMR Spectroscopy as a Structural Tool

NMR plays an important role in the process of identification of fragment hits and developing them into high affinity and selective compounds. Protein-detected NMR methods can be implemented to provide critical 3D structural information in a timely manner to advance compounds through the fragment hit-to-lead stage. NMR offers broad capabilities that can suit the type of information needed. Protein based NMR methods compares the changes in the NMR parameters of the protein resonances in the presence and the absence of compounds. Such methods are not only capable of detecting the ligand binding but also provide structural information on where a ligand binds on a protein. Protein-detected experiments usually require isotope-labelled (^{15}N-labelled or ^{15}N/^{13}C labelled) target protein at higher concentrations (0.1 - 1 mM), but can afford high-resolution structural data about the protein and the complex. Unfortunately, there is an upper size limit for proteins (30-60 kDa, depending on the isotope labelling and the spectrometer) whose resonances can be observed and assigned by NMR.[35-37]

Basic concepts for Protein detected NMR methods

Isotope Labeling

As the NMR phenomenon relies on the existence of nuclear spin, nuclei with an even mass and atomic number are NMR inactive (not visible in a NMR spectrum).

For study of biomolecules, the most important nuclei are with spin quantum number, *I* = 1/2 are ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P. For proteins, the overlap caused by the number of ^{1}H resonances can be resolved by using heteronuclear correlation NMR methods (e.g. ^{1}H-^{15}N and ^{1}H-^{13}C HSQCs) achieved by the isotopic enrichment of nitrogen and carbon isotopes in a protein sample. Another important nucleus often employed in protein sample preparation is the deuteron, ^{2}H. Replacing protons with the deuterons in a protein slows down the relaxation process and simplifies the NMR spectrum by reducing the proton density.

Resonance Assignment

The prerequisite to any detailed protein-detected NMR study is the resonance assignment of the target protein. The resonance assignment refers to associating each resonance signal in a NMR spectrum to a specific nuclear spin. Multidimensional NMR experiments allow carrying out the backbone and side-chain assignment procedure by making through bond connections between the protons, nitrogen and carbon atoms, thereby linking all the atoms of the entire protein amino acid chain. An example of an assigned 2D [^{1}H, ^{15}N] HSQC spectrum of N-terminal ATPase domain of Hsp90 in which each resonance signal is associated to backbone nuclei (^{1}H and ^{15}N) of an amino acid residue in the protein sequence. This work was carried out in Chapter 4 of the thesis and HSQC spectrum with assignments is shown in Figure 2. The combination of ^{15}N, ^{13}C and ^{2}H isotope labeling has expanded the size of proteins amenable to NMR analyses.[36,37]

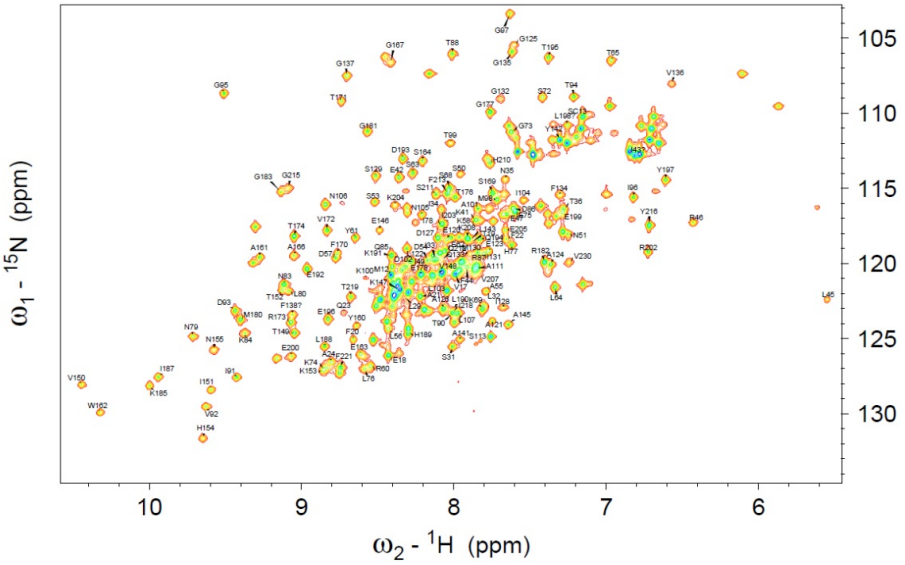


Figure 2. A 2D [^{1}H,^{15}N] HSQC spectrum of N-terminal ATPase domain of Hsp90 displayed with backbone (^{1}H and ^{15}N) resonance assignments.

The following sections describe various NMR methods that are capable of delivering 3D structural information on a protein ligand complex and also mention some recent developments in NMR that will give an overview of the research work performed in this thesis.

Low Resolution Structural Information by Protein Observed NMR

Advancing fragment hits with high micromolar to low millimolar binding affinities can be achieved efficiently if there are readily available 3D structures, X-ray or NMR, of the target protein and fragment hits. One particularly powerful and efficient method to map the ligand binding site at low resolution is through the use of NMR based chemical shift perturbation (CSP) data. The CSPs are typically obtained using a heteronuclear single quantum coherence spectroscopy (HSQC) experiment. Some details of the process are provided below.

The [^{1}H-^{15}N HSQC] spectrum of a protein contains resonances which correspond to the amide group of each amino acid residue except proline. Of all the different NMR parameters that change upon complex formation, CSPs are easiest to measure and are routinely employed to determine the binding location of the ligand. When a ligand is titrated, the amide group of amino acids within the protein that are close to the ligand will experience a change in their local chemical environment. This change (a CSP) is translated by a shift in the position of resonance signal for the respective amide group of the amino acid. A chemical shift map is generated which represents CSPs in a quantitative manner (the combined magnitude of ^{1}H and ^{15}N chemical shift differences for an affected residue into one parameter). This map allows identifying those amide groups whose environment is most affected due to the binding of a ligand. The map will also include those residues that are indirectly affected by ligand induced changes in the protein. These CSPs result mainly from the increased sensitivity of amide groups to pH and/or small changes that occur in the hydrogen bonding patterns of protein backbones upon ligand binding. If a three dimensional structure (X-ray or an NMR structure) and resonance assignments for the target protein are available, CSP mapping will show a clear surface patch of affected residues on to the protein structure. This can become a reliable indicator of the binding location of a ligand.[36,37,43]

The CSP data usually shows clearly where a ligand binds to the target but structural information obtained by such an approach is often of low resolution and not of sufficient detail to calculate precisely the orientation of a ligand in the binding site, the level of detail that would be obtained by a complete structure determination of the protein –ligand complex. Still, there is valuable information, and using the CSP information to perform restrained docking (inclusion of CSPs as ambiguous restraints

in a docking program) can be a first step towards utilizing shift information for a protein–ligand costructure.[62,63] Many computational programs have been implemented to localize a ligand binding site based on purely on the CSP analysis. McCoy and Wyss have developed a program *j*-surf based on the fact that since many drug molecules have aromatic rings, chemical shift perturbations are in part caused due to ring current shifts induced by ligand.[64] By quantifying the spatial dependence of ligand ring current effects and local magnetic fields of the neighboring spins, ligand binding site on the protein can be accurately characterized.[43,63,64]

A nice extension of chemical shift mapping called "SAR (structure-activity relationships) by NMR" has been developed by Fesik and co-workers.[65] In this approach a library of fragments can be screened to identify those molecules that bind to two distinct but neighboring binding locations. These weakly binding molecules (or their analogs) can be linked into one that binds with much more higher affinity to the protein. This approach was successful in the finding of high affinity small molecule inhibitors of drug targets like BACE-1, FKBP, stromelysin, urokinase and many more.[20,23,65]

High Resolution Structural Information on Protein-Ligand Complexes by NOEs

NMR has become a firmly established method for determining the three dimensional structures not only of proteins but also of protein-protein, protein-nucleic acid and protein-small molecule complexes. The focus in this section is to introduce recently developed NOE based applications of NMR to obtain structural information on the location of the binding site and the conformation of the bound ligand.

Structure determination by high resolution NMR has traditionally relied on the

use of Nuclear Overhauser Enhancement (NOEs) derived distance restraints. Structures of proteins up to 20-30 kDa and protein-small molecule complexes can be determined successfully by NMR. NOEs provide a mechanism for both inter- and intramolecular magnetization transfers. The magnitude of the NOE enhancement between two nuclei spins is inversely related to the internuclear distance (r^{-6}) between them. Therefore, NOE related experiments have been widely used for determining three dimensional structures of protein and protein–ligand complex as well as for deriving dynamic information for the protein-ligand interactions.[43,36,37,66]

trNOEs, ILOEs and INPHARMA

One of the ways to obtain the conformation of the bound ligand is by measurement of transferred NOEs (trNOEs).[66] Protein targets have long correlation times due to large molecular weight. This allows rapid build up of NOE and extensive spin diffusion. By contrast, ligands are small molecules and have slow NOE buildup and negligible spin diffusion. This implies that if a NOESY experiment is carried out on a protein-ligand complex, in the presence of excess amount of ligand, NOEs within the free ligand develop very slowly, whereas NOEs within the bound ligand develop much more rapidly as it is in complex with the protein. The exchange of the ligand between the bound and free state will produce free ligand (with intense signals and chemical shifts at the positions of free ligand) displaying NOEs characteristic of the bound state. This is a very useful experiment, since it provides conformational information on the bound ligand but the information is measured from the easily observed and assigned free ligand signals, and is unambiguous.[67] Clear advantages of this approach include 1) no requirement of prior information about the target protein 2) consumption of less amount of protein (~20-50 μM), and 3) ease of

spectral analysis as the observed ligand resonances are of much higher intensity compared to the protein resonances. This is due to presence of ligand concentration of at least 5-10 fold excess to that of the protein. Another experiment, also called an ILOE (interligand overhauser effect) relies on the transfer of the magnetization between two nuclei on different ligands which are known to occupy adjacent pockets on the protein. ILOEs were demonstrated first for a ternary complex of coenzyme A, chloramphenicol and chloramphenicol acetyltransferase.[69,70] Sledz *et al*. have demonstrated the ILOE approach in fragment based inhibitor design.[70] The main limitation of ILOE experiments is that magnetization transfer pathway caused by spin diffusion to protons of the protein (ligand1-protein-ligand2 instead of ligand1-ligand2) might cause two ligands to appear closer than is in reality. Hence the distance restraints obtained by ILOE measurements should be treated with caution otherwise they may lead to inaccurate structural information.[37]

In contrast to trNOE and ILOEs, a method which uses spin diffusion as a way to determine the relative orientation of two competing ligands in the binding site is INPHARMA (Interligand NOEs for PHARmacophore MApping). Here, the cross-relaxation (magnetization of two different spins that are close (< 5 Å) to each other) is transferred *via* spin diffusion between ligand A and protein protons when ligand A is bound. When ligand A dissociates and ligand B binds in the same binding site, the magnetization is transferred from protein protons to ligand B. In this way structural information on one of the ligands bound to the protein can be obtained provided the structure of the other ligand is available.[71] It should be realized that NOESY experiments such as trNOEs, ILOEs and INPHARMA are only applicable when the given protein-ligand interaction is in the fast exchange regime.

The clearest information on the binding site and orientation of the ligand can be obtained by direct observation of the resonances of the bound ligand. When isotopically labeled protein or a ligand is available, ligand and protein resonances can be observed separately by measurement of isotope-edited or filtered NOESY experiments. These experiments take advantage of the presence of natural isotopic abundance of the ligand and can be applied to protein-ligand systems that are in both fast- and slow-exchange regimes. Intermolecular NOEs are observed by observing one proton dimension which is filtered to protons attached to the ^{12}C to observe ligand resonances and other proton dimension which is ^{13}C edited to select for protein resonances. The structure calculations can be performed using the distance restraints obtained from intermolecular NOEs between the complex, if the 3D structure of protein is known.[36,37,43]

Despite its promise, it is to be realized that NMR structure determination process remains nontrivial as well as laborious and time consuming. It requires acquisition of a suite of multidimensional NMR experiments. It could take about 1-2 months to obtain a complete 3D structure of a protein-ligand complex provided other considerations like sample preparation and stability are optimized. It is also to be realized that the application of NMR is usually constrained due to molecular weight limit of the target protein. The NMR spectra for large proteins typically are of poor sensitivity (broader spectral resonances caused by faster relaxation properties) and resolution (higher spectral complexity arising from the increased number of nuclear spins). Recent adaptations such as (a) the development of cryogenically cooled NMR probes, (b) protein deuteration which result in narrower resonances, (c) implementation of TROSY (Transverse relaxation-optimized spectroscopy), CRINEPT, CRIPT based NMR pulse sequences,[72,73] (d) increase in the lifetime of an

NMR signal by selective ^{13}C labeled methyl group labeling of isoleucine, leucine and valine groups (caused by the long T_2 relaxation times exhibited by the methyl groups) and (e) advances in paramagnetic NMR have partly addressed the limitations imposed by higher molecular weight proteins.[36,37]

Paramagnetic NMR

Paramagnetic based relaxation enhancement is a technique that is based on the interaction of an unpaired electron with a nearby nuclear spin. The electronic relaxation time, (the longitudinal relaxation time of the unpaired electron spin) is much shorter than for protons and typically ranges from microseconds down to picoseconds. There are two ways in which paramagnetic effects can be observed on the nuclear spin. One is from isotropic paramagnetic centers which give only an increase in the transverse relaxation rate (PRE) of nuclei. The other is from anisotropic centers that cause a shift in the resonance of the nuclei (Pseudo Contact Shifts, PCS).

Paramagnetic Relaxation Enhancement (PRE)

Paramagnetic centres with slow electronic relaxation cause strong nuclear relaxation and thus broadening of the resonance signal. This is called paramagnetic relaxation enhancement (PRE).[37,74] Paramagnetic effects are measured by differences between the NMR spectra of a target molecule bound to paramagnetic probe and bound to diamagnetic probe. A paramagnetic center containing an unpaired electron, e.g. a nitroxide radical, is attached via a disulphide linkage to an engineered cysteine residue and invariably causes broadening of the resonances due to the enhanced transverse relaxation rate (T_2) of the nuclei in close proximity. The PRE effect is distance dependent and proportional to r^{-6}, where r is the distance between the

unpaired electron and the nuclear spin.

Pseudocontact Shifts (PCS)

The anisotropy of the paramagnetic effect is described by the magnetic susceptibility tensor, and causes pseudocontact shifts. The PCS is angle/orientation dependent and proportional to r^{-3} distance. The r^{-3} distance dependence of the PCS in comparison to the r^{-6} dependence of the PRE, allows PCSs to be measured for nuclear spins that are far away (> 20 Å). The paramagnetic effects can be converted into distance restraints which can be used to dock the binding partners. This enables new possibilities for the analysis of protein-protein and protein-small molecule interactions.[75]

Application of Paramagnetic NMR Methods

The application of paramagnetic NMR in structural biology is increasingly becoming important as it can provide different levels of structural information.

Application of PRE to Study Ligand Binding

Paramagnetic NMR can be used to obtain low resolution information on the binding site of the fragments as demonstrated by Janke and co-workers.[76] They have demonstrated an PRE-based approach called SLAPSTIC (spin labels attached to protein sidechains as a tool to identify interacting fragments) to obtain 3D structural information. In this approach, paramagnetic, organic radicals such as TEMPO, are covalently linked to the side chain of specific amino acids (lysine, cysteine, methionine, histidine or tyrosine). This approach uses a spin labeled compound as a first-site ligand. Screening of this complex allows identification of compounds that bind simultaneously with the first spin-labeled compound. This is of special interest in

drug discovery and optimization process, because linking of two compounds that bind in proximity can result in compounds with significantly higher affinity. The principal advantages for this approach are its robustness to identify second site binders, low protein requirements and high spectral sensitivity. However, SLAPSTIC requires considerable knowledge of the protein 3D structure, confirmation that spin label attachment does not compromise the binding site on the protein target and the relaxation enhancement only results if two ligands bind simultaneously to the protein target and at neighboring sites.[76] A further extension is also presented in chapter 3, where a PRE based approach is developed for a GTPase to obtain 3D structural information on the binding site of biologically active compounds.

Application of PCS to Study Ligand Binding

Recently, a paramagnetic NMR based approach has been developed by our group in which 3D co-structures of small molecule bound to a protein can now be readily determined using paramagnetic PCS data.[77] Here, a co-structure of the target protein, FKBP12 bound to initial fragment hit was obtained using PCS datasets. These structures allowed determination of the binding site and the orientation of the ligand. The PCS-driven result was then compared and found in close agreement with independently NOE-derived structure of the same FKBP12-ligand complex. A major advantage of this method is no labeled protein is required. Thus, it can be applied to larger molecular weight protein targets that are suboptimal for the resonance assignment procedure.

Selective Isotope Labeling of Methyl Groups

Recent advances in NMR spectroscopy of high molecular weight proteins

have been strongly connected to the development of optimal isotope labeling techniques. Of particular interest are the experimental protocols that have been developed to obtain protein samples that are deuterated but selectively protonated at specific sites (amide protons and certain methyl group containing residues). There are several advantages of using methyl groups as probes, (i) the residues containing methyl groups occur frequently throughout the protein sequence and at ligand binding sites[81] (ii) methyl groups have favorable NMR relaxation properties so that even for protein targets with high molecular weights, NMR spectra are of higher sensitivity and resolution[80] (iii) data interpretation is simplified due to less overlap of the resonances in the NMR spectra and (iv) methods to produce ILV selectively labeled samples in *E. coli* are robust and economical.[79,80]

Applications of Methyl Group Labeling

Tugarinov and co-workers[80] have shown that by selective labeling it is possible to obtain sidechain methyl and amide resonance assignments and calculate the global fold of a protein, Malate Synthase G (MSG), which has a molecular weight of 82 kDa.[80] The global fold of MSG was calculated using NOEs between methyl-methyl, methyl-amide and amide-amide groups. In such an application Hajduk and co-workers[81] have shown that the methyl groups of Leu, Val and Ile residues in a protein can be selectively ^{13}C labeled and $^{13}C/^{1}H$ chemical shift perturbations can be monitored to detect ligand binding. Both these examples show that the selective labeling procedure extends the size of the molecular systems that can be investigated by NMR and methyl groups can be used as probes to detect ligand binding. To broadly extend the approach to study ligand binding by selective labeling, Otten and coworkers[82] have demonstrated an economical way to label methyl groups

of all methyl containing residues (leucine, valine, isoleucine methionine, threonine and alanine) in a protein. This type of labeling achieves a better coverage of the binding site of the protein. However, it should be noted that methods to produce selectively labeled protein samples are currently applicable using *E. coli*.

In order to increase the efficiency of utilizing methyl groups as probes to detect ligand binding, steps to automate the methyl group assignment procedure have been undertaken. These methods are either based on the availability of the crystal structure of the protein and make use of the NOESY experiment or paramagnetic NMR to define the methyl group network. Appropriate paramagnetic tags are placed on the protein surface through engineered cysteine residues to get complete coverage of the PRE effects for the methyl groups within the protein structure. Methods also have been developed that demonstrate the use of through-space paramagnetic effects combined with NOESY experiments to rapidly obtain methyl assignments, if a crystal structure of the protein target is known. [83-87] Similarly, Otting and co-workers[84] as well as Skinner and co-workers[87] have developed tools that use paramagnetic pseudo-contact shifts to directly obtain ILV sidechain assignments.[78,83-87]

Various academic groups have demonstrated that selectively isotope labeled groups of the protein can be used to collect intermolecular NOEs with its binding partner in order to obtain structural information. However, these methods have one of the following limitations: 1) they require prior information of the binding site so as to appropriately label residues on the target, 2) resonance assignment is based on the pattern of observed CSPs induced by ligand binding 3) requirement for large number of intermolecular NOEs and/or 4) extensive computer calculations to generate structures that match the experimental data. To address these limitations, a method

to generate protein-ligand structures that is based on a combination of selective amino acid labeling and collection of only few intermolecular NOEs is presented in chapter 3 of this thesis.[89]

Scope and Outline of Thesis

NMR methods that provide a better understanding of protein-ligand interactions are critical in the early stages of drug discovery. Traditional NMR methods used to obtain 3D structural information tend to be slow and labor intensive. The main area of this thesis was to develop and implement efficient solution based NMR approaches that provide 3D structural information on the protein-ligand complexes and could be readily applied in early stages of preclinical drug discovery. Below is the brief overview of all the chapters that describe the research work performed in this thesis.

Chapter 2 describes the application of TINS screening to discover small molecule ligands that bind the ETS-domain of TEL (TEL_{ETS}). TEL is a DNA binding protein and involved in the transcriptional regulation of the other proteins and a therapeutic target for tumorigenesis.[90-92] Biochemical and structural analyses were performed using protein observed NMR, SPR (Surface Plasmon Resonance) and gel-shift assay to demonstrate DNA binding activity of TEL. Three fragment hits generated by the TINS methodology were then validated by protein observed NMR to obtain low-resolution structural information on the binding site of these fragments. Interestingly, these primary hits occupy the same binding spot on the protein as the DNA and when used at high concentrations in gel-shift assay have the potential to disrupt the DNA binding capability of TEL_{ETS}. These novel fragments represent valuable starting points for further elaboration and hit development against TEL.

In Chapter 3, a fragment based drug discovery approach is applied to Rit1 GTPase, a validated target for Rheumatoid Arthritis.[93,94] In this chapter the results obtained from fragment based screen, crystallization, analoging, hit development and structural study on the most potent compound are discussed. One main obstacle in the project was to obtain 3D structural information on the Rit1-ligand complexes. Substantial efforts to crystallize the complex were not successful due to low solvent content of the protein crystals and cracking of protein crystals in the presence of the most potent compound. To address this issue, a solution based approach was necessary and a paramagnetic NMR based approach was sought, whereby a spin label was introduced on GDP. Paramagnetic studies using the GDP-spin label followed by docking calculations propose a novel mechanism by which the compound inhibits GDP-GTP exchange of Rit1. The PRE based method in this chapter presents an alternative to obtain binding site information on the protein-ligand complex when other high resolution techniques fail.

In Chapter 4, a solution NMR method was developed to obtain 3D structures of a protein-small molecule complex in rapid and efficient manner. The NMR method makes use of a small molecule that binds in the ATP binding pocket of the N-terminal domain of Hsp90 that was discovered by TINS NMR screening.[40,80,95] The main goal was to use this protein-ligand system to develop an efficient way to obtain 3D structural information on protein-ligand complexes. This chapter demonstrates how a combination of selective methyl group labeling, standard NMR experiments and computational docking can be used to rapidly determine the 3D structure of a small molecule bound “weakly” to a protein target. The approach requires only a sparse set of intermolecular NOEs and is an alternative to traditional NMR approaches that involve uniform isotope labeling.

Finally, concluding remarks about the research work in this thesis are presented in Chapter 5. Each molecular system investigated is overviewed and the different approaches that were employed are presented. The prospects and possible applications of the study are also discussed.

REFERENCES

1. Tupper, K.W. (2012). Psychoactive substances and the English language: "Drugs," discourses, and public policy. *Contem. Drug Prob.* **2012** *39*(3), 461-492.
2. Kotz, J. Phenotypic screening, take two. *SciBX.* **2012**, 5(15).
3. Swinney, DC., Anthony, J. How were new medicines discovered? *Nat. Review Drug Disc*. **2011**, 10, 507-519.
4. Hughes, JP., Rees, S., Kalindjian, SB., Philpott, KL. Principles of early drug discovery, *Br. J. Pharmacol*. **2011**,162(6), 1239–1249.
5. Chen, XP., Du, GH. Target validation: A door to drug discovery. *Drug Discov Ther.* **2007**, 1(1), 23-29.
6. Campbell, SF. Science art and drug discovery: a personal perspective. *Clinical Science.* **2000**, 99, 255–260.
7. Boppana, K., Dubey, PK., Jagarlapudi, SARP., Vadivelan, S., Rambabu, G. Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models. *Eur J Med Chem*. **2009**, 44, 3584–3590.
8. Keseru, GM., Makara, GM. Hit discovery and hit-to-lead approaches. *Drug Discov Today*. **2006** ,15-16,741-748.
9. Goodnow Jr, R., Gillespie, P. Hit and Lead Identification: Efficient Practices for Drug Discovery. *Prog Med Chem*. **2007**,45, 1-61.
10. Liu, B., Li, S., Hu, J. Technological advances in high-throughput screening. *Am J Pharmacogenomics*. **2004**, 4(4), 263-276.
11. Hann, MM., Leach, AR., Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.,* **2001**, 41, 856-864.
12. Kobayashi, M., Retra, K., Figaroa, F., Hollander, JG., Ab, E., Heetebrij, RJ., Irth, H., Siegal, G. Target immobilization as a strategy for NMR-based fragment screening: comparison of TINS, STD, and SPR for fragment hit identification. *J Biomol Screen*. **2010,** 15(8), 978-989.
13. Erlanson, DA., McDowell, RS., O'Brien, T. Fragment-Based Drug Discovery, *J. Med. Chem.*, **2004**, 47,14.
14. Rishton, GM. Reactive compounds and in vitro false positives in HTS, *Drug Discov Today*, **1997**, 2, 382-384.
15. McGovern, SL., Caselli, E., Grigorieff, N., Shoichet, BK. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening, *J. Med. Chem.*, **2002**, 45, 1712-1722.
16. Popa-Burke, IG., Issakova, O., Arroway, JD., Bernasconi, P., Chen, M., Coudurier, S., Galasinski, Jadhav, AP., Janzen, WP., Lagasca, D., Liu, D., Lewis, RS., Mohney, RP., Sepetov, N., Sparkman, D., Hodge, CN. Streamlined system for purifying and quantifying a diverse library of compounds and the effect of compound concentration measurements on the accurate interpretation of biological assay results. *Anal Chem*, **2004**, 76, 7278-7287.
17. Lipinski, CA., Lombardo, F., Dominy, BW., Feeney, PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv Drug Deliv Rev*, **2001**, 46, 3-26.
18. Scott, DE., Coyne, AG., Hudson, SA., Abell, C. Fragment-Based Approaches

in Drug Discovery and Chemical Biology**,** *Biochemistry* **2012,** 51 (25), 4990-5003.
19. Rees, DC., Congreve, Miles., Murray, CW., Carr, R. Fragment based lead discovery. **2004**, 3, 661-672.
20. Erlanson, DA. Introduction to Fragment-Based Drug Discovery, *Top Curr. Chem.* **2011**,317, 1-32.
21. Jencks, WP. On the attribution and additivity of binding energies, *Proc Natl Acad Sci U S A,* **1981**, 78, 4046-4050.
22. Nakamura, CE., Abeles, RE. Mode of interaction of beta-hydroxy-beta-methylglutaryl coenzyme A reductase with strong binding inhibitors: compactin and related compounds. *Biochemistry*, **1985**, 24, 1364-1376.
23. Baker, M. Fragment based drug discovery grows up. *Nat. Reviews Drug. Disc,* **2013**, 12, 5-7.
24. Hajduk, PJ., Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nature reviews Drug discovery*, *2007*, 6, 211-219.
25. Congreve, M., Carr, R., Murray, C., Jhoti, H. A 'rule of three' for fragment-based lead discovery?, *Drug Discov Today*, **2003**, 8, 876-877.
26. Dalvit, C. NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Disc.Today*, **2009**, 14(21/22), 1051-1057.
27. Stockman, BJ., Dalvit, C. NMR screening techniques in drug discovery and drug design. *Prog. NMR Spectrosc.*, **2002**, 41, 187–231.
28. Ewing, T.J.A., Kuntz, I.D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem*. **1997**, 18, 1175-1189.
29. Bohm, H.J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des*. **1992**, 6, 61-78.
30. Rarey, M., Kramer, B., Lengauer, T., Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol*. **1996**, 261, 470-489.
31. Mizutani, M.Y., Tomioka, N., Itai, A. Rational automatic search method for stable docking models of protein and ligand. *J. Mol. Biol*. **1994**, 243, 310-326.
32. Maly, D. J., Choong, I. C., Ellman, J. A. Combinatorial target-guided ligand assembly: Identification of potent subtype-selective c-Src inhibitors. *Proc*. *Natl*. *Acad*. *Sci*. *U.S.A*. **2000**, 97, 2419−2424.
33. Jhoti, H., Williams, G., Rees, DC., Murray, CW. The 'rule of three' for fragment-based drug discovery: where are we now?. *Nat. Reviews Drug Disc.* **2013**, 12, 644
34. Schade, M. NMR fragment screening: Advantages and applications. *IDrugs*. **2006** 9(2), 110-3.
35. Malcolm, H. Levitt: *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd edition, John Wiley & Sons, New York 2008, Section 11.9.2
36. Cavanagh, J., Fairbrother, WJ., Palmer, AG (III)., Rance, M., Skelton, NJ. Protein NMR spectroscopy, principles and practice, 2nd edition, **2006**. Elsevier Academic Press, Amsterdam.
37. Roberts, G., Lian, LY. Protein NMR spectroscopy: principal techniques and applications, **2011**.John Wiley and Sons Ltd. United Kingdom.
38. Carr, H. Y., Purcell, E. M. *Phys. Rev.* **1954**, 54, 630– 638.
39. Meiboom, S., Gill, D. *Rev. Sci. Instrum.* **1958**, 29, 688– 691.
40. Vanwetswinkel, S., Heetebrij, R.J., Van Duynhoven, J., Hollander, J.G., Filippov, D.V., Hajduk, P.J, Siegal, G. TINS, target immobilized NMR

screening: an efficient and sensitive method for ligand discovery. *Chem. Biol.,* **2005**, 12, 207-216.
41. Dalvit, C., Pevarello, P., Tatò, M., Veronesi M., Vulpetti, A., Sundström M. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J Biomol NMR*. **2000**, 18(1):65-8.
42. Mayer, M., Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem Int Ed Engl*, **1999**, 38, 1784-1788.
43. Lepre, CA., Moore, JM., Peng, JW. Theory and applications of NMR-based screening in pharmaceutical research. *Chem Rev*. **2004**, 104(8), 3641-3676.
44. Siegal, G., AB, E., and Schultz, J. Integration of fragment screening and library design *Drug Discov. Today*, **2007**, 12, 1032-1039.
45. Marquardsen, T., Hofmann, M., Hollander, JG., Loch, CM., Kiihne, SR., Engelke, F., Siegal, G. Development of a dual cell, flow-injection sample holder, and NMR probe for comparative ligand-binding studies. *J Magn Reson*. **2006**, 182(1), 55-65.
46. Mayer, M., Meyer, B. 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*, **2001**, 123, 6108-6117.
47. James Tsai, John T. Lee, Weiru Wang, Jiazhong Zhang, Hanna Cho, Shumeye Mamo, Ryan Bremer, Sam Gillette, Jun Kong, Nikolas K. Haass, Katrin Sproesser, Ling Li, Keiran S. M. Smalley, Daniel Fong, Yong-Liang Zhu, Adhirai Marimuthu, Hoa Nguyen, Billy Lam, Jennifer Liu, Ivana Cheung, Julie Rice, Yoshihisa Suzuki, Catherine Luu, Calvin Settachatgul, Rafe Shellooe, John Cantwell, Sung-Hou Kim, Joseph Schlessinger, Kam Y. J. Zhang, Brian L. West, Ben Powell, Gaston Habets, Chao Zhang, Prabha N. Ibrahim, Peter Hirth, Dean R. Artis, Meenhard Herlyn, and Gideon Bollag. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *PNAS* **2008,** 105 (8), 3041-3046.
48. Murray, CW., Blundell, TL. Structural biology in fragment based drug design. *Curr.Opn.Struc.Bio.* **2010**, 20(4), 497-507.
49. Oltersdorf, T., Elmore, SW., Shoemaker, AR., Armstrong, RC., Augeri, DJ., Belli, BA., Bruncko, M., Deckwerth, TL., Dinges, J., Hajduk, PJ., Joseph, MK., Kitada, S., Korsmeyer, SJ., Kunzer, AR., Letai, A., Li, C., Mitten, MJ., Nettesheim, DG., Ng, S., Nimmer, PM., O'Connor, JM., Oleksijew, A., Petros, AM., Reed, JC., Shen, W., Tahir, SK., Thompson, CB., Tomaselli, KJ., Wang, B., Wendt, MD., Zhang, H., Fesik, SW., Rosenberg, SH. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*. **2005**, 2; 435(7042):677-81.
50. Wendt, MD., Sun, C., Kunzer, A., Sauer, D., Sarris, K. Discovery of a novel small molecule binding site of human survivin. Bioorg Med Chem Lett, **2007**, 17, 3122–3129.
51. Klebe, G. Recent developments in structure based drug design. *J. Mol. Med. (Berl).* **2000**, 78(5),269-281.
52. Lesuisse, D., Lange, G., Deprez, P., Benard, D., Schoot, B., Delettre, G., Marquette, J. P., Broto, P., Jean-Baptiste, V., Bichet, P., Sarubbi, E., Mandine, E. SAR and X-ray. A new approach combining fragment-based screening and rational drug design: Application to the discovery of nanomolar inhibitors of Src SH2. *J. Med. Chem*. **2002**, 45, 2379−2387.
53. Lange, G., Lesuisse, D., Deprez, P., Schoot, B., Loenze, P., Benard, D.,

Marquette, J. P., Broto, P., Sarubbi, E., Mandine, E. Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of (pp60)Src are identical to those for high affinity binding of full length inhibitors. *J. Med. Chem*. **2003**, 46, 5184−5195.
54. Congreve, MS., Davis, DJ., Devine, L., Granata, C., O'Reilly, M., Wyatt, P. G., Jhoti, H. Detection of ligands from a dynamic combinatorial library by X-ray crystallography. *Angew. Chem., Int. Ed*. **2003**, 42, 4479−4482.
55. Brough, P.A., Barril, X., Borgognoni, J., Chene, P., Nicholas, GD., Davis, B.; Drysdale, M.J., Dymock, B., Eccles, S.A., Garcia-Echeverria, C., Fromont, C., Hayes, A., Hubbard, R.E., Jordan, A.M., Jensen, R.M., Massey, A., Merrett, A., Padfield, A., Parsons, R., Radimerski, T., Raynaud, F., Robertson, A., Roughley, S., Schoepfer, R., Simmonite, H., Sharp, S., Surgenor, A., Valenti, M., Walls, S., Webb, P., Wood, M., Workman, P., Wright, L. *J. Med. Chem.* **2009,** 52 (15), 4794-4809.
56. Erlanson, DA., McDowell, RS., O'Brien, T. Fragment based drug discovery. *J. Med. Chem*. **2004**, 47(14), 3463-3482.
57. Chilingaryan, Z., Yin, Z., Oakley, A.J. Fragment based screening by protein crystallography: successes and pitfalls. *Int. J. Mol. Sci.* **2012**, 13, 12857-12879.
58. Davis, A.M., St-Gallay, S.A., Kleywegt, G.J. Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discov. Today* **2008**, *13*, 831–841.
59. Malde, A.K., Mark, A.E. Challenges in the determination of the binding modes of non-standard ligands in X-ray crystal complexes. *J. Comput. Aided Mol. Des.* **2011**, *25*, 1–12.
60. Kleywegt, G.J., Henrick, K., Dodson, E.J., van Aalten, D.M. Pound-wise but penny-foolish: How well do micromolecules fare in macromolecular refinement? *Structure* **2003**, *11*, 1051–1059.
61. Ward, R.A., Brassington, C., Breeze, A.L., Caputo, A., Critchlow, S., Davies, G., Goodwin, L., Hassall, G., Greenwood, R., Holdgate, G.A., Design and synthesis of novel lactate dehydrogenase a inhibitors by fragment-based lead generation. *J. Med. Chem.* **2012**, *55*,3285–3306.
62. Wyss DF, Wang YS, Eaton HL, Strickland C, Voigt JH, Zhu Z, Stamford AW Combining NMR and X-ray crystallography in fragment-based drug discovery: discovery of highly potent and selective BACE-1 inhibitors.*Top Curr Chem*. **2012**, 317, 83-114.
63. Hajduk, P. J., Dinges, J., Miknis, G. F., Merlock, M., Middleton, T., Kempf, D. J., Egan, D. A., Walter, K. A., Robins, T. S., Shuker, S. B., Holzman, T. F., Fesik, S. W. NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein. *J. Med. Chem*. **1997**, 40, 3144−3150.
64. McCoy, M., Wyss, DF. Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations, *J Am Chem Soc*, **2002**, 124, 11758-11763.
65. Shuker, SB., Hajduk, PJ., Meadows, RP., Fesik, SW. Discovering high-affinity ligands for proteins: SAR by NMR, *Science*, **1996**, 274, 1531-1534.
66. Campbell AP, Sykes BD. The two-dimensional transferred nuclear Overhauser effect: theory and practice. *Annu Rev Biophys Biomol Struct*. **1993**, 22, 99-122.
67. Williamson, MP. Applications of the NOE in molecular biology. *Ann Rep NMR Spec.***2009,** 65, 77-109.

68. Post CB. Exchange-transferred NOE spectroscopy and bound ligand structure determination. *Curr Opin Struct Biol.* **2003**,13(5), 581-588.
69. Barsukov, IL., Lian, LY., Ellis, J., Sze, KH., Shaw, WV., Roberts GC. The conformation of coenzyme A bound to chloramphenicol acetyltransferase determined by transferred NOE experiments. *J Mol Biol*. **1996**, 262(4), 543-558.
70. Śledź, P., Silvestre, H.L., Hung, A.W., Ciulli, A., Blundell, T.L., and Abell, C. Optimization of the Interligand Overhauser Effect for fragment linking: application to inhibitor discovery against Mycobacterium tuberculosis pantothenate synthetase. *J. Am. Chem. Soc.* **2010**, 132, 4544–4545.
71. Orts J, Griesinger C, Carlomagno T. The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method. *J Magn Reson***.** **2009**, 200(1), 64-73.
72. Pervushin, K., Riek, R., Wider, G., Wüthrich, K. Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *PNAS* **1997**, 94, 12366-12371.
73. Riek, R., Wider,G., Pervushin, K., Wüthrich,K. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *PNAS*, **1999,** 96, 4918-4923.
74. Bloembergen, N., Morgan, LO. Proton Relaxation Times in Paramagnetic Solutions. Effects of Electron Spin Relaxation. *J. Chem. Phys*. **1961,** 34, 842.
75. John, M., Otting, G. Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy. Chem.phys. chem. **2007**, 8, 2309–2313.
76. Jahnke, W., Perez, LB., Paris, CG., Strauss, A., Fendrich, G., Nalin, CM. Second-site NMR screening with a spin-labeled first ligand. *J Am Chem Soc*, **2000**, 122, 7394-7395.
77. Guan, JY., Keizers, PHJ., Liu, WL., Löhr, F., Skinner, Sp., Heeneman, E., Schwalbe, H., Ubbink, M., Siegal, G. Small-Molecule Binding Sites on Proteins Established by Paramagnetic NMR Spectroscopy *J. Am. Chem. Soc.*, 2013, 135 (15), 5859-5868.
78. Otting, G. Protein NMR using paramagnetic ions. *Annu Rev Biophys*. **2010**, 39, 387-405.
79. Tugarinov, V., Kay, LE. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods, *J Am Chem Soc*, **2003**, 125, 13868-13878.
80. Tugarinov, V., Choy, WY., Orekhov, VY., Kay, LE. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc Natl Acad Sci U S A*. **2005,** 18, 102(3), 622-627.
81. Hajduk, PJ., Augeri, DJ., Mack, J., Mendoza, R., Yang, JG., Betz, SF., Fesik, SW. NMR-based screening of proteins containing ^{13}C-labeled methyl groups, *J Am Chem Soc,* **2000**, 122, 7898-7904.
82. Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J., Mulder, F. A. A. Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins.J. Am. Chem. Soc., **2010**, 132 (9), 2952 – 2960.
83. Venditti, V., Fawzi, N.L., Clore, G.M. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. *J. Biomol. NMR.* **2011**, 51, 319-328.
84. John, M., Schmitz, C., Park, A., Dixon, N., Huber, T., Otting, G. Sequence-

specific and stereospecific assignment of methyl groups using paramagnetic lanthanides. *J. Am. Chem. Soc.* **2007**, 129, 44, 13749-13757.
85. Xu, Y., Liu, M., Simpson, PJ., Isaacson, R., Cota, E., Marchant, J., Yang, D., Zhang, X., Freemont, P.,, Matthews, S. Automated assignment in selectively methyl-labeled proteins. *J. Am. Chem. Soc.* **2009**, 131, 9480–9481.
86. Chao, FA., Shi, L., Masterson, L.R., Veglia, G. FLAMEnGO: a fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. *J. Magn.Reson.,* **2012,** 214 (1), 103-110.
87. Skinner, SP., Moshev, M., Hass, MA., Ubbink, M. PARAssign-paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts. *J. Biomol.NMR.* **2013**, 55(4),379-389.
88. Sharrocks, AD. The ETS-domain transcription factor family. *Nat Rev Mol Cell Bio.l* **2001,** 2, 827-837.
89. Shah,DM., AB, E, Diercks, T., Hass, M., Van Nuland, N., Siegal, G. Rapid protein-ligand co-structures by sparse NOE data. J Med Chem, **2012,** 55 (23), 10786-10790.
90. Lopez, RG., Carron, C., Oury, C., Gardellin, P., Bernard, O., Ghysdael, J. TEL is a sequence- specific transcriptional repressor. *J. Biol. Chem.* **1999**, 274: 30132–30138.
91. Green, SM., Coyne, HJ, 3rd., McIntosh, LP., Graves, BJ. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem.* **2010,** 285, 18496–18504.
92. Roukens, MG., Alloul-Ramdhani, M., Baan, B., Kobayashi, K., Peterson-Maduro, J., van Dam H., Schulte-Merker, S., Baker, DA. Control of endothelial sprouting by a Tel–CtBP complex. *Nat Cell Biol*. **2010**, 12, 933–942
93. Konstantinopoulos, P., Karamouzis, MV., Papavassiliou., AG. Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets. *Nature Reviews Drug Discovery*, **2007,** 6, 541-555.
94. Methods for Identification, and Compounds useful, for the treatment of degenerative and inflammatory diseases, Ser. No. 11/52,366, filed Jun.14, 2005 Galapagos B.V. US patent 7,919,259 B2.
95. Miura, T., Fukami, T., Hasegawa, K., Ono, N., Suda, A., Shindo, H., Yoon, D.O., Kim, S.J., Na, Y.J., Aoki, Y., Shimma, N., Tsukuda, T., Shiratori, Y. Lead generation of heat shock protein 90 inhibitors by a combination of fragment-based approach, virtual screening, and structure-based drug design. *Bioorg. Med. Chem. Lett*., **2011**, 21, 5778-5783

Chapter 2

Discovery of small molecule fragments that inhibit the DNA binding of the TEL-ETS domain

Dipen M. Shah,[‡,†] Eiso AB,[†] Ruta K. Nachane,[†] Johan G. Hollander, [†] David A. Baker,[§] and Gregg Siegal [‡,†]*

[‡] Leiden Institute of Chemistry, Leiden University, Leiden 2300RA, the Netherlands.

[†] ZoBio BV, Leiden 2300RA, the Netherlands.

[§] Leiden University Medical Center (LUMC), Section of Growth Control and Transcription, Department of Molecular Cell Biology, 2300 RC Leiden, the Netherlands.

Manuscript to be submitted to the PLOS One journal

Abstract

The discovery of small molecule ligands help in the design of inhibitors against a protein target involved in a particular disease. Here, we present a study in which we have applied a fragment based screen against a potential tumor angiogenesis target, TEL. TEL and its associated network proteins are involved in control of the development of blood circulatory system. TEL is a transcription repressor that functions through bind of its C-terminal ETS domain to DNA which sterically blocks the access of transcription factors to the promotor. The development of specific inhibitors which would disrupt DNA binding is desirable as this would lead to loss of downstream signaling cascade. Here, we have characterized the DNA binding of TEL-ETS domain using protein-observed NMR, gel retardation and surface plasmon resonance. Of a number of fragment hits, three were further validated using protein-observed NMR. Chemical shift mapping revealed that these fragments share the same binding interface as the DNA. Furthermore, we see that the presence of fragments at a high concentration in the gel retardation assay disrupts the DNA binding capability of TEL. These fragments provide valuable starting entities to develop potent compounds against TEL with a novel mechanism which in turn could lead to loss of formation of blood vessels in tumors.

Introduction

Activation and repression of the transcription of eukaryotic genes during cellular development and/or response to extracellular stimuli, is a highly organized event involving assembly of many protein complexes at promoters and enhancers.[1] Amongst various signal dependent transcriptional regulators, the ETS-family of proteins is one of the largest known involved in proliferation, differentiation and tumorigenesis.[2] ETS1 is the founding member of this family and the name "ETS" is derived from the avian erythroblastosis virus E26, which carried "*v-ets' (E twenty-six)* oncogene.[3] Many of the members are known to activate transcription but the ETS family also includes transcriptional repressors.[2] One of the known repressors is TEL which was identified due to its fusion to the 3' -half of the gene encoding the platelet derived growth factor β receptor in chronic myelomonocytic leukemia.[4] The transcriptional repression may involve one of the following mechanisms. The repressor may act by 1] directly binding to the activation domain of a transcriptional activator, 2] by displacing an activator from activation site, 3] indirectly by altering the chromatin structure and accessibility of a region on the DNA or 4] interfering with the basic transcriptional machinery.[2]

Structurally, TEL possesses an evolutionarily conserved ETS-domain which binds DNA sequences with a GGAA/T core motif.[5] In fact, most of the ETS proteins recognize a consensus 5' -GGAA/T-3' motif within the context of a 9- to 10-bp DNA sequence. [3,6,7] TEL also carries a PNT (Pointed or SAM) domain at the N-terminus which is observed to influence not only the DNA binding but its interaction with other regulatory proteins. The PNT domain of TEL is capable of forming a stable head to tail polymer and hence most likely the full length TEL also has the ability to polymerize. This ability of the PNT domain to cause self-association of the protein

also results in the constitutive activation of the tyrosine kinase activity of fusion protein TEL-ABL, TEL-platelet derived growth factor β receptor and TEL-JAK2, a property that is essential to their transforming and leukemogenic properties.[4,8-11] On other hand, the role of the ETS-domain of TEL, here after refered as TEL_{ETS}, in DNA binding is influenced by an autoinhibitory mechanism in which a flanking C-terminal helix sterically blocks the monomeric ETS-domain. A conformational change is then likely to be necessary in order to carry out DNA binding. It has also been proposed that TEL displays a conformational equilibrium between two structural states 1] where the C-terminal helix is in the inhibitory conformation and 2] where it is in an uninhibited conformation thus not interfering with the DNA binding.[4]

In addition to structural features, the functional importance of TEL is realized by the fact that it is also a therapeutic target to inhibit tumor angiogenesis. TEL is indispensable for endothelial sprouting and normal development of the *Danio rerio* blood circulatory system. TEL acts by regulating the transcription of various angiogenesis inhibitors and executes its function in conjunction with CtBP (Carboxy terminal binding protein). The complex of TEL-CtBP helps in the conditioning of the endothelial cells for angiogenesis. This conditioning is achieved by regulating the balance between stimulatory and antagonistic sprouting cues. One of the plausible hypotheses is also considered where the interaction of CtBP with the PxEIM motif of C-terminal helix of TEL leads to increased DNA binding affinity of TEL. By this way, TEL is able to block the transcription of other essential proteins thought to control the development of blood vessels to tumorigenic tissues. Clearly, the development of specific inhibitors that would either disrupt the DNA or CtBP binding of TEL would open the way for the development of therapeutic strategies to inhibit pathological angiogenesis.[12]

In this work we have demonstrated the binding of the TEL_{ETS} domain to DNA by protein observed NMR, a gel retardation assay and surface plasmon resonance. With a view to find small molecule inhibitors for TEL_{ETS} that would potentially disrupt the DNA binding and hence downstream signaling cascade, we have applied a fragment based screen to generate a set of primary hits. In fragment based drug discovery (FBDD) approach, a larger portion of chemical structure space is explored with a smaller number of compounds. Since the fragments are small and much simpler than drug-like molecules, the fragment methodology enables to discover molecules which bind more efficiently on the protein surface. A fragment based screen was carried out using TINS screening of a fragment library consisting of commercially available compounds.[13-16] Three fragments hits from the screen were further validated using a protein observed NMR approach and gel retardation assay. Using chemical shift mapping through protein-observed NMR experiments, we were able to obtain the binding site of the fragments. Interestingly, the binding surface for the DNA and fragments mapped onto the 3D structure of TEL_{ETS} available in the protein data bank (PDB 2DAO) revealed that the fragments bind at a similar binding interface as the DNA. Furthermore, we have shown that the presence of high concentration of the fragments in the gel retardation assay disrupts the DNA binding capability of TEL_{ETS} domain.[13]

Results

Target immobilized NMR screening of TEL-ETS domain:

We employed TINS (Target Immobilized NMR screening) screening as a primary screen for hit generation on TEL_{ETS} domain.[14] TEL_{ETS} and the reference protein, Akt PH domain were purified and immobilized *via* primary amine coupling as described in the experimental section. The PH domain of Akt was chosen as the reference protein

because it represents a typical protein surface, yet devoid of any specific small molecule binding sites. Further, the use of reference protein in TINS eliminates very weak and non-specific binders which otherwise might lead to false positives in the hit identification process. The functionality of immobilized TEL was determined by repetitive injection of a fragment that appeared as a hit during initial part of the screening procedure.[15]

A fragment library consisting of 1364 commercially available fragments was screened against TEL_{ETS}.[16] The fragment library was screened by repeated cycles of injection of various mixes of fragments, which consisted of 3 to 5 compounds per mix, into both the cells of a dual-cell packed with immobilized TEL_{ETS} domain and the reference protein. After a mix was injected, the flow was stopped and the NMR data was acquired. TINS NMR experiments consisted of using spatially selective Hadamard pulse sequences and the results were analyzed as described.[14,15] The fragments were washed out prior to the next injection of a new fragment mix. The binding of a fragment to a protein molecule immobilized on the solid support causes broadening of the resonances from that fragment. As a result a fragment binding specifically to TEL_{ETS} domain could be detected by simple reduction in the height of all the NMR signals from that fragment in the presence of TEL_{ETS} relative to that in the presence of a reference protein (Figure S1). The fragment screen resulted in the identification of 103 unique hits for TEL_{ETS} domain resulting in a hit rate of 7.5%.

Sequential Backbone resonance assignments

The sequential backbone assignment for the TEL_{ETS} was achieved by acquiring standard triple resonance NMR experiments. Although the solution structure of the TEL_{ETS} domain is deposited in the pdb (Figure S2), there are no published resonance

assignments that are available. The sequential backbone assignments were obtained by correlation of Cα and Cβ chemical shifts of i and i-1 residues to the amide ^{1}H and ^{15}N resonances using HNCACB, CBCAcoNH, HNCA and HNcoCA spectra. Overall, the spectra were of good quality and 82% of the backbone sequential assignment was obtained.

Characterization of TEL-DNA binding

The DNA binding of TEL_{ETS} was characterized using three different approaches. We have used protein-observed NMR, gel-retardation and surface plasmon resonance.

DNA binding by Protein-observed NMR

We employed a frequently used protein observed 2D-heteronuclear single quantum coherence (HSQC) NMR method wherein increasing concentrations of the compound is titrated into uniformly ^{15}N labeled protein and changes in the NMR spectra are followed in presence and absence of the compounds.[17,18] In order to assess the DNA binding capability of TEL_{ETS} , an oligonucleotide consisting of 10 base pairs containing the TEL_{ETS} recognition site was titrated into the ^{15}N isotopically labelled TEL_{ETS}. A short oligo nucleotide sequence (5'-ACA**GGA**AGTG-3') was purposely chosen for the NMR study. The oligonucleotide was titrated at three protein to oligo ratios of 1:2, 1:4 and 1:6. Significant chemical shift perturbations (CSPs) are observed in the 2D [^{1}H,^{15}N] HSQC NMR spectrum upon addition of the DNA as seen in Figure 1.[17,18] It is also observed that at a protein to oligo ratio of 1:4, the binding is saturated as the position of the CSP does not change beyond that concentration. The upper limit for the binding affinity is estimated to be around 80-100 μM. The availability of the sequential backbone assignment of the protein allowed us to map the specific residues involved in DNA binding onto the available solution structure of the protein (PDB 2DAO). Figure 2 shows the DNA binding

interface of TEL_{ETS} domain based on the observed CSPs. Figure 2A shows the binding map of all the residues that show significant chemical shift perturbations in the presence of DNA. The key protein residues involved in the binding are I40, L46, G51, K54, S65, R66, Y71, Y72, R84, F87, T101, L112 and S113. The numbering of the residues here is same as PDB 2DAO. Overall, the protein-DNA complex seems to be in the fast exchange regime on an NMR time scale. Some broadening of the resonance peaks is observed when concentration of DNA is gradually increased.

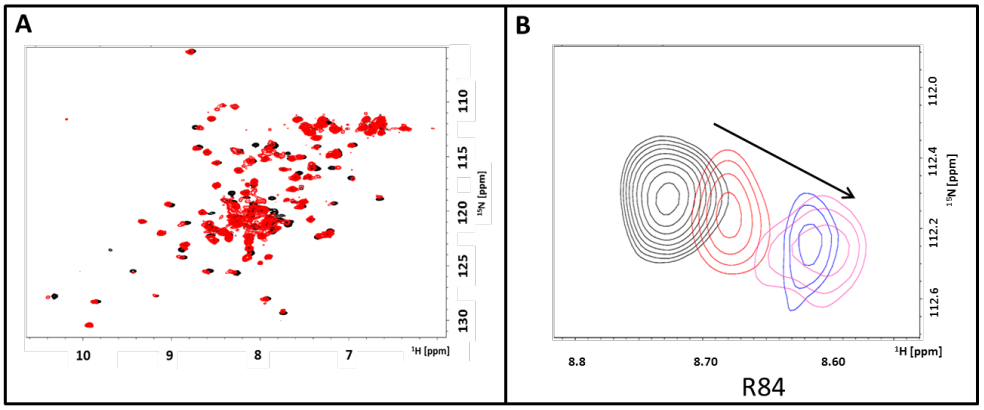


Figure 1. The characterization of DNA binding using protein-observed NMR. [A] The overlay of the 2D [^{1}H, ^{15}N] HSQC spectra of TEL_{ETS} in the absence of DNA (black) and in the presence of the DNA (red). The concentration of the protein used was 60 μM. [B] The concentration dependent chemical shift perturbation observed for residue R84 in the presence of the increasing protein to DNA ratios is shown. In black is the free protein. The red, blue and pink represent protein to DNA ratios of 1:2, 1:4 and 1:6 respectively.

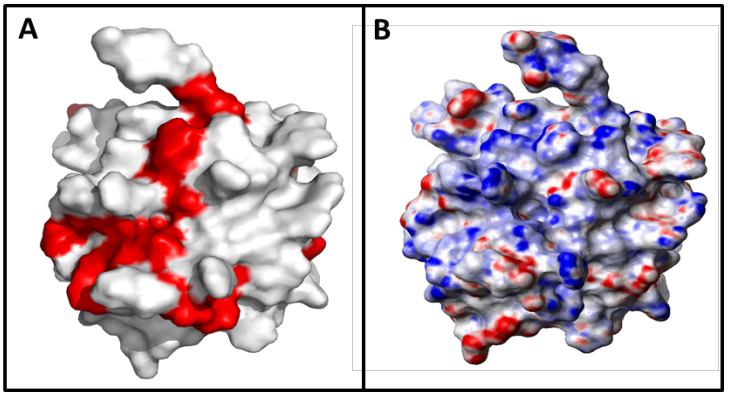


Figure 2. The determination of protein-DNA binding interface. A] Chemical shift mapping of the residues involved in the DNA binding as obtained by titration of increasing concentrations ($\Delta\delta$, ($CSP_{difference}$) > two times standard deviation + $\Delta\delta_{avg}$) of the DNA into ^{15}N labeled TEL and observed by 2D HSQC NMR. The mapping of the residues was performed using the available structure of TEL (2DAO) B] The electrostatic surface potential (blue-positive charge, red- negative charge) of the protein. The figure was created in PyMOL [19]

Gel retardation assay

The DNA binding to TEL_{ETS} was characterized using a gel retardation assay. To investigate DNA binding using gel retardation, increasing amounts of the protein were titrated into a ^{32}P-labelled- DNA oligonucleotide containing three ETS-binding sites (5'-AAACA**GGA**AGTGAGAACA**GGA**AGTGAGAACA**GGA**AGTGG-3'). Subsequently, DNA binding was analyzed using a gel retardation assay. As seen in Figure 3, the migration of the labeled oligonucleotide was retarded indicating that the protein binds to the DNA. The DNA binding of the protein is only visible upon titrating higher concentration of protein indicating that TEL_{ETS} on its own binds comparatively more weakly than the full length TEL harboring the N-terminal PNT-domain.[4] It is also observed in Figure 3 that two bands (a lower band representing unbound oligonucleotide and a higher band representing protein-DNA complex) are seen for DNA to protein ratios between 1:40 to 1:200. To quantitate the binding affinity of the protein-oligonucloetide interaction, we further characterized the complex by SPR

(Surface Plasmon Resonance) assay.

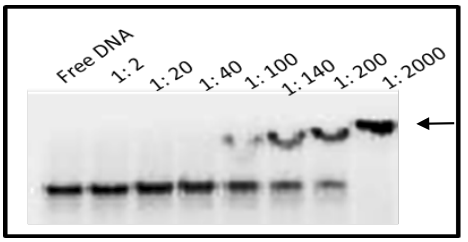


Figure 3. The DNA binding by TEL_{ETS} using gel retardation assay. TEL_{ETS} was titrated into a constant amount of ^{32}P labeled oligonucleotide at indicated ratios. The position of the DNA-protein complex is denoted by arrow. The retardation of the DNA can be clearly seen upon titrating higher concentration of the protein, indicative of the complex formation.

Surface Plasmon Resonance

The characterization and affinity determination of DNA binding to TEL_{ETS} by SPR was achieved by immobilizing the protein on the chip surface to an immobilization level of around 9000 RU. Following the immobilization, increasing concentrations of the same oligonucleotide used in the gel retardation assay (with three recognition sites) were injected in a single cycle manner. The binding response at each concentration was then plotted against the DNA concentration and K_D was determined by a 1:1 binding model using GraphPad Prism software (Figure 4). The fit yielded a K_D of 1.9 μM. The affinity appears to be tighter than that estimated from the NMR titration. This difference in the K_D is likely due to the presence of three -GGAA-recognition sites on the oligonucleotide used in the SPR assay compared to one in the oligonucleotide used in the NMR titration.

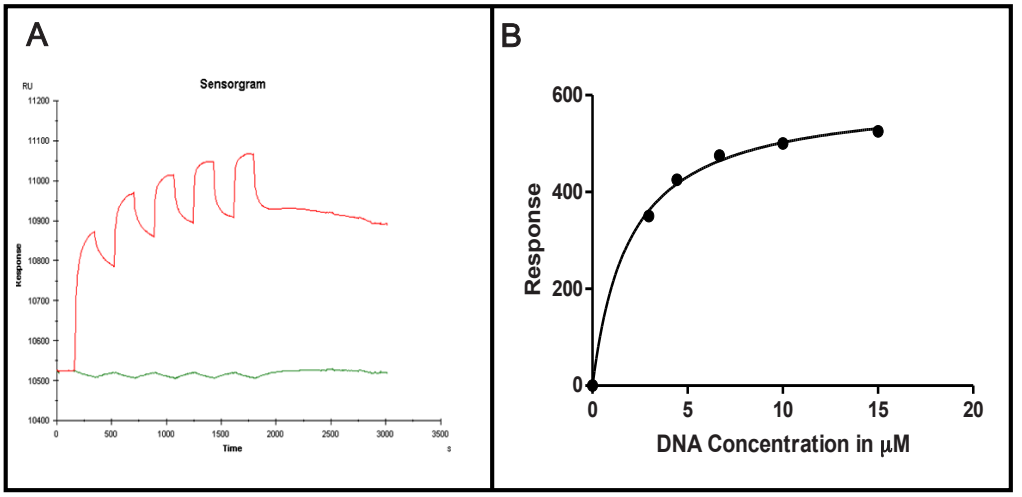


Figure 4. The determination of the DNA binding affinity of TEL_{ETS} using SPR. [A] The sensorgram obtained by single cycle injections of increasing concentrations of DNA (1 µM, 5 µM, 7.5 µM, 10 µM, 15 µM) onto immobilized TEL_{ETS} . The green line corresponds to the background response. [B] Shows the graph obtained by a fit to 1:1 binding model as analyzed by GraphPad prism software and exhibited a KD of 1.9 µM.

Validation of fragment hits by HSQC

The validation of fragment binding to TEL_{ETS} was carried out by protein observed NMR using HSQC experiment (described earlier). The main goal at this stage was to validate a limited number of fragments that would be easier to follow-up by medicinal chemistry (Fragments for which commercial analogs were available and those that can be easier to undergo synthetic chemistry). Three fragments as shown in Figure 5 were selected from a set of hits for this purpose. The resulting CSPs observed upon addition of the fragments were then mapped onto the available TEL_{ETS} solution structure (PDB 2DAO). Interestingly, a similar region of TEL_{ETS} is affected upon addition of all three fragments. As seen in Figure 6, CSPs for these fragments are localized mainly on the protein surface which is common to DNA binding (Figure 3A). The key residues that show significant chemical shift perturbations are N44, R56, K63, R66, A67, R69, I76, F87, T101 and S113 for fragments **1** and **2.** For fragment **3**,

the key residues that gave significant chemical shift perturbations were V15, G51, S66, I77, E80, F87, T101, L112 and S113.

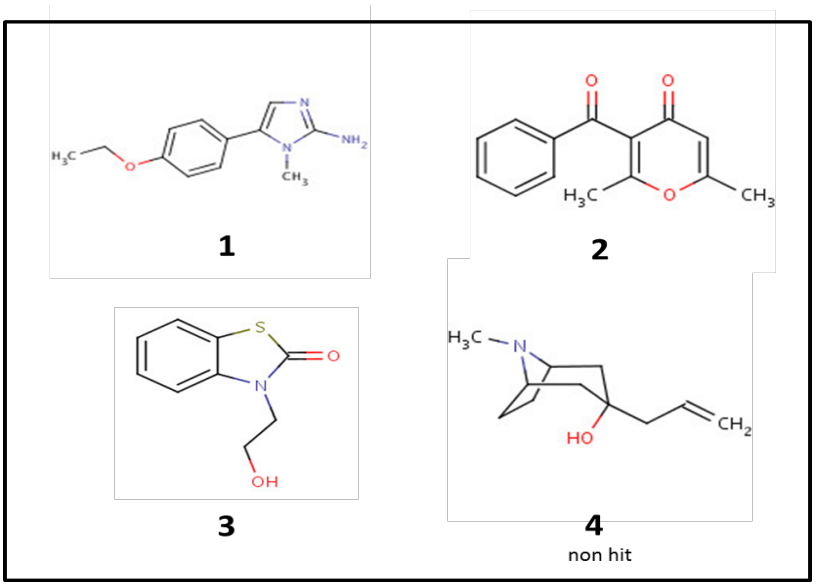


Figure 5. The chemical structures of fragment hits that were employed for follow-up studies after the initial TINS screen on TEL_{ETS}. Fragment 4 is not a hit and was used as a control for gel retardation assay.[16]

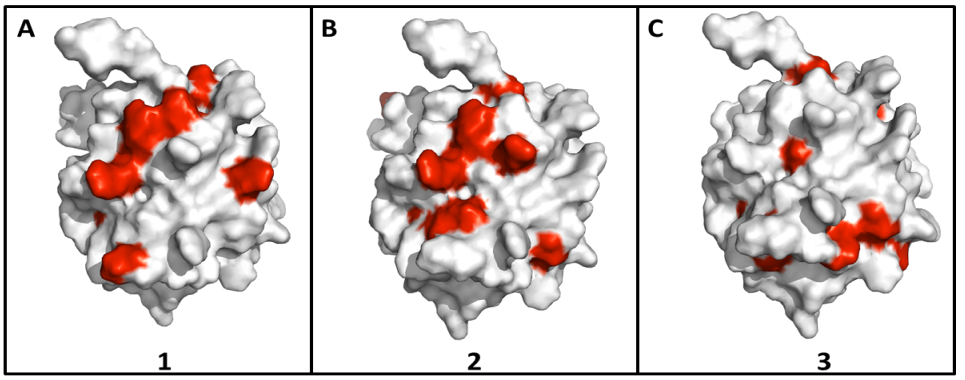


Figure 6. Chemical shift mapping of the protein residues involved in fragment binding to TEL_{ETS}. The chemical shift perturbations (in red) caused by the presence of the indicated fragment hit as determined by [^{1}H,^{15}N] HSQC NMR are mapped onto the available PDB structure of TEL_{ETS} (2DAO). The most significant chemical shift perturbations ($\Delta\delta$ > two times standard deviation + $\Delta\delta_{avg}$) are shown.[17,18] The figure was created in PyMOL.[19]

Disruption of the DNA binding to TEL_{ETS} ***by fragments in gel retardation assay***

In order to investigate the effect of fragments on oligonucleotide binding, we carried out gel retardation assay. First, TEL_{ETS} and oligonucleotide were incubated to form a stable complex at a DNA to protein ratio of 1:200. Subsequently, fragments **2** and **3** were added and the reaction mixture was further incubated for 2 hrs at room temperature. The reaction mixture for fragment **1** showed signs of precipitation and as a result was not considered for further characterization. As a control, a fragment that did not appear as a hit in the TINS screen was also used. As seen in Figure 7, in the presence of 10 mM of the fragment, DNA binding to TEL_{ETS} is significantly reduced. In the presence of control fragment, the DNA binding is not affected indicating the specificity of the fragments that bind to the TEL-ETS domain.

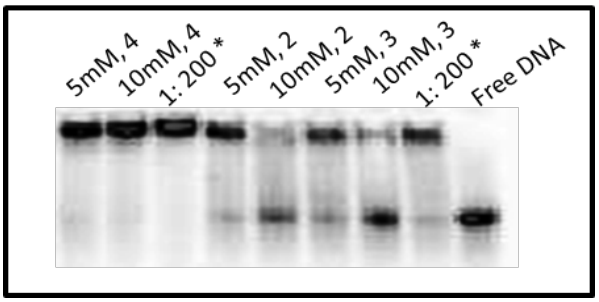


Figure 7. Disruption of DNA binding observed in the presence of validated fragment hits 2 and 3. The final concentration of the respective fragment in the reaction mixture is indicated. Fragment 4 was not a hit in the TINS screen and was used as a control. The asterisk (*) indicates a reaction without added fragment with equal percentage of DMSO. The disruption of oligonucleotide binding is clearly seen at fragment concentrations of 10 mM and comparatively less at a lower fragment concentration (5 mM).

Discussion and Conclusions

Here we report a study that focuses on the interaction of small molecules with TEL_{ETS}. To date there are no known inhibitors that disrupt the TEL-DNA binding. To identify potential fragment hits, we have applied a fragment based screen by NMR using a library of 1364 molecules.[14,16] The advantage of screening molecular

fragments rather than drug-sized molecules is that a dramatically larger portion of chemical structure space is explored with a smaller number of compounds.

In the study presented here, the interaction of the DNA oligonucleotide with TEL_{ETS} was analyzed using protein observed NMR, gel retardation assay and SPR. The binding affinity for the DNA was determined to be 1.9 μM**.** To find small molecule inhibitors that would target the DNA binding site, it would be first necessary to obtain a structural basis for the binding of TEL_{ETS} -DNA complex. We have used protein-observed HSQC NMR to obtain low resolution information to define the DNA binding interface of TEL_{ETS}. The CSP data obtained from the [^{1}H,^{15}N] HSQC experiment clearly demonstrate that DNA binding to TEL_{ETS} is mediated by a highly positively charged protein interface. Of a number of fragments hits discovered from the TINS NMR screen, three fragments were further selected from medicinal chemistry point of view (fragments that can be easily chemically modified). These were further characterized and validated for binding using protein-observed NMR experiments as described earlier. Interestingly, the ligand induced chemical shift changes clearly illustrate that all three fragments bind at the DNA binding interface. The commonality of the binding site between the fragments and the DNA is also reflected in the gel retardation assay. The presence of fragments **2** and **3** at high concentrations in the assay caused significant disruption of TEL_{ETS} -DNA binding. The use of high concentrations was necessary and indicates that the fragments weakly interact with TEL_{ETS}. This is not surprising as these are starting points and not elaborated compounds.

Inhibitors that would specifically disrupt DNA binding should lead to loss of the downstream signaling cascade and influence the transcription repressor function of TEL. This would allow transcription of other proteins that might play a critical role in

the regulation of angiogenesis in tumors tissues.[12] The functional inhibition of the TEL_{ETS}-DNA complex exhibited by the validated fragments **2** and **3** in gel retardation assay suggests they might be starting points to develop tool compounds to enable further target validation studies. However, the selectivity of these fragment hits against the ETS- domain of other proteins and in the presence of CtBP, remains to be tested.

Experimental Section

Protein expression and purification:

C-terminal hexahistidine tagged- TEL_{ETS} (1-118 as per numbering in the PDB-2DAO; 334-452, the numbering as per UniProt ID P41212) was cloned into pET28a and cultures in LB medium containing kanamycin (50 µg/mL) at 37°C in *E.coli* strain BL21 (DE3) (Stratagene). Protein expression was induced at an OD600 of 0.6 with 0.5 mM IPTG and growth was continued for 4 hours at 30°C. The cell pellets were collected by centrifugation at 5000 rpm for 30 minutes. The cell pellets were resuspended in 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol and stored at -80°C. Cells were lysed by addition of 50µl lysozyme (50 mg/mL) followed by incubation for 1 hour at 4°C and passing through French press at 1500 psig twice. The recovered lysate was centrifuged at 35000 rpm at 4°C for 45 minutes using a Beckman Ti35 rotor. The supernatant was applied to a 5 mL HisTrap HP Ni^{2+} affinity column (GE healthcare) equilibrated in the manufacturer's suggested binding buffer supplemented with protease inhibitor PMSF. The column was washed with 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 200 mM imidazole, 1 mM β-mercaptoethanol and protein was eluted with 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 10 mM imidazole and 1 mM β-mercaptoethanol. The eluted fractions were further purified on a Superdex G75 equilibrated with buffer

of 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA and 1 mM β-mercaptoethanol.

TINS NMR screening

C-terminal hexahistidine tagged- TEL_{ETS} (334-452) was cloned into pET28a and expressed in *E.coli* strain BL21(DE3) (Stratagene). The protein was first purified on Ni-Hitrap FF (GE lifescience) followed by purification on Superdex G75 with buffer of 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA and 1 mM β-mercaptoethanol. TEL_{ETS} (334-452) and Akt PH domain (aa 1-123) were immobilized via amine-coupling to 500 µL Actigel-ALD resin (Sterogene) in 25 mM Hepes pH 7.5, 100 mM NaCl and 2 mM $MgCl_2$ at 4°C using the coupling reagent provided by the manufacturer. The immobilization efficiency was above 90% and the final concentrations of the immobilized targets were typically in the range of 100 µM. TINS NMR experiments were performed on a 500 MHz Bruker NMR spectrometer using spatially selective Hadamard pulse sequences and analyzed as described previously.[14-16,20,21]

Enzyme Mobility Shift Assay

The DNA binding to TEL_{ETS} was detected using a gel retardation assay. The indicated amount of protein was diluted in a buffer consisting of 10 mM HEPES, pH 7.8, 2 mM $MgCl_2$, 0.1 mM EDTA, 100 µg/mL bovine serum albumin, 15% glycerol, (0-0.8) µg/mL poly(dI-dC) (Boerhringer Mannheim) and 2 mM dithiothreitol in a total volume of 15 µL. 20 fmol of the ^{32}P-labeled DNA oligonucleotide in 5 µL was added, incubated on ice for 30 minutes and applied to a non-denaturing 8% Tris-glycine acrylamide gel containing 2% glycerol. Electrophoresis was performed at 80V for 20 minutes and subsequently 120V for 40 minutes at 4°C in 25 mM Tris-HCl, pH 8.5 and

200 mM glycine. The gel was dried and radioactivity was detected using a phosphorimager (Biorad).

The three validated fragments determined as hits to TEL_{ETS} domain from TINS NMR screening were subjected to gel shift assay analysis. The indicated concentrations of the fragment were added to the reaction mixture and incubated for 2 hours at room temperature. The amount of DMSO in the controls was matched to the equivalent amount in the presence of a fragment.

NMR backbone sequential assignment

Uniform ^{13}C, ^{15}N labeling of TEL-ETS domain was achieved by expression of C-terminally hexahistidine tagged TEL_{ETS} (aa 334-452) in *E.coli* BL21 (DE3) cell grown in M9 minimal medium supplemented with $^{15}NH_4Cl$ and ^{13}C-D-Glucose (CIL) as a sole nitrogen and carbon source. The protein was purified as described above. The protein was concentrated down to 0.7 mM in 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA and 1 mM β-mercaptoethanol. In order to perform sequential assignment of backbone amide, the following experiments were acquired at 296K on a 600MHz BRUKER DMX NMR spectrometer equipped with a TXI cryo-probe; [^{1}H, ^{15}N]-HSQC, HNCO, HNCaCb, CbCaCONH, HNCaCO, HNCA, and HNCOCA . The acquired data was processed using nmrpipe and visualized on Sparky.[22,23] The assignment process was guided by the predicted chemical shifts calculated by SHIFTX in automatic assignment program MARS using PDB-2DAO as an input structure.[24,25]

Chemical shift perturbation and generation of binding site

^{15}N labeled C-terminal hexahistidine tagged TEL_{ETS} (aa 334-452) in *E.coli* BL21

(DE3) cell grown in M9 minimal medium supplemented with $^{15}NH_4Cl$ as a sole nitrogen source, and purified as described above. The [^{1}H,^{15}N]-HSQC were acquired at 296K on a 600MHz Bruker DMX NMR spectrometer equipped with a TXI cryo-probe. The NMR sample was prepared in 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA and 1 mM β-mercaptoethanol. The typical NMR sample contained 0.130 mM of the protein, fragments at various concentrations (see text) and 5% d_6-DMSO. The pH of the samples was adjusted carefully within +/- 0.05 units after addition of the compound. During the experiment, total of 128 indirect increments with 16 scans per increments were acquired. The data was processed using Topspin 1.2/2.1 (Bruker) and visualized on Sparky.[23] Chemical shift perturbations in [^{1}H, ^{15}N]-HSQC were calculated based on the ($\Delta\delta$ > two times standard deviation + $\Delta\delta_{avg}$) for change in ^{1}H and ^{15}N ppm value between in the presence and absence of a compound. The potential binding sites of a compound were mapped onto the surface of the TEL-ETS domain (PDB-2DAO) structure using chemical shift perturbation data and viewed in PyMOL.[19]

Surface Plasmon Resonance studies

DNA binding to TEL_{ETS} domain was tested using surface plasmon resonance on a T200 biacore instrument (GE healthcare). Initially, pH scouting was performed to determine the optimum pH for protein immobilization on the CM5 chip surface. Following immobilization of the protein to required response units (6000 RU), increasing concentrations of DNA were titrated in a single cycle kinetic mode. The buffer conditions used was 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA and 1 mM β-mercaptoethanol + 3% DMSO. The analysis was performed using a Biacore evaluation software and GraphPad Prism software.

Contributions

The clones for the production of the protein were provided by David A. Baker (Leiden Univeristy Medical Centrum). Ruta K. Nachane was responsible for protein and compound preparations for TINS NMR screening. Johan G. Hollander helped in the NMR set-up of the TINS screen and Eiso AB performed the computational analysis of the fragment screen.

SUPPLEMENTRY INFORMATION

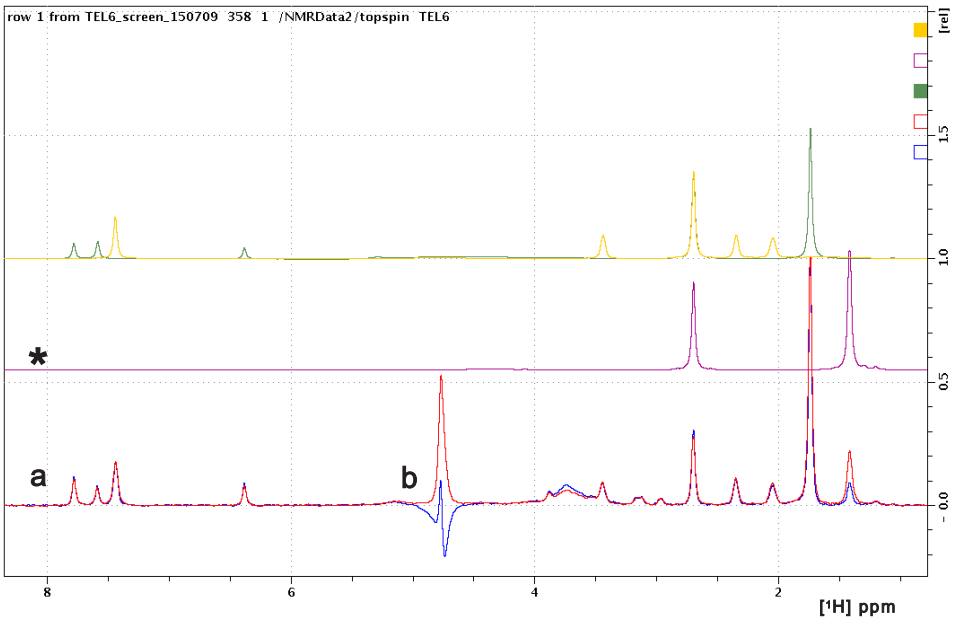


Figure S1. The screening of fragment library using TINS for binding to immobilized TEL_{ETS}.[14,15] The figure shows detection of ligand binding to the protein in a fragment mix. The overlaid ^{1}H resonances spectrum at the bottom represented by **a** shows the binding of a fragment to TEL_{ETS} as seen by the reduction in the peak height of the blue spectrum as compared to the red one which is in the presence of reference protein, PH domain of AKT. The asterisk (*****) indicates the spectrum (in purple) of fragment that preferentially binds TEL_{ETS} in solution. The residual signal from water is indicated by **b**. All other resonances belong to the other fragments in the mix.

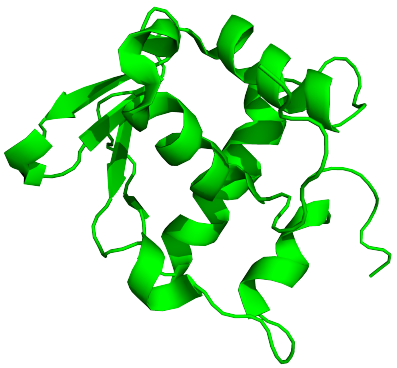


Figure S2. The NMR structure of TEL_{ETS} domain (PDB-2DAO)

REFERENCES

1. Verger A, Duterque-Coquillaud M. (2002) When Ets transcription factors meet their partner. Bioessays 4: 362-70.
2. Mavrothalassitis G, Ghysdael J. (2000) Proteins of the ETS family with transcriptional repressor activity. 19 (55): 6524-6532.
3. Sharrocks, AD. (2001) The ETS-domain transcription factor family. Nat Rev Mol Cell Biol 2; 827-837.
4. Lopez RG, Carron C, Oury C, Gardellin P, Bernard O, et al. (1999) TEL is a sequence- specific transcriptional repressor. J. Biol. Chem. 274: 30132–30138.
5. Karim FD, Urness LD, Thummel CS, Klemsz MJ, McKercher SR, et al. (1990) The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. Genes Dev 4: 1451-1453.
6. Green SM, Coyne HJ, 3rd, McIntosh LP, Graves BJ. (2010) DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. J Biol Chem 285: 18496–18504.
7. Graves, B.J, Petersen J.M. (1998) Advances in Cancer Research (Woude G.V., Klein G., editors. eds) pp 1-55, Academic Press, San Diego.
8. Golub TR, Goga A, Barker GF, Afar DE, McLaughlin J, et al. (1996) Oligomerization of the ABL tyrosine kinase by the ETS protein TEL in human leukemia. Mol. Cell. Biol. 16: 4107-4116.
9. Lacronique V, Boureux A, Valle VD, Poirel H, Tran Quang, C, et al. (1997) A TEL-JAK2 Fusion Protein with Constitutive Kinase Activity in Human Leukemia. Science 278: 1309-1312.
10. Jousset C, Carron C, Boureux A, Tran Quang C, Oury, C, et al. (1997) A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TEL-PDGFR beta oncoprotein EMBO J. 16: 69-82.
11. Carroll M, Tomasson MH, Barker GF, Golub TR, Gilliland DG. (1996) The TEL/platelet-derived growth factor b receptor (PDGFβR) fusion in chronic myelomonocytic leukemia is a transforming protein that self-associates and activates PDGFβR kinase-dependent signaling pathways. Proc. Natl. Acad. Sci. 93: 14845-14850.
12. Roukens MG, Alloul-Ramdhani M, Baan B, Kobayashi K, Peterson-Maduro J, et al. (2010) Control of endothelial sprouting by a Tel–CtBP complex. Nat Cell Biol 12: 933–942.
13. Hajduk PJ, Greer J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6(3): 211-219.
14. Vanwetswinkel S, Heetebrij RJ, van Duynhoven J, Hollander JG, Filippov DV, et al. (2005) Chem Biol 12: 207-216.
15. Marquardsen T, Hofmann M, Hollander JG, Loch CMP, Kiihne SR, et al. (2006) Development of a dual cell, flow-injection sample holder and NMR probe for comparative ligand-binding studies. J. Mag. Reson 182: 55-65.
16. Siegal G, AB E, Schultz J. (2007) Integration of fragment screening and library design. Drug Discov Today 12: 1032-1039.
17. Sklenar V, Piotto M, Leppik R, Saudektim V. (1993) Gradient-Tailored Water Suppression for ^{1}H - ^{15}N HSQC Experiments Optimized to Retain Full Sensitivity. J Magn Reson 102 (A): 241 -245.

18. Bodenhausen G, Ruben, DJ. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem Phys Lett 69 (1): 185-189.
19. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
20. Kobayashi M, Retra K, Figaroa F, Hollander JG, Ab E, et al. (2010) Target immobilization as a strategy for NMR-based fragment screening: comparison of TINS, STD, and SPR for fragment hit identification. J Biomol Screen 15 (8): 978-989.
21. Murali N, Miller WM, John BK,Avizonis DA, Smallcombe SH. (2006) Spectral unraveling by space-selective Hadamard spectroscopy. J Magn Reson 179: 182-189.
22. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3): 277-93.
23. Goddard TD, Kneller DG. SPARKY 3, University of California, San Francisco.
24. Neal S, Nip AM, Zhang H, Wishart DS. (2003) Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. J. Biomol. NMR 26: 215-240.
25. Jung, YS and Zweckstetter, M. (2004) Mars-robust automatic backbone assignment of proteins. J Biomol NMR 30(1): 11-23.

Chapter 3

Inhibition of Small GTPases by Stabilization of the GDP Complex, a Novel Approach applied to Rit1, a Target for Rheumatoid Arthritis

Dipen M. Shah[1], Masakazu Kobayashi[2], Peter H. J. Keizers[1], Adriaan W.Tuin[1], Eiso AB[5], Linda Manning[3], Anna A. Rzepiela[4], Martin Andrews[2], Flip J. Hoedemaeker[3] and Gregg Siegal[1,5]*

[1] Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands.
[2] Galapagos NV, Mechelen, Belgium
[3] KeyDP, Amsterdam, the Netherlands
[4] Pyxis Discovery, Delft, the Netherlands
[5] ZoBio, Leiden, the Netherlands

Manuscript to be submitted to the Journal of Medicinal Chemistry

Abstract

A fragment based drug discovery approach has been used to discover phthalimide-based inhibitors of a novel Rheumatoid Arthritis (RA) protein target Rit1, a member of the Ras superfamily of small GTPases. NMR screening and a nucleotide exchange biochemical assay were used to discover and validate fragment hits with IC_{50}'s in the millimolar range (**1,2**) that specifically inhibit the interconversion between the GDP bound "inactive form" and the GTP bound "active form" of Rit1. Elaboration of fragment hits yielded compounds with IC_{50}'s of ~180 μM (**16e**, **17e** and **18c**) in the biochemical assay. A crystal structure of the binary Rit1:GDP complex was solved to 2.3 Å resolution, however, it proved impossible to obtain crystals with any of the Rit1 inhibitors. Instead alternative, NMR-based methods in combination with molecular docking were employed to obtain 3D structural information on the ternary complex of Rit1·GDP and inhibitor **18c**. The docking solutions show that **18c** exchanges between two binding sites that lie very close to the GDP binding pocket of Rit1. The docked structures further reveal a novel mechanism of action, where in one of the binding modes, **18c** causes a steric occlusion of the GDP and likely inhibits its dissociation.

INTRODUCTION

Rheumatoid Arthritis (RA) is a chronic degenerative disease, characterized by inflammation and erosion of the joint structures. When not treated, it leads to substantial disability and pain in the joints. The onset of RA is characterized by the inflammation of the synovial membrane ('Synovitis") which correlates with increased expression of a specific matrix-metalloproteinase (MMP) MMP1, by synovial fibroblasts (SFs). Numerous treatments for aspects of the disease are available e.g. *NSAIDS* (Non-Steroidal Anti-inflammatory Drugs) to limit inflammation and *DMARDS* (Disease Modifying Anti-Rheumatic Drugs), but most of them are found to have limited efficacy for the treatment RA in the longer term.[1] In an effort to find new targets with therapeutic value for RA, an siRNA screening strategy was undertaken. The screen uncovered a group of potential protein targets, of which a specific protein, Rit1, was involved in an early stage of the onset of RA (US patent 7,919,259 B2). Rit1 is an interesting therapeutic target as it regulates the expression and secretion of MMP1 leading to the onset of collagen degradation. Rit1 activity is upregulated in synovial fibroblasts.[2] This observation suggests that inhibition of Rit1 activity could have therapeutic benefits for RA patients.

Rit1 (Ras-like protein in all tissues) is a member of the Ras superfamily of GTPases.[3] Proteins of the Ras family, which comprises more than 150 members in humans, are small (~25 kDa) and monomeric. Ras members function as molecular switches that govern a wide variety of signal transduction pathways that regulate e.g. cellular growth, proliferation, differentiation and apoptosis.[4,5,6] GTPases carry out their cellular function by cycling between an inactive GDP bound form and an active, GTP bound form.[7] Shuttling between active and inactive forms is facilitated by two classes of regulatory proteins, namely Guanine nucleotide exchange factors (GEFs)

and GTPase activating proteins (GAPs). GEFs function by catalyzing the exchange of GDP for GTP thereby activating the G protein. In the active form, G proteins interact with downstream effectors and thus propagate further signaling events such as transcription regulation that lead to the desired biological response. The low intrinsic GTPase activity of G proteins is enhanced by interaction with GTPase activating proteins (GAPs), that down regulate signaling activity by speeding conversion to the inactive form.[8] Numerous studies have shown that proteins from the Ras family are overexpressed in about one third of human cancers and thus are implicated as oncology drug targets.[5]

Ras GTPases undergo extensive post-translational modifications that regulate their interaction with other proteins, protect them from proteolytic degradation, facilitate association with the cell membrane or determine their subcellular localization and function. Previous attempts to target this family pharmacologically have focused mainly on inhibition of post-translational modifications such as *farnesylation* or *geranylgeranylation.*[9,10,11] These inhibitors prevent membrane association of the GTPase and hence cause the loss of cellular function. However, the performance of such compounds in clinical trials has not been as good as hoped, mainly due to adverse effects and toxicity related issues.[12] Many other therapeutic approaches such as targeting prenylation and post-prenylation modifications, targeting GEFs and GAPs and the functional regulation of GDP/GTP exchange are exciting alternatives and are reviewed elsewhere .[12]

In contrast to other Ras GTPases, Rit1 lacks a known recognition signal for C terminal lipidation, which is necessary for membrane association.[13] It has been reported by a number of groups that Rit1 signals to a variety of Ras responsive transcription factors, weakly transforms NIH 3T3 cells and binds to and activates

RGL3, a novel RalGEF.[14] However, these studies also show that Rit1 fails to activate the mitogen-activated protein (MAP) kinase or phosphatidylinositol 3-kinase (PI3K)/Akt kinase signaling cascades in NIH3T3 cells, suggesting that Rit1 uses novel effector pathways to regulate proliferation, transformation and differentiation.[15,16] These alternative approaches to inhibition of Rit1 are therefore not directly available.

With a view towards finding small molecule inhibitors of Rit1, we have employed a fragment based drug discovery approach.[17] Given the large number of GTPases present in the cell and the apparent lack of post-translational modification, we sought compounds that would specifically bind to and stabilize the Rit1-GDP complex, thereby inhibiting Rit1 dependent transcription. Validated screening hits were elaborated by analogue searching and synthetic chemistry which generated phthalimide based inhibitors exhibiting a modest micromolar affinity as judged by a biochemical nucleotide exchange assay.[14] Although we were able to elucidate a 2.3 Å resolution X-ray crystal structure of Rit1 bound to GDP, attempts to soak the inhibitors into the Rit1-GDP crystals were fruitless. Therefore, we have used NMR based methods to obtain structural information to guide further compound elaboration efforts. The resulting docking model suggests that the compound binds adjacent to the GDP. The structural information is consistent with the mode of inhibition observed in the biochemical assay.

Results & Discussion

Crystal Structure of Rit1 bound to GDP

The crystal structure of Rit1 bound to GDP (Figure 1) to 2.3 Å resolution by molecular replacement using the ras-like protein TC21 as a search model (PDB: 2ERY). The crystals contained an exceptionally low solvent content (~20%) and the

protein exhibited multiple crystal contacts apparently contributing to a rather rigid overall structure as judged by the low B-factors. Not unexpectedly, the structure of Rit1 is similar to other small GTPases and consists of the canonical fold comprising a β sheet with one anti-parallel and four parallel strands surrounded by four α-helices.[18,19] Electron density was not found for the switch II loop formed by residues 79-87, likely due to flexibility. Based on structural similarity to other GTPases, the P-lobe (orange), which forms the characteristic phosphate binding loop, consists of residues Leu 27 – Lys 34 in Rit1, while the switch I loop extends from Phe 46 to Ala 57. The orientation of the GDP nucleotide is conserved and the magnesium ion is coordinated by the oxygens of the β phosphate of the nucleotide. The electron density for the GDP and magnesium ion was clearly interpretable. Further commentary on the structure is provided in the Supplementary Information (Table S1).

Screening for Rit1-GDP Specific Ligands

A strategy was employed to search for ligands specific for GDP bound Rit1, in which. Ligand screening was accomplished using TINS (Target Immobilized NMR Screening) on both the apo and GDP bound forms of the protein, screening against a fragment collection and hits that were unique for the complex were further investigated.[20,21] The PH domain of human Akt1 was used as a reference protein. Both Rit1 and the Akt1 PH domain were immobilized using Schiff's base chemistry, which at the pH used primarily targets the amino terminus. The functionality of immobilized Rit1 was determined using a nucleotide exchange assay (Supplementary Figure S1).[14] Immobilized GDP-free Rit1 was generated by incubation with alkaline phosphatase (AP) in the presence of a non-hydrolysable analogue of GTP, followed by incubation with phosphodiesterase (PDE).[22] Since

nucleotide free Rit1 rapidly precipitates, ligand screening could only be accomplished using an approach such as TINS where the protein is immobilized on a solid support.[22]

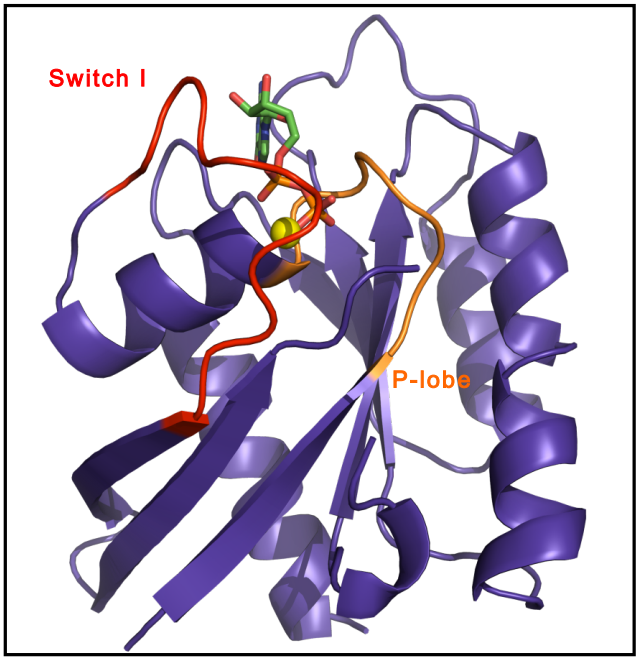


Figure 1. The crystal structure of Rit1. The switch I loop is shown in red and the phosphate recognition region, also called the P-lobe, in orange. These loops are defined based on the structural similarity of Rit1 to other GTPases.[18,19] The yellow sphere indicates the magnesium ion coordinated by the oxygens of the β phosphate of the nucleotide. The electron density for residues that form the switch II loop is missing and hence not represented in the crystal structure. The GDP is shown in sticks (green, orange, red and blue).

A library of nearly 1,000 commercially available fragments was screened against both the GDP-bound and apo-form of Rit1. The design philosophy and characteristics of the library have been previously described.[23] The hit rate observed was 3.4% and 8.8% for GDP bound- and Apo-forms of Rit1 respectively, consistent with the idea that the primary ligand binding site is the nucleotide binding pocket. The screens

resulted in the identification of 33 and 99 fragments with affinity for the GDP-bound and apo-forms of Rit1, respectively. 18 of the fragments bound both forms of Rit1, while the remaining hits bind specifically to either apo or GDP-bound Rit1.

Characterization of primary fragment hits

All hits from the TINS screen were assayed for inhibition of nucleotide exchange using purified protein.[14] Rit1 has an intrinsically slow off-rate for GDP and, since no GEF specific for Rit1 is known, it is not practical to perform the assay under equilibrium conditions.[24] Each hit from the screen was assayed at 4 mM and subsequently the IC_{50} of hits that exhibited a percentage of inhibition (PIN) above 60% was determined where possible. Since it did not prove possible to determine the IC_{50} of all weakly binding fragment hits, the fragments were ranked according to the PIN (See supplementary Figure S2). The five fragments, which exhibited a PIN greater than 60% (Figure 2) were selected for further characterization.

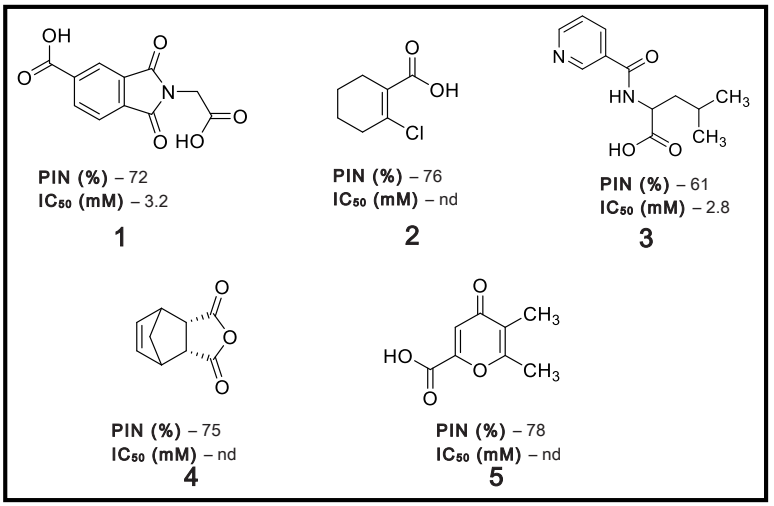
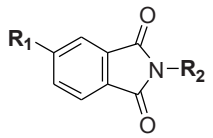


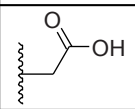
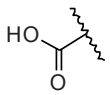
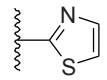
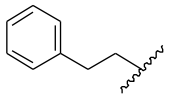
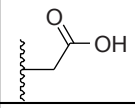
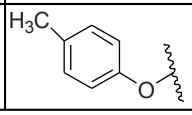
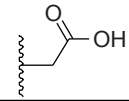
Figure 2. The five most potent fragment hits in the nucleotide exchange assay. The percentage of inhibition of the nucleotide exchange (PIN) and IC_{50} values are indicated. For fragment hits where the hill slope was too steep (>2.0), IC_{50} value could not be determined (nd).

Our repeated attempts to either soak **1** into existing crystals of Rit1·GDP or co-crystallize the ternary complex were unsuccessful. In previous reports a series of hydroxylamine bearing inhibitors were shown to inhibit the intrinsic nucleotide exchange of Ras.[25,26] The NMR based structure of a Ras-inhibitor (SCH-54292) complex revealed that the conformation of the rather flexible switch II region may be locked by the binding of the inhibitor to this region. The switch II region is known to play a role in the rate of nucleotide exchange.[25] The naphthyl moiety of SCH-54292 is buried in a hydrophobic pocket in the vicinity of switch II, while the hydroxylamine at the other end of SCH-54292 likely chelates the Mg^{2+} ion. As seen in the crystal structure of Rit1, the presence of the Mg^{2+} ion is critical for nucleotide binding and is co-ordinated by the oxygens of the β phosphate of the GDP. Based on the Ras-inhibitor (SCH-54292) NMR structure, it was hypothesized that **1** might bind to Rit1 in a pocket adjacent to GDP and in the vicinity of the Mg^{2+} atom. The carboxylic acid moiety together with one of the carbonyl groups of the phthalimide ring, was hypothesized to chelate the metal. However, with the inability to obtain a crystal structure, no conclusive experimental data was available to identify the binding site of **1**. In order to test the binding hypothesis, we sought commercially available analogues of **1**. Substructure searches were used to first find analogues containing additional molecular features on the phenyl ring of **1**. A second set of analogues was selected to replace the acetate motif on the N of the phthalimide with groups that would either ligate the Mg^{2+} tighter or provide a conformation more optimal for ligation. If metal ligation was indeed important for binding, some of these analogues would be expected to show improvement in potency in the nucleotide exchange assay. Each of the analogues was assayed for inhibition of nucleotide exchange and ranked based on the IC_{50} (Table 1).[14] Compound **6** was moderately more potent

than **1**, (~1.5 fold) while compound **7** resulted in a 4 fold increase in the potency. However, introduction of a phenyl group at R1 as in compounds **8** and **9,** results in a roughly six fold improvement in potency compared to **1**. The SAR data suggested that the carboxylic acid functionality was required at only one end of the molecule and that addition of a hydrophobic, preferably aromatic, moiety at the other end could lead to enhanced potency, although the ligand efficiency remained rather low (~0.14 pIC_{50}/Heavy atom count).[27] Interestingly, it did not appear to matter which end of the molecule contained the carboxylic acid or the aromatic ring. However, the data did not give a clear picture of whether or not metal ligation was important for binding.

Table 1. Representation of the expansion process from primary fragment hit. The table shows the features that were explored by modification of R_1 and R_2 groups resulting in the four most potent commercial analogues of **1**.



No.	R_1	R_2	Hill Slope	IC_{50} (mM)
6			1.13	2
7			1.15	0.97
8			1.58	0.520
9			1.12	0.600

Since crystallography failed to yield structures of Rit1 with any ligand other than GDP, we turned to protein observed NMR to obtain structural information. It is well known that the 2D [^{1}H,^{15}N]-HSQC NMR experiment can provide low resolution structural information on ligand binding if the backbone resonance assignment of the protein is available. Using standard through-bond, triple resonance NMR experiments, the sequential assignment for about 72% of the backbone of Rit1 was determined (see Experimental Section). Subsequently, compound **8** was titrated into uniformly ^{15}N labeled Rit1 and residues with significant chemical shift perturbations (CSP) of the amide cross peak were identified (δH > 0.01 ppm). Potential binding sites were mapped onto the Rit1 crystal structure using the program *j-surf*.[28] Note that the NMR titration experiments were carried out at pH 9.1 as the stability of the protein was significantly greater than at pH 7.5, the pH at which the TINS screen was performed. Significant CSPs were clearly observed close to the the GDP binding site and on the opposite side of the protein close to the diphosphate moiety (Figure 3). Although this data gave us confidence that the ligand bound close to the nucleotide, it was not sufficient to precisely define the site.

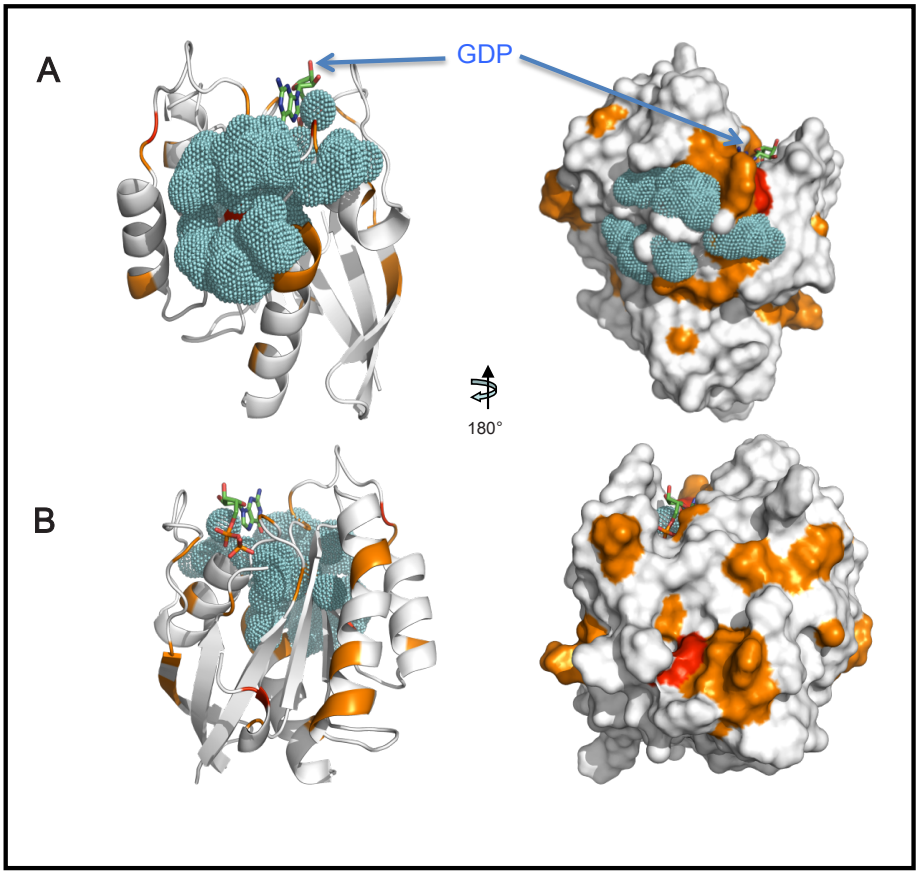
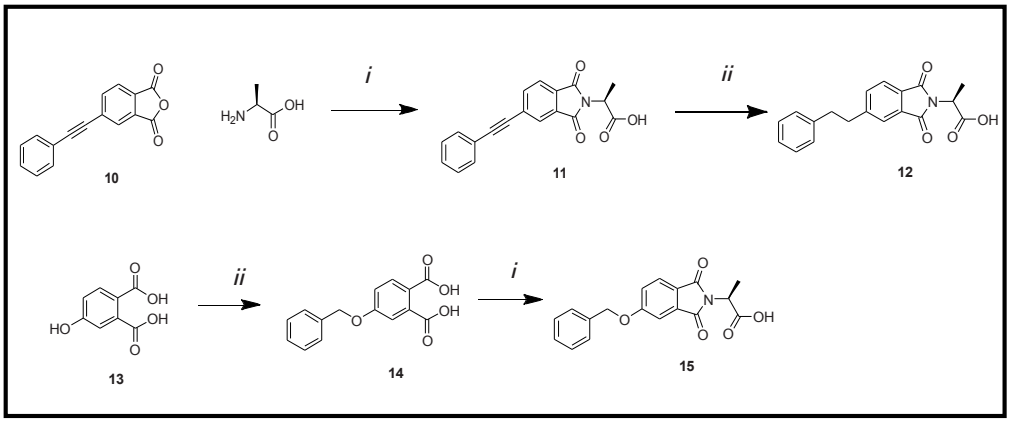


Figure 3. Mapping of the chemical shift perturbations induced by binding of **8** onto Rit1. The program j-surf was used to calculate potential binding sites.[28] The small dots forming spheres indicate the likely binding site of the compound as determined by j-surf. The red and orange colors represent the location and magnitude (larger and smaller respectively) of significant chemical shift perturbations observed on the protein. A and B show the two possible binding sites as obtained from the CSP data. This and subsequent structure figures were created in PyMOL.[29]



Scheme 1. Synthetic approach to elaborate compounds based **8.** Three different series of compounds **16a-e**, **17a-e** and **18a-c** (see Table 1) were accessible via this scheme. See experimental section for more information. *Reagents and conditions; i*) AcOH, MW 165°C, 5h. *ii*) EtOH, Pd/C, 2h. *iii*) BnCl, KOH, H_2O, reflux, 16h.

Elaboration of Compound 8

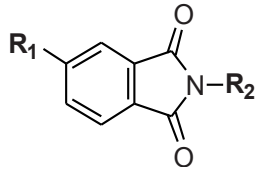
The chemical shift binding data, while not precisely defining the mode of interaction of **8** with Rit1, did suggest the presence of a shallow, hydrophobic patch adjacent to one of the binding sites. Based on this model, we explored the SAR around **8** using the approach outlined in scheme 1. For optimization of compound **8**, we focused on modifications of the phenyl (R_1-) and acidic (R_2-) ends of the molecule, keeping the carbon functionalized phthalimide ring structure intact. Replacement of the phenylethyl group with either a benzyl-oxy or a phenyl-ethynyl moiety resulted in compounds (**16a-d and 17a-d**) with no significant gain in potency (Table 2). The linker to the carboxylic group was modified to include a second methylene carbon to give compounds **16b** and **17b** with 2-methyl – acetic acid functionalized carbon phthalimides, however extension of the linker did not result in improved potency. The linker was then modified to include a branched methyl group with either (*S*) or (*R*) stereochemistry (**16c-d,17c-d and 18a-b**). All analogs containing the *S* stereoisomer led to reduced or unchanged potency. However, the *R* enantiomer led to significant improvement in potency with respect to the *S*

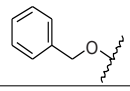
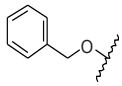
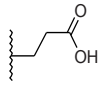
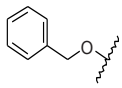
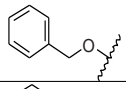
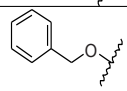
enantiomer, except in the case where the phenyl ring is connected *via* an ethyne spacer to the phthalimide core. In this latter case the reduced flexibility may affect the interaction of the acidic end of the molecule with the protein. A further significant improvement in potency (about 4 fold) was achieved by addition of a phenyl ring to the branched (*R*) methyl group resulting in compounds **16e**, **17e** and **18c**. As **18c** was found to be the most soluble of this series, it was pursued for biophysical characterization.[30,31,32]

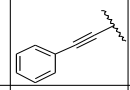
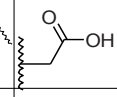
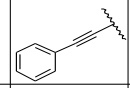
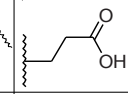
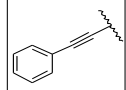
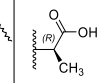
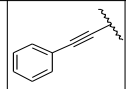
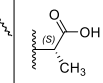
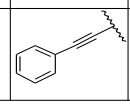
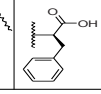
Determining the binding site of compound 18c

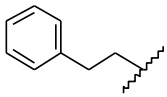
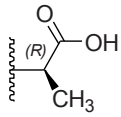
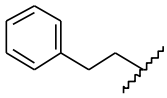
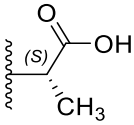
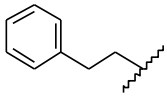
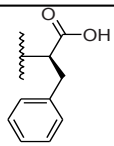
Compound **18c** was titrated into ^{15}N labeled Rit1 and the chemical shift perturbations in the [^{1}H,^{15}N] HSQC experiment were again mapped onto the crystal structure (PDB 4KLZ).[33,34] The key protein residues involved in binding are Tyr 167, Tyr 169, Tyr 170 and Asp 137, which are adjacent to the GDP binding pocket. In addition to these residues, CSPs were also observed for residues 107, 110, 116, 113, 90 and 94, similar to what was observed for compound **8,** indicative of a secondary binding site. However, it is to be noted that this secondary site is close to the switch II loop of Rit1 that is missing from the crystal structure. This might suggest that these CSPs could be either due to a conformational rearrangement caused upon binding of **18c** or influenced by the flexibility of this part of the protein. As the data suggests the possibility of two binding sites, further structure restraints were necessary in order to more precisely define the nature of the binding of **18c** to the protein and to shed light on the mechanism of inhibition of nucleotide exchange. Since sidechain resonance assignments were not readily available due to the marginal quality of the NMR spectra, we chose to investigate the use of paramagnetic NMR effects for structure elucidation by introducing an organic radical (MTSL) at the 2' position of the GDP(referred to as GDP-SL).

Table 2. Analogue compounds synthesized in house by SAR on compound **8**.



No.	R_1	R_2	Hill slope	IC_{50} (mM)
16a			13.2	0.829
16b			2.1	0.575
16c			2.38	0.398
16d			1.87	1.525
16e			3.6	0.136

No.	R_1	R_2	Hill Slope	IC_{50} (mM)
17a			1.9	0.717
17b			2.6	0.922
17c			2.9	0.819
17d			3.3	0.672
17e			1.0	0.173

No.	R_1	R_2	Hill Slope	IC_{50}
18a			1.3	0.632
18b			1.0	3557
18c			1.6	0.177

The presence of the unpaired electron of MTSL results in Paramagnetic Relaxation Enhancement (PRE) of affected NMR resonances. The PRE effect is dependent (r^{-6}) on the distance between the nucleus and the unpaired electron and can be readily calibrated by observing the effect on resonances of the protein. Determination of the PRE effects thus provides unique, long range distance information, typically from 12 to 22 Å.[35] In order to calibrate the PRE effects, a 2D [^{1}H,^{15}N] HSQC spectrum was recorded of ^{15}N labeled Rit1 with either the paramagnetic GDP-SL or diamagnetic GDP bound. For each observed amide proton, the intensity ratio (I_{para}/I_{dia}) of the resonances in the paramagnetic and diamagnetic spectra was determined, from which the PRE was converted into a distance as described.[35] The observed distances were then used to back calculate the position of the radical using the Rit1-GDP crystal structure. The average position of the nitroxide radical of the GDP-SL relative to the protein backbone could be determined with high reliability and was found to be close to the bound GDP in the crystal structure. The back-calculated nitroxide to amide proton distances match the experimentally determined distances very well, as judged by the low Q-value of 0.059 (Supplementary Figure S3). Q value is a statistical parameter in which the distances from the docking model are compared to the experimentally obtained. The smaller the value, the better the fit of back calculated versus experimental data. A model of the nitroxide spin labeled GDP bound to Rit1 was generated based on the experimental distances using XPLOR-NIH (For more details refer to the experimental section).[36] The model was subsequently used to characterize the binding of **18c**.

The PRE effect on the ^{1}H resonances of **18c** was determined by comparing the intensity of peaks in the presence of either GDP or GDP-SL bound Rit1. As seen in

Figure 4, resonances of **18c** are broadened in the presence of diamagnetic Rit1·GDP, confirming that it binds to the binary complex. Most, but not all, of the ^{1}H resonances of **18c** are further broadened beyond detection in the presence of paramagnetic Rit1·GDP-SL, indicating that they are strongly affected by the presence of the spin label. The observed PRE's were converted into distances as was done for the protein, taking into account the fractional occupancy. These intermolecular distance restraints were then used to calculate a model of the ternary complex of Rit1·GDP-SL·**18c** by docking.

Molecular Docking using PRE distance restraints

All observable protons of **18c** experience a paramagnetic effect of varying degrees due to the nitroxide radical of the GDP-SL, indicating that all protons from the compound must approach the GDP-SL for at least some part of the time (Figure 4). The docking calculations were performed using the PRE restraints, CSPs and saturation transfer effects. Saturation transfer measurements (STD) were performed to obtain structural information on the spatial proximity of ligand protons to the protein surface. This approach, also known as group epitope mapping, was originally proposed by Mayer and Meyer, 2001.[37,38,39] In the case of **18c**, the saturation transfer curves show similar kinetics for all protons suggesting that either the ligand binds in different orientations or is approximately equally surrounded by the protein on all sides. As a result, the STD data was included as ambiguous restraints in the docking procedure (Figure S6). The resulting structures exhibited a poor correlation between experimental restraints and back calculated distances with a Q-value of 0.07 (Figure S7). Specifically, the average distances in the ensemble of docked structures is consistently less than the corresponding experimental restraint with large violations.[35]

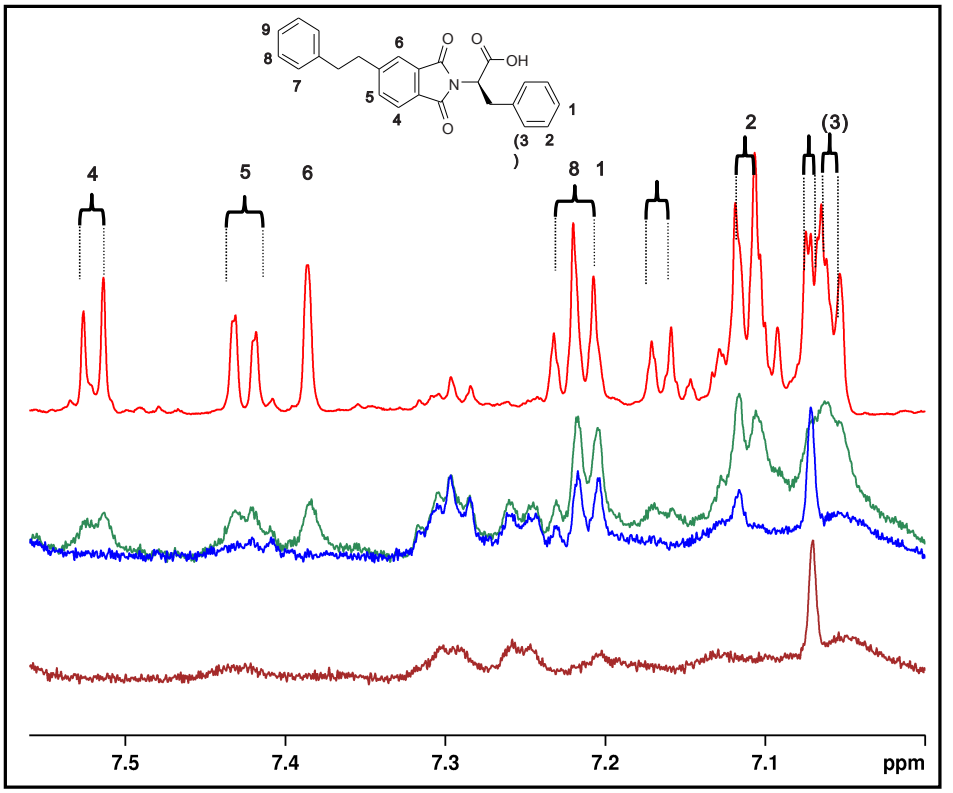


Figure 4. Paramagnetic relaxation enhancement (PRE) effect on the ^{1}H resonances of **18c** bound to Rit1·GDP-SL. The structure of compound **18c** is indicated with proton assignments. The corresponding resonances in the aromatic portion of the ^{1}H NMR spectrum are indicated. A reference spectrum of **18c** is provided (red). Peak broadening in the presence of Rit1·GDP (green) is indicative of binding. Further broadening due to PREs is clearly observed for many resonances in the presence of Rit1·GDP-SL (blue). A reference spectrum of Rit1 ·GDP-SL in the absence of **18c** is provided (brown). Due to considerable overlap in the resonances caused by the overall symmetry of **18c**, one proton (3) could not be definitively assigned and resonances arising from protons 8 and 1 could not be distinguished. The resonances from the aliphatic region of the 1D spectrum are shown in the SI (Figure S5).

The pattern of violations observed in the docking based on a single set of restraints suggests the possibility of averaging over multiple orientations or possibly binding sites. Given the symmetry of the compound, with the two phenyl groups attached to the phthalimide moiety, it is conceivable that the compound binds in two orientations. Therefore, it was decided that the PRE-based distance restraints should

be divided into two groups, i) one PRE set representing the protons of the acidic side of the compound and ii) a second PRE set representing the protons of the non-acidic side of the compound. Subsequent docking calculations were performed with either of the two sets of PRE restraints, CSPs and the ambiguous restraints from STD. The five lowest energy solutions from each of the two calculations show two minima (Figure 5). In one binding mode, defined by the restraints from the non-acidic end, **18c** lies in a shallow groove parallel to the sugar and phosphates of GDP and crosses over to occlude the base. In the second binding mode, the compound binds in the groove above the β3-strand formed by residues 160:163. Within the five lowest energy structures for each of the binding modes, **18c** is found in two orientations, rotated by approximately 180 degrees. In contrast to the calculations in which all of the PRE restraints were simultaneously used, the intermolecular distance restraints for the two ensembles of five structures fit very well to the experimentally determined distances, with a Q-value of 0.04 (supplementary Figure S7).

Discussion and Conclusion

In this work we aimed to develop small molecule inhibitors of the small GTPase Rit1, a novel target for rheumatoid arthritis.[2] We report here for the first time the crystal structure of Rit1 bound to GDP solved to 2.3 Å. This was a useful starting point, but since it proved impossible to obtain structures with the inhibitors bound did not directly aid in driving design. Thus, the elaboration of the initial 3 mM fragment hit was based on structural information obtained by NMR methods. The NMR methods were selected for the ability to rapidly generate reliable constraints for molecular docking studies and included chemical shift perturbations derived from 2D [^{1}H,^{15}N] HSQC spectra, epitope mapping by saturation transfer difference spectroscopy and paramagnetic relaxation enhancement derived intermolecular distances. The latter

two types of restraints can also be employed in cases where isotope labeling of the target is not possible. This demonstrates that NMR can provide structural information with sufficient detail and timeliness to support medicinal chemistry efforts when X-ray crystallography is not available. Although the directly modulation of GTPase activity with small molecules has proven challenging, we have been moderately successful in elaborating the phthalimide based inhibitors to an IC_{50} of approximately 180 μM. Importantly, the inhibitors have a novel mode of action.

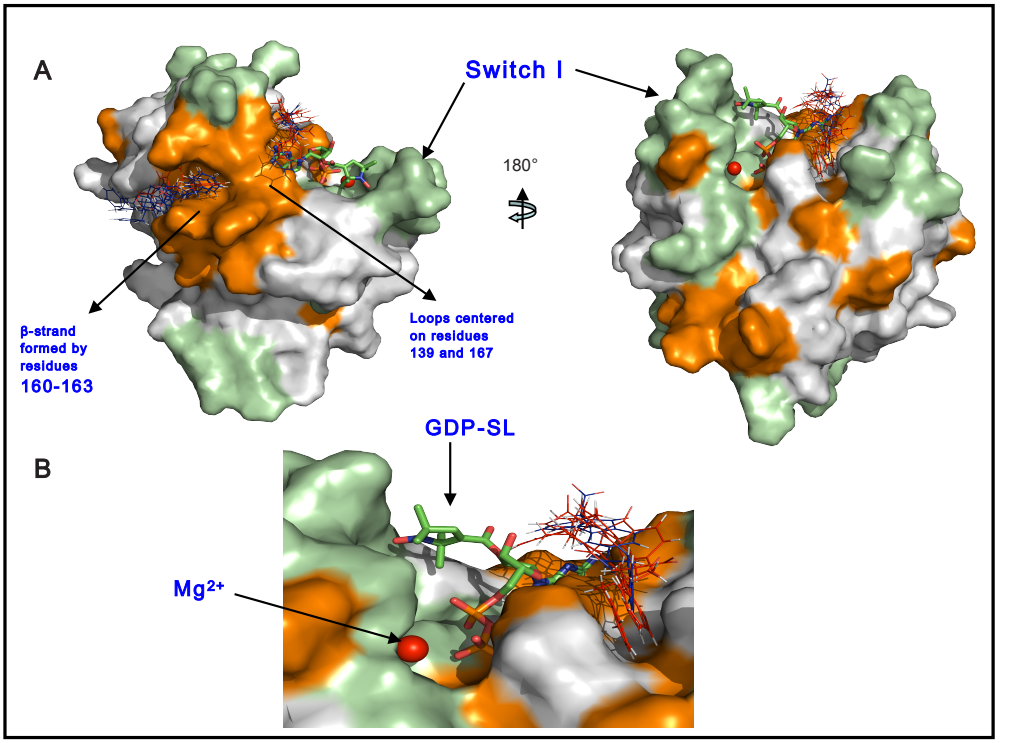


Figure 5. Experimentally restrained docking of **18c** onto the Rit1·GDP-SL complex. **A]** The lowest energy binding solutions as determined by docking using restraints from PRE, CSP and saturation transfer. The surface of Rit1 is presented and the bound GDP-SL (green) is shown in bond representation. The presence of magnesium ion is indicated by a red sphere. Compound **18c** is also shown in bond representation (narrow lines) where the two different orientations, rotated by 180°, are shown in red and blue. Amino acids with significant chemical shift perturbations observed in the presence of compound are shown in orange. The protein residues for which there are no assignments available are shown in pale green. **B]** Closeup view of one of the binding modes of compound **18c** binding adjacent to the GDP-SL (green). This binding site likely stabilizes the Rit1·GDP complex by steric hindrance of nucleotide exchange.

The Rit1·GDP·18c complex

The PRE based data and the subsequent docking results reveal that compound **18c** does not bind Rit1 in one preferred orientation or even within a single binding site. Rather, **18c** explores a large surface near the GDP binding pocket of Rit1. The two specific binding modes found in the restrained docking indicate that **18c** binds in a shallow groove. These two binding grooves are physically separated in the Rit1 crystal structure by the bulk of the loops consisting of residues 135 - 145 and 165 – 170, although at the closest points the two sites are within 6.5 Å and translocation from one site to the other could be accomplished by a simple flip (Figure 5). Interestingly, the Hill slope of 1.6, which suggests somewhat greater than 1:1 stoichiometry, is consistent with the two binding modes from the docking. In Figure 4, it is clear that protons from both phenyl rings of **18c,** which are some 12 Å distant, experience a considerable paramagnetic effect. This data means that both phenyl rings of **18c** must closely approach the MTSL during the complex formation, at least for some part of the time. For the base occluded binding mode, both ends of 18c are sufficiently close to MTSL that in principle a single orientation could explain the data, however, two orientations are found in the docking ensemble. For the binding mode close to the β3-strand, 18c must bind in both orientations (i.e. flipped 180 degrees) in order to satisfy the data. As the complex is in rapid exchange on the NMR timescale, it seems likely that the two orientations are exchanging rapidly as well and indeed, there may be additional rapid exchange between the two binding modes. To our knowledge, this is the first documented case of a small molecule exchanging between multiple binding poses, although this behavior is frequently invoked in fragment based drug discovery as a possible explanation for the failure to observe ligands in crystal structures of the target.

Neither of the binding sites of the phthalimide compounds comprises a deep

groove and in particular, the base occluded binding mode is unusual. In principal the addition of the organic radical at the 2' carbon of the sugar of GDP could perturb the structure of Rit1 and create an artificial binding site. For the β3-strand binding site, this seems particularly unlikely since the site is remote from the sugar and there is no evidence from the PRE data that the crystal structure is perturbed in this region. On the other hand, the base occluded site is adjacent to the GDP and could possibly be influenced by the presence of the MTSL. However, three observations argue against this possibility. First, the MTSL is on the other side of the GDP and does indeed perturb the structure of Rit1, but in a flexible loop distal to the proposed binding site of **18c** (Figure S4). Second, there are extensive CSPs induced by the binding of **18c** to the native Rit1:GDP complex in the base occluded binding site. This suggests that **18c** binds at this site even when the MTSL is absent. Thirdly, a truncated-driven NOE experiment suggested that the aromatic protons of a closely related compound were within 5Å of the imino proton of the base (data not shown).[40] The base occluded binding mode is most consistent with this data, even though no restraints based on the data were included in the calculation. In the base occluded model, the Mg^{2+} ion is on the opposite side of the phosphates of GDP from **18c**, therefore, it seems unlikely that the compound binds *via* metal chelation.

Novel mode of nucleotide inhibition

Recently the crystal structure of an "inactive" form of GTP bound HRas has been elucidated.[41] In this structure the Switch I loop is significantly displaced from the nucleotide, theoretically creating a potential small molecule binding site. One could speculate that if a similar Switch I displacement occurred in the GDP bound state a similar binding site might exist. However, if it does, it is on the opposite side of the nucleotide from the occluded binding mode of **18c** and we do not have

any data to suggest that **18c** nor any of the other compounds in this study might bind at such a hypothetical site. A small molecule binding site with allosteric properties has been found on KRas.[42] Interestingly, small molecules bind this site in the presence of either GDP or GTP analogues. Apparently, binding at this site inhibits the interaction with the GEF SOS. This allosteric site however is remote from the base occluded binding site (Figure 6), suggesting that 18c acts with a novel mechanism, that is steric inhibition of GDP release. This new binding site may point towards a new approach to inhibit this pharmaceutically important yet challenging class of targets.

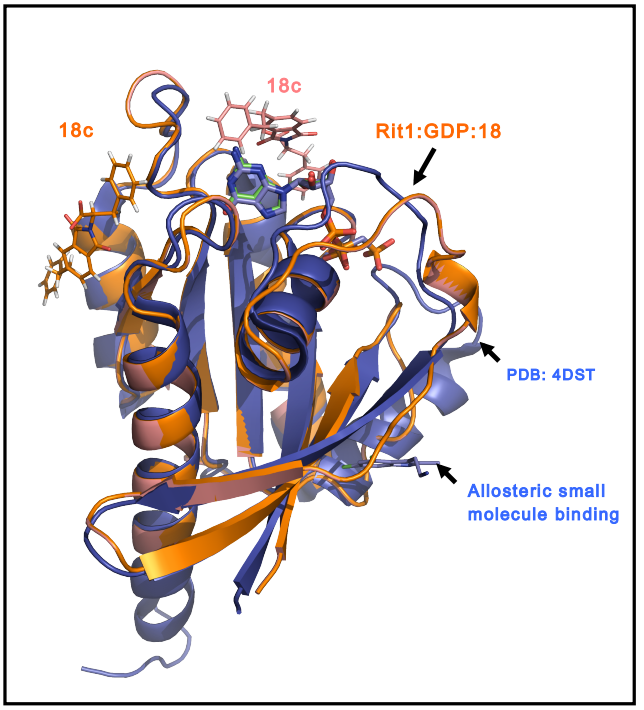


Figure 6. The comparison of the docked model of Rit1·GDP·**18c** complex with $KRas_m$·GCP:small molecule complex. The docked model of the ternary complex, Rit1·GDP-SL·**18c** (orange) is shown overlaid with and the structure of $KRas_m$·GCP:small molecule complex (blue, PDB:4DST). The two binding sites for **18c** are shown in orange and dark pink color. It is seen that **18c** binding sites are distal from the pocket 1 and switch II region but lie close to pocket 2. The allosteric binding site found for small molecule bound to $KRas_m$ lies remote from the binding sites of **18c.**[41,42]

Experimental Section
Protein Production

For X-ray crystallography, an N-terminal cleavable hexahistidine tagged-Rit1 construct (aa 19-189) was cloned into pQTEV vector and expressed in *E. coli* BL21 (DE3) codon plus (RP) (Stratagene). The protein was purified on Ni-Hitrap FF (GE lifescience) followed by purification on a superdex G75 gel filtration column (GE lifescience). The hexahistidine tag was removed by the proteolytic digestion with TEV protease (Invitrogen) according to the manufacturer recommended buffer condition. The cleaved tag and the TEV protease were removed by affinity purification on Ni-Hitrap FF, and subsequently the buffer of Rit1 (19-189) was exchanged to 20 mM Tris pH9, 5 mM NaCl, 2 mM $MgCl_2$ and 1 mM DTT.

For NMR studies, a C-terminal hexahistidine tagged-Rit1 (aa 19-189) was cloned into pET20b and expressed in *E. coli* strain BL21 (DE3) Codon Plus (RP) (Stratagene). The protein was first purified on Ni-Hitrap FF (GE lifescience) followed by purification on Superdex G75 with buffer of 20 mM Tris pH 9, 5 mM NaCl, 2 mM $MgCl_2$ and 1 mM DTT. To obtain sequential backbone resonance assignments, uniform ^{13}C, ^{15}N labeling of Rit1 was achieved by expression of C-terminally hexahistidine tagged Rit1 (aa 19-189) in BL21 (DE3) Codon Plus (RP) cells grown in M9 minimal medium supplemented with $^{15}NH_4Cl$ and ^{13}C-D-Glucose (CIL) as sole nitrogen and carbon sources. The protein was purified as described above. For protein observed HSQC studies, ^{15}N labeled C-terminal hexahistidine tagged Rit1 (aa 19-189) was produced in M9 minimal medium supplemented with $^{15}NH_4Cl$ as a sole nitrogen source and purified as described above.

For nucleotide exchange assay, a N-terminal GST (Glutathione S-transferase) tagged-Rit1 (aa 1-201) was expressed and purified as described previously.[14,43]

Crystallization and structure determination of the Rit1·GDP complex

The protein was concentrated to 10 mg/mL concentration prior to crystallization. The Rit1 crystals were grown in the presence of 11-14% w/v PEG 4000, 5-8% v/v Jeffamine M600 and immidazole buffer pH 7.0-7.2 using sitting drop vapour diffusion method. The crystals were harvested with a nylon loop and transferred to cryo-conditions by briefly passing the crystals through mother liquor augmented with approximately 30% glycerol. The crystals were then flash-frozen by exposing them to the cold-stream of the X-ray source. All the diffraction data was collected using a Cu-Kα rotating anode X-ray generator and a CCD area detector. The best diffracting crystals grew on the plastic surface of the crystallization well and diffracted up to 2.3 Å on a rotating anode X-ray beam. The diffraction data were integrated and scaled with the PROTEUM*plus* Crystallographic Software Suite by Bruker AXS and processed by CCP4 (Collaborative Computational Project, Number 4. 1994).[44] The structure of the protein was solved by Molecular Replacement, using the ras-like protein TC21 (PDB: 2ERY); 55% sequence identity) as a search model. Molecular replacement was carried out with the *MOLREP* routine in *CCP4* by searching for one molecule in the asymmetric unit. The refinement of the MR solution was carried out using *REFMAC5* from *CCP4*. Model building for the structure of Rit1 was carried out with the program *COOT.* [44,45]. The co-crystallization and soaking trial experiments for protein compound complex were unsuccessful and caused protein crystals to crack.

TINS NMR screening

Rit1 (aa 19-189) and Akt PH domain (aa 1-123) were immobilized via amine-coupling to 500 µL Actigel-ALD resin (Sterogene) in 25 mM Hepes pH 7.5, 100 mM

NaCl and 2 mM $MgCl_2$ at 4°C using the coupling reagent provided by the manufacturer. Due to low solubility of Rit1 at pH 7.5, the protein was immobilized in consecutive addition at 50 μM. The immobilization efficiency was above 90% and the final concentrations of the immobilized targets were typically in the range of 500 μM. The immobilized Rit1 was shown functional by the nucleotide exchange activity (Shao, H, et al 2000).[14,43] TINS NMR experiments were performed on 500 MHz Bruker NMR spectrometer using spatially selective Hadamard pulse sequences and analyzed as described previously.[20,46,47] GDP-bound Rit1 was immobilized as described above. The composition of the fragment library screened has been reported previously.[23] The typical fragment mix applied to the immobilized proteins consisted of 3 ~ 5 fragments which were each at 500 μM. The buffer used in the screen was comprised of 25 mM Tris-d_{11} pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$ for GDP-bound from of Rit1.

In vitro nucleotide exchange assay / High concentration bioassay

The nucleotide exchange was determined by the rapid filtration technique. The exchange buffer comprised of 1 μM of GST-Rit1 (aa 1-201), 2 μM [^{32}S] GTPγS (2 Ci/mmol, PerkinElmer) in 25 mM Tris pH 7.5, 100 mM NaCl, 1 mM DTT and 0.5 mM $MgCl_2$ (on ice), and then the spontaneous nucleotide exchange was initiated by incubation at 30°C. After 1 hour incubation at 30°C, the [^{32}S] GTPγS bound GST fusion protein was isolated onto 96-well Unifilter GF/B using Filteramate harvester (PerkinElmer) with cold wash buffer of 25 mM Tris pH 7.5, 100 mM NaCl, 1 mM DTT and 10 mM $MgCl_2$. The filter was soaked in Microsint scintillation fluid and the bound [^{32}S] GTPγS was detected using TopCount (PerkinElmer). All the fragments determined as hits specific to Rit1 from TINS NMR screening were subjected to high

concentration bioassay. Initially, the total nucleotide exchange was monitored in the presence of fragment or in DMSO. The amount of DMSO in the controls was matched to 1.5% (v/v), which is equivalent to the amount in the presence of a fragment. In 96 wells, 8 wells containing either DMSO or 400 μM of GDP were employed as internal controls. The percentage of inhibition (PIN) was determined as follows: PIN (%) = $[1 - (A_{fragment} - A_{positive\ control}) / (A_{negative\ control} - A_{positive\ control})] \times 100\%$, where $A_{fragment}$ and $A_{negative\ control}$ are the radioactive counts measured in the presence of fragments and DMSO, respectively, whereas $A_{positive\ control}$ is the radioactivity measured in the presence of 400 μM of GDP that competed out [^{35}S] GTPγS for the nucleotide site in Rit1. The diagrammatic representation of the nucleotide exchange assay is shown in Figure S10.

NMR backbone sequential assignment

The following NMR experiments were acquired at 293K on an 800 MHz Varian Inova spectrometer; [^{1}H,^{15}N] -HSQC, HNCO, HNCA, HNCACB, CBCACONH and 3D [^{1}H,^{15}N] NOESY-HSQC.[33,34] The protein sample consisted of 0.5 mM Rit1 (aa 19-189) in 25 mM d_{11}-Tris pH9, 5 mM NaCl and 5 mM $MgCl_2$. The data was processed using NMRPipe and visualized using Sparky.[48,49] The assignment process was guided by the automatic assignment program MARS using PDB 4KLZ as an input.[50]

Chemical shift perturbation and generation of binding site

The [^{1}H, ^{15}N]-HSQC were acquired at 298K on 600 MHz Bruker DMX NMR spectrometer equipped with a TXI cryo-probe.[33,34] The NMR sample was prepared in 25 mM Tris pH 9, 5 mM NaCl and 5 mM $MgCl_2$. The typical NMR sample contained 0.125 mM of the protein, compounds at various concentrations (see text) and 5% d_6-

DMSO. The pH of the samples was adjusted carefully within +/- 0.05 units after addition of the compound. During the experiment, total of 128 indirect increments with 16 scans per increments were acquired. The data was processed using Topspin 1.2/2.1 (Bruker) and visualized on Sparky.[49] Chemical shift perturbation in [^{1}H, ^{15}N]-HSQC was calculated for change in ^{1}H ppm value between in the presence and absence of a compound. The potential binding sites of a compound were mapped onto the surface of the Rit1 structure by *jsurf* using chemical shift perturbation data (delta ^{1}H ppm) as input.[28]

Saturation Transfer Experiments

Saturation transfer NMR experiments were conducted using pulse sequences as described.[37,38,39] The protein concentration was 40 μM and compound was present at a 20 fold in molar ratio. The pH of the sample was adjusted carefully within +/- 0.05 units. The on-resonance irradiation of the protein was performed at -0.5 ppm and off-resonance irradiation was performed at 100 ppm where no protein signals are present. The saturation time was varied between 50 ms to 2 sec in order to determine the buildup rate. No significant differences in the buildup rate between the protons were observed (see supplementary figure S6).

Paramagnetic NMR Studies

Synthesis of GDP-Spin label

The procedure followed was modified from literature.[51,52,53] One equivalent each of 3-Carboxy-2,2,5,5-tetramethyl-3-pyrrolin-1-yloxy, Free Radical (Toronto Research Chemicals) and N,N-carbonyldiimizdazole (dissolved in dimethyl formamide) were mixed together and incubated for 1 hour. To this one equivalent of

guanosine diphosphate sodium salt (Sigma) was added and incubated further for 1 hour at 38°C water bath while stirring and pH adjusted to 9.1 using 1 M NaOH. The reaction mixture was extracted twice with chloroform and pH of the aqueous phase was adjusted to pH 7.5 with 1N acetic acid. The solution was filtered and was further subjected to LCMS analysis using reversed-phase HPLC purification using an automated HPLC system supplied with a semi-preparative C_{18} column (10.0 mmD × 250 mmL, 5 μ particle size) was used. Mass spectra were recorded for elutes and yielded the product of expected size. Purified elutes were further lyophilized and gave a typical pale yellow color, characteristic of the presence of nitroxide radical. The presence of radical was further confirmed by acquiring an EPR spectrum. The X-band cw EPR measurements were performed using an ELEXSYS E680 spectrometer (Bruker, Rheinstetten, Germany) equipped with a rectangular cavity. A modulation frequency of 100 kHz was used. Measurements were done at temperature of 20°C, using 6.31 mW of microwave power and modulation amplitude of 0.5 G (See Figure S8 in supplementary).

PRE studies on Protein

The removal of GDP was performed by buffer exchange into 20 mM HEPES pH 7.5, 1 mM EDTA, 1 M $(NH_4)_2SO_4$, 20 % glycerol.[54] The GDP removal was checked by disappearance of a peak at 260 nm in the UV absorption spectrum. The ability of apo-Rit1 to bind to chemically modified GDP was checked by a fluorescent based- assay. In this assay, MANT-GDP [(2'-(or-3')-O-(N-methylanthraniloyl) guanosine 5'-diphosphate] (Biolog, Germany) was titrated into apo-Rit1 and change in the intensity of the MANT fluorescence caused by binding to Rit1 was measured in a concentration dependant manner. The results showed that MANT-GDP bound with

a 1:1 stoichiometry to apo-Rit1. As MANT-GDP carries a similar chemical modification at the ribose moiety of GDP as in the GDP-SL, we assumed that this would lead only to a minimal perturbation of Rit1-GDP-SL interaction (data not shown). The protein was exchanged back into the buffer composed of 25 mM Tris pH 9, 5 mM NaCl and 5 mM $MgCl_2$. The paramagnetic GDP-SL was added at a 1.5 molar excess over protein, incubated on ice for 30 min and then washed three times with 25 mM Tris pH 9, 5 mM NaCl and 5 mM $MgCl_2$ using a Centricon centrifugal concentrator with MWCO 10 kDa (Millipore). The same procedure was followed during preparation of diamagnetic control sample except that diamagnetic GDP (SIGMA) was added. The protein concentrations used were identical during NMR measurements. NMR samples contained, a paramagnetic sample consisting of 0.1 mM Rit1 bound to GDP-SL and a diamagnetic control consisting of 0.1 mM Rit1 bound to GDP in 25 mM Tris pH 9, 5 mM NaCl and 5 mM $MgCl_2$, 5%D_2O. The measurements were performed at 298 K on 600 MHz Bruker DMX NMR spectrometer equipped with a TXI cryo-probe. During the experiment, total of 128 indirect increments with 16 scans per increments were acquired. The data was processed using Topspin 2.1 (Bruker) and visualized using Sparky.[49]

PRE studies on compound 18c

C-terminally hexahistidine tagged Rit1 (aa 19-189) was expressed and purified as described above. The paramagnetic and diamagnetic protein samples were made as described above. The compound was added to the protein sample at a ratio of 1:1. 1D proton NMR spectra were acquired at 298 K on 600 MHz Bruker DMX NMR spectrometer equipped with a TXI cryo-probe. A CPMG delay of 40 ms was used to suppress residual protein signals. The data was processed using Topspin 2.1

(Bruker). For each observed amide proton, the intensity ratio (I_{para}/I_{dia}) of the resonances in the paramagnetic and diamagnetic spectra was determined, from which the PRE was converted into a distance as described.[35] The minor variations in the intensities for the compound signals in the 1D spectrum diamagnetic and paramagnetic sample of **18c** in the presence of the protein were referenced to the signal from DMSO.

Docking of Rit1·GDP·18c ternary complex

PREs of amide protons of the protein were obtained from peak intensities in HSQC spectra of paramagnetic and diamagnetic protein and subsequently used to determine the position of the nitroxide radical relative to the protein crystal structure. For this, the interatomic proton to radical distances were determined from the PREs as described,[35,55] using an experimentally determined correlation time of 11.5 ns. The correlation time was determined using [^{1}H,^{15}N] HSQC analysis with varying T1 and T2 delays The distances were divided into three different classes. The protons for which the resonances were broadened beyond detection, were given an upper limit of 14 + 3/-8 Å, protons of which the resonances were unaffected, were given a lower limit of 21 + 100/-3 Å. The distances for 27 protons, lower limits for 25 and upper limits for 5 protons were used to obtain the energy minimum of the nitroxide radical using XPLOR-NIH.[36]

The compound was docked onto the GDP-SL·Rit1 complex using XPLOR-NIH using experimental restraints from the CSPs data by HSQC NMR titration (some CSPs are shown in Figure S9), saturation transfer and paramagnetic experiments.[36] Based on the measured PREs, 10 distances between ^{1}Hs of **18c** including two aliphatic protons and the nitroxide were obtained. Large CSPs of 13 amides as

determined by titrating the compound into the protein as determined by HSQC, were translated into ambiguous distance restraints of 6 +/- 3 Å to any compound proton while averaging over $<r^{-6}>$. Similarly, based on the saturation transfer experiments, ambiguous distance restraints of 5 + 4/-3 Å were applied between any compound proton and any protein carbon atom. Protons were subsequently added manually. In a typical simulation, the compound approached the protein 10,000 times, where the compound was placed at a random position every four approaches the energy did not drop. The compound was treated flexible and was initially docked on the backbone of the protein including Cα and Cβ followed by inclusion and optimization of the sidechains. (The docking script is provided in the Supplementary Information; S12). The docking results are evaluated by a Q-value.[35] In this statistical parameter distances from the docking model are compared to the experimentally obtained. The smaller the value, the better the fit of back calculated versus experimental data.

Model Structure of GDP-Spin Label

A structure of 3-carboxy-2,2,5,5-tetramethyl-3-pyrrolin-1-yloxy was prepared in Spartan (www.wavefun.com) and parameterized using the Prodrugserver.[56] The chemical structure of the GDP spinlabel probe is shown is Figure S11. Additional dihedral and angle parameters were added to keep the nitroxide in plain with the pyrroline ring. The molecule was added to the protein and attached to the GDP using XPLOR-NIH. To this end, an ester linkage was created between the carboxylate and the 2'- hydroxyl of the GDP and the nitroxide oxygen atom was moved in the direction of the predetermined minimum energy position. The spin-labeled GDP was energy minimized, before optimizing the loop consisting of residues 42:57, surrounding the GDP based on 15 PREs determined for these residues

(Supplementary Figure S4).

Synthesis of elaborative compounds based on compound 8:

Carbon functionalized phthalimides are synthesized analogues to the literature procedure (Scheme 1).[i] Commercially available 4-phenylenthynylphthalic anhydride **10** is condensed with the appropriate amino acid in acetic acid under microwave irradiation at 165°C for four to five hours yielding the desired products in yields ranging from 80% to 97% after recrystallization or HPLC purification. Hydrogenation of the triple bond using palladium on charcoal afforded saturated phthalimides similar to **12** in yields ranging from 74% to 99% after HPLC purification. Oxygen functionalized phthalimides such as **15** were synthesized from 4-benzyloxyphthalic acid[ii] (**14**) in yields ranging from 43% to 78%.[30,31]

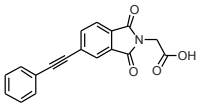
General: PE with a boiling range of 40 - 60°C was used. THF and Et_2O were distilled over $LiAlH_4$ prior to use. DCM (dichloromethane) was distilled over CaH_2 prior to use. All other solvents used under anhydrous conditions were stored over molecular sieves (4Å) except for methanol which was stored over 3Å molecular sieves. Solvents used for work-up and column chromatography were of technical grade and distilled before use. Microwave assisted reaction were carried out in an Emrys Optimizer (Biotage AB, formerly Personal Chemistry). Wattage was adjusted automatically so as to maintain the desired temperature. Unless stated otherwise, solvents were removed by rotary evaporation under reduced pressure at 40°C. Reactions were monitored by TLC-analysis using DC-fertigfolien (Schleicher & Schuell, F1500, LS254) with detection by spraying with 20% H_2SO_4 in EtOH, $(NH_4)_6Mo_7O_{24}\cdot 4H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4\cdot 2H_2O$ (10 g/L) in 10% sulfuric acid or by spraying with a solution of ninhydrin (3 g/L) in EtOH / AcOH (20/1 v/v), followed

by charring at ~150°C. Column chromatography was performed on Fluka silicagel (0.04 – 0.063 mm). For LC/MS analysis, an JASCO HPLC-system (detection simultaneously at 214 and 254 nm) equipped with an analytical C_{18} column (4.6 mmD × 250 mmL, 5μ particle size) in combination with buffers A: H_2O, B: MeCN and C: 0.5% aq. TFA and coupled to a mass instrument with a custom-made Electronspray Interface (ESI) was used. For reversed-phase HPLC purification of the final compounds, an automated HPLC system supplied with a semi-preparative C_{18} column (10.0 mmD × 250 mmL, 5 μ particle size) was used. The applied buffers were A: H_2O, B: MeCN and C: 1.0% aq. TFA. High resolution mass spectra were recorded by direct injection (2 μL of a 2 μM solution in water/acetonitrile; 50/50; v/v and 0.1% formic acid) on a mass spectrometer (Thermo Finnigan LTQ Orbitrap) equipped with an electrospray ion source in positive mode (source voltage 3.5 kV, sheath gas flow 10, capillary temperature 250°C) with resolution R = 60000 at m/z 400 (mass range m/z = 150-2000) and dioctylpthalate (m/z = 391.28428) as a "lock mass".[iii] The high resolution mass spectrometer was calibrated prior to measurements with a calibration mixture (Thermo Finnigan). ^{1}H- en ^{13}C-NMR spectra were measured on a Joel JNM-FX-200 (200/50 MHz). Chemical shifts are given in ppm (δ) relative to TMS (0 ppm) or MeOD (3.30 ppm) and coupling constants are given in *Hz*.[32,57,58]

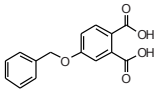
General condensation procedure: Phthalic acid or anhydride and the appropriate amino acid (1.05 equiv) are suspended in acetic acid (5 mL) and the reaction mixture is heated in a sealed tube for five hours at 165°C in the microwave. The reaction mixture is concentrated and purified by HPLC or by means of recrystalization.[57,58,59]

General hydrogenation procedure: The appropriate alkyne is dissolved in ethanol (2 ml) and treated with Pd/C (10 mg) and hydrogen gas for 2 hours at r.t. after which the catalyst is removed by filtration over a pad of celite. The filtrate is

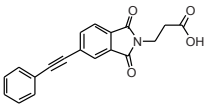
concentrated and the residue is, if required, purified by HPLC.



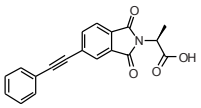
17a: Prepared from **10** (1.24 g, 5 mmol) and glycine (394 mg, 5.25 mmol) according to the general condensation procedure and purified by recrystallization from hot water. Yield: 1.48 g, 4.85 mmol, 97% off white solid. ^{1}H NMR (400 MHz, DMSO) δ 13.18 (bs, 1H, CO_2H), 8.05 (s, 1H), 8.01 (dd, *J* = 1.3, 7.7, 1H), 7.96 (d, *J* = 7.7, 1H), 7.67 – 7.60 (m, 2H), 7.50 – 7.43 (m, 3H), 4.33 (s, 2H). ^{13}C NMR (101 MHz, DMSO) δ 168.74, 166.52, 166.43, 137.34, 131.99, 131.71, 130.53, 129.60, 128.84, 128.60, 125.77, 123.78, 121.36, 93.51, 87.88, 39.02. LC-MS: *t*R: 7.91 min (linear gradient 10-90% in 13.5 min), m/z = 611.1 $[2M+H]^+$, 306.1 $[M+H]^+$, 259.9 $[M-CO_2+H]^+$



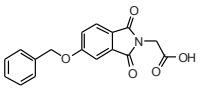
14: Prepared according to literature procedure. Yield: 5.66 g, 20.8 mmol, 76%. ^{1}H NMR (400 MHz, MeOD) δ 7.94 (d, *J* = 8.5, 1H), 7.49 – 7.26 (m, 6H), 7.13 (d, *J* = 8.6, 1H), 5.15 (s, 2H). ^{13}C NMR (101 MHz, MeOD) δ 172.33, 170.53, 162.56, 138.18, 137.83, 133.95, 129.86, 129.58, 129.12, 128.68, 124.84, 117.32, 116.89, 71.30.



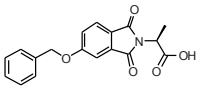
17b: Prepared from **10** (496 mg, 2 mmol) and β-alanine (187 mg, 2.1 mmol) according to the general condensation procedure and recrystallized from EtOAc/PE. Yield: 549 mg, 1.7 mmol, 85% as white solid. ^{1}H NMR (400 MHz, DMSO) δ 12.38 (bs, 1H), 7.99 – 7.94 (m, 2H), 7.89 (d, *J* = 7.6, 1H), 7.65 – 7.60 (m, 2H), 7.49 – 7.44 (m, 3H), 3.80 (t, *J* = 7.4, 2H), 2.62 (t, *J* = 7.4, 2H). ^{13}C NMR (101 MHz, DMSO) δ 172.04, 166.92, 166.83, 136.96, 132.24, 131.68, 130.86, 129.55, 128.83, 128.14, 125.37, 123.39, 121.39, 93.20, 87.95, 33.72, 32.25. LC-MS: *t*R: 7.91 min (linear gradient 10-90% in 13.5 min), m/z = 319.9 $[M+H]^+$, 302.1 $[M-H_2O+H]^+$, 259.9 $[M-CH_2CO_2+H]^+$



17c: Prepared from **10** (496 mg, 2 mmol) and alanine (187 mg, 2.1 mmol) according to the general condensation procedure and purified by recrystallization from hot water yielding the title compound as white solid (522 mg, 1.6 mmol, 81%).^{1}H NMR (400 MHz, DMSO) δ 13.14 (s, 1H), 8.05 – 7.99 (m, 2H), 7.94 (dd, *J* = 0.6, 7.7, 1H), 7.67 – 7.61 (m, 2H), 7.50 – 7.45 (m, 3H), 4.89 (q, *J* = 7.3, 2H), 1.56 (d, *J* = 7.3, 3H). ^{13}C NMR (101 MHz, DMSO) δ 170.89, 166.45, 166.35, 137.31, 131.89, 131.71, 130.46, 129.60, 128.85, 128.51, 125.66, 123.69, 121.35, 93.43, 87.91, 47.11, 14.74. LC-MS: *t*R: 8.33 min (linear gradient 10-90% in 13.5 min), m/z = 320.0 $[M+H]^+$, 274.0 $[M-CO_2+H]^+$

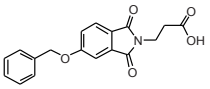


16a: Prepared from **14** (516 mg, 2 mmol) and glycine (157 mg, 2.1 mmol) according to the general condensation procedure to yield to title compound as white solid (440 mg, 1.41 mmol, 70%). ^{1}H NMR (400 MHz, DMSO) δ 13.13 (s, 1H), 7.85 (d, *J* = 8.3, 1H), 7.52 (d, *J* = 2.2, 1H), 7.48 (d, *J* = 7.1, 2H), 7.44 – 7.39 (m, 3H), 7.38 – 7.32 (m, 1H), 5.32 (s, 2H), 4.28 (s, 2H). ^{13}C NMR (101 MHz, DMSO) δ 168.92, 166.89, 166.80, 163.59, 136.02, 134.06, 128.54, 128.14, 127.82, 125.25, 123.37, 120.95, 109.26, 70.22. LC-MS: *t*R: 7.10 min (linear gradient 10-90% in 13.5 min), m/z = 312.2 $[M+H]^+$, 265.9 $[M-CO_2+H]^+$

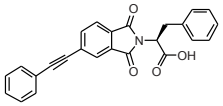


15: Prepared from **14** (516 mg, 2 mmol) and alanine (187 mg, 2.1 mmol) according to the general condensation procedure to yield to title compound as white solid (509 mg, 1.56 mmol, 78%). ^{1}H NMR (400 MHz, DMSO) δ 13.04 (s, 1H), 7.82 (d, *J* = 8.3, 1H), 7.51 – 7.45 (m, 3H), 7.44 – 7.38 (m, 3H), 7.38 – 7.32 (m, 1H), 5.32 (s, 2H), 4.83 (q, *J* = 7.3, 1H), 1.54 (d, *J* = 7.3, 3H). ^{13}C NMR (101

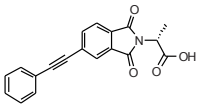
MHz, DMSO) δ 171.09, 166.80, 166.75, 163.56, 136.04, 133.94, 128.81, 128.54, 128.46, 128.42, 128.12, 127.78, 127.57, 125.17, 123.28, 120.96, 109.14, 70.19, 46.93, 14.87. LC-MS: *t*R: 7.56 min (linear gradient 10-90% in 13.5 min), m/z = 326.1 $[M+H]^+$, 280.0 $[M-CO_2+H]^+$



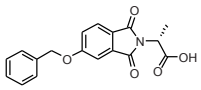
16b: Prepared from **14** (516 mg, 2 mmol) and β-alanine (187 mg, 2.1 mmol) according to the general condensation procedure to yield to title compound as white solid (500 mg, 1.54 mmol, 77%). ^{1}H NMR (400 MHz, DMSO) δ 12.35 (s, 1H), 7.77 (d, *J* = 8.3, 1H), 7.49 – 7.31 (m, 7H), 5.29 (s, 2H), 3.76 (t, *J* = 7.3, 2H), 2.59 (t, *J* = 7.3, 2H). ^{13}C NMR (101 MHz, DMSO) δ 172.09, 167.19, 163.32, 136.08, 134.20, 128.82, 128.53, 128.40, 128.11, 127.77, 127.55, 124.87, 123.59, 120.52, 108.94, 70.14, 33.58, 32.43. LC-MS: *t*R: 7.26 min (linear gradient 10-90% in 13.5 min), m/z = 326.0 $[M+H]^+$, 308.2 $[M-H_2O+H]^+$, 266.0 $[M-CH_2CO_2+H]^+$



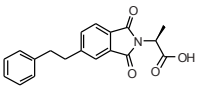
16e: Prepared from **10** (496 mg, 2 mmol) and phenylalanine (347 mg, 2.1 mmol) according to the general condensation procedure and purified by HPLC yielding the title compound as off white foam (639 mg, 1.6 mmol, 80%). ^{1}H NMR (400 MHz, DMSO) δ 13.30 (bs, 1H), 7.91 (d, *J* = 6.6, 2H), 7.83 (d, *J* = 8.0, 1H), 7.62 – 7.56 (m, 2H), 7.43 (d, *J* = 4.8, 3H), 7.19 – 7.14 (m, 4H), 7.14 – 7.07 (m, 1H), 5.16 (dd, *J* = 4.7, 11.5, 1H), 3.52 (dd, *J* = 4.7, 14.1, 1H), 3.43 – 3.31 (m, 1H). ^{13}C NMR (101 MHz, DMSO) δ 169.99, 166.43, 166.34, 137.43, 137.27, 131.70, 131.29, 129.80, 129.58, 128.81, 128.73, 128.33, 126.59, 125.77, 123.77, 121.38, 93.66, 87.77, 53.17, 33.98. LC-MS: *t*R: 9.42 min (linear gradient 10-90% in 13.5 min), m/z = 396.1 $[M+H]^+$, 350.0 $[M-CO_2+H]^+$



17c: Prepared from **10** (496 mg, 2 mmol) and D-alanine (187 mg, 2.1 mmol) according to the general condensation procedure and purified by recrystalization from hot water yielding the title compound as white solid (506 mg, 1.6 mmol, 80%). ^{1}H NMR (400 MHz, DMSO) δ 13.15 (s, 1H), 8.04 – 7.98 (m, 2H), 7.93 (d, *J* = 7.7, 1H), 7.66 – 7.60 (m, 2H), 7.50 – 7.45 (m, 3H), 4.89 (q, *J* = 7.2, 1H), 1.56 (d, *J* = 7.3, 3H). ^{13}C NMR (101 MHz, DMSO) δ 170.90, 166.46, 166.36, 137.30, 131.89, 131.71, 130.46, 129.59, 128.84, 128.53, 125.66, 123.69, 121.36, 93.44, 87.91, 47.13, 14.75. LC-MS: *t*R: 8.31 min (linear gradient 10-90% in 13.5 min), m/z = 320.0 $[M+H]^+$, 273.9 $[M-CO_2+H]^+$

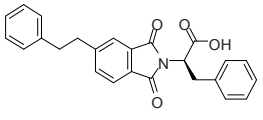


16c: Prepared **14** (496 mg, 2 mmol) and D-alanine (187 mg, 2.1 mmol) according to the general condensation procedure and purified by HPLC to yield the title compound as white solid (280,7 mg, 0.86 mmol, 43%). ^{1}H NMR (400 MHz, DMSO) δ 13.05 (bs, 1H), 7.82 (d, *J* = 8.3, 1H), 7.51 – 7.45 (m, 3H), 7.44 – 7.38 (m, 3H), 7.38 – 7.32 (m, 1H), 5.32 (s, 2H), 4.83 (q, *J* = 7.3, 1H), 1.53 (d, *J* = 7.3, 3H). ^{13}C NMR (101 MHz, DMSO) δ 171.06, 166.78, 166.73, 163.55, 136.03, 133.93, 128.53, 128.11, 127.77, 125.16, 123.26, 120.97, 109.13, 70.17, 46.90, 14.85. LC-MS: **16c** did not ionize, preventing acquisition of the LC-MS specrum

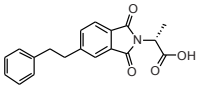


18b: Prepared from **17c** (133 mg, 0.35 mmol) according to the general hydrogenation procedure and purified by HPLC to yield the title compound (84 mg, 0.26 mmol, 74%) ^{1}H NMR (400 MHz, DMSO) δ 7.74 (d, *J* = 7.7, 2H), 7.63 (d, *J* = 8.5, 1H), 7.28 – 7.18 (m, 4H), 7.15 (t, *J* = 6.9, 1H), 4.85 (q, *J* = 7.2, 1H), 3.05 (t, *J* = 7.8, 2H), 2.91 (t, *J* = 7.8, 2H), 1.56 (d, *J* = 7.3, 3H). ^{13}C NMR

(101 MHz, DMSO) δ 171.24, 167.37, 167.23, 149.55, 140.87, 134.75, 131.84, 129.29, 128.52, 128.36, 126.09, 123.38, 123.26, 47.12, 37.19, 36.75, 14.93. LC-MS: *t*R: 7.97 min (linear gradient 10-90% in 13.5 min), m/z = 324.1 $[M+H]^+$, 278.1 $[M-CO_2+H]^+$



18c Prepared from **17e** (138 mg, 0.35 mmol) according to the general hydrogenation procedure to yield the title compound (141 mg, 0.35 mmol, quant). ^{1}H NMR (400 MHz, DMSO) δ 7.69 (d, *J* = 7.5, 2H), 7.60 (d, *J* = 8.2, 1H), 7.28 – 7.06 (m, 10H), 5.12 (dd, *J* = 4.8, 11.7, 1H), 3.50 (dd, *J* = 5.1, 14.9, 2H), 3.42 – 3.34 (m, 1H), 3.05 – 2.96 (m, 2H), 2.93 – 2.84 (m, 2H).^{13}C NMR (101 MHz, DMSO) δ 170.24, 167.28, 167.14, 149.66, 140.84, 137.41, 134.81, 131.21, 129.10, 128.74, 128.65, 128.44, 128.33, 128.29, 126.57, 126.02, 123.41, 123.27, 52.99, 37.08, 36.57, 34.06. LC-MS: *t*R: 9.00 min (linear gradient 10-90% in 13.5 min), m/z = 400.0 $[M+H]^+$,354.0 $[M-CO_2+H]^+$



18a Prepared from **17c** (99.5 mg, 0.31 mmol) according to the general hydrogenation procedure to yield the title compound (93 mg, 0.29 mmol, 94%).^{1}H NMR (400 MHz, DMSO) δ 13.14 (s, 1H), 7.95 – 7.52 (m, 3H), 7.39 – 7.10 (m, 5H), 4.86 (d, *J* = 4.8, 1H), 3.06 (s, 2H), 2.92 (s, 2H), 1.55 (d, *J* = 4.5, 3H). ^{13}C NMR (101 MHz, DMSO) δ 171.13, 167.25, 167.11, 149.46, 140.79, 134.67, 131.68, 129.13, 128.43, 128.27, 125.99, 123.31, 123.19, 46.99, 36.99, 36.58, 14.88. LC-MS: *t*R: 7.97 min (linear gradient 10-90% in 13.5 min), m/z = 324.1 $[M+H]^+$,278.1 $[M-CO_2+H]^+$

Supplementary Information

Rit1 Crystallization

The data were integrated and scaled with the PROTEUM*plus* Crystallographic Software .

Table S1: Data collection and integration statistics for Rit1. The statistics in parentheses are for the highest resolution shell.

X-ray source	Cu-Kα rotating anode			
Wavelength (Å)	1.54146			
Detector	Bruker Nonius CCD area detector			
Temperature (°C)	-174°C (99K)			
Resolution range (Å)	40.52 – 2.3 (2.4-2.3)			
Crystal to film distance (mm)	60			
Exposure time (s)	10			
Oscillation per image (°)	0.5			
Space group	$P2_1$			
Mosaic spread (°)	~0.36			
Unit cell parameters (Å)	a (Å)	40.278	α (°)	90.0
	b (Å)	37.048	β (°)	95.90
	c (Å)	40.736	γ (°)	90.0
Total reflections	82115			
Unique reflections	5464			
Multiplicity	14.6 (3.75)			
Completeness (%)	98.0 (89.6)			
R_{sym}	0.0547 (0.163)			
$I/\sigma(I)$	36.45 (8.25)			

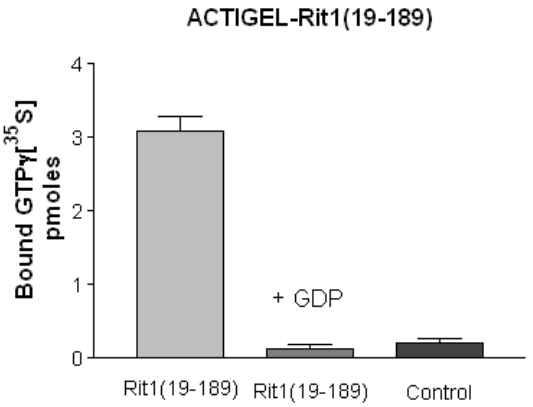


Figure S1. Nucleotide binding activity of the nucleotide-free form of immobilized Rit1 was determined using biochemical assay as described.

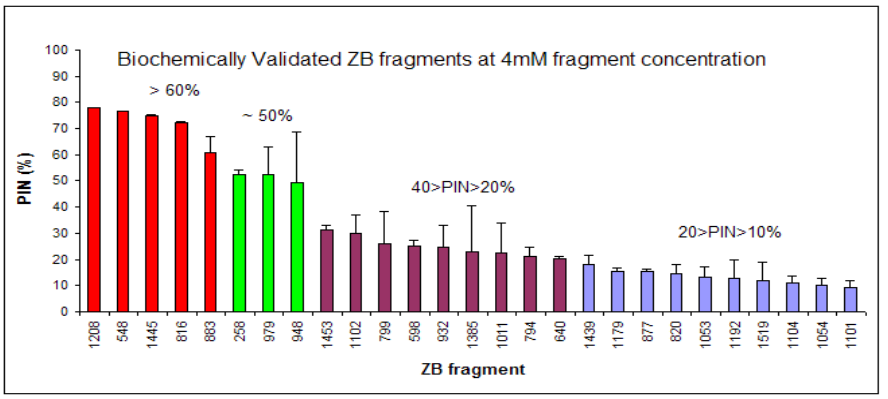


Figure S2. The ranking of primary fragment hits obtained from TINS NMR screening according to PIN % (percentage of inhibition) determined at high concentration in vitro biochemical assay.

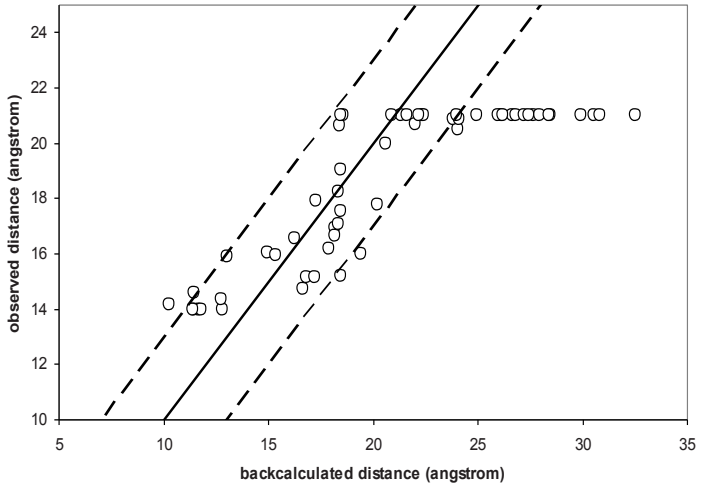


Figure S3. The experimentally observed nitroxide to amide proton distances (Å) are plotted against the back-calculated distances after optimizing the nitroxide position in XPLOR-NIH (see Methods). The solid line indicates the ideal correlation, dashed lines represent ± 3 Å error limits (Q =0.059). Amide protons of which the resonances were not affected by the nitroxide, were given a lower limit of 21 Å.

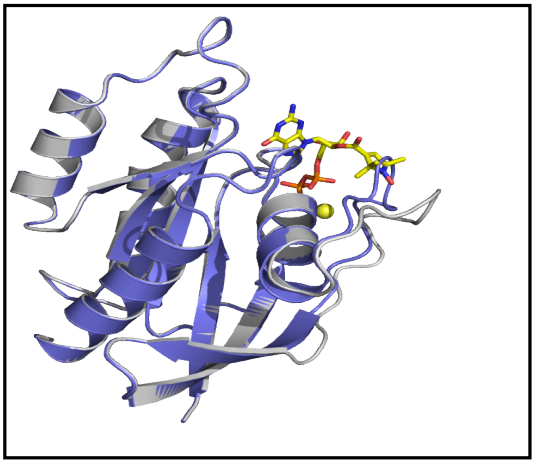


Figure S4. The loop consisting of residues 42-57 was optimized using PRE's observed from the nitroxide radical. In blue is the original crystal structure, yellow stick representing the position of the GDP-spin label, yellow sphere shows the magnesium ion and in the grey is the protein structure with the optimized loop position used for docking process.

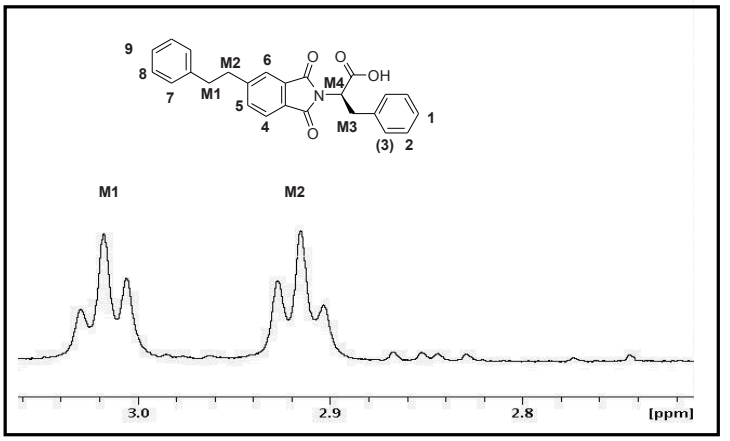


Figure S5. The two aliphatic resonances, M1 and M2 of **18c** are shown. Other resonances M3 and M4 could not be assigned due to overlap with the other buffer components.

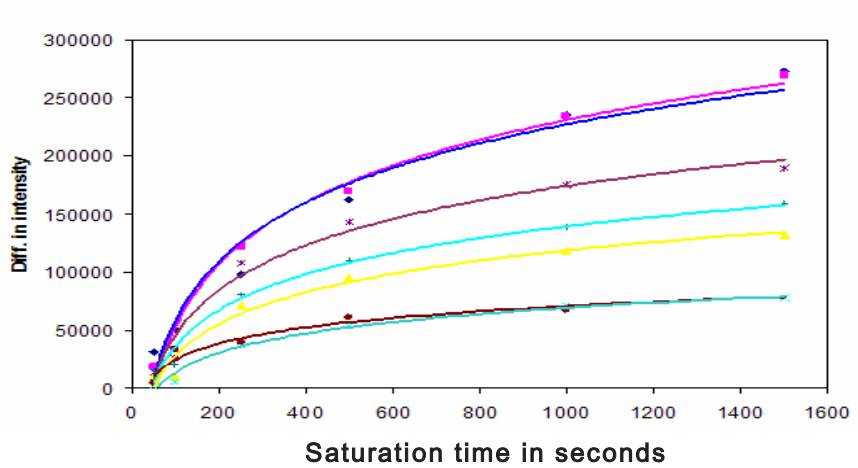


Figure S6. The saturation transfer experiments were performed as described in the experimental section. The plot of the intensity vs saturation time is depicted showing the saturation transfer rate for various protons of **18c**. The reduction in the intensity was caused by saturation transfer by irradiating at -0.5 ppm. The saturation time was varied from 50 ms to 2 sec.

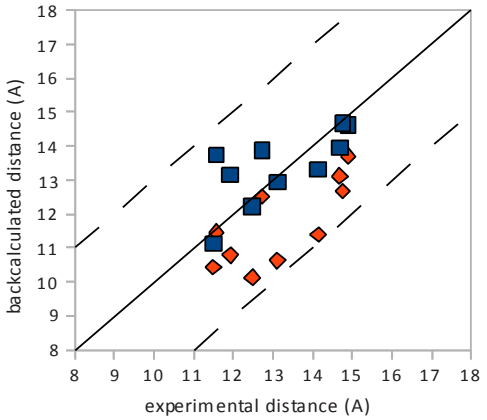


Figure S7. Comparison of back calculated and observed distances averaged over top 10 structures with error limits of ± 3A shown by dashed lines. Solid line represents the perfect correlation. The blue squares indicate the lowest docking solutions obtained by treating the intermolecular restraints as two separate sets. The red diamonds shows the distances obtained by simulation consisting of all inter-molecular PRE distance restraints in one set. As only two PRE restraints were available for the acidic moiety of **18c**, the force constant for PREs was increased from 3.0 to 90.0 to get a significant energy penalty whenever the PRE restraints were not fulfilled.

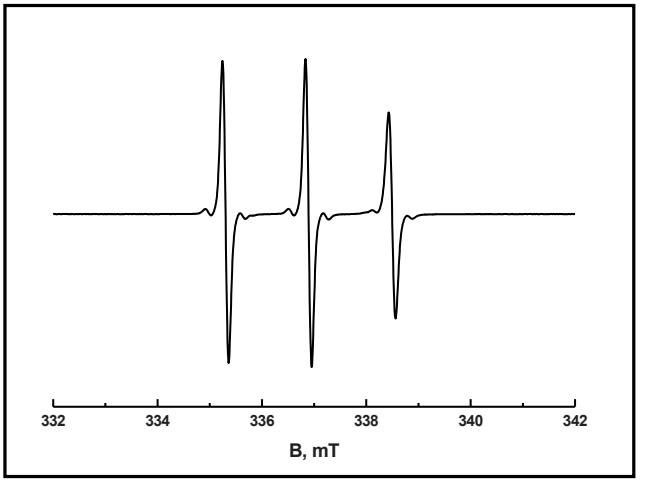


Figure S8. The X-band cw EPR measurements on GDP-spin label have been performed using an ELEXSYS E680 spectrometer (Bruker, Rheinstetten, Germany) equipped with a rectangular cavity. A modulation frequency of 100 kHz was used. Measurements were done at temperature of 20°C, using 6.31 mW of microwave power and modulation amplitude of 0.5 G.

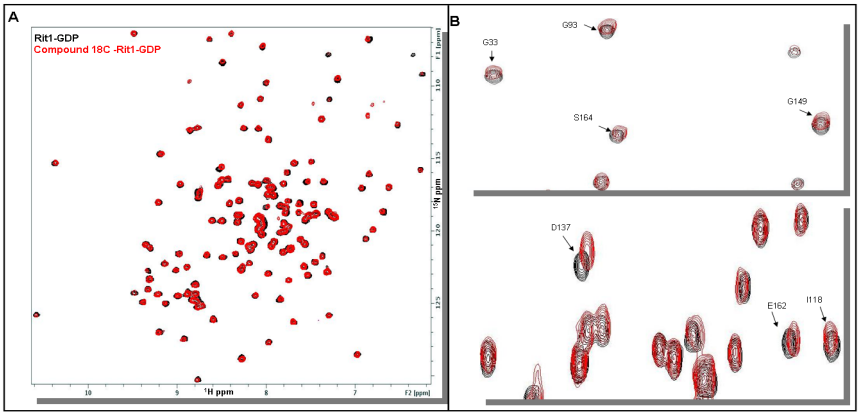


Figure S9. A] The overlay of the [^{1}H,^{15}N] HSQC spectra of Rit1-GDP (black) and in the presence of 1mM compound **18c** (red). B] The zoomed region of the overlaid [^{1}H, ^{15}N] HSQC NMR spectra shown in (A). The residues that show significant CSPs are labeled.

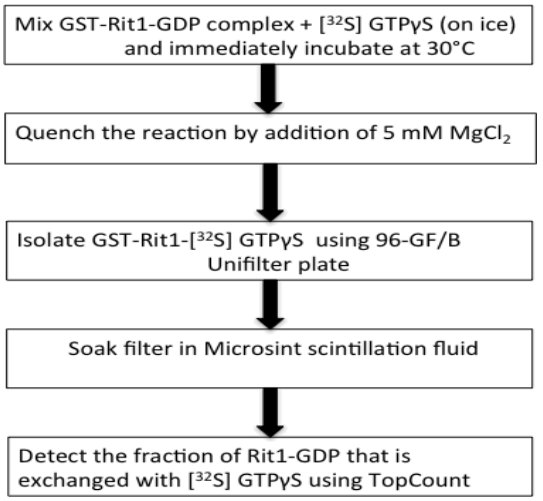


Figure S10. The diagrammatic representation of the nucleotide exchange assay used to determine IC_{50}'s of the inhibitory compounds.

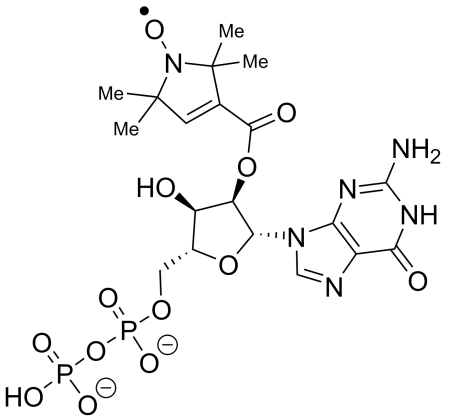


Figure S11. The depiction of the chemical structure of the GDP-spin label probe used in the paramagnetic NMR studies.

S12. Docking script used for the calculations in XPLOR-NIH:

```
!-----------------variables-------------------------------------
 eval ($a01 = 1)            ! cycle counter
 eval ($a02 = 20.0)         ! random distance max added
 eval ($a03 = 0.0)          ! prevents to throw two times
 eval ($a04 = 40.0)         ! structure writing threshold 27
 eval ($a05 = 10000)                ! nr cycles
 eval ($a06 = "cycle_1.pdb")        ! initial value saved structure
 eval ($a08 = 30.0)                 ! fbeta, frictinal constant
 eval ($a09 = 1.5)          ! velocity factor
 eval ($a10 = 180.0)                ! random rotation max +/-0.5*
 eval ($a11 = 1)            ! number of structures output per run
 eval ($a12 = $a11-1)               ! count saved structures
 eval ($a13 = $a04)         ! initial lowest energy filter
 eval ($a14 = 1)            ! initial value system energy
 eval ($a15 = 4000)         ! initial value system energy
 eval ($a16 = 4000)         ! initial value system energy
 eval ($a17 = 4000)         ! initial value system energy
 eval ($a18 = 1000)         ! initial value lowest energy from throw
 eval ($a19 = -1)           ! initial value count within throw
 eval ($a20 = 1)            ! count throws from minimum
 eval ($a21 = $cpu*1)       ! starting point random numbers
 eval ($a22 = 3.0)          ! force constant PRE
 eval ($a23 = 5.0)          ! force constant CSP
```

```
 eval ($a24 = 10.0)          ! force constant STD
 eval ($a25 = 4000)          ! initial value system energy
 eval ($a26 = 10)            ! Etot
 eval ($a27 = 10)            ! Evdw
 eval ($a28 = 10)            ! Enoe

!----------------ligand structure-------------------------------
set echo=off end
set mess=off end

 topology
   @/afs/gorlaeus.net/users/p/pkeizers/xplor/dipen/ligand/lig.top
 end

 param
   @/afs/gorlaeus.net/users/p/pkeizers/xplor/dipen/ligand/lig.par
 end

 segment
   name="lig"
   molecule number=1 name="lig" end
 end

 vector do (resi="500") (resn = "lig")
 vector do (segid="    ") (resn = "lig")
 coor @/afs/gorlaeus.net/users/p/pkeizers/xplor/dipen/ligand/lig.pdb
 vector do (segid="lig") (resi="500")

 flags exclude * include bonds angles impropers dihedrals end

 parameters
   learn initiate sele=(segid="lig") mode=nostatistics end
   learn accumulate end
   learn terminate end
 end

 parameters
   BOND   (segid "lig") (segid "lig") 400. TOKEN
   ANGLE  (segid "lig") (segid "lig") (segid "lig") 60. TOKEN
   IMPR  (segid "lig") (segid "lig") (segid "lig") (segid "lig") 50. TOKEN TOKEN
 end

 flags include vdw elec pvdw pele end
```

```
 hbuild
    ACCEptor=(all)
    phistep=10.0
 end

shake
  molecule (segid "lig")
end


!-----------------protein structure----------------------------
set echo=off end
set mess=off end

structure
  @/afs/gorlaeus.net/users/p/pkeizers/xplor/dipen/sl_struc/dip_sl.psf
end

param
  @/afs/gorlaeus.net/users/p/pkeizers/xplor/dipen/sl_struc/dip_sl.par
end

 if ($a01=1) then
  coor @/afs/gorlaeus.net/users/p/pkeizers/xplor/dipen/sl_struc/dip_sl.pdb
 end if

 while ($a01 LE $a05) loop calc
 eval ($a21=$cpu*1)
 set seed=$a21 end

   if ($a14 GE $a15) then
    if ($a14 GE $a16) then
     if ($a14 GE $a17) then
      if ($a14 GE $a25) then
       eval ($a20 = $a20 + 1) end
        if ($a01 > $a03) then
         eval ($a03 = $a01 + 1.0)
          coor @/afs/gorlaeus.net/users/p/pkeizers/xplor/dipen/sl_struc/dip_sl.pdb
          vector show element (X) (NAME CAW)
            eval ($UX=($result))
            vector show element (Y) (NAME CAW)
            eval ($UY=($result))
            vector show element (Z) (NAME CAW)
```

```
                eval ($UZ=($result))
                coor trans sele= (segid "lig") vect =(-$UX -$UY -$UZ) end
                !eval ($a20 = $a20 + 1) end
                vector do (fbeta = $a08) (segid="lig")
                vector do (vx = $a09) (segid="lig")
                vector do (vy = $a09) (segid="lig")
                vector do (vz = $a09) (segid="lig")
                eval ($dist1=(40 + (RANDOM()*$a02)))
                eval ($ang1=(0.5 - RANDOM())*$a10)
                eval ($v1=(0.5 - RANDOM())*20)
                eval ($v2=(0.5 - RANDOM())*20)
                eval ($v3=(0.5 - RANDOM())*20)
                coor trans SELE=(segid="lig") VECT= ($v1 $v2 $v3) DIST = $dist1
end
                coor rota SELE=(segid="lig") cent= (0 0 0) axis ($v3 $v1 $v2) $ang1
end
            end if
           end if
          end if
         end if
        end if

       parameter
         nbonds
           inhi = 0.75
           ctofnb = 7.5 !5.0 default settings no sign. effect Etot
           ctonnb = 6.5 !4.0 default settings no sign. effect Etot
           repe1 = 0.8 !0.9 has severe effects Etot
           rexp = 2
           irex = 2
           rcon = 0.4 !0.1 may do the trick, adapt threshold too 0.3
         end
       end

      !-----------------DEFINE RESTRAINTS-----------------------------------

      !----------------- PRE restraints---------------------------------------
      if ($a01=1) then
        noe reset end
         NOE
           nrestraints=100000
         end
```

```
NOE
  class CL1
end

NOE
  assign (segid="lig" and name 1HAA) (segid= "dip" and name OBI) 13.11
3.0 3.0
  assign (segid="lig" and name 1HAD) (segid= "dip" and name OBI) 12.48
3.0 3.0
  assign (segid="lig" and name 1HAB) (segid= "dip" and name OBI) 11.92
3.0 3.0
  assign (segid="lig" and name 1HBA) (segid= "dip" and name OBI) 14.14
3.0 3.0
  assign (segid="lig" and name 1HBD) (segid= "dip" and name OBI) 14.76
3.0 3.0
  assign (segid="lig" and name 1HAZ) (segid= "dip" and name OBI) 14.68
3.0 3.0
 assign (segid="lig" and name 1HZ)  (segid= "dip" and name OBI) 14.89 3.0
3.0
 assign (segid="lig" and name 1HE2) (segid= "dip" and name OBI) 12.73 3.0
3.0
  assign (segid="lig" and name 1HAW) (segid= "dip" and name OBI) 11.56
3.0 3.0
  assign (segid="lig" and name 1HAX) (segid= "dip" and name OBI) 11.50
3.0 3.0
end

NOE
  averaging CL1 R-6
  potential CL1 square
  sqconstant CL1 0.005
  sqoffset CL1 0.0
  scale CL1 $a22
  sqexponent CL1 2
end

!--------------------CSP restraints----------------------------
NOE
   class CL2
end

NOE
   assign (atom "lig" * *H*) (resid 137 and name HN)  6.0 3.0 3.0
```

```
    assign (atom "lig" * *H*) (resid 167 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 170 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 162 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 136 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 173 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 169 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 168 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 164 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 163 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 139 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 116 and name HN)  6.0 3.0 3.0
    assign (atom "lig" * *H*) (resid 113 and name HN)  6.0 3.0 3.0
  end

  NOE
    averaging CL2 R-6
    potential CL2 square
    sqconstant CL2 0.005
    sqoffset CL2 0.0
    scale CL2 $a23
    sqexponent CL2 2
  end

!--------------------STD restraints----------------------------
  NOE
   class CL3
  end

  NOE
   assign (atom "lig" * *H*) (atom "dip" * C*)  5.0 3.0 4.0
  end

  NOE
   averaging CL3 R-6
   potential CL3 square
   sqconstant CL3 0.005
   sqoffset CL3 0.0
   scale CL3 $a24
   sqexponent CL3 2
  end
end if

!---------------------dock ligand--------------------------------
```

```
constraints
 interactions (segid="dip" or segid="lig")
 (segid="lig") end
constraints
 FIX (not (segid="lig")) end

flag exclude * include bond angle dihe impro noe vdw end

minimize powell
 nstep = 200
 nprint = 200
 drop = 10
end

constraints
 interactions ((segid="dip" and (name CA or name C or name O or name N
or name
 HN or name CB or name CG or resi 186)) or segid="lig")
 (segid="lig") end
constraints
 FIX (not (segid="lig")) end

flag exclude * include noe bond angle impro dihe vdw end

vector do ( fbeta = 6.657235 ) ( segid="lig" )
dynamics langevin
 timestep = 0.001
 nstep = 2500
 ilbfrq = 2500
 nprint = 2500
 tbath = 300.
 iasvel = maxwell
 rbuf = 0.0
 origin = ( 0. 0. 0. )
end

!------------optimize sidechains-------------------------------
constraints
 interactions (segid="dip" or segid="lig")
 (segid="lig" or segid="dip") end
constraints
 FIX (segid="dip" and (name CA or name C or name O or name N or name
```

```
  HN or name MG or name CM or resi 186)) end
 flag exclude * include noe bond angle impro dihe vdw end

 vector do ( fbeta = 6.657235 ) ( segid="lig" )
  dynamics langevin
   timestep = 0.001
   nstep = 1500
   ilbfrq = 1500
   nprint = 1500
   tbath = 300.
   iasvel = maxwell
   rbuf = 0.0
   origin = ( 0. 0. 0. )
  end

!------------calculate energies and write structures-------------
 flag exclude * include noe vdw end
 energy end

 set disp=ener.dat end
  if ($a01=1) then
        display file: ener.dat          $DATE $TIME
        display
        display cycle Etot    Evwd  Enoe  structure       number         time
  end if
  eval ($a25 = $a17)
  eval ($a17 = $a16)
  eval ($a16 = $a15)
  eval ($a15 = $a14)
  eval ($a14 = $ENER)
  eval ($a07 = ($a20 - $a19))
   if ($a07 = 1) then
    coor swap end
    write coord output = $a06 end
    display $a01      $a26   $a27   $a28   $a06   $a11   $TIME
    set disp=OUTPUT end

    eval ($a12 = $a11)
    eval ($a11 = $a11 +1)
    coor swap end
   end if
    if ($a14 LE $a04) then
     if ($a07 LE 0) then
```

```
      if ($a14 LE $a18) then
       coor copy end
       eval ($a26 = $ener)
         eval ($a27 = $vdw)
         eval ($a28 = $noe)
         eval ($a18 = $a14)
       eval ($a12 = $a12 + 1)
        if ($a14 LE $a13) then
         eval ($a13 = $a14)
        end if
       eval ($a19 = $a20)
      end if
     end if
      if ($a07 > 0) then
         coor copy end
       eval ($a26 = $ener)
         eval ($a27 = $vdw)
         eval ($a28 = $noe)
         eval ($a06 = "cycle_"+encode($a11)+".pdb")
        if ($a14 LE $a13) then
         eval ($a13 = $a14)
        end if
       eval ($a18 = $a14)
       eval ($a12 = 0)
         eval ($a19 = $a20)
      end if
    end if
  eval ($a01 = $a01 + 1)
 end loop calc

!-----------------------------------------------------------------
!-----------------------FINISH-----------------------------------
!-----------------------------------------------------------------
stop
```

Contributions

Target identification, validation and production of protein for crystallography, nucleotide exchange assays and NMR screening were performed by Masakazu Kobayashi (Galapagos NV). NMR screening was performed at ZoBio BV. Johan Hollander (ZoBio BV) helped to set-up the NMR screen. Eiso AB and Caroline Loch (ZoBio BV) performed the computational analysis to generate fragments hits. Francis Figaroa (ZoBio BV) assisted with the compound preparations for TINS NMR screen. Anna A. Rzepiela (Pyxis discovery) performed the computational studies for analogs. Crystallography studies on the target and target-ligand complexes were performed by F. Hoedemaker and Linda Manning (KeyDP BV). Synthesis of compounds at various stages was done by Adriaan W.Tuin (Leiden University). Peter H.J. Keizers (Leiden University) performed the computational docking for paramagnetic restraints.

REFERENCES

(1) Quan, L., Thiele, G., Tian, J. and Wang, D. The development of novel therapies for Rheumatoid Arthritis. *Expert Opin Ther Pat*, **2008**, 18(7), 723–738.

(2) Methods for Identification, and Compounds useful; for the treatment of degenerative andnflammatory diseases, Ser. No. 11/52,366, filed Jun.14, **2005** Galapagos B.V. US patent 7,919,259 B2.

(3) Colicelli, J. Human RAS Superfamily Proteins and Related GTPases *Sci. STKE*, **2004**, 250, 13.

(4) Wennerberg, K., Rossman. L.K., Channing, J D. The Ras superfamily at a glance *J. Cell. Sci.* **2005,** 118, 843-846**.**

(5) Gorfe, A. A. Mechanisms of allostery and membrane attachment in Ras GTPases: implications for anti-cancer drug discovery**.** *Curr Med Chem*, **2010**, 171-9**.**

(6) Hall, B., Bar-Sagi, D., Nassar, N. The structural basis for the transition from Ras-GTP to Ras-GDPPNAS, *Proc Natl Acad Sci*. **2002**, 99(19), 12138-12142.

(7) Kinbara, K., Goldfinger,LE., Hansen, M., Chou, F and Ginsberg, M. Ras GTPases: integrins' friends or foes *Nature Rev. Mol Cell Biol.* **2003**, 4, 767-778.

(8) Walker, K., Olson, M. Targeting Ras and Rho GTPases as opportunities for cancer therapeutics. Curr. Opinion in Gen. & Development **2005**, 15(1), 62-68.(9) Bell, I.M., Gallicchio, S.N., Abrams, M., Beese, L., Beshore, D., Bhimnathwala, H., Bogusky, M., Buser, C., Culberson, J., Davide, J. 3-Aminopyrrolidinone farnesyl transferase inhibitors: Design of macrocyclic compounds with improved pharmacokinetics and excellent cell potency *J Med Chem*, **2002**,45, 2388–2409.

(10) Hunt, J., Ding, C., Batorsky, R., Bednarz, M., Bhide, R., Cho, Y., Chong, S., Gullo-Brown, J., Guo P. Discovery of (R)-7-cyano-2,3,4, 5-tetrahydro-1-(1H-imidazol-4-ylmethyl)-3- (phenylmethyl)-4-(2-thienylsulfonyl)-1H–1,4-benzodiazepine (BMS-214662), a farnesyl transferase inhibitor with potent preclinical antitumor activity. *J Med Chem*, **2000**, 43, 3587–3595.

(11) Zhang, FL and Casey, PJ., Protein prenylation: molecular mechanisms and functional consequences. *Annu Rev Biochem*, **1996**, 65, 241–269.

(12) Konstantinopoulos, P., Karamouzis, M., Papavassiliou, A. Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets. *Nat Rev Drug Disc,* **2007**, 6, 541-555.

(13) Hynds, D., Spencer, ML., Andres, D., Snow, D. Rit promotes MEK-independent neurite branching in human neuroblastoma cells. *J Cell Sci*, **2004**, 116 (10), 1925–35.

(14) Shao, H., and Andres, D. A novel RalGEF-like protein, RGL3, as a candidate effector for rit and Ras. *J Biol Chem*, **2000**, 275 , 26914–26924.

(15) Andres, D., Rudolph, J., Sengoku, T., and Shi, GX. Analysis of Rit Signaling and Biological Activity. *Meth Enzymol*, **2006**,407, 499-512.

(16) Rusyn, E., Reynolds, E., Shao, H., Grana, T., Chan, T., Andres, D., and Cox, A. Rit, a non-lipid-modified Ras-related protein, transforms NIH3T3 cells without activating the ERK, JNK, p38 MAPK or PI3K/Akt pathways. *Oncogene*, **2000**,19, 4685–4694.

(17) Hajduk,P.J., Greer, J. A decade of fragment-based drug design: Strategic advances and lesson learned. *Nat Rev Drug Disc*, **2007**, 6, 211-219.

(18) Scheffzek, K., Ahmadian, M., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F and Wittinghofer, A. The Ras-RasGAP Complex: Structural Basis for GTPase Activation and Its Loss in Oncogenic Ras. *Mut Sci*, **1997**, 277, 333-339.

(19) Karassek, S., Berghaus, C., Schwarten, M., Emans, C., Ohse, N., Kock, G., Jockers, K., Neumann, S., Gottfried, S., Herrmann, C., Heumann, R., and Stoll, R. Ras homolog enriched in brain (Rheb) enhances apoptotic signaling. *J Biol Chem*, **2010**, 285 (44) 33979-33991.

(20) Vanwetswinkel, S., Heetebrij, RJ., van Duynhoven, J., Hollander JG., Filippov DV., Hajduk PJ., Siegal, G.TINS: Target Immobilized NMR Screening – An efficient and sensitive new method for ligand discovery. *Chem Biol*, **2005**, 12, 207-216.

(21) Marquardsen, T., Hofmann, M., Hollander, J., Loch, C., Kiihne, S., Engelke, F., and Siegal, G. Development of a dual cell, flow-injection sample holder and NMR probe for comparative ligand-binding studies. *J Mag Reson*, **2006**, 182, 55-65.

(22) John, J., Sohmen, R., Feuerstein, J., Linke, R., Wittinghofer, A., Goody, R. Kinetics of interaction of nucleotides with nucleotide-free H-ras p21.*Biochem*, **1990**, 29(25), 6058–6065.

(23) Siegal, G., AB, Eiso and Schultz, J. Integration of Fragment Screening and Library Design, *Drug Disc Today*, **2007**, 12,1032-1039.

(24) Taveras, AG., Remiszewski, SW., Doll, RJ., Cesarz, D., Huang, EC., Kirschmeier, P., Pramanik, BN., Snow, ME., Wang, YS., del Rosario, JD., Vibulbhan, B., Bauer, BB., Brown, JE., Carr, D., Catino, J., Evans, CA., Girijavallabhan, V., Heimark, L., James, L., Liberles, S., Nash, C., Perkins, L., Senior, MM., Tsarbopoulos, A., Webber, SE. *Bioorg Med* Chem,**1997**, 5(1), 125-133.

(25) Ganguly, AK., Wang, YS., Pramanik, BN., Doll, RJ., Snow, ME., Taveras, AG., Remiszewski, S., Cesarz, D., del Rosario, J., Vibulbhan, B., Brown, JE., Kirschmeier, P., Huang, EC., Heimark, L., Tsarbopoulos, A., Girijavallabhan, VM., Aust, RM., Brown, EL., DeLisle, DM., Fuhrman, SA., Hendrickson, TF., Kissinger, CR., Love, RA., Sisson, WA., Webber, SE. Interaction of a novel GDP exchange inhibitor with the Ras protein. *Biochem*, **1998**, 37(45):15631-7.

(26) Peri, F., Airoldi, C., Colombo, S., Mari, S., Jiménez-Barbero, J., Martegani, E., and Nicotra F. Sugar-derived Ras inhibitors: group epitope mapping by NMR spectroscopy and biological evaluation. *Euro J Org Chem*, **2006**, 3707-3720.

(27) Hajduk, PJ., Huth, JR., and Fesik, SW. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem*, **2005**,48, 2518-2525

(28) McCoy, MA., Wyss, DF. Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J Am Chem Soc*, **2002** 124, 11758–11763.

(29) The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC

(30) Olsen, J., De Godoy, L., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A.,
Lange, O., Horning, S., Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteo*, **2005**, 4, 2010–2021.

(31) Zav'yalov, S., Dorofeeva, O., Rumyantseva, E., Kulikova, L., Ezhova, G. Synthesis of N-Phthaloyl Derivatives of Amino Acids, *Phar Chem Jour,* **2002**, 36, 440 – 442

(32) Nakamura, T., Noguchi, T., Kobayashi, H., Miyachi, H., and Hashimoto, Y. Mono- and dihydroxylated metabolites of thalidomide: synthesis and TNF-alpha production-inhibitory activity. *Pharm. Bull*, **2006**. 54, 1709—1714

(33) Bodenhausen, G., and Ruben, D. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett*, **1980**, 69 (1), 185-189.

(34) Sklenar, V., Piotto, M., Leppik, R., and Saudektim, V. Gradient-Tailored Water Suppression for ^{1}H - ^{15}N HSQC Experiments Optimized to Retain Full Sensitivity. *J Magn Reson*, **1993**, (A) 102, 241 -245.

(35) Bashir, Q., Volkov, A., Ullmann, GM., and Ubbink, M. Visualization of the encounter ensemble of the transient electron transfer complex of cytochrome *c* and cytochrome *c* peroxidase *J Am Chem Soc*, **2010,** 132**,** 241-247.

(36) Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson*, **2003**, 160, 65– 73.

(37) Mayer, M., and Meyer, B. Group Epitope Mapping (GEM) by STD NMR to identify segments of a ligand in direct contact with a protein receptor, *J Am Chem Soc*, **2001**, 123, 6108- 6117.

(38) Mayer., M and Meyer, B. A fast and sensitive method to characterize ligand binding by saturation transfer difference NMR spectra, *Angew Chem*, **1999**, 111, 1902-1906.

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

(40) Wüthrich K. NMR of Proteins and Nucleic Acids. New York: Wiley and Sons, **1986**.

(41) Muraoka, S., Shima, F., Araki, M., Inoue, T., Yoshimoto, A., Ijiri, Y., Seki, N., Tamura, A., Kumasaka, T., Yamamoto, M, Kataoka, T. Crystal structures of the state 1 conformations of the GTP-bound H-Ras protein and its oncogenic G12V and Q61L mutants. *FEBS Lett*, **2012**, 586(12), 1715-8.

(42) Maurer, T., Garrenton, LS., Oh, A., Pitts, K., Anderson, DJ., Skelton, NJ., Fauber, BP., Pan, B., Malek, S., Stokoe, D., Ludlam, MJ., Bowman, KK., Wu, J., Giannetti, AM., Starovasnik, MA., Mellman, I., Jackson, PK., Rudolph, J., Wang, W., Fang, G. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proc Natl Acad Sci*, **2012,**109(14), 5299-304.

(43) Shao, H., Kadono-Okuda, K., Finlin, B. S. and Andres, D. A**.**. Biochemical characterization of the Ras-related GTPases Rit and Rin. *Arch Biochem Biophys*, **1999,** 371*,* 207 -219.

(44) The CCP4 Suite: Programs for Protein Crystallography. Collaborative Computational Project, Number 4. *Acta Cryst*, **1994**, D50, 760-763.

(45) Emsley, P., Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr*, **2004**, D60, 2126-2132.

(46) Kobayashi, M., Retra, K., Figaroa, F., Hollander, JG., Ab, E. Heetebrij, RJ., Irth, H., Siegal, G. Target immobilization as a strategy for NMR based fragment screening. Comparison if TINS, STD and SPR for fragment hit identification. *J Biomol Screen*, **2010**, 15(8), 978-989.

(47) Murali, N., Miller, WM., John, BK., Avizonis, DA. Smallcombe SH spectral unraveling by space-selective Hadamard spectroscopy*. J Magn Reson,* **2006***,*179*,*182-189*.*

(48) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Delaglio, F., Grzesiek, S., Vuister, GW., Zhu, G., Pfeifer, J., and Bax, A. *J Biomol NMR*, **1995**, 6(3), 277-93.

(49) Goddard, TD., and Kneller, DG. SPARKY 3, University of California, San Francisco.

(50) Jung, YS., and Zweckstetter, M. Mars-robust automatic backbone assignment of proteins. *J Biomol NMR*, **2004**, 30(1), 11—23.

(51) Lin, B., Covalle, KL., Maddock, JR.The Caulobacter crescentus CgtA protein displays unusual guanine nucleotide binding and exchange properties, *J Bacteriol*, **1999**,181(18), 5825–5832.

(52) Lenzen, C., Cool, RH., Wittinghofer, A. Analysis of intrinsic and CDC25-stimulated guanine nucleotide exchange of p21ras-nucleotide complexes by fluorescence measurements. *Methods Enzymol*. **1995**;255:95-109

(53) Streckenbach, B., Schwarz, D., and Kurt, R.H. Repke Analysis of phosphoryl transfer mechanism and catalytic centre geometries of transport ATPase by means of spin-labelled ATP *Biochimic et Biophysic Acta Biomem*, **1980**, 601, 34-46.

(54) Ferguson, K., Higashijima, T., Smigel, MD., and Gilman, A. The influence of bound GDP on the kinetics of Guanine Nucleotide binding to G proteins. *J Biol Chem*, **1986**, 261(16), 7393-7399.

(55) Keizers, P., Saragliadis, A., Hiruma, Y., Overhand, M., and Ubbink, M. Design, synthesis, and evaluation of a lanthanide chelating protein probe: CLaNP-5 yields predictable paramagnetic effects independent of environment, *J Am Chem Soc*, **2008**,130 (44),14802–14812.

(56) Schüttelkopf, AW., and van Aalten, D. PRODRG - a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallogr,* **2004** D60, 1355–1363.

Chapter 4

Rapid protein ligand costructures from sparse NOE data

*Dipen M. Shah,[‡,†] Eiso AB,[†] Tammo Diercks,[φ] Mathias A. S. Hass,[‡] Nico A. J. van Nuland,[§] and Gregg Siegal[‡,†]**

[‡]Leiden Institute of Chemistry, Leiden University, Leiden 2300RA, the Netherlands.

[†] ZoBio BV, Leiden 2300RA, the Netherlands.

[φ] Centre for Co-operative Research in Biosciences, Derio 48160, Spain.

[§] Jean Jeener NMR Centre, Structural Biology, Vrije Universiteit Brussel and Molecular Recognition Unit, VIB Department of Structural Biology, Brussels 1050, Belgium.

This chapter was published as a research article:
Shah, D.M.; Ab, E.; Diercks, T.; Hass, M.A.S.; van Nuland, N.A.J.; Siegal, G.
Journal of Medicinal Chemistry, Volume 55, Number 23, pp. 10786-10790 (2012)

ABSTRACT

An efficient way to rapidly generate protein-ligand co-structures based on solution-NMR using sparse NOE data combined with selective isotope labeling is presented. A docked model of the 27 kDa N-terminal ATPase domain of Hsp90 bound to a small molecule ligand was generated using only 21 intermolecular NOEs which uniquely defined both the binding site and the orientation of the ligand. The approach can prove valuable for the early stages of fragment based drug discovery.

INTRODUCTION

The availability of 3D structural information on protein-ligand complexes has become an important driver to guide the pre-clinical stages of drug discovery. Structure based drug design (SBDD) has enabled the development of a wide array of drugs that are currently on the market such as inhibitors of the HIV protease and kinases.[1,2] In the past 15 years fragment-based drug discovery (FBDD), which uses very small, soluble drug fragments as starting points to develop new medicines, has become widespread. Due to their small size, fragments typically bind the target with low affinity (K_D > 10 μM). The development of weak fragment hits to more potent lead-like structures is 2-3 times more successful when 3D structural information is available.[3] Thus the success of both SBDD and FBDD is heavily dependent upon the availability of structural information. Presently both approaches are primarily driven by X-ray crystallography. Crystallography has the advantage that when successful, it is both rapid and high resolution. However, there are a number of cases in which crystallography is not successful such as when crystals are not available, the crystal packing precludes access to the active site or most commonly, weakly binding fragments simply do not give rise to electron density. In the latter case there can be many causes including insufficient occupancy of the binding site or multiple possible binding orientations. In principle, NMR-based solution methods can also generate atomic resolution structural information and NMR has indeed successfully supported FBDD campaigns.[4] Since NMR is extremely sensitive to weak protein-ligand interactions, it should be applicable exactly where crystallography is least effective, i.e. for complexes of weakly binding ligands. However, traditional NMR approaches involving uniform isotopic labeling are labor intensive and limited to proteins of moderate size (e.g. < 30 kDa) and have therefore not been widely adopted in drug

discovery.

In drug discovery it is often the case that the 3D structure of a target or a homologous protein is available. If the resonance assignment of such a protein is available, mapping of chemical shift perturbations (CSPs) onto the 3D structure is a simple, fast procedure that can provide low resolution information about a small molecule binding site.[5] CSP mapping is most commonly accomplished through analysis of ^{15}N- or ^{13}C-edited heteronuclear correlation spectra of the protein. However, the interpretation of ^{1}H-^{15}N or ^{1}H-^{13}C HSQC CSPs becomes ambiguous when the chemical shift perturbations are caused by changes in protein dynamics or a shift in equilibrium between two (or more) conformations. Various computational methods primarily based on chemical shift perturbation analysis are most successful at determining the location of the binding site to low resolution. The precise nature of the intermolecular contacts are well beyond the capabilities of structural analysis based on CSPs.[6,7] In principle a limited set of intermolecular NOEs could provide sufficient information to determine the orientation and binding mode of a small molecule bound to a protein. Indeed, a number of approaches that use a combination of amino acid selective labeling and intermolecular NOEs have been proposed.[8-10] While these are quite powerful, they suffer from one or more of the following limitations: (i) requirement for *a priori* knowledge of the ligand binding site in order to be able to efficiently select the right combination of residues to label, (ii) requirement for a large number of intermolecular NOE contacts, (iii) the resonance assignment uses the pattern of chemical shift perturbations induced by ligand binding and therefore may not be robust and (iv) requirement for extensive calculations to generate structures that match the experimental data. We sought an NMR based method capable of providing structures of sufficient resolution and reliability to

support elaboration of “weakly” binding fragments to target proteins that met the following criteria: widely applicable to small and large proteins, rapid (2-3 weeks per complex), standard solution conditions and NMR experiments, requires samples that can be easily made and the data interpretation is unambiguous. Here, we demonstrate such a method based on selective ILV methyl-labeling and a sparse set of intermolecular NOEs.

RESULTS AND DISCUSSION

The approach as outlined in Figure 1, requires an NMR sample of the protein in which the target protein is highly deuterated and where NMR visible, isotopic labels have been introduced along the backbone and selectively at ILV sidechains as described by Tugarinov and Kay.[11] Standard, through-bond double and triple-resonance NMR spectra are used to obtain the backbone and ILV sidechain resonance assignments of the protein. The ligand is then titrated into the protein in small increments such that the shift in both backbone and sidechain resonance positions can be readily followed using 2D heteronuclear correlation spectra. Once binding of the compound to the protein is saturated, ^{13}C- and ^{15}N-edited NOESY-HSQC spectra are recorded to generate a set of intermolecular NOEs between the ligand resonances and ILV methyl groups as well as the backbone amide protons of the protein. A known 3D structure or structures of the protein target (or a homology model) is then used for molecular docking using HADDOCK to obtain the protein-ligand structure based solely on the intermolecular NOE restraints.[12] Given that the isotopic labeling method has been used to obtain assignments for large proteins, it is likely that the proposed scheme can also be used to determine experimentally derived molecular models based on sparse NOE data for large protein-small

molecule complexes in an efficient yet reliable manner.[11,13-15] Here the approach is illustrated using the moderately sized 27 kDa N-terminal domain of the molecular chaperone heat shock protein 90 (Hsp90), a known cancer target.

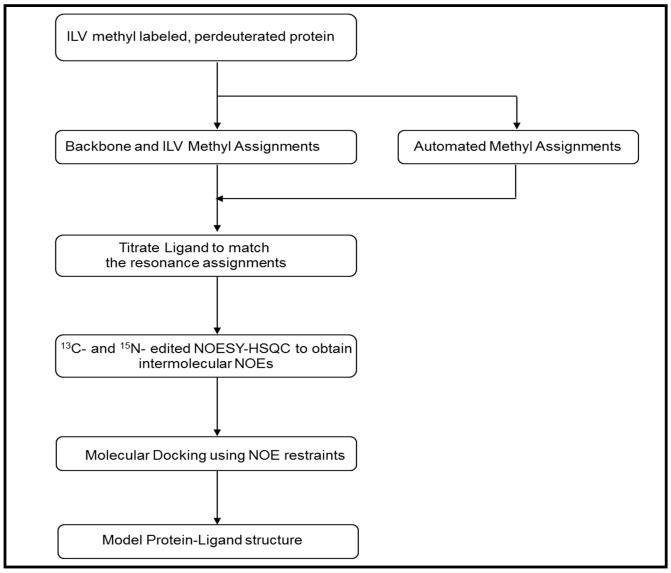


Figure 1. A flowchart scheme to illustrate the number of steps involved in the determination of the model protein-ligand complex by solution NMR based on the sparse NOE data using ILV methyl labeling approach.

The N-terminal domain of Hsp90 (9-233) was isotopically labeled in *E. coli* according to the published procedure.[11] Heteronuclear correlation experiments yielded highly resolved spectra with excellent sensitivity in which all expected methyl peaks can be observed (Figure 2). The triazine **1** had been discovered as a ligand of Hsp90 in a target immobilized NMR screen (TINS) of a fragment library constructed from commercially available compounds.[16,17] The binding of **1** to Hsp90 was confirmed using surface plasmon resonance technology and the equilibrium K_D was determined to be 58 μM *via* fitting to a 1:1 binding model (data not shown). We investigated the effect of **1** binding to the protein by recording a high-resolution CT-[^{1}H,^{13}C] HSQC spectrum in the presence and absence of the compound (Figure

2).[18,19] The methyl chemical shift perturbations are clear evidence of the binding of **1** to Hsp90. The peak pattern in the bound state is suggestive of a folded protein indicating that the binding is specific and does not result in protein denaturation or other undesirable affects. Surprisingly, the majority of the methyl resonances are affected by the binding of **1** and a similar pattern was observed in the [^{1}H, ^{15}N]-HSQC (not shown). Using the sequential assignment determined below, the methyl and backbone amide chemical shift perturbations (CSPs) have been mapped onto the crystal structure of Hsp90 (Figure 3). Although the CSPs surround the known ATP binding site, residues far from the site are also significantly affected. The widespread changes in the spectrum of Hsp90 could be indicative of conformational changes in the protein induced or stabilized by the binding of compound **1.** Indeed, conformational changes have been observed for a number of ligands binding to Hsp90.[20] Thus it can be difficult to determine the ligand binding site simply on the basis of CSPs (see below). Furthermore, the orientation of the ligand within the binding site, and therefore the nature of the protein-ligand interactions, is not defined by the CSP information.

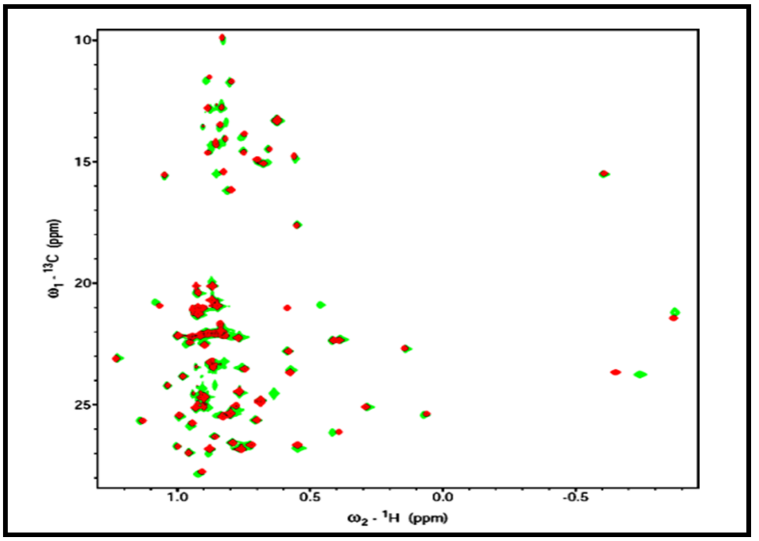


Figure 2. CT- [^{1}H,^{13}C] HSQC spectra of an ILV methyl protonated sample of deuterated N-terminal domain of Hsp90 recorded in the absence (in red) and in the presence of **1** (in green) at a protein to ligand ratio of 1:6. Significant chemical shift perturbations for various methyl groups are seen clearly in the presence of **1**, indicative of binding.

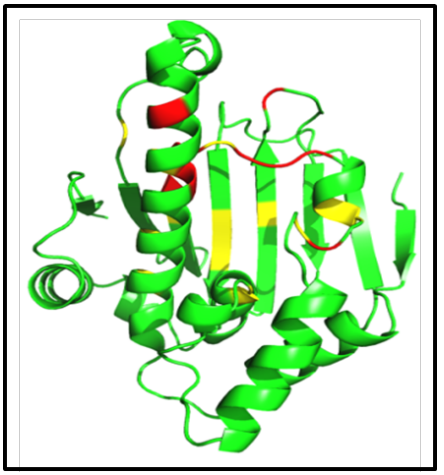


Figure 3. The methyl (yellow) and amide (red) chemical shift perturbations caused by the presence of **1** are mapped on the crystal structure of Hsp90 (PDB 1YER). The figure was created in PyMOL.[21]

The sequential backbone assignment of Hsp90 is available (e.g. BMRB 5355) but was confirmed by analysis of TROSY-based HNCACB and HNcoCACB NMR spectra. In total, 76% of the HN, N, Cα and Cβ chemical shifts were sequentially

assigned, including 87% of the ILV residues. The methyl resonances were subsequently correlated with the intraresidue Cα and Cβ assignments by a CCH-TOCSY experiment. Of the ILV residues that had been sequentially assigned, 92% of the ILV methyl assignments were obtained in a straightforward manner. It is not essential to obtain all ILV methyl assignments since only those within the binding site will give rise to intermolecular NOEs. Based on the crystal structure of a complex of **1** with Hsp90 (see below), all methyl groups that are within NOE distance (8 Å using a highly deuterated protein) to the ligand have been assigned using only data from these three experiments. Stereospecific methyl assignments for Leu and Val residues were obtained by producing a 10% ^{13}C labeled sample and CT-HSQC analysis as described previously.[22]

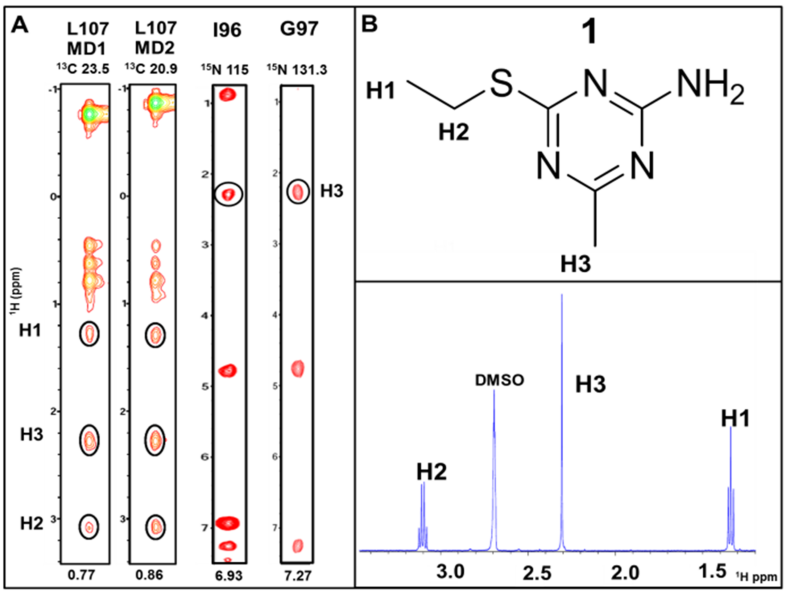


Figure 4. A] Representative strips from the 3D ^{13}C- and ^{15}N-edited NOESY-HSQC spectra of methyl protonated {I(13CH_3, δ1 only), L($^{13}CH_3$,$^{12}CD_3$), V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] Hsp90 in the presence of **1**. The intermolecular NOEs between the methyl groups of the protein (L107MD1, L107MD2) and the ligand H1, H2, H3 groups are circled. The strips for residues I96 and G97 from a ^{15}N-edited NOESY-HSQC spectrum are shown from which intermolecular NOEs to the H3 group of **1** are circled. The frequencies (ppm) of ^{1}H and ^{13}C or ^{15}N nuclei are shown at the bottom and the top of the strips respectively. B] The structure (top) and 1D ^{1}H spectrum (bottom) of **1** in D_2O with the resonance assignment.

3D ^{13}C-edited and ^{15}N-edited NOESY-HSQC spectra were recorded on the protein-ligand complex in order to detect intermolecular NOEs.[23,24] Due to the selective labeling scheme employed, there was minimal overlap amongst the protein resonances and therefore intermolecular NOEs were unambiguously assigned. Selected strips from the ^{13}C-edited NOESY-HSQC and ^{15}N-edited NOESY-HSQC spectra exhibiting intermolecular NOEs from the ligand are shown in Figure 4A. A total of 21 intermolecular NOEs were identified between resonances of **1** and protein residues including three to amide protons (I96, G97 and M98) and 18 NOEs to methyl groups (L107, L103, V150, V186 and I96). The peak intensities from the intermolecular NOEs were converted to distances and used as restraints to carry out the docking of the ligand using HADDOCK (See experimental section for more information).[12]

Typical of many pharmaceutical targets, a crystal structure of apo-Hsp90 (PDB 1YER) is available in the PDB. This structure was the starting point for the experimentally restrained docking procedure. We performed HADDOCK calculations using only the apo-Hsp90 structure and 21 intermolecular-NOE distances as unambiguous restraints (Table 1).[12] The lowest energy cluster exhibited a HADDOCK score and minimal restraint violation energy of -7.0 and 3.8 kcal/mol respectively, whereas the next ligand cluster exhibited a HADDOCK score of 3.4 and minimal restraint violation energy of 20.0 kcal/mol. The difference between the clusters indicates that the input data defines a single set of structures.

Table 1. Intermolecular NOE distance restraints included into HADDOCK calculations are shown. Note that a lower and upper distance margin of 1-2 Å were employed to above target distances.

Ligand atom	Protein residue	Protein residue atom	Target distance (Å)
H1	I96	HD1	3.8
H1	L103	HD+	4
H1	L107	HD1	4.3
H1	L107	HD2	4
H1	V150	HG1	3.8
H1	V150	HG2	3.8
H1	V186	HG2	4.3
H2	I96	HD1	4
H2	L103	HD+	5
H2	L107	HD1	4.6
H2	L107	HD2	4
H2	V150	HG1	5.3
H2	V186	HG2	5.3
H3	L103	HD+	4.9
H3	L107	HD1	3.9
H3	L107	HD2	3.6
H3	V150	HG1	4.8
H3	L186	HG2	5
H3	I96	HN	4
H3	G97	HN	4
H3	M98	HN	4

A crystal structure of the complex of **1** with Hsp90 was available in the PDB (3B24).[25] In 3B24 there are two protein molecules in the asymmetric unit and although **1** is bound in the same site in each, it is rotated by approximately 180° in the different structures (Figure 5A). Superposition of the protein in the three structures (NMR model and two crystal structures) indicates that the sparse NOE method defines the binding site of the ligand uniquely and accurately. Most Hsp90 ligands form critical hydrogen bonds to residues D93 and T184 and these are present in both the X-ray structures and the NMR model. The orientation of **1** in the NMR based model is quite similar to one of the orientations in the asymmetric unit of the crystal structure (1.7 Å rmsd). In the NMR model, the orientation of **1** is defined by a network of 3 intermolecular NOEs from the ligand H3 group to the amide protons of residues I96,

G97 and M98 (Figure 5B) and is thus unambiguous. Interestingly, PDB entries 2WI2 and 2WI3, are structures of a similar triazine bound to Hsp90 in which two orientations differing by a 180° axial rotation are found. In this case the different orientation was dependent on whether soaking or co-crystallization was used to form the complex.[26] Taken together the crystal structures suggest that there are two or more low energy conformations for the complex of these simple triazines with Hsp90. In contrast, the NMR data indicate that in solution, there is one predominant conformation and this is likely to be the most physiologically relevant. Importantly, despite the moderate number of intermolecular restraints, the NMR model defines precisely the same binding site as the crystal structure and the critical intermolecular hydrogen bonds suggesting that in addition to being fast, the method is robust.

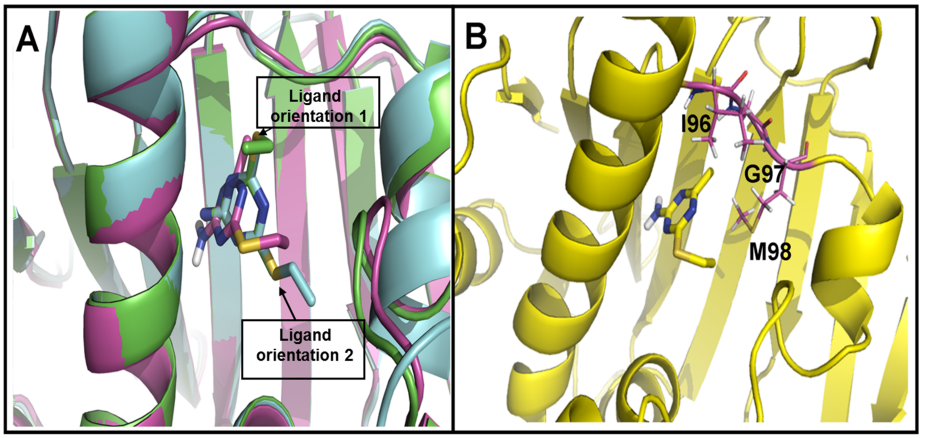


Figure 5. A] Overlay of the lowest energy HADDOCK model of **1** bound to Hsp90 with a crystal structure of the complex (PDB 3B24) in which **1** binds in two orientations (docked model- magenta, 3B24: green and sky blue). The orientation of **1** in the NOE-based model is similar to one of the binding modes (seen in sky blue color) of the ligand in PDB entry 3B24. B] The orientation of **1** in the docked model is largely determined by the three intermolecular NOEs observed between the H3 group of the ligand and HN of I96, G97 and M98 shown in pink within the protein. The figure was created in PyMOL.[21]

Numerous groups have attempted to use CSPs to determine the structure of protein-ligand complexes.[7,27] We used the ability of HADDOCK to include CSPs as ambiguous restraints for docking and modeled the structure of the Hsp90-**1** complex

using the most significant methyl and backbone amide CSPs (Figure 6). The HADDOCK calculations generated 2 clusters whose HADDOCK scores were similar and significantly lower than others. Interestingly, the clusters are both located at the binding site of **1** defined in both the NOE model and the crystal structures. However, despite the fact that electrostatics were used during the HADDOCK calculation, the hydrogen bonds to D93 and T184 are not present. In fact, the amine group of **1** is pointing away from these residues.

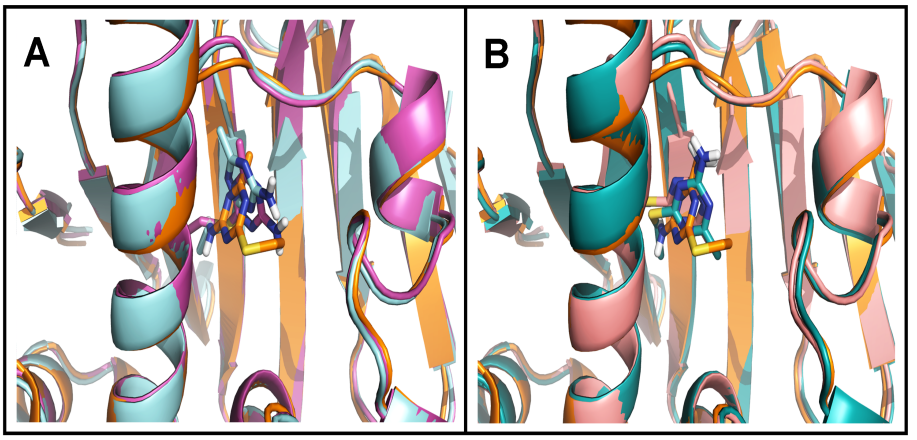


Figure 6.Overlay of the lowest energy HADDOCK model structures of **1** bound to Hsp90 obtained using the most significant methyl and backbone amide chemical shift perturbations ($\Delta\delta_i > \Delta\delta_{avg}$ + 2 x standard deviation). Two clusters of **1** are observed which have similar HADDOCK score. A] Overlay of the two structures from one of the clusters (magenta and sky blue) with NOE-derived docked model shown in orange. B] Overlay of the two structures from the other cluster (pink and dark green) with NOE-derived docked model shown in orange. The figure was created in PyMOL.[21]

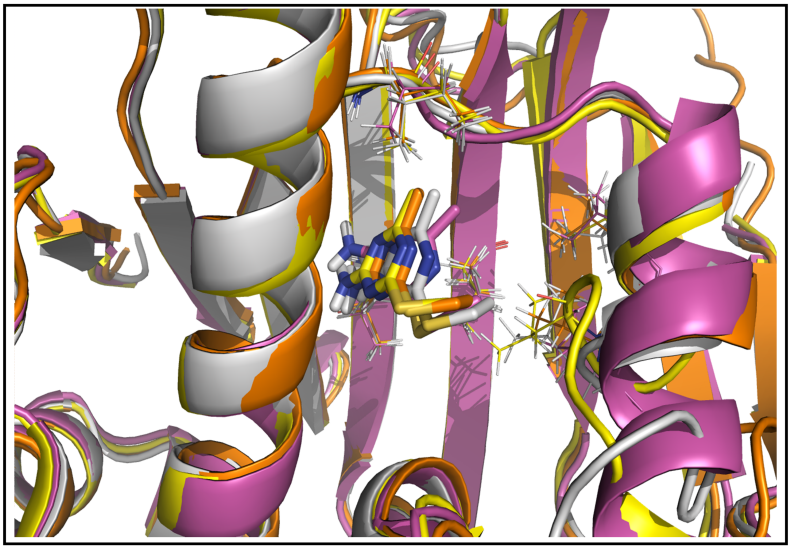


Figure 7. Overlays of the lowest energy HADDOCK model structures are shown when **1** is docked with various N-term Hsp90 structures with different sidechain conformations (shown as lines) in the PDB (1YER- yellow, 1OSF-orange, 2QF6-magenta and 3B26-grey). The ligand is found to bind in the same binding spot in comparison with the X-ray structure, irrespective of the conformational flexibility of the helix formed by residues 100-124 seen in the various PDB structures. The figure was created in PyMOL.[21]

Frequently ligand binding is accompanied by a conformational change in the protein. Conformational rearrangement of residues 100-124, which are in close proximity to the ligand binding site of Hsp90, has been observed for a number of ligands including in the two structures in the unit cell of 3B24 (Figure 5A). To investigate any possible influence of conformational rearrangement on the structures calculated from NMR data, we selected 3 different Hsp90 structures which differ in the conformation of residues 100-124 for docking using HADDOCK.[12] The binding site and orientation of the ligand cluster in each of the 3 different Hsp90 structures was very similar to that determined using the apo-protein (Figure 7). Importantly, the intermolecular hydrogen bonds to D93 and T184 are conserved in all 4 experimentally constrained docking structures. This data suggests that, at least for the case of Hsp90, binding induced conformational changes do not preclude

determining the essential features of a small molecule-protein complex even if those changes are unknown.

CONCLUSION

The method presented here appears to be capable of generating reliable protein-ligand structures in a quick and efficient manner. We have shown that a universal selective labeling scheme can be used to rapidly identify sufficient numbers of restraints for a small molecule ligand weakly bound to protein. Moreover, universal labeling precludes the necessity for *a priori* knowledge of the ligand binding site. However, the present method depends on the availability of selectively labeled, deuterated protein which can only be produced in *E. coli*. Note that the 3 NOEs involving backbone amide protons of the protein were critical for orienting **1** in the protein-ligand complex. This observation suggests that, at least in some cases, the use of exclusively methyl labeled protein could result in the loss of important constraints for the calculation procedure. Another obvious limitation is the requirement for methyl groups at the ligand binding site. Although ILV residues tend to be well located in proteins and clearly Hsp90 is a good example of this, it is not always the case.[13] However, it is also possible to selectively label all methyl containing residues providing even more complete coverage of protein structures and ligand binding sites.[28] While we have used standard, through-bond NMR techniques for resonance assignment here, recently steps towards automating the assignment of methyl resonances based either on intramolecular NOEs and/or through-space paramagnetic effects have been taken.[29] Implementation of selective methyl labeling in conjunction with automated resonance assignment should enable structure determination of complexes involving reasonably large proteins up to 75 kDa. We feel the present method could prove valuable for the early stages of FBDD by

providing 3D structure information on weakly binding fragment-protein complexes.

Experimental Section:

Sample Preparation

A methyl protonated {I($^{13}CH_3$, δ1 only), L($^{13}CH_3$,$^{12}CD_3$), V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] sample of N terminal domain (9-233) of Hsp90 was obtained by protein overexpression from a culture of *E. coli* BL21 (DE3) cells transformed with the plasmid pQTEV as described.[11] The protein was expressed in 1 L D_2O M9 medium supplemented with 2 g/L of U-[^{13}C, ^{2}H]-glucose (Sigma) as the main source of carbon and 0.3 g/L of $^{15}NH_4Cl$ (Sigma) as the nitrogen source. One hour prior to induction at ~0.5 OD, 50 mg of 2-keto-3,3-d^2-1,2,3,4-^{13}C-butyrate and 100 mg of 2-keto-3-methyl-d^3-3-d^1-1,2,3,4-^{13}C butyrate (^{13}C labeled α-ketoisovalerate deuterated at the β-position and with one of the methyl groups - $^{12}CD3$) were added to growth medium. The expression of the protein was induced by the addition of 1 mM IPTG for 16 hours at 18°C. The protein was purified on Ni-Hitrap FF (GE Lifescience) followed by purification on a superdex G75 gel filtration column (GE Lifescience). The seven-histidine tag was removed by proteolytic digestion with TEV protease (Invitrogen) according to the manufacturer recommended buffer condition. The cleaved tag and the TEV protease were removed by affinity purification on Ni-Hitrap FF, and subsequently the buffer of Hsp90 (9-233) was exchanged to 20 mM sodium phosphate pH 7.5, 50 mM NaCl. The NMR sample was 0.7 mM in protein, 95% H_20, 5% D_20, 20 mM sodium phosphate, pH 7.5, 50 mM NaCl. The NMR sample on the Hsp90-**1** complex was obtained by addition of 6 equivalents of **1** to the protein. The pH of the sample was adjusted carefully within +/- 0.05 units of 7.5 after addition of the ligand. Stereospecific methyl assignments were obtained by producing a 10% ^{13}C

labeled sample and CT-HSQC analysis as described. [22]

Compound **1** was purchased from Maybridge, catalog number RF 02707.

NMR backbone sequential assignments and ILV methyl assignments

In order to detect binding of **1** to Hsp90, a CT- [^{1}H,^{13}C] HSQC spectrum of 0.7 mM protein and protein plus ligand (1:6 ratio) was recorded on a Varian 800 MHz spectrometer equipped with a cryogenic probe. The chemical shifts of **1** in the bound state were also followed by acquiring 1D ^{1}H proton spectrum in the presence of protein at different protein plus ligand ratios (1:0, 1:2, 1:4 and 1:6).

In order to perform sequential assignment of the backbone resonances and correlate them to the intraresidue ILV methyl resonances of the Hsp90-**1** complex, the following experiments were acquired at 298K on 600MHz four-channel Bruker DMX NMR spectrometer equipped with a TXI cryo-probe; [^{1}H,^{15}N]-HSQC, trHNCACB recorded with (2048x80x110) in (^{1}HN,^{15}N,^{13}C) dimensions, trHNcoCACB recorded with (2048x80x180) in (^{1}HN,^{15}N,^{13}C) dimensions with a recycle delay of 1s and with 16 scans/FID.[30,31] To correlate the backbone Cα and Cβ frequencies to the intraresidue methyl resonances of ILV residues, a CCH-TOCSY experiment was recorded with (2048x80x152) in (^{1}H$_{m}$,^{13}C$_{m}$,^{13}C) dimensions with a recycle delay of 1s and with 8 scans/FID. The CCH-TOCSY pulse sequence was obtained by modification of a standard (H)CCH-TOCSY pulse sequence by removal of the initial INEPT pulses, addition of deuterium decoupling during the *t1* and *t2* period, starting with a direct carbon excitation pulse, followed by a DIPSI-3 TOCSY transfer sequence and final detection on the methyl protons. The optimal mixing time for the TOCSY sequence was found to be 23.3 ms.[32]

A series of 2D ^{1}H-^{1}H NOESY planes with mixing time between 100-600 ms were acquired in order to determine the optimal mixing time based on the decay of cross

peak intensity. The optimal mixing time was defined to be a balance of being able to detect long range NOEs and spin diffusion and was found to be 400 ms. In order to observe intermolecular NOE cross peaks between the ligand and the protein, a ^{13}C-edited NOESY-HSQC (1024x40x160 points in ^{1}H,^{13}C,^{1}H respectively) and a ^{15}N-edited NOESY-HSQC (1024x44x160 points in ^{1}H,^{15}N,^{1}H respectively) were recorded with , mixing time of 400 ms and a recycle delay of 1s and with 16 scans/FID respectively.[33-36] The NOESY experiments were acquired at 298K on a 900 MHz four-channel Bruker DMX spectrometer equipped with pulsed field gradient, conventional room-temperature triple resonance probe. The cross-peak volumes/peak intensities from the intermolecular NOEs were converted to distances based on the equation $V=C/b^{n}$ (V= cross-peak volume/peak intensity, C is a constant, b is the upper distance bound and *n* is an exponent 4).[37] The acquired NMR data was processed using the NMRpipe/NMRDraw software package and visualized in Sparky.[38,39] The backbone assignment process was guided by the predicted chemical shifts calculated by SHIFTX in the automatic assignment program MARS using PDB 1YER as input.[40,41] The use of ILV methyl labeling is attractive for a number of reasons including: (i) the favorable relaxation properties of methyl groups so that NMR spectra, even for larger systems, are of higher sensitivity and resolution,[14,42] (ii) the well-dispersed distribution of the methyl groups throughout the protein structure and particularly at ligand binding sites,[43] (iii) restriction of the NMR analysis to the backbone and sidechain resonances of ILV methyl groups of the protein simplifies the process and allows unambiguous identification of intermolecular NOEs and (iv) methods to produce ILV labeled samples in *E. coli* are robust and economical.[11,44,45]

Protein-Ligand Docking

Throughout the docking procedure using HADDOCK, 1000 protein structures were used for initial rigid body docking. Explicit solvent refinement was performed on 200 structures after initial docking iterations. The final docking solutions were selected by clustering of the structures and analysis of the HADDOCK scores. The Hsp90 structures used in HADDOCK calculations to investigate the conformational rearrangement of the helix formed by residues 100-124 are derived from a number of different complexes for which the ligand was removed prior to docking to compound **1**.[13]

Contributions

Tammo Diercks (CiCBiogune, Bilbao), Mathias A. S. Hass (Leiden University) and Nico A. J. van Nuland (Vrije Universiteit, Brussels) helped to set-up NMR experiments at the respective NMR facilities. Eiso AB (ZoBio BV, Leiden) assisted with the NMR data analysis and molecular docking. We also thank Dr. Hans Wienk for his assistance and the SONNMRLSF in Utrecht. We also acknowledge the contribution to Hsp90 crystallography from Thomas A. Ceska at UCB, UK.

REFERENCES

(1) Volarath, P.; Harrison, R.W.; Weber, I.T. Structure based drug design for HIV protease: from molecular modeling to cheminformatics. *Curr. Top. Med. Chem.* **2007**, 7(10), 1030-1038.

(2) Hardy, L. W.; Malikayil, A. The impact of structure-guided drug design on clinical agents. *Curr. Drug Discov.* **2003**, 3, 15–20.

(3) Hajduk, P.J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **2007**, 6, 211-219.

(4) Oltersdorf, T.; Elmore, S.W.; Shoemaker, A.R.; Armstrong, R.C.; Augeri, D.J.; Belli, B.A.; Bruncko, M.; Deckwerth, T.L.; Dinges, J.; Hajduk., P.J; Joseph, M.K.; Kitada, S.; Korsmeyer, S.J.; Kunzer, A.R.; Letai, A.; Li, C.; Mitten, M.J.; Nettesheim, D.G.; Ng, S.; Nimmer, P.M.; O'Connor, J.M.; Oleksijew, A.; Petros, A.M.; Reed, J.C.; Shen, W.; Tahir, S.K.; Thompson, C.B.; Tomaselli, K.J.; Wang, B.; Wendt, M.D.; Zhang, H.; Fesik, S.W.; Rosenberg, S.H. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*, **2005**. 435, 677-681.

(5) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science*. **1996**, 274, 1531-1534.

(6) Krishnamoorthy, J.; Yu, VCK.; Mok, Y-K. Auto-FACE: An NMR based binding site mapping program for fast chemical exchange protein-ligand systems. *PLoS ONE.* **2010,** 5(2), e8943.

(7) Schieborr, U.; Vogtherr, M.; Elshorst, B.; Betz, M.; Grimme, S.; Pescatore, B.; Langer, T.; Saxena, K.; Schwalbe, H. How much NMR data is required to determine a protein-ligand complex structure? *Chembiochem.* **2005**, 6 (10), 1891-1898.

(8) Pellecchia, M.; Meininger, D.; Dong, Q.; Chang, E.; Jack, R.; Sem, D.S. NMR-based structural characterization of large protein-ligand interactions. *J. Biomol. NMR.* **2002**, 22, 165-173.

(9) Constantine, K.L.; Davis M.E.; Metzler, W.; Mueller, L.; Claus, B.L. Protein-ligand NOE matching: a high-throughput method for binding pose evaluation that does not require protein NMR resonance assignments. *J. Am. Chem. Soc.* **2006**, 128, 7252-7263.

(10) Yu, L.; Sun, C.; Song, D.; Shen, J.; Xu, N.; Gunasekera, A.; Hajduk, P.J.; Olejniczak, E.T. Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex. *Biochemistry.* **2005**, 44**,** 15834–15841.

(11) Tugarinov, V.; Kay, L. E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc*., **2003**, 125, 13868-13878.

(12) De Vries, S.J.; Van Dijk, M.; Bonvin, A.M.J.J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **2010**, 5, 883-897.

(13) Sprangers, R.; Kay, L.E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*. **2007**, 445, 618–622.

(14) Sibille, N.; Hanoulle, X.; Bonachera, F.; Verdegem, D.; Landrieu, I.; Wieruszeski, J.M.; Lippens, G. Selective backbone labelling of ILV methyl labelled proteins. *J. Bio. Mol.* **2009**, 43 (4), 219-227.

(15) Tugarinov, V.; Choy, W.Y.; Orekhov, V.Y; Kay, L.E. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci.* **2005**,102, 622–627.
(16) Vanwetswinkel, S.; Heetebrij, R.J.; Van Duynhoven, J.; Hollander, J.G.; Filippov, D.V.; Hajduk, P.J; Siegal, G. TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery. *Chem. Biol.,* **2005**, 12, 207-216.
(17) Siegal, G.; AB, E.; Schultz, J. Integration of fragment screening and library design. *Drug Discov. Today*, **2007**, 12, 1032-1039.
(18) Palmer, A.G. III; Cavanagh, J.; Wright, P.E.; Rance, M. Sensitivity improvement correlation NMR-spectroscopy. *J. Magn. Reson.* **1991**, 93, 151-170.
(19) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S.J.; Sorensen, O.W.; Griesinger, C. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR.* **1994**, 4, 301-306.
(20) Murray, C. W.; Carr, M.G.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S.; Coyle, J.E.; Downham, R.; Figueroa, E.; Frederickson, M.; Graham, B.; McMenamin, R.; O'Brien, A.M.; Patel, S.; Phillips, T.R.; Williams, G.; Woodhead, A.J.; Woolford, A.J.-A. Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. *J. Med. Chem*. **2010**, 53(16), 5942-5955.
(21) The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
(22) Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wuthrich, K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry.* **1989**, 28, 7510-7516.
(23) Kay, L.E.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc*. **1992**, 114, 10663-10665.
(24) Davis, A.L.; Keeler, J.; Laue, E.D.; Moskau, D. Experiments for recording pureabsorption heteronuclear correlation spectra using pulsed field gradients. *J. Magn. Reson.* **1992**, 98, 207-216.
(25) Miura, T.; Fukami, T.; Hasegawa, K.; Ono, N.; Suda, A.; Shindo, H.; Yoon, D.O.; Kim, S.J.; Na, Y.J.; Aoki, Y.; Shimma, N.; Tsukuda, T.; Shiratori, Y. Lead generation of heat shock protein 90 inhibitors by a combination of fragment-based approach, virtual screening, and structure-based drug design. *Bioorg. Med. Chem. Lett.*, **2011**, 21, 5778-5783.
(26) Brough, P.A.; Barril, X.; Borgognoni, J.; Chene, P.; Nicholas G. M. Davies, Davis, B.; Drysdale, M.J.; Dymock, B.; Eccles, S.A.; Garcia-Echeverria, C..; Fromont, C.; Hayes, A.; Hubbard, R.E.; Jordan, A.M.; Jensen, R.M.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F.; Robertson, A.; Roughley, S.; Schoepfer, R.; Simmonite, H.; Sharp, S.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L. Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *J. Med. Chem*. **2009**, 52(15), 4794-4809.
(27) Stark, J.; Powers, R. Rapid protein-ligand costructures using chemical shift perturbations. *J. Am. Chem. Soc.* **2008**, 130, 535-545.

(28) Otten, R.; Chu, B.; Krewulak, K. D.; Vogel, H. J.; Mulder, F. A. A. A comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. *J. Am. Chem. Soc*. **2010**, 132(9), 2952 – 2960.
(29) Venditti, V.; Fawzi, N.L.; Clore, G.M. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. *J. Biomol. NMR.* **2011**, 51, 319-328.
(30) Salzmann, M.; Wider, G.; Pervushin, K.; Senn, H.; Wuethrich,K. TROSY-type triple- resonance experiments for sequential NMR assignments of large proteins. *J. Am. Chem. Soc.* **1999**, 121, 844-848.
(31) Eletsky, A.; Kienhoefer, A.; Pervushin,K. TROSY NMR with partially deuterated proteins. *J. Biomol. NMR.* **2001**, 20, 188-180.
(32) Kay, L.E.; Xu, G.Y.; Singer, A.U.; Muhandiram, D.R.; Forman-Kay, A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H2O samples of proteins J.D. *J. Magn. Reson.* **1993**, B 101, 333 – 337.
(33) Davis, A.L.; Keeler, J.; Laue, E.D.; Moskau, D. Experiments for recording pureabsorption heteronuclear correlation spectra using pulsed field gradients. *J. Magn. Reson.* **1992**, 98, 207-216.
(34) Palmer, A.G. III; Cavanagh, J.; Wright, P.E.; Rance, M. Sensitivity improvement correlation NMR-spectroscopy. *J. Magn.Reson.* 1991, 93, 151-170.
(35) Kay, L.E.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc*. **1992**, 114, 10663-10665.
(36) Güntert, P.; Braun, W., Wüthrich, K. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **1991**, 217, 517-530.
(37) Delaglio, F.; Grzesiek, S.; Vuister, G.W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol NMR*. **1995**, 6(3), 277-93.
(38) Goddard, T.D.; Kneller, D.G. SPARKY 3, University of California, San Francisco.
(39) Neal, S.; Nip, A.; Zhang, H.; Wishart. D. Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. *J. Biomol. NMR*, **2003**, 26, 215-240.
(40) Jung, Y.S.; Zweckstetter, M. Mars -- robust automatic backbone assignment of proteins. *J. Biomol.NMR.* **2004**, 30(1), 11-23.
(41) Karagöz, G.E.; Duarte, A.M.; Ippel, H.; Uetrecht, C.; Sinnige, T.; van Rosmalen, M.; Hausmann, J.; Heck, A.J.; Boelens, R.; Rüdiger, S.G. N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. *Proc. Natl. Acad. Sci*. **2011**, 108(2), 580-585.
(42) Hajduk, P.; Augeri, D.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S.; Fesik, S. NMR-based screening of proteins containing 13C-labeled methyl groups. *J. Am. Chem. Soc.* **2000**, 122, 7898–7904.
(43) Sibille, N.; Hanoulle, X.; Bonachera, F.; Verdegem, D.; Landrieu, I.; Wieruszeski, J.M.; Lippens, G. Selective backbone labelling of ILV methyl labelled proteins. *J. Bio. Mol.* **2009**, 43 (4), 219-227.

(44) Tugarinov, V.; Kanelis, V.; Kay, L.E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* **2006**, 1, 749-754.
(45) De Vries, S.J.; Van Dijk, M.; Bonvin, A.M.J.J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc*. **2010**, 5, 883-897.

Chapter 5

General Discussion

NMR Spectroscopy to Investigate Protein-Ligand Interactions

NMR is a powerful and versatile tool with methods that are used for a variety of applications in fragment based drug discovery (FBDD). Common applications involve 1. ligand screening to generate hit matter (Chapters 2 and 3) 2. hit confirmation or hit validation (Chapters 2 and 3) and 3. obtaining atomic resolution 3D structures of the protein-ligand complexes to guide hit to lead development (Chapters 3 and 4). FBDD involves the discovery of weak affinity fragments, which are later optimized into more potent lead compounds. The availability of 3D structural information about the interactions that a small molecule fragment makes within the binding site of a protein target is critical during the fragment hit-lead optimization stage. SBDD (structure based drug design) is considered to be the main driver for the development of many marketed drugs.[1]

The goal of the work described in this thesis was to develop and implement efficient NMR methods that are capable of providing 3D structural information on protein-ligand complexes and applicable in the early stages of the drug discovery projects. Both X-ray crystallography and NMR based methods generate atomic resolution structural information and successfully support SBDD and FBDD. Especially as NMR based methods are sensitive towards weak protein-ligand interactions, they should be applicable where X-ray crystallography is least effective i.e. for weakly bound ligands. However, traditional NMR approaches involving uniform isotopic labeling are labor intensive and limited to proteins of moderate size (e.g., < 30-40 kDa) and have therefore not been widely adopted in drug discovery. In this thesis, NMR methods were sought primarily to address these limitations.[2]

Chemical Shift Perturbation Analysis on Protein-Ligand Complexes

NMR based chemical shift perturbation (CSP) analysis offers a fast and robust means to obtain low resolution structural information on the protein-ligand complex. CSPs are routinely employed to determine the binding location of the ligand. The pre-requisite for determination of the ligand binding site from CSP analysis is the availability of the resonance assignment and the 3D structure (NMR or X-ray) of the target protein. When a ligand is titrated, the amide group of amino acids within the protein that are close to the ligand will experience a change in their local chemical environment. A chemical shift map is generated which allows identifying those amide groups whose environment is most affected due to the binding of a ligand. The map will also include those residues that are indirectly affected by ligand induced changes in the protein. These CSPs result mainly from the sensitivity of amide groups to pH and/or small changes that occur in the hydrogen bonding patterns of protein backbones upon ligand binding.[3,4,5]

CSP based determination of the binding site was employed in Chapters 2 and 3. CSP analysis resulted in locating the binding sites of a DNA oligonucleotide and three fragments (Chapter 2). The CSP mapping clearly showed that the binding site for the fragment hits was the same as that of the oligonucleotide. In the absence of high resolution co-structures of TEL_{ETS}-fragments and based solely on CSP data, we were able to deduce that fragments were able to bind to the DNA binding interface. This subsequently explained the disruption of the protein-DNA complex by fragments in gel-shift assays.

The absence of co-structures of Rit1 with inhibitory compounds (Chapter 3), necessitated the use of the CSP based approach to understand the mechanism of inhibition of nucleotide exchange. NMR based CSP analysis was applied to

determine the binding site of the ligand with the highest affinity for Rit1. Interestingly, the CSP mapping for the ligand indicated multiple binding sites on Rit1. In this case, the primary binding site was close to the GDP nucleotide binding pocket and a secondary site was close to a loop that was missing from the Rit1-GDP crystal structure (PDB 4KLZ). The secondary site might have been a result of a conformational rearrangement caused upon binding of the ligand or influenced by the flexibility of the loop of the protein.[6] As the CSP data suggested the possibility of two binding sites, further structure restraints (described in the paramagnetic NMR section below) were necessary in order to precisely define the ligand binding to the protein and to shed light on the mechanism of inhibition of nucleotide exchange.

In conclusion, the use of CSP data analysis in TEL_{ETS} small molecule discovery work assisted in the determination of the binding site for ligands and the DNA oligonucleotide. CSP data also explained the disruption of the protein-DNA complex by fragments. These fragments represent valuable starting points for developing potent lead compounds. Although the CSP data on the Rit1-inhibitor complexes described in Chapter 3 was informative, they gave ambiguous results. This was attributed to the presence of two distinct binding sites for the inhibitor on the protein surface. In Chapter 4, CSP analysis was used to determine the binding of a fragment hit to a selective methyl group labeled protein using both ^{15}N amide and ^{13}C methyl CSPs. CSP results seem to be unambiguous in cases where a ligand binds at a defined binding site.

In general, CSP analysis can be routinely applied to molecular weight systems up to 40-50 kDa. However, in some cases the resonance assignments for large molecular weight systems (> 60 kDa) cannot be obtained due to their poor NMR spectral characteristics. For such proteins, other alternatives such as selective

isotope labeling and/or paramagnetic NMR methods can be quite useful.[3,4] Sidechain methyl CSPs are comparatively more sensitive than the amide groups as long sidechain containing methyl groups extend out into the binding pocket.[1] Hajduk and coworkers have demonstrated that ^{13}C methyl CSPs can also be used on large molecular weight protein targets to screen a collection of compounds.[7] Spectra of such proteins also have minimal overlap and therefore the interpretation of CSP data becomes easier.[8] Using both the amide and/or sidechain methyl CSPs, it may be possible to classify the compounds in groups that bind to the same binding site on the protein target if they show CSPs for the same resonance, even if the assignment for those resonances is not available. This type of information may prove useful at early stages of the drug discovery. Based on the CSP results obtained for different molecular systems investigated in this thesis, it would be recommended, wherever possible, to observe both the amide and sidechain methyl CSPs to give a better coverage of the ligand binding site (Chapter 4). It should be kept in the mind that the structural information obtained by the CSP approach is of low resolution and does not contain sufficient detail to calculate precisely the orientation of a ligand in the binding site. To obtain the level of detail that includes the precise orientation of a ligand in the binding site, other structural restraints such as intermolecular NOEs should be used.

Paramagnetic NMR to Investigate Protein-Ligand Interactions

Paramagnetic NMR is a widely used technique to study protein-protein interactions.[3,4] As described earlier for the Rit1 work, the CSP analysis resulted in two binding sites for the compound and additional restraints were necessary to define the Rit1-inhibitor interaction. Paramagnetic studies using the GDP-spin label followed

by docking calculations proposed a novel mechanism by which the compound inhibits GDP-GTP exchange of Rit1. The docking generated model structures of Rit1 bound to an inhibitor show that the Rit1-ligand complex is relatively dynamic. The ligand occupies two binding sites on the protein surface. In one of the binding sites, the ligand sterically blocks the GDP by binding over the top of the nucleotide binding site of the protein. This explains that by physically trapping the GDP, the ligand stabilizes the GDP bound form of the protein and nucleotide exchange is inhibited. Using the PRE based NMR method, we were able to understand the mechanism of inhibition of nucleotide exchange caused by the compound. This new binding site may point towards a new approach to inhibit this pharmaceutically important, yet challenging, class of targets.

The Rit1 PRE-work demonstrates how NMR-based techniques can provide structural information even on dynamic protein-ligand complexes when other biophysical approaches to obtain binding site information are unsuccessful or ambiguous. The main advantage of the PRE based approach over using NOEs is that isotope labeled protein is not required. Here, a PRE approach was chosen over conventional NOEs as sidechain assignments for Rit1 were not readily available due to the marginal quality of the NMR spectra. NOE based studies require isotope labeled protein in large quantities and complete protein backbone and sidechain resonance assignments are necessary. Obtaining the required assignments can be challenging. For PRE-based work, only the ligand assignments are required and these are straightforward to obtain.

The Rit1-PRE work shows a small molecule exchanging between multiple binding poses. It is conceivable that a ligand with a weak affinity towards a protein target can adopt multiple binding poses. Various computational studies have shown

distinct multiple binding poses of small molecule ligands in binding sites of T4 lysozyme, neutrophil elastase, estrogen receptor inhibitors, FKBP inhibitors, biotin - streptavidin and cytochrome P450cam.[9-15] Constantine *et al* have observed multiple binding modes of fragment-like kinase inhibitors using computational modeling and NMR studies.[16] As a follow-up study to the Rit1 PRE work, it would be interesting to use paramagnetic pseudo-contact shifts to determine the same protein-ligand co-structure. PCS provide long-range distance restraints (in the range of 10-40 Å) as well as relative orientations on the basis of the anisotropy of the magnetic susceptibility of paramagnetic metals. It would be interesting to see if the inclusion of pseudocontact shift datasets into docking calculations could also confirm the dynamic behavior of the Rit1-inhibitor complex. This should be readily possible as the application of pseudocontact shifts (PCS) in solving 3D atomic resolution structures of protein-small molecule complexes has been recently demonstrated by John *et al* and Guan *et al*.[17,18] It would be interesting to combine the PCS data with molecular dynamic simulations to study dynamic protein-ligand complexes. This may help to differentiate the relative populations of the compound within the multiple binding poses. An extension of the Rit1 PRE-work can be applied to kinases, which are an important class of drug targets against oncology. There is an urgent need for selective kinase inhibitors that bind proximal to the ATP-binding site. The ATP site is highly conserved among protein kinases. The compounds that bind proximal to the ATP site might offer better selectivity. Here, spin labeled ATP can be used for the detection of compounds that bind outside the ATP site. A similar approach was shown by Janke *et al*.[19] Methods like these can be broadly applied to screen for non-ATP site binders. Overall, paramagnetic NMR methods present an alternative to obtain binding site information when other high resolution techniques fail.

Sparse NOE Data Leads to High Resolution Protein-Ligand Co-structures

The availability of protein-ligand co-structures allows the identification of essential interactions made by small molecule with the amino acid residues within the binding site of the target protein.[1] These key interactions can then be improved upon by medicinal chemistry approaches leading towards compounds with better specificity, potency and ligand efficiency. The Hsp90 work in Chapter 4 demonstrates an NMR method that is based on the production of a protein target that is selectively isotopically labeled with the methyl groups of isoleucine, leucine and valine residues and acquisition of a limited number of intermolecular NOE restraints.[13] In this method standard NOESY experiments were used to obtain a set of intermolecular NOEs between the protein-ligand complex. The NOE-based 3D co-structures of the Hsp90-fragment complex obtained using our method revealed interesting difference when compared with the available crystal structures for the identical ligand with Hsp90 in the PDB. The crystal structures show that two or more low energy conformations of the ligand are present in the complex. In contrast, the NMR data indicate that in solution, there is one predominant ligand conformation. Importantly, despite the moderate number of intermolecular restraints, the NMR defines precisely the same binding site as the crystal structure and the critical intermolecular hydrogen bonds.[2]

The method developed in Chapter 4 shows that a universal selective labeling scheme can be used to rapidly identify sufficient numbers of restraints for a small-molecule ligand weakly bound to protein. This chapter demonstrates how a combination of selective methyl group labeling, standard NMR experiments and computational docking can be used to rapidly determine the 3D structure of a small molecule bound weakly to a protein target. Usually, NMR studies of large molecular

weight systems are restrained due to fast signal decay and severe peak overlap. The method utilizes selective labeling and advanced NMR experiments that have allowed structural studies on proteins up to ~100 kDa.[20] Various groups have also undertaken steps towards achieving the resonance assignments using paramagnetic NMR. These were demonstrated to work on various molecular systems previously considered difficult for the resonance assignment. Implementation of selective methyl group labeling in conjunction with automated resonance assignment assisted by paramagnetic NMR should enable structure determination of complexes on reasonably large proteins.[17,21-25] As the method requires only a sparse set of intermolecular NOEs, it presents an alternative to time consuming traditional NMR approaches that involve uniform isotope labeling and a large number of structural restraints. However, the present method depends on the availability of the structure of the target or a good homology model. The method also relies on the production of selectively labeled, deuterated protein, which can be performed only in *E. coli*. Another obvious limitation is the requirement for methyl groups at the ligand binding site. However, it is also possible to selectively label all methyl containing residues (Ala, Met, Thr, Ile, Leu and Val) providing even more complete coverage of protein structures and ligand binding sites.[26] I think that the method combined with above adaptations could be even more powerful and prove valuable for the early stages of FBDD by reliably providing 3D structure information on weakly binding fragment-protein complexes.

In my current position at ZoBio BV, a company that provides tools for FBDD in the pharmaceutical industry, the NMR methods described in Chapter 2, 3 and 4 are being implemented routinely in FBDD projects. The results have led to a significant impact in commercial drug discovery projects. The methods appear to

have sufficient speed and precision to support fragment hit to lead medicinal chemistry efforts.

REFERENCES

1. Hajduk, PJ.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat rev Drug discov*, **2007**, 6, 211-219.
2. Shah, DM.; AB, E.; Diercks,T.; Hass, MA.; Van Nuland, NA.; Siegal, G. Rapid protein-ligand co-structures by sparse NOE data. J Med Chem, **2012,** 55 (23), 10786-10790.
3. Cavanagh, J.; Fairbrother, WJ.; Palmer, AG (III).; Rance, M.; Skelton, NJ. Protein NMR spectroscopy, principles and practice, 2nd edition, **2006**. Elsevier Academic Press, Amsterdam.
4. Roberts, G.; Lian, LY. Protein NMR spectroscopy: principal techniques and applications, **2011**.John Wiley and Sons Ltd. United Kingdom.
5. Lepre, CA;. Moore, JM.; Peng, JW. Theory and applications of NMR-based screening in pharmaceutical research. *Chem Rev.* **2004**, 104(8), 3641-3676.
6. McCoy, MA.; and Wyss, DF. Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations, *J Am Chem Soc*, **2002**, 124, 11758-11763.
7. Hajduk, PJ., Augeri, DJ.; Mack, J.; Mendoza, R.; Yang, JG.; Betz, SF.; Fesik, SW. NMR-based screening of proteins containing ^{13}C-labeled methyl groups, *J Am Chem Soc*, **2000**, 122, 7898-7904.
8. Tugarinov, V.; Kay, L. E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc*. **2003**, 125, 13868-13878.
9. Mobley, DL.; Chodera, JD.; Dill, KA. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J Chem Phys*. **2006**,125, 084902.
10. Mobley, DL.; Chodera, JD.; Dill, KA. Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *J Chem Theory Comput*. **2007**, 3, 1231–1235.
11. Steinbrecher, T.; Case, DA.; Labahn, AA. Multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J Med Chem.* **2006**, 49, 1837–1844.
12. Oostenbrink, C.; Van Gunsteren, WF. Free energies of binding of polychlorinated biphenyls to the estrogen receptor from a single simulation. *Proteins*. **2004**, 54, 237–246.
13. Jayachandran, G.; Shirts, MR.; Park, S; Pande, VS. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J Chem Phys*. **2006**, 125, 084901.
14. Lazaridis, T.; Matsunov, A.; Gandolfo, F. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins*. **2002**, 47, 194–208.
15. Paulsen, MD.; Ornstein, RL. Predicting the product specificity and coupling of cytochrome P450cam. *J Comput Aided Mol Des.* **1992**, 6, 449–460.
16. Constantine, K.; Mueller, L.: Metzler, WJ.; McDonnell, PA.; Todderud, G.; Goldfarb, V.; Fan, Y.; Newitt, JA.; Keifer, SE.; Gao, M. Multiple and single binding modes of fragment-like kinase inhibitors revealed by molecular modeling, residue type-selective protonation, and nuclear overhauser effects. *J Med Chem*. **2008**, 51, 6225–6229.

17. Guan, JY.; Keizers, PH.; Liu, W.; Löhr, F.; Skinner, SP.; Heeneman, E.; Schwalbe, H.; Ubbink, M.; Siegal, G. Small-Molecule Binding Sites on Proteins Established by Paramagnetic NMR Spectroscopy *J. Am. Chem. Soc.*, **2013**, 135 (15), 5859-5868.
18. John, M.; Pintacuda, G.; Young Park, A.; Dixon, NE.; Otting, G. Structure Determination of Protein−Ligand Complexes by Transferred Paramagnetic Shifts. *J. Am. Chem. Soc.,* **2006**, 128 (39), 12910–12916.
19. Jahnke, W.; Blommers, MJ.; Fernández, C.; Zwingelstein, C.; Amstutz, R. Strategies for the NMR-based identification and optimization of allosteric protein kinase inhibitors. *Chembiochem.* **2005**, 6(9),1607-1610.
20. Tugarinov, V.; Choy, WY.; Orekhov, VY; Kay, LE. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci.* **2005**,102, 622–627.
21. Otting, G. Protein NMR using paramagnetic ions. *Annu Rev Biophys.* **2010**, 39, 387-405.
22. Venditti, V.; Fawzi, N.L.; Clore, G.M. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. *J. Biomol. NMR.* **2011**, 51, 319-328.
23. John, M.; Schmitz, C.; Park, A.; Dixon, N.; Huber, T.; Otting, G. Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides. *J. Am. Chem. Soc.* **2007**, 129, 44, 13749-13757.
24. Xu, Y.; Liu, M.; Simpson, PJ.; Isaacson, R.; Cota, E.; Marchant, J.; Yang, D.; Zhang, X.; Freemont, P.; Matthews, S. Automated assignment in selectively methyl-labeled proteins. *J. Am. Chem. Soc.* **2009**, 131, 9480–9481.
25. Skinner, SP.; Moshev, M.; Hass, MA.; Ubbink, M. PARAssign-paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts. *J. Biomol.NMR.* **2013**, 55(4),379-389.
26. Otten, R.; Chu, B.; Krewulak, KD.; Vogel, HJ.; Mulder, F. Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. *J. Am. Chem. Soc*. **2010**, 132(9), 2952 – 2960.

English Summary

NMR methods are used for a variety of applications in a drug discovery process to guide hit to lead development and ultimately to generate potential drug candidates. The availability of 3D structural information on protein-ligand complexes is crucial during various stages of drug development. NMR based methods are sensitive toward weak protein-ligand interactions and are applicable where X-ray crystallography is least effective i.e. for weakly bound ligands. However, traditional NMR approaches that involve uniform isotopic labeling are labor intensive and limited by the size of protein targets (e.g., < 30-40 kDa). Therefore, these have not been widely adopted in drug discovery. The main focus of the work described in this thesis was to develop and implement efficient NMR methods that are capable of providing 3D structural information on protein-ligand complexes in the early stages of the drug discovery projects. The work presented in this thesis covers a wide range of classical and newly developed NMR techniques applied to diverse molecular systems.

Chapter 1 provides general introduction to the drug discovery process, in particular, to the fragment-based drug discovery. The chapter also describes NMR methods that are often used for screening purposes and to obtain 3D structural information on protein target-small molecule complexes.

In Chapter 2, a fragment based- small molecule discovery approach was sought to discover inhibitors that disrupt the DNA binding capability of TEL, a target for oncology. TEL is a DNA binding protein involved in transcriptional regulation of many cellular proteins that are involved in the regulation of angiogenesis. TEL is considered to be a therapeutic target for pathological angiogenesis in tumor cells. Chapter 2 describes the techniques and approaches that were undertaken to discover small molecule ligands that specifically bind the DNA binding ETS-domain of TEL. In order to discover the primary hit matter, a TINS NMR fragment screen was

applied. To demonstrate the DNA binding capability of TEL_{ETS}, different array of techniques were used. Biochemical and structural analysis were performed using protein observed NMR, Surface Plasmon Resonance (SPR) and gel-shift assay to characterize DNA binding activity of TEL. The availability of the resonance assignments allowed determining the binding site of DNA and fragments hits. Based on the chemical shift perturbation analysis and CSP mapping, it was clearly observed that the binding site for the DNA and three fragments was common on the protein surface. To investigate whether these fragments were capable of disrupting the TEL_{ETS}-DNA complex, a gel retardation assay was performed. The assay results indicated that the TEL_{ETS}-DNA complex could be disrupted in the presence of high (millimolar) concentrations of 2 fragments. Clearly, the affinity for these fragments appears to be very low and the specificity, in relation to the other ETS-containing proteins remains to be tested. However, these fragments represent valuable starting points for further stages of drug discovery on TEL_{ETS}. This work demonstrates how the application of NMR screening can be efficiently used to discover fragment hits and protein-observed NMR studies can be used to obtain low resolution structural information on protein-ligand complexes.

In Chapter 3, a fragment based- small molecule discovery approach was initiated to discover small molecule inhibitors for Rit1 GTPase, a target for rheumatoid arthritis. Cellular studies carried out by Galapagos BV showed that inactivation of Rit1 reduced the levels of matrix metalloproteinases (MMPs) in bone cartilage tissue and that increased expression of Rit1 caused MMP levels to increase. MMPs are responsible for degradation of cartilage tissue and hence cause RA. Therefore, development of small molecule inhibitors that could inactivate Rit1 is important for a potential RA treatment. Chapter 3 describes the results obtained from

fragment based screening, crystallization, analoging (compounds with similar chemical structures), hit development and a structural study on the most potent compound. A diverse range of techniques and methods was employed. Fragment hits were discovered using TINS NMR. The fragment hits were further tested in an *in vitro* biochemical assay. These hits and analog compounds were able to inhibit the GDP-GTP nucleotide exchange. In other words, compounds were able to stabilize the GDP bound inactive state of Rit1. Despite their functional inhibition, the mechanism of action of compounds remained elusive. High resolution 3D structures of the protein-compound would give a direct insight into the mechanism. The crystallography efforts yielded a high resolution 3D structure of the Rit1 bound to GDP nucleotide but efforts to obtain 3D structures on protein-ligand complexes failed. Substantial efforts to crystallize the complex were not successful due to the low solvent content of the protein crystals and cracking of protein crystals in the presence of the most potent compound. To address this issue, a solution based NMR approach was necessary and a paramagnetic NMR based approach was sought, whereby a spin label was introduced on GDP to obtain GDP-spin label. The paramagnetic NMR approach was combined with CSP analysis to obtain low resolution information on the binding site of ligands. Paramagnetic studies using the GDP-spin label followed by docking calculations proposed a novel mechanism by which the compound inhibits GDP-GTP exchange of Rit1. In one of the binding sites, the ligand sterically blocks the GDP nucleotide by binding over the top of the nucleotide binding site of the protein. This explains that by physically trapping the GDP nucleotide, the ligand stabilizes the GDP bound form of the protein and the nucleotide exchange with GTP is inhibited. This work shows how NMR-based techniques can provide structural information when other biophysical approaches to

obtain binding site information are unsuccessful. Here, using the PRE based NMR method, we were able to understand the mechanism exhibited by the compounds to achieve the inhibition of nucleotide exchange. PRE based methods present an alternative to obtain binding site information when other high resolution techniques fail. In principle, GDP-SL can be used to investigate any other GTPase.

In Chapter 4, a solution NMR method was developed to obtain 3D structures of protein-small molecule complexes in rapid and efficient manner. The method was developed using the Hsp90-fragment complex as a model system. Hsp90 interacts with a large number of client proteins and it is thought that ATP hydrolysis is the main driver for these interactions and for this reason many drug discovery approaches have focused on targeting the ATP binding pocket in the N-terminal domain of Hsp90. The method described in Chapter 4 makes use of a small molecule fragment that binds in the ATP binding pocket of the N-terminal Hsp90 discovered by TINS NMR screening. The main goal was to use this protein-ligand system to develop an efficient way to obtain 3D structural information on protein bound to a ligand. In Chapter 4, an NMR method that is based on the production of a protein target that is selectively isotopically labeled with the methyl groups of isoleucine, leucine and valine residues is described. In this method standard NOESY experiments were used to obtain a set of intermolecular NOEs between the protein-ligand complex. The intermolecular NOEs were used as distance restraints to obtain 3D co-structures by molecular docking. The crystal structures were available for the identical ligand with Hsp90 in the PDB. The crystal structures show that two or more low energy conformations of the ligand are present in the complex. In contrast, the NMR data indicate that in solution, there is one predominant ligand conformation. Importantly, despite the moderate number of intermolecular restraints, the NMR defines precisely

the same binding site as the crystal structure and the critical intermolecular hydrogen bonds, suggesting that in addition to being fast, the method is robust. The method shows that a universal selective labeling scheme can be used to rapidly identify sufficient numbers of restraints for a small-molecule ligand weakly bound to protein. This chapter demonstrates how a combination of selective methyl group labeling, standard NMR experiments and computational docking can be used to rapidly determine the 3D structure of a small molecule bound “weakly” to a protein target. The approach requires only a sparse set of intermolecular NOEs and is an alternative to time consuming traditional NMR approaches that involve uniform isotope labeling. This method is also amenable to large molecular weight targets.

Chapter 5 presents the general discussion of the work and provides future perspectives for NMR methods developed and implemented in this thesis.

Nederlandse Samenvatting

In de ontwikkeling van geneesmiddelen worden NMR-methoden gebruikt voor tal van toepassingen in het proces van *hit-to-lead* en de uiteindelijke ontwikkeling van potentiële medicijnkandidaten. De beschikbaarheid van 3D structuurinformatie in eiwit-ligand complexen is cruciaal tijdens de verschillende fasen van medicijnontwikkeling. Methoden die op NMR zijn gebaseerd, zijn gevoelig voor zwakke eiwit-ligand interacties, in het gebied waar Röntgenkristallografie het minst effectief is. Helaas zijn traditionele NMR-benaderingen, die uniforme isotoopverrijking vereisen, arbeidsintensief en beperkt tot kleinere eiwitten (< 30-40 kDa) en daarom worden deze methoden niet op grote schaal gebruikt in medicijnontwikkeling. Het belangrijkste doel van het in dit proefschrift beschreven werk was daarom het ontwikkelen en implementeren van efficiënte NMR-technieken die 3D structuurinformatie van eiwit-ligand complexen kunnen opleveren voor het gebruik in de vroege stadia van medicijnontwikkeling. Het werk bestrijkt een groot gebied van klassieke en nieuw-ontwikkelde NMR-technieken, toegepast op diverse moleculaire systemen.

Hoofdstuk 1 geeft een algemene introductie in het proces van medicijnontwikkeling, in het bijzonder de *fragment based drug discovery*. Dit hoofdstuk beschrijft ook de NMR-methoden die vaak worden gebruikt voor *screening* doeleinden en om 3D structuurinformatie van complexen tussen *target* eiwitten en kleine moleculen te verkrijgen.

In hoofdstuk 2 wordt een *fragment based small molecule discovery* benadering getoetst om remmers te vinden die het bindend vermogen van TEL, een *target*-eiwit in de oncologie, beïnvloeden. TEL is een eiwit dat aan DNA bindt en betrokken is bij

de transcriptieregulatie van veel cellulaire eiwitten die een rol spelen bij angiogenese. TEL wordt beschouwd als een therapeutisch *target* voor pathologische angiogenese in tumoren. In hoofdstuk 2 worden de technieken en benaderingen beschreven die zijn toegepast om kleine moleculen te vinden die specifiek associëren met het DNA-bindende ETS-domein van TEL. Een TINS NMR *fragment screen* werd uitgevoerd om de primaire *hit*-materie (alle bindende moleculen) te vinden. Om het resterende DNA-bindende vermogen van TEL_{ETS} te bepalen zijn verschillende technieken gebruikt. Biochemische analyse en structuuranalyse werden uitgevoerd met behulp van NMR-spectroscopie aan het eiwit en *Surface Plasmon Resonance* (SPR) en een *gel-shift assay* werden gebruikt om de DNA-binding van TEL te beschrijven. De reeds beschikbare resonantietoekenningen maakten het mogelijk om de bindingsplaats van zowel het DNA als de *hit*-fragmenten op het eiwitoppervlak vast te stellen. Met behulp van *chemical shift perturbation* analyse kon duidelijk worden vastgesteld dat de bindingsplaats voor het DNA overeenkwam met die van drie fragmenten. Om te onderzoeken of deze fragmenten het TEL_{ETS}-DNA complex kunnen verstoren werd een gelretardatie-test uitgevoerd. De resultaten toonden aan dat het TEL_{ETS}-DNA complex kon worden verstoord in de aanwezigheid van hoge (millimolair) concentraties van twee fragmenten. De affiniteit voor deze fragmenten is duidelijk erg laag en de specificiteit ten opzichte van andere ETS-bevattende eiwitten moet nog worden onderzocht. Toch vertegenwoordigen deze fragmenten waardevolle uitgangspunten voor medicijnontwikkeling met betrekking tot TEL_{ETS}. Dit werk laat zien dat de toepassing van NMR *screening* op een efficiënte manier kan worden gebruikt om *hit*-fragmenten te vinden en dat NMR aan het eiwit kan worden toegepast om structuurinformatie van eiwit-ligand complexen met een lage resolutie te verkrijgen.

In hoofdstuk 3 staat beschreven hoe een *fragment based small molecule discovery* benadering werd geïnitieerd om kleine-molecule remmers te vinden voor Rit1 GTPase, een *target* voor reumatoïde artritis. Cellulaire studies, uitgevoerd door Galapagos BV, lieten zien dat het inactiveren van Rit1 leidde tot lagere hoeveelheden van matrix metalloproteinen (MMPs) in kraakbeen en dat toegenomen expressie van het gen voor Rit1 die hoeveelheden deed stijgen. MMPs zijn verantwoordelijk voor vermindering van kraakbeenweefsel en veroorzaken dus RA. Daarom is de ontwikkeling van kleine-molecule remmers die Rit1 kunnen inactiveren belangrijk voor een mogelijke behandeling van RA. Hoofdstuk 3 beschrijft de verkregen resultaten van *fragment based screening*, kristallisatie, *analoging* (evaluatie van moleculen met vergelijkbare structuur), *hit development* en een structuuronderzoek van de meest potente verbinding. Een reeks van diverse technieken en methoden werd toegepast. *Hit*-fragmenten werden gevonden met behulp van TINS NMR. Deze fragmenten werden nader getest in een *in vitro* biochemische test. De *hits* en analoge verbindingen konden de GDP/GTP nucleotide-uitwisseling verhinderen. Met andere woorden, de verbindingen waren in staat om de inactieve, GDP-gebonden toestand van Rit1 te stabiliseren. Het mechanisme van deze remmende werking was onduidelijk. Hoge-resolutie 3D structuren van het eiwit-molecule complex zouden inzicht in dit mechanisme kunnen geven. Kristallisatie leverde weliswaar een hoge-resolutie 3D structuur van Rit1 gebonden aan GDP nucleotiden op, maar 3D structuren van eiwit-ligand complexen konden niet worden verkregen. Veel pogingen om het complex te kristalliseren bleven zonder succes, vanwege het lage gehalte aan oplosmiddel in de eiwitkristallen en omdat de eiwitkristallen braken in de aanwezigheid van de meest potente verbinding. Vanwege deze problemen was een NMR-benadering

noodzakelijk. Er werd een paramagnetische NMR methode toegepast, waarbij een *spin label* op GDP werd aangebracht om een *GDP-spin label* te verkrijgen. De paramagnetische NMR-benadering werd gecombineerd met CSP analyse om lage-resolutie informatie over de bindingsplaats van de liganden te verkrijgen. Op grond van de paramagnetische studies met de *GDP-spin label*, gevolgd door *docking* berekeningen, is een nieuw mechanisme voorgesteld voor het afremmen van de GDP/GTP-uitwisseling van Rit1 door de verbinding. Het ligand blokkeert de GDP-nucleotide aan de 'bovenkant' van de nucleotide-bindingsplaats op het eiwit. Door fysiek afschermen van de nucleotide door het ligand wordt de GDP-gebonden toestand van het eiwit gestabiliseerd en de nucleotide-uitwisseling met GTP verhinderd. Dit werk laat zien hoe op NMR-gebaseerde methoden, in dit geval gebruik makend van paramagnetische relaxatie, structuurinformatie kunnen geven in gevallen waar andere biofysische technieken zonder resultaat blijven. In principe kan het *GDP-spin label* worden gebruikt om elke andere GTPase te bestuderen.

In hoofdstuk 4 staat beschreven hoe een andere methode voor vloeistof-NMR is ontwikkeld om op een snelle en efficiënte manier 3D structuurinformatie van complexen tussen eiwitten en kleine moleculen te verkrijgen. Bij de ontwikkeling van deze methode werd gebruik gemaakt van een Hsp90-fragment complex als modelsysteem. Hsp90 gaat interacties aan met een groot aantal andere eiwitten. Verondersteld wordt dat de hydrolyse van ATP de drijvende kracht is achter deze wisselwerkingen. Dit is de reden waarom veel medicijnontwikkeling gericht is op het ontwikkelen van moleculen die binden in de ATP-bindingsplek in het N-terminale domein van Hsp90. De methode maakt gebruik van een fragment dat bindt in de ATP-bindingsplek, gevonden door gebruikmaking van TINS NMR *screening*. Het

belangrijkste doel was om met behulp van dit eiwit/ligand-systeem een efficiënte manier te ontwikkelen om 3D structuurinformatie van een aan een ligand gebonden eiwit te verkrijgen. De methode is gebaseerd op het gebruik van eiwit dat selectief is verrijkt met isotopen in de methylgroepen van isoleucine, leucine en valine residuen. Bij deze methode worden standaard NOESY experimenten gebruikt om een reeks intermoleculaire NOEs in het eiwit-ligand complex te verkrijgen. Deze NOEs werden gebruikt om 3D co-structuren te verkrijgen door moleculaire *docking*. De kristalstructuren waren voor hetzelfde ligand gebonden aan Hsp90 waren reeds beschikbaar in de PDB. Deze structuren laten zien dat er twee of meer lage-energie conformaties van het ligand aanwezig zijn in het complex. In tegenspraak hiermee tonen de NMR data aan dat er in oplossing slechts één overheersende ligandconformatie aanwezig is. Het is een belangrijke vinding dat, ondanks het bescheiden aantal intermoleculaire contacten, de NMR-methode precies dezelfde bindingsplaats en kritische intermoleculaire waterstofbruggen definieert als de kristalstructuur laat zien. Dit suggereert dat de methode niet alleen snel maar ook robuust is. De methode laat zien dat een universele selectieve verrijkingsstrategie kan worden gebruikt om snel een voldoende aantal contacten te identificeren voor een klein molecule dat zwak gebonden is aan een eiwit. Het vormt ook een alternatief voor tijdrovende traditionele NMR benaderingen die uniforme isotoopverrijking met zich mee brengen. De methode is ook toe te passen op eiwitten met hoge moleculegewichten.

In hoofdstuk 5 wordt een algemene discussie van het werk gegeven en er worden toekomstperspectieven besproken voor de ontwikkelde en toegepaste NMR-methoden die in dit proefschrift zijn behandeld.

List of publications

1. De Groote, FH., Jansen, JG., Masuda, Y., **Shah, DM.**, Kamiya, K., de Wind, N., Siegal, G.
The Rev1 translesion synthesis polymerase has multiple distinct DNA binding modes.
DNA Repair (Amst), **2011**, 10(9):915-25

2. **Shah, DM.**, AB, E., Diercks, T., Hass, MAS., van Nuland, NAJ., Siegal, G.
Rapid protein-ligand costructures from sparse NOE data.
J Med Chem. **2012**, 55(23):10786-90.

3. **Shah, DM.**, Kobayashi, M., Keizers, PHJ., Tuin, AW., AB, E., Manning, L.,
Rzepiela, AA., Andrews, M., Hoedemaeker, FJ., Siegal, G.
Inhibition of small GTPases by stabilization of the GDP Complex, a novel approach applied to Rit1, a target for rheumatoid arthritis.
Manuscript to be submitted

4. **Shah, DM.,** AB, E., Nachane, RK., Hollander, JG., Baker, DA., Siegal, G.
Discovery of small molecule fragments that inhibit the DNA binding of the TEL-ETS domain.
Manuscript to be submitted

Curriculum Vitae

Dipen Shah was born on 26th of January 1985 in Mangalore, India. He obtained his Bachelor's degree in Pharmaceutical Sciences from ICT (Institute of Chemical Technology), University of Mumbai in 2006. After his graduation, he was awarded "J.N.Tata Endowment for the Higher Education of Indians" and "Nagher Samaj Educational Foundation" scholarships to pursue MRes degree (Master of Research) in Biomedical Sciences at University of Glasgow, Scotland, U.K. During the master's course, he had the opportunity to carry out research projects under the supervision of Dr. Niall Fraser and Prof. Dr. Graeme Milligan. The research work involved expression and functional characterization of adenosine and dopamine GPCRs (G-protein coupled receptors). In October 2007, he started his PhD in the Protein Chemistry group under the supervision of Dr. Gregg Siegal. The research focused on developing and implementing NMR based methods to investigate protein-small molecule complexes. Part of the work described in the thesis was presented at the Gordon Research Conference on DNA damage and mutations in 2010 at Ventura, CA, USA and CHAINS 2011 (Chemistry as Innovating Science) in Maarssen, the Netherlands. Since January 2012, he is employed as Postdoctoral Scientist at ZoBio BV, a company that provides tools for fragment based drug discovery.