NMR structural studies of protein-small molecule interactions

Shah, D.M.

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]

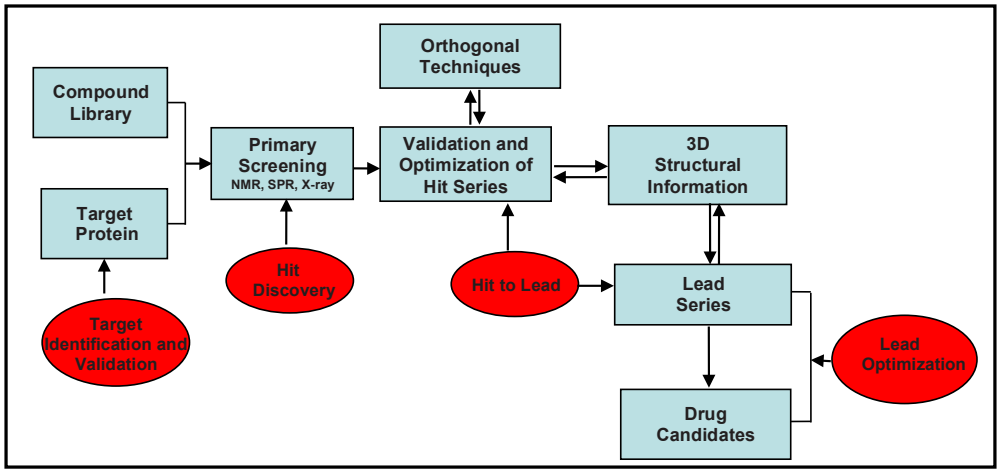

Figure 1. Outline of initial stages involved in a target based drug development.

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 ^{1}H 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]

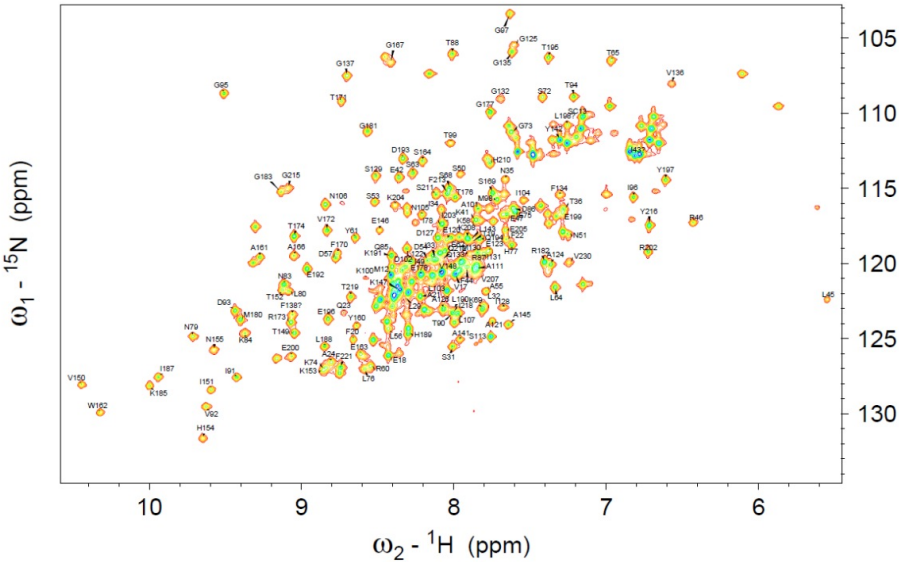

Figure 2. A 2D [^{1}H,^{15}N] HSQC spectrum of N-terminal ATPase domain of Hsp90 displayed with backbone (^{1}H and ^{15}N) resonance assignments.

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