NMR structural studies of protein-small molecule interactions

Shah, D.M.

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.