Studies of Protein Structure, Dynamics and Protein-ligand Interactions using NMR Spectroscopy

Tobias Tengel
Title
Studies of protein structure, dynamics and protein-ligand interactions using NMR spectroscopy.

Author
Tobias Tengel, Department of Medical Biochemistry and Biophysics, Umeå University, SE – 901 87, Umeå, Sweden.

Abstract
In the first part of the thesis, protein-ligand interactions were investigated using the chaperone LcrH, from *Yersinia* as target protein. The structure of a peptide encompassing the amphipathic domain (residue 278-300) of the protein YopD from *Yersinia* was determined by NMR in 40% TFE. The structure of YopD278-300 is a well defined α-helix with a β-turn at the C-terminus of the helix capping the structure. This turn is crucial for the structure as peptides lacking the residues involved in the turn are unstructured. NMR relaxation indicates that the peptide is not monomeric. This is supported by intermolecular NOEs found from residue Phe280 to Ile288 and Val292 indicative of a multimeric structure with the helical structures oriented in an antiparallel manner with hydrophobic residues forming the oligomer. The interaction with the chaperone LcrH was confirmed by $^1$H relaxation experiments and induced chemical shift changes in the peptide.

Protein-ligand interactions were investigated further in the second paper using a different approach. A wide range of substances were used in screening for affinity against the chaperones PapD and FimC from uropathogenic *Escherichia coli* using $^1$H relaxation NMR experiments, surface plasmon resonance and $^{19}$F NMR. Fluorine NMR proved to be advantageous as compared to proton NMR as it is straightforward to identify binding ligands due to the well resolved $^{19}$F NMR spectra. Several compounds were found to interact with PapD and FimC through induced line-broadening and chemical shift changes for the ligands. Data corroborate well with surface plasmon resonance and proton NMR experiments. However, our results indicate the substances used in this study to have poor specificity for PapD and FimC as the induced chemical shift is minor and hardly no competitive binding is observed.
Paper III and IV is an investigation of the structural features of the allergenic 2S albumin Ber e 1 from Brazil nut. Ber e 1 is a 2S albumin previously identified as the major allergen of Brazil nut. Recent studies have demonstrated that endogenous Brazil nut lipids are required for an immune response to occur \textit{in vivo}. The structure was obtained from 3D heteronuclear NMR experiments followed by simulated annealing using the software ARIA. Interestingly, the common fold of the 2S albumin family, described as a right-handed super helix with the core composed of a helix bundle, is not found in Ber e 1. Instead the C-terminal region is participating in the formation of the core between helix 3, 4 and 5. The dynamic properties of Ber e 1 were investigated using $^{15}$N relaxation experiments and data was analyzed using the model-free approach. The analysis showed that a few residues in the loop between helix 2 and 3 experience decreased mobility, compared to the rest of the loop. This is consistent with NOE data as long range NOEs were found from the loop to the core region of the protein. The anchoring of this loop is a unique feature of Ber e 1, as it is not found in any other structures of 2S albumins. Chemical shift mapping of Ber e 1 upon the addition of lipid extract from Brazil nut identified 4 regions in the protein where chemical shift perturbations were detected. Interestingly, all four structural clusters align along a cleft in the structure formed by helix 1-3 on one side and helix 4-5 on the other. This cleft is big enough to encompass a lipid molecule. It is therefore tempting to speculate whether this cleft is the lipid binding epitope in Ber e 1.

\textbf{Keywords}
$^{19}$F NMR, 2S albumin, allergy, amphipathic helix, Ber e 1, Brazil nut, dynamics, FimC, LcrH, NMR screening, PapD, protein-ligand interaction, structure, TFE, \textit{Yersinia}, YopD,
1. List of Papers

This thesis is based on the papers listed below in chronological order, which will be referred to in the text by the corresponding Roman numerals.


IV  Tengel, T.; Alcocer, M. J.; Larsson, G.; Schleucher, J., Solution NMR structure and backbone dynamics of the allergenic 2S albumin Ber e 1 from Brazil nut (*Manuscript*)

Reprinted with kind permission of the publishers

Projects where a contribution has been made (not included in the thesis)


Przygodzka, P.; Olausson, B.; Tengel, T.; Larsson, G.; Wilczynska, M., Bomapinis a redox-regulated serpin which stabilizes retinoblastoma protein during apoptosis and increase proliferation of leukemia cells. (*Manuscript*)


Tengel, T.; Lundqvist, M.; Sethson, I., NMR solution structure of the B-linker of DmpR (*Manuscript*)
## 2. List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIA</td>
<td>Ambiguous Restraints in Iterative Assignment</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>FimC</td>
<td><em>E. coli</em> chaperone associated with type 1 pilus</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum correlation</td>
</tr>
<tr>
<td>Lcr</td>
<td>low calcium response</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear Overhauser enhancement</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser enhancement spectroscopy</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PapD</td>
<td><em>E. coli</em> chaperone associated with P pilus</td>
</tr>
<tr>
<td>SAR</td>
<td>structure activity relationship</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>TFE</td>
<td>2,2,2-trifluoroethanol</td>
</tr>
<tr>
<td>TROSY</td>
<td>transverse relaxation optimized spectroscopy</td>
</tr>
<tr>
<td>Yop</td>
<td><em>Yersinia</em> outer protein</td>
</tr>
</tbody>
</table>
3. Introduction

3.1. NMR in a Historical Perspective

Knowledge about the structure of biological macromolecules is crucial in understanding the function of biological processes. Since the first protein structure in 1985\textsuperscript{1}, technological breakthroughs have brought nuclear magnetic resonance to the front of structural biology.

While NMR 20 years ago accounted for only a few percent of the depositions in the Protein Data Bank\textsuperscript{2}, they currently represent 15\% of the total structures with about 900 new structures added each year of the total 6000 yearly depositions. The protein structures in the Protein Data Bank are mostly X-ray structures, 86\%, as compared to 14\% for NMR structures. However when looking at the structures of nucleic acids NMR accounts for 44\% of the deposited structures which is an impressive number considering X-ray has a 25 years head start of NMR\textsuperscript{3}.

NMR has long been an important technique for obtaining the structure of peptides and proteins but over the last ten years, NMR spectroscopy has evolved into an important discipline in drug discovery. Biological has traditionally been used as a technique to provide structural information of protein drug targets and target-ligand interactions. More recently, it has been shown that NMR may be used as an alternative method for identification of small molecules that bind to protein drug targets. Although it has been known for many years that NMR is a sensitive method for detecting the binding of small ligands to proteins, the use of NMR as a tool for screening large libraries has only been applied in recent years.

The applications of NMR to structural biology are greatly facilitated by hardware developments during the last ten years such as the use of cryoprobe technology to substantially increase the signal-to-noise ratio. Flow probes make NMR tubes and their time-consuming handling obsolete\textsuperscript{4} and also enable the direct coupling with separation techniques such as (HP)LC. In addition, the constant development in computer technology greatly facilitates the study of biological macromolecules as the speed of data analysis is increased.
Unlike other methods, NMR screening is a “blind method”, that is, no prior knowledge of target function is necessary. Consequently, NMR has the possibility to detect ligands with affinity for target proteins at a very early stage in the drug discovery process when no functional assay has been developed or in the study of proteins with unknown functions. In addition to information whether a ligand is binding a target or not,\(^5\) information about bound structure,\(^6\) binding epitope and the binding site in target can be obtained using various NMR techniques, reviewed in\(^7\), \(^8\). In the case of NMR screening, which is mainly applied to proton NMR resulting in complex data with risk of extensive chemical overlap when large screening libraries are used. Large libraries could be used in the case of target observed screening\(^9\). However, that approach requires the use of labelled protein samples.

It should be stressed, before I upset too many people that NMR can never compete with X-ray crystallography in the speed of structure determination. However it can provide additional information regarding for example molecular dynamics and competitive binding and NMR can solve the structure of highly flexible molecules.

### 3.2. Structural Biology

To be able to understand functions of complex biological processes, information regarding the structure and dynamics of macromolecules are of great importance. The best example where structural information has been crucial for the understanding of biological function is the DNA molecule.\(^10\) Structural information of this double helical structure composed of only a few different base pairs facilitated the understanding of genetics. Understanding the structural properties of DNA gave information of two important processes, the ability of DNA to replicate and the ability to determine the structure of proteins. Molecular genetics since these findings has evolved remarkably fast and at this point the whole human genome is known. The importance of structural data to understand biological processes has lead to an increase in research to determine the structure of biomolecules.

However, in comparison with genetics, progress in protein structure determination has been painfully slow. As proteins are composed of 20 different amino acids, compared to only four base pairs in DNA, the
numbers of possible combinations are enormous. Although numerous protein structures are known today, almost 50 000 in the Protein Data Bank, there is still impossible to predict the three-dimensional structure using only the amino acid sequence. Structural biology has relied on experimental methods as X-ray crystallography and NMR to obtain structural data of biomolecules. While crystallography is a good method, protein crystals are sometimes difficult to obtain for certain proteins. The study of protein structure by NMR also suffers drawbacks as all proteins are not possible to solubilize in concentrations required for NMR. Protein expression techniques have also evolved making it possible for crystallographers to study biologically interesting proteins, and not only proteins that are naturally present in large amounts. NMR benefit from the same advances as large amount of proteins can be obtained and if necessary also be labelled with the NMR active nuclei $^{13}$C and/or $^{15}$N. As mentioned, a large part of the structures in the Protein data Bank are determined by X-ray crystallography and to understand structural biology, a static structure is not always enough, although it is a good starting point. Even though each protein has a specific structure, this structure is by no means static in reality. The flexibility and dynamics of biomolecules are of great importance in the understanding of protein-protein and protein-ligand interactions and should be studied to obtain a complete picture.

NMR has evolved into a powerful tool in structural biology as it has the ability to provide information about dynamics as well as the structure for active molecules in solution. The fact that the structural features are studied in solution gives the opportunity to vary a wide range of conditions, like pH, temperature and different solvents. In addition, NMR has the ability to study protein folding as well as competitive binding between, for example, different drugs and a target protein.

### 3.3. NMR as a Tool for Structural Investigation

NMR has been established as a valuable method for determining the structure of protein and protein-ligand complexes. NMR produces signals which can be assigned to specific atoms in a molecule, for example to the amide proton in a certain residue in a protein. The use of proton NMR is most common and yield most information about a molecule due to the abundance and sensitivity of this nucleus. However,
when studying larger molecules this also results in overlap between
different resonances. Multidimensional NMR methods\textsuperscript{12} may resolve this
problem and in the case of large proteins, introduction of other nuclei
such as $^{13}$C and $^{15}$N will raise dimensionality, further increasing the
number of molecules for which detailed structural information could be
obtained.

Structure determination by NMR always requires the assignment of all
resonances. When all nuclei have been assigned, NMR experiments give
information about for example inter-nuclear distances, dihedral angles
and hydrogen bonds in a molecule. The parameters are then used as
restraints in the structure calculations which result in an ensemble of
structures that has to be validated to ensure that obtained structures are in
agreement with the experimental data. The following section gives a brief
overview of different NMR restraints and methods used for structure
calculations and validation.

3.3.1. Assignment

Assignment of all resonances in a molecule is the first step in order to
obtain structural restraints. It is also the most time consuming step in the
process of getting from spectra to structure. Standard two-dimensional
NMR experiments used for structural information have been reviewed
extensively.\textsuperscript{12} The spectral assignment of large proteins may be a difficult
task due to spectral overlap and 2D-NMR experiment will not be
sufficient in this case. However, the introduction of three- and four-
dimensional experiments and the availability of $^{13}$C and $^{15}$N labelled
proteins have provided methods to determine the structure of large protein
complex, reviewed in.\textsuperscript{13} For unlabelled molecules the resonance
assignment is a combination of nuclear Overhauser enhancement (NOE)
and through-bond connectivities.\textsuperscript{11} The full benefits of labelling were not
recognized until the arrival of the triple resonance experiment.\textsuperscript{14, 15}
Experiments that allowed through bond correlations of spins made it
possible to assign large proteins. Using these experiments a continuous
unambiguous assignment of the entire backbone of large proteins is
possible. Experiments used for backbone assignment of Bere 1 (\textbf{Paper
III}) are illustrated in Figure 3.1. The resonance assignment of
multidimensional experiments is in principle an extension of Wüthrich’s
strategy which rely on homonuclear $^1$H NMR experiments. Equivalent experiments are available for backbone and side chain assignment are reviewed in. In addition, semi-automated approaches are available for the resonance assignment which could reduce the work load.

\[ \text{Figure 3.1 NMR experiments used for backbone assignment of Ber e 11 (Paper III).} \]
3.3.2. NMR Restraints

The traditional structure restraints in NMR contain information about distances and dihedral angles. The structure of proteins is mainly determined by distance restraints derived from cross peaks in a NOESY\(^{17}\) experiment. The NOE is a consequence of dipole-dipole cross relaxation in nuclear spin systems which allows one to estimate the intramolecular distance between for example two protons in a protein. The NOE between two nuclei is proportional to the inverse sixth power of the distance between them. This strong distance dependence selects for distances less than \(\sim 6\) Å in the molecule. Dihedral angles are often used as a complement to distance restraints in structural investigations. Dihedral angles can be derived from J-coupling constants using the Karplus equation\(^{18}\) or alternatively from chemical shift data.\(^{19}\) In addition to the two restraints mentioned above other parameters like hydrogen bond restraints, chemical shift restraints, hydrodynamic calculations from NMR relaxation\(^{20}\) and residual dipolar couplings\(^{21}\) are available.

The NMR restraints used in this thesis consist of the two classic ones, NOEs were used in Paper I, III and IV where Paper I relied on two dimensional experiments whereas isotope labelled material was available for Ber e 1, hence \(^{13}\)C and \(^{15}\)N-edited NOESY experiments were used. Dihedral angles were obtained using the HNHA experiment\(^{13}\) (Paper IV). Other restraints such as chemical shift restraints were used indirectly as the software TALOS\(^{19}\) was used to derive the dihedral angles (Paper I, III and IV). In addition, relaxation data has been collected for Ber e 1 (Paper IV) but has not been incorporated into the structure calculations yet, although it is feasible today.\(^{22}\)

3.3.3. Structure Calculations and Validations

Once the NMR restraints have been collected they are used as input in the structure calculation. The most used methods to obtain a model of proteins by NMR are distance geometry\(^{23}\) and \(ab\ initio\) simulated annealing.\(^{24,25}\) Once a reasonable model is obtained the goal is to identify the global minimum of the molecule, which consists of chemical and experimental restraints. Conducting a conformational search to find the global minimum is not straightforward as the target function defining the
minimum consists of several local minima that have to be overcome. In order to solve this issue, simulated annealing refinement is commonly used. The method involves raising the temperature of the system followed by slow cooling. This makes it possible to overcome local minima as energy barriers, which would be impossible to cross in room temperature can be crossed.

One software commonly used to calculate NMR data are X-PLOR\(^{26}\) (Paper I). An extended template structure is used as input in the calculations and the NMR restraints are used in molecular dynamics simulated annealing (SA)\(^{24}\). This generally consists of a high temperature search phase when atoms are allowed to pass through each other, two cooling phases and a refinement phase.

After conducting a structure calculation it is equally important to back determine the quality of the structure, especially to check that the obtained structures are in agreement with experimental data. When analysing NMR structures, there is a large number of criteria that can be considered. Some important ones include:

- The number of restraints used in the calculation
- The number of restraints violations. This is generally conducted by choosing a certain threshold (typically somewhere between 0.1 and 0.5 Å for distance restraints, and 1 to 5 degrees for dihedral angle restraints) above which the violations are reported.
- R.m.s. deviations from experimental restraints.
- X-PLOR energies. (Assuming that this program was used) Divided into different groups such as NOE-, dihedral energies etc.
- Cartesian positional root-mean-square differences. Values are often calculated separately for backbone atoms and all heavy atoms. It can also be informative to calculate local rmsd's to illustrate local flexibilities or locally rigid regions by plotting the rmsd as a function of the residue sequence.
Stereochemical quality of the generated structures. This includes deviations from standard bond lengths and angles, hydrogen bonding and dihedral angle distribution.

Programs are available to perform these quality checks, for example PROCHECK-NMR⁷ and WHATIF.⁸

The assignment of NOEs is a time-consuming step in structure calculations. Protons often have the same chemical shifts and due to this they can not be assigned to a specific target. These NOEs are omitted from the calculation and unambiguous assigned NOEs are used to calculate the structure. Based on the structure obtained one tries to assign the NOEs discarded in the first iteration. Calculate a new structure, assign some more NOEs and so it continues. However, there are automated methods available for structure calculations for example ARIA (Paper IV).⁹,¹⁰ ARIA automates this process as it can assign the NOESY spectrum, based on a frequency list of the atoms. In addition, it identifies NOEs that does not provide structural information, either spectral artefacts or simple human error, and rejects them from the calculations. With ARIA it is possible to calculate structures using unambiguous NOEs which is done in an iterative manner, that is you specify how many calculations you want ARIA to perform. This greatly facilitates the process as instead of running a number of X-PLOR calculations you can run one calculation in ARIA (and take a coffee break).
# 4. Protein-ligand Interactions

NMR spectroscopy has become a valuable tool in the identification of ligand-target binding and offers some key advantages over other techniques. For example, NMR can detect weak ligand-target interactions. Unlike several bio-assays, NMR requires no knowledge about protein function to be studied. In contrast to other techniques used in structural biology, NMR enables the determination of binding constants. The ability of NMR to separate individual components allows the screening of complex mixture from natural sources or combinatorial chemistry. In addition to gaining information whether a compound is interacting with the target or not, structural information can be obtained for both the target and the ligand.

Numerous NMR parameters could change even upon temporary binding and those effects may be monitored either on the target or on the ligand itself, Figure 4.1. However, all are not equally valuable in the process of identifying binding compounds. Assuming the structure of the ligand changes upon binding, the scalar coupling can be studied to gain information about dihedral angles. On the other hand, these effects are almost always too small to be useful. Changes in chemical shift upon binding are expected, but the magnitude of the induced shift cannot easily be predicted, making the method less useful to make a clear identification of binding ligands. The most informative parameters in NMR screening are those that can be predicted with certainty. Upon interaction with a larger target molecule, the ligand will always experience different relaxation and diffusion rates compared to the free form, Figure 4.1.

![Figure 4.1. Consequences for NMR parameters of a ligand upon interacting with target protein.](image)
The use of NMR to detect ligands with affinity for targets can be performed in several ways, using techniques such as chemical shift perturbation, differential line-broadening, transferred NOE, and diffusion based methods. However, the detection of binding can be divided into two main categories, those that detect the signals from the protein or from the free ligand. There are advantages and disadvantages for both categories which are presented below with some of the techniques.

4.1. Ligand-observed Detection

As the name ligand-observed detection suggests, detection takes place on the free ligand which has several advantages compared to the target protein. As one observes the resonances of all small molecules in a mixture, binding ligands is easily identified using for example the inversion of NOE signals in a ligand interacting with a larger target. In addition, there is no requirement of labelled protein, no limitations in size for the target molecule or any requirement of knowledge about target structure. However, many of these techniques depend on size-dependent effects between ligand and target and can be difficult to resolve for weak binders and small relative size difference between ligand and target. In addition, the information for structure-based ligand optimisation obtainable through these techniques is more limited than for target observed techniques as structural information regarding the binding site is obtained using the latter. Ligand-observed screening was the method of choice in Paper II and was used to verify the interaction of YopD$_{278-300}$ and LcrH in Paper I.

4.2. Target-observed Detection

Observation on the target offers some advantages compared to ligand-observed screening. This method is not restricted by the ligands’ size or by an upper affinity limit. In addition, it reveals different binding sites making it possible to distinguish between specific and non-specific binding. However, in order use these methods, spectral assignment and the structure of the target has to be known. Another requirement is the use of isotope-labelled protein, although the amounts of protein can be
reduced by employing cryoprobe technology and screening larger mixtures of compounds. There are limitations in size for the target, typically ~40 kDa, incorporation of deuterium, prolonging relaxation times of heteronuclear signals with narrower linewidths as a result, can push this limit to ~60 kDa. The incorporation of TROSY techniques can, in favourable cases, further increase the limit to >100 kDa. It has become possible obtain some structural information for systems up to 900 kDa. Target observed screening was used in the Ber e 1 project (Paper IV) to map the lipid binding epitope using chemical shift mapping (see below).

4.2.1. Chemical shift Mapping

Studying induced chemical shift changes for the target upon ligand binding is an efficient method to identify the binding site. Chemical shift mapping is commonly based on the $^1$H-$^{15}$N HSQC, which yields a two-dimensional spectrum containing one cross peak for each amide in the protein. As a consequence the spectrum contains one signal for each amino acid except prolines. On addition of the ligand, the signals of those amides whose environments are affected by ligand binding will change position. After assignment, affected residues can be mapped upon the protein structure to reveal the binding site/s in the protein, Figure 4.2. However, this will include residues in direct contact with the ligand as well as residues that are affected indirectly by induced conformational changes in the protein. This approach has been widely used to study both protein–ligand and protein–protein interactions. In addition to screening, this method has been used later in the design process to ensure that new compounds interact with the same site as the ‘parent’ compound.
Studies of Protein Structure, Dynamics and Protein-ligand Interactions using NMR Spectroscopy

**Figure 4.2.** Concept of chemical shift mapping. Chemical shift changes are used as indicators of intermolecular binding. The resonances that experience an induced chemical shift upon addition of ligand are mapped in the protein structure, revealing the binding site/s of the compound.

This approach was used for the 2S albumin Ber e 1 where a crude lipid extract from Brazil nut was titrated to a sample of Ber e 1, Figure 4.3 (Paper IV). For clarity only two residues are shown. The resonances affected by the addition of ligand were then mapped unto the structure of Ber e 1, Figure 4.3.

**Figure 4.3** Chemical shift mapping of Ber e 1. HSQC experiment of Ber e 1 with and without lipid (left). Structure of Ber e 1 (right). Residues where chemical shift perturbations could be detected are illustrated in blue.
4.2.2. SAR by NMR

In 1996, scientists unveiled a powerful new technique in drug discovery. This method is called structure-activity relationships by nuclear magnetic resonance (SAR by NMR). SAR by NMR is in principle an extension of chemical shift mapping. Initially, a library of molecules is screened using chemical shift mapping to identify ligands with affinity for different sites in the target molecule. When these ‘lead’ molecules are found, analogues of these are tested in the same manner to optimize the affinity for each site. When two compounds with acceptable affinity, interacting at different sites in the molecule, have been identified, their location and orientation in the complex are determined experimentally by NMR methods. Finally, the two compounds are linked together to create a new molecule that will bind to the target protein more tightly than the two separate ligands, Figure 4.4.

![Figure 4.4](image)

Figure 4.4 Basic principles for SAR by NMR. The binding sites for two compounds are identified by chemical shift mapping. The affinity for each site is then optimized constructing analogues of the two identified ligands. Using the knowledge about target structure, a linker is synthesised connecting the two ligands, yielding a compound with higher affinity than the initial substances.

Traditionally, an enormous number compounds had to be synthesized and tested to find a lead compound that was active against the disease target. Then, this lead compound had to be optimized by synthesizing...
various structurally similar analogues. SAR by NMR drastically reduces the number of molecules that need to be made and tested, because the design of the linked compound is based on extensive information, both about the structure of the target and the small molecules that have a binding affinity for them. In this way, it is possible to build high affinity drug molecules out of pieces that have been shown to bind to the protein target.
5. Molecular Dynamics and Relaxation

In recent years an impressive number of three dimensional biomolecular structures have been solved by NMR and X-ray diffraction. Although this work has lead to a better understanding about structural architecture, reactions sites and biomolecular interactions it has become clear that a static structure is not sufficient to understand the function of biomolecules. There are many examples of that protein undergo significant motion on a variety of time scales and that processes such as enzyme catalysis, cellular signalling cascades, protein folding and protein-ligand interactions all involve motion that are essential for the biological function.47

It has been known since the early days of protein NMR that information on protein dynamics could be obtained. The effect of molecular motion on linewidths was used to explain the narrow lines observed in liquids as compared to solids.48 The concept expanded to study internal motions in solids and later in the determination of internal motions in proteins. The longitudinal relaxation ($T_1$), transverse relaxation ($T_2$) and NOE were developed for an idealized two-spin system$^{49,50}$ where the internuclear vector was assumed to reorient randomly and isotropically. However, it was realized that a vector attached to a tumbling sphere was not a realistic model in any molecular system. The model was extended to a tumbling ellipsoid and local motions were considered as the vector was allowed to move.$^{51,52}$ As the studied systems grew larger a more informative interpretation of relaxation were needed. Instead of merely reporting relaxation times a correlation between the measured relaxation parameter and dynamics of the system in question was needed. The link was provided by the “model-free” approach by Lipari and Szabo$^{53}$ which supplied information about fast internal motions (nano- to picoseconds dynamics).

Most commonly, three different relaxation parameters, $T_1$, $T_2$ and $^{15}$N-$^1$H-NOE, are used in the model-free approach. The dynamic parameters obtained from such an analysis are the generalized order parameter, $S^2$, which is a measure of local rigidity (where a value of 1 implies that the nucleus is completely rigid and 0 that it tumbles without any correlations) and the effective local correlation time, $\tau_e$, which measures the rate of the internal or local motions.

An analysis of the values of $S^2$ and $\tau_e$ with a reasonable model one has the opportunity to generate a physical picture of the motion, Paper IV.
6. Fluorine NMR

With a natural abundance of 100%, a sensitivity of 83% compared to that of proton and a high sensitivity of $^{19}\text{F}$ chemical shift to the surrounding environment, $^{19}\text{F}$ NMR could be an extremely valuable technique for NMR screening. As fluorine has very large chemical shift dispersion it should be possible to screen large libraries without chemical overlap becoming a significant problem. Due to the fact that fluorine is not naturally found in protein there are no background resonances, thereby simplifying the interpretation of the data. In addition, there are only a few resonances for each compound, dependent on the number of fluorine atoms, further simplifying the analysis. Fluorine labelled compounds is now readily available as building blocks in chemical synthesis and labelled amino acids for insertion in peptides.

Using $^{19}\text{F}$-labelled ligands the fluorine nucleus can be used as a probe to study protein-ligand interactions.\textsuperscript{55, 56} Fluorine is advantageous to use compared to proton as it has a very large chemical shift dispersion making it possible to screen large libraries without the requirement of labelled protein and without chemical overlap becoming a significant problem as one only observe a few resonances for each ligand, depending on the number of incorporated fluorine atoms. Since fluorine is not naturally occurring in proteins background resonances are lacking, making the data easier to interpret.

As stated earlier, as a ligand interacts with a larger target it should move and tumble at a different rate and therefore display a change in line-broadening compared to the “free” ligand. Unlike proton NMR, induced $^{19}\text{F}$ chemical shift is often a more accurate method to identify binding ligands due to the high sensitivity of fluorine to the surrounding environment. Fluorine NMR could be a valuable tool to screen large libraries using the $^{19}\text{F}$ linewidth and chemical shift to directly identify binding ligands, \textbf{Paper II}. In addition, as $^{19}\text{F}$ has the same spin as $^{1}\text{H}$, nuclear Overhauser effect between these nuclei could be one way to identify binding sites or binding epitopes.

In addition, working with fluorine labelled drugs, $^{19}\text{F}$ can be used clinically as a probe to study the pharmacokinetics as well as metabolism of drugs.\textsuperscript{57}
7. Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a technique for monitoring macromolecular interactions in real time. SPR can provide information about affinity, binding kinetics, concentration determinations and binding specificity. The principle of SPR is illustrated below in figure 7.1.

**Figure 7.1:** Principle of SPR. A molecule is immobilized on a dextran matrix coupled to a gold surface. The interaction partner is injected into the flow cell. Monochromatic light is totally reflected on the gold layer and the reflected light displays a minima at a defined angle, the resonance angle. Binding interaction partners changes the refractive index near the surface layer and the resulting shift of the intensity minima is a measure of the amount of bound molecules.

SPR allows real time measurement of binding kinetics between two or more molecules, by measuring the changes in refractive index at the surface layer of the sensor chip, caused by binding. During a binding analysis SPR changes occur as a solution is passed over the surface of a sensor chip. To perform an analysis, one molecule is immobilized on a sensor surface. The interaction partner is injected over this surface in a precisely controlled flow. Fixed wavelength light is directed at the sensor surface and biomolecular binding events are detected as changes in the particular angle where SPR creates extinction of light. This change is measured continuously to form a sensorgram, which provides a complete record of the progress of association or dissociation of the interactants. Binding events can be monitored between two or more molecules, between proteins, peptides, nucleic acids, carbohydrates, lipids and even low molecular weight molecules, such as drug candidates or signaling substances. SPR was used to verify NMR screening of ligands for the chaperones PapD and FimC (Paper II) and proved to be a good complement.
8. Circular Dichroism

Circular Dichroism (CD) is a well established technique for studying protein structure in solution. CD is observed when optically active matter absorbs left and right hand circular polarized light slightly differently. The CD spectra of proteins are generally divided into three classes of wavelength, Figure 8.1. These are (1) the far UV (below 250-nm), where the peptide contributions dominate, (2) the near UV (250–300-nm), where aromatic side chains contribute, and (3) the near UV–visible region (300–700-nm), where extrinsic chromophores contribute. CD spectra are sensitive to the secondary structure present in the molecule, Figure 8.1. The analysis of CD spectra are mainly used to yield valuable information about the secondary structure of biological macromolecules. As the structure elements are different CD spectroscopy is often used to estimate the amount and/or type of secondary structure in peptides and proteins. CD is also a useful method when studying the stability of macromolecules by chemical denaturants or temperature stability. One drawback of CD is that it reports a mean value for all structures in a sample. Complex samples where one observe changes in the CD spectrum could be all of the species being affected or only one. CD was used to investigate the secondary structure of peptides (Paper I). In addition, the method was also used to optimize conditions for NMR spectroscopy as it is a fast method and small amount of material is needed (in comparison to NMR).

Figure 8.1. Far UV CD spectra associated with various types of secondary structure. Solid line, α-helix; long dashed line, anti-parallel β-sheet; dotted line, type I β-turn; short dashed line, irregular structure.
9. Results and Discussion

Paper I

The protein YopD is a crucial component in the type III secretion system (TTSS) during a *Yersinia* infection as it is essential for the injection of Yop-effector proteins into target cells. While the mechanism of YopD is unknown, it is dependent on the interaction with the chaperone LcrH to maintain its function. This interaction is essential for stabilization and efficient secretion of YopD.

YopD possesses LcrH binding domains in the N-terminal and in an amphipathic domain near the C-terminal. As full length YopD has been shown to aggregate we chose to utilize peptides representing the amphipathic domain for structure characterisation.

<table>
<thead>
<tr>
<th>YopD&lt;sub&gt;278-292&lt;/sub&gt;</th>
<th>DNFMKDVLRLIEQYV</th>
</tr>
</thead>
<tbody>
<tr>
<td>YopD&lt;sub&gt;271-292&lt;/sub&gt;</td>
<td>EEAMNYNDNFMKDVLRLIEQYV</td>
</tr>
<tr>
<td>YopD&lt;sub&gt;278-300&lt;/sub&gt;</td>
<td>DNFMKDVLRLIEQYVSSHTHAMK</td>
</tr>
</tbody>
</table>

CD experiment proved YopD<sub>278-300</sub> to contain α-helical elements while the C-terminal truncated peptides were unstructured, indicating the importance of the C-terminal part of the peptide for structural stability. Initial NMR experiments on YopD<sub>278-300</sub> in aqueous solution experienced severe aggregation problems. However, these problems were overcome by the use of TFE which was used to break up the aggregate. A concentration of 40% was used in the experiments with a temperature of 40 °C to minimize aggregation. Structure calculations revealed a well defined helical structure for the peptide, Figure 9.1. In addition, a β-turn was identified involving residues Val292 to His295 which appear critical for the stability of the helical structure as the peptides lacking those residues were unstructured. Nor can a well defined helical structure be induced in those truncated peptides by the addition of TFE.

![Figure 9.1] Superposition of the backbone atoms for the 25 lowest energy structures. Aligned for the best overlap for residues 280-295.
However, this stabilisation of the helical structure may be through intermolecular interactions. Several observations suggest the peptide to form smaller aggregates even in 40% TFE. NOEs were found between the aromatic residues of Phe280 to Ile288 and Val292, this NOE distance is too large to be explained by intramolecular interactions. In addition, $^1$H $T_1$ and $T_2$ relaxation experiments resulted in a rotational correlation time, $\tau_c$, of 8 ns is not consistent with a monomeric structure. These findings indicate that the peptide aggregate through hydrophobic interactions, possibly as a dimer or a tetramer, Figure 9.2. There are indications of stabilization of the helical structure through formation of peptide aggregates. As the temperature is lowered, relaxation measurements indicate an increase in aggregate size and an upfield shift are observed for the $\alpha$-protons of residues involved in helical structure. An upfield shift of $\alpha$-proton chemical shifts are indications of inducement of helical structure, and could in this case be interpreted as the formation of a more well defined helical structure.

The interaction between the peptide and the chaperone LcrH was investigated by NMR methods. Relaxation-edited NMR experiments proved this peptide to interact with LcrH. In addition, two residues Tyr291 and Val292, had a significant induced chemical shift upon the addition of LcrH suggesting them to be directly involved in the binding to the protein.
Studies of Protein Structure, Dynamics and Protein-ligand Interactions using NMR Spectroscopy

Paper II

The chaperones PapD and FimC from uropathogenic *Escherichia coli* were used as target proteins in the study. The chaperones are involved in the assembly of hair-like protein structures termed pili on the surface of the bacterium. Functional pili and thereby functional PapD/FimC are essential for the ability of *E. coli* to adhere to host cell tissue, making these chaperones potential targets for developments of new antibacterial agents.

The purpose was to identify ligands with affinity for these proteins using $^{19}$F NMR screening. A library of various $^{19}$F labelled compounds, 2-Pyridinones and amino acid derivatives were used in this study, along with two control substances, Figure 9.3.

![Figure 9.3](image)

Due to the large chemical shift dispersion of fluorine, and the fact that only one resonance is observed for each compound, it is possible to screen large libraries without chemical overlap becoming a significant problem as compared to proton NMR, Figure 9.4. In addition, as $^{19}$F is not naturally occurring in proteins there are no background resonances, simplifying the analysis of the data.
The use of fluorine NMR proved to be a simple method for identifying ligands with affinity for the chaperone PapD as a significant line-broadening were obtained for several compound proving them to interact with the chaperone. The equivalent analysis was done for the chaperone FimC showing that the compounds with affinity for PapD interact with FimC as well. However, all ligands appear to have less affinity for this protein as a decrease in induced line-broadening as well as induced chemical shift are observed throughout the spectrum.

The analysis of the data for competitive binding indicates that it is minimal if present at all. Rather, several ligands appear to bind more and
more for each titration step indicative of several binding sites in PapD and FimC, or even non-specific binding.

Although indicating non-specific binding, the results obtained for binding to PapD using $^{19}$F NMR screening, in general, agreed very well with independent binding studies performed using surface plasmon resonance. Screening by $^{19}$F NMR spectroscopy should therefore be a valuable addition to existing NMR techniques for investigation chemical libraries in medical chemistry projects.

**Paper III and IV**

Paper III and IV present the solution structure of the allergenic 2S albumin Ber e 1. Ber e 1 has previously been identified as the major food allergen in Brazil nut, however recent studies have demonstrated endogenous Brazil nut lipids are required for optimal immune response *in vivo*. This finding is in contrast to previous experiments with known food allergens and raises questions about the approach for identifying potentially allergenic proteins. Consequently a high resolution structure of Ber e 1 is crucial for further understanding the allergenic properties of Ber e 1 and the possible interaction with lipids.

NMR data was collected for $^{15}$N and $^{13}$C/$^{15}$N labelled Ber e 1. The assignment of Ber e 1 is presented in Paper III and the secondary elements predicted from chemical shift and dihedral angles derived from HNHA experiment are comparable with the secondary structure of the homologous 2S albumin 8 from sunflower seeds.

The structure calculations were made with the software ARIA using a partially assigned $^{15}$N-edited NOESY and two unassigned $^{13}$C-edited NOESY spectra, one optimized for the methionine side chains. Dihedral angles and J-couplings supplemented the NOEs in the calculation.

The obtained structure differs compared to structures of other 2S albumins in that the core of the protein is not a helix bundle. Instead the C-terminal region is participating in the core and anchors a loop structure between helix 2 and 3 believed to be mobile. To verify these results we used $^{15}$N relaxation experiments followed by model-free simulations. The order parameters ($S^2$) are low for the loop with the exception of Arg 40 which is the residue having long range NOEs to the C-terminal part of
Ber e 1. The restricted mobility for this loop is a unique feature compared to other 2S albumins and needs to be investigated further.

The requirement of lipids for an optimal immune response to Ber e 1 was investigated using lipid extracts from Brazil nut, purified by our collaborators in UK. The chemical shift perturbations in Ber e 1 upon the addition of lipids were followed by HSQC experiments. Residues affected in chemical shift were mapped onto the structure of Ber e 1 (see Figure 5.3). The mapped residues cluster in 4 different regions: the N-terminal region (residue 2 and 8), the hypervariable loop (residue 32-37) the methionine rich helix 4B (residue 63, 67-69) and the C-terminal region (residue 107-109). All four structural clusters align along a cleft in the structure formed by helix 1-3 on one side and helix 4-5 on the other side. This cleft is big enough to encompass a lipid molecule. It is therefore tempting to speculate that this cleft can be the lipid binding epitope in the Ber e 1 structure.
10. Conclusions and perspectives

NMR has evolved as a powerful tool in structural biology and although NMR can never compete with X-ray crystallography in the speed of structure determination, it can provide additional information regarding for example molecular dynamics and competitive binding of which I have tried to illustrate in this thesis.

A conformational study of different peptides was described in paper I using mainly NMR spectroscopy and dynamics. More knowledge about the structure of the peptide would be of interest to determine which factors are important in stabilising the helical structure of the peptide. Our data suggest the turn to stabilise the helical structure but intermolecular interactions may be of equal importance. Attempts to obtain acceptable $^1$H NMR spectra of LcrH were unsuccessful, however with the use of labelled protein this may be possible to achieve, and if so map the binding site of YopD$_{278-300}$ in LcrH.

$^{19}$F NMR proved to be an efficient screening method compared to proton NMR in paper II as binding ligands are easily identified through induced line-broadening compared to the free form. However, our result indicate the compounds used in this study to have poor specificity for PapD and FimC as hardly no competitive binding is observed together with a small induced chemical shift.

The structural features of Ber e 1 are interesting as they do not match any other 2S albumin structure. The fold is different and the presence of an anchored loop is unique as well. From chemical shift mapping a potential binding epitope for the lipid was identified in the cleft of Ber e 1. This will be further investigated using pure lipid fractions from Brazil nut.

The project concerning Ber e 1 is, as always, ongoing.
11. Acknowledgements

Jag vill framförallt tacka mina handledare i Umeå vilka varit ett par under den här tiden. Dom står helt enkelt angivna i den ordning de ”avverkats”.

1. Hans Wolf-Watz

2. Ingmar Sethson. Som tack vara sina stora pedagogiska talanger (lite ironi) gjorde mig intresserad av NMR spektroskopi. Tack för all hjälp under den här tiden, alla intelligenta idéer och förslag, med vissa reservationer som tex (här har jag ett doktorandprojekt till dig Tobias, du kan få renovera min vind).


4. Jürgen Schleucher som varit både min handledare och biträddande handledare. Tack för all hjälp med framför allt NMR experiment, hade inte du varit där hade jag haft sönder NMReden för länge sen. All hjälp med NMR teori samt att du alltid ser till att parken fungerar som den ska.

5. Göran Larsson som var min handledare den sista tiden. Tack för all hjälp med strukturberäkningar, tolkning av data etc (kul att du alltid har pipe script som man kan köra och få ”snabba” resultat). Ett stort tack för allt hjälp med relaxationsbiten. Samt naturligtvis för att du är en jättebra kompis, speciellt när man vill låna verktyg, stegar, tapetserarbord, släp etc☺

Thanks to Matt Francis at the department of Molecular Biology for all the help during the YopD collaboration. For doing all the things I was happy to miss out on, protein expression, purification etc and for correcting my English ”grammar” in paper I.


Alla anställda på organisk kemi som gjorde det till en trevlig plats att jobba på.
Min före detta rumskompis nere i bunker n, Mattias Hedenström, för ett trevligt NMR samarbete och skitsnack på rummet.

Det gamla gänget med Pjocken, Hitman, Beizan & Mattis, tack för att ni gjorde det så kul att vara i kalla norrland.

Peter Stenlund som jobbade på Biokemi för värdefull hjälp med CD mätningar.

Alla anställda på medicinsk kemi och biofysik som gör att det är roligt att gå till jobbet.

Min gamla rumskompis Tobias Sparrman. Jürgens lilla lärling som också ser till att NMR parken är igång. Eftersom du undervisat mig i NMR spektroskopi så får jag som vanlig dödlig (de som inte läser Journal of Chemical Physics på skithuset) be om ursäkt för eventuella förenklingar/felaktigheter i NMR relaxationen.

HV71 för all underhållning. Är inte lika spännande i år när alla andra lag är så usla men vad kan man begära när bottenskrap som Skellefteå är i högsta serien.

Resten av NMR gruppen Björn, Katja, Beavis & Butthead (samt resten av Butheads personligheter, har inte namn på dom och jag har dessutom bara 2 sidor att skriva på)

Thomas Borén som fixade en espressomaskin till avdelningen, årets höjdpunkt.

Whiskyklubben Uishge Beatha. För alla trevliga möten med många goda whisky sorter, med undantag för vissa vietnamesiska häxblandningar, tur att Hasse fanns och kunde hinka i sig det (eller var det någon som stal det ur hans bil Hmmmm).


Skulle även vilja tacka familj och vänner för allt annat roligt som inträffar utanför arbetstid. Ibland ser man trots allt indikationer på att det finns en värld utanför universitetet

Stort tack till Viktoria, för att du stått ut med mina arbetstider under denna tid och är min personliga lilla väckarklocka.

Och sist och minst (fast ganska tjock). Gustav för att du hållit mig vaken om nätterna så jag hunnit skriva ihop detta. Det innebär dessutom att alla stavfel och andra grodor inte är mitt fel.
12. References


Studies of Protein Structure, Dynamics and Protein-ligand Interactions using NMR Spectroscopy


41. williamson, r. a.; carr, m. d.; frenkieL, t. a.; feeney, j.; freedman, r. b., mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR Chemical Shift Perturbation. Biochemistry 1997, 36, (45), 13882-13889.
42. lian, l. y.; barsukov, i.; golovanov, a. p.; hawkins, d. i.; badii, r.; szé, k. h.; keep, n. h.; bokoCh, g. m.; roberts, g. c., mapping the binding site for the GTP-binding protein Rac-1 on its inhibitor RhoGDI-1. Structure. Fold. Des. 2000, 8, (1), 47-55.