Technische Universität München
Institut für Organische Chemie und Biochemie

Max-Planck-Institut für Biochemie
Abteilung Strukturforschung (NMR-Arbeitsgruppe)

Monitoring the effects of antagonists on protein-protein interactions with NMR spectroscopy and structural characterization of the major intermediate in the oxidative folding of the leech carboxypeptidase inhibitor

Loyola D'Silva

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. St. J. Glaser
Prüfer der Dissertation: 1. Univ.-Prof. Dr. H. Kessler
2. apl. Prof. Dr. L. Moroder

Die Dissertation wurde am 01.12.2005 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 10.01.2006 angenommen.

Acknowledgements

First and foremost I would like to thank Prof. Dr. Horst Kessler for being my Doktorvater, and Prof. Dr. Robert Huber for giving me the opportunity to work as a graduate student in his department.

I wish to express my deepest gratitude to Dr. Tad A. Holak, my thesis advisor, for supporting my research from start to finish. He has consistently guided me on how to explore the deep issues while still keeping the broad perspective. Over the years with him, I have learnt about the problems that are interesting and challenging to pursue. This thesis would not have been possible without his guidance and wisdom.

My PhD experience would have been less satisfying without the company of the NMR group members (*the Holakies, as they are called*), both past (Till, Igor, Madhu) and present (Mahavir, Sudipta, Marcin, Joma, Grzegorz, Ulli, Przemek, Ola and Tomek). Thanks to Joan for the LCI project. Each of my projects has benefited from the inputs and the expertise of the others in the group.

A special thanks is due to friends who provided support from outside: Bhumi, Shruti, and Sathish. Klaus and Ingo for the music sessions in the "Glass House".

I like to thank my family, specially my parents and my in-laws for their love, support and understanding through these years. I hope that they will find some pride in my achievements through this thesis.

With the completion of this thesis, I am again at a crossroad in my life. I look back with a sense of pride and satisfaction, mixed, with a sense of renaissance. I think of the many people who came into my life, some by choice and some by chance, people who shaped and influenced what I am today. For better or worse, it certainly has been an experience.

Finally, a special thanks to Joma, my wife, for being there.

Publications

Parts of this thesis have already been published or will be published in due course:

Loyola D'Silva, Przemyslaw Ozdowy, Marcin Krajewski, Ulli Rothweiler, Mahavir Singh and Tad Holak. *Monitoring the effects of antagonists on protein-protein interactions with NMR spectroscopy*. J. Am. Chem. Soc., 127 (38), 13220-13226, 2005.

Joan L. Arolas, **Loyola D'Silva**, Grzegorz Popowicz, Francesc Aviles, Tad Holak, and Salvador Ventura. *NMR structural characterization and computational predictions of the major intermediate in the oxidative folding of the leech carboxypeptidase inhibitor.* Structure, 13 (8), 1993-1202, 2005.

Marcin Krajewski, Przemyslaw Ozdowy, **Loyola D'Silva**, Ulli Rothweiler and Tad Holak. *NMR indicates that the small molecule RITA does not block the p53-MDM2 binding in vitro.* Nature Medicine, 11(11), 1-3, 2005.

Igor Siwanowicz, Grzegorz Popowicz, **Loyola D'Silva,** Joma Joy, Sudipta Majumdar, Magdalena Wisniewska, Sue Firth, Louis Moroder, Robert C. Baxter, Robert Huber, and Tad Holak. *Molecular architecture of the insulin-like growth factor binding proteins (IGFBPs).* J. Biol. Chem., submitted, 2005.

Marcin Krajewski, Ulli Rothweiler, **Loyola D'Silva**, Sudipta Majumdar, Christian Klein, and Tad Holak. *Targeting ligand-protein and protein-protein interactions in the NMR competition binding experiments* (manuscript in preparation).

Scope of this work

This thesis is a summary of projects conducted at the Department of Structural Research at the Max Planck Institute for Biochemistry from 2002 until 2005. NMR is not only an excellent tool to select proteins amenable to structural analysis, but also to screen for inhibitors of protein-protein interactions.

Chapter 1 gives an overview of the use on NMR spectroscopy in screening of small molecules as applied in the field of drug discovery.

Chapter 2 describes a new methodology to monitor the effects of antagonists on protein-protein interactions with NMR spectroscopy, in particular the p53-MDM2 system. This methodology combined with the "***power of NMR***" shows how certain small molecules that were claimed to inhibit certain protein-protein interactions, using other techniques, did not in fact do so.

NMR is the most widely used technique for solution structure determination of small and medium sized proteins. Chapter 3 describes the use of this method to study the structure of the intermediate in the oxidative folding of the leech carboxypeptidase inhibitor.

Contents

Chapter 1

Introduction

1.1 Landmarks in NMR spectroscopy

Eleven years after NMR was originally identified as an experimental technique (Bloch, 1946; Purcell et al., 1946), the first NMR spectrum of a protein was recorded, (Saunders et al., 1957) and the spectra of amino acids were analyzed in 1957 (Takeda and Jardetzky, 1957). By this time, chemical shift had been discovered (Arnold et al., 1951; Knight,

1949) and coupling had been observed (Gutowsky et al., 1951; Hahn and Maxwell, 1951), but it took another twenty five years before the technology had advanced to the point where the spin systems of an entire protein could be assigned (Wagner and Wuthrich, 1982). The complete assignment of a protein however, required fundamental methodological advances and substantial increase in resolution and sensitivity over the original 40 MHz spectrum of ribonuclease recorded by Saunders. Key advances included the development of Fourier transform spectroscopy, two-dimensional spectroscopy, superconducting magnets, and new computational methods. The additional insight that Nuclear Overhauser Effect (NOE) at short mixing times could be used to obtain interproton distances within a protein, made the determination of a complete solution structure of a protein by NMR seem feasible (Gordon and Wuthrich, 1978; Wagner and Wuthrich, 1979). However, the actual completion of a structure had to await the development of distance geometry methods that could use the distance data to derive the fold of a protein (Havel and Wuthrich, 1984). Although NMR spectroscopy had been applied to some biological studies of DNA hydration and metal chelation while still in its infancy, it began to make major contributions to biology only when it developed the capacity to solve macromolecular structures.

Since the publication of the first complete solution structure of a protein in 1985 (Williamson et al., 1985), tremendous technological advances have brought nuclear magnetic resonance spectroscopy to the forefront of structural biology. Innovations in magnet design, electronics, pulse sequences, data analysis, and computational methods have combined to make NMR an extremely powerful technique for studying biological macromolecules at atomic resolution (Bax and Grzesiek 1993; Clore and Gronenborn,

1998; Sattler et al., 1999; Diercks et al, 2001). Most recently, new labeling and pulse techniques have been developed that push the fundamental line-width limit for resolution in NMR spectroscopy, making it possible to obtain high field spectra with better resolution than ever before (Doetsch and Wagner, 1998). These methods are facilitating the study of systems of ever-increasing complexity and molecular weight. As of September 2005, over 4000 NMR structures have been deposited in the protein data bank, which accounts for about 25% of all structures till date.

Biomolecular NMR spectroscopy has expanded dramatically in recent years and is now a powerful tool for the study of structure, dynamics, and interactions of biomolecules. Previous limitations with respect to molecular size are no longer a primary barrier, and systems as large as 900 kDa were recently studied. NMR spectroscopy is already well established as an efficient method for ligand screening. A number of recently developed techniques show promise as aids in structure-based drug design, for example, in the rapid determination of global protein folds, the structural characterization of ligand–protein complexes, and the derivation of thermodynamic parameters. An advantage of the method is that all these interactions can be studied in solution—time-consuming crystallization is not necessary.

From the point of view of structure-based drug design, structure determination is only a part of the process that gives rise to suitable lead compounds. The process of design, synthesis, evaluation, and structural analysis involves the determination of the structure of not only the protein target but also a significant number of ligand–receptor complexes. Moreover, a suitable screening method is required to identify “active ligands”, and ideally an understanding of the molecular basis of ligand affinity and

specificity is required. It is not surprising that given these requirements, a number of "bottlenecks" can occur in this entire process. With the advent of very powerful new methods based on NMR spectroscopy, the possibility now exists to overcome many of these difficulties (Bax and Grzesiek 1993; Clore and Gronenborn, 1998; Sattler et al., 1999; Prestegard, 2001; Diercks T. et al, 2001; Pellecchia et al., 2002).

1.2 Determination of protein structure

A number of very significant advances in both protein crystallography and biomolecular NMR spectroscopy have taken place in recent years. Apart from the problems with protein overexpression, it is very difficult to predict a timescale for the production of diffraction-quality crystals of a given protein, and in the case of certain membrane proteins it is possible that this may not be achieved at all. Although NMR-based approaches to protein-structure determination do not suffer from this limitation, there are a number of significant obstacles in the use of this technique in a robust and automated manner. Foremost amongst these is the need to assign a substantial fraction of the resonances observed in NMR spectra in order to generate a list of pairwise NOE distance restraints in order to determine the 3D structure of the molecule. It is well known that ideally ten or more such restraints per residue (amino acid) are required to obtain a "good" 3D structure. This restraints list will include side-chain- as well as backbone derived restraints, and thus by definition side-chain resonance assignments are required—a time-consuming task that is not well-suited to automation. A global fold of the protein is clearly insufficient for lead optimization, for which atomic-resolution structures are generally required, low resolution structures may be adequate for "SAR by NMR" type

screening, in which the protein is assigned but no 3D structure is available. A lot of literature has been published about using NOE constraints and Residual Dipolar Coupling (RDC) in order to determine the global fold of the proteins, which will not be discussed here (Kramer et al, 2004; Freudenberger et al, 2005).

1.3 Screening of ligands

Ligand design for HTS, validation and NMR screening is a complicated work. One reason for this is that, despite many advances in techniques for structure determination, our knowledge of the factors that govern the affinity and specificity of biomolecular interactions is very elementary. Fundamental thermodynamics tells us that the affinity is not governed by structure alone, but is a complex interplay between structure and dynamics. Although a high-resolution structure of the ligand–protein complex is unquestionably thought provoking in the context of ligand design, the probability that an effective lead compound can be designed on the basis of this structure is very small. One way to overcome this deficiency is to synthesize not one ligand, but many diverse ligands based on a common scaffold, that is, a focused combinatorial library. This gives rise to the need for a suitable screening assay. The NMR chemical shift is a probe of the chemical environment and a perturbation in chemical shifts is a sufficiently sensitive method to study biomolecular interaction. Based on this method, a number of effective NMR-based screening techniques have been described in recent years (Meyer and Peters, 2003). The two commonly practiced techniques are: monitoring protein resonances (chemical shift changes) and monitoring ligand resonances.

1.4 Monitoring protein resonances (Chemical shift changes)

The output from chemical shift perturbation studies is a collection of ^{1}H-^{15}N or ^{1}H-^{13}C correlation spectra. These chemical shifts are environment sensitive and are reporters for ligand binding events (Figure 1.1). Ligand binding induced chemical shift perturbations are easily observed in amide (^{1}H-^{15}N) resonances. Backbone amides in or near the site of ligand binding are particularly affected. Most of the ^{1}H-^{15}N resonances in a protein are the backbone amides, but side chains, such as Trp, Lys, His, Gln, Asn, or Arg, which are functionally important in a diverse range of proteins, also can be observed in ^{1}H-^{15}N spectra. Tryptophan side chains are especially good reporters due to their extreme downfield proton chemical shifts. The effect of binding on other sidechain resonances can vary widely. If the side-chain amide is not normally observed, it may become observable upon ligand binding, on formation of a hydrogen bond. Side-chain resonances which are directly involved in binding will be greatly perturbed by the binding event. Small or no perturbations are observed for those residues that are distant from the ligand binding site. If the binding is nonspecific or induces a conformational change, then the effect will be global, resulting in the shifting of a large majority of the resonances. This being said, the vast majority of peaks in a ^{1}H-^{15}N correlation spectrum are little affected upon ligand binding and add spectral complexity without providing substantive additional information. In the original "SAR by NMR" report, only 8 out of 107 resonances (7.5%) were noted as showing significant perturbations upon compound binding (Shuker et al., 1996). The complexity of ^{1}H-^{15}N correlation spectra increases quickly as the number of amino acids increases. For example, a single spectrum resulting from a moderate-size protein (30 kDa) can have 300 ^{1}H-^{15}N correlation peaks. The advent of TROSY increased

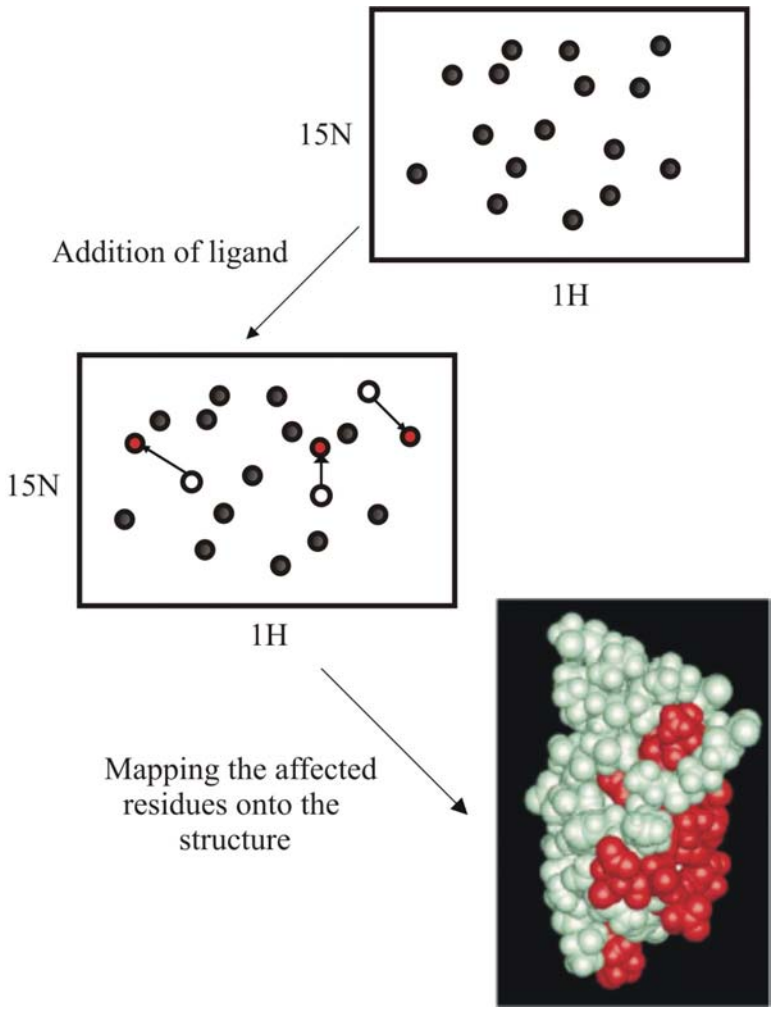


Figure 1.1: The principle of 'chemical shift mapping'. Resonances in a HSQC spectrum that are affected by ligand binding are shown on a surface representation of the protein. This commonly defines a 'patch' of surface residues that form the binding site.

the size limit of protein NMR from 400 amino acids to more than 1000 amino acids for a per-deuterated protein (Pervushin et al., 1997) or 1000 observable resonances. For simple cases of small proteins, it is easy to detect peaks that have shifted. Detecting a subset of peaks, 30 of 500 for a large protein, which shift upon addition of a compound can be difficult, especially if the spectrum is crowded (severe overlap); doing this for tens or

hundreds of spectra each with hundreds of peaks is almost impossible. A great deal of computational effort has been put forth to conquer this problem (Ross et al., 2000). Reduced spectral complexity using methods such as SEA-TROSY (Pellecchia et al., 2001) could speed up "SAR by NMR"; still only one protein can be screened at a time.

A related approach for ligand screening involves ^{13}C isotopic enrichment of methyl groups on protein side chains. Although ^{13}C isotopic enrichment would typically be considered prohibitively expensive for screening purposes, given the relatively large amounts of protein required, an alternative ingenious approach for the selective ^{13}C labeling of side-chain methyl groups by using inexpensive $^{13}CH_3I$ was devised (Hajduk et al., 2000). A major advantage of this $^{13}CH_3$-based screening is that $^{1}H,^{13}C$-HSQC is considerably more sensitive than $^{1}H,^{15}N$-HSQC for a given sample concentration. Typically, threefold higher sensitivity is obtained for proteins <30 kDa, enabling $^{1}H,^{13}C$-HSQC spectra to be acquired on a 50 μm sample of protein in 10 min. Moreover, the use of selective methyl labeling in combination with deuterium enrichment enables screening of protein targets in excess of 100 kDa. The availability of cryoprobe technology at intermediate to high fields promises to widen the scope for NMR based screening further, with the possibility of natural-abundance ^{13}C spectroscopy on the horizon.

1.5 Monitoring ligand resonances

The approach described above provides information about a ligand binding to the target protein only when added sequentially or in trivial cases when only one ligand is present in a given sample. Moreover, no information is obtained on the "active" chemical groupings of the ligands that are present, and those that interact with the binding site of

the target protein. However, this information can be derived with NMR spectroscopy by using NMR techniques that detect ligand resonances. One simple approach to detect which ligand undergoes binding is to exploit the difference in the relaxation properties or translational diffusion coefficient of the free ligand with that when it binds to the target protein. This can be achieved by using simple 1D NMR experiments designed to measure T_2 relaxation times or diffusion coefficients (Hajduk et al., 1997). A related technique exploits the fact that NOEs for bound ligands are typically large and negative whereas for free ligands they are very small (Meyer et al., 1997). In the fast-exchange regime, that is, relatively weak binding, these so-called transferred NOEs (TRNOEs) can readily be measured by using homonuclear 2D NMR techniques. Importantly, it is not necessary to isotopically enrich either the ligand or the protein in order to record these data. This is because it is possible to work with relatively high ligand/protein ratios (typically 15:1), whereby the relatively narrow ligand resonances can easily be observed above the much broader background protein resonances.

To determine the "active" regions of a given ligand, a simple 1D NMR experiment termed saturation transfer difference (STD) can be applied (Mayer and Meyer, 1999). This technique relies on the fact that in large complexes, proton magnetization is very efficiently transferred throughout the molecule or molecules owing to a process known as spin diffusion. Hence, if a region of the NMR spectrum of the complex containing resonances for the protein (but not for the ligand) is saturated (i.e. the magnetization is attenuated) by application of a radio frequency field, this saturation will be efficiently transferred through the protein and also to resonances of the ligand that are located within the protein binding site. This technique has proven successful at mapping

the active regions of ligands and can also be used as a simple screening tool. One situation in which the above approaches might fail concerns ligands that bind strongly to the protein of interest. Under these circumstances, the chemical-exchange phenomena on which the above methods depend are not on the appropriate timescales. However, under these circumstances, NMR "reporter screening" can be applied (Figure 1.2).

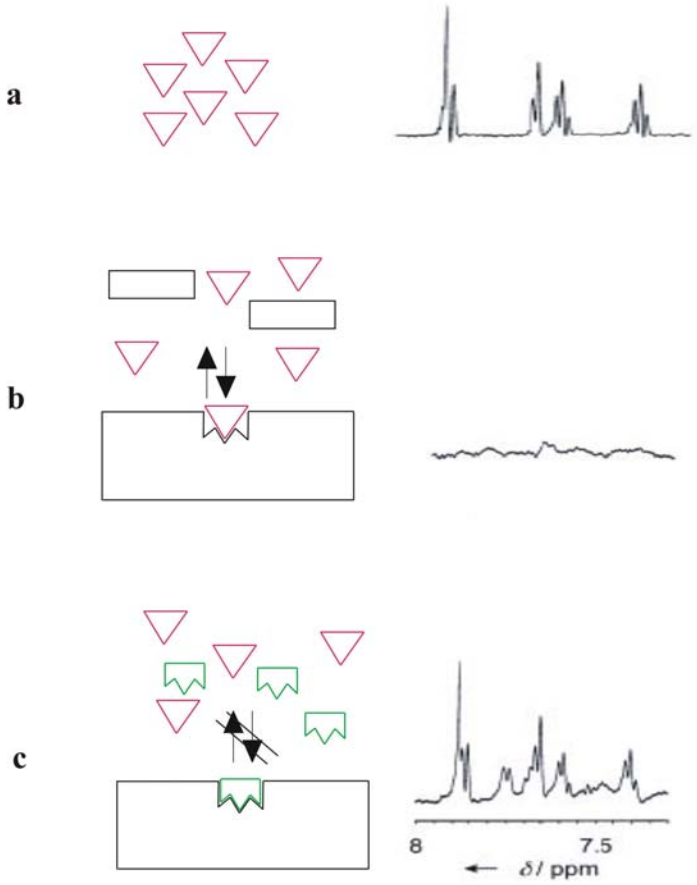


Figure 1.2: Principle of ligand (reporter) screening. NMR reporter screening detects the resonance signals of a reporter ligand (triangles) in the presence of target protein and test compounds by simple 1D proton spectra. a) Without protein target, the reporter ligand has sharp signals typical for a small molecule. b) Broad signals of the reporter ligand indicate that the reporter ligand is bound to the target, and that no test compound has comparable binding affinity. c) Sharp signals of the reporter ligand indicate unbound reporter ligand: At least one of the compounds binds more tightly to the protein target and thus displaces the reporter ligand.

The basis of this approach is straightforward: During acquisition of spectra of the free protein and of the protein in the presence of ligand(s), a "reporter-ligand" that is known to bind weakly to the protein is also included. When a ligand that binds more tightly than the reporter ligand is present in the sample, the binding event can be detected by changes in intensity and line-width of the latter—these will change as the fraction of free reporter ligand in solution increases as it is displaced from the binding site of the protein (Jahnke et al., 2002).

Recently, a new method called **RAMPED-UP NMR** (**R**apid **A**nalysis and **M**ulti**p**lexing of **E**xperimentally **D**iscriminated **U**niquely Labeled **P**roteins using **NMR**) has been described that generates simple spectra which are easy to interpret and allows several proteins to be screened simultaneously (***multiplexing of proteins***) (Zartler et al., 2003). In this method, the proteins to be screened are uniquely labeled with one amino acid type. There are several benefits derived from this unique labeling strategy: the spectra are greatly simplified, resonances that are most likely to be affected by binding are the only ones observed, and peaks that yield little or no information upon binding are eliminated, allowing the analysis of multiple proteins easily and simultaneously. This method should have significant impact in the use of NMR spectroscopy for both the lead generation and lead optimization phases of drug discovery by its ability to increase screening throughput and the ability to examine selectivity. To the best of our knowledge, this is the first time in any format that multiple proteins can be screened in one tube. This method however can be employed only with the aprior knowledge of the structure of the protein/target receptor in question or the complete ^{1}H-^{15}N assignment, so that selectively labeled proteins can be prepared for a suitable ligand. The only advantage in this method

being that the spectra are less crowded and more than one non-interacting protein mixtures can be studied at the same time for a given ligand or different ligands. However, a weak point of this method is the high protein concentration required for performing the screening.

Another new method called **T**arget **I**mmobilized **N**MR **S**creening (TINS) (Vanwetswinkel et al., 2005) has been developed that reduces the amount of target required for the fragment based approach to drug discovery. Binding is detected by comparing 1D NMR spectra of compound mixtures in the presence of a target immobilized on a solid support to a control sample. The method has been validated by the detection of a variety of ligands for protein and nucleic acid targets (K_D from 60 to 5000 μM). The ligand binding capacity of a protein was undiminished after 2000 different compounds had been applied, indicating the potential to apply the assay for screening typical fragment libraries. TINS can be used in competition mode, allowing rapid characterization of the ligand binding site. TINS may allow screening of targets that are difficult to produce or that are insoluble, such as membrane proteins. TINS works by immobilizing the sample on a solid support that is compatible with high-resolution, static NMR methods. Mixes of compounds from the library can then be applied and binding detected by comparison of a simple, 1D NMR spectrum with that of an appropriately prepared control sample. TINS can in principle be used to screen an entire library with a single sample of 3–5 mg of the target.

TINS makes use of NMR spectroscopy in order to study the ligand screening, by observing changes in the ligand resonances. These changes may be manifested directly as a change in the diffusion constant (Hajduk et al., 1997; Bleicher et al., 1998; Anderson et

al., 1999) or indirectly via an effect on the correlation time τ_c. Numerous techniques have been elaborated that detect changes in τ_c, including line broadening (Fejzo et al., 1999*)*, the presence of intramolecular transfer NOEs (trNOE) (Henrichsen et al., 1999), or intermolecular trNOEs (Chen et al., 2000; Dalvit et al., 2000), which are collectively referred to as small molecule NMR methods. These techniques alleviate the requirement for isotopically labeled protein while, at the same time, reducing the amount of target needed to screen a compound library. An additional benefit is that these techniques have the potential to directly identify the binder even in complex mixtures and can therefore alleviate subsequent deconvolution steps. Small molecule NMR methods do not provide the structural information that the SAR by NMR method does. However, they do provide at least a limited characterization of the ligand binding site with modifications that allow the assay to be performed in a competition mode (Dalvit et al., 2002), or when combined with molecular modeling (Bhunia et al., 2004).

Many small molecule NMR methods have been put to good use in actual drug discovery programs, yet they too have limitations with respect to the range of targets to which they can be applied. In particular, screening complete compound libraries using targets that are only available in low mg quantities (as is typically the case with integral membrane proteins) is not possible using presently available methods. One way to further reduce the amount of target required for small molecule NMR methods would be to use a single sample of the target to screen an entire compound library. This could be accomplished through equalizing the target and utilizing a flow-injection NMR instrument. The group of Meyer was indeed able to detect binding of one oligosaccharide from a mixture of seven in the presence of a lectin bound to controlled pore glass (CPG)

beads (Klein et al., 1999). However, in this study, it was necessary to use magic angle spinning (MAS) NMR in order to overcome the line broadening induced by the magnetic field gradients resulting from the difference in magnetic susceptibility of the glass beads and the surrounding aqueous solution. The requirement for MAS necessitates a batch mode that is not well suited to screening even moderately large compound libraries. TINS, allows one to record high-resolution spectra in the static mode. This arrangement should allow the assay to be performed in a flow through manner using only one sample of target. Target-ligand interactions with a wide range of affinities can be detected using TINS, which, like other small molecule NMR methods, can provide limited structural information when used in a competition assay with a known ligand.

1.6 Future perspectives

Fragment-based drug discovery is becoming an important additional tool in the search for fundamentally new lead compounds. This approach generally produces lead compounds that are chemically very different than those derived from high-throughput screening, which should result in higher success rates for finding new drugs. However, fragment-based methods have been limited to pharmaceutical targets that are readily available in large quantities.

New NMR experiments that exploit ^{2}H, ^{13}C and ^{15}N isotopic enrichment, combined with instrumentation improvements and increased magnetic field strengths, will continue to increase the size and scope of proteins and protein/ligand complexes that can be studied by NMR spectroscopic methods. As the molecular weight and the number of resonances increases, isotope-editing and isotope-filtering techniques will become

increasingly important in order to focus on the active site functional domains that may be more amenable to detailed NMR studies. The full range of biochemical and chemical methods to uniformly or selectively label ligands and/or active site residues will be required in order for NMR spectroscopy to contribute to structure-based drug design projects in a timely manner. The seemingly disparate techniques of structure based drug design, combinatorial chemistry and high-throughput screening will continue to become more intertwined. The capacity for NMR spectroscopy to enhance each individual technique and their synergistic application to a given biochemical problem presents exciting new opportunities to the NMR spectroscopist. NMR success stories will undoubtedly justify the further development of NMR resources in academia and the pharmaceutical industry. NMR spectroscopy should also begin to enhance the roles that genomics and bioinformatics play in drug discovery and design. Genomics and bioinformatics offer the potential to discover new therapeutic targets and to combine information already known about similar targets to jumpstart the drug discovery process. Starting with a novel target identified using a genomics approach, high-throughput screening of a molecularly diverse set of compounds created using combinatorial techniques can lead to the rapid identification of high affinity lead compounds. Subsequent structure-based drug design experiments, in the context of genomics and bioinformatics data, can then be used to enhance ligand specificity and pharmacological properties. Given the finite number of protein folds, homologous resonance assignment of new proteins based on sequence alignment with a previously assigned protein may facilitate structural studies. Assignment databases, such as BioMagResBank, will become increasingly useful resources for the NMR spectroscopist. Bioinformatics can also be

used to reduce the size of larger proteins to smaller functional domains that may be more amenable to detailed NMR studies.

1.7 Conclusions

There are now a large number of reports on small-molecule inhibitors of protein-protein interaction. It seems that the approaches for initial compound discovery- HTS, computational screening, fragment discovery- have many of the same advantages and encounter the same hurdles for protein-protein systems, for more traditional therapeutic targets, as opposed to the SAR by NMR technique. It is likely that as with traditional targets a variety of discovery approaches will be needed. On the basis of the data so far, good targets for small-molecule inhibition are those that have small hot spots that can be covered by a drug-sized molecule. In general, a novel molecule can be described as "validated" when it has been shown to bind with a 1:1 stoichiometry to the target of interest. Protein-protein interactions are clearly more challenging than drug targets that naturally bind small molecules. However there have been clear inroads into these targets. The drug discovery community understands more about what kinds of binding sites might be more tractable than others. In the recent years we have developed better ways of screening diversity space for new compounds that interact with these sites. This field is still in its infancy, but given the progress that has been made and the importance of this target class, it is likely to receive increased attention in the future.

The NMR screening studies for lead compounds concentrated so far on binary interactions of lead compounds with small to middle size domains of target proteins and very little has been done in order to understand the effect of inhibitors on protein-protein

complexes, especially by NMR spectroscopy. What happens when multiplexing of two proteins leads to a complex formation? How does one monitor the effect of inhibitors on such a complex? The principle purpose of an antagonist compound discovery, however, is determining whether a lead compound inhibits or dissociates protein-protein interactions. This thesis mainly deals with answering these issues. The next chapter describes a new NMR method that allows for direct monitoring of the influence of a ligand on protein-protein binding. We call this technique **DIMAPPS- DI**rect **M**onitoring of **A**ntagonists on **P**rotein-**P**rotein interaction**S**. Importantly, this method shows whether a small molecule compound is capable of releasing proteins in their wild-type folded states or whether it induces their denaturation, partial unfolding or precipitation.

Chapter 2

*DI*rect *M*onitoring of *A*ntagonists on *P*rotein-*P*rotein interaction*S* (*DIMAPPS*) by NMR spectroscopy

2.1 Introduction

Protein–ligand and protein–protein interactions are of primary interest in structural biology and drug discovery, and often provide clues to the physiological functions of newly discovered proteins. In this context, NMR spectroscopy has long been an attractive technique for characterizing both high- and low-affinity macromolecular complexes. A

common approach is to label the target receptor protein with the stable isotopes ^{15}N and/or ^{13}C, and observe the changes in chemical shift upon interactions (Markus et al., 1994; Shuker et al., 1996; Kriwacki et al., 1996; Prestegard et al., 2001; Dehner et al., 2003; Fernandez et al., 2004; Seidel et al., 2004; Pellecchia et al., 2002; Stockman and Dalvit, 2002; Coles et al., 2003; Lepre et al., 2004). In general, the aprior knowledge of assignment of the NMR spectrum of the protein would be helpful, or at least the nuclei that exhibit chemical shift changes should be known in order to map these interactions, although a method has recently been described that allows mapping interfaces of protein complexes without the knowledge of chemical shift assignments provided that the 3D structures are known (Reese and Doetsche, 2003). If the only purpose of the NMR experiment is detecting the binding, then the assignment is not needed. The NMR screening studies for lead compounds concentrated so far on binary interactions of lead compounds with small to middle size domains of target proteins. This chapter describes a new method to directly monitor the effects of antagonists on protein-protein interactions. As an example of this application we use the p53-MDM2 interaction.

2.2 A brief biological background on the p53-MDM2 system

The oncoprotein MDM2 (human murine double minute clone 2 protein) inhibits the tumor suppressor protein p53 by binding to the p53 transactivation domain (Oliner et al., 1992; Kussie et al., 1996; Lane and Hall, 1997; Lozano and Montes, 1998). The p53 gene is inactivated in many human tumors either by mutations or by binding to oncogenic proteins. In some tumors, such as soft tissue sarcomas, overexpression of MDM2 inactivates an otherwise intact p53, disabling the genome integrity checkpoint and

allowing cell cycle progression of defective cells (Momand et al., 1998; Juven-Gershon and Oren, 1999). Studies comparing MDM2 overexpression and p53 mutation concluded that these are mutually exclusive events, supporting the notion that the primary impact of MDM2 amplification in cancer cells is the inactivation of the resident wild-type p53. It has been shown that a peptide homologue of p53 is sufficient to induce p53-dependent cell death in cells overexpressing MDM2 (Wasylyk et al., 1999). The rescue of the impaired p53 function by disrupting the MDM2-p53 interaction offers thus new avenues for anticancer therapeutics (Figure 2.1) and several lead compounds have recently been reported to inhibit the p53-MDM2 interaction in assays based on tumor cell cultures or immunoprecipitation techniques (Chene, 2003; Fischer and Lane, 2004; Klein and Vassilev, 2004).

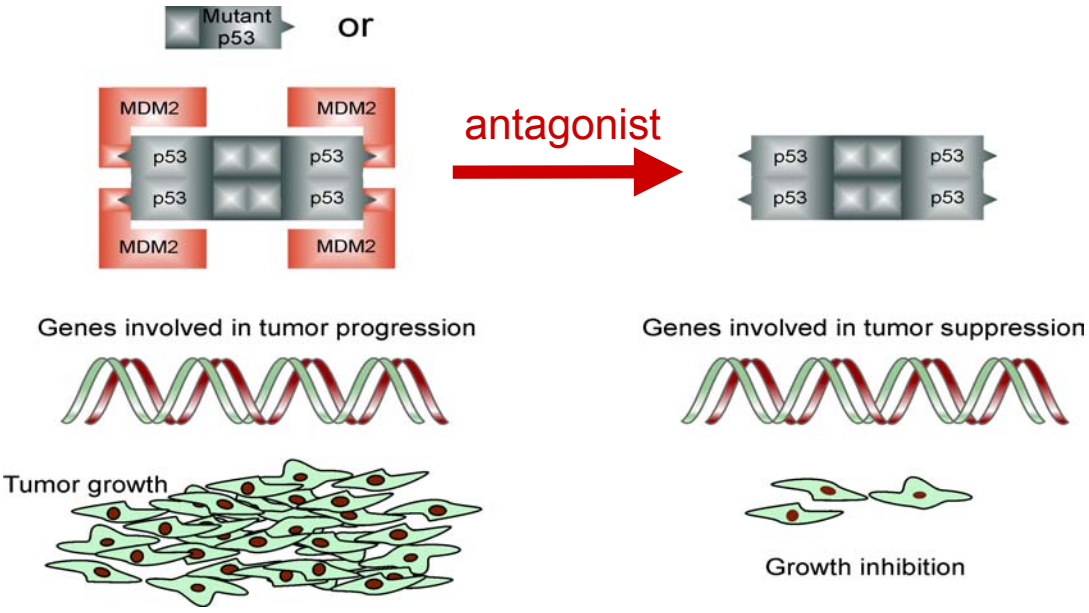


Figure 2.1: Releasing p53 from the MDM2 inhibition in order to carry out its function in tumor cells.

2.3 A methodology- DIMAPPS

This methodology is based on observation of a two-protein complex and the schematic is shown in Figure 2.2. In this study, two protein components that make up a complex are monitored, with one component being small enough (less than ca. 15 kDa) to provide a good quality HSQC spectrum after ^{15}N or ^{13}C labeling of the protein (Figure 2.2a). The size of the second component should be large enough so that the molecular weight of the preformed complex is larger than ca. 40 kDa. On forming a complex, the smaller protein no longer behaves like before, but begins to exhibit characteristics of the larger one. The observed $1/T_2$ transverse relaxation rates of the bound protein in the complex increase significantly and broadening of NMR resonances results in the disappearance of most of the cross peaks in the HSQC spectrum (Riek et al., 2000; Rehm et al., 2002; Wüthrich, 2003) (Figure 2.2b). In order to restore the spectrum of the smaller component, one would have to add an antagonist that would dissociate this complex, as shown in Figure 2.2c, whereas a non-binder would not affect the HSQC spectrum of the complex as seen in Figure 2.2d. However, a weak inhibitor could partially release the labeled protein when added in large excess. A simple calculation using standard formulas (Wang, 1995) shows that an inhibitor of K_D 5 μM would release about 40% of the protein at 250 μM concentration, assuming 100 μM each of proteins and a K_D of 0.7 μM for the complex.

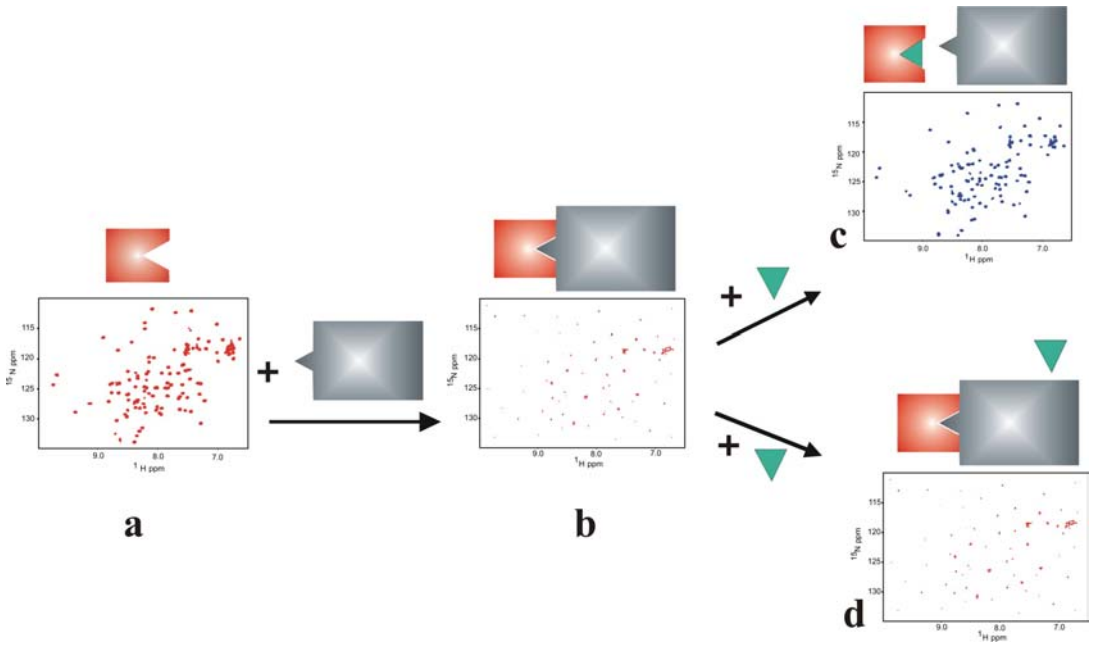


Figure 2.2: Schematic representation of our method for studying the effect of an antagonist on the interface between two proteins. (**a**) An ^{15}N HSQC spectrum of a ca. 10 kDa uniformly ^{15}N labeled protein (each amino acid gives a cross peak for the N-H pair. The side chain N-H resonances are observed at around 7 ppm ^{1}H and at 120 ppm ^{15}N chemical shifts). (**b**) The cross peaks disappear on addition of a large protein (ca. 35 kDa) that forms a complex with the smaller one. (**c**) The cross peaks reappear on addition of a strong inhibitor ligand that dissociates the complex. (**d**) A weak inhibitor does not dissociate the complex.

In order to test our methodology, we made use of the p53-MDM2 interaction. Structure based screens for this interaction utilized an N-terminal domain of MDM2 of ca. 100 amino acids and short peptides of p53 (Kussie et al. 1996; Blommers et al. 1997; Stoll et al., 2001; Vassilev et al., 2004; Fry et al., 2004). The primary binding site of MDM2 on the p53 protein has been mapped to residues 18-26 (Chen et al., 1993; Picksley et al., 1994; Bottger et al.; 1996; Bottger et al., 1997; Schon et al., 2002; Schon et al., 2004). A 118 amino acid N-terminal domain of MDM2 and the N-terminal 312-residue fragment of p53, encompassing the transactivation and DNA binding domains

was used in this study. Figure 2.3 shows the schematic of the full-length proteins used for this purpose. The binary complex has a total molecular weight of 45 kDa. Isothermal Titration Calorimetry (ITC) measurements indicated a K_D of 0.77 μM in the buffer solution used in the NMR experiments, which agrees well with that of ca. 0.7 μM reported in the literature (Kussie et al., 1996; Blommers et al., 1997; Stoll et al., 2001; Dawson et al., 2003; Schon et al. 2004; Fry et al. 2004).

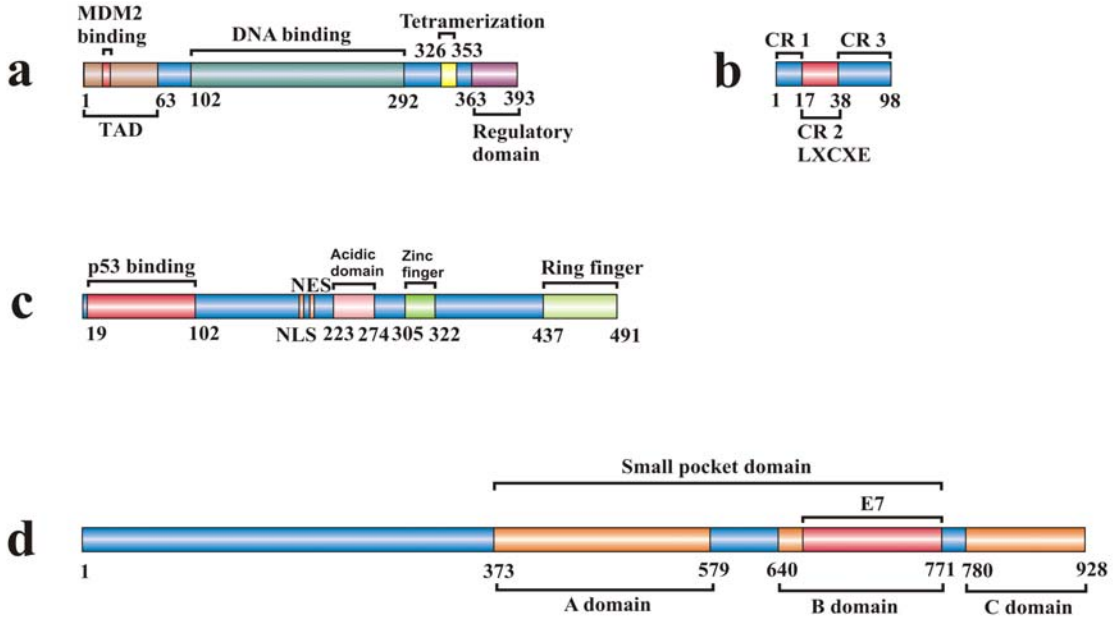


Figure 2.3: Schematic representations of the full-length proteins (**a**) p53, (**b**) HPV16 E7, (**c**) MDM2, and (**d**) pRb.

To start with, we decided to use MDM2 that is isotopically enriched with ^{15}N. Figure 2.4a shows the 2D ^{15}N-^{1}H spectrum of ^{15}N-MDM2. The complex formation was observed by the stepwise addition of unlabeled p53 to ^{15}N-MDM2. The disappearance of

most of the MDM2 peaks, as seen in Figure 2.4b indicates a complete complex formation between the two.

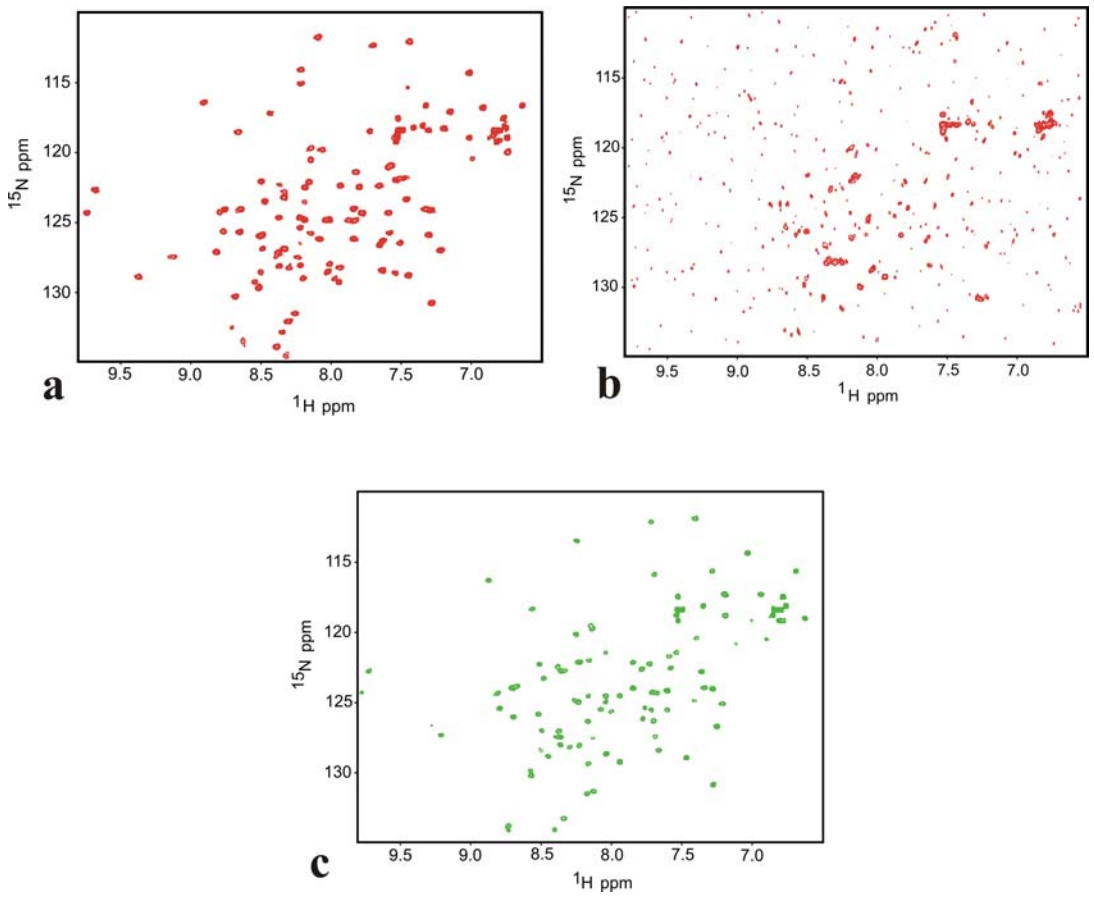


Figure 2.4: Spectra of the ^{15}N uniformly labeled MDM2. (**a**) ^{1}H-^{15}N HSQC spectrum of ^{15}N-MDM2. (**b**) ^{1}H-^{15}N HSQC spectrum of ^{15}N MDM2 complexed with p53, as seen, most of the cross peaks disappear. (**c**) ^{1}H-^{15}N HSQC spectrum of MDM2 in complex with nutlin-3. Some cross peaks are shifted (compared to Figure 2.4a) due to binding of nutlin to MDM2.

As expected, majority of the backbone ^{15}N-^{1}H resonances of the structured regions of the molecule broadened and/or disappeared, while the leftover peaks originate from flexible residues of the complex and/or free MDM2. Following our previous calculations, we found that, for the complex of MDM2-p53 having a K_D of 0.77 μM and

a protein concentration of 0.1 mM there is still 8.4% free MDM2 present and the residual observed sharp signals could arise from this free protein. All these signals are located in the spectrum at the "central 8.3 ppm NH amide" region, diagnostic for unstructured residues, plus flexible side chains at 7 and 7.5 ppm.

In the search for high affinity inhibitors, a recent research report showed that Nutlins (Figure 2.5), a class of cis-imidazoline compounds were indeed capable of antagonising the p53-MDM2 interaction (Vassilev et al., 2004). The most potent among them, however being nutlin-3, was reported to displace recombinant p53 protein from its complex with MDM2 and had an inhibitory concentration (IC_{50}) value of 0.09 μM (Vassilev et al., 2004*)*. An inhibitor with such high inhibitory activity should displace p53 from its complex with MDM2. In order to test this we added nutlin-3 in a stepwise manner to the MDM2/p53 complex and indeed found that it restores the MDM2 spectrum, as seen in Figure 2.4c, with the sites involved in binding to nutlin being however shifted. A 1D spectrum of the dissociated complex revealed that the freed p53 is folded (the core domain) as judged by NMR. The experiment also shows that the MDM2/nutlin complex is soluble, and that nutlin did not induce precipitation of MDM2.

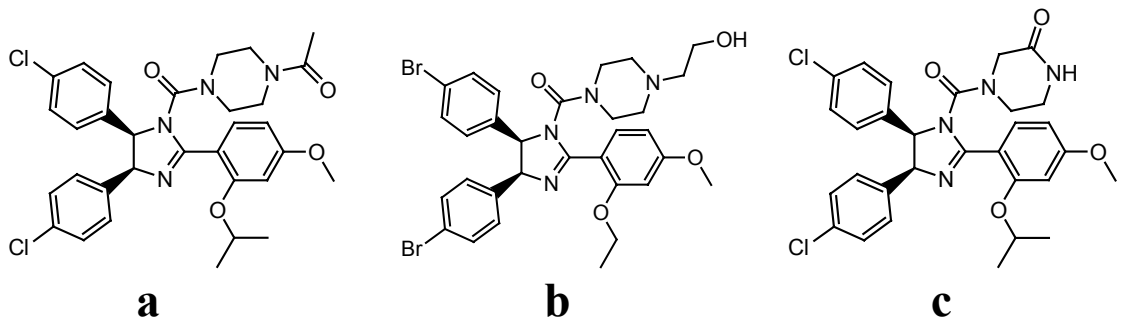


Figure 2.5: Nutlins, a class of cis-imidazoline compounds that was reported to antagonise the p53-MDM2 interactions with high affinity. (a) Nutlin-1, (b) Nutlin-2, and (c) Nutlin-3.

To be sure of these results and to establish our methodology, we performed experiments on ^{15}N-labeled p53 and unlabeled MDM2. Figure 2.6a shows the ^{1}H-^{15}N HSQC spectrum of the uniformly labeled p53.

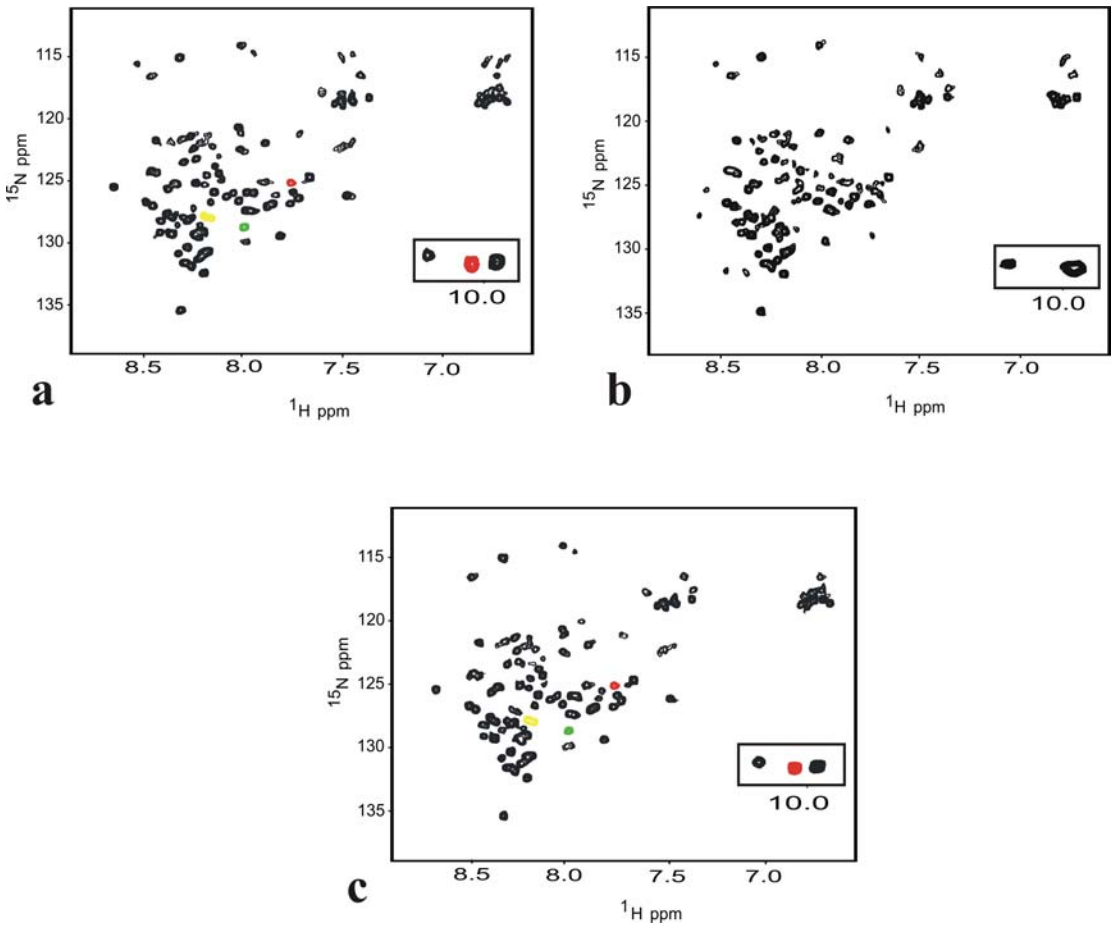


Figure 2.6: Spectra of the ^{15}N uniformly labeled p53. (**a**) ^{1}H-^{15}N HSQC spectrum of p53 (for assignments see (Ayed et al., 2001; Lee et al., 2000). The cross peak of Phe19 is shown in yellow and Leu26 in green. The two cross peaks in red are from Trp23, with the side chain cross peak at the ^{1}H chemical shift of 10.10 ppm (inset) (**b**) ^{1}H-^{15}N HSQC spectrum of ^{15}N-p53 in complex with MDM2. The complex formation is monitored by the disappearance of Phe19, Trp23 and Leu26 peaks. (**c**) ^{1}H-^{15}N HSQC spectrum of free p53 after addition of nutlin-3. The spectrum corresponds to that of Figure 2.6a, with all the "bound" peaks being restored, indicating the dissociation of the complex.

The spectrum indicates that the first 93 residues are flexible and mostly unstructured. These flexible residues are the ones that are mainly seen in the spectrum owing to their narrower line widths as compared to the p53 DNA core domain that have broad linewidths, and hence unobservable. The primary binding sites of p53 to MDM2 are known and the complex formation can be monitored from the NMR spectrum by observing the three binding sites (Phe19, Trp23 and Leu26 (Kussie et al., 1996)). The knowledge of the assignment of p53 helped in the process (Ayed et al., 2001; Lee et al., 2000). A stepwise addition of MDM2 resulted in the disappearance of these cross peaks and a complete disappearance indicated complex formation, as seen in Figure 2.6b. An interesting observation was that the N-terminal residues were still not structured when bound to the MDM2 domain, with the exception of a 10 residue-binding site (residues 17 to 26). As before, nutlin-3 was then added to the MDM2/p53 complex in a stepwise manner and the reappearance of the three peaks (Phe19, Trp23 and Leu26) was monitored. Figure 2.6c shows the spectra of ^{15}N-p53 and the reappearance of the three binding sites indicating that nutlin-3 dissociates the p53-MDM2 complex by binding to MDM2.

After the success with nutlins, we decided to check our methodology for compounds that had a considerable inhibitory effect on the p53-MDM2 interaction, but with lower affinity compared to nutlins. This would help us extend the methodology to low affinity inhibitors as well. To do this we made use of another small molecule compound, a sulfonamide compound NSC 279287, (Figure 2.7) that was recently reported to inhibit the MDM2/p53 interaction with IC_{50} of 32 μM (Galatin and Abraham, 2004) was used to check our methodology.

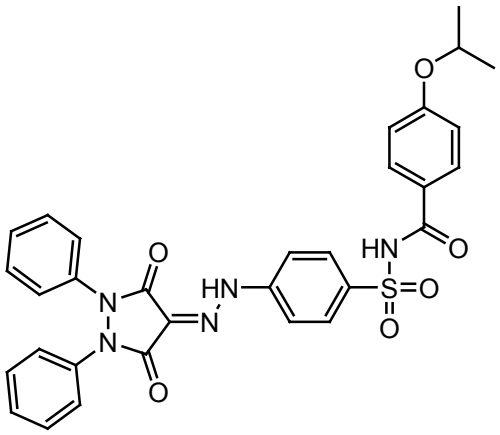


Figure 2.7: Sulfonamide compound (NSC 279287) was claimed to inhibit the p53-MDM2 interactions with an IC_{50} value of 32 μM (Galatin and Abraham, 2004).

Following the previous protocol for titrations, a stepwise addition of the sulfonamide to p53/^{15}N-MDM2 did not result in the recovery of the ^{15}N-MDM2 spectrum, but resulted in the release of the folded p53. The release of p53 could also be monitored from the 1D proton NMR spectra. Figure 2.8 shows the 1D spectra of the region where the side chain of Trp23 resonates at 10.10 ppm.

Before drawing conclusions about the inhibitory effect of sulfonamide we decided to test if preincubation of MDM2 with sulfonamide would help inhibit the interaction with p53. We added sulfonamide to the free ^{15}N-MDM2 with the intention of adding p53 after the preincubation of MDM2 with the compound, however we observed only very small, insignificant induced chemical shift changes in the ^{15}N HSQC spectrum of MDM2 on titration with sulfonomide. A closer inspection of the 1D proton spectra indicated that sulfonamide precipitated MDM2, as no signals from MDM2 were present in the final 1D proton spectrum. MDM2 begins to precipitate at about 0.3 mM sulfonamide concentration, which is about three times the protein concentration, and the precipitation was complete at about 1 mM, which is about ten times the protein concentration.

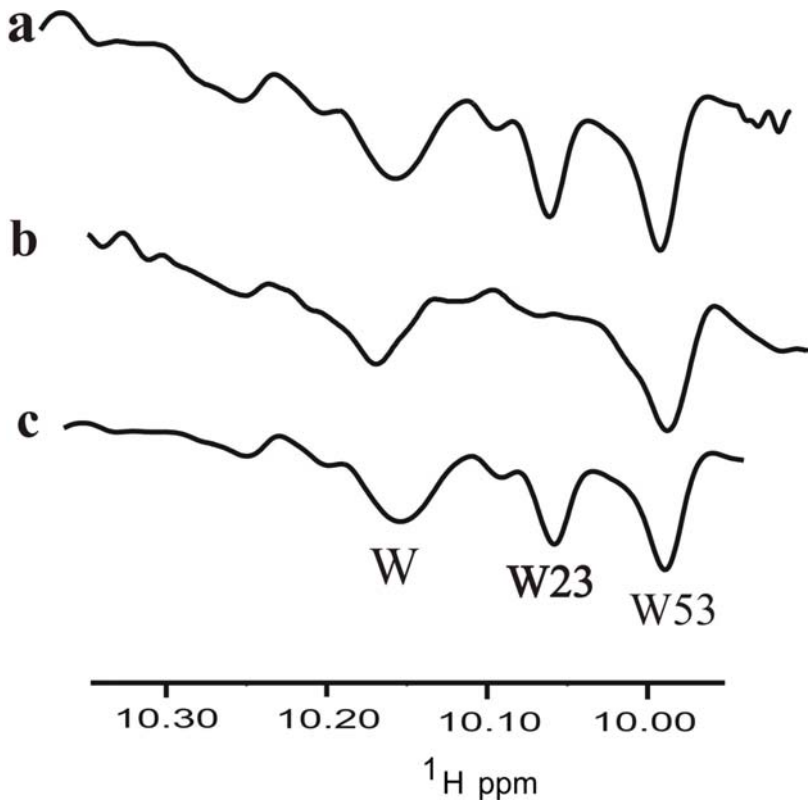


Figure 2.8: 1D proton spectra of the side chains of tryptophans (W) (for assignment refer to Ayed et al., 2001). (**a**) Side chain peaks of free p53, showing W23, W53. W is the side chain peak from the tryptophan of the DNA binding region that is not assigned. (**b**) On forming a complex with MDM2, the side chain of Trp23 disappears (as in Figure 2.6b). The W53 peaks broaden as seen in the 2D spectrum also. (**c**) The peak reappears on addition of the sulfonamide, indicating that p53 has been released.

With these interesting results, we decided to test the inhibitory effect of another small molecule, a boronic–chalcone, that was described as MDM2 antagonists with antitumor effect against cultured tumour cells (Kumar et al., 2003). Employing our methodology by using ^{15}N labeled MDM2, the derivative 3b of the boronic chalcone (Table 1, Kumar et al., 2003) was titrated in order to study its effect on the MDM2/p53

complex. This boronic chalcone did not dissociate the MDM2-p53 complex even at concentrations of 2 mM, i.e. 20 times that of the protein complex (the HSQCs were all equivalent to those of Figure 2.4b). At this concentration of the ligand, the complex completely precipitated in the NMR tube. However, direct titration of ^{15}N-MDM2 with these chalcone showed that they bind to the tryptophan-binding subsite of the p53-binding cleft of human MDM2 (Kussie et al., 1996; Blommers et al., 1997; Stoll et al., 2001) with very low, i.e. high micromolar, affinity, indicating that these compounds are extremely weak inhibitors for the MDM2/p53 interaction, and exhibit their activity in cell lines via a mechanism that is different from the primary p53-MDM2 interaction.

It was recently reported by Issaeva et al. that a small molecule RITA (reactivation of p53 and induction of tumor cell apoptosis), restores apoptosis-inducing function of p53 in tumour cells (Issaeva et al., 2004). The authors in their in vitro studies using FCS (Flourescence Correlation Spectroscopy) reported that RITA binds to the N-terminal domain of p53 and inhibits the complex formation with MDM2. They also found that RITA activates p53 function after its release from the complex. Using our methodology we found that that RITA does not block the formation of the complex in vitro between p53 (residues 1-312) and the N-terminal p53-binding domain of MDM2 (residues 1-118). For these experiments we used, a 118 amino acid N-terminal domain of ^{15}N-MDM2, a 312-residue N-terminal fragment of p53, which encompasses the transactivation and DNA binding domains of p53 (Lee et al., 2000; Ayed et al., 2001; Bell et al., 2002) and a ^{15}N-GST-p53 (residues 1-75) construct. MDM2 interacts through its 100 residue amino terminal domain with the N-terminal transactivation domain of p53 (residues 1-75). The

primary binding site of MDM2 on the p53 protein was mapped to residues 18-26 (Picksley et al., 1994; Kussie et al., 1996).

In the first experiment the isotopically enriched ^{15}N-MDM2 domain was titrated against unlabeled p53. The details of the experiment are summarised in Figure 2.9. In order to get consistent results, we decide to test our experiments in the presence and absence of RITA. Figure 2.9a,d show the reference ^{15}N-MDM2 spectra. A stepwise addition of p53 resulted in the disappearance of the MDM2 (Figure 2.9b,e) spectra just as before (Figure 2.4b). The visible peaks originate from the remaining 8.4% free MDM2 as before. RITA was added to this complex up to a maximum of a 5-fold molar excess relative to p53 and it did not prevent the MDM2/p53 complex formation (Figure 2.9b). RITA was administered to the complex using two protocols. p53 was pre-incubated with RITA at 37^{o}C for 80 min and then MDM2 was added to the mixture (Figure 2.9b). In the second protocol RITA was titrated into MDM2/p53 complex.

Since RITA did not give the desired results of dissociating the complex, experiments were carried out by adding our positive control, nutlin-3, to the complex. Addition of nutlin-3 restores the MDM2 spectrum, as seen in Figure 2.9c,f, with the signals from MDM2 sites involved in binding to nutlin being however shifted. This was a clear indication that nutlin-3 releases p53 from the complex by competing with p53 for binding to MDM2, whereas RITA has no effect on this interaction.

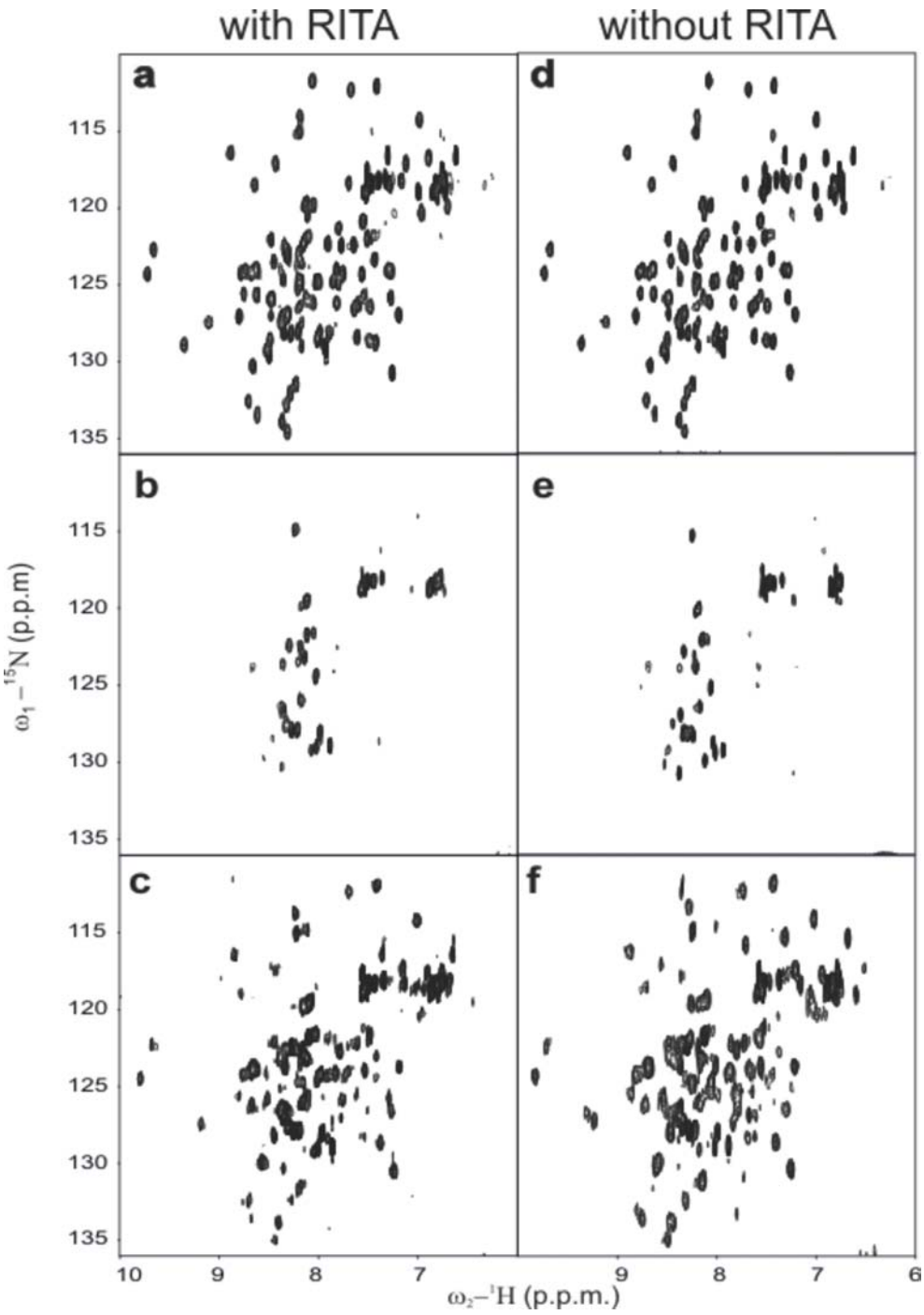


Figure 2.9: HSQC spectra of the ^{15}N labeled MDM2 (residues 1-118). (**a-c**) with RITA, (**d-f**) without. (**a,d**) HSQC spectra of the ^{15}N labeled MDM2. (**b,e**) Addition of p53 results in the disappearance of most of the peaks indicating complex formation. (**c,f**) Additions of nutlin-3 dissociates the p53-MDM2 complex.

To dispel any questions, we decided to test the effect of RITA by using ^{15}N-p53 and unlabeled MDM2. Figure 2.10a shows the ^{1}H-^{15}N HSQC spectrum of ^{15}N-p53 with RITA. A comparison of this spectrum with that of ^{15}N-p53 (without RITA) did not show any significant changes to conclude binding. Since RITA did not perturb these signals, we could conclude that RITA does not bind to the N-terminal domain of p53. A detailed inspection of HSQC spectra of p53 with and without RITA however revealed small changes in the signals originating from the folded core part of p53. To seal any conflict, the best way was to add unlabeled MDM2 to this mixture and observe the three primary binding sites (Phe19, Trp23 and Leu26). The cross peaks of these residues begin to disappear on stepwise addition of MDM2 to the p53/RITA mixture, and completely disappeared on complex formation as seen in Figure 2.10b.

These results clearly indicated that RITA, in vitro, did not inhibit the interaction of the two p53 constructs (residues 1-75 and 1-312) with the N-terminal domain of MDM2. RITA may still be an important discovery, but a possibility exists that RITA selectively targets cells expressing p53 by other mechanisms than direct binding to the N-terminal domain of p53. These might include: interactions through other binding proteins, co-isolation of co-factors, or stabilization of the complex against ubiquitin driven proteolysis. The observed differences in our NMR and Issaeva et al. FCS results could arise if their measurements had an increased sensitivity to a minor sub-population of p53 conformation, which binds to RITA. Such a small population would have little contribution to the total binding and might not be detectable by NMR (for example, NMR monitors a population-averaged mean conformation of a protein for fast exchanging populations on the NMR time scale).

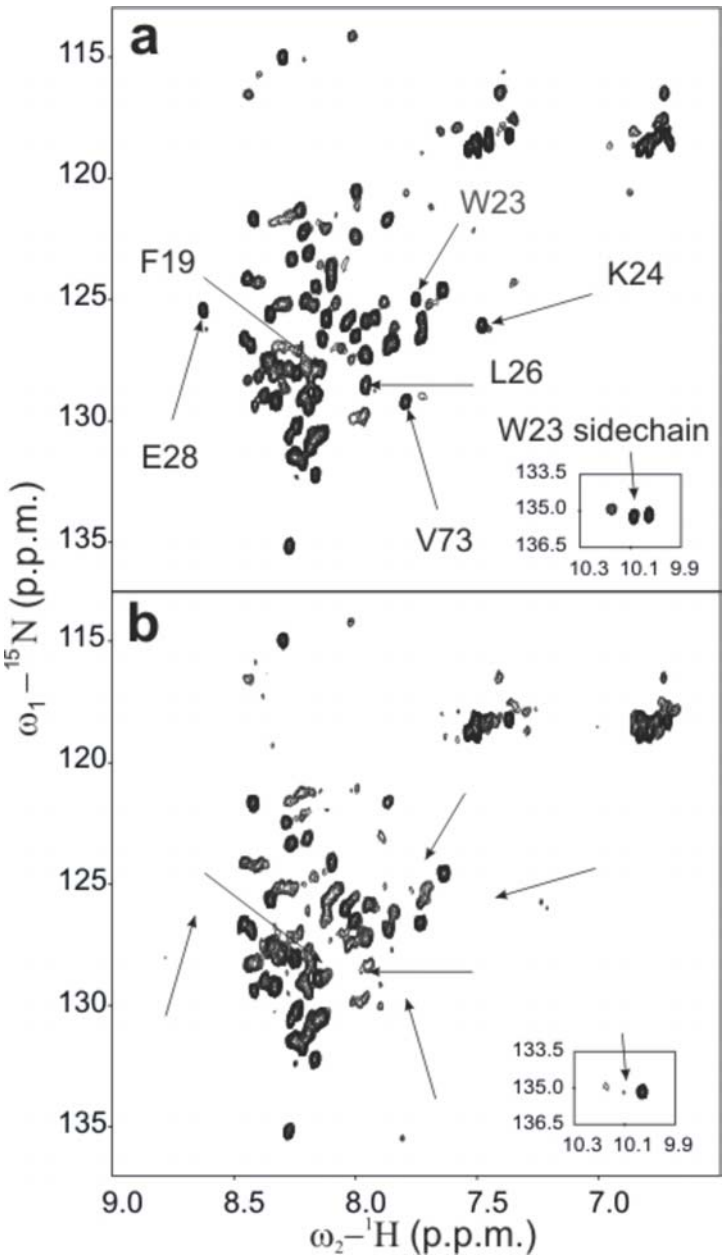


Figure 2.10: HSQC spectra of the flexible N-terminal residues (1-93) of the ^{15}N labeled p53 (the residues of the p53 DNA core domain are mostly not seen because of their broad linewidths). (**a**) At 5% DMSO and saturated solution of RITA. (**b**) After addition of MDM2: the signals of the residues involved in the primary MDM2 binding site disappeared or had dramatically reduced intensities.

2.4 Discussion and the use DIMAPPS

In the course of this methodology, we tested four lead compounds that were reported to inhibit the p53-MDM2 interaction: nutlin-3, a sulfonamide compound NSC 279287, a boronic chalcone, and RITA. Of these four candidates, only nutlin-3 was found to be an inhibitor of the p53-MDM2 interaction and therefore is a potential candidate for drug development in cancer treatment. The findings from our experiments with nutlin-3 showed that nutlin-3 releases p53 from the complex by competing with p53 for binding to MDM2. By monitoring the 1D/2D spectra we could conclude that the freed p53 is folded, the MDM2/nutlin complex is soluble, and that nutlin did not induce precipitation of MDM2. The dissociation constant (K_D) of nutlin-3 for MDM2/p53 complexes estimated from NMR is in the low micro-molar range (< 1 μM), indicating tight binding. The sulfonamide compound precipitates MDM2 and releases folded p53. The boronic chalcone precipitated p53 and MDM2 at high ligand concentrations. This boronic chalcone is an extremely weak binder for MDM2, inconsequential for the MDM2/p53 interaction, and probably expresses its antitumor effect in tumor cell lines via inhibition of interactions other than the primary p53-MDM2 interaction. In case of sulfonamide, the precipitation was selective for MDM2 but not for p53. Direct treatment of MDM2 with sulfonamide gave the same results as that of the complex. RITA was found to be a non-binder and hence cannot be used as an antagonist for the primary p53-MDM2 interaction.

The studies with sulfonamide and the boronic chalcone provides for a more rigorous detection of inhibition of protein-protein interactions than the approaches based on affinity chromatography pull down assays, immunoprecipitation, as these methods are known to give false positive results. In addition these methods provide very limited

information about the structural status of proteins. The correct structure of a protein is a universal requirement for its function. An interesting side-result of these experiments is the determination of the folding status of the p53 that is free and bound to MDM2. Characterization of this status has been a subject of several recent studies (Lee et al., 2000; Ayed et al., 2001; Bell et al., 2002; Dyson and Wright, 2005*)*. Our NMR spectra indicate that the first 93 residues are flexible and unstructured in agreement with the findings which showed that the full-length p53 contains large unstructured N- and C-terminal regions in its native state (Lee et al., 2000; Ayed et al., 2001; Bell et al., 2002). Since the HSQC spectrum of these 93 residues is almost identical to that of the isolated N-terminal domain (Lee et al., 2000), the conformations of these residues have to be the same as those found in (Lee et al., 2000) i.e. although the p53 transactivation domain does not have tertiary structure it is nevertheless populated by a nascent helix and turns (Uesugi and Verdine, 1999; Lee et al., 2000; Ayed et al., 2001*)*. A new finding is that our NMR spectra unequivocally show that, with the exception of a 10 residue-binding site (residues 17 to 26), the bulk of the N-terminal residues are still not structured when bound to the MDM2 domain.

This method should provide an important extension to the traditional "SAR by NMR" technique. One weakness of the SAR approach is that small structured fragments of large proteins have to be found although only larger fragments are usually available at initial studies of protein-protein bindings. Also, many important minimal domains of proteins are about 300 amino acids in length. For large proteins of ca. 30 kDa, the HSQC spectra are normally too crowded to be of practical use in these types of experiments. A remedy would be to prepare a ^{15}N selectively labeled protein with one amino acid type

and obtain a less crowded NMR "subspectrum". Another weakness of the "SAR by NMR" approach arises when NMR cross peaks of several resonances disappear upon ligand binding. This happens for intermediate exchanges when the lifetime of the free and bound states is approximately equal to the differences in chemical shift and/or transverse relaxation rates between the free and ligand-bound forms. These peaks are usually those residues located at the binding interface, however, it is difficult in practice to interpret these data, unless additional information from the structure of the protein is available.

The problem of spectral overlap can however be overcome. An example of spectral overlap is provided by the spectra of the retinoblastoma protein (pRb) shown in Figure 2.11. The smallest functional fragment of pRb, which is also structured, is the so-called small pocket domain of 361 residues (pRb-AB) (Lee et al., 1998). The retinoblastoma tumor suppressor protein is a fundamental negative regulator of cell proliferation that is frequently targeted in human cancer. Many viral oncoproteins (for example, HPV E7, E1A) are known to bind to the pRb pocket domain via a LXCXE binding motif (Hu et al., 1990; Kaelin et al., 1990; Lee et al., 1998). A schematic of these proteins is shown in Figure 2.3. The HSQC spectrum of the $^{15}N/^{2}H$ uniformly labeled pRb-AB contains NH resonances from 361 residues (except prolines) and is too crowded for detecting interactions with the pRb binding peptides. Figure 2.11a shows the spectra of ^{15}N labeled pRb small pocket domain titrated with the HPV E7 peptide of 20 amino acids. Figure 2.11b shows that the complex formation E7-pRb can easily be detected with NMR if a larger fragment of the E7 protein is used.

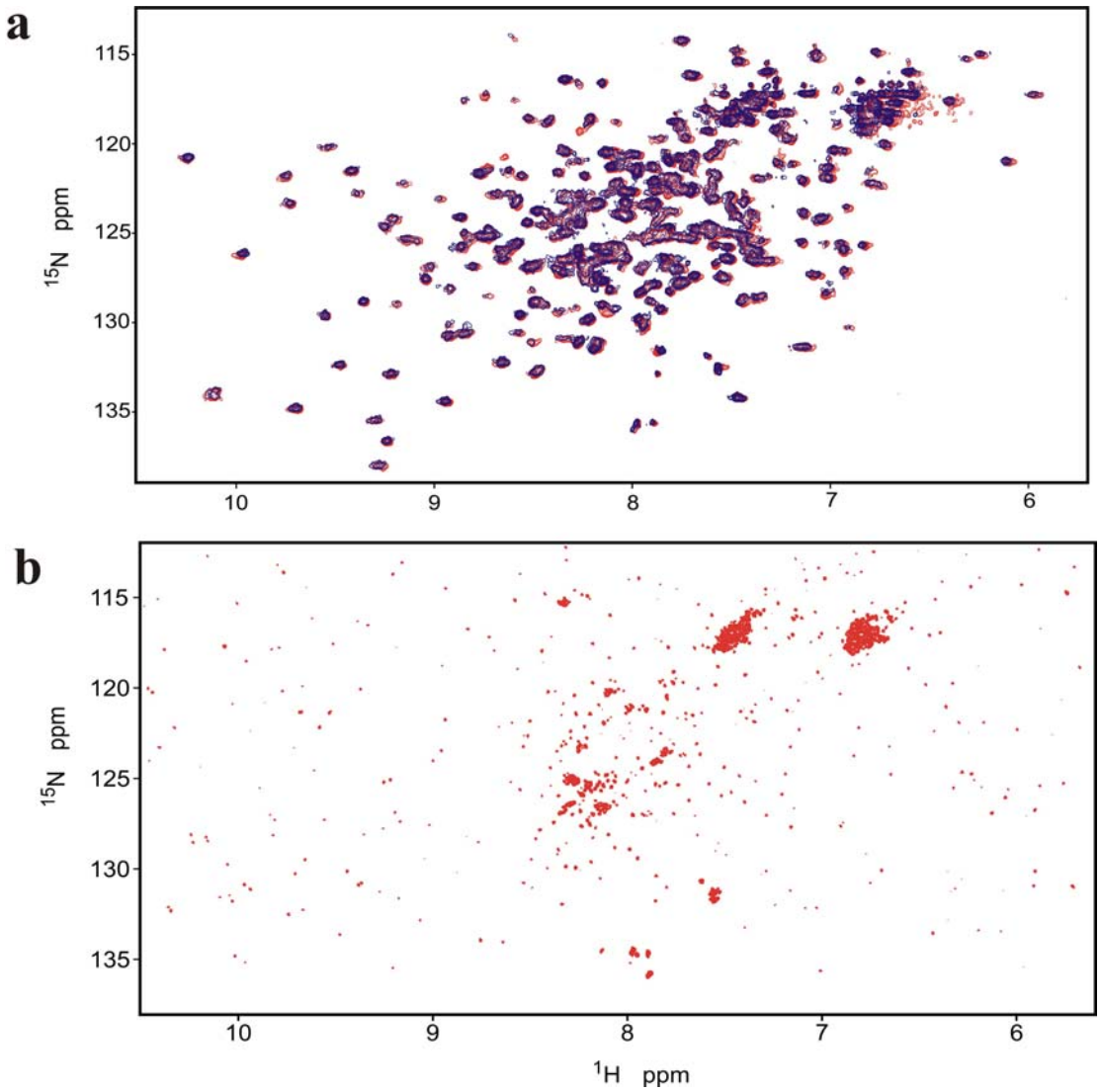


Figure 2.11: (**a**) ^{15}N-HSQC spectra of the small pocket domain of [^{15}N, ^{2}H]-pRb. The reference spectrum is shown in red and the one in blue is after titration with the E7 peptide. (**b**) On addition of the tetrameric full-length E7 protein (104 residues for a monomer) the pRb spectrum disappears due to the formation of the complex.

Several variants of our methodology are possible. In few favorable cases (flexible residues), 1D proton NMR spectra may suffice for monitoring the states of proteins in complexes upon treatment with ligands. The regions of the proton NMR spectra at ppm 8.7 to 12 and 0.0 to -0.5 could be used for these purposes. We have used 1D spectra of

the NH side chains of Trp residues of p53. In general, however, the highly flexible feature of the N-terminal domain of p53 is an exception than a rule.

A protocol that would start first with the titration of the small ^{15}N labeled component of the complex would correspond to the traditional "SAR by NMR" approach. Adding the second larger protein would then follow. Finally, a part of the large protein fragment could be replaced by GST, which has 226 amino acids. For example, a GST tagged N-terminal 93-residue p53 domain could replace the 1-312 residue p53 fragment.

2.5 Experimental methods

2.5.1 Protein expression and purification

The recombinant human MDM2 (residues 1-118) was overexpressed at 30°C in *E. coli* BL21 (DE3) using the pQE-40 vector (Qiagen). The protein was renatured from *E. coli* inclusion bodies as previously published (Stoll et al., 2001) Refolded MDM2 was first applied to the butyl Sepharose 4 Fast Flow (Amersham) and second to HiLoad 16/60 Superdex75 gel filtration (Amersham) columns. The recombinant human p53 protein (residues 1-312) was overexpressed at 37^{o}C overnight in *E. coli* BL21 (DE3) using a modified pQE-40 with N-terminal His-tag and T5 promotor as described in (Bell et al., 2002) The protein was purified under denaturing conditions using a NiNTA (Qiagen) column, refolded and further purified using a Heparin Sepharose 6 Fast flow (Amersham) column. Final purification was done by HiLoad 16/60 Superdex75 gel filtration column. The uniformly ^{15}N enriched protein samples were prepared by growing the bacteria in minimal media containing ^{15}N-NH_4Cl (Stoll et al., 2001).

The small pocket domain of pRb (pRb-AB) was cloned into the pET30 LIC/Xa vector and consisted of an N-terminal 11 amino acid long 6-His-tag followed by the pocket domain without spacer region (from amino acid 379-578 and 642-791). For expression of the protein, the E. coli strain BL21 STARTM (DE3) (Invitrogen) was used. Uniformly ^{15}N and specifically ^{15}N lysine labeled pRb-AB were prepared following standard procedures, (Senn et al., 1987) except that protein induction was carried out at 18^{o}C for overnight. For the perdeuterated sample preparation, cells were first adapted by growing them on small culture with 30%, 60%, 75% and 90% before growing them in 99% $^{2}H_20$ containing media. In this case induction was carried out for 18 h at 18^{o}C.

In the first step of purification the crude cell lysate, after the sonication, was passed through the Ni-NTA column. The eluent from the Ni column was subjected to the MonoQ (Amersham–Pharmacia) anionic exchange chromatography column. The final buffer of the protein solution was 50 mM NaH2PO4.H2O, 10 mM βME, pH 7.8. The identity of protein was confirmed with the SDS-PAGE, Western blot, N-terminal sequencing and mass spectrometry. Final protein samples used for all the studies were more then 95% pure as judged by SDS-PAGE analysis. In case of the perdeuterated sample the protein was >95% perdeuterated as judged by mass-spectrometry.

2.5.2 NMR spectroscopy

All NMR spectra were acquired at 300 K on a Bruker DRX 600 MHz spectrometer equipped with a cryoprobe. Typically, NMR samples contained up to 0.1 mM of protein in 50 mM KH_2PO_4, 50 mMNa_2HPO_4, 150 mM NaCl, pH 7.4, 5 mM DTT, 0.02 % NaN_3, and protease inhibitors. For the ^{1}H-^{15}N HSQC spectrum, (Mori et al., 1995) a total of

1024 complex points in t_2 and 128 t_1 increments were acquired. Water suppression was carried out using the WATERGATE sequence. NMR data were processed using the Bruker program Xwin-NMR version 3.5.

2.5.3 Ligand Binding

Ligand binding experiments were carried out in an analogous way to that described in Stoll et al (Stoll et al., 2001) 500 µl of the protein sample containing 10% D_2O, at a concentration of about 0.1 mM, and a 20 mM stock solution of each compound in DMSO-d_6 was used in all the experiments. Titrations were carried out with three inhibitors, namely, nutlin-3 (nutlin-3 was purchased from Cayman Chemical, MI, USA (Catalog No. 10004372) and comprised an enantiomer with the most potent binding activity towards MDM2 (Vassilev et al., 2004), sulfonamide (obtained from the National Cancer Institute, NSC 279287) and a boronic chalcone (derivative 3b, Table 1, Kumar et al., 2003). In case of the experiments with nutlin-3, the titration was carried out until no further change in the 2D spectrum was observed, thus indicating saturation. The maximum concentration of DMSO at the end of titration experiments was about 2-3%. The pH was maintained constant during the entire titration.

As controls, in order to check the effect of DMSO on the proteins, we titrated the protein complex and proteins with DMSO. We found no significant changes in terms of chemical shifts, precipitation or denaturation of the proteins for DMSO concentrations used in the compound titrations (up to 3%). Figure 2.12 shows the ^{1}H-^{15}N HSQC spectra of MDM2 titrated with DMSO. The p53/MDM2 complex titrated up to 20% with DMSO partially precipitated (40% after 10 h). Preincubation of the p53/MDM2 complex with

15% DMSO at 310 K for 1.5 h, 10 degrees higher than the temperature at which NMR experiments were carried out (300 K), resulted in increased precipitation (50%), but no other changes in the spectrum of the soluble fraction were observed. The soluble complex was reacting with nutlin-3. The p53 peaks of the ^{15}N labeled p53 shift little with higher concentrations of DMSO (c.a. 0.1 ppm in proton dimension for the Trp 23 sidechain signal at 20% DMSO). We thus conclude that the DMSO effect on the proteins while titrating with different compounds can be neglected.

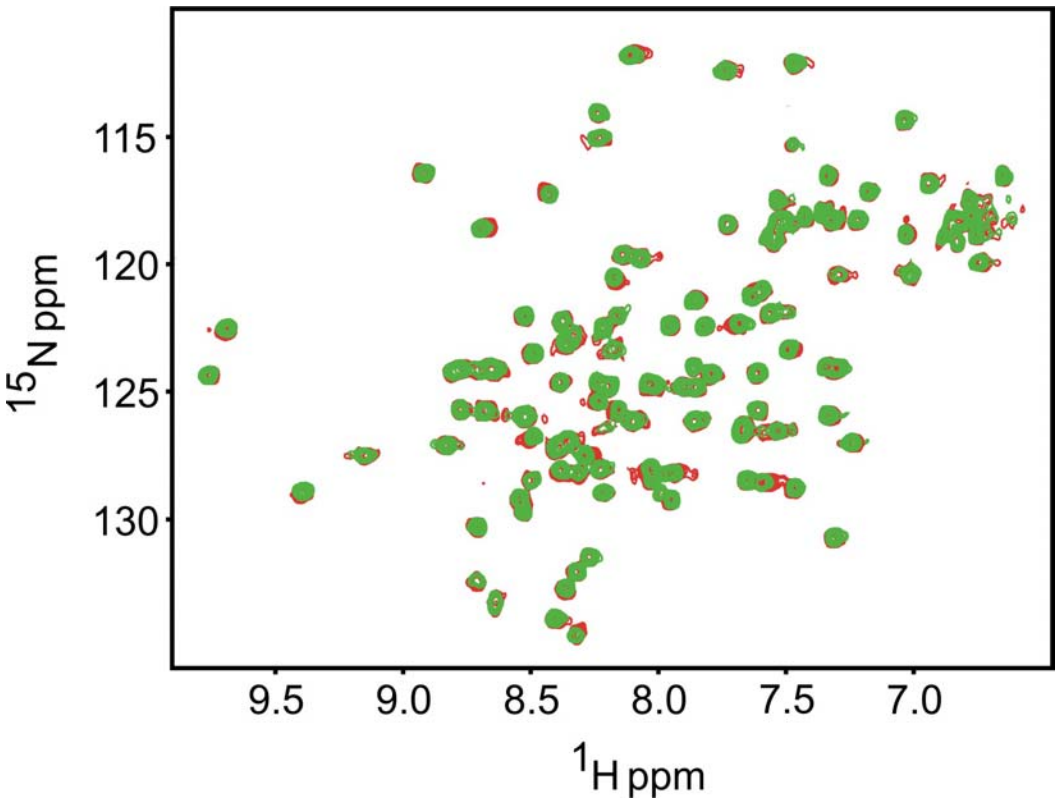


Figure 2.12: ^{1}H-^{15}N HSQC spectra of MDM2 (in red) titrated with DMSO (in green). There were no significant chemical changes when DSMO is added upto about 5% in the NMR tube containing the protein sample. DMSO at this amount did not induce precipitation/denaturation of the protein as seen from the 2D spectra.

Chapter 3

NMR structural characterization and computational predictions of the major intermediate in the oxidative folding of the leech carboxypeptidase inhibitor

3.1 Introduction

Much effort has gone into identifying intermediates that are assumed to be necessary for rapid protein folding (Creighton et al., 1996; Creighton, 1997). However, folding intermediates are usually difficult to isolate and characterize due to their short half-life and highly flexible/unfolded structure. A recent breakthrough in folding has come from

the theoretical field because several groups have succeeded in developing computational approximations that are able to capture the major features of folding processes (Alm, 1999; Galzitskaya et al., 1999; Munoz et al., 1999; Guerois et al., 2000). In general these algorithms can predict accurately the folding nucleus of proteins by exploiting a tight dependence between native state topology and folding mechanisms. A good qualitative agreement between predictions and experiments has been found not only for two-state folding proteins, but also for larger proteins that exhibit transient on-pathway intermediates (Clementi et al. 2000a; Clementi 2000b). Nevertheless, much of our knowledge about folding intermediates comes from studies on the oxidative folding of disulfide-containing proteins, in which transient forms have been trapped using the disulfide acid-quenching approach, analyzed by reversed-phase high performance liquid chromatography (RP-HPLC), and further characterized (Creighton 1986). Folding intermediates of several proteins have been analyzed in this way; among them are those from bovine pancreatic trypsin inhibitor (BPTI) (Weissman et al., 1991), bovine pancreatic ribonuclease A (RNase A) (Scheraga et al., 2001), and lysozyme (Radford et al., 1992).

Leech carboxypeptidase inhibitor (LCI) is a 67-residue inhibitor of A/B metallo-carboxypeptidases isolated from the medical leech *Hirudo medicinalis* (Reverter et al., 1998). It folds in a compact domain consisting of a five-stranded antiparallel β–sheet and a short α–helix that are tightly stabilized by four disulfide bonds (Reverter et al., 2000). LCI is a potent inhibitor of plasma carboxypeptidase B, also called thrombin-activatable fibrinolysis inhibitor (TAFI), which is a well-known attenuator of fibrinolysis (Wang et al., 1998; Bouma et al., 2003). Recent tests *in vitro* on the pro-fibrinolytic activity of LCI

prove its possible use as an enhancer of the tissue plasminogen activator therapy of thrombosis. Knowledge about the folding determinants of this molecule constitutes a basis for the development of variants with enhanced stability. In this regard we have described the unfolding pathway and conformational stability of LCI showing that this protein has slow kinetics of unfolding and is highly stable (Salamanca et al., 2002). The LCI oxidative folding pathway has been extensively characterized using the structural and kinetic analysis of the intermediates trapped by acidification (Salamanca et al., 2003; Arolas et al., 2004). These studies showed that reduced and denatured LCI refolds through a rapid and sequential flow of 1- and 2-disulfide intermediates to reach a rate-limiting step in which two 3-disulfide species (III-A and III-B) strongly accumulate. These two intermediates contain only native disulfide bonds and act as kinetic traps that need major structural rearrangements through the formation of a heterogeneous population of 4-disulfide (scrambled) isomers to render native LCI. III-A and III-B intermediates also populate transiently the reductive unfolding pathway of LCI. The preliminary structural characterization of these two intermediates showed that III-A is a properly folded and stable species that might have high content of native structure, while III-B is a less structured form.

This chapter reports a detailed NMR structural analysis of the III-A folding intermediate of LCI. This intermediate has been directly purified from the oxidative folding reaction using RP-HPLC, and its structure characterized by NMR spectroscopy and compared to that of native LCI. The obtained results allow deciphering why this species acts as a major kinetic trap in the folding process of LCI. In addition, the integration of structural data with theoretical folding predictions provides insights into

the crucial role played by secondary structure in directing the folding of LCI and has general implications for understanding the folding of other disulfide-rich proteins.

3.2 Results and Discussion

3.2.1 Isolation of the III-A folding intermediate

In the past years several folding intermediates of protein models such as BPTI, RNase A and lysozyme have been structurally characterized. However, most of these studies have been centered on the analysis of mutant analogues of the intermediates lacking a specific disulfide bond (van Mierlo et al., 1991; van Mierlo et al., 1993; Noda et al., 2002; Yokota et al., 2004). These analogues can be obtained in substantial quantities and in homogeneous form, but they are not "true" intermediates in the sense that they do not posses any reactive free cysteine. Many analogues in which the same cysteine residues have been substituted by different amino acids or chemically blocked differ in stability and/or functionality among them (Talluri et al., 1994; Shimotakahara et al., 1997; Iwaoka et al., 1998). The difficulty to obtain and characterize "true" folding intermediates is reflected in the fact that, to the best of our knowledge, no high-resolution structures of intermediates directly purified from a folding reaction have been reported so far, and only partial structural characterizations have been attained (Gehrmann et al., 1998; van den Berg et al., 1999; Eliezer et al., 2000; Sato et al., 2000; Cemazar et al., 2003).

This work was aimed to characterize in detail the structure of the major folding intermediate of LCI, the III-A. For this task, LCI was expressed in *E. coli* as uniformly ^{15}N-labeled protein and purified to homogeneity (purity >98%). Oxidative refolding of

fully reduced ^{15}N LCI was carried out to generate III-A. The RP-HPLC profiles of acid-trapped intermediates at selected time points during folding in the absence (control -) and presence (control +) of 2-mercaptoethanol as thiol catalyst are shown in Figure 3.1a. A high degree of heterogeneity of 1- and 2-disulfide intermediates is observed at the beginning of the folding reaction, with identical RP-HPLC profiles in both refolding conditions (control - and control +). This initial stage of folding is followed by the accumulation of two major intermediates (III-A and III-B) that contain three native disulfide bonds and act as strong kinetic traps. At this point (8 h), the RP-HPLC patterns are still similar regardless of the presence of the reducing agent. The last stage of the folding process is characterized by accumulation of a heterogeneous population of 4-disulfide (scrambled) isomers that is more pronounced when the refolding is performed in the absence of 2-mercaptoethanol. This reducing agent strongly promotes disulfide reshuffling and conversion of scrambled forms to the native conformation.

To isolate III-A, the reaction was acid-quenched approximately after 8 h of refolding in the absence of 2-mercaptoethanol to avoid the presence of possible traces of scrambled isomers. The intermediate was separated by RP-HPLC and freeze-dried after verification of its purity by treatment with vinylpyridine and analysis by matrix-assisted laser desorption/ionization - time of flight mass spectrometry (MALDI-TOF MS). The ^{1}H NMR spectrum of III-A also confirmed the homogeneity of the preparation (see Figure 3.1b). The chemical shift dispersion of the III-A ^{1}H spectrum at pH 3.5 is typical for globular proteins and suggests a properly folded conformation at this pH. A low pH is a necessary condition to maintain the III-A folding intermediate in a “quenched” state.

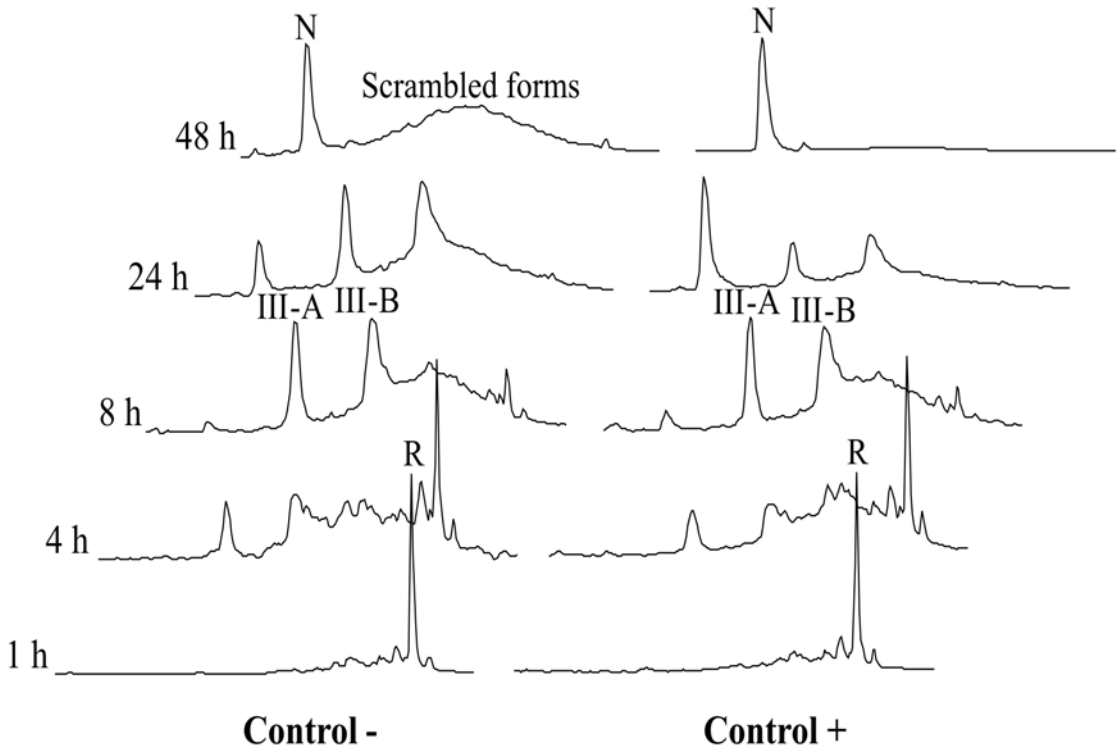


Figure 3.1a: RP-HPLC analysis of the folding intermediates of LCI trapped by acidification. Oxidative folding was carried out at 22ºC in Tris-HCl buffer (0.1 M, pH 8.4) with (Control +) or without (Control -) the redox agent 2-mercaptoethanol (0.25 mM). The protein concentration was 0.5 mg/mL. Folding intermediates were trapped at the noted times by acidification and analyzed by RP-HPLC using the conditions described in "Experimental Procedures". *N* and *R* indicate the elution positions of the native and fully reduced species of LCI, respectively. *III-A* and *III-B* are two major intermediates containing three native disulfides and s*crambled forms* are isomers with four disulfides.

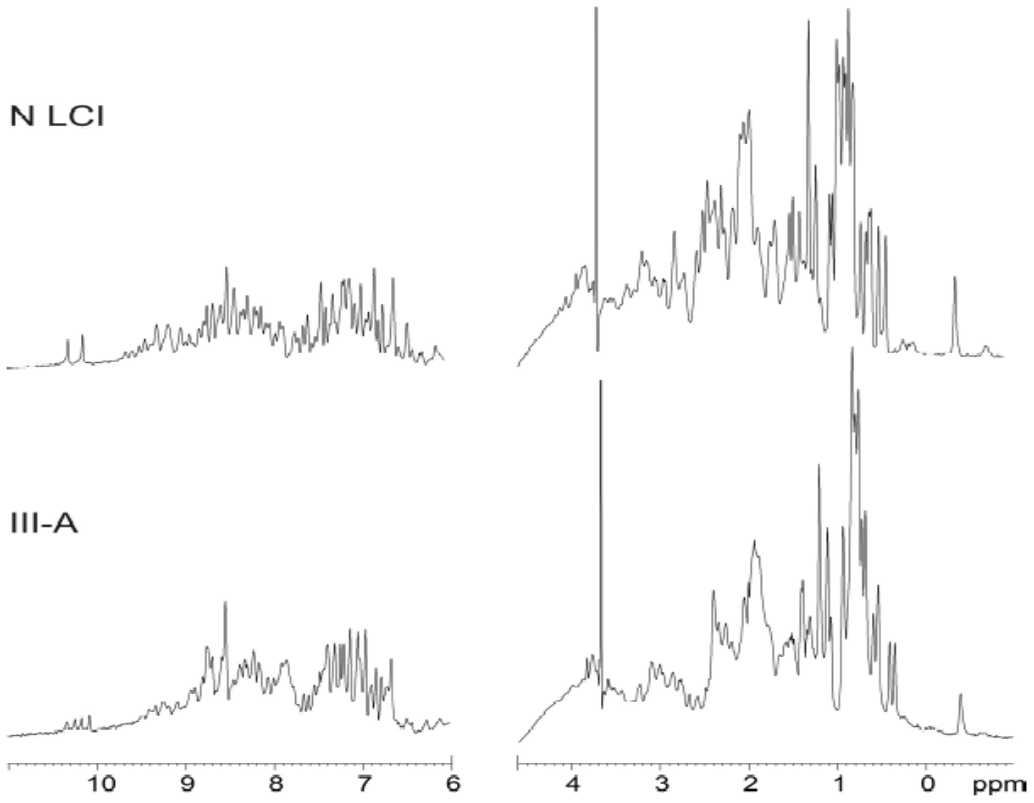


Figure 3.1b: ^{1}H-NMR analysis of native LCI and III-A intermediate. The spectra were recorded on a Bruker DRX 600-MHz spectrometer at 300 K and at a protein concentration of 1 mM. Samples were prepared in H2O/D2O (9:1) at pH 3.5.

3.2.2 Resonance assignments

^{1}H and ^{15}N NMR assignments were performed by the standard sequence-specific method on the basis of heteronuclear 2D and 3D spectra of the uniformly ^{15}N-labeled native LCI and the III-A folding intermediate recorded in H_2O at 27ºC and pH 3.5. The 2D ^{1}H-^{15}N HSQC spectra of both native LCI and III-A show excellent chemical shift dispersion with few overlaps (Figure 3.2). Excluding the N-terminal residues 1-4, which are highly flexible, and the nine Pro residues, all 54 backbone NH cross-peaks could be observed in

the ^{1}H-^{15}N HSQC spectrum of the native LCI. For III-A, there are 53 cross-peaks and only residue Val51 is not observed, probably due to an increased flexibility in this region. All backbone cross-peaks were readily assigned to their corresponding spin system after sequential assignment.

Spin systems were sequentially connected throughout $d_{\alpha N(i,i+1)}$, $d_{NN(i,i+1)}$, and/or $d_{\beta N(i,i+1)}$ nuclear Overhauser enhancements (NOEs) observed in the 2D NOESY and 3D NOESY-HSQC. The whole stretch of sequential connectivities at the spectra was only broken at the nine Pro residues for native LCI and also at residue Val51 for III-A. The ^{1}H and ^{1}H-^{15}N HSQC spectra of the III-A intermediate did not change along the experiment, indicating that, as expected, intra-molecular disulfide rearrangements did not occur to a significant extent at the working pH.

Position of most cross-peaks in the ^{1}H-^{15}N HSQC spectra of native LCI and III-A was similar (Figure 3.2). However, on the overall smaller cross-peak dispersion was observed for the folding intermediate. This indicates a less compact conformation as compared to the native protein. Individual protein chemical shifts depend strongly on local chemical and electronic environments. Therefore, direct comparisons of chemical shift data between native LCI and III-A can elucidate local areas within the intermediate where differences are present. Accordingly, a comparison of assigned backbone NH and $C^{\alpha}H$ was carried out (Figure 3.3). As might be expected, the largest differences are centered in the two regions encompassing Cys22-Arg23 and Gln57-Thr60. This can be unequivocally attributed to the absence of the Cys22-Cys58 disulfide bond in the intermediate.

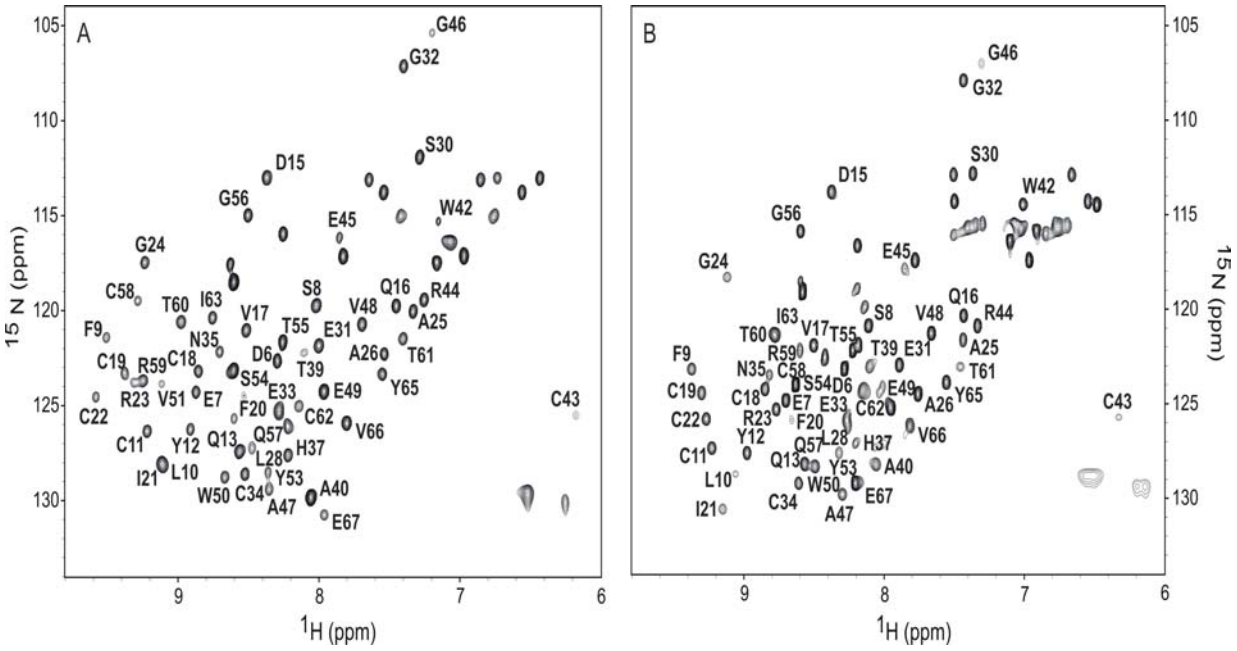


Figure 3.2: 2D ^{1}H-^{15}N-HSQC spectra of uniformly ^{15}N-labeled native LCI and III-A intermediate. Backbone cross-peak assignments are indicated for both native LCI (A) and III-A (B). The experiments were carried out with 1 mM protein concentration in H_2O/D_2O (9:1) at pH 3.5 and 27ºC.

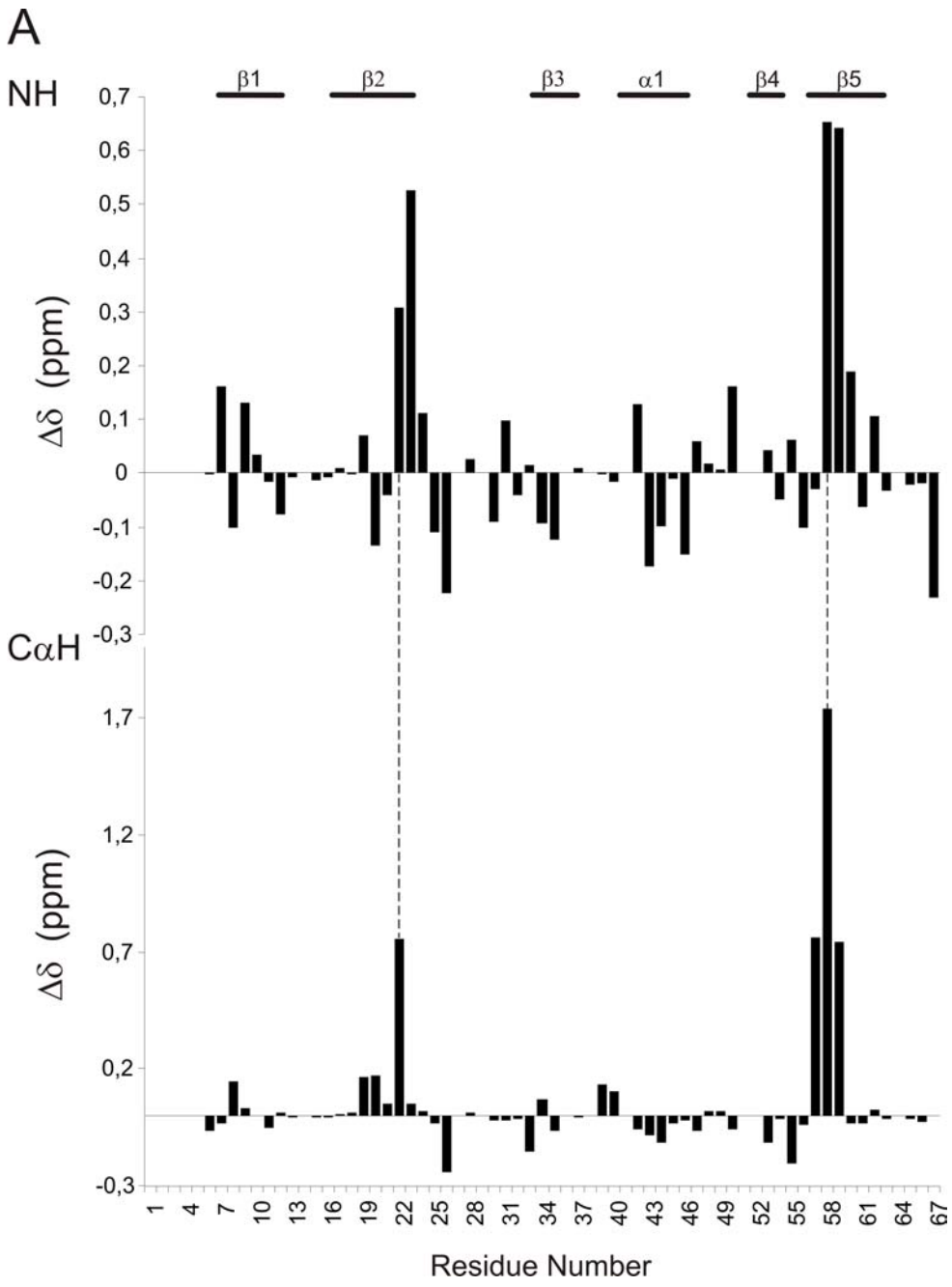


Figure 3.3: Chemical shift differences (Δδ) of backbone $C^{\alpha}H$ and NH resonances between Native LCI and III-A Intermediate $\Delta\delta = \delta_{native} - \delta_{III\text{-}A}$ as a function of the residue number. Backbone β-sheet and α-helix structures indicated above the plots are based on the structure of native LCI.

3.2.3 Secondary structure analysis

The elements of secondary structure in native LCI and III-A were first delineated using information provided by the NOE data (Wüthrich et al., 1984; Wüthrich, 1986). For native LCI, a five-stranded antiparallel β-sheet was identified by strong long-range interstrand $d_{\alpha\alpha(i,j)}$, $d_{\alpha N(i,j)}$ NOEs, medium intensity interstrand $d_{NN(i,j)}$ NOEs, and strong sequential $d_{\alpha N(i,i+1)}$ NOEs with weak or absent $d_{NN(i,i+1)}$ NOEs. This antiparallel β-sheet involves residues Glu7-Gln13 (β1), Gln16-Arg23 (β2), Glu33-His37 (β3), Val51-Tyr53 (β4) and Gly56-Ile63 (β5) (Figure 3.4). A short stretch of strong sequential $d_{NN(i,i+1)}$ NOEs, weak or absent $d_{\alpha N(i,i+1)}$ NOEs, and medium range $d_{NN(i,i+2)}$, $d_{\alpha N(i,i+3)}$, and $d_{\alpha N(i,i+4)}$ NOEs indicated the presence of a short α-helix (α1) located between residues Pro41 and Gly46.

For the III-A intermediate, a four-stranded antiparallel β-sheet was identified by strong sequential $d_{\alpha N(i,i+1)}$ NOEs with weak or absent $d_{NN(i,i+1)}$ NOEs involving residues Glu7-Gln13 (β1), Gln16-Arg23 (β2), Glu33-His37 (β3), and Cys58-Ile63 (β5') (Figure 3.3). Strong long-range interstrand $d_{\alpha\alpha(i,j)}$ contacts were found between β3-β1 (Cys34-Cys11), β1-β2 (Ser8-Ile21, Leu10-Cys19), and β2-β5' (Cys18-Cys62). Strong long-range interstrand $d_{\alpha N(i,j)}$ NOEs were detected between β3-β1 (His37-Phe9), β1-β2 (Ser8-Cys22, Cys11-Cys19), and β2-β5' (Gln16-Tyr65, Cys18-Ile63). Finally, medium intensity interstrand $d_{NN(i,j)}$ NOEs were identified between β3-β1 (Glu33-Tyr12, Asn35-Leu10), β1-β2 (Glu7-Cys22, Phe9-Phe20, Cys11-Cys18, Gln13-Gln16), and β2-β5' (Val17-Ile63, Cys19-Thr61, Ile21-Arg59). All these contacts were also present in the spectra of the native LCI defining the antiparallel strands β3-β1-β2-β5.

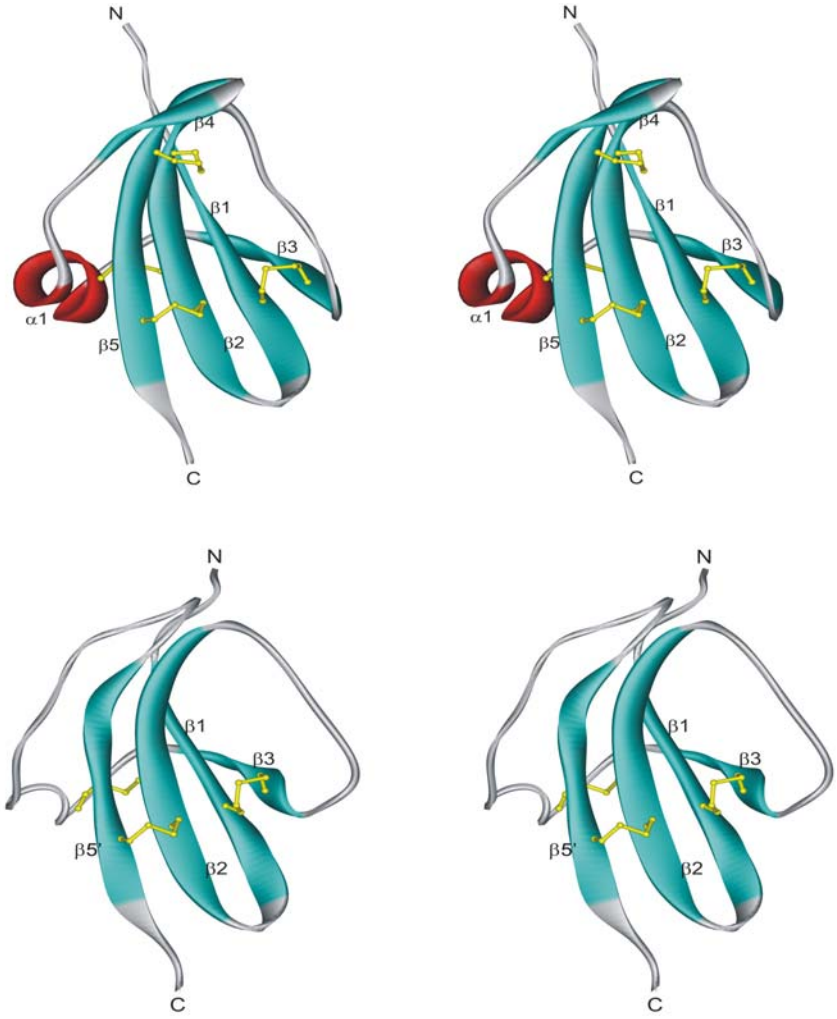


Figure 3.4: Stereoview of the structure of native LCI and III-A intermediate. Ribbon representation of the calculated structure for native LCI (upper panel) and III-A (lower panel). β-strands are shown in blue and the α-helix in red. N and C indicate the location of the N- and C-terminal tails of both proteins. The disulfide bonds are shown in yellow.

Three main differences in terms of secondary structure could be observed between native LCI and the III-A intermediate. First, an increased structural disorder is present in the region between residues Glu49-Thr55 as compared to that in the native protein. Proton resonances of residue Val51 cannot be observed in the spectra of III-A. In addition, NH-NH contacts observed in native LCI between β4-β5 are lost in the

intermediate (e.g. Val51-Cys58 and Tyr53-Gly56). This probably precludes the formation in III-A of the short strand present in the native form (β4).

Second, the strand β5' of III-A (located in the C-terminus) is significantly shorter than the corresponding strand in the native LCI (β5); in the intermediate some contacts are lost between β2 and β5' (e.g. the NH-NH contact between Arg23-Gln57). Despite these missing contacts, the C-terminus of the intermediate is far from being free; the β5' strand in the intermediate is large and stable enough to keep the C-terminus tail in a proper orientation, providing the conformational rigidity needed for the binding to the carboxypeptidase. This explains why the inhibitory capability of III-A is indistinguishable from that of the native LCI for all tested carboxypeptidases (Arolas et al., 2004). The last remarkable difference is the absence of a well-defined α-helix in the intermediate. However, the region between residues Pro41-Gly46 is very similar to a short α-helix. Some strong $d_{NN(i,i+1)}$ and medium $d_{NN(i,i+2)}$ NOEs are found here (e.g. Cys43-Arg44, Arg44-Glu45, Trp42-Arg44), together with weak or absent $d_{\alpha N(i,i+1)}$ NOEs.

3.2.4 Three-dimensional structure calculation

The backbone structures of native LCI and III-A intermediate were calculated using the simulated annealing method with the program CNS (Brünger et al., 1998). The parameters describing these structures are given in Table 3.1. For native LCI, with the exception of the five N-terminal and the three C-terminal residues, the ensemble of 20 calculated structures is well defined (Figure 3.5), with an average backbone root mean square deviation (rmsd) of 1.40 Å (residues 6-64). Figure 3.6 indicates the rmsd of each residue in the bundle of 20 structures, thus showing the most disordered regions in native

LCI. The three-dimensional structure of this molecule at pH 3.5 is similar to the one we reported previously at high resolution at pH 6.5 (Reverter et al., 2000), with a rmsd of 1.31 Å between average structures. Native LCI consist of a five-stranded antiparallel β-sheet with a β3-β1-β2-β5-β4 topology, and a short α-helix that packs onto the most compact part of the β-structure interacting with the end and the beginning of the β1 and β2 strands, respectively (Figure 3.4 and 3.5). A high percentage of residues belong to regular secondary structure elements (nearly 45%), which are cross connected and stabilized by the presence of four disulfide bonds: Cys11-Cys34 (β1-β3), Cys18-Cys62 (β2-β5), Cys19-Cys43 (β2-α1), and Cys22-Cys58 (β2-β5). They provide high stability and compactness to the protein. The calculation of the structure of the III-A intermediate confirms that it also possesses a well-defined globular conformation that includes a four-stranded antiparallel β-sheet with a β3-β1-β2-β5' topology (Figure 3.4). However, some parts of the intermediate appear to be more disordered than in native LCI (Figure 3.5). This is shown by the higher average backbone rmsd value for residues 6-64 of the 20 calculated structures: 2.47 Å. The rmsd of each residue is shown in Figure 3.6 and provides evidence for the presence of highly disordered regions, mainly between residues Arg23-Gly32 and Arg44-Gln57. The disulfide pairings of III-A were unambiguously determined during three-dimensional structure calculations and were in complete agreement with a previous assignment carried out by digestion of the vinylpyridine-derivatized intermediate with thermolysin and analysis of the resulting disulfide-containing peptides by MALDI-TOF MS and automated Edman degradation (Arolas et al., 2004): Cys11-Cys34 (β1-β3), Cys18-Cys62 (β2-β5'), and Cys19-Cys43 (β2-"α"). The missing disulfide bond present in the native form between Cys22 and Cys58 seems

to account for the lower compactness of such intermediate as compared to that of the native protein. However, the structural similarities between the native protein and the III-A intermediate are striking.

	Native LCI	III-A intermediate
rms deviation from ideal geometry		
Bond lengths [Å]	0.0063	0.0060
Bond angles [°]	0.610	0.590
rms deviation from the mean structure [Å]		
Backbone atoms (residues 6-64)	1.40	2.47
Ramachandran analysis [%]		
Residues in favored regions	60.5	48.5
Residues in allowed regions	33.2	40.2
Residues in generously allowed regions	5.4	9.5
Residues in disallowed regions	0.9	1.8
Distance restraints		
Total NOE distance restraints	316	325
Short range	133	128
Medium range	168	184
Long range	15	13
Hydrogen bond restraints	30	18
Number violations > 0.5 Å	0	0

Table 3.1: Parameters for the NMR structure calculation of native LCI and III-A intermediate

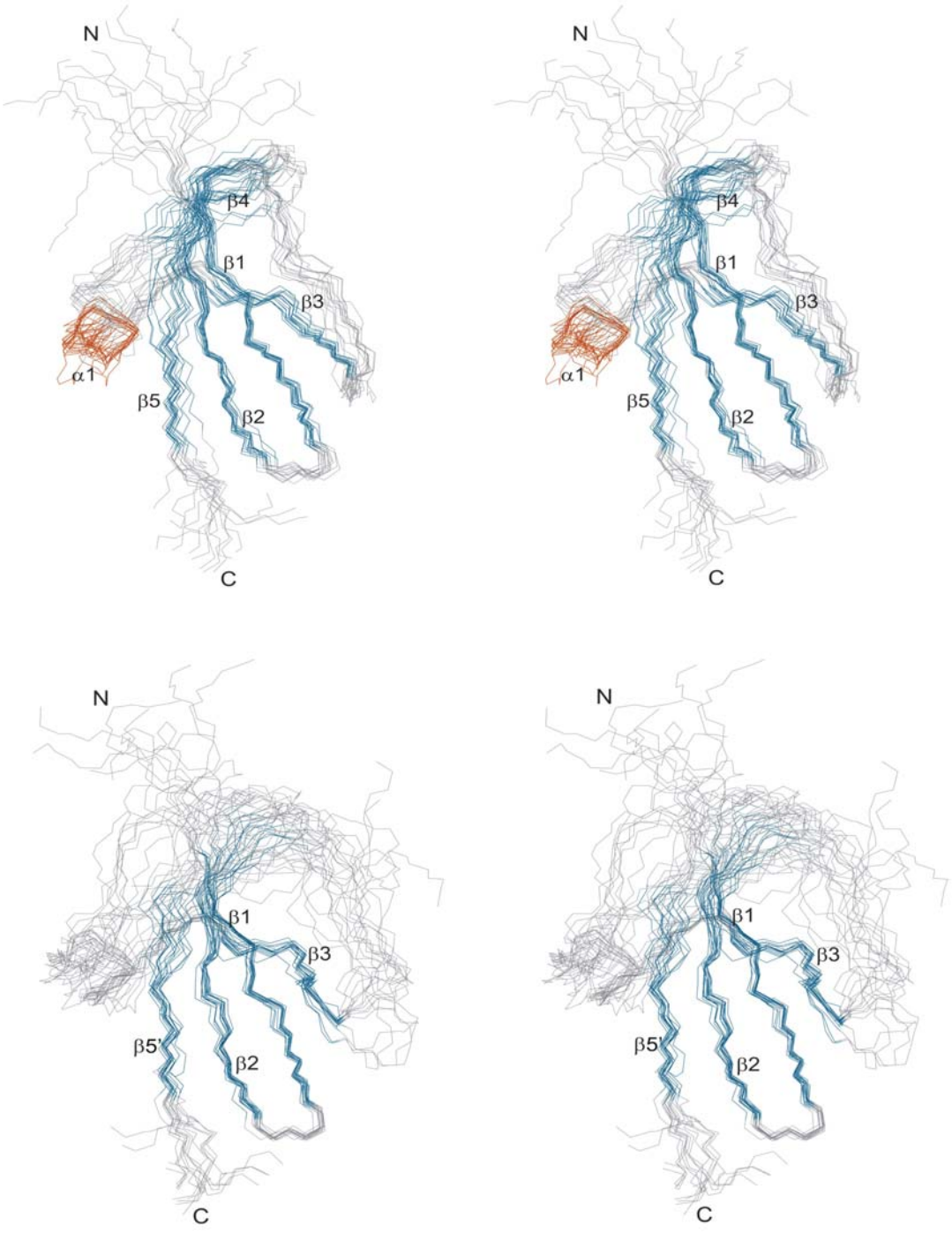


Figure 3.5: Stereoview with the superposition of the calculated structures for native LCI and III-A intermediate. The pictures show the backbone atoms of the bundle of lowest energy structures calculated for native LCI (upper panel) and III-A (lower panel). β-strands are indicated in blue and the α-helix in red. The N- and C-terminal tails are labeled.

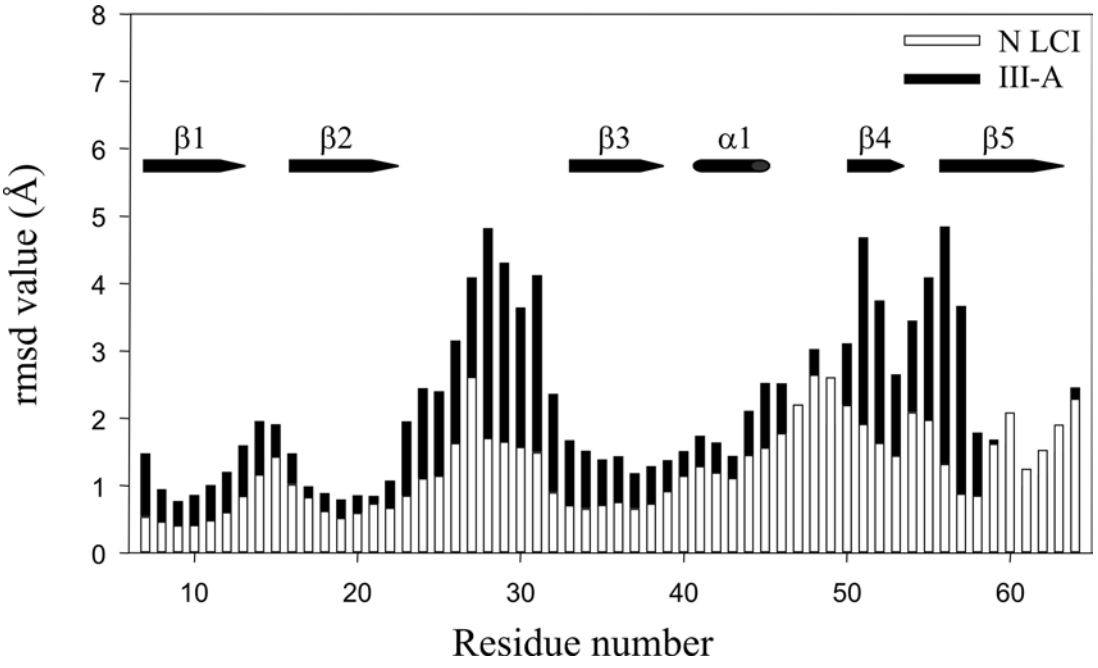


Figure 3.6: Comparison of local rmsd values for backbone atoms of native LCI and III-A. The rms deviations of the backbone atoms from the 20 calculated structures are plotted vs the residue number (residues 6-64). Secondary structure elements for native LCI are indicated inside the graphic.

3.2.5 Understanding the role of the III-A intermediate in the folding of LCI

In a recent study, it was hypothesized that the way the oxidative folding pathway of LCI proceeds depends on the ability of its secondary structure elements to protect progressively native disulfide bonds from rearrangement in the interior of a totally or partially folded structure (Arolas et al., 2004). This would result in a final native structure in which disulfide bonds should be highly protected. This view is strongly reinforced by the results of amide proton exchange experiments on native LCI and III-A intermediate.

Both forms were dissolved in D_2O at pH 3.5 and several ^{1}H-^{15}N HSQC spectra were recorded over time. For native LCI, maximum exchange was achieved after approximately 10 h and the spectra did not significantly change after that time point. The following residues were found protected: Phe9-Gln13 (β1), Gln16-Arg23 (β2), Gly32, Asn35 and His37 (β3), Thr39, Cys43 (α1), Val51 (β4), Gln57-Arg59 and Thr60-Ile63 (β5), and Tyr65 (see Figure 3.7). Thus, the protected residues are located within all secondary structure elements around the cysteine residues and this fact clearly indicates that in native LCI the four disulfide bonds are buried and not solvent-accessible.

For the III-A intermediate, after 15 min of exchange 35 residues from all secondary structure elements were found protected (in the native form were 36 residues). As expected, residues located in highly disordered regions were already exchanged at this early time point (Figure 3.7). Similarly to the native form, in the intermediate maximum exchange was achieved after about 10 h. The protected residues were Phe9-Gln13 (β1), Gln16-Cys19 and Ile21-Cys22 (β2), Asn35 (β3), Thr60 and Ile63 and Tyr65 (β5'). Thus, in III-A both free cysteines (Cys 22 and 58) and the three disulfide bonds are located in protected regions or close to them and therefore are not solvent-accessible or have limited accessibility. The final number of protected residues for native LCI and the intermediate (27 and 15 residues, respectively) is in good agreement with those previously found measuring the global D/H exchange by MALDI-TOF MS (26 and 16 residues, respectively; (Arolas et al., 2004). This approximately 40% of decrease in protected residues between native LCI and III-A, mainly localized in the "missing" secondary structure elements of the latter (α1, β4 and β5'), is a reflection of its lower level of conformational packing.

Analogs of BPTI and RNase A intermediates that lack only one native disulfide bond and act as major kinetic traps along the folding process display structures similar to the native ones (Hurle et al., 1992; van Mierlo et al., 1994; Laity et al., 1997). It is seen that this is also the case of the III-A intermediate of LCI. However, critical differences among the folding pathways of these proteins are seen. For both BPTI and RNase A, the last step in their folding reactions consists of a direct oxidation of two free cysteines to form the native structure. In LCI, however, a major disorganization of the acquired tertiary structure (in III-A) to form a heterogeneous ensemble of non-native 4-disulfide (scrambled) isomers seems to be necessary before the native structure can be gained. This was previously shown by stop/go experiments using the III-A intermediate and scrambled isomers (Arolas et al., 2004). In these studies we demonstrated that III-A mainly converts into scrambled isomers prior to the formation of native LCI. This reaction is not affected by the presence of an excess of redox reagents, which confirms that the formation of the native form by direct oxidation of the two free cysteines in the intermediate is very unlikely. In addition, the isolated scrambled isomers were shown to reshuffle directly into the native form. The presence of scrambled isomers as a last productive step has been observed in the folding of several small disulfide-rich proteins (Chang et al., 2001).

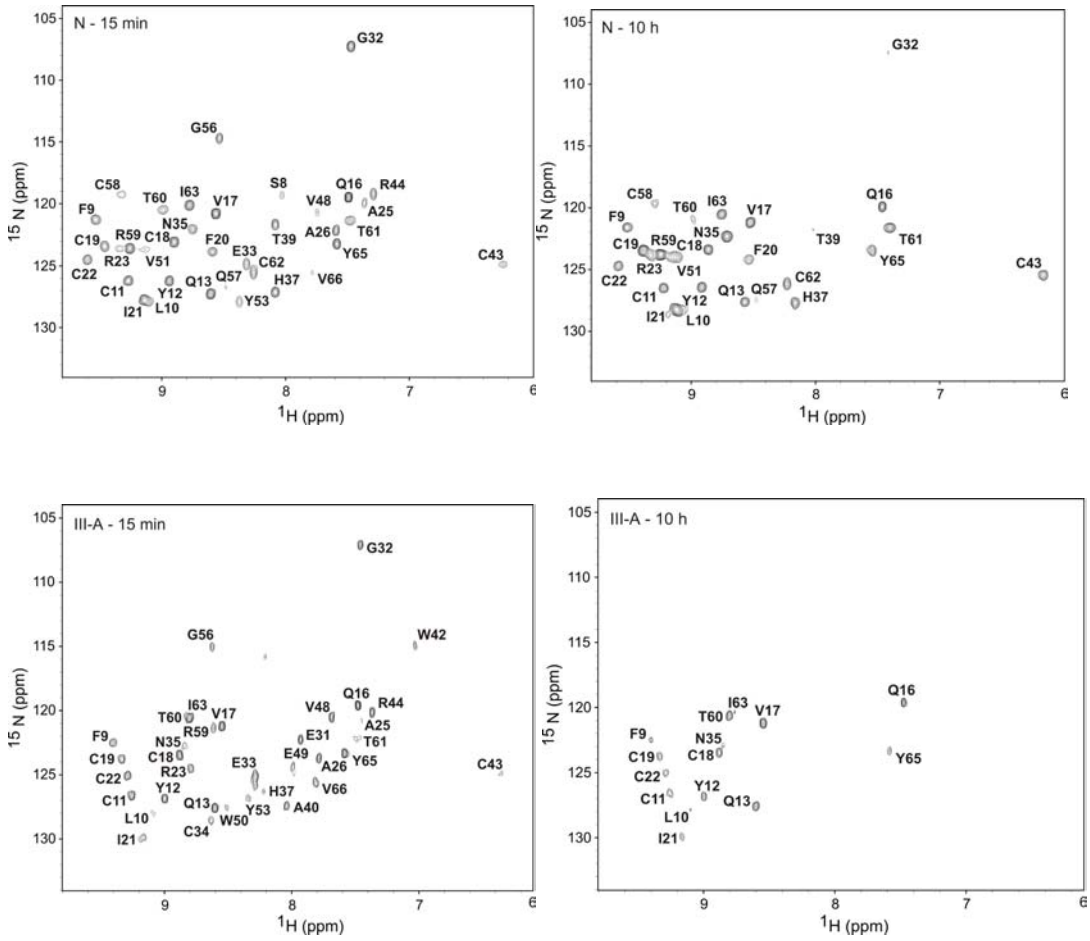


Figure 3.7: Amide exchange experiments on Native LCI (top) and III-A intermediate (bottom). The protected residues are shown in the 2D ^{1}H-^{15}N-HSQC spectra of uniformly ^{15}Nlabeled native LCI and III-A intermediate recorded after 15 min and 10 h of exchange. The experiments were carried out with 1 mM protein concentration in D_2O at pH 3.5 and 300 K.

It has been previously postulated that III-A is a metastable form similar to what Scheraga and co-workers had defined as *disulfide-insecure* intermediates (Welker et al., 2001). In these forms, the free thiol groups are as well protected as the disulfide bonds, and cannot be simply exposed and oxidized by a local unfolding process as it happens with BPTI or RNase A on-pathway intermediates. Structural fluctuations that expose the thiol groups are also likely to expose the disulfide bonds and promote their reshuffling

instead of oxidation of the free thiols to the native pairings. The data from this work, together with information from the stop/go assays confirm that indeed III-A acts as a *disulfide-insecure* intermediate. Unlike in III-A, the four disulfide bonds of the unstructured scrambled isomers are solvent-accessible, therefore the addition of reducing agent strongly accelerates their kinetics of reshuffling to attain native disulfide pairings and the native structure.

In BPTI or RNase A analogs, deletion of the disulfide bond introduces only small and localized structural perturbations with respect to the native form, in the vicinity of the eliminated bond (Hurle et al., 1992; van Mierlo et al., 1994; Laity et al., 1997). In contrast, long-range effects are detected in III-A: the α-helix and β4 strand present in the native form disappear in the intermediate and the loop connecting strands β2 and β3 is distorted. In addition, a higher degree of disorder, probably indicating higher backbone flexibility, is detectable in some protein regions when compared to those of native LCI. One of these flexible stretches is the region around Cys43, even if in the intermediate this residue is bonded to Cys19. Interestingly, the other major kinetic trap of LCI folding, III-B, lacks this disulfide bond. We have shown by stop/go experiments that there is an inter-conversion of intermediates III-A and III-B to each other, reaching equilibrium prior to the formation of scrambled isomers. This disulfide interchange is an internal process in which all reaction groups are protected from the exterior, since neither the rate of inter-conversion nor the concentration of the species at the equilibrium are affected by the presence of external thiols. Minor local fluctuations due to an increase of flexibility in the backbone of the involved regions, such as the one observed here, may be well behind this solvent-independent disulfide interchange.

3.2.6 Predicting the folding pathway of LCI

As concluded above, the III-A intermediate, representing the major kinetic trap in the oxidative folding process of LCI, is a highly structured molecule. Several groups have hypothesized that the folding of disulfide-rich proteins hinges critically on their secondary structure organization. In order to provide additional evidences for this view, we tested whether a folding predicting algorithm, which only uses the native structure of the protein as an input and whose calculations are mainly based on topological considerations, may be able to predict the presence and the structural properties of the highest energy barrier in the folding of LCI. For this task we used the Fold-X software developed by Serrano's group at the EMBL (Guerois et al., 2002; Kiel et al., 2004). This algorithm, which takes into account also sequence features underlying protein topology, has been shown to predict even subtle differences within protein folding pathways. Figure 3.8 shows the probability for LCI residues to be ordered during the folding process. The first β-hairpin, formed by β1, β2, and the turn connecting them, exhibits the highest probability to be ordered at the beginning of the folding reaction. When the number of ordered residues increases (up to 55%) folding of β3 and β5 also reaches high probability. The association of these two regions corresponds sharply to the native-like regions in the III-A intermediate. Conversely, the fourth β-strand as well as the α-helix exhibit very low propensity to fold.

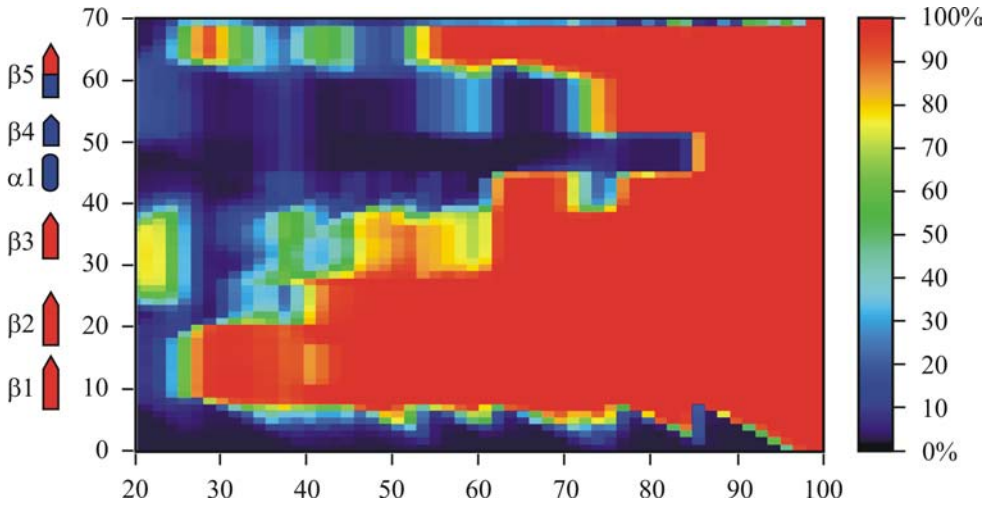


Figure 3.8: Folding pathway prediction for LCI. Hierarchy of structure formation for LCI computed with Fold-X software. The residue number is reported in the vertical axis. The horizontal axis corresponds to percentage of ordered residues. The secondary structure elements of the native protein are shown on the left side. The color code reflects the probability for a residue to be ordered during the folding process (dark blue corresponds to 0% and red to 100%).

Thus, the hierarchy of structure formation predicted using the Fold-X algorithm provides a description of the folding process of LCI consistent with the experimental data and predicts the main structural features of the highest energy barrier in the folding reaction. Since topology is characterized by the way secondary structure elements are connected in the sequence and are organized together, and the Fold-X algorithm does not use the disulfide bonds as an input, this successful prediction clearly indicates that secondary structure plays a crucial role in the folding of LCI.

3.3 Conclusions

The III-A intermediate constitutes the major kinetic trap in the oxidative folding of LCI. The NMR structure determined here represents one of the few available structural characterizations of folding intermediates isolated directly from a folding reaction. Despite the fact that this intermediate lacks a native disulfide bond, we show that it is a highly structured molecule with striking structural similarity to the native state. Comparison of native and intermediate structures allows deciphering why III-A accumulates along the folding reaction: it acts as a *disulfide insecure* intermediate, which protects both their native disulfide bonds and free cysteine residues from rearrangement and direct oxidation, respectively, in the interior of a highly folded protein conformation. Although III-A is a fully functional form that is formed quickly and efficiently in the LCI folding process, a conformational search for the formation of the last disulfide bond takes place while losing most of the tertiary structure already gained in this intermediate. The results of this study, together with previous stability data demonstrate that the fourth disulfide bond restricts conformational flexibility allowing a net gain in stability and structural specificity to the native form. This assumption makes sense taking into account that LCI is a protease inhibitor from leech saliva evolved to act in blood. In addition, here we show that theoretical approximations based on topological constrains predict accurately the main characteristics of the folding pathway of LCI. The overall data provide direct evidence for the importance of native-like interactions between elements of secondary structure in directing the folding of this protein and suggest that the forces governing the folding of this disulfide-rich protein do not differ much from those governing the folding of small none disulfide-rich proteins.

3.4 Experimental procedures

3.4.1 Sample preparation

The synthetic gene for LCI (Reverter et al., 1998) was cloned into the pBAT4 plasmid (Peranen et al., 1996), fused in frame to the OmpA signal sequence for periplasmic expression. Recombinant ^{15}N-labeled LCI was obtained by heterologous expression in *Escherichia coli* strain TG1. Fresh transformed cells were precultured in LB media containing 0.1 mg/mL carbenicillin at 37ºC; after 5 h, 0.5 mL of culture were centrifuged at 3000 rpm for 5 min, and cells were resuspended in 25 mL of M9 media containing $^{15}NH_4Cl$ as the only nitrogen source and 0.1 mg/mL carbenicillin. This second preculture was continued overnight, and the cells contained in 10 mL were used to inoculate 1 L of the same minimal media. Protein expression was induced in late phase (OD_{600} = 1.0) by adding IPTG (1 mM final concentration). LCI was purified from supernatant as described (Reverter et al., 1998). In summary, the protein was initially purified using a Sep-Pak C_{18} Cartridge (Waters), followed by anion-exchange chromatography on a TSK-DEAE 5PW column (Tosohaas), and by RP-HPLC on a 4.6 mm Protein C4 column (Vydac). The ^{15}N LCI labeling was almost heterogeneous (>99%) as deduced by MALDI-TOF MS analysis on a Bruker Ultraflex spectrometer.

The III-A folding intermediate was produced by oxidative refolding of LCI as previously reported (Arolas et al., 2004). Briefly, native ^{15}N-labeled LCI was reduced and denatured in 0.1 M Tris-HCl (pH 8.4) containing 8 M guanidine hydrochloride and 50 mM dithiothreitol, at 22ºC for 2 h. To initiate folding, the sample was passed through a PD-10 column (Sephadex-25, Amersham Biosciences), previously equilibrated with 0.1

M Tris-HCl (pH 8.4). Reduced and denatured LCI was recovered and immediately diluted to a final protein concentration of 0.5 mg/ml in the same Tris-HCl buffer. Folding intermediates of LCI were trapped after approximately 8 h of refolding by mixing aliquots of the sample with 2% trifluoroacetic acid (TFA). The trapped III-A intermediate was purified by RP-HPLC using the following conditions: solvent A was water containing 0.1% TFA and solvent B acetonitrile containing 0.1% TFA. A linear 20-40% gradient of solvent B was applied over 50 min, with a flow rate of 0.75 ml/min. The column used was a 4.6 mm Protein C4 (Vydac).

3.4.2 NMR experiments and structure calculation

Protein samples for NMR experiments were prepared by dissolving lyophilized ^{15}N LCI and ^{15}N III-A in either H_2O/D_2O (9:1 ratio by volume) or D_2O, at a concentration of 1 mM and pH 3.5. All experiments were carried out on a Bruker DRX-600 spectrometer, at 27ºC. The spectrometer was equipped with a triple resonance, triple gradient probe head. The TOCSY experiments (Rance, 1987) were performed with different mixing times between 20 and 40 ms, while the NOESY experiments (Kumar et al., 1980) were carried out with a mixing time of 120 ms. 4096 complex data points were recorded in the time domain t_2 and 700 increments in the t_1 domain. Water suppression was achieved using the WATERGATE pulse sequence (Piotto et al., 1992). The ^{1}H-^{15}N HSQC spectra (Bodenhausen et al., 1980) were also recorded at the same temperature with 2048 complex data points in the t_2 domain and 128 points in the t_1 domain, with 256 scans. The 3D NOESY-HSQC spectra (Marion et al., 1989) were performed with a mixing time of 100 ms, and 4096 complex data points were recorded in the t_3 domain. For the amide

proton exchange experiments lyophilized samples of ^{15}N native LCI and III-A were dissolved in D_2O at pH 3.5, 300 K. A series of consecutive 2D heteronuclear ^{1}H-^{15}N HSQC experiments were acquired with increased delays for up to 3 days.

The collected spectra were processed using the standard XWinNMR software package of Bruker and analyzed with the SPARKY software (Goddard and Kneller, University of California, San Francisco, 2000). Chemical shifts were assigned applying a combination of TOCSY/NOESY techniques. Peaks were classified according to their intensities as weak (3.8-5 Å), medium (2.8-3.8 Å), and strong (2.0-2.8 Å). A total of 20 structures were calculated by the simulated-annealing method with the program CNS (Brünger et al., 1998). Structure calculations were carried out essentially according to the basic protocol described previously (Holak et al., 1989; Schnuchel et al., 1993). For the final refinement, NOE tables were supplemented with constraints for several hydrogen bonds identified from the determined secondary structure. The atomic coordinates of the resulted 20 energy-minimized conformers for native LCI and III-A intermediate have been deposited with the Protein Data Bank, accession codes 1ZF1 and 1ZFL, respectively.

3.4.3 Folding pathway prediction

Fold-X is a program developed by Serrano's Group at the EMBL-Heidelberg to calculate the folding pathways of proteins and the effect of a point mutation on the stability of a protein. Fold-X exploits the previously described correlation between protein topology and folding mechanisms (Plaxco et al., 1998). The method is based on two simple assumptions: (1) only native interactions contribute significantly to the folding process;

and (2) each individual residue has two accessible states, native and disordered. As a novelty, the algorithm energy function includes terms that account for the difference in solvation energy for apolar and polar groups, the free energy difference between the formation of an intra-molecular H-bond compared to inter-molecular H-bond formation, and local and loop entropy. As described by the authors, the different energy terms taken into account in Fold-X have been weighted using empirical data obtained from protein engineering experiments. Fold-X uses a full atomic description of the structure of the proteins. We used the available online version of the software at http://foldx.embl.de employing the default software settings. The protein data bank accession number used for the calculations was 1DTV (average NMR structure of LCI; Reverter et al., 2000).

Summary

The work presented in this thesis was carried out in the Department of Structural Research at the Max Planck Institute for Biochemistry, Martinsried, between May 2002 and December 2005. The first part of this thesis involved the understanding of effects of inhibitors on protein-protein interactions -- the p53-MDM2 system, and the second part involved the structural characterization of an intermediate of the leech carboxypeptidase inhibitor.

The NMR chemical shift perturbation methods have been successfully used for mapping binding interfaces in proteins and for screening small molecule lead compounds in ligand-protein interactions using ^{15}N-labeled proteins. The NMR screening studies for lead compounds concentrated so far on binary interactions of lead compounds with small to middle size domains of target proteins, but not on two protein complexes. In order to study the effect of anatagonists on a two protein complex, we devised a new methodology based on observation of the $1/T_2$ relaxation rates of the proteins forming the complex. The size of one component should be small enough (less than ca. 15 kDa) to provide a good quality HSQC spectrum after ^{15}N or ^{13}C labeling of the protein. The size of the second component should be large enough so that the molecular weight of the

preformed complex is larger than ca. 40 kDa. The methodology was put to test using the p53-MDM2 complex.

The oncoprotein MDM2 (human murine double minute clone 2 protein) inhibits the tumor suppressor protein p53 by binding to the p53 transactivation domain, and thus inhibiting the G_1 cell cycle arrest and apoptosis. MDM2 deregulation has been observed to be very high in a variety of human tumors, except in cells that retain wild type p53. The disruption of the p53-MDM2 interaction with inhibitors has been shown to stabilize wild type p53 in cells and offers new therapeutics for cancer therapy. Three compounds that were reported to inhibit this interaction were tested using this methodology and only one of them was found to be a good inhibitor of the p53-MDM2 complex. Apart from monitoring inhibition we also can monitor the folded state of the protein on being released. This method should provide an important extension to the traditional "SAR by NMR" technique.

In a structurally oriented project in this thesis we characterized the III-A intermediate of the LCI (leech carboxypeptidase inhibitor) in order to understand the folding pathway of the native LCI. Leech carboxypeptidase inhibitor (LCI) is a 67-residue inhibitor of A/B metallo-carboxypeptidases isolated from the medical leech *Hirudo medicinalis*. LCI is a potent inhibitor of plasma carboxypeptidase B, also called thrombin-activatable fibrinolysis inhibitor (TAFI), which is a well-known attenuator of fibrinolysis The pro-fibrinolytic activity of LCI prove its possible use as an enhancer of the tissue plasminogen activator therapy of thrombosis. This intermediate was directly purified from the oxidative folding reaction using RP-HPLC, and its structure characterized by NMR spectroscopy and compared to that of native LCI. The obtained

results allow deciphering why this species acts as a major kinetic trap in the folding process of LCI. In addition, the integration of structural data with theoretical folding predictions provides insights into the crucial role played by secondary structure in directing the folding of LCI and has general implications for understanding the folding of other disulfide-rich proteins.

Zusammenfassung

Die vorliegende Arbeit wurde in der Abteilung Strukturforschung am Max-Planck-Institut für Biochemie Martinsried zwischen Mai 2002 und Dezember 2005 angefertigt. Der erste Teil der Arbeit behandelt die Effekte von Inhibitoren auf Protein-Protein Wechselwirkungen am Beispiel des p53-MDM2 Systems. Der zweite Teil behandelt die strukturelle Untersuchung eines Intermediats des Leech Carboxypeptidase Inhibitors.

Die NMR Methode der induzierte Änderung der chemische Verschiebung wurde erfolgreich eingesetzt um die Bindungsstelle in Proteinen zu bestimmen und für ein Suchverfahren für kleine Moleküle in Ligand-Protein Wechselwirkung mit Hilfe von ^{15}N-markierten Proteinen. Das NMR Suchverfahren für Schlüsselverbindungen beschränkte sich bis jetzt auf die binäre Interaktion der Schlüsselverbindung mit kleinen bis mittelgroßen Domänen von Zielproteinen. Proteinkomplexe wurden nicht untersucht. Um den Effekt eines Antagonists auf einen binären Protein-Protein Komplex zu untersuchen haben wir eine neue Methode, basierend auf der Beobachtung der $1/T_2$ Relaxation Rate der Komplexbildenden Proteine, entwickelt. Die Größe der ersten Proteinkomponente sollte klein genug (kleiner 15 kDa) sein um eine gutes HSQC-Spektrum nach ^{15}N oder ^{13}C Markierung der Proteine zu liefern. Die zweite Komponente sollte groß genug sein, damit der sich bildende Komplex eine Molekulargewicht von mehr als 40 kDa besitzt. Die Methode wurde am p53-MDM2 Komplex getestet.

Das Onkoprotein MDM2 (*murine double minute clone 2 protein*) inhibiert die Funktion des Tumor Supressor Proteins p53 in dem es an die Transaktivierungsdomäne von p53 bindet und damit zum Einen die Arretierung in der G_1 Phase des Zellzyklus und zum Anderen die durch p53 vermittelte Apoptose unterbindet. Die Deregulation von MDM2 wurde in vielen Tumoren beobachtet. Es konnte gezeigt werden, dass die Zerrstörung der p53-MDM2 Interaktion, in Tumoren die über funktionales Wildtyp p53 verfügen, mit Inhibitoren Wildtyp p53 in der Zelle stabilisiert und dieser Weg einen neuen therapeutischen Ansatz für die Behandlung von Krebs liefert. Drei Komponenten, von denen berichtet wurde, dass sie die Interaction zwischen p53 und MDM2 stören, wurden mit dieser Methode getestet, aber nur eine hat sich als ein guter Inhibitor für den p53-MDM2 Komplex herausgestellt. Neben der Inhibition kann auch der Faltungszustand des aus dem Komplex verdrängten Proteins beobachtet werden. Diese Methode stellt eine gute Erweiterung zur traditionellen „SAR by NMR“ Technik dar.

In dem mehr strukturell orientierten Teil der Arbeit haben wir das III-A Iintermediate von LCI (leech carboxypeptidase inhibitor) untersucht, um den Faltungsweg von nativen LCI zu verstehen. Leech Carboxypeptidase Inhibitor (LCI) ist ein 67-Reste Inhibitor von A/B Metallo-Carboxypeptidasen, isoliert aus *Hirudo medicinalis*. LCI ist ein potenter Inhibitor der Plasma Carboxypeptidase B, besser bekannt als *Thrombin-activatable fibrinolysis inhibitor* (TAFI), welcher ein bekannter Dämpfer der Fibrinolyse ist. Die Profibrinolytische Aktivität von LCI zeigt dessen Bedeutung als ein möglicher Verstärker der Gewebe *Pasminogen Aktivator Therapy* von Trombose.

Das Intermediat wurde direkt aus der oxidativen Faltungsreaktion mit Hilfe von RP-HPLC gereinigt und die Struktur mit NMR Spektroskopie charakterisiert und mit der Struktur von naticen LC1 verglichen. Die erhaltenen Resultate entschlüssen warum diese Spezie als eine kinetische Falle im Faltungsprozess von LCI wirkt. Die Verbindung der Strukturdaten mit theoretischen Faltungsvorhersagen gibt zusätzlich Anhaltspunkte für die entscheidende Rolle, die die Sekundärstruktur in der zielgerichteten Faltung von LCI spielt und erlaubt allgemeine Schlußfolgerungen für das Verständnis von anderen disulfidreichen Proteinen.

References

Alm E. and Baker D.; Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures, *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 11305-11310.

Anderson R. C. et al.; Affinity NMR: Decoding DNA binding, *J. Comb. Chem.* 1999, **1**, 69-72.

Arnold J. T. et al.; Chemical effects on nuclear induction signals from organic compounds, *J. Chem. Phys*. 1951, **19**, 507.

Arolas J. L. et al.; Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor, *J. Biol. Chem* 2004, **279**, 37261-37270.

Ayed A. et al.; Latent and active p53 are identical in conformation, *Nat. Struct. Biol.* 2001, **8**, 756-760.

Bax A. and Grzesiek S.; Methodological advances in protein NMR, *Acc. Chem. Res*. 1993, **26**, 131-138.

Bell S. et al.; p53 contains large unstructured regions in its native state, *J. Mol. Biol.* 2002, **322**, 917-927.

Bhunia A., et al.; Saturation transfer difference NMR and computational modeling of a sialoadhesin–sialyl lactose complex, *Carbohydr. Res.* 2004, **339**, 259-267.

Bleicher K. et al.; Diffusion edited NMR: Screening compound mixtures by affinity NMR to detect binding ligands to vancomycin, *J. Org. Chem.* 1998, **63**, 8486-8490.

Bloch F.; Nuclear induction, *Phys. Rev.* 1946, **70**, 460-474.

Blommers M. J. J. et al.; On the interaction between p53 and MDM2: transfer NOE study of a p53-derived peptide ligated to MDM2, *J. Am. Chem. Soc.* 1997, **119**, 3425-3426.

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

Bottger V. et al; Identification of novel mdm2 binding peptides by phage display *Oncogene* 1996, **13**, 2141-2147.

Bottger A. et al.; Design of a synthetic Mdm2-binding mini protein that activates the p53 response in vivo, *Curr. Biol.* 1997, **7**, 860-869.

Bouma B. N. and Meijers J. C. M.; Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U), *J. Thromb. Haemost* 2003, **1**, 1566-1574.

Brünger A. T. et al.; Crystallography & NMR system: a new software suite for macromolecular structure determination, *Acta Crystallog. D* 1998, **54**, 905-921.

Cemazar M. et al.; Oxidative folding intermediates with nonnative disulfide bridges between adjacent cysteine residues, *Proc. Natl. Acad. Sci. USA* 2003, **100**, 5754-5759.

Chang J. Y. et al.; A major kinetic trap for the oxidative folding of human epidermal growth factor, *J. Biol. Chem.* 2001, **276**, 4845-4852.

Chen A. and Shapiro M. J.; NOE pumping. A high-throughput method to determine compounds with binding affinity to macromolecules by NMR, *J. Am. Chem. Soc.* 2000, **122**, 414-415.

Chen J. et al.; Mapping of the p53 and MDM2 interaction domains, *Mol. Cell. Biol.* 1993, **13**, 4107-4114.

Chene P.; Inhibiting the p53–mdm2 interaction: an important target for cancer therapy *Nat. Rev. Cancer* 2003, **3**, 102-109.

Clementi C. et al.; How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta, *Proc. Natl. Acad. Sci. U.S.A.* 2000a, **97**, 5871-5876.

Clementi C. et al.; Topological and energetic factors: what determines the structural details of the transition state ensemble and 'en-route' intermediates for protein folding? An investigation for small globular proteins, *J. Mol. Biol.* 2000b, **298**, 937-953.

Clore G. M. and Gronenborn A. M; NMR structure determination of proteins and protein complexes larger than 20 kDa, *Curr. Opin. Chem. Biol.* 1998, **2**, 564-570.

Coles M. et al.; NMR-based screening technologies, *Drug Discov. Today* 2003, **8**, 803-810.

Creighton T. E.; Disulfide bonds as probes of protein folding pathways, *Methods Enzymol* 1986, **131**, 83-106.

Creighton T. E. et al.; The roles of partly folded intermediates in protein folding, *Faseb J.* 1996, **10**, 110-118.

Creighton T. E.; How important is the molten globule for correct protein folding?, *Trends Biochem. Sci.* 1997, **22**, 6-10.

Dalvit C. et al.; Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water, *J. Biol. NMR* 2000, **18**, 65-68.

Dalvit C. et al.; Fluorine-NMR competition binding experiments for high- throughput screening of large compound mixtures, *Comb. Chem. High Throughput Screen.* 2002, **5**, 605-611.

Dawson R. et al.; The N-terminal domain of p53 is natively unfolded, *J. Mol. Biol.* 2003, **332**, 1131-1141.

Dehner A. et al.; NMR Chemical Shift Perturbation Study of the N-Terminal Domain of Hsp90 upon Binding of ADP, AMP-PNP, Geldanamycin, and Radicicol, *Chembiochem* 2003, **4**, 870-877.

Diercks T. et al.; Applications of NMR in drug discovery, *Curr. Opin. Chem. Biol.* 2001, **5,** 285-91.

Doetsch V. and Wagner G.; New approaches to structure determination by NMR spectroscopy, *Curr. Opin. Str. Biol.* 1998, **8**, 619-623.

Dyson H. J. and Wright P. E.; Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol.* 2005, **6**, 197-208.

Eliezer D. et al.; Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin, *Biochemistry*, 2000, **39**, 2894-2901.

Fejzo J. et al.; The SHAPES strategy: an NMR-based approach for lead generation in drug discovery *Chem. Biol.* 1999, **6**, 755-769.

Fernandez C. O. et al.; NMR of α-synuclein−polyamine complexes elucidates the mechanism and kinetics of induced aggregation, *EMBO J.* 2004, **23**, 2039-2046.

Fischer P. M. and Lane D. P.; Small-molecule inhibitors of the p53 suppressor HDM2: have protein–protein interactions come of age as drug targets?, *Trends Pharmacol. Sci.* 2004, **25**, 343-346.

Freudenberger J. C. et al; Stretched poly(vinyl acetate) gels as NMR alignment media for the measurement of residual dipolar couplings in polar organic solvents, *Angew. Chem. Int. Ed.* 2005, **44**, 423-426.

Fry D. C. et al.; NMR structure of a complex between MDM2 and a small molecule inhibitor, *J. Biomol. NMR* 2004, **30**, 163-173.

Galatin P. S. and Abraham D. J.; A nonpeptidic sulfonamide inhibits the p53-mdm2 interaction and activates p53-dependent transcription in mdm2-overexpressing cells, *J. Med. Chem.* 2004, **47**, 4163-4165.

Galzitskaya O. V. and Finkelstein A. V.; A theoretical search for folding/unfolding nuclei in three-dimensional protein structures, *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 11299-11304.

Gehrmann J. et al.; Structure determination of the three disulfide bond isomers of alpha-conotoxin GI: a model for the role of disulfide bonds in structural stability, *J. Mol. Biol.* 1998, **278**, 401-415.

Goddard T. D. and Kneller D. G.; "SPARKY 3". (University of California, San Francisco) 2000.

Gordon S. L. and Wuthrich K.; Transient proton-proton Overhauser effects in horse ferrocytochrome c, *J. Am. Chem. Soc.* 1978, **100**, 7094-7096.

Guerois R. and Serrano L.; The SH3-fold family: experimental evidence and prediction of variations in the folding pathways, *J. Mol. Biol.* 2000, **304**, 967-982.

Guerois R. et al; Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations, *J. Mol. Biol.* 2002, **320**, 369-387.

Gutowsky S. H. et al.; Coupling among Nuclear Magnetic Dipoles in Molecules, *Phys. Rev.* 1951, **84**, 589-590.

Hahn E. L. and Maxwell D. E.; Chemical shift and field independent frequency modulation of the spin echo envelope, *Phys. Rev.* 1951, **84**, 1246-1247.

Hajduk P. J. et al.; One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules, *J. Am. Chem. Soc.* 1997, **119**, 12257-12261.

Hajduk P. J. et al.; NMR-based screening of proteins containing ^{13}C-labeled methyl groups, *J. Am. Chem. Soc.* 2000, **122**, 7898-7901.

Havel T. and Wuthrich K.; A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^{1}H-^{1}H proximities in solution, *Bull. Math. Biol.* 1984, **46**, 673-698.

Henrichsen D. et al.; Bioaffinity NMR spectroscopy: identification of an E-Selectin antagonist in a substance mixture by transfer NOE, *Angew. Chemie Int. Ed.* 1999, **38**, 98-102.

Holak T. A. et al.; Determination of the complete 3-dimensional structure of the trypsin-inhibitor from squash seeds in aqueous-solution by nuclear magnetic-resonance and a combination of distance geometry and dynamical simulated annealing, *J. Mol. Biol.* 1989, **210**, 635-648.

Hu Q. J. et al.; The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations, *EMBO J.* 1990, **9**, 1147-1155.

Hurle M. R. et al.; Comparison of solution structures of mutant bovine pancreatic trypsin inhibitor proteins using two-dimensional nuclear magnetic resonance, *Protein Sci.* 1992, **1**, 91-106.

Issaeva N. et al.; Small molecule RITA binds to p53, blocks p53-HDM-2 interaction and activates p53 function in tumors, *Nat. Med.* 2004, **10**, 1321-1328.

Iwaoka M. et al.; Regeneration of three-disulfide mutants of bovine pancreatic ribonuclease A missing the 65-72 disulfide bond: characterization of a minor folding pathway of ribonuclease A and kinetic roles of Cys65 and Cys72, *Biochemistry* 1998, **37**, 4490-4501.

Jahnke W. et al.; NMR reporter screening for the detection of high-affinity Ligands, *Angew. Chem. Int. Ed.* 2002, **41**, 3420-3423.

Juven-Gershon T. and Oren M.; Mdm2: The ups and downs, *Mol. Med.* 1999, **5,** 71-83.

Kaelin W. G. Jr. et al.; Definition of the minimal simian virus 40 large T antigen- and adenovirus E1A-binding domain in the retinoblastoma gene product, *Mol. Cell. Biol.* 1990, **10**, 3761-3769.

Kiel C. et al.; A detailed thermodynamic analysis of ras/effector complex interfaces, *J. Mol. Biol.* 2004, **340**, 1039-1058.

Klein C. and Vassilev L. T.; Targeting the p53-MDM2 interaction to treat cancer, *Brit. J. Cancer.* 2004, **91**, 1415-1419.

Klein J., et al.; Detecting binding affinity to immobilized receptor proteins in compound libraries by HR-MAS STD NMR, *J. Am. Chem. Soc.* 1999, **121**, 5336-5337.

Knight W. D.; Nuclear Magnetic Resonance Shift in Metals, *Phys. Rev.* 1949, **76**, 1259-1260.

Kramer F. et al; Residual dipolar coupling constants: An elementary derivation of key equations, *Concepts in Magnetic Resonance* 2004, **21A**, 10-21.

Kriwacki, R. W. et al.; Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: Conformational disorder mediates binding diversity, *Proc. Natl. Acad. Sci. U.S. A.* 1996, **93**, 11504-11509.

Kumar A. et al.; A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules, *Biochem. Biophys. Res. Commun* 1980, **95**, 1-6.

Kumar S. K. et al.; Design, synthesis, and evaluation of novel Boronic-chalcone derivatives as antitumor agents, *J. Med. Chem.* **2003**, 46, 2813-2815.

Kussie P. H.; et al.; Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain, *Science* 1996, **274**, 948-953.

Laity J. H. et al.; Structural characterization of an analog of the major rate-determining disulfide folding intermediate of bovine pancreatic ribonuclease A *Biochemistry* 1997, **36**, 12683-12699.

Lane D. P. and Hall P.A.; MDM2 - arbiter of p53's destruction, *Trends Biochem. Sci.* 1997, **22,** 372-374.

Lee H. et al.; Local structural elements in the mostly unstructured transcriptional activation domain of human p53, *J. Biol. Chem.* 2000, **275**, 29426-29432.

Lee J. O. et al.; Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7, *Nature* 1998, **391**, 859-865.

Lepre C. A. et al.; Theory and applications of NMR-based screening in pharmaceutical research, *Chem. Rev.* 2004, **104**, 3641-3676.

Lozano G. and Montes de Oca Luna; MDM2 Function, *Biochim. Biophys. Acta* 1998, **1377**, M55-M59.

Marion D. et al.; Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1 β, *Biochemistry* 1989, **28**, 6150-6156.

Markus, M. A. et al.; Solution structure of villin 14T, a domain conserved among actin severing proteins, *Protein Sci.* 1994, **3**, 70-81.

Mayer M. and Meyer B.; Characterization of ligand binding by saturation transfer difference NMR spectroscopy, *Angew. Chem. Int. Ed.* 1999, **38**, 1784-1788.

Meyer B. et al.; Screening mixtures for biological activity by NMR, *Eur. J. Biochem.* 1997, **246**, 705-709.

Meyer B. and Peters T.; NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors, *Angew. Chem. Int. Ed.* 2003, **42**, 864-890.

Momand J. et al.; The *mdm-2* oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation, *Cell* 1998, **69**, 1237-1245.

Mori S. et al.; Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new Fast HSQC (FHSQC) detection scheme that avoids water saturation, *J. Mag. Res. B* 1995, **108**, 94-98.

Munoz V. and Eaton W. A.; A simple model for calculating the kinetics of protein folding from three-dimensional structures, *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 11311-11316.

Noda Y. et al.; NMR structural study of two-disulfide variant of hen lysozyme: 2SS [6-127, 30-115] – A disulfide intermediate with a partly unfolded structure, *Biochemistry* 2002, **41**, 2130-2139.

Oliner J. D. et al.; Wild-type p53 activates transcription in vitro, *Nature* 1992, **358**, 80-83.

Oliner J. D. et al.; Oncoprotein MDM2 conceals the activation domain of tumor suppressor p53, *Nature* 1993, **362**, 857-860.

Pellecchia, M. et al.; SEA-TROSY (Solvent Exposed Amides with TROSY): A method to resolve the problem of spectral overlap in very large proteins, *J. Am. Chem. Soc.* 2001, **123**, 4633-4634.

Pellecchia M. et al.; NMR in drug discovery, *Nat. Rev. Drug Discovery* 2002, **1**, 211-219.

Peranen J. et al.; T7 vectors with modified T7 lac promoter for expression of proteins in *Escherichia coli*, *Anal. Biochem* 1996, **236**, 371-373.

Pervushin K. et al.; 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, *Proc. Natl. Acad. Sci. U.S.A.* 1997, **94**, 12366-12371.

Picksley S. M. et al; Immunochemical analysis of the interaction of p53 with MDM2;--fine mapping of the MDM2 binding site on p53 using synthetic peptides, *Oncogene* 1994, **9**, 2523-2529.

Piotto M. et al.; Gradient-tailored excitation for single quantum NMR spectroscopy in aqueous solution, *J. Biomol. NMR* 1992, **6**, 661-665.

Plaxco K. W. et al.; Contact order, transition state placement and the refolding rates of single domain proteins, *J. Mol. Biol.* 1998, **277**, 985-994.

Prestegard J. H. et al.; Nuclear magnetic resonance in the era of structural genomics, *Biochemistry* 2001, **40**, 8677-8685.

Purcell E. M. et al.; Resonance absorption by nuclear magnetic moments in a solid, *Phys. Rev*. 1946, **69**, 37-38.

Radford S. E. et al.; The folding of hen lysozyme involves partially structured intermediates and multiple pathways, *Nature* 1992, **358**, 302-307.

Rance M.; Improved techniques for homonuclear rotating frame and isotropic mixing experiments, *J. Magn. Reson.* 1987, **74**, 557-564.

Rehm T. et al.; Application of NMR in structural proteomics: screening for proteins amenable to structural analysis, *Structure* 2002, **10**, 1613-1618.

Riek R. et al.; TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution, *Trends Biochem. Sci.* 2000, **25**, 462-468.

Reese M. L. and Doetsche V.; Fast mapping of protein-protein interfaces by NMR spectroscopy, *J. Am. Chem. Soc.* 2003, **125**, 14250-14251.

Reverter D., et al.; A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization, *J. Biol. Chem.* 1998, **273**, 32927-32933.

Reverter D. et al.; Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2, *Nat. Struct. Biol.* 2000, **7**, 322-328.

Ross A. et al.; Automation of NMR measurements and data evaluation for systematically screening interactions of small molecules with target proteins, *J. Biol. NMR* 2000, **16**, 139-146.

Salamanca S. et al.; The unfolding pathway of leech carboxypeptidase inhibitor, *J. Biol. Chem.* 2002, **277**, 17538-17543.

Salamanca S. et al.; Major kinetic traps for the oxidative folding of leech carboxypeptidase inhibitor, *Biochemistry* 2003, **42**, 6754-6761.

Sato A. et al.; Three-dimensional solution structure of a disulfide bond isomer of the human insulin-like growth factor-I, *J. Pept. Res.* 2000, **56**, 218-230.

Sattler M. et al.; Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients**,** *Prog. Nucl. Magn. Reson. Spectrosc.* 1999, **34**, 93-158.

Saunders M. et al.; The nuclear magnetic resonance spectrum of ribonuclease, *J. Am. Chem. Soc.* 1957, **79**, 3289-3290.

Scheraga H. A. et al.; Bovine pancreatic ribonuclease A: oxidative and conformational folding studies, *Methods Enzymol* 2001, **341**, 189-221.

Schnuchel A. et al.; Structure in solution of the major cold shock protein from *Bacillus Subtilis*, *Nature* 1993, **364**, 169-171.

Schon O. et al.; Molecular mechanism of the interaction between MDM2 and p53, *J. Mol. Biol.* 2002, **323**, 491-501.

Schon O. et al.; Binding of p53-derived ligands to MDM2 induces a variety of long range conformational changes, *J. Mol. Biol.* 2004, **336**, 197-202.

Seidel R. D. et al.; Structural perturbations in human ADP ribosylation factor-1 accompanying the binding of phosphatidylinositides, *Biochemistry* 2004, **43**, 15393-15403.

Senn H. et al.; 15N-labeled P22 c2 repressor for nuclear magnetic resonance studies of protein-DNA interactions, *Eur. Biophys. J.* 1987, **14**, 301-306.

Shimotakahara S. et al.; NMR structural analysis of an analog of an intermediate formed in the rate-determining step of one pathway in the oxidative folding of bovine pancreatic ribonuclease A: automated analysis of 1H, 13C, and 15N resonance assignments for wild-type and [C65S, C72S] mutant forms, *Biochemistry* 1997, **36**, 6915-6929.

Shuker S. B. et al.; Discovering high-affinity ligands for proteins: SAR by NMR, *Science* 1996, **274**, 1531-1534.

Stockman B. J. and Dalvit C.; NMR screening techniques in drug discovery and drug design, *Prog. NMR Spectrosc.* 2002, **41**, 187-231.

Stoll R. et al.; Chalcone derivatives antagonize interactions between the human oncoprotein MDM2 and p53, *Biochemistry* 2001, **40**, 336-344.

Takeda M. and Jardetzky O.; Proton magnetic resonance of simple amino acids and dipeptides in aqueous solution, *J. Chem. Phys.* 1957, **26**, 1346-1347.

Talluri S. et al.; Structural characterization of a three-disulfide intermediate of ribonuclease A involved in both the folding and unfolding pathways. Biochemistry 1994, **33**, 10437-10449.

van den Berg B. et al.; Characterisation of the dominant oxidative folding intermediate of hen lysozyme, *J. Mol. Biol.* 1999, **290**, 781-796.

van Mierlo C. P. M. et al.; Two-dimensional ^{1}H nuclear magnetic resonance study of the (5-55) single-disulfde folding intermediate of bovine pancreatic trypsin inhibitor, *J. Mol. Biol.* 1991, **222**, 373-390.

van Mierlo C. P. M. et al.; Partially folded conformation of the (30-51) intermediate in the disulfde folding pathway of bovine pancreatic trypsin inhibitor, ^{1}H and ^{15}N resonance assignments and determination of backbone dynamics from ^{15}N relaxation measurements, *J. Mol. Biol.* 1993, **229**, 1125-1146.

van Mierlo C. P. et al.; 1H NMR analysis of the partly-folded non-native two-disulphide intermediates (30-51,5-14) and (30-51,5-38) in the folding pathway of bovine pancreatic trypsin inhibitor, *J. Mol. Biol.* 1994, **235**, 1044-1061.

Vanwetswinkel S. et al.; TINS, Target Immobilized NMR Screening: An efficient and sensitive method for ligand discovery, *Chem. Biol.* 2005, **12**, 207-216.

Vassilev L. T. et al.; In vivo activation of the p53 pathway by small-molecule antagonists of MDM2, *Science* 2004, **303**, 844-848.

Wagner G. and Wuthrich K.; Truncated driven nuclear overhauser effect (TOE). A new technique for studies of selective 1H-1H Overhauser effects in the presence of spin diffusion, *J. Mag. Res.* 1979, **33**, 675-680.

Wagner G. and Wuthrich K.; Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra : Basic pancreatic trypsin inhibitor, *J. Mol. Biol.* 1982, **155**, 347-366.

Wang W. et al.; A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor, *J. Biol. Chem.* 1998, **273**, 27176-27181.

Wang Z. X.; An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule, *FEBS Lett.* 1995, **360**, 111-114.

Wasylyk C. et al.; p53 mediated death of cells overexpressing MDM2 by an inhibitor of MDM2 interaction with p53, *Oncogene* 1999, **18**, 1921-1934.

Weissman J. S. and Kim P. S. Reexamination of the folding of BPTI: predominance of native intermediates, *Science* 1991, **253**, 1386-1393.

Welker E. et al.; Structural determinants of oxidative folding in proteins, *Proc. Natl. Acad. Sci. U.S.A.* 2001, **98**, 2312-2316.

Williamson M. P. et al.; Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry, *J. Mol. Biol.* 1985, **182**, 295-315.

Wüthrich K. et al.; Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances, *J. Mol. Biol.* 1984, **180**, 715-740.

Wüthrich K.; NMR of Proteins & Nucleic Acids (Wiley, New York), 1986.

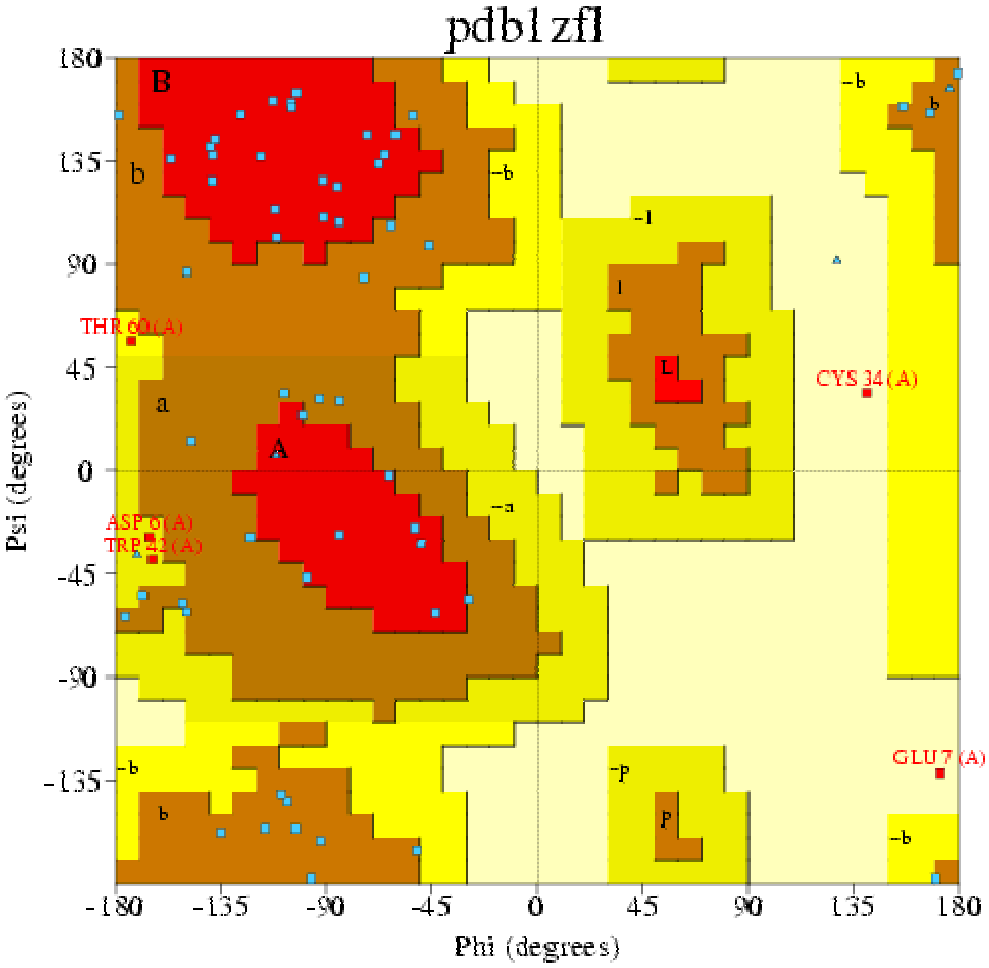
Wüthrich K.; NMR Studies of Structure and Function of Biological Macromolecules (Nobel Lecture), *Angew. Chem. Int. Ed.* 2003, **42**, 3340-3363.

Yokota A. et al.; NMR characterization of three-disulfide variants of lyzozyme, C64A/C80A, C76A/C94A, and C30A/C115A – A marginally stable state in folded proteins, *Biochemistry* 2004, **43**, 6663-6669.

Zartler E. R. et al.; RAMPED-UP NMR: Multiplexed NMR-based screening for drug discovery, *J. Am. Chem. Soc.* 2003, **125**, 10941-10946.

Appendix

Ramachandran plot



1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	23	44.2%**
Additional allowed regions	[a,b,l,p]	24	46.2%
Generously allowed regions	[~a,~b,~l,~p]	3	5.8%
Disallowed regions	[XX]	2	3.8%*
Non-glycine and non-proline residues		52	100.0%
End-residues (excl. Gly and Pro)		1	
Glycine residues		5	
Proline residues		9	
Total number of residues		67	

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

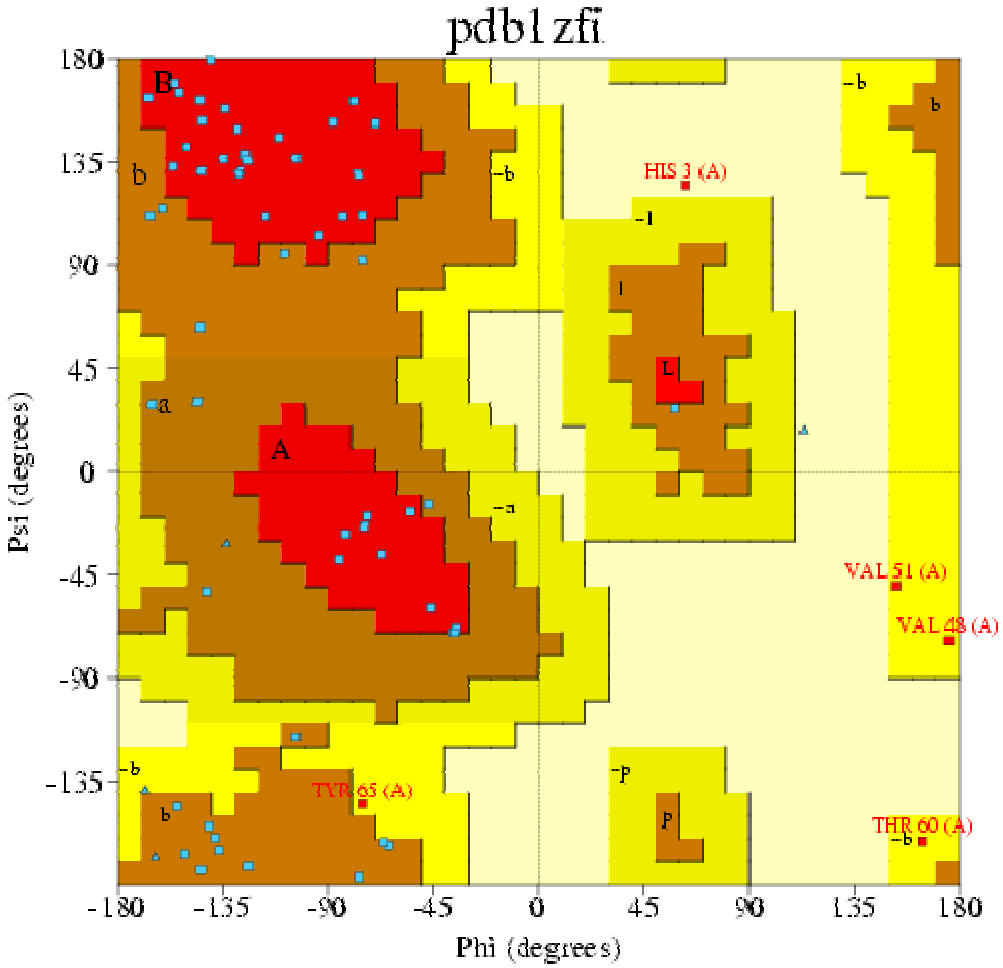
2. G-Factors

Parameter	Score	Average Score
Dihedral angles:-		
Phi-psi distribution	**-2.04****	
Chi1-chi2 distribution	**-1.46****	
Chi1 only	-0.51*	
Chi3 & chi4	0.42	
Omega	0.61	
		-0.70*
Main-chain covalent forces:-		
Main-chain bond lengths	0.54	
Main-chain bond angles	0.53	
		0.53
OVERALL AVERAGE		-0.22

G-factors provide a measure of how **unusual**, or out-of-the-ordinary, a property is.

Values below -0.5* - unusual
Values below **-1.0**** - highly unusual

Ramachandran plot



1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	30	57.7%**
Additional allowed regions	[a,b,l,p]	17	32.7%
Generously allowed regions	[~a,~b,~l,~p]	4	7.7%
Disallowed regions	[XX]	1	1.9%*
Non-glycine and non-proline residues		52	100.0%
End-residues (excl. Gly and Pro)		1	

Glycine residues	5
Proline residues	9
Total number of residues	67

Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and *R*-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favoured regions [A,B,L].

2. G-Factors

Parameter	**Score**	**Average Score**
Dihedral angles:-		
Phi-psi distribution	**-1.70****	
Chi1-chi2 distribution	**-1.70****	
Chi1 only	-0.67*	
Chi3 & chi4	0.26	
Omega	0.56	
		-0.68*
Main-chain covalent forces:-		
Main-chain bond lengths	0.52	
Main-chain bond angles	0.51	
		0.52
OVERALL AVERAGE		-0.22

G-factors provide a measure of how **unusual**, or out-of-the-ordinary, a property is.

Values below -0.5* - unusual
Values below **-1.0**** - highly unusual

TABLE 1: ^{15}N, HN and αH backbone assignments (in ppm) for the native LCI and the III-A intermediate

	N LCI			III-A		
Residue	^{15}N	HN	αH1/2	^{15}N	HN	αH1/2
P5	-	-	4.59	-	-	4.46
D6	122.7	8.30	5.00	122.4	8.29	5.06
E7	124.4	8.88	4.61	124.2	8.71	4.65
S8	119.6	8.02	5.75	120.0	8.13	5.61
F9	121.4	9.52	5.29	122.5	9.38	5.27
L10	128.5	9.11	5.02	128.4	9.08	5.03
C11	126.6	9.22	5.47	126.9	9.24	5.53
Y12	126.5	8.92	4.59	127.2	8.99	4.57
Q13	127.8	8.57	4.91	127.9	8.58	4.92
D15	112.5	8.38	5.09	112.4	8.39	5.11
Q16	119.6	7.46	4.62	119.4	7.45	4.62
V17	121.0	8.53	4.79	121.1	8.52	4.79
C18	123.3	8.86	5.64	123.6	8.86	5.62
C19	123.4	9.38	5.82	123.8	9.31	5.66
F20	124.7	8.53	5.10	125.5	8.67	4.93
I21	128.5	9.13	4.75	130.4	9.17	4.70
C22	124.7	9.59	6.09	125.3	9.28	5.34
R23	123.9	9.31	4.91	124.7	8.78	4.86
G24	117.2	9.24	4.02/3.83	117.2	9.14	4.03/3.81
A25	120.0	7.34	4.37	120.8	7.45	4.41
A26	122.4	7.55	4.20	123.9	7.77	4.46
L28	128.9	8.37	4.58	126.9	8.34	4.57
S30	111.4	7.29	4.18	111.4	7.38	4.20
E31	121.9	8.01	4.45	122.2	7.91	4.47
G32	106.3	7.40	3.66/4.45	106.1	7.45	3.68/4.44
E33	125.5	8.30	4.47	125.4	8.28	4.62
C34	129.0	8.53	6.04	129.0	8.62	5.97
N35	122.2	8.71	5.44	122.8	8.83	5.51
H37	128.0	8.23	3.88	126.6	8.20	3.92

T39	122.2	8.12	4.03	122.3	8.12	3.90
A40	130.3	8.06	3.75	127.9	8.07	3.66
W42	114.9	7.16	4.22	113.1	7.04	4.28
C43	125.5	6.16	3.81	125.2	6.35	3.91
R44	119.3	7.26	3.64	120.0	7.35	3.76
E45	115.9	7.85	3.98	116.8	7.86	4.02
G46	104.4	7.20	3.77/4.33	105.2	7.35	3.79/4.21
A47	129.9	8.37	4.35	129.6	8.32	4.40
V48	120.7	7.70	4.03	120.4	7.68	4.02
E49	124.4	7.97	4.56	124.6	7.97	4.54
W50	129.2	8.68	4.38	128.0	8.51	4.44
V51	123.9	9.12	4.89	-	-	-
Y53	129.0	8.37	4.19	127.2	8.33	4.30
S54	123.2	8.61	3.76	123.3	8.66	3.76
T55	121.7	8.26	4.10	121.1	8.20	4.31
G56	114.6	8.51	3.89/4.31	114.7	8.62	3.92/3.79
Q57	127.6	8.48	5.53	127.7	8.50	4.76
C58	119.4	9.30	6.17	123.2	8.64	4.43
R59	123.8	9.26	4.75	121.4	8.61	4.01
T60	120.5	8.99	4.87	120.5	8.80	4.90
T61	121.5	7.40	4.60	122.3	7.47	4.63
C62	125.2	8.14	5.79	123.6	8.03	5.76
I63	120.3	8.76	4.60	120.6	8.79	4.61
Y65	123.4	7.55	4.47	123.2	7.57	4.48
V66	126.2	7.81	4.03	125.7	7.83	4.05