Technische Universität München
Institut für Organische Chemie und Biochemie

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

NMR high resolution spectroscopy of medically relevant proteins and their antagonists

Marcin Krajewski

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. Steffen J. Glaser
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Horst Kessler
	2. apl. Prof. Dr. Luis Moroder

Die Dissertation wurde am 08.02.2006 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 23.05.2006 angenommen.

To my parents, and my wife Kasia.

Acknowledgements

I would like to thank all the people who have contributed to this work.

My supervisor Dr. Tad A. Holak for his scientific guidance, and many corrections to this thesis.

I have to greatly acknowledge all my groupmates. It is impossible to define or overestimate their input to this work, particularly protein expression and purification.

I wish to thank Dr. Joma Kanikadu Joy and Dr. Sudipta Majumdar for correcting my language pitfals in this thesis, and Ulli Rothweiler and my wife Kasia for translating necessary parts to german.

My thanks to Dr. hab. Christian Renner, who was always ready to answer my questions about NMR.

To Dr. Markus Seifert for intruducing and teaching the practice of NMR spectroscopy.

To Dr. Till Rehm for many practical comments about NMR.

To Prof. Dr. Horst Kessler, for being my doktorfather.

To Dr. Reiner Haessner and PD Dr. Gerd Gemmecker for technical assistance with the spectrometers at TU Garching and Ms. Beate Diaw, for scheduling time on the spectrometers there.

My apologies to all others who I have not mention by name I am indebted to them for their help.

Publications

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

Marcin Krajewski, Przemyslaw Ozdowy, Loyola D'Silva, Ulli Rothweiler, and Tad A. Holak. NMR indicates that the small molecule Rita does not block p53-mdm2 binding in vitro. *Nature Medicine*, 11(11):1135–1137, November 2005.

Loyola D'Silva, Przemyslaw Ozdowy, **Marcin Krajewski**, Ulli Rothweiler, Mahavir Singh, and Tad A. Holak. Monitoring the effects of antagonists on protein–protein interactions with NMR spectroscopy. *Journal of the American Chemical Society*, 127(38):13220–13226, September 2005.

Mahavir Singh, **Marcin Krajewski**, Alexandra Mikolajka, and Tad A. Holak. Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences. *The Journal of Biological Chemistry*, 280(45):37868–37876, November 2005.

Grzegorz Dubin, **Marcin Krajewski**, Grzegorz Popowicz, Justyna Stec-Niemczyk, Matthias Bochtler, Jan Potempa, Adam Dubin, and Tad A. Holak. A novel class of cysteine protease inhibitors: Solution structure of staphostatin A from *Staphylococcus aureus*. *Biochemistry*, 42:13449–13456, 2003.

Grzegorz Dubin, Grzegorz Popowicz, **Marcin Krajewski**, Jan Potempa, Adam Dubin, and Tad A. Holak. Letter to the editor: ^{1}H ^{15}N and ^{13}C NMR resonance assignments of staphostatin A, a specific *Staphyloccocus aureus* cysteine proteinase inhibitor. *Journal of Biomolecula NMR*, 28:295–296, 2004.

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

Contents

Chapter 1

Theoretical basics of biomolecular NMR

1.1 Introduction

1.1.1 NMR phenomenon

The magnetic resonance phenomenon can be observed when a nucleus with a nonzero spin is placed in an external magnetic field. The equilibrium distribution of spin states results in a net value of the macroscopic magnetization along the external field, traditionally the $\vec{z}$ direction. If by some operation the magnetization possesses a component perpendicular to the $\vec{z}$ direction, its evolution can be described as a precession around the $\vec{z}$ axis, with some dumping due to energy dissipation. The precession frequency depends on the gyromagnetic ratio of the particle and the local magnetic field. The latter acts as a perfect probe for local environment. For nuclei important in biomolecular NMR, the effect of the local magnetic field is usually within 200 ppm (parts per million) for ^{13}C, 20 ppm for ^{1}H, and 250 for ^{15}N.

1.1.2 CW-NMR and pulsed FT-NMR

There has been two different approaches to observe NMR resonance. One can irradiate a sample by sweeping frequencies and monitoring energy absorbed

by the sample in the so-called *continuous-wave* NMR experiment.

Pulsed Fourier transform NMR uses a different measurement paradigm. A short pulse, or a series of pulses, is applied to the sample. The method requires a pulse producing transmitter and a receiver together with a computer for FT. This is now the most popular method because it offers the highest signal to noise ratio in a given time. Additionally it allows for higher flexibility in the experiment design.

1.1.3 Solution vs. solid-state NMR

Although NMR on proteins is dominated by solution state measurements, the solid state NMR is becoming increasingly popular. Since ever larger proteins are of interest in structural biology, the advantage of solid-state NMR (ssNMR), like for example no limitation due to T_2 on protein size or no problems with protein solubility (important for membrane proteins) ssNMR starts to compete with the solution state NMR despite strict necessity of ^{13}C labeling, a less sensitive nucleus used for detection, and mechanical requirements for the equipment. The possibility of a complete de novo structure determination purely based on the magic angle spinning (MAS) ssNMR recoupling techniques was first demonstrated for small insoluble peptides [1, 2]. The techniques have recently been extended to soluble proteins that have been studied in the solid-state [3]. Recently, a study on protein-ligand interactions was also reported using a ^{13}C 1D MAS and ^{13}C-^{13}C dipole-dipole correlation spectra [4].

1.1.4 The NMR spectrometer design

The NMR spectrometer consists of a strong magnet (typically above 10 T, which produces a homogeneous magnetic field in the sample volume), a transmitter/receiver system, and a computer for data processing. The spectrometer is usually equipped with a gradient system, which allows for application

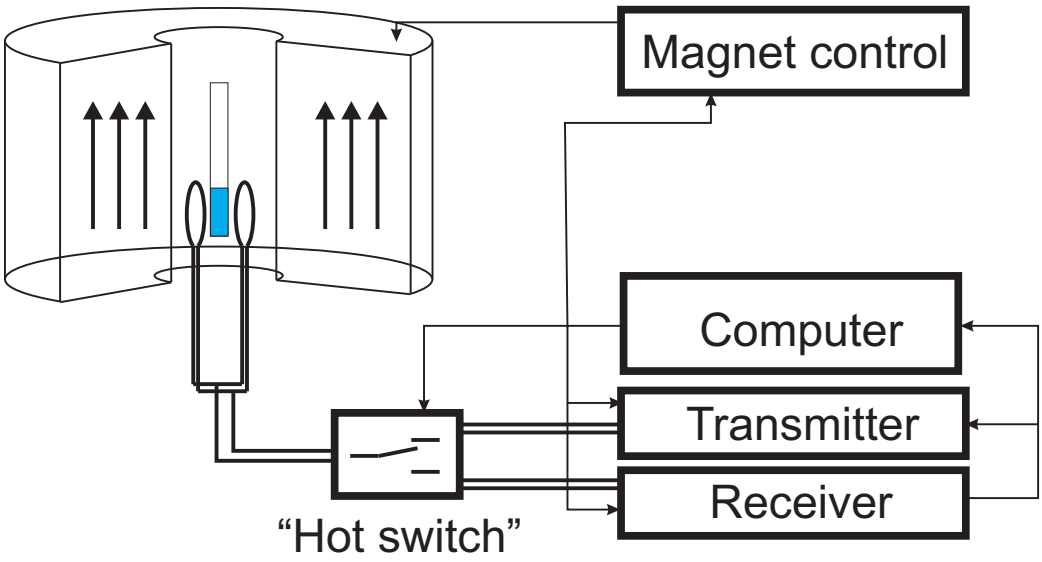

Figure 1.1: The schematics of a high resolution NMR spectrometer.

of a well-defined spatial field inhomogeneity at the time of the experiment. To obtain such high fields, the magnet consists of a superconducting coil, or a set of coils, immersed in liquid helium. The magnet is charged at the time of installation, then the superconducting loop is closed and the power supply disconnected usually for years. Since the coil current decays slowly (magnet drift), to ensure accuracy and referencing of the spectrum, a “lock” system is installed. Also a user-controllable set of room-temperature shim coils is present to compensate for small inhomogeneities of the magnetic field for different samples. Since an NMR frequency often is temperature dependent, a temperature control unit is also integrated into the system.

Gradients

Introduction of gradients to NMR spectrometers has opened a new world for NMR. The main achievement has been magnetic resonance imaging.

Modern probeheads allow the user to apply field gradients of duration as short as hundreds of microseconds and strength of 50 G/cm or more. Gradient coils are incorporated into the probehead. Most probeheads offer at least

z-gradients. Currents induced in the conducting elements of the probehead are side effects of gradients, and usually delays of few microseconds are required afterwords to avoid unwanted effects. They are more prominent for x or y gradients because of strong interference with transmitter/receiver coils.

Gradients are useful not only at the time of the NMR experiment, but also can be used for an automated shimming procedure, thus shortening the time of the experiment setup.

Cryo-probehead

The cryo-probehead replaces the "traditional" probehead and consists of at least one superconducting coil, an integrated low operating temperature (77 K) preamplifier, and a hot switch, plus an external helium compressor/liquidizer. The S/N is up to four-fold higher than that of "traditional" probeheads.

1.1.5 Determinants of spectrum resolution

Relaxation processes

An intrinsic limitation for the spectral resolution is the natural linewidth. Larger molecules give broader lines. Main relaxation mechanisms for macromolecules are magnetic dipolar and quadrupolar interactions, chemical shift anisotropy (CSA), and scalar coupling interactions; the dipolar interactions and CSA being dominant for spin $\frac{1}{2}$ nuclei in diamagnetic macromolecules.

Tumbling. Random Brownian rotations of molecules in solution sharpens resonance lines, because it averages to zero the CSA and D-D interaction Hamiltonians. The faster the molecule tumbles, sharper the NMR resonance lines are. In consequence of the Boltzmann energy distribution and energy equipartition rule, larger molecules tumble slower. With larger masses the orientational auto-correlation functions drop slower with time, what is

manifested in shorter transverse relaxation times, and consequently broader spectral lines. Larger proteins also contain more interacting spins, causing dramatic line broadening for molecules larger than ca. 30 kDa. Not only are their resonance broader, but more crosspeaks appear in the spectrum, resulting in spectral overlap.

Local magnetic field gradients

Since NMR detects the mean value of magnetization, differences in the local magnetic field that are caused by field inhomogeneity can severely obscure the spectrum. Although shimming usually minimizes field inhomogeneity to negligible levels, in cases of inhomogeneous samples (bubbles, precipitates, huge thermal gradients) broaden linewidths may be observed.

Acquisition time

The resolution of a FT spectrum is proportional to the total acquisition time of the FID signal. Longer acquisition results in higher resolution. Nevertheless there is no use of sampling the FID when the signal is indistinguishable from noise. In such cases the relaxation rate limits the resolution. It is common in multidimensional experiments that NMR signals cannot be sampled long, and some techniques, like for example *linear prediction*, are used at the time of data processing to gain resolution.

1.1.6 Basics for analysis of NMR experiments

Density matrix

Let $c_m = \langle m \mid \Psi \rangle$ denote an m-th component of a state vector $\mid \Psi \rangle$ of the NMR system. The knowledge of a projection operator

$$P_{nm} = \langle n \mid P \mid m \rangle = c_n c_m^* \tag{1.1}$$

is sufficient to calculate the expectation value of an operator A in state $\mid \Psi\rangle$:

$$\langle A\rangle = \sum_{nm} c_n c_m^* \langle m \mid A \mid n\rangle = Tr\{PA\} \tag{1.2}$$

The sample/molecule is divided into spin systems, which are treated separately, and the mutual interaction is hidden in a magic "lattice", which represents everything apart from the spin-system considered. The spin-system is described by a state $\mid \Psi\rangle$ and $\mathcal{P}(\Psi)$, the probability that a particular state is realized by the spin-system of interest. The average expectation value of A will be given by:

$$\langle A\rangle = \int \mathcal{P}(\Psi) \sum_{nm} c_n c_m^* \langle m \mid A \mid n\rangle = Tr\{\sigma A\} \tag{1.3}$$

where the matrix elements of a new, Hermitian, *density operator* σ

$$\sigma_{nm} = \int \mathcal{P}(\Psi) c_n c_m^* \tag{1.4}$$

form the so-called *density matrix*.

The time evolution of a density operator is given by the Liouville-von Neuman equation:

$$\frac{d}{dt}\sigma(t) = i\left[\sigma(t), H(t)\right] \tag{1.5}$$

If the Hamiltonian is time independent the solution has the form:

$$\sigma(t) = \exp(-iHt)\sigma(0)\exp(iHt) \tag{1.6}$$

Product operator formalism

If one expresses $\sigma = \sum_k s_k S_k$ in terms of angular momentum operators, the calculation becomes straightforward

$$\sigma(t) = \sum_k s_k(0)\exp(-iHt)S_k\exp(iHt) \tag{1.7}$$

and limits to successive application of transformation rules for S_k. The effect of spin evolution, an RF pulse, spin-spin interaction can be expressed as

a set of rules. Analysis of pulse sequences is then simplified to a mechanical application of these rules at each step of a pulse sequence. This can still lead to very large expressions, however, majority of important experiments can be analyzed this way. Exceptions include experiments where a trail of pulses and delays is applied and the delays between pulses are similar in duration to pulse-length; thus pulses can no longer be treated as infinitesimally short and this must be taken into account. Examples are: WATERGATE-5 pulse-sequence, TOCSY-type experiments and long pulses.

The rules for calculations

- The operators for different spins commute
- $[I_l, I_k] = iI_m$ if $\{l, k, m\} = \{x, y, z\}, \{y, z, x\}, \{z, x, y\}$
- The effect of a Hamiltonian often can be calculated from:

$$\exp(-iHt)S_k \exp(iHt) = S_k \cos(Ht) + [H, S_k] \sin(\theta)$$

  providing that $[S_k, H] = iM_k$ for all cyclic permutations of S_k, H, M_k.

NMR interaction Hamiltonians

An RF pulse. The Hamiltonian of a short pulse of phase x for nuclei described by I is given by:

$$H = \alpha I_x \tag{1.8}$$

where angle α is a parameter depending on the power and time of irradiation.

Scalar coupling. Scalar coupling is observed between nuclei connected by one or more bonds, the coupling strength being strongly dependent on the bond type and number of bonds. It can be expressed by angular momentum operators as:

$$H = 2\pi J_{IS} I \cdot S = 2\pi J_{IS}(I_x S_x + I_y S_y + I_z S_z) \tag{1.9}$$

If $J_{IS} \ll \mid \omega_I - \omega_S \mid$, it can be well approximated by:

$$H = 2\pi J_{IS} I_z S_z \tag{1.10}$$

Under this interaction the initial coherence of I_x will oscillate with the angular frequency of $2\pi J_{IS}$, change with time to $2\mathrm{I}_y\mathrm{S}_z$ and back, if left undisturbed. This oscillation will produce splitting of the NMR signal to a doublet if two spins are coupled or multiplets if more than one coupling is active. The frequency of these oscillations depends on the coupling strength (J_{IS}), and for H-N pair, the doublet separation is in the range of 80-100 Hz. This effect is commonly used to transfer magnetization from one nucleus to another by series of pulses and delays. The transfer techniques include COSY, INEPT, refocused INEPT, DEPT, TOCSY, etc. The value of the coupling constant can also give information about the structure, for example, about the dihedral angles in the polypeptide chain [5].

Dipolar coupling. The dipole-dipole interaction is observed between spins that are near in space. The strong distance dependence makes this interaction of an ultimate power in NMR, particularly in the structure determination process or backbone dynamics measurements. The dipole-dipole interaction can be measured in NOESY-type experiments.

The dipole-dipole interaction Hamiltonian is:

$$H_{1,2}^{DD} = -\mu_0 \frac{\gamma_1 \gamma_2 \hbar}{4\pi r_{1,2}^3} \left\{ \frac{3}{r_{1,2}^2} \left(\mu_1 r_{1,2}\right) \left(\mu_2 r_{1,2}\right) - \mu_1 \mu_2 \right\} \tag{1.11}$$

In the secular approximation when only frequency independent terms are retained this can be simplified with the use of angular momentum operators to:

$$H_{I,S}^{DD} = -\mu_0 \frac{\gamma_1 \gamma_2 \hbar}{4\pi r_{I,S}^3} \frac{3\cos^2(\beta) - 1}{2} \left(3 I_z S_z - I \cdot S\right) \tag{1.12}$$

for a homonuclear case, and

$$H_{I,S}^{DD} = -\mu_0 \frac{\gamma_1 \gamma_2 \hbar}{4\pi r_{I,S}^3} \frac{3\cos^2(\beta) - 1}{2} 2 I_z S_z \tag{1.13}$$

for a heteronuclear case.

Chemical shift. The chemical shift Hamiltonian:

$$H_I^{CS} = \gamma_I \left(I_x \sigma_{xx} + I_y \sigma_{yy} + I_z \sigma_{zz} \right) B_0 \tag{1.14}$$

can be expressed as:

$$\omega_I = \omega_0 (1 - \sigma_{iso}) + \frac{\delta}{2} \left[3\cos^2(\beta) - 1 - \eta \sin^2(\beta) \cos(2\gamma) \right] \tag{1.15}$$

where:

$$\sigma_{iso} = \frac{1}{3} \left(\sigma_{xx} + \sigma yy + \sigma_{zz} \right) \tag{1.16}$$

$$\delta = -\omega_0 \left(\sigma_{zz} - \sigma_{iso} \right) \tag{1.17}$$

$$\eta = \frac{\sigma_{yy} - \sigma_{xx}}{\sigma_{zz} - \sigma_{iso}} \tag{1.18}$$

For the solution NMR, only the first (isotropic) part is important, and anisotropic terms are averaged to zero by molecular Brownian motions.

Cross-relaxation and TROSY

Although the chemical shift anisotropy do not contribute to the observed chemical shift in solution NMR, it is one of the two major relaxation mechanisms in biomolecular NMR, the second being dipole-dipole (DD) interactions. Because both interactions are averaged by the same Brownian motion of the molecule, they are correlated, resulting in the interference effects between them. These interference manifests itself, in the case of scalar-coupled spins, in different linewidths of multiplet components. The theoretical consideration of two unlike $\frac{1}{2}$ spins, using the density-matrix formalism, was published in [6], with wide references to earlier works.

The CSA is field dependent, and the maximum of TROSY effect for an ^{15}N-^{1}H interaction can be observed around 1 GHz. In the traditional

NMR experiments the decoupling procedure effectively averages the doublet components. However, a series of experiments (TROSY) were designed to profit from the line narrowing of one component. The pulse-sequence selects only the narrowest component of each doublet in the ^{15}N and ^{1}H directions. As a result the crosspeaks appear in the position shifted by J_{NH} in both directions relatively to the corresponding HSQC spectrum, the linewidths are narrower, but the sensitivity of the experiment is lower for small proteins.

The cross-relaxation can also be used for magnetization transfer, because of the difference in relaxation rate after a given time period, there will be excess of the TROSY component over anti-TROSY (effectively there was a transfer from I_x to $2I_xS_z$). Several modified magnetization transfer techniques were developed to benefit from these effects [7, 8].

1.1.7 A 1D NMR experiment

A one-dimensional NMR experiment consists of three parts:

1. preparation
2. spin evolution
3. acquisition

The preparation step produces an xy magnetization, for example, by a simple 90° pulse on the magnetized sample, then the magnetization evolves for a defined period of time and finally is acquired. In principle to sample a complete signal, the experiment should be repeated for different evolution times. In fact this is not necessary, since acquisition usually does not disturb the sample magnetization to detectable levels, and after first sampling event, the second, third and more can follow, Fig. 1.2. This does not mean that the instrumentation does not interfere with the sample, radiation dumping being a clear example of such influence, but the sampled value is identical despite

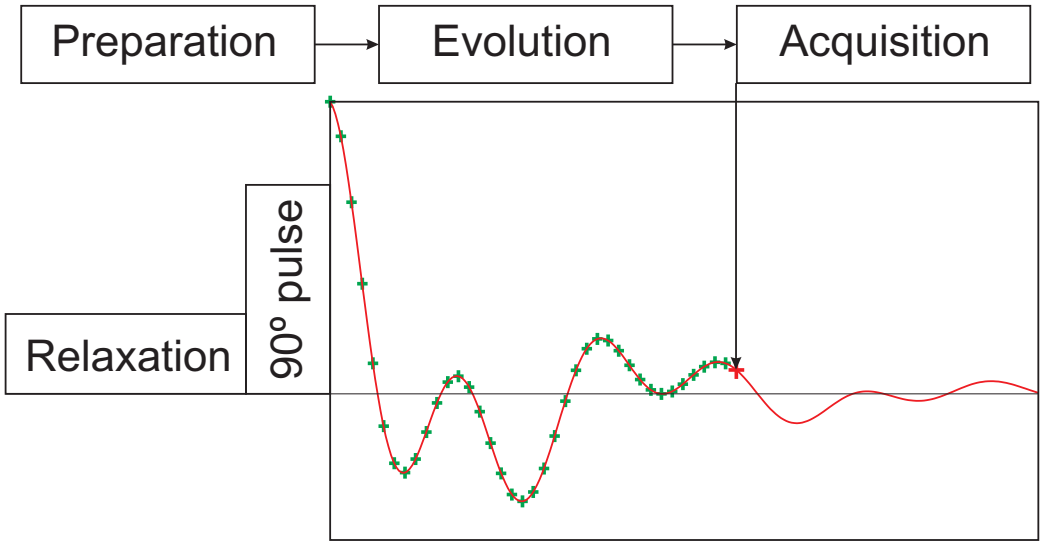

Figure 1.2: Schematics of the 1D NMR experiment. The red curve represents the x-component of the sample magnetization, the green crosses are points sampled before the currently sampled point (marked in red). Note that all points can be sampled during one experiment if the sampling procedure do not affect the magnetization.

of any sampling events taking place before. The resulting set of values forms a digital representation of a free induction decay (FID), Fig. 1.3a, and is subjected to Fourier transform to obtain a spectrum (Fig. 1.3b).

1.1.8 Multidimensional NMR

Spins are not isolated in molecules. Interactions with other spins vary in strength with a spatial distance, dynamics of the molecule, and chemical bond configurations. These interactions cause splitting of resonance lines into multiplets corresponding to different states of the coupled "neighbors".

How is a multidimensional spectrum acquired? From the point of the data acquisition it is nothing more than an ordered series of 1D spectra. To understand the preparation procedure of each FID, one can consider two parallel 1D experiments where a "magic" block, named *mixing*, causes a magnetization transfer from spin 1 to spin 2 (Fig. 1.4). The mixing procedure

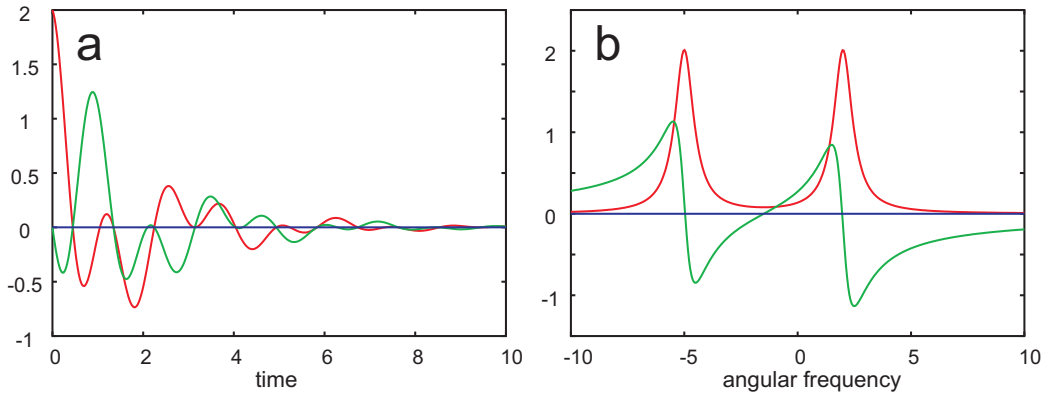

Figure 1.3: The FID (a) and a corresponding spectrum after Fourier transform (b). The x-component (real part) is shown in red and the (imaginary part) y-component in green.

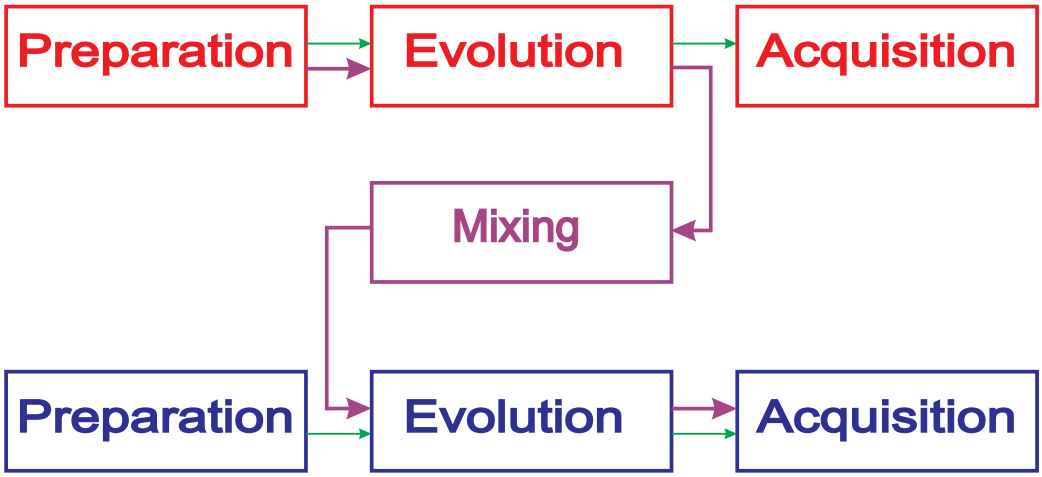

Figure 1.4: A two-dimensional NMR experiment can be build from two combined 1D experiments with the use of a mixing block.

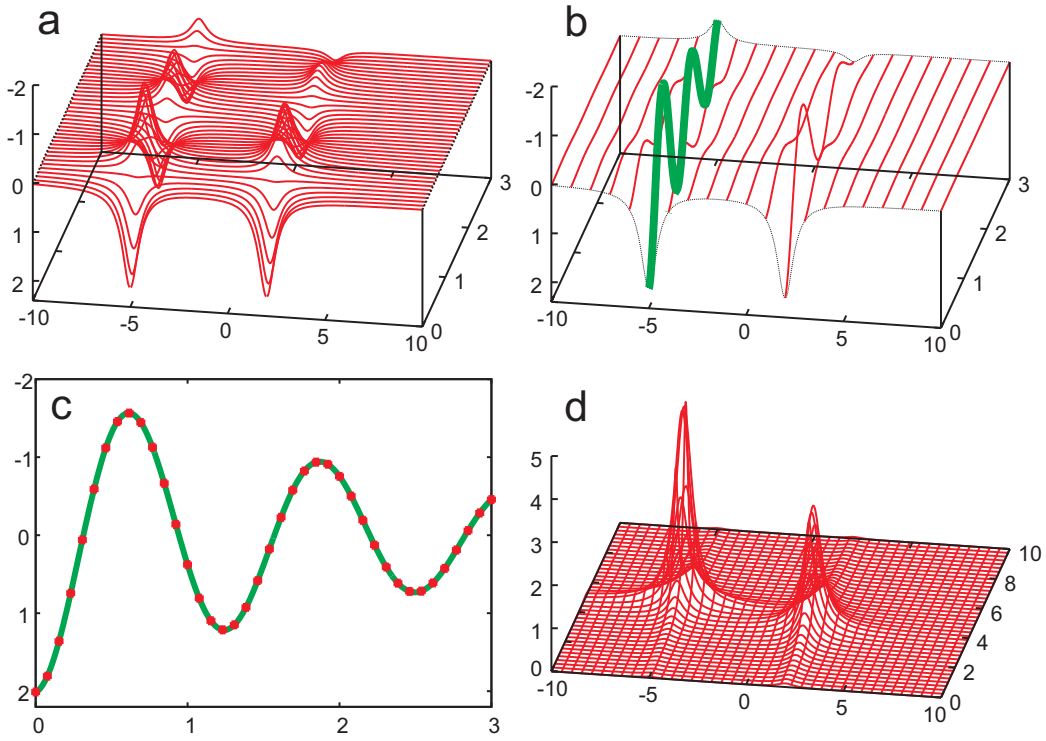

Figure 1.5: From 1D to 2D step by step.

stands for an acquisition event on spin 1 and excitation/preparation on spin 2. The resulting intensity of the signal of spin 2 is proportional to the x or y component of the nucleus 1 magnetization just when it entered the mixing procedure. Each consecutive FID recorded on the second spin is acquired with different evolution time of the first spin, resulting in the modulation of the peak intensity between 1D spectra (Fig. 1.5a). The spectra are aligned row by row. Selection of a column (Fig. 1.5b – green for a particular frequency) forms an FID in an indirect dimension (Fig. 1.5c), analogous to being acquired on the first nucleus, the signals, however, are separated according to the second spin frequency. As the next step the data is subjected to a Fourier transformation column by column, resulting in a 2D spectrum (Fig. 1.5d).

The protocol presented above can be extended to 3D or more dimen-

sions. In practice, however, more transfer steps means lower sensitivity, and each extra dimension requires more experimental time, thus the number of acquired indirect points must be reduced to keep the experiment within a realistic length. The golden rule is to use the lowest possible dimensionality that still allows to resolve signal overlaps. On the other hand, signal overlaps are more likely for larger molecules, with higher relaxation rates, what limits the pulse sequence duration. This explains why while 3D experiments are commonly in use for protein assignments and structure determination, 4D are rare.

1.2 Common problems in NMR of proteins

A major, generic problem arises from the fact that many proteins cannot be concentrated to more than 1 mM, and many proteins precipitate or aggregate even at lower concentrations. In addition higher protein concentrations cause dramatic viscosity change, resulting in extra line broadening. The temperature range for NMR measurements of proteins is limited to liquid state, ca. 0-100 °C, but often the requirement for physiological conditions limits it further to ca. 10-40°C.

1.2.1 Water suppression techniques

For biomolecules, water is the most widely used solvent. The molar concentration of water in the buffer is around[1] 50 M. Not only is the concentration ratio of protein to water lower than 1:50000, but the line-width of a protein is broadened by relaxation. Under this condition NMR signals of interest are near to or below the noise or digital quantization[2].

A similar problem arises when small molecules are present in a buffer, like DMSO, DTT, β-mercaptoethanol. Although they result in signals that are

[1] Usually 5-10% of D_2O is added.

[2] The signal is sampled digitally with a finite resolution, usually not higher than 18 bits.

usually less intense than water, often they are still several orders of magnitude higher than protein signals. Additionally, due to slower transverse relaxation rates such signals can persist until the end of an acquisition period and distinctly obscure the spectrum.

Using perdeuterated solvents

One solution to suppressing the water signal is to use D_2O instead of H_2O. This approach, however, has a severe drawback for protein NMR: intensities of signals from exchangeable protons of the protein are reduced, or completely suppressed. On the other hand, these effects can be used to get specific information about the protein, for example, from the $^1H^N \rightleftharpoons {}^2H^N$ exchange experiment one can determine the resonances of the protein core.

Presaturation

During the relaxation period, prior to each measurement, a weak long selective pulse on the water line (or a sequence of pulses) is applied to the sample. As a result a steady state magnetization of the solvent is reduced heavily. Of course, signals which are near this frequency are also affected, as well. The signals from protons which exchange with solvent or are coupled in another way to an irradiated nuclei are affected. This effect can be significant for nuclei which exchange on the time scale of the duration of the single experiment or faster.

The jump-return pulse sequence

The jump-return pulse sequence is a simple and powerful method when radiation dumping effects can be neglected. The disadvantage is a nonuniform excitation profile, with a relatively wide bandwidth of reduced intensities. Thus peak intensities cannot be directly used for analysis, and incorporation into longer pulse sequences requires caution. This method is suitable for suppression of only one signal frequency.

WATERGATE type pulse sequences

A gradient can be used to spatially change the Larmor frequency. With time the sample magnetization will obtain different phases in different parts and the overall signal will decrease. This effect can, however, be reverted if another gradient of an opposite sign is used or, alternatively, an effective 180° pulse is applied in between two identical gradients. The latter idea is used in the WATERGATE type of water suppression. Several modifications were developed, so one can choose among a relatively broad saturating traditional WATERGATE or W3, W5, or even more complicated sequences, which offer narrower saturation profiles [9].

On the first sight it would be beneficial to always use the W5 or "higher order" modifications, but as the water resonance becomes broader at strong magnetic fields and the use of wider saturating sequences may be beneficial, especially for the experiments that filter ^{1}H-^{15}N correlations, as the signals are far enough from the affected region. My experience is that WATERGATE-5 works efficiently for a well-adjusted proton pulse length and also when the shimming is good. For protein NMR it is sometimes necessary to record a spectrum on a limited sample volume, which may be insufficient for perfect shimming. For proteins, the quality of the spectrum may be good enough even with worse shimming, since the line-widths are large.

Selective water flip-back pulses

Another technique, which is based on selective pulses, is the so-called *water flip-back pulse*. Right after or before a nonselective pulse, a selective pulse is applied at the water resonance to align the water magnetization along the z axis. This idea is, however, implemented with the price of adjusting each selective pulse separately. On the other hand it gives a lot of control of the solvent signal, and selective pulses can be applied in the transfer periods in multidimensional experiments, thus not introducing extra delays. In com-

bination with other techniques, like for example, WATERGATE, selective pulses can improve the experiment sensitivity even when not perfectly adjusted.

An ingenious solution was used in the *fast HSQC* experiment [10]. In the "traditional" HSQC experiment, the WATERGATE sequence at the end of a pulse-sequence led to saturation of the water signal, and consequently to saturation transfer to exchangeable protons, limiting sensitivity for short inter-scan delays. Since the pulse-sequence allows for freedom in the choice of pulse phases, in the fast-HSQC they were chosen in that way that the sum of a total rotation angle for an on-resonance signal (water), is a multiplicity of 360°, serving as an effective water flip-back pulse. The WATERGATE effectively does not affect the longitudinal component of the water signal and only the transverse part is dephased. Alignment of water magnetization along the z-direction speeds up the measurement. Saturation of the solvent is avoided, and the experiment can be repeated without extra delays caused by a long water relaxation, or saturation of amide protons due to exchange with (saturated) water.

1.2.2 Overcoming size limitation

Protein perdeuteration

The simplest method to combat increased relaxation in large proteins is perdeuteration [11, 12]. In this procedure the protein is obtained from an organism grown on a medium containing only or partial deuterium instead of protons. Due to significant isotope effects the organisms usually have to accommodate for new conditions and thus they grow slower. The cost of a protein sample can increase dramatically, particularly when a high perdeuteration level is required.

Partial random perdeuteration, eventually followed by the amide proton exchange with solvent, can be beneficial for ^{1}H-^{15}N correlation experiments.

In partially, randomly perdeuterated proteins ^{1}H-^{1}H NOESY crosspeaks from aliphatic protons are broadened or even split into multiplets, since different isotopomers are present. A partial non-random perdeuteration may be an even more expensive approach. In fully perdeuterated proteins ^{1}H-^{1}H NOESY correlations to aliphatic protons cannot be measured.

Low viscosity solvents

Several methods have been tried to overcome the molecule size limitation in NMR. One of them is to change the tumbling rate of protein molecules. This cannot be easily done by raising the temperature because proteins often would degrade. Also, higher temperatures introduce extra intramolecular motions and reduces the equilibrium magnetization.

A solution would be to encapsulate the hydrated protein in micelles, which will protect the inner environment against the solvent. This process by itself can damage the protein. However, a structure determined in [13] was identical to a "normal solution-state". As a next step micelles are dissolved in low viscosity solvents, like alkanes [14], liquid CO_2 [15].

Encapsulation itself increases the weight/size of the tumbling molecule, thus the viscosity difference must be big enough to compensate for this drawback. The micelle shell can be perdeuterated independently of the protein. Apart from decreased relaxation rate, this techniques opens new possibilities to study protein dynamics and structure.

1.2.3 Overcoming sensitivity limitation

The NMR experiment is usually repeated several times to obtain better signal to noise ratio. The signal intensity is proportional to the number of scans, while the intensity of an uncorrelated noise increases with the square root of this number. A four-time longer experiment results in a $\sqrt{4} = 2$ times better S/N. In principle, this allows to increase sensitivity of the instrument

infinitely, providing that the quantization step is smaller than the noise level. In practice, experiment times longer than 2 weeks are usually not acceptable.

Proteins or nucleic acids, as any biological materials, are subjected to degradation. It is possible to avoid microbial contamination of the sample, but some traces of co-purified proteases or auto-degradation are not rare. Also it is often difficult to avoid denaturation or precipitation of the sample on the longer timescale (months or years). Some proteins denature slowly in contact with the tube glass and many protein samples are not stable on a timescale of a week.

To overcome sensitivity limitation, one can use bigger sample volumes. A bigger volume, however, brings higher requirements for shimming, thus the sensitivity does not increase proportionally to the sample volume. Use of the cryo-probehead is a highly advisable solution. Also the higher magnetic field corresponds to higher sensitivity. On the other hand, in many experiments not the signal-to-noise but the signal-to-artifact ratio is a limiting factor, thus improvements of the NMR experiments can often be crucial for the spectrum quality.

1.3 NMR assignment

The assignment of NMR resonances is a key step in the structure determination protocol of biomolecules. A variety of experiments correlating different nuclei in the protein molecule were developed. In addition several computer programs were developed that assign automatically simpler biomolecules. In the case of larger proteins or limited number of spectra/samples the assignment must be carried out manually.

Fig. 1.6 exemplifies the kind of information extracted from spectra used for assignment purposes. In the NOESY-HSQC of an α-helical protein (Fig. 1.6a), two $^{1}H^{N}$ connectivities between adjacent residues in the amino acid sequence can be found. In the HNCA spectrum (Fig. 1.6b), the $^{1}H^{N}$-^{15}N

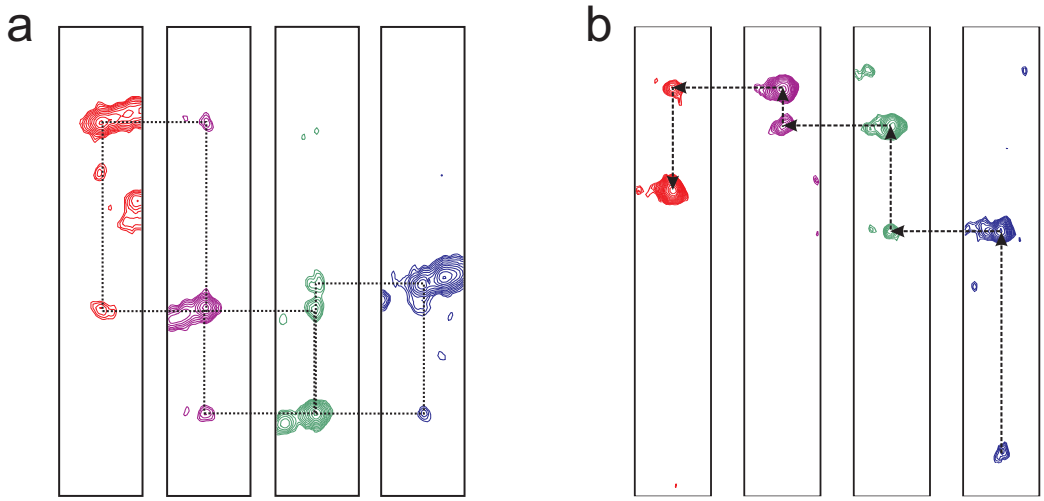

Figure 1.6: Illustration of the assignment procedure (a) using the H^N-H^N NOESY-HSQC correlations, (b) using an HNCA spectrum, correlating sequential HN-HSQC crosspeak through C^α. The stripes present ^{15}N planes. The horizontal axes are in the ^1H direction and the vertical are for ^{15}N and ^{13}C.

crosspeak is correlated with the preceding residue in the amino acid sequence through a $^{13}C^\alpha$ frequency. In both cases several sequential residues are found, and when the aminoacid type is known, often an alignment with a protein sequence is unique. Assignment of resonances to an aminoacid type and information on residues which could not be identified from such procedures must be obtained from different spectra, for example, TOCSY-HSQC, COSY.

1.4 NMR and chemical equilibrium

1.4.1 Exchange processes in NMR

McConnell equations

The "classical" evolution of a net magnetization of a sample is given by Bloch equations. Here we will consider their extensions by including an influence of chemical exchange. These are the so-called McConnell equations [16].

Let the NMR sample contain N states of an observed nucleus, the first order reaction kinetics for concentration vector $\vec{P}$ of each state satisfies:

$$\frac{d\vec{P}(t)}{dt} = V\vec{P}(t) \tag{1.19}$$

where the rate matrix elements V_{ij} describe the rate in which the i'th state changes to j'th:

$$P_i \overset{V_{ij}}{\rightarrow} P_j \tag{1.20}$$

Conservation of mass requires:

$$V_{ii} = -\sum_{i \neq j} V_{ij} \tag{1.21}$$

Evolution of the magnetization can be described by McConnell equations [16]:

$$\frac{dM_z^j(t)}{dt} = \gamma(1-\sigma_j)\{\vec{M}_j(t) \times \vec{B}(t)\}_z - R_1^j\{M_z^j(t) - M_0^j(t)\} + \sum_k V_{jk}M_z^k(t) \tag{1.22}$$

$$\frac{dM_x^j(t)}{dt} = \gamma(1-\sigma_j)\{\vec{M}_j(t) \times \vec{B}(t)\}_x - R_2^j M_x^j(t) + \sum_k V_{jk}M_x^k(t) \tag{1.23}$$

where

$$M_0^j(t) = \frac{M_0[P_j](t)}{\sum_j [P_j](t)} \tag{1.24}$$

and the equation for $M_y^j(t)$ is identical to eq. 1.23 but the subscript x is changed to y.

The solution is straightforward when Fourier transformation is used. Since the transverse magnetization $M_{x,y}$ is a product of an oscillatory, exponentially dumped function and the $\Theta(0)$ function[3], the following property can be

[3] $\Theta(x) = \{0 for x < 0; 1 for x \geq 0\}$. Here we assume a simple "pulse-acquire" experiment. Differentiation of the $\Theta(0)$ function leads to Dirac's $\delta(0)$, resulting finally in the $M(0)$ component. In the case of a more complicated pulse-sequence more components of this type may be necessary in analysis.

easily obtained:

$$\mathcal{F}\left(\frac{d}{dt}\vec{M}(t)\right) = \vec{M}(0) + i\omega\vec{M}(\omega) \tag{1.25}$$

Assuming $\vec{B} = \{0, 0, B_z\}$ one can find it convenient to use $M = M_x + iM_y$ to describe the transverse magnetization. Then:

$$M^j(0) = i(\omega_j - \omega)M^j(\omega) - R_2^j M^j(\omega) + \sum_k K_{jk}M^k(\omega) \tag{1.26}$$

where $\omega_j = -\gamma(1 - \sigma_j)B_z$ is the Larmor frequency of the observed nucleus in the j'th state.

This is a general solution. We will now concentrate on the three or two site exchange, $N = 2, 3$.

Two cases are of interest in protein NMR. The first one is of the conformational exchange. The second case is the chemical exchange. In this second case the population of a sub-state P_i of P can be influenced by controlling the concentrations of interacting compounds. Let us consider a model of the following coupled reactions:

$$X + P \overset{k_{on}^X}{\rightleftharpoons} XP, \quad Y + P \overset{k_{on}^Y}{\rightleftharpoons} YP, \tag{1.27}$$

which implements competitive binding.

We would like to connect the rate constant matrix, V, with the chemical equilibrium. First, it should be remembered that on average the same number of molecules associate and dissociate in equilibrium. Thus this is a stronger requirement than to satisfy the steady state condition:

$$\frac{d\,[A_i]}{dt} = 0 \quad i = 1, \cdots, N \tag{1.28}$$

$[A_i]$ represents the concentration of the i'th sample ingredient. It is not unusual to have a steady state instead of an equilibrium, or not even reaching the steady state. The trivial cases are:

- precipitation

- aggregation
- auto-degradation

A non-trivial case would be a cycle of reactions, powered for example by light, buffer ingredient (like ATP or reducing agents) etc.

The number of associating molecules is proportional to $[X][P]k_{on}^X$, while dissociating by $[XP]k_{off}^X$. k_{off}^X is inversely proportional to the lifetime of the PX complex. The meaning of k_{on}^X is slightly more complicated. It reflects the conditional probability of forming the complex in the case of substrate collision. For further discussion of correlation between K_D and exchange rates it will be useful to note that an upper limit for k_{on}^X exists[4]. In the equilibrium state:

$$K_D = \frac{[X][P]}{[XP]} = \frac{k_{off}^X}{k_{on}^X} \tag{1.29}$$

This equation, together with the upper limit on k_{on}, explains why the tight binding usually corresponds to slow exchange rate. Weakly binding compounds can in principle be both in slow and fast exchange regimes while the tight binders usually realize limit on k_{on}. It is often assumed that the association rate is diffusion controlled [17]. Then the only way to reach low K_D is by increasing stability of the complex, and this implicates a slow exchange rate. It seems to be more intuitive to use k_{off} and K_D for further considerations – k_{on} being replaced by a function of these two parameters.

The "natural" time unit

Let us first consider the case of no exchange. Using Rayleigh's criterion as a determinant for two spectral lines to be separated, one can immediately see that the reciprocal of the linewidth is a natural unit of time, since it is a determinant of the maximum spectral resolution. The spectrum can be scaled in

[4]Theoretically, the limit is reached when the complex is always formed upon substrate collision. Practically, a factor, like activation energy or the necessary orientation, may exist.

units of a linewidth instead of ppm's. For simplicity I ignore the complication when different resonances have different relaxation rates. This difference can be dramatic for biomolecules and the calculation should then include these factors. For simulations of line-shape in cases of different exchange rates, or different chemical shift differences, a reciprocal of the linewidth will be used as a time unit.

Exchange rates and line-shape

There are two extreme cases when the spectrum can be easily predicted. These are: infinitely fast exchange and extremely slow exchange. In the first case, the resonance position will be a weighted average of the "free" and "bound" states. To prove this, one needs to consider the angle of precession in a short period of time. The assumption behind this is an instantaneous jump, without any intermediate steps. Upon addition of a ligand, the resonance will shift towards a "fully bound" position, and the frequency will reflect the free/bound ratio (see also 1.4.2). On the opposite side of exchange rate, when the exchange is very slow, NMR peak positions will be unchanged; the population of bound and free states will be manifested in respective peak intensities. These effects were shown experimentally in the temperature dependent NMR spectra of cyclohexane [18].

Particularly interesting are the cases of intermediate exchanges and the coalescence point where the fast exchange changes to slow exchange regime.

From the experimental data the exchange rate is said to be fast if a peak shifts smoothly upon addition of a ligand (Fig. 1.7a), slow if a new peak grows and the original one disappears slowly (Fig. 1.7b), and moderate exchange if it cannot be approximated by these two cases.

The "line-shape" criterion. The exchange is assumed to be fast when the line-shape of the NMR signal has only one maximum (Fig. 1.8a) for

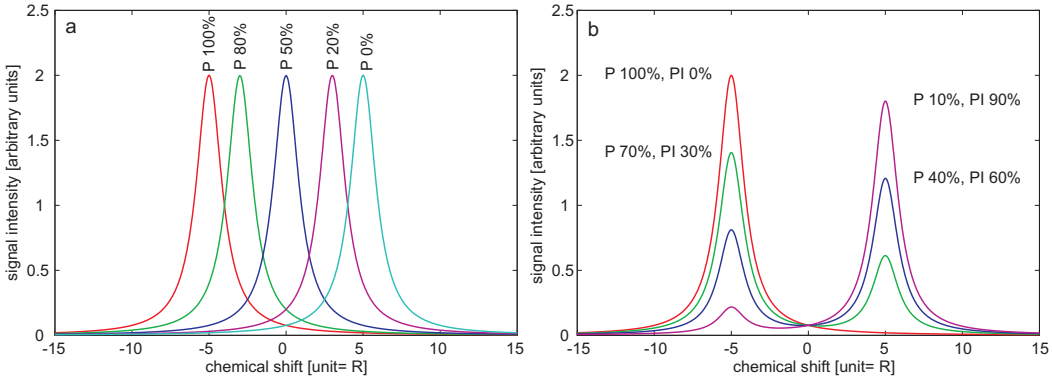

Figure 1.7: Simulation of line-shapes for extremely fast (a) and extremely slow (b) exchange rates for different abundance of free protein "P" and protein-inhibitor complex "PI".

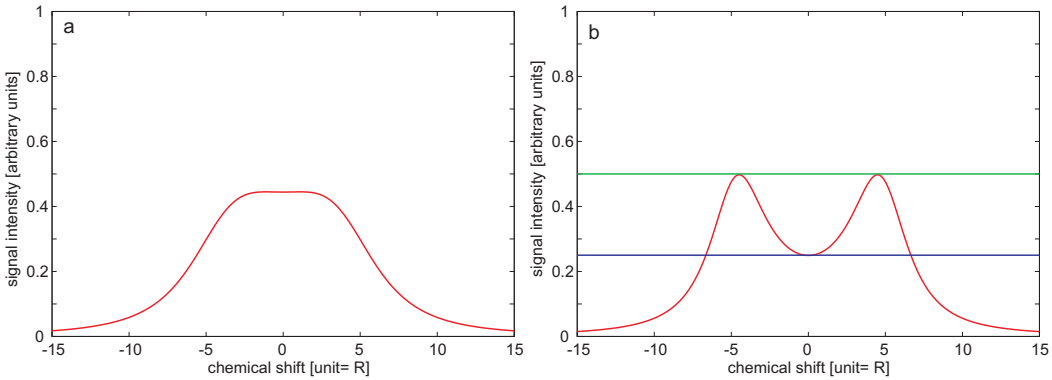

Figure 1.8: Illustration of the "line-shape" criterion between fast/moderate (a) and moderate/slow (b) exchange rates. Populations of the states are 50%.

any ratio of bound to free populations, and slow exchange in the case when the minimum between the two maxima is lower than a half of the maximum value at equally populated speces (Fig. 1.8b). In this way the region of moderate exchange is underestimated, but the border between different exchange regimes is well defined.

Finally, it will be interesting to see how the rate of exchange will be correlated with the chemical shift difference of bound and free states (Fig. 1.9). The conclusion is not surprising, but not trivial. The timescales of both processes have been scaled to the relaxation time units. Linear dependence

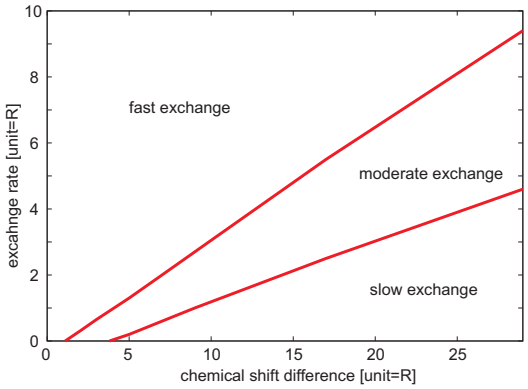

Figure 1.9: Exchange rate diagram calculated for the "line-shape" criterion.

of the limits for exchange regimes on chemical shift difference allows one to think of exchange in terms of $\Delta\omega$. The effect of linewidth being important only if the changes in the spectrum are small. Two conclusions can be drawn:

1. When the change in the peak position is similar to the linewidth, which is not rare for biomolecules, the estimation of exchange rate must be done with caution.

2. It is possible to observe in one experiment different exchange regimes for different resonances affected by binding to the same molecule, even if the ligand molecule is rigid and binds to only one binding site.

The "intensity" criterion. For practical reasons not only line-shapes are important but signal intensities have to be taken into account also. In the Fig. 1.10, when the two states are equally occupyed, the signal intensity dropps below 25% of the initial peak intensity. A plot of the minimum intensity of the NMR signal during a titration experiment versus $\Delta\omega$, at exchange rates on the border of fast to moderate, (Fig. 1.11), shows a dramatic drop for a bigger chemical shift difference.

This suggests to use a different criterion: Let us consider the minimum intensity of a signal during the experiment. Let us arbitrary require the

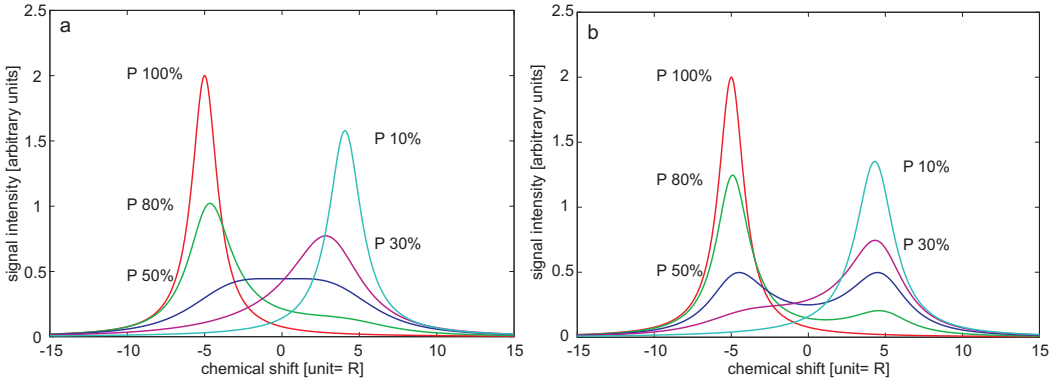

Figure 1.10: Simulation of line-shapes with different populations at the border of fast/moderate (a) and moderate/slow (b) exchange regimes.

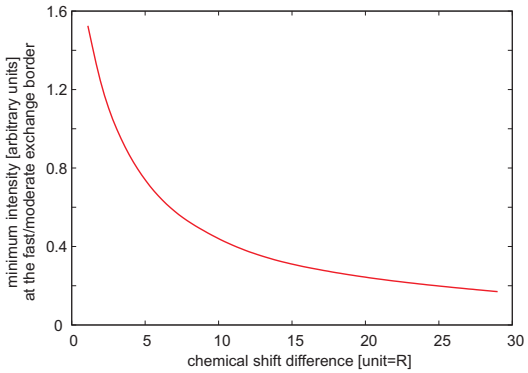

Figure 1.11: Calculated minimum intensity of a signal during titration experiments

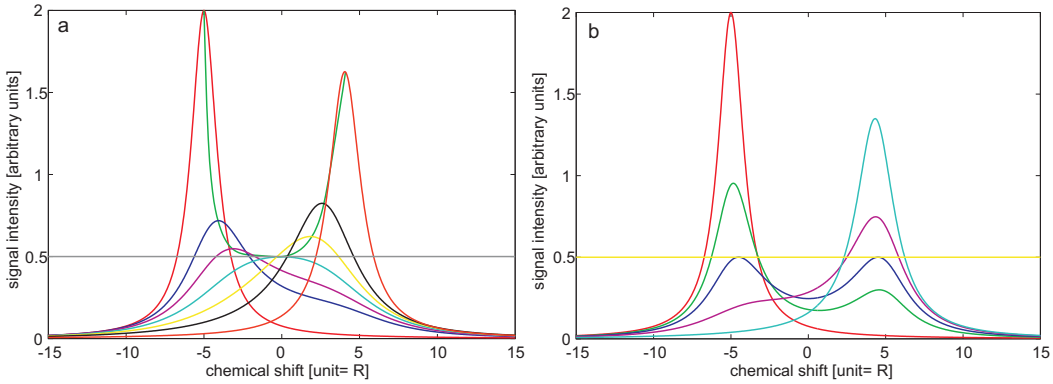

Figure 1.12: The "Intensity" criterion for exchange rate classification

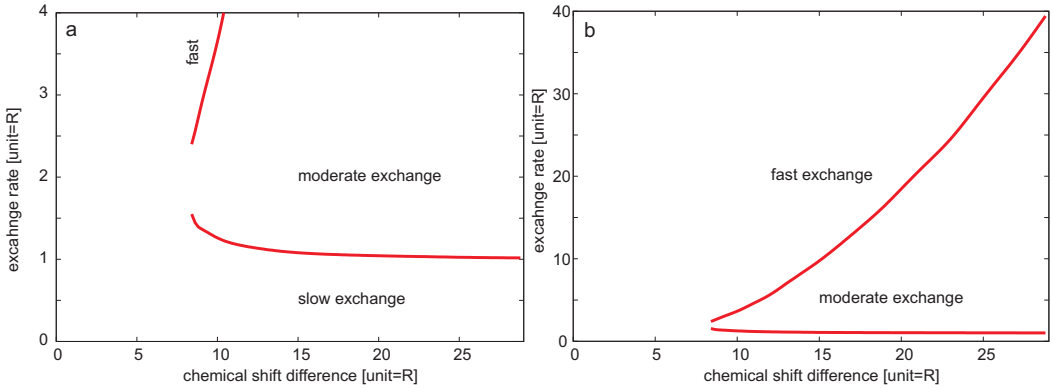

Figure 1.13: Exchange rate diagram calculated for the "Intensity criterion".

signal intensity to stay above 25% of the initial intensity. When $\Delta\omega$ is small, it will be fulfilled for any exchange rate, for bigger differences, however, two limiting cases can be found. They are shown in the Fig. 1.12.

The diagram of different exchange rates obtained with the "intensity" criterion shows no longer linear behavior with frequency difference. There is a striking increase in the exchange rate required for the fast exchange regime upon increase of the chemical shift difference. The slow-to-middle exchange rate asymptotically reaches 1 in the relaxation rate units.

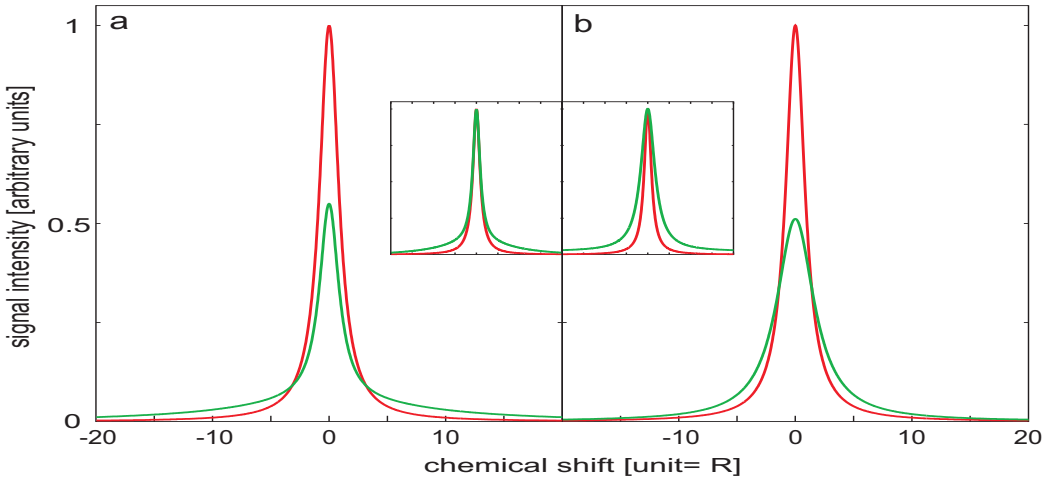

Figure 1.14: Tight binding to a large MW target (a), or exchange broadening (b). The red curve presents the line-shapes of free protein (relaxation rate is 1.0) and the green curve in the case of 50% free and 50% bound. The chemical shifts of free and bound protein are 0.0 (a) the exchange rate is 0.0 and the relaxation rate of a bound state 10.0. (b) the exchange rate is 10.0 and the relaxation rate of a bound state 3.0. The inserts show the same result but scaled to the intensity of the free protein peak (red).

Large molecular weight of the complex or exchange broadening?

An initially "good" spectrum of the smaller protein disappears upon addition of the binding partner in our method [19]. This can happen for two reasons. Either the complex has a significantly larger molecular weight as compared to that of a smaller component and the relaxation in the complex suppress the signal, or the complex is in a moderate exchange regime, where the lines are broad due to the exchange. Classification can be established from additional NMR experiments, like perdeuteration, change of temperature, or by other methods, for example ITC. But it can also be found by analyzing the line-shape at 50% bound states. In Fig. 1.14 two line-shapes were calculated. One for no exchange, but a fast relaxation rate Fig. 1.14a, and Fig. 1.14b, when the relaxation rate was 3.3 time smaller than in Fig. 1.14a, but also the exchange process was going on. Frequencies of free and bound forms

were at the same position. The inserts show the same picture scaled to the intensity of the original resonance. The relaxation rate and exchange parameters were chosen to obtain the same intensity of the peak at 50% bound state. The difference is intuitively clear. In the case of a large molecular weight, the line is a superposition of c.a. 50% of the initial intensity of the free protein plus a broad (thus much less intense) line from the complex. The line-shape is not Lorentzian. The base of the signal can be broader, but the scaled peak should be of the same linewidth as the original in higher parts, whereas the exchange-affected peak is broader, since the linewidth is an average width of free and bound states. If free and bound states have different Larmor frequencies also a shift, or anisotropic broadening, is expected for the exchange-broaden peak.

1.4.2 K_D determination by NMR

In principle an exact measurement of populations of species (e.g. bound and free) as a function of ligand concentration can yield K_D. In practice, the accuracy of the intensity measurements is too low to exactly determine K_D in the case of slow exchange. An upper limit and/or stoichiometry can, however, be estimated. The possibility to measure K_D in the fast exchange limit stems from the fact that the concentrations of species are measured not directly from the intensities, but indirectly by the frequency shift, which is a weighted average of their frequencies.

Basics

Let us consider a first order reaction:

$$[P] + [B] \rightleftharpoons [PB] \tag{1.30}$$

Only Larmor frequency of P compound will be of interest here, assuming signals from B being unobservable or easy to filter out. At the beginning of

the experiment, the V_0 volume at the C_p concentration of P was present, and the observed frequency was ω_P. During the experiment a controlled amount V_b of the solution of compound B at concentration C_b was added resulting in a set of pairs: $\{\omega(V_b),V_b\}$.

The assumptions are:

- first order kinetics
- no solvent/concentration influence on ω_P
- fast exchange between bound and free forms of P
- no other interaction of P or B, no degradation of them etc.

The first assumption can be verified in most cases at the end of the experiment, considering parameters like molar ratio, line-width (against possible oligomerization) etc.

The second usually holds, unless extremely high concentrations of P are used. Note that the protein can influence pH in the sample. In such a case the pH of the sample can change at the time of ligand addition, even if it is dissolved in "the same" buffer. This is because the ligand buffer lacks the protein, the protein buffer has no or less of the ligand.

The third assumption is verified during the experiment. If the frequency of interest shifts continuously the condition is fulfilled, otherwise two frequencies are observed with intensities varying with V_b in the case of slow exchange, or extremely broadened peaks indicate moderate exchange rates.

Constraints

A definition of K_D is:

$$K_D = \frac{[P]\,[B]}{[PB]} \tag{1.31}$$

Conservation of mass (conservation of number of particles) for both compounds:

$$([B] + [PB])(V_0 + V_b) = C_b V_b \tag{1.32}$$

$$([P] + [PB])(V_0 + V_b) = C_p V_0 \tag{1.33}$$

In the case of fast exchange, the observed frequency is a weighted average of frequencies in bound ω_{PB} and unbound ω_P states. Thus in this situation it can be written as a function of V_b.

$$\omega(V_b) = \frac{\omega_P [P] + \omega_{PB} [PB]}{[P] + [PB]} \tag{1.34}$$

where the dependence of $[P]$ and $[PB]$ on V_b was not written explicitly for clarity. To derive this equation one has to consider the total rotation angle of the magnetization for a short period of time. The partial times of evolution in states P and PB are directly proportional to their concentrations.

There are two unknown parameters in equations 1.31-1.34: K_D and ω_{PB}. The first is an object of interest, the latter can be determined experimentally only when high concentration of B can be achieved, thus at the final step mostly PB will be present: $\frac{[P]}{[PB]} \ll 1$. This is satisfied when:

$$\frac{K_D (V_0 + V_b)}{C_b V_b - C_p V_0} \ll 1 \quad and \quad C_b V_b > C_p V_0 \tag{1.35}$$

In practice, it is not always possible to achieve this condition with a small compound, and almost never with protein–protein interactions. The quality of such approximation is directly proportional to $\frac{[P]}{[PB]}$ ratio and to $\Delta\omega$. Neglecting $[P]$ in the denominator of eq. 1.34:

$$\omega(V_b) \approx \frac{[P]}{[PB]} \omega_P + \omega_{PB} \tag{1.36}$$

Solution to the set of equations 1.31-1.34

First let us eliminate $[B]$ from eq. 1.31 using eq. 1.32.

$$K_D = \frac{[P]}{[PB]} \cdot \frac{C_b V_b}{V_0 + V_b} - [P] \tag{1.37}$$

Then using eq. 1.33, $[PB]$ can be eliminated from eq. 1.34:

$$\omega(V_b) = \omega_{PB} + \frac{V_0 + V_b}{C_p V_0} (\omega_P - \omega_{PB}) [P] \tag{1.38}$$

and eq. 1.37:

$$K_D = \frac{[P] C_b V_b}{C_p V_0 - (V_0 + V_b) [P]} - [P] \tag{1.39}$$

The last is a quadratic equation in $[P]$:

$$(V_0 + V_b) [P]^2 + (C_b V_b - C_p V_0 + K_D V_0 + K_D V_b) [P] - K_D C_p V_0 = 0 \tag{1.40}$$

A determinant Δ is given by:

$$\Delta = (C_b V_b - C_p V_0 + K_D V_0 + K_D V_b)^2 + 4 (V_0 + V_b) K_D C_p V_0 \tag{1.41}$$

and a solution for $[P]$ can be obtained as:

$$[P] = \frac{\sqrt{\Delta} + (C_b V_b + C_p V_0 - K_D V_0 - K_D V_b)}{2 (V_0 + V_b)} \tag{1.42}$$

where a positive solution for $[P]$ was chosen.

Finally, equations 1.38, 1.41 and 1.42 define the function $\omega(V_b)$, where C_0, V_0, C_b, V_b, ω_P are *a-priori* known parameters, and ω_{PB}, K_D are fitting parameters.

Approximate methods for K_D determination

Although computers allow for an easy function fitting and such procedures are the best [17], it is sometimes useful to quickly estimate the K_D range. This can be done by determining the amount of the compound necessary to cause a shift to the middle between ω_P and ω_{PB}. Than $[P] = [PB]$ and:

$$K_D = [B] = \frac{C_b V_b - \frac{1}{2} C_p V_0}{V_0 + V_b} \tag{1.43}$$

This equation also allows to quickly estimate the error caused by concentration uncertainties. It can be quite significant especially for low K_D.

Problems

One interesting possibility is the observation of different K_D for different residues. Such effects can be expected when the protein binding sites or binding partners are flexible, particularly for longer peptides binding to protein. Small chemical compounds can bind in more than one binding site on a protein, particularly when they bind unspecifically.

1.4.3 A binary mixture of a protein and a ligand

At first, since concentrations of the sample and the binding partner/inhibitor can vary, it is convenient to express K_D in the units of the final sample concentration $C_0 = \frac{C_p V_0}{V_p + V_b}$ for binary or $C_0 = \frac{C_p V_0}{V_p + V_b + V_a}$ for ternary mixtures. This operation yields more general results. It also allows to extrapolate the results obtained on NMR samples to those at physiological concentrations.

The plot of the percentage of the complex formed, $[PB] / ([P] + [PB]) \cdot 100\%$, at a certain molar ratio of ligand to protein, $([P] + [PB]) / ([B] + [PB])$, for several K_D values, is shown in Fig. 1.15.

Fig. 1.16 shows how much of P is left free at different ratios of K_D to sample concentration for ligand to protein molar ratio, of 1:1, 1:1.25 and 1:1.5.

1.4.4 The binary mixture with limited solubility of ligand

An assumption was made that the ligand solubility is irrelevant for binding effects. In fact in the drug discovery process weakly soluble in water substances are frequent.

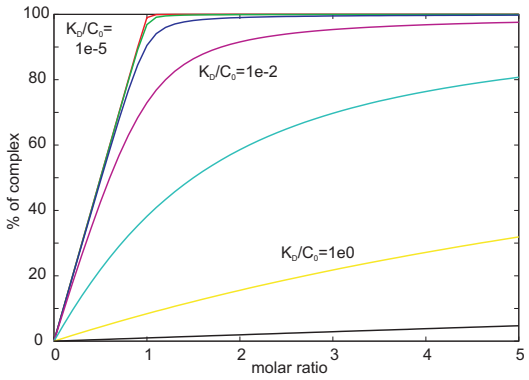

Figure 1.15: A binary mixture. The calculated percentage of the protein-ligand complex versus the ligand-to-protein molar ratio for different dissociation constants, K_D. K_D was normalized to protein concentration C_0.

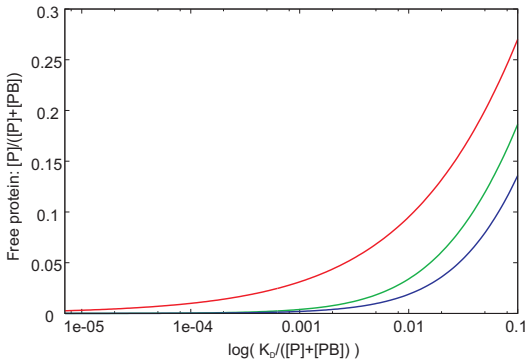

Figure 1.16: The free protein in the sample for 1:1, 1:1.25 and 1:1.5 molar ratio versus K_D.

Let us consider eq. 1.34 where the protein concentration will be replaced using eq. 1.31 yielding:

$$\omega\left(V_b\right) = \frac{\omega_P K_D + \omega_{PB}\left[B\right]}{K_D + \left[B\right]} \tag{1.44}$$

The dependence of $[B]$ on V_b was already solved in 1.4.2 in the case where B was soluble. This result holds as long as $[B] < [B]_{max}$. Once the concentration of B reaches the limit no further shift in the NMR spectrum will be observed. This result is independent of the protein concentration!

1.4.5 A ternary mixture of a protein and two ligands – competitive binding

The situation considered in 1.4.3 is relatively simple to solve and serves well for many situations. Nevertheless sometimes more interactions must be taken into account, especially when an inhibitor, which competes with another ligand for binding to the protein is added. The calculations complicate accordingly, but an analytical solution can be found in this case as well. This problem was solved [20] and extensively analyzed for ITC purposes [21]. In a more complicated case a numerical solution can also be found. The binding of B to P in the presence of a weaker inhibitor is presented in Fig. 1.17a and in the case of an inhibitor which binds to P stronger than B is shown in Fig. 1.17b.

It can also be interesting to asses the amount of PI for a week inhibitor during the titration as shown in Fig. 1.18. Finally, it would be interesting to know what excess of an inhibitor is necessary to replace 30%, 50%, or 70% of P from the PB complex for different inhibitor strengths. This relationship is shown in Fig. 1.19.

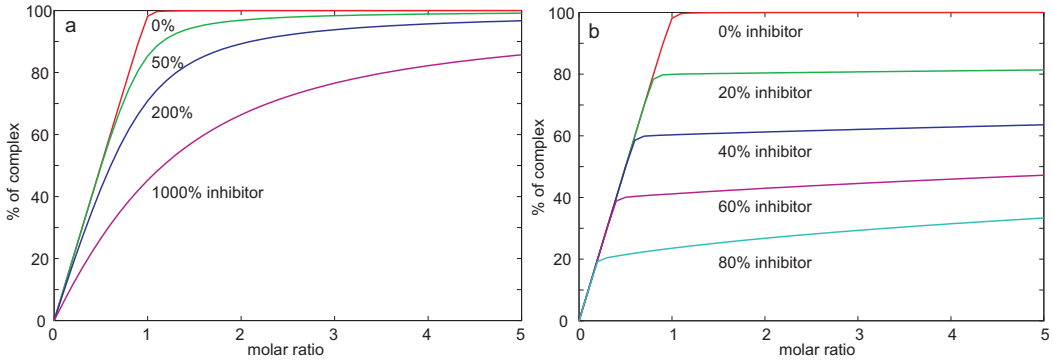

Figure 1.17: The ternary mixture of protein (P), ligand (B), and an inhibitor (I). Formation of the complex PB ($K_D/C_0 = 3.5E-4$) in the presence of different amounts of inhibitor (a) PI complex is weaker than PB ($K_D^{PI}/K_D^{PB} = 14.3$), (b) PI complex is stronger than PB ($K_D^{PI}/K_D^{PB} = 1.4e-2$).

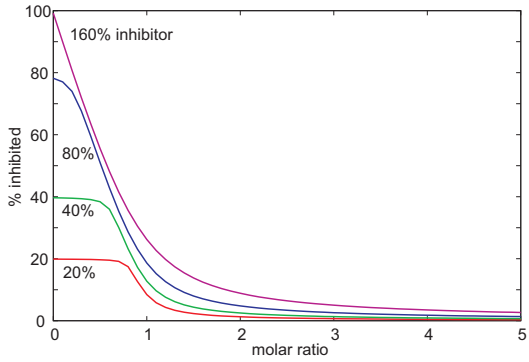

Figure 1.18: Levels of protein–inhibitor complex for conditions as in Fig. 1.17b

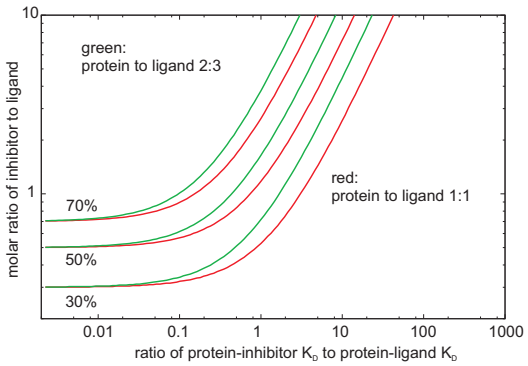

Figure 1.19: Amount of an inhibitor necessary to replace 30%, 50%, or 70% of P from the PB complex for different inhibitor strengths, for $K_D^{PB}/C_0 = 0.1$.

1.5 Applications of NMR to binding studies

NMR-based methods for screening for drug compounds that bind protein targets can be divided into two major classes. In the first class, the NMR signals of low molecular-weight molecules are monitored by utilizing NOE effects or differential relaxation rates of free and bound states [22, 23, 24, 25]. The second approach focuses on the changes in the ligand-induced NMR chemical shifts of the protein [26]. The latter class of methods requires larger amounts of isotopically labeled proteins but can provide more detailed information on the ligand-protein interaction, in particular the method not only allows to determine whether a compound binds, but also where it binds on its target.

1.5.1 Ligand-site methods

Chemical shift or peak intensity: one would observe a simple proton 1D spectrum for changes either in chemical shifts or in peak intensities.

Transverse relaxation rates: can be measured either from the linewidth or with a simple 1D NMR experiment designed to measure T_2 relaxation times [27]. Since most of ligand based methods were optimized for high throughput screening (HTS), they usually assume a fast-exchange regime for binding between ligand and protein, tightly binding compound, a rare case in HTS, being a "false negative" result. A method called "reporter screening" has been developed to overcome this problem [28]. The idea is to perform screening in the presence of a known ligand, weakly binding to the target protein, and monitor the reporter ligand resonances.

Longitudinal relaxation rates: To determine the "active" regions of a given ligand, a 1D NMR experiment termed saturation transfer difference (STD) can be applied [29]. This technique relies on spin diffusion processes.

Transfer NOE effect: a NOESY peak intensity, or sign. This method is based on the the fact that NOEs for bound ligands are typically large and negative, whereas for free ligands they are very small [30].

Diffusion rate: interaction with a larger molecule changes the linear diffusion rate of the compound [27, 31].

1.5.2 Methods that monitor target proteins

For proteins of small size (e. g. less than 20 kDa) the most popular protocol has been based on the use of chemical shift perturbations in 2D ^{1}H-^{15}N HSQC spectra of uniformly ^{15}N-labeled proteins [26, 32, 33, 34] For larger proteins, when ^{15}N labeling is not sufficient to resolve spectral overlap in 2D ^{1}H-^{15}N spectra, selective amino acid labeling (^{15}N or ^{13}C) or protein perdeuteration may be necessary and a TROSY [35, 36] type of experiments are advantageous over "traditional" HSQCs.

In principle the assignment of NMR resonances is not required if the only purpose of the NMR experiment is to detect the binding of ligand to the target protein. To monitor a specific binding site on a protein, active site residues must be identify, for example, by experiments with known ligands [37], or for a more detailed structural interpretation the interface can be determined on the basis of a known 3D structure [38], or by a more laborious but well-established NMR assignment procedure [16].

In the original "SAR by NMR" report [33], only 8 out of 107 resonances (7.5%) showed significant perturbations upon compound binding [33]. The complexity of ^{1}H-^{15}N correlation spectra increases quickly with molecular weight. A remedy would be an ^{15}N selectively labeled protein with one amino acid type, and thus to obtain a less crowded NMR "subspectrum". Reduced spectral complexity can also be obtained by methods like SEA-TROSY [39]. Although ^{13}C isotopic enrichment is typically considered too expensive for

screening purposes, an alternative by using inexpensive $^{13}CH_3I$ was devised for labeling of methyl groups [40].

Binding site mapping

One of the most interesting information that can be extracted from the binding studies is a binding site. The direct method would observe NOESY correlations to the ligand. In the case of chemical shift monitoring, the residues mostly affected upon ligand binding are considered as the binding site. Allosteric effects or distinct conformational change can complicate the analysis.

One of the important steps in spectrum analysis is assignment of the ligand induced chemical shifts to the initial protein resonances. This may be straightforward in the case of fast exchange since the shifts can be mapped quasicontinuously from starting to final positions in a series of experiments. In the case of slow exchange (tight binding), however, the assignment is usually a challenge, and the nearest-new peak is often chosen.

1.5.3 DIMAPPS: DIrect Monitoring of Antagonists on Protein-Protein interactionS by NMR spectroscopy

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 little has been done in order to understand the effect of inhibitors on protein-protein complexes, especially by NMR spectroscopy. The principle purpose of an antagonist compound discovery, however, is determining whether a lead compound inhibits or dissociates protein-protein interactions.

The method we suggest and have successfully applied to several cases is to directly monitor the dissociation of the complex of interest. A spectacular effect of a strong inhibitor allows for an unambiguous interpretation of the data even with low resolution spectra. The essence of the method origi-

nates from the fact that protein complexes larger than ca. 30 kDa show line broadening due to the increase T_2 relaxation rates.

We have illustrated our method with the p53-MDM2 interaction in chapter 2. The binary complex of the N-terminal domain of MDM2 (residues 1-118) and the fragment of p53 (residues 1-312) has a total molecular weight of 45 kDa. Isothermal titration calorimetry (ITC) measurements indicated a K_D of 0.77 μM, which agrees well with that of ca. 0.7 μM reported in the literature [41, 42, 43, 44, 45, 46].

A protocol for DIMAPPS

DIMAPPS requires the protein complex to be formed between proteins with considerably varying molecular weights, one of which should give a good quality S/N NMR spectrum and would serve as a NMR reporter protein. Formation of the complex should result in higher transverse relaxation rates R_2 of the reporter protein, which causes a decrease in its NMR intensities to a point that they may completely disappear. This clear indication of interaction can be easily detected and this part of the experiment must be performed only once, at the time of the system and conditions choice.

The spectrum of the protein-protein complex changes dramatically if an inhibitor that breaks the complex is added. Two outcomes are possible in case of the complex dissociation: either the reporter spectrum is recovered unchanged, when an inhibitor binds to the larger protein, or the recovered spectrum displays modifications caused by the inhibitor binding. It is noteworthy to point out that the dissociation is detected irrespective of the protein it acts on; thus there are two targets checked simultaneously and the selection of the right binding site and inhibition strength is obtained in principle without any prior knowledge of these parameters. Thus our approach targets protein-protein interactions and not a single protein.

Discussion on DIMAPPS

DIMAPPS is neither competitive to the traditional "SAR by NMR" [33, 26, 32] nor it can be considered as its variant. These two methods are based on essentially different principles, thus they nicely supplement each other.

While DIMAPPS directly selects the inhibitor of a complex, the complex inhibition is does not intrinsically result from the "SAR by NMR" approach. A final ligand has to be tested in a separate experiment for its inhibitory potency.

A 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 longer than those used in SAR by NMR (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. Such systems are often perfect for the method we present.

Another weakness of the "SAR by NMR" approach arises when NMR cross peaks of several resonances disappear upon ligand binding. This happens for intermediate exchange rates between the free and ligand-bound forms. These peaks are usually those residues located at the ligand binding interface, however, it is difficult in practice to interpret these data without additional information from the structure, for example. This problem simply does not exist in our method, since we directly monitor the desired effect. 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.

The direct monitoring of dissociation can give negative results in case of weak binders, which still might be a valuable substrate for engineering a strong inhibitor. A weak inhibitor could, however, partially release the labeled protein when added in large excess.

Several variants of our approach are possible. A 1D version of an HSQC

experiment or even a proton 1D spectrum may suffice for monitoring the states of proteins in complexes upon treatment with ligands. The regions of 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. 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 this stage.

Chapter 2

MDM2, p53, GST-p53, p63, and Rita

2.1 Biological context

p53

The human p53 protein is a tumor suppressor transcription factor and loss of p53 function through mutation is involved in 50% of human cancers. The remainder retains wild-type p53 but the p53 pathway is inactivated through, for example, interaction with the overexpressed MDM2 protein. MDM2 is a primary negative regulator of p53 activity [47]. The major role of p53 is to induce a G1/S and G2/M checkpoints, and cause the cell cycle arrest or apoptosis [48, 49].

The structure of p53 protein can be divided into five domains [50]:

residues 1-70: the transactivation domain (TAD), responsible for binding to MDM2; it is natively unfolded [44]. The primary binding site of MDM2 on the p53 protein has been mapped to residues 18-26 [41, 51, 52, 53, 54, 49]. The ^{1}H-^{15}N resonance assignment of TAD was published and the binding site to MDM2 was mapped [50].

residues 60-97: a flexible proline-rich stretch containing five copies of PXXP motif.

residues 100-300: a well-structured DNA binding domain responsible for binding to target DNA sequences. The DNA binding domain in solution forms a dimer [31].

residues 320-360: the oligomerization domain responsible for forming a tetramer in a wild-type protein; can be mutated to allow dimerization only without considerable modulation in DNA binding activity [50].

residues 360-393: the C-terminal domain was found to play a regulatory function for the protein. Constructs which lack this domain, or if this domain is bound to a monoclonal antibody or post-translationally modified, exhibit enhanced DNA binding. There is an experimental evidence, based on NMR, suggesting that this domain is not folded into a regular structure [50, 44].

p63

Much less is known about p63 [55] in terms of structure-activity relationship, although it often behaves similarly to p53 in (in vitro?) experimental systems. The NMR resonance assignment of the DNA binding domain of p63 was published [56].

MDM2

MDM2 is a principal cellular antagonist of p53 that interacts primarily through its 100 residue amino terminal domain with the TAD of p53. The rescue of the impaired p53 function by disrupting the MDM2-p53 interaction offers new avenues for anticancer therapeutics [50, 57, 41] and several lead compounds have recently been reported to inhibit the p53-MDM2 interaction in assays based on tumor cell cultures or immunoprecipitation techniques. Structure based screens for this interaction utilized an N-terminal domain of MDM2 of ca. 100 amino acids and short peptides of p53 [45, 52, 58, 54, 47].

The ^{1}H-^{15}N NMR resonance assignment of the N-terminal domain of MDM2 and a titration with E17-N29 p53 peptide were published [43].

2.2 Interaction study

In our study we have used a 118 amino acid N-terminal domain of MDM2 and a GST fusion N-terminal 75-residue fragment of p53 which encompasses the TAD, or a 1-312 residue p53 fragment, which contains also a DNA binding domain. In the HSQC spectrum resonances of folded DNA binding domain or GST are broad and of low intensity due to fast relaxation rates, the spectrum shows mostly the flexible TAD residues. The molecular weight of the folded part of p53 is above 30 kDa.

2.2.1 MDM2-p53 complex

The N-terminal domain of MDM2 is small enough (ca. 13 kDa) to provide a good quality HSQC spectrum (Fig. 2.1a). This ^{15}N-MDM2 was titrated against unlabeled GST-p53, and the complex formation was observed by the disappearance or severe reduction in intensity of most of the MDM2 peaks, as seen in Fig. 2.1b. The leftover peaks originate from flexible residues of the complex that are located in the spectrum at the central 8.3 ppm NH amide region, diagnostic for unstructured residues, plus side chains at 7 and 7.5 ppm (for details how to use NMR to assess the extent of folding in proteins see [53]).

We also performed the experiment with the ^{15}N-labeled p53 and unlabeled MDM2. Complex formation can be monitored from the NMR spectrum as the three primary binding sites of p53 to MDM2 are known Phe19, Trp23 and Leu26 [41, 50]. The cross peaks of these residues begin to disappear on stepwise addition of MDM2, and completely disappeared on the complex formation. Particularly well resolved tryptophan sidechain signals can be observed at ca. 10.10 ppm. The complex formation can be monitored from

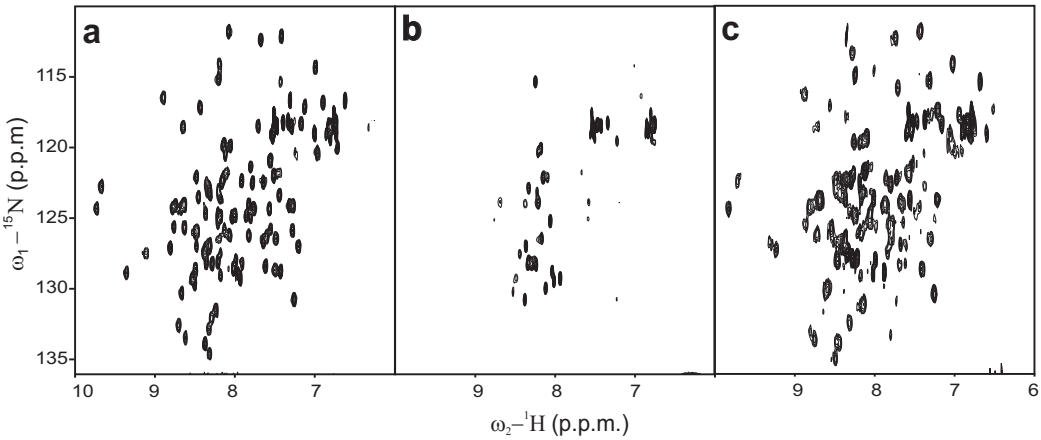

Figure 2.1: DIMAPPS: a spectrum of MDM2 (a), the spectrum after addition of GST-p53, most of the MDM2 peaks disappeared (b), after addition of nutlin-3 the MDM2 spectrum is recovered (c).

a proton 1D spectrum since the signal from Trp23 disappears on complex formation, as seen in Fig. 2.6. The experiment unequivocally shows that the N-terminal residues are still not structured when bound to the MDM2 domain, with the exception of a 10 residue-binding site (residues 17 to 26).

2.2.2 Nutlin-3: DIMAPPS in practice.

Nutlins are a class of cis-imidazoline compounds that were recently reported to inhibit the p53-MDM2 interaction in nano molar range (nutlin-3 IC_{50} value of 0.09) [52]. Addition of nutlin-3 to the MDM2/p53 complex restores the MDM2 spectrum, as seen in Fig. 2.1c, with the sites involved in binding to nutlin-3 being however shifted. The freed GST-p53 is not precipitated as judged by 1D NMR spectrum. Since we measured a 1D proton spectrum of p53 before the experiment, observation of the tryptophan sidechain at 10.10 ppm allowed us to show, that a "soluble" p53 was recovered by nutlin-3. The experiment also shows that the MDM2/nutlin complex is soluble, and that nutlin-3 did not induce precipitation of MDM2. Our assay thus showed that nutlin-3 releases p53 from the complex by competing with p53

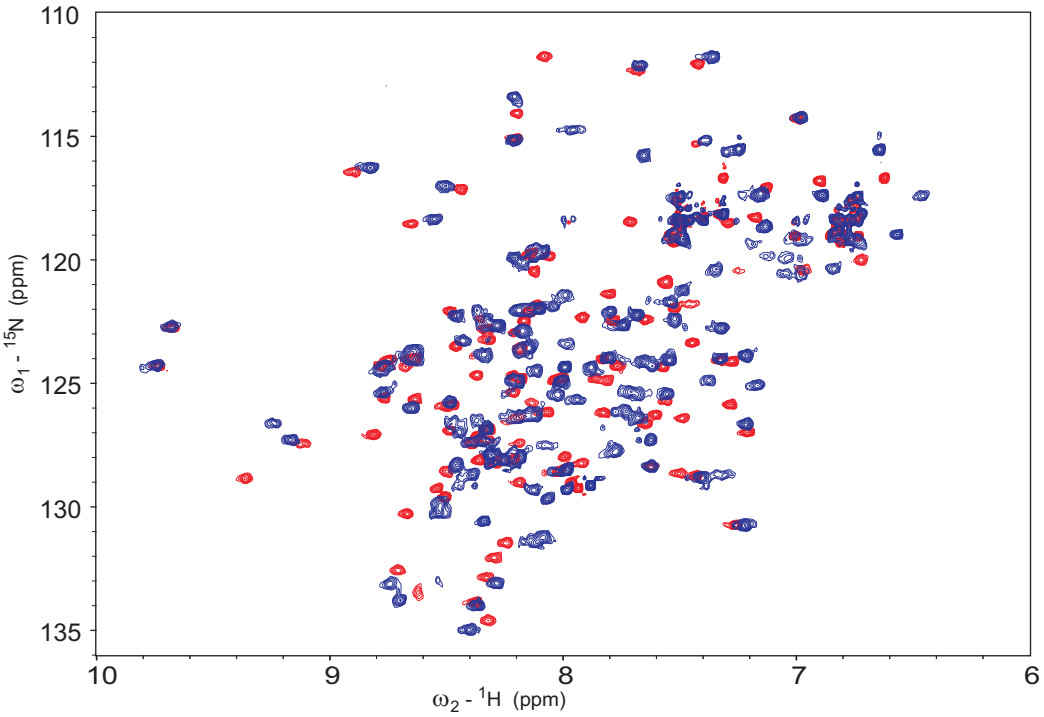

Figure 2.2: The spectrum of MDM2 (red) and MDM2 with nutlin-3 (blue).

for binding to MDM2. The important outcome of our experiment is that it shows that the freed p53 is in a native form.

2.2.3 RITA (NCS652287)

A recently discovered small molecule, RITA (NSC652287), has been reported to reactivate the p53 function in tumors by binding to the N-terminal domain of p53 and blocking the interaction of this domain with MDM2 [59]. Using NMR, we have unequivocally shown that RITA does not in fact block the formation of the complex between p53 and the p53-binding domain of MDM2 in vitro [60]. We tested RITA on MDM2 and the 312-residue domain of p53 in detail; GST-p53 produced analogous results.

^{15}N labeled MDM2 with unlabeled p53. In one variant of the experiment, to check if RITA might bind to MDM2, 3 μl of 60 mM RITA

in DMSO was added to 500 μl of the MDM2 sample at this stage, resulting in a saturated RITA solution and a fine RITA precipitate. No major changes in the MDM2 (Fig. 2.3a,b) spectrum were observed, thus RITA does not bind to MDM2.

A stepwise addition of p53 followed. The disappearance of most of the MDM2 cross peaks in the 2D ^{1}H-^{15}N HSQC spectrum indicated MDM2/p53 complex formation Fig. 2.3c,d.

The DMSO stock solution of RITA was added gradually, up to the concentration of 15% DMSO. After each step an HSQC experiment was recorded. Finally a high excess of RITA (compared to the signals of the proteins) was seen in the proton 1D NMR spectrum, but no free MDM2 released from the complex was detected. Thus RITA, up to at least 5-fold molar excess relative to p53, did not prevent the MDM2/p53 complex formation (Fig. 2.3b).

We then continued the experiments by adding nutlin-3 as a positive control. Addition of nutlin-3 restore the MDM2 spectrum, as seen in Fig. 2.3e,f, with the signals from MDM2 sites involved in binding to nutlin being however shifted. This is a clear indication that whereas nutlin-3 releases p53 from the complex by competing with p53 for binding to MDM2, RITA has no effect on this interaction. Figure 2.3 includes also the control titration of the MDM2/p53 complex with nutlin-3 only.

^{15}N labeled p53 and unlabeled MDM2. In the second protocol RITA was titrated into p53. First DMSO was added to the ^{15}N p53 sample, up to the concentration of 5%/volume, and an HSQC was acquired. Fig. 2.4(red) shows the ^{1}H-^{15}N HSQC spectrum of the uniformly ^{15}N-labeled p53. Only the first 93 N-terminal residues that are flexible and mostly unstructured [57, 50, 61] give rise to strong cross peaks (Fig. 2.4a). We observed the strong signals to check the binding of RITA to the N-terminal part of p53.

Next, 4 μl of 60 mM DMSO solution of RITA was added and an HSQC

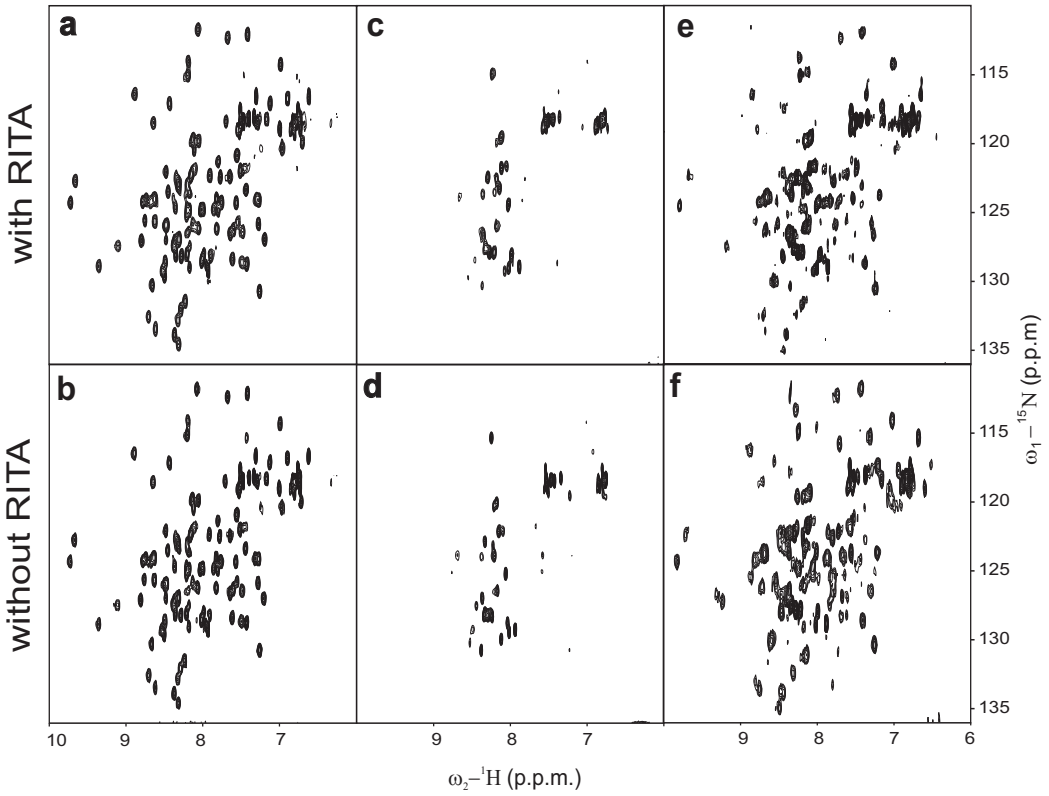

Figure 2.3: HSQC spectra of the ^{15}N labeled MDM2 (residues 1-118). (a,c,e) with RITA, (b,d,f) without. (a,b) HSQC spectra of the ^{15}N labeled MDM2. (c,d) After addition of the unlabeled p53 most of the signals broadened to the noise level. The leftover peaks originate from flexible residues of the complex. (e,f) When nutlin-3 was added, the HSQC spectra of MDM2 recovered as a result of dissociation of the p53-MDM2 complex.

spectrum recorded (Fig. 2.4(blue)). Afterwords DMSO or RITA were added stepwise to a high excess of free RITA (observed in a 1D NMR spectrum). In one variant we also pre-incubated p53 with RITA at 37°C for 80 min, with no effect observed, apart from some p53 precipitate which we assigned to the increased temperature[1]. NMR measurements consisted of monitoring changes in chemical shifts and line widths of the backbone proton and nitrogen resonances of ^{15}N-p53 upon addition of RITA. 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 revealed, however, small changes in the signals originating from the folded core part of p53 (Fig. 2.4b). Considering DMSO already present in the sample of the reference spectrum, the small changes in the folded part (Fig. 2.4b) are unlikely to originate from DMSO. The peaks are too crowded and the shifts are small for an unambiguous interpretation of the data. We cannot, however, exclude binding of RITA to other than the N-terminal residues of p53.

We next studied the effect of RITA on the interaction between ^{15}N-p53 and MDM2 by monitoring NMR signals of the three primary binding sites of p53 to MDM2 (Phe19, Trp23 and Leu26; [41]) upon addition of MDM2. The cross peaks of these residues begin to disappear on stepwise addition of MDM2 to the p53/RITA, and completely disappeared on the complex formation as seen in (fig. 2.5b).

The status of p53 could also be checked by following the proton NMR signals of the NH side chains of tryptophans in 1D proton NMR spectra. Fig. 2.6 shows the 1D spectra of the region where the side chain of Trp23 resonates. The experiments with the ^{15}N labeled p53 confirmed the results of the first series of our measurements on ^{15}N-MDM2.

[1]Precipitation of p53 was also observed for long NMR measurements at 298K.

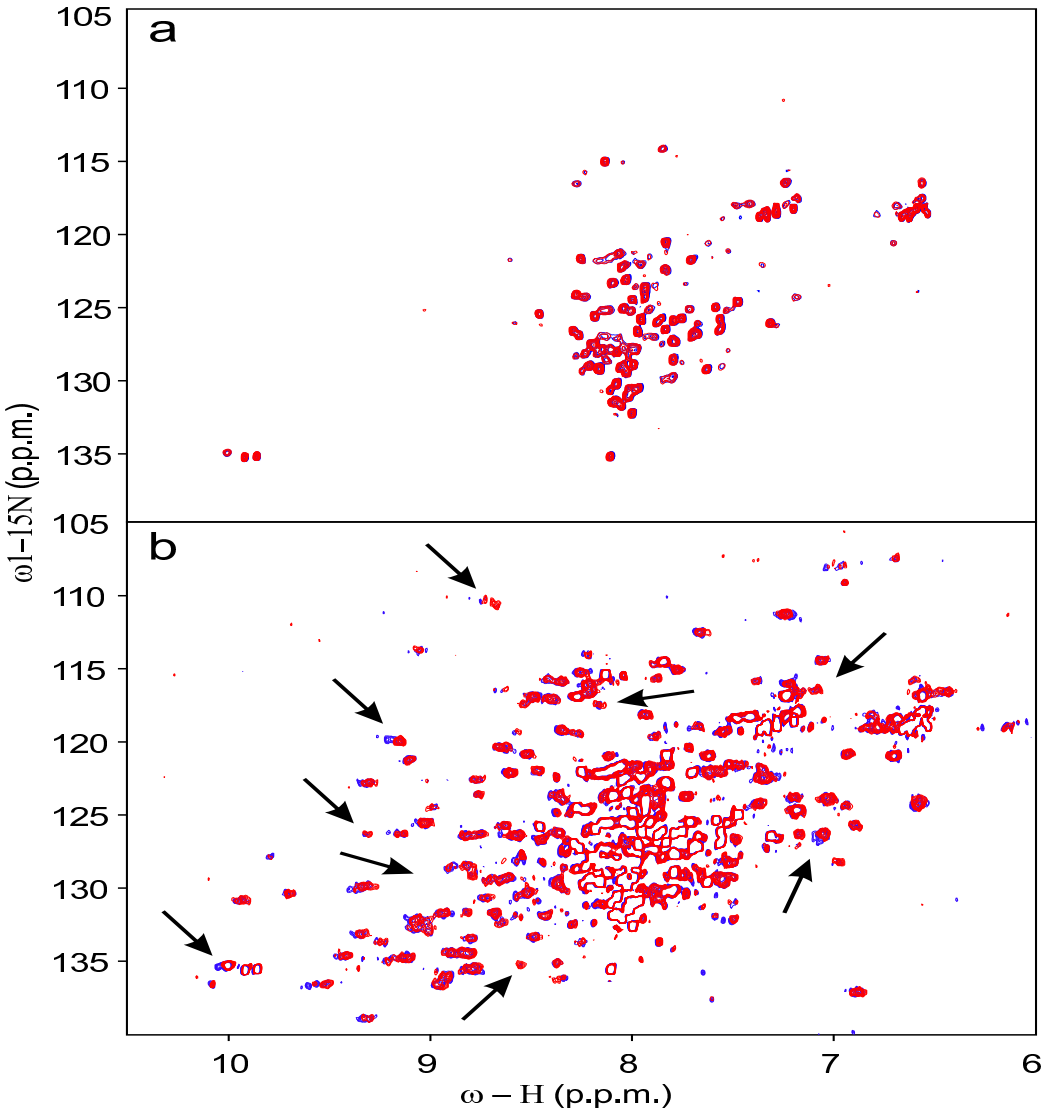

Figure 2.4: Superposition of HSQC spectra of ^{15}N labeled p53 in 5% DMSO (red) and after addition of RITA (blue). (a) The spectrum displays changes at places indicated by arrows in the folded core part of p53 (this domain gives broad and weak but observable signals). (b) No changes were detected in the flexible part of p53, which contains the MDM2 binding site. Cross peaks of the folded part of p53 are not seen because only higher contour levels were plotted.

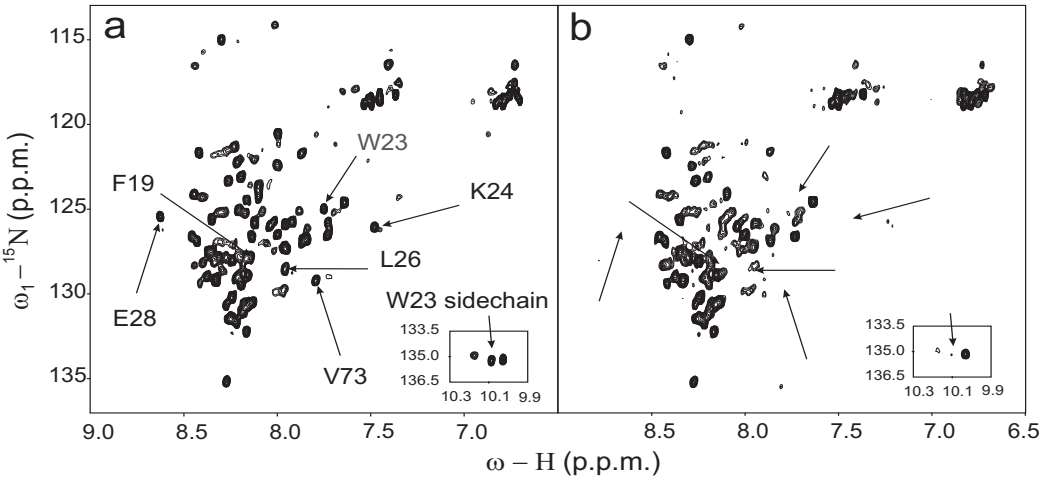

Figure 2.5: HSQC spectra of the flexible N-terminal residues (1-93) of the ^{15}N labeled p53 (the residues of the p53 DNA core domain are mostly not seen because of their broad linewidths). (a) At 5% DMSO and saturated solution of RITA. (b) After addition of MDM2: the signals of the residues involved in the primary MDM2 binding site disappeared or had dramatically reduced intensities (indicated with arrows; the assignment taken from [50, 61]).

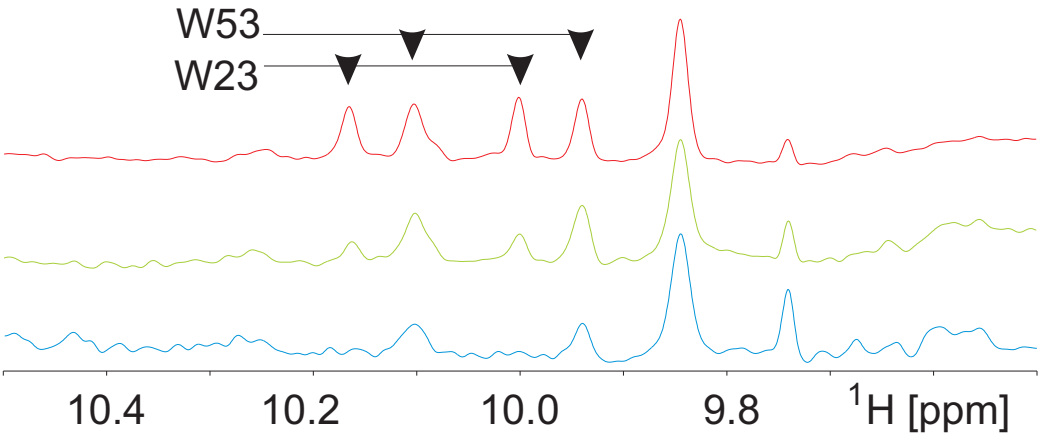

Figure 2.6: A tryptophan region of ^{15}N labeled p53 (a) sample titrated with MDM2 (b,c).

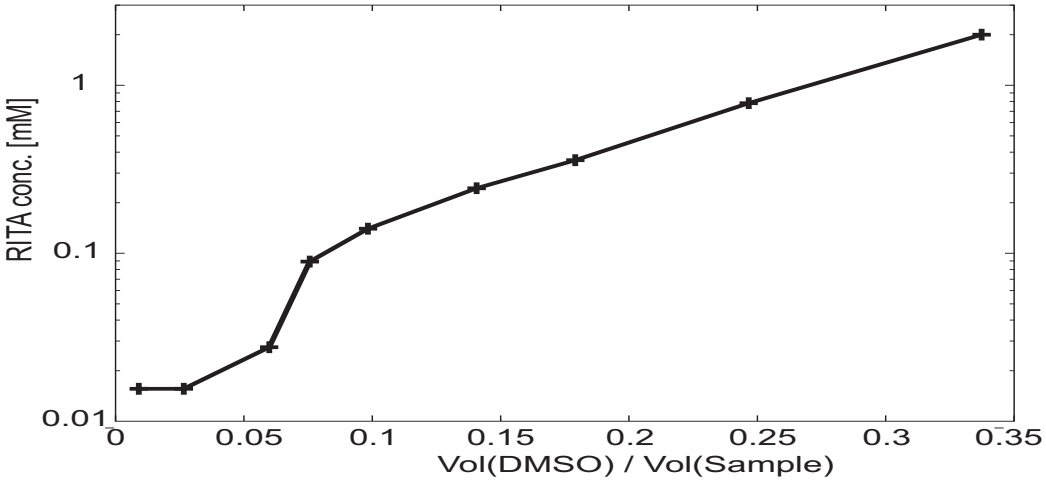

Figure 2.7: Solubility of RITA in the MDM2 buffer as a function of DMSO concentration.

RITA solubility. Since RITA is soluble in the buffer used for NMR samples at around 0.15 mM (as judged from the proton 1D NMR spectrum), the DMSO-d_6 stock solution of RITA was used for NMR measurements. The effect of DMSO on the proteins and the protein complex was checked to be negligible up to 15%/volume of DMSO.

2.2.4 Other compounds from the NCI library

We have also tested other compounds from the NCI library reported to have the p53 dependent effect on cancer cell-lines. They included: NSC146109, NSC154829, NSC188491, NSC348401, NSC676693 [62].

A stock DMSO solution of each compound (ca. 20 mM) was added to the ^{15}N labeled MDM2 complexed with p53, and the complex dissociation was monitored by HSQC and 1D spectra. In the case of NSC154829, NSC188491, NSC348401 the ligand was seen in the proton 1D spectra of the mixture, but no effect on MDM2-p53 complex was observed. The remaining two compounds also did not cause complex dissociation, but their existence in the buffer had to be checked in separate experiments, dissolving each of them in

the buffer without MDM2 or p53. The ^{1}H 1D spectra show that NSC146109 and NSC676693 are hardly soluble in water.

Although we did not test the compounds for their binding to MDM2 or p53 only in a separate experiments, we can conclude that none of them can inhibit MDM2-p53 complex with high potency. In the case of NSC146109 and NSC676693 poor ligand solubility could led to negative results even if the compounds are weak inhibitors of the complex.

2.2.5 Compounds: c11 and c7

Here we show the results for two compounds, named c11 and c7, obtained from a biotech company, that we have tested for their inhibitory effects on the MDM2-p53 interaction. The structure and chemical formula of them was not disclosed to us. They are small molecular weight compounds (MW_{c11} is ca. 410 Da, MW_{c7} is ca. 480 Da).

Compound c11

The interaction of c11 with MDM2 (1-118). The DMSO solution of c11 when administered to the MDM2 sample induced changes in the HSQC spectrum of Fig. 2.8. Corresponding induced chemical shift changes are shown in Fig. 2.9 and Table 2.1. The structural illustration is presented in Fig. 2.10. The compound seems to bind to a well-defined binding site on MDM2, near to the p53 binding site, with low affinity. Its concentration could not be increased due to solubility limit (see also Fig. 2.16 for details).

Interaction of c11 with the MDM2-GSTp53 complex. Addition of c11 to the MDM2-p53 complex did not dissociate the complex. Some changes, however, are observed in the ^{1}H NMR spectrum of the complex, in the tryptophan region of p53, thus a weak interaction with the complex cannot be excluded.

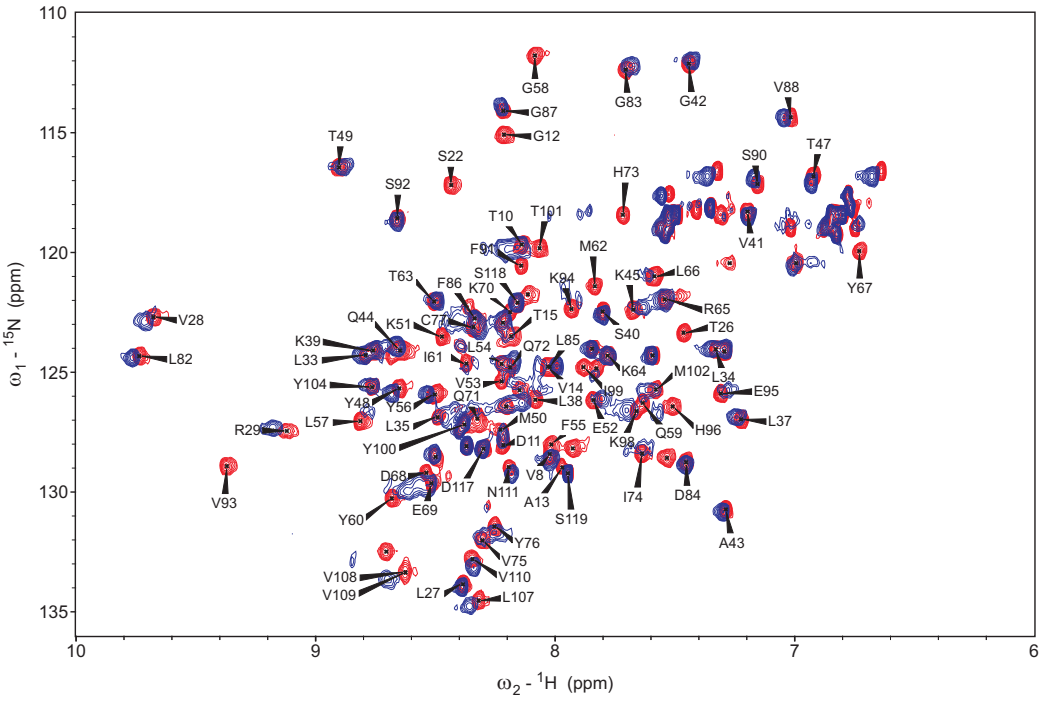

Figure 2.8: The spectrum of MDM2 (red) and MDM2 with c11 (blue).

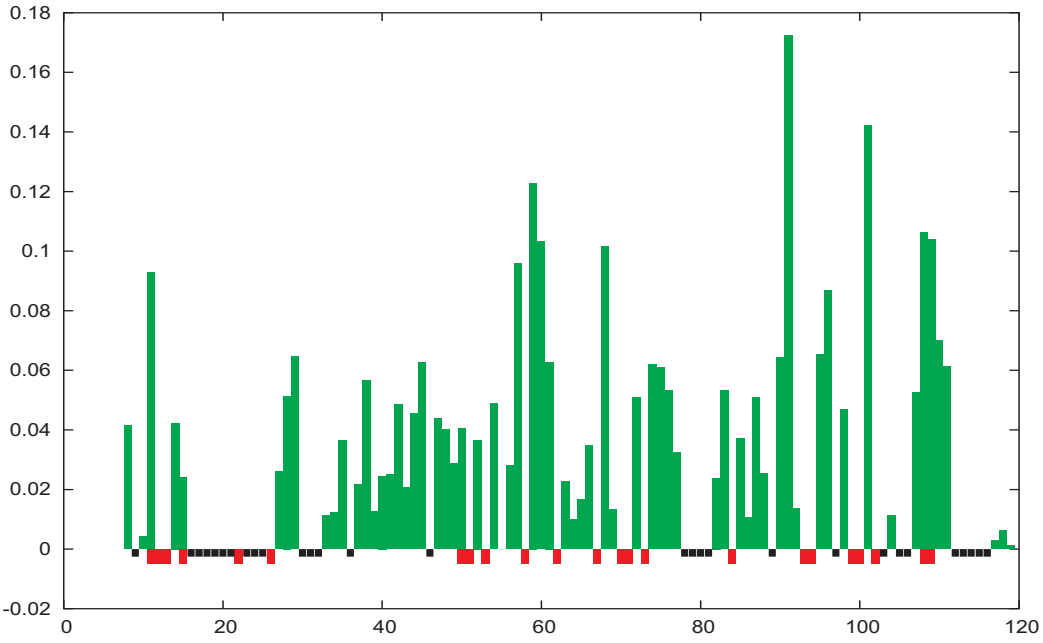

Figure 2.9: The induced chemical shifts caused by c11. Red bars mean: the peak cannot be traced or a mistake is possible, black bars: no assignment of MDM2.

Table 2.1: Expected minimum chemical shifts of MDM2 upon addition of c11.

Residue number	Expected minimum shift
26	0.17
51	0.11
53	0.075
58	0.18
62	0.27
67	0.28
71	0.06
73	0.14
93	0.08
94	0.5
99	0.16

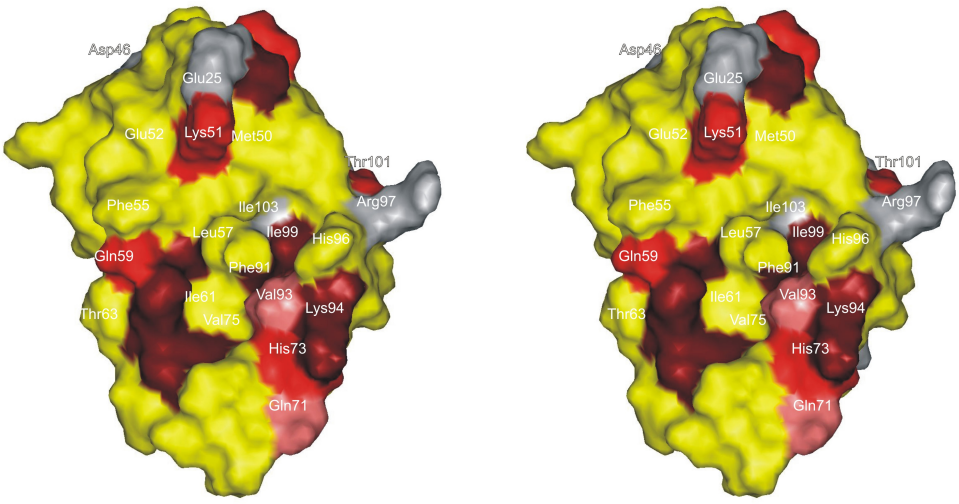

Figure 2.10: The stereo view of the c11 effects on MDM2. Dark red: shift bigger than 0.15; light red: shift between 0.1 and 0.15; pink: expected minimum shift smaller than 0.1 but the peak is missing; gray: no assignment of MDM2.

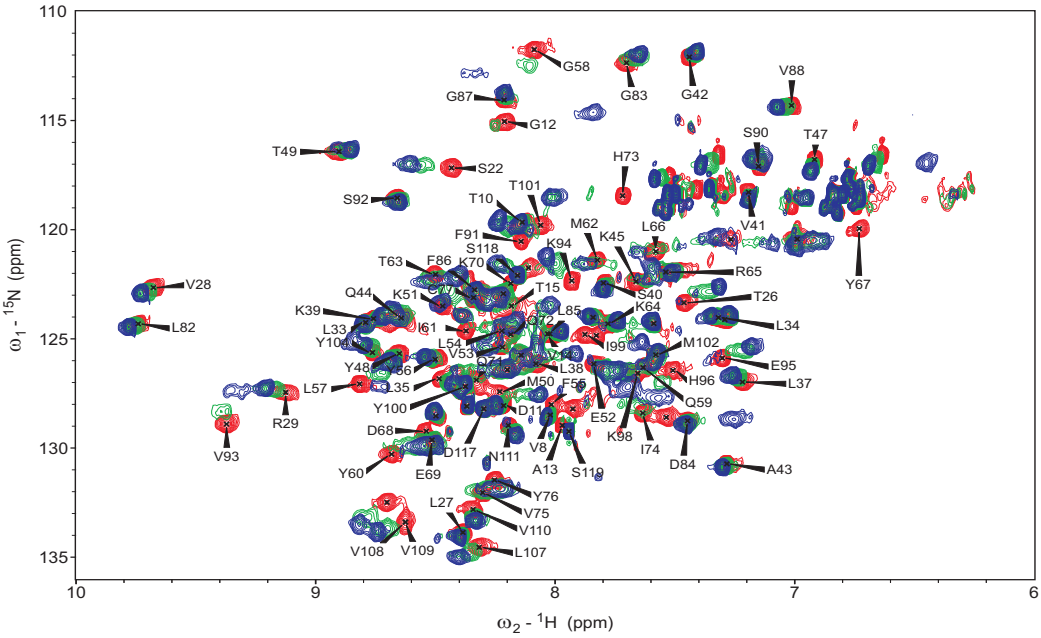

Figure 2.11: The MDM2 spectrum (red), and MDM2 with c7 at the maximum ligand concentration (blue) and a middle step (cyan).

Compound c7

Interaction of c7 with MDM2 (1-118). The DMSO solution of c7 caused changes in the HSQC spectrum of the MDM2 shown in Fig. 2.11. The induced chemical shifts are summarized in Fig. 2.12 and Table 2.2.

These spectral changes can be mapped on the MDM2 structure, shown in Fig. 2.13, revealing well defined binding site. The binding site coincides with p53 binding site.

Interaction of c7 with the MDM2-GSTp53 complex. c7 added to the MDM2-p53 complex binds to MDM2 and releases measurable amounts of p53. One can see in the HSQC presented in Fig. 2.14 that peaks reappear in the HSQC spectrum of ^{15}N labeled MDM2 complexed with p53. The c7 recovers c.a. 30% of p53 - as estimated from the tryptophan sidechain signal in Fig. 2.15.

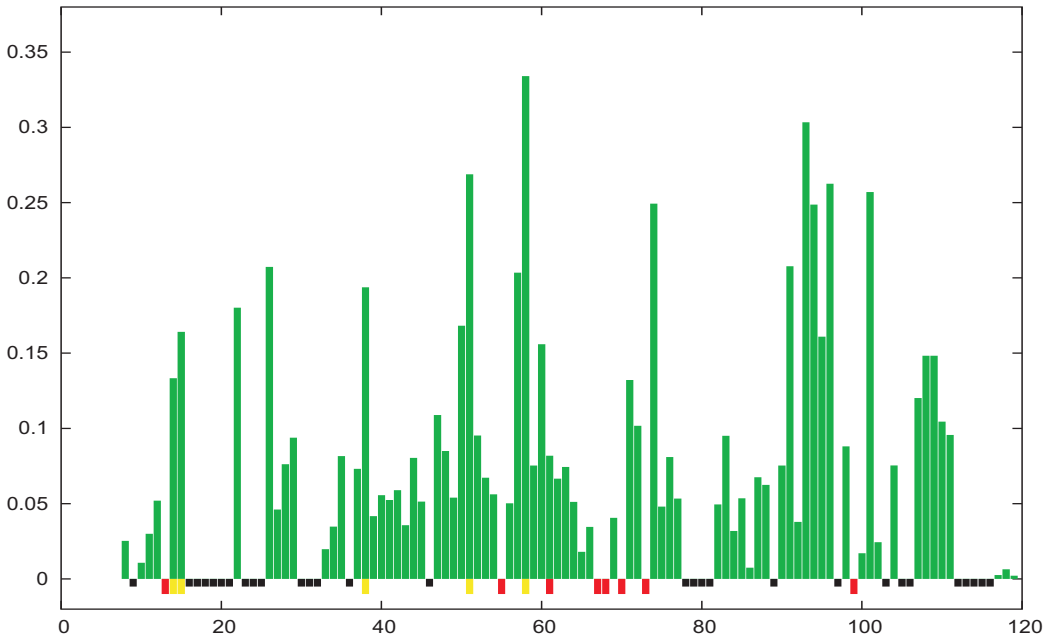

Figure 2.12: Induced chemical shift change in MDM2 spectrum upon addition of c7. Black boxes: no assignment, yellow boxes: the peak position/value can be ambiguous, red boxes: peak is missing after compound addition.

Table 2.2: Expected minimum chemical shifts for c7.

Residue number	Expected minimum shift
55	0.11
67	0.18
68	0.15
70	0.08
73	0.225
99	0.104

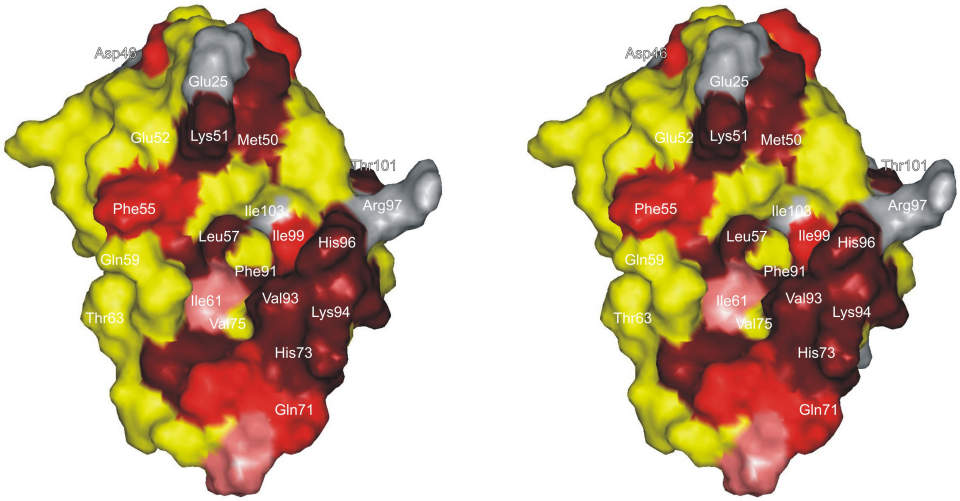

Figure 2.13: The stereo view of the c7 effects on MDM2. Dark red: shift bigger than 0.15; light red: shift between 0.1 and 0.15; pink: expected minimum shift smaller than 0.1 but the peak is missing; gray: no assignment of MDM2.

Both observations consistently confirm inhibitory potency of c7. The excess of the compound could not be increased due to the limit in the c7 solubility. The MDM2-p53 concentration was 0.2 mM, while c7 was added in the amount which should cause a 3-fold molar excess, but precipitation was observed and the maximum of the c7 solubility was reached, thus the concentration of c7 is lower. This explains why no significant c7 signals over protein signals could be detected in a proton 1D spectrum.

Solubility of the c11 and c7 compounds

The result is that The c7 compound was worse soluble than c11 in the protein buffer. This consequently means that the weaker effect of c11 cannot be explained on the basis of solubility and is probably a result of weaker binding affinity. Both compounds seems to bind quite near to the p53 binding site of MDM2. The solubility does not allow for significant excess of compound over MDM2, unless bigger amounts of DMSO are present in the protein sample.

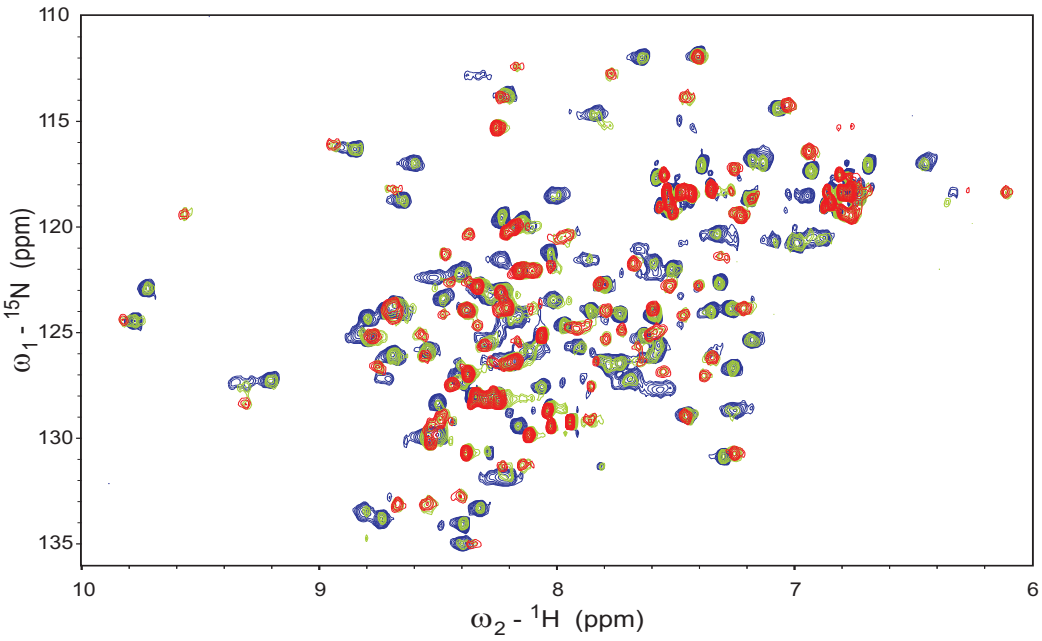

Figure 2.14: The effect of c7 on MDM2-p53 observed in HSQC of (^{15}N labeled) MDM2. In the MDM2-p53 reference spectrum only few MDM2 peaks are observed due to increase relaxation rate (red). After addition of c7, MDM2 spectrum is partially recovered, with changes however (green). These changes are caused by binding of c7 to MDM2. The final spectrum from MDM2/c7 titration experiment (blue).

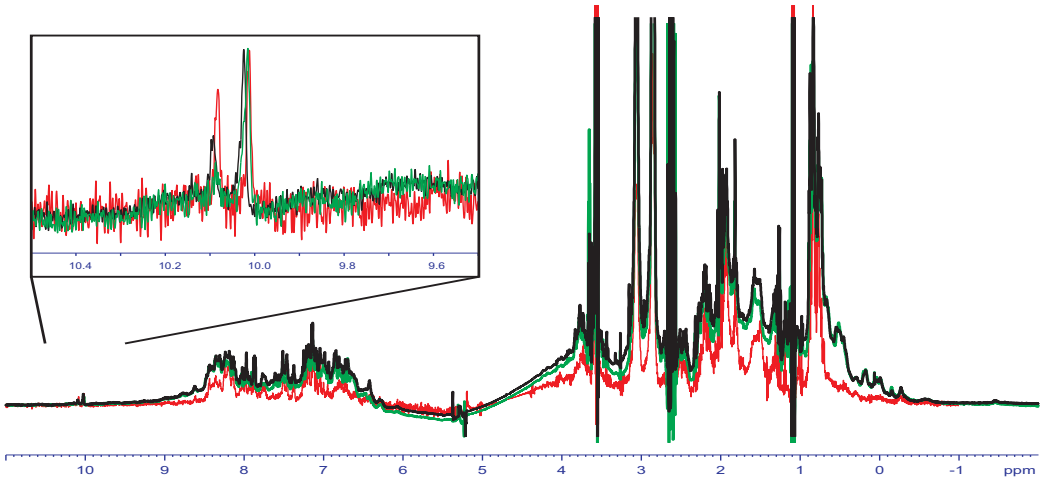

Figure 2.15: The 1D proton spectrum of GST-p53 (red), MDM2p53 complex (green), MDM2p53 complex after addition of c7 (black). The insert shows the Trp sidechain region.

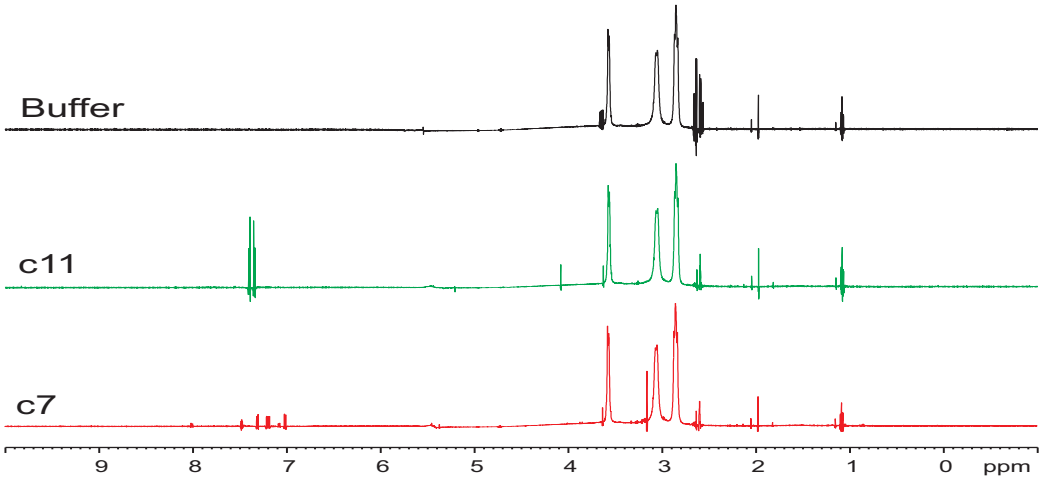

Figure 2.16: The solubility check: The protein buffer spectrum (black), the buffer after addition of c11 (green), the buffer after addition of c7 (red). After addition of each compound a fine precipitate was observed. The signal at 2.6 ppm, different in intensity, is a reducing reagent.

2.2.6 MDM2-p63

A 1D-NMR proton spectrum of p63 has many characteristics of a spectrum of an unfolded protein but the molecule is a tetramer of the total weight of 4×63 kDa, and the transverse relaxation rate is large enough to suppress the folded parts in the spectrum. The outcome is that p63, like p53, contains large unfolded regions.

We have tested p63 for binding to MDM2 (1-118). The formation of the MDM2-p63 complex was observed in the same way as for p53 (see: 2.2.1), however, the addition of nutlin-3 did not restore the MDM2 even when nutlin-3 was in ten fold excess.

2.3 Conclusions

Nutlin-3. Nutlin-3 has proved indeed to be an inhibitor of the p53-MDM2 complex and therefore is a potential candidate for drug development in cancer treatment. An interesting side-result of our experiments is the determination

of the folding status of our p53 free and bound to MDM2. Characterization of this status has been a subject of several recent studies. [42, 43, 46, 44] Our NMR spectra indicate that the first 93 residues are flexible and unstructured in agreement with the findings that showed that the full-length p53 contains large unstructured N- and C-terminal regions in its native state. [43, 42, 46, 44] The conformations of these residues have to be the same as those found in [44]; i. e. although the p53 transactivation domain does not have tertiary structure it is nevertheless populated by a nascent helix and turns[46, 44]

RITA. In summary, our results show lack of inhibition by RITA of the in vitro interaction of 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. Interactions through other binding partners or stabilization of the complex against ubiquitin driven proteolysis (or modulation of gene expression levels of proteins) might be a probable mode of action of RITA. We note that structurally similar tiophene derivatives were cytotoxic particularly, but not only, after irradiation with near UV light. The action included production of singlet oxygen, but also other, not identified mechanisms [63, 64].

We also showed that RITA does not bind to the N-terminal domains of p53 and MDM2. The possibility that RITA targeted other regions of p53 or MDM2 is unlikely because we do not expect that a ligand that does not interact in vitro with the primary binding sites of target proteins (N-terminal domains of p53 and MDM2) could show important binding in vivo.

The high sensitivity of fluorescence correlation spectroscopy (FCS) is known to result in false positive results. We have contacted the authors that reported the action of RITA on the MDM2/p53 complex [59] and they

made FCS measurements on the protein produced by our group. The FCS effect is also present, distinctly smaller, however. The observed differences in our NMR and Issaeva et al. [59] FCS results could arise if their measurements had increased sensitivity to minor sub-populations of the 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). Radical differences between in vitro and in vivo experiments, should they exist, will always have a molecular basis. Should functional interactions radically differ, origins will be the presence of unidentified additional binding partners, "surprising" co-factors, such as specific ions etc.

Compounds c11 and c7. Our experiments showed the inhibitory potency of c7, and using DIMAPPS, classified this compound as a better inhibitor than c11. By "traditional" SAR by NMR such a decision would be much more difficult.

In both cases it is difficult to quantify the K_Ds, since the ligand solubility was reached at relatively low compound concentrations. We could classify the K_D of c7 compound in the range of 3-15 μM.

It is also the first observation of dissociation of the MDM2-p53 complex by a "fast-exchanging molecule" (c7). Thus, the method performed well in the case of fast-exchanging weakly-soluble compound.

2.4 Methods

Protein expression and purification. The recombinant human MDM2 (residues 1-118) was overexpressed at 30° C in *Escherichia coli* BL21 (DE3) using the pQE-40 vector (Qiagen). The protein was renatured from *E. coli* inclusion bodies as previously published [65]. Refolded MDM2 was first ap-

plied 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 room temperature in *E. coli* BL21 (DE3) using a modified pQE-40 with N-terminal His-tag and T5 promoter as described in [43]?. 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-NH4Cl [33].

NMR spectroscopy. All NMR spectra were acquired at 298 K on a Bruker DRX 600 MHz spectrometer equipped with a cryoprobe. Typically, NMR samples contained up to 0.15 mM of protein in 50 mM KH_2PO_4, 50 mM Na_2HPO_4, 150 mM NaCl, pH 7.3, 5 mM DTT, 0.02 % NaN_3. Titration experiments were performed using a series of proton 1D and ^{1}H-^{15}N HSQC spectra of labeled p53 or MDM2 along with the unlabeled partner. For the ^{1}H-^{15}N HSQC spectrum, a total of 1024 complex points in t_2 and 128 t_1 increments were acquired (96 when a high concentration of DMSO in the ^{15}N MDM2 sample was used e.g. when nutlin-3 was added). Water suppression was carried out using the WATERGATE sequence. NMR data were processed using the Bruker program Xwin-NMR version 3.5 and analysed with a program Sparky [66].

Structural illustration of chemical shift changes The induced chemical shift changes were calculated from the formula:

$$\Delta = \sqrt{(\Delta\omega_H)^2 + (0.2 * \Delta\omega_N)^2} \tag{2.1}$$

and plotted versus the amino acid sequence of MDM2 (the same procedure as in [43]).

Some peaks cannot be traced after addition of the compound. This can be the result of overlap or large chemical shift differences caused by binding. Since these residues may be the strongest interacting residues, in such cases we have estimated the *expected minimum shift* measuring the chemical shift to the nearest neighbor. The results are summarized in the tables, and marked on the MDM2 structure pictures according to the estimated value. In some cases this procedure likely underestimates the shift and classifies the residue to be yellow (neutral). In such cases pink color was used.

Val108 and Val109 are assigned to one peak in the MDM2 spectrum [43]. This peak is involved in binding, so both V108 and V109 were treated the same. However, it must be kept in mind, that the information content about this residues is limited.

Chapter 3

Cyclin-dependent kinase 2, cyclin A2, and p27

3.1 Biological context

The cell division cycle in eucaryotic cells consists of several phases: a G_1 phase (the first gap), the S phase (DNA synthesis), the G_2 gap phase, and M (mitotic) phase. The family of proteins named cyclin-dependent kinases (CDKs) were identified to play a critical role in different phases of the cell division cycle. The activation of CDKs requires binding to one of the regulatory proteins called cyclins and subsequent phosphorylation by a CDK activating kinase [67].

In the G1 phase the pRb protein is bound to the E2F/DP family transcription factors, and the pRb/E2F/DP complex acts as a transcriptional repressor. Upon receiving mitogenic signals, repression by pRb is removed by its phosphorylation by CDK4/6-cyclin D complexes followed by hyperphosphorylation by CDK2-cyclin E/A complex. The activity of the latter complex is negatively regulated by CIP/KIP protein family (p21, p27, and p57) of inhibitors. The CDK2 cyclin E/A complex is thought to be the major regulator of entry to and progression through the S-phase [68]. Later on inhibitory CDK2 dephosphorylation on Y15 accompanies mitosis and progress through G2-M checkpoint.

3.2 Results and discussion

3.2.1 p27-CDK2 interaction

The first experiment was performed on the uniformly ^{15}N labeled CDK2. The uniformly ^{15}N labeled CDK2 sample was titrated with increasing amounts of unlabeled p27. Splitting of the signals, showing that reaction is in slow exchange, indicated tight binding. The experiment was continued until excess of p27 could be observed in the proton 1D NMR spectrum.

In the second experiment changes in the HSQC spectrum of the ^{15}N labeled p27 protein upon addition of unlabeled CDK2 were measured. The results are presented in the Fig. 3.1. The effect could also be observed in proton 1D spectra.

It was expected that unfolded (in its free state) p27 may fold upon binding to the CDK2 complex as in the case of p21-CDK2 interaction [69]. However no folding of p27 was detected in our measurements.

3.2.2 CDK2 interaction with roscovitine

Roscovitine is a known inhibitor of CDK2, which binds in the ATP binding site of CDK2 [70]. We have measured the CDK2 chemical shift changes in 2D HSQC spectra caused by roscovitine (Fig. 3.2). The results confirm the interaction, and several well resolved peaks sensitive to roscovitine could be identified. Further analysis, however, needs assignment of CDK2 resonances, which is not available until now.

3.2.3 p27/CDK2 and roscovitine

Since the binding site for p27 probably does not spatially interfere with the roscovitine binding site, both molecules should bind independently. The changes in the spectrum caused by p27 binding in general (see: 3.2.1) do not coincide with those observed for the roscovitine titration (see: 3.2.2), which supports the previously published results.

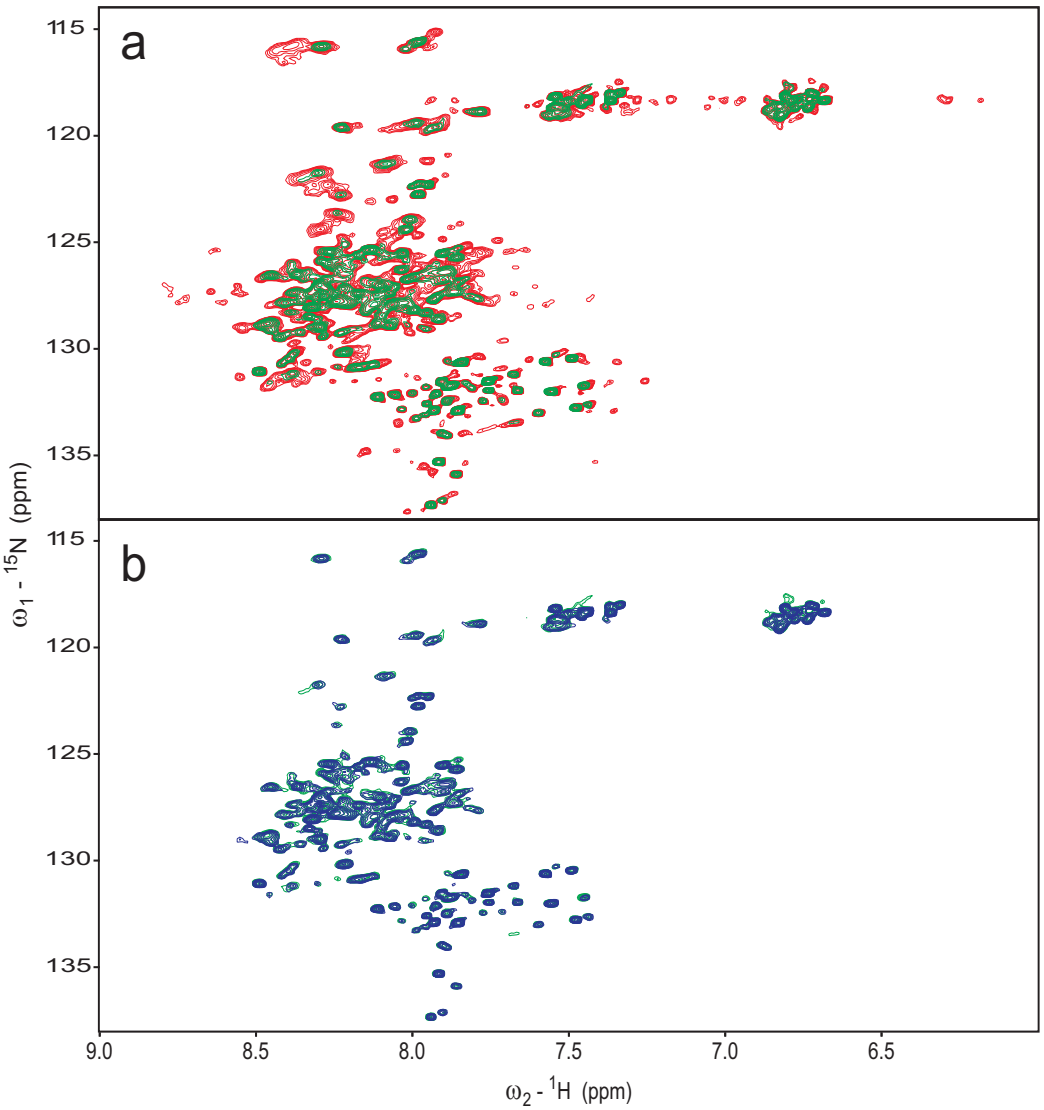

Figure 3.1: Titration experiments of ^{15}N labeled p27 with CDK2 (a) and succesive titration with unlabeled p27. A reference spectrum of p27 (red), the spectrum after addition of (unlabeled) CDK2, several peaks disappear (green), the spectrum after addition of unlabeled p27 (blue), no significant changes were observed.

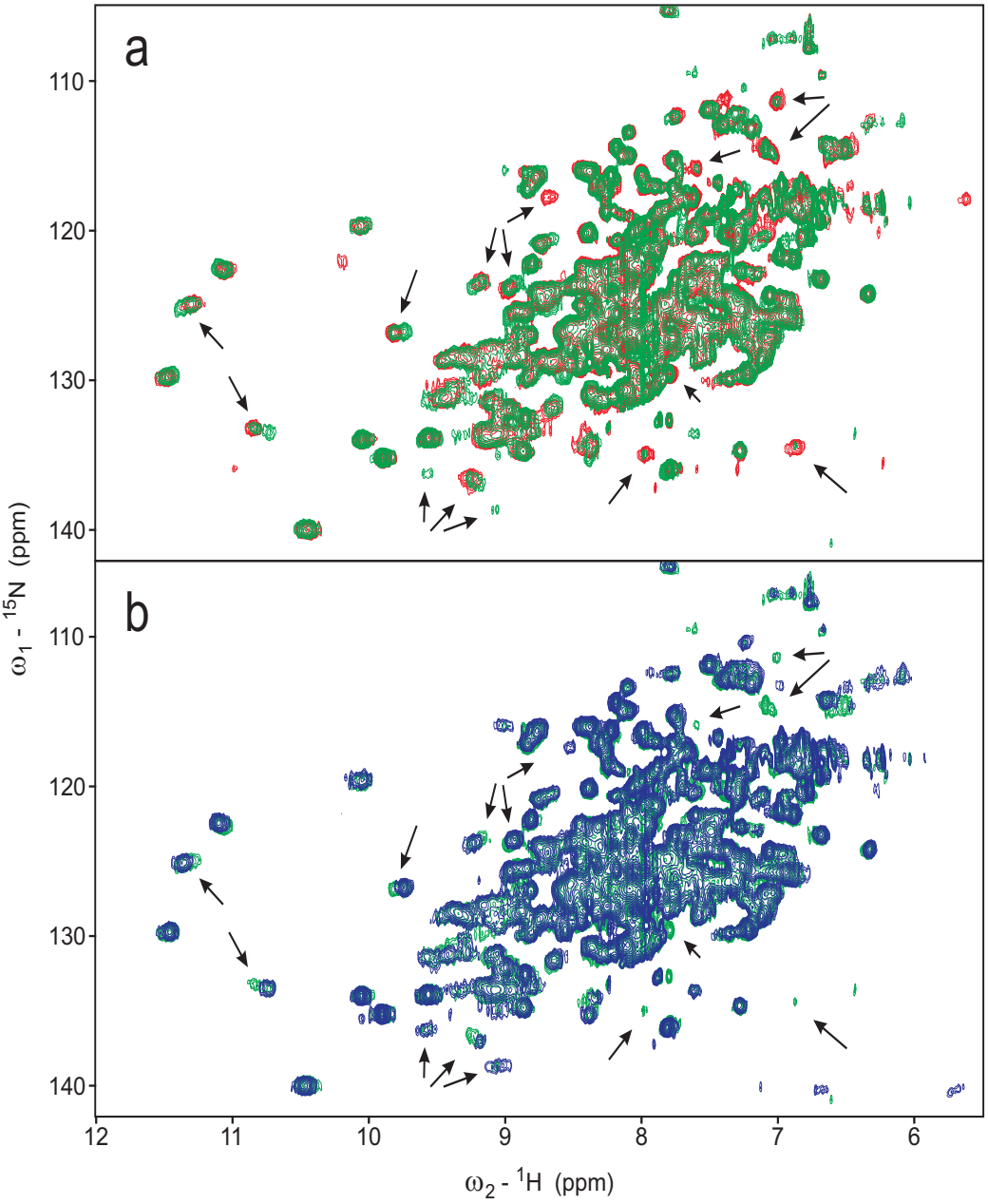

Figure 3.2: Titration experiments of ^{15}N labeled CDK2 with roscovitine. (a) a reference spectrum of CDK2 (red), after addition of roscovitine several well resolved peaks display splitting, or decrease in intensities (green), (b) the spectrum with an excess of roscovitine (blue). Arrows indicate most distinct changes in the spectrum.

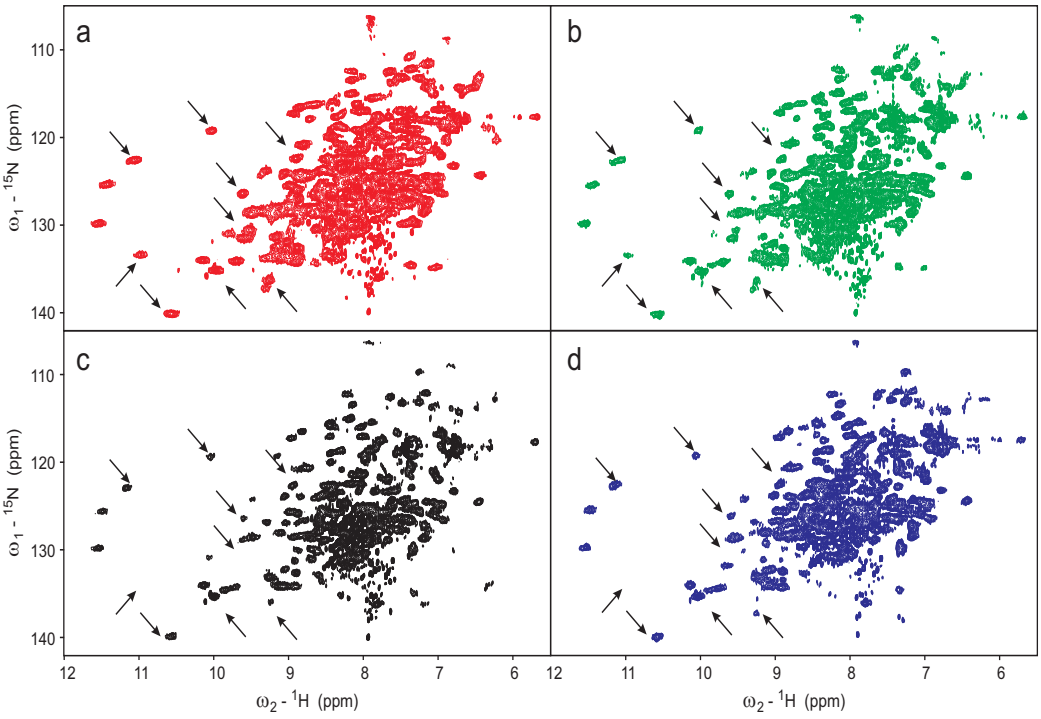

Figure 3.3: Titration experiments of ^{15}N labeled CDK2 with unlabeled p27. (a) reference spectrum of CDK2 (b) the spectrum after addition of (unlabeled) p27 several peaks display splitting and decrease in intensity, (c) the spectrum after addition of an excess of p27 several peaks disappear or have changed position. (d) the spectrum after addition of roscovitine other peaks are affected.

To confirm this, we added roscovitine to the CDK2-p27 complex (CDK2 titration with p27) as described in 3.2.1. New changes in the spectrum were clearly seen, which correspond well to the results from the CDK2/roscovitine titration. This proves that roscovitine interacts with the p27/CDK2 complex. No dramatic release of p27 could be observed as judged from a proton 1D spectrum. Apart from one peak, which reappears after addition of roscovitine there were no reverse changes in CDK2 peaks due to p27 binding, thus it can be excluded that p27 and roscivitin bind simultaniously to CDK2, the binding strength of p27 might, however, depend on the presence of roscovitine.

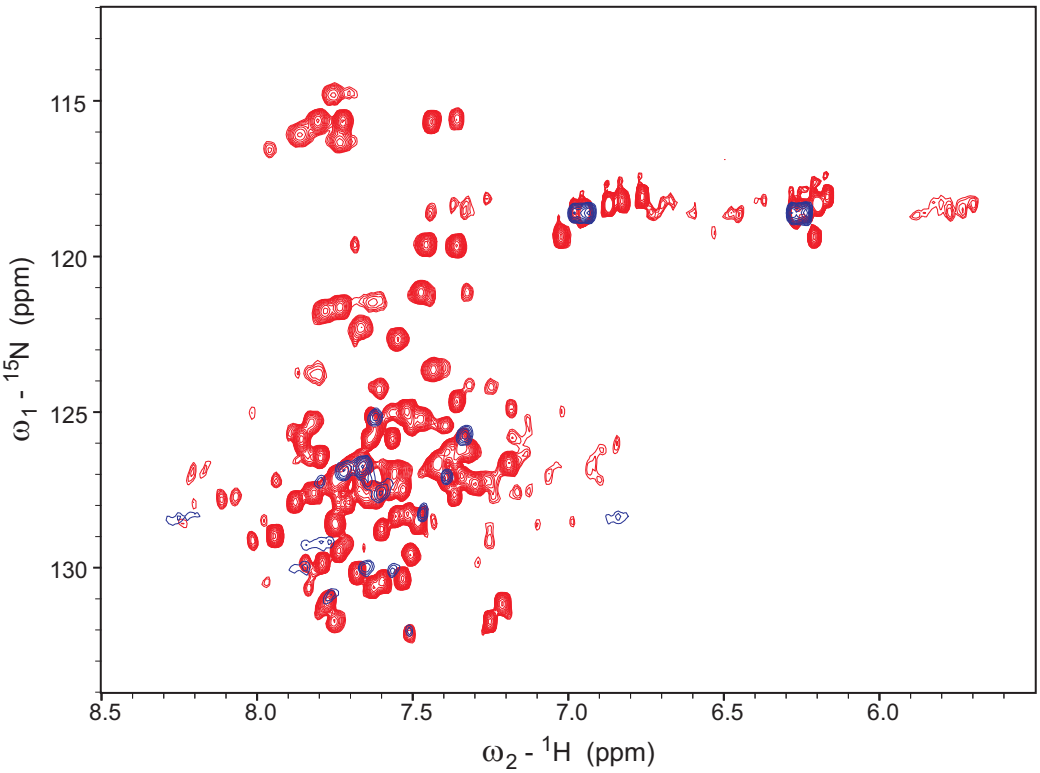

Figure 3.4: The HSQC spectrum of ^{15}N labeled p27 (red), and after addition of CDK2/cyclin A complex (blue).

3.2.4 Interaction of p27 with the CDK2/cycA complex

In this experiment interaction between p27 and CDK2/cycA complex was monitored. The ^{15}N labeled p27 was titrated with unlabeled CDK2/cycA complex. Formation of the ternary complex was manifested in the HSQC spectrum by considerably increased relaxation rates, and consequently disappearence of most of the peaks (Fig. 3.4). To test whether the bound p27 exchanges, unlabeled p27 was added and measurements were repeated. However, no significant change was observed in the spectra indicating strong interaction between p27 and the CDK2/cycA complex.

3.2.5 Can p27 be displaced from its complex with CDK2 only?

Since the interaction of p27 with the CDK2/cycA complex was so strong that exchange of bound ^{15}N labeled p27 with unlabeled p27 added afterwards was not observed, we decided to use only CDK2 for the test experiment, to check its binding with p27.

From the 1D spectra of ^{15}N-labeled p27 (Fig. 3.5a) the molar ratio of p27 to CDK2 was estimated to be nearly 1:1 with an excess of p27 less than 10%, (Fig. 3.5b). Addition of unlabeled p27 did not restore the original HSQC spectrum, and no significant change could be found. From the ^{1}H 1D NMR spectrum (the tryptophan region) one can see excess of free p27. The excess protein is however not labeled, since no J_{HN} splitting of the tryptophan peak was observed (Fig. 3.5c). The sample was also heated and incubated at 40°C for 80 minutes and the NMR measurement was repeated, giving the same result.

3.3 Conclusions

Our results show that p27 is not structured in solution. We also showed that p27 could bind to CDK2 alone. In our experiments the binding was tight even when only CDK2 was used, the exchange rate of p27 was very slow, undetectable even after several hours. We have also shown a simultaneous binding of p27 and roscovitine to CDK2, indicating that they interact at separate sites on CDK2.

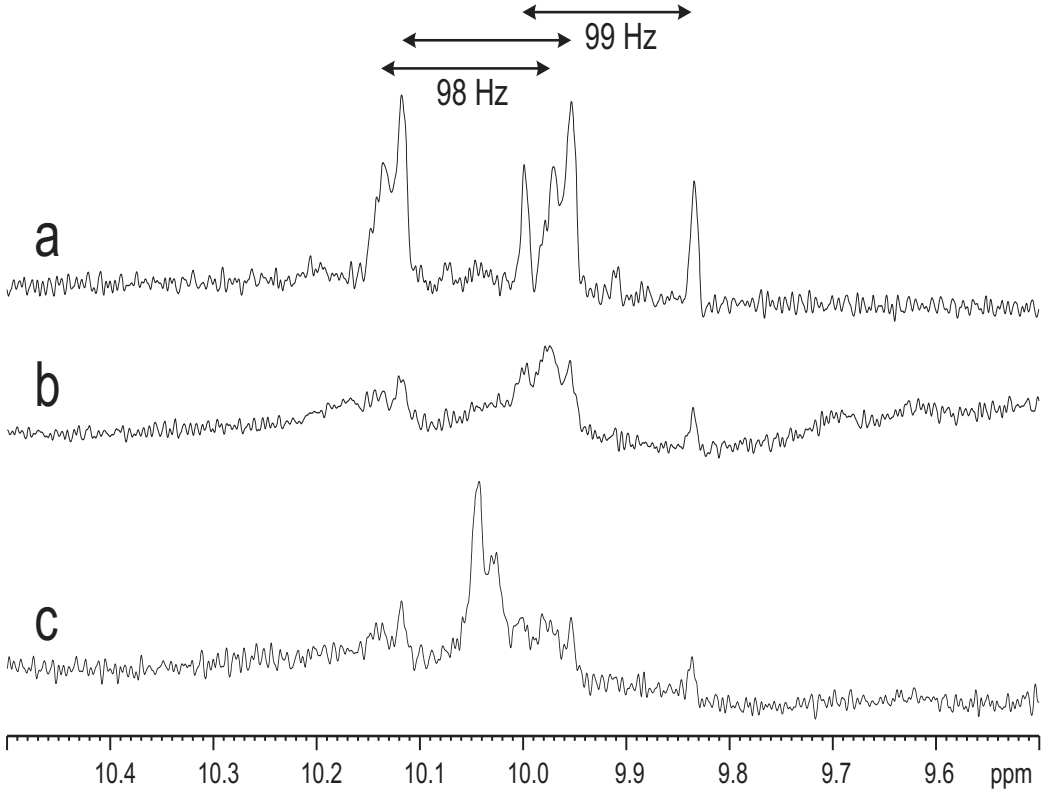

Figure 3.5: A tryptophan region of the proton spectrum of ^{15}N labeled p27 (a), the spectrum after addition of CDK2 (b), and the spectrum after addition of unlabeled p27 (c).

Chapter 4

pRb, E7, E2F

4.1 Biological context

The retinoblastoma protein (pRb) belongs to the family of so-called pocket proteins. It is a 928 aminoacid long protein composed of N-terminal, A, B and C domains. The central "small pocket", formed by A and B domains (residues 379-578 and 641-791 respectively) connected by a linker region, is of particular interest as it is found mutated in many cancers. The linker is flexible and not necessary for protein integrity [71].

The hypophosphorylated pRb protein causes the cell cycle arrest in G1 phase, and its inactivation weakens the cell cycle control. The activity of pRb is regulated by phosphorylation – active, underphosphorylated protein is mostly present in G1 phase and binds E2F transcription factors and inhibits them. Hyperphosphorylation of pRb inactivates it leading to release of transcriptional competent E2F. Pocket domain together with C-domain (large pocket domain) harbour the binding sites for many proteins. In many human cancers pRb is inactivated either indirectly or directly by mutations or by interaction with viral oncoproteins.

Structures of pRb small pocket domain bound to various peptides have been solved by X-ray crystallography. The fold of A and B-domains of pRb is cyclin-like, mostly α-helical in nature [71, 72].

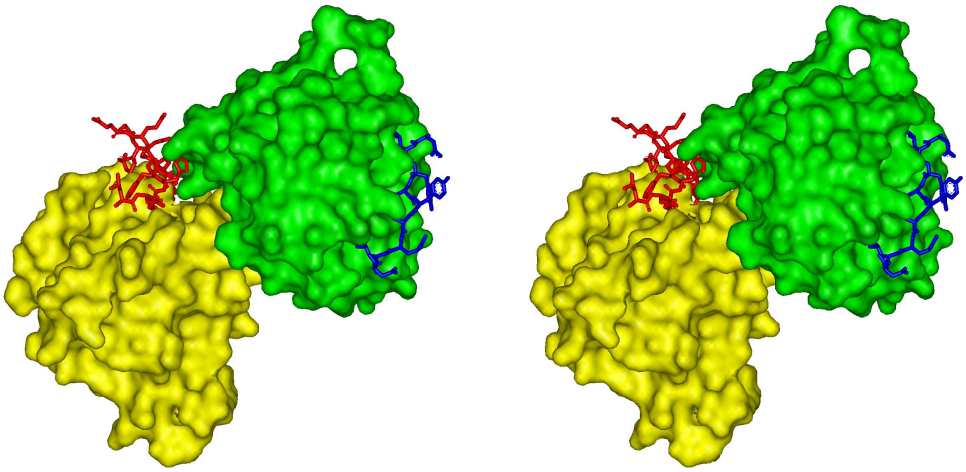

Figure 4.1: The solvent accessibility surface of pRb (probe diameter 1.4 Å). Yellow: A domain, green: B domain, blue: E7 peptide (based on [71]), red: superimposed E2F peptide (based on [72]).

The small pocket of pRb protein includes the primary binding site for E2F proteins, as well as LXCXE-sequence-containing proteins. This LXCXE binding site is a target for many viral oncoproteins, like for example human papiloma virus (HPV) E7 and SV40 large T antigen. Also many cellular proteins posses LXCXE motif, including histone deacetylases 1 and 2 (HDAC1, HDAC2) and BRG1. More than a hundred proteins are reported to interact with pRb [73, 74]. It was not clear whether the cellular proteins interacts with pRb by LXCXE motif or by different mechanism, involving different regions. A particular result could be also a false positive due to extreme sensitivity of the assey used. We have used direct in vivo experiments to test several peptides containing LXCXE sequences from different proteins for binding to pRb.

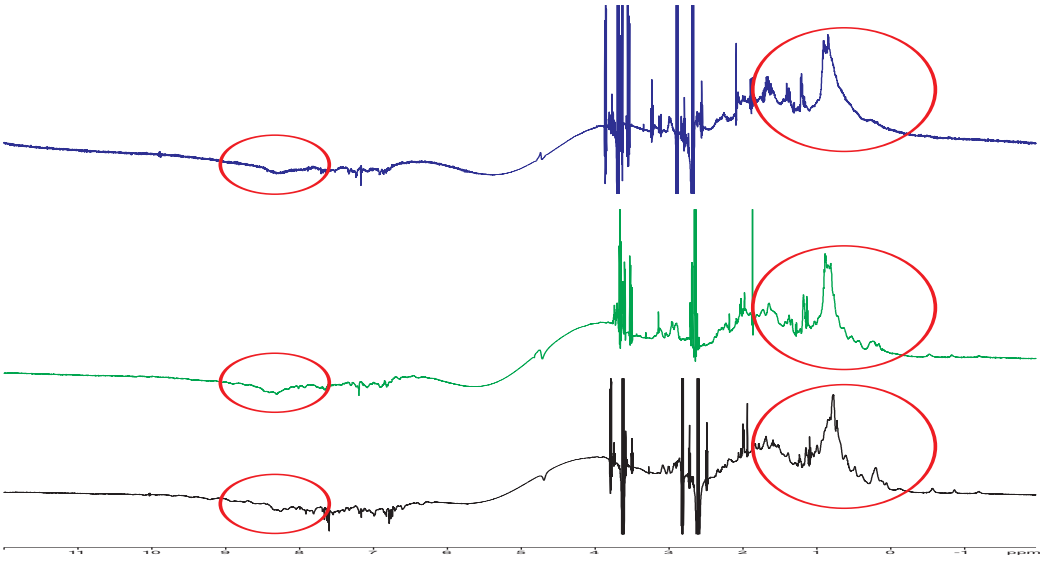

Figure 4.2: Comparison of 1H 1D spectra of different constructs of pRb. The ABC construct (blue), the AB (green) and AB without loop (black). The regions diagnostic for folding are indicated.

4.2 Results and discussion

Construct selection

In our group we have purified and measured with NMR several pRb constructs. Most successfull tructs were:

pRb-ABC , residues 398-928

pRb-ABC-C, residues 398-791

pRb-AB, residues 398-578 and 642-791

K713A mutant of pRb-AB

The results of NMR measurements show that only A and B domains are well folded. Our results showed that the linker region is flexible and disorded, as well as C domain. Whether an in vitro results of folding states of the linker and C domain represents the in vivo situation is not clear. The expression

of B-subdomain in *E. coli* as soluble fraction was unsuccessfull as well as refolding attempt from insoluble inclusion bodies. It was suggested that B domain needs A domain to stabilize the fold [71] which would explain our results.

4.2.1 Overcoming limited resolution

Interaction of pRb with selected peptides can be detected from ^{15}N uniformly labeled sample, however the details are mostly hidden in the crowded center of the spectrum (pRb-AB is composed of 378 amino-acids). Since the resolution of HSQC spectrum was too poor to observe binding in details and the sample precipitation was observed after few days, we decided to proceed with binding studies on selectively labeled samples. On the basis of the crystal structure, lysine was chosen for selective labeling since several lysine residues are involved in binding.

Perdeuteration was also used to enhance the resolution of the HSQC, and indeed the HSQC spectrum improved, however a lot of overlapped peaks could not be resolved.

4.2.2 Assignment attempt

The NMR resonance assignment of pRb is not known. K713 was assigned directly by a one site mutation. The K713A mutant is well folded and displays very small changes in the NMR spectrum in comparison with the pRb-AB spectrum (the excellent overlap of lysine spectra being a nice illustration).

Several selective samples were prepared and measured. The spectra of selective samples are well resolved, also NOESY-HSQC spectrum of the amide region of pRb was recorded on perdeuterated sample. On the basis of crystal structure one can expect that several sequential connectivities between amide protons can be extracted due to α-helical secondary structure. In fact we could find several H_N-H_N connectivities, however no information on residues

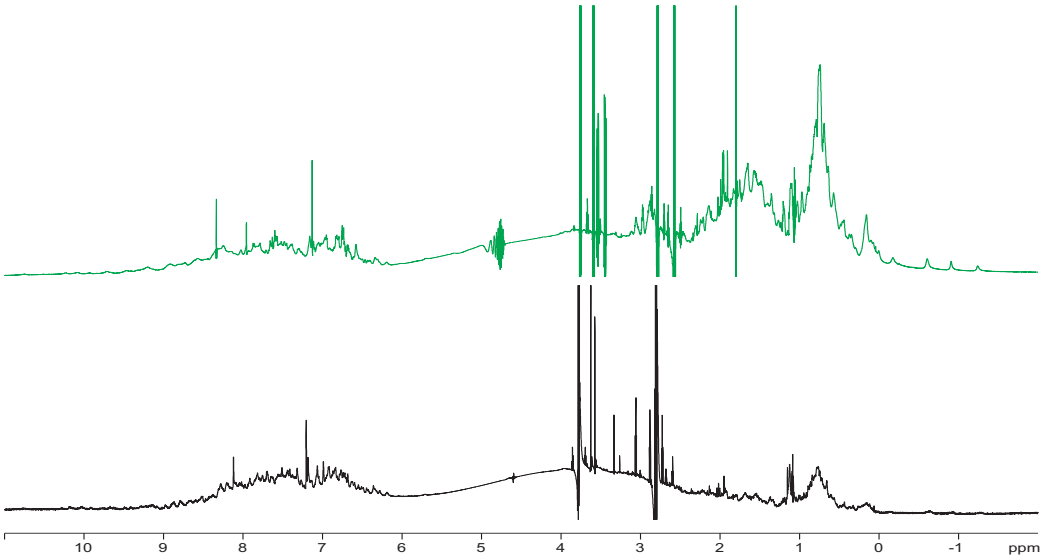

Figure 4.3: Comparison of ^{1}H 1D spectra of ^{15}N-Lys labeled (red) and uniformly ^{15}N ^{2}H labeled sample (black) displays triking difference in the relative intensity of aliphatic (-1.0–4.0) to amide (5.5-10.5) regions.

affected by E7 peptide binding could be extracted. The quality and resolution also did not allow for unambiguous assignment of any other residues and no further effort was made on the assignment.

4.2.3 Identification of the binding sites

Experiments with E7 and E2F peptides on ^{15}N-lysine sample allowed us to distinguish two different binding sites. The number of affected peaks quite well corresponds to the expected number on the basis of structure.

E7 binding site

The titration experiment with ^{15}N-lysine labeled sample and E7 peptide is presented on the (Fig. 4.5). Distinct changes point out six residues, one being extremely sensitive to sample conditions and not seen in all spectra. In the crystal structure E7 peptide is shown to bind to B domain.

The chemical shift differences between free and bound forms are compa-

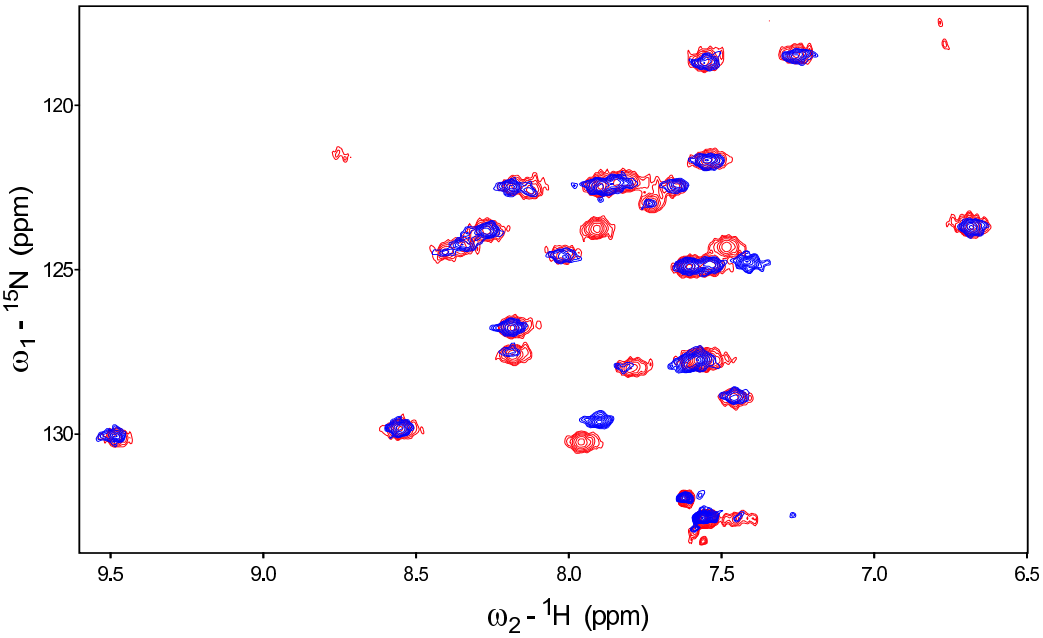

Figure 4.4: Assignment of K713. Superposition of a reference spectrum of ^{15}N-Lys labeled sample (green) and ^{15}N-Lys labeled sample of K713A mutant.

rable to the line-width, two peaks can be clearly seen at the middle step of the titration experiment indicating tight binding.

The K713A mutant was also tested for binding to E7 peptide. The result is positive, the mutation did not abrogate the E7 peptide interaction. The binding was in the slow exchange regime, indicating tight complex formation.

E2F binding site

After E7 peptide titration a short E2F peptide was added to the sample and no significant change in the spectrum was found. This is in agreement with structure [72], since these residues in E2F peptide form a loop which has no, or only week contact with pRb.

Next we used a longer E2F peptide for the titration. The E2F peptide binds to the pocket at the A and B domain interface. The result is shown in

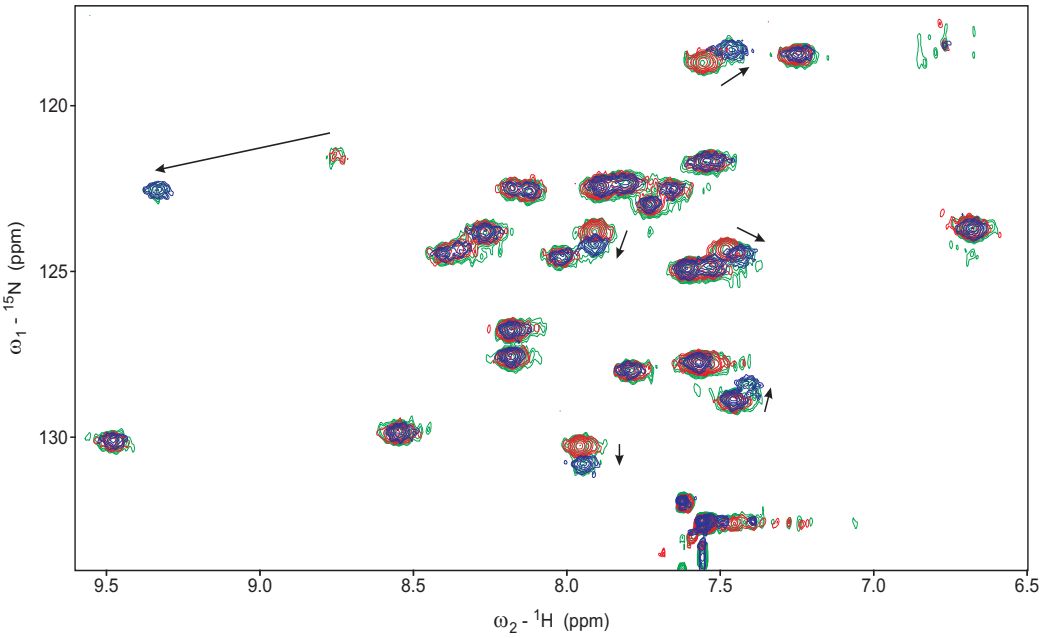

Figure 4.5: E7 peptide titration on pRb ^{15}N-Lys labeled. The resonances in the initial spectrum (red) split into two after addition of small amount of E7 peptide (green). In the final step they are at new position (blue).

the Fig. 4.6. Changes in the spectrum caused by E2F peptide binding are pointed by arrows. This experiment well agrees with the result described in [72] which showed that E7 protein and shorter constructs of E2F or longer E2F peptides can bind simultaneously to the pRb, since their binding sites are different (ca. 30Å apart each other).

4.2.4 LXCXE peptides titrations

HDAC peptide The first outcome from the HDAC peptide titration is that the binding can be classified as fast exchange thus weaker as compared to E7 or E2F peptides. The estimated K_D value for pRb-Ab/HDAC peptide was in the range of 40-50 uM.

HDAC mutants peptides An interesting result was obtained for HDAC RxxxD mutant peptide. This mutation increased the binding strength of the peptide, almost to the extent of tight binding.

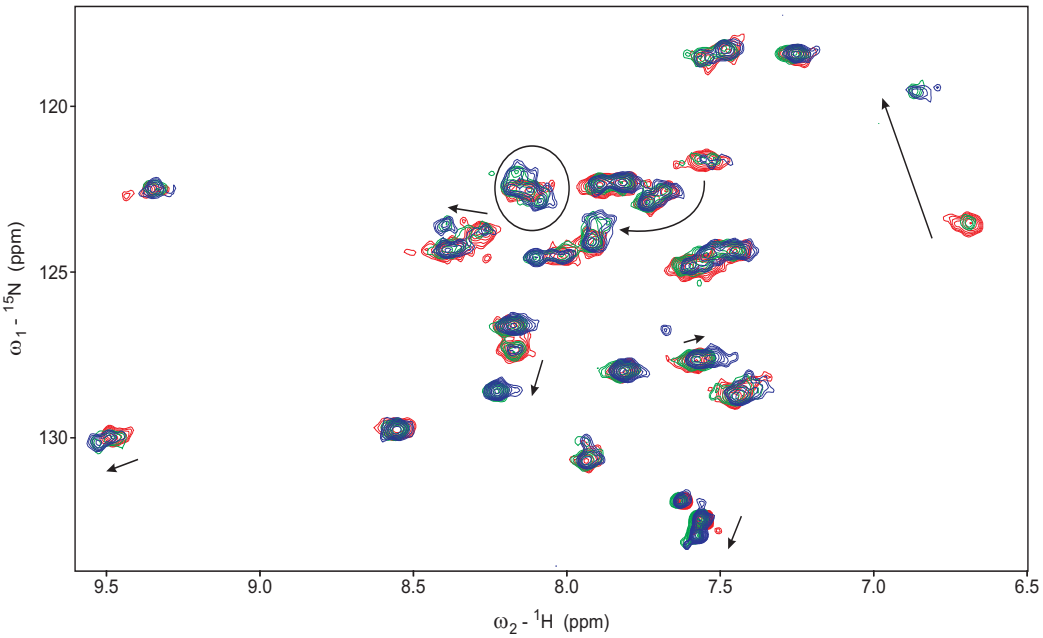

Figure 4.6: A superposition of pRb ^{15}N-Lys labeled reference spectrum (red), after addition of a dose of E2F-long peptide (green), and after addition of more E2F peptide

E7 mutant peptide Analogous, but reverse mutation in E7 peptide, did not weaken it's bindiing affinity beyond tight binding limit, and the K_D could not be determined by NMR, while the ITC measurements have shown dramatic decrease in binding strength.

4.2.5 pRb interaction with E7 protein

Since ^{15}N labeling of E7 protein was possible, we decided to proceed with E7 protein. The 1D spectrum indicates, that E7 contains large unstructured regions. During purification in gel-filtration process it elutes in two peaks, as a dimer and as an oligomer. For measurements the dimer was chosen, and the measurements were conducted on freshly purified protein. The molecular mass for the dimer corresponds to c.a. 28kDa. The HSQC spectrum (Fig. 4.7) is typical for unstructured, flexible polypeptides [53].

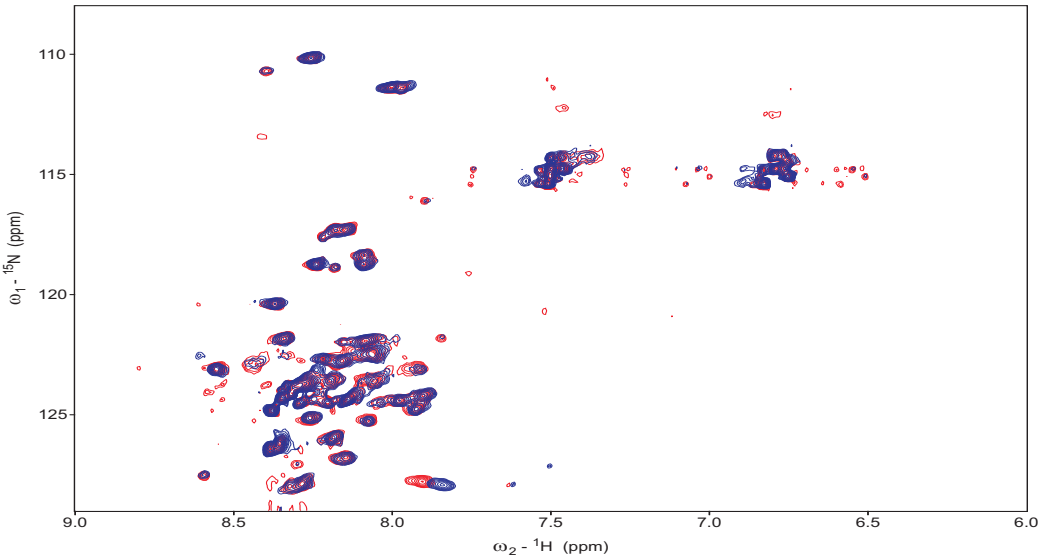

Figure 4.7: The HSQC spectrum of ^{15}N labeled E7 protein (red) and after addition of A and B domains of pRb (blue).

4.2.6 Interaction of pRb with E2F and E7 proteins

The interaction of pRb with E7 and E2F proteins were studied using per-deuterated and uniformly ^{15}N labeled pRb sample. First E2F-1 (243-437) protein was added. The resulting spectrum displays differences: several peaks disappear and some new peaks appeared.

An interesting observation was made with one of lysines. Initially a sharp and intense resonance was present at 7.9 (^{1}H), 132 (^{15}N) ppm (let us denote it K^{E2F}). This peak splits into two of similar but smaller than the initial intensity, a behavior identical to that observed in the case of E2F peptide titration (Fig. 4.6).

Afterwords E7 protein was added, which resulted in line-broadening, unequivocally confirming interaction of E7 protein with pRb-AB in the presence of E2F. The interesting outcome from this experiment was that the K^{E2F} was seen afterwords and hence was not affected by E7 binding. The immediate suggestion is that E2F remains bound even after formation of a complex between pRb and E7.

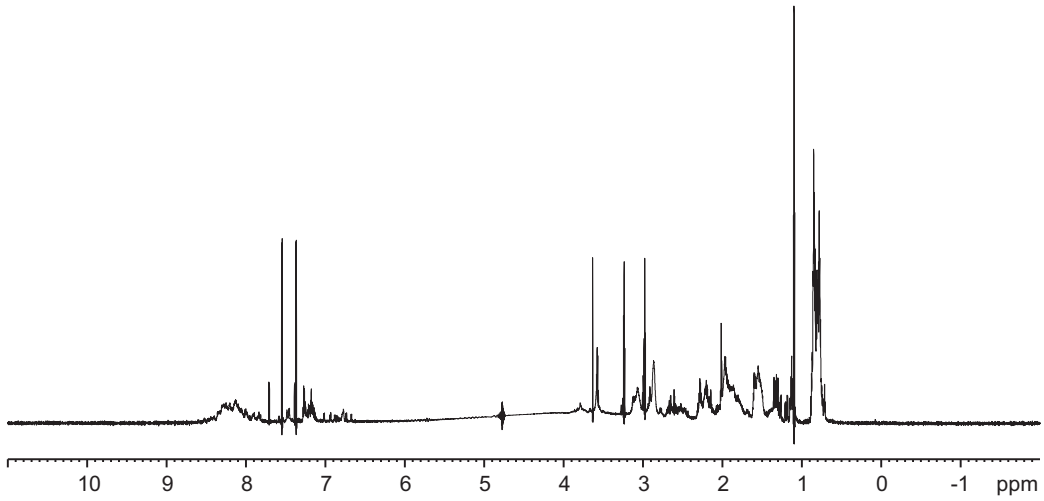

Figure 4.8: A 1D spectrum of E2F construct.

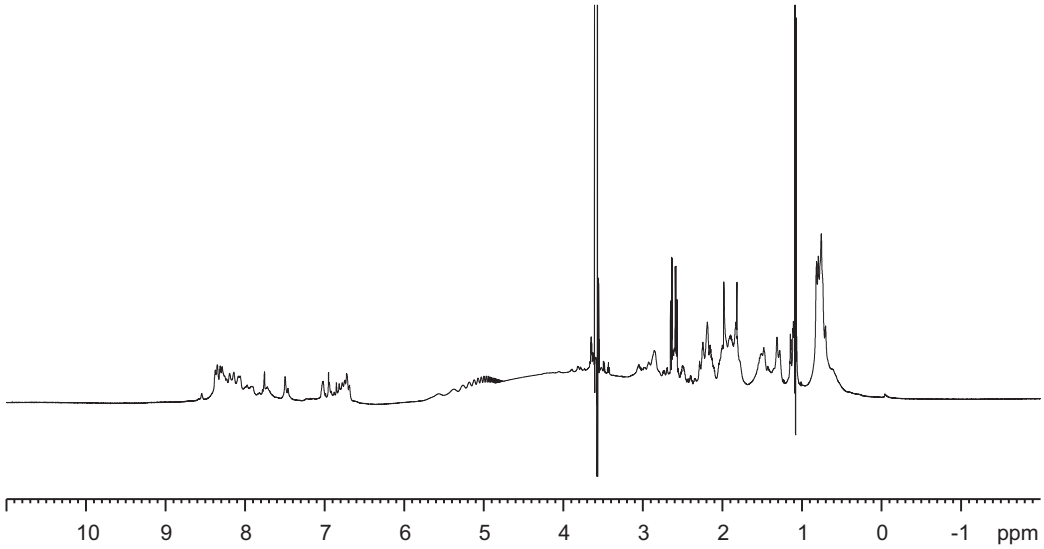

Figure 4.9: A 1D spectrum of E7 protein (dimer).

The E2F protein, especially it's shorter fragments, was already reported to bind to pRb simultaneously with E7 protein [72], and the effect of E7 manifests in deterioration of E2F binding strength, not in suppression of the binding. Our observation is consistent with this result. This experiment would be even more interesting if the full length E2F protein is used, however preparations of that were unsuccessful so far.

4.3 Conclusions

Our study on pRb and it's interaction supplement and add on the results published by other groups. Use of NMR gave the opportunity to monitor the binding directly in in vitro experiments at the conditions much more similar to natural than crystallography. The conclusion from our binding studies on pRb can be that the simple LXCXE motif may not be sufficient to effectively influence the pRb function because of several reasons:

- As it was already suggested the binding of E7 to B-domain will probably not result in E2F replacement, other mechanism must be present, most probably involving C-terminal domain of pRb. So far there is no evidence that LXCXE binding site is influencing the binding of pRb to E2F.

- The E7 protein is mostly unfolded as judged by a proton 1D spectrum and HSQC spectra. The LXCXE-sequence containing protein may be folded in a way that no interaction of the motif will be possible.

- Our results suggests that not only LXCXE motif is important, but also the flanking residues are critical for binding strength. If the E7 action mode is to weaken the E2F binding, different binding strength of other LXCXE-carrying proteins may render them irrelevant for E2F-pRb interaction. This conclusion is in agreement with already reported fact

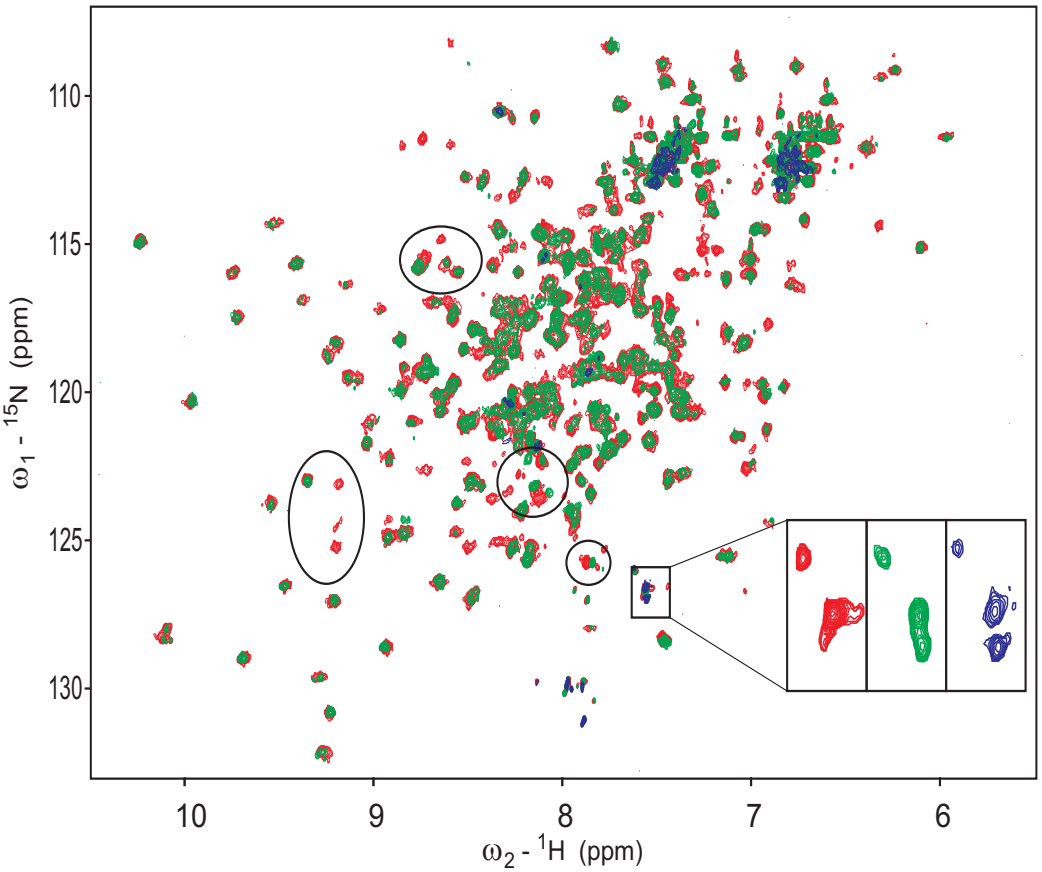

Figure 4.10: The experiment of E2F and E7 binding to A and B domains of pRb. A reference spectrum (red) after addition of E2F protein displays differences indicated on the spectrum (green). Subsequent addition of E7 protein leads to formation of big molecular weight complexes, resulting in dramatic increase in relaxation rate. Only flexible residues (especially sidechains) are recordable. The insert shows a magnification of one K^{E2F}.

that E7 protein binds more than an order of magnitude tighter than E7 peptides [72].

4.4 Materials and methods

The expression and purification of pRb ware done as in [37]. For perdeuterated sample the medium was made with 99% D_2O instead of H_2O, and the culture was gradually accommodated to new medium.

The E7 protein was purified using Ni-NTA chromatography under native condition followed by S75 gel filtration chromatography. Only the peak corresponding to E7 dimer was collected for experiments.

The E2F protein (residues 243-437) was over-expressed as a GST-fusion protein at 37 °C in *E. coli* BL21 (DE3) using a pGEX vector. The protein was purified using a GST Sepharose Fast Flow column (Amersham). The protein was eluted from the column after the digestion of the GST tag.

NMR samples contained usually 0.5 mM pRb protein in 50 mM $NaH_2PO_4 \cdot H_2O$, 10 mM β-mercaptoethanol, pH 7.3 and 5-10% D_2O. For titration experiments with peptides ^{15}N-Lys labeled samples were used. The peptides were dissolved in the same buffer up to 5 mM. The pH of the peptide solution was checked and set to 7.3 prior to the titration.

NMR measurements were carried on AV900MHz with Cryo-probe(R) or DRX600 with cryo- or standard triple resonance probe-heads or DMX750MHz with standard TXI probe-head. 128 increments for uniformly ^{15}N labeled or 72 for selective Lys labeled samples were recorded in indirect dimension. The sweep-width and carrier frequency for ^{15}N-Lys samples were adjusted to obtained the same maximum evolution time in ^{15}N direction.

Aminoacid sequences

Peptide sequences:

```
E7 peptide:                              DLYCYEQLN
E7 D1R mutant peptide:                   RLYCYEQLN
E2F short peptide:                       GLEEGEGIR
E2F large peptide:                LDYHFGLEEGEGIRDLFD
HDAC1 peptide (17-mer):              DKRIACEEEFSDSEEEG
HDAC1 short peptide:                     RIACEEEFS
HDAC1 short F8L mutant peptide:          RIACEEELS
HDAC1 short R1D mutant peptide:          DIACEEEFS
HDAC1 short R1D,E6Y mutant peptide:      DIACYEEFS
```

ABC sequence:

```
     (MHHHHHHHIEGR)-MN TIQQLMMILN SASDQPSENL ISYFNNCTVN PKESILKRVK 420
  DIGYIFKEKF AKAVGQGCVE IGSQRYKLGV RLYYRVMESM LKSEEERLSI QNFSKLLNDN 480
  IFHMSLLACA LEVVMATYSR STSQNLDSGT DLSFPWILNV LNLKAFDFYK VIESFIKAEG 540
  NLTREMIKHL ERCEHRIMES LAWLSDSPLF DLIKQSKDRE GPTDHLESAC PLNLPLQNNH 600
  TAADMYLSPV RSPKKKGSTT RVNSTANAET QATSAFQTQK PLKSTSLSLF YKKVYRLAYL 660
  RLNTLCERLL SEHPELEHII WTLFQHTLQN EYELMRDRHL DQIMMCSMYG ICKVKNIDLK 720
  FKIIVTAYKD LPHAVQETFK RVLIKEEEYD SIIVFYNSVF MQRLKTNILQ YASTRPPTLS 780
  PIPHIPRSPY KFPSSPLRIP GGNIYISPLK SPYKISEGLP TPTKMTPRSR ILVSIGESFG 640
  TSEKFQKINQ MVCNSDRVLK RSAEGSNPPK PLKKLRFDIE GSDEADGSKH LPGESKFQQK 900
  LAEMTSTRTR MQKQKMNDSM DTSNKEEK-(Stop)                            928
```

ABC-C sequence

```
     (MHHHHHHHIEGR)-MN TIQQLMMILN SASDQPSENL ISYFNNCTVN PKESILKRVK 420
  DIGYIFKEKF AKAVGQGCVE IGSQRYKLGV RLYYRVMESM LKSEEERLSI QNFSKLLNDN 480
  IFHMSLLACA LEVVMATYSR STSQNLDSGT DLSFPWILNV LNLKAFDFYK VIESFIKAEG 540
  NLTREMIKHL ERCEHRIMES LAWLSDSPLF DLIKQSKDRE GPTDHLESAC PLNLPLQNNH 600
```

```
TAADMYLSPV RSPKKKGSTT RVNSTANAET QATSAFQTQK PLKSTSLSLF YKKVYRLAYL 660
RLNTLCERLL SEHPELEHII WTLFQHTLQN EYELMRDRHL DQIMMCSMYG ICKVKNIDLK 720
FKIIVTAYKD LPHAVQETFK RVLIKEEEYD SIIVFYNSVF MQRLKTNILQ YASTRPPTLS 780
PIPHIPRSPY K-(Stop)                                              791
```

ABC-C-loop (pRb-AB) sequence

```
     (MHHHHHHIEGR)-MN TIQQLMMILN SASDQPSENL ISYFNNCTVN PKESILKRVK 420
DIGYIFKEKF AKAVGQGCVE IGSQRYKLGV RLYYRVMESM LKSEEERLSI QNFSKLLNDN 480
IFHMSLLACA LEVVMATYSR STSQNLDSGT DLSFPWILNV LNLKAFDFYK VIESFIKAEG 540
NLTREMIKHL ERCEHRIMES LAWLSDSPLF DLIKQSKD
                                          LKSTSLSLF YKKVYRLAYL 660
RLNTLCERLL SEHPELEHII WTLFQHTLQN EYELMRDRHL DQIMMCSMYG ICKVKNIDLK 720
FKIIVTAYKD LPHAVQETFK RVLIKEEEYD SIIVFYNSVF MQRLKTNILQ YASTRPPTLS 780
PIPHIPRSPY K-(Stop)                                              791
```

E2F-1 (243-437)

```
  LAYVTCQD LRSIADPAEQ MVMVIKAPPE TQLQAVDSSE NFQISLKSKQ GPIDVFLCPE 300
ETVGGISPGK TPSQEVTSEE ENRATDSATI VSPPPSSPPS SLTTDPSQSL LSLEQEPLLS 360
RMGSLRAPVD EDRLSPLVAA DSLLEHVRED FSGLLPEEFI SLSPPHEALD YHFGLEEGEG 420
IRDLFDCDFG DLTPLDF                                               437
```

Chapter 5

Structure and dynamics of staphostatin A from *Staphylococcus aureus*

5.1 Biological context

Staphylococcus aureus is responsible for a wide variety of human diseases and its ability to infect almost every tissue, together with increasing antibiotic resistance, renders it very dangerous. In response to these problems, scientists have looked for novel strategies of treatment. Although the infection process still hides a lot of secrets, current knowledge already suggests some interesting targets.

Two cysteine proteases of *S. aureus*, staphopain A and B, show evidence that they are important virulence factors in this bactrium. Two specific inhibitors, staphostatin A and SspC, were an interesting discovery. Staphostatin A was found to be active only against staphopain A, and analogously SspC exclusively targets staphopain B.

Amino acid sequences of staphostatin A and B are not similar to any known protease inhibitors. The crystal structure of staphostatin B is not related to cystatins, a superclass of cysteine peptidase inhibitors with a unique inhibition mode.

5.2 NMR measurements

A 1D proton NMR spectrum

A proton 1D NMR spectrum of staphostatin A indicated that the protein was well folded, highly concentrated, and peaks were well dispersed, suggesting mostly the β secondary structures (Fig. 5.1). The sample was prepared after lyophilization, thus proton-deuterium exchange experiments were feasible.

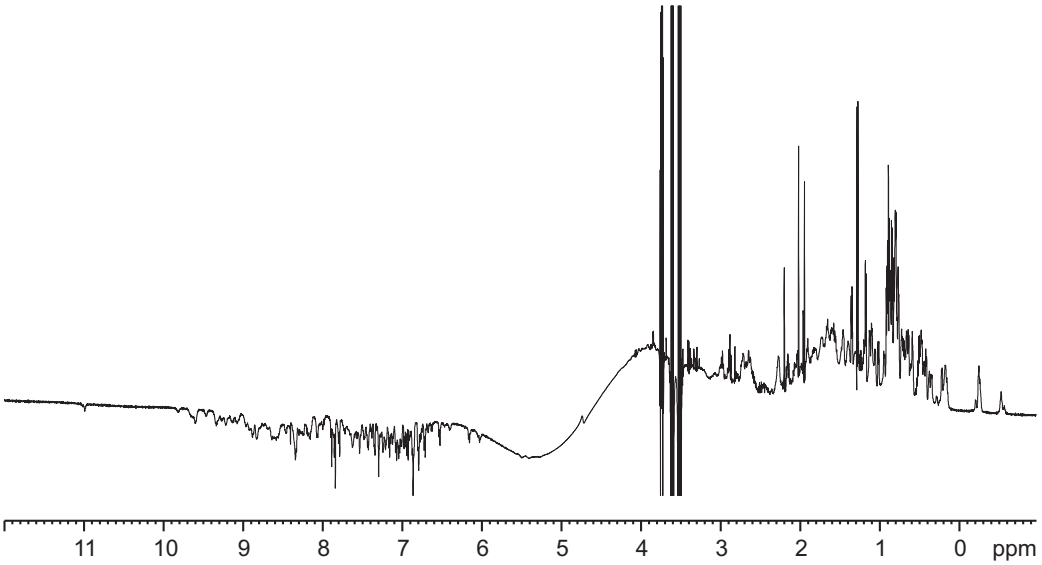

Figure 5.1: The first proton 1D NMR spectrum of staphostatin A at 300K.

A detailed inspection of the Trp side-chain at 11 ppm suggested two forms of the protein (Fig. 5.2a). Also, methyl signals displayed inhomogeneity. This inhomogeneity turned out to be the different oxidation states of Cys16 and Cys53. At this stage also a refolding protocol was established to produce homogeneous reduced or oxidized forms (Fig. 5.2b,c). The sample of reduced Cys residues was slowly oxidizing, and staphostatin A is and intracellular protein, thus oxidized form was used for further experiments.

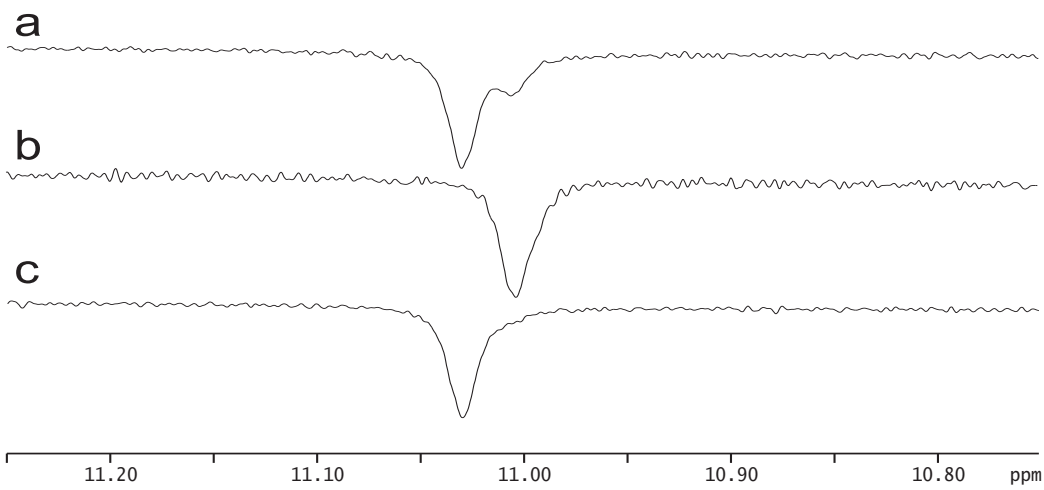

Figure 5.2: 1D NMR spectrum of staphostatin A, the tryptophan sidechain region. (a) the initial preparation, (b) a fully reduced form, (c) the oxidized form.

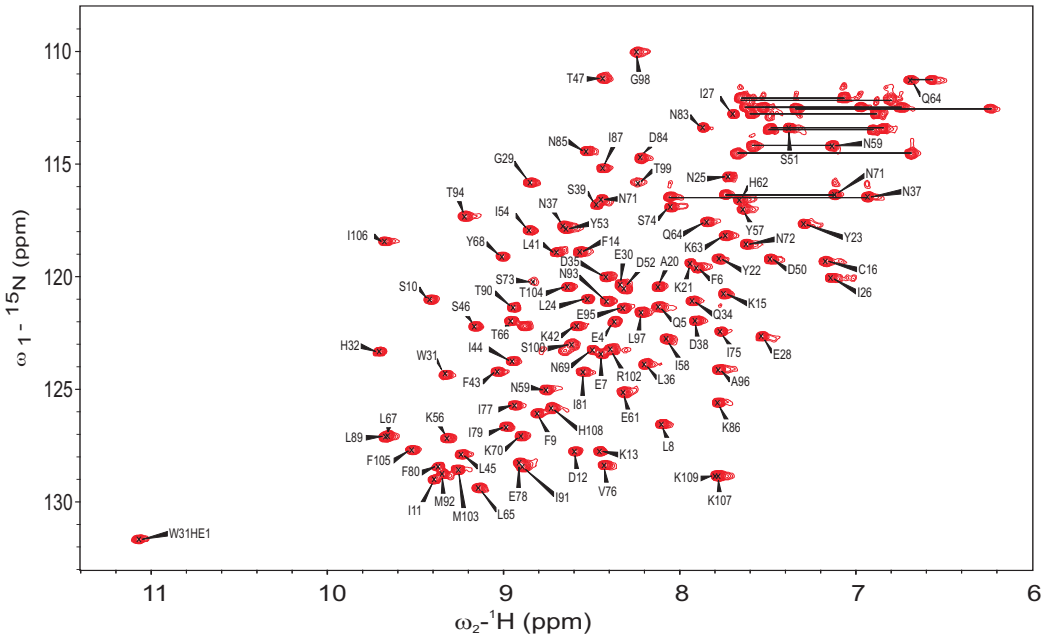

Figure 5.3: The ^{1}H-^{15}N HSQC spectrum of staphostatin A.

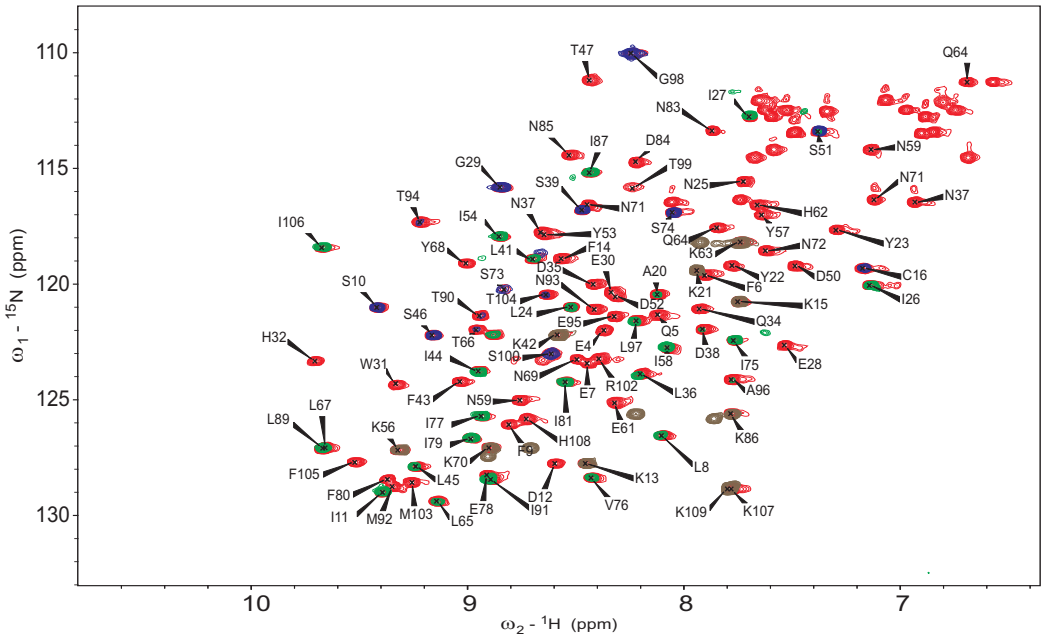

Figure 5.4: The superposition of ^{1}H-^{15}N HSQC spectrum of uniformly (red) and selective ^{15}N-labeled: Lys (brown), Gly/Ser (blue), Leu/Iso (green) samples of staphostatin A.

2D ^{1}H-^{15}N HSQC spectra

The HSQC spectrum of a uniformly ^{15}N-labeled sample shows good spectral dispersion of the signals (Fig. 5.3). A subspectrum can be obtained with the use of selectively ^{15}N labeled amino-acids. Fig. 5.4.

2D NOESY spectra

2D NOESY spectra of staphostatin A in water were of good quality. Additionally, a 2D NOESY spectrum in 100% D_2O was recorded. The 1D spectrum of staphostatin A in D_2O is shown in the Fig. 5.5.

3D spectra

The HNCA and CBCA(CO)NH spectra were recorded on the ^{15}N, ^{13}C-labeled sample. Together with the ^{15}N selectively labeled samples they were sufficient for backbone assignment.

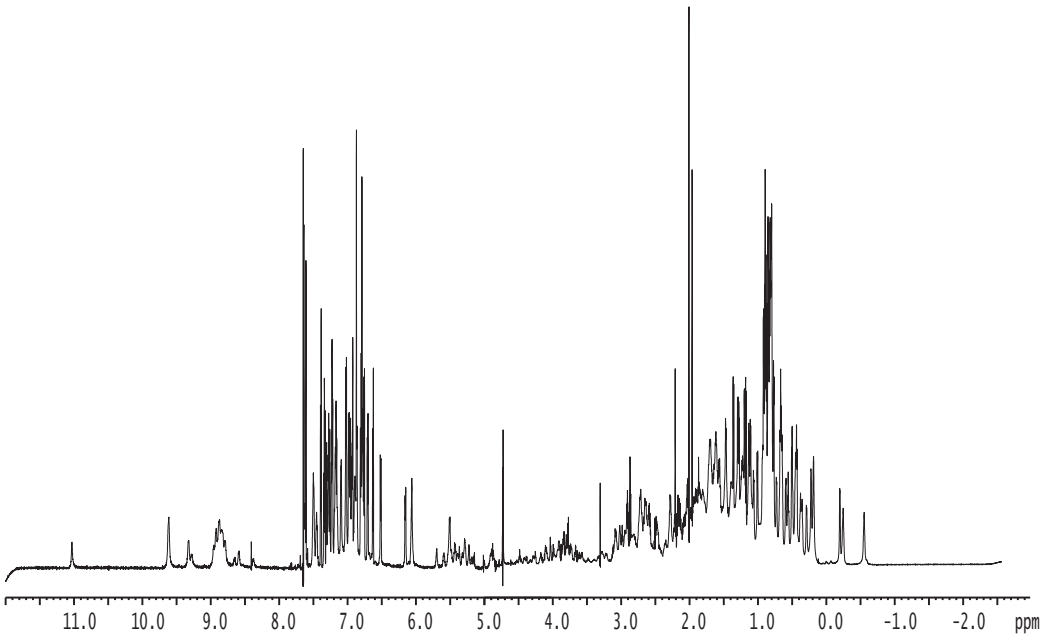

Figure 5.5: 1D spectrum of staphostatin A in D_2O.

NOESY-HSQC and TOCSY-HSQC spectra were recorded on a uniformly ^{15}N-labeled sample and used to verify the backbone assignments and to assign side-chain protons. 2D NOESY and NOESY-HSQC spectra also supplied information about inter-proton distances used later in the structure calculation process.

For determination of $^3J_{H_N H_A}$ [75] an HNHA experiment was recorded.

5.2.1 The structure of the staphostatin A

Structure calculation

The structure was calculated with the use of

- 546 approximate distance restrains derived from NOESY spectra.
- 82 torsion angle restrains from $^3J_{H_N H_\alpha}$.
- 1 disulphide bridge between Cys16 and Cys53.

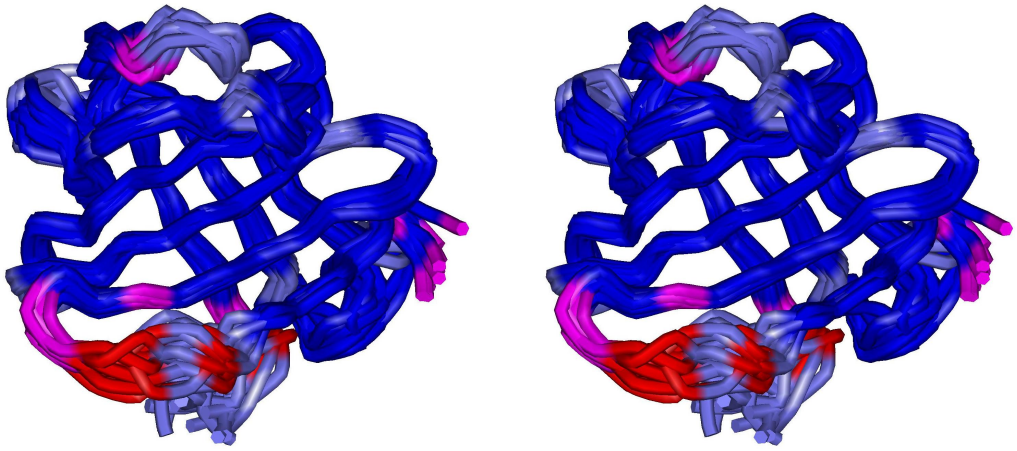

Figure 5.6: The staphostatin A structure. A bundle of 20 representative structures is shown with ^{1}H-^{15}N heteronuclear-NOEs indicated by colors: blue—value above 0.6, rigid structure; magenta—value between 0.4 and 0.6; red—value below 0.4 indicating flexibility; gray—no reliable data available.

with the simulated annealing protocol, without starting distance geometry step. Using the total energy criterion, from a set of several hundreds structures, a family of 20 best was selected [76]. All of these 20 structures fully satisfy experimental constraints, thus we could conclude that the final model well describes the protein structure in solution.

Description of the structure

The overall fold of staphostatin A is identical to that of staphostatin B unequivocally classifying both proteins to the same family. The staphostatin A fold can be best described as a slightly deformed, eight-stranded β-barrel. The barrel-closing between C- and N-terminal parts consists of a disulphide bridge and hydrogen bonds.

5.2.2 NMR relaxation measurements

Heteronuclear-NOE measurements

The ^{1}H-^{15}N heteronuclear-NOE results correspond to a rigid molecule, since most residues give the value above 0.6, the theoretical limit being around 0.85 for 500 MHz. Values lower than 0.4 are observed only at the termini of the protein and for a region of residues 97-100, corresponding to the active loop of the protein. This correlates with the loop being less well defined in the 20-structure bundle.

T_1, longitudinal relaxation time measurements

The T_1 data confirm the characteristics of motions in staphostatin A inferred from the heteronuclear-NOE measurements (data not shown).

T_2, the transverse relaxation time measurements

The T_2 relaxation measurements recorded resulted in a low signal to noise spectrum, too low for proper analysis of the data. Since the detailed dynamical study was not intended and the backbone dynamics seemed to be well explained on the basis of heteronuclear-NOE and T_1, further measurements were not carried out.

5.3 Discussion

Out of the total of 107 residues (103 without prolines) of staphostatin A, 95 residues were assigned. From the HNHA correlation experiment $^{3}J_{H_N H\alpha}$ coupling constants were measured. The assignment has been published [77], and the chemical shifts and $^{3}J_{H_N H\alpha}$ coupling constants where deposited in BioMagResBank under accession number 5810 (http://www.bmrb,wisc.edu).

Despite low sequence similarity, the NMR structure of staphostatin A is similar to the staphostatin B structure, thus classifying both proteins into the

same and a new class of protease inhibitors. Staphostatin A differs little from staphostatin B, mainly by the presence of a disulphide bridge. The backbone dynamics measured by NMR shows a flexible loop between residues L97 and S100. This flexibility well correlates with the inhibition mechanism of the protein.

Characterization of staphostatin in terms of its structure and inhibition mechanism show distinct characteristics from those of other known papain-like cysteine proteases inhibitors. Whether our results have direct therapeutic implications is not clear, due to limited knowledge of the *S. aureus* infection mechanism.

5.4 Materials and methods

Sample preparation

The NMR samples contained ca. 0.7 mM staphostatin A in 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, 0.05% NaN_3, pH 7.4, and 10% D_2O or 100% D_2O. In order to obtain the staphostatin A sample in 100% D_2O a ^{15}N labeled sample was liophilized and dissolved in D_2O, then the procedure was repeated again, yielding a deuterated sample with only traces of water.

NMR spectroscopy

NMR experiments were performed on Bruker AMX500, DRX600 and DMX750 spectrometers at 300 K. ^{1}H,^{15}N,$^{13}C^{\alpha}$,$^{13}C^{\beta}$ resonance assignments were carried out from 2D NOESY, ^{1}H-^{15}N HSQC (of the uniformly or selectively ^{15}N Tyr, Lys, Val, Ile, Gly/Ser, Leu, Phe labeled proteins), HNCA, CBCA(CO)NH and ^{15}N-edited 3D-NOESY and TOCSY [78]. The NMR relaxation measurements were carried on a Bruker AMX500 spectrometer, equipped with a triple resonance probehead. A modified ^{1}H-^{15}N heteronuclear-NOE experiment [79] was repeated three times to minimize the errors. To measure the

T_1 values, six ^{1}H-^{15}N planes were recorded in an interleaved manner, with delays of 12.4, 384.4, 756.4, 1128.4, 1500.4 and 12.4 ms, and peak heights were taken for analysis. The spectra were processed without linear prediction to avoid intensity distortions.

Summary

The work presented in this thesis was carried out in the Department of Structural Research at the Max Planck Institute for Biochemistry, Martinsried, from July 2002 and December 2005. The first part of this thesis involved the understanding of effects of inhibitors on protein-protein interactions. Specifically, we studied the interactions in cellular proteins related to cancer or anticancer therapy.

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 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 protein complexes. In order to study the effect of anatagonists on a protein complex, we devised a new methodology based on a dramatic difference in 1/T2 relaxation rates of the protein upon forming a large molecular weight complex. The size of one protein 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. This method should provide an important extension to the traditional "SAR by NMR" technique. The methodology was illustrated using the p53-MDM2 interaction.

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 G1 cell cycle arrest and apoptosis. p53 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. We have tested several small molecular weight compounds that were reported to inhibit this interaction. Only two of them from the totla 20 tested were found to be potent inhibitors of the p53-MDM2 binding.

The interactions of cyclin-dependent kinase 2 (CDK2) with p27 and CDK2-cyclin A2 with p27 have been tested using in vitro NMR experiments. The purine type of inhibitors of CDK2 were also studied with NMR spectroscopy. New insights were gained from these experiments into the involvement of CDKs and their cellular partners in controlling the cell cycle.

The second major project involved functional investigations on the small pocket of the retinoblastoma protein (pRb) and its interaction with binding partners. The retinoblastoma tumor suppressor protein (pRb) is a key negative regulator of cell proliferation that is frequently deregulated in human cancer. More than 130 proteins have been reported to bind to pRb directly. Many of these interactions were reported with the pRb small pocket, which is the major focus of tumorigenic mutations in pRb. However, results found in the literature were ambiguous and contradictory. Many viral oncoproteins (for example, HPV E7, E1A) are known to bind to the pRb pocket domain via a LXCXE binding motif. There are also some 20 cellular proteins that contain a LXCXE motif and have been reported to associate with the pocket domain of pRb. Using the in-vitro methods including NMR spectroscopy, I have shown that LXCXE peptides of viral oncoproteins bind strongly to the pocket domain of pRb. Additionally, I showed that the LXCXE-like peptides of histone deacetylase 1 (HDAC1) binds to the same site on pRb, with however a weak (micromolar) and transient association. Systematic substitution of residues other than conserved L, C, and E show that the residues flanking the LXCXE are important for the binding. The interactions of pRb with the following proteins were also characterized using this approach: E2F1, HDAC1, HPV E7.

In a structurally oriented project in this thesis we characterized the structure of staphostatin A from Staphylococcus aureus. Staphostatin A belongs to a new class of cysteine protease inhibitors.

Zusamenfassung

Die vorliegende Arbeit wurde in der Abteilung Strukturbiologie am Max-Planck-Institut für Biochemie in Martinsried zwischen Juli 2002 und Dezember 2005 angefertigt. Der erste Teil der Arbeit beinhaltet die Beschreibung der Inhibitoren von Protein-Protein-Wechselwirkungen. Im Speziellen wurden die Wechselwirkungen von zellulären Proteinen untersucht, die beim Krebs oder bei der Krebs Therapie eine Rolle spielen.

Die NMR Methode der induzierten Änderung der chemischen 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 mittelgroβen Domänen von Zielproteinen. Proteinkomplexe wurden nicht untersucht. Um die Effekte von Antagonisten auf Proteinkomplexe zu testen, haben wir eine Methode entwickelt, die auf dem deutlichen Unterschied der $1/T_2$ Relaxationszeiten des Proteins, bei der Bildung von hochmolekularen Komplexen beruht. Die Gröβe der ersten Proteinkomponente sollte klein genug (kleiner 15 kDa) sein, um ein gutes HSQC-Spektrum nach ^{15}N oder ^{13}C Markierung der Proteine zu liefern. Die zweite Komponente sollte groβ genug sein, damit der sich bildende Komplex eine Molekulargewicht von mehr als 40 kDa besitzt. Die Methode sollte eine wichtige Ergänzung zu der traditionellen "SAR by NMR" Technik darstellen. 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 G1 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 Zerstörung 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. Wir haben eine Reihe von niedermolekularen Substanzen getestet, von denen eine Inhibierung der MDM2 p53 Wechselwirkung berichtet wurde. Nur zwei von zwanzig getesteten Substanzen konnten als potentielle Inhibitoren verifiziert werden.

Die Interaktion der Cyclin-abhängigen Kinase 2 (CDK2) mit p27 und CDK2/Cyclin A2 mit p27 wurden in vitro mit NMR untersucht. Ebenso wurde Purine-Typ Inhibitoren für CDK2 mit NMR untersucht. Neue Erkenntnissen wurden aus diesen Experimenten gewonnen über die Rolle von CDK2 und dessen Bindungspartnern bei der Kontrolle des Zellzyklus.

Der zweite Teil befasst sich mit der Untersuchung der kleinen Taschen Domäne des Retinoblastoma Proteins und dessen Bindungspartnern. Das Retinoblastoma Tumor Suppressor Protein (pRb) ist ein negativer Schlüsselregulator der Zellproliferation, der häufig in menschlichen Tumoren dereguliert ist. Von mehr als 130 Proteine wurde eine direkte Bindung an pRb nachgesagt. Viele dieser Wechselwirkungen sind mit der kleinen Pocket Domäne von pRb umfasst, in der bevorzugt Mutationen auftreten, die bei der Krebsentstehung beobachtet wurden. Viele virale Onkoproteine (wie z. B.: HPV E7, E1A) binden an die kleine Pocket Domäne von pRb über ihr LXCXE Bindungsmotiv. Es wurde ebenfalls von ca. 20 zelluläre Proteine berichtet, die ein LXCXE Motiv haben, dass sie mit der kleinen Pocket Domäne von pRB assoziiert sind. Mit Hilfe der NMR Spektroskopie konnten wir zeigen, dass die LXCXE Peptide von viralen Onkoproteinen stark an die kleine Pocket Domäne von pRb binden. Zusätzlich konnten wir zeigen, dass das LXCXE ähnliche Peptid der Histon Deacetylase 1 (HDAC1) an die gleiche Bindungsstelle an pRb bindet, jedoch mit einer schwächeren (im

micromolaren Bereich) und transienten Bindung. Systematischer Ersatz von Aminosäureresten auβer den konservierten Resten L, C und E zeigte, dass die flankierende Reste des LXCXE Motivs die Bindung an pRb wichtig für sind. Die Wechselwirkung von pRb mit den folgenden Proteinen wurde ebenfalls untersucht: E2F1, HDAC1, HPV E7.

In einem mehr strukturell orientierten Projekt dieser Doktorarbeit haben wir die Struktur von Staphostatin A von *Staphylococcus aureus* untersucht. Staphostatin A gehört zu einer neuen Klasse von Cystein Protease Inhibitoren.

Appendix A

NMR resonance assignment of RITA [2,5-bis(5-hydroxymethyl-2-thienyl)furan]

The compound is at least 98% pure in terms of the proton containing impurities. The resonances have been assigned based on the 2D ^{1}H COSY Fig. A.3 and water exchange experiments of Fig. A.4.

The assignment of three signals from quaternary carbons (shown separately in Fig. A.7 can also be assigned by analysis of multiplet changes in these experiments (data not shown). The assignment of the OH proton resonance was also confirmed since no change to the carbon spectrum was observed upon decoupling of the OH proton signal.

Conclusion: Since natural abundance of ^{13}C is 1% and no signals other than those from RITA, H_2O, and DMSO are visible in the spectrum, concentrations of impurities must be below 1%, if any at all.

Conclusions: The conclusion from our physicochemical analysis of RITA used in our NMR experiments on the p53/MDM2 binding is that the compound NSC652287 received from the National Cancer Institute corresponds to the formula of RITA [2,5-bis(5-hydroxymethyl-2-thienyl)furan] and is pure to at least 98%. We have also measured NMR spectra of another batch of

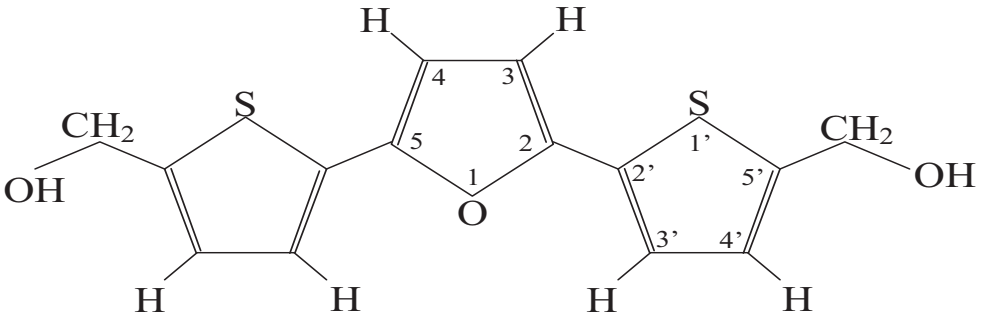

Figure A.1: Chemical formula of RITA. The molecule in solution at room temperature is symmetrical with the symmetry axis going through the central oxygen.

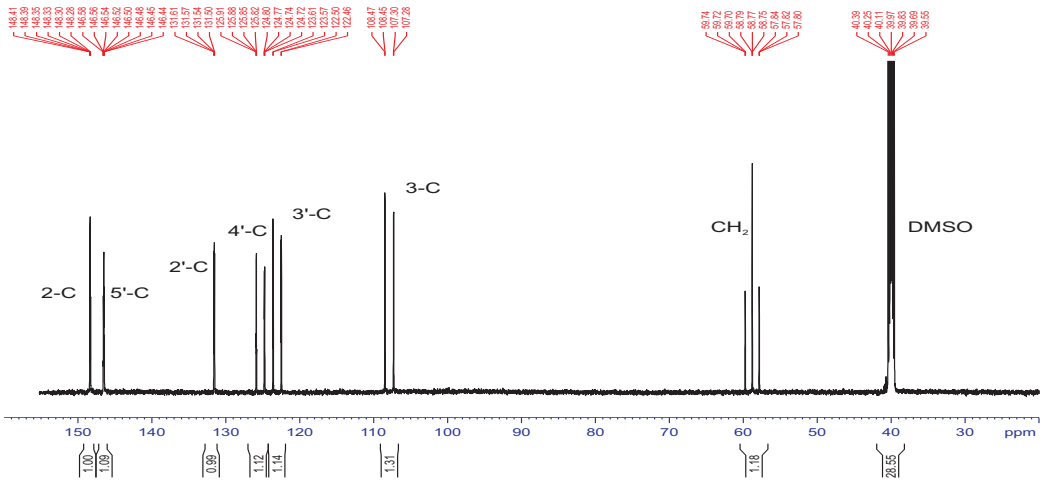

Figure A.2: The proton spectrum of RITA in DMSO. The ^{1}H 1D NMR spectrum of RITA in 100% DMSO-d6 recorded on the DRX-600 MHz spectrometer at 298 K. The asterisks denote an example of ^{13}C satellites from the ^{1}H-^{13}C coupling.

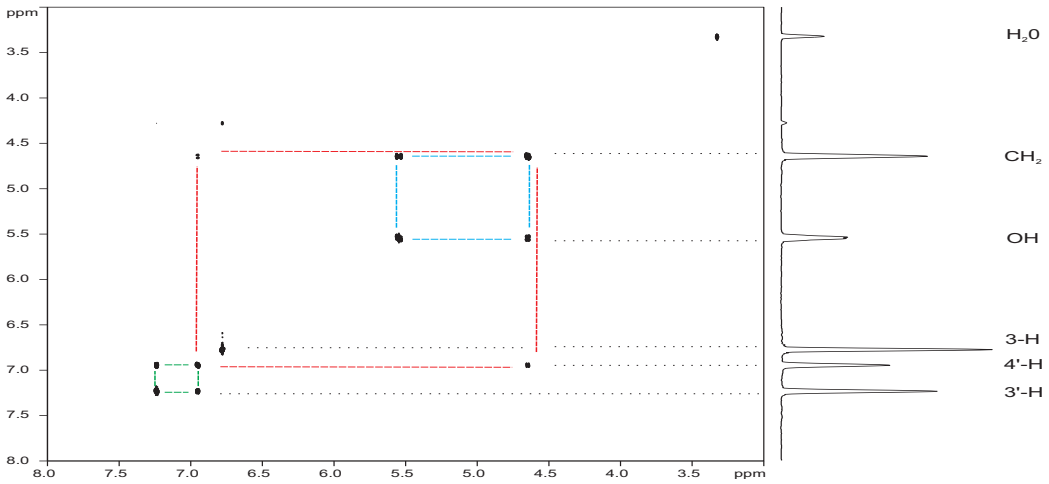

Figure A.3: A COSY spectrum of RITA in DMSO. On the basis of the COSY spectrum (AMX 400MHz, 300K) RITA resonances can be easily assigned: DMSO at 2.5 ppm, H_2O at 3.3 ppm, CH_2 at 4.65 ppm (connectivity to OH, blue lines), OH at 5.52 ppm (connectivity to CH_2, blue), 3-H at 6.8 ppm (no cross peaks present), 4'-H at 6.95 ppm (cross peaks to 3-H, green, and to CH_2, red), 3'-H at 7.25 ppm (connected to 4-H, green).

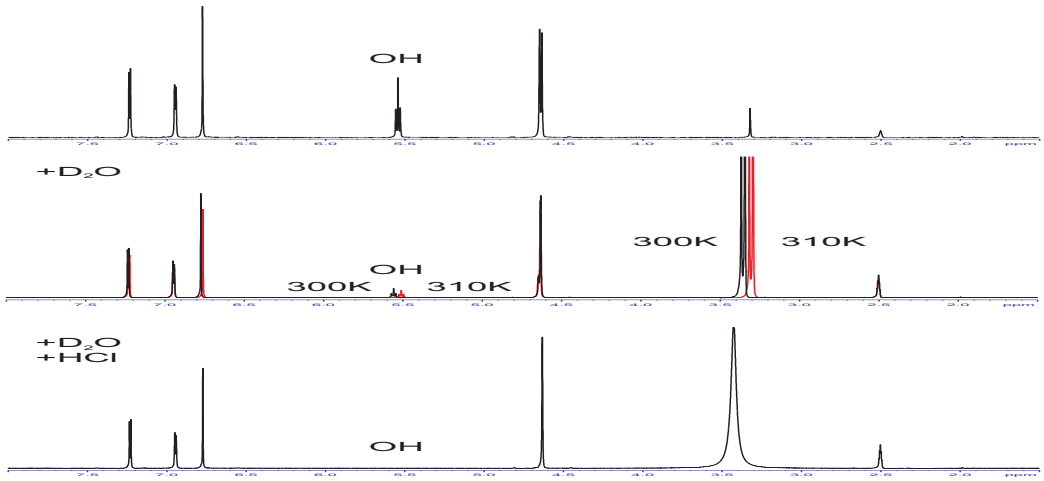

Figure A.4: Verification of the OH assignment. Since the OH resonance should be sensitive to the chemical exchange with water, we have carried out the following experiments: (a) 5 μl (1% vol.) of D_2O was added to the sample of RITA in DMSO (100%), T = 300 K. The OH triplet is reduced, CH_2 is an unsymmetric doublet. (b) The temperature was changed to 310 K. A shift of H_2O and OH signals was observed. (c) T= 300 K, added 1 μl (0.2% vol.) of 3.7% HCl. The OH signal dissapeared, CH_2 is a singlet.

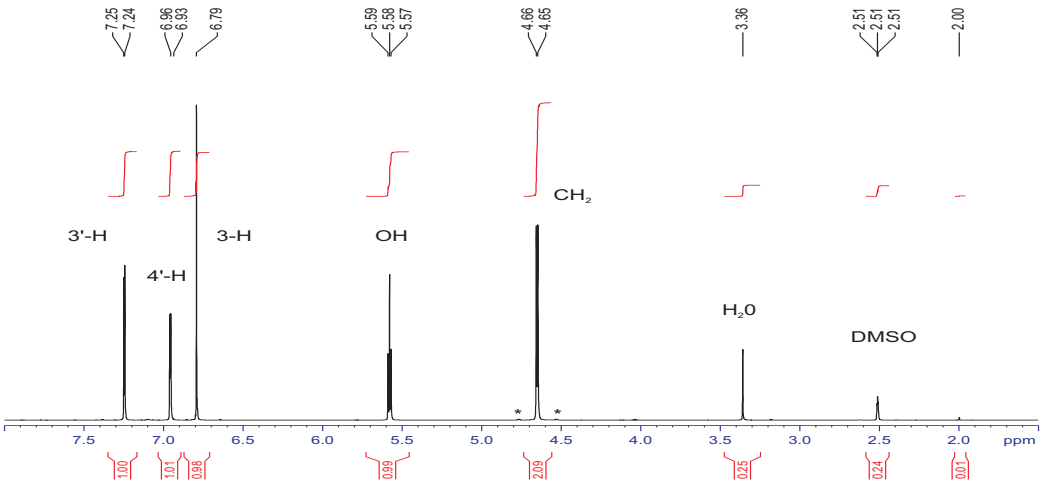

Figure A.5: A natural abundance ^{13}C spectrum. A ^{13}C 1D NMR spectrum without decoupling and without NOESY enchancement was recorded on the DRX 600 MHz spectrometer. Nine different carbons (apart from the DMSO peak at 40 ppm) are seen. These are: CH_2 at 58.7 ppm, 3-C at 108 ppm, 3'-C at 123 ppm, 4'-C at 125.8 ppm, 2'-C at 131.5 ppm, 5' at 146.5 ppm, 2 at 148.3 ppm. The assignment was accomplished by selective decoupling at certain proton frequencies (Fig. A.6).

RITA obtained from the National Cancer Institute about 3 months after the first shipment. This most recent RITA gave identical NMR spectra to those of the first sample.

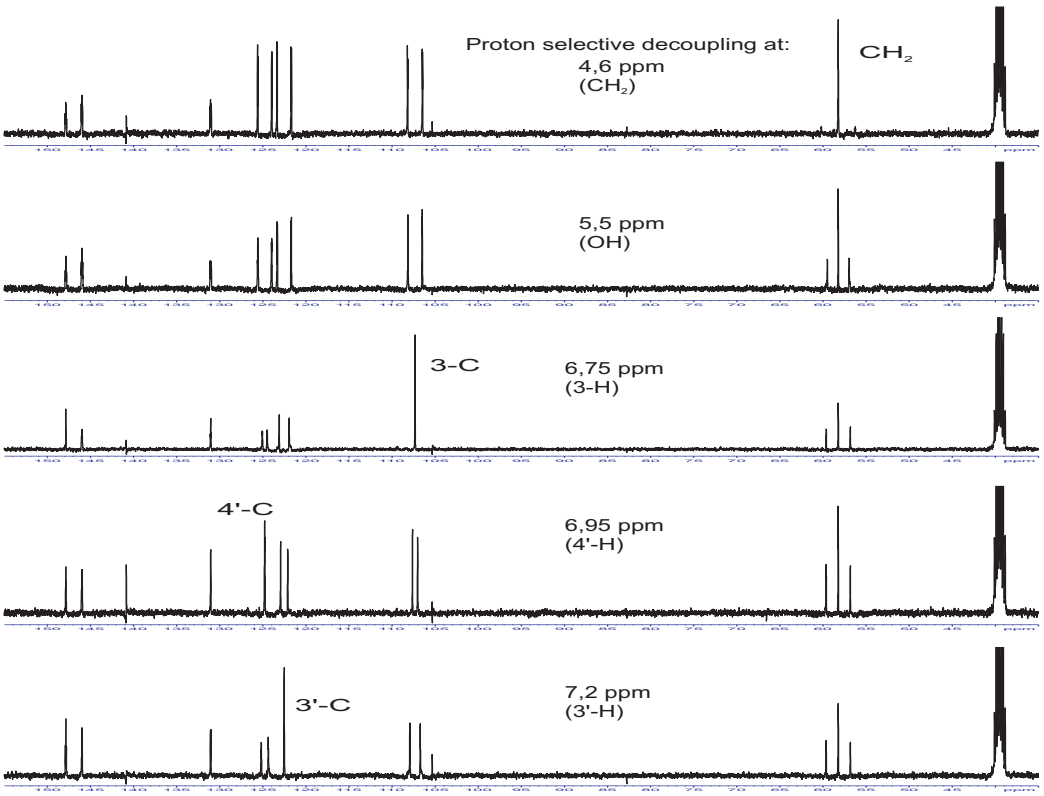

Figure A.6: Selective proton-decoupling experiments. ^{13}C spectra were recorded on the AMX 400 MHz at 300 K, with the broadband proton presaturation prior to acquisition (NOESY enhancement) and selective CW irradiation at a single proton frequency indicated in ppm above the ^{13}C spectrum.

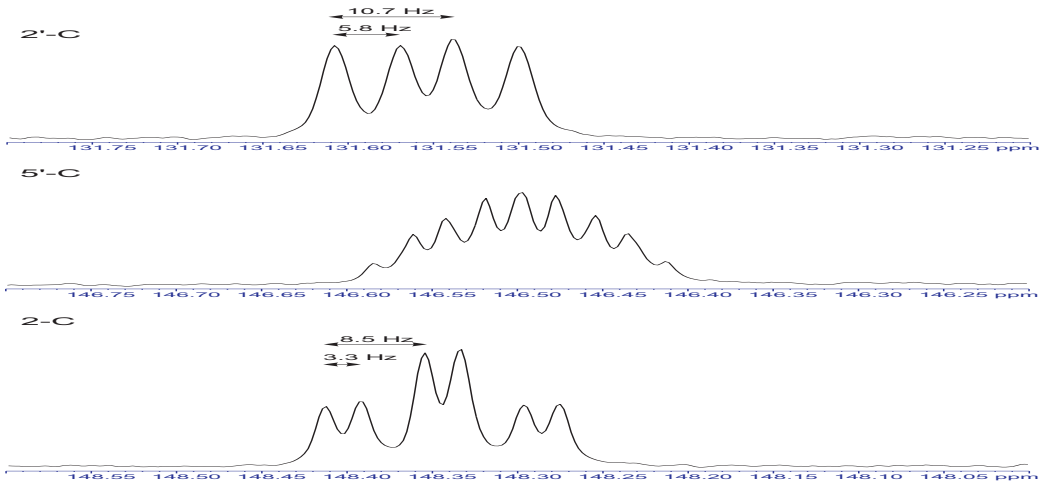

Figure A.7: Multiplet structures of quaternary carbons in RITA.

Bibliography

[1] C. P. Jaroniec, C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson, and R. G. Griffin. High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 101:711–717, 2004.

[2] C. M. Rienstra, L. Tucker-Kellogg, C. P. Jaroniec, M. Hohwy, B. Reif, M. T. McMahon, B. Tidor, T. Lozano-Perez, and R. G. Griffin. De novo determination of peptide structure with solid-state magic-angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 99:10260–10265, 2002.

[3] F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbeln, and H. Oschkinat. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature*, 420:98–102, 2002.

[4] S. G. Zech, E. Olejniczak, P. Hajduk, J. Mack, and A. E. McDermott. Characterisation of protein–ligand interactions by high-resolution solid-state NMR spectroscopy. *J. Am. Chem. Soc.*, 126:13948–13953, 2004.

[5] A. Pardi, M. Billeter, and K. Wüthrich. Calibration of the angular dependence of the amide proton-C^{α} proton coupling constants, ${}^{3}J_{HN\alpha}$, in a globular protein. *J. Mol. Biol.*, 180:741–751, 1984.

[6] M. Goldman. Interference effect in the relaxation of a pair of unlike spin-$\frac{1}{2}$ nuclei. *J. Mag. Reson.*, 60:437–452, 1984.

[7] M. Salzmann, G. Wider, K. Pervushin, and K. Wüthrich. Improved sensitivity and coherence selection for [^{15}N,^{1}H]-TROSY elements in triple resonance experiments. *J. Biomol. NMR*, 15:181–184, 1999.

[8] Roland Riek, Jocelyne Fiaux, Eric B. Bertelsen, Arthur L. Hortwich, and Kurt Wüthruch. Solution NMR techniques for large molecular and supramolecular structures. *J. Am. Chem. Soc.*, 124:12144–12153, May 2002.

[9] M. Liu, X. Mao, C. Ye, H. Huang, J. K. Nicholson, and J. C. Lindon. Improved WATERGATE pulse sequences for solvent suppresion in NMR spectroscopy. *J. Mag. Reson.*, 132:125–129, 1998.

[10] S. Mori, C. Abeygunawardana, M. O. Johnson, and P. C.M. van Zijl. 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. Reson. B*, 108:94–98, 1995.

[11] D. Nietlispach, R. T. Clowes, R. W. Broadhurst, Y. Ito., J. Keeler, M. Kelly, J. Ashurst, H. Oschkinat, P. J. Domaille, and E. D. Laue. An approach to the structure determination of larger proteins using triple resonance NMR experiments in conjunction with random fractional deuteration. *J. Am. Chem. Soc.*, 118:407–415, 1996.

[12] S. Grzesiek, P. Wingfield, S. Stahl, J. D. Kaufman, and A. Bax. Four-dimensional ^{15}N-separated NOESY of slowly tumbling perdeuterated ^{15}N-enriched proteins. Application to HIV-1 Nef. *J. Am. Chem. Soc.*, 117:9594–9595, 1995.

[13] C. R. Babu, P. F. Flynn, and A. J. Wand. Validation of protein structure from preparation of encapsulated proteins in low viscosity fluids. *J. Am. Chem. Soc.*, 123:2691–2692, 2001.

[14] P. F. Flynn, M. J. Milton, C. R. Babu, and A. J. Wand. A simple and effective NMR cell for studies of encapsulated proteins dissolved in lov viscosity solvents. *J. Biomol. NMR.*, 23:311–316, 2002.

[15] S. Gaemers, C. J. Elsevier, and A. Bax. NMR of biomolecules in lov viscosity, liquid CO_2. *Chem. Phys. Lett.*, 301:138–144, February 1999.

[16] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, and N. J. Skelton. *Protein NMR Spectrosopy Principles And Practice.* Academic Press, 1996.

[17] L. Fielding. NMR methods for the determination of protein-ligand dissociation constants. *Current Topics in Medicinal Chemistry*, 3:39–53, 2003.

[18] P. W. Rabideau, editor. *The Conformational Analysis of Cyclohexanes, Cyclohexadienes and Related Hydroaromatic Compounds*, Chapter 1. VCH Publishers, New York, 1987.

[19] L. D'Silva, P. Ozdowy, M. Krajewski, U. Rothweiler, M. Singh, and T. A. Holak. Monitoring the effects of antagonists on protein–protein interactions with NMR spectroscopy. *J. Am. Chem. Soc.*, 127(38):13220–13226, September 2005.

[20] Z. Wang. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.*, 360:111–114, 1995.

[21] B. W. Sigurskjold. Exact analysis of competition ligand binding by displacement isothermal titration calorymetry. *Analytical Biochem.*, 277:260–266, 2000.

[22] W. Jahnke and H. Widmer. Protein NMR in biomedical research. *Cell. Mol. Life Sci.*, 61:580–599, 2004. Review.

[23] C.A. Lepre, J.M. Moore, and J.W. Peng. Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.*, 104:3641–3675‘, 2004.

[24] B.J. Stockman and C. Dalvit. NMR screening techniques in drug discovery and drug design. *Progr. NMR Spectrosc.*, 41:187–231, 2002. Review.

[25] M. Coles, M. Heller, and H. Kessler. NMR-based screening technologies. *DDT*, 8:803–809, 2003. review.

[26] J. H. Prestegard, H. Valafar, J. Glushka, and F. Tian. Nuclear magentic resonance in the era of structural genomics. *Biochemistry*, 40:8677–8685, 2001.

[27] P. J. Hajduk, E. T. Olejniczak, and S. W. Fesik. One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *J. Am. Chem. Soc.*, 119:12257–12261, 1997.

[28] W. Jahnke, P. Floerheim, C Ostermeier, Z Zhang, R. Hemmig, K. Hurth, D. P. Uzunov. NMR reporter screening for the detection of high-affinity ligands. *Angew. Chem. Int. Ed.*, 41:3420–3423, 2002.

[29] M. Mayer and B. Meyer. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed.*, 38:1784–1788, 1999.

[30] B. Meyer, T. Weimar, T. Peters. Screening mixtures for biological activity by NMR. *Eur. J. Biochem.*, 246:705–709, 1997.

[31] A. Dehner and H. Kessler. Diffusion NMR spectroscopy: Folding and aggregation of domains in p53. *ChemBioChem*, 6:1550–1565, 2005. Review.

[32] A. Dehner, J. Furrer, K. Richter, I. Schuster, J. Buchner, and H. Kessler. NMR chemical shift perturbation study of the N-terminal domain of Hsp90 upon binding fo ADP, AMP-PNP, geldanamycin, and radicicol. *ChemBioChem*, 4:870–877, 2003.

[33] S. B. Shuker, P. J. Hajduk, R. P. Meadows, and S. W. Fesik. Discovering high-affinity ligands for proteins: SAR by NMR. *Science*, 274:1531–1534, 1996.

[34] M. Pellecchia, D.S. Sem, and K. Wüthrich. NMR in drug discovery. *Nat. Rev. Drug Discov.*, 1:211–219, 2002.

[35] K. V. Pervushin, G. Wider, and K. Wüthrich. Single transition-to-single transition polarization transfer (ST2-PT) in [^{15}N, ^{1}H]-TROSY. *J. Biomol. NMR*, 12:345–348, 1998.

[36] G. Zhu, X. M. Kong, and K. H. Sze. Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *J. Biomol. NMR.*, 13:77–81, 1999.

[37] M. Singh, M. Krajewski, A. Mikolajka, and T. A. Holak. Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences. *J. Biol. Chem.*, 280(45):37868–37876, November 2005.

[38] M. L. Reese and V. J. Dötsch. Fast mapping of protein-protein interfaces by NMR spectroscopy. *J. Am. Chem. Soc.*, 125:14250–14251, 2003.

[39] M. Pellecchia, D. Meininger, A. L. Shen, R. Jack, C. B. Kasper, D. S. Sem. SEA-TROSY (solvent exposed amides with TROSY): A method to resolve the problem of spectral overlap in very large proteins. *J. Am. Chem. Soc.*, 123:4633–4634, 2001.

[40] P. J. Hajduk, D. J. Augeri, J. Mack, R. Mendoze, J. Yang, S. F. Betz, and S. W. Fesik. NMR-based screening of proteins containing ^{13}C-labelled methyl groups. *J. Am. Chem. Soc.*, 122:7898–7904, 2000.

[41] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, 274:948–953, November 1996.

[42] M. J. J. Blommers, G. Fendrich, C. Garcia-Echeverria, and P. Chene. On the interaction between p53 and mdm2: transfer NOE study of a p53 derived peptide ligated to MDM2. *J. Am. Chem. Soc.*, 119:3425–3426, 1997.

[43] R. Stoll, C. Renner, S. Hansen, S. Palme, C. Klein, A. Belling, W. Zeslawski, M. Kamionka, T. Rehm, P. Mühlhahn, R. Schumacher, F. Hesse, B. Kaluza, W. Voelter, R. A. Engh, and T. A. Holak. Chalcone derivatives antagonize interactions between the human oncoprotein MDM2 and p53. *Biochemistry*, 40:336–344, 2001.

[44] R. Dawson, L. Müller, A. Dehner, C. Klein, H. Kessler, and J. Buchner. The N-terminal domain of p53 is natively unfolded. *Journal of the Molecular Biology*, 332:1131–1141, 2003.

[45] O. Schon, A. Friedler, S. Freund, and A. R. Fersht. Binding of p53-derrived ligands to MDM2 induces a variety of long distance conformational changes. *J. Mol. Biol.*, 336:197–202, 2004.

[46] D. C. Fry, S. D. Emerson, S Palme, B. T. Vu, C. Liu, F. Podlaski . NMR structure of a complex between MDM2 and a small molecule inhibitor. *J. Biomol. NMR*, 30:163–173, 2004.

[47] J. Chen, V. Marechal, and A. J. Levine. Mapping of the p53 and MDM-2 interaction domains. *Molecular and Cellular Bioliogy*, 13(7):4107, July 1993.

[48] A. J. Levine. p53, the cellular gatekeeper for growth and division. *Cell*, 88:323–331, February 1997. Review.

[49] C. Klein, and L. T. Vassilev. Targeting the p53-MDM2 interaction to treat cancer. *British Journal of Cancer*, 91:1415, 2004. Minireview.

[50] A. Ayed, F. A. A. Mulder, G. Yi, Y. Lu, L. E. Kay, and C. H. Arrowsmith. Latent and active p53 are identical in conformation. *Nat. Struc. Biol.*, 8(9):756–760, September 2001.

[51] S. M. Picksley, B. Vojtesek, A. Sparks, and D. P. Lane. Immunochemical analysis of the interaction of p53 with MDM2; fine mapping of the mdm2 binding site on p53 using synthetic piptides. *Oncogene*, 9:2523–2529, 1994.

[52] L. T. Vassilev, B. T. Vu, B. Graves, D. Carvanjal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, and E. A. Liu. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*, 303:844–848, February 2004.

[53] T. Rehm, R. Huber, and T. A. Holak. Application of NMR to structural proteomics: Screening for proteins amenable to structural analysis. *Structure*, 10:1613–1618, 2002.

[54] K. Wüthrich. NMR studies of structure and function of biological macromolecules (Nobel lecutre). *Angew. Chem. Int. Ed.*, 42:3340–3363, 2003. Review.

[55] A. Yang and F. McKeon. p63 and p73: p53 mimics, menaces and more. *Nature Reviews Molecular Cell Biology*, 1:199–207, 2000. Review.

[56] J. Furrer, A. Enthart, A. Kühlewein, A. Dehner, C. Klein, S. Hansen, M. Schwaiger, H. Kessler, and G. Gemmecker. Letter to the editor: Backbone ^{1}H, ^{13}C and ^{15}N resonance assignment for the 25.8 kda dna binding domain of the human p63 protein. *J. Biomol. NMR*, 26:377–378, 2003.

[57] H. Lee, K. H. Mok, R. Muhandiram, K. Park, J. Suk, D. Kim, J. Chang, Y. Sung, K. Y. Choi, and K. Han. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J. Biol. Chem.*, 275:29426–29432, 2000.

[58] B. L. Grasberger, T. Lu, C. Schubert, D. J. Parks, T. E. Carver, H. K. Koblish, M. D. Cummings, L. V. LaFrance, K. L. Milkiewicz, R. R. Calvo, D. Maguire, J. Lattanze, C. F. Franks, S. Zhao, K. Ramachandren, G. R. Bylebyl, M. Zhang, C. L. Manthey, E. C. Petrella, M. W. Pantoliano, I. C. Deckman, J. C. Spurlino, A. C. Maroney, B. E. Tomczuk, C. J. Molly, and R. F. Bone. Discovery and cocrystali structure of benzodiazepinedione hdm2 antagonists that activate p53 in cells. *J. Med. Chem*, 48:909–912, 2005.

[59] N. Issaeva, P. Bozko, M. Enge, M. Protopopova, L. G. G. C. Verhoef, M. Masucci, A. Pramanik, and G. Selivanova. Small molecule RITA binds to p53, blocks p53-HDM-2 interaction and activates p53 function in tumors. *Nat. Med.*, 10(12):1321–1328, December 2004.

[60] M. Krajewski, P. Ozdowy, L. D'Silva, U. Rothweiler, and T. A. Holak. NMR indicates that the small molecule RITA does not block p53-MDM2 binding in vitro. *Nat. Med.*, 11(11):1135–1137, November 2005.

[61] S. Bell, C. Klein, L. Müller, S. Hansen, and J. Buchner. p53 contains large unstructured regions in its native state. *J. Mol. Biol.*, 322:917–927, 2002.

[62] R. Bergson, J. J. Hollick, N. J. Westwood, J. A. Woods, D. P. Lane, and S. Lain. Pilot screening programme for small molecule activators of p53. *Int. J. Cancer*, 115:701–710, 2005.

[63] D. S. H. L. Kim, C. L. Ashendel, Q. Zhou, C. Chang, E. Lee, and C. Chang. Novel protein kinase C inhibitors: α-terthiophene derivatives. *Bioorganic & Medicinal Chemistry Letters*, 8:2695–2698, 1998.

[64] R. J. Marles, J. B. Hudson, E. A. Graham, C. Soucy-Breau, P. Morand, R. L. Compadre, C. M. Compadre, G. H. N. Towers, and J. T. Arnason. Structure-activity studies of photoactivated antiviral and cytotoxic tricyclic thiophenes. *Photochemistry and Photobiology*, 56(4):479–487, 1992.

[65] R. Jaenicke and R. Rudolph. Refolding and association of oligomeric proteins. *Methods in Enzymology*, 131:218–50, 1986.

[66] T. D. Goddard and D. G. Kneller. Sparky 3. University of California, San Francisco, 2000.

[67] A. A. Russo, P. D. Jeffrey, and N. P. Pavletich. Structural basis of cyklin-dependent kinase activation by phosphorylation. *Nature Structural Biology*, 3(8):696–700, August 1996.

[68] M. Ruas and G. Peters. The p16^{INK4a}/cdkn2a tumor suppressor and its relatives. *Biochimica et Biophysica Acta*, 1378:F115–F177, 1998. Review.

[69] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, and P. E. Wright. Structural studies pf p21$^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. USA*, 93:11504–11509, October 1996.

[70] W. F. De Azwedo, S. Leclerc, L. Meijer, L. Havlicek, M. Strnad, and S. Kim. Inhibition of cyclin-dependent kinases by purine analogoues crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.*, 243:518–526, 1997.

[71] J. Lee, A. A. Russo, and N. P. Pavletich. Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature*, 391:859–865, February 1998.

[72] B. Xiao, J. Spencer, A. Clements, N. Ali-Khan, S. Mittnacht, C. Brocenõ, M. Burghammer, A. Perrakis, R. Marmorstein, and S. J. Gamblin. Crystal structure of the retinablastoma tumor suppressor protein bound to E2F and the molecular bais of its regulation. *Proc. Natl. Acad. Sci. USA*, 100(5):2363–2368, March 2003.

[73] E. J. Morris and N. J. Dayson. Retinoblastoma protein partners. *Advances in Cancer Reserach*, 82:1–54, 2001. review.

[74] D. Cobrinik. Pocket proteins and cell cycle control. *Oncogene*, 24:2796–2809, 2005.

[75] G. W. Vuister and A. Bax. Quantitative J correlation: A new approach for measuring homoneclear three-bond J($H^N H^\alpha$) coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.*, 115:7772–7777, 1993.

[76] G. Dubin, M. Krajewski, G. Popowicz, J. Stec-Niemczyk, M. Bochtler, J. Potempa, A. Dubin, and T. A. Holak. A novel class of cysteine protease inhibitors: Solution structure of staphostatin A from *Staphylococcus aureus*. *Biochemistry*, 42:13449–13456, 2003.

[77] G. Dubin, G. Popowicz, M. Krajewski, J. Potempa, A. Dubin, and T. A. Holak. Letter to the editor: 1H 15N and 13C NMR resonance assign-

ments of staphostatin a, a specific *Staphyloccocus aureus* cysteine proteinase inhibitor. *J. Biomol. NMR*, 28:295–296, 2004.

[78] C. Dhalluin, J. M. Wierusyeski, and G. Lippens. An improved tocsy experiment with minimal water saturation. *J. Mag. Reson. B*, 111:168–179, 1996.

[79] N.A. Farrow, R. Munhandiram, A. U. Singer, S. M. Pascal, C. M. Kay , G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. Backbone dynamics of a free and a phosphopeptide-complexed SRC homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, 33:5984–6003, 1994.