New Avenues in bio-NMR of Nucleic Acids and Membrane Systems

Inaugural-Dissertation

zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Jan Borggräfe
aus Dachau

Düsseldorf, 17. September 2020

aus dem Institut für Physikalische Biologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Dr. Manuel Etzkorn
Korreferent: Prof. Dr. Holger Gohlke

Tag der mündlichen Prüfung: 31. März 2021

Für meine Eltern

Contents

Summary

The main scope of this thesis is a comprehensive study of the structure and function of the RNA-cleaving 10-23 DNAzyme (Dz), which has a high therapeutic potential but it is limited in its activity *in vivo*. It could be demonstrated that the Dz is well accessible by solution NMR in all states of its catalytic cycle and under suitable conditions also high-resolution spectra of a pre-catalytic complex of Dz with its RNA target could be obtained. Using state-of-the-art techniques, such as exact NOEs, paramagnetic spinlabels and residual dipolar coupling, as well as novel approaches, like ^{19}F experiments and homology restraints, the first structure of Dz could be determined giving unique insights into its function. A key feature of the calculated structure is that Dz's catalytic loop is winding around the RNA at height of its cleavage site and brings catalytically relevant nucleotides in close proximity to the scissile bond. Furthermore, the impact of metal ion cofactors was investigated, revealing that they are necessary to stabilize the Dz in an active conformation as well as being directly involved in the cleavage reaction. At least one Mg^{2+}-binding site is located in the Dz's catalytic loop close to the RNA cleavage site and is only formed in the catalytically active state. Temperature-gradient and real-time NMR experiments unveiled that product dissociation after cleavage might also be a limiting step in Dz's catalysis. These findings are in agreement with additionally performed functional kinetic assays. These results provides a framework for further studies of other DNAzymes or variants and may lay the foundation for engineering theurapeutically active Dz variants.

Moreover, studies on membrane-associated systems such as epidermal growth factor receptor (EGFR) and melanocortic-4 receptor (MC4R) were performed. The interplay between EGFR, which is linked to cancer development, and its potential modulators were investigated and the results point towards an additional regulation mechanism in EGFR signalling, which could be the focus of further *in vivo* studies. MC4R regulates food intake and is targeted by a variety of agonistic and antagonistic hormones. Studies performed on MC4R ligands revealed their strong interaction with negatively charged phospholipid bilayers, which might influence receptor activation and should be taken into account for future studies on MC4R-hormone interactions.
Furthermore, NMR experiments could be performed on the transmembrane receptor bacterioopsin and it could be shown that incorporation of selectively ^{13}C labelled and partially deuterated Leucines (Leu-meth$_{LD}$) significantly improve linewidths in methyl NMR spectra. Leu-meth$_{LD}$ and ^{19}F might serve as useful and widely applicable NMR-sensitive probes which can be introduced into protein or nucleic acid systems, respectively.
This work provides comprehensive insights into several biologically relevant systems and describes novel methods which might open new avenues in biomolecular NMR.

Zusammenfassung

Ein wesentlicher Schwerpunkt dieser Arbeit sind Untersuchungen zur Struktur und Funktion des RNA-spaltenden 10-23 DNAzyms (Dz), welches ein hohes therapeutisches Potenzial besitzt, jedoch nur begrenzt *in vivo* aktiv ist. Es konnte gezeigt werden, dass das Dz allgemein in allen Zustände seines Katalysezyklus für die Lösungs-NMR gut zugänglich ist und unter geeigneten Bedingungen auch hochauflösende Spektren eines vorkatalytischen Komplexes mit seinem RNA-Substrat aufgenommen werden können. Mit modernsten Techniken, wie *exact* NOEs, paramagnetischen *spinlabels* und *residual dipolar coupling* sowie neuartigen Ansätzen, wie ^{19}F-Experimenten und *homology restraints*, konnte die erste Struktur eines Dz ermittelt werden, welche einzigartige Einblicke in seine Funktion liefert. Ein Schlüsselmerkmal der berechneten Struktur ist, dass sich die katalytische Schleife des Dz in der Höhe der Schnittstelle um die RNA windet und damit katalytisch relevante Nukleotide in unmittelbarer Nähe der zu spaltenden Bindung bringt. Zusätzlich wird gezeigt, das Metall-Kationen das Dz in einer aktiven Konformation stabilisieren sowie direkt an der Spaltungsreaktion beteiligt sind. Mindestens eine Mg^{2+}-Bindungsstelle befindet sich in der katalytischen Schleife des Dz und nahe der RNA-Schnittstelle und bildet sich nur im katalytisch aktiven Zustand des Komplexes aus. Temperaturgradienten- und Echtzeit-NMR-Experimente zeigten, dass Produktdissoziation nach dem RNA-Schnitt ebenfalls ein limitierender Schritt in der Katalyse von Dz sein könnte. Die Ergebnisse konnten mit zusätzlich durchgeführten funktionellen kinetischen Untersuchungen verifiziert werden. Diese Arbeit dient als Vorlage für weitere Studien an anderen DNAzymen oder Dz-Varienten und legt möglicherweise den Grundstein für die Entwicklung einer theurapeutisch aktiven Dz-Variante.

Darüber hinaus wurden Studien an membranassoziierten Systemen wie dem epidermalen Wachstumsfaktorrezeptor (EGFR) und dem Melanocortin-4-Rezeptor (MC4R) durchgeführt. Das Zusammenspiel zwischen EGFR, welcher mit Krebsentstehung zusammenhängt, und seinen möglichen Modulatoren wurden untersucht und die Ergebnisse deuten auf einen zusätzlichen Regulationsmechanismus bei der EGFR-Signalübertragung hin, der Schwerpunkt weiterer *in-vivo*-Studien sein könnte. MC4R reguliert die Nahrungsaufnahme und wird von einer Vielzahl von agonistischen und antagonistischen Hormonen angesteuert. Studien an MC4R-Liganden zeigten starke Interaktion mit negativ geladenen Phospholipid-Membranen, welche die Rezeptoraktivierung beeinflussen könnten und bei zukünftigen Studien zu MC4R-Hormon-Wechselwirkungen berücksichtigt werden sollten. Zusätzlich wurden NMR-Experimente and dem Transmembranrezeptor Bacterioopsin durchgeführt und gezeigt, dass der Einbau von selektiv ^{13}C-markierten und teilweise deuterierten Leucinen (Leu-meth$_{LD}$) die Linienbreiten in Methyl-NMR-Spektren signifikant verbesserte. Leu-meth$_{LD}$ und ^{19}F könnten als günstige und breit einsetzbare NMR-Sonden für Protein- bzw. Nukleinsäuresysteme dienen.
Diese Arbeit liefert detailierte Einblicke in verschiedene biologisch relevante Systeme und erläutert neuartige Methoden, welche neue Wege in der biomolekularen NMR eröffnen.

List of Figures

List of Tables

1 Basics of NMR Spectroscopy

The main focus of this thesis lays on nuclear magnetic resonance spectroscopy (NMR) experiments in solution, which exploit the interaction between atom nuclei and a strong magnetic field. NMR is frequently used in analytics of synthesis products or biological mixtures, structure determination of (bio-)polymers and molecular interaction studies, as well as in dynamic measurements. The application of various NMR experiments on different biological systems are described in the experimental sections 3 and 4 and to provide a deeper understanding of the performed experiments a brief introduction of the main concepts of NMR is presented in this section.

1.1 Physical Background

NMR measurements rely on the presence of a strong external magnetic field (B_0), usually in the magnitude of a few Tesla. Every nucleus has a physical attribute called the spin, which can be characterized by the spin quantum number I. Nuclei with I = 0 do not interact with the external field, while all other nuclei align parallel or antiparallel relative to the field vector (Figure 1.1a+b), generating $2 \cdot I+1$ discrete energy levels. The energy states formed by this *Zeeman splitting* can be described by the magnetic quantum number m in the range of

$$m = -I, -I+1, ..., I-1, I$$

Nuclei with a spin quantum number of ½, as for ^{1}H, ^{13}C, ^{15}N, ^{31}P, and ^{19}F, are especially interesting for NMR, because they split into two distinct energy levels (m = +½, -½), usually named α and β state. The transition between α and β (Δm = 1) is an allowed transition in quantum mechanics, and therefore can it be triggered and observed during NMR experiments.

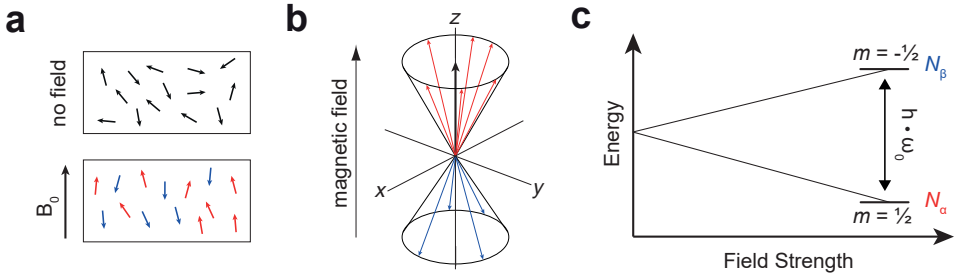


Figure 1.1: The Zeeman Effect. (a) Nuclear spins are magnetic dipols without any preferred orientation (black arrows), but can align with or against an applied magnetic field B_0 (red and blue arrows, respectively) (b) Since more spins are aligned with the B_0 field, a net magnetisation vector in direction of B_0 is formed. (c) The distribution between parallel and anti-parallel aligned spins (N_α and N_b, respectively) is directly proportional to the field strength.

The distribution of the spins occupying α and β states (N_α and N_β) and the energy difference between these states ΔE are described by Boltzmann's law (Equation 1.1) and is depending on the field strength of B_0 and from the gyromagnetic ratio γ, an intrinsic feature of the nuclei. The specific frequency, which matches the energy for state transition, is called Larmor frequency ω_0 and can be derived from division of the energy difference by Planck's constant h (Equation 1.2, Figure 1.1c).

$$\frac{N_\alpha}{N_\beta} = e^{(\frac{\Delta E}{k_B \cdot T})} \tag{1.1}$$

$$\Delta E = h \cdot \omega_0 = \hbar \cdot \frac{\gamma \cdot B_0}{2\pi} \tag{1.2}$$

The difference in the population of α and β state is much smaller than suggested in Figure 1.1a and is in the range of 1% or lower, but due to the high number of spins in a sample (e.g. 10^{20}) the net magnetisation in direction of the external field (z-direction) becomes a relevant (and detectable) dipole, while magnetization in x- and y-direction cancels out (Figure 1.1b). Application of radiofrequency (RF) pulses in the range of the Larmor frequency can tip the magnetisation into the xy-plane and the external field is generating a magnetic moment leading the magnetisation vector to precess around the external field vector. These Larmor precession can be detected by the instrument's receiver coil giving a so called free induction signal (Figure 1.2a).

1.2 Relaxation Processes

The recorded signal is decaying over time (Free Induction Decay, FID) in an exponential manner. Reason for this decay is the transverse relaxation, which is driven by one main process: the individual spins experience different local magnetic fields (B_{loc}) due to the various orientations of the individual molecules. After applying a RF pulse and generating magnetisation in the xy-plane, all spin vectors are pointing in the same direction, which is called phase coherence. Since the Larmor frequency is depending on the experienced magnetic field, which is B_0 plus B_{loc}, the spins will lose phase coherence due to different Larmor precession. This leads to a decrease of phase coherence, net magnetisation and detectable signal (Figure 1.2b). The process is significantly smaller for fast tumbling small molecules, but it becomes severe for larger (bio-)polymers with a rotational correlation time τ_C of several ns.

Another important relaxation mechanism is the longitudinal relaxation, which describes the regeneration of the Boltzmann equilibrium and therefore how fast an experiment can be repeated (Figure 1.2c). This process relies on random spin flips due to dipolar interactions with other spins, which are caused by magnetic field fluctuations of other nearby nuclei. Since the solvent has always the potential to flip a spin from α to β, and *vice versa*, the whole system will relax back into equilibrium over time. The longitudinal and transverse relaxation rates, R_1 and

R_2 respectively, can be descripted by the Bloch equations (Equations 1.3-1.5).

$$dM_z(t) = 1 - dM_z(0)e^{-t/T_1} \tag{1.3}$$

$$dM_{xy}(t) = dM_{xy}(0)e^{-t/T_2} \tag{1.4}$$

with

$$T_{1,2} = 1/R_{1,2} \tag{1.5}$$

where $M_z(t)$ and $M_{xy}(t)$ are the magnetisations in z- and xy-direction at a given timepoint t.

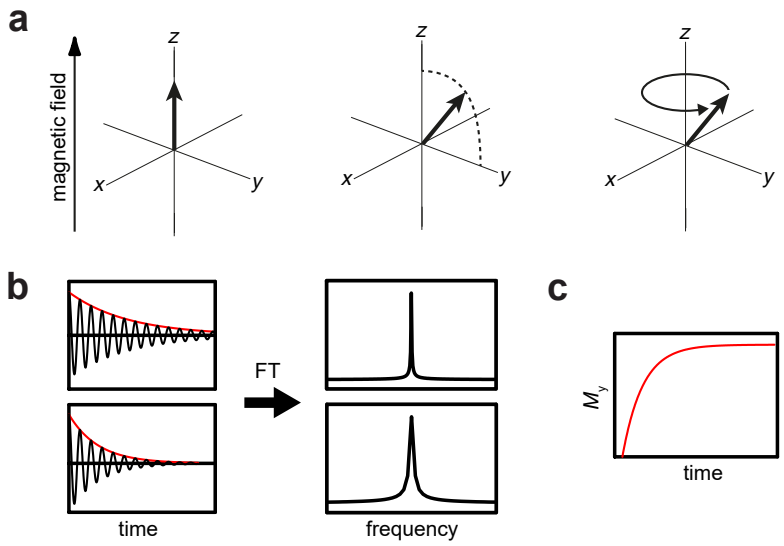


Figure 1.2: Origin and Relaxation of the Free Induction Decay. (a) Net magnetisation along the external magnetic field (left) can be rotated into the xy-plane, e.g. by a selective RF pulse (middle). The external field generates a magnetic moment leading the magnetisation vector to precess around the z-axis (right). (b) Precession of the magnetisation vector can be detected by the instrument's receiver coil, leading to an oscillating signal (left). The signal is decaying according to the transverse relaxation R_2 (Equation 1.4), and leads after Fourier-transformation (FT) to a broadened signal in the resulting spectrum for high R_2 (right). (c) Longitudinal relaxation R_1 (Equation 1.3) describes the regeneration of magnetisation in z-direction after equilibrium was disturbed by a RF pulse.

1.3 The Chemical Shift

As mentioned above, local magnetic fields influence the Larmor precession. Different local fields are mainly generated by electron orbitals in close proximity to the nucleus, whereby protons with low electronegative binding partners, as in methyl groups, experience a lower field and are called shielded. On the other hand, protons with high electronegative binding partners, as in amid, hydroxyl or aromatic groups, experience a higher field and are called deshielded. Since the change of the Larmor frequency in Hz ν_0 (Equation 1.6) by the local field is rather

small (Hz) compared to the effect of the external field (MHz), the recorded signal is usually corrected by the spectrometer's transmitter frequency and multiplied by 10^6 (parts per million, ppm). This normalization also allows to compare spectra recorded with different basic field strengths. However, it is more accurate to use a reference signal of a compound with highly shielded protons, e.g. trimethylsilylpropanesulfonate (DSS) or tetramethylsilane (TMS), rather than solely using the transmitter frequency (Equationn 1.7).

$$\nu_0[Hz] = \frac{\omega_0[rads^{-1}]}{2\pi} \tag{1.6}$$

$$\delta[ppm] = 10^6 \cdot \frac{\nu - \nu_{Ref}}{\nu_{Ref}} \tag{1.7}$$

δ is called the chemical shift, and is undeniably the key feature of NMR spectroscopy. Since it is highly sensitive to the chemical surrounding, nuclei can be assigned individually and be used as individual probes. Alterations of δ due to the environment are called chemical shift perturbation (CSP) and are highly useful to observe structural changes and molecular interactions on atomic level. For instance, they can be used to detect protein secondary structure, ligand binding sites or solvent-induced changes.

1.4 Two Spin Systems

Being magnetic entities, spins can interact with each other via different interactions, which are leading to an observable modulation of the spins' FID. This NMR feature is termed correlation and allows selective spin filtering and to set up multidimensional NMR experiments. The correlation of spins can either derive from direct dipole-dipole interaction, which is described in detail in section 1.5, or indirectly via covalent bonds. The latter is called scalar coupling and will be described in the following section.

1.4.1 The Scalar Coupling

Scalar coupling, or J coupling, is a through-bond interaction. A polarized spin perturbates the paired electrons of covalent bond orbitals, which can themselves pertubate the spin of another nucleus. Scalar coupling is relatively weak and its strength is decreasing corresponding to the amount of bonds involved. However, the weak polarisation leads to slight energy differences between the allowed transitions depending on the state of the coupled spin (Figure 1.3a+b). This results in a signal splitting of a few Hz in the corresponding spectrum (Figure 1.3c), characterized by its multiplicity, which arises from the number of bound half spin nuclei, and the coupling constant *J*, which is the magnitude of the splitting. The J coupling constant between nuclei A and B is usually written as ${}^nJ_{AB}$, where *n* is giving the number of bonds between both atoms. Usually only directly bond (${}^1J_{AB}$), geminal (${}^2J_{AB}$), and vincial (${}^3J_{AB}$) nuclei give measureable coupling constants. Despite the number of involved bonds, also the atom type, the bond length and the bond angle are influencing the magnitude of *J*. The latter one can be used

to experimentally derive dihedral angles due to the Karplus relation [101] (Equation 1.8)

$$J(\Psi) = A\, cos^2\Psi + B\, cos\Psi + C \tag{1.8}$$

with *A*, *B* and *C* being semi-empirical constants. Despite providing useful parameters, splitting due to J coupling can lead to crowded spectra, which are hard to interpret. Therefore decoupling pulse sequences and detection methods are implemented, in order to suppress this effect.

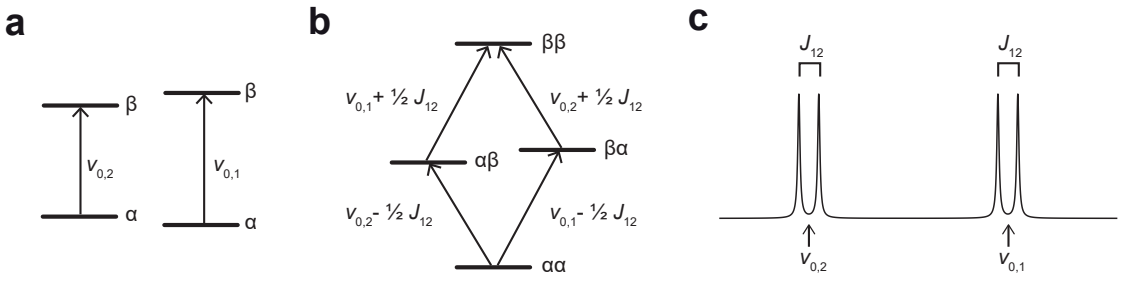


Figure 1.3: Scalar Coupling. (a) Energy levels of two isolated spins 1 and 2 with the eigenfrequencies $\nu_{0,1}$ and $\nu_{0,2}$. (b) Energy levels of a coupled two-spin system. The transitions of the same spin differ slightly by the coupling constant J_{12}, depending on the state of the coupled spin. (c) These energy differences results in line splitting with the magnitude of J_{12}, usually a few Hz.

1.4.2 Multidimensional Experiments

The energy transfer by dipolar and scalar coupling can be use to perform multidimensional experiments, in which a signal derived from two coupled spins *I* and *S* is modulated by the frequencies of both, ν_I and ν_S respectively. Since the modulation of the signal is time-depending, two evolution times (t_1 and t_2) have to be introduced in the experimental setup, which leads to a spectrum with two frequency dimensions (F_1 and F_2) after Fourier-transformation (FT). The signal can only be recorded in one time dimension during acquisition (t_2), therefore the other time-dimension is introduced by an variable delay t_1. This delay is stepwise varied during different one-dimensional sub-experiments (increments). FT along t_2 gives an array of 1D spectra with signals oscillating according to their modulation during t_1. Subsequential FT along t_1 leads to a twodimensional spectrum with crosspeaks for correlated nuclei. A scheme of this procedure is shown in Figure 1.4.
Since multiple increments for every dimension are needed, multidimensional experiments are quiet time-expensive. A typical 1D experiment can be performed within a few seconds, but it needs minutes for a two-dimensional and hours for a three-dimensional version. Since higher-dimensional experiments easily can take many days or weeks, they are usually not performed. However, data acquisition methologies like non-uniform sampling (NUS) are invented [131],

which allows to record the experiments in a fraction of the time normally needed, but the complexity and data size of high-dimensional spectra still limits their usability.

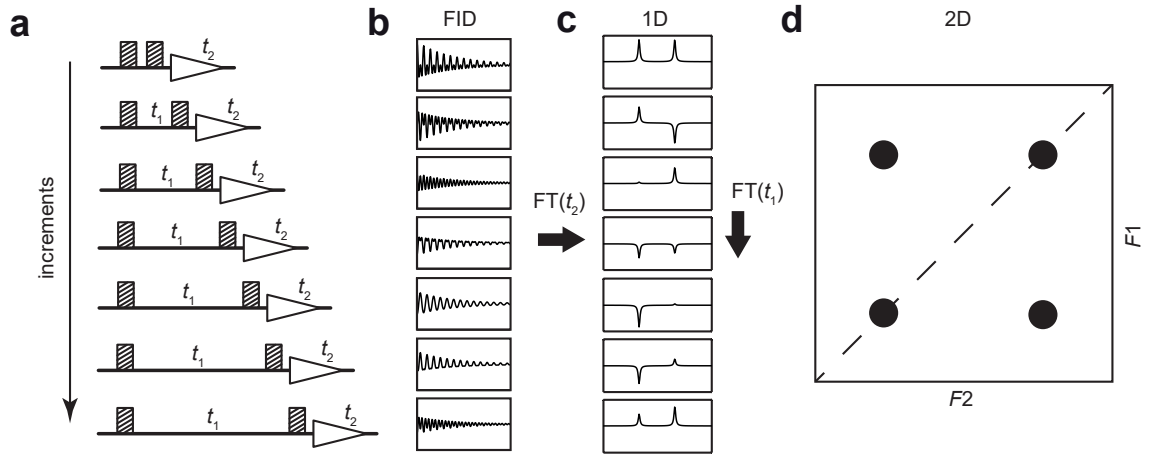


Figure 1.4: Two-dimensional Experiment. 2D experiments needs signal evolution in two time dimensions, once directly during the acquisition time t_2 and once by an evolution time t_1, which is varied in each 1D experiment, or increment (a). FT of the directly recorded FIDs (b) leads to 1D spectra (c), which are oscillating with the increments, or time t_2. A second FT along t_2 results in a 2D spectrum with crosspeaks for correlated nuclei (c).

1.5 Dipole-Dipole Interactions

As described in section 1.2, an important relaxation mechanism is the magnetisation exchange between two nuclear spins by dipolar interactions, which is termed the Nuclear Overhauser Effect (NOE). NOE as well as dipole-dipole interactions between nuclear and electron spins often serve as distance restraints in solution NMR-based structure calculations. The energy levels and possible transition for a homonuclear two-spin systems are shown in Figure 1.5. It can be seen, that despite the usually allowed single-quantum transitions (W_{1I} and W_{1S}) also double-quantum (W_{2IS}) and zero-quantum transitions (W_{0IS}) are possible relaxation pathways. Transitions W_{0IS} and W_{2IS} can exchange energy between the two spins, therefore relaxation of spin I is affecting the magnetisation of S, and *vice versa*. The change of magnetisation for both spins can be descripted by the Solomon equations:

$$\frac{dI(t)}{dt} = -\rho_I(I(t) - I_0) - \sigma_{IS}(S(t) - S_0) \tag{1.9}$$

$$\frac{dS(t)}{dt} = -\rho_S(S(t) - S_0) - \sigma_{IS}(I(t) - I_0) \tag{1.10}$$

with

$$\rho_I = W_{0IS} + 2W_{1I} + W_{2IS} \tag{1.11}$$

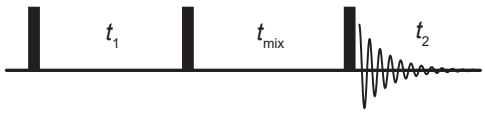
$$\rho_S = W_{0IS} + 2W_{1S} + W_{2IS} \tag{1.12}$$

$$\sigma_{IS} = -(W_{0IS} - W_{2IS}) \tag{1.13}$$

The auto-relaxation rate ρ_I (or ρ_S) are the sum of all transitions involving spin I (or S) and is driving z-magnetisation back into equilibrium, while σ_{IS} is describing the magnetisation addition from W_{0IS} and W_{2IS} and is therefore termed cross-relaxation. In a quantum-mechanical description it can be derived, that σ_{IS} is inverse proportional to the sixth power of the distance between the two nuclei r_{IS}:

$$\sigma_{IS} = (\frac{\mu_0}{4\pi})^2 \frac{{\gamma_I}^2 {\gamma_S}^2 \hbar^2 \tau_c}{10} \frac{1}{{r_{IS}}^6} [\frac{6}{1 + 4{\omega_0}^2 {\tau_c}^2} - 1] \tag{1.14}$$

with γ_I and γ_S are the gyromagnetic ratios of nucleus I and S, respectively. In a typical Nuclear Overhauser Effect Spectroscopy (NOESY) experiment, transverse magnetisation for both (homonuclear) spins is produced by a 90° RF pulse, followed by a first evolution t_1. Another 90° pulse regenerates again magnetisation in z-direction, followed by the mixing time t_{mix}, in which NOE transfer occurs. A last 90° pulse generates again transverse magnetisation, which is recorded during the evolution time t_2. The following scheme depicts the most basic NOESY pulse sequence with 90° pulses as black boxes:



Due to the first evolution time the z-magnetisation for both spins, I_z and S_z, is somewhat different. Exemplarily it is assumed, that I_z is close to its initial magnetisation $I_z(0)$, while S_z varies more from $S_z(0)$:

$$I_z - I_z(0) = 0 \tag{1.15}$$

$$S_z - S_z(0) \neq 0 \tag{1.16}$$

For very short mixing times (t_{mix}) of a few ms the magnetisation $I_z(t_{mix})$ and $S_z(t_{mix})$ are still close their initial values, and the solution for Equations 1.9 gives a linear dependency:

$$\frac{I_z(t_{mix})}{dt_{mix}} = 2\sigma_{IS} {S_z}^0 \tag{1.17}$$

For longer mixing times auto-relaxation processes become relevant. Assuming, that $\rho = \rho_I = \rho_S$ the solution for Equations 1.9 becomes

$$\frac{I_z(t_{mix})}{dt_{mix}} = {S_z}^0 (e^{-(\rho+\sigma_{IS})t_{mix}} - e^{-(\rho-\sigma_{IS})t_{mix}}) \tag{1.18}$$

The impact of different auto-relaxation rates on the NOE transfer is shown in Figure 1.5b.

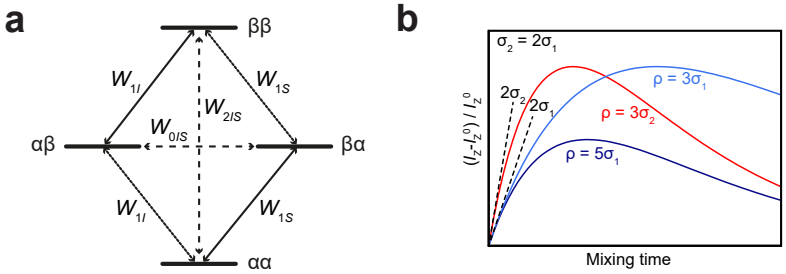


Figure 1.5: The Nuclear Overhauser Effect. (a) The diagram shows possible transitions during relaxation of a two-spin system with the individual transition probabilities *W*. (b) Nuclear Overhauser enhancement predicted by Equation 1.18 for a two-spin system with $\rho = \rho_I = \rho_S$. The NOE intensity is linear to the cross-relaxation rate σ_{IS} for very short mixing times t_{mix}, but strongly depending on individual ρ for longer t_{mix}.

1.5.1 Conventional Nuclear Overhauser Effect

To generate distance restraints from NOESY crosspeaks, e.g. for structure calculation, one usually integrates all assigned crosspeaks in a NOESY spectrum and normalizes their intensities by the intensity of a crosspeak derived from a known distance. According to their relative intensities, crosspeaks are grouped into strong, intermediate, and weak NOE contacts, with certain ranges for distance restraints. Nevertheless, this approach has several issues, which are making it unreliable to get accurate distances:

a) The accuracy of this approach is maximal, if the distance used for normalization is close to the unknown distances. However, due to conformational changes within a molecule, there are not many distances which are known within a certain confidence. Therefore, especially short distances are used for normalization, which makes longer distances more erratic.

b) The application of very short mixing times, at which the NOE dependency is linear, is often not possible, since weak crosspeaks are not measurable due to low signal-to-noise ratio.

c) For systems in slow motion with $\omega_0\tau_c >> 1$, e.g. macromolecules, energy transfer by dipolar interactions is increased. Since every molecule has many more magnetic active nuclei than a two-spin system, magnetisation can be transferred further to additional nuclei. This *spin diffusion* renders the NOE build-up far more complex, and can make long-distance NOE contacts appear stronger, or short-distance contacts weaker.

The first two points can be addressed by recording NOE data with varying t_{mix} and fitting them using equation 1.18. Normalization of σ_{IS} is far more accurate to generate distance restraints than using intensities from one data set alone. Several attempts were performed to overcome the severe effect of spin diffusion, of which one is described below.

1.5.2 Exact Nuclear Overhauser Effect

A macromolecule consisting of *N* half-spins is given by a *N*-spin system. The full relaxation processes between all spins can be described in good approximation by a relaxation matrix ***R***,

with the autorelaxation rates forming the diagonal elements, while the crossrelaxation rates between the individual spins are given by the respective off-diagonal elements [17]. The evolution of the magnetisations can then be described by the following Solomon equation:

$$\frac{d}{dt}\begin{pmatrix} I_1(t) \\ I_2(t) \\ \dots \\ \dots \\ I_N(t) \end{pmatrix} = -\begin{pmatrix} \rho_1 & \sigma_{12} & \dots & \dots & \dots & \sigma_{1N} \\ \sigma_{12} & \rho_2 & \dots & \dots & \dots & \sigma_{2N} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{1N} & \sigma_{2N} & \dots & \dots & \dots & \rho_N \end{pmatrix} \begin{pmatrix} I_1(t) \\ I_2(t) \\ \dots \\ \dots \\ I_N(t) \end{pmatrix} \tag{1.19}$$

$$\frac{d}{dt}I(t) = -\boldsymbol{R}I(t)$$

With $t_{mix} > 0$ the magnetization vector $I(t_{mix})$ becomes a NxN matrix, describing the peak intensities of a 2D NOESY recorded with the respective mixing time. The matrix diagonal correspond to the diagonal of the 2D spectrum, while the off-diagonal gives the intensities of the corresponding crosspeaks. Theoretically, a NOESY spectrum can be generated from a known structure, and *vice versa*. However, experimentally it is rarely possible, since an accurate, fully determined matrix for $I(t_{mix})$ can often not be derived from measurements due to high signal overlap and/or low signal-to-noise ratio. In a Full Relaxation Matrix Approach, $\boldsymbol{R}$ is usually generated using a preliminary structure calculated in conventional ways, and then $I(t_{mix})$ is supplemented by calculated values. This approach is done iteratively until the solution converges. The approach used in this work applies a similar method published by the Vögeli group [112, 119]. Firstly, all crosspeaks are treated as an isolated two-spin system (as explained above) and the individual spin diffusion impact for every (determined) nucleus is calculated from a preliminary model. Then, the apparent cross-relaxation rates are corrected by the calculated spin diffusion factor and used to generate distance restraints for structure calculations. The newly calculated structure is then again used to determine spin diffusion factors, and the process is repeated.

1.5.3 Saturation Transfer Difference

Spin diffusion is depending on spins in close proximity and also can be used as a tool to study the proximity of different molecules or molecular parts. The saturation transfer difference (STD) experiment exploits spin diffusion effects and is well-established to analyse weak binding of ligands on proteins [171]. Proton spins of a protein are saturated by selective irradiation for several seconds at a frequency, where the ligand protons are not affected, e.g. the methyl region. Due to the efficient NOE transfer in large molecules, the saturation is rapidly transferred over the whole protein (Figure 1.6a). A bound ligand will also be saturated and eventually dissociates, which lowers the overall NMR signal of unbound ligand. In a resulting spectrum (on-resonance), the ligand protons would show slightly decreased intensities compared to a reference spectrum, in which the saturation pulse is applied far away from any frequency of protein or ligand (off-resonance). The difference spectrum between the on- and off-resonance

spectra makes this effect visible (Figure 1.6b left). This approach can be transferred to structural studies by substitution of ^{1}H with ^{19}F. Both nuclei have comparable gyromagnetic ratios, but their Larmor frequency differs largely. Therefore, ^{19}F can be selectively saturated and spin diffusion effects on nearby protons can be observed (Figure 1.6b). However, an exact distance determination like in the NOESY experiments is not possible, but it is helpful to find initial low resolution restraints in otherwise too crowded spectra.

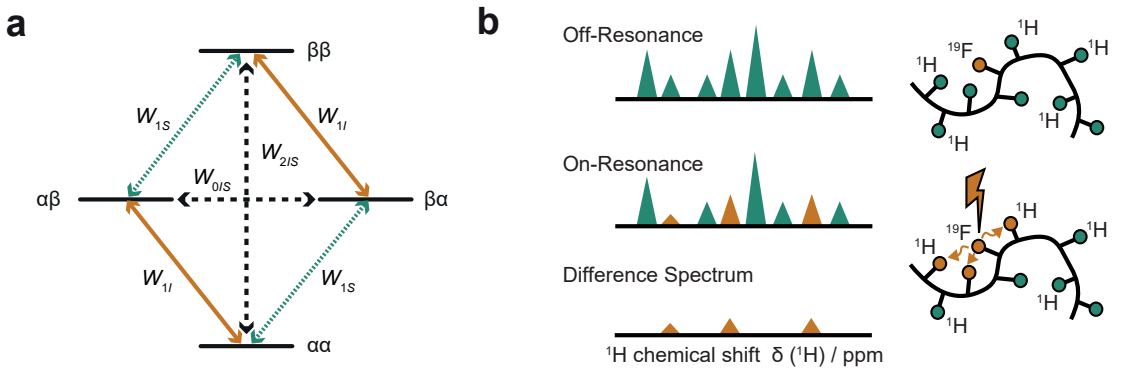


Figure 1.6: Saturation Transfer Difference. (a) The diagram shows a two-spin system with selective saturated spin *I*. During saturation the populations of the states are $N_{\alpha\alpha} = N_{\alpha\beta}$ and $N_{\beta\alpha} = N_{\beta\beta}$, therefore relaxation to equilibrium has to occur via transitions W_{0IS} and W_{2IS}. (b) Schematic representation of a STD experiment. The fully green spectrum represents a proton spectrum without any saturation effects (off-resonance). Selective saturation of a fluorine (orange sphere) leads to spin diffusion towards nearby protons (green spheres if unaffected, orange spheres if affected). This results in a slight intensity decrease of the affected protons in the on-resonance spectrum (affected signals are in orange). Subtracting of the on- from the off-resonance spectrum gives the difference spectrum.

1.5.4 Paramagnetic Relaxation Enhancement

Paramagnetic centres, like certain transition metals and radical compounds, are also highly magnetic active, due to the 1/2-spin character of their unpaired electrons. Despite their quantum-mechanical formalism is different, electrons can be seen as spins with a very high gyromagnetic ratio and very short transverse relaxation time. The correct description of the longitudinal (Γ_1) and transverse (Γ_2) paramagnetic relaxation rates is given by the Solomon-Bloembergen equations [156]:

$$\Gamma_1 = \frac{2}{5}(\frac{\mu_0}{4\pi})^2\gamma_I{}^2 g^2 \mu_B{}^2 r^{-6} S(S+1)\frac{\tau_C}{1+(\omega\tau_C)^2)} \tag{1.20}$$

$$\Gamma_2 = \frac{1}{15}(\frac{\mu_0}{4\pi})^2\gamma_I{}^2 g^2 \mu_B{}^2 r^{-6} S(S+1)\{4\tau_C + 3\frac{\tau_C}{1+(\omega\tau_C)^2}\} \tag{1.21}$$

$g{\cdot}\mu_B$ is a factor describing the magnetic susceptibility of electrons, comparable to γ of nuclei. Due to the high "γ" of electrons, coupling between diamagnetic nuclei and electrons can act up to 35 Å, much higher than for the NOE transfer (up to 5-6 Å). The correlation time τ_C is

dominated by the relaxation rate of the electron, meaning that the coupled nucleus is relaxing much faster and shows severe line broadening (Figure 1.7a). This effect is called paramagnetic relaxation enhancement (PRE) and became an useful tool in structural biology. Paramagnetic spinlabels, such as TEMPO or MTSL (Figure 1.7a top), can be introduced site-specifically into a protein or nucleic acid by utilizing functional groups and the resulting line broadening effects can be translated into (loose) distance restraints. Also paramagnetic Mn^{2+} is often used as a mimic to locate Mg^{2+} binding sites, since increasing amounts of Mn^{2+} during a titration experiment leads to specific line broadening.

1.5.5 Residual Dipolar Coupling

Similar to scalar coupling (see 1.4.1), also dipolar coupling leads to different energy levels of the coupled states, which results in peak splitting by the dipolar coupling constant *D*. While this effect is severe in solid-state NMR spectroscopy, the isotropic tumbling in solution renders *D* effectively zero. If molecules are partially aligned by adding a suitable co-solvent into the buffer, this effect is not completely averaged out, which leads to peak splitting described by a residual dipolar coupling (RDC) constant (Figure 1.7b). Large systems, which can orientate along an magnetic field and present a substantial interaction surface, are used as an alignment medium, such as bicelles and phages [30]. The magnitude of the RDC constant is depending on γ of the involved nuclei, the orientation of the vector connecting both nuclei to the main magnetic field as well as the vector magnitude. Since precise atom distances are usually difficult to obtain and *D* is maximal between nuclei in close proximity, usually dipolar couplings between heteronuclei ^{13}C or ^{15}N and directly bond protons are analysed. From known distances and experimentally derived RDC values, angles of the bond vector towards the main magnetic field can be calculated [169]. RDC constants do not provide distance information, but relative orientations between different bond vectors in a biopolymer.

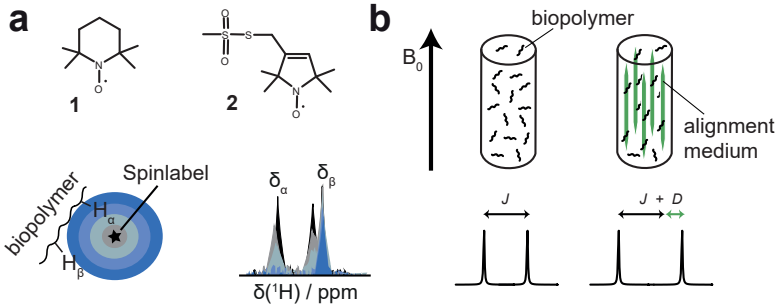


Figure 1.7: Paramagnetic Relaxation Enhancement and Residual Dipolar Coupling. (a) Paramagnetic compounds like (2,2,6,6-tetramethylpiperidin-1-yl)oxyl (TEMPO, **1**) and *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methylmethanesulfonate (MTSL, **2**) are often applied as spinlabels, which broaden NMR signal in a distance-depending manner. (b) Biopolymers partially align towards the external magnetic field in presence of an alignment medium. This leads to peak splitting by the amount of the residual dipolar coupling constant *D*, which is added to the scalar coupling constant *J* in spectra without decoupling.

1.6 Isotope Labeling

The NMR active ^{1}H and ^{31}P are found nearly exclusively in biomolecules, while ^{13}C and ^{15}N only have a natural abundance of around 1% and 0.4%, respectively. Although 2D heteronuclear experiments with ^{13}C and ^{15}N can be recorded with a high number of scans or high sample concentration without previous isotope-enrichment, correlation spectra between two heteronuclei are nearly impossible to acquire. Complete labelling of proteins with ^{15}N and ^{13}C is well-established for bacterial, yeast and insect cell cultures, however deuteration is commonly done in prokaryotic expression systems. ^{2}H enrichment greatly reduces spin diffusion effects and improves line width of large protein complexes [142]. Selective labelling with NMR active isotopes often simplifies overlapping spectra and can be achieved by adding labelled amino acid precursors for in-cell [53, 70, 167, 168] or directly labelled amino acids for cell-free expression [82].
Isotope-labelled nucleic acids can also be produced enzymatically [90, 110], however these techniques are not widely used. Usually, nucleic acid oligomers are synthesized on solid-phase, since it is cheap for unlabelled samples, but isotope-enrichment increases the production costs notably. On the other hand, synthesis of non-native fluorinated nucleic acids are inexpensive and NMR active ^{19}F has a 100% natural abundance, which makes it interesting for bio-NMR studies.

1.7 Dynamics accessible by solution NMR

In general, molecules show randomized diffusion and intramolecular dynamics due to Brownian motion. Covalent bonds oscillate and rotate very fast (ps-ns), while local reorientations, e.g. nucleobase flipping or helix movements in proteins, often occur at slower timescales (µs-ms). Large structural changes, like refolding events, instead, take place in seconds or even hours. NMR spectroscopy is one of the few methodologies able to access dynamics in all these time regimes and in the best case with atomic resolution [95]. However, different experimental setups are required depending on the timescale of interest (Figure 1.8).

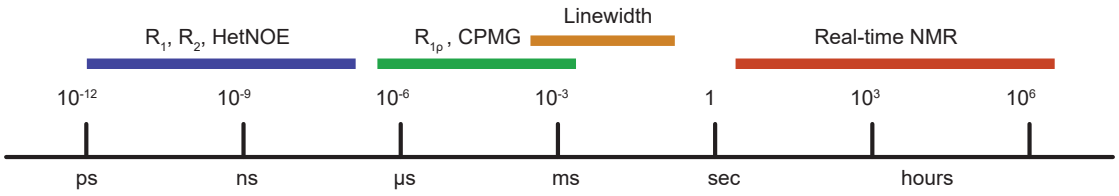


Figure 1.8: Dynamic Range accessible by NMR. NMR spectroscopy can access (bio-)molecular dynamics in nearly all time regimes, from thermal vibrations (ps-ns timescale) over domain movements (ns-ms timescale) to refolding processes (seconds or longer). A subset of different NMR experiments for each time domain are shown.

1.7.1 Fast Dynamic Processes

To get insights into the dynamic behaviour of a system, the observable of an experiment has to be significantly altered by the dynamic process itself. As seen from section 1.2, transverse and longitudinal relaxation is very sensitive to frequencies close to the Larmor frequency (in the MHz regime), which correspond to fluctuations in the ps-ns regime. Therefore, accurate determination of the relaxation rates R_1 and R_2 will give insights into dynamics at this time scale. To avoid spin diffusion effects, usually the relaxation rates of low-γ nuclei like ^{13}C or ^{15}N are recorded. Standard experiments for determination of R_1 and R_2 are the inverse-recovery and the spin-echo experiments, respectively. In the inverse-recovery experiment, a 180° pulse is inverting the spin's z-magnetisation (I_z). During a time t, the magnetisation goes back to equilibrium (from -I_z to I_z) and after t, a 90° pulse transforms I_z into detectable xy-magnetisation. For determination of R_2 the loss of transverse magnetisation is of interest, therefore just one 90° pulse is applied and the signal is detected after time t. However, since also inhomogeneity of the external field can lead to dephasing, a 180° pulse is applied after t/2 to compensate this effect, a pulse sequence called spin-echo. The inverse-recovery and the spin-echo experiments gives time-depending signal intensities, which can be fitted with the Bloch equations 1.3 and 1.4.

Another method for analysing fast dynamics is the HetNOE experiment and depends on the magnetisation transfer efficiency between a proton (I) and a heteroatom (S) as a function of correlation time τ_C. Since, as already discussed, the magnetisation transfer is especially depending on the distance between the nuclei, directly bound protons and heteroatoms are used, e.g. amid or methylene groups. The proton magnetisation is, like in the STD experiment, saturated by a series of weak pulses. The NOE transfer is now in steady-state, which means magnetisation of the heteronucleus is not changing over time ($S_{z,ss}$). From equation 1.8 derives

$$S_{z,ss} = \frac{\sigma_{IS}}{\rho_S} I_z(0) + S_z(0) \tag{1.22}$$

By using a reference experiment in absence of saturation, the NOE enhancement η_{ss} can be calculated (Equation 1.23), which is directly proportional to the cross-relaxation rate σ_{IS}, and therefore, to the rotational correlation time τ_C.

$$\eta_S S = \frac{S_{z,ss} - S_{z,ref}}{S_{z,ref}} = \frac{\sigma_{IS}}{\rho_S} \frac{I_z(0)}{S_z(0)} \tag{1.23}$$

1.7.2 Intermediate Dynamic Processes

As introduced in chapter 1.2, the linewidth is depending on the transverse relaxation time R_2, but it can also be influenced by other processes. Since the FID is usually recorded for several tens or hundreds of ms, chemical exchange at this time range can have an impact on the FID. For simplifications a system with two states A and B is assumed, each with a specific Larmor frequency ν_A and ν_B. Both states are interchanging with an exchange rate k_{ex}. In the first case, the exchange rate is far higher than the difference between the two Larmor frequencies (k_{ex}

>> ν_A-ν_B, Figure 1.9a top). The FID shows the averaged signal for both states and a peak will appear at a frequency between both Larmor frequencies (according to the relative population of both states). In the second case, the exchange rate is much lower (k_{ex} << ν_A-ν_B, Figure 1.9a bottom), and both states will appear as separate signals with their corresponding frequencies. However, the relative peak intensities are depending on the relative population of the states. In the last case, the chemical exchange rate is close to the frequency difference ($k_{ex} \sim \nu_A$-ν_B, Figure 1.9a middle) and only one peak is observed. The linewidth of this peak is (severely) broadened due to the interconversion between state A and B during the detection period. The exchange rates can be altered by temperature changes, and ν_A-ν_B can be altered by changing the field strength (Figure 1.9a). Despite useful for studying underlying dynamics, line broadening due to intermediate exchange creates often a problem in NMR titration experiments. The signal might "broaden away" during the titration, which can lead to an insufficient dataset to determine binding constants.

A more in-depth analysis of dynamics in the µs-ms timescale can be done by Carr-Purcell-Meiboom-Gill (CPMG) [111] relaxation dispersion or rotating-frame ($R_{1\rho}$) [43] experiments, since they can provide exact exchange rates and uncover normally invisible low-populated conformations.

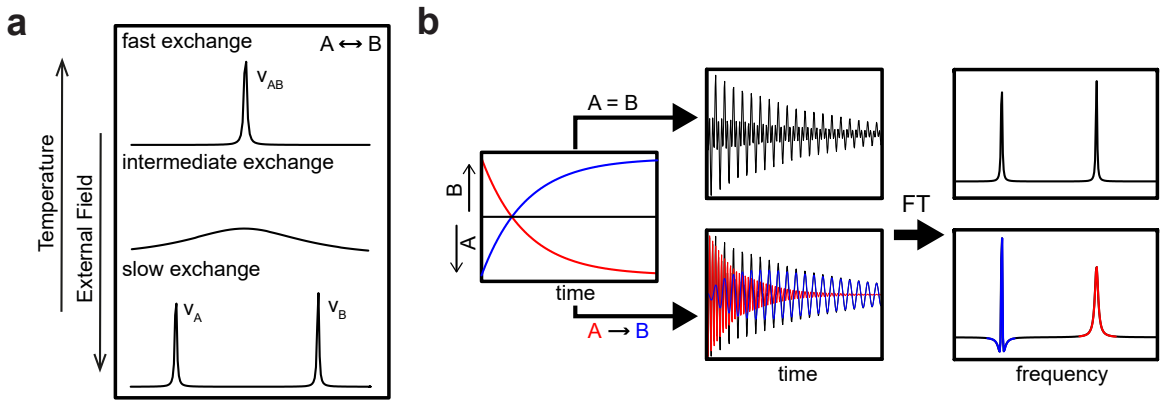


Figure 1.9: Dynamics and Lineshape. (a) Signal lineshapes of interchanging states with different exchange rates. Slow exchange results in two signals each for one state (bottom), while fast exchange results in one signal at a frequency between both states (top). Intermediate exchange leads also to one but severely broadened signal (middle). Exchange regimes can be altered, either by increase of exchange rates with increasing temperature, or by increase of the frequency difference between both states with increasing field strength. (b) Real-time (RT) NMR experiments are recorded during the transition of state A to state B (left). State A is disappearing during acquisition resulting in a faster decaying FID and a broadened peak (red). The signal of state B is building up at the begin of the experiment, and decaying later due to transverse relaxation, which leads to a sharper peak with characteristic flanking dips (blue). Lineshape analysis of the RT-NMR spectrum and comparison with lineshapes of the system in equilibrium (black) yields rate constants for the individual peaks.

1.7.3 Slow Dynamic Processes

The recording of dynamic processes in the range of minutes or hours, such as chemical reactions or folding events, is termed Real-time (RT) NMR. Again two states A and B are assumed, but instead of being in equilibrium, only state A is present at the beginning of the experiment (Figure 1.9b). During the acquisition, A is slowly converting into B, while the total acquisition time is fully covering the process. In praxis, 2D experiments are chosen, where the indirect dimension *F*2 is modulated by the interconversion. During the reaction time and acquisition, the educt signal is not only decaying due to relaxation, but also by the conversion of A into B according to the reaction rate. This leads to a broadened educt peak compared to a reference spectrum. On the other hand, the product signal is first increasing due to the interconversion, but soon decreasing due to transverse relaxation. This leads to a product peak with decrease linewidth and characteristic flanking dips. The lineshapes of educt and product peaks can be simulated using different rate constants, and the simulated curves are compared with the experimental data [9], which gives insights into slow dynamic processes with atomic resolution.

1.8 NMR-derived Structure Calculation

On molecular level life is highly organized by different biopolymers, most prominently by proteins, ribonucleic (RNA) and deoxyribonucleic acids (DNA). A main field of structural biology is solving their atomic structure to understand the underlying mechanistic and interactive features. Despite high computational power and experimentally derived data-sets, also understanding of the elementary building blocks of biopolymers is necessary.

1.8.1 Protein Structure Elements

Proteins are build up by a chain of up to 22 different proteinogenous amino acids, each with different physical and chemical properties, such as charges, polarity, hydrophobicity, size and rotational freedom, facilitated by their individual sidechains. This allows proteins to adopt a variety of different structures and functions, for instance as bio-catalysators, receptors or parts of the cytoskeleton. Two amino acids are connected by an aminoester between their carboxylic and amino group, the peptide bond (Figure 1.10a). Due to the partial double bond character of these aminoesters, free rotation around this bond and therefore the flexibility of the polymer chain is hindered. Only a handful of conformations are possible, which can be displayed by a Ramachandran plot (Figure 1.10b). The most abundant features of these secondary structure elements are α-helices and β-sheets, since these structures are also stabilized by a large amount of hydrogen bonds (Figure 1.10c). Different secondary structures show distinct different chemical shift patterns for the amino acid nuclei, which is termed the Chemical Shift Index (CSI) [179]. Typical dihedral angles and hydrogen bond networks of a polypeptide strand can be derived from the CSI, and even exact angles can be calculated from a full set of chemical shift data by using software like TALOS [151]. One or several peptide chains can fold into

higher complexity, called ternary structure, and is usually driven by interaction between different amino acid sidechains or sidechains with the solvent. Long-range NOE contacts are the main NMR-derived restraints for calculating the ternary structure of proteins, since they are usually abundant due to sidechain-sidechain interactions. Protons of methyl groups are highly shielded, which makes them easy to distinguish in NOESY spectra and therefore NOE contacts including methyl protons are often utilized as structural restraints. Amino acids with methyl groups, such as Leucine, Valine and Isoleucine, are also often use as targets for selective labelling strategies [53, 70, 167, 168].

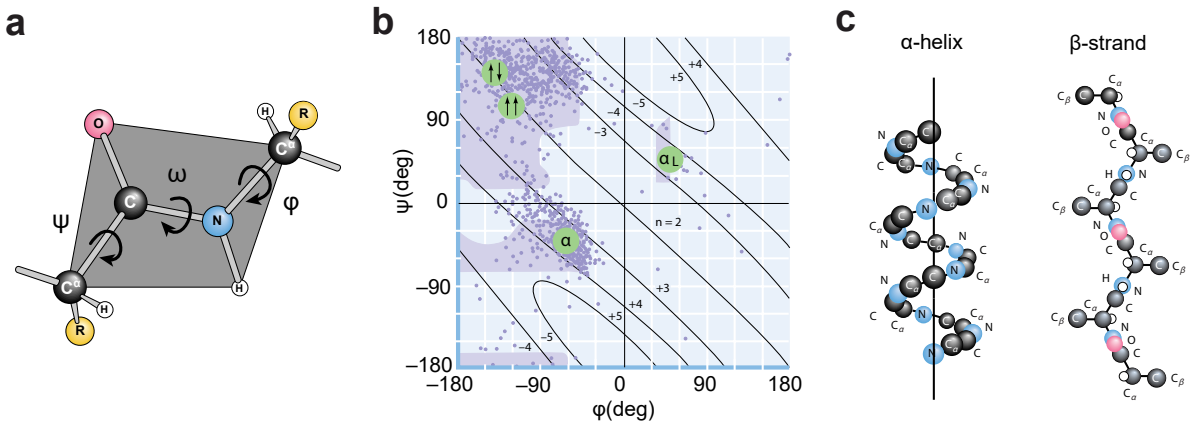


Figure 1.10: Secondary Structure of Proteins. (a) Dihedral angles ω, φ, and ψ of the peptide bond. sp^2-Hybridization of N, O, and C is restraining angle ω in a plane. (b) Ramachandran plot shows allowed ψ and φ conformations for amino acids excluding glycines (purple region). Dots represent actually measured angles for 1000 amino acids. Conformations of parallel (↑↑) and anti-parallel β-sheets (↑↓) and right- (α) and left-handed α-helices (α_L) are highlighted. (c) Scheme of the most abundant secondary structure elements α-helix and β-strand. Figures adapted from *Biochemistry*, Garret RH, Grisham CM [54].

1.8.2 Nucleic Acid Structure Elements

In contrast to proteins, nucleic acids mainly function as repository and regulators of genetic information, despite some have also catalytic activity [91]. RNA and DNA are chains of up to four different nucleotides, connected by a phosphate ester between the 5' and 3' hydroxyl group of the ribose moieties (Figure 1.11a). The ribose unit of a nucleotide is linked with a purine (guanine G, adenine A) or pyrimidine (cytosine C, thymine T in DNA or uridine U in RNA) nucleobase bound at 1' position of the ribose unit. In DNA the 2'-hydroxyl group of the sugar moiety is replaced by a proton, which makes the polymer more resistant against hydrolysis. Nucleic acids are mainly stabilized by hydrogen bonds between different nucleobases or nucleobases and the phosphate backbone, where Watson-Crick base-pairing between G and C or A and T/U is the most prominent, since it allows hybridization of two complementary strands. Nucleic acids have an up to threefold lower hydrogen content than proteins of comparable size,

and chemical shift dispersion is severely lower, which renders classical NMR approaches more complicated. The nucleic acid structure on atomic level is described by several torsion angles, of which α, β, γ, δ, ε, and ζ define the chain's backbone conformation, while the dihedral angle χ restraints the relative orientation of nucleobase and sugar moiety (Figure 1.11a). The latter can often be determined from the intensity of intra-nucleotide NOE contacts between sugar and base protons. Ribose is a five-atom ring system and can only adapt a reduced number of stable conformations due to the sp^3-hybridization of the involved carbons. These distorted ring conformations are called sugar puckering and can be described by a pseudo-rotation phase P and an amplitude Φ_m, from which the dihedral angles of the ring system ($\nu_{1\text{-}5}$) can be back-calculated:

$$\nu_j = \Phi_m cos[P + 0.8\phi(j-2)] \tag{1.24}$$

The most favourable ribose conformations are 2'-*endo* and 3'-*endo* puckers, where the name refers to the ring carbon which is oriented towards the nucleobase (Figure 1.11b). Coupling constants between ^{31}P and ^{1}H can be used to derive dihedral angles of the phosphate backbone, while J^3_{HH} gives insights into sugar puckering. However, for large nucleic acid system the splitting patterns in decoupled spectra can highly overlap, which might complicate this approach to a not solvable level.

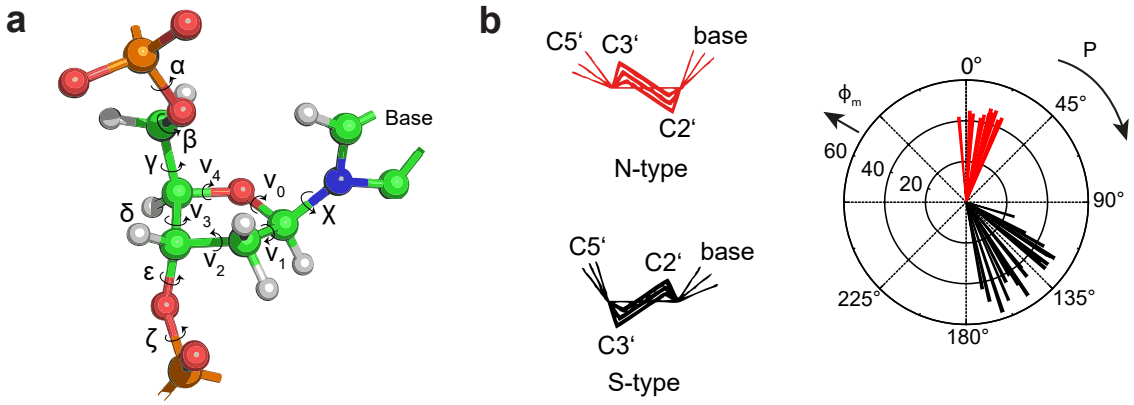


Figure 1.11: Dihedral Angles of Nucleotides. (a) Dihedral angles α, β, γ, δ, ε, and ζ define the backbone structure, $\nu_{1\text{-}5}$ the sugar pucker, and χ the nucleobase orientation of nucleic acids. (b) Most favourable puckering states of ribose are C3'-*endo* (left upper) and C2'-*endo* (left lower). Puckering can be described by pseudo-rotational phase P and amplitude Φ_m. In a common polar plot (right), C3'-*endo* conformation are facing north (N-type) and C2'-*endo* south (S-type).

1.8.3 Restraint Molecular Dynamics / Simulated Annealing

NMR-based structure calculations depend on the minimization of a target function, which describes the agreement between a conformation and a (experimentally) derived set of restraints.

Most commonly used are distance geometry approaches and restraint molecular dynamic (rMD) calculations, where in the latter all given restraints are described as an energy function, which is solved numerically [147, 148]. In the rMD approach, the total potential function E_{total}, which is defining the system, is called force field and consists of potentials for all bonding and non-bonding atom-atom interactions (Equation 1.25-1.30).

$$E_{total} = E_{bond} + E_{angle} + E_{dihed} + E_{non-bond} + E_{repel} \tag{1.25}$$

$$= \Sigma[k_{bond}(b(r) - b_0)^2] \tag{1.26}$$

$$+ \Sigma[k_{angle}(\theta(r) - \theta_0)^2] \tag{1.27}$$

$$+ \Sigma[\frac{k_\Phi}{2}(1 + cos(n\Phi(r) - \Phi_0))] \tag{1.28}$$

$$+ \Sigma[\frac{A_{ij}}{r_{ij}(r)^{12}} - \frac{B_{ij}}{r_{ij}(r)^6} + \frac{q_i q_i}{\epsilon_i \epsilon_i r_{ij}(r)}] \tag{1.29}$$

$$+ \Sigma \begin{cases} 0, & \text{if } r \geq s \cdot r_0 \\ k_{rep}(s^2 {r_0}^2 - r^2)^2, & \text{if } r < s \cdot r_0 \end{cases} \tag{1.30}$$

where E_{bond}, E_{angle}, E_{dihed}, $E_{\text{non-bon}}$, and E_{repel} are potentials for bond lengths, bond angles, dihedral angles, non-bonding interactions, and Pauli repulsion, and k_{bond}, k_{angle}, k_{dihed}, k_{rep}, A_{ij} and B_{ij} are semi-empirical parameter specific for the individual atoms and interactions. $b(r)$, $\theta(r)$, $\Phi(r)$, $r_{\text{ij}}(r)$ and b_0, θ_0, Φ_0, r_0 are the bond length and angle, dihedral angle and atomic distance at Cartesian coordinate r and at equilibrium, respectively. $\frac{q_{\text{i}} q_{\text{i}}}{e_{\text{i}} e_{\text{i}} r_{\text{ij}}}$ is the Coulomb addition of the Lennard-Jones potential for non-bonding interactions, n an integer, and s a fraction between 0 and 1.

Experimentally derived restraints are also added as potential functions to the force field and are usually introduced with upper (u_{ij}) and lower limits (l_{ij}). A common potential for NOE contacts is

$$E_{NOE} = \Sigma \begin{cases} k_{NOE}(r_{ij} - u_{ij})^2, & \text{if } r_{ij} > u_{ij} \\ 0, & \text{if } l_{ij} \leq r_{ij} \leq u_{ij} \\ k_{NOE}(l_{ij} - r_{ij})^2, & \text{if } r_{ij} < l_{ij} \end{cases} \tag{1.31}$$

with E_{NOE} and k_{NOE} being the potential and the scaling factor for the NOE energy term. The equation of Newtonian motion derived from the potential energies are solved by numerical integration methods like Monte-Carlo, typically in time steps much smaller than the fastest local motion. High potential barriers might trap the simulation in an artificial conformation, therefore additional kinetic energy terms are often introduced into the calculation. In this Simulated Annealing (SA) approach the structure is heated up *in-silico*, which allows to sample a broad range of conformation, followed by a step-wise cooling process (Figure 1.12). As starting point for the calculation a model template or randomized atomic coordinates (*ab initio*) can be used, whereby the latter is more reliable for unknown systems. The calculation is repeated several

tens or hundreds of times, leading to an ensemble of structural conformations with lowest energy terms. One or several low-energy structures are subsequently used as templates for structure refinement procedures.

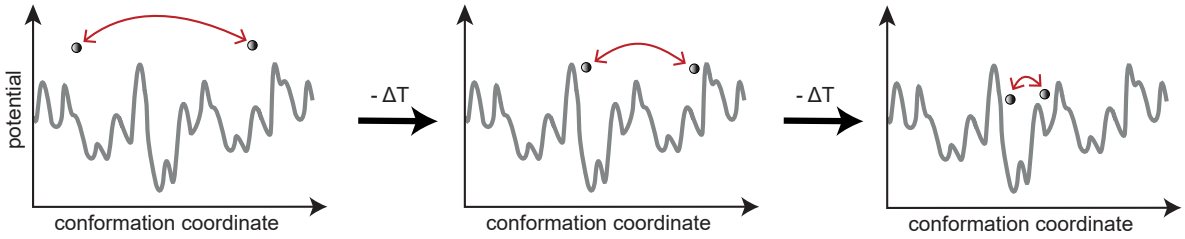


Figure 1.12: Simulated Annealing. Force field and restraint potentials are forming a complex energy landscape with several local minima along the coordination axis. High temperatures allow to overcome energy barriers and to sample all possible conformations (spheres). By stepwise cooling (-ΔT) the simulation is able to find the global minimum of the energy landscape, which represents the conformation best supported by the introduced restraints.

2 Aims of the Thesis

The aim of this thesis is to demonstrate applications of (novel) NMR techniques on different biological systems, which are a catalytically active DNA and (membrane-associated) proteins.

The specific aims in respect to DNA-mediated catalyses are to:

- generate the first high-resolution insights of the 10-23 DNAzyme
- characterize the different states occurring during the catalytic cycle
- understand the role of metal ion cofactors for structure and function of the 10-23 DNAzyme

In respect to membrane-associated protein systems the aims are to:

- investigate the interaction network between epidermal growth factor receptor and potential modulators
- improve selective labelling strategies for NMR studies of large protein systems
- study ligands of the melanocortic-4 receptor and the role of lipid bilayers in receptor activation

3 The 10-23 DNAzyme - A RNA-cleaving DNA enzyme

This part of the thesis summarizes studies performed with the 10-23 DNAzyme and reflects, in parts, results published in Rosenbach *et al.* (2019) [133] and Rosenbach *et al.* (submitted). Experiments and analysis performed by co-workers are mentioned in the Material and Methods section.

3.1 Introduction

DNAzymes (or deoxyribozymes) are artificial single-stranded DNA molecules, which are capable to catalyse a wide variety of reactions, such as DNA or RNA cleavage [28, 139] and ligation [39, 50], photoreversion of thymine dimers [32], as well as peptide modifications [29, 35, 124, 154, 180]. In contrast to their ribozyme analogs, DNAzymes have not been identified in nature yet, instead they are isolated from a randomized library by *in vitro* selection (Figure 3.1a) [20]. DNAzymes can have applications in biosensing, diagnostics, synthesis [11, , 152, 165, 181, 182] and gene silencing, where the latter makes RNA-cleaving DNAzymes a promising therapeutic for diseases such as amyloidosis, cancer, and virus infections [10, 33, 55, 66, 79, 188]. Despite their various functions and many performed studies, the mechanisms of DNAzyme-mediated catalysis are poorly understood, especially due to lack of structural data. Extensive effort in the past two decades led to only two crystal structures comprising the RNA-ligating 9DB1 [3] and the RNA-cleaving 8-17 DNAzymes [88].

The most prominent RNA-cleaving DNAzyme is the 10-23 DNAzyme (Dz) [140], which is depending on divalent metal ions and shows under optimal conditions comparable RNA-cleaving activity as the hammerhead ribozyme under native conditions [19]. Dz consists of a catalytic loop of 15 nucleotides flanked by two target-specific binding arms, which are capable to hybridize with a complementary RNA sequence with high specificity, leaving one purine ribonucleotide in the middle unpaired [25]. The addition of divalent ions leads to RNA cleavage between the unpaired ribonucleotide and its 3' neighbour, which has to be exclusively a pyrimidine (Figure 3.1b). The length of the Dz's binding arms can be designed to match a melting temperature, which allows to hybridize with its RNA target but to dissociate after cleavage [159, 160]. This enables multiple turnover reactions and makes Dzs an interesting therapeutic by targeting disease-related RNA. However, despite it cleaves RNA with high efficiency *in vitro*, in-cell activity is drastically lowered [170]. One hypothesis for this behaviour is the interplay between Dz and endogenous metal ion cofactors. Several mutations and modifications of the Dz were performed, leading to gain, reduction or loss of function (see [134] for a review). Despite many studies were already performed to unveil Dz's function, no definite conclusion could be drawn yet, due to lack of high-resolution structural data. Solution NMR spectroscopy reveals to be a

suitable tool to analyse the structure and function of Dz, as well as the role of ion cofactors, and may lead to knowledge-based engineering of new therapeutic relevant Dz variants.

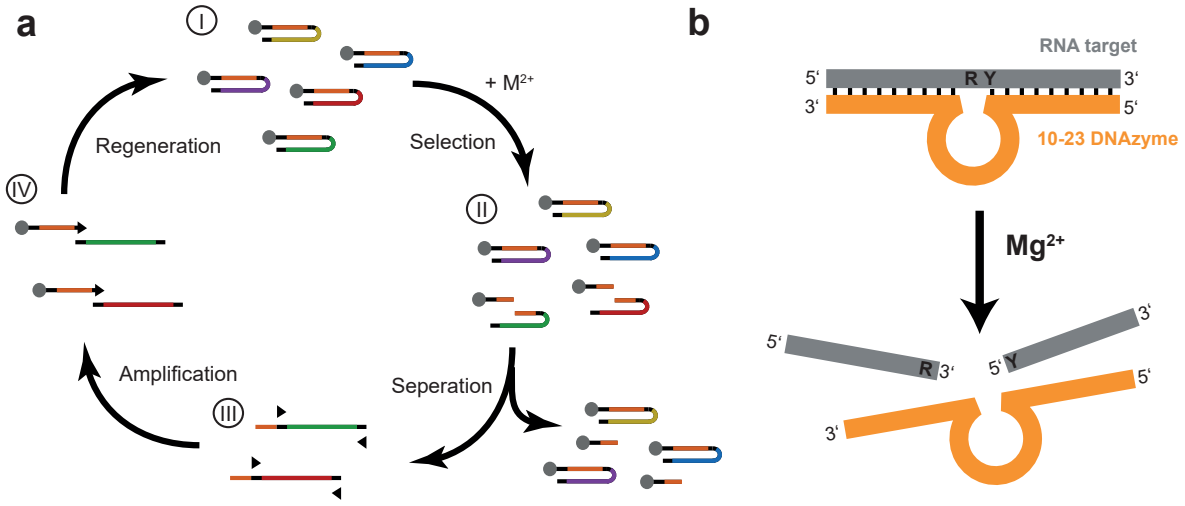


Figure 3.1: *In vitro* selection of RNA-cleaving DNAzymes. (a). (I) A library of 10^{12} randomized DNAzyme sequences (coloured regions) are fused to their RNA target (orange region). (II) After addition of divalent ions (M^{2+}), competent DNA sequences cleave the RNA. Incompetent DNAzymes and cleavage products are immobilized via a biotin resin (grey sphere). (III) Active DNAzyme sequences are amplified by primers (arrows) hybridizing with conserved primer binding sites (black regions). (IV) Amplified DNAzymes are fused with their biotinylated targets by polymerase chain reaction and the cycle is started again until the most active sequences are isolated. (b) The 10-23 DNAzyme (Dz) has a 15nt catalytic loop and cleaves a hybridized RNA target exclusively between purine (R) and pyrimidine (Y) residues in presence of divalent ions.

3.2 Results and Discussion

3.2.1 DNAzymes are well-accessible for NMR studies

The Dz used in this study is a 33 nucleotide long oligomer which is targeting mRNA of the Prion protein, the key protein involved in Creutzfeldt-Jakob disease. Although this DNAzyme shows high catalytic activity *in vitro*, no in-cell activity could be observed [170]. The nucleotides of the catalytic loop region are numbered by indices from 1 to 15, while the (deoxy)ribonucleotides of the binding arms are numbered with indices from -9 to -1 (or from +1 to +9) when they are laying upstream (or downstream) of the loop sequence and the unpaired guanidine ribonucleotide at the cleavage site is labeled with index 0 (Figure 3.2a). The palindromic sequence of loop nucleotides 2-7 (GCTAGC) causes an artificial four-way junction in a published crystal structure [113], and may have a similar effect at µM concentrations used in NMR studies. Mutations at position 5 can disrupt this palindrome, but also have an impact on Dz's catalytic activity (*vide infra*). Most of the performed NMR experiments rely on the signals of non-exchangeable

protons and their numbering convention is shown in Figure 3.2b. Using high-field spectrometer, homonuclear NOESY and TOCSY spectra could be recorded in good resolution for several states of the Dz catalytic cycle, as shown exemplarily for Dz in complex with its RNA target (Figure 3.2c). The crosspeaks cluster in different regions of the 2D spectra and strongly overlap in certain parts. This is a consequence of the low chemical shift difference between protons of different nucleotides, as well as of the size of the complex. However, TOCSY crosspeaks of pyrimidine protons H6 and H5 are well-separated and can be used as probes to follow changes of the system (Figure 3.2c, boxed area). To prevent Dz-mediated cleavage of the RNA in presence of Mg^{2+} during acquisition, a fluorine substitution at 2' position of rG_0 was introduced (referred to as $RNA^{2'F}$). Chemical shift perturbations (CSP) due to the ^{19}F modification could only be observed for the adjacent rU_{-1}, suggesting that it has no significant impact on the Dz structure (Figure 3.2d).

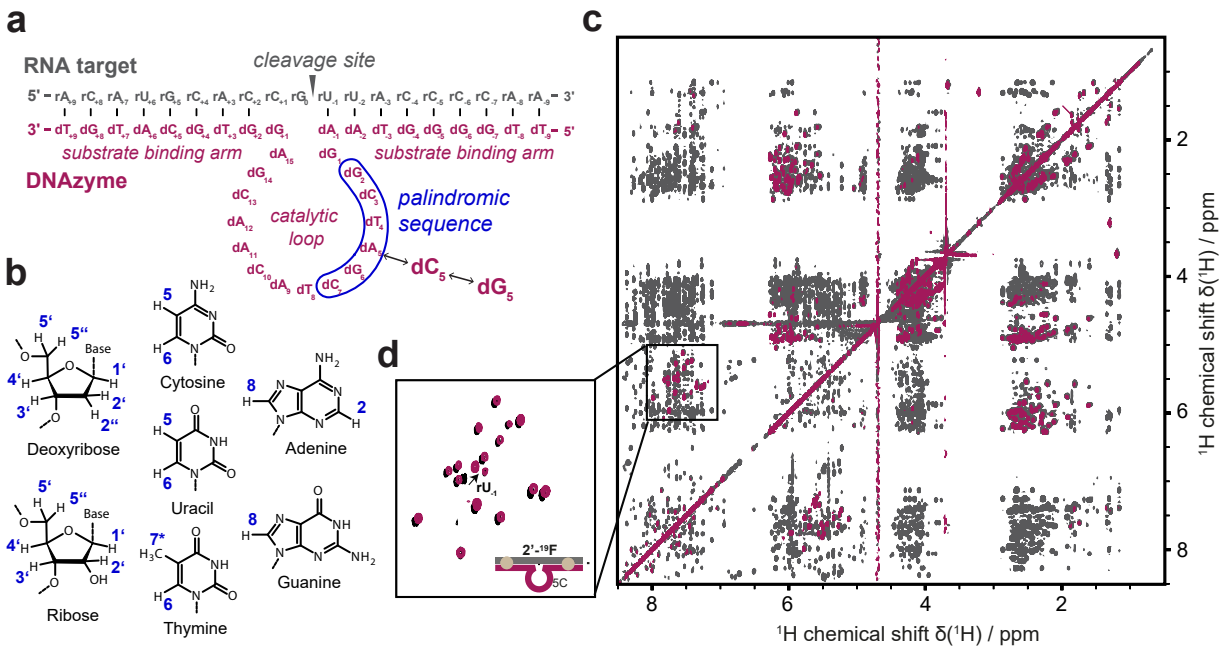


Figure 3.2: Complex of Dz and RNA is accessible via solution NMR spectroscopy. (a) Sequence of PrP-mRNA targeting Dz (purple) and the respective RNA target (grey). The catalytic loop is numbered from 1 to 15, while the binding arms are numbered from -9 to +9 in respect to their position from the cleavage site. At the blue indicated sequence two DNAzymes (Dz^{5A}) are able to hybridize leading to a catalytic irrelevant X-ray structure [113]. The self-hybridization can be disrupted by the mutants Dz^{5C} and Dz^{5G}. (b) Lewis structures of (deoxy)ribose and nucleobases are shown and the non-exchangeable protons are indicated. (c) [^{1}H,^{1}H]-NOESY (black) and [^{1}H,^{1}H]-TOCSY spectra (purple) of Dz^{5C}:$RNA^{2'F}$ in 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$ at 37 °C show well-resolved areas as well as regions with high signal overlap. (d) Crosspeaks of U and C base protons are well-separated in the TOCSY spectrum (purple, boxed area of c) and serve as probes for structural changes. A ^{19}F substitution of the 2'-OH group at the scissile bond (rG_0) protects the RNA from cleavage (referred to as $RNA^{2'F}$). Only the adjacent nucleotide rU_{-1} shows ^{19}F-induced CSP compared to a spectrum of Dz^{5C}:RNA (black), suggesting that no structural changes are cause by the fluorination.

Different states of the Dz's catalytic cycle could be detected by solution NMR

TOCSY crosspeaks of pyrimidines were used as a fingerprint to characterize different states of the Dz's catalytic cycle, since their chemical shift is highly sensible to changes of the chemical environment (Figure 3.3). Single-stranded Dz and RNA (Figure 3.3 I+II) have significant different peak patterns as compared to the Dz:RNA complex, where the signals belonging to the hybridized binding arms showing a characteristic upfield shift (Figure 3.3 III-V). Complexed Dz:RNA shows more TOCSY crosspeaks in the pyrimidine region than pyrimidines in the sequence (Figure 3.3 III), which is caused by multiple, slowly interchanging conformational states. Addition of cations such as Na^+ is stabilizing one conformation and leads to a homogenous spectrum (Figure 3.3 IV). By addition of Mg^{2+} the unstabilized RNA gets cleaved by the Dz resulting in a still hybridized complex of Dz and cleaved RNA (Dz:products) at NMR concentrations (Figure 3.3 V). Interestingly, the product complex shows slowly interchanging states, comparable to the pre-catalytic complex in absence of cations (Figure 3.3 III). The product complex is melted above 60 °C (Figure 3.3 VI) and shows similar signal pattern as single-stranded Dz or melted pre-catalytic complex (Figure A1a). Pyrimidine fingerprints are also useful to identify sample integrity (Figure A2) and to follow the effects of ion cofactor binding (*vide infra*).

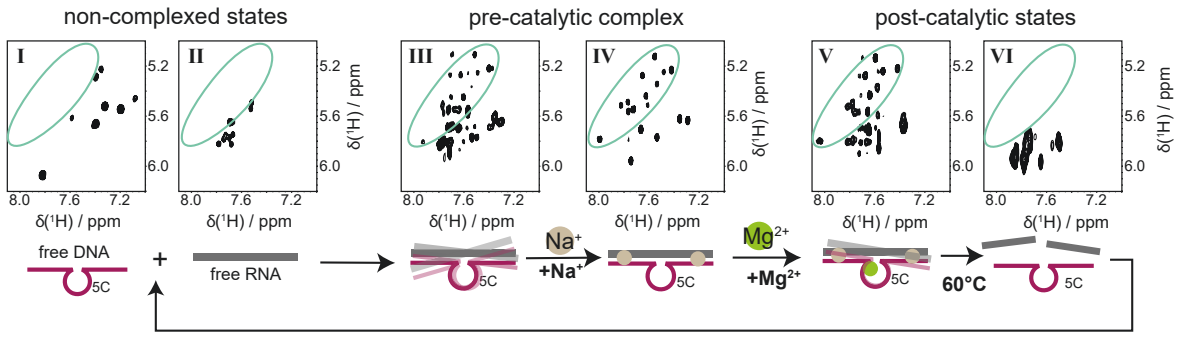


Figure 3.3: States of Dz's catalytic cycle. Pyrimidine TOCSY crosspeaks serve as fingerprints for the different states of Dz^{5C} and RNA. Peak positions of single-stranded Dz (I) and RNA (II) significantly differ from complexed states (III-V), where at the latter the signals belonging to the hybridized arms show characteristic upfield shifts (green ellipse). Dz:RNA in absence of cationic cofactors (III) is slowly interchanging between different conformation, which results in multiple signals in the spectra. Addition of 100 mM NaCl stabilizes the complex (IV), while addition of Mg^{2+} leads to RNA cleavage if the target is not ^{19}F-stabilized (V). Cleaved RNA is still hybridized with Dz under NMR conditions and exists in different conformations leading to a spectrum comparable to state III. The product complex is melted at 60 °C (VI) and shows a peak pattern similar to single-stranded RNA and Dz at this temperature (compare to Figure A1).

Mutations at position 5 affect structure and activity of Dz

Previous X-ray crystallographic studies of Dz have failed due to dimerization by the palindromic sequence within the catalytic loop of the original 10-23 DNAzyme sequence, referred to as Dz^{5A} (Figure 3.1a) [113]. Mutation of the adenosine at loop position 5 to cytosine (Dz^{5C}) or guanine

(Dz^{5G}) disrupts the palindrome, but leads to reduction or loss of activity, respectively [186]. The activity of the different Dz mutants can be determined using denaturing polyacrylamide gel electrophoresis (SDS-PAGE) or a Förster resonance energy transfer (FRET) based technique (Figure 3.4a+b). The original Dz^{5A} sequence is able to fully cleave its RNA target within 30 min in a single-turnover experiment at 37 °C and in presence of 1 mM $MgCl_2$. Dz^{5C} cleavage rates are slower and complete cleavage is obtained after ∼2-3 h under the same condition, while Dz^{5G} does not show any activity in the performed assays. The binding arm signals of all three complexes Dz^{5A}:$RNA^{2'F}$, Dz^{5C}:$RNA^{2'F}$ and Dz^{5G}:$RNA^{2'F}$ present similar TOCSY peak pattern for the binding arms, while most peaks belonging to the loop region differ significantly (Figure 3.4c). Notably, the peak position of rU_{-1}, which is directly adjacent to the scissile bond, differs strongly in all three mutants. This suggests, that all three Dz sequences are forming similar complexes with RNA, but different loop conformations lead to severe structural variations at the cleavage site, which is affecting Dz's mode of action. Addition of 1 mM $MgCl_2$ leads to strong signal broadening for the loop region of Dz^{5A}:$RNA^{2'F}$, which is an effect due to intermediate chemical exchange processes upon Mg^{2+} binding (Figure 3.4 d). This renders NMR studies of Dz^{5A}:$RNA^{2'F}$ in the most relevant pre-catalytic Mg^{2+}-bound state very difficult, therefore, the slower cleaving Dz^{5C} mutant was used for a more detailed NMR study.

Resonance assignment strategy for Dz and Dz:RNA

The key feature of proton resonance assignment for DNA:RNA complexes lays in the 2'OH substitution of ribose, which leads to downfield shifted 2' protons (4-5 ppm) compared with 2' and 2" deoxyribonucleic protons (1-3 ppm). This allows to assign DNA independently from the RNA strand, since only DNA signals can be found in this part of the spectrum (Figure A2a). Proton resonance assignment of nucleic acids is usually performed by exploiting short-range NOE contacts between one nucleotide and its 5' neighbour (see [176] for a review on nucleic acid assignment). The base protons H6 (in pyrimidines) or H8 (in purines) form strong NOE contacts with protons H1', H2'/H2", and H3' of its own sugar moiety and slightly weaker with the same protons of its 5' neighbour (Figure 3.5a+b). By searching for matching NOE patterns sequentially following spin systems can be identified, while TOCSY crosspeaks between H1', H2'/H2", and H3' were utilized to identify which protons belong to the same spin system. Comparison of NOESY and TOCSY crosspeaks between H6 and H5 (or H7*) were used to identify if one spin system belongs to a pyrimidine nucleotide, and are helpful landmarks to assign a sequence of spin systems to the nucleic acid sequence. Some sequential neighbours were difficult to identify due to high signal overlap or weak NOE intensity, however additional indicators were used, like base-base contacts of adjacent nucleotides (Figure A3d). Full Dz assignment greatly reduced the number of unassigned crosspeaks and eased the assignment of the RNA target using the same strategy. Additional (deoxy)ribose protons as well as H2 of adenine could be identified afterwards due to their relative distance to already assigned protons. Usually, H2' of deoxyribonucleotide appear more shielded as H2" and builds more intensive NOESY crosspeaks with its own base proton H6 or H8. However, also relative NOE build-up rates of H1'-H2' and H1'-H2" were compared to identify protons unambiguously (see

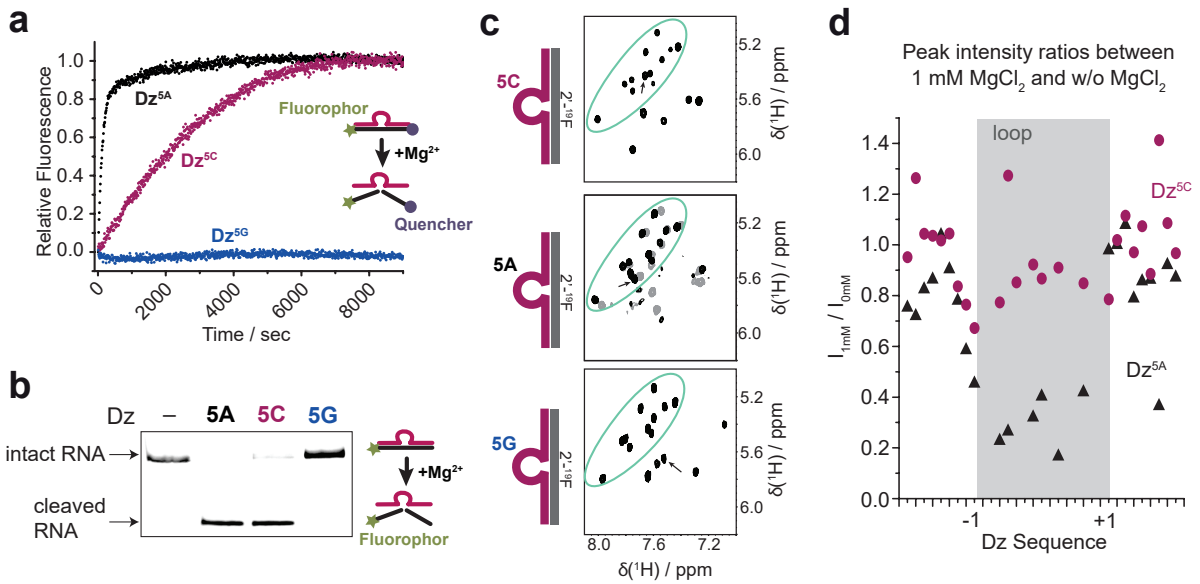


Figure 3.4: Mutations at loop position 5 have an impact on Dz structure and activity. (a) In an assay with RNA tagged with an fluorophore and a quencher on separate termini cleavage activity can be followed by an increase of fluorescence. Dz^{5A} completely cleaves RNA within 30 min in a single turnover experiment at 37 °C and with 1 mM $MgCl_2$, while Dz^{5C} is approximately 5-fold slower at same conditions. (b) Denaturing polyacrylamide gel electrophoresis (SDS-PAGE) performed with fluorescein-labelled RNA (RNA^{FAM}) after 12 h of cleavage reaction shows only activity for Dz^{5A} and Dz^{5C}, but none for Dz^{5G}. (c) The binding arms of the variants Dz^{5C} (top), Dz^{5A} (middle), and Dz^{5G} (bottom) in complex with its RNA target have the same structure, while the loop region and the signal for rU_{-1} (indicated by arrow) strongly differs. Black spectra were recorded in presence of 100 mM NaCl and 1 mM $MgCl_2$, while the grey spectrum was recorded without $MgCl_2$. (d) Peak intensities of Dz^{5A}:$RNA^{2'F}$ are strongly lowered in presence of 1 mM $MgCl_2$ compared to a spectrum recorded in absence of $MgCl_2$. This renders NMR studies of a pre-catalytic state of Dz^{5A}:$RNA^{2'F}$ very challenging, therefore the slower cleaving Dz^{5C} was used for most experiments (see c, top and middle spectra).

also Figure 3.10d). Only a few exchangeable protons could be assigned, due to high exchange rates with the solvent at 37 °C. Experiments at lower temperatures improves their visibility, but severly broadens the other peaks in the spectrum. Due to the high overlap only a few sugar protons at postion 4' and 5' could be assigned. Carbon atoms of Dz^{5C} were assigned using 2D [^{1}H,^{13}C]-HSQC-TOCSY and 3D [^{1}H,^{13}C]-HSQC-NOESY experiments with an isotope-enriched ^{15}N,^{13}C-Dz^{5C}:$RNA^{2'F}$ complex. However, due to the high signal overlap only C1' nuclei could be fully assigned (Figure 3.5c). Assigned resonances for carbons and non-exchangeable protons of Dz^{5A}:$RNA^{2'F}$ can be found in Table A1+A2.

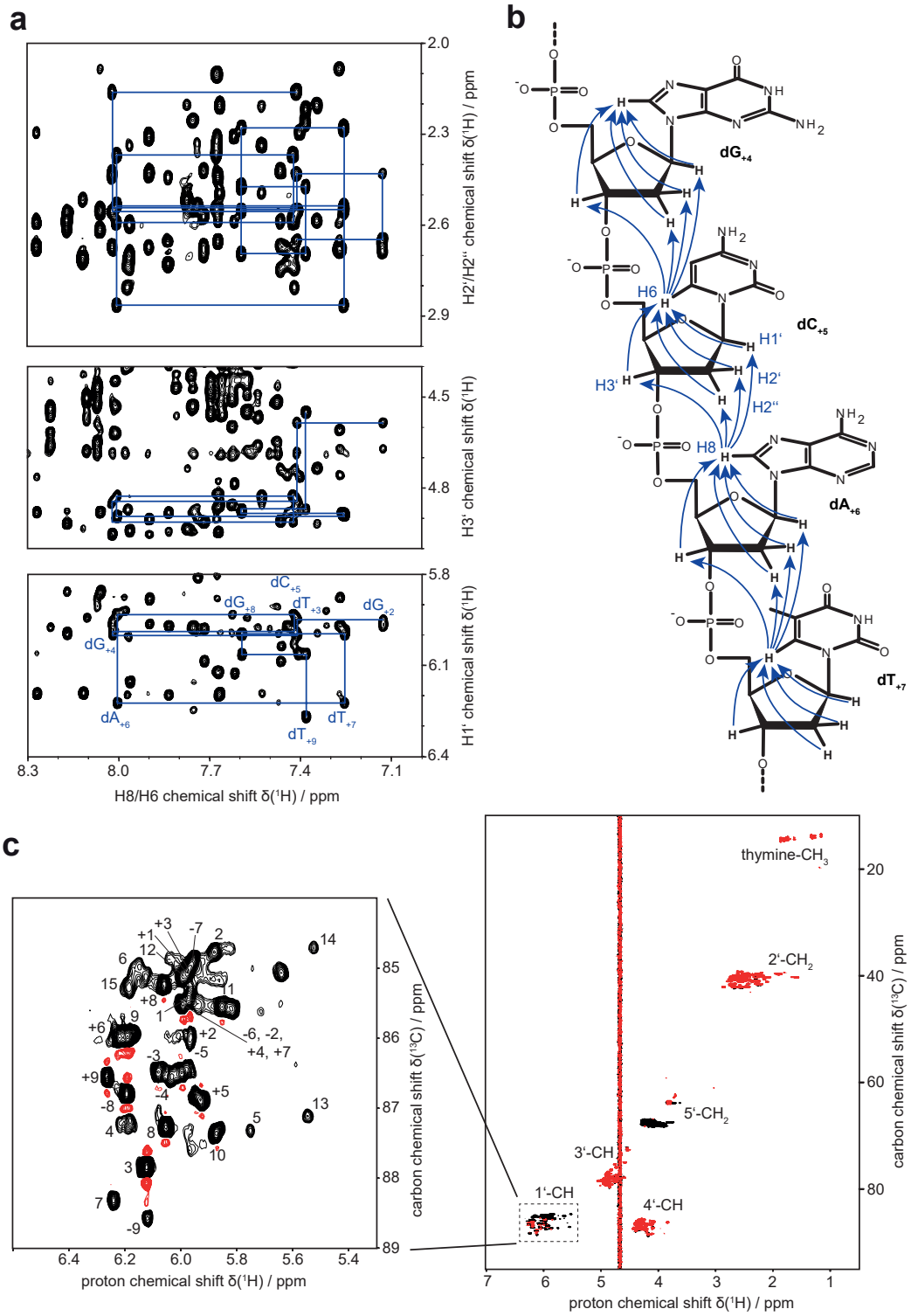


Figure 3.5: Resonance assignment of Dz^{5C}:RNA$^{2'F}$. (a) In most nucleic acid structures nucleobase protons H6 or H8 of a nucleotide are in close proximity to the sugar protons H1', H2', H2", and H3' of their 5' adjacent neighbours. The figure show sections of a [^{1}H,^{1}H]-NOESY spectra with H6/H8 resonances in direct and H2'/H2" (top), H3' (middle), and H1' (bottom) resonances in indirect dimension. Blue lines indicate the sequential assignment of Dz sequence $G_{+2}T_{+3}G_{+4}C_{+5}A_{+6}T_{+7}G_{+8}T_{+9}$. (b) The scheme shows the NOE contacts between adjacent nucleotides, which are used for sequential assignment. (c) [^{1}H,^{13}C]-HSQC spectrum of Dz^{5C}:RNA$^{2'F}$ shows highly overlapping regions for ribose methylenes at position 3', 4', and 5' (right). CH at position 1' show the largest dispersion and could be fully assigned (left).

3.2.2 Mono- and divalent cations have a severe impact on Dz's structure and function.

Influence of divalent cations on Dz's structure and activity

The catalytic propensity of the Dz is depending on divalent cations, but it is unclear if they are important for structural stability or if they are directly involved in the catalysis. Previous studies have investigated the effect of different divalent metal ions on the 10-23 DNAzyme-mediated RNA cleavage and it was found that the metal ions enhance the reaction in the following order: Mn^{2+} > Mg^{2+} > Ca^{2+} >> Ba^{2+} in tris(hydroxymethyl)aminomethane (Tris) buffer [161]. Increasing $MgCl_2$ concentrations can accelerate RNA cleavage of both Dz^{5C} and Dz^{5A}, but the first one is 10-fold less affected (Figure 3.6a+b). NMR titration experiments with Dz^{5C}:$RNA^{2'F}$ and Dz^{5A}:$RNA^{2'F}$ shows that nearly all pyrimidine signals are influenced by increasing amounts of $MgCl_2$. In the Dz^{5C}:$RNA^{2'F}$ complex strong CSP could be observed for all pyrimidine signals in the complex, despite loop positions 5 and 10. For Dz^{5A}:$RNA^{2'F}$ the same behaviour is observed for the binding arms, but nucleotides close to the cleavage site and the loop region are strongly affected by peak broadening induced by intermediate exchange rates (Figure 3.6c-e). Mg^{2+} dissociation coefficients (K_D) for the binding arms in both mutants are between 2-20 mM without showing any site-specific preference (Figure A4). However, the maximum Mg^{2+}-induced CSP in

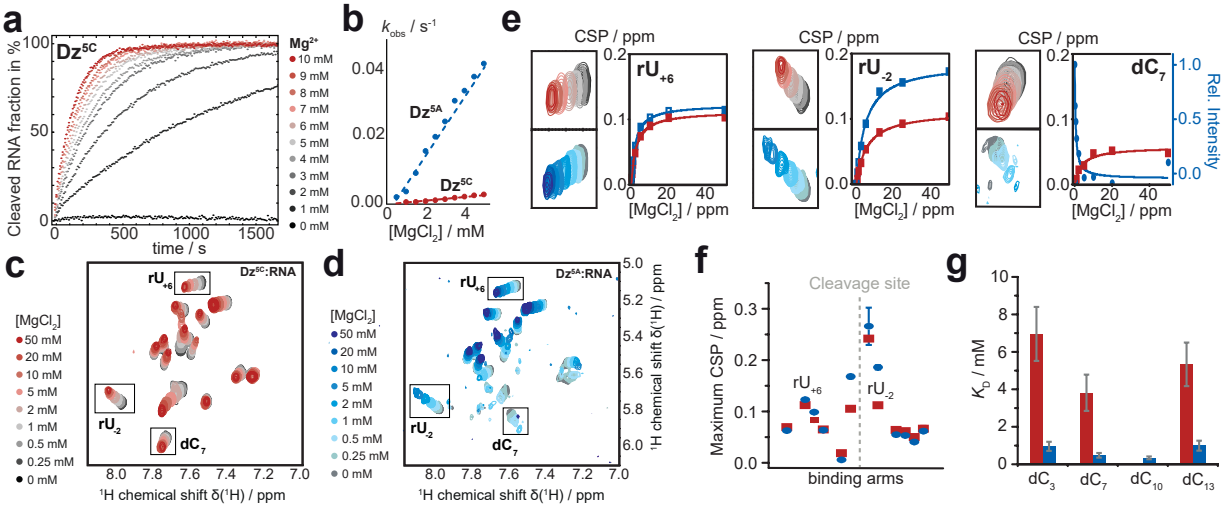


Figure 3.6: Mg^{2+} influences structure and activity of Dz^{5C} and Dz^{5A} differently. (a) Activity of Dz^{5C} is increased linearly with $MgCl_2$ concentration in presence of 100 mM NaCl. (b) The increase is 10-fold higher for Dz^{5A} than for Dz^{5C}. (c+d) NMR titration experiments with $MgCl_2$ in presence 100 mM NaCl shows identical shifts for binding arm signals of Dz^{5C}:$RNA^{2'F}$ (c) and Dz^{5A}:$RNA^{2'F}$ (d). The cytosines 3, 7 and 13 of the catalytic loop of Dz^{5C}:$RNA^{2'F}$ presents also Mg^{2+}-induced CSP, but dC_5 and dC_{10} seem to be unaffected. For Dz^{5A}:$RNA^{2'F}$, all cytosines of the loop region, as well as nucleotides close to the cleavage site, show CSP and line-broadening due to intermediate chemical exchange rates. (e-g) K_D values and maximal shifts are shown for selected nucleotides of Dz^{5C}:$RNA^{2'F}$ (red). In (e), example peak shifts and curve fits for binding arms, cleavage site and catalytic loop (rU_{+6}, rU_{-2}, and dC_7, respectively) are shown. Mg^{2+} binding induces identical effects on the binding arms (e left), while the maximum CSP for signals close to the cleavage sites are larger for Dz^{5A} than for Dz^{5C} (e middle, and f). Due to the severe line-broadening the maximum shift for loop signals of Dz^{5A}: $RNA^{2'F}$ could not be determined (e right), therefore, K_D values for Dz^{5A} loop signals were estimated by their intensity decrease, despite for dC_{13}. For the loop region of Dz^{5A}:$RNA^{2'F}$ all K_D values lay below 1 mM and are significantly lower than for Dz^{5C}.

Dz^{5A} are higher for nucleotides adjacent to the cleavage site compared to Dz^{5C} (Figure 3.6e+f). Due to the peak broadening, CSPs belonging to the loop region of Dz^{5A} could not be completely analysed despite of cytosine 3, therefore K_D values of the other nucleotides were estimated from the decay of peak intensity. All pyrimidine signals in the catalytic loop of Dz^{5A} have a K_D of 1 mM or lower, while their counterparts in Dz^{5C} have K_D values comparable to the binding arms or are not affected at all (Figure 3.6e+g). CSP and line broadening follow changes in the chemical environment, and therefore do not allow to distinguish between direct Mg^{2+} binding and Mg^{2+} binding-induced structural changes. However, most of the Mg^{2+}-induced events on Dz^{5C}:$RNA^{2'F}$ can be located close to the cleavage site (*vide infra*, Figure 3.14e+h).

Influence of monovalent cations on Dz's structure and activity

Unlike the hammerhead ribozyme [137], Dz is not active in the sole presence of monovalent cations, but monovalent ions have a clear effect on its catalytic efficiency. In this respect, we observed that increasing amounts of NaCl significantly lower the Mg^{2+}-induced RNA cleavage for Dz^{5A} and Dz^{5C} (Figure 3.7a). To unveil the role of monovalent cations in Dz-mediated catalysis, [^{1}H,^{1}H]-TOCSY experiments of complexed Dz were performed in absence and presence of NaCl (Fig. 3.7b+c). Dz^{5A}:$RNA^{2'F}$ in absence of NaCl shows more peaks than expected in the pyrimidine region, suggesting multiple slow interchanging states for the complex. Addition of 100 mM NaCl leads to one (or multiple fast interchanging) states, which can be seen by a single peak per pyrimidine nucleotide in the resulting spectrum. Further addition of 1 mM $MgCl_2$ leads to the shifts and line broadening effects, which were already shown in Figure 3.4c+d, but interestingly addition of 400 mM NaCl instead of $MgCl_2$ results in similar changes, which leads to the assumption that Na^+ and Mg^{2+} are occupying similar binding sites and inducing the same structural transitions (Figure 3.7c). The interaction of Mg^{2+} with the Dz^{5A}:$RNA^{2'F}$ complex was also analysed using isothermal titration calorimetric (ITC) measurements, which is an approach for determining the heat energy associated with a molecular interaction (Figure A5). Therefore, a $MgCl_2$ titration to the preformed complex Dz^{5C}:$RNA^{2'F}$ complex pre-equilibrated in the absence and or presence of 100 mM NaCl was carried out and the changes in heat energy were measured. Assuming a model with multiple Mg^{2+} binding sites in a n:1 binding model, the ITC data shows that Mg^{2+} binding to the complex occurs with high µM binding affinities (K_D = (178 ± 14) µM) and that binding is about 2-fold weaker in the presence of 100 mM NaCl (K_D = (430 ± 30) µM) with approximately four cooperative binding sites. To further analyse this effect, [^{1}H,^{1}H]-TOCSY spectra of Dz^{5C}:$RNA^{2'F}$ complex in absence of NaCl but with increasing $MgCl_2$ concentration were performed (Figure 3.7d) and compared with the corresponding titration performed in presence of 100 mM NaCl (Figure 3.6c). Overall, the K_D values derived from CSP data of the pyrimidine crosspeaks are also decreased by around half in the absence of NaCl (from 1.9-2.8 mM to 1.1-1.6 mM), which agrees with the results for Dz^{5A} recorded by ITC and strengthens the assumption that Mg^{2+} can be displaced by higher amounts of monovalent Na^+. The chemical shift in absence of NaCl and low amounts of $MgCl_2$ (0.25-1 mM) are similar to the one in presence of solely 100 mM NaCl, which again suggests that Na^+ indeed binds to similar binding sites and induces similar structural changes as Mg^{2+} (Figure A6). It

is noteworthy, that the peaks for the additional state in absence of NaCl, as seen exemplarily for rU_{-4}, is disappearing during $MgCl_2$ titration, showing that low Mg^{2+} concentrations have an effect on homogenising the multiple states as seen for increasing amounts of Na^+ (Figure 3.7c, left). Taken together the data derived from NMR and ITC, the inhibiting effect of Na^+ can be explained by a lowered binding affinity of Mg^{2+} in presence of NaCl. Catalytic active, divalent Mg^{2+} ions are displaced by inactive, monovalent Na^+, but both ions inducing similar structural changes on the Dz:RNA complex. This may lead to the conclusion, that mono- and divalent cations are able to stabilize an active conformation of Dz:RNA, but only divalent cations can induce cleavage either directly or by coordinating essential functional groups towards the cleavage site.

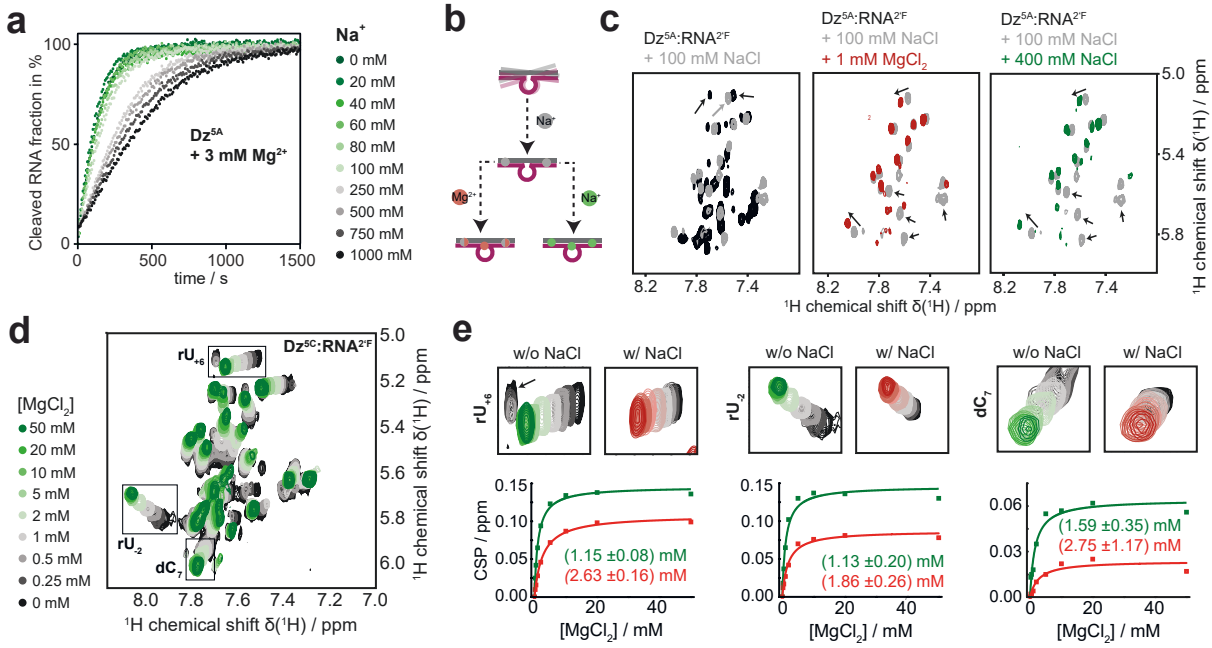


Figure 3.7: Monovalent Na^+ and divalent Mg^{2+} share the same binding sites. (a) Increasing amounts of NaCl lower the cleavage activity of Dz^{5A} in presence of 3 mM $MgCl_2$. (b) Scheme of the performed NMR experiments in c. Dz^{5A}:$RNA^{2'F}$ were first titrated with 100 mM NaCl, which leads to a homogenized complex. Further titration with either 1 mM $MgCl_2$ (red) or 400 mM NaCl (green) yield similar structural effects. (c) Dz^{5A}:$RNA^{2'F}$ in absence of ions shows multiple states in slow exchange in a TOCSY spectrum (black spectrum). Addition of 100 mM NaCl stabilizes a single complex state (grey spectrum). Further addition of 400 mM NaCl leads to similar chemical shifts and peak disappearance (green spectrum) as does addition of 1 mM $MgCl_2$ (red spectrum). (d) NMR titration experiment with Dz^{5C}:$RNA^{2'F}$ and increasing $MgCl_2$ concentrations was performed in absence of NaCl. (e) Enlarged sections of peaks belonging to rU_{+6}, rU_{-2}, and dC_7 of Dz^{5C}:$RNA^{2'F}$ in absence and presence of NaCl indicated as boxes in d and Figure 3.6c, respectively, show similar shifts. Binding arm signal rU_{+6} (left) shows the same disappearance of the additional state during the addition of Mg^{2+} as seen for the addition of Na^+ (see c, left). The starting chemical shift of the titration in presence of NaCl resembles the chemical shift in absence of NaCl and 0.5 to 1 mM $MgCl_2$, but the end position are similar. Therefore, the maximum CSP in presence of 100 mM NaCl is significantly reduced. K_D values derived from the titration in presence of NaCl are lower by up to a factor of two as compared to the titration in absence of NaCl.

3.2.3 Structure determination of Dz^{5C}:$RNA^{2'F}$ by solution NMR.

To gain insights into the mechanistic role of the different parts of the 10-23 DNAzyme, structural relevant NMR data was recorded with a pre-catalytic Dz state. Dz^{5C}:$RNA^{2'F}$ was used for the structure determination, since the 5C mutant showed more favourable NMR properties in presence of Mg^{2+} and the stabilizing 2'-^{19}F modification protects RNA from cleavage during data acquisition (see section 3.2.1). To avoid different conformational states, the measurements were performed in presence of 100 mM NaCl and 1 mM $MgCl_2$ at 37°C. Conventionally, structure calculation by solution NMR depends on semi-experimentally derived angular restraints and an abundancy of long-range NOE contacts. Because NMR spectra of larger nucleic acids like the Dz^{5C}:$RNA^{2'F}$ complex show many unresolved peaks, long-range NOE contacts are not trivial to find without prior structural knowledge.

Long-range restraints by paramagnetic spinlabels.
Initial long-range restraints could be derived by side-specific introduction of paramagnetic spinlabels and analysis of signal loss of all accessible protons due to the paramagnetic relaxation enhancement (PRE) (Figure 3.8). Protons in close proximity to the paramagnetic centre show higher relaxation rates than these ones more distant, which allows to use the relative intensity loss of crosspeaks as loose distant restraints up to 35 Å. Spinlabels were introduced at positions dT_{-7}, dT_8, and dT_{+8}, which are located at the end of the binding arms and the centre of the catalytic loop. Three sequences were ordered with a 5-Ethynyl-2'-deoxyuridine (EdU) residue instead of a thymidine at the respective position and the ethynyl group was coupled with a paramagnetic 2,2,6,6-tetramethyl-1-piperidinyloxy (TEMPO) spinlabel via copper click-chemistry (see Material and Methods). The functionalized sequences show similar cleavage efficiencies as the original sequences (Figure A7). NOESY and TOCSY spectra of the spinlabelled DNAzymes in complex with its stabilized target were acquired in absence and presence of ascorbic acid, a reactant for reducing the spinlabel into a PRE-inactive state (Figure 3.8a-c). Peak intensity ratios between both spectra were analysed for all accessible crosspeaks and mean values per proton were calculated from crosspeaks belonging to the same proton. Figure 3.8d-f shows site-specific signal reduction in the three different samples, which subsequently were translated into low-resolution distance restraints. In total 126 PRE restraints could be identified.

Unique distance restraints by ^{19}F-derived experiments.
^{19}F substitutions are suitable probes in nucleic acid NMR spectroscopy, since synthesis is cost-efficient, substitutions are nearly invasive on the structure and ^{19}F has a gyromagnetic ratio comparably high as ^{1}H (Figure A8). Six fluorinations were introduced into a single Dz^{5C} on positions dG_{14}, dA_{11}, dC_7, dG_2, dG_{-5}, and dG_{-4}. In complex with the ^{19}F-stabilized RNA target, the in total seven fluorines lead to four intensive signals in a 1D ^{19}F spectrum (Figure 3.9a). A homonuclear [^{19}F,^{19}F]-NOESY experiment does not show any correlation between the fluorine nuclei, but heteronuclear [^{1}H,^{19}F]- and [^{19}F,^{1}H]-HOESY spectra show some signal. However,

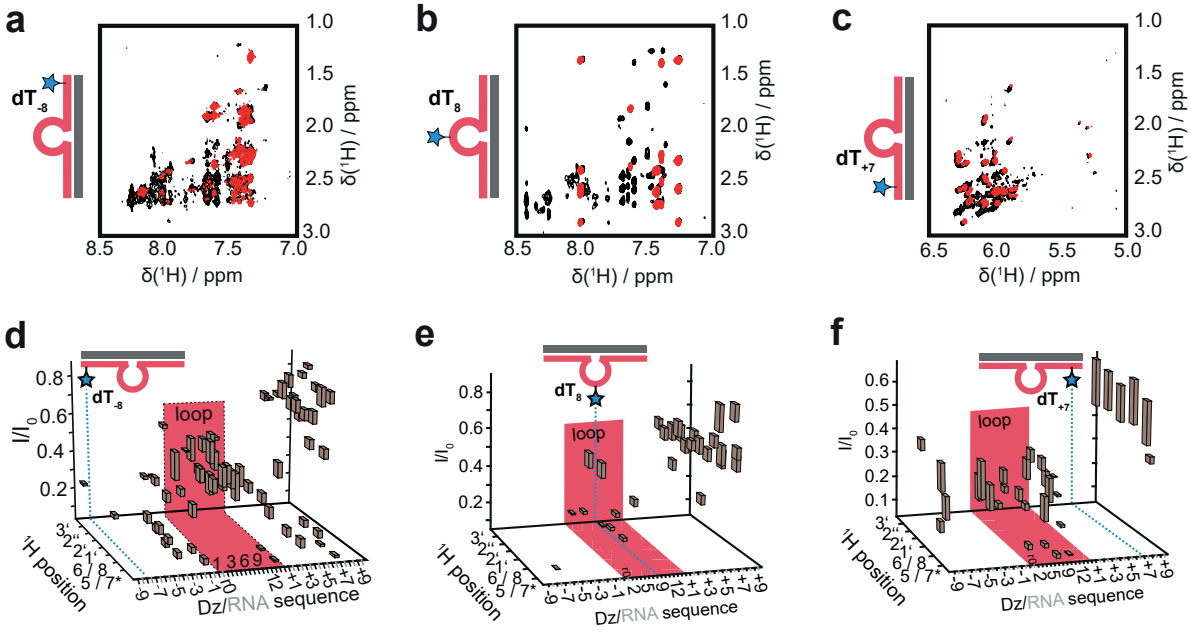


Figure 3.8: Paramagnetic spinlabelling generates low-resolution long-range restraints. (a-c) Thymine bases on positions dT_{-8} (a), dT_8 (b), and dT_{+7} (c) were substituted by 5-ethynyl-2'-deoxyuridine and functionalized with a paramagnetic TEMPO spinlabel (blue star). [^{1}H,^{1}H]-NOESY (a+b) and TOCSY spectra (c) show selective line-broadening due the PRE effect (purple spectra) compared to reference spectra with a reduced spinlabel (black spectra). (d-f) Mean intensity ratios for individual protons were calculated from all accessible [^{1}H,^{1}H]-NOESY and TOCSY crosspeaks, which can be translated in loose long-range distance restraints. The bar position in the plots shows the relative signal loss of individual protons, while the bar size is indicating the respective standard deviation. The blue dotted line shows the sequence position of the spin label.

the ^{19}F-detected version is lacking resolution, while with ^{1}H detection signal-to-noise levels are insufficient (Figure A9). The most promising approach was the application of ^{19}F-saturated STD experiments. Saturation was applied at frequencies of the different signals in the 1D ^{19}F spectrum, while proton difference spectra were recorded, which give rise to several STD signals (Figure 3.9b+c), where the strongest signals derive from vicinal and geminal bound protons at 1' and 2' position of the fluorinated ribose units. These peaks show characteristic splitting due to J^2_{FH} and J^3_{FH} of 10-18 Hz and 50-55 Hz, respectively. The corresponding peaks can been directly identified in a [^{1}H,^{1}H]-NOESY spectrum by their peak splitting, and therefore assigned to their sequential positions (Figure 3.9b). Several other STD peaks can be assigned, if they are not laying in a too crowded spectral region (Figure 3.9c), which can be translated into distance restraints. Since the saturation effect is not directly proportional to distance, STD-derived restraints have to be more loose than conventional NOE restraints and therefore were introduced in the structure calculation with 8 Å upper restraint limit. Due to broadness of the saturation pulse (50-80 Hz) also fluorines close to the saturation frequency are excited, which leads to the same STD signals with different saturation frequencies (e.g. signal at 5.7 ppm in Figure 3.9c). The STD intensity decreases, when the saturation point moves away from the ^{19}F resonance frequency of interest. Careful interpretation of the STD signals in respect to different saturation frequencies and strength allows to assign these signals to their respective ^{19}F

frequency. However, this problem can be avoided by using only one fluorination as performed for the Dz^{5C}:RNA$^{2'F}$ (Figure A9c). Interestingly, 2' fluorinations show decreased linewidth when the nucleotide is involved in strand hybridization, which might make ^{19}F also a suitable probe for *in situ* NMR studies (Figure A9d).

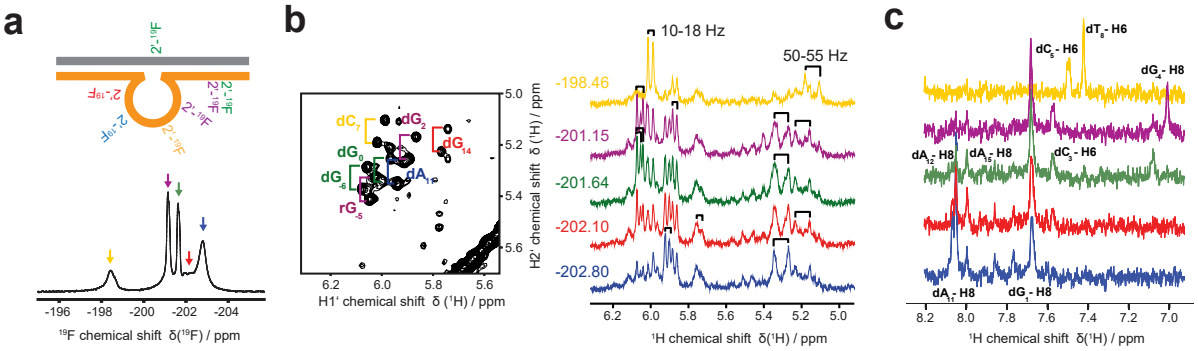


Figure 3.9: Selective fluorination enables novel NMR strategies to generate ^{19}F-derived distance restraints. (a) Additional to the stabilized RNA$^{2'F}$, six other 2' fluorinations were introduced at Dz^{5C} positions -6, -5, 2, 7, 11, and 14 (top). A 1D ^{19}F spectrum gives four sharp and one broad signal. Selective saturation pulses at the different peak frequencies (arrows) result in different STD spectra (indicated by the respective colours). (b) H1' and H2' protons of fluorinated ribose moieties can be identified in a [^{1}H,^{1}H]-NOESY spectrum by their characteristic peak splitting due to the strong J_{HF} coupling (left). The same splitting pattern can also be observed in the respective difference spectrum and therefore ^{19}F resonances can be assigned to their nucleotide position (right). (c) Additional STD signals were arising for protons in close proximity to the fluorine atom, and can be translated into distance restraints.

Angular restraints by structural homology and NOE build-up rates.

Dihedral angle restraints for nucleic acids are commonly extracted from *J* coupling data, which can be derived using the peak splitting in a double-quantum filtered (DQF-)COSY spectrum. Unfortunately, the complexity of the coupling and the high peak density in the spectrum of Dz^{5C}:RNA$^{2'F}$ makes a simple analysis very challenging. An extensive peak deconvolution strategy might enable to extract *J* coupling constants from this data-set (Figure A10). An exclusive COSY (E.COSY) experiment gives more simple splitting patterns, but requires isotope-labelled sample. Due to insufficient signal-to-noise levels and strong overlap only three coupling constants of nucleotide dT$_{-9}$ could be derived from a H(C)CH-TOCSY-E.COSY spectrum of ^{13}C,^{15}N-Dz^{5C}:RNA$^{2'F}$ (Figure 3.10a). On the other hand, a straight-forward approach to access angular restraints for the binding arms is a simple comparison of chemical shifts between Dz^{5C}:RNA$^{2'F}$ and a complex of target RNA and a complementary DNA lacking the catalytic loop (DNA^{C}:RNA$^{2'F}$). Direct comparison of pyrimidine TOCSY crosspeaks between Dz^{5C}:RNA$^{2'F}$ and DNA^{C}:RNA$^{2'F}$ gives the same pattern for nearly all peaks of the binding arms (Figure 3.10b). An analysis of the chemical shift differences of all accessible protons of the binding arms reveals, that only the four inner nucleotides of both arms show a significant difference between both complexes (Figure 3.10c). A DNA:RNA hybrid adopts an A-form helical structure

in solution [6, 132], and since the binding arms of Dz^{5C}:$RNA^{2'F}$ show identical chemical shifts, a similar structure can be assumed. Therefore, dihedral angles of the nucleotides at positions +9 to +5 and -5 to -9 were restrained with values typical for an A-form helix. The dihedral angle χ defines the relative orientation of the nucleobase to the sugar moiety. If the base is oriented towards the sugar, the conformation is named as *syn*, and if the base is pointing away as *anti*. In *syn* conformation the base proton H6 or H8 is forming a stronger NOE contact with the H1' of its own sugar (Figure 3.10d). Comparison of NOE build-up curves of the individual nucleotides show stronger build-up rates for H6-H1' and H8-H1' distances of nucleotides dA_{-1} and dC_{13} and these nucleotides were therefore restrained as *syn*, while data for nucleotides dT_4, dC_5, dT_8, dA_{11}, rU_{-2}, rU_{-1}, rG_0, rC_{+1}, and rC_{+2} was ambiguous and their χ angles were left unrestrained. All other remaining nucleobases were restrained in *anti* conformation.

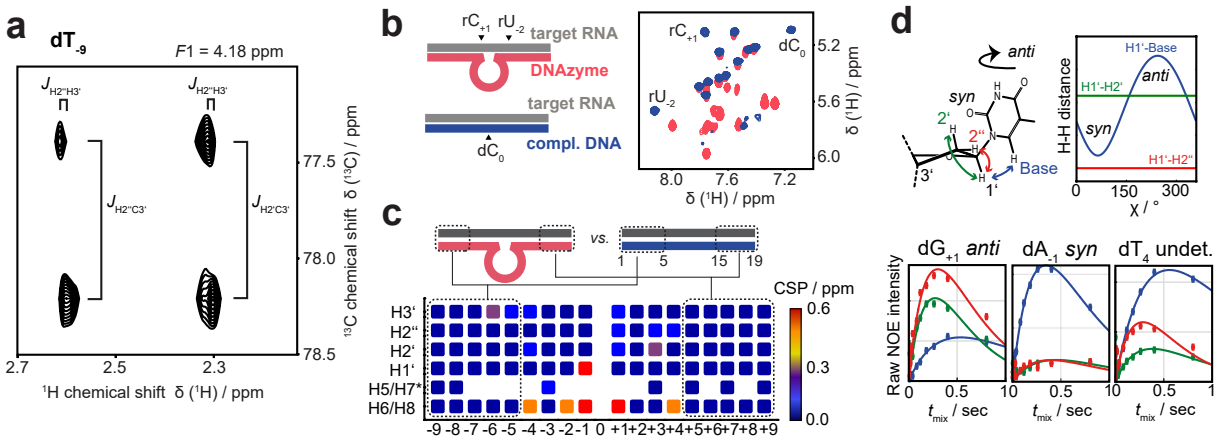


Figure 3.10: Dihedral angular restraints for Dz^{5C}:$RNA^{2'F}$. (a) H(C)CH-TOCSY-E.COSY spectrum of ^{15}N,^{13}C-Dz^{5C}:$RNA^{2'F}$ provides resolved J coupling for nucleotide dT_{-9} which can be calculated into dihedral angles by the Karplus relation (Equation 1.8). However, additional J coupling constants could not have been extracted. (b) Comparison of Dz^{5C}:$RNA^{2'F}$ with a complex of $RNA^{2'F}$ and complementary DNA (DNA^{C}:$RNA^{2'F}$) yields fast insights into secondary structure. Overlay of TOCSY pyrimidine regions of Dz^{5C}:$RNA^{2'F}$ (red) and DNA^{C}:$RNA^{2'F}$ (blue) shows matching peak positions for nearly all binding arm peaks. (c) Analysis of individual proton shift differences over the full (binding arm) sequences gives maximum deviations for nucleotides close to the cleavage site. (d) Dihedral angle χ between C1' and N_{Base} defines the relative position of nucleobase to sugar moiety. Atomic distances between H1' and H6 or H8 (H1'-H_{Base}), as well as H1'-H2' and H1'-H2", are not altered by ribose puckering, and therefore only depending on angle χ (top). Relative NOE build-up rates of H1'-H_{Base}, H1'-H2', and H1'-H2" can be used to define *syn* and *anti* states and for differentiating between H2' and H2" during peak assignment (bottom).

Structure calculation was supplemented with exact NOE and RDC data.

Relative NOE intensities derived from a single mixing time do not provide accurate atomic distances due to different relaxation and spin diffusion effects of proximal protons. In conventional NMR-derived structure calculations, NOE intensities are grouped into strong, medium and weak restraints in relation to NOE peak intensities with a known distance. With an abundancy of long-range NOE contacts, even loose restraints derived from classical NOEs lead to

a successful structure calculation. Because of the strong signal overlap in the Dz^{5C}:$RNA^{2'F}$ spectra and the reduced number of protons in nucleic acids compared to proteins, long-range NOESY crosspeaks are sparse. The determination of cross-relaxation rates from NOE build-up curves using multiple mixing periods are more reliable, but they still suffer from the influence of spin diffusion. To compensate this effect a relaxation matrix approach based on the eNORA software is used [112, 119]. An initial structure was calculated with the Xplor-NIH software [148, 148] using the restraints described above as well as distance restraints derived from conventional NOEs (Figure 3.11a). NOE build-up data was used to extract individual auto- and crossrelaxation rates and the lowest energy structure from the initial calculation was used to determine spin diffusion effects. Using this approach, cross-relaxation rates were corrected by the effects of spin diffusion and exact NOE (eNOE) distance restraints could be derived, which are replacing the conventional, more loose restraints (Figure 3.10a inlet, see Figure A11 for a larger subset of derived eNOE plots). Furthermore, the initial structure model was used to find still unassigned long-range NOE contacts in the NOESY spectra. The newly generated restraints were used as an input for the next structure calculation, and outcoming structures serve as templates to search for eNOE and long-range contacts. This process was repeated, until no new long-range or eNOE could been determined. Additionally, residual dipolar coupling (RDC) constant were used as restraints in the structure calculations. If molecules are partially aligned with the magnetic field by addition of a suitable cosolvent, RDC between two nuclei leads to peak splitting, which magnitude is directly related to the orientation of the bond vector between both nuclei and the field vector. RDC constants of ^{13}C,^{15}N-Dz^{5C}:$RNA^{2'F}$ could be extracted from [^{1}H,^{13}C]-HSQC spectra without decoupling in absence and presence of phages Pf1 as alignment medium (Figure A12). The molecule's alignment tensor was calculated from an initial structure model and 52 determined RDC values (Table A3), however 20% of the RDC constants were excluded from the subsequent structure calculation and only included for later verification of the calculated structure. RDC constants are rather small and signal intensity levels were complicating accurate determination of RDC values, therefore an error up to 3 Hz was assumed.

Calculated structure ensemble was clustered into different folds.
A random starting structure for Dz and RNA was used for the initial runs of the structure calculation and yielded mainly models, in which the RNA strand follows its A-form helical turn at the cleavage site, while the Dz's loop region was laying up- or downstream of the cleavage site. However, several low-energy structures showed also a fold, in which the RNA strand is passing through the Dz's loop, which corresponds with several findings presented in this study (*vide infra*). Since root mean square deviations of atom positions between structures representing two different folds are fairly small, the different conformations were sorted manually into different clusters (Figure A13 and A14). One third of the structures with lowest energy profile show the conformation, in which the RNA strand is passing through the Dz's loop and winding around both Dz's binding arms (cluster I) in a calculation run excluding RDC restraints (Figure A13), however, including these restraints more than 50% of structures in the ensemble adopt this

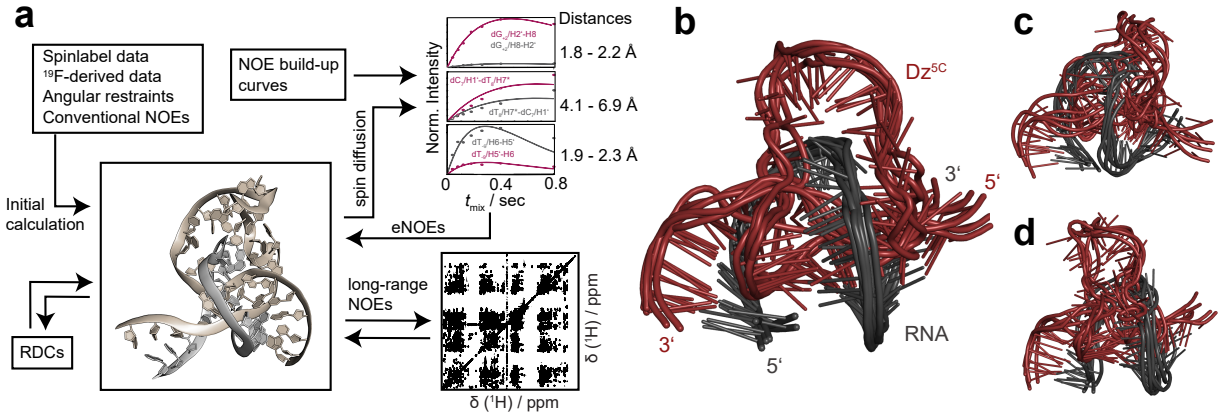


Figure 3.11: Iterative structure calculation of Dz^{5C}:$RNA^{2'F}$. (a) An initial structure was calculated using PRE data from selective spinlabelling, ^{19}F-derived distances, angular restraints and conventional NOE contacts. Output structures were used to derived spin-diffusion effects, to find former unassigned long-range NOESY crosspeaks and to calculate alignment tensor. Exact NOE contacts (eNOEs) give more accurate distance restraints and were calculated by the eNORA software using NOE build-up data and structure-derived spin diffusion effects (upper right inlet). New restraints were fed into the next round of structure calculation. (b-d) Ensemble of the 10 lowest-energy structures after structural refinement were clustered into three distinct folds: Six of the 10 structures with lowest energy profile fall into cluster I, in which the RNA passes through the Dz loop (b). In other structures the loop region is located down- (c, cluster II) or upstream (d, cluster III) of the RNA's cleavage site. Five structures with lowest energy profile in the respective cluster are presented.

specific fold (Figure A15). The RDC restraints led to a slight reduction of the other violations, although several RDC restraints were violated itself (see Table A4), but later back-calculation of RDC constants using the lowest-energy structure of cluster I (Figure A14) showed that they are generally in agreement with the observed values (Figure A12c). RDC restraints led to several artificial kinks and distortions in helices of the binding arms, therefore a refinement calculation was performed with a starting model selected from cluster I, in which the kinetic component in the calculation was strongly reduced. The final structure calculations showed the three different folds described above, whereas the cluster I type is mostly occupied. Tables 3.1 and A4 shows a summary of all used restraints and violations of the structure calculations.

3.2.4 Dz's catalytic loop is stabilized by winding around the RNA target.

It is an open question, if Dz forms a stable conformation before performing the cleavage reaction or if the loop region remains mainly unstructured. To investigate the dynamic behaviour on a ps-ns timescale, HetNOE and T_1/T_2 relaxation experiments were performed using isotope-labelled ^{13}C,^{15}N-Dz^{5C}:$RNA^{2'F}$. {^{13}C}-^{1}H NOE enhancement and T_1 relaxation rates of C1' in deoxynucleotides do not show any significant difference between residues belonging to the catalytic loop or binding arms, suggesting no faster dynamics for the loop region than the overall tumbling rate (Figure 3.12a). T_2 relaxation rates do not provide conclusive data, mainly due to the strong scalar coupling of the neighbouring 2' carbon (Figure A15). Base-pairing is a main factor for stabilisation in nucleic acid structures and in a Watson-Crick (WC) base pair the imino and amino protons are protected from exchange with the solvent. For a cytosine base, which is involved in WC base pairing, strong NOE signals between the base protons H5

Table 3.1: Restraints and violations of Dz^{5C}:$RNA^{2'F}$ refinement structure calculation.

Restraints		**Violations**[a]	
	Number	**Number**[b]	**RMSD**[c]
conventional NOE			
intra-nucleotide	691	54.2(3.2)	0.241(0.009)
inter-nucleotide	274		
long-range	43	1.2(0.4)	0.166(0.001)
exact NOE	108	20.5(1.2)	0.516(0.034)
^{19}F-derived distances	44	0.3(0.5)	0.079(0.042)
Spinlabel-derived distances	126	5.1(1.4)	0.295(0.021)
H bonds	143	0.0(0.0)	0.055(0.005)
Dihedral angles	208	13.6(4.2)	0.886(0.042)
Residual dipolar coupling	42	28.9(2.6)	4.331(0.265)
Ensemble RMSD / Å		**Lowest energy**[a]	**Cluster I**[d]
		3.329	2.431

[a] Calculated from the 10 lowest energy structures out of 100 calculation runs.
[b] Mean number of violations in the ensemble (standard deviation within the ensemble). Thresholds for violations are 0.5 Å for distances, 5° for angles and 2.5 Hz for RDC.
[c] Mean RMSD of violations in the ensemble (standard deviation within the ensemble). Distance violations in Å, angular violations in °, and RDC violations in Hz.
[d] 6 out of 10 lowest energy structures (see Figure 3.11b).

and H41 are visible, due to the slower exchange. While these could be detected for nearly all cytosines in the hybridized binding arms, none were appearing for dC_3, dC_5, dC_7, dC_{10}, and dC_{13} of the loop region (Figure 3.12b), despite there are potential WC partners available within the catalytic loop. A TOCSY spectrum recorded at acidic conditions (pH 5.7) shows effects on the chemical shift of solely the catalytic loop, compared to neutral pH 7.5 (Figure 3.12c). Santoro and Joyce already published a linear behaviour of the catalytic rate within the pH range 6.5 to 8.5 and explained it with a necessary deprotonation event at the 2'OH next to the scissile bond [140]. However, below this pH regime the catalytic activity is dramatically reduced, which might be explained by destabilizing protonation events, leading to a disruption of the active loop conformation (Figure A16). Temperature-gradient NMR experiments on Dz^{5C}:$RNA^{2'F}$ gives individual melting curves for each accessible nucleotide and are suitable to localize stabilized regions, therefore [^{1}H,^{1}H]-TOCSY spectra were performed with temperatures ranging from 15 to 65℃ and the pyrimidine peak intensities were analysed (Figure 3.12d+e). For unstructured single-stranded Dz^{5C} the crosspeak intensity is linearly increasing with the temperature (Figure 3.12d), while nucleotide peaks of the Dz^{5C}:$RNA^{2'F}$ complex, which are involved in WC

base pairing, show a characteristic intensity dip around 52 °C due to interchange of paired and unpaired state. Interestingly, nucleotides adjacent to the cleavage site and of the 5' site of the catalytic loop (positions 3, 5, 7) show the same or similar behaviour of stabilization, while nucleotides of the 3' site of the loop (positions 17, 19, 22) behave like single-stranded nucleic acids (Figure 3.12e, see Figure A1 for a full analysis). Although the principle of the stabilizing effects are unclear, only nucleotides 1-7 of the catalytic loop seem to be involved in it, which have, notably, also a significant impact on Dz activity when being mutated (*vide infra*).

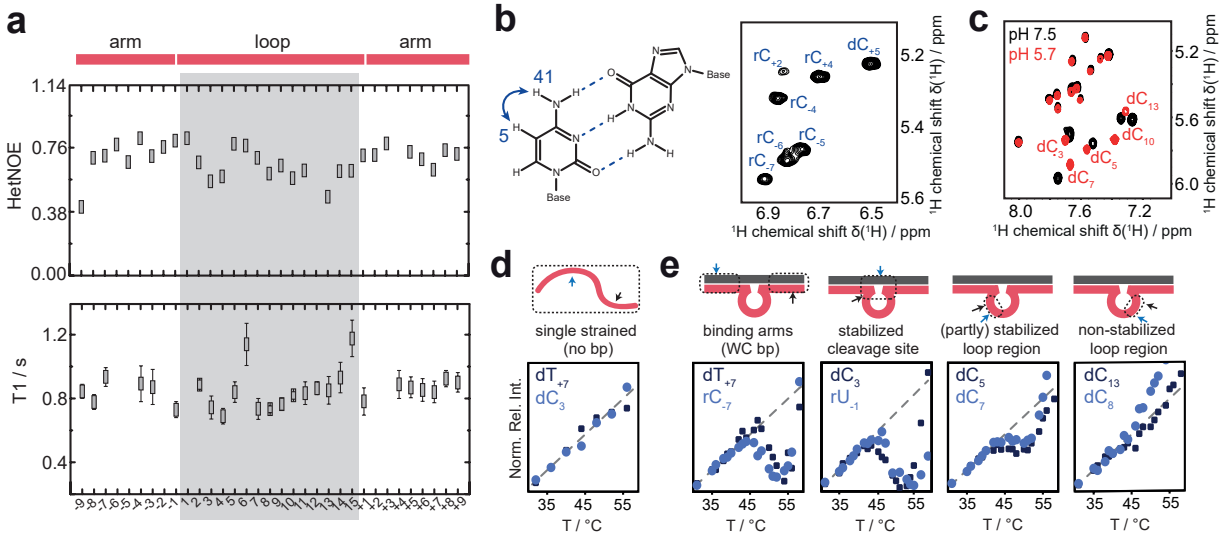


Figure 3.12: The catalytic loop is stabilized in absence of permanent WC base pairing. (a) HetNOE (top) and T_1 relaxation data of ^{13}C1' from Dz^{5C}:RNA$^{2'F}$ are sensitive for ps-ns dynamics and do not show any significant differences between binding arms and loop region (grey area). (b) NOESY crosspeaks between H5 and amino H41 protons are visible for cytosines which are protected from solvent exchange by WC base pairing. Only amino crosspeaks belonging to the hybridized binding arms are visible, but non for the loop region. (c) Dz is less active at acidic pH (see Fig. A14). TOCSY spectra recorded in phosphate buffer at pH 7.5 (black) and pH 5.7 (red) show no differences for the binding arms, while the loop region is highly affected by the pH change. (d) TOCSY spectra of single-stranded Dz^{5C} and Dz^{5C}:RNA$^{2'F}$ were recorded with different temperatures and individual peak intensities are plotted against the temperature. Single-stranded Dz^{5C} is not stabilized and show a linear increase in intensity. Nucleotides involved in WC base pairs show an intensity dip at 45-58 °C due to the conformational exchange between paired and unpaired nucleotides. Residues close to the cleavage site show a similar stabilizing effect, while parts of the 5' site of the loop are partially stabilized. The 3' site of the loop is unstabilized and shows similar behaviour like ssDz^{5C} (from left to right). A full dataset is shown in Figure A1.

The most striking feature of the calculated Dz model (Figure 3.11b) is, that the catalytic loop region is twisted around the RNA strand at height of the cleavage site. This sterically limits loop rearrangements and might explain the loop's stability even in absence of base pairing. To allow this structural element to form, both binding arms are crossing each other, which brings the phosphate backbones of both Dz arms in close proximity (Figure 3.13a). The dG_{+1} nucleotide is flipped out to allow the Dz strand of the 5' arm to pass by, which can be seen by missing sequential NOESY crosspeaks between dG_{+1} and dG_{+2} (Figure 3.5a and 3.13c). This prohibits

to form a stable interstrand WC basepair at position +1 and is reshaping the 20 year-old model of the 10-23 DNAzyme. The results could be further verified by functional assays, where mutations at Dz postion +1 did not hinder its activity, while mutations at potentially base-paired Dz position -1 completely abolished its cleavage capability (Figure 3.13b). Interestingly, NOESY crosspeaks between dG_{+2} and dG_{-5} become visible at higher Mg^{2+} concentrations, which are clear indicators for the crossing of both Dz arms (Figure 3.13a+c). It can be speculated, that Mg^{2+} (or high amounts of Na^{+}) are neutralizing the repulsive interaction between both negatively charged phosphate backbones and are stabilizing the pre-catalytic complex (compare Figure 3.7).

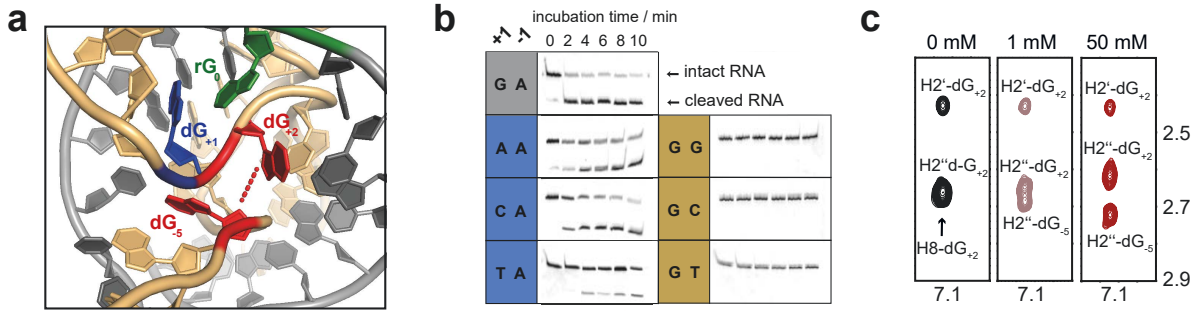


Figure 3.13: Position +1 is not WC base paired in order to form pre-catalytic complex. (a) Section of the Dz^{5C} (orange) in complex with $RNA^{2'F}$ (grey) with view from Dz's 3' terminus. Unpaired dG_{+1} (blue) is flipped out in respect to dG_{+2} (red), while dG_{+2} comes in close proximity to dG_{-5} (both red). (b) Cleavage activity recorded by SDS-PAGE of non-mutated Dz^{5C} (grey), and with mutations at position +1 (blue), and position -1 (orange). (c) Sections of NOESY spectra at 0, 1, and 50 mM are showing a crosspeak between H8-dG_{+2} and H2"-dG_{-5}, which is clearly separated at high Mg^{2+} concentrations (indicated as dotted red line in a).

3.2.5 Dz^{5C}:$RNA^{2'F}$ has up to three distinct Mg^{2+} binding sites.

Localisation of Mg^{2+} binding sites onto the novel three-dimensional model of Dz might reveal its role in catalysis. Since Mg^{2+}-induced CSPs do not allow to distinguish between direct binding or structural rearrangements due to Mg^{2+} binding, a NMR titration experiment using $MnCl_2$ was applied. Due to its comparable chemophysical property to Mg^{2+} and its line-broadening effect on NMR signals, paramagnetic Mn^{2+} is often employed as a mimic in NMR experiments to probe for Mg^{2+} binding sites. Mn^{2+} is also able to accelerate Dz-mediated RNA cleavage more than 10-fold compared to Mg^{2+} (Figure 3.14a). High amounts of $MnCl_2$ would severely broaden all NMR signals to an undetectable level, therefore Dz^{5C}:$RNA^{2'F}$ was preincubated with 100 mM NaCl and 1 mM $MgCl_2$ and subsequently titrated with increasing amounts of $MnCl_2$ to a final concentration of 10 µM in the presence of 200 µM complex. Mn^{2+} and Mg^{2+} compete for similar binding sites and PRE effects could be observed proton-specifically depending on the occupancy level of Mn^{2+} ions. PRE rates could be extracted from the individual peaks by analysing their peak intensities with increasing Mn^{2+} concentration (Figure 3.14b+c). A mean value of PRE rates for all crosspeaks involving the same proton were calculated and assigned

to the respective proton. Display of the PRE effects on the Dz^{5C}:$RNA^{2'F}$ sequence and structure shows, that especially the DNAzyme binding arms at position +1, +2 and -5 to -7 were affected by Mn^{2+} binding (Figure 3.14d+h). In the structure model the binding arms are crossing each other at this position bringing the phosphate backbones in close proximity. This is generating a high density of negative charges and a high number of cationic counterions is needed to neutralize the resulting electrostatic repulsion. This might be a crucial step for Dz:RNA complex stabilisation and leads presumably to the state homogenising effect seen with Na^{+} and Mg^{2+} (see Figure 3.7b-e). In the loop region also nucleotides 1-3, 11 and 14 show elevated PRE rates, which suggests two binding sites in the loop. Mn^{2+} shows a binding preference for guanine nucleotides in Dz^{5C}:$RNA^{2'F}$, however this could not be observed for DNA^{C}:$RNA^{2'F}$ (Figure A17). Interestingly, Mn^{2+}-induced PRE effects in the loop do not match with most of the Mg^{2+}-derived CSP, which either concludes Mn^{2+} and Mg^{2+} have different binding sites, or that binding of divalent cations causes global rearrangements of the Dz^{5C}:$RNA^{2'F}$ complex (Figure 3.14d,e+h).

Another mimic for exclusively hexahydrated Mg^{2+} is cobalt hexamine $Co(NH_3)_6$, whose amine protons can form NOE contacts to protons of the Dz^{5C}:$RNA^{2'F}$ complex, and therefore act as an additional probe for magnesium binding sites. Unlike Mg^{2+}, $Co(NH_3)_6$ is not able to induced RNA cleavage, but increasing amounts of $Co(NH_3)_6$ reduces Mg^{2+}-induced Dz activity, suggesting that they both share similar binding sites (Figure 3.14f). [^{1}H,^{1}H]-NOESY spectra of Dc^{5C}:$RNA^{2'F}$ were recorded with different concentrations of $Co(NH_3)_6$ up to one 1 mM. Several crosspeaks between the amine protons of cobalt hexamine (3.36 ppm) and Dz^{5C} base protons (8.5-6.5 ppm) appear with increasing titration steps. One peak can be unambiguously assigned to H8 of dG_{+1}, four other signals are matching to base protons of dG_{-4}, dG_{-3}, dT_{-2} and dT_{+3}. Another crosspeak can be assign to either dA_{11} or dA_{12} (Figure 3.14g+h). For the Dz region, in which both binding arms crossing each other, binding effects derived from PRE, CSP and $Co(NH_3)_6$ are visible and lead to the conclusion that most binding events are occurring at this site. Interestingly, the nucleotides close to the cleavage site as well as loop nucleotides 3 to 6 present the largest Mg^{2+}-induced CSP, but do not show any binding in $MnCl_2$ and $Co(NH_3)_6$ titration experiments. Possibly, structural rearrangements are induced by binding events at another position like at nucleotides dG_1 and dG_2 (Figure 3.14h). It is noteworthy, that the loop positions with the highest CSP also show the largest stabilising effects during melting of the complex (Figure 3.12e). Mn^{2+}-induced PRE data as well as a NOE contact with $Co(NH_3)_6$ give another binding patch for divalent metals at loop nucleotides 11 to 14. No significant CSP during $MgCl_2$ titration could be observed, and therefore no structural changes seem to be linked with this binding site. Presumably, binding at this position is only possible after the Dz^{5C}:$RNA^{2'F}$ adopts a favourable conformation. The phosphate ester of the scissile bond is exposed towards this binding site, possibly to allow RNA cleavage mediated by a bound divalent ion. That $Co(NH_3)_6$ is also able to bind here and is able to displace catalytic active Mg^{2+} supports the hypothesis, that loop positions 11 to 14 are the site of action for Dz-mediated RNA cleavage (Figure 3.14f).

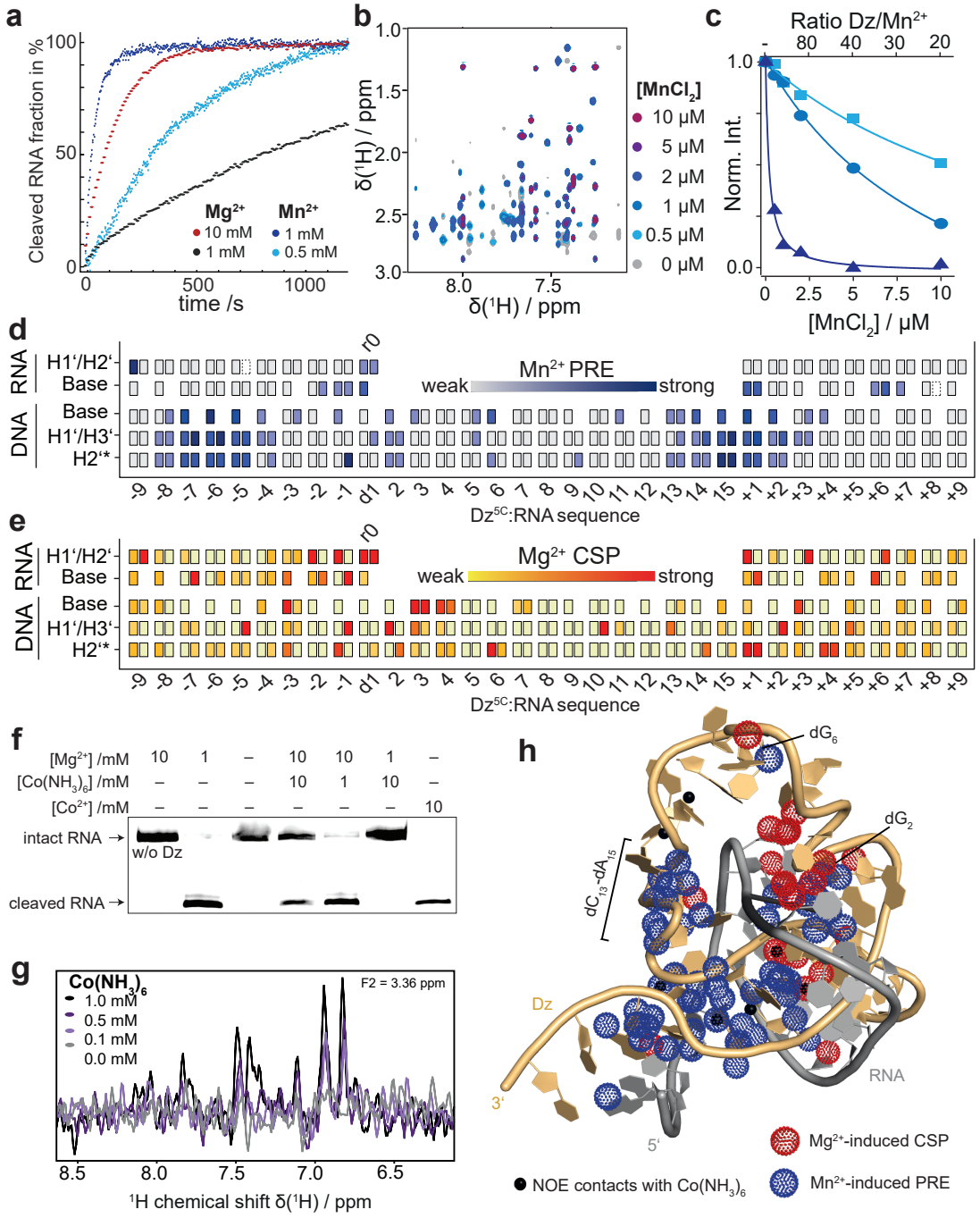


Figure 3.14: Mn^{2+} and $Co(NH_3)_6$ titration experiments reveal at least three different binding sites for divalent ions. (a) Mn^{2+} is activating Dz 10-fold more efficient than Mg^{2+}. (b) $MnCl_2$ titration of Dz^{5C}:$RNA^{2'F}$ selectively broadens proton-proton crosspeaks. (c) Individual PRE rates were extracted by fitting of peak intensities. (d) Average PRE rates for individual protons are plotted along the nucleotide sequence. Mn^{2+} binding can be located at positions -8 to -5, 2, 6, 13 to 15 and +1 to +3. Dotted boxes display protons for which no value could be determined. (e) Average CSP induced by addition of 50 mM $MgCl_2$ for individual protons are plotted along the nucleotide sequence. (f) Co^{2+} is activating Dz, while $Co(NH_3)_6$ is not able to induce RNA cleaving, but inhibits Mg^{2+}-depending Dz catalysis. (g) Amine protons of $Co(NH_3)_6$ have their chemical shift at 3.36 ppm and can form NOE contacts with proximal protons of Dz^{5C}:$RNA^{2'F}$. Cross-sections taken form 2D NOESY data at 3.36 ppm in the area of the base protons are shown and crosspeaks could be assigned to dG_{-5}, dG_{-4}, dT_{-3}, dG_{+1} and dT_{+3}. The signal at 7.9 ppm cannot be distinguished between dA_{11} and dA_{12}. (h) The effects of Mg^{2+}, Mn^{2+}, and $Co(NH_3)_6$ binding are plotted on the structural model of Dz^{5C}:$RNA^{2'F}$. Mg^{2+}-induced CSP above a threshold of 0.05 ppm are displayed as red dotted spheres; Mn^{2+} PRE rates above a threshold of 0.7 rate constant are displayed as blue dotted spheres; Protons involved in NOE contacts with amines of $Co(NH_3)_6$ are shown as black spheres. The RNA cleavage site is also highlighted in black.

3.2.6 Catalytic loop structure is in agreement with functional studies

Zaborowska *et al.* already performed mutation studies for each nucleotide within the catalytic loop and observed their impact on the catalytic activity of Dz [186]. They found out that nucleotides dG_1, dG_2, dT_4, dG_6, and dG_{14} cannot be mutated without complete loss of activity, while $dC_{3,}$, dC_{13} and dA_5 can only be substituted with A and C, respectively. Since the Dz catalytic loop is twisted around the RNA target in the favoured model, especially these nucleotides are in close proximity to the cleavage site (Figure 3.15a). Surprisingly, exchange of dG_1, dG_2 and dT_4 to abasic nucleotides or only a C3 spacer have no significant impact on the activity of 10-23 DNAzyme, which suggests that possible interaction of these residues is facilitated via their phosphate groups [174]. dC_3 and dT_4 showed the highest Mg^{2+}-induced CSP, but are probably not directly involved in Mg^{2+} binding, since no positive binding data could be observed from Mn^{2+} and $Co(NH_3)_6$ titration experiments (Figure 3.14h). Since an abasic substitution at position 4 does not significantly affect Dz's activity, it can be hypothesized that dT_4 is flipped out in a Mg^{2+}-bound state. NOE stacking contacts between this thymine and cytosines at position 3 and 5 seem indeed to disappear with elevated $MgCl_2$ concentrations, but unfortunately are overlapping with the strong H2'-H1' crosspeak of dC_5 at 50 mM $MgCl_2$ (Figure 3.15b). Interestingly, no other parts in the loop show significant CSP during titration experiments with $MgCl_2$, despite binding of divalent cations could be observed for loop position 11-14. Especially dG_6 and dG_{14} are well-know to be crucial for Dz's activity [174, 186], which are positioned on opposing sites of the cleavage site (Figure 3.15a). Since loop positions 11-14 show divalent ion binding in the performed NMR experiments and dG_{14} is located in close proximity to the scissile bond in the calculated model, it is most likely directly involved into the coordination of a catalytic relevant Mg^{2+} (Figure 3.14h). Proton H8 of dG_6 is also strongly affected by Mn^{2+}-induced PRE, but is located more distant from dG_{14} in most of the models in the calculated structure ensemble. The loop rearrangements, which are indicated by the CSP at elevated Mg^{2+}, might relocate the nucleobase of dG_6 in close proximity to dG_{14} and the scissile bond and they are forming together a binding site for a divalent ion, which is directly involved in the cleavage mechanism. The loop nucleotides 5-13 show high similarity with single-stranded Dz in stacking contacts and chemical shift, which might point towards an already preformed 'minimal' structure of the loop regions stabilized by π-π stacking and hydrogen bonds (Figure A3). This might explain, why mutations in nearly all loop nucleotides have an impact on Dz's cleavage rate albeit not dramatically, because this 'minimal' structure might be disrupted. All published mutations were based on the Dz^{5A} loop sequence, therefore activity assays with the given mutations were also performed based on the Dz^{5C}, which lead to the expected activity loss for mutated positions 1-4, 13, and 14 (Figure 3.15c).

3.2.7 Cleavage reaction can be followed by real-time NMR

The less active Dz^{5C} mutant cleaves its RNA target completely within 12 hours in a single turnover experiment at 37°C and in presence of 1 mM $MgCl_2$. Since most of the conversion happens in the first two hours of reaction time, the process can be well followed during NMR

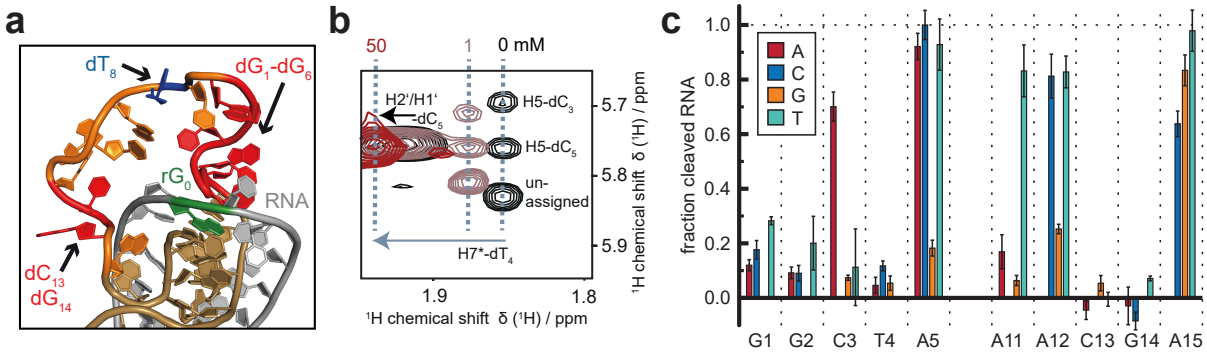


Figure 3.15: The Dz structural model is in accordance with mechanistic data. (a) Dz loop region (orange) and RNA (grey) with highlighted cleavage site (green) and nucleotides according to their catalytic relevance in mutational studies [186]: no effect (blue), moderate effect (orange), strong effect (red). (b) Overlaid [^{1}H,^{1}H]-NOESY section of Dz^{5C}:$RNA^{2'F}$ with 100 mM NaCl and no (black), 1 mM (light purple), and 50 mM $MgCl^2$ (red). Indicated crosspeaks between H7*-dT_4 and H5-dC_3 or H5-dC_5 seem to disappear with increasing Mg^{2+} concentration. However, their position is masked by the strong crosspeak of H2'/H1'-dC_5 at 50 mM $MgCl_2$. (c) Relative cleavage activity of Dz^{5C} with single-mutations at the indicated position measured by FRET-based kinetic assay.

acquisition. In a time-resolved experiment, identical 1D experiments are performed in row to follow changes in the system by spectral changes. 1 mM of $MgCl_2$ was added to a pre-formed Dz^{5C}:RNA complex and a series of 280 ^{1}H 1D spectra of approximately 2.5 min duration each were recorded during the reaction. During acquisition, several peaks corresponding to the initial state were disappearing, while new peaks are appearing with a similar rate (Figure 3.16a+b). Analysis of the first recorded spectrum showed that equilibrium between Mg^{2+}-bound and unbound states was established on a fast time-scale.

To further analyse the individual behaviour of the single peaks, a real-time NMR experiment was performed. Therefore, a standard 2D [^{1}H,^{1}H]-TOCSY experiment was started immediately after initiation of the cleavage reaction and recorded over a time of 3h. Before and one day after start of the reaction, additional TOCSY experiments were recorded as a educt and product reference (Figure 3.16c). The 2D real-time spectrum shows peaks corresponding to the educt and product states, but the educt peaks are shifted due to the addition of 1 mM $MgCl_2$ (Figure 3.16c). Educt peaks in the real-time spectrum are broadened compared to their reference, since the initial state is disappearing during the course of the reaction. Contrary, peaks corresponding to the product state are sharper and show characteristic flanking dips. The cleaved RNA fragments are still hybridized with the DNAzyme binding arms, as seen from the typical pattern of pyrimidine TOCSY crosspeaks (compare Figure 3.3 V). Interestingly, the spectrum of Dz^{5C}:products after cleavage shows multiple states, similar to the spectra recorded for uncleaved Dz^{5C}:RNA in absence of ions, while one product state (B_1) is higher populated than the other state (B_2) (Figure 3.16d). The post-reaction complex was melted in the NMR machine by step-wise increasing of the temperature and subsequent spectra acquisition (Figure 3.16e, Figure A18). The high-occupied state B_1 was melting above 40 °C, while the lower occupied state (B_2) was appearing to be more stable with a melting temperature of 45-50 °C. The lineshapes of product and educt peaks of the RT spectrum can be simulated with different rate constants and compared to the recorded spectrum (Figure 3.16f) [9]. The rate constants *k* of

the best fits are spreading between 0.0002-0.001 sec^{-1} for all analysed peaks, while the rates belonging to arms and the catalytic loop are lower than the overall rate derived from the kinetic assay of 0.0004 sec^{-1} and the rates belonging to nucleotides close to the cleavage site are above (Figure 3.16g+h). The rate difference between catalytic loop and cleavage site might point towards two different processes during Dz-mediated RNA cleavage, which is not visible in fluorescence-based assays.

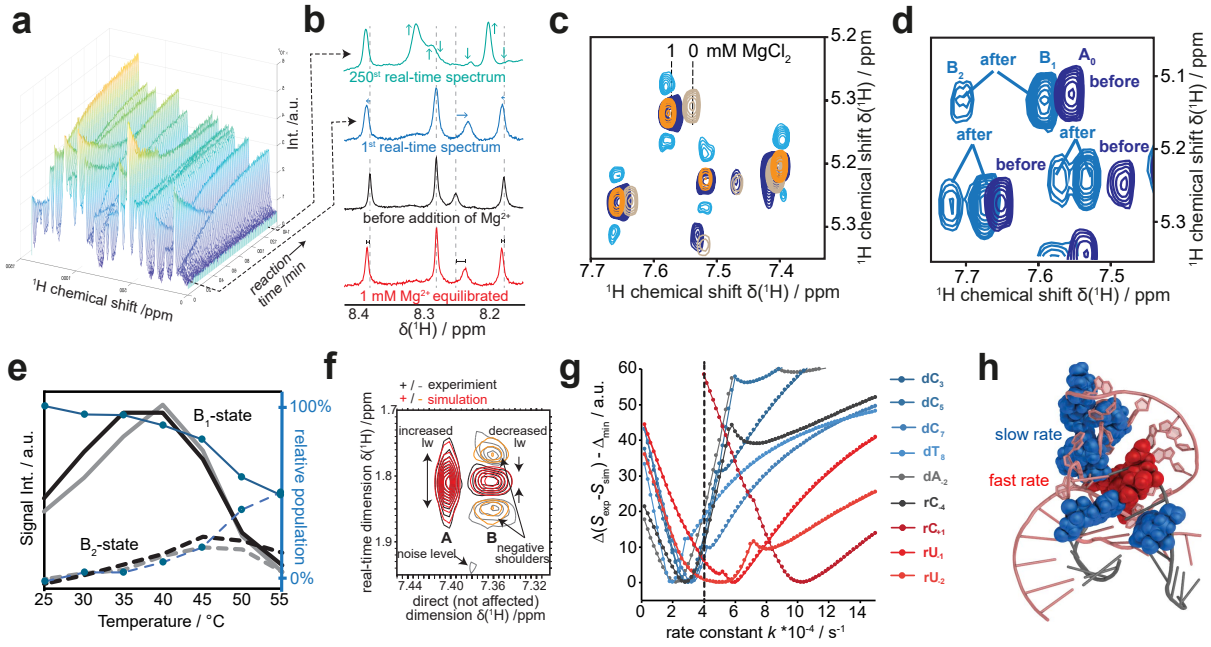


Figure 3.16: Real-time NMR spectroscopy allows to follow the cleavage reaction nucleotide-specifically. (a) A series of [^{1}H] 1D spectra were recorded during the cleavage reaction. Peaks belonging to the educt state are disappearing during acquisition, while product peaks are appearing. (b) First recorded spectrum (blue) shows peak shifts compared to the educt reference (black), which matches the peak position as seen in a Mg^{2+}-equilibrated spectrum (red spectrum). 250th spectrum shows increased product and decreased educt peaks (green). (c) Realtime (RT) [^{1}H,^{1}H]-TOCSY spectrum (blue spectrum) is consisting of signals belonging to the product and educt state. Educt peaks are broader than in a pre-catalytic reference spectrum without $MgCl_2$ (light brown spectrum) and are shifted due to the Mg^{2+}-induced CSP. Product peaks are sharper than in a reference recorded 24 h after reaction (orange spectrum) and show flanking negative dips. (d) The [^{1}H,^{1}H]-TOCSY spectrum of the cleavage products show additional peaks if displayed with lower contour levels. One state (B_1) is higher populated than the other (B_2), but both show the typical peak pattern for hybridized binding arms. (e) Complexed cleavage products can be dehybridized at temperatures above 50 °C, but state B_2 shows a slightly higher resistance to elevated temperature. (f) Overlay of simulated educt and product peak for dT_8 (red) on top of the acquired RT spectrum (black). (g) Product peak shapes were fitted with the two-state model using different rate constants *k* and the deviation to the experimental data was calculated. Nucleotides belonging to the loop region (positions 3, 5, 7, 8; blue curves) show *k* values below 0.0004 sec^{-1}, which is the rate determined by the fluorescence-based assay (vertical line). A higher *k* could be observed for residues directly adjacent to the cleavage site (positions -2,-1,+1; red curves). (h) Nucleotides with slow (blue spheres) and fast (red spheres) catalytic rates are mapped on the Dz^{5C}:$RNA^{2'F}$ structure.

3.3 Conclusion

This research provides the first high-resolution insights into the structure and the catalytic mechanism of the 10-23 DNAzyme (Dz) as well as the role of metal ion cofactors. NMR spectroscopy serves as a suitable tool for studying Dz in its different pre- and postcatalytic states and gives insights into dynamic behaviour and cation interaction. A pre-catalytic structure was determined of a complex with RNA stabilized against cleavage by a ^{19}F substitution at the cleavage site and Dz^{5C}, which has a lower cleavage rate than the original Dz^{5A} mutant, but shows favourable NMR features. To overcome intrinsic difficulties of nucleic acid NMR spectroscopy, several approaches were introduced to yield structural restraints, such as ^{19}F- and spinlabel-derived distances, exact NOEs, homology restraints and residual dipolar coupling (RDC).
In complex with its RNA substrate, the Dz's catalytic loop does not show fast random structural reorganisation, but unlike to the also RNA-cleaving 8-17 DNAzyme no permanent WC base pairs could been detected. Instead, the Dz loop region is twisting around the RNA strand in height of the cleavage site, while the binding arms are crossing each other. The deoxynucleotide at position +1 has to flip out to allow the bypassing of the opposite strand of the Dz binding arm. Their negatively charged phosphate backbone come in close proximity to each other, and counterions are needed to neutralize their electrostatic repulsion and to bring the Dz:RNA complex in a homogenious conformation. Binding data derived from titrations with $MgCl_2$, $MnCl_2$ and $Co(NH_3)_6$ are in agreement with this hypothesis and shifts of NOESY cross-peaks between H8-dG$_{+2}$ and H2"-dG$_{-5}$ with increasing Mg^{2+} concentration are a strong indication for the ionic stabilizing at the crossing point of the binding arms (Fig. 3.17 I-IV).
The first two nucleotides of the catalytic loop are also involved in metal ion binding according to Mn^{2+}-induced PRE rates and, presumably, bring Dz loop and RNA cleavage site into the correct spatial position for an active pre-catalytic state (Figure 3.17 V). The nuclei close to the scissile bond and the 5' site of the Dz loop show the largest CSP during Mg^{2+} titration, however only dG$_1$ and dG$_2$ show direct binding of Mn^{2+}. Elevated $MgCl_2$ and NaCl concentration on NMR spectra of Dz:RNA$^{2'F}$ have similar impacts, therefore the structural reorganisations to adopt an active conformation might not dependent on the valency of metal ion cofactors and the necessary interaction are mainly charge-driven. It can be assumed that the relative occupancy of active Dz:RNA states is fairly low, and increasing amounts of cationic metal ions are the main stabilizing factor.

Mutation of dG$_6$ and dG$_{14}$ have severe effects on Dz's activity in already published data [109, 174, 186] and both also show high PRE rates during $MnCl_2$ titration. Since no significant structural rearrangements with increasing ion concentrations could be detected in the 3' part of the Dz loop, it can be assumed that the binding site is only formed after adapting the catalytic active state. In the NMR-derived model the phosphate ester of the scissile bond is orientated towards nucleotide dC$_{13}$ and dG$_{14}$, and binding of an divalent metal, which is directly involved in hydrolysis, can be located at this position. $Co(NH_3)_6$ NOE contacts with dA$_{11}$ or dA$_{12}$ also support divalent ion interaction at this site of the loop. Interestingly, elevated concentrations of

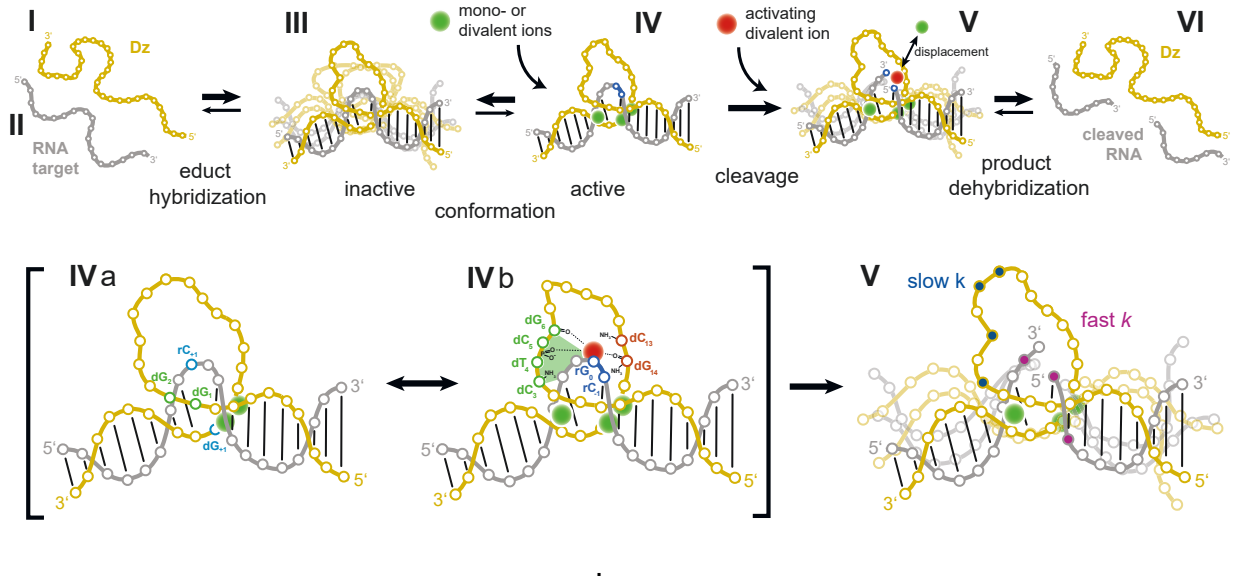


Figure 3.17: Hypothesized mechanism of 10-23 DNAzyme-mediated cleavage reaction. (top) Under optimal thermal conditions Dz (light brown) is forming a complex with its RNA target (grey) by hybridization of the binding arms (**I**+**II**). Dz:RNA complexes are forming slow-interchanging, mostly inactive states (**III**). An excess of metal cations stabilizes the active conformation, forming a binding site for a divalent ion, which is directly involved in the cleavage reaction, and coordinates the cleavage site (blue) in the correct position (**IV**). Binding of an additional divalent ion initiates cleavage of the scissile bond. Mg^{2+} at the active site can be displaced by an excess of inactivating (monovalent) ions (**V**). If the reaction temperature is higher than the melting temperature of the post-catalytic complex, RNA fragments are dissociating from Dz to allow another catalytic turnover (**VI**). (bottom) Dz:RNA is stabilized by two (not necessary sequential) cationic metal binding events. To form the Dz:RNA complex, Dz binding arms has to cross each other, which is enable by a broken GC basepair at position +1 (**IV** a, blue circles). Cations neutralize repulsive electrostatic interaction between negative charged phosphate backbones at the crossing point (**IV** a, green spheres). Nucleotides dG_1 and dG_2 are also involved into metal binding (**IV** a, green circles), presumably leading to a reorganisation of loop residues 2 to 6 (**IV** b, green circles). These nucleotides might bring the catalytic loop and the cleavage site (**IV** b, blue circles) into the active spatial arrangement (**IV** b, green area), which also forms an addition binding site involving nucleotides dC_{13} and dG_{14} (**IV** red circles). Functional groups, which are crucial for catalysis in functional assays, are shown in black, and possible coordination of the activating divalent metal ion (**IV** b, red sphere) is displayed as dotted lines. RT-NMR experiments reveal lower rate constants *k* for deoxynucleotides within the catalytic loop (**V**, blue filled circles) and higher *k* for ribonucleotides close to the cleave site (**IV**, purple filled circles).

Na^+ are inhibiting Mg^{2+}-induced catalysis of Dz, while they increase Dz activity in presence of Mn^{2+} ions (data not shown). Based on the Pearson concept, or Hard and Soft Acid and Base (HSAB) model [122], the Mn^{2+} ion has a dramatically higher affinity to form ionic interactions with nitrogen and oxygen atoms of nucleic acids and is more capable of loosing its hydration shell than Mg^{2+} ions [51,62]. Therefore, Mg^{2+} might be easier displaceable by Na^+ at the active site than Mn^{2+}, which explains the contradicting effects of NaCl on Dz catalysis in presence of the two different divalent ions. Activity studies were performed using the different properties of Mg^{2+} and Mn^{2+} in the HSAB concept by replacing systematically ”hard” oxygen with ”soft” sulfur atoms and reveal, that especially oxigens of guanine keto groups dG_6 and dG_{14} as well as of the phosphate ester between dT_4 and dA_5 [109] are crucial for Dz activity and are probably involved in the coordination of a divalent ion. However, Mn^{2+} binding could only be verified for dG_6 and dG_{14} in NMR experiments.

The active loop structure is presumably mediated by non-WC nucleotide contacts like base-phosphate (BPh) or base-base stacking interactions. Although the atomic interplay is unclear, loop stabilization effects could be especially located at nucleotides close to the binding site and at loop positions 1 to 6, whose importance is also shown by mutational studies of Zaborowski *et al* [186]. These deoxynucleotides do not seem to be directly involved in the catalytic mechanism, but they might be important for the correct positioning of loop and cleavage site. This can be exemplary shown for residue 5, which leads to an active Dz conformation with an adenosine (Dz^{5A}), cytosine (Dz^{5C}) or abasic nucleotide, but renders Dz inactive when substituted by guanine (Dz^{5G}), thymine or a C3 spacer. Notably, Dz^{5C} shows a reduced catalytic rate compared to Dz^{5A}, while the 5A substitution leads to severe line broadening in NMR spectra recorded in presence of Mg^{2+}. This line-broadening effect suggests a slower exchange between in-active and active state of Dz^{5A}:RNA in respect to the Dz^{5C}:RNA complex, which might indicate a more persistent cleavage-competent conformation and therefore a higher catalytic rate. RT-NMR experiments performed with Dz^{5C} also shows higher rates for residues at the cleavage site than in the Dz loop, which might be explained with slower unwinding process after cleavage.
The performed studies reveal a complex picture of the Dz-mediated catalytic mechanism with several interchanging conformational states. The novel NMR-derived structure shows a distinct fold in which Dz is twisting around the RNA strand and highlights the proximity of catalytic relevant residues. It could be demonstrated, that Dz activity is limited due to the low occupation of active structures and that this is highly modulated by electrostatic effects of cationic cofactors. This dependency might be the main obstacle for creating therapeutic Dz variants, which show sufficient activity *in vivo*. The presented structural model and mechanistic insights of the 10-23 DNAzyme might serve as foundation for future engineering of highly active and therapeutically relevant DNAzyme variants.

3.4 Material & Methods

Nucleic acid sequences. Oligomers for kinetic assays and ITC were ordered from Biomers, Ulm, Germany, NMR-grade as well as all modified nucleic acids from BioSpring Frankfurt (Main), Germany and isotope-enriched Dz oligomers from Silantes, München, Germany. When DNA and RNA oligomers of the same complex were deriving from different distributers, nucleic acid concentration was determined by absorption at 260 nm using extinction coefficients of 31922 $M^{-1}cm^{-1}$ and 176800 $M^{-1}cm^{-1}$ for Dz^{5C} and RNA target, respectively. Table 3.2 shows all used nucleic acid sequences and their abbreviations.

Denaturing polyacrylamid gel electrophoresis (SDS-PAGE). Integrity of NMR samples was analyzed by Hannah Rosenbach and Dr. Julian Victor using denaturing SDS-PAGE. Oligomers were denatured at 73 ℃ for 5 min, followed by an incubation for 10 min at room temperature. Analysis of the samples was carried out on 18% polyacrylamide gels with 7 M urea buffered with Tris-borate EDTA buffer (TBE) for 1 h at 20 W. For visualization the gel was incubated in

Table 3.2: Used oligomers in these studies.

Oligomer	Sequence
Dz^{5A}	5'-TTGGGGTAAGGCTAGCTACAACGAGGTGCATGT-3'
Dz^{5C}	5'-TTGGGGTAAGGCTCGCTACAACGAGGTGCATGT-3'
Dz^{5G}	5'-TTGGGGTAAGGCTGGCTACAACGAGGTGCATGT-3'
Dz^{6xF}	5'-TTGG^{F}G^{F}GTAAGG^{F}CTCGC^{F}TACA^{F}ACG^{F}AGGTGCATGT-3'
Dz^{EdU-8}	5'-TEGGGGTAAGGCTCGCTACAACGAGGTGCATGT-3'
Dz^{EdU8}	5'-TTGGGGTAAGGCTCGCEACAACGAGGTGCATGT-3'
Dz^{EdU+7}	5'-TTGGGGTAAGGCTCGCTACAACGAGGTGCAEGT-3'
DNA^{C}	5'-TTGGGGTAAGGTGCATGT-3'
RNA	5'-acaugraccguuaccccaa-3'
$RNA^{2'F}$	5'-acaugraccg^{F}uuaccccaa-3'
RNA^{FAM}	5'-FAM-acaugraccguuaccccaa-3'
RNA^{FRET}	5'-FAM-acaugraccguuaccccaa-BHQ-3'

Deoxynucleotides in upper case, ribonucleotides in lower case
^{F}: Fluorine substitution at 2' position
E: 5-ethynyl-2'-deoxyuridine (EdU); FAM: fluorescein; BHQ: black hole quencher

Tris-borate EDTA (TBE) buffer containing a 1:10,000 dilution of GelRed (Fremont, California, USA) for 1 hour. Images were acquired using the ChemiDoc MP System (Bio-Rad, Hercules, California, USA).

Cleavage assays performed with electrophoretic mobility shift. All assays were carried out by Hannah Rosenbach and Dr. Julian Victor. Gel-based kinetic assays were performed for rather slow hydrolysis reactions. Those kinetic assays were performed with 0.4 µM RNA^{FAM} substrate and 0.4 µM Dz in a total volume of 10 µl in 50 mM Tris-HCl pH 7.5 (or Na_2HPO_4/NaH_2PO_4 for pH dependent reactions) with 0.1 mM EDTA and 100 mM NaCl in the presence and absence of Mg^{2+}, Co^{2+}, or $Co(NH_3)_6$ for up to 260 min at 37°C. RNA and Dz were denatured in buffer in the absence of Mg^{2+} at 73°C for 5 min, followed by an incubation for 10 min at room temperature. Samples were taken at defined time points and the reaction was stopped in 95% formamide with 25 mM EDTA followed by boiling the sample for 5 min at 95°C. Separation of the samples was carried out on 18% polyacrylamide gels with 7 M urea buffered with Tris-borate EDTA buffer (TBE) for 1 h at 20 W. Visualization of RNA^{FAM} substrates was carried out by fluorescence detection. Images were acquired using the ChemiDoc MP System (Bio-Rad, Hercules, CA, USA).

Fluorophore-quencher based kinetic assays. All kinetic assays were carried out by Hannah Rosenbach, Dr. Julian Victor or Dr. Manuel Etzkorn. Hydrolysis reactions were carried out with 0.1 µM RNA and Dz in 50 mM Tris-HCl pH 7.5 with 0.1 mM EDTA and with different concentrations of Na^{+}, Mg^{2+} or Mn^{2+} at 37°C. Control experiments were performed either in the absence of the Dz but in the presence of Mg^{2+} or Mn^{2+}, or in the absence of the divalent metal ion but in the presence of the Dz. For the hydrolysis assay 0.8 µM of RNA^{FRET} substrate

and Dz were denatured in buffer with 0.8 mM EDTA in the absence of Na^+ and Mg^{2+} or Mn^{2+} at 73°C for 5 min, cooled down to room temperature for 15 min, and 20 µl of the solution were pipetted into the wells of a 384 well non-binding microplate (Greiner Bio-One, Kremsmünster, Austria). The plate was sealed with tape (Polyolefine Acrylate, Thermo Scientific, Waltham, MA, USA), placed inside the plate reader (CLARIOStar, BMG LABTECH, Ortenberg, Germany) and equilibrated to 37°C for 30 min. The reaction was started by the injection of 10 µl $MgCl_2$ in buffer. Data points were obtained every 5, 3, or 2 s, depending on the $MgCl_2$ concentration and therefore on the speed of the reaction. Excitation and emission wavelength were 484 nm and 530 nm, respectively.

ITC measurements. All measurements, data analysis and plotting were carried out by Dr Wolfgang Hoyer. ITC was performed in 50 mM Tris buffer pH 7.5, with or without addition of 100 mM NaCl, at 30°C on a Microcal iTC200 calorimeter (GE Healthcare, Chicaco, IL, USA). Affinities were determined with the pre-formed complex of Dz and RNA substrate as titrant in the cell at a concentration of 100 µM, and 10 mM $MgCl_2$ as titrant in the syringe. Dissociation constants were obtained from a nonlinear least-squares fit to either a 1:1 binding model or a n:1 binding model with n identical Mg^{2+} binding sites per Dz:RNA complex using the Origin software (MicroCal) provided with the calorimeter.

NMR data acquisition, processing and analysis. For NMR samples, oligomer stocks were solvated in water, mixed together with a 2-fold concentrated measurement buffer and heated at 73 °C for 5 min. If not stated otherwise, 200 µl samples were measured in 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, and 10% D_2O at 37°C in a 3 mm sample tube. Experiments were performed on Bruker spectrometers (Bruker, Billerica, USA) with 600, 700, and 900 MHz basic carrier frequencies using a H/N/C triple-resonance, and in some cases a H/N/C/P quadruple-resonance cryo-cooled probehead. [^{1}H,^{1}H]-TOCSY experiments were measured with 80 ms mixing times and [^{1}H,^{1}H]-NOESY experiments with 300 ms mixing time, if not used for NOE build-up determination. NMR data were processed using Topspin 4.0.6 (Bruker) or NMRpipe 8.9 [41] and spectra were analysed and plotted with Topspin 4.0.6, Cara 1.9.1.5 [71] or Sparky 3.114 [83]. Data analysis and plotting was performed using OriginPro 9.0G (OriginLab, Northampton, USA), Matlab R2019a (MathWorks, Natick, USA) or in-house Python-based scripts. Samples with higher D_2O content were prepared as described, lyophilized over night and resuspended in water containing the desired grade of deuteration. Tables A5-8 list all performed NMR experiments with their relevant acquisition parameters.

Peak assignment. High-resolution spectra of most of the Dz states and variants as well as $RNA^{2'F}$ and DNA^{C}:$RNA^{2'F}$ were recorded using sample concentration between 100 and 750 µM. However, most chemical shifts were assigned for Dz^{5C}:$RNA^{2'F}$ in presence of 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, and 10% D_2O at 37°C and are listed in Tables A1+A2. Sequential assignment of DNA was performed by finding matching NOESY crosspeaks between base protons H6 or H8 and protons H1', H2', H2", and H3' of the own and

the 5' adjacent sugar moiety. TOCSY crosspeaks between H1' and H2', H2", or H1' were used to distinguish between different spin systems. NOE contacts between H6 or H8 of adjacent nucleotides were also used for sequential assignment. Crosspeaks between H6 and H5, or H7*, were used to map spin systems to pyrimidines C/U, or T, in the nucleic acid sequence. RNA peaks were assigned after Dz assignment due to the highly reduced set of unassigned crosspeaks in the chemical shift region of H6/H8, H1' and H3'. Since peaks are highly overlapping for H4', H5',H5" and ribonuclear H2', only few of them could be assigned, as well as amino and imino peaks, due to their lacking signal-to-noise ratio. Adenosine H2 could only assign unambiguously with preliminary structural knowledge and after full assignment of H6 and H8. H2' and H2" as well as H5' and H5" could be distinguished via their different NOE build-up rates towards H1' of the same nucleotide. Carbon nuclei were assigned using 2D [^{15}N,^{13}C]-HSQC-TOCSY and 3D [^{15}N,^{13}C]-HSQC-NOESY experiments with 500 µM ^{15}N,^{13}C-Dz^{5C}:RNA$^{2'F}$. Only deoxyribunucleic C1' could be fully assigned.

NMR titration experiments. For NMR titration experiments 2 µl aliquots of stock solutions were subsequently added into an NMR tube containing 200 µl sample and solutions were mixed by 10 times inverting the tube. Repetition of same experiments showed that this method gives reproducible data. Mg^{2+} titration was performed using 200 µM Dz^{5C}:RNA$^{2'F}$ and Dz^{5A}:RNA$^{2'F}$ complex in 50 mM Tis/HCl buffer pH 7.5, 100 mM NaCl, and 10% (v/v) D_20 and aliquots of $MgCl_2$ stock solution with concentrations of 25.25, 25.5, 51.5, 104, 315, 530, 1.070, and 3.240 mM were added. The experiment with Dz^{5C}:RNA$^{2'F}$ was also performed in absence of NaCl. The chemical shift and intensity of separated NOESY and TOCSY peaks were extracted from recorded spectra. K_D values were calculated by fitting CSP data with Equation 3.1 or estimated from peak intensities by fitting with Equation 3.2.

$$CSP = \frac{CSP_{\text{max}} \cdot [Mg^{2+}]}{K_K + [Mg^{2+}]} \tag{3.1}$$

$$Int. = Int_{\text{max}} - \frac{Int_{\text{max}} \cdot [Mg^{2+}]}{K_K + [Mg^{2+}]} \tag{3.2}$$

Additionally, maximum CSP values were extracted from peak shift differences of spectra of Dz^{5C}:RNA$^{2'F}$ at 0 and 50 mM $MgCl_2$. Mn^{2+} titration were performed using 200 µM Dz^{5C}:Dz$^{2'F}$ complex in 50 mM Tis/HCl buffer (pH 7.5), 100 mM NaCl, 1 mM $MgCl_2$ and 10% (v/v) D_20 and aliquots of $MnCl_2$ stock solution with concentrations of 50.5, 51, 103, 312, and 525 µM were added. The intensity of separated NOESY and TOCSY peaks were extracted from recorded spectra and relaxation rates were calculated using a simple exponential decay function (Equation 3.3). Mean PRE rates for individual protons were calculated from crosspeak PRE rates involving the respective proton.

$$Int = \exp -k \cdot [Mn^{2+}] \tag{3.3}$$

^{19}F-derived experiments. ^{19}F-derived experiments were performed by Dr. Helena Kovacs at 700 basic proton carrier frequency with sample concentrations of 100 µM Dz^{6xF}, 750 µM

Dz^{6xF}:$RNA^{2'F}$ and 500 µM Dz^{5C}:$RNA^{2'F}$. ^{19}F 1D spectra were recorded for Dz^{6xF}, Dz^{6xF}:$RNA^{2'F}$ and Dz^{5C}:$RNA^{2'F}$, ^{19}F-detecting [^{19}F,^{1}H]-HOESY and ^{1}H-detecting [^{1}H,^{19}F]-HOESY experiments were performed with Dz^{6xF}, Dz^{6xF}:$RNA^{2'F}$. In the proton-detecting version fluorine shifts for adjacent nucleotides dG_{-6} and dG_{-5} could be assigned directly. Fluoro-saturated STD experiment of Dz^{5C}:$RNA^{2'F}$ were performed with 3 s saturation of 250 Hz pulse width at -206.6 ppm ^{19}F frequency. For STD experiments with Dz^{6xF}:$RNA^{2'F}$ 50 and 80 Hz saturation pulse width at ^{19}F frequencies of -198.46, -201.15, 201.635, -202.1, and -202.8 ppm were used. ^{19}F-substituted nucleotides could be assigned by using strong STD peaks of geminal H2'and vincial H1' peaks split by J_{FH} coupling, which can also be detected in homonuclear [^{1}H,^{1}H] spectra. Same signals also appear in STD spectra with saturation frequencies close-by, but with decreased intensities. This effect is reduced by using saturation pulses with smaller width as 50 instead of 80 Hz, but leads to weaker STD signals. Analysis of the different intensities of the known H1' and H2' peaks at various saturation frequencies and the two different pulse width enabled to assign fluorine signals specifically. Long-range STD signals could only be assigned for protons in less crowded chemical shift regions at 1-3 ppm and 6-8 ppm.

Paramagnetic spinlabels. Spinlabelling of Dz nucleic acids was performed by Dr. Julian Victor. Three sequences were ordered from BioSpring with a substitution of thymidine by 5-ethynyl-2'-deoxyuridine (EdU) at positions dT_{-8}, dT_{8}, and dT_{+7} (Dz^{EdU-8}, Dz^{EdU8}, Dz^{EdU+7}). dEdU nucleotides were modified with a 4-azido-2,2,6,6-tetramethylpiperidinyl-1-oxy (TEMPO) spinlabel via Cu(I) click chemistry. Briefly, a 5fold excess of TEMPO was mixed with 2 mM EdU-modified Dz to a total of 10 µl reaction volume and 3 µl of freshly prepared 0.1 M catalyst solution (1 part CuBr, 2 parts TBTA) were added. The solution was mixed vigorously and incubated at 25 °C for 4 h. For purification, 100 µl 0.3 M NaOAc and 1 ml cold ethanol were added, incubated on ice for 20 min, and centrifuged 1h at max rpm (I have to check). The supernatant was discarded and the pellet washed twice with cold 70% (v/v) ethanol. Homonuclear TOCSY and NOESY spectra were recorded for 220 µM Dz^{EdU-8}:$RNA^{2'F}$ and 350 µM Dz^{EdU8}:$RNA^{2'F}$, and only a TOCSY spectrum was recorded for 100 µM Dz^{EdU+7}:$RNA^{2'F}$. 2 mM ascorbic acid was added and the samples were incubated for 1 h at 37 °C to reduce and inactivate the paramagnetic spinlabel. Spectra of the complexes with inactivated spinlabel were recorded under the same conditions as the spectra before. Crosspeaks of spectra for complexes with activated and inactivated TEMPO were assigned by comparison with a high-resolved spectra of Dz^{5C}:$RNA^{2'F}$, peak intensity ratios of all accessible crosspeaks were extracted and intensity ratios of all crosspeaks derived from the same proton were averaged. Averaged intensities only derived from one crosspeak or with a higher standard deviation of 0.8 were excluded as restraints for structure calculations.

T_1, T_2 relaxation and hetNOE. T1 and T2 relaxation experiments were conducted with 500 µM ^{13}C,^{15}N-Dz^{5C}:$RNA^{2'F}$ using *constant-time* (ct) modified versions of pulse programs *hsqct1etgpsi3d* and *hsqct2etgpsi3d* (Topspin 4.0.6). Delay for inversion recovery were set as 20, 100, 250, 500, and 1000 ms for T1 determination and 1, 2, 3, 4, and 6 spin-echo cycles

of 22.4 µs each was used for T2 determination. H1'-C1' crosspeak intensities were extracted from each spectrum and fitted with Bloch equations 1.3 and 1.4. HetNOE spectra could not be recorded as a ct-version, therefore splitting due to C-C coupling lead to severe signal overlapping. Data was processed with maximum entropy convolution using J^2_{CC} of 145 Hz to suppress the splitting effects. HetNOE enhancement was derived by calculating the H1'-C1' crosspeak intensity ratios of saturated spectrum and unsaturated reference.

NMR melting experiments. TOCSY spectra of single-stranded Dz^{5C}, Dz^{5C}:$RNA^{2'R}$ and Dz^{5C}:products were recorded at temperatures between 60 and 15 °C. with decreasing gradient. Peaks were assigned at temperatures close to 37 °C, their individual temperature shifts were followed and peak intensities extracted. Comparison of recorded Dz^{5C}:products spectra before an after melting showed that the procedure is reversible (Figure A18b). Pyrimidine crosspeak intensities were normalized to their intensity close to 37 °C.

NOE build-up and eNOE. NOE build-up data was recorded with an array of [^{1}H,^{1}H]-NOESY spectra of 500 µM Dz^{5C}:$RNA^{2'}$ at 600 MHz basic transmitter frequency using mixing times of 0.5, 5, 10, 20, 30, 40, 80, 120, 180, 260, 400, and 800 ms. NOE crosspeak intensities extracted by peak integration were fitted using a double-exponential build-up function derived from the Solomon equation:

$$I(t_{mix}) = I_{max} e^{-\rho\sigma t_{mix}} (1 - e^{-2\sigma t_{mix}}) \tag{3.4}$$

where $I(t_{mix})$ is the peak intensity at a given mixing time, I_{max} the theoretical maximal intensity, σ the cross-, and ρ the autorelaxation rate of the NOE build-up. A precise σ can be derived with the eNORA software [119], since it takes into account relative intensities, individual ρ, transfer mechanism between upper and lower diagonal peaks, and especially proton-proton spin diffusion effects based on a structural model. Libraries of the published eNORA scripts were modified to work with nucleic acids and Xplor-NIH 2.49 [148, 148] structure files. Since lower mixing times lead to unreliable data, only intensities derived from t_{mix} between 40 and 800 ms were analysed. A 'full-matrix' approach was used and an overall rotational time τ_C of 3.5 µs was assumed.

RT-NMR experiments. RT-NMR experiments were carried out with 200 µl Dz^{5C}:RNA with 100 mM NaCl after addition of Mg^{2+}. The lag time between sample mixing and start of acquisition was approximately 2 min. A series of 280 [^{1}H] 1D experiments was recorded over a time course of 12 h after addition of 1 mM $MgCl_2$, resulting in a complex of Dz^{5C} and cleaved RNA (Dz^{5A}:products). For a 2D version of 200 µM complex a homonuclear TOCSY of 3h duration with 64 increments was started after reaction initiation, followed by seven identical experiments subsequently. A longer [^{1}H,^{1}H]-TOCSY and a [^{1}H,^{1}H]-NOESY experiment were also recorded 24 h after reaction start as a product reference. Analysis of RT spectra and extraction of rate constants was performed by Dr. Manuel Etzkorn. Briefly, linewidths in direct (λ_2) and indirect dimension (λ_1) of isolated peaks for educt (A) and product state(s) (B) were extracted from

reference spectra. Time-dependent changes of the NMR signal ($S(t)$) by state-transition were simulated for the indirect dimension (t_1) before FT. The direct dimension was simulated in frequency space (ω_2) using Gaussian peak shapes ($G(\omega_{n,2},\lambda_{n,2})$). Spectra were simulated with three different conformations (n) and varying rate constants (k) using equation 3.5:

$$S(t_1, \omega_1) = \sum_n P_n(k, t) \cdot e^{(2\pi i \omega_{1,n})t} \cdot G(\omega_{2,n}, \lambda_{2,n}) \tag{3.5}$$

where the population profiles (P) of the simulated transition in a 2-state model had the form:

$$P_A = A \cdot e^{-kt} \tag{3.6}$$

$$P_{B_1} = (1 - B_2) \cdot (1 - e^{-kt}) \tag{3.7}$$

$$P_{B_2} = B_2 \cdot (1 - e^{-kt}) \tag{3.8}$$

Simulation itself were carried out using the Fast Fourier Transformation (FFT) algorithm implemented in Matlab R2018b (MathWorks) and the respective population profiles. To account for the occurrence of multiple B-states with variable intensity ratios, B_2 (in Equations. 6b,c and 7b,c) was included as another variable parameter, in addition to k, via nested loops in the Matlab script. For each set of parameters (B_2, k) the simulated spectral extract was subtracted from the experimental data after normalizing both data sets to their maximal intensity. The parameter set for which this difference reaches its minimum was selected as best fit condition (see Figure 3.16g). In this analysis only a 2-state model was used, however simulation according to a 3-state model resembles the lineshapes slightly better (data not shown).

Residual Dipolar Coupling. RDC constants could be extracted from decoupled [^{1}H,^{13}C]-HSQC spectra of 500 mM ^{15}N,^{13}C-Dz^{5C}:RNA$^{2'F}$ recorded at 20 °C in presence and absence of 10 mg/ml phage Pf1 as an alignment medium. RDC of align Dz^{5C}:RNA$^{2'F}$ adds on top of peak splitting due to scalar coupling, therefore the differences in peak splittings between both spectra yield the dipolar coupling constants. For peak assignment at 20 °C, also [^{1}H,^{13}C]-HSQC experiments without decoupling were acquired. The extracted RDC constants and the structure with lowest-energy profile from the initial structure calculation were used as input for the Redcat [169] software of *calcTensor* of Xplor-NIH [148, 148] to calculate a final field tensor of 3.23 magnitude (Da) and a 0.364 rhombicity (R). The calculated field tensor and a by 20% reduced set of RDC constants were used as input of the final structure calculation. Theoretical RDC values were calculated using *calcTensor* from the resulted lowest-energy structure of the *ab-initio* calculation and correlated with the experimental data.

Structure calculation. Backbone dihedral angles α, β, γ, δ, ε, and ζ were restrained to A-form, sugar dihedral angles $\nu_{1\text{-}3}$ to 3'-*endo* pucker and χ to *anti* conformation for nucleotides which do not show significant CSP difference between Dz^{5C}:RNA$^{2'F}$ and DNA^{C}:RNA$^{2'F}$ for these residues. Additionally, nucleotides were set to *syn* or *anti* conformation according to their

relative NOE build-up rates. All angular restraints used for structure calculation are listed in Table 3.3.

Table 3.3: Dihedral angular restraints for Dz^{5C}:$RNA^{2'F}$.

Name	Involved atoms	Restraint	Conformation	Restraint nucleotides
Backbone dihedrals				
α	O3'-P-O5'-C5'	-60° ±40°	A form	
β	P-O5'-C5'-C4'	-180° ±50°	A form	DNA: -9 to -5 / +5 to +9
γ	O5'-C5'-O4'-C3	60° ±30°	A form	RNA: -9 to -5 / +5 to +9
ϵ	C4'-C3'-O3'-P	-160° ±50°	A form	
ζ	C3'-O3'-P-O5'	-70° ±50°	A form	
Sugar puckering				
ν_1	O4'-C1'-C2'-C3'	-20° ±10°	3'-*endo*	DNA: -9 to -5 / +5 to +9
ν_2	C1'-C2'-C3'-C4'	35° ±5°	3'-*endo*	RNA: -9 to -5 / +5 to +9
ν_3/δ	C5'-C4'-C3'-O3'	80° ±20°	3'-*endo*	
ν_1	O4'-C1'-C2'-C3'	35° ±5°	2'-*endo*	
ν_2	C1'-C2'-C3'-C4'	-35° ±5°	2'-*endo*	none restraint
ν_3/δ	C5'-C4'-C3'-O3'	145° ±20°	2'-*endo*	
Nucleobase orientation				
χ	O4'-C1'-N1-C2	180° ±90°	*anti*	DNA: -9 to 3 / 6,7,9 / 12 to +9
	O4'-C1'-N9-C4			RNA: -9 to -3 / +3 to +9
χ	O4'-C1'-N1-C2	0° ±90°	*syn*	DNA: 4,5,8,11
	O4'-C1'-N9-C4			RNA: -2 to 0

Conventional NOE restraints were derived from crosspeak integration of a [^{1}H,^{1}H]-NOESY with 300 ms mixing time. Peak intensities were divided by the average intensity of all H6-H5 crosspeaks and grouped into weak, intermediate or strong contacts if the ratio is below 0.1, between 0.1 and 0.5, or high than 0.5, respectively. Weak contacts were set as restraints of 4.1 (2.8-6.7) Å distance, intermediate contacts as 3.2 (2.0-5.6) Å, and strong contacts as 2.4 (1.2-3.6) Å, where the values in the brackets represents the lower and upper restraint limits. Restraints derived from heteronuclear fluorine-proton NOE contacts were set as 5.0 (1.0-7.0) Å distance, and as 5.0 A (1.0-8.0) Å if derived from STD experiments. PRE ratios of the spinlabels were group in strong, intermediated, or weak contacts, if ratios are below 0.2, between 0.2 and 0.4, or above 0.4, respectively, and distance restraints of 10.0 (1.0-30.0) Å, 25.0 (10.0-45.0) Å, or 50.0 (25.0-149.0) Å were used. The distances should also take possible spinlabel rotation into account. Additionally, hydrogen bond distances for WC base pairs -9 to -1 and +2 to +9 were restrained according to [60] and restraints for base pair planarity were used.
Structure calculations were performed with the restraint molecular dynamics software Xplor-NIH 2.49 [148, 148] using the *fold.py* script and 200 runs starting from a randomized initial structure including or excluding RDC restraints. The 20 output structures with the lowest overall

energy potential were selected and further analysed to find new long-range NOE restraints. eNOE contacts were calculated from promising structures and were replacing their conventionally derived counterparts. The structure calculation was re-run with the new restraint data and the process was repeated iteratively, until no new long-range NOE and eNOE could be found. Refinement structure calculation was performed using the script *refine.py* and the lowest energy structure from the *ab initio* calculation with RDC restraints. Fore refinement runs the starting temperature of the simulated annealing stage was reduced from 3500 K to 350 K.

4 NMR Spectroscopy of Protein Systems

This part of the thesis covers NMR studies performed with protein systems. Section 4.1 and 4.2 are published or submitted manuscripts, and the full authors affiliations and contributions are listed in the appendix. Section 4.3 summarizes studies performed with MC4R ligands and the data is unpublished, yet. Experiments and analyses of section 4.3, which are not originally performed by the thesis' author, are also shortly described in Material and Methods and the contributed co-worker is mentioned.

4.1 Molecular Architecture of a Network of Potential Intracellular EGFR Modulators: ARNO, CaM, Phospholipids, and the Juxtamembrane Segment

Aldino Viegas, Dongsheng M. Yin, Jan Borggräfe, Thibault Viennet, Marcel Falke, Anton Schmitz, Michael Famulok, Manuel Etzkorn

4.1.1 Introduction

The epidermal growth factor receptor (EGFR) is a major regulator of proliferation in epithelial cells. Since its misregulated activation can lead to hyperproliferation and the development of cancer, an intricate regulatory network to control EGFR activity has evolved comprising systemic and cell-autonomous elements [36]. Key to the regulation of EGFR activity is the receptor´s intrinsic autoinhibition, on which the regulatory network is built. The importance of this autoinhibition is evident from several mutations which disrupt the autoinhibition and are linked to specific types of cancer [84]. In general, the autoinhibited state can be released by formation of an asymmetric EGFR dimer in which one kinase domain activates the other one [187]. This asymmetric dimer is stabilized by the juxtamembrane (JM) segments of both intracellular domains of the involved monomers. The C-terminal part of the JM segment of the activated kinase functions as a "latch" or "cradle" for the activator kinase [67, 128]. In addition, the N-terminal parts of both JM segments are thought to form an antiparallel coiled-coil enhancing the affinity of the monomers for each other [67]. The formation of the antiparallel coiled-coil requires the C-termini of the transmembrane helices to be separated from each other which on its turn is coupled to the ligand-bound conformation of the extracellular domain and thus confers regulation by EGF. In the inactive state, the basic JM segment binds to acidic phospholipids of the inner leaflet of the plasma membrane, and this interaction contributes to the autoinhibition of the receptor [149].
In addition to its function in relaying the conformational changes induced by EGF-binding from the extracellular domains to the kinase domains, the JM segment is a site of modulation of EGFR activity by intracellular factors. Previous studies identified a number of hot spots associated e.g. with kinase interaction (T654 and T669) [63, 106], activation of the receptor (V665 and L680) [128] or a possible conformational constraining of the receptor (R645 – R657) [123]. When e.g. threonine-654 in the JM segment is phosphorylated by protein kinase C the activity of the EGFR is attenuated [37,40,52,63] probably due to inhibition of EGFR dimerization [164]. Phosphorylation of another threonine (T669) in the JM segment by extracellular signal-regulated kinase (ERK) reduces EGFR downregulation [86]. Recently the tumor necrosis factor receptor associated factor 4 (TRAF4) has been reported to interact with the C-terminal part of EGFR-JM to

promote receptor dimerization [24]. The cytosolic protein calmodulin (CaM) binds to the N-terminal part of the JM segment in a calcium-dependent manner [2, 94] enhancing EGFR activation [85, 96].
In this work we report on the ability of ARNO (ADP ribosylation factor nucleotide binding-site opener), a member of the cytohesin family of guanine nucleotide exchange factors, to bind EGFR's JM domain. While ARNO has been proposed to function as activator of the EGFR [16, 120, 121], the underlying mechanism has not been determined. Here, we provide a comprehensive *in vitro* analysis of the determinants that define the ARNO-JM interaction and characterize the interaction at the molecular level in the absence and presence of a membrane environment. We identify the JM-binding site in the Sec7 domain of ARNO and show that JM's interaction with ARNO-Sec7 displays large similarities to its interaction with CaM, pointing to the speculation that ARNO and CaM may modulate EGFR in a similar manner. Our data also reveal that ARNO-Sec7, CaM as well as lipid bilayers containing anionic phospholipids compete for overlapping binding sites on the JM segment. Moreover, we show that additional factors including auto inhibition for ARNO, Ca^{2+} availability for CaM and lipid composition for JM's membrane association, are capable to further regulate this competitive network of EGFR-JM interaction partners. While we here focus on the structural and biophysical characterization of this network under defined *in vitro* conditions, our findings are consistent with previous findings in living cells and in tumor tissue [16, 120, 121] and should stimulate future studies of this important aspects in EGFR signaling.

4.1.2 Results and Discussion

The JM segment of the EGFR interacts with the Sec7 domain of ARNO
To investigate whether ARNO interacts with the EGFR we carried out microscale thermophoresis (MST) measurements of selected isolated domains. Since ARNO is a cytosolic protein, only EGFR constructs comprising the intracellular domain were considered (Figure 4.1a). Our MST data show that the EGFR intracellular domain (ICD) indeed interacts with the Sec7 domain of ARNO (Sec7) (Figure 4.1b, black). The other major domain of ARNO, the pleckstrin homology domain (PH), did not show binding under the applied conditions (data not shown). Interestingly the purified juxtamembrane segment (JM) alone displays a comparable binding behavior to ARNO-Sec7 as EGFR-ICD (Figure 4.1b, blue). While EGFR-ICD and EGFR-JM bind ARNO-Sec7 with similar affinity (K_D of about 50 µM), an EGFR-ICD construct lacking the first 27 amino acids of the JM segment (EGFR-ICD$\Delta JM_{1\text{-}27}$) does not show interaction (Figure 4.1c, grey), which is also true for a scrambled version of JM (JM_{SC}) containing the same amino-acids, but randomly redistributed (Figure 4.1c, pink, see Methods for full sequence). Our data demonstrate that ARNO-Sec7 interacts with the EGFR *in vitro* and strongly suggest that this interaction is on the EGFR side mainly driven by the JM

segment.
Due to JM's key role in EGFR regulation [45,56,67,123,128,164] and its high potential for EGFR signaling modulation [2], we carried out a comprehensive NMR study to characterize the molecular architecture of the interaction of ARNO-Sec7 and EGFR-JM. Following the full resonance assignments of both domains (see Methods Table 4.1 and 4.2 and Figure A19 for details on data acquisition, resonance assignment and structural features) we performed NMR-based titration studies with Sec7 and JM to identify the interacting regions based on the chemical shift perturbations (CSP) induced by their binding partner. Figure 4.1d-g summarize the data from the point of view of the (^{15}N-isotope-labeled) JM segment. The presence of increasing amounts of (unlabeled) Sec7 induces characteristic concentration-dependent chemical shift perturbations for certain residues (Figure 4.1d). Plotting these chemical shift perturbations along the JM sequence clearly identifies the N-terminal half of JM, i.e. the JM-A segment [67], as the one involved in the interaction with Sec7. Furthermore, the NMR chemical shifts continuously change with increasing concentration of Sec7, revealing a rather transient interaction (NMR fast-exchange regime) with residue specific binding affinities (K_D) in the high µM range (see Figure A19e for more details).
In general, NMR chemical shifts, in particular of carbon Cα and Cβ nuclei, are robust indicators of secondary structure [14]. Analysis of the respective chemical shifts of the isolated JM segment points to the absence of a clear secondary structure when free in solution (Figure 4.1f). This observation is in good agreement with previous results [34, 102] in which JM was shown to behave mainly as random coil in the absence of membrane mimetics.

Addition of Sec7 did not lead to detectable ^{13}C chemical shift perturbations (data not shown), which suggests that the interaction with Sec7 does not induce a stable secondary structure in JM. However, the rather uniform shift of the affected peaks in the [^{1}H,^{15}N]-HSQC spectrum (Figure 4.1d) towards lower ^{1}H and ^{15}N frequencies would be in line with an increase in transient α-helical propensity upon Sec7 binding [14]. The JM segment of EGFR contains a high number of charged residues (see Figure A19c). In particular the JM-A segment comprises an unusually high number of positively charged residues (i.e. 9 out of 19 residues). To test whether the interaction with Sec7, which contains both negatively and positively charged regions (Figure A20), is driven by nonspecific electrostatic interactions we used again the scrambled JM construct (JM_{SC}) containing the same total amino acid composition but randomly redistributed. In line with the MST measurements (Figure 4.1c, pink), the NMR measurement (using ^{15}N-labeled JM_{SC}, Figure 4.1g) shows that the scrambled JM does not interact with Sec7, in clear contrast to wild-type JM under identical conditions (e.g. Figure 4.1d, purple). Our data therefore demonstrate that the absolute charge of JM is not key to the interaction and imply that the primary sequence of JM promotes a specific recognition

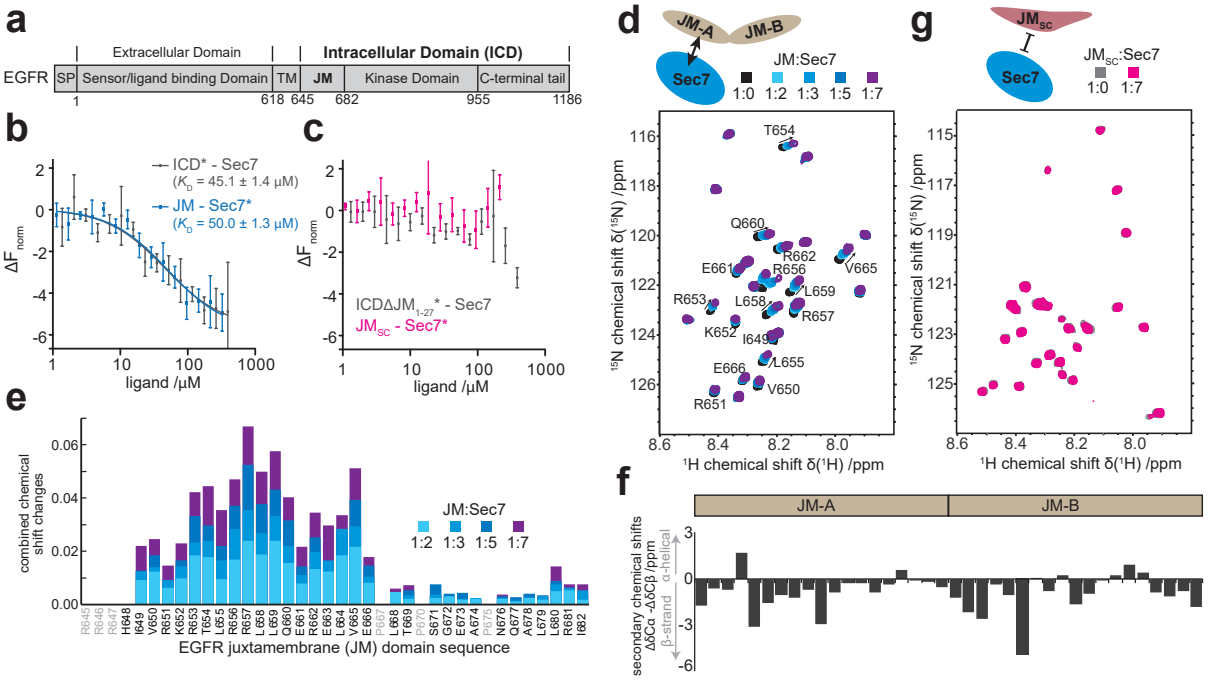


Figure 4.1: JM-Sec7 interaction as seen from the EGFR-JM side. (a) Schematic representation of EGFR's domain architecture (SP: signal peptide; TM: transmembrane domain; JM: juxtamembrane). (b+c) MST data for indicated EGFR and ARNO constructs. Here and in the subsequent figures the EGFR construct is always mentioned first and the fluorophore-labeled molecule is labeled with asterisk (n=3, mean±SD). (d) [^{1}H,^{15}N]-HSQC NMR spectra of ^{15}N-labeled EGFR-JM in the presence of increasing amounts of unlabeled ARNO-Sec7. (e) Chemical shift perturbations along the EGFR-JM sequence induced by the presence of indicated amounts of ARNO-Sec7. Grey labels indicate residues that were not observed. (f) Secondary chemical shifts (see Material and Methods for definition) as indicator for secondary structure of EGFR-JM when free in solution. (g) [^{1}H,^{15}N]-HSQC NMR spectra of a ^{15}N-labeled scrambled version of EGFR-JM (JM_{SC}) in the absence (grey) or presence of 7-fold excess (magenta) of ARNO-Sec7.

by Sec7.

Due to its good NMR properties [15] ARNO-Sec7 offers the appealing opportunity to investigate the interaction also from the cytohesin point of view. Consequently, we recorded a series of NMR experiments using ^{15}N-isotope labeled ARNO-Sec7 and non-labeled EGFR-JM (Figure 4.2). In line with the data obtained from the JM point of view (Figure 4.1), the presence of increasing amounts of EGFR-JM induced chemical shift perturbation for specific Sec7 residues (Figure 4.2b+c) reproducing the transient interaction of the two domains (NMR fast exchange regime) and pinpointing a specific JM-binding site of Sec7. Following the resonance assignment of the 21 kDa Sec7 construct (Figure A19, BMRB deposition code: 27761) distinct regions of the Sec7 domain can be identified that interact with the isolated JM segment (Figure 4.2d). The affected residues mainly cluster around helices E (5), F (6), G (7), H (8) and I (9) and the loop connecting helices I (9) and J (10) (helix nomenclature as in Mossessova *et al.* [108] and, in brackets, according to Betz *et al.* [15]. Highlighting the most affected residues in the Sec7 structure reveals a well-defined JM-binding interface (Figure 4.2e). While

the affected region partially overlaps with the negatively charged surface of Sec7 (see Figure A20), it also involves a high number of hydrophobic residues (about 13 in the central binding interface of Sec7 and 7 in JM), suggesting that ARNO-Sec7 interacts with EGFR-JM, in part, through an extended hydrophobic surface. In particular, a surface-exposed hydrophobic patch of residues in Sec7's helix H appears to be in the center of this interaction. Reducing the hydrophobicity of this patch by alanine substitutions of Y186, F190, I193 and M194, i.e. ARNO-Sec7(4A), indeed inhibits binding to JM as determined by MST (Figure 4.2f).
Of note, the observed binding site is also located in a region populated by residues crucial for the interaction of Sec7 with ARF1 [15, 31, 108]. ARF1 binding is prevented in the autoinhibited state in all cytohesin members when helix H forms intramolecular contacts with the linker and the polybasic region (pbr). Accordingly, ARNO lacking the polybasic region (ARNOΔpbr) loses this autoinhibition [44]. To test whether this autoinhibitory mechanism also plays a role for an interaction of ARNO with the EGFR, we carried out MST measurements using EGFR-ICD and either full-length ARNO or ARNOΔpbr (Figure 4.2g+h). Indeed, full length (autoinhibited) ARNO did not bind EGFR-ICD (Figure 4.2g, black) whereas for ARNOΔpbr the interaction was restored (Figure 4.2g, green). This data supports the importance of Sec7's helix H in the interaction and suggests that ARNO's autoinhibitory mechanism may also regulate its interaction with the EGFR.

EGFR-JM's interaction with membranes shares common features and distinct differences to ARNO-Sec7

We have shown that ARNO-Sec7 binds the JM segment of the EGFR. As it is known that the JM segment also interacts with CaM and anionic phospholipids of the inner leaflet of the plasma membrane [1, 2, 57, 93, 138, 149] we subsequently investigated similarities and/or differences in the binding mode of these interactors. To obtain the desired high-resolution information into the effect of the membrane surface, the interactions of JM with phospholipids in the form of phospholipid-bilayer nanodiscs (NDs) were characterized by NMR spectroscopy.
Our data show that the presence of NDs containing only the neutral POPC phospholipid does not induce noticeable chemical shift perturbations in EGFR-JM (Figure 4.3a, yellow), indicating that this domain on its own does neither interact with neutral phospholipids nor with the membrane scaffold proteins (MSP) used to assemble the NDs. While the latter corroborates usage of MSP-derived nanodiscs as suitable membrane mimetic for the system, the absence of interactions with POPC lipids differs to previous findings in which a strong interaction of JM to DPC micelles was observed [34]. Since DPC detergent molecules and POPC lipids both comprise the same phosphocholine head group, our results suggest that the overall assembly of the membrane mimetic (detergent-free lipid bilayers vs. detergent monomer-micelle equilibrium) has a strong

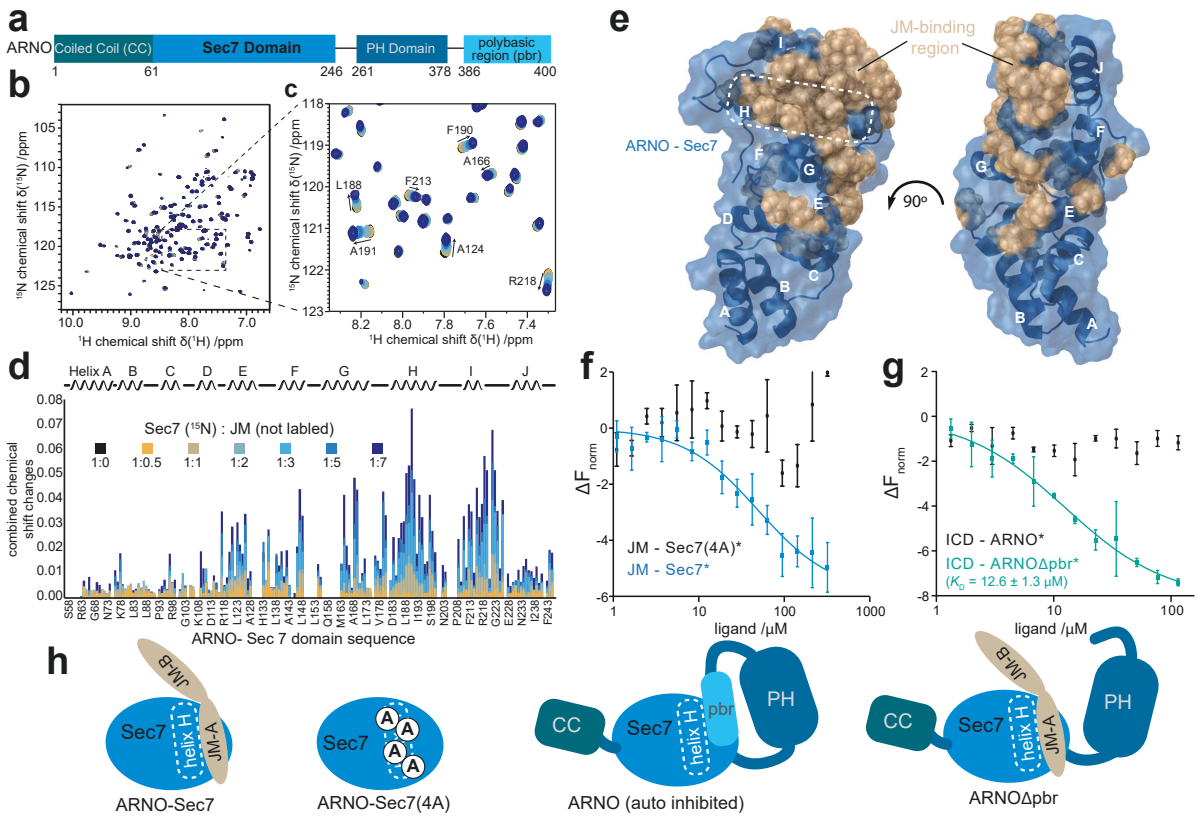


Figure 4.2: JM-Sec7 interaction as seen from the ARNO-Sec7 side. (a) Schematic representation of ARNO's domain architecture. (b+c) [^{1}H,^{15}N]-HSQC NMR spectra of ^{15}N-labeled ARNO-Sec7 in the presence of increasing amounts of unlabeled EGFR-JM (color code as in d). (d) Chemical shift perturbations along the ARNO-Sec7 sequence induced by the presence of indicated molar ratios of EGFR-JM. (e) Mapping of most affected residues on the 3D structure of ARNO-Sec7 (pdb code: 4JMI [136]) indicating the EGFR-JM binding site of ARNO-Sec7. (f+g) MST data showing disruption of JM's interaction with ARNO-Sec7 due to mutations (f) or autoinhibition (g). (f) Alanine substitutions of surface-exposed hydrophobic residues of helix H of Sec7, i.e. Sec7(4A), lead to disruption of the interaction with JM. The JM-Sec7 data (blue) is identical to data shown in Figure 4.1b and serves as reference. (g) While the presence of the autoinhibitory polybasic region (pbr) in full-length ARNO inhibits interaction with EGFR-ICD (black), deletion of the polybasic region (ARNOΔpbr) restores the interaction (n=3, mean±SD). h) schematic summary of MST and NMR results.

influence on the interaction with JM. At this point it can only be speculated that the nanodiscs better reflect the physiologically-relevant membrane interaction of JM. However, in any case, the observed difference between detergent micelles and nanodiscs highlights the importance of the choice of a suitable membrane mimetic for structural studies of membrane interactions.

The strengths of the nanodisc system include its homogeneity, stability, the absence of detergents and near native bilayer arrangement as well as a the possibility to accurately change their lipid composition without modifying other parameters and use NMR-spectroscopy to determine lipid specific interaction with single amino acid resolution [172, 173]. In the following we made use of these features to investigate

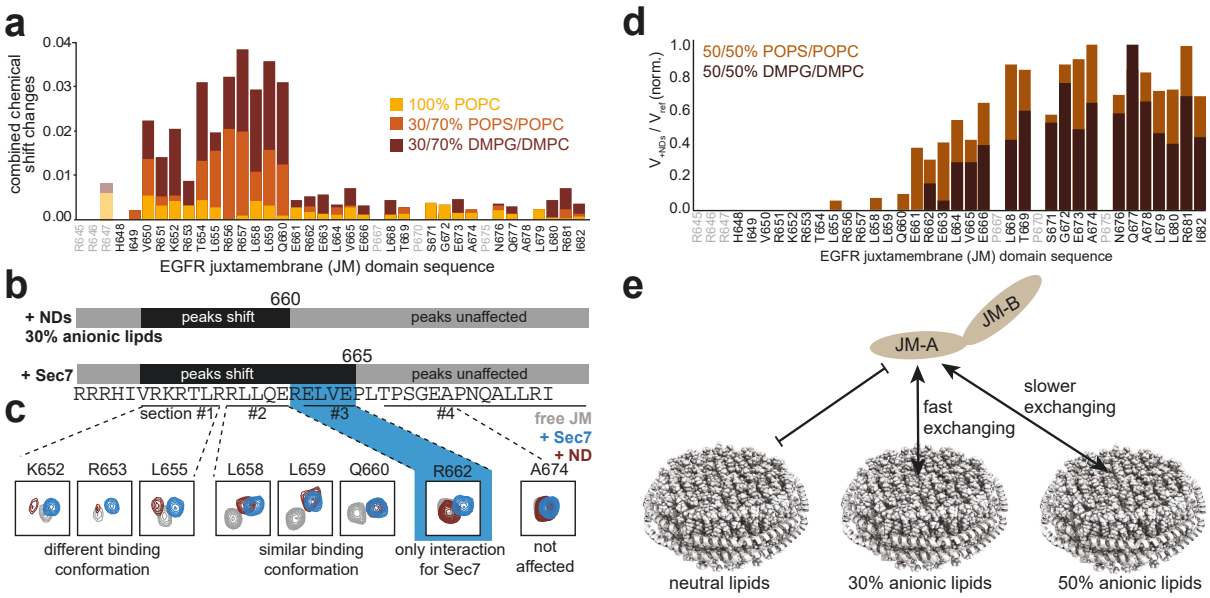


Figure 4.3: JM-membrane interaction depends on anionic lipid content and follows a similar pattern as Sec7 binding with distinct differences. (a) Chemical shift perturbations along the EGFR-JM sequence induced by the presence of nanodiscs with the indicated lipid composition. (b) Schematic comparison of EGFR-JM binding behavior to lipid bilayers containing 30% anionic lipids (upper chart) or ARNO-Sec7 (lower chart, according to data shown in Figure 4.1e). (c) NMR signals of selected residues representative of JM regions with different behavior induced by the presence of Sec7 (blue peaks, also see Figure 4.1) or NDs with 30% anionic lipids (red peaks). Blue area highlights residues showing interaction exclusively with Sec7 and not the used NDs. (d) Effects of addition of NDs with 50% anionic lipid content. Unlike to the peak shifts visible for ND interaction with 30% anionic lipid content (a) or Sec7 binding (Figure 4.1), addition of NDs containing 50% POPS and 50% POPC (brown) or 50% DMPG and 50% DMPC (dark brown) lipids predominantly leads to disappearance of peaks for residues in JM-A (see. Figure A19 and A20 for comparison of spectra, peak shifts and volumes for all used lipid mixtures). The observed peak disappearance is indicative of prolonged contact times of this region with the lipids (i.e. NMR medium or slow exchange regime for 50% anionic lipids and NMR fast exchange regime for 30% anionic lipids or Sec7). e) Schematic summary of EGFR-JM's interaction with different NDs. In a and d, grey labels indicate residues that were not observed.

the interaction of JM with NDs containing 30% anionic phospholipids via NMR spectroscopy. Two different phospholipid mixtures were used, i.e. 30% anionic DMPG lipids with 70% neutral DMPC lipids as well as anionic POPS lipids (30%) with neutral POPC lipids (70%). In both cases clear changes in the NMR spectrum induced by the presence of the respective nanodiscs can be observed (Figure 4.3a, Figure A21 and Figure A22). Similar to the interaction with ARNO-Sec7, the residues affected the most by the presence of the anionic membrane surface are confined to the JM-A region. However, a closer look also reveals that the phospholipid interacting region is a few residues shorter than the Sec7-binding region.

A comparison of the results obtained on the frequently used model phospholipids DMPC/DMPG (Figure 4.3a, maroon and Figure A22) to the more physiologically relevant POPC/POPS phospholipids (Figure 4.3a, orange and Figure A22) reveals that the observed effects are slightly elevated for the DMPC/DMPG system, suggesting that the different position of the negative charge in the head group and/or the presence of

unsaturated fatty acids may affect the interaction with JM.
A comparison of EGFR-JM's interaction with anionic lipids or ARNO-Sec7 highlights four different sections in JM (Figure 4.3b+c). While residues V650-Q660 show considerable chemical-shift perturbations induced by both interaction partners, the first half of these residues (V650-L655, Figure 3c, section #1) show clearly different chemical shifts upon binding to lipids or Sec7, whereas the second half (R656-Q660, Figure 4.3c, section #2) experience an almost identical variation in chemical shift. The third section (R662-V665, Figure 4.3c, section #3) is only affected by Sec7 and not by the lipids. The fourth section (L667-I682, Figure 4.3c, section #4) is not affect by the presence of either interaction partner. Consequently, EGFR-JM's interaction with anionic phospholipids shares some common features with the interaction with ARNO-Sec7, but also displays distinct differences. While the presence of NDs with 30% content of anionic phospholipids leads to chemical shift perturbations, indicative of fast exchange processes, increasing the membrane charge density to 50% anionic phospholipids alters the interaction kinetics and leads to considerable peak broadening, indicative of intermediate exchange processes (see Figure A22 for full experimental data). Considering the size of the ND system, a tight binding (in the slow exchange regime) could also explain this observation. In any case, it can be assumed that the JM-membrane interaction becomes stronger with increased negative charge density of the membrane. When plotting the peak intensity instead of the chemical shift changes it can be seen that also under these conditions, the JM-A region is the driving force of the interaction (Figure 4.3d).
Overall our data shows that despite JM-A being mainly involved in the interaction with lipids and Sec7, the interaction with Sec7 occurs over an extended binding region that involves a number of additional JM residues, as compared to JM's interaction with the membrane surface. In addition, an increase of the anionic lipid content from 30% to 50% slows down the otherwise fast bound-to-free exchange processes, revealing the possibility of modulating JM's membrane interaction kinetics by variations in lipid composition.

The interplay of lipids, CaM and Sec7 as intracellular modulators of EGFR-JM

To directly compare the observed interaction of EGFR-JM with ARNO-Sec7 to the known cytoplasmic EGFR modulator CaM, we carried out additional MST and NMR-based experiments. Unsurprisingly, our MST data shows that binding of CaM to EGFR-ICD is calcium- and JM-dependent (Figure 4.4a). When recording NMR spectra of EGFR-JM in the presence of unlabeled CaM, a set of peaks disappear from the spectrum (in line with an interaction in the NMR intermediate exchange regime). As expected [2, 163], plotting the decrease in intensity along the JM sequence again reveals that predominantly JM-A interacts with CaM (Figure 4.4b). Looking at the

affected JM residues it can be seen that the CaM binding region of JM is again a few residues longer than its membrane binding region. Interestingly the CaM and the Sec7 binding regions of JM are essentially identical. However, in line with a higher binding affinity seen in the MST data (K_D of about 1 µM), the NMR data also suggest that CaM interacts less transiently with JM as compared to Sec7 or membranes with 30% negative charge content.

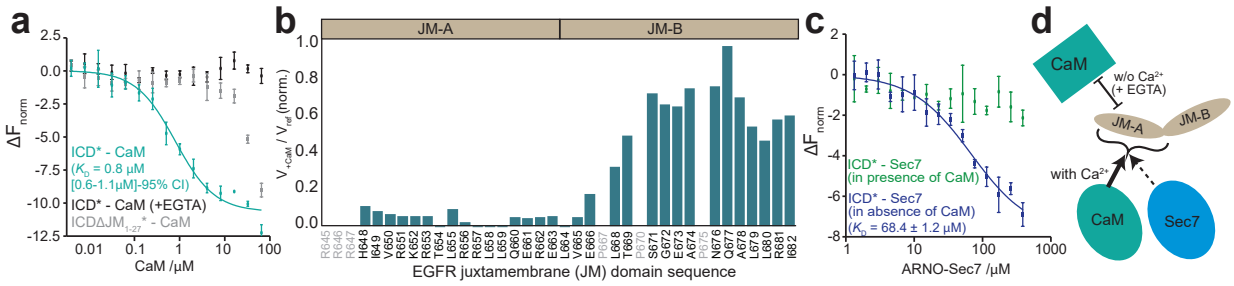


Figure 4.4: Calmodulin (CaM) and ARNO-Sec7 share same binding site and compete for EGFR-JM binding *in vitro*. (a) MST data of the interaction of CaM and EGFR-ICD. Removal of accessible calcium via EGTA (black) as well as deletion of the first 27 residues of the JM segment (grey) largely reduces the binding of calcium-activated CaM to EGFR-ICD (green; n=3, mean±SD). (b) Changes in EGFR-JM residue specific peak volumes upon addition of CaM. Peak disappearance reports on interaction between the effected JM residues and CaM (NMR intermediate exchange regime). Grey labels indicate residues that were not observed. (c) MST data of the interaction between ARNO-Sec7 and EGFR-ICD in the absence (blue) or presence of 30 µM CaM (green; n=3, mean±SD). (d) Schematic comparison of the observed CaM and Sec7 binding behavior of EGFR-JM.

Having found that CaM and ARNO-Sec7 bind to an essentially identical binding site on EGFR-JM we investigated a possible competition of CaM and ARNO-Sec7 for binding to the EGFR using MST. In line with the higher binding affinity of CaM, when EGFR-ICD (200 nM) was preincubated with a saturating concentration of CaM (30 µM), the binding of ARNO-Sec7 was completely prevented (Figure 4.4c), confirming a competitive binding of CaM and ARNO-Sec7 *in vitro*.

Our data shows that CaM and ARNO-Sec7 interact with the same JM region. This fact hinders a reliable NMR investigation of the competition between these two proteins. In contrast, the binding regions of JM to phospholipid nanodiscs or Sec7 sufficiently differ to allow distinction between the binary JM-nanodisc and JM-Sec7 complexes. In particular, residues (E661-V665) in the center of the JM segment can be used as reporters since they are not affected by binding to phospholipids but are part of the Sec7-interacting region (Figure 4.3b and A22). Indeed, when adding unlabeled ARNO-Sec7 to the JM peptide preincubated with NDs containing high amounts of anionic lipids (50/50% POPC/POPS) distinct chemical shift perturbations are visible for the

'Sec7-specific-reporter residues' E661-V665 as compared to free JM or to JM in the presence of just NDs (Figure 4.5a and Figure A22). The observed peak shift is consistent with the perturbations expected due to formation of a JM-Sec7 complex. Interestingly, JM residues directly at the edge of the membrane binding interface (Q660 and R662) show stronger or different chemical shift perturbation when both binding partners are present (as compared to the individual pairwise interactions, Figure 4.5a). This behavior is indicative of cooperative effects and/or different structural alterations. While our data does not allow to distinguish between a ternary JM-membrane-Sec7 complex or an exchanging 3-state equilibrium (free JM, membrane-bound JM, Sec7-bound JM), the NMR data show that ARNO-Sec7 can interact with the JM peptide even in the presence of NDs containing a high amount of anionic lipids and thus support the notion that ARNO can interact with the EGFR at the plasma membrane.

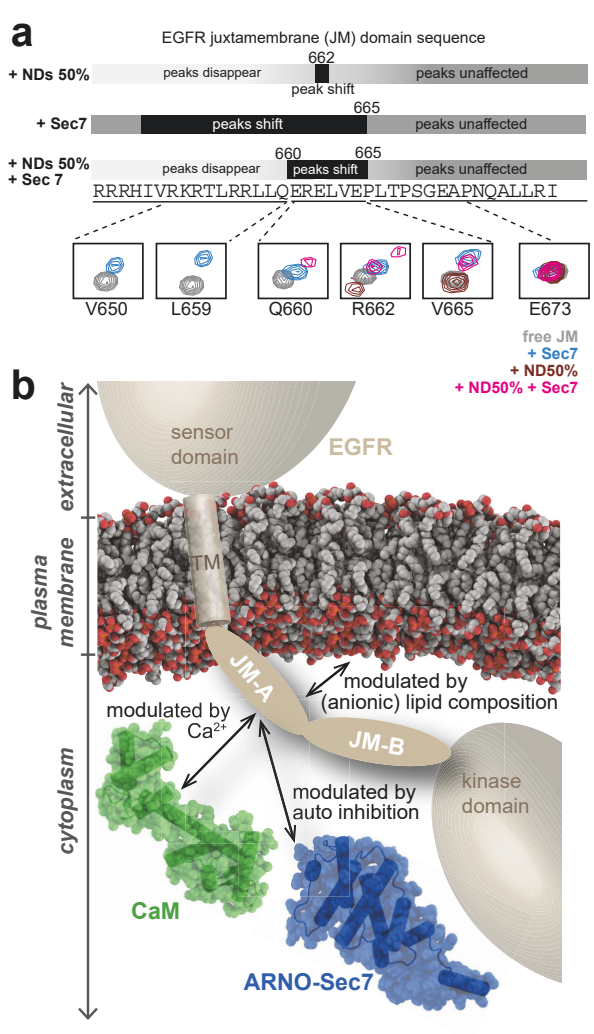


Figure 4.5: The interplay between possible modulators acting on EGFR-JM as central interface in the intracellular interaction network of the EGFR. (a) Comparison of NMR results for EGFR-JM in the presence of NDs with 50% anionic lipids (POPS, top), or in the presence of ARNO-Sec7 (middle), or in the presence of both interaction partners (lower schematic). Overlay of residue specific NMR signals in the absence (grey) or presence of indicated interaction partner(s). Selected residues, representative of the three differently affected regions, are shown. (b) Schematic summary of EGFR-JM interaction partners, shown in this study to interact with the JM-A segment *in vitro*, and their individual modulators.

4.1.3 Conclusions

Using solution NMR spectroscopy and microscale thermophoresis (MST), supported by site-directed mutagenesis techniques, we show that ARNO-Sec7 and EGFR-JM interact *in vitro*. The residues of both Sec7 and JM involved in binding were identified (Figure 4.1 and 4.2). From the JM side, NMR data showed that Sec7 specifically recognizes the first half of the segment (i.e. JM-A, Figure 4.1e), which has been shown to be of importance for EGFR activation [45, 56, 67, 123, 164]. JM-A is also the region recognized by calmodulin (Figure 4.4), a major regulatory protein of the EGFR [2, 85, 138, 163]. Furthermore, JM-A tethers JM to the plasma membrane (Figure 4.3), stabilizing the inactive conformation of the EGFR [5, 78].
The isolated JM peptide in solution seems to exist mainly as random coil (Figure 4.1f), with some propensity to form a transient α-helix at the JM-A region [5, 47, 67, 102]. Our data reveals that, upon binding to Sec7, the amide resonances of JM-A shift upfield (Figure A22c), which is indicative of a higher helical propensity in the Sec7-bound conformation. The JM segment also interacts with negatively charged phospholipids of the inner leaflet of the membrane [78, 149]. By using phospholipid nanodiscs (NDs) we demonstrate that JM-A is also the region responsible for binding to membranes containing anionic lipids. While the membrane-binding and Sec7-binding regions of JM largely overlap, a closer analysis of the NMR data shows that some residues experience different chemical shift perturbations upon binding to either partner and that the binding interface to Sec7 is elongated as compared to the lipid binding interface (Figure 4.3b and Figure A22).
From the Sec7 side, our data shows that the accessibility of the JM binding site is restricted when ARNO is in the autoinhibited state which is common to all cytohesin members [44] (Figure 4.2). Release from the autoinhibition requires the binding of an already activated, membrane-attached ARF molecule or of phosphoinositides (PIPs) to the PH domain and maximum activation requires both steps resulting in exposure of the Sec7 domain [38, 158]. As PIPs cluster around the EGFR [1, 57, 96, 98, 175], binding of ARNO to these PIPs would bring it near to the EGFR and simultaneously activate it for JM binding (see Figure A23 for schematic visualization). The PIPs-driven co-localization and/or insufficient Ca^{2+} availability could also counteract the higher affinity observed for Ca^{2+}-activated CaM as possible competitor of the ARNO-EGFR interaction. The interplay between ARNO´s expression level and state of activation, the plasma membrane's lipid composition and its arrangement, as well as the available Ca^{2+}- and CaM levels could therefore provide a further layer of modulation of EGFR signaling (Figure 4.5b).
Whether binding of ARNO to the JM segment of the EGFR occurs in the living cell and whether this binding would indeed modulate EGFR signaling is however currently unknown. There is indirect evidence for ARNO modulating EGFR activity [16, 120, 121] but

the mechanism has not been elucidated. Due to our findings that *in vitro* ARNO interacts with the JM segment in a similar way as CaM does, it is tempting to speculate that ARNO and CaM could modulate EGFR activity also by a similar mechanism. Although the mechanism by which CaM modulates EGFR activity has not yet been exactly determined, it appears to involve the weakening of JM´s interaction with phospholipids of the membrane [96, 141, 149]. Our data obtained in the presence of phospholipid nanodiscs are consistent with this view and with a model in which CaM and ARNO could contribute to the activation of the EGFR by releasing one of several autoinhibition mechanisms of the EGFR, namely the immobilization of the JM segment on the surface of the membrane.

4.1.4 Material and Methods

Protein constructs and expression. Human EGFR-ICD (amino acids 645-1186, numbering according to UniProt P00533 without the 24 amino acids of the signal peptide) was equipped with a 6xHis tag and a TEV cleavage site and cloned into pFastBac-1 (Invitrogen) such that after TEV cleavage the protein contained two additional amino acids (Gly, Ala) at the N-terminus. EGFR-ICDΔJM$_{1\text{-}27}$ (amino acids 672-1186) was constructed by inserting a TEV cleavage site between amino acids 671 and 672 of EGFR-ICD and cloned into pACEBac-1 (ATG:biosynthetics) such that after TEV cleavage the protein had no additional amino acids. Recombinant baculoviruses were generated using the MultiBacTurbo Expression System (ATG:biosynthetics) and proteins expressed for 3 days in SF9 cells (Invitrogen). EGFR-JM (amino acids 645-682) was fused to maltose binding protein followed by a TEV site such that after TEV cleavage the unmodified JM peptide was obtained. It was cloned into pET-28a (Novagen) and expressed for 3 h at 37 °C in *E. coli* BL21(DE3) (Stratagene). EGFR-JM*SC* was obtained by scrambling amino acids 645-682 resulting in the sequence: RELKHIQVRL RTERQLEPLE IRAVNRSRLT PRLAGLPR. Otherwise it was treated the same way. Human ARNO (UniProt Q99418), ARNOΔpbr (amino acids 1-386), ARNO-Sec7 (amino acids 61-246), ARNO-Sec7(4A) (Y186, F190, I193 and M194 changed to Ala) and human CaM (UniProt P0DP23, amino acids 2-149) were equipped with a 6xHis tag and a TEV cleavage site, cloned into pET-28a and expressed at 20 °C overnight in *E. coli* BL21(DE3). Except for CaM, the constructs contain additional Gly and Ser at the N-terminus after TEV cleavage.

Protein purification and labeling. All cell pellets were homogenized via French press in lysis buffer (50 mM Tris/HCl, pH 7.8, 300 mM NaCl, 10% glycerol, 25 mM imidazole), except for calmodulin in different lysis buffer (50 mM HEPES/KOH, pH 7.8, 300 mM NaCl, 10% glycerol, 25 mM imidazole). EGFR-ICD, EGFR-ICDΔJM$_{1\text{-}27}$, all ARNO constructs and calmodulin were purified via Ni-NTA affinity chromatography

(Macherey-Nagel). Eluted samples were buffer exchanged to remove imidazole, before TEV cleavage overnight at 4 °C. Protein samples were then subjected to reverse Ni-NTA chromatography (Macherey-Nagel), and concentrated using Vivaspin Turbo (Sartorius) followed by size exclusion chromatography either on HiLoad 16/600 Superdex 200pg (GE Healthcare) for EGFR-ICD and EGFR-ICDΔJM$_{1\text{-}27}$, or on HiLoad 16/600 Superdex 75pg (GE Healthcare) for ARNO-Sec7, ARNO-Sec7(4A) and calmodulin. In addition, during TEV cleavage of EGFR-ICD, 0.5 µM His-tagged YopH and 0.5 mM $MgCl_2$ was added for dephosphorylation of the kinase. During calmodulin purification, cleared lysate was heated for 5 min at 80 °C, then cooled down on ice for 10 min, followed by centrifugation to remove denatured proteins. Furthermore, 1 mM of $CaCl_2$ was supplemented to the sample immediately before size exclusion chromatography. MBPT-EGFR-JM and MBPT-EGFR-JM$_{SC}$ were purified via amylose affinity chromatography (New England Biolabs), followed by TEV cleavage at room temperature for 48 h. Afterwards, digested sample was applied to size exclusion chromatography on HiLoad 16/600 Superdex 30pg (GE Healthcare). All the gel filtration runs were monitored at 280 nm, except for EGFR-JM, EGFR-JM$_{SC}$ and calmodulin at 214 nm. All the collected peak samples were concentrated in buffer H (20 mM HEPES/KOH, pH 7.8, 150 mM NaCl), using Vivaspin Turbo (Sartorius). For the fluorescence labeling of ARNO-Sec7 and ARNO-Sec7(4A), 10 µM proteins were mixed with 100 µM Alexa Fluor 647 NHS Ester (Thermo Fisher) in labeling buffer T (20mM HEPES/KOH, pH 7.8, 150 mM NaCl, 100 mM $NaHCO_3$). The labeling reactions were carried out on ice in the dark for 1 h. For the labeling of EGFR-ICD and EGFR-ICDΔJM$_{1\text{-}27}$, 10 µM proteins were mixed with 30 µM RED-NHS 2nd generation (NanoTemper) in labeling buffer N (20 mM HEPES/KOH, pH 7.8, 150 mM NaCl). The mixture was incubated on ice in darkness for 30 min. All labeling reactions were terminated by addition of 100 mM Tris/HCl, pH 8. Afterwards, samples were applied to pre-equilibrated illustra Nap-5 columns (GE Healthcare) to remove free dye, followed by elution with buffer H. Protein concentrations and degrees of labeling were quantified on NanoDrop 2000c Spectrophotometer (Thermo Fisher), before aliquoting and flash freezing.

MST measurements. For each MST assay, unlabeled protein was used to prepare 15-step serial dilution with final volume of 5 µL in assay buffer (20 mM HEPES/KOH, pH 7.8, 150 mM NaCl, 0.005% Triton X-100, 10 µM BSA). Next, 5 µL of 200 nM fluorescence-labeled protein was added to each dilution. For measurements including CaM (except for that with EGTA), 2 mM $CaCl_2$ was added to the assay buffer. The calmodulin titration was carried out in 1:2 serial dilution, while the others were performed in 1:3 dilution. For the calmodulin competition assay, 30 µM calmodulin was premixed with 200 nM labeled protein, before being added to 15 serial dilutions. Mixed samples were loaded into Monolith NT.115 Premium Capillaries (NanoTemper) and MST measurements were performed on Monolith NT.115 system (NanoTemper). For

assays using labeled ARNO-Sec7 and ARNO-Sec7(4A), samples were pre-incubated at room temperature for 10 min and measured with 60% LED power, 50% MST power. For assays using labeled EGFR-ICD and EGFR-ICDΔJM$_{1\text{-}27}$, samples were pre-incubated at room temperature for 5 min and measured with 20% LED power, 40% MST power. Each sample preparation and measurement was carried out in triplicate. Data analysis was performed using the K_D fitting function of MO.Affinity Analysis v2.3 (NanoTemper) and graphs were prepared using Prism 5.0f (GraphPad). For the calculation of F_{norm}, hot cursor was set at 5 seconds for assays involving labeled ARNO-Sec7 and ARNO-Sec7(4A), while for assays involving labeled EGFR-ICD and EGFR-ICDΔJM$_{1\text{-}27}$, hot cursor was set at 2.5 seconds.

Nanodiscs production. As reported before [13], *E. coli* BL21 (DE3) were transformed with MSP1D1 plasmid DNA in vector pET28a. Cells were grown in LB medium, induced by 1 mM IPTG at an optical density of 0.7, incubated 5-6 hours at 37 ℃ and pelleted down. Cells were resuspended in buffer B (50 mM Tris/HCl, pH 8.0, 500 mM NaCl) supplemented with 6 M GdnHCl and EDTA-free Complete protease inhibitors (Roche) lysed by sonication (Bandelin Sonopuls MS72 probe), centrifuged at 17000 g for 1 h (Beckman J2-21 rotor JA-20.1) and incubated 1 h with previously equilibrated 2.5 ml Ni-NTA agarose resin/3 L culture (Macherey-Nagel). Column was washed with 4 CV buffer B, 4 CV buffer B supplemented with 1% Triton X-100, 4 CV buffer B + 60 mM Na-cholate, 4 CV buffer B, 4 CV buffer B + 20 mM imidazole. Four fractions of 1 CV were eluted with 250 mM imidazole. The whole process was kept at 4 ℃ in a cold room. The elution fractions were pooled and dialyzed against 100-fold dialysis buffer (200 mM Tris/HCl, pH 7.5, 100 mM NaCl). N-terminal His-tag was cleaved using TEV protease incubated overnight at 4 ℃. ΔHis-MSP was separated from MSP by IMAC and concentrated to the desired molarity using a Vivaspin centrifugal device of 10 kDa MWCO.
Nanodiscs were assembled according to established protocols [42,130]. In short, lipids' chloroform stocks were dried under nitrogen flow to obtain a lipid film and stored under vacuum overnight. ΔHis-MSP1D1 and the appropriate amount of lipids (Avanti Polar Lipids) solubilized in 60 mM Na-cholate were mixed together in lipid buffer (20 mM Tris/HCl, pH 7.5, 100 mM NaCl, 5 mM EDTA). Four different batches were prepared: one using 100% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) as a non-charged control; one using 30% 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS) and 70% POPC containing 30% net negative charge and similar properties as native membranes; one using 50% POPS and 50% POPC with a higher density of negative charges; one using 50% 1,2-dimyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DMPG) and 50% 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) containing 50% negative charge and different head group and hydrocarbon chain properties (see main text for more information). The scaffold-to-lipids molar ratio was calculated from

geometrical considerations. 20% (w/v) of previously washed Biobeads SM-2 (Biorad) were added and the mixture incubated at room temperature overnight. The Biobeads were removed by centrifugation and once again 20% (w/v) were added for an additional 4-5 h. Finally, they were purified by SEC on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with SEC buffer (20 mM sodium phosphate, pH 7.4, 50 mM NaCl) using a Äkta pure device at a flow rate of 1 ml/min. The quality of NDs preparation was check by the SEC chromatogram as well as by DLS (PSS Nicomp). NDs were concentrated to the desired molarity using a Vivaspin centrifugal device of 10 kDa MWCO.

NMR Spectroscopy. All NMR experiments were performed on Bruker Avance III HD+ spectrometers operating either at 600 or 700 MHz, both equipped with 5 mm inverse detection triple-resonance z-gradient cryogenic probes. Data was collected at 32 or 15℃ and processed with TOPSPIN 3.2 (Bruker BioSpin). 4,4-dimethyl-4-silapentanesulfonic acid (DSS) was used as a chemical shift standard, and ^{13}C and ^{15}N data were referenced using frequency ratios as previously described [178].
For the resonance assignment of Sec7 and JM, triple (U[^{2}H,^{13}C,^{15}N]) and double-labeled (^{13}C,^{15}N) samples were prepared, respectively. The U[^{2}H,^{13}C,^{15}N]-Sec7 sample was prepared at a concentration of 360 µM in 20 mM sodium phosphate buffer pH 7.4 containing 300 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 µM DSS. The ^{13}C,^{15}N-JM sample was prepared at a concentration of 270 µM in 20 mM sodium phosphate buffer pH 5.5 containing 100 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 µM DSS. The lower pH in this sample was used in order to avoid residue-amide exchange with the solvent. TROSY versions (Tr) of ^{15}N-edited HSQC and three-dimensional HNCO, HN(CA)CO, HN(CO)CACB (or CBCA(CO)NH, for JM) and HNCACB experiments were performed to obtain the chemical shift assignments of the backbone atoms of Sec7, while the standard versions were used for JM. Furthermore, for the assignment of the sidechain resonances of JM we also acquired a ^{13}C-edited HSQC and a 3D hCCH-TOCSY. The assignment of the ^{1}H, ^{13}C, and ^{15}N signals in the spectra was performed using CARA 1.9.24a [71]. Data was acquired at 32 and 15℃ for U[^{2}H,^{13}C,^{15}N]-Sec7 and ^{13}C,^{15}N-JM, respectively. Tables 4.1 and 4.2 summarize the acquisition parameters for Sec7 and JM, respectively.
The residues of Sec7 responsible for binding were identified by titrating a sample of ^{15}N-labeled Sec7 with increasing amounts of non-labeled JM and acquiring a [^{1}H,^{15}N]-HSQC spectrum at each titration point. The concentration of protein was maintained at 60 µM and the concentration of JM varied from 0 to 420 µM (using seven individual samples at 0.0, 0.5, 1.0, 2.0, 3.0, 5.0 and 7.0 molar equivalents). The ^{1}H,^{15}N-HSQC spectra were acquired with 2048 × 128 points and 256 scans. Spectral widths were 14 ppm for ^{1}H and 36 ppm for ^{15}N. The central frequency for proton was set on the solvent signal (4.704 ppm) and for nitrogen was set on the center of the amide region

Table 4.1: Aquisition parameters of the spectra used for Sec7 resonance assignment.

	Number of points			Spectral width (ppm)			Central frequency (ppm)			NS
2D	F3	F2	F1	F3	F2	F1	F3	F2	F1	
^{1}H,^{15}N-TrHSQC	-	2048	128	-	14	36	-	4.704	116	64
3D										
TrHNCO	2048	40	128	14	36	22	4.704	116	176	8
TrHN(CA)CO	2048	40	128	14	36	22	4.704	116	176	24
TrHN(CO)CACB	2048	40	128	14	36	75	4.704	116	39	24
TrHNCACB	2048	40	128	14	36	75	4.704	116	39	32

Table 4.2: Aquisition parameters of the spectra used for JM resonance assignment.

	Number of points			Spectral width (ppm)			Central frequency (ppm)			NS
2D	F3	F2	F1	F3	F2	F1	F3	F2	F1	
^{1}H,^{15}N-HSQC	-	2048	256	-	13	30	-	4.697	119.5	32
^{1}H,^{13}C-HSQC	-	1024	512	-	13	75	-	4.696	42	32
3D										
HNCO	2048	40	128	13	30	22	4.697	119.5	176	8
HN(CA)CO	2048	40	128	13	30	22	4.697	119.5	176	16
CBCA(CO)NH	2048	40	128	13	30	80	4.697	119.5	42	32
HNCACB	2048	40	128	13	30	75	4.697	119.5	42	32
hCCH-TOCSY	2048	40	128	13	75	75	4.697	42	42	16

(116 ppm). The data was acquired in 20 mM sodium phosphate buffer containing 100 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 µM DSS, pH 7.4. All data was acquired at 32℃.

The residues of JM responsible for binding were identified in a similar way as described above, using ^{15}N-labeled JM and non-labeled Sec7. The concentration of JM was maintained at 40 µM and the concentration of Sec7 varied from 0 to 280 µM (using five individual samples at 0.0, 2.0, 3.0, 5.0 and 7.0 molar equivalents). The ^{1}H,^{15}N-HSQC spectra were acquired with 2048 × 128 points and 8 scans. Spectral widths were 13 ppm for ^{1}H and 30 ppm for ^{15}N. The central frequency for proton was set on the solvent signal (4.695 ppm) and for nitrogen was set on the center of the amide region (119.5 ppm). The data was acquired in 20 mM sodium phosphate buffer containing 100 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 µM DSS, pH 5.5. All data was acquired at 15 and 32℃.

To study the interaction of JM with the different NDs we measured ^{15}N-edited HSQC spectra of the free ^{15}N-JM (40 µM) and in the presence of 20 µM of NDs, containing the different lipids (note that this will result on average in one JM per membrane leaflet). The [^{1}H,^{15}N]-HSQC spectra were acquired with 2048 × 128 points and 8 scans. Spectral widths were 15 ppm for ^{1}H and 30 ppm for ^{15}N. The central frequency for proton was set on the solvent signal (4.703 ppm) and for nitrogen was set on the center of

the amide region (119.5 ppm). The data was acquired in 20 mM sodium phosphate buffer containing 100 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 µM DSS, pH 5.5. All data was acquired at 32°C.
The interaction between JM and calmodulin (CaM) was measured using ^{15}N-edited HSQC experiments with 100 µM ^{15}N-labeled JM in absence and presence of 400 µM CaM in 20 mM sodium phosphate buffer, pH 5.5, with 150 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 µM DSS. The spectra were acquired with 2048 x 128 points and the central frequency for protons were set on the solvent signal (4.690 ppm) and for nitrogens on 119.5 ppm. The spectral widths for ^{1}H and ^{15}N were set to 13 ppm and 30 ppm, respectively. Both spectra were acquired with 16 scans at 15 °C. Data for JM's three N-terminal Arginines was not unambiguous and, where shown, could reflect either only on Arg647 or also on Arg646 and/or Arg645. Signal for His648 was considerably weaker as for all other assigned residues and not always clearly distinguishable from spectral noise. In unclear cases, the residue was removed from analysis.
To investigate the effect of the overall charge of JM in binding we prepared a scrambled version of JM, JM_{SC}, containing a redistributed but overall identical amino acid composition with the sequence: RELKHIQVRLRTERQLEPLEIRAVNRSRLTPRLAGLPR. We measured a ^{15}N-edited HSQC spectrum of the free JM_{SC} (40 µM) and in the presence of 7.0 equivalents of Sec7 (280 µM). The [^{1}H,^{15}N]-HSQC spectra were acquired with 2048 × 128 points and 8 scans. Spectral widths were 13 ppm for ^{1}H and 30 ppm for ^{15}N. The central frequency for proton was set on the solvent signal (4.701 ppm) and for nitrogen was set on the center of the amide region (119.5 ppm). The data was acquired in 20 mM sodium phosphate buffer containing 100 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 µM DSS, pH 5.5. All data was acquired at 32°C.

Combined Chemical Shift, $\Delta\delta_{comb}$. For the evaluation of the behavior of individual amino acids upon addition of increasing amounts of ligand we calculated the combined amide proton and nitrogen chemical shift differences using Equation 4.1 [144]:

$$\Delta\delta_{comb} = \sqrt{(\Delta\delta_H)r + (0.1x\Delta\delta_N)r} \tag{4.1}$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shifts of proton and nitrogen, respectively. In order to decide whether a given residue belongs to the class of interacting or non-interacting residues, we have calculated a corrected standard deviation to zero (σ_0^{corr}) [144].

Sec7 and JM Resonance Assignment. Despite existence of an NMR structure of Sec7 [15], the experimental assignments are not available. As such, a *de novo* assignment was carried out. The backbone assignment of the amide resonances of Sec7 and JM has been performed using a standard triple resonance approach [183]. For

Sec7, the amide resonances of amino acids S1, E2, T3, R4, Q5, R6, Y44, G48, K51, T52, F73, D74, L75, H76, R88, S93, F94, R95, L96, A100, Q101, K102, I103, D104, R105, M106, T125, N144, R148, D149 and L150 could not be assigned (possibly due to exchange with the solvent). The Chemical Shift Index (CSI) [179] was used to identify protein secondary structure and compare it with the deposited structures [15, 136] (Figure A19). The secondary structure of Sec7 was predicted for each assigned amino acid residue using Equation 4.2:

$$CSI = \Delta\delta(C_\alpha) - \Delta\delta(C_\beta) \tag{4.2}$$

where CSI is the Chemical shift index and $\Delta\delta C_\alpha$ and $\Delta\delta C_\beta$ are the variations of the measured C_α and C_β chemical shifts with respect to random coil values. Three or more consecutive negative values indicate β-strand while three or more positive values indicate a α-helical structure.

4.2 Local deuteration enables NMR observation of methyl groups in proteins from eukaryotic and cell-free expression systems

Abhinav Dubey, Nikolay Stoyanov, Sandeep Chhabra, Thibault Viennet, Shantha Elter, Jan Borggräfe, Aldino Viegas, Radosław P. Nowak, Nikola Burdziev, Ognyan Petrov, Eric S. Fischer, Manuel Etzkorn, Vladimir Gelev, Haribabu Arthanari

4.2.1 Introduction

Methyl-NMR spectroscopy represents a unique source of information about protein-protein and protein-ligand interactions, and macromolecular motions on a microsecond to millisecond timescale [4, 53, 75, 92, 129, 145, 166]. The rapid rotation of the methyl hydrogen around the connected C–C bond introduces favorable relaxation properties and when combined with the three-fold symmetry of the methyl group results in a single intense signal. This is further enhanced by the selection of the slow relaxing ^{13}C-H_3 coherences using HMQC experiment, also referred as methyl TROSY. When combined with most effective isotope labeling NMR spectroscopy of megadalton-size proteins become accessible. Observation of the [^{13}C,^{1}H]-methyl signal is largely facilitated when surrounding protons are replaced with deuterons to reduce signal loss due to dipole-dipole relaxation [70]. The most common method for protein methyl labeling is to grow bacteria in perdeuterated medium supplemented with the appropriate $^{13}C^{1}H_3$-methyl-biosynthetic precursors [53, 70, 167, 168]. However, $^{13}CH_3$-methyl labeling of deuterated proteins remains a challenge in eukaryotic and *in vitro* expression systems due to inaccessibility of suitable isotope labeled precursors. Therapeutically relevant targets, including most protein kinases, integral membrane proteins (GPCRs, ABC transporters and TCRs), and therapeutic antibodies often cannot be functionally expressed in bacteria. Expression in insect cells is often the method of choice for these and other proteins that have stringent requirements for proper folding and post-translational modifications. Isotope labeling in higher eukaryotic cells is mainly restricted to usage of pre-synthesized isotope labeled amino acids. In addition, these cells are not viable in $^{2}H_2O$ concentrations higher than 30%. Up to 75% protein deuteration was nevertheless achieved, by the addition of deuterated amino acids to the growth medium [100, 117, 155]. The methyl groups of Met [76, 114, 117] and Ala [76] have been labeled by addition of the corresponding amino acid, and were used to probe G-protein coupled receptors (GPCR) dynamics. The methyl signal was boosted by judicious inter-residue "local" deuteration, i.e. addition of deuterated amino acids of the types that occur in the spatial vicinity of the methyl groups of interest. *In vitro* protein expression with ribosomal extracts represents an increasingly viable option for obtaining functional proteins; however the [^{13}C,^{1}H,^{2}H]-labeled amino acids required for production

of methyl-labeled proteins by this method remain expensive [69, 82, 87] and/or require elaborate experimental protocols [82,87]. Lazarova *et al.* used Leu and Val deaminated precursors together with exogenous trans-aminase to achieve stereoselective methyl labeling of cell-free expressed proteins. The cost of the leucine precursor (20 €/mg) was reported as the major obstacle [82]. Furthermore, expressing functional GPCRs *in vitro* remains a challenge [58]. Yeast strains able to produce functional GPCRs, survive deuteration, and process the metabolic precursor from Ile have been developed [104]. Chemical modifications such as ^{13}C-methylation of Lys [18, 157] and disulfide bonding of Cys with trifluoroethanethiol [46, 184] have also been employed to observe methyls in GPCRs. Despite progress, the need remains for simple and broadly applicable methods for labeling of methyl groups [76, 77]. Here we report on such a method, based on a 'locally' deuterated, and stereo-selectively methyl ^{13}C-labeled leucine (Leu-meth$_{LD}$, Figure 4.6). Sole use of this amino acid significantly increased the NMR spectral resolution and sensitivity even in otherwise protonated proteins. We successfully test incorporation of Leu-meth$_{LD}$ into the 42-kDa maltose binding protein, the 12-kDa human RBM39$^{245\text{-}332}$ encapsulating RRM2 domain, and the 25-kDa membrane protein bacteriorhodopsin using *E. coli*, insect cell and cell-free protein expression, respectively, without modification of the original expression protocols.

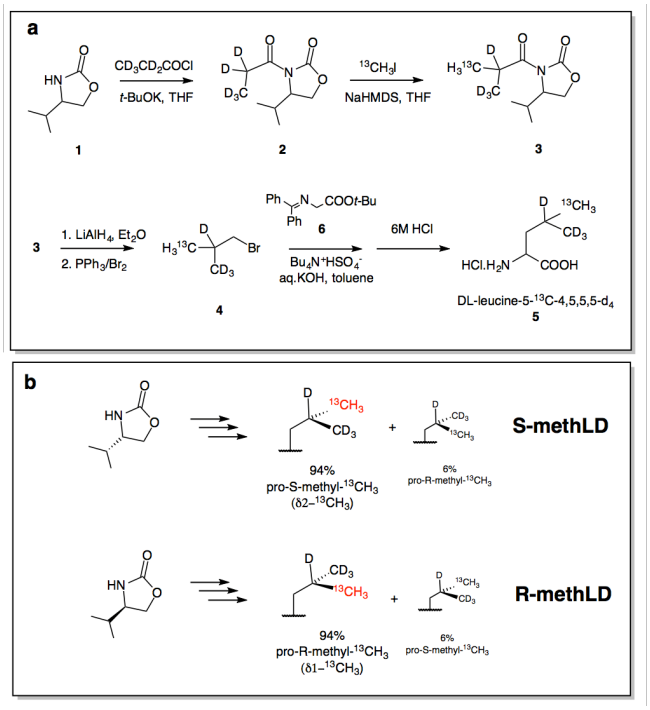


Figure 4.6: (a) Simple and inexpensive synthesis of $^{13}CH_3$-methyl leucine that is "locally" deuterated (Leu-meth$_{LD}$, **5**) to enhance the NMR spectra of non-deuterated proteins; (b) The pro-*R*-$^{13}CH_3$ or pro-*S*-$^{13}CH_3$ methyl configuration of **5** is obtained with 94% stereoselectivity starting from the corresponding Evans' chiral auxilliary **1**.

4.2.2 Results and Discussion

The main source of transverse relaxation (and thus, NMR line broadening) of methyl groups is ^{1}H-^{1}H dipole interactions with surrounding protons. To evaluate the potential of a single-residue labeling approach, we calculated the theoretical contributions of intra- vs. inter-residue interactions to transverse relaxation using a published NMR structure of MBP (PDB ID: 1EZO). The dipole-dipole relaxation contributions of all inter-proton pairs between Leu methyl protons and other protons were computed and averaged over all protein conformations in 1EZO (Material and Methods). We found that for a given methyl group, intra-residue interactions with the other (geminal) methyl group and the adjacent Hγ contributed on average to 37% of the total dipole-dipole relaxation (Figure A24). Interestingly, this number matches the dipole-dipole relaxation contribution of all inter-residue interactions with the rest of the protein (i.e. 38%). Thus, a stereo-selective, intra-residue "local deuteration" of Leu will provide a cost-effective strategy to reduce transverse relaxation sufficiently to generate usable NMR spectra from large, otherwise non-deuterated protein complexes. This strategy is applicable to all commonly used protein expression systems.

The stereo-selective isotope enrichment of leucine and other branched amino acids is amply documented in the literature [7, 26, 49, 61, 72, 115, 116, 150]. The culmination of these efforts is the set of "SAIL" [^{13}C,^{15}N,^{2}H]-labeled amino acids designed by Kainosho's and co-workers [68, 69]. A simpler version of SAIL Leu was shown by the same group to label methyl groups of perdeuterated proteins in bacteria [103]. However, the complex chemistry and expensive starting materials hamper the use of these amino acids in applications other than cell-free protein expression. In our hands the most economical route was a simplified version of a synthetic route by Siebum *et al.* [153] (Figure 4.6a). The isopropyl oxazolidinone (**1**) was acylated with deuterated propionyl chloride, and the product was stereo-selectively methylated with small excess of $^{13}CH_3$-methyl iodide, with an enantiomeric excess of 94%. Reductive cleavage with $LiAlH_4$ ($LiAlD_4$ can be used for deuteration of the β positions), followed by treatment with PPh_3/Br_2 yielded the stereoselective methyl-labeled isobutyl bromide (**4**). Chiral alkylation of activated glycine (**6**) yielded L-Leu, but the reaction proved sluggish unless stoichiometric quantities of expensive Maruoka catalysts [74] were used. To save cost we opted for the racemic DL-Leu which retained the stereochemistry of the labeled methyl groups (Figure 4.6a). The final Leu-meth$_{LD}$ (5-^{13}C-5,5,5,4-d$_4$-DL-Leu, **5**) contained 94% of the desired $^{13}C^{1}H_3$-methyl DL-leucine and 6% of the methyl-inverted configuration (Figure 4.6b). We synthesized and tested the pro-*R* isotopomer (*R*-meth$_{LD}$) in overall yield of 24% from d5-propionic acid and 18% from $^{13}CH_3$-methyl iodide. The cost of reagents per milligram of *R*-meth$_{LD}$ was less than €1 (see Material and Mehtods for more details).

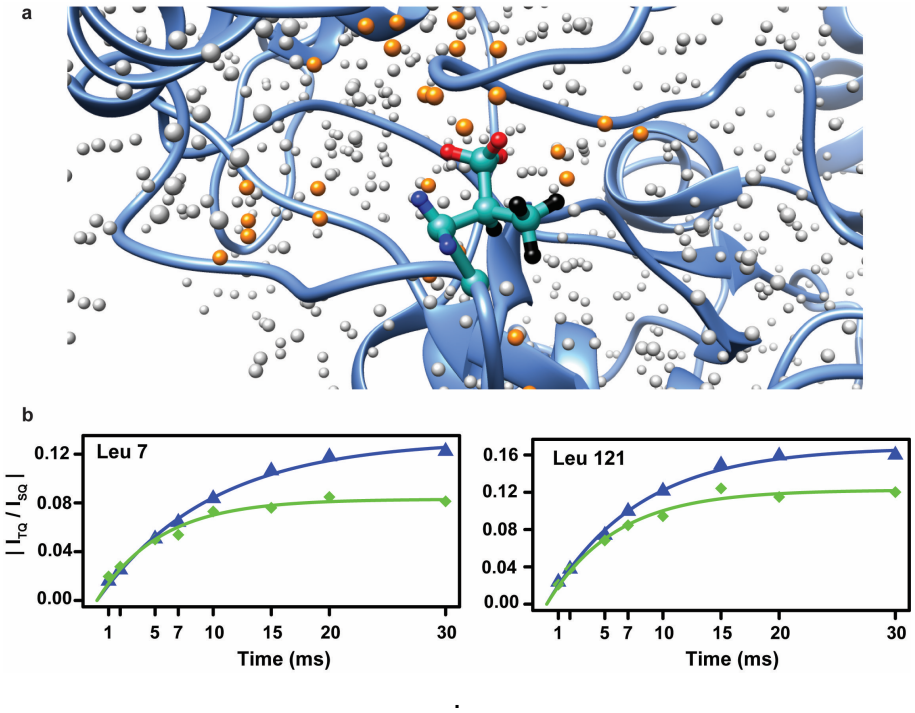


Figure 4.7: Local deuteration reduces the transverse relaxation rate of Leu methyl groups. (a) Enlarged view of the NMR structure of maltose binding protein (MBP, PDB:1EZO) around Leu121. Leu121 is shown as a ball and stick model with deuterated positions in black, Hβ protons in blue, and methyl protons in red. The orange spheres indicate inter-residue protons within 6 Å of the Cδ1 atom of Leu121, and the grey spheres indicate other protons in MBP. (b) Experimentally determined FCT curves of Leu7 δ1[$^{13}CH_3$] and Leu121 δ1[$^{13}CH_3$] in natural hydrogen abundance MBP (green curve) and MBP with δ1[$^{13}CH_3$] that is "locally" δ2 and γ-deuterated with *R*-meth$_{LD}$ (blue curve). The lower the value to which the FCT curve converges, the faster the transverse relaxation.

To experimentally quantify the effect of local deuteration, we labeled MBP in bacteria with the *R*-meth$_{LD}$ and performed forbidden coherence transfer (FCT) experiments [105] (Figure 4.7). FCT experiments measure the ratio of proton signal intensities between the multiple quantum and single quantum coherences as a function of evolution time. This ratio is plotted as a build-up curve in the form of a hyperbolic tangent, where the slope reports on local dynamics (order parameter S^2) and the plateau reports on the local ^{1}H density that contributes to relaxation (Figure A26). A lower plateau indicates faster relaxation resulting from higher ^{1}H density in the vicinity of the methyl proton [105]. We compared the build-up curves of leucine residues in MBP that was (i) δ2-[^{13}C,^{1}H]-labeled and fully protonated, and (ii) δ1-[^{13}C,^{1}H]-δ2-[^{12}C,^{2}H]-γ-[^{2}H]-labeled and otherwise fully protonated. The samples were produced in *E. coli* with the addition of ^{13}C-acetolactate [53] or *R*-meth$_{LD}$ to M9 minimal medium 1 h before induction (Figure A27). The ratio of values to which the two FCT curves of each residue (Leu 7 and Leu 121) in Figure 4.7b plateaus, gives an estimate of the linewidth reduction (5̃0%) due to the reduced dipole-dipole interaction. Similar results were obtained for the other Leu methyl groups in MBP (Figure A25).

To validate the use of Leu-meth$_{LD}$ in other expression systems, we expressed the seven-transmembrane-helix protein bacteriorhodopsin (bR) in an *E. coli*-based cell-

free system. Uniformly ^{13}C-labeled *L*-Leu (8.5 mg) or *R*-meth$_{LD}$ (Figure 4.6b) (17 mg) was used in dialysis mode 3 ml reactions. No difference in expression yield was observed between the samples. bR was refolded and purified in *n*-dodecyl-β-D-maltoside (DDM) micelles, leading to a final protein-detergent complex of approximately 100 kDa [48]. Heteronuclear multiple quantum correlation (HMQC) spectra were acquired on both samples (Figure 4.8a, black vs. red lines). The spectra directly reveal that the usage of *R*-meth$_{LD}$ (i) simplifies the spectral crowding from the 38 Leu residues in bR (Figure 4.8b) by reducing signals from the pro-*S* (δ2) methyl groups and (ii) strongly boosts the signals of the desired pro-*R* (δ1) nuclei. Accordingly, a projection of the 2D-HMQC (along ^{1}H dimensions, Figure 4.8c, e) showed a 30% increase in resolution and 5.5-fold increase in intensity. The absence of ^{13}C-^{13}C ^{1}J-coupling (Figure 4.8d) additionally eliminates the use of constant-time evolution [12] in the ^{13}C dimension, further increasing sensitivity for large molecular weight proteins and allowing high-resolution spectra to be recorded by extended sampling in the ^{13}C dimension using non-uniform sampling.

Overall, the use of Leu-meth$_{LD}$ in the otherwise natural abundance expression system largely improved the quality of the spectrum tipping the scales for many Leu residues to become valuable NMR probes suitable to report on binding interactions [69, 82], dynamic and allosteric mechanisms [87] in large membrane protein complexes.

To test incorporation of Leu-meth$_{LD}$ in insect cells, we expressed the 12-kDa human RBM39$^{245\text{-}332}$ encapsulating RRM2 domain in Sf-9 cells in medium prepared with 75 mg/L of either uniformly ^{13}C labeled Leu or *R*-meth$_{LD}$. The HSQC spectrum with *R*-meth$_{LD}$ (Figure 4.8f, red) displayed the expected eight Leu pro-*R* (δ1) methyl peaks (Figure 4.8g). The quality of the spectrum was sufficient to detect the much weaker signals arising from the residual 6% pro-*S* (δ2) ^{13}C^{1}H$_3$-methyl groups. In line with the results on bR, Leu-meth$_{LD}$ RBM39$^{245\text{-}332}$ exhibited a six-fold improvement in the signal intensity ratio and a two-fold improvement in resolution in the proton dimension, compared to the uniformly ^{13}C-labeled sample (Figure 4.8h-j). While further improvements in the protein spectra are expected if Leu-meth$_{LD}$ is additionally deuterated at the β protons by using LiAlD$_4$ (Figure 4.6a, an additional 20-40% increase in the cost of reagents), the significant improvement seen here in the absence of β-deuteration are well in line with the expected dipolar contributions (Figure A24) and suggests that the local deuteration would be equally successful for methyl labeling of valine and isoleucine.

4.2.3 Conclusion

The enhanced resolution of methyl-NMR has established methyl groups as a unique probe of protein structure, dynamics and function. Spectroscopic advances including NOESY-based resonance assignment [126, 127], solid-state NMR applications [80],

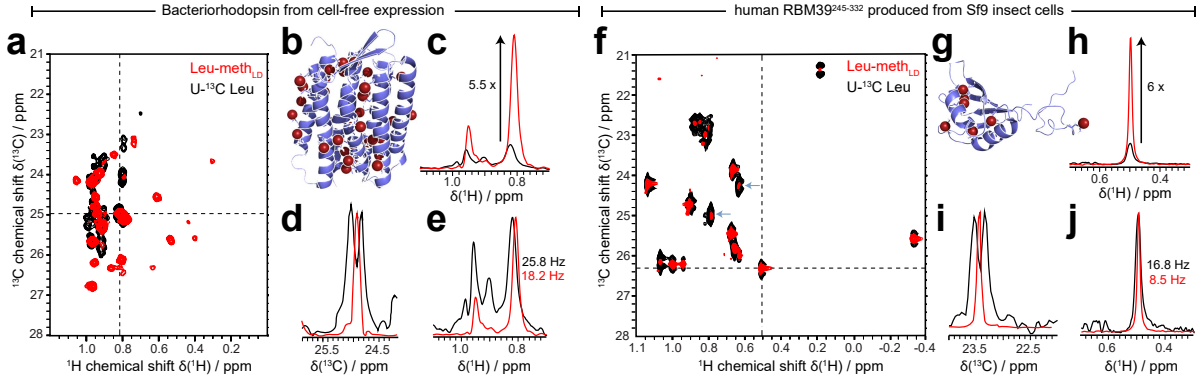


Figure 4.8: Stereo-specific $CD_3/^{13}CH_3$-labeled Leu with Leu-meth$_{LD}$ facilitates NMR characterization of proteins expressed *in-vitro* and in insect cells. (a) [^{13}C,^{1}H]-HMQC spectra of cell-free expressed bR in DDM micelles with [U-^{13}C, ^{1}H] Leu (black) or Leu-meth$_{LD}$ (red). (b) Structure of bacteriorhodopsin (bR, PDB ID:1R84) with ^{13}C, ^{1}H-labeled Leu methyl groups highlighted as red spheres. One-dimensional slices at the indicated frequency in the ^{1}H/^{13}C dimension are shown in c-e. (c) Signal intensity is increased 5.5 fold. Comparison of linewidths of normalized data in the ^{1}H/^{13}C. (d) A singlet is obtained instead of doublet for high resolution acquisition in ^{13}C dimension. (e) Linewidth at half-height is reduced by 29% in the ^{1}H dimension. (f) [^{13}C,^{1}H]-HSQC spectra of the human RBM39$^{245\text{-}332}$ encapsulating RRM2 domain expressed in insect cells with [U-^{13}C, ^{1}H] Leu (black) or Leu-methLD (red). (g) Structure of the CAPER-RRM2 domain (PDB ID: 2JRS) with ^{13}C, ^{1}H-labeled Leu methyl groups highlighted as red spheres. Some of the residual 6% of inverted methyl labeled leucines are also observable (an example marked by a blue arrow). One-dimensional slices at the indicated frequency are shown in h-j. (h) Signal intensity is increased six fold. (j) Linewidth at half height is reduced by 49% in the ^{1}H dimension.

and non-uniform data sampling [131] are constantly enhancing the utility of methyl-NMR. However, challenges in suitable isotope labeling in eukaryotic or cell-free protein expression systems limited the potential of methyl-NMR for many therapeutically relevant biological systems. Here we present the first methyl labeling method that is compatible with most protein expression systems, significantly expanding the range of proteins amenable to methyl-NMR. Our simulations and experiments further clarify the effects of local deuteration on transverse relaxation rates. The fixed deuteration pattern of Leu-meth$_{LD}$ in combination with FCT analysis of surrounding proton density may in addition find new uses as a sensor of protein structure and dynamics.

4.2.4 Materials and Methods

Synthesis of Leu-meth$_{LD}$ Chiral oxazolidinone **1** and NaHMDS (sodium bis(trimethylsilyl)amide, 2 M in THF) were purchased from Fluorochem Ltd (fluorochem.co.uk). d5-propionic acid was purchased CDN Isotopes (cdnisotopes.com). $^{13}CH_3I$ was purchased from Cambridge Isotopes (isotope.com). Diethyl ether and THF were dried by distillation from $LiAlH_4$ and stored over 3A molecular sieves. ESI-MS analysis was performed on Waters Micromass ZQ 2000. NMR spectra were recorded on Bruker AVIII 500 MHz instrument.

Synthesis of (4*R*)-DL-leucine-5-^{13}C-(4,5,5,5)-d4 hydrochloride (5, *R*-meth$_{LD}$)
(*R*)-4-isopropyl-3-(propanoyl-d5)oxazolidin-2-one (2). 1.23 g (11 mmol) tBuOK was suspended in 20 ml dry THF in a dry 100 ml flask. After cooling the suspension to 0 °C (ice bath) a solution of 1.3 g (10 mmol) (4*R*)-isopropyl-2-oxazolidinone (**1**) in 20 ml dry THF was added drop-wise (5 minutes). The reaction mixture was stirred for 30 minutes at 0 °C, then a solution of 1.1 g (11 mmol) C_2D_5COCl (prepared by mixing d5-propionic acid and PCl_5 in inert atmosphere for 3 hours, followed by distillation; yield 90%) in 5 ml dry THF was added drop-wise. Stirring was continued for 1 hour at 0 °C. The mixture was poured into saturated aqueous $NaHCO_3$ solution, the organic phase was separated, and the aqueous phase was extracted three times with 30 ml Et_2O. The combined organic extracts were dried over anhydrous Na_2SO_4 and the solvent was removed under reduced pressure to yield 1.9 g (10 mmol, 91%) **2** as a clear oil. ^{1}H NMR ($CDCl_3$, 500 MHz): δ 0.88 (d, J = 6.9 Hz, 3H, CH_3), 0.92 (d, J = 7.1 Hz, 3H, CH_3), 2.38 (dtd, J_1 = 13.7 Hz, J_2 = 7 Hz, J_3 = 3.9 Hz, 1H, isopropyl CH), 4.22 (dd, J_1 = 9.1 Hz, J_2 = 3 Hz, 1H, H^5), 4.28 (t, J = 8.7 Hz, 1H, H^5), 4.44 (dt, J_1 = 8 Hz, J_2 = 3.8 Hz, 1H, H^4); LRMS (ESI): m/z calcd for $C_9H_{10}D_5NO_3$+H^+: 191.14; found 191.33.
(*R*)-4-isopropyl-3-((*R*)-2-(methyl-^{13}C)propanoyl-2,3,3,3-d4)oxazolidin-2-one (3). A solution of 0.77 g (4 mmol) 2 in 5 ml dry THF was cooled to -78 °C under argon atmosphere and 2.5 ml (4.8 mmol) 1.9 M NaHMDS solution in THF was added via syringe. Stirring was continued under these conditions for 1.5 hours, then 0.97 g (6.8 mmol) $^{13}CH_3I$ dissolved in 2 ml THF was added via syringe. After 15 hours at -78°C the reaction was quenched with 0.5 ml of D_2O and the mixture was warmed to room temperature and poured into saturated aqueous $NaHCO_3$ solution. The organic phase was separated and the aqueous phase was extracted three times with 30 ml Et_2O. The combined organic extracts were dried over anhydrous Na_2SO_4 and the solvent was removed under reduced pressure. The resulting oil was purified by flash chromatography (5% EtOAc in hexanes) to give 0.49 g (2.4 mmol, 60%) of **3** (additional 0.1 g of 3 was found in mixed fractions with unreacted starting material **2**). ^{1}H NMR ($CDCl_3$, 500 MHz): δ 0.86 (d, J = 6.9 Hz, 3H, CH_3), 0.9 (d, J = 7.1 Hz, 3H, CH_3), 1.2 (d, J = 128 Hz, 3H, $^{13}CH_3$), 2.34 (dqd, J_1 = 18.2 Hz, J_2 = 7.1 Hz, J_3 = 4 Hz, 1H, isopropyl CH), 4.2 (dd, J_1 = 9.1 Hz, J_2 = 3.1 Hz, 1H, H^5), 4.26 (t, J = 8.7 Hz, 1H, H^5), 4.43 (tt, J_1 = 7.7 Hz, J_2 = 3.8 Hz, 1H, H^4). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 18.1 ((*R*)-methyl-^{13}C), 19.4 ((*S*)-methyl-^{13}C). Enantiomeric excess of 12/1 was determined based on the intensity of ^{13}C-NMR (*R*)- and (*S*)-methyl-^{13}C signals. LRMS (ESI): m/z calcd for $C_9{}^{13}CH_{13}D_4NO_3$+H^+: 204.15; found 204.24.
(*R*)-1-bromo-2-(methyl-^{13}C)propane-2,3,3,3-d4 (4). A solution of 400 mg (2 mmol) **3** in 2 ml dry Et_2O was added to a cooled (0 °C, ice bath) suspension of 152 mg (4 mmol) $LiAlHt_4$ in 2 ml dry Et_2O, then the reaction mixture was allowed to warm to room temperature with continued stirring for one hour. The Et_2O was removed under vacuum, toluene (2ml) was added to the dry residue and the excess $LiAlH_4$ was

decomposed with 0.2 ml water. The resulting mixture was subjected to atmospheric distillation, collecting 7 ml of distillate until the constant boiling point of toluene (111 °C) was reached. The distillate containing toluene, water, and isobutyl alcohol was dried over anhydrous K_2CO_3, decanted from the drying agent and added to 2.2 mmol Ph_3PBr_2 in 2 ml toluene and 1 ml dimethylacetamide, chilled on ice bath. (The Ph_3PBr_2 was prepared from 0.6 g (2.2 mmol) Ph_3P and 0.35 g (2.2 mmol) Br_2.) After standing for 20 hours at room temperature the reaction mixture was distilled at atmospheric pressure, collecting 5 ml of distillate until the boiling temperature of toluene (111 °C) was reached. The distillate containing toluene and the desired isobutyl bromide **4** was used in the next step without further purification.

(4*R*)-DL-leucine-5-^{13}C-(4,5,5,5)-d4-hydrochloride (5). To the solution of butyl bromide **4** in toluene, 0.75 g (2.5 mmol) *tert*-butyl 2-((diphenylmethylene)amino)acetate, 0.45 g (1.7 mmol) tetrabutyl ammonium HSO_4^- and 3 ml 10 M NaOH was added. The mixture was stirred for 15 hours at room temperature and extracted three times with EtOAc. The organic extracts were washed with water and brine, dried over anhydrous Na_2SO_4 and the solvent was removed under reduced pressure. The product was purified by flash chromatography (4% EtOAc in hexane) to yield 390 mg of white solid which was dissolved in 3 ml toluene. 3 ml 6 M HCl was added and the mixture was refluxed for 2 hours. The aqueous layer was separated, the organic phase was washed with 3 ml of water and the combined aqueous phases were evaporated to dryness on a rotary evaporator. The residue was dissolved in 0.5 ml of absolute ethanol and the product **5** was precipitated with 5 ml of Et_2O. Yield: 170 mg (1 mmol, 50% from 3, 27% from 1). ^{1}H NMR (D_2O, 500 MHz): δ 0.87 (dd, J_1 = 125.3 Hz, J_2 = 6.2 Hz, 3H, $^{13}CH_3$), 1.65 (ddd, J_1 = 13.8 Hz, J_2 = 8.8 Hz, J_3 = 4.5 Hz, 1H), 1.76 (dt, J_1 = 14.3 Hz, J_2 = 4.7 Hz, 1H), 3.94 (dd, J_1 = 8.1 Hz, J_2 = 6.1 Hz, 1H). ^{13}C NMR (D_2O, 125 MHz): δ 20.6, 21.4. LRMS (ESI): m/z calcd for $C_5{}^{13}CH_9D_4NO_2+H^+$: 137.12; found 137.19.

Overall yield of leucine **5** from d5-propionate and from $^{13}CH_3I$ was 24% and 18%, respectively. At the fairly standard prices at which these reagents were obtained ($66/gram and $100/gram, respectively) the cost of labeled leucine in terms of isotope labeled compounds is 0.75 USD per milligram (0.66 euro). Use of $LiAlD_4$ to further deuterated the 3-position would raise this to about 1 euro per milligram. The following table lists the amount of leucine **5** used to produce the protein samples discussed in this work.

Preparation of maltose binding protein in *E. coli* BL21 (DE3). *E. coli* were transformed with plasmid encoding the protein MBP (pMALC4X), Cultures were grown at 37 °C in a shaker incubator in M9 minimal medium prepared in 100% H_2O containing 100 µg/l Carbenicillin, 1 g/l $^{15}NH_4Cl$ and 2 g/l ^{12}C-^{1}H-glucose as the sole carbon source. For uniform or stereoselective methyl labeling of leucine residue, 75 mg/l ^{13}C-acetolactate,

Table 4.3: Amounts of Leu-meth$_{LD}$ used for different protein expressions.

Purified and refolded protein (mg)	Expression system / liters	Leu-meth$_{LD}$ (mg)
MBP (100 mg)	*E. coli* / 1 l	75 mg
RBM39 (2 mg)	Sf9 / 1 l	75 mg
bR (1 mg)	Cell-free / 3 ml	17 mg

or Leu-meth$_{LD}$ amino acid was added to the culture one hour prior to induction. At an OD_{600} of 0.7–0.8, the cells were induced with 1 mM IPTG and grown at 28 ℃ for an additional 16 h. Cells were pelleted by centrifugation for 30 min at 4 ℃ at 3,500g. Cell pellets were resuspended in 40 ml MBP lysis buffer (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA) per liter of original bacterial culture, lysed by sonication, and clarified by centrifugation for 40 min at 4 ℃ at 33,000g. The lysate was loaded onto 5 ml amylose resin. After the resin was washed with resuspension buffer, MBP was eluted with the addition of 40 mM d(+)-maltose. MBP was further purified by size-exclusion chromatography (GE Healthcare Life Sciences, Superdex 75 10/300 GL prepacked size-exclusion chromatography column) into NMR buffer (10 mM HEPES, pH 6.5, 1 mM EDTA, 1 mM β-cyclodextrin).

Cell free synthesis of bacteriorhodopsin. A construct of bacteriorhodopsin (bR) containing a N-terminal deca-Histidine tag was cloned into a pIVEX2.4d vector for the production of plasmid DNA. Bacterioopsin (bO) was expressed in an *E. coli*-based cell-free expression system following established protocols [4,53,75,92,129,145,166]. Custom-made mix of 19 unlabeled amino acids (8.25 mg of Ala, Ser, Met, Val, Thr, Pro, Arg, Ile, His, Tyr, Lys, Phe, Asp, Gly, Glu; 7.95 mg of Cys; 13.3 mg of Trp; 9.77 mg of Asn; 9.51 mg of Gln) supplemented with 8.5 mg of uniformly labeled [U-^{13}C-^{15}N]-L-Leu or 17 mg of racemic, stereo-specifically labeled Leu-meth$_{LD}$ was used. 3 ml dialysis mode reactions were carried out at 28℃ for 16 h using 2.2 µg of plasmid DNA in the 300 µl reaction mix and in the absence of retinal or detergents. The reaction mix was centrifuged for 1 h at 21000 g and 4℃. The resulting pellet was then washed with 10 volumes of buffer containing 50 mM sodium phosphate (NaH_2PO_4/Na_2HPO_4) pH 7.5, 150 mM sodium chloride (NaCl) and Complete protease inhibitors. Refolding of bR was initiated by resuspending 3 ml of CFE protein pellet in an equal volume of refolding buffer (1 M NaCl; 5% *n*-dodecyl-β-D-maltoside (DDM) (w/v); 100 µM *all-trans* retinal; 40 mM NaH_2PO_4/Na_2HPO_4, pH 7.5). The sample was incubated for 1 h protected from light at RT. Purification of refolded bR was performed using a Micro BioSpin™ gravity-flow column (Bio-Rad, Hercules, USA) prepared with Ni-NTA Agarose (Macherey-Nagel, Germany) for a total column volume (CV) of 800 µl. The column was equilibrated with 10 CV of buffer P (150 mM NaCl; 0.4 mM DDM; 50 mM NaH_2PO_4/Na_2HPO_4 pH 7.5) before loading the sample. After washing the column with

3 x 2 CV of buffer P supplemented with 20 mM imidazole and Complete protease inhibitors, bR was eluted with 4 x 1 CV buffer P supplemented with 250 mM Imidazole and Complete protease inhibitors. Elution fractions containing bR were combined and dialysed over night against 2 L of NMR buffer (50 mM NaCl; 0.4 mM DDM; 20 mM NaH_2PO_4/Na_2HPO_4 pH 7.5). Sample was then concentrated (10 kDa MWCO, Vivaspin 4, PES membrane, Sartorius, Germany) to a volume of roughly 200 µl. The final concentration of bR was measured using a Spectrophotometer (V-650, Jasco, Germany) using the extinction coefficient of fully folded bR at 555 nm of ε= 55000 L cm^{-1} mol^{-1}.

Human RBM39$^{245-332}$ expression in the insect cells. Human RBM39$^{245-332}$ encapsulating RRM2 domain was cloned in pAC-derived vectors42 and recombinant RBM39$^{245-332}$ was expressed as N-terminal His6 fusion in *Spodoptera frugiperda* Sf-9 insect cells using the baculovirus expression system (Invitrogen). Sf-9 cells for protein expression and production were routinely maintained in ESF 921 medium. For isotopic labeling ESF 921 amino acid deficient media was used. Expression media was first supplemented with custom-made mix of 19 unlabeled amino acids (225 mg L-Ala, 200 mg β-Ala, 700 mg L-Arg HCl, 350 mg each L-Asp and L-Asn, 25 mg L-Cys, 600 mg each L-Glu and L-Gln, 650 mg Gly, 2500 mg L-His, 50 mg L-Ile, 625 mg L-Lys, 50 mg L-Met, 150 mg L-Phe, 350 mg L-Pro, 1100 mg L-Ser, 175 mg L-Thr, 100 mg L-Trp, 50 mg L-Tyr, and 100 mg L-Val) along with 70 mg/l uniformly labeled [U-^{13}C-^{15}N]-L-Leu or racemic, stereo-specifically labeled leucine (Leu-meth$_{LD}$), followed by filtering through a 0.22 µm Corning disposable bottle-top filter. Expression media was exchanged for isotopic labeling. Cells in ESF 921 media were centrifuged at 500 g for 5 min, supernatant was discarded, and cells were resuspended in isotopic labeling ESF 921 media. Cells were infected at density of 1-2x106 cells per ml with 1.5% (v/v) of P3 virus. For purification cells were harvested and resuspended in buffer containing 50 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) pH 8.0, 200 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1x protease inhibitor cocktail (Sigma) and lysed by sonication. Following ultracentrifugation, the soluble fraction was passed over Ni Sepharose 6 Fast Flow affinity resin (GE Healthcare) and eluted with wash buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM TCEP) supplemented with 100 mM imidazole (Fischer Chemical). The affinity-purified protein was subjected to size exclusion chromatography in 50 mM HEPES pH 7.4, 200 mM NaCl and 1 mM TCEP. The protein-containing fractions were concentrated using ultrafiltration (Millipore) and flash frozen in liquid nitrogen and stored at -80℃.

Simulating proton density around leucine methyl protons in the protein MBP. The NMR structure of Maltose Binding Protein (PDB: 1EZO) was used in the analysis. The inverse sixth power of the Euclidean distance for each of the HD11, HD12,

and HD13 nuclei of all thirty leucines in MBP from the rest of ^{1}H nuclei in the MBP was computed. The total ^{1}H-^{1}H dipole-dipole relaxation was categorized in three non-overlapping categories. These categories are a) intra-residue HD21, HD22, HD23 and HG nuclei of leucine, b) intra-residue HB2, HB3, HN and HA nuclei of leucine, and c) the rest of the inter-residue protons from the other amino acids. The percentage contribution of intra-residue leucine HD23, HD22, HD23 and HG nuclei towards the dipole-dipole relaxation (DDR) to a protons HD11, HD12 and HD13 of leucine was calculated using equation 4.3 and was averaged to simulate the fast rotation of methyl protons. These contributions were then averaged over all the ten conformations of MBP deposited in PDB. These percentage contribution from the three categories are plotted as the stacked bar graphs in Figures A24 and A25.

$$\textit{Percentage contribution to DDR} = 100 \cdot \frac{\sum \frac{1}{r^6_{\text{intra,HD2/HG}}}}{\sum \frac{1}{r^6_{\text{intra,HD2/HG}}} + \sum \frac{1}{r^6_{\text{intra,HB/HA}}} + \sum \frac{1}{r^6_{\text{intra,H}}}} \tag{4.3}$$

Acquisition and processing of NMR spectra. The FCT NMR experiments were recorded on Bruker Avance III spectrometer operating at 800 MHz, equipped with a triple resonance TCI cryoprobe. Data was collected at 25℃ in 10 mM HEPES, pH 6.5, 1 mM EDTA, 1 mM β-cyclodextrin and 5% (v/v) deuterium oxide. [^{13}C-^{1}H]-Heteronuclear Triple Quantum Coherence (3QC) and [^{13}C-^{1}H]-Heteronuclear Single Quantum Coherence (1QC) spectra were acquired under the same acquisition parameters for both uniform and stereo-specific leucine samples (parameters given in Table A9). Nine interleaved experiments with the coherence evolution time of 1, 2, 5, 7, 10, 15, 20, and 30 ms were acquired for both 3QC and 1QC. The spectra were processed using Nmrpipe and analysed using CCPNMR. The signal was apodized with shifted sine function and zero-filled to 2048 and 256 complex points in the direct and indirect dimension respectively. For bR, NMR experiments were performed on a Bruker Avance III HD+ spectrometer operating at 750 MHz, equipped with a triple resonance TCI (^{1}H, ^{13}C, ^{15}N) cryoprobe. Data was collected at 25℃ in 20 mM NaH_2PO_4/Na_2HPO_4 pH 7.5, 50 mM NaCl, 0.4 mM DDM, 0.3% (v/v) sodium azide and 5% (v/v) deuterium oxide. [^{3}C-^{1}H]-Heteronuclear Multiple Quantum Correlation (HMQC) spectra was acquired under the same acquisition parameters for both uniform and stereo-specific leucine samples. That is, 336 scans with 512 points (11.4 ms) in the direct (^{1}H) dimension with a spectral width of 3 ppm and 206 points (2.7 ms) in the indirect (^{3}C) dimension with a spectral width of 20 ppm. Spectra were processed with TOPSPIN 3.2 (Bruker) and plotted with Sparky [70]. Both were zero-filled to 2048 points in the direct dimension and 1024 points in the indirect dimension. Shifted sine-bell apodization and polynomial baseline correction was applied to both dimensions, as well as linear prediction in the indirect dimension. For human RBM39$^{245\text{-}332}$, NMR experiments were acquired on a Bruker Avance III spectrometer operating at 800 MHz

and equipped with a triple resonance TXO cryoprobe. Data was collected at 25℃ in 20 mM NaH_2PO_4/Na_2HPO_4 pH 7.5, 50 mM NaCl, 0.3% (v/v) sodium azide and 5% (v/v) deuterium oxide. [^{13}C-^{1}H]-Heteronuclear Single Quantum Correlation (HSQC) spectra was acquired under the same acquisition parameters for both uniform and stereo-specific leucine samples (parameters given in Table A9). Spectra were processed with TOPSPIN 3.2 (Bruker) and analysed using CCPNMR. Both were zero filled to 8096 complex points in the direct dimension and 2048 complex points in the indirect dimension. Shifted quadratic-sine bell apodization was applied to both the dimension with no linear prediction.

Forbidden Coherence Transfer experiment to estimate local proton density. In a Forbidden Coherence Transfer (FCT) NMR experiment, a set of two [^{13}C, ^{1}H] HMQC type experiments are recorded. The triple quantum coherence and single quantum coherence are evolved in each set respectively for a defined delay (T). The ratio of intensities of the two coherence at given time T can be determined by the equation 4.4, where the parameters η and δ are defined in equations 4.5 and 4.6, respectively. The delays used in this study are 1, 2, 5, 7, 10, 15, 20, and 30 ms. Equation 4.5 was used to fit the curve in the Figure 4.7b. The parameters in the equations are: μ_0 – the magnetic permeability of free space; Θ_{axis} – the angle between a carbon-hydrogen bond and the three-fold symmetry axis of the methyl group; γ_H - the gyromagnetic ratio of proton; τ_C - the rotational tumbling time and $\hbar$ - the Planck constant divided by 2π. We used P_2 (cos $\Theta_{axis,HH}$) = -0.5.

$$\frac{I_{3qc}}{I_{1qc}} = \frac{-0.75 \cdot \mu \cdot tanh(\sqrt{\nu^2 + \delta^2} \cdot T)}{\sqrt{\nu^2 + \delta^2} - \delta \cdot tanh(\sqrt{\nu^2 + \delta^2} \cdot T)} \tag{4.4}$$

$$\nu = \frac{9}{10}(\frac{\mu_0}{4\pi})^2[P_2(cos\Theta_{axis,HH})]^2 \frac{S^2 {\gamma_H}^4 \hbar^2 \tau_C}{{r_{HH}}^6} \tag{4.5}$$

$$\delta = -\frac{1}{5}(\frac{\mu_0}{4\pi})^2 \sum \frac{{\gamma_H}^4 \hbar^2 \tau_C}{r_{HH}^6} \tag{4.6}$$

According to equation 4.6, δ contains the cumulative sum of the inverse six power of all protons, which is reflective of the dipolar relaxation caused by protons. Thus, we simulated the effect of variation in the value of δ on the shape of FCT curves, as shown in Figure A24. The curves demonstrate that the higher the magnitude of δ, the lower the FCT curve plateaus, indicating the relatively higher proton density (effectively causing more dipole-dipole relaxation). The η in the equation 4.5 doesn't capture the dipole-dipole contribution from other protons and reflects the mobility of methyl group. Interestingly, the ratio of δ values from two FCT curves can capture the dipole-dipole contribution to transverse relaxation of methyl protons. For example, if the magnitude of the FCT curve 'A' plateaus to a value of 0.15 and another FCT curve 'B' plateaus to

a value of 0.1, it translates to 50% less transverse relaxation, hence narrower linewidth of ‘A’ compared to ‘B’ due to the dipole-dipole relaxation.

4.3 Hormones of the Melanocortin Receptor-4 System

4.3.1 Introduction

The melanocortin-4 receptor (MC4R), a member of the G protein-coupled receptor (GPCR) family, is expressed in hypothalamic neurons and plays a key role in energy homeostasis and food intake [64, 89, 162]. MC4R is targeted by various ligands including the neuropeptide agonists adrenocorticotropic hormone (ACTH) and α-melanocyte stimulation hormone (α-MSH), which trigger anorexigenic signals [125], and its inverse agonist Agouti-related protein (AgRP), which leads to orexigenic signalling [118]. The amino acid sequence of α-MSH is identical with the first 13 amino acids of ACTH, since both peptides derive from the precursor protein propiomelanocortin (POMC) [125]. Both agonists are unstructured in solution, but may adopt a structure upon receptor binding [59]. However, MC4R activation efficiency is dramatically increased using analogs with acetylated N-terminus, and substitutions by norleucin and *D*-phenylalanin at position 4 and 7, respectively (NDP-α-MSH, NDP-ACTH) [143]. AgRP consists of a N-terminal signalling sequence, which is cleaved after translation leading to a 15 kDa protein with an unstructured N-terminus and a Cys-rich C-terminal domain. The ten cystein residues within the C-terminal region are forming five intramolecular disulfide bonds and stabilizing three loops, which are modulating the deactivation of MC4R [27, 97]. Especially residue triplet R111, F112, and F113 of the middle loop are crucial for the reverse agonistic effect, while the other two loops are influencing activity and selectivity of AgRP (Figure 4.9a) [99, 177]. Surprisingly, the bone-derived lipocalin 2 (LCN2) is also reported to activate melanocortic-4 signalling [107]. A recent X-ray crystallographic structure of MC4R gives insights into the binding of artificial antagonist SHU9119 and secondary messenger Ca^{2+} modulation of the receptor [185]. Based on the crystal structure of MC4R, NMR spectroscopy may give site-specific information about binding behaviour of other ligands. Although (isotope-labeled) production of the receptor and the preparation of a NMR-accessible sample is clearly the largest obstacle, expression and purification strategies for isotope-labeled MC4R hormones have to be developed, too. Neuropeptides, generally, are in contact with different membranes including synaptic vesicles and pre- and postsynaptic cell walls *in vivo*, and with several membrane mimetics *in vitro*. It is important to understand the interplay between ligands, membranes and divalent Ca^{2+} ions firstly, in order to interpret MC4R binding studies correctly.
In this study, a purification strategy for antagonistic hormone AgRP was developed as well as NMR analysis of different LCN2 variants were performed. To analyse hormone-membrane interactions, fluorescence spectroscopy of a fluorophore-labeled ACTH variant and NMR studies with a ^{15}N-α-MSH were applied.

4.3.2 Results and Discussion

Purification and Refolding of AgRP$^{23\text{-}132}$.
Full-length AgRP (AgRP$^{23\text{-}132}$) as well as only the C-terminal domain (AgRP$^{83\text{-}132}$) could be successfully over-expressed in an *E. coli* cell system, also in isotope-enriched minimal medium (Figure A28a). After cell lysis and centrifugation of the expression culture, no AgRP could be detected in the supernatant, but as sediment with the cell debris. During overexpression the high amount of cysteines in AgRP's amino acid sequence leads to random intra- and intermolecular crosslinks and result in the formation of misfolded protein as well as dimers and higher oligomers, which are usually forming together with other endogenous biomolecules dense particles, called inclusion bodies (IB). Insoluble IB were separated from the residual sediment, such as membranes and nucleic acids, by applying several washing steps with detergent and high-salt containing buffers (Figure A28b). IB could be solubilized with 6 M chaotropic guanidine hydrochloride (GdnHCl) and a high amount of reducing agents (50 mM dithiothreitol (DTT)) to break sulfur-sulfur bonds. Subsequently immobilized metal affinity chromatography (IMAC) after solubilisation leads to highly pure and concentrated AgRP in unfolded state. To avoid reduction of the column matrix and rebuilding of disulfide bridges, solubilized IB were diluted 5-fold to an end concentration of maximal 10 mM DTT.
During refolding, proteins undergo a complex pathway of several intermediate, semi-stable states, which often leads to non-native insoluble conformations. It is necessary for all refolding attempts to change from a denaturing environment into an environment favourable for the native state. The refolding conditions for AgRP were adapted from published procedures [65, 118, 135] and the refolding buffer contained 2 M GndHCl, 5% (v/v) glycerol, 1 mM reduced glutathione (GSH), and 0.2 mM oxidized glutathione (GSSG). Moderate GdnHCl concentrations are considered to not unfold stable proteins, but to stabilize intermediate, partially folded conformations [21]. However, refolding assays, in which misfolded aggregates were observed via scattering at 370 nm, did not show any influence of glycerol or GndHCl concentrations between 0.5 and 2 M, but samples without GdnHCl show high amounts of insoluble particles (Figure A29b). For disulfide bridged proteins a pair of reductant and oxidant, as monomeric (GSH) and dimeric glutathione (GSSG), is essential, since it allows efficient reshuffling of cystein bonds. Several refolding strategies are described, such as dialysis or on-column refolding (see Busgess [21] for an overview). In this study highly concentrated unfolded AgRP was slowly dripped into a large volume of strongly stirring refolding buffer to exclude oligomerization of semi-stable AgRP conformations. Another IMAC was applied to concentrate AgRP solution after refolding and buffer was exchanged using dialysis. Two bands in a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) under non-reducing conditions are visible around 15 kDa hight, where

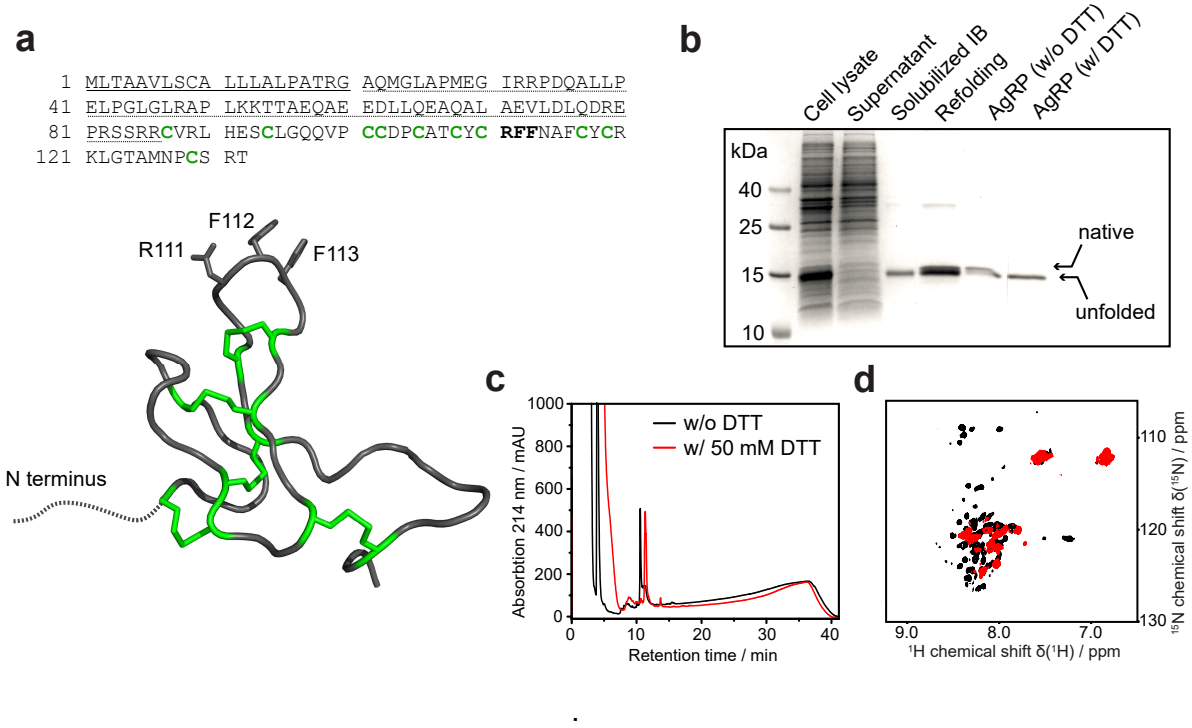


Figure 4.9: Production and analysis of Agouti-related Protein (AgRP). (a) AgRP consists of a N-terminal signalling sequence (straight underlined sequence), which is cleaved off after translation, followed by an unstructured part (dotted underlined sequence). The C-terminus is structured into three loops and stabilized by disulfide bridges involving ten cysteins (green in structure and sequence). A RFF triplet within the middle loop is crucial for AgRP's antagonistic function (bold in sequence, sticks in structure). (b) Non-reducing SDS-PAGE shows the different steps of AgRP production. After expression cells were lysated (lane 1) and insoluble parts were separated by centrifugation, whereas no AgRP could be found in the supernatant (lane 2). Inclusion bodies (IB) were purified from other cell fragments in several washing steps and solubilized by high amounts of GndHCl and DTT (lane 3). Solubilized AgRP was refolded in buffer containing 10 mM Tris/HCl pH 8, 6 M GdnHCl, 50 mM DTT (lane 4). Refolded disulfide-bridged AgRP (lane 5) shows an electrophoretic mobility of a 2 kDa larger protein than reduced unfolded AgRP (lane 6). (c) Refolded AgRP can be separated from un- and misfolded states by reversed phase high performance liquid chromatography (RP-HPLC), where its folded form shows a shorter retention time. (d) Mixtures of folded and unfolded ^{15}N-AgRP show better signal-to-noise ratio at pH 5.0 (black) than at pH 7.6 (red) in a [^{1}H,^{15}N]-HSQC spectrum.

the lower one matches the size of unfolded AgRP (Figure 4.9b). Rosenfeld *et al.* [135] have already shown that dislufide bridged AgRP appears approximately 2 kDa larger in a SDS-PAGE than its reduced variant. Interestingly, during dialysis or some additional incubation time, the ratio between native and unfolded AgRP often increased, however this was not reproducible. Figure 4.9b shows a denaturing polyacrylamide gel of the different purification and refolding steps.

During IMAC, predominantly unfolded AgRP elutes in later elution steps (Figure A29a), but complete separation of both states could only be achieved by reversed phase high performance liquid chromatography (RP-HPLC). Refolded AgRP$^{23\text{-}132}$ elutes in analytical RP-HPLC runs earlier and nearly baseline-separated from its reduced form (Figure 4.9c). However, preparative-scale HPLC runs to acquire NMR-sufficient concentrations were lacking sufficient elution peak separation, yet. A preparation of ^{15}N-enriched AgRP showed approximately 50% native conformation after refolding and [^{1}H,^{15}N]-

HSQC spectra of this sample were recorded at neutral and acid pH (Figure 4.9d). While peak linewidth were unfavourable broad at neutral pH, the peak pattern at low pH showed a similar distribution as published by Jackson *et al.* [65] Most of the signals lay around the region for unstructured proteins around 8.5-7.5 ppm ^{1}H chemical shift, but it is unclear if it derives from unfolded protein or the unstructured N-terminus of AgRP.

MC4R hormones interact with charged membranes.
Peptides, in general, often interacting with membrane surfaces by electrostatic or hydrophobic interactions [8, 73]. This might play an important role for hormone signalling, but complicates *in vitro* studies of receptor-ligand interaction, since membrane and receptor binding has to be distinguishable. Nanodiscs are lipid bilayers surrounded by two membrane scaffold proteins (MSP) and are suitable membrane mimetics to study membrane-associated proteins by solution NMR [172, 173]. To study the interplay of membranes and hormones, [^{1}H,^{15}N]-correlation spectra of ^{15}N-enriched MC4R agonist α-MSH were recorded in presence of different nanodisc systems. No interaction could be detected by solution NMR in presence of nanodiscs consisting only out of overall neutrally charged phospholipids (100% DMPC). In presence of moderately and highly negative charged nanodiscs (70% POPC/ 30% POPS and 50% POPC/ 50% POPS, respectively) binding-induced CSP could be observed, whereas the shifts were larger for the higher charged 50% POPC/ 50% POPS nanodiscs (Figure 4.10a). Interestingly, in presence of 10 mM $CaCl_2$, while Na^{+} concentration were reduced to keep a comparable ionic strength, α-MSH dissociates from the nanodisc surface, resulting in a spectrum identical to free α-MSH. This might indicate, that Ca^{2+}-induced release of hormone from vesical membranes is another regulative mechanism for MC4R signalling, additionally to its well-studied function as a secondary messenger [162] or as cofactor [185]. The membrane-bound state of α-MSH was further analysed by recording ^{13}C-correlated spectra of ^{15}N,^{13}C-labeled peptide in presence of 50% POPC/ 50% POPS nanodiscs. [^{1}H,^{13}C]-TROSY and [^{13}C,^{13}C]-FLOPSY experiments were applied to assign chemical shifts of Hα, Hβ, Cα, and Cβ nuclei (Figure A30). A secondary chemical shift analysis of Cα, and Cβ gives positive values for α-helix and negative for β-sheet elements [179], but α-MSH does not adopt a clear secondary structure in free or membrane-bound form, which is unusual for membrane-interacting peptides (Figure 4.10b).
It is important to distinguish between unspecific membrane interaction and receptor activation, and therefore a binding assay was developed using *TNAO38* cell membrane extracts containing and not containing MC4R receptor. Since all residues of α-MSH seem to be significantly affected by membrane-binding, a C-terminal shortened variant of the highly MC4R-activating NDP-ACTH (NDP-ACTH23C) was covalently link with

a fluorophore at its artificial C-terminal cystein (Figure A31). Membranes with and without MC4R were diluted to different membrane concentrations and were incubated with an excess of fluorophore-labeled ligand in a filter plate in presence of $CaCl_2$ containing buffer. Unbound ligand was removed by centrifugation and the membranes were washed with the same buffer, while the content of ligand in the flow-through was followed by fluorescence. Figure 4.10c shows the fluorescence of the flow-through at different membrane concentrations of MC4R-overexpressed *TNAO38* membranes corrected by the fluorescence of *TNAO38* membranes without MC4R expression. Specific receptor binding of the hormone could be observed, whereas the effect is largest for higher membrane content.

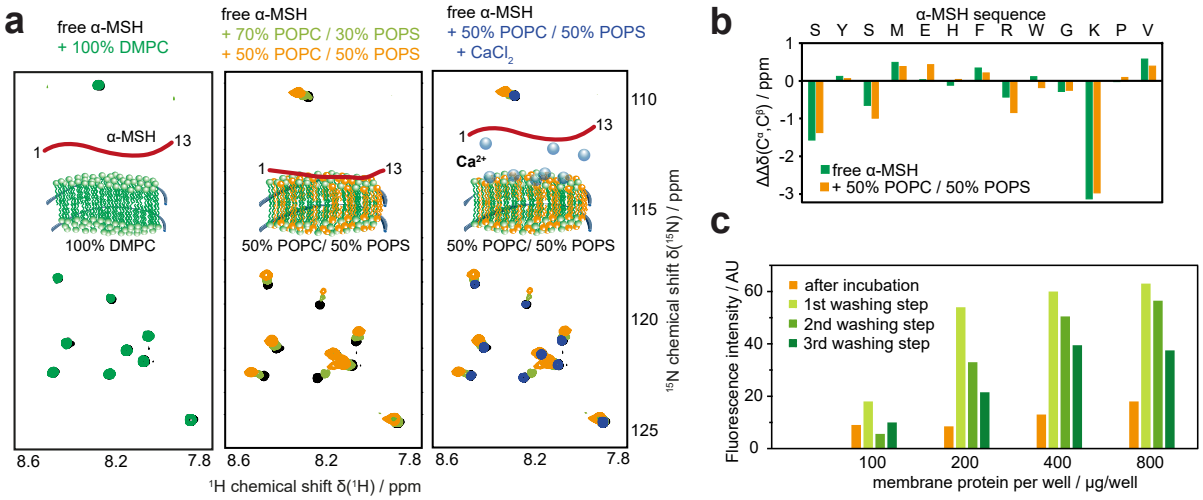


Figure 4.10: Membrane Interaction of MC4R Hormones. (a) [^{1}H,^{15}N]-HSQC spectra of free ^{15}N-α-MSH in absence (black) and presence of 100% DMPC nanodiscs (green), and bound states in presence of nanodiscs consisting of 70% POPC/ 30% POPS (light green) or 50% POPC/ 50% POPS (orange). Spectra of of ^{15}N-α-MSH with 50% POPC/ 50% POPS nanodiscs and 10 mM $CaCl_2$ (blue) are identical with the free form. (b) Secondary chemical shift for Cα and Cβ (ΔΔδ(Cα,Cβ)) of ^{15}N,^{13}C-α-MSH bound to 50% POPC/ 50% POPS (orange) indicates no change in secondary structure compared to its free form (green). (c) Membrane preparations from MC4R-overexpressing and non-expressing *TNAO38* cells were incubated with fluorophore-labeled NDP-ACTH23C in filtering plates and washed several times. The graph shows the fluorescence of the flow-through of MC4R-containing membrane samples corrected by the reference samples without MC4R at different membrane concentrations. NDP-ACTH23C binding to MC4R-containing membranes do not show high difference to reference membranes (orange), while unspecific bound peptides are washed off already in the first step (light green) as seen by the saturating effect at high membrane concentrations.

LCN2 does not activate MC4R in *in vitro* assays.

Mosialou *et al.* published [107] an activation of MC4R by bone-derived lipocalin-2 (LCN2) in mouse models and a cell-based activity assay. This is surprising, since no

further experimental evidence so far linked hormones expressed in bone tissue to appetite regulation. However, LCN2 is a globular shaped protein with a molecular size of approximately 20 kDa, and therefore a suitable target for NMR studies [23]. High-resolution [^{1}H,^{15}N]-correlation experiments of isotope-enriched human (hLCN2) and mouse LCN2 (mLCN2) yield highly dispersed spectra for both variants (Figure 4.11a). The amide chemical shifts of hLCN2 and mLCN2 have similar patterns with some clear differences most likely caused by the 38% non-identical amino acids (see Material and Methods).

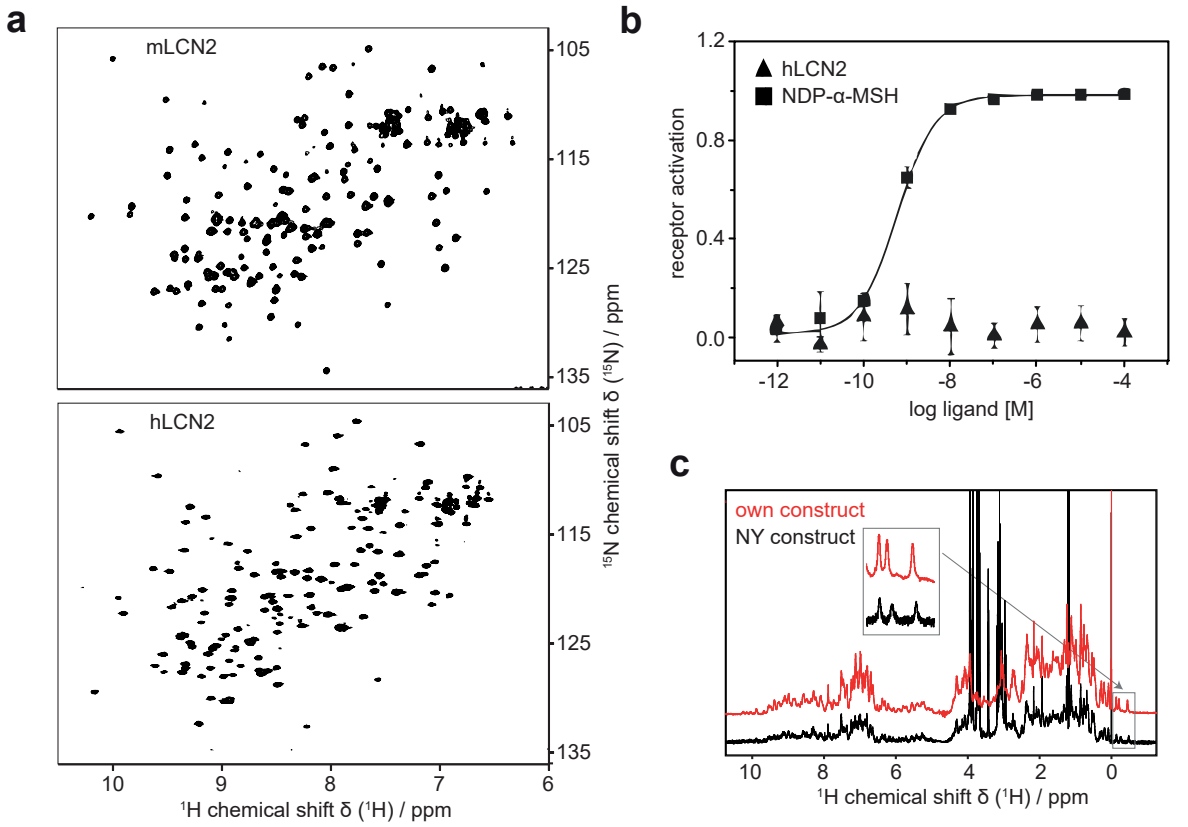


Figure 4.11: Lipocalin-2 (LCN2) Characterisation. (a) [^{1}H,^{15}N]-TROSY spectra of ^{15}N-labeled mouse (mLCN2) and human LCN2 (hLCN2) show similarly dispersed peak patterns. (b) Dose-response curve of MC4R-overexpressing HEK293F cells incubated with different concentrations of ligands NDP-α-MSH and hLCN2. The receptor activation was followed by directly measuring GCPR-depended cAMP concentrations. (c) ^{1}H 1D spectra of own produced hLCN2 variant and variant derived from Kousteni lab (Columbia University, New York) show high similarities (red and black, respectively). Inlet shows region of high-shielded methyl groups which are indicative for ternary structure.

Ligand-induced MC4R activity was observed in over-expressing HEK293F cells by measuring the concentration of cyclic adenosine monophosphate (cAMP) in an assay

based on cAMP-specific antibodies and Förster resonance energy transfer (FRET). This assay allows to directly quantify the amount of intracellular cAMP which is formed due to GPCR response after a certain incubation period with ligands (see Material and Methods). No MC4R activity could be detected for any of the tested LCN2 variants, while the highly activating agonist NDP-α-MSH triggered strong receptor response (Figure 4.11b). [^{1}H,^{15}N]-correlation spectra suggest, that LCN2 structure and interaction is not affected by the presence of membrane mimetics (Figure A32), the oxidation state or the buffer used in the activation assay (data not shown). A LCN2 sample used in the published activity report [107] from Columbia University, New York City, USA, did also not activate MC4R in the used assay, and ^{1}H 1D NMR spectra indicate high structural similarity to the own LCN2 construct (Figure 4.11c). One explanation for the contrary results which are presented here and the one published by Mosialou *et al.* is the difference in the methodology to detect receptor activation. While the assay used here directly detects the GPCR-depending cAMP concentration, the setup used by Mosialou *et al.* relies on down-stream activation of an encoded Luciferase gene. The output of both methods might differ for weakly activating ligands, however a convincing explanation of the contradicting results is still missing.

4.3.3 Conclusion

MC4R is regulating energy homeostasis and is targeted by a variety of agonistic and antagonistic hormones. While the availability of soluble and active receptor is indisputable the biggest obstacle for structural research by solution NMR, also production strategies for isotope-labeled ligands in high yield are needed. While own expression and purification protocols for MC4R agonists α-MSH and ACTH were already established, a strategy for antagonist AgRP was developed in this work. Unfolded AgRP could be expressed in *E. coli* cells in high yields, and subsequently purified and refolded from IB. However, even after condition screening the refolding efficiency was not reproducible and lays between 50% to 100%. Separation of folded and un- or misfolded protein can be achieved by RP-HPLC, but could not be established for preparative scales yet. An important factor for receptor-hormone interaction studies is unspecific ligand binding to membrane surfaces, whose relevancy could be shown for α-MSH and ACTH by NMR binding studies as well as fluorescence-based assays. The peptides did not interact with membranes of neutrally charged surface, but with negatively charged lipid head groups. Divalent Ca^{2+} is known to act as a secondary messenger in MC4R signalling, but was shown here to also release membrane-bound hormones, which might serve as an additional regulatory effect at the synaptic cleft. Recently, LCN2 was postulated to activate MC4R signalling, but no response could be detected in an assay directly detecting cAMP levels. Structural integrity of the used LCN2 variants could be

confirmed and negative impacts on LCN2 by buffer or membrane systems could be excluded using solution NMR. The presented results reveal the importance of NMR spectroscopy as a complementary method in sample analysis and interaction studies.

4.3.4 Material and Methods

Used constructs of MC4R ligands. Isotope-labeled α-MSH and human or mouse LCN2 were kindly provided by Maria Dahlhaus and Claire Ortmann, respectively. NDP-ACTH23C and NDP-α-MSH were ordered from peptides&elephants, Hennigsdorf, Germany, and ACTH23C from Tocris, Wiesbaden-Nordenstadt, Germany. AgRP was expressed and purified as described below. Amino acid sequences of hormone constructs used in this study are shown in Table 4.4.

Table 4.4: Sequences of used peptides.

10xHis-AgRP$^{23-132}$	MHHHHHHHHHHSSGENLYFQGMKAPMEGIRRPDQALLPELPGLGLRAPLKKTTAEQAEEDLLQEAQALAEVLDLQDREPRSSRRCVRLHE
	SCLGQQVPCCDPCATCYCRFFNAFCYCRKLGTAMNPCSRT
10xHis-AgRP$^{83-132}$	MHHHHHHHHHHSSGENLYFQSSRRCVRLHESCLGQQVPCCDPCATCYCEFFNAFCYCRKLGTAMNPCSRT
ACTH	SYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAEAFPLEFC
ACTH^{1-23C}	SYSMEHFRWGKPVGKKRRPVKVYC
α-MSH	SYSMGHFRWGKPV
hLCN2	QDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK--CDYWIRTFVPGCQ
mLCN2	QDSTQNLIPAPSLLTVPLQPDFRSDQFRGRWYVVGLAGNAVQKKTEGSFTMYSTIYELQENNSYNVTSILVRDQDQGCRYWIRTFVPSSR
	PGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNREYFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDG
	AGQFTLGNMHRYPQVQSYNVQVATTDYNQFAMVFFRKTSENKQYFKITLYGRTKELSPELKERFTRFAKSLGLKDDNIIFSVPTDQCIDN

Protein concentration determination. Protein concentrations were measured by absorption at 280 nm using NanoDrop 2000 (Thermo Scientific). Due to the low extinction coefficient of AgRP, some concentrations were also determined using the BCA assay kit (BioRad).

Sodium dodecyl polyacrylamide gel electrophoresis. Discontinuous dodecyl polyacrylamide gel electrophoresis (SDS-PAGE) was performed at constant 30 mA using a BioRad electrophoresis system with 16% (v/v) acrylamide/bisacrylamide (19:1, Roth) for the resolving gel according to Laemmli [81]. Reducing and non-reducing loading buffer were used to distinguish between different states of AgRP, and a Spectra^{TM} Multicolor Low-Range Protein Ladder (Thermo Scientific) was used as protein standard. Gels were stained by shaking for 10 min in heated solution with 0.0025% (m/v) Comassier Blue G250, 10% (v/v) acetic acid and destained by shaking one hour in

warm water. Gel photos were taken with BioRad Doc System.

Expression and purification of AgRP. $AgRP^{23-132}$ or $AgRP^{83-132}$ (numbering includes signal sequence) was cloned into pETEV16b vector, which allows expression with a N-terminal 6xHis tag and a TEV cleavage site. *E. coli* BL21 (DE3) were transformed with the plasmids and grown at 37°C and 160 rpm in 1 l LB media, or in case of ^{15}N-enriched $AgRP^{23-132}$ in M9 minimal medium with $^{15}NH_4Cl$ as solely nitrogen source. Expression was induced with 2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) after an OD_{600} of 0.6-0.8 was reached and cultures were incubated overnight at 25°C and 160 rpm. Cells were harvested by centrifugation with 6500 g at 4°C and resuspended in lysis buffer (20 mM tris(hydroxymethyl)aminomethane (Tris)/HCl pH 7.5, 500 mM NaCl, 0.1% (v/v) Triton X-100, 10% (v/v) glycerol, 20 µg/ml DNase I (Roche), 5 mM dithiothreitol (DTT)) supplemented with a complete Protease Inhibitor pill (Roche). Cell debris and inclusion bodies (IB) were sedimented by centrifugation at 4°C with 15000 g. The pellet was resuspended once again in lysis buffer and twice in washing buffer (20 mM Tris/HCl pH 7.5, 500 mM NaCl, 10 mM DTT, 1 mM EDTA) supplemented with a complete Protease Inhibitor pill with a sonification and centrifugation step after each resuspension. Washed IB were solubilized in solubilizing buffer (10 mM Tris/HCl pH 8, 6 M guanidinium hydrochloride (GdnHCl) and up to 50 mM DTT) by shaking overnight at room temperature. Supernatant was separated by a last centrifugation step, diluted to less than 10 mM DTT and applied on a gravity-flow column with 2 ml pre-equilibrated Ni^{2+}-NTA sepharose resin (Roche). Column was washed with 20 ml solubilizing buffer (10 mM DTT) and eluted with solubilizing buffer (10 mM DTT, 500 mM imidazole) in 5 ml steps.
Concentrated $AgRP^{83-132}$ solution after immobilized metal column chromatography (IMAC) was slowly dripped into 50-fold higher volume of fast stirring refolding buffer (100 mM Tris/HCl pH 8, 2 M GdnHCl, 5% (v/v) glycerol, 1 mM reduced glutathione (GSH), 0.2 mM oxidized glutathione (GSSG)). 2 ml pre-equilibrated Ni^{2+}-NTA sepharose resin was added and the solution was slowly stirred overnight at 4°C. Mixture was applied onto a gravity-flow column and the collected resin was washed with 10 ml 100 mM Tris/HCl pH 8, 15 mM imidazole, and eluted with 100 mM Tris/HCl pH 8, 500 mM imidazole in 3 ml steps. Buffer was exchanged by two times dialysis against 100-fold volume of the desired buffer for several hours at 4°C. The expression, purification and refolding efficiency was followed by SDS-PAGE under reducing and non-reducing conditions.

Refolding assay. 5 µl droplets of 0.8 mg/ml unfolded $AgRP^{83-132}$ in solubilizing buffer (10 mM DTT) were placed into wells of 96-well plate with optical bottom (Nunc, ThermoFisher). 245 µl of different refolding buffer, water or solubilizing buffer were added on top and both solutions were rapidly mixed by pipetting up and down. The refolding

buffers consisted out of 100 mM Tris/HCl ph 8 or 9, 1 mM GSH, 0.2 mM GSSG, and different concentrations of GdnHCl (0, 0.5, 1, 2 M) and glycerol (0, 5, 10% (v/v)). After 3 h incubation at room temperature, scattering of formed aggregates was followed by absorption measurements at 370 nm with a Tecan Spark plate reader. Observed absorption values were corrected by values obtained from reference experiments with solubilizing buffer in absence of protein.

Reversed-Phase High-performance liquid chromatography. Analytical reversed phase high-performance liquid chromatography (RP-HPLC) was performed using a Zorbax 300SB-C8 column (4.6 x 250cm, Agilent) with a Knauer WellChrom HPLC system. Used injection volumes were 100 or 500 µl, and proteins were eluted with a increasing gradient of acetonitrile and water with 0.1% (v/v) trifluoroacetic acid (TFA). Elution of $AgRP^{83\text{-}132}$ was observed by 220 nm around 70% acetonitrile gradient at room temperature. Preparative RP-HPLC runs of 20 µl 1 mg/ml $AgRP^{83\text{-}132}$ were performed using (9.4 x 250cm, Agilent) with an Agilent HPLC system at 40 °C.

Nanodisc assembly of isotope-enriched α-MSH. Nanodiscs were kindly prepared by Maria Dahlhaus using established protocols [42, 130]. In short, lipid in chloroform were dried under a nitrogen flow and resuspended in lipid buffer (20 mM Tris/HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 60 mM Na-cholate) and mixed in desired ratios. Three different batches of lipid solutions were prepared: one using 100% 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC), one using 30% 1- palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS) and 70% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), and one using 50% POPS/ 50% POPC. The theoretical amount of scaffold protein MSP1D1 was calculated according to [130] and added to the respective batches. 20% (w/v) of Biobeads SM-2 (Biorad) were washed with methanol, water and 20 mM Tris/HCl, pH 7.5, 100 mM NaCl and added to the lipid/MSP1D1 mixtures. After incubation at room temperature overnight, the Biobeads were removed by centrifugation and once again 20% (w/v) of Biodbeads were added and incubated for an additional 2 h. Biobeads were again separated from the mixtures by centrifugation, and the solutions were additionally centrifuged at 16000 g for 10 min. Finally, they were purified by size-exclusion chromatography (SEC) on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with SEC buffer (20 mM sodium phosphate, pH 7.4, 50 mM NaCl) using a Äkta pure device at a flow rate of 1 ml/min. NDs were concentrated to the desired molarity using a Vivaspin centrifugal device of 10 kDa MWCO.

Solution NMR spectroscopy. Data was acquired at 600 MHz Bruker Avance III spectrometer equipped with a inverse detection HNC-triple resonance cryogenic probe, if not stated otherwise. Spectra were processed using Topsin 4.0.6 (Bruker), analysed

using Cara 1.9.1.5 [71] and plotted using Sparky 3.114 [83].

^{15}N-enriched AgRP$^{23-132}$. [^{1}H,^{15}N]-HSQC spectra of 100 µM ^{15}N-labeled AgRP$^{23-132}$ were recorded at 37 °C, using 2048x128 data points, 32 scans, 12 ppm and 36 ppm spectral width and 4.7 ppm and 117 ppm transmitter offset for ^{1}H and ^{15}N dimension, respectively. Samples were measured in phosphate-buffered saline (PBS buffer) for neutral pH and 50 mM acetate buffer for pH 5.

Human and mouse LCN2. Measurements with LCN2 variants were performed in 20 mM NaH_2PO_4/Na_2HPO_4 pH 7.4, 50 mM NaCl. Interaction between ^{15}N-hLCN2 and membrane mimetics was checked by recording [^{1}H,^{15}N]-TROSY spectra of 50 µM protein at 30 °C in presence of 25 µM nanodiscs containing 100% DMPC or 50% POPC/ 50% POPS, as well as 25 µM ^{15}N-hLCN2 and 50 mM *n*-dodecyl β-D-maltoside (DDM), with 16, 384, and 56 scans, respectively. 2048x128 data points were recorded each, with 13 ppm and 31 ppm spectral width and 4.7 ppm and 119 ppm transmitter offset for ^{1}H and ^{15}N dimension, respectively. High-resolution spectra were recorded with 200 µM ^{15}N-hLCN2 and ^{13}C,^{15}N-mLCN2 at 25 °C at 700 or 750 MHz Bruker Avance III with 16 scans and 4.7 ppm and 118.5 ppm transmitter offset for ^{1}H and ^{15}N dimension. The recorded data size was 2048 points in direct dimension, and 128 and 256 in indirect dimension for hLCN2 and mLCN2, respectively. Spectral width used for hLCN2 and mLCN2 of ^{1}H dimension was 13 and 16 ppm, and of ^{15}N dimension was 31 and 35 ppm, respectively.

Isotope-enriched α-MSH in presence of nanodiscs. Interaction studies were performed with 50 µM ^{15}N-enriched α-MSH and 25 µM nanodiscs with in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.0, 50 mM NaCl and 10% (v/v) D_2O at 10 °C. [^{1}H,^{15}N]-HSQC experiments were recorded for five samples of ^{15}N-α-MSH: in absence of nanodiscs, with nanodiscs consisting out of 100% DMPC, 50%POPC/ 50% POPS or 70% POPC/ 30% POPS, and 50%POPC/ 50% POPS with additional 10 mM $CaCl_2$ but only 30 mM NaCl. All spectra were recorded with 2014x128 points of FID, 32 scans, 13 ppm and 30 ppm spectral width and 4.7 ppm and 119.5 ppm transmitter offset for ^{1}H and ^{15}N dimension, respectively.

For carbon resonance assignment, [^{1}H,^{13}C]-HSQC, [^{1}H,^{13}C]-FLOPSY, and 2D [^{1}H,^{13}C]-HSQC-TOCSY experiments were performed at 30 °C with 200 µM ^{13}C,^{15}N-α-MSH or 50 µM ^{13}C,^{15}N-α-MSH in presence of 200 µM 50%POPC/ 50% POPS nanodiscs. ^{13}C-detecting experiments were performed at 800 MHz Bruker Avance III spectrometer equipped with a ^{13}C-optimized TXO cryogenic probe. Acquisition parameters are listed in Table 4.5.

4,4-dimethyl-4-silapentanesulfonic acid (DSS) was used as a chemical shift standard, and ^{13}C and ^{15}N data were referenced using frequency ratios as previously described [178]. The secondary chemical shift of Cα and Cβ was used to identify protein secondary structure by Equation 4.7. Reference chemical shifts were obtained from the

Table 4.5: Acquisition parameters for [^{13}C,^{15}N]-α-MSH carbon assignment experiments.

Experiment	Sample	Data points	Scans	Spectral width / ppm	Transmitter Offset / ppm
[^{13}C,^{13}C]-FLOPSY	- ND	2048x128	128	90 / 90	42 / 42
	+ ND	2048x128	288	90 / 90	42 / 42
[^{1}H,^{13}C]-HSQC	- ND	2048x256	32	13 / 165	4.7 / 75
	+ ND	2048x128	32	12 / 75	4.7 / 42
[^{1}H,^{13}C]-HSQC-TOCSY	- ND	1024x256	32	13 / 166	4.7 / 75
	+ ND	1024x128	32	13 / 75	4.7 / 42

Biological Magnetic Resonance Data Bank [146].

$$\Delta\Delta\delta_{C\alpha,C\beta} = (\delta_{C\alpha,exp} - \delta_{C\alpha,BMRB}) - (\delta_{C\beta,exp} - \delta_{C\beta,BMRB}) \tag{4.7}$$

Fluorophore labeling of the ACTH. 1 mM ACTH23C, 1 mM TCEP and a 1.5-fold excess of Atto647 in 20 mM Tris/HCl pH 7.4, 50 mM NaCl were incubated at room temperature for 1.5 h in the dark. 600 µM of the reaction volume were injected for RP-HPLC using a Zorbax 300SB-C8 column (4.6 x 250cm, Agilent) with a Knauer WellChrom HPLC system. The run was performed with a linear gradient of acetonitrile against water with 0.1% (v/v) TFA at room temperature. HPLC runs with only peptide and only fluorophore in the reaction buffer were performed to identify the elution peak of ACTH23C-Atto647. The respective peak was collected and verified by an absorption spectra.

Membrane binding assay with fluorescence-labeld ligands. Cell membrane preparations were kindly provided by Dr. Marcel Falke and membrane binding assay were performed by Claire Ortmann. Membrane protein concentrations were measured and membranes diluted to concentrations ranging from 20 g to 800 g total membrane protein. As the membrane is stored in a different buffer than the binding buffer (20 mM HEPES pH 7, 50 mM NaCl, 2 mM $CaCl_2$) the volumes were adjusted with the membrane storage buffer to ensure the same NaCl and buffer concentrations for all conditions. The reaction volume per well was 100 µl. Glass fibrefilter plates (Corning) were coated with 0.125% polyethylene for 30 minutes, before three initial washing steps were performed. All procedures were carried out at 4 °C. After loading the membrane and incubating it with ligand for one hour at 4 °C, the plates were centrifuged for 5 minutes at 1000 g. The flow-through was collected in a transparent (Thermo Fisher) or black 96 well plate (Corning) and measured at a Tecan Spark 10 M plate reader. Subsequently, the membranes on the filter plate were washed between four to six times, depending on when the fluorescence intensity of the flow-through would not change any further.

cAMP assay. Activity assays were kindly performed by Claire Ortmann. The used assay (Cisbio Bioassay) relays on FRET between a Eu^{3+} cryptate-labeled monoclonal anti-cAMP antibody and d2-labeld cAMP. cAMP-d2 competes with native cAMP produced by cells and the ratio of 665 and 620 nm emission is directly correlated to endogenous cAMP. 5000 HEK293F cells overexpressing hMC4R per well were incubated for 30 minutes with FreeStyle™ 293 Expression Medium (gibco, Life Technologies) supplemented with 0.5 mM 3-isobutyl-1-methylxanthine (IBMX) which inhibits cellular phosphodiesterases. Subsequently the cells were incubated with the ligand in question for a minimum of 30 minutes. Finally, cAMP-d2 (solved in lysis-detection buffer) and anti-cAMP Eu^{3+} cryptate (solved in lysis-detection buffer) were added to the cell-ligand mixture and incubated for one hour before fluorescence measurements were performed.

5 General Conclusions

Solution NMR for complex biological systems often requires expensive and elaborated isotope-labeling strategies. For instance, $^{13}CH_3$ are often introduced into large protein complexes by isotope-enriched metabolic precursors to only record methyl signals in HMQC experiments, which greatly simplifies the spectra. Additionally, protein deuteration greatly reduces the linewidth of recorded spectra due to the reduced spin interaction. However, ^{13}C-labelled precursors are expensive and not universal applicable for several expression systems, and eukaryotic expression cultures are often not viable at high D_2O concentrations. In section 4.2 a cost-efficient synthesis and labelling strategy with specifically labelled leucine (Leu_{LD}) is described. Leu_{LD} has one $^{13}C^1H_3$ group and a vincial $^{12}C^2H_3$ and was successfully incorporated into the proteins MBP, $RBM39^{245\text{-}332}$, and bR using *E. coli*, Sf9 insect cell, and cell-free expression systems, respectively. The ”local” deuteration of Leu_{LD} significantly reduces dipole-dipole relaxation of the $^{13}C^1H_3$ methyl magnetisation, which could be shown for MBP by FTC experiments. Also [^{1}H,^{13}C]-HMQC spectra of Leu_{MD}-labeled bO and $RBM39^{245\text{-}332}$ showed decreased linewidths compared to variants with uniform ^{13}C-labeled leucine and no deuteration. This new method of methyl-labeling is compatible with most of the expression systems and is most likely transferable to other amino acids like Valine and Isoleucine. Despite several elaborated labeling strategies are available for proteins, isotope-enrichment in nucleic acids, and especially DNA, is very expensive and most laboratories cannot perform it themselves. In section 3.2.3 it could be shown, that ^{19}F substitutions are a cheap and easily available option to introduce NMR-active probes site-specifically in nucleic acids. It could be confirmed, that fluorine substitution has minimal effect on the structure and function of 10-23 DNAzyme and that ^{19}F-derived signals of selectively saturated STD experiments serve as suitable distance restraints for structure determination.

NMR is a widely used method to analyse interaction between proteins and binding partners, since isotope-labeling is barely affecting structural and interactive features of biopolymers and binding effects can be residue-specifically localized by observing CSP and peak intensities of amide signals in [^{1}H,^{15}N]-correlation spectra. In section 4.1, the interplay between the juxtamembrane segment of EGFR and possible modulators were studied using NMR titration experiments, in which spectra of labeled protein were recorded in presence of potential binding partners in different concentrations. It could be shown that the N-terminal half of JM is interacting with the Sec7 domain of ARNO, with Calmodulin, as well as with nanodiscs containing anionic phospholipids, which may suggest that this interplay plays a role for *in-vivo* EGFR signaling. The findings could be also supported with microscale thermophoresis experiments and mutagenesis studies, which is showing that NMR is highly complementary to other biophysical

and biochemical techniques. Additionally, structural features of protein states can be easily characterized by the chemical shifts of Cα and Cβ, as it has been demonstrated that JM has a higher α-helical propensity when bound to Sec7 or lipid membranes. The results point towards an additional regulation mechanism of EGFR, which should be the focus of further *in vivo* studies.

Membrane interaction was also investigated for (potential) MC4R activating hormones α-MSH and LCN2 by NMR (section 4.3), which is providing explanations for previous findings regarding receptor activation. For instance, receptor activation with α-MSH and ACTH is significantly altered by divalent ion concentrations of the buffer (data not shown). NMR studies and fluorescence-based assays revealed that the peptides strongly interact with negative charged lipid bilayers and can be released by addition of Ca^{2+}, which is altering the amount of available ligands in an activity assay. This effect should be taken into account for future studies on MC4R-hormone interaction. On the other hand, LCN2 did not trigger detectable MC4R response and no membrane interaction was observed. It could be assured that LCN2 is indeed structured at the assay conditions visible by the dispersive pattern of amide and methyl signals in NMR, which leaves little explanation why the activity assay contradicts previous findings. Chemical shift dispersion was also used to characterize MC4R's inverse agonist AgRP after refolding, and allowed to detect different states within 10-23 DNAzyme's catalytic cycle, as described in the following paragraph.

The main scope of this thesis lays on the structural and functional description of the RNA-cleaving 10-23 DNAzyme (Dz). Although mainly unlabeled samples were used in this research, ^{1}H resonance assignment of Dz:RNA complex could be performed due to the differently shielded protons at ribose position 2'. Only crosspeaks of deoxyribose H2' and H2" (and thymine methyls) appear at a low chemical shift range below 3 ppm, which allowed to assign Dz separately and simplified subsequent assignment of its RNA target. Following this strategy, nearly all analysed states have been completely sequentially assigned. TOCSY crosspeaks of pyrimidine base protons turned out to be excellent probes for characterising Dz states, since the are highly dispersed, separated from other signals, and the content of nucleotides U, C and T is widely distributed over the Dz's and RNA's sequences. Analysis of pyrimidine signal patterns of free and complexed Dz under various conditions allowed to identify the different states of Dz's catalytic cycle, which are (i) single stranded Dz, (ii) hybridized Dz:RNA complex with slowly interchanging conformations, which (iii) can be homogenized by addition of moderate amounts of Na^{+} or Mg^{2+}, (iv) a presumable pre-catalytic conformation in the presence of high Mg^{2+} concentration, and (v) cleavage products in a still complexed state.

Cation binding is stabilizing complex conformation, presumably because it neutralizes opposing negative charges of phosphate backbones in close proximity. Interestingly, addition of NaCl induces similar effects as $MgCl_2$, suggesting an unspecific electro-

static interaction of ionic cofactors, which is in agreement with kinetic measurements that are showing a decreasing Mg^{2+}-induced cleavage activity at higher concentrations of NaCl. Three different mutants of Dz were analysed with A, C, or G (Dz^{5A}, Dz^{5C}, and Dz^{5G}) substitutions at loop position 5, which are severely differing in their activity. While Dz^{5G} showed no activity in the performed assays, Dz^{5C}'s cleavage rate is up to 5-fold slower than Dz^{5A}. To analyse Dz:RNA complexes without cleavage during acquisition, the RNA was stabilized by a 2'-^{19}F substitution at nucleotide next to the scissile bond. While for stabilized Dz^{5A}:$RNA^{2'F}$ increasing Mg^{2+} lead to severe line broadening for signals belonging to the catalytic loop, only CSP without significant intensity decrease could be observed for the binding arms and in the loop region of Dz^{5C}:$RNA^{2'F}$, which is indicating a fast exchange of bound ions. This result agrees with high K_D values (mM range) observed for Mg^{2+} binding, and that already a 20-fold lower amount of paramagnetic Mn^{2+} could completely reduce some NMR signals of Dz^{5C}:$RNA^{2'F}$ even in the presence of 1 mM $MgCl_2$.

The Dz^{5C}:$RNA^{2'F}$ complex was further structurally analysed using NOE contacts, as well as distance restraints derived from six additional 2'-^{19}F substitutions and three paramagnetic TEMPO spinlabels at the Dz oligomer. ^{19}F-derived experiments (especially STD) allowed to find unique distance relationships in otherwise too crowded spectra, while the PRE effect of TEMPO enabled to find restraints for far longer distances up to 35 Å. The usability of spinlabels might be enhanced by using sterically fixed spinlabel variants, e.g. paramagnetic nucleobases. Electron paramagnetic resonance experiments may also give additional and unique distance and orientation data between two spinlabels or a spinlabel and an introduced *19*F nucleus. Since conventional determination of J coupling constants failed due to the large size of the system, a homology approach was used in which Dz^{5C}:$RNA^{2'F}$ resonance shifts were compared with shifts derived from an A-form helical hybrid of $RNA^{2'F}$ and a complementary DNA strand. Additionally, NOE build-up curves were recorded, which were giving insights into the nucleobase conformation and could be used to derive exact NOEs by correcting measured cross-relaxation rates with calculated spin diffusion effects. The experimentally collected restraints were used for structure determination yielding two distinct models for Dz^{5C}:$RNA^{2'F}$, of which only one is in full agreement with performed and published mechanistic data. The most critical feature of this structure is, that the catalytic loop is winding around the RNA strand at height of the cleavage site, while the Watson-Crick (WC) base pair at position +1 of the binding arm is not formed in order to let the opposing Dz strand pass by. No classical WC base pairing could be found within the loop region, but T1 relaxation and HetNOE data suggest no high flexibility in this part of the Dz sequence. Loop mobility is probably hindered by the through-passing RNA strand and additionally in some means stabilized at its 5' site, which could be detected using NMR melting experiments. However, the origin of this stabilization remains unclear. Based on the calculated model, also Mg^{2+} binding sites could be determined using

titration experiments with Mg^{2+}-mimicking Mn^{2+} and $Co(NH_2)_6$, where after binding the former leads to specific paramagnetic line broadening and the latter to NOE contacts with close-by nucleic acid protons. Three potential binding sites could be determined, of which only one at the 3' site of the loop seems to bind a divalent ion which is crucial for RNA hydrolysis. The rate of Dz^{5C}-mediated RNA cleavage allowed to cover the reaction by time-resolved NMR experiments, which unveiled a slow loop rearrangement after cleavage leading to two differently stabilized product states. The unique insights into the pre-catalytic Dz complex are reshaping the 20 years old model and might serve as a ground work for knowledge-based Dz-derived therapeutics.

In this work, several novel and established NMR techniques were applied to different biological systems, such as the EGFR and MC4R signaling pathways, as well as the RNA-cleaving 10-23 DNAzyme. The presented result gives not only new insights into mechanisms of the studied biomolecules, but also demonstrated the application and significance of new techniques, which may push the frontiers of biomolecular solution NMR.

6 Publication List

- Jan Borggräfe, Julian Victor, Hannah Rosenbach, Aldino Viegas, Christoph Gertzen, Christine Wübben, Ralf Biel, Helena Kovacs, Detlev Riesner, Gerhard Steger, Olav Schiemann, Holger Gohlke, Ingrid Span, Manuel Etzkorn: High-resolution and time-resolved insights into 10-23 DNAzyme mediated catalysis. *in preparation*

- Hannah Rosenbach, Julian Victor, Jan Borggräfe, Christine Wübben, Olav Schiemann, Wolfgang Hoyer, Gerhard Steger, Manuel Etzkorn, Ingrid Span: Influence of ionic strength on metal binding and catalytic activity of the 10-23 DNAzyme. *submitted*

- Hannah Rosenbach, Julian Victor, Jan Borggräfe, Ralf Biehl, Gerhard Steger, Manuel Etzkorn, Ingrid Span: Expanding crystallization tools for nucleic acid complexes using U1A protein variants. *Journal of Structural Biology* **2020**; 210(2):107480. doi:10.1016/j.jsb.2020.107480

- Aldino Viegas, Dongsheng M. Yin, Jan Borggräfe, Thibault Viennet, Marcel Falke, Anton Schmitz, Michael Famulok, Manuel Etzkorn: Molecular Architecture of a Network of Potential Intracellular EGFR Modulators: ARNO, CaM, Phospholipids, and the Juxtamembrane Segment. *Structure* **2019**; 28(1):54-62.e5. doi:10.1016/j.str.2019.11.001

- Abhinav Dubey, Nikolay Stoyanov, Sandeep Chhabra, Thibault Viennet, Shantha Elter, Jan Borggräfe, Aldino Viegas, Radosław P. Nowak, Nikola Burdziev, Ognyan Petrov, Eric S. Fischer, Manuel Etzkorn, Vladimir Gelev, Haribabu Arthanari: Local deuteration enables NMR observation of methyl groups in proteins from eukaryotic and cell-free expression systems. *submitted*

- Sabine Schriek, Aldino Viegas, Jan Borggräfe, Marcel Falke, Ci Chu, Thibault Viennet, Maria Dahlhaus, Laetitia Heid, Steffen Mielenbrink, James Yu, Gerhard Wagner, Manuel Etzkorn: Membranes and Calcium as possible modulators in endocrine ACTH and neuropeptide a-MSH signaling. *in preparation*

- Falke M, Schloesser R, Tripailo A, Borggräfe J, Ortmann C, Etzkorn M: Lipocalin 2 does not activate MC4R *in situ*. *in preparation*

- Jakub Kubiak, Neha Verma, Peter Dollinger, Florian Bleffert, Aldino Viegas, Jan Borggräfe, Manuel Etzkorn, Filip Kovacic, Karl-Erich Jaeger, Holger Gohlke, Claus A.M. Seidel: Functional dynamics of a structurally minimalistic chaperone. *in preparation*

Bibliography

[1] K. B. Abd Halim, H. Koldsø, and M. S. Sansom. Interactions of the EGFR juxtamembrane domain with PIP2-containing lipid bilayers: Insights from multiscale molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):1017–1025, may 2015.

[2] S. Aifa, K. Johansen, U. K. Nilsson, B. Liedberg, I. Lundström, and S. P. Svensson. Interactions between the juxtamembrane domain of the EGFR and calmodulin measured by surface plasmon resonance. *Cellular Signalling*, 14(12):1005–1013, dec 2002.

[3] J. Aranda, M. Terrazas, H. Gómez, N. Villegas, and M. Orozco. An artificial DNAzyme RNA ligase shows a reaction mechanism resembling that of cellular polymerases. *Nature Catalysis*, 2(6):544–552, jun 2019.

[4] L. W. Arbogast, F. Delaglio, J. E. Schiel, and J. P. Marino. Multivariate Analysis of Two-Dimensional 1 H, 13 C Methyl NMR Spectra of Monoclonal Antibody Therapeutics To Facilitate Assessment of Higher Order Structure. *Analytical Chemistry*, 89(21):11839–11845, nov 2017.

[5] A. Arkhipov, Y. Shan, R. Das, N. F. Endres, M. P. Eastwood, D. E. Wemmer, J. Kuriyan, and D. E. Shaw. Architecture and Membrane Interactions of the EGF Receptor. *Cell*, 152(3):557–569, jan 2013.

[6] S. Arnott, R. Chandrasekaran, R. Millane, and H.-S. Park. RNA-RNA, DNA-DNA, and DNA-RNA Polymorphism. *Biophysical Journal*, 49(1):3–5, jan 1986.

[7] R. A. August, J. A. Khan, C. M. Moody, and D. W. Young. Stereospecific synthesis of (2S,4R)-[5,5,5-2H3]-leucine. *Tetrahedron Letters*, 33(32):4617–4620, aug 1992.

[8] F. G. Avci, B. S. Akbulut, and E. Ozkirimli. Membrane Active Peptides and Their Biophysical Characterization. *Biomolecules*, 8(3):77, aug 2018.

[9] J. Balbach, C. Steegborn, T. Schindler, and F. X. Schmid. A protein folding intermediate of ribonuclease T1 characterized at high resolution by 1D and 2D real-time NMR spectroscopy. *Journal of molecular biology*, 285(2):829–42, jan 1999.

[10] S. Basu, B. Sriram, R. Goila, and A. C. Banerjea. Targeted cleavage of HIV-1 coreceptor-CXCR-4 by RNA-cleaving DNA-enzyme: inhibition of coreceptor function. *Antiviral Research*, 46(2):125–134, may 2000.

[11] D. A. Baum and S. K. Silverman. Deoxyribozyme-Catalyzed Labeling of RNA. *Angewandte Chemie International Edition*, 46(19):3502–3504, may 2007.

[12] A. Bax, A. Mehlkopf, and J. Smidt. Homonuclear broadband-decoupled absorption spectra, with linewidths which are independent of the transverse relaxation rate. *Journal of Magnetic Resonance (1969)*, 35(1):167–169, jul 1979.

[13] T. H. Bayburt, J. W. Carlson, and S. G. Sligar. Reconstitution and Imaging of a Membrane Protein in a Nanometer-Size Phospholipid Bilayer. *Journal of Structural Biology*, 123(1):37–44, sep 1998.

[14] M. V. Berjanskii and D. S. Wishart. Unraveling the meaning of chemical shifts in protein NMR. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1865(11):1564–1576, nov 2017.

[15] S. F. Betz, A. Schnuchel, H. Wang, E. T. Olejniczak, R. P. Meadows, B. P. Lipsky, E. A. S. Harris, D. E. Staunton, and S. W. Fesik. Solution structure of the cytohesin-1 (B2-1) Sec7 domain and its interaction with the GTPase ADP ribosylation factor 1. *Proceedings of the National Academy of Sciences*, 95(14):7909–7914, jul 1998.

[16] A. Bill, A. Schmitz, K. König, L. C. Heukamp, J. S. Hannam, and M. Famulok. Anti-Proliferative Effect of Cytohesin Inhibition in Gefitinib-Resistant Lung Cancer Cells. *PLoS ONE*, 7(7):e41179, jul 2012.

[17] R. Boelens, T. Koning, and R. Kaptein. Determination of biomolecular structures from proton-proton NOE's using a relaxation matrix approach. *Journal of Molecular Structure*, 173:299–311, jan 1988.

[18] M. P. Bokoch, Y. Zou, S. G. F. Rasmussen, C. W. Liu, R. Nygaard, D. M. Rosenbaum, J. J. Fung, H.-J. Choi, F. S. Thian, T. S. Kobilka, J. D. Puglisi, W. I. Weis, L. Pardo, R. S. Prosser, L. Mueller, and B. K. Kobilka. Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature*, 463(7277):108–112, jan 2010.

[19] R. R. BREAKER. A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA*, 9(8):949–957, aug 2003.

[20] R. R. Breaker and G. F. Joyce. A DNA enzyme that cleaves RNA. *Chemistry & Biology*, 1(4):223–229, dec 1994.

[21] R. R. Burgess. Chapter 17 Refolding Solubilized Inclusion Body Proteins. pages 259–282. 2009.

[22] L. Büttner, F. Javadi-Zarnaghi, and C. Höbartner. Site-Specific Labeling of RNA at Internal Ribose Hydroxyl Groups: Terbium-Assisted Deoxyribozymes at Work. *Journal of the American Chemical Society*, 136(22):8131–8137, jun 2014.

[23] A.-I. Cabedo Martinez, K. Weinhäupl, W.-K. Lee, N. A. Wolff, B. Storch, S. Żerko, R. Konrat, W. Koźmiński, K. Breuker, F. Thévenod, and N. Coudevylle. Biochemical and Structural Characterization of the Interaction between the Siderocalin NGAL/LCN2 (Neutrophil Gelatinase-associated Lipocalin/Lipocalin 2) and the N-terminal Domain of Its Endocytic Receptor SLC22A17. *Journal of Biological Chemistry*, 291(6):2917–2930, feb 2016.

[24] G. Cai, L. Zhu, X. Chen, K. Sun, C. Liu, G. C. Sen, G. R. Stark, J. Qin, and X. Li. TRAF4 binds to the juxtamembrane region of EGFR directly and promotes kinase activation. *Proceedings of the National Academy of Sciences*, 115(45):11531–11536, nov 2018.

[25] M. J. Cairns. Optimisation of the 10-23 DNAzyme-substrate pairing interactions enhanced RNA cleavage activity at purine-cytosine target sites. *Nucleic Acids Research*, 31(11):2883–2889, jun 2003.

[26] R. Cardillo, C. Fuganti, D. Ghiringhelli, P. Grasselli, and G. Gatti. Pattern of incorporation of leucine samples asymmetrically labelled with 13C in the isopropyl unit into the C5-isoprenoid units of echinuline and flavoglaucine. *Journal of the Chemical Society, Chemical Communications*, (13):474, 1977.

[27] B.-X. Chai, R. R. Neubig, G. L. Millhauser, D. A. Thompson, P. J. Jackson, G. S. Barsh, C. J. Dickinson, J.-Y. Li, Y.-M. Lai, and I. Gantz. Inverse agonist activity of agouti and agouti-related protein. *Peptides*, 24(4):603–609, apr 2003.

[28] M. Chandra, A. Sachdeva, and S. K. Silverman. DNA-catalyzed sequence-specific hydrolysis of DNA. *Nature Chemical Biology*, 5(10):718–720, oct 2009.

[29] J. Chandrasekar and S. K. Silverman. Catalytic DNA with phosphatase activity. *Proceedings of the National Academy of Sciences*, 110(14):5315–5320, apr 2013.

[30] K. Chen and N. Tjandra. The use of residual dipolar coupling in studying proteins by NMR. *Topics in current chemistry*, 326:47–67, 2012.

[31] J. Cherfils, J. Ménétrey, M. Mathieu, G. Le Bras, S. Robineau, S. Béraud-Dufour, B. Antonny, and P. Chardin. Structure of the Sec7 domain of the Arf exchange factor ARNO. *Nature*, 392(6671):101–105, mar 1998.

[32] D. J.-F. Chinnapen and D. Sen. A deoxyribozyme that harnesses light to repair thymine dimers in DNA. *Proceedings of the National Academy of Sciences*, 101(1):65–69, jan 2004.

[33] E.-A. Cho, F. J. Moloney, H. Cai, A. Au-Yeung, C. China, R. A. Scolyer, B. Yosufi, M. J. Raftery, J. Z. Deng, S. W. Morton, P. T. Hammond, H.-T. Arkenau, D. L. Damian, D. J. Francis, C. N. Chesterman, R. S. C. Barnetson, G. M. Halliday, and L. M. Khachigian. Safety and tolerability of an intratumorally injected DNAzyme, Dz13, in patients with nodular basal-cell carcinoma: a phase 1 first-in-human trial (DISCOVER). *The Lancet*, 381(9880):1835–1843, may 2013.

[34] K. Choowongkomon, C. R. Carlin, and F. D. Sönnichsen. A Structural Model for the Membrane-bound Form of the Juxtamembrane Domain of the Epidermal Growth Factor Receptor. *Journal of Biological Chemistry*, 280(25):24043–24052, jun 2005.

[35] C.-C. Chu, O. Y. Wong, and S. K. Silverman. A Generalizable DNA-Catalyzed Approach to Peptide-Nucleic Acid Conjugation. *ChemBioChem*, 15(13):1905–1910, sep 2014.

[36] A. Citri and Y. Yarden. EGF–ERBB signalling: towards the systems level. *Nature Reviews Molecular Cell Biology*, 7(7):505–516, jul 2006.

[37] C. Cochet, G. N. Gill, J. Meisenhelder, J. A. Cooper, and T. Hunter. C-kinase phosphorylates the epidermal growth factor receptor and reduces its epidermal growth factor-stimulated tyrosine protein kinase activity. *The Journal of biological chemistry*, 259(4):2553–8, feb 1984.

[38] L. A. Cohen, A. Honda, P. Varnai, F. D. Brown, T. Balla, and J. G. Donaldson. Active Arf6 Recruits ARNO/Cytohesin GEFs to the PM by Binding Their PH Domains. *Molecular Biology of the Cell*, 18(6):2244–2253, jun 2007.

[39] B. Cuenoud and J. W. Szostak. A DNA metalloenzyme with DNA ligase activity. *Nature*, 375(6532):611–614, jun 1995.

[40] R. J. Davis and M. P. Czech. Tumor-promoting phorbol diesters cause the phosphorylation of epidermal growth factor receptors in normal human fibroblasts at threonine-654. *Proceedings of the National Academy of Sciences*, 82(7):1974–1978, apr 1985.

[41] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3), nov 1995.

[42] I. G. Denisov, Y. V. Grinkova, A. A. Lazarides, and S. G. Sligar. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *Journal of the American Chemical Society*, 126(11):3477–3487, mar 2004.

[43] E. A. Dethoff, K. Petzold, J. Chugh, A. Casiano-Negroni, and H. M. Al-Hashimi. Visualizing transient low-populated structures of RNA. *Nature*, 491(7426):724–728, nov 2012.

[44] J. P. DiNitto, A. Delprato, M.-T. Gabe Lee, T. C. Cronin, S. Huang, A. Guilherme, M. P. Czech, and D. G. Lambright. Structural Basis and Mechanism of Autoregulation in 3-Phosphoinositide-Dependent Grp1 Family Arf GTPase Exchange Factors. *Molecular Cell*, 28(4):569–583, nov 2007.

[45] A. Doerner, R. Scheck, and A. Schepartz. Growth Factor Identity Is Encoded by Discrete Coiled-Coil Rotamers in the EGFR Juxtamembrane Region. *Chemistry & Biology*, 22(6):776–784, jun 2015.

[46] M. T. Eddy, T. Didenko, R. C. Stevens, and K. Wüthrich. β 2 -Adrenergic Receptor Conformational Response to Fusion Protein in the Third Intracellular Loop. *Structure*, 24(12):2190–2197, dec 2016.

[47] N. F. Endres, R. Das, A. W. Smith, A. Arkhipov, E. Kovacs, Y. Huang, J. G. Pelton, Y. Shan, D. E. Shaw, D. E. Wemmer, J. T. Groves, and J. Kuriyan. Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. *Cell*, 152(3):543–556, jan 2013.

[48] M. Etzkorn, T. Raschle, F. Hagn, V. Gelev, A. J. Rice, T. Walz, and G. Wagner. Cell-free Expressed Bacteriorhodopsin in Different Soluble Membrane Mimetics: Biophysical Properties and NMR Accessibility. *Structure*, 21(3):394–401, mar 2013.

[49] M. D. Fletcher, J. R. Harding, R. A. Hughes, N. M. Kelly, H. Schmalz, A. Sutherland, and C. L. Willis. Three approaches to the synthesis of L-leucine selectively labelled with carbon-13 or deuterium in either diastereotopic methyl group. *Journal of the Chemical Society, Perkin Transactions 1*, (1):43–51, 2000.

[50] A. Flynn-Charlebois, Y. Wang, T. K. Prior, I. Rashid, K. A. Hoadley, R. L. Coppins, A. C. Wolf, and S. K. Silverman. Deoxyribozymes with 2'-5' RNA Ligase Activity. *Journal of the American Chemical Society*, 125(9):2444–2454, mar 2003.

[51] E. Freisinger and R. K. Sigel. From nucleotides to ribozymes—A comparison of their metal ion binding properties. *Coordination Chemistry Reviews*, 251(13-14):1834–1851, jul 2007.

[52] B. Friedman, A. R. Frackelton, A. H. Ross, J. M. Connors, H. Fujiki, T. Sugimura, and M. R. Rosner. Tumor promoters block tyrosine-specific phosphorylation of the epidermal growth factor receptor. *Proceedings of the National Academy of Sciences*, 81(10):3034–3038, may 1984.

[53] P. Gans, O. Hamelin, R. Sounier, I. Ayala, M. A. Durá, C. D. Amero, M. Noirclerc-Savoye, B. Franzetti, M. J. Plevin, and J. Boisbouvier. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angewandte Chemie International Edition*, 49(11):1958–1962, mar 2010.

[54] C. M. Garret, Reginald H., Grisham. Biochemistry. *Wadsworth Publishing Co Inc*, 2012.

[55] R. Goila and A. C. Banerjea. Sequence specific cleavage of the HIV-1 coreceptor CCR5 gene by a hammerhead ribozyme and a DNA-enzyme: inhibition of the coreceptor function by DNA-enzyme. *FEBS Letters*, 436(2):233–238, oct 1998.

[56] L. He and K. Hristova. Consequences of replacing EGFR juxtamembrane domain with an unstructured sequence. *Scientific Reports*, 2(1):854, dec 2012.

[57] G. Hedger, M. S. P. Sansom, and H. Koldsø. The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. *Scientific Reports*, 5(1):9198, aug 2015.

[58] E. Henrich, F. Löhr, J. Mezhyrova, A. Laguerre, F. Bernhard, and V. Dötsch. Synthetic Biology-Based Solution NMR Studies on Membrane Proteins in Lipid Environments. pages 143–185. 2019.

[59] V. J. Hruby, S. D. Sharma, K. Toth, J. Y. Jaw, F. Al-Obeidi, T. K. Sawyer, and M. E. Hadley. Design, Synthesis, and Conformation of Superpotent and Prolonged Acting Melanotropins. *Annals of the New York Academy of Sciences*, 680(1 The Melanotro):51–63, may 1993.

[60] K. Huang. Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. *The EMBO Journal*, 19(11):2615–2628, jun 2000.

[61] J. M. Humphrey, J. A. Hart, and A. Richard Chamberlin. Efficient syntheses of diastereoselectively labeled (3S)-[4-^{13}C]-l-valine, and regioselectively labeled [3-$^{13}CH_3$]-l-isoleucine hydrochlorides. *Bioorganic & Medicinal Chemistry Letters*, 5(12):1315–1320, jun 1995.

[62] L. Hunsicker. Activities and relative affinities of divalent metals in unmodified and phosphorothioate-substituted hammerhead ribozymes. *Journal of Inorganic Biochemistry*, 80(3-4):271–281, jul 2000.

[63] T. Hunter, N. Ling, and J. A. Cooper. Protein kinase C phosphorylation of the EGF receptor at a threonine residue close to the cytoplasmic face of the plasma membrane. *Nature*, 311(5985):480–483, oct 1984.

[64] D. Huszar, C. A. Lynch, V. Fairchild-Huntress, J. H. Dunmore, Q. Fang, L. R. Berkemeier, W. Gu, R. A. Kesterson, B. A. Boston, R. D. Cone, F. J. Smith, L. Campfield, P. Burn, and F. Lee. Targeted Disruption of the Melanocortin-4 Receptor Results in Obesity in Mice. *Cell*, 88(1):131–141, jan 1997.

[65] P. J. Jackson, N. R. Douglas, B. Chai, J. Binkley, A. Sidow, G. S. Barsh, and G. L. Millhauser. Structural and Molecular Evolutionary Analysis of Agouti and Agouti-Related Proteins. *Chemistry & Biology*, 13(12):1297–1305, dec 2006.

[66] M. R. Jakobsen, J. Haasnoot, J. Wengel, B. Berkhout, and J. Kjems. Efficient inhibition of HIV-1 expression by LNA modified antisense oligonucleotides and DNAzymes targeted to functionally selected binding sites. *Retrovirology*, 4(1):29, 2007.

[67] N. Jura, N. F. Endres, K. Engel, S. Deindl, R. Das, M. H. Lamers, D. E. Wemmer, X. Zhang, and J. Kuriyan. Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. *Cell*, 137(7):1293–1307, jun 2009.

[68] M. Kainosho and P. Güntert. SAIL – stereo-array isotope labeling. *Quarterly Reviews of Biophysics*, 42(4):247–300, nov 2009.

[69] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono, and P. Güntert. Optimal isotope labelling for NMR protein structure determinations. *Nature*, 440(7080):52–57, mar 2006.

[70] L. E. Kay. NMR studies of protein structure and dynamics. *Journal of Magnetic Resonance*, 213(2):477–491, dec 2011.

[71] R. Keller. The computer aided resonance assignment tutorial. In Institute for Molecular Biology and Boiphysics (Zurich: The Swiss Federal Institute of Technology), ISBM 3-85600-112-3. 2004.

[72] N. M. Kelly, R. G. Reid, C. L. Willis, and P. L. Winton. Methods for the Synthesis of L-Leucine Selectively Labelled with Carbon-13 or Deuterium in either Diastereotopic Methyl Group. *Tetrahedron Letters*, 36(45):8315–8318, nov 1995.

[73] H. Khandelia, J. H. Ipsen, and O. G. Mouritsen. The impact of peptides on lipid membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1778(7-8):1528–1536, jul 2008.

[74] M. Kitamura, S. Shirakawa, and K. Maruoka. Powerful Chiral Phase-Transfer Catalysts for the Asymmetric Synthesis of ?-Alkyl- and ?,?-Dialkyl-?-amino Acids. *Angewandte Chemie International Edition*, 44(10):1549–1551, feb 2005.

[75] J. L. Kitevski-LeBlanc, T. Yuwen, P. N. Dyer, J. Rudolph, K. Luger, and L. E. Kay. Investigating the Dynamics of Destabilized Nucleosomes Using Methyl-TROSY NMR. *Journal of the American Chemical Society*, 140(14):4774–4777, apr 2018.

[76] Y. Kofuku, T. Ueda, J. Okude, Y. Shiraishi, K. Kondo, T. Mizumura, S. Suzuki, and I. Shimada. Functional Dynamics of Deuterated β 2 -Adrenergic Receptor in Lipid Bilayers Revealed by NMR Spectroscopy. *Angewandte Chemie International Edition*, 53(49):13376–13379, dec 2014.

[77] Y. Kofuku, T. Yokomizo, S. Imai, Y. Shiraishi, M. Natsume, H. Itoh, M. Inoue, K. Nakata, S. Igarashi, H. Yamaguchi, T. Mizukoshi, E.-i. Suzuki, T. Ueda, and I. Shimada. Deuteration and selective labeling of alanine methyl groups of β2-adrenergic receptor expressed in a baculovirus-insect cell expression system. *Journal of Biomolecular NMR*, 71(3):185–192, jul 2018.

[78] E. Kovacs, J. A. Zorn, Y. Huang, T. Barros, and J. Kuriyan. A structural perspective on the regulation of the epidermal growth factor receptor. *Annual review of biochemistry*, 84:739–64, 2015.

[79] N. Krug, J. M. Hohlfeld, A.-M. Kirsten, O. Kornmann, K. M. Beeh, D. Kappeler, S. Korn, S. Ignatenko, W. Timmer, C. Rogon, J. Zeitvogel, N. Zhang, J. Bille, U. Homburg, A. Turowska, C. Bachert, T. Werfel, R. Buhl, J. Renz, H. Garn, and H. Renz. Allergen-Induced Asthmatic Responses Modified by a GATA3-Specific DNAzyme. *New England Journal of Medicine*, 372(21):1987–1995, may 2015.

[80] V. Kurauskas, E. Crublet, P. Macek, R. Kerfah, D. F. Gauto, J. Boisbouvier, and P. Schanda. Sensitive proton-detected solid-state NMR spectroscopy of large proteins with selective CH_3 labelling: application to the 50S ribosome subunit. *Chemical Communications*, 52(61):9558–9561, 2016.

[81] U. K. LAEMMLI. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259):680–685, aug 1970.

[82] M. Lazarova, F. Löhr, R.-B. Rues, R. Kleebach, V. Dötsch, and F. Bernhard. Precursor-Based Selective Methyl Labeling of Cell-Free Synthesized Proteins. *ACS Chemical Biology*, 13(8):2170–2178, aug 2018.

[83] W. Lee, M. Tonelli, and J. L. Markley. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, 31(8):1325–1327, apr 2015.

[84] M. A. Lemmon, J. Schlessinger, and K. M. Ferguson. The EGFR Family: Not So Prototypical Receptor Tyrosine Kinases. *Cold Spring Harbor Perspectives in Biology*, 6(4):a020768–a020768, apr 2014.

[85] H. Li, S. Panina, A. Kaur, M. J. Ruano, P. Sánchez-González, J. M. la Cour, A. Stephan, U. H. Olesen, M. W. Berchtold, and A. Villalobo. Regulation of the Ligand-dependent Activation of the Epidermal Growth Factor Receptor by Calmodulin. *Journal of Biological Chemistry*, 287(5):3273–3281, jan 2012.

[86] X. Li, Y. Huang, J. Jiang, and S. J. Frank. ERK-dependent threonine phosphorylation of EGF receptor modulates receptor downregulation and signaling. *Cellular signalling*, 20(11):2145–55, nov 2008.

[87] R. Linser, V. Gelev, F. Hagn, H. Arthanari, S. G. Hyberts, and G. Wagner. Selective Methyl Labeling of Eukaryotic Membrane Proteins Using Cell-Free Expression. *Journal of the American Chemical Society*, 136(32):11308–11310, aug 2014.

[88] H. Liu, X. Yu, Y. Chen, J. Zhang, B. Wu, L. Zheng, P. Haruehanroengra, R. Wang, S. Li, J. Lin, J. Li, J. Sheng, Z. Huang, J. Ma, and J. Gan. Crystal structure of an RNA-cleaving DNAzyme. *Nature Communications*, 8(1):2006, dec 2017.

[89] L. A. Lotta, J. Mokrosiński, E. Mendes de Oliveira, C. Li, S. J. Sharp, J. Luan, B. Brouwers, V. Ayinampudi, N. Bowker, N. Kerrison, V. Kaimakis, D. Hoult, I. D. Stewart, E. Wheeler, F. R. Day, J. R. Perry, C. Langenberg, N. J. Wareham, and I. S. Farooqi. Human Gain-of-Function MC4R Variants Show Signaling Bias and Protect against Obesity. *Cell*, 177(3):597–607.e9, apr 2019.

[90] K. Lu, Y. Miyazaki, and M. F. Summers. Isotope labeling strategies for NMR studies of RNA. *Journal of Biomolecular NMR*, 46(1):113–125, jan 2010.

[91] L. Ma and J. Liu. Catalytic Nucleic Acids: Biochemistry, Chemical Biology, Biosensors, and Nanotechnology. *iScience*, 23(1):100815, jan 2020.

[92] P. Macek, R. Kerfah, E. B. Erba, E. Crublet, C. Moriscot, G. Schoehn, C. Amero, and J. Boisbouvier. Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Science Advances*, 3(4):e1601601, apr 2017.

[93] R. Maeda, T. Sato, K. Okamoto, M. Yanagawa, and Y. Sako. Lipid-Protein Interplay in Dimerization of Juxtamembrane Domains of Epidermal Growth Factor Receptor. *Biophysical Journal*, 114(4):893–903, feb 2018.

[94] J. Martín-Nieto and A. Villalobo. The Human Epidermal Growth Factor Receptor Contains a Juxtamembrane Calmodulin-Binding Site †. *Biochemistry*, 37(1):227–236, jan 1998.

[95] M. Marušič, J. Schlagnitweit, and K. Petzold. RNA Dynamics by NMR Spectroscopy. *ChemBioChem*, 20(21):2685–2710, nov 2019.

[96] S. McLaughlin, S. O. Smith, M. J. Hayman, and D. Murray. An Electrostatic Engine Model for Autoinhibition and Activation of the Epidermal Growth Factor Receptor (EGFR/ErbB) Family. *Journal of General Physiology*, 126(1):41–53, jul 2005.

[97] J. C. McNulty, D. A. Thompson, K. A. Bolin, J. Wilken, G. S. Barsh, and G. L. Millhauser. High-Resolution NMR Structure of the Chemically-Synthesized Melanocortin Receptor Binding Domain AGRP(87-132) of the Agouti-Related Protein † , ‡. *Biochemistry*, 40(51):15520–15527, dec 2001.

[98] I. E. Michailidis, R. Rusinova, A. Georgakopoulos, Y. Chen, R. Iyengar, N. K. Robakis, D. E. Logothetis, and L. Baki. Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation. *Pflügers Archiv - European Journal of Physiology*, 461(3):387–397, mar 2011.

[99] G. L. MILLHAUSER, J. C. McNULTY, P. J. JACKSON, D. A. THOMPSON, G. S. BARSH, and I. GANTZ. Loops and Links: Structural Insights into the Remarkable Function of the Agouti-Related Protein. *Annals of the New York Academy of Sciences*, 994(1):27–35, jun 2003.

[100] Y. Minato, S. Suzuki, T. Hara, Y. Kofuku, G. Kasuya, Y. Fujiwara, S. Igarashi, E.-i. Suzuki, O. Nureki, M. Hattori, T. Ueda, and I. Shimada. Conductance of P2X 4 purinergic receptor is determined by conformational equilibrium in the transmembrane region. *Proceedings of the National Academy of Sciences*, 113(17):4741–4746, apr 2016.

[101] M. J. Minch. Orientational dependence of vicinal proton-proton NMR coupling constants: The Karplus relationship. *Concepts in Magnetic Resonance*, 6(1):41–56, jan 1994.

[102] K. S. Mineev, S. V. Panova, O. V. Bocharova, E. V. Bocharov, and A. S. Arseniev. The Membrane Mimetic Affects the Spatial Structure and Mobility of EGFR Transmembrane and Juxtamembrane Domains. *Biochemistry*, 54(41):6295–6298, oct 2015.

[103] Y. Miyanoiri, M. Takeda, K. Okuma, A. M. Ono, T. Terauchi, and M. Kainosho. Differential isotope-labeling for Leu and Val residues in a protein by E. coli cellular expression using stereo-specifically methyl labeled amino acids. *Journal of Biomolecular NMR*, 57(3):237–249, nov 2013.

[104] M. Miyazawa-Onami, K. Takeuchi, T. Takano, T. Sugiki, I. Shimada, and H. Takahashi. Perdeuteration and methyl-selective 1H, 13C-labeling by using a Kluyveromyces lactis expression system. *Journal of Biomolecular NMR*, 57(3):297–304, nov 2013.

[105] Y. Mizukoshi, K. Takeuchi, M. Arutaki, Y. Tokunaga, T. Takizawa, H. Hanzawa, and I. Shimada. Improvement of Ligand Affinity and Thermodynamic Properties by NMR-Based Evaluation of Local Dynamics and Surface Complementarity in the Receptor-Bound State. *Angewandte Chemie International Edition*, 55(47):14606–14609, nov 2016.

[106] P. Morrison, A. R. Saltiel, and M. R. Rosner. Role of Mitogen-activated Protein Kinase Kinase in Regulation of the Epidermal Growth Factor Receptor by Protein Kinase C. *Journal of Biological Chemistry*, 271(22):12891–12896, may 1996.

[107] I. Mosialou, S. Shikhel, J.-M. Liu, A. Maurizi, N. Luo, Z. He, Y. Huang, H. Zong, R. A. Friedman, J. Barasch, P. Lanzano, L. Deng, R. L. Leibel, M. Rubin, T. Nickolas, W. Chung, L. M. Zeltser, K. W. Williams, J. E. Pessin, and S. Kousteni. MC4R-dependent suppression of appetite by bone-derived lipocalin 2. *Nature*, 543(7645):385–390, mar 2017.

[108] E. Mossessova, J. M. Gulbis, and J. Goldberg. Structure of the Guanine Nucleotide Exchange Factor Sec7 Domain of Human Arno and Analysis of the Interaction with ARF GTPase. *Cell*, 92(3):415–423, feb 1998.

[109] B. Nawrot, K. Widera, M. Wojcik, B. Rebowska, G. Nowak, and W. J. Stec. Mapping of the functional phosphate groups in the catalytic core of deoxyribozyme 10-23. *The FEBS Journal*, 274(4):1062–1072, feb 2007.

[110] F. H. Nelissen, M. Tessari, S. S. Wijmenga, and H. A. Heus. Stable isotope labeling methods for DNA. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 96:89–108, aug 2016.

[111] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundstrom, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay. Structure of an Intermediate State in Protein Folding and Aggregation. *Science*, 336(6079):362–366, apr 2012.

[112] P. Nichols, A. Born, M. Henen, D. Strotz, J. Orts, S. Olsson, P. Güntert, C. Chi, and B. Vögeli. The Exact Nuclear Overhauser Enhancement: Recent Advances. *Molecules*, 22(7):1176, jul 2017.

[113] J. Nowakowski, P. J. Shim, G. S. Prasad, C. D. Stout, and G. F. Joyce. Crystal structure of an 82-nucleotide RNA-DNA complex formed by the 10-23 DNA enzyme. *Nature structural biology*, 6(2):151–6, feb 1999.

[114] R. Nygaard, Y. Zou, R. O. Dror, T. J. Mildorf, D. H. Arlow, A. Manglik, A. C. Pan, C. W. Liu, J. J. Fung, M. P. Bokoch, F. S. Thian, T. S. Kobilka, D. E. Shaw, L. Mueller, R. S. Prosser, and B. K. Kobilka. The Dynamic Process of β2-Adrenergic Receptor Activation. *Cell*, 152(3):532–542, jan 2013.

[115] M. Oba, K. Ohkuma, H. Hitokawa, A. Shirai, and K. Nishiyama. Convenient synthesis of deuterated glutamic acid, proline and leucine via catalytic deuteration of unsaturated pyroglutamate derivatives. *Journal of Labelled Compounds and Radiopharmaceuticals*, 49(3):229–235, mar 2006.

[116] M. Oba, T. Terauchi, A. Miyakawa, H. Kamo, and K. Nishiyama. Stereoselective deuterium-labelling of diastereotopic methyl and methylene protons of L-leucine. *Tetrahedron Letters*, 39(12):1595–1598, mar 1998.

[117] J. Okude, T. Ueda, Y. Kofuku, M. Sato, N. Nobuyama, K. Kondo, Y. Shiraishi, T. Mizumura, K. Onishi, M. Natsume, M. Maeda, H. Tsujishita, T. Kuranaga, M. Inoue, and I. Shimada. Identification of a Conformational Equilibrium That Determines the Efficacy and Functional Selectivity of the μ-Opioid Receptor. *Angewandte Chemie International Edition*, 54(52):15771–15776, dec 2015.

[118] M. M. Ollmann, B. D. Wilson, Y.-K. Yang, J. A. Kerns, Y. Chen, I. Gantz, and G. S. Barsh. Antagonism of Central Melanocortin Receptors *in Vitro* and *in Vivo* by Agouti-Related Protein. *Science*, 278(5335):135–138, oct 1997.

[119] J. Orts, B. Vögeli, and R. Riek. Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs. *Journal of Chemical Theory and Computation*, 8(10):3483–3492, oct 2012.

[120] T. Pan, J. Sun, J. Hu, Y. Hu, J. Zhou, Z. Chen, D. Xu, W. Xu, S. Zheng, and S. Zhang. Cytohesins/ARNO: The Function in Colorectal Cancer Cells. *PLoS ONE*, 9(3):e90997, mar 2014.

[121] T. PAN, J. SUN, J. ZHOU, Z. FU, Y. HU, S. ZHENG, and S. ZHANG. Function and mode of action of cytohesins in the epidermal growth factor pathway in colorectal cancer cells. *Oncology Letters*, 5(2):521–526, feb 2013.

[122] R. G. Pearson. Hard and soft acids and bases, HSAB, part 1: Fundamental principles. *Journal of Chemical Education*, 45(9):581, sep 1968.

[123] H. M. Poppleton, G. J. Wiepz, P. J. Bertics, and T. B. Patel. Modulation of the Protein Tyrosine Kinase Activity and Autophosphorylation of the Epidermal Growth Factor Receptor by Its Juxtamembrane Region. *Archives of Biochemistry and Biophysics*, 363(2):227–236, mar 1999.

[124] P. I. Pradeepkumar, C. Höbartner, D. A. Baum, and S. K. Silverman. DNA-Catalyzed Formation of Nucleopeptide Linkages. *Angewandte Chemie International Edition*, 47(9):1753–1757, feb 2008.

[125] L. Pritchard, A. Turnbull, and A. White. Pro-opiomelanocortin processing in the hypothalamus: impact on melanocortin signalling and obesity. *Journal of Endocrinology*, 172(3):411–421, mar 2002.

[126] I. Pritišanac, M. T. Degiacomi, T. R. Alderson, M. G. Carneiro, E. AB, G. Siegal, and A. J. Baldwin. Automatic Assignment of Methyl-NMR Spectra of Supramolecular Machines Using Graph Theory. *Journal of the American Chemical Society*, 139(28):9523–9533, jul 2017.

[127] I. Pritišanac, J. M. Würz, T. R. Alderson, and P. Güntert. Automatic structure-based NMR methyl resonance assignment in large proteins. *Nature Communications*, 10(1):4922, dec 2019.

[128] M. Red Brewer, S. H. Choi, D. Alvarado, K. Moravcevic, A. Pozzi, M. A. Lemmon, and G. Carpenter. The Juxtamembrane Region of the EGF Receptor Functions as an Activation Domain. *Molecular Cell*, 34(6):641–651, jun 2009.

[129] T. L. Religa, R. Sprangers, and L. E. Kay. Dynamic Regulation of Archaeal Proteasome Gate Opening As Studied by TROSY NMR. *Science*, 328(5974):98–102, apr 2010.

[130] T. Ritchie, Y. Grinkova, T. Bayburt, I. Denisov, J. Zolnerciks, W. Atkins, and S. Sligar. Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. pages 211–231. 2009.

[131] S. Robson, H. Arthanari, S. G. Hyberts, and G. Wagner. Nonuniform Sampling for NMR Spectroscopy. pages 263–291. 2019.

[132] E. J. Romainczyk O, Endeward B, Prisner TF. The RNA-DNA hybrid structure determined by EPR, CD and RNase H1. *Mol Biosyst.*, 7(4):1050–1052, 2011.

[133] H. Rosenbach, J. Victor, J. Borggräfe, B. Ralf, G. Steger, E. Manuel, and I. Span. Expanding crystallization tools for nucleic acid complexes using U1A protein variants. *Journal of Structural Biology*, 210(2):107480, 2019.

[134] H. Rosenbach, J. Victor, M. Etzkorn, G. Steger, D. Riesner, and I. Span. Molecular Features and Metal Ions That Influence 10-23 DNAzyme Activity. *Molecules*, 25(13):3100, jul 2020.

[135] R. D. Rosenfeld, L. Zeni, A. A. Welcher, L. O. Narhi, C. Hale, J. Marasco, J. Delaney, T. Gleason, J. S. Philo, V. Katta, J. Hui, J. Baumgartner, M. Graham, K. L. Stark, and W. Karbon. Biochemical, Biophysical, and Pharmacological Characterization of Bacterially Expressed Human Agouti-Related Protein. *Biochemistry*, 37(46):16041–16052, nov 1998.

[136] J. Rouhana, F. Hoh, S. Estaran, C. Henriquet, Y. Boublik, A. Kerkour, R. Trouillard, J. Martinez, M. Pugnière, A. Padilla, and A. Chavanieu. Fragment-Based Identification of a Locus in the Sec7 Domain of Arno for the Design of Protein–Protein Interaction Inhibitors. *Journal of Medicinal Chemistry*, 56(21):8497–8511, nov 2013.

[137] M. Roychowdhury-Saha and D. H. Burke. Distinct reaction pathway promoted by non-divalent-metal cations in a tertiary stabilized hammerhead ribozyme. *RNA*, 13(6):841–848, apr 2007.

[138] P. Sánchez-González, K. Jellali, and A. Villalobo. Calmodulin-mediated regulation of the epidermal growth factor receptor. *FEBS Journal*, 277(2):327–342, jan 2010.

[139] S. W. Santoro and G. F. Joyce. A general purpose RNA-cleaving DNA enzyme. *Proceedings of the National Academy of Sciences*, 94(9):4262–4266, apr 1997.

[140] S. W. Santoro and G. F. Joyce. Mechanism and Utility of an RNA-Cleaving DNA Enzyme †. *Biochemistry*, 37(38):13330–13342, sep 1998.

[141] T. Sato, P. Pallavi, U. Golebiewska, S. McLaughlin, and S. O. Smith. Structure of the Membrane Reconstituted Transmembrane-Juxtamembrane Peptide EGFR(622-660) and Its Interaction with Ca 2+ /Calmodulin †. *Biochemistry*, 45(42):12704–12714, oct 2006.

[142] M. Sattler and S. W. Fesik. Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure*, 4(11):1245–1249, nov 1996.

[143] T. K. Sawyer, P. J. Sanfilippo, V. J. Hruby, M. H. Engel, C. B. Heward, J. B. Burnett, and M. E. Hadley. 4-Norleucine, 7-D-phenylalanine-alpha-melanocyte-stimulating hormone: a highly potent alpha-melanotropin with ultralong biological activity. *Proceedings of the National Academy of Sciences*, 77(10):5754–5758, oct 1980.

[144] F. H. Schumann, H. Riepl, T. Maurer, W. Gronwald, K.-P. Neidig, and H. R. Kalbitzer. Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions. *Journal of Biomolecular NMR*, 39(4):275–289, nov 2007.

[145] A. K. Schütz, E. Rennella, and L. E. Kay. Exploiting conformational plasticity in the AAA+ protein VCP/p97 to modify function. *Proceedings of the National Academy of Sciences*, 114(33):E6822–E6829, aug 2017.

[146] S. Schwarzinger, G. J. Kroon, T. R. Foss, P. E. Wright, and H. J. Dyson. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *Journal of biomolecular NMR*, 18(1):43–8, sep 2000.

[147] G. M. Schwieters, Charles and Kuszewski, John and Clore. Using Xplor–NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 48:47–62, 2006.

[148] G. M. Schwieters, Charles D; Kuszewski, John J; Tjandra, Nico; Clore. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, 160(1):65–73, 2003.

[149] P. Sengupta, E. Bosis, E. Nachliel, M. Gutman, S. O. Smith, G. Mihályné, I. Zaitseva, and S. McLaughlin. EGFR Juxtamembrane Domain, Membranes, and Calmodulin: Kinetics of Their Interaction. *Biophysical Journal*, 96(12):4887–4895, jun 2009.

[150] J. C. Shattuck and J. Meinwald. The preparation of L-(2*S*,*3*S)-4,4,4-[$^2H_3$] valine. *Tetrahedron Letters*, 38(49):8461–8464, dec 1997.

[151] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4):213–223, aug 2009.

[152] Z. Shen, Z. Wu, D. Chang, W. Zhang, K. Tram, C. Lee, P. Kim, B. J. Salena, and Y. Li. A Catalytic DNA Activated by a Specific Strain of Bacterial Pathogen. *Angewandte Chemie International Edition*, 55(7):2431–2434, feb 2016.

[153] A. H. G. Siebum, W. S. Woo, and J. Lugtenburg. Preparation and Characterization of [5-13C]-(2S,4R)-Leucine and [4-13C]-(2S,3S)-Valine- Establishing Synthetic Schemes to Prepare Any Site-Directed Isotopomer ofL-Leucine,L-Isoleucine andL-Valine. *European Journal of Organic Chemistry*, 2003(23):4664–4678, dec 2003.

[154] S. K. Silverman. Pursuing DNA Catalysts for Protein Modification. *Accounts of Chemical Research*, 48(5):1369–1379, may 2015.

[155] A. Sitarska, L. Skora, J. Klopp, S. Roest, C. Fernández, B. Shrestha, and A. D. Gossert. Affordable uniform isotope labeling with 2H, 13C and 15N in insect cells. *Journal of Biomolecular NMR*, 62(2):191–197, jun 2015.

[156] I. Solomon. Relaxation Processes in a System of Two Spins. *Physical Review*, 99(2):559–565, jul 1955.

[157] R. Sounier, C. Mas, J. Steyaert, T. Laeremans, A. Manglik, W. Huang, B. K. Kobilka, H. Déméné, and S. Granier. Propagation of conformational changes during μ-opioid receptor activation. *Nature*, 524(7565):375–378, aug 2015.

[158] D. Stalder, H. Barelli, R. Gautier, E. Macia, C. L. Jackson, and B. Antonny. Kinetic Studies of the Arf Activator Arno on Model Membranes in the Presence of Arf Effectors Suggest Control by a Positive Feedback Loop. *Journal of Biological Chemistry*, 286(5):3873–3883, feb 2011.

[159] G. Steger. Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. *Nucleic Acids Research*, 22(14):2760–2768, 1994.

[160] N. Sugimoto, S.-i. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, and M. Sasaki. Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry*, 34(35):11211–11216, sep 1995.

[161] N. Sugimoto, Y. Okumoto, and T. Ohmichi. Effect of metal ions and sequence of deoxyribozymes on their RNA cleavage activity. *Journal of the Chemical Society, Perkin Transactions 2*, (7):1381–1386, 1999.

[162] Y.-X. Tao. The Melanocortin-4 Receptor: Physiology, Pharmacology, and Pathophysiology. *Endocrine Reviews*, 31(4):506–543, aug 2010.

[163] F. Tebar, P. Villalonga, T. Sorkina, N. Agell, A. Sorkin, and C. Enrich. Calmodulin Regulates Intracellular Trafficking of Epidermal Growth Factor Receptor and the MAPK Signaling Pathway. *Molecular Biology of the Cell*, 13(6):2057–2068, jun 2002.

[164] K. W. Thiel and G. Carpenter. Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. *Proceedings of the National Academy of Sciences*, 104(49):19238–19243, dec 2007.

[165] S.-F. Torabi, P. Wu, C. E. McGhee, L. Chen, K. Hwang, N. Zheng, J. Cheng, and Y. Lu. In vitro selection of a sodium-specific DNAzyme and its application in intracellular sensing. *Proceedings of the National Academy of Sciences*, 112(19):5903–5908, may 2015.

[166] V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, and L. E. Kay. Cross-Correlated Relaxation Enhanced 1 H-13 C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *Journal of the American Chemical Society*, 125(34):10420–10428, aug 2003.

[167] V. Tugarinov, V. Kanelis, and L. E. Kay. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature Protocols*, 1(2):749–754, aug 2006.

[168] V. Tugarinov and L. E. Kay. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *Journal of the American Chemical Society*, 125(45):13868–13878, nov 2003.

[169] H. Valafar and J. H. Prestegard. REDCAT: a residual dipolar coupling analysis tool. *Journal of Magnetic Resonance*, 167(2):228–241, apr 2004.

[170] J. Victor, G. Steger, and D. Riesner. Inability of DNAzymes to cleave RNA *in vivo* is due to limited Mg^{2+} concentration in cells. *European Biophysics Journal*, 47(4):333–343, may 2018.

[171] A. Viegas, J. Manso, F. L. Nobrega, and E. J. Cabrita. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *Journal of Chemical Education*, 88(7):990–994, jul 2011.

[172] A. Viegas, T. Viennet, and M. Etzkorn. The power, pitfalls and potential of the nanodisc system for NMR-based studies. *Biological Chemistry*, 397(12):1335–1354, dec 2016.

[173] T. Viennet, M. M. Wördehoff, B. Uluca, C. Poojari, H. Shaykhalishahi, D. Willbold, B. Strodel, H. Heise, A. K. Buell, W. Hoyer, and M. Etzkorn. Structural insights from lipid-bilayer nanodiscs link α-Synuclein membrane-binding modes to amyloid fibril formation. *Communications Biology*, 1(1):44, dec 2018.

[174] B. Wang, L. Cao, W. Chiuman, Y. Li, and Z. Xi. Probing the Function of Nucleotides in the Catalytic Cores of the 8-17 and 10-23 DNAzymes by Abasic Nucleotide and C3 Spacer Substitutions. *Biochemistry*, 49(35):7553–7562, sep 2010.

[175] Y. Wang, J. Gao, X. Guo, T. Tong, X. Shi, L. Li, M. Qi, Y. Wang, M. Cai, J. Jiang, C. Xu, H. Ji, and H. Wang. Regulation of EGFR nanocluster formation by ionic protein-lipid interaction. *Cell Research*, 24(8):959–976, aug 2014.

[176] S. S. Wijmenga and B. N. van Buuren. The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 32(4):287–387, jun 1998.

[177] A. M. Wilczynski, C. G. Joseph, and C. Haskell-Luevano. Current trends in the structure—activity relationship studies of the endogenous agouti-related protein (AGRP) melanocortin receptor antagonist. *Medicinal Research Reviews*, 25(5):545–556, sep 2005.

[178] D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley, and B. D. Sykes. ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR*, 6(2):135–140, sep 1995.

[179] D. S. Wishart, B. D. Sykes, and F. M. Richards. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, 31(6):1647–1651, feb 1992.

[180] O. Y. Wong, P. I. Pradeepkumar, and S. K. Silverman. DNA-Catalyzed Covalent Modification of Amino Acid Side Chains in Tethered and Free Peptide Substrates. *Biochemistry*, 50(21):4741–4749, may 2011.

[181] P. Wu, K. Hwang, T. Lan, and Y. Lu. A DNAzyme-Gold Nanoparticle Probe for Uranyl Ion in Living Cells. *Journal of the American Chemical Society*, 135(14):5254–5257, apr 2013.

[182] Y. Xiao, V. Pavlov, T. Niazov, A. Dishon, M. Kotler, and I. Willner. Catalytic Beacons for the Detection of DNA and Telomerase Activity. *Journal of the American Chemical Society*, 126(24):7430–7431, jun 2004.

[183] T. Yamazaki, W. Lee, C. H. Arrowsmith, D. R. Muhandiram, and L. E. Kay. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ^{15}N, ^{13}C, ^{2}H Labeled Proteins with High Sensitivity. *Journal of the American Chemical Society*, 116(26):11655–11666, dec 1994.

[184] L. Ye, N. Van Eps, M. Zimmer, O. P. Ernst, and R. Scott Prosser. Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature*, 533(7602):265–268, may 2016.

[185] J. Yu, L. E. Gimenez, C. C. Hernandez, Y. Wu, A. H. Wein, G. W. Han, K. McClary, S. R. Mittal, K. Burdsall, B. Stauch, L. Wu, S. N. Stevens, A. Peisley, S. Y. Williams, V. Chen, G. L. Millhauser, S. Zhao, R. D. Cone, and R. C. Stevens. Determination of the melanocortin-4 receptor structure identifies Ca2+ as a cofactor for ligand binding. *Science (New York, N.Y.)*, 368(6489):428–433, 2020.

[186] Z. Zaborowska, J. P. Fürste, V. A. Erdmann, and J. Kurreck. Sequence Requirements in the Catalytic Core of the “10-23” DNA Enzyme. *Journal of Biological Chemistry*, 277(43):40617–40622, oct 2002.

[187] X. Zhang, J. Gureasko, K. Shen, P. A. Cole, and J. Kuriyan. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell*, 125(6):1137–1149, jun 2006.

[188] X. Zhang, Y. Xu, H. Ling, and T. Hattori. Inhibition of infection of incoming HIV-1 virus by RNA-cleaving DNA enzyme. *FEBS Letters*, 458(2):151–156, sep 1999.

Appendix

Supplementary Information for section 3

The 10-23 DNAzyme - A RNA-cleaving DNA enzyme

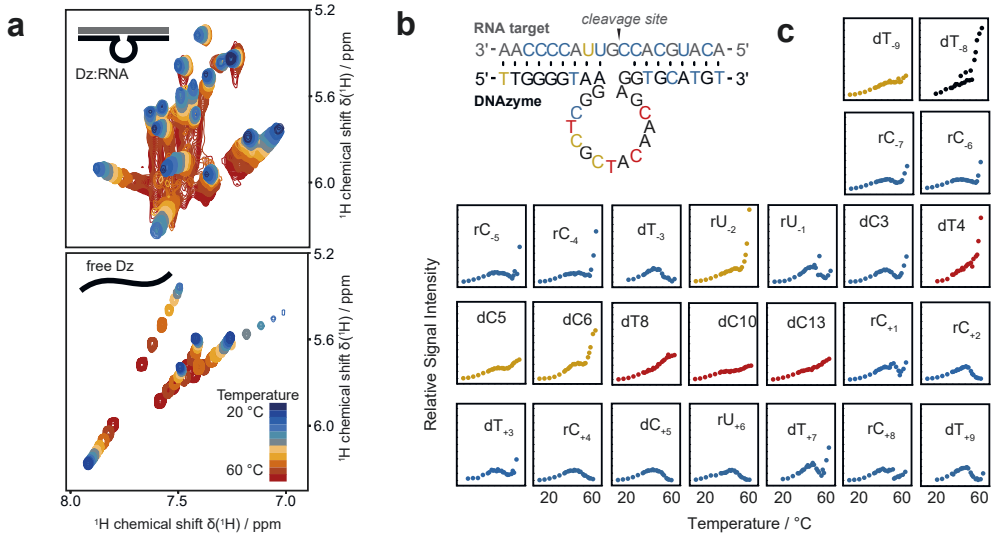


Figure A1: NMR melting experiments of Dz^{5C}:$RNA^{2'F}$ and Dz^{5C}. (a) [^{1}H,^{1}H]-TOCSY spectra of Dz^{5C}:$RNA^{2'F}$ (top) and Dz^{5C} (bottom) recorded at temperatures between 65 and 15 °C. (b) Scheme of Dz^{5C}:$RNA^{2'F}$ with stabilized nucleotides highlighted in blue, partially stabilized in yellow, and unstabilized in red. (c) Pyrimidine crosspeak intensities of Dz^{5C}:$RNA^{2'F}$ (a, top) are plotted against temperature derived from two performed melting experiments. Stabilized nucleotides (e.g. by WC base pairing) show a characteristic intensity dip around 58 °C (blue), while unstabilized nucleotides show a steady increase (red). Partially stabilized nucleotides show a plateau instead of a dip (orange). Data of dT_{-8} is inconsistent in both performed experiments (black).

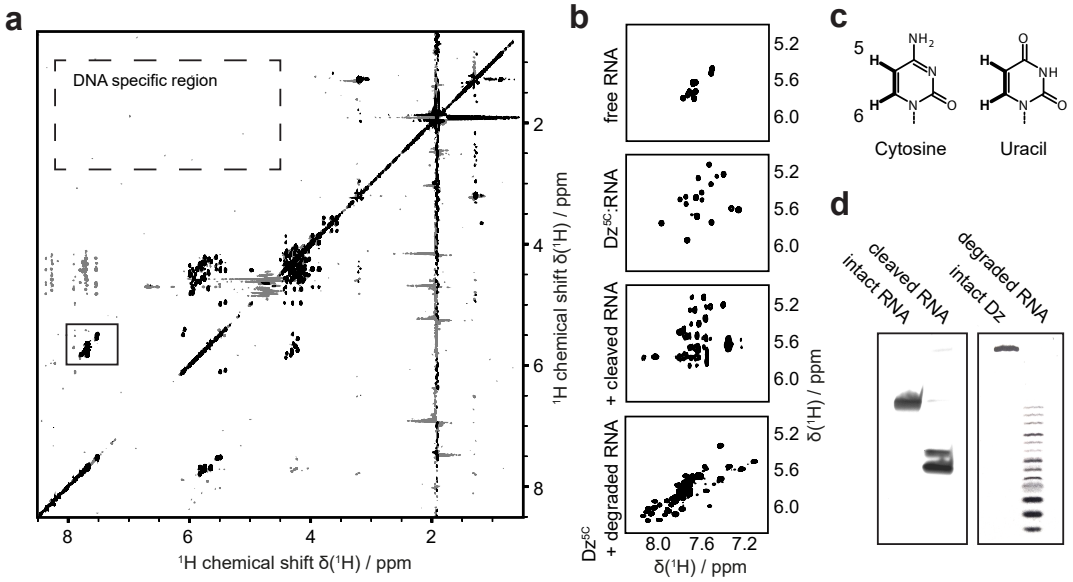


Figure A2: Pyrimidine TOCSY region allows to follow RNA's integrity. (a) [^{1}H,^{1}H]-NOESY (grey) and TOCSY (black) spectra of free intact RNA. The dotted box indicates the spectra region, where only DNA specific signals appear, and the straight box indicates the pyrimidine fingerprint region. (b) Enlarged fingerprint region of free intact RNA (top), stabilized Dz^{5C}:RNA$^{2'F}$ (middle upper), cleaved RNA in complex with Dz^{5C} (middle lower), and Dz^{5C} with degraded RNA (bottom). (c) Lewis structures of cytosine and uracil with indicated protons H6 and H5. The TOCSY transfer occurs via the bond indicated in bold and gives arise to the signals in the fingerprint region. (d) SDS-PAGE of samples with intact RNA, Dz and cleaved RNA, Dz, and degraded RNA (from left to right).

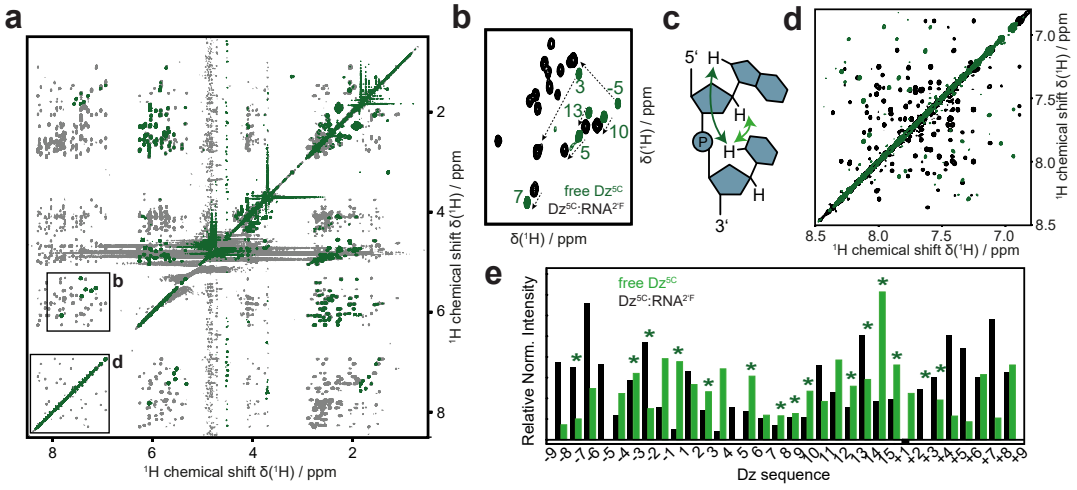


Figure A3: Single-stranded Dz^{5C} adopts a similar structure for loop positions 5-13 as Dz^{5C}:RNA$^{2'F}$. (a) [^{1}H,^{1}H]-NOESY (grey) and TOCSY (green) spectra of single-stranded Dz^{5C}. (b) Enlarged TOCSY section of single-stranded Dz^{5C} (green, boxed in a) and Dz^{5C}:RNA$^{2'F}$ (black). The Dz pyrimidine signals are indicated, and cytosines 5, 7, 10, and 13 show similar CSP in both spectra. (c) Scheme of stacking contacts between base proton H6/H8 and H1' (light green) or H6/H8 (dark green) of the 5' adjacent nucleotide. (d) Enlarged NOESY section of single-stranded Dz^{5C} (green, boxed in a) and Dz^{5C}:RNA$^{2'F}$ (black) shows less base proton stacking contacts (dark green arrow in c) for single-stranded Dz^{5C} than complex. (d) Relative intensities normalized to dC$_{+5}$ pyrimidine crosspeak of stacking contacts between H6/H8 and H1' (light green arrow in b) of single-stranded Dz^{5C} (green) and Dz^{5C}:RNA$^{2'F}$ (black) show high similarity for nucleotides 6-13 and low similarity for the binding arms. Green asterices represent existing base proton stacking contacts in single-stranded Dz^{5C} (dark green arrow in c, green spectrum in d), which are more abundant in the loop region.

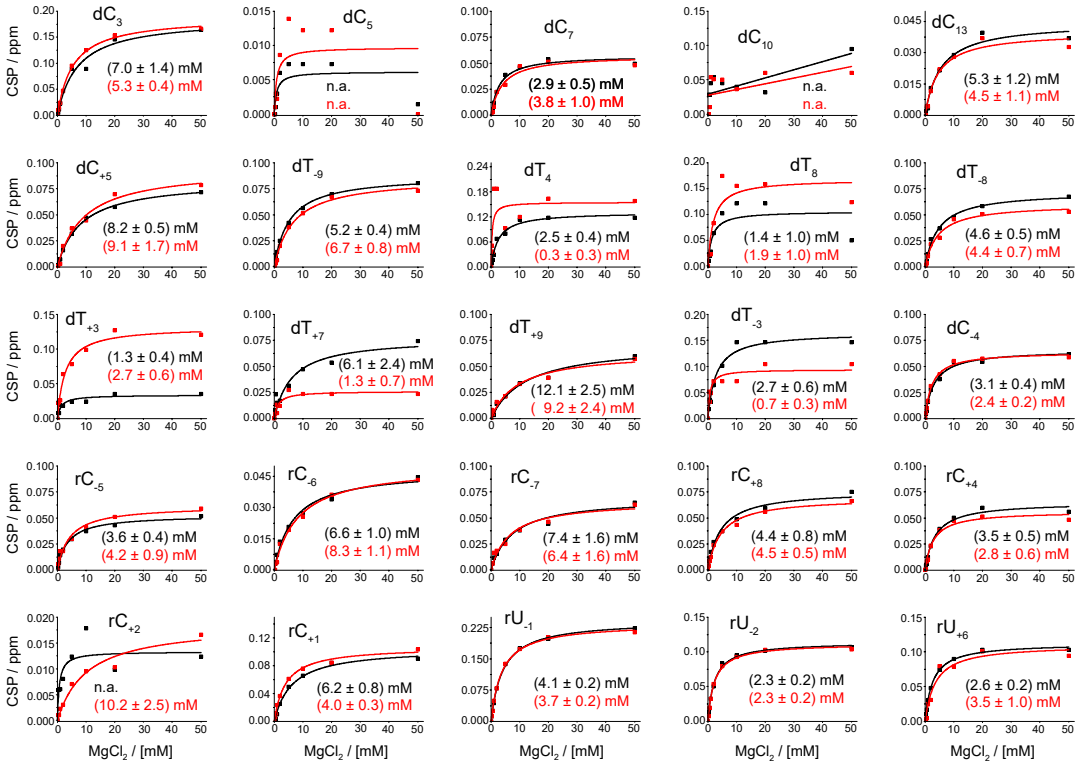


Figure A4: Mg^{2+} binding coefficients derived from pyrimidine signals. Mg^{2+}-depending CSP signal of pyrimidine crosspeaks (Figure 3.6 c) upper (black) and lower (red) the diagonal with respective binding fits and K_D values.

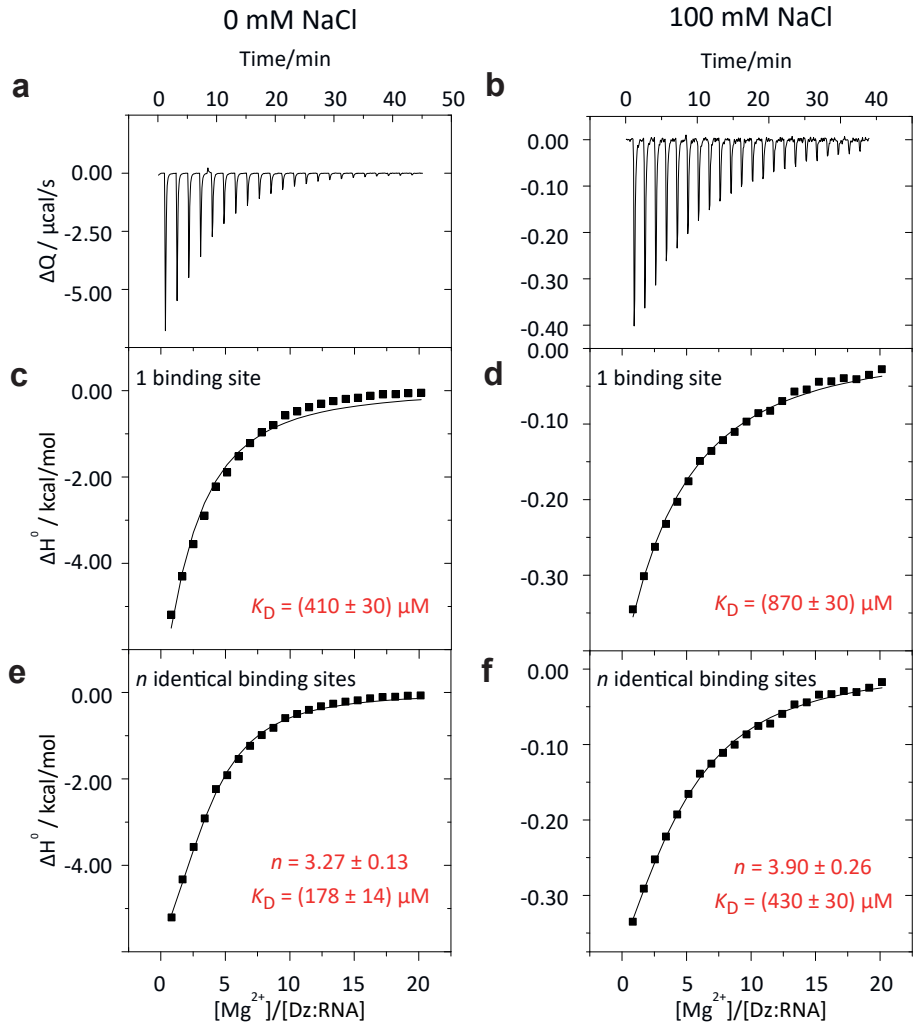


Figure A5: Interaction between Mg^{2+} and the Dz^{5A}:$RNA^{2'F}$ complex at low and high ionic strength determined by ITC measurements. (a, b) Raw ITC data for the titration of Mg^{2+} and Dz^{5A}:$RNA^{2'F}$ (100 µM) at 30 °C in absence (a) and presence of 100 mM NaCl (b) after subtraction of the integration baseline. (c-f) Titration plots derived from the integrated raw data shown in (a) and (b). The solid lines represent the best fit to the data according to a single (c,d) and a multiple non-interacting (e, f) Mg^{2+} binding site model, respectively. Both analysis show a K_D lowered approximately by half in absence of NaCl.

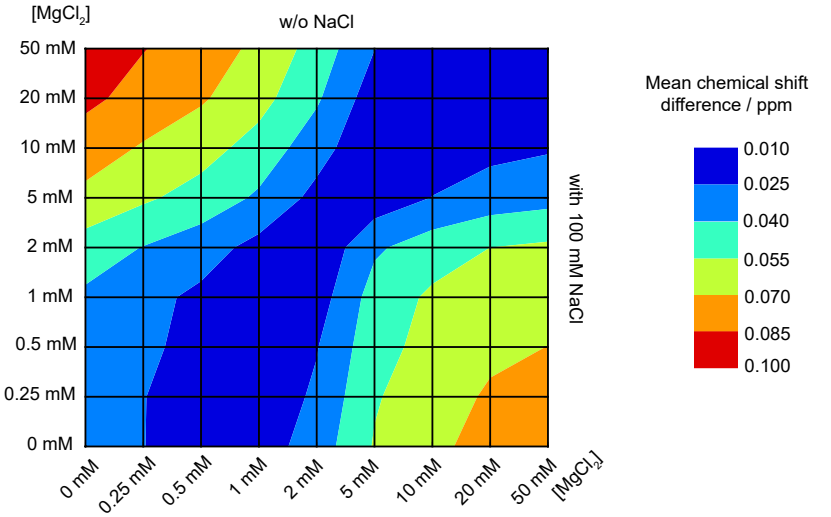


Figure A6: Mean chemical shift difference between Dz^{5C}:$RNA^{2'F}$ spectra with different NaCl and $MgCl_2$ concentration. Chemical shift differences of pyrimidine crosspeaks between one spectrum of Dz^{5C}:$RNA^{2'F}$ in absence of NaCl and with different $MgCl_2$ concentrations (0, 0.25, 0.5, 1, 2, 5, 10, 20, 50 mM) and one spectrum in presence of 100 mM NaCl and with different $MgCl_2$ concentrations were calculated. The mean value of all chemical shift differences was built for each pair of spectra and are shown as a contour plot. Spectra with high ionic strength are most similar, as well as spectra with 100 mM NaCl in absence of $MgCl_2$ and with 0.25-1 mM $MgCl_2$ in absence of NaCl.

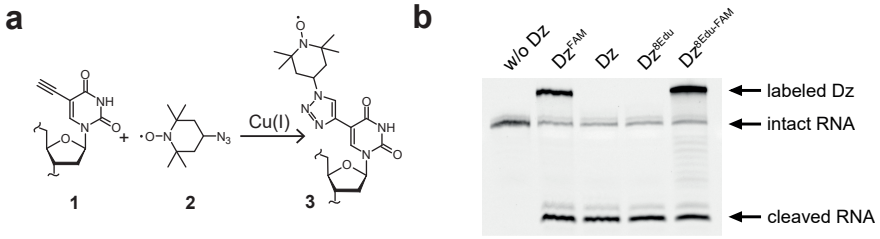


Figure A7: Spinlabeling of Dz^{5C}. (a) Thymidines at position -8, 8, and +7 were substituted by 5-Ethynyl-2'-deoxyuridine (EdU, **1**) and functionalized with a paramagnetic 4-azido-2,2,6,6-tetramethylpiperidinyl-1-oxy (TEMPO) spinlabel (**2**) via Cu(I) click chemistry leading to EdU-TEMPO (**3**). (b) SDS-PAGE of RNA and modified Dz after cleavage relaxtion: (from left to right) negative control without Dz, with Dz modified with FAM at its 3' end, with unmodified Dz, with Edu-substituted Dz at position 8, with Dz modified with Edu-FAM at position 8. Only FAM-modified RNA and Dz gives a signal in the gel.

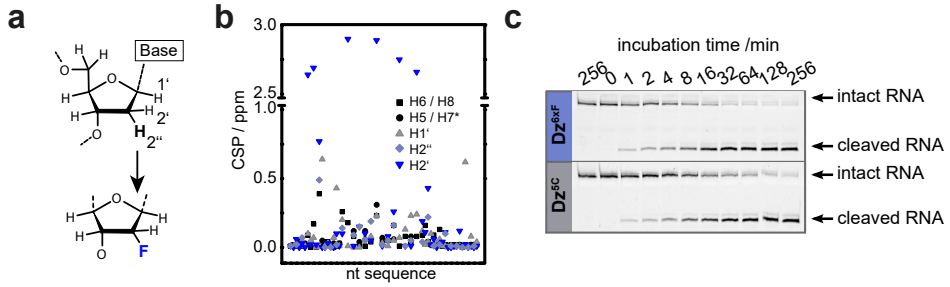


Figure A8: 2'-^{19}F substitutions of deoxynucleotides do not affect Dz's structure and function. (a) Scheme of fluorination at 2' position of deoxyribose. (b) Comparison of proton resonances between Dz^{6xF}:RNA$^{2'F}$ and Dz^{5C}:RNA$^{2'F}$ shows only high CSP for protons geminal to ^{19}F indicating no structural impact caused by the fluorinations. (c) Dz^{6xF} has the same cleavage efficiency as Dz^{5C} in a SDS-PAGE assay. The first lane is a negative control in absence of any Dz.

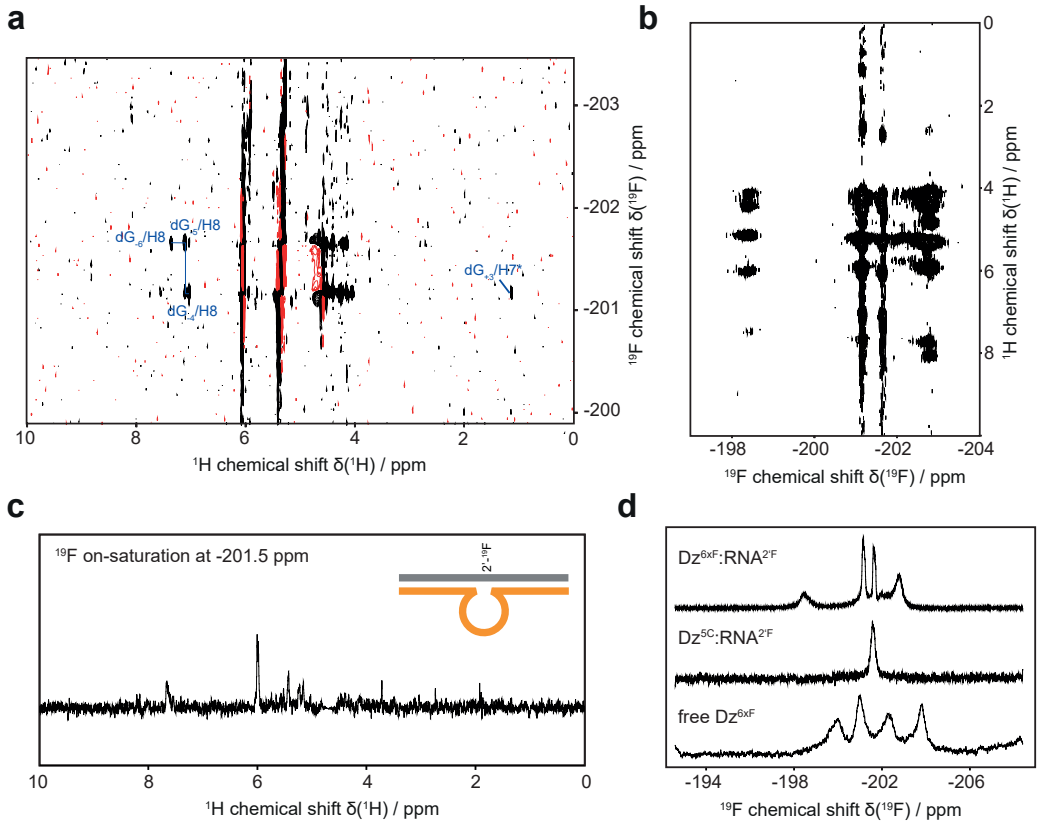


Figure A9: HOESY experiments performed with Dz^{6xF}:RNA$^{2'F}$. (a) ^{1}H-detected [^{1}H,^{19}F]-HOESY spectrum gives strong crosspeaks for short-range distances (sequential crosspeaks of dG$_{-6}$-dG$_{-4}$ shown) but to low signal to noise ratio masks relevant long-range NOE contacts (only assigned crosspeak with dT$_{+3}$/H7*). (b) ^{19}F-detected [^{19}F,^{1}H]-HOESY spectrum shows strong signal intensity but lacks resolution. (c) ^{19}F-derived STD experiment of Dz^{5C}:RNA$^{2'F}$. (d). ^{19}F 1D spectra of Dz^{6xF}:RNA$^{2'F}$, Dz^{6xF}:RNA$^{2'F}$ and single-stranded Dz^{6xF}. Fluorine signals of hybridized nucleotides dG^{-6} and dG^{-5} show significant smaller linewidths (100 Hz) compared to other peaks (200-250 Hz).

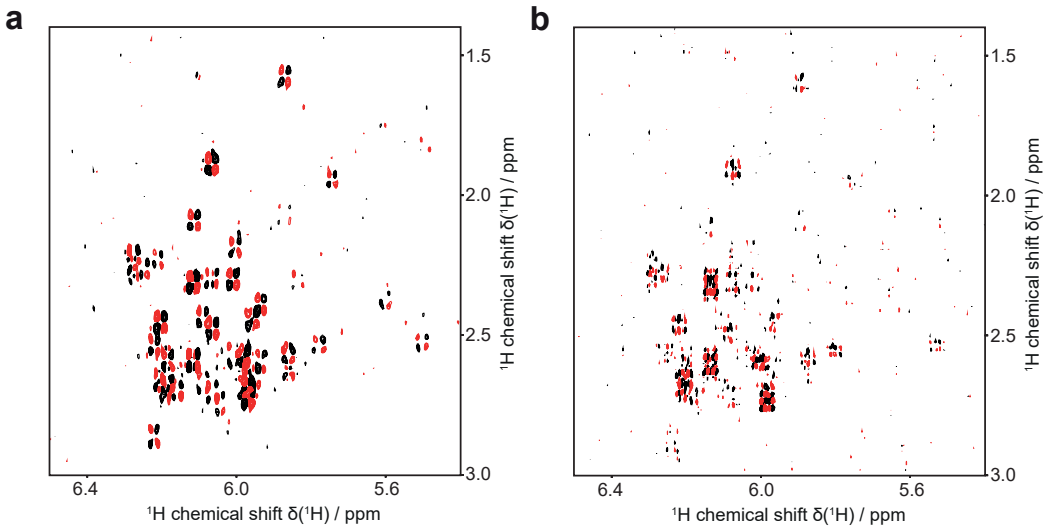


Figure A10: [^{1}H,^{1}H]-DQF-COSY are not usable to determine J coupling constants of Dz^{5C}:RNA$^{2'F}$. (a) H1'-H2' and H1'-H2" section of [^{1}H,^{1}H]-DQF-COSY spectrum recorded with 2048 points in direct dimension shows only splitting due to large geminal proton-proton coupling (J_{HH}^2). (b) H1'-H2' and H1'-H2" section of [^{1}H,^{1}H]-DQF-COSY spectrum recorded with 4096 points in direct dimension shows complex overlapping peak splitting patterns.

Figure A11: Subset of determined NOE build-up curves.. Normalized NOE intensities recorded with mixing times between 40 and 800 ms (filled circles) and build-up curves (lines) determined by eNORA [112, 119]. [^{1}H,^{1}H]-NOESY crosspeaks above (purple) and below (grey) the diagonal were considered for most NOE contacts. Notably, even short and fixed distances, such as H2'-H2", show differences in their curve behaviour, which demonstrates the necessity of correction by spin diffusion.

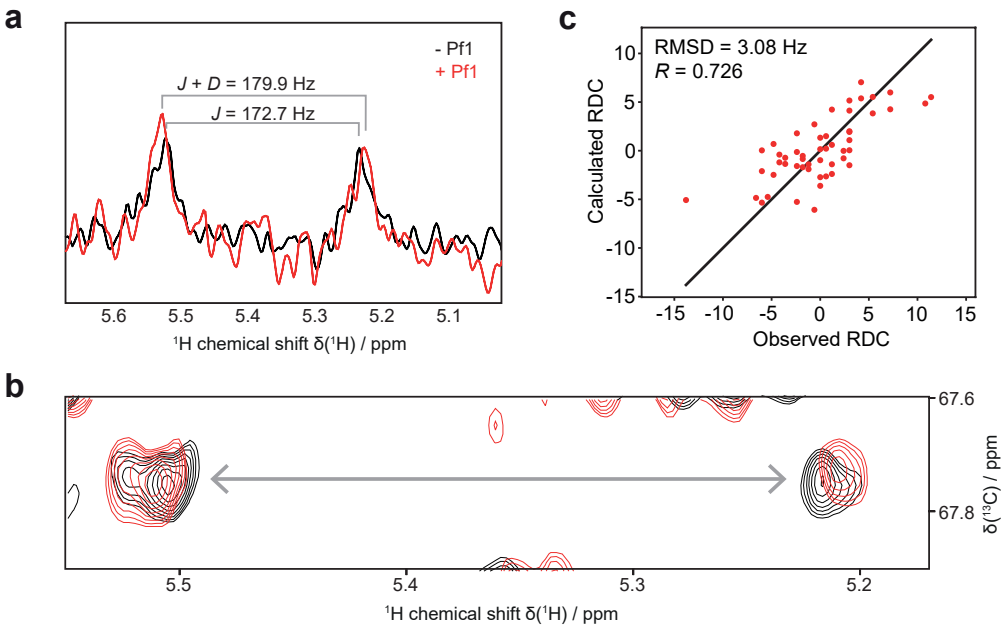
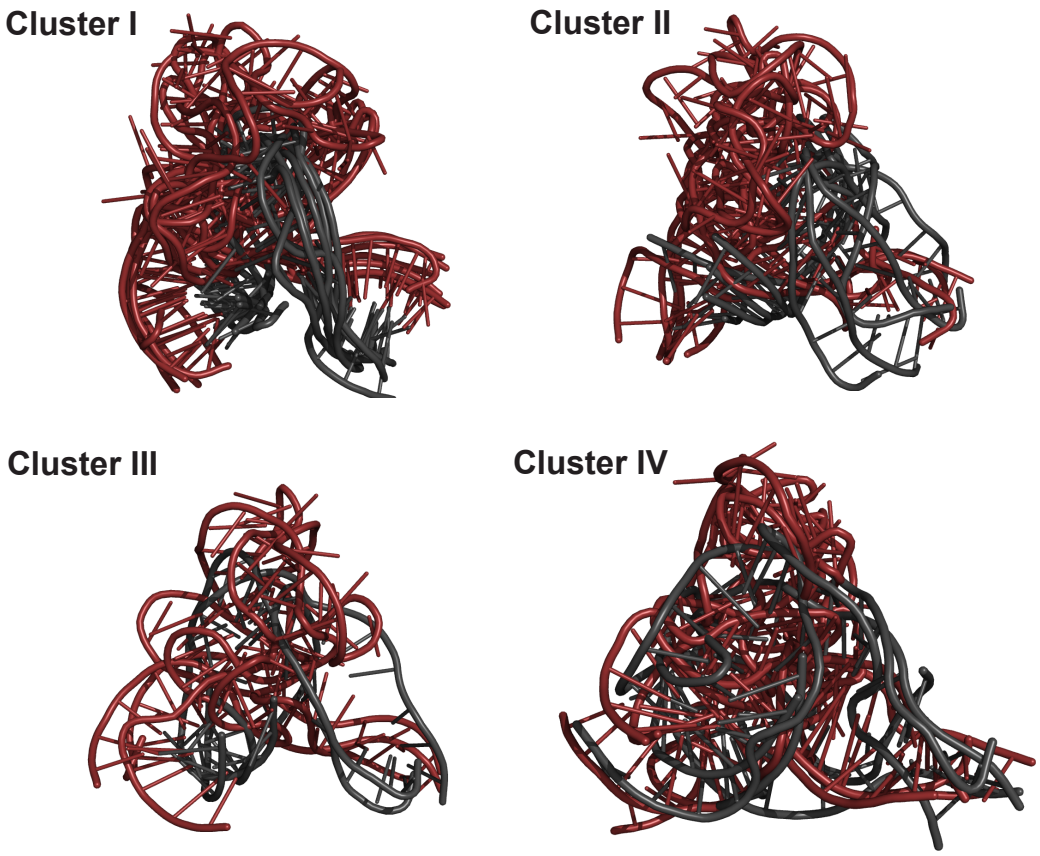


Figure A12: Residual dipolar coupling of $^{1}H,^{13}C$-Dz^{5C}:RNA$^{2'F}$ at 20 °C. (a) RDC constants *D* were determined by subtraction of determined coupling constants of a sample in presence (red) and absence (red) of Pf1. *D* is rather small compared to scalar coupling $J_{HC}{}^{1}$. 1D slices of H5-C5 peaks of dC_{13} are shown (seen in b). (b) Section for H5-C5 peaks of dC_{13} of [$^{1}H,^{13}C$]-HSQC spectra without ^{1}H decoupling are shown in presence (red) and absence (red) of Pf1. Grey arrow indicates the 1D slices shown in a. (c) Correlation plot of observed and back-calculated RDC constants for lowest-energy structure of cluster I in Figure A14.



.

Figure A13: Cluster analysis of the *ab initio* structure calculation for Dz^{5C}:$RNA^{2'F}$ without RDC restraints. 20 lowest-energy structures out of a *ab initio* simulated annealing rMD calculation (200 runs) were sorted into four different clusters: (Cluster I) The RNA strand passes through the Dz's loop and is winding around both Dz's binding arms in 7 structures. (Cluster II+III) The RNA strands winds around the Dz without passing through the loop, while the loop is positioned downstream of the cleavage site in 5 structures (II) and upstream of the cleavage site in 3 structures (III). (Cluster IV) 5 structures which are deviating from clusters I-III.

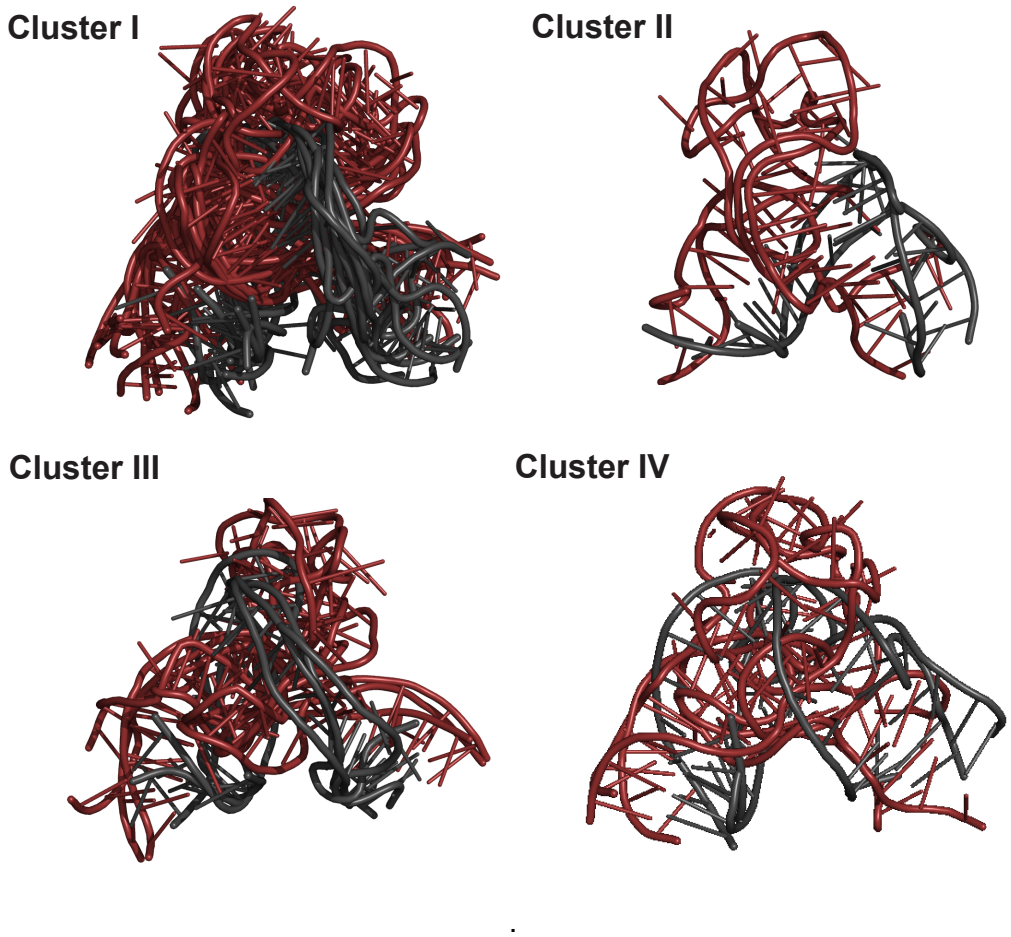
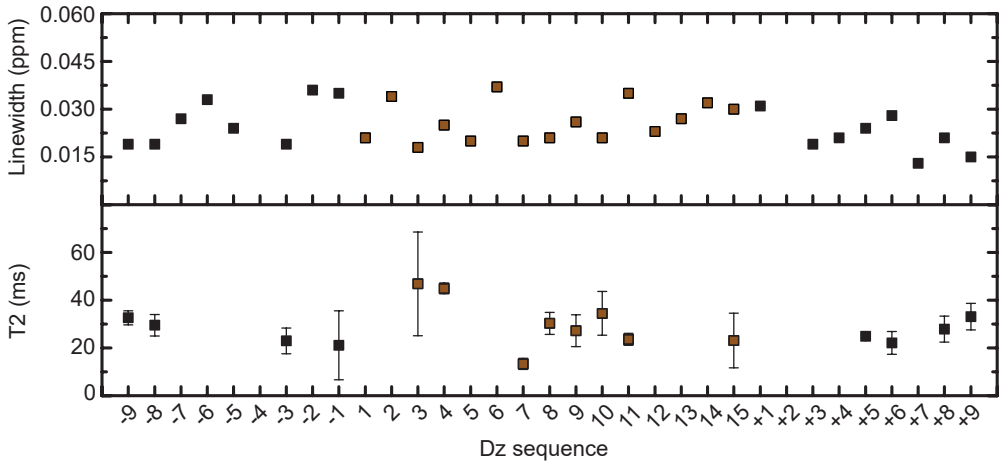
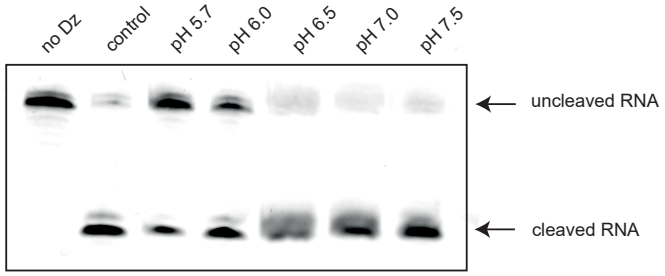


Figure A14: Cluster analysis of the *ab initio* structure calculation for Dz^{5C}:$RNA^{2'F}$ with RDC restraints. 20 lowest-energy structures out of a *ab initio* simulated annealing rMD calculation (200 runs) were sorted into four different clusters: (Cluster I) The RNA strand passes through the Dz's loop and is winding around both Dz's binding arms in 11 structures. (Cluster II+III) The RNA strands winds around the Dz without passing through the loop, while the loop is positioned downstream of the cleavage site in 2 structures (II) and upstream of the cleavage site in 4 structures (III). (Cluster IV) 3 structures which are deviating from clusters I-III.



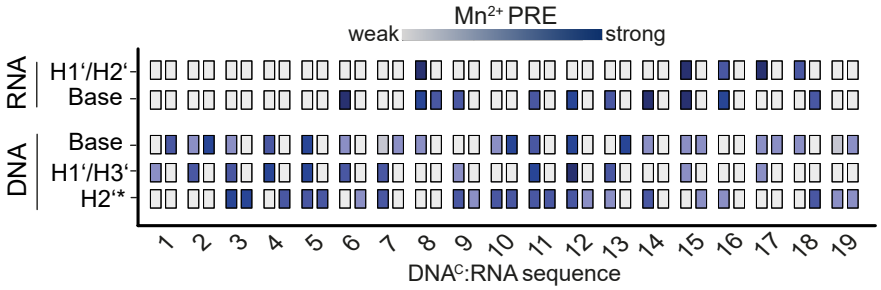
.

Figure A15: Linewidth and inconsistent T2 relaxation data of $^{15}N,^{13}C$-Dz^{5C}:RNA$^{2'F}$. (top) H1'-C1' peak linewidths extracted from [$^{1}H,^{13}C$]-HSQC spectrum of $^{15}N,^{13}C$-Dz^{5C}:RNA$^{2'F}$ does not show differences between binding arms (black) and loop region (brown). (bottom) T2 relaxation data of $^{15}N,^{13}C$-Dz^{5C}:RNA$^{2'F}$ is insufficient probably due to C-C coupling of adjacent C2'.



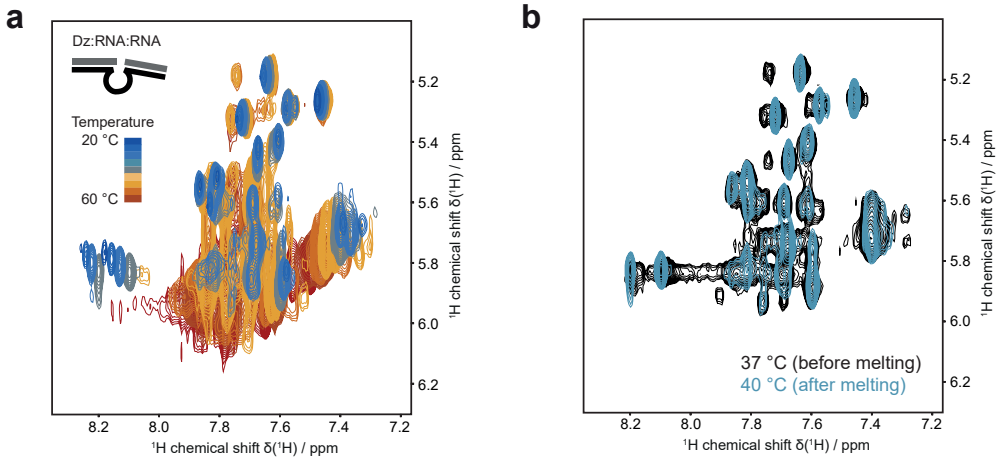
.

Figure A16: Dz^{5C} activity is pH dependent. Cleavage activity recorded after 3h reaction time by SDS-PAGE with 10 mM $MgCl_2$ at 37 °C in absence (no Dz) and presence of Dz in 50 mM Tris/HCl pH 7.5 (control) or 50 mM Na_2HPO_4/NaH_2PO_4 at pH 5.7, 6.0, 6.5, 7.0, and 7.5.



.

Figure A17: Mn^{2+} is not specifically binding at DNA^{C}:RNA$^{2'F}$. Mean PRE rates for each proton plotted against the sequence of DNA^{C}:RNA$^{2'F}$ does not reveal a specific binding site.



.

Figure A18: Melting of the product complex is reversible. (a) Overlay of [^{1}H,^{1}H]-TOCSY spectra of Dz^{5C}:products after cleavage at temperatures between 60 and 20 °C (5 °C stepsize). (b) [^{1}H,^{1}H]-TOCSY spectra of product complex before up-heating to 60 °C (black) and during cool-down at 40 °C (light blue) are identical and show that the melting process is reversible.

Table A1: Deoxyribonucleotide resonance assignment of Dz^{5C}:$RNA^{2'F}$.

Nucleotide	H6/H8[a]	H2/H5/H7*[b]	H1'	H2'	H2''	H3'	H4'	H5'	H5''	C1'	C2'	C3'	C4'	C5'
dT_{-9}	7.613	1.741	6.122	2.303	2.582	4.778	4.177	3.791	3.838	88.55	40.30	77.77	88.50	63.77
dT_{-8}	7.667	1.863	6.193	2.434	2.649	4.941				86.76	39.94	78.65	86.82	63.76
dG_{-7}	7.832		5.967	2.725[c]	2.725[c]	4.875	4.390	4.232[c]	4.232[c]	84.94	41.59	76.82	86.36	
dG_{-6}	7.439		5.964	2.678[c]	2.678[c]	4.752	4.337			85.53	42.01	78.58	85.81	
dG_{-5}	7.267		5.966	2.691[c]	2.691[c]	4.667	4.326			85.96	42.02			
dG_{-4}	7.416		6.035	2.581	2.737	4.734	4.379			86.52	41.77	75.93		67.66
dT_{-3}	7.461	1.264	6.083	2.415	2.662	4.905	4.325	4.235	4.188	86.45	40.76	77.28		
dA_{-2}	7.715	6.880	5.975	2.568	2.676	4.907	4.379	4.197	4.238	85.50	40.48	78.66		
dA_{-1}	7.411	7.556	5.425	2.651	2.728	4.831				84.13	41.50	77.97		
dG_{1}	8.016		5.957	2.429	2.668	4.948				85.41	42.02	78.36	86.27	
dG_{2}	7.895		5.876	2.306	2.422	4.899	4.297	4.037	4.120	84.76	40.29	78.22		
dC_{3}	7.673	5.705	6.121	2.098	2.541	4.812	4.350	4.034	4.085	87.81	41.14	79.09		
dT_{4}	7.661	1.868	6.189	2.200	2.488	4.854	4.307	4.040	4.107	87.16	39.82	79.58		
dC_{5}	7.524	5.758	5.748	1.920	2.334	4.823		4.074[c]	4.074[c]	87.29	39.66	77.97	86.11	
dG_{6}	7.959		6.150	2.695	2.593	4.935	4.394			84.97	40.52	78.64	87.81	67.87
dC_{7}	7.751	5.968	6.242	2.250	2.547	4.872	4.249	4.098	4.039	88.29	40.98	77.51		
dT_{8}	7.400	1.809	6.054	1.903	2.290	4.757	4.218	4.019	4.051	87.25	39.68	75.38	86.86	
dA_{9}	8.267	8.015	6.190	2.590	2.668	4.874	4.316	4.039[c]	4.039[c]	85.94	40.30	78.53		
dC_{10}	7.268	5.614	5.870	1.586	2.080	4.603	4.316	3.752	3.827	87.34	40.28	77.82		67.45
dA_{11}	8.057	7.781	5.853	2.547	2.610	4.874	4.220	3.872	3.948	85.52	39.24	79.26	87.65	
dA_{12}	8.167		5.989	2.580[c]	2.580[c]	4.912	4.335	4.097	4.097	85.22	40.22	78.56		
dC_{13}	7.341	5.605	5.538	1.808	2.200	4.731	4.334	4.005	4.086	87.09	39.39	77.43		
dG_{14}	7.769		5.523	2.512[c]	2.512[c]	4.928	4.263	3.961	4.046	84.68	39.68	79.10		
dA_{15}	8.114	7.901	6.188	2.706[c]	2.706[c]	5.038	4.420	4.119	4.174	85.24	40.33	78.43	87.40	
dG_{+1}	7.967		6.002	2.799	2.735	4.865	4.392	4.197	4.246	85.10	41.03	77.49	86.68	
dG_{+2}	7.123		5.952	2.424	2.640	4.579	4.280	4.084	4.217	85.88	41.47	75.03		
dT_{+3}	7.407	1.157	5.994	2.156	2.536	4.839		4.115[c]	4.115[c]	85.08	40.62	77.96	85.70	
dG_{+4}	8.016		5.988	2.550[c]	2.550[c]	4.902	4.262	4.083	4.126	85.50	40.06	77.34		
dC_{+5}	7.419	5.222	5.929	2.364	2.586	4.821	4.272	4.194[c]	4.194[c]	86.87	41.02	76.39		
dA_{+6}	8.002	7.470	6.220	2.532	2.861	4.887				85.97	42.09	77.13	86.27	67.54
dT_{+7}	7.253	1.312	5.992	2.275	2.547	4.871	4.275	4.159	4.229	85.48	40.29	78.48		
dG_{+8}	7.590		6.058	2.468	2.689	4.866				85.21	42.01	78.55	86.51	67.07
dT_{+9}	7.378	1.326	6.264	2.212	2.257	4.544	4.367	4.097	4.219	86.53	41.57	72.36	86.50	

[a] Base proton H6 or H8 of pyrimidine or purine, respectively.
[b] Base proton H2, H5 or H7* of adenosine, cytosine/uracil or thymine methyl, respectively.
[c] Assignment of H2'/H2" or H5'/H5" ambiguous.

Table A2: Ribonucleotide resonance assignment of Dz^{5C}:$RNA^{2'F}$.

Nucleotide	H6/H8[a]	H2/H5/H7*[b]	H1'	H2'	H3'	H4'	H5'	H5"
rA_{-9}	7.980	7.968	5.917	4.332	4.403	4.268	4.085	4.155
rA_{-8}	8.067	7.312	5.869	4.350	4.681		4.144[c]	4.144[c]
rC_{-7}	7.752	5.542	5.408	4.302	4.524	4.419	4.070[c]	4.070[c]
rC_{-6}	7.806	5.489	5.403	4.360	4.552		4.058[c]	4.058[c]
rC_{-5}	7.757	5.458	5.497	4.349	4.462			
rC_{-4}	7.537	5.313	5.287	4.311	4.571	4.385	4.242[c]	4.242[c]
rA_{-3}	8.222	7.242	5.969	4.537	4.647	4.444	4.215[c]	4.215[c]
rU_{-2}	8.008	5.743	5.653	4.518	4.613	4.445	4.207[c]	4.207[c]
rU_{-1}	7.660	5.425	5.513	4.525		4.389		
rG_{0}	7.624		5.987	5.195				
rC_{+1}	7.601	5.485	5.569	4.356	4.629		4.056[c]	4.056[c]
rC_{+2}	7.474	5.242	5.391	4.345	4.513	4.167	4.048[c]	4.048[c]
rA_{+3}	8.067	7.490	5.900	4.464	4.638	4.423	4.100[c]	4.100[c]
rC_{+4}	7.657	5.255	5.342	4.450	4.714			
rG_{+5}	7.693		5.744	4.445	4.479		4.126[c]	4.126[c]
rU_{+6}	7.574	5.117	5.426	4.543	4.642	4.431	4.111	4.176
rA_{+7}	8.102		5.938	4.567	4.671		4.181[c]	4.181[c]
rC_{+8}	7.622	5.413	5.337	4.500	4.838		4.181[c]	4.181[c]
rA_{+9}	8.373	8.020	5.928	4.591	4.852	4.421	3.940	4.080

[a] Base proton H6 or H8 of pyrimidine or purine, respectively.
[b] Base proton H2 or H5 of adenosine or cytosine/uracil, respectively.
[c] Assignment of H5'/H5" ambiguous.

Table A3: Determined RDC constants.

Dz nt	Atom bond	RDC / Hz	Dz nt	Atom bond	RDC / Hz
-9	C4'-H4'	0.0	10	C2'-H2"	-4.2
-8	C1'-H1'	0.0	10	C5-H5	1.2
-7	C1'-H1'	-1.2	11	C1'-H1'	5.4
-7	C3'-H3'	-1.8	11	C2'-H2'	1.2
-6	C1'-H1'	-6.6	11	C2'-H2"	-2.4
-4	C1'-H1'	10.8	12	C1'-H1'	-2.4
-4	C3'-H3'	7.2	13	C1'-H1'	0.0
-3	C1'-H1'	4.2	13	C2'-H2'	-5.4
1	C1'-H1'	3	13	C2'-H2"	3.0
2	C1'-H1'	-4.8	13	C5-H5	7.2
3	C1'-H1'	-0.6	14	C1'-H1'	5.4
3	C5-H5	-6.0	15	C1'-H1'	-0.6
4	C1'-H1'	11.4	+1	C1'-H1'	-3.6
5	C2'-H2'	3.0	+2	C3'-H3'	-4.2
5	C2'-H2"	3.0	+3	C1'-H1'	2.4
5	C5-H5	-13.8	+4	C1'-H1'	-2.4
6	C1'-H1'	2.4	+5	C1'-H1'	3.0
7	C1'-H1'	0.6	+5	C3'-H3'	-3.6
7	C5-H5	-2.4	+5	C5-H5	-6.0
8	C1'-H1'	-6.0	+6	C1'-H1'	0.0
8	C2'-H2'	1.2	+7	C1'-H1'	0.6
8	C2'-H2"	1.2	+8	C1'-H1'	-1.2
8	C3'-H3'	3	+9	C1'-H1'	-1.8
9	C1'-H1'	3	+9	C2'-H2'	4.2
10	C1'-H1'	0.6	+9	C2'-H2"	-4.8
10	C2'-H2'	-1.8	+9	C3'-H3'	0.0

Table A4: Restraints and violations of Dz^{5C}:$RNA^{2'F}$ *ab-initio* structure calculation.

		without RDC		with RDC	
Restraints	**Number**	**Violations[a]** **Number[b]**	**RMSD[c]**	**Number[b]**	**RMSD[c]**
conventional NOE					
intra-nucleotide	691	50.3(6.8)	0.32(0.03)	43.5(6.3)	0.28(0.03)
inter-nucleotide	274				
long-range	43	0.1(0.4)	0.13(0.06)	0.0(0.0)	0.08(0.02)
exact NOE	108	22.0(2.1)	0.69(0.14)	18.8(2.3)	0.60(0.08)
^{19}F-derived distances	44	0.1(0.2)	0.06(0.04)	0.0(0.0)	0.04(0.02)
Spinlabel-derived distances	126	2.8(1.9)	0.36(0.16)	1.7(1.0)	0.21(0.10)
H bonds	143	0.0(0.0)	0.05(0.02)	0.0(0.0)	0.04(0.01)
Dihedral angles	208	7.7(2.4)	0.76(0.06)	4.7(2.4)	0.70(0.05)
Residual dipolar coupling	42	–	–	16.8(2.4)	3.34(0.21)
Ensemble RMSD / Å		**Lowest energy[a]**	**Cluster I[d]**	**Lowest energy[a]**	**Cluster I[e]**
		9.019	5.951	7.408	6.261

[a] Calculated from the 20 lowest energy structures out of 200 calculation runs.
[b] Mean number of violations in the ensemble (standard deviation within the ensemble). Thresholds for violations are 0.5 Å for distances, 5° for angles and 2.5 Hz for RDC.
[c] Mean RMSD of violations in the ensemble (standard deviation within the ensemble). Distance violations in Å, angular violations in °, and RDC violations in Hz.
[d] Seven out of 20 lowest energy structures (see Figure A13).
[e] Eleven out of 20 lowest energy structures (see Figure A13).

Table A5: NMR parameters for recorded spectra on states of Dz^{5C} catalytic cylce.

Sample	Experiment	NaCl / mM	$MgCl_2$ / mM	Transmitter offset F2/F1 / ppm	Spectral width F2/F1 / ppm	Data points F2xF1	Number of scans	Basic frequency / MHz	Figures
Single-stranded oligomers									
200 µM ssDz^{5C}	TOCSY	100	1	4.697/4.697	10/10	2048x512	64	599.30	3.3I, A13
	NOESY				20/20	2048x1024	48		A13
45 µM ssRNA$^{2'F}$	TOCSY[a]	100	1	4.702/4.702	20/10	2048x512	128	599.86	3.3II, A2
	NOESY[a]								A2
Stabilized complex									
	TOCSY	100	0	4.688/4.688	10/10	2048x512	32	899.84	3.3III
	NOESY				20/20	2048x1024	128		3.13c, 3.15b
200 µM Dz^{5C}:RNA$^{2'F}$	TOCSY	100	1	4.688/4.688	10/10	2048x512	56	899.84	3.2c+d, 3.4c, 3.5a
	NOESY				20/20	2048x1024	64		3.12b+c, 3.13c, 3.15b
	TOCSY	100	50	4.690/4.690	10/10	2048x512	48	899.84	3.13c, 3.15b
	NOESY				20/20	2048x1024	56		
pH 5.7	TOCSY[a]	100	1	4.697/4.697	10/10	2048x512	64	599.35	3.12c
Complex with unstabilized RNA									
200 µM Dz^{5C}:RNA	TOCSY	100	0	4.696/4.696	10/10	2048x256	16	599.35	3.2d, 3.3IV, 3.16c+d
200 µM Dz^{5C}:RNA	TOCSY[b]	100	1	4.700/4.700	10/10	2048x256	24	750.20	3.16c+f
200 µM Dz^{5C}:products	TOCSY	100	1	4.700/4.700	10/10	2048x256	16	750.20	3.3V, 3.16c+d
	NOESY					2048x512	128		
Melting experiments[c]		**T / °C**							
500 µM Dz^{5C}:RNA$^{2'F}$	TOCSY	58-15 (4 °C steps) 60-38 (2 °C steps)		4.700/4.700	10/10	2048x256	32	599.35	A1
200 µM Dz^{5C}:products	TOCSY	60-20 (5 °C steps)		4.692/4.692	10/10	2048x256	8	899.84	3.3VI, A18
200 µM ssDz^{5C}	TOCSY	56-16 (4 °C steps)		4.697/4.697	10/10	2048x256	32	599.30	A1

All spectra recorded in 50 mM Tris/HCl pH 7.5, 10% (v/v) D_20 at 37 °C, if not stated otherwise.
[a] Recorded in 50 mM Na_2HPO_4/NaH_2PO_4, 10% (v/v) D_20.
[b] Used for RT-NMR.
[c] All spectra recorded with 100 mM NaCl, 1 $MgCl_2$.

Table A6: NMR parameters for recorded titration experiments.

[^{1}H,^{1}H]-TOCSY spectra recorded for Mg^{2+} titration

Sample	NaCl / mM	$MgCl_2$ / mM	Transmitter offset F2/F1 / ppm	Spectral width F2/F1 / ppm	Data points F2xF1	Number of scans	Basic frequency / MHz	Figures
200 µM Dz^{5C}:$RNA^{2'F}$	100	0, 0.25, 0.5, 1, 2, 5, 10, 20, 50	4.696/4.696	10/10	2048x256	16	599.35	3.6c
100 µM Dz^{5A}:$RNA^{2'F}$	100		4.696/4.696	8/8	2048x256	32	599.86	3.6d
200 µM Dz^{5C}:$RNA^{2'F}$	0		4.699/4.699	10/10	2048x256	16	750.20	3.7d

Spectra recorded for Mn^{2+} titration[(a)]

Sample	Experiment	$MnCl_2$ / µM	Transmitter offset F2/F1 / ppm	Spectral width F2/F1 / ppm	Data points F2xF1	Number of scans	Basic frequency / MHz	Figures
200 µM Dz^{5C}:$RNA^{2'F}$	NOESY	0, 0.5, 1, 2, 5, 10	4.703/4.703	10/10	2048x512	32	700.30	3.14b
	TOCSY		4.703/4.703	10/10	2048x512	64	700.30	
100 µM DNA^{C}:$RNA^{2'F}$	NOESY		4.703/4.703	10/10	2048x512	16	700.30	

[^{1}H,^{1}H]-NOESY spectra recorded for $Co(nH_3)_6$ titration

Sample	NaCl	$MgCl_2$ / mM	$Co(nH_3)_6$	Transmitter offset F2/F1 / ppm	Spectral width F2/F1 / ppm	Data points F2xF1	Number of scans	Basic frequency / MHz	Figures
200 µM Dz^{5C}:$RNA^{2'F}$	100	0	0	4.693/4.692	20/20	2048x256	16	*599.86	3.14g
	100	0	0.2	4.692/4.694	20/20	2048x256	16	*599.86	
	100	0	0.5	4.694/4.694	20/20	2048x256	16	599.86	
	100	0	1	4.694/4.694	20/20	2048x256	16	599.86	
	100	1	1	4.693/5.695	20/20	2048x256	16	599.86	

All spectra recorded in 50 mM Tris/HCl pH 7.5, 10% (v/v) D_20 at 37 °C, if not stated otherwise.
[(a)] All spectra recorded with 100 mM NaCl, 1 mM $MgCl_2$.

Table A7: NMR parameters for other recorded homonuclear spectra on other oligomers.

Sample	Experiment	NaCl / mM	$MgCl_2$ / mM	Transmitter offset F2/F1 / ppm	Spectral width F2/F1 / ppm	Data points F2xF1	Number of scans	Basic frequency / MHz	Figures
200 µM Dz^{5G}:$RNA^{2'F}$	TOCSY	100	1	4.692/4.692	10/10	2048x512	32	750.20	3.4c
	NOESY			4.692/4.692	20/20	2048x512	64	750.20	
100 µM Dz^{5A}:$RNA^{2'}$	TOCSY	0	0	4.701/4.701	8/8	2048x256	32	599.86	
100 µM Dz^{5A}:$RNA^{2'}$	TOCSY	0	1	4.702/4.702	8/8	2048x256	120	599.86	
100 µM Dz^{5A}:$RNA^{2'}$	TOCSY	100	0	4.701/4.701	8/8	2048x256	40	599.86	3.4c, 3.7c
	NOESY			4.701/4.701	8/8	2048x512	32	599.86	
100 µM Dz^{5A}:$RNA^{2'}$	TOCSY	500	0	4.701/4.701	8/8	2048x256	32	599.86	3.7c
100 µM Dz^{5A}:$RNA^{2'}$	TOCSY	100	1	4.701/4.701	8/8	2048x256	120	599.86	3.4c, 3.7c
220 µM Dz^{Edu-8}:$RNA^{2'F}$	TOCSY[a]	100	1	4.713/4.713	10/10	2048x512	128	750.20	
	NOESY[a]			4.713/4.713	10/10	2048x512	208	750.20	3.8a
350 µM Dz^{Edu8}:$RNA^{2'F}$	TOCSY[a]	100	1	4.689/4.689	10/10	2048x352	32	899.84	
	NOESY[a]			4.689/4.689	10/10	2048x512	64	899.84	3.8b
100 µM Dz^{Edu+7}:$RNA^{2'F}$	TOCSY[a]	100	1	4.694/4.694	10/10	2048x512	88	899.84	3.8c
750 µM Dz^{6xF}:$RNA^{2'F}$	TOCSY	100	1	4.697/4.697	10/10	2048x256	64	599.35	
	NOESY			4.697/4.697	20/20	2048x1024	64	599.35	3.9b
200 µM DNA^{C}:$RNA^{2'F}$	TOCSY	100	0	4.705/4.705	20/10	2048x512	32	599.86	3.10b
	NOESY			4.705/4.705	20/20	2048x512	48	599.86	
200 µM Dz^{5C}:$RNA^{2'F}$	DQF-TOCSY[b]	100	1	4.706/4.706	10/10	2048x512	128	700.30	A10
				4.706/4.706	10/10	4096x512	64	599.40	

All spectra recorded in 50 mM Tris/HCl pH 7.5, 10% (v/v) D_20 at 37 °C, if not stated otherwise.
[a] All spectra recorded in presence and absence of 2 mM ascorbic acid.
[a] Recorded in 100% (v/v) D_20.

Table A8: NMR parameters for carbon-correlated experiments.

Sample	Experiment	Transmitter offset (F3/)F2/F1 / ppm	Spectral width (F3/)F2/F1 / ppm	Data points (F3x)F2xF1	Number of scans	Basic frequency / MHz	Figures
^{13}C assignment							
500 µM ^{13}C,^{15}N-Dz^{5C}:RNA$^{2'F}$	[^{13}C,^{15}N]-HSQC	4.699/50	10/100	2048x512	64	750.20	3.5c
	2C CCH-TOCSY	4.698/50	10/100	2048x512	32	599.86	
	[^{13}C,^{15}N]-HSQC-NOESY	4.700/50/4.700	10/100/10	2048x64x320	16	750.20	
J coupling							
500 µM ^{13}C,^{15}N-Dz^{5C}:RNA$^{2'F}$	HcCH-E.COSY[(a)]	4.703/60/4.703	10/70/7	2048x176x256	8	750.20	3.10a
Residual dipolar coupling							
500 µM ^{13}C,^{15}N-Dz^{5C}:RNA$^{2'F}$	[^{13}C,^{15}N]-HSQC	4.695/72.5	10/31	2048x512	80	599.86	A12
+/- 10 mg/ml Pf1	[^{13}C,^{15}N]-HSQC	4.695/72.5	10/31	2048x512	240	599.86	
Dynamics							
200 µM ^{13}C,^{15}N-Dz^{5C}:RNA$^{2'F}$	[^{13}C,^{15}N]-ctHSQCT1	4.696/72.5	10/31	2048x256	104	750.20	
	[^{13}C,^{15}N]-ctHSQCT2	4.699/72.5	10/31	2048x256	136	750.20	
500 µM ^{13}C,^{15}N-Dz^{5C}:RNA$^{2'F}$	[^{13}C,^{15}N]-HSQCnoe	4.705/72.5	10/31	2048x256	128	599.35	

All spectra recorded in 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 10% (v/v) D_20 at 37 °C, if not stated otherwise.
[a] Recorded in 100% D_20.

Supplementary Information for section 4.1

Molecular Architecture of a Network of Potential Intracellular EGFR Modulators: ARNO, CaM, Phospholipids, and the Juxtamembrane Segment

Aldino Viegas[1,2,a#], Dongsheng M. Yin[3,4#], Jan Borggräfe[1,2,5], Thibault Viennet[1,2,b], Marcel Falke[1], Anton Schmitz[3,4*], Michael Famulok[3,4], Manuel Etzkorn[1,2,5,6*]

[1] Institute of Physical Biology, Heinrich-Heine University, Düsseldorf, Germany

[2] Institute of Complex Systems (ICS-6), Forschungszentrum Jülich, Wilhelm-Jonen-Strasse, 52425 Jülich, Germany

[3] Max Planck Fellow Chemical Biology, Center of Advanced European Studies and Research (caesar), Ludwig-Erhard-Allee 2, 53175 Bonn, Germany

[4] Department of Chemical Biology, Life and Medical Sciences (LIMES) Institute, Rheinische Friedrich-Wilhelms-Universität Bonn, Gerhard-Domagk-Str.1, 53121 Bonn, Germany

[5] JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich, Wilhelm Jonen Strasse, 52425 Jülich, Germany

[6] Lead contact

[#] equal contribution

Corresponding Authors

[*] anton.schmitz@uni-bonn.de

[*] manuel.etzkorn@hhu.de

Present Addresses

[a] UCIBIO-REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

[b] Dana-Farber Cancer Institute, Cancer Biology, 360 Longwood Ave, Boston, MA 02215, USA

Contribution

Jan Borggräfe performed and analysed NMR experiments on JM, nanodiscs and CaM.

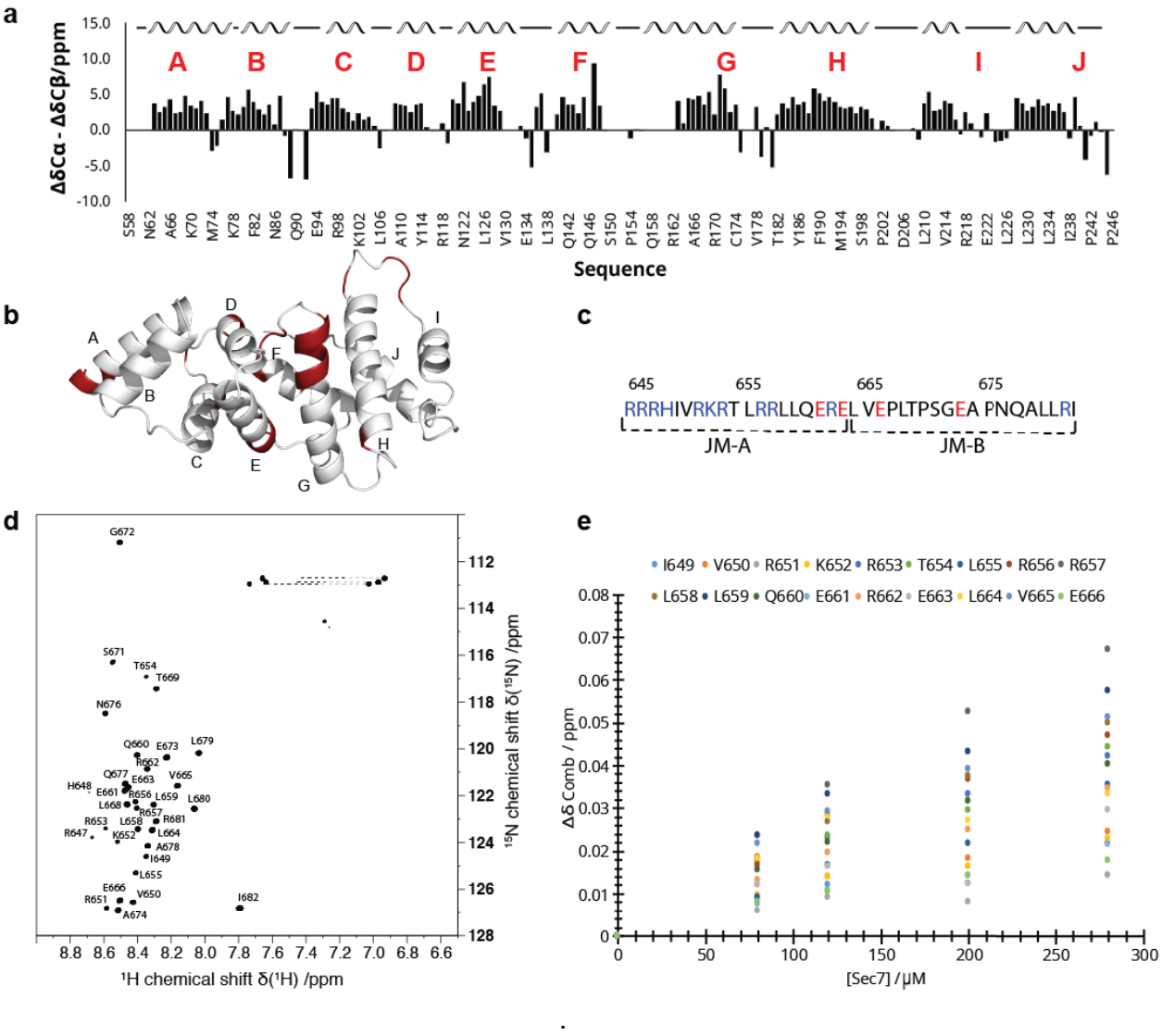


Figure A19: Sec7 structure and assignment, related to Figures 4.1+4.2. (a) CSI-predicted secondary structure of Sec7 based on ^{13}Cα and ^{13}Cβ, chemical shifts (Chemical Shift Index). Three or more consecutive negative values indicate β-strand while three or more positive values indicate a α-helical structure. On top of the plot the secondary structure as defined by the crystal structure is shown (PDB code: 4JMI [136]) (b) Crystal structure of Sec7 depicting the non-assigned residues in red. Helix nomenclature is according to [108]. Data was acquired at 32 °C. (c) Sequence of JM highlighting positively and negatively charged residues in blue and red, respectively. JM can be divided into 2 segments: JM-A (residues R645 to E663) and JM-B (residues L664 to 682) [67]. For JM, nearly complete backbone resonance assignment was obtained and only the amide resonances of R645 and R646 could not be assigned. (d) [^{1}H,^{15}N]-HSQC with assignment of JM in aqueous solution. Data was acquired at pH 5.5 and 15 °C. (e) Combined chemical shift for the JM titration with Sec7, as a function of the sequence. Only the residues affected (corresponding to JM-A) are represented. The concentration of JM was maintained at 40 μM and the concentration of Sec7 varied from 0 to 280 μM (0.0, 2.0, 3.0, 5.0 and 7.0 molar equivalents). Data was acquired at 15 °C in 20 mM sodium phosphate buffer containing 100 mM NaCl, 10% (v/v) D_2O, 0.01% sodium azide and 100 μM DSS, pH 5.5. Due to the absence of a clear plateau (before reaching the solubility limit of the titrant) no residue specific affinities were calculated. The behavior is however in line with K_D values in the high μM range.

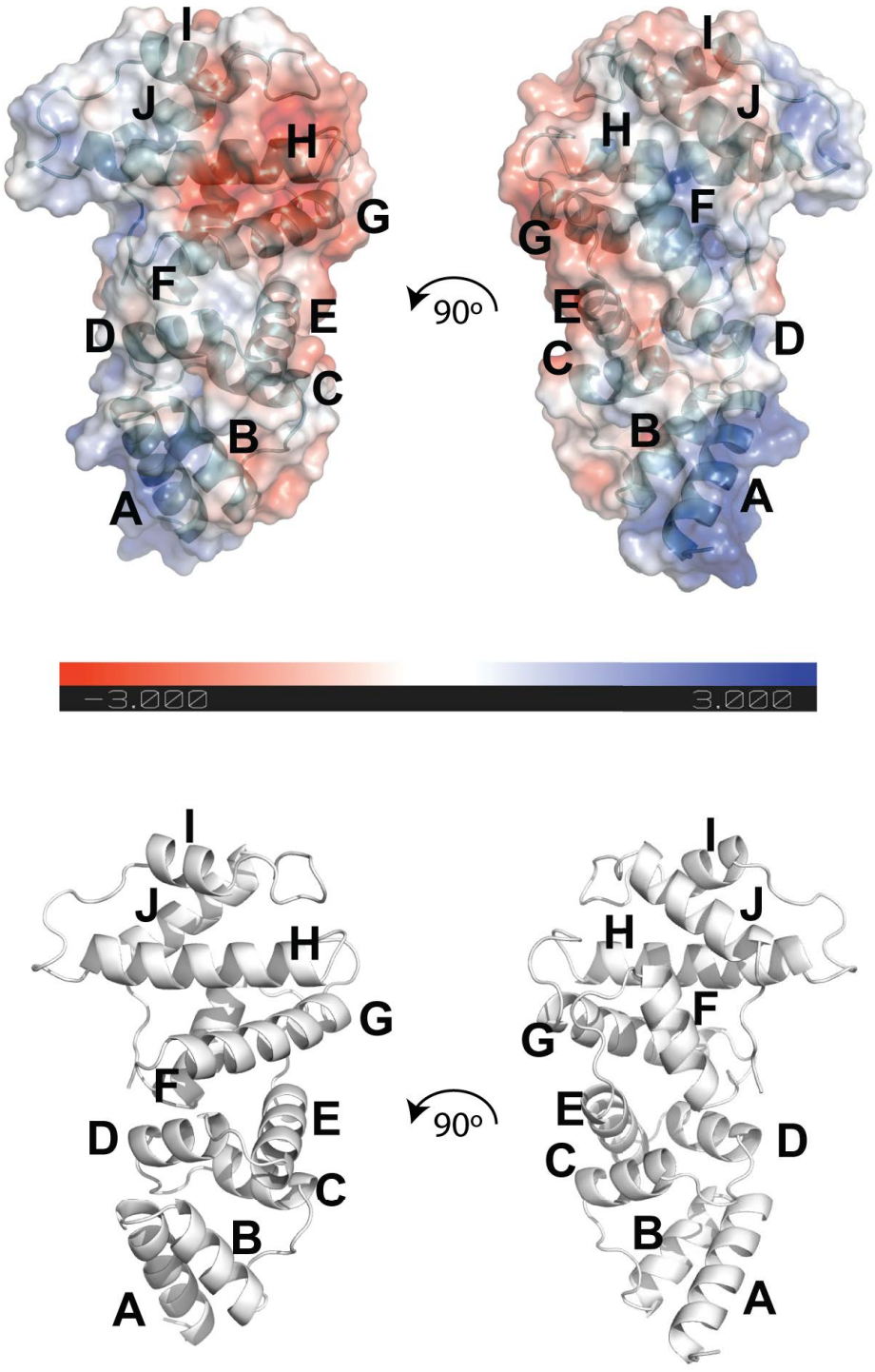


Figure A20: Electrostatic surface of Sec7, related to Figure 4.2. Electrostatic surface of Sec7 according to the APBS Electrostatics module of PyMol. Helix nomenclature is according to [108].

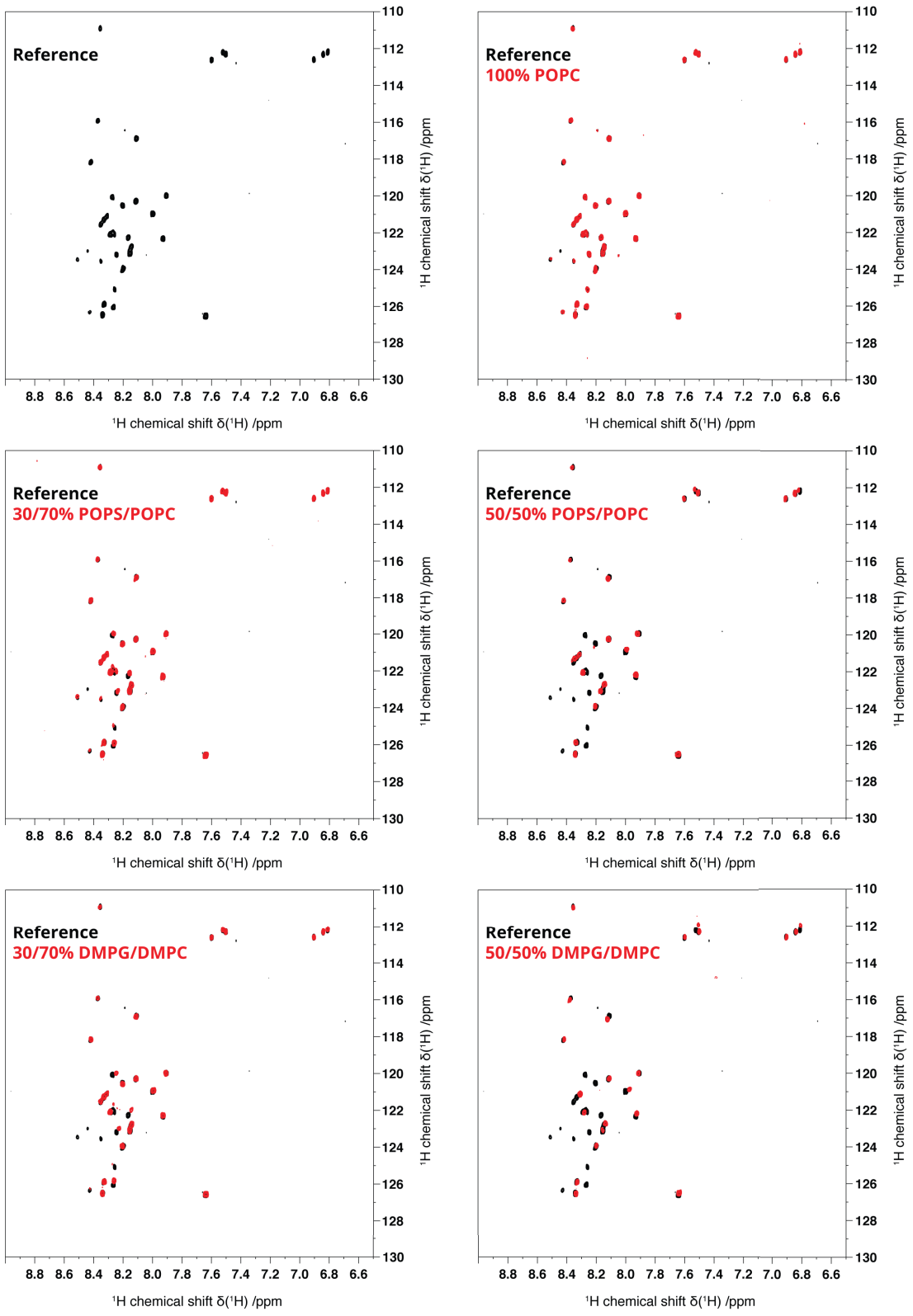


Figure A21: [^{1}H,^{15}N]-HSQC spectra with NDs containing different lipids, related to Figure 4.3. JM ^{1}H,^{15}N-HSQC spectra in the absence (black) and presence of NDs containing the indicated different lipid compositions (red).

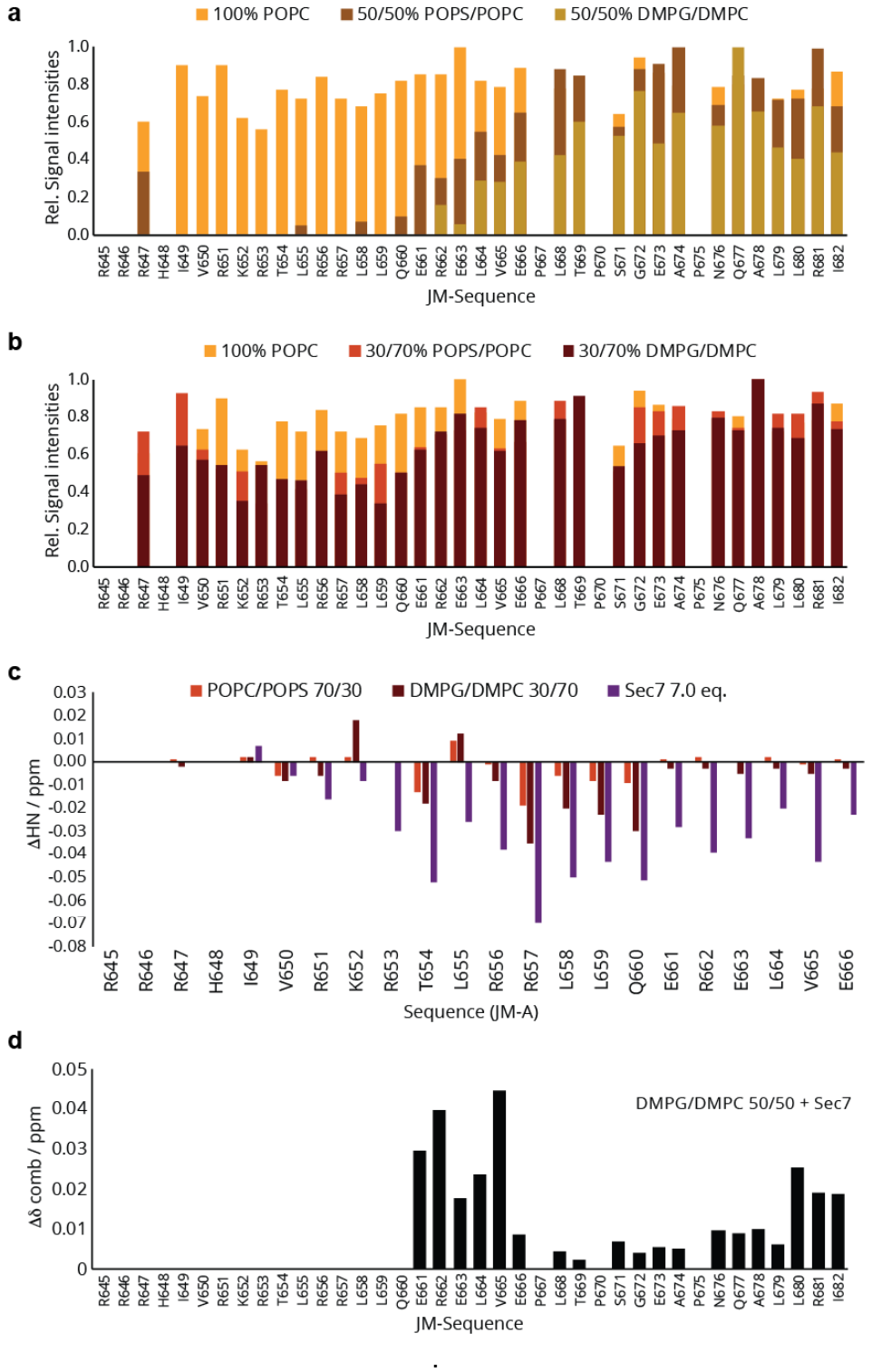


Figure A22: Interactions of JM with NDs containing different lipids, related to Figure 4.3. Relative signal intensities in absence and presence of nanodiscs containing (a) 100% POPC, 30/70% POPS/POPC and 30/70% DMPG/DMPC; (b) 100% POPC, 50/50% POPS/POPC and 50/50% DMPG/DMPC. (c) JM-A amide proton chemical shift deviations upon addition of different interaction partners. The rather uniformly shift towards lower ^{1}H and ^{15}N frequencies would be in line with an increase in transient α-helical propensity upon Sec7 interaction [14]. (d) Combined chemical shift ($\Delta\delta_{comb}$) perturbations of JM in the presence of 1 equivalent of NDs containing a mixture of 50% DMPG/50% DMPC and of 7 equivalents of Sec7. The resonances of the amide groups up to Q660 are bleached from the [^{1}H,^{15}N]-HSQC spectrum upon addition of the NDs (Fig. A16) and are not recovered upon addition of Sec7. However, the chemical shift perturbations displayed by residues E661-V665 demonstrate that even in the presence of NDs, Sec7 is able to interact with JM.

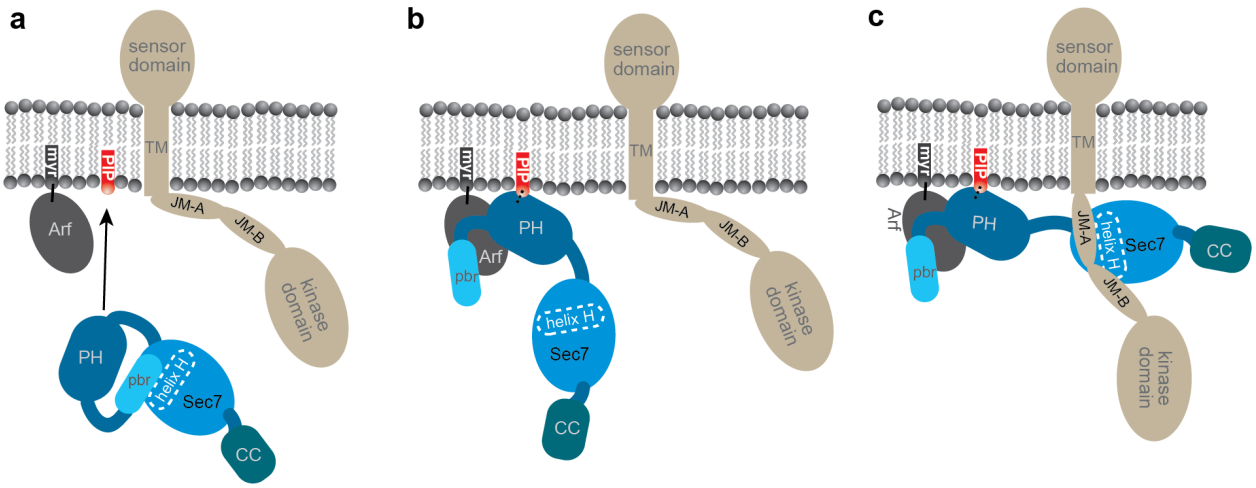


Figure A23: Speculative model of known factors that could promote a potential ARNO-EGFR interaction *in vivo*, related to Figures 4.2+4.5. While autoinhibited ARNO does not interact with EGFR-JM (this study), ARNO is recruited to the membrane via an interaction of ARNO-PH and PIP ([141]) and autoinhibition is reduced via this interaction as well as via interactions of ARNO-PH with membrane located Arf (a,b) ([96], [144], [172]). The increased PIP levels found in the proximity of the EGFR ([102], [141], [149]) could increase co-localization of activated ARNO and EGFR. Previously suggested high linker flexibility in activated ARNO ([149]) could further facilitate interaction of ARNO-Sec7 with EGFR-JM in a similar manner as found in our *in vitro* studies of the isolated domains

Supplementary Information for section 4.2

Local deuteration enables NMR observation of methyl groups in proteins from eukaryotic and cell-free expression systems

Abhinav Dubey[a], Nikolay Stoyanov[b], Sandeep Chhabra[a], Thibault Viennet[c], Shantha Elter[c], Jan Borggräfe[c], Aldino Viegas[c] , Radosław P. Nowak [a], Nikola Burdziev[b], Ognyan Petrov[b], Eric S. Fischer [a], Manuel Etzkorn[c], Vladimir Gelev[b], and Haribabu Arthanari[a]

[a] Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue Boston MA, 02215 (USA) and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston MA, 02215 (USA)

[b] Faculty of Chemistry and Pharmacy, Sofia University, 1 James Bourchier Blvd., 1164 Sofia, Bulgaria

[c] Institute of Physical Biology, Heinrich-Heine-University, Universitätsstr. 1, 40225, Düsseldorf (Germany) and Institute of Biological Information Processing (IBI-7) Forschungszentrum Jülich, Wilhelm-Jonen-Strasse, Jülich (Germany)

Contribution

Jan Borggräfe performed NMR measurements on uniform-labeled bO.

Table A9: Acquisition parameters for different NMR experiments done in the present study.

Experiments	Forbidden Coherence Transfer		2D [^{13}C,^{1}H] T2 Measurement	2D [^{13}C,^{1}H] HSQC	2D [^{13}C,^{1}H] HMQC
	Triple Quantum Coherence	Single Quantum Coherence		Insect Cell Sample	Cell free
Complex points (F2 x F1)	2048 x 104	2048 x 104	2048 x 128	4096 x 2048	512 x 206
Acquisition Time (ms)	98.3 and 21.5	98.3 and 21.5	49.1 and 26.5	196.7 and 63.6	113.3 and 27.3
# Scans	200	80	80	16	336
Recycle Delay (s)	1.2	1.2	1.2	1	1
Experiment Time (hh:mm)	74:56	30:00	47:13	11:20	23:55

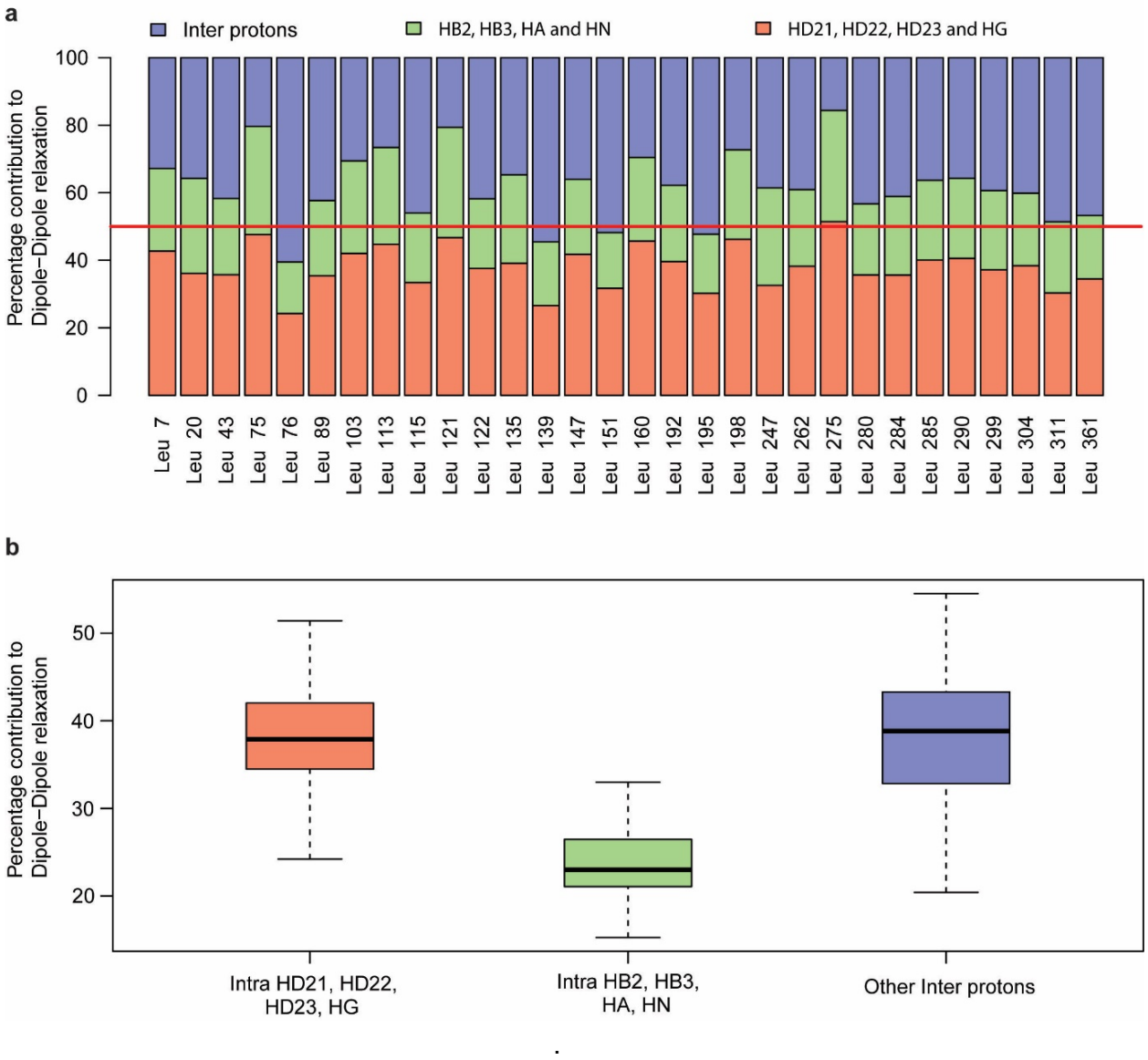
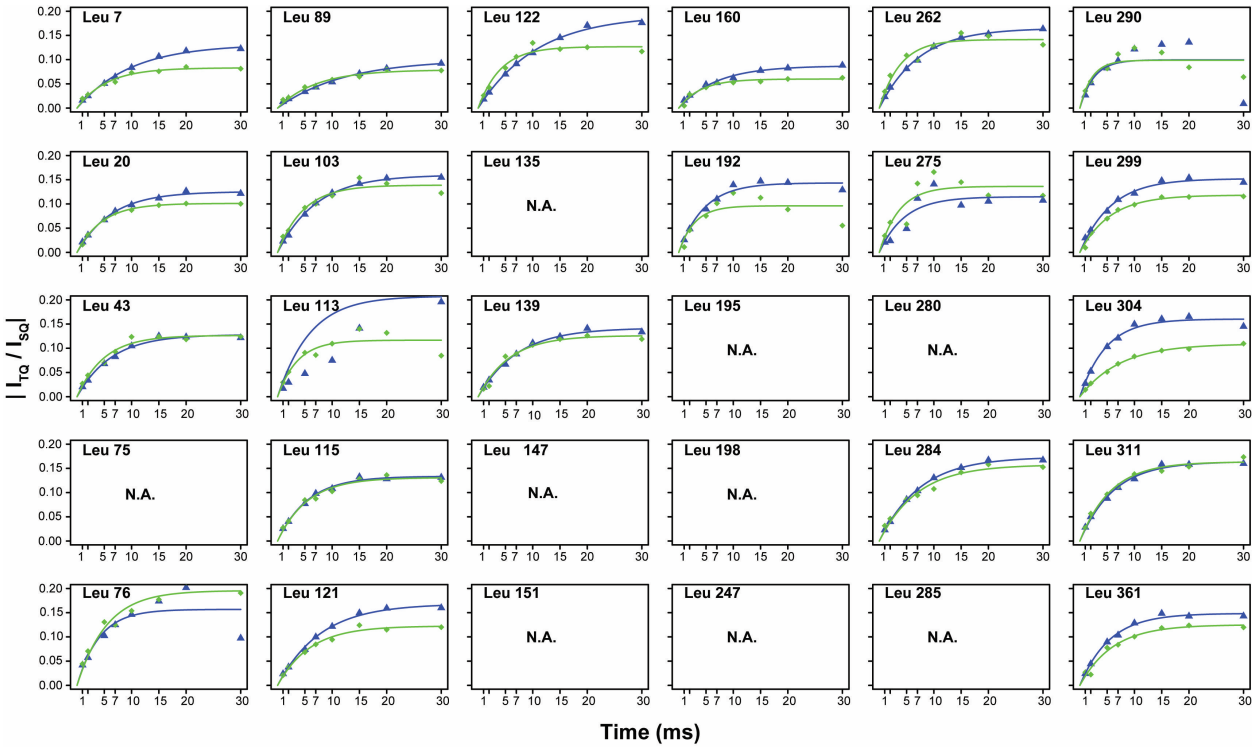
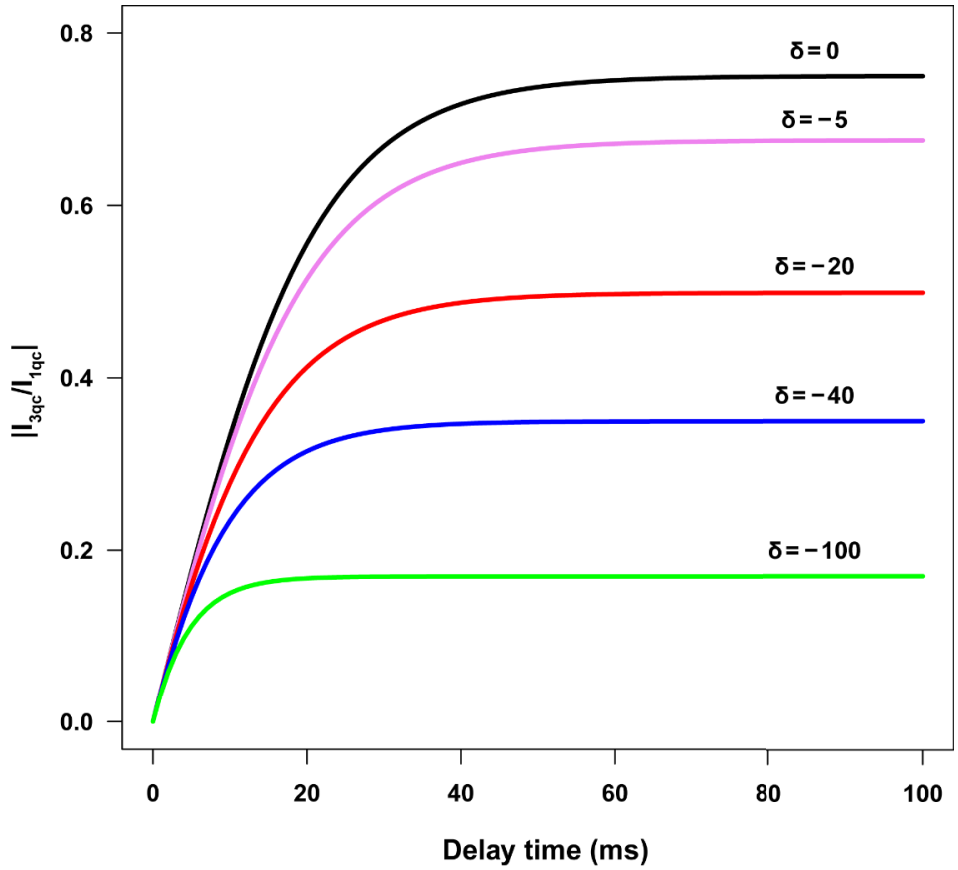


Figure A24: Estimation of the contribution of protons to the dipole-dipole relaxation of leucine methyl protons in MBP. (a) The stacked bars show the dipole-dipole contribution of three categories of protons to each leucine proton attached to Cδ1 in MBP. The orange bar is for intra-residue HD21, HD22, HD23 and HG atoms of leucine. The green bar is for intra-residue HB2, HB3, HA and HN atoms of leucine. The blue bar is for the rest of inter-residue protons. The red line indicates 50%. (b) The statistical summary of the bar plot in (a) presented for 30 Leu residues in MBP as box plot.



.

Figure A25: The Forbidden Coherence Transfer curve for all thirty Leu in MBP. The blue curve is FCT measured on the protein prepared using Leu-meth$_{LD}$ and green curve is FCT measured on fully protonated protein. 'N.A.' depicts data not available for the particular leucine either due to inability to unambiguously assign the peaks or inability to fit the FCT curve due to poor signal to noise ratio. The curves for Leu 76, and Leu 275 is unexpected and we attribute to peak integration errors due to poor sensitivity of triple quantum FCT experiment.



.

Figure A26: Simulation of FCT build-up curves. All the curves were simulated with identical parameters in equation S1, except for the values of δ. The δ value is indicative of the protons contributing towards the dipole-dipole relaxation of the methyl protons. The fitted value of δ on experimental data can be used for relative comparison of surrounding proton density.

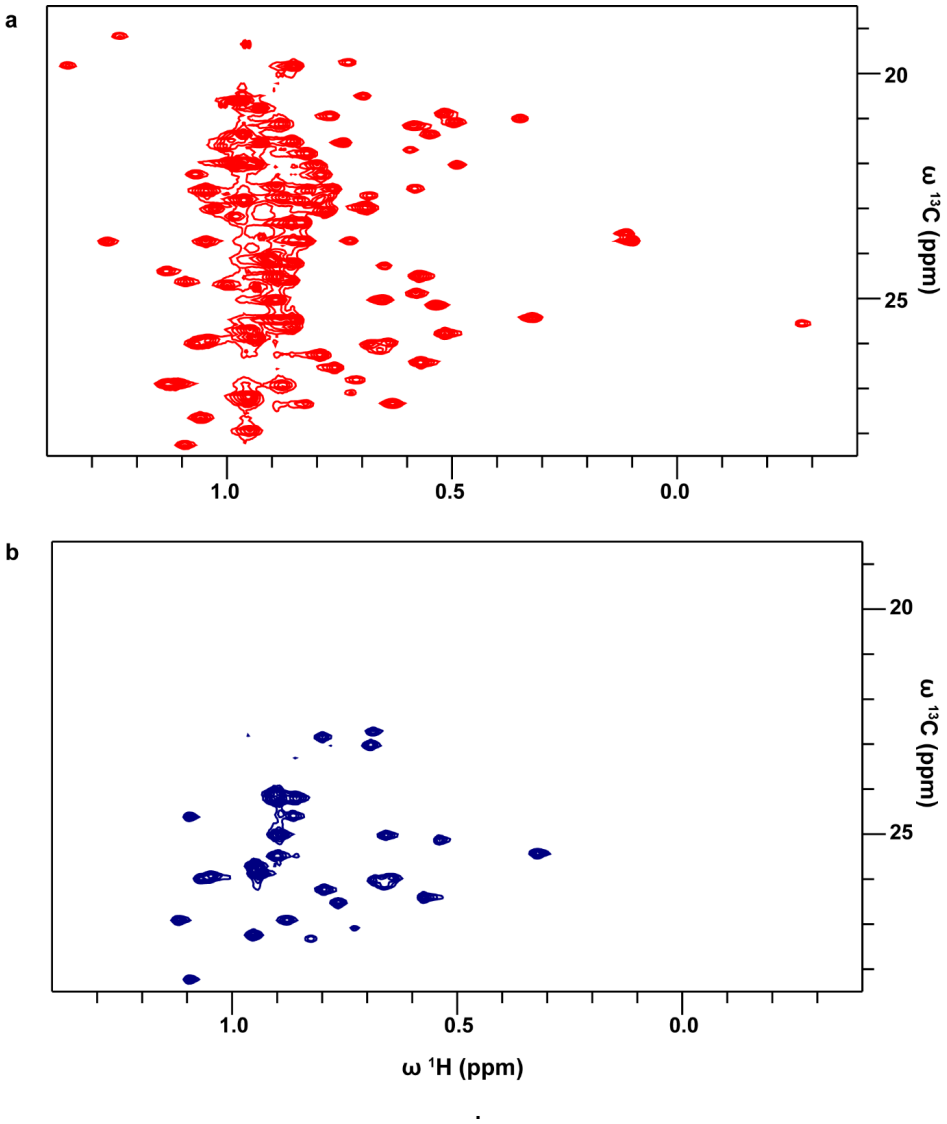


Figure A27: 2D [^{13}C,^{1}H]-HSQC of MBP expressed in *E.coli* using (a) ^{13}C-acetolactate as precursor, which uniformly labels Leu and Val methyl and (b) Leu-meth$_{LD}$.

Supplementary Information for section 4.3

Hormones of the Melanocortin Receptor-4 System

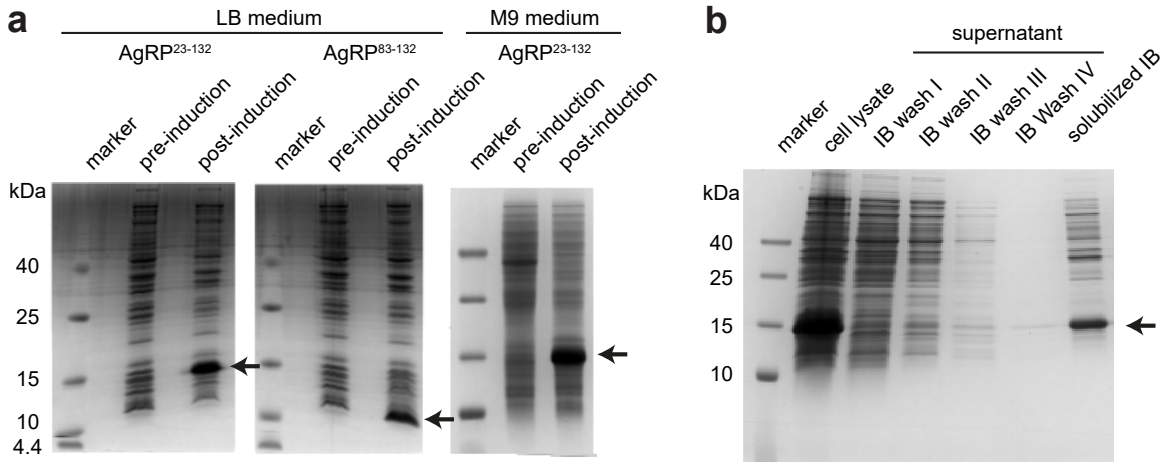


Figure A28: Expression and solubilized of AgRP. (a) SDS-PAGE for pre- and post-induction culture of $AgRP^{23-132}$ (left) and $AgRP^{83-132}$ in LB medium (middle), as well as $AgRP^{23-132}$ in $^{15}NH_4Cl$-supplemented M9 minimal medium (right). (b) SDS-PAGE of $AgRP^{23-132}$-overexpressing cell lysate, supernatant of different washing steps with lysate buffer (I, II) and high-salt buffer (III, IV), as well as solubilized $AgRP^{23-132}$ in 6 M GdnHCl and 50 mM DTT. Arrows indicated the height of AgRP.

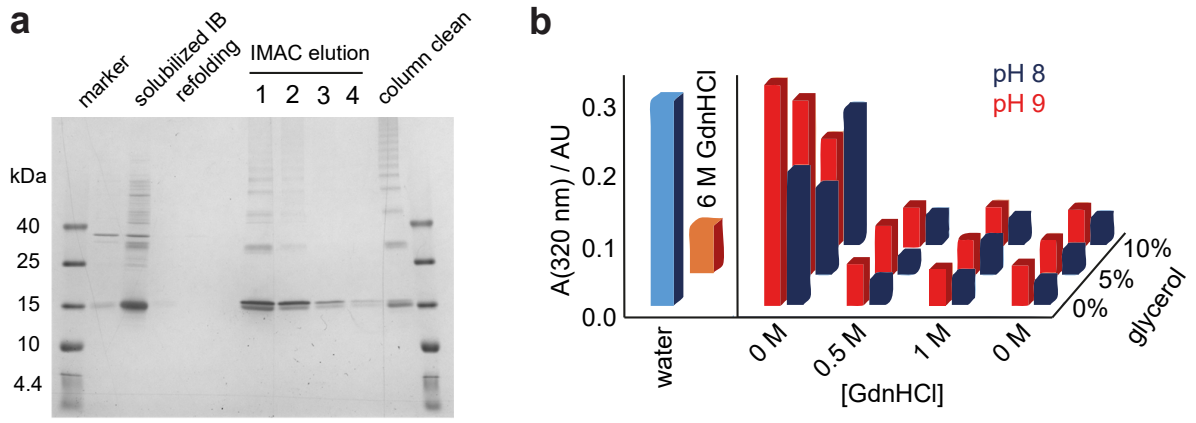


Figure A29: Refolding of AgRP$^{23\text{-}132}$. (a) SDS-PAGE of solubilized AgRP$^{23\text{-}132}$ before and after refolding. Refolding solution was concentrated by IMAC and four elution steps with 500 mM imidazole. Un- and misfolded AgRP$^{23\text{-}132}$ (lower band) shows higher affinity to the column resin and can be eluted with 1 M imidazole and 6 M GdnHCl (column clean). (b) Refolding assay was performed by measuring absorbance at 320 nm with refolding buffer containing different concentrations of GdnHCl and glycerol at pH 8 (blue) and pH 9 (red). Pure water (light blue) and solubilizing buffer (6 M, orange) serve as controls.

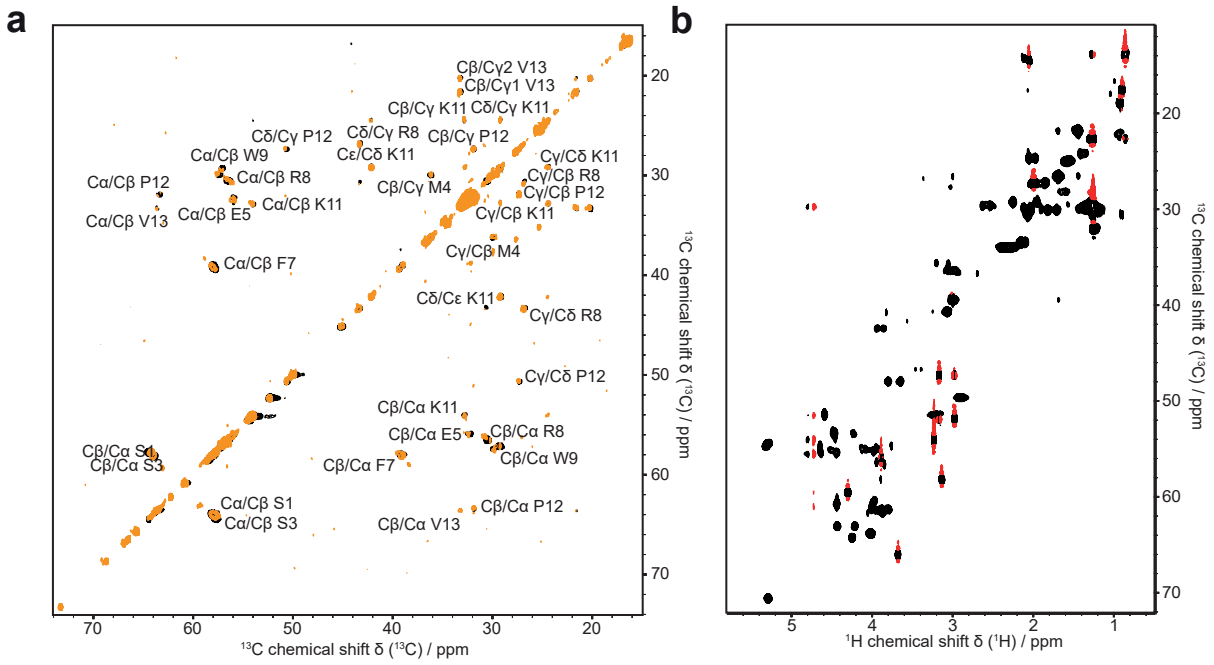


Figure A30: Carbon assignment of α-MSH. A. [^{13}C,^{13}C]-FLOPSY spectra of free isotope-labeled α-MSH (black) and in presence of nanodisc containing 50% POPC/ 50% POPS (orange). Assigned carbon crosspeaks are indicated. B. [^{1}H,^{13}C]-HSQC spectrum of free isotope-labeled α-MSH.

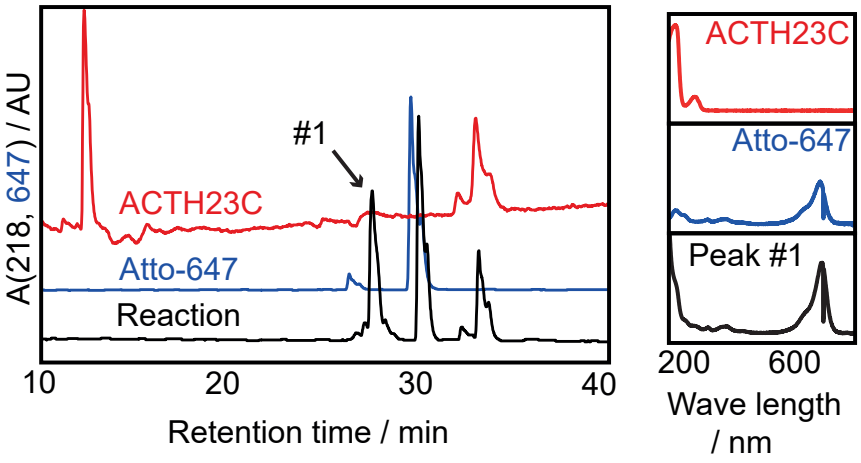


Figure A31: Atto647 labeling of ACTH23C. RP-HPLC runs (left) of ACTH23C (red), Atto647 (blue), and reaction mix (black). The additional elution peak #1 represents labeled ACTH23C-Atto647, and shows absorbance signals (right) for fluorophore (647 nm) and peptide backbone (190 nm).

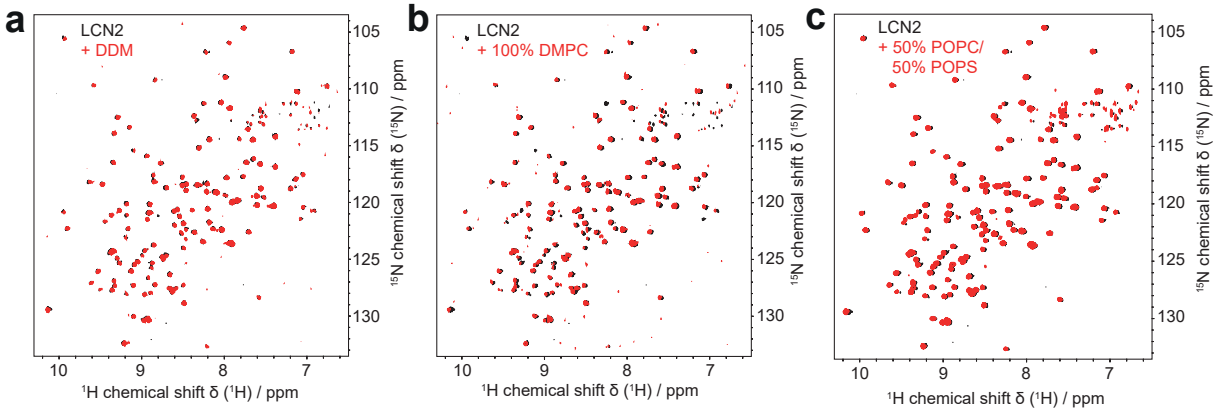


Figure A32: hLCN2 does not interact with membrane mimetics. [^{1}H,^{15}N]-TROSY spectra of 25 µM ^{15}N-hLCN2 alone (a-c, black) and in presence of 50 mM DDM (a, red), 50 µM 100% DMPC nanodiscs (b, red), and 50 µM 50% POPC/ 50% POPS (c, red).

Glossary

1QC	Single Quantum Coherence
3QC	Triple Quantum Coherence
ABC	ATP binding cassette
ACTH	Adrenocorticotropic hormon
ADP	Adenosine diphosphat
AgRP	Agouti-related protein
α-MSH	α-melanocyte stimulation hormone
ARF1	ADP ribosylation factor 1
ARNO	ADP ribosylation factor nucleotide binding-site opener
ATP	Adenosine triphosphat
B_0	External magnetic field
BHQ	Black hole quencher
B_{loc}	Local magnetic field
BMRB	Biological Magnetic Resonance Bank
bO	Bacterioopsin
BPh	Base-phosphate interaction
bR	Bacteriorhodopsin
BSA	Bovine serum albumin
C3 spacer	Three carbon unit spacer
CaM	Calmodulin
cAMP	Cyclic adenosine monophosphate
CAPER	Co-activator of AP-1 and ER
CFE	Cell-free expression
CPMG	Carr-Purcell-Meiboom-Gill
CSI	Chemical shift index
CSP	Chemical shift perturbation
CV	Column volume
D	Dipolar coupling constant
d2	Fluorescent dye d2
dA	Deoxyribose adenosine
Da	Tensor magnitude
dC	Deoxyribose cytosine
DDM	*n*-Dodecyl-b-D-maltoside
DDR	Dipole-dipole relaxation
δ	Chemical shift / ppm
$\Delta\delta_{comb}$	Combined chemical shift
dG	Deoxyribose guanosine
DLS	Dynamic light scattering
DMPC	1,2-Ditetradecanoyl-sn-glycero-3-phosphocholine
DMPG	1,2-Dimyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
DNA	Deoxyribonucleic acid
DNA^c	Complementary DNA to $RNA^{2'F}$
DPC	Dodecylphosphocholine
DSS	4,4-Dimethyl-4-silapentanesulfonic acid
dT	Deoxyribose thymidine
DTT	Dithiothreitol
Dz	PrP-mRNA targetign 10-23 DNAzyme
$Dz^{5A,5C,5G}$	Original Dz sequence (5A) and with C, G mutation at loop position 5
Dz^{6xF}	Dz with six fluoro substitutions
$Dz^{Edu-8,8,+7}$	Dz with EdU modification at position -8, 8, +7
E. coli	*Escherichia coli*
E.COSY	Exclusive correlation spectroscopy
EDTA	Ethylenediaminetetraacetic acid
EdU	5-Ethynyl-2'-deoxyuridine
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EGTA	Ethylene glycol-bis(β-aminoethylether) tetraacetic acid
eNOE	exact NOE
eNORA	exact NOE by Relaxation matrix Analysis
ERK	Extracellular signal-regulated kinase
ESI-MS	Electrospray ionization mass spectrometry
Et	Ethyl
Et_2OAc	Ethyl acetate
FAM	Fluorescein
FCT	Forbidden coherence transfer
FFT	Fast Fourier-transformation
FID	Free induction decay
FLOPSY	Flip-flop spectroscopy
FRET	Förster resonance energy transfer
FT	Fourier-transformation

g	Electron spin g factor
γ	Gyromagnetic ratio
$\Gamma 1$	Longitudinal paramagnetic relaxation enhancement
$\Gamma 2$	Transverse paramagnetic relaxation enhancement
GdnHCl	Guanidine hydrochloride
GPCR	G-protein coupled receptor
GSH	Monomeric glutathione
GSSG	Dimeric glutathione
h	Planck's constant
$\hbar$	$h/2\pi$
HEK293F	Human Embryonic Kidney 293F cells
HEPES	4-(2-Hydroxyethyl)-1-piperazine-ethanesulfonic acid
hLCN2	human LCN2
HMQC	Heteronuclear multiple bond correlation
HOESY	Heteronuclear NOESY
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
I	Spin quantum number
IBMX	3-Isobutyl-1-methylxanthine
ICD	EGFR intracellular domain
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl b-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetric
I_z	z-Magnetisation of nucleus I
J	Scalar coupling constant
JM	Juxtamembrane
JM_{SC}	Scrambled JM variant
K_D	Binding coefficient
k_{ex}	Exchange rate
λ	Linewidth
LCN2	Lipocalin 2
LED	Light-emitting diode
Leu-meth$_{LD}$	selectively ^{13}C-labeled Leucine methyls with partial deuteration
LRMS	Low resolution mass spectrometry
m	Magnetic quantum number
m/z	Mass-to-charge ratio
μ_B	Bohr magneton
MBP	Maltose binding protein
MC4R	Melanocortic-4 receptor
mLCN2	mouse LCN2
MST	Microscale thermophoresis
MWCO	Molecular weight cut-off
M_{xy}	Magnetisation in transverse plane
M_z	Magnetisation in z-direction
N_α	Spin population in α state
N_β	Spin population in β state
NDP	Norleucin and D-phenylalanin modification
NDs	Nanodiscs
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NTA	Nitrilotriacetic acid
ν_0	Larmor frequency / Hz
NUS	Non-uniform sampling
ω_0	Larmor frequency / rads^{-1}
P	Pseudo-rotation phase Population of state(s)
pbr	Polybasic region
PBS	Phosphate-buffered saline
PDB	Protein database
Ph	Phenyl
PH	Pleckstrin homology domain
Phim	Pseudo-rotation amplitude
PIPs	Phosphoinositides
PMSF	Phenylmethylsulfonyl fluoride
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
ppm	parts per million

PRE	Paramagnetic relaxation enhancement
R	Tensor rhombicity
R_1	Longitudinal relaxation rate
$R_{1\rho}$	Relaxation time under spin-lock conditions
R_2	Transverse relaxation rate
rA	Ribose adenosine
rC	Ribose cytosine
RDC	Residual dipolar coupling
RED-NHS	Amine reactive fluorescent dye
RF	Radio frequency
rG	Ribose guanosine
ρ	Auto-correlation rate
r_{IS}	Internuclei distance
rMD	Restraint molecular dynamic
RNA	Ribonucleic acid
$RNA^{2'F}$	Target RNA with fluoro stabilisation
RNA^{FAM}	Target RNA with fluorescein label
RNA^{FRET}	Target RNA with fluorophore-quencher labeling
RRM2	Ribonucleoside-diphosphate reductase subunit M2
RP	Reversed-phase
RT	Real-time Room temperature
rU	Ribose uridine
S^2	Ordo parameter
SAIL	Stereo-array isotope labeling
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate poly-acrylamide gel electrophoresis
SEC	Size-exclusion chromatography
Sec7	Guanine nucleotide exchange factor domain of ARNO
σ	cross-relaxation rate
SP	Signale peptide
STD	Saturation transfer difference
S_z	z-Magnetisation of nucleus S
$S_{z,ss}$	z-Magnetisation of heteronucleus under steady-state conditions
T_1	Longitudinal relaxation time
T_2	Transverse relaxation time
τ_C	Rotational correlation time
TBE	Tris-borate EDTA buffer
TBTA	Tris((1-benzyl-4-triazolyl)methyl)amine
TCEP	Tris(2-carboxyethyl)phosphine
TCRs	T-cell receptors
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TM	Transmembrane domain
t_{mix}	Mixing time
TMS	Tetramethylsilane
*TNAO*38	*Trichoplusia ni* cell line
TOCSY	Total correlation spectroscopy
TRAF4	Tumor necrosis factor receptor associated factor 4
Tris	Tris(hydroxymethyl)aminomethane
TROSY (Tr)	Transverse relaxation optimized spectroscopy
W^{0IS}	Zero-quantum transition between spins I and S
$W^{1I,S}$	Single-quantum transition of spin I, S
W^{2IS}	Double-quantum transition between spins I and S
WC	Watson-Crick
YopH	Tyrosine-protein phosphatase

Acknowledgments

I thank Dr. Manuel Etzkorn for the excellent supervision and Prof. Dr. Gohlke for his mentorship. I also thank all colleagues of the Institut für physikalische Biology at Heinrich-Heine university and the Institut für Biologische Ingormationsprozesse (IBI-7) at the Forschungszentrum Jülich (FZJ) for the support and friendly working atmosphere. Especially acknowledged are Dr. Julian Victor, Dr. Hannah Rosenbach and Dr. Aldino Viegas for their deep and helpful engagement in the DNAzyme project.

I thank Prof. Dieter Willbold and the FZJ for granting access to the NMR spectrometers and Dr. Rudolf Hartmann and Kevin Bochinsky for the technical support.

Furthermore, I thank my family and friends for the patience and encouragement.

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other German or foreign examination board.

Date Signature