

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Biomolekulare NMR-Spektroskopie,
Department Chemie

NMR Investigation of Protein-RNA Interactions Involved in Ribosomal Gene Silencing.

Irina Anosova

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Univ. – Prof. Dr. St. Sieber

Prüfer der Dissertation: 1. Univ. – Prof. Dr. M. Sattler

2. Univ. – Prof. Dr. P. Cramer,
Ludwig – Maximilians – Universität München

Die Dissertation wurde am 21.12.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 21.03.2012 angenommen.

ABSTRACT.....	3
INTRODUCTION.....	5
1. Introduction I. Biological background.....	5
1.1. Regulation of eukaryotic gene expression.....	6
1.2. Pol I transcription machinery.....	7
1.3. Organization of rRNA genes.....	9
1.4. Regulation of rRNA transcription.....	10
1.5. NoRC – the key epigenetic repressor.....	12
1.6. Regulation of NoRC function. The pRNA.....	14
1.7. TIP5 protein.....	17
1.8. TAM domain of TIP5.....	18
2. Introduction II. Solution state NMR Spectroscopy.....	23
2.1. Nuclear magnetic resonance spectroscopy.....	24
2.2. Basic principles of NMR.....	24
2.3. NMR spectroscopy experimental setup.....	32
2.4. Calculating structures in solution.....	39
2.5. Protein dynamics by NMR.....	45
2.6. Determination of molecular size by NMR.....	46
2.7. Ligand binding studies by NMR.....	46
SCOPE OF THE THESIS.....	51
MATERIALS AND METHODS.....	53
3. Materials.....	53
3.1. Sequences.....	54
3.2. Chemicals.....	54
3.3. Enzymes.....	56
3.4. Bacterial strains and vectors.....	56
3.5. Recipes.....	57
4. Methods I. Molecular Biology and Biochemistry.....	59
4.1. Cloning, expression and purification of <i>H. sapiens</i> TIP5 constructs.....	60
4.2. Alignment media for measurement of residual dipolar couplings (RDCs).....	65
4.3. Attachment of a paramagnetic relaxation enhancement spin label.....	69
4.4. <i>In vitro</i> RNA transcription.....	69
4.5. Complex formation of TIP4N and 1wt RNA.....	72
4.6. Circular dichroism.....	72
5. Methods II. Nuclear Magnetic Resonance Spectroscopy.....	73
5.1. Assignment of backbone resonances of the TIP4N protein construct.....	74
5.2. Resonance assignment and structure determination of the TAM ₅₁₆₋₆₂₃	74
5.3. Calculation of secondary chemical shifts.....	76
5.4. Conservation analysis of the TAM domain residues.....	76
5.5. Relaxation and dynamics analysis of the proteins.....	76
5.6. Residual dipolar couplings.....	78
5.7. Paramagnetic relaxation enhancement analysis.....	79
5.8. NMR diffusion measurements.....	80
5.9. Hydrogen to deuterium exchange experiments (H-D exchange).....	80
5.10. Mapping of the pRNA binding. NMR chemical shift perturbation assays.....	81
5.11. NMR spectroscopy of pRNA constructs.....	82
5.12. Multiple sequence alignment.....	83

RESULTS.....	85
6. Results I. Defining TAM domain boundaries.....	85
6.1. Multiple sequence alignment.....	86
6.2. NMR studies of TIP4N construct.....	88
6.3. NMR relaxation studies of TIP4N.....	89
6.4. Diffusion coefficient measurements by NMR.....	91
6.5. TAM domain constructs.....	92
6.6. EMSA of TIP5 truncation constructs.....	93
7. Results II. Solution structure of the TAM domain from <i>H. sapiens</i>.....	95
7.1. NMR ensemble of structures.....	96
7.2. Electrostatic potential.....	101
7.3. Residual dipolar couplings.....	101
7.4. Structure validation using paramagnetic relaxation enhancement (PRE).....	104
7.5. H-D exchange.....	106
7.6. NMR relaxation analysis.....	107
7.7. Residue conservation.....	108
8. Results III. NMR studies of pRNA.....	111
8.1. Defining minimal TAM interaction sequence.....	112
8.2. Structural studies of 1wt construct.....	116
8.3. The “short” construct.....	119
9. Results IV. pRNA interaction with the TAM domain.....	123
9.1. NMR titration of TIP4N with 1wt RNA.....	124
9.2. Mutational analysis I, based on residue conservation.....	125
9.3. Complex reconstitution of TIP4N/1wt.....	128
9.4. Mutational analysis II, based on titration results 9.3.....	130
9.5. Structural integrity of TAM domain. TIP5WY/GA mutant.....	133
9.6. Short conclusion.....	136
10. Discussion.....	137
10.1. Structure of the human TAM domain.....	138
10.2. pRNA.....	140
10.3. pRNA/TAM interaction.....	141
10.4. DNA/TAM interaction.....	145
10.5. pRNA mediated gene silencing.....	147
REFERENCES.....	151
APPENDIX.....	165
ABBREVIATIONS I.....	181
ABBREVIATIONS II.....	182
LIST OF FIGURES.....	183
LIST OF TABLES.....	185
ACKNOWLEDGEMENTS.....	187
CURRICULUM VITAE.....	189

Abstract.

NMR Investigation of Protein-RNA Interactions Involved in Ribosomal Gene Silencing.

Genes encoding ribosomal RNA (rRNA) exist in two distinct states of chromatin, an ‘open’ one that is permissive to transcription and a ‘closed’ one that is transcriptionally refractive. Even in metabolically active mammalian cells a significant part of the tandemly repeated rRNA genes is epigenetically silenced. Silencing of rDNA is mediated by NoRC, a SNF2h-containing chromatin remodeling complex with TIP5 as its largest subunit, which recruits chromatin-modifying activities to the rDNA promoter, thereby triggering heterochromatin formation. The evolutionary conserved TAM domain of TIP5 shows sequence homology to MBD domains of methyl-CpG-binding proteins. However, in contrast to MBD domains, TIP5 TAM domain specifically binds to a phylogenetically conserved RNA hairpin. This small non-coding pRNA (promoter RNA) is produced from Pol I promoters in the intergenic spacer regions (IGS) that separate rDNA repeats, and matches the rDNA promoter sequence. Interaction of TIP5 with pRNA was shown to be required for nucleolar localization of NoRC and heterochromatin formation in the rDNA regions. Thus, the TAM/pRNA interaction plays a major role in the regulation of eukaryotic gene expression. However, the molecular details of this interaction have not been described yet. The goal of the thesis was to perform structural and biochemical studies of the TAM/pRNA interaction, by starting from an analysis of the individual binding partners.

In Chapter 1 of this thesis the biological background of rDNA silencing and the role of the NoRC complex are presented. Chapter 2 reviews the basic principles of NMR spectroscopy and its application for structure calculation and investigation of biomolecules in solution. Chapters 3-5 present a summary of materials and methodology used for biochemical experiments and structural analysis. In Chapter 6 the boundaries of the functional TAM domain of human TIP5 are defined based on sequence analysis of homologous MBD and TAM domains and NMR and biochemical studies of the derived constructs. Chapter 7 presents the three-dimensional solution structure of the TIP5 TAM domain. The structure was solved using state-of-the-art NMR methods and examined using information from paramagnetic relaxation enhancement and hydrogen-deuterium exchange. The conformational dynamics of the TAM domain was analyzed using ^{15}N NMR relaxation experiments. In Chapter 8 the minimal pRNA sequence sufficient for TIP5 interaction and

suitable for NMR studies is defined and investigated by NMR and biochemically. In Chapter 9 the TAM/pRNA interaction is characterized using NMR titrations, biochemical experiments and mutational analysis. The pRNA binding interface was mapped onto the structure of the TAM domain. Residues that are crucial for the TAM/pRNA interaction were identified by mutations and tested experimentally in electrophoretic mobility shift and filter binding assays. Biochemical and NMR studies of a previously reported TIP5 mutant, deficient in pRNA binding are shown as well. In Chapter 10 the results of the thesis are discussed.

The results of this thesis show, that the TAM domain adopts a novel variant of the MBD-domain fold, which is extended by additional structural elements at the N- and C-termini. The 35 nucleotide stem of pRNA is sufficient to promote TIP5 binding *in vitro*. NMR analysis suggests that the free RNA is dynamic and may adopt multiple conformations. Therefore, further optimization of the RNA construct will be required for structural analysis by NMR. The most distinct feature of the TAM domain structure is an extended antiparallel β -sheet, which is not conserved within the MBD domains, but mediates the pRNA binding to the TAM domain. The TAM/pRNA interaction is dependent on the residues W546 and R545, as point mutation of these amino acids to alanine abolishes the pRNA binding activity of TAM or considerably decreases its specificity to pRNA respectively. In contrast, the double mutation W551G/Y552A, which has been previously shown to abolish the silencing activity of TIP5, strongly affects the tertiary fold of the TAM domain as indicated by NMR and biochemical analysis, thus demonstrating the importance of the native TAM domain fold for its functional activity.

The structural and biochemical data presented in this thesis provide novel insight into the molecular function of the TIP5 TAM domain and indicate a novel mode of RNA binding by the TAM domain fold, distinct to the DNA binding by MBD domains. Given that the TAM/pRNA interaction is important for the functional activity of TIP5 and the NoRC for heterochromatin formation and transcriptional silencing of eukaryotic genes, the results provide new insight into molecular details of the processes involved in genome stability and cancer development.

Introduction.

1. Introduction I. Biological background.

1.1. Regulation of eukaryotic gene expression.

A living cell, as long as it's alive, is an ever functioning unit. Even if finally differentiated, the cell exists in a dynamic equilibrium. Numerous processes run in parallel to ensure proper cell functions, signaling, defense, energy supply and response to external stimuli. Metabolic processes occurring at the very same moment make up a complex network that ensures cell survival and fulfillment of its function. These processes are mediated by proteins and, consequently, are dependent on protein synthesis and on cell supply with ribosomes.

All cells from prokaryotes to vertebrates produce around 1-2 million ribosomes per cell cycle. This consumes a tremendous amount of energy and is adapted to various needs of the cell. It is linked to cell growth and proliferation on one hand and tightly bound to the intracellular available energy on the other (Moss et al, 2007; Murayama et al, 2008).

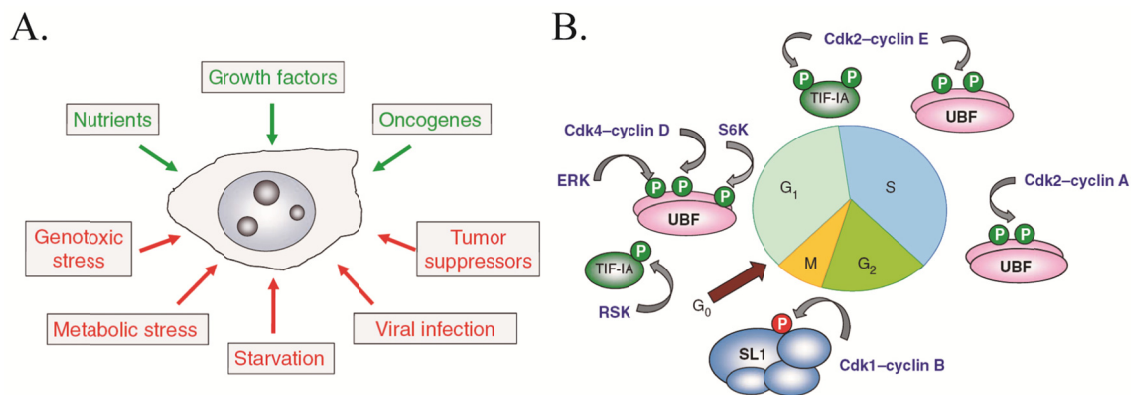


Fig. 1 Regulation of rRNA synthesis. (A) Extracellular signals that have an effect on rRNA synthesis, converging at Pol I. Signaling pathways that upregulate Pol I transcription are shown as green arrows; pathways downregulating Pol I transcription rate are presented as red arrows. (B) Regulation of Pol I transcription during the cell cycle progression. In the G₁-phase and S-phase, UBF (upstream binding factor) is activated by phosphorylation of several serines by Cdk4-cyclin D and S6K and Cdk2-cyclin E/A respectively. In addition, mTOR-dependent and ERK-dependent pathways activate TIF-IA by a serine - phosphorylation. At entry into mitosis, Cdk1-cyclin B phosphorylates the TAF_I-TBP complex SL1 at a threonine, leading to repression of Pol I transcription during mitosis. At the exit from mitosis, Cdc14B dephosphorylates SL1, leading to recovery of its activity. Activating phosphorylations are marked in green and inhibiting ones in red. Adapted from (Grummt, 2010).

Since synthesis of rRNA is the first step in this energy consuming process, it is very tightly regulated. Almost all signaling pathways, that effect cell growth and proliferation, directly regulate rRNA synthesis, their downstream effectors often converging at the DNA dependent polymerase Pol I machinery (Fig. 1A). For example, cdc2/cyclin B dependent

phosphorylation of TAF_I, a subunit of SL1 that nucleates preinitiation complex assembly causes shutdown of rDNA transcription during mitosis and the phosphatase Cdc14B resets it in the interphase (Kuhn et al, 1998; Mailand et al, 2002). Upon mitogenic stimulation the extracellular signal regulated kinase ERK phosphorylates UBF (upstream binding factor) (Fig. 1B), an important component of the Pol I complex at two residues, increasing its interaction with rDNA and facilitating Pol I transcription through chromatin (Stefanovsky et al, 2006; Stefanovsky et al, 2001). Also mTOR, a key regulator of cell growth and proliferation, is acting on UBF, stimulating Pol I activity through phosphorylation (Hannan et al, 2003; Stefanovsky et al, 2001). Pol I machinery responds as well to products of oncogenes and tumor suppressors. For review see (Drygin et al, 2010; Grummt, 2010).

Minor mistakes in this tight regulation of Pol I mediated rRNA synthesis process lead to severe failures in cell functionality and to development of cancer. Mutations in signaling pathways that lead to excessive ribosome synthesis to support the proliferation of malignant cells are often detected in case of cancer (Montanaro et al, 2008). Loss of regulation at rDNA loci is correlated with rDNA instability, nucleolar disintegration and cellular senescence (Kobayashi et al, 2004; Peng & Karpen, 2007). Indeed, recent evidence suggest that deregulation of rRNA synthesis may be among the most important alterations in a cancer cell (Ghoshal et al, 2004; White, 2008; Williamson et al, 2006).

1.2. Pol I transcription machinery

Transcription initiation at rDNA genes is a complex process that requires assembly of a specific multiprotein complex – the preinitiation complex - at the rDNA promoter. It contains Pol I and a large number of associated factors that promote Pol I mediated transcription (Fig. 2A). For review see (Drygin et al, 2010; Grummt, 2010).

In mammals the assembly of the preinitiation complex is mediated by the action of two Pol I specific transcription factors, that bind the rDNA promoter, the UBF (upstream binding factor) and the SL1 (promoter selectivity factor, also termed TIF-IB in mice). UBF is a member of the high mobility group (HMG) protein family and contains five HMG boxes. Since this motif is known to interact with and bend DNA, the HMG boxes of UBF enable it to loop ca. 140bp of DNA into a single turn, forming a nucleosome-like structure called enhanceosome. UBF further activates rDNA transcription by recruiting Pol I to the rDNA promoter and stabilizing specific binding of transcription factors, like SL1/TIF-IB (Kuhn & Grummt, 1992)

while simultaneously displacing nonspecific DNA binding proteins such as histone H1 (Kuhn et al, 1993). UBF was also shown to regulate promoter escape of Pol I (Panov et al, 2006) and transcription elongation (Stefanovsky et al, 2006).

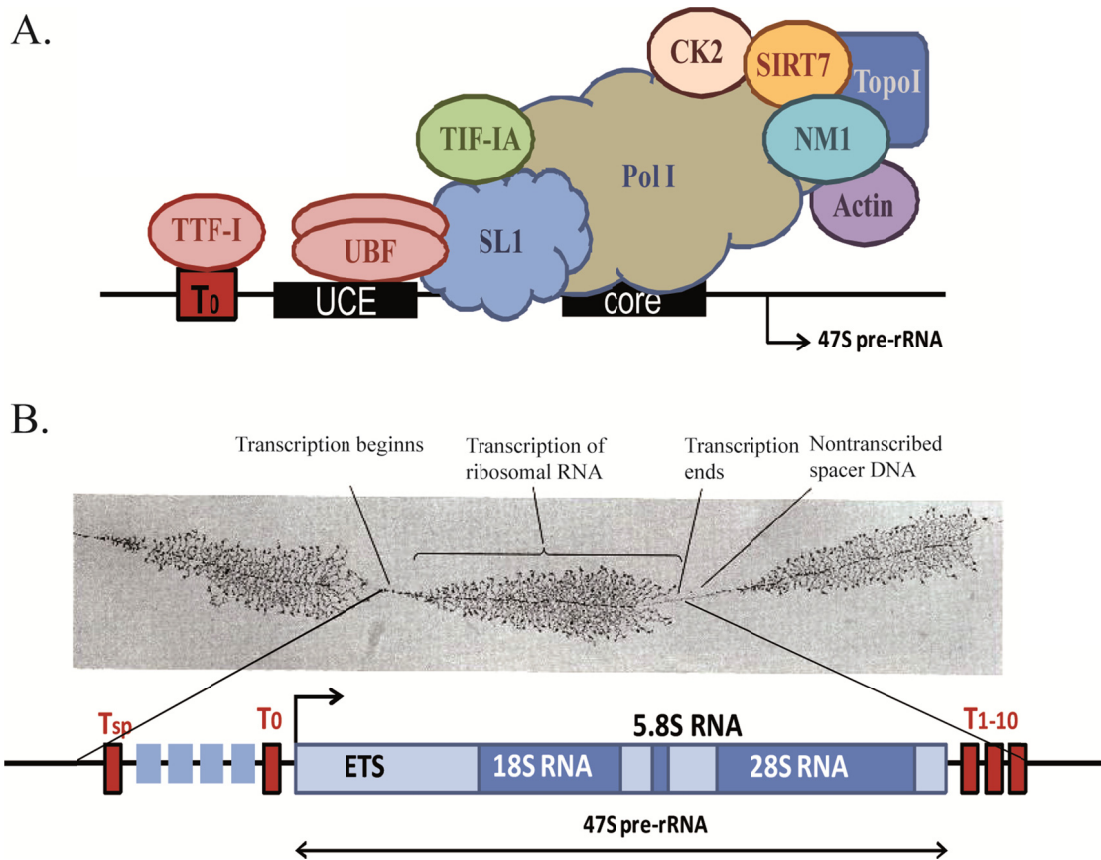


Fig. 2 Expression of rRNA genes. (A) Basal factors required for Pol I transcription initiation. The site of transcription initiation of 47S pre-rRNA is indicated as black arrow. The upstream control element (UCE) and the core rDNA promoter are depicted as black boxes. Binding site for the transcription termination factor (TTF-I) upstream of the gene promoter (T_0) is shown as red box. Factors that are associated with the rDNA promoter and Pol I are depicted by ellipsoids. (B) Organization of the rRNA genes. An electron microscopic image showing active amphibian rRNA genes. It reveals the tandem head-to-tail arrangement of rRNA genes that are separated by ‘nontranscribed spacers’ and the characteristic Christmas tree appearance of active transcription units. From (Miller & Beatty, 1969). A close up of a single rDNA repeat. The coding sequence for the long 47S precursor rRNA including the sections coding for the 18S, 5.8S, 28S rRNA and the external transcribed spacer (ETS) is shown. The site of transcription initiation of 47S pre-rRNA is indicated by a black arrow. Binding sites for the transcription termination factors located downstream of the transcription unit (T_{1-10}), downstream of the spacer promoter (T_{sp}) and upstream of the gene promoter (T_0) are indicated by red boxes. Repetitive enhancer elements located between the spacer promoter and major gene promoter are indicated by blue boxes. After (Grummt, 2010).

SL1/TIF-1B determines the promoter specificity of Pol I. It is a complex of proteins consisting of the TATA binding protein (TBP) and four TBP-associated factors (TAF_{IS}).

TAF_I subunits mediate specific interactions between the rDNA promoter and Pol I, thereby guiding Pol I - together with associated factors - to rDNA. SL1 acts synergistically with UBF and formation of a functional preinitiation complex requires interaction of SF1 with UBF (Bell et al, 1988; Hempel et al, 1996; Tuan et al, 1999).

Furthermore, in mammals Pol I exists in two distinct forms and only Pol I β is able to assemble into productive transcription initiation complexes. In addition to UBF and SL1, it is further associated with numerous proteins, including growth-dependent transcription initiation factor TIF-IA, protein kinase CK2, nuclear actin and nuclear myosin 1, chromatin modifiers PCAF, SIRT7, as well as proteins involved in DNA repair and replication and Cockayne syndrome group B protein (CSB) (Bradsher et al, 2002).

The commonly discussed mechanism of transcription suggests that Pol I is recruited to the rDNA promoter as a huge complex, that coordinates rRNA synthesis and maturation, chromatin modification and DNA repair. Studies show that Pol I machinery is highly dynamic, assembling sometimes individually and sometimes in subcomplexes (Dundr et al, 2002). Quantitative single cell imaging reveals that upregulation of transcription is accompanied by prolonged retention of Pol I transcription factors at the rDNA promoter (Gorski et al, 2007), underlining that transcription complex assembly is a decisive step in regulation of expression of the rRNA genes.

1.3. Organization of rRNA genes.

The number of rRNA genes varies greatly among the organisms, from fewer than 100 to more than 10 000. For reviews see (Grummt, 2010; McStay & Grummt, 2008). Eukaryotic rRNA genes (rDNA) are arranged in clusters of tandem repeats, known as nucleolar organizer regions (NORs), which are located on short arms of acrocentric chromosomes; in humans these are chromosomes 13, 14, 15, 21 and 22 (Henderson et al, 1972). The repetitive structure of NOR chromatin makes it very difficult to sequence those regions. However, several techniques demonstrate that NORs contain most probably solely rDNA and suggest that most human nucleolar organizer centers are composed of ca. 70 rDNA repeats (Sakai et al, 1995). The positioning of the NORs isolates them from genes transcribed by Pol II and Pol III. This isolation is reinforced by adjacent heterochromatic repetitive satellite DNA.

Within NORs in mammalian cells, rRNA genes are composed of identical transcription units comprising ca. 43kb in humans and 45kb in mice (Gonzalez & Sylvester, 1995; Grozdanov et al, 2003) (Fig. 2B). Until recently they were thought to be canonically organized as uniform head-to-tail arrays, however, single-DNA molecule analysis showed that they are composed of a mosaic of canonical and non-canonical rDNA repeats, which are arranged head-to-head and comprise up to one third of the array (Caburet et al, 2005). rDNA transcription units harbor ca 13-14kb of coding sequence encoding long precursor rRNA (47S pre-rRNA in humans) that is processed and post-transcriptionally modified to generate one molecule of each 18S, 5.8S and 28S rRNA. The coding sequences are separated by ca. 30kb of intergenic spacers (IGS). Regulatory elements, including gene promoters, repetitive enhancer elements and transcription terminators are located in the IGS. The rDNA promoter includes the upstream control element (UCE) and the core promoter, directly adjacent to the transcription start site, that serve as docking platform for UBF and Pol I respectively. Transcription units are flanked at their 5' and 3' sites by one or more terminator sequences that are recognized by transcription termination factor I (TTF-I), that serves an important role in rRNA transcription regulation (Grummt et al, 1986a; Grummt et al, 1986b; Henderson & Sollner-Webb, 1986).

1.4. Regulation of rRNA transcription.

Given the repetitive nature of rRNA genes, the rRNA synthesis process can be controlled in two ways: by changing the transcription rate of the active genes or by adjusting the number of genes being actively transcribed (Fig. 3). There is evidence for both events. For review see (Moss, 2004; Russell & Zomerdijk, 2006).

1.4.1. Short-term regulation.

Short-term regulation takes place preliminary during cell cycle as response to external stimuli, like nutrient availability, cellular stress, growth factors etc., as well as oncogenes and tumor suppressors. Basal Pol I transcription factors are targeted by multiple signaling pathways. The transcription rate on active euchromatic rDNA is altered by reversible modification of Pol I transcription machinery, the Pol I associated transcription factors. It affects for example the efficiency of transcription initiation, preinitiation complex assembly, promoter escape, transcription elongation and termination.

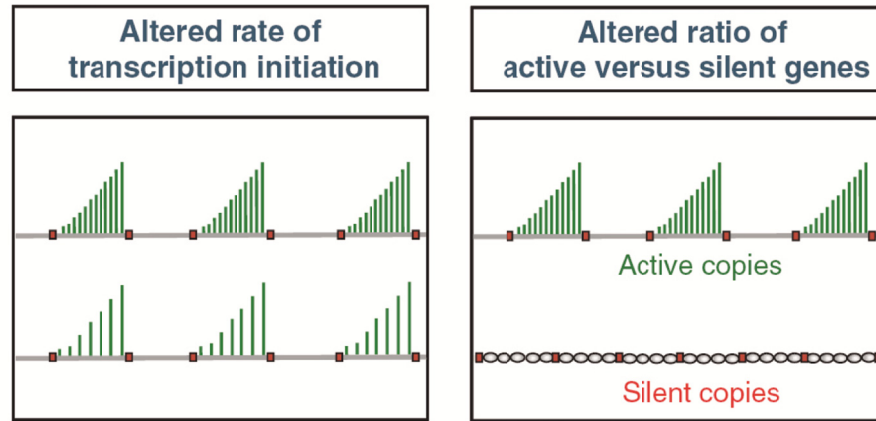


Fig. 3 Two ways of regulation of rRNA synthesis. Left panel: short-term transcription regulation. Cells regulate rDNA transcription by modulating the rate of transcription initiation, thereby controlling the number of nascent pre-rRNA molecules (green lines) that are generated from active genes. Right panel: long-term transcription regulation. Subsets of rDNA repeats are turned either ‘on’ or ‘off’ as required. The gray ellipses indicate the more compact, heterochromatic conformation of silent rRNA genes; the red boxes represent transcription terminator elements located upstream and downstream of the rDNA transcription units. After (Grummt, 2010).

1.4.2. Epigenetic mechanisms.

Long-term regulation during development and differentiation is achieved by epigenetic mechanisms that alter the ratio of active to silent copies of rRNA genes, thereby regulating the number of genes being transcribed. In each cell rRNA genes exist in two distinct epigenetic states and active and inactive nucleolar organizer regions have very distinct chromatin structure.

Active NORs remain undercondensed during mitosis, which is evident by secondary constriction on metaphase chromosomes (Heliot et al, 1997). They are associated with Pol I and nascent pre-rRNA and harbor distinctive euchromatic marks, which are typical for actively transcribed chromatin. It was shown that they are characterized by DNA hypomethylation at the rDNA promoter (Brown & Szyf, 2008), acetylation of histones H4 and H3 and methylation at K4 of histone H3 (H3K4me3) (Earley et al, 2006; Lawrence et al, 2004). Little is known about mechanisms that promote establishment of euchromatic state. Recently it was suggested that the Cockayne syndrome group B protein (CSB) together with the H3K9 histone methyltransferase G9a may promote Pol I mediated transcription elongation by depositing a specific histone modification pattern (Yuan et al, 2007).

Approximately half of the several hundred copies of rRNA genes exhibit heterochromatic chromatin structure and are transcriptionally silent. Even in proliferating cells a significant fraction of rRNA genes is epigenetically silenced. On inactive NORs chromatin is tightly packaged. They can be visualized as condensed foci of rDNA lacking Pol I and Pol I associated factors. This state is maintained throughout the cell cycle and propagated to the next cell generation. There is evidence suggesting that in organisms containing multiple NORs, there must be mechanisms that silence entire NORs (McStay & Grummt, 2008).

In mice, methylation of a single key cytosine at position -133 in the upstream control region of the rDNA promoter abrogates rDNA transcription (Santoro & Grummt, 2001). The promoter of human rDNA genes shows a mosaic pattern of DNA methylation meaning it is neither completely methylated nor demethylated (Ghoshal et al, 2004). Silent rRNA genes are marked by DNA hypermethylation, whereas insufficient methylation of rRNA genes correlates with decreased genomic stability and profound disorganization of the nucleolus indicating that it is also required for maintenance of proper nucleolar structure (Espada et al, 2007). rDNA promoter was shown to be hypomethylated in tumors (Ghoshal et al, 2004; Majumder et al, 2006).

On the histone level, silent rRNA genes are hypoacetylated at histone H4, carry methylation of H3K9, H4K20 and H3K27, and are associated with heterochromatin protein (HP1). Active and silent rDNA copies are characterized by specific nucleosome positions (Grummt, 2007; Li et al, 2006).

1.5. NoRC – the key epigenetic repressor.

The long-term epigenetic repression of rRNA genes is mediated by the chromatin remodeling complex termed NoRC (nucleolar remodeling complex), a member of the ATP-dependent chromatin remodeling machines. It comprises of two subunits, the ATPase SNF2h and a big ca. 205kDa protein TIP5 (TTF-I interacting protein 5) (Strohner et al, 2001).

NoRC was discovered while searching for an rDNA transcription activator. It was known for some time, that binding of TTF-I to its recognition sequence upstream of the rDNA promoter is not only required for initiation of rDNA transcription, but was also accompanied by changes in chromatin architecture, suggesting that it was recruiting some remodeling machinery to the rDNA promoter. In a yeast two-hybrid screen, TIP5 was identified as a

protein interacting with the TTF-I. TIP5 co-localized with UBF in nucleolus, was associated with SNF2h-ATPase and induced nucleosome sliding in an ATP and histone H4 tail dependent manner (Strohner et al, 2001).

Further experiments revealed, however, that NoRC is indeed the key epigenetic repressor complex (Fig. 4). Chromatin immunoprecipitation assays pointed out that NoRC was precipitating with promoter of hypermethylated rRNA genes associated with HP1 and hypoacetylated histone tails. Cell transfections revealed that TIP5 was directly initiating heterochromatin formation. It was facilitating DNA *de novo* methylation and histone deacetylation by recruiting DNMT1, DNMT3 and HDAC1 and repressing rDNA transcription from a reporter plasmid (Santoro et al, 2002). In mouse NoRC-dependent transcriptional silencing lead to methylation of the critical CpG at -133 position (Grummt, 2010).

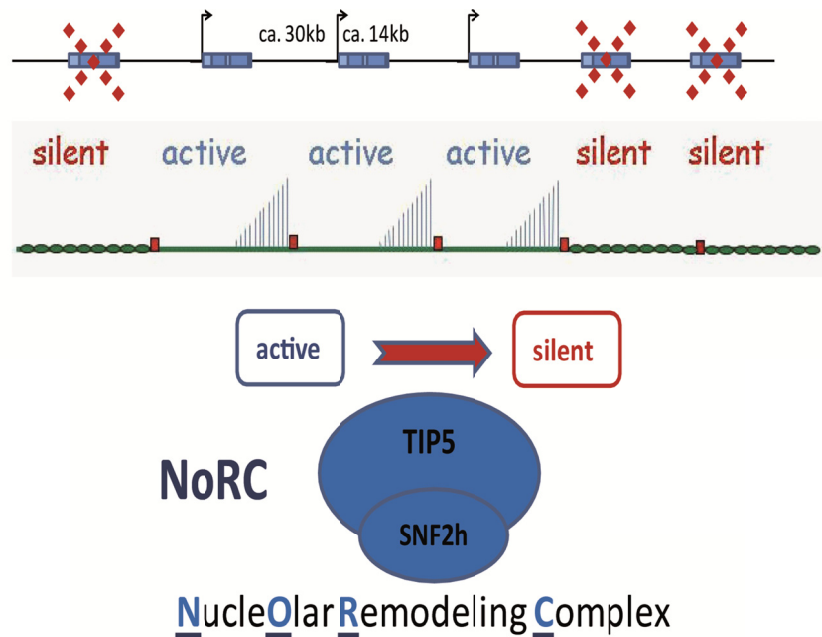


Fig. 4 Epigenetic silencing of rRNA genes mediated by NoRC (nucleolar remodeling complex). On top a schematic drawing of arrays of rRNA genes is presented. The coding sequence for pre-rRNA (ca. 14kb in human) is shown as blue boxes, the intergenic spacer regions (ca.30kb) are shown as black line, and the start of transcription is indicated by black arrows. Upon action of NoRC rRNA genes are being silenced. In the middle, the same arrays are depicted. From active genes nascent pre-rRNA is being produced (blue lines). The silent genes have a compact chromatin structure, indicated by ellipses and lack Pol I mediated transcription. The transition from active to silent state of rRNA genes is promoted by the nucleolar remodeling complex. NoRC acts as a long-term regulator, employing epigenetic mechanisms to alter the ratio of active to silent rRNA genes.

NoRC is recruited to rDNA by the interaction with the N-terminal part of TTF-I and this is also required for TTF-I binding to its recognition sequence. After association with rDNA promoter NoRC represses Pol I transcription effectively before the pre-initiation complex assembly and has no effect on post-initiation effects (Strohner et al, 2004). Later, it was revealed that active and silent rRNA genes differ in the position of promoter-bound nucleosome, and that nucleosome sliding induced by NoRC shifts the nucleosome into the silent position (Li et al, 2006).

NoRC is a complex machine. It is the key mediator of a number of rDNA epigenetic silencing processes. It exerts at least two important functions: shifting of nucleosome into silent position at the rDNA promoter and coordinating the action of numerous chromatin-modifying enzymes that establish the closed heterochromatic state of the rRNA genes. Above that it was shown to control the timing of rDNA replication and the epigenetic inheritance of the rRNA genes (Li et al, 2005).

Given the important role of stringent control of rRNA transcription in the life of a healthy cell, it is easy to suggest that NoRC has an indispensable role in the proper regulation of eukaryotic gene expression and maintenance of the cell stability. And indeed it was recently shown that NoRC mediated heterochromatin is crucial for stability of rRNA genes and that depletion of NoRC upregulates rDNA transcription levels and induces malignant cell transformation (Guete et al, 2010).

1.6. Regulation of NoRC function. The pRNA.

To exert its function NoRC requires interaction of TIP5, its largest subunit, with an ncRNA (Mayer et al, 2006).

Evidence from several systems demonstrates an important role of noncoding RNAs (ncRNA) in regulation of gene expression. Though mechanistic details of ncRNA and chromatin connection are not very clear yet, there is evidence that epigenetic regulation may depend on the interplay of these components (Buhler, 2009).

From immunofluorescence assays it was clear that binding of RNA was required for association of NoRC with nucleolar chromatin. Simple RNase A treatment of eukaryotic cells lead to disappearance of NoRC staining from nucleolus, without affecting the staining for UBF. As well mutations in TIP5 component of NoRC, that abolished RNA interaction, lead

to failures of NoRC dependent induction of heterochromatic hallmarks. Further investigations identified a 150-250 nucleotide long RNA interacting specifically with TIP5, that was required for NoRC mediated gene silencing (Mayer et al, 2006).

It was suggested that NoRC mediated response to the intracellular energy status is as well regulated via the interaction of TIP5 with pRNA. Acetyltransferase MOF acetylates TIP5, impairing its interaction with pRNA, promoting NoRC dependent shifting of the promoter-bound nucleosome into the silent position. NAD⁺ dependent deacetylases SIRT-1 removes the acetyl group upon glucose deprivation, restoring the interaction of NoRC with pRNA and enhancing heterochromatin formation at rDNA loci (Zhou et al, 2009).

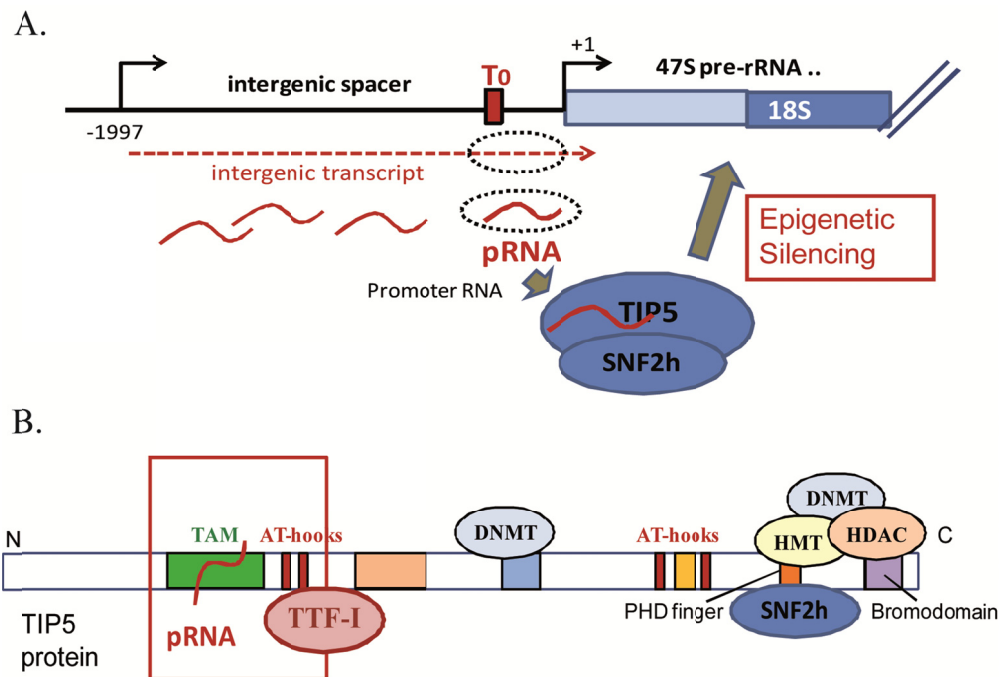


Fig. 5 Model depicting the role of pRNA in NoRC mediated rDNA silencing. (A) Intergenic transcripts (dashed line) are synthesized from spacer promoter, located ca. 2kb upstream of the major pre-rRNA promoter. They are degraded or processed by the yet-unknown mechanism into shorter sequences. Transcripts of 150-250 nucleotides that match the rDNA promoter (pRNA) are bound by TIP5 subunit of NoRC. The association of pRNA with TIP5 is required for NoRC dependent epigenetic silencing of rRNA genes. (B) Modular organization of TIP5. Schematic drawing, illustrating localization of sequence motifs in TIP5, known to be associated with functions in chromatin structure and function. The N- and C- terminus of the protein are indicated as N and C respectively. Important chromatin remodeling enzymes interacting with TIP5 are shown as ellipses at their interaction sequence. HMT – histone methyltransferases. The unique region of TIP5, harboring sequences responsible for specificity of NoRC interaction is boxed red. TAM domain is shown interacting with pRNA. The structure of pRNA is important for its interaction with NoRC, however the structural aspects of pRNA/TAM interaction are not studied yet and thus pRNA structure is left out from the cartoon. Adapted from (McStay & Grummt, 2008).

This ncRNA originates from intergenic spacer regions. It is complementary to the rDNA promoter and, thus, it is termed promoter RNA or pRNA. It has been known since a long time, that intergenic spacer regions, that separate the coding sequences of rRNA genes often contain one or more Pol I promoters, that are similarly organized as gene promoter (Moss et al, 1980). Spacer promoters have been known to be variously active in producing transcripts. The intergenic transcripts were usually of low abundance and did not accumulate *in vivo* (Kuhn & Grummt, 1987). However, intergenic transcripts are highly enriched in the nucleoli and enhanced levels of NoRC are accompanied by elevated amounts of pRNA, suggesting that NoRC physically interacts with pRNA and that this accounts for stability of the pRNA (Mayer et al, 2006).

pRNA originates from the Pol I promoter within the intergenic spacer ca. 2kb upstream of the 47S pre-rRNA coding region (Fig. 5A). It is discussed, that presumably larger intergenic transcripts are processed into shorter intermediates that are rapidly degraded, unless they are bound with the TIP5 subunit of NoRC, as it happens in case of the pRNA (McStay & Grummt, 2008). Recently it was shown, that intergenic spacer transcripts and pre-rRNA are differently regulated, with promoters being differently sensitive to mitogenic signals. IGS RNA seems to be transcribed from a specific subfraction of rRNA genes and function in *trans*, supporting the idea that mammalian rDNA arrays are functionally distinct (Santoro et al, 2010).

pRNA is predicted to fold into a stem-loop structure. Though mammalian rDNA promoters share little homology, pRNA fold was predicted to be common and conserved in several mammals. Mutations that disrupt the stem-loop formation and disturb the pRNA structure impair its interaction with TIP5 and abolish targeting of NoRC to nucleoli. Thus, pRNA stem-loop structure is recognized by TIP5 and is indispensable for NoRC function. (Mayer et al, 2008)

Interaction with pRNA is crucial for NoRC promoted epigenetic silencing of rRNA genes. The interaction is mediated by the TIP5 subunit of the complex and any factors that abrogate TIP5/pRNA interaction on the protein or pRNA lead to impairment of NoRC function, and elevated Pol I transcription (Mayer et al, 2008).

1.7. TIP5 protein.

TIP5 protein (TTF-I interacting protein 5), the largest subunit of NoRC has a molecular weight of ca. 205kDa. It is predicted to harbor a number of domains that are known to be involved in protein-protein and protein-nucleic acid interactions (Fig. 5B). Some of those domains are shared with large subunits of other SNF2-containing chromatin remodeling complexes, ACF/WCRF and CHRAC that have been identified in *Drosophila* (Ito et al, 1999; Varga-Weisz et al, 1997) and humans (Bochar et al, 2000; Poot et al, 2000) and WICH complex (Bozhenok et al, 2002). These domains include a PHD and a bromodomain at the C-terminus, as well as several BAZ motifs.

The C-terminus of TIP5 is responsible for recruitment of chromatin modification enzymes. The PHD finger interacts with histone methyltransferases (HMT) (McStay & Grummt, 2008) and the bromodomain was shown to recruit histone deacetylases HDAC1 and HDAC2 and contact the H4K16 acetylated lysine. It also recruits DNA methyltransferases DNMT1 and DNMT2b and contacts the SNF2h-ATPase. Interestingly, it was observed that deletion of the PHD/bromodomain region of TIP5 turns TIP5 into transcriptional activator and increases Pol I transcription from a reporter plasmid and the same region fused to a DNA binding unit could autonomously mediate rDNA silencing (Zhou & Grummt, 2005; Zhou et al, 2002) DNA methyltransferases are recruited by both the C-terminal and internal part of TIP5 (Santoro et al, 2002).

The N-terminus of TIP5 harbors distinct modules that are not present in other chromatin remodeling complexes. This part of the protein would be responsible for the specificity of NoRC function and specific recruitment of NoRC to the rDNA promoter. The N-terminal region of TIP5 contains the TTF-I interaction sequence, as well as two motifs known as AT-hooks and a TAM domain (Strohner et al, 2001). AT-hooks are peptide motifs that were first described in the high mobility group proteins and are known to mediate binding to the minor groove in DNA. They act as auxiliary protein motifs cooperating with other DNA binding activities (Aravind & Landsman, 1998). Parenthetically, two more AT-hooks are predicted in the C-terminal region of TIP5.

The region comprising the TAM domain and the two AT-hooks is very important. The critical lysine residue (K633 for murine TIP5), that is accountable for NoRC mediated response to the intracellular energy status, is also contained within the second AT-hook

(Zhou et al, 2009). TAM domain and AT-hooks synergize in mediating TIP5 interaction with DNA, as tested on a synthetic oligonucleotide (Strohner et al, 2001). More importantly, the interaction of NoRC with pRNA itself, which determines the functionality of the complex, is mediated by the TIP5 region comprising TAM domain. Deletion of the TAM domain region and mutations in it abrogate the interaction of NoRC with RNA, impair NoRC binding to rDNA promoter and prevent heterochromatin formation (Mayer et al, 2006).

1.8. TAM domain of TIP5.

TAM domain is an MBD-like domain. It is termed by the first letters of the proteins TIP5, ARBP and MBD (Hendrich & Tweedie, 2003).

MBD, the methyl-CpG binding domain recognizes and binds to methylated cytosine in the context of 5'-CG-3'. It is contained within several proteins involved in reading the epigenetic marks on silenced DNA. In humans there are several MBD-containing proteins known. They show high degree of sequence similarity and are conserved across the eukaryotic lineage (Hendrich & Tweedie, 2003). The founding protein is termed MeCP2. It was isolated and characterized in the 1990s. It proved to have a motif, the 70 amino acid MBD domain, which was necessary and sufficient to direct interaction with methyl-CpG-containing DNA fragments (Hendrich & Bird, 1998; Nan et al, 1993).

Based on this motif a number of other proteins containing an MBD domain were identified. The phylogenetic tree of MBD/TAM domains is shown in Fig. 6. Not all of MBD-proteins bind methylated CpG or there is difference in function between mammals and other organisms. For example, MBD2 and MBD3 proteins are the only vertebrate methyl-CpG binders for which homologs were identified in invertebrate genomes (Hendrich & Tweedie, 2003). MBD2 is binding methylated DNA; however, mammalian MBD3 does not. By contrast MBD3 in frogs does bind methyl-CpG (Hendrich & Bird, 1998; Wade et al, 1999).

Most of the MBD-containing proteins are involved in heterochromatin dependent transcriptional repression and, thus, in regulation of a number of metabolic processes. MBD3 is a component of NuRD co-repressor complex (Zhang et al, 1999), which together with MBD2 constitutes the MeCP1 complex. Depletion or deletion of NuRD components result in abnormal embryonic development (Ahringer, 2000). MeCP2, the first described MBD containing protein is an architectural component of chromatin (Dhasarathy & Wade, 2008). It

plays a crucial role in neurons and is required for normal brain function. Mutations in *MeCP2*-gene were linked to development of the Rett syndrome, an X-linked neurodevelopmental disorder (Amir et al, 1999).

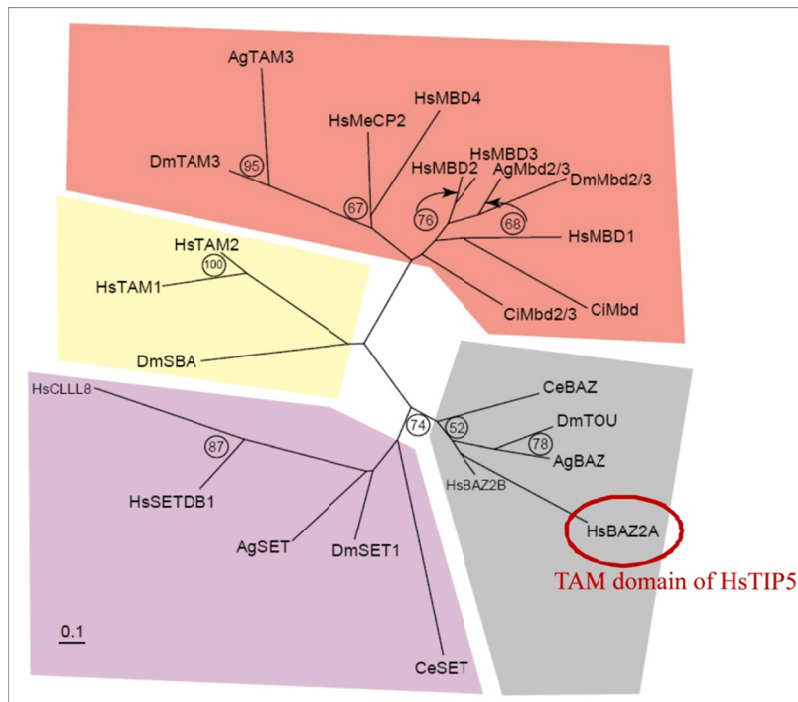


Fig. 6 Canonical MBDs and the evolutionary relationship between MBD and MBD-like domains. Human TAM domain of TIP5 is termed here BAZ2A. This neighbor-joining tree was generated from an alignment of the MBD/TAM sequences encoded by the draft genome sequences of human (Hs), *C. intestinalis* (Ci), *D. melanogaster* (Dm), *A. gambiae* (Ag), and *C. elegans* (Ce) using clustal X (with correcting for multiple substitutions). The tree was constructed with 1000 bootstrap replicates and values of >50% are shown. The scale bar indicates the number of amino acid substitutions per site. The different branches of the tree are shown within different colored boxes: Proteins, containing canonical MBD domains are boxed in orange, the BAZ proteins in grey, SET domain-containing proteins in purple, and SBA/TAM1/TAM2 in yellow. Only MBD1, MBD2, MBD4 and MeCP2 have been shown to be true methyl-CpG binding proteins. Adopted from (Hendrich & Tweedie, 2003).

The structure of the MBD domain is much conserved. Structures of several canonical MBD domains from different mammals have been solved by NMR and crystallography, alone and in complex with target methylated CpG DNA sequence (Fig. 7). (Ho et al, 2008; Ohki et al, 2001; Ohki et al, 1999; Scarsdale et al, 2011; Wakefield et al, 1999). They all adopt very similar folds, forming a wedge-shaped structure. The front side is formed by an antiparallel β -sheet with 3 or 4 strands. The other face of the wedge contains a short α -helix. Despite the lack of apparent sequence similarity, the structural organization of MBDs is close to those of other DNA binding proteins, which show an α/β sandwich structure, formed by a three-

stranded β -sheet and an α -helix (e.g. Tn916 integrase and AtERF1). The contact area to the DNA is, however, rather small (Ohki et al, 2001). Contacts to the DNA backbone are mediated by the loop L_1 and the N-terminal part of the short α -helix α_1 (Fig. 7). Recognition of methylated CpGs is performed by major groove contacts mediated by residues on the β -sheet. Methyls are recognized by a hydrophobic patch, whereas key amino acids are two arginines that are conserved within the family of canonical MBDs (Dhasarathy & Wade, 2008). Further contacts between the MBD and DNA, including two hydrogen bonds of a conserved tyrosine hydroxyl group to methylated cytosine, are mediated by highly structured water molecules, as deduced from the crystal structure of MeCP2 in complex with methylated DNA (Ho et al, 2008).

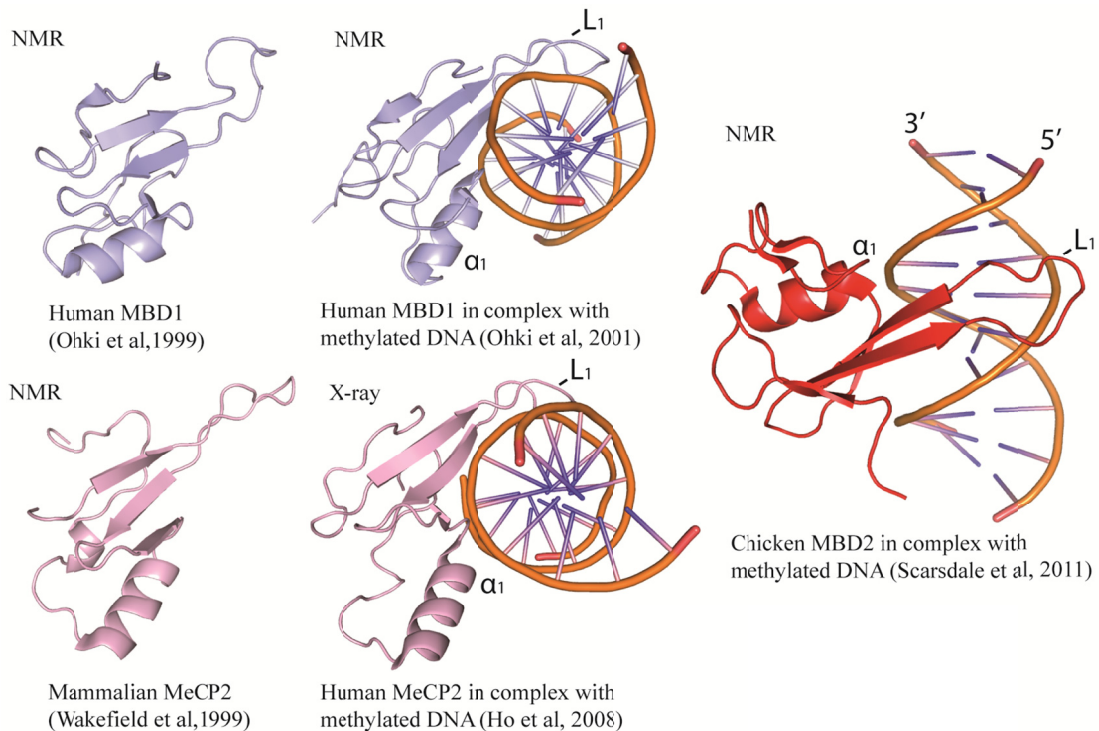


Fig. 7 Structures of canonical MBD domains alone and in complex with their target methylated DNA sequence. The method of determination is indicated on top of each structure. The name of the corresponding protein is indicated at the bottom. The position of loop L_1 and the helix α_1 is indicated for each MBD/DNA complex.

Since identification of the classical MBD family members, several new proteins have been identified in different organisms. The degree of similarity of these sequences to the original MBDs is sometimes modest; however they are related to MBD proteins by phylogenetic

analysis using the MBD domain and to some extent by virtue of other domains (Hendrich & Tweedie, 2003; Mabuchi et al, 2001; Schultz et al, 2002). These domains are termed MBD-like domains and TAM is one of them (Fig. 6). Based on other conserved domains, MBD-like containing proteins are predicted to be chromatin associated as well.

Though MBD domains have been extensively biochemically studied, there is not much information on the TAM domains so far. TAM of TIP5 is the only exception from this rule, and there is some functional and biochemical characterization of it in context of NoRC. TAM domain of TIP5 shows sequence homology and is evolutionary related to functional MBD domains. It was also proven to weakly bind DNA containing a CpG; however it didn't show preference for methylated CpG-probe (Strohner et al, 2001).

Still, it has another very unique feature that strikingly discriminates TAM domain from canonical MBDs and makes it especially interesting. Except for the DNA binding potential, mammalian TAM domain of TIP5 was shown to bind specifically pRNA (Mayer et al, 2006). Until now this has not been shown for any other of the mammalian methyl-CpG binding domains. Though MBD2 and MeCP2 full length protein was shown to bind RNA *in vitro*, this interaction was mapped on another domain (Jeffery & Nakielny, 2004). Even more important is that the interaction of pRNA with TAM has a regulatory effect on the whole NoRC complex function of heterochromatin formation and regulation of the rRNA transcription (see 6).

The structural basis of MBD/DNA interaction is well understood, however, there is no structural information on the TAM domain. Neither the structure of the domain has been investigated, nor its unique specific interaction with the pRNA.

TAM/pRNA interaction is a basic event in regulation of rDNA gene expression by NoRC and consequently in regulation of the whole eukaryotic gene expression. A lot of malfunctions are connected to failures in this process and thus, structural studies of the TAM domain and TAM/pRNA interaction will promote our understanding not only of heterochromatin formation and transcriptional silencing of eukaryotic genes but also of the processes involved in genome stability and cancer development.

2. Introduction II. Solution state NMR Spectroscopy.

2.1. Nuclear magnetic resonance spectroscopy.

Nuclear magnetic resonance (NMR) spectroscopy is one of the principal techniques used to obtain physical, chemical and structural information about biomolecules. It is the most powerful tool to provide detailed information on the three-dimensional structure and dynamics of biological molecules in solution. Additionally, it allows studies of biomolecular interactions by identifying potential binding partners and mapping the interaction interface on molecular structures. It relies on detection of resonance signals coming from atomic nuclei with certain intrinsic properties. Until recently, it was very much limited by the size of the molecules of up to 20kDa. However, nowadays modern isotope enrichment schemes (deuteration of molecules, selective labeling of amino acid groups), new acquisition methods (TROSY (Pervushin et al, 1997), SOFAST HMQC (Schanda & Brutscher, 2005), non uniform-sampling (Rovnyak et al, 2004) and other advanced techniques (residual dipolar couplings and paramagnetic relaxation enhancement) have shifted this border up to 2.3MDa in solution (Christodoulou et al, 2004). With recent advances it is gaining more and more importance in investigation of biological processes at atomic resolution.

2.2. Basic principles of NMR

2.2.1. Theoretical background.

NMR spectroscopy relies on the fact that nuclei have an intrinsic property called spin. The spin is defined by the fourth quantum number I for any given wavefunction, obtained by solving the Schrödinger equation:

$$H\Psi = E\Psi$$

in which Ψ is the wavefunction of the particle, H the Hamiltonian operator and E the energy eigenvalue for the system.

I can have different values: integral, fractional or zero. If the spin value is nonzero a nucleus, that represents a spinning charge in terms of classical mechanics, is associated with an own magnetic field. The resulting spin-magnet has a spin angular momentum I (a vector quantity) and an associated magnetic moment μ , proportional to the spin. By convention the value of the z-component of I is specified by the following equation:

$$I_z = \hbar m ,$$

in which \hbar is the Planck's constant divided by 2π and m is the magnetic quantum number.

The spin magnetic moment μ is collinear with the vector representing the nuclear spin angular momentum and is given by

$$\mu = \gamma I, \text{ or}$$

$$\mu_z = \gamma I_z = \gamma \hbar m,$$

in which γ is a nucleus dependent value called the gyromagnetic ratio. The magnitude of γ determines the receptivity of a nucleus in NMR spectroscopy.

μ is quantised. It is restricted to the values $-I$ to I in integer steps and depends on the orientation of the spin in an external magnetic field.

The nuclei of many isotopes have characteristic spins. The ones that are mostly used in biological NMR are ^{15}N , ^{13}C , ^1H , ^{31}P and ^2H . These isotopes have I -values different from 0 and are either present in biological molecules in high amounts or can be relatively easily enriched by incorporation through different uniform or selective labeling techniques.

The ^{15}N , ^{13}C , ^1H , ^{31}P isotopes have $I = \frac{1}{2}$. The interaction of these nuclei with an external magnetic field gives rise to two spin states and thus to two energy levels, characterized by $I_z = +\frac{1}{2}$ and $I_z = -\frac{1}{2}$ (Fig. 8). The magnetic moment of the lower energy state $+\frac{1}{2}$ is aligned with the external magnetic field; the one of the higher energy with $-\frac{1}{2}$ spin state is opposed to the external magnetic field. The energy of each state in the conventional laboratory coordinate system is given by

$$E = -\mu_z B_0 = -\gamma I_z B_0 = -\gamma \hbar m B_0,$$

in which B_0 is the strength of the external magnetic field. The difference in energy levels of two spin states is thus

$$\Delta E = \gamma \hbar B_0$$

and is dependent on the type of the nuclei (given by γ) and the strength of the external magnetic field (B_0). Considering this, the nuclei with higher γ will give rise to a stronger signal. The same is true for stronger external magnetic fields, which will increase the energy gap. That is why the advances in nuclear magnetic resonance became possible partially due to the new technological achievements in producing NMR spectrometers able to maintain stable extremely high magnetic fields (at the moment up to 1GHz).

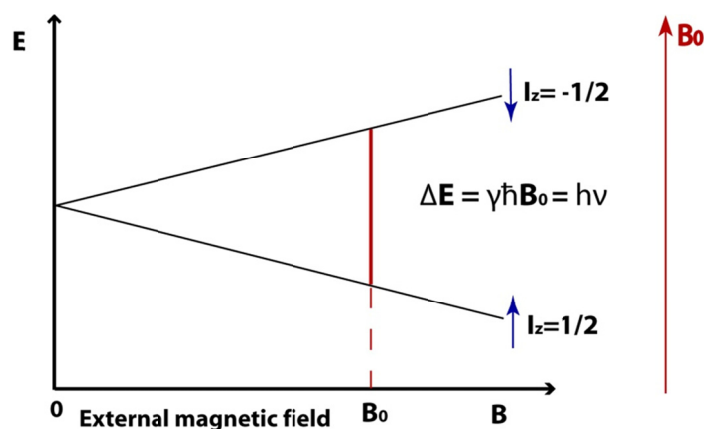


Fig. 8 Splitting of the energy levels for nuclei with $I = \frac{1}{2}$ in an external homogeneous magnetic field. I is the nuclear spin, B is the strength of external magnetic field, γ is the gyromagnetic ratio of the nucleus, h is the Planck's constant and \hbar is the Planck's constant divided by 2π . The energy term is indicated as E ; ΔE is the energy difference between the corresponding spin levels. Blue arrows indicate the alignment of the spins with respect to the vector of external magnetic field (B_0 , red arrow).

In an ensemble of spins both energy levels are nearly equally populated, because the energy gap is in the order of magnitude of the thermic movements. Still, at a given temperature, the lower $I_z = +\frac{1}{2}$ energy level will be slightly more populated than the higher one with $I_z = -\frac{1}{2}$.

The ratio of the parallel to antiparallel spins is described by the Boltzmann distribution:

$$N_{upper}/N_{lower} = e^{-\Delta E/kT}.$$

Here N_{upper} and N_{lower} represent the number of spins on the upper and lower energy levels respectively and ΔE is the energy difference between the two levels, T is the temperature in Kelvin and k represents the Boltzmann constant. In the presence of an external magnetic field the magnetic moments of the individual spins will thus sum up to a macroscopically observable equilibrium magnetization M_0 along B_0 . M_0 can be described by vector aligned with the external magnetic field and the classic theory of NMR deals with this quantity.

Unfortunately the bulk magnetization is small and is the major cause of the sensitivity problem in NMR.

An irradiation of a sample with radiofrequency energy, corresponding exactly to the energy gap between the two spin states of the specific nuclei, will cause excitation of the spins from the $+\frac{1}{2}$ state to the higher $-\frac{1}{2}$ state and, thus, energy absorption at a specific radiofrequency.

For the irradiation frequency this energy is defined as

$$E = \Delta E = h\nu,$$

where ΔE is the energy value for the gap between two spin states, h is the Planck's constant and ν is the irradiation frequency in Hz. In case of NMR ν lies typically in the range of 100-1000MHz, corresponding to normal FM radio frequency band. When the absorption occurs, the nucleus is described as being at resonance. The resonance radiofrequency is detected in NMR spectroscopy giving rise to a nuclear magnetic resonance spectrum.

The resonance method for recording the magnetic properties of atomic nuclei was invented by Isador Rabi, who got the Nobel Prize for his work 1944. The effect of nuclear magnetic resonance was first described by Felix Bloch and Edward Mills Purcell in 1946 (Bloch, 1946; Bloch et al, 1946a; Bloch et al, 1946b; Purcell et al, 1946), both of whom shared the Nobel Prize in physics for their discovery in 1952.

2.2.2. Chemical shift and couplings.

The effect of what is now understood as the "chemical shift" was discovered around 1950 by W.G. Proctor and F.C.Yu (Proctor & Yu, 1951) during determination of the magnetic moments of various nuclei, who noticed that for some isotopes the resonance frequencies were dependent on the compound that was used.

The energy difference between the two spin states is dependent not only on the strength of the external magnetic field but also on the local magnetic field experienced by the spins. Hence, the resonance frequency is specific for each nuclide. Still, nuclei of the same isotope will not resonate at the same frequency. Local electron effects shield each nucleus differently from the external magnetic field. The differences in the effective magnetic field experienced by each nucleus will give rise to different energy gaps and thus to many peak frequencies in a

NMR spectrum. The variation of the spin resonance frequencies of the same kind of nucleus due to the variations in the local chemical environment is called the chemical shift. It is one of the major parameters of NMR spectroscopy, since it causes the differences in positions of the signals in a NMR spectrum, and correlates the changes of the local chemical environment of the nuclei directly. The value of the chemical shift (δ) in ppm (parts per million) is defined as:

$$\delta = \frac{\nu_{\text{signal}} - \nu_{\text{reference}}}{\nu_{\text{reference}}} \times 10^6,$$

where ν_{signal} and $\nu_{\text{reference}}$ are the resonance frequencies measured for the excited nucleus and for the reference component respectively. The size of the chemical shift is given with respect to a reference frequency or reference sample in ppm units instead of Hz, as it is then independent of the external magnetic field strength. The reference component is usually a molecule with a barely distorted electron distribution. In protein NMR, the reference frequency is usually the signal of the methyl groups of DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid).

Another important aspect of NMR is that magnetization can be transferred from one nucleus to another through covalent bonds. This interaction is known as scalar coupling (J) and Ramsey and Purcell suggested that it is mediated by electrons forming the bond (Ramsey & Purcell, 1952). The strength of the interaction for two nuclei a and b is measured by the scalar coupling constant ${}^nJ_{\text{ab}}$ (in Hz), where n is the number of covalent bonds separating the nuclei. Scalar coupling modifies the energy levels of the system and the NMR spectrum is modified correspondingly. With current methods magnetization can be transferred over scalar coupling as far as over one to four covalent bonds. In proteins transfer over three bonds is very often utilized in various NMR experiments. J-coupling is a very important phenomenon for NMR, since it allows to measure not only the frequency of one nucleus, but also of the ones directly bound to it and to deduce local connectivity within a molecule.

Nuclei can also interact through space and one more type of interaction important for NMR is the dipolar coupling (D), also measured in Hz. It is caused by direct dipolar interactions between two magnetic dipole moments and goes through space. Many NMR experiments make use of it; among others, for example the NOESY experiments, performed to identify nuclear neighbors that are not connected by chemical bonds and to determine distances

between them. Dipolar coupling is an anisotropic quantity and for a molecule in isotropic solution it averages to zero due to rotational diffusion. It carries a lot of additional valuable information and is often measured as residual dipolar coupling (RDC) in a solution artificially partially deprived of isotropy (discussed later).

2.2.3. Vector formalism.

Bloch has formulated a simple semiclassical model to describe the behavior of a sample with noninteracting nuclei of a spin $-\frac{1}{2}$ in a static external magnetic field (Bloch, 1946). Here, the bulk magnetization is represented as a vector quantity that experiences a torque in presence of an external magnetic field. In the laboratory frame

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{M}(t) \times \gamma\mathbf{B}(t)$$

where $\mathbf{M}(t)$ is the bulk magnetization, $\mathbf{B}(t)$ is the strength of magnetic field and γ is the gyromagnetic ratio of the nuclei.

Without going into mathematical details, the bulk magnetization, put into a Cartesian coordinate system, represents a vector \mathbf{M} , pointing along the z-axis, collinear with B_0 (Fig. 9A). If a short radio pulse is applied along the x-axis (Fig. 9B), the magnetization vector will turn towards the $-y$ -axis. The angle α will depend on the length of the pulse. The direction of the rotation is determined by the right hand rule, known from physics of electromagnetism. The magnetization will start precessing around the z-axis or the external magnetic field B_0 (Fig. 9C) with an angular frequency ω , generating the signal in NMR detection coil. For a static B_0 the precession frequency equals

$$\omega = -\gamma B_0.$$

It is called the Larmor frequency and, as can be obtained from the equation, it has different signs for nuclei with positive and negative gyromagnetic ratios. The magnitude of the precession frequency ω is identical to the frequency of electromagnetic radiation required to excite transitions between the spin energy levels and this is the actual readout of an NMR experiment.

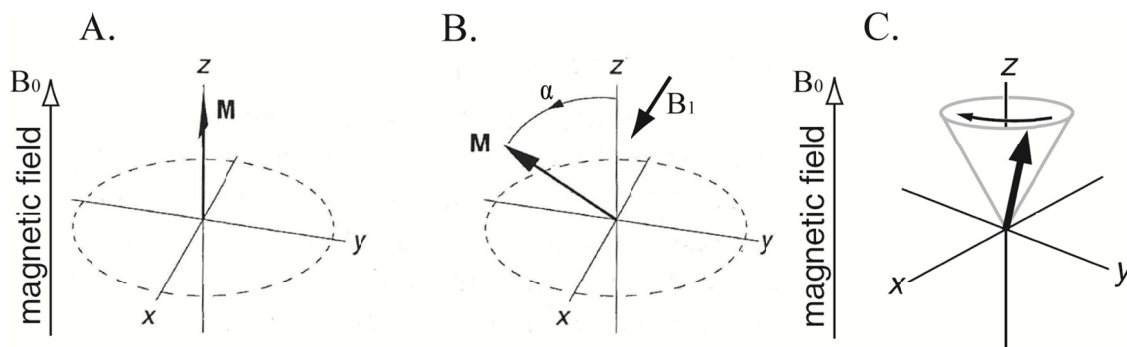


Fig. 9 The vector model of NMR. (A) The bulk magnetization vector M in the presence of external magnetic field B_0 . (B) Effect of a radiofrequency pulse B_1 along x on the magnetization vector M . α is the flip angle, enforced by the pulse. Adapted from (Hore et al, 2001) (C) The bulk magnetization vector M precessing around the z -axis, collinear with the external magnetic field B_0 , after a radiofrequency pulse. From (Keeler, 2005).

Due to relaxation effects, which include the ultimate loss of spin alignment (transverse relaxation T_2) and simultaneous return of the system to the thermodynamic equilibrium state (longitudinal relaxation T_1) the bulk magnetization will return to the ground state and the NMR signal will decay.

2.2.4. Product operator formalism.

Unfortunately, the Bloch vector model can only explain simple NMR experiments on isolated spins. To understand more complex phenomena like couplings, product operators were introduced. Product operators are based on quantum mechanics of the populated energy states and can handle coupled nuclei. They have a clear physical meaning. The product operator formalism allows a complete quantum mechanical description of NMR experiments and the calculation of their outcome.

In brief, the three components of the spin angular momentum along z -, x - and y -axis, can be represented by Cartesian spin angular momentum operators: I_z , I_x , and I_y , and the entire state of the spins or spin system can be described by the wave function $\psi(t)$ or the density operator $\sigma(t)$. Disregarding relaxation, the time evolution of the density operator is described by the Liouville-von Neumann equation:

$$\frac{d\sigma(t)}{dt} = -i[H(t), \sigma(t)]$$

where $H(t)$ is the Hamiltonian operator, that includes chemical shift terms, coupling terms and so on, acting on the density operator. In terms of Cartesian product operators, the density operator for a single spin $\frac{1}{2}$ can be described as a sum of different amounts of the three product operators:

$$\sigma(t) = a(t)Ix + b(t)Iy + c(t)Iz.$$

The amounts of three operators will vary with time during pulses and delays. At equilibrium, the density operator is proportional to I_z . The constant of proportionality is often dropped, so it is written as $\sigma_{\text{eq}} = I_z$. During an NMR experiment this state is sequentially transformed and, hence, product operators evolve during that time.

Product operators evolve under pulses. For example the effect of 90° and 180° pulses on I_z is:

$$\begin{aligned} I_z &\xrightarrow{90^\circ I_x} -I_y & I_z &\xrightarrow{90^\circ I_y} I_x & I_z &\xrightarrow{90^\circ I_z} I_z \\ I_z &\xrightarrow{180^\circ I_x, I_y} -I_z & I_z &\xrightarrow{180^\circ I_z} I_z \end{aligned}$$

This can be interpreted as rotation along different axes and corresponds to what is expected from the vector model. The evolution can be calculated for any degree of rotation; however, a rotation of operator about its own axis leaves it untouched.

The chemical shift evolves with the offset Ω during the time t of precession with the offset (Ω) being the difference between a signal and a reference value, i.e. $\nu - \nu_{\text{ref}}$:

$$\begin{aligned} I_x &\xrightarrow{\Omega t I_z} I_x \cos \Omega t + I_y \sin \Omega t \\ I_y &\xrightarrow{\Omega t I_z} I_y \cos \Omega t - I_x \sin \Omega t \\ I_z &\xrightarrow{\Omega t I_z} I_z \end{aligned}$$

So far only uncoupled nuclei were considered. The product operator approach comes to its own when considering the coupled spin systems. To explain the J-coupling, a second spin S is introduced with product operators S_x , S_y and S_z describing it. The S-spin in biomolecular NMR is typically ^{15}N or ^{13}C . Due to large difference in resonance frequency, pulses on either of the spins will leave the other one untouched. However, due to the J-coupling the states of the I and S spins will mix. The result is a product operator for two spins $2IS$ (the factor 2 is

needed for normalization purposes). The operators for two spins evolve under offsets and pulses the same way as operators for a single spin. The rotations, however, have to be applied separately for each spin and do not affect each other. Operators I_x , I_y , S_x and S_y evolve under coupling, whereas I_z and S_z do not. For the evolution of product operators under coupling see Fig. 10.

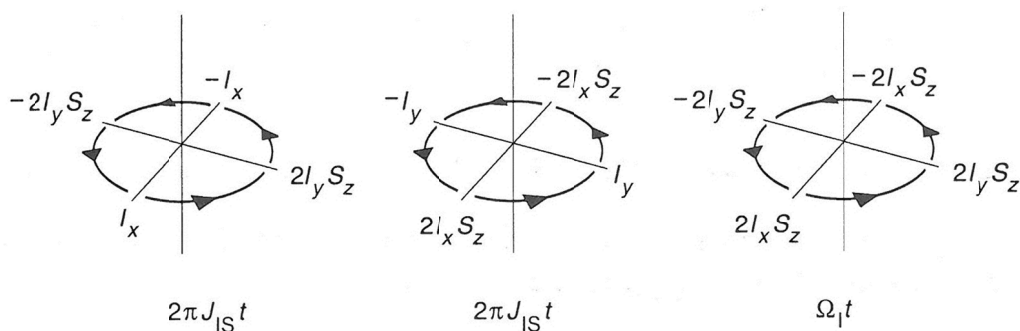


Fig. 10 Evolution of the product operators in a coupled spin system. The diagrams illustrate transformations of the product operators for a two-spin system IS, produced by the J-coupling (the left panel for I_x and the middle panel for I_y) and by the resonance offset Ω_1 during period t . Adapted from (Hore et al, 2001).

The final state of an NMR experiment is calculated by applying sequentially these evolution rules to the product operators of the initial state.

2.3. NMR spectroscopy experimental setup.

2.3.1. Acquisition methods

Throughout the first decades NMR utilized a technique of continuous-wave spectroscopy in which either the sample was held in a constant external magnetic field and the radiofrequency of the source was slowly scanned to chart the on-resonance positions of the spectrum or, more frequently, the source radiofrequency was held constant and the magnetic field was scanned. Unfortunately, apart from limitations in that one had to probe each frequency individually, that technique had also a poor signal-to-noise ratio that had to be improved by signal averaging.

The modern spectrometers utilize the technique of pulsed Fourier transform nuclear magnetic resonance spectroscopy (FT-NMR). This field was pioneered by Richard Ernst (Ernst et al, 1989), who was awarded the Nobel Prize 1991. Pulsed FT-NMR works by irradiating the

sample, held in a strong static external magnetic field, with a short burst of radiofrequency electromagnetic radiation containing all radiofrequencies in the range of interest (pulse), usually in the order of several microseconds in duration. Thereby the equilibrium magnetization M_0 along z-axis is transferred to the xy detector plane. The polarized nuclear magnets start precessing around the external magnetic field vector creating a radiofrequency signal that is detected. The relaxation effects cause a decay of the signal, known as FID, the free induction decay. This time-dependent signal pattern contains information on all excited nuclei. It is recorded and deconvoluted into a frequency-dependent pattern of nuclear resonances using the Fourier transformation, producing a typical nuclear magnetic resonance spectrum.

Fig. 11 shows the basic hardware of an experimental setup. The sample is placed into the core of a superconducting magnet that generates strong homogeneous external magnetic field (B_0). To manipulate the spins specially designed sequences of radiofrequency pulses with different pulse shapes, orientations and durations are used (an NMR pulse sequence). Such pulse sequences are sent to the radiofrequency (RF) synthesizer, which executes them. The electric pulses are converted into radio pulses by coils of the probe within the magnet and transferred to the sample. Directly after execution of the pulse program the same coils are used to detect the NMR signal (FID). The signal is then intensified by an amplifier, which is inevitable due to low signal intensity. After amplification, the oscillating FID is digitized, and can be recorded and processed on a computer. The processing involves the deconvolution of the FID by the Fourier-transform.

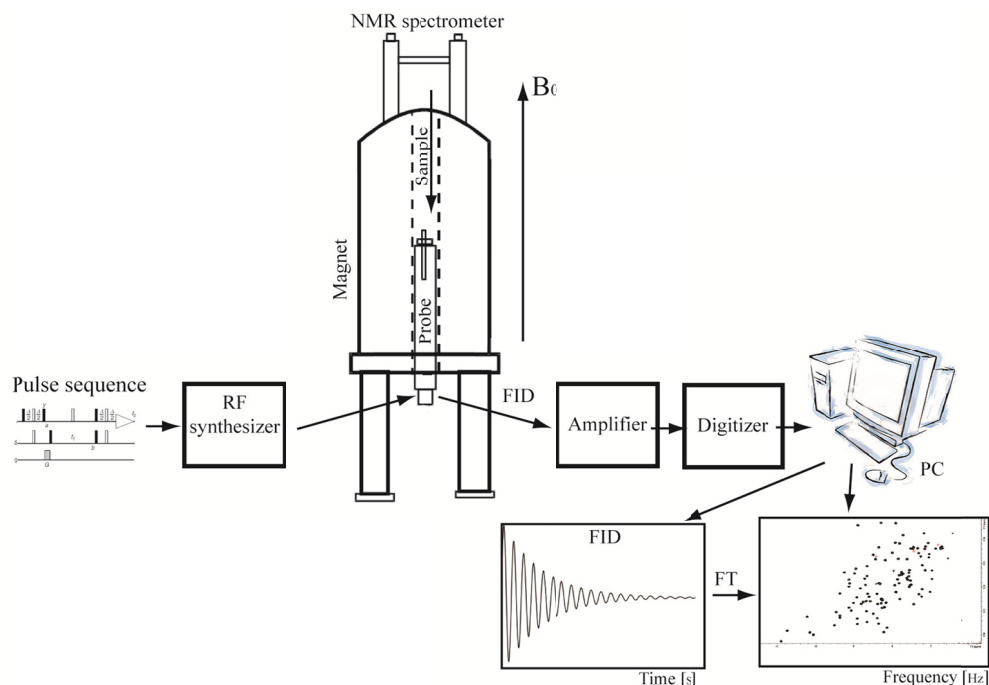


Fig. 11 Outline of the setup of an NMR experiment. B_0 shows the direction of the external magnetic field, produced by the NMR spectrometer. RF stands for radiofrequency, FID for the free induction decay, PC for personal computer.

The use of pulse sequences and application of the multidimensional FT-NMR experiments on unlabelled and isotope labeled samples allows specific excitation of different types of nuclei. Thus, selective extraction of many different types of information about a molecule in a relatively short period of time is possible, increasing the resolution and sensitivity of NMR spectroscopy and making it applicable to studies of proteins and nucleic acids that contain a huge number of nuclei.

Richard Ernst and Kurt Wüthrich – in addition to many others – developed two-dimensional and multi-dimensional FT-NMR into a powerful technique for studying biochemistry and structure of biomolecules (Wüthrich, 1986). 2002 Kurt Wüthrich shared a Nobel Prize in chemistry for his work.

2.3.2. One-dimensional FT-NMR spectroscopy

In the pulsed FT-NMR each one-dimensional experiment consists of two parts: preparation and detection (Fig. 12A). During the preparation, the spin system is set to a defined state, mostly by applying a radiofrequency pulse that would rotate the M_0 90° from the equilibrium state into the detection plane of the spectrometer (a 90° pulse). After this pulse the spins

precess around the initial axis. The system is allowed to return to the initial state, whereby the arising FID is detected and recorded. Usually, the experiment is repeated several times to increase the signal-to-noise ratio. After the summation, the data are Fourier transformed to yield the final NMR spectrum.

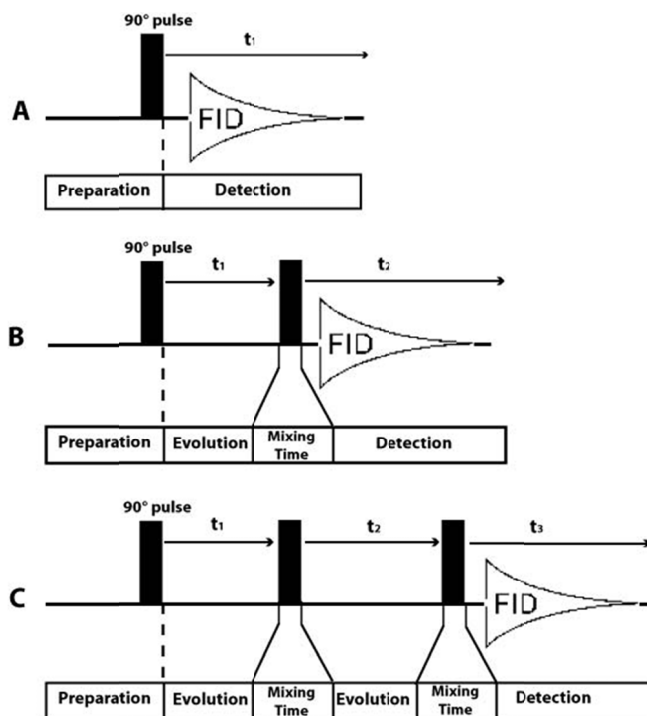


Fig. 12 Schematic representation of pulse sequences for FT-NMR experiments. Black squares indicate blocks of pulse; t_1 , t_2 and t_3 is the evolution time between them. The first 90°, pulse that transfers the bulk magnetization into the xy-plane is indicated at the top of the first pulse train. (A) A one-dimensional FT-NMR experiment. (B) A two-dimensional FT-NMR experiment. (C) A three-dimensional FT-NMR experiment.

The arising 1D spectrum has a unique pattern for every protein. It is far too complex for interpretation as most signals overlap heavily. Still it already permits first analysis of the protein structure. A ^1H one-dimensional spectrum of a folded protein is indicated by a high dispersion of peaks within the -0.5ppm to ca. 12ppm scale range. The presence of the signals above 8.5ppm coming from the protein amide protons and signals below 0.5ppm arising from the protons of the aliphatic methyl groups are the most diagnostic for the presence of a folded protein.

2.3.3. Two-dimensional FT-NMR spectroscopy. The HSQC.

In addition to preparation and detection, a two-dimensional FT-NMR experiment has an indirect variable evolution time and a mixing sequence (Fig. 12B). After the preparation of a system with a 90° pulse, or a series of pulses, the spins are allowed to precess freely for a given evolution time t_1 . During this time, the magnetization is labeled with the chemical shift of the first nucleus. With the help of the mixing sequence, the magnetization is then transferred from the first nucleus to the second one either through bonds (J-coupling) or over dipolar interaction (Nuclear Overhauser effect). Afterwards the FID is acquired during the detection time (direct evolution time t_2). The experiment is repeated several times, whereby the evolution time is not held constant but is incremented in equal intervals starting with 0 until sufficient data is recorded. Thus the FID is acquired as a function of t_2 for each value of t_1 . The resulting data are Fourier transformed twice, yielding a two dimensional spectrum which correlates the resonance frequencies of two types of nuclei.

The information found in the spectrum is determined by the exact nature of the preparation and the mixing periods. If the correlation is recorded between the nuclei of the same isotope the spectra are called homonuclear. In the other case they are called heteronuclear.

One of the most important two-dimensional heteronuclear NMR experiments for the proteins is the $^1\text{H}^{15}\text{N}$ -HSQC (Heteronuclear Single Quantum Coherence). It has the ^1H and ^{15}N frequency axes, and correlates the nitrogen atom of an NH_x group with the directly attached proton. The resulting spectrum contains the signals of the H^{N} protons in the protein backbone, signals from the NH_2 groups of the side chains of Asn and Gln and of the aromatic H^{N} protons of Trp and His, giving a unique and disperse pattern for a folded protein. The $^1\text{H}^{15}\text{N}$ -HSQC spectrum is often referred to as a protein fingerprint spectrum, since it is unique for every protein. Amide groups not involved in secondary structure give rise to signals in the middle region of the spectrum around 8ppm. Thus, apart from other applications, the $^1\text{H}^{15}\text{N}$ -HSQC can be used to monitor the folding state of proteins under certain conditions.

Unfortunately, for larger proteins the HSQC has a limited application due to significant overlap in the middle region of the spectrum, as well as high transverse relaxation (T_2) rates. The relaxation can be reduced by replacement of the major source of T_2 , the omnipresent hydrogen atoms, by deuterons [reviewed by (Gardner & Kay, 1998)]. Still, protein targets remained limited to 50 kDa until in the late nineties, when TROSY (Transverse Relaxation Optimized SpectroscopY) experiment was introduced (Pervushin et al, 1997; Salzman et al,

1998). A $^1\text{H}^{15}\text{N}$ -TROSY correlates the same atoms as HSQC, however it reduces the relaxation effect such, that better line widths and sensitivity can be achieved for larger molecules. It works best for deuterated proteins and is especially suited for applications to protonated amide groups. It has greatly extended the size limit of molecules possible to study by NMR [reviewed by (Fernandez & Wider, 2003)].

2.3.4. Three-dimensional FT-NMR spectroscopy. Backbone assignment

A three dimensional NMR experiment can be constructed from a two dimensional experiment by inserting an additional indirect evolution time and a second mixing period between the first mixing period and the direct data acquisition (Fig. 12C). Each of the different indirect time periods is incremented separately in equal intervals starting with 0 until sufficient data is recorded. The resulting data are Fourier transformed three times, yielding a three dimensional spectrum which correlates the resonance frequencies of three types of nuclei. The introduction of three-dimensional heteronuclear experiments and the availability of ^{15}N - and ^{13}C -labelled proteins allowed determination of their structure in solution.

Prior to extracting all available information about interatomic distances and torsion angles, the initial step of structure investigation by NMR spectroscopy requires assigning of the observed resonances of the investigated molecule. For proteins, the sequential assignment of the backbone atoms is completed first. One of the strategies for the backbone assignment employs HNCA, HNCACB and HN(CO)CACB triple resonance experiments (Sattler et al, 1999; Shan et al, 1996).

During the HNCA experiment, the magnetization starts at an amide proton (H^{N}) and is then transferred to the directly attached nitrogen atom (N) which is measured as the first spectral dimension. Then the magnetization is transferred to the nearby $\text{C}\alpha$ nuclei which are measured as the second spectral dimension. After that, the magnetization is transferred back the same way to the amide proton which is measured as the third (direct) dimension. Thus the HNCA experiment correlates the H^{N} of each amino acid (residue i) with its own $\text{C}\alpha$ and the $\text{C}\alpha$ of the previous residue (residue $i-1$). A $^1\text{H}^{15}\text{N}$ projection of an HNCA looks like an HSQC plane. The HNCACB correlates in a similar way the H^{N} of each amino acid with its own backbone carbon and the carbon of the previous residue. However this time the magnetization is also transferred to the $\text{C}\beta$ atoms. The way the magnetization is transferred between $\text{C}\alpha$ and $\text{C}\beta$ also causes the correlations for these atoms to have opposite signs, which is one of the ways of distinguishing between them in the arising HNCACB spectrum. To differentiate the

resonances coming from the i and $i-1$ amino acids the HN(CO)CACB experiment is often recorded. In this case the magnetization is transferred over the carbonyl carbon (CO) and thus only the interresidual cross signals can be observed.

Taken together the HNCA, HNCACB and HN(CO)CACB experiments unambiguously correlate the frequencies of the $C\alpha$ and $C\beta$ backbone carbons with the HN resonances of the same amino acid. Since the $C\alpha$ and $C\beta$ frequencies of the amino acids are related to the identity of the side chains (Cavanagh J., 1996), it is possible to distinguish between the different amino acids and, in the context of the neighboring residues, unambiguously sequentially assign the $C\alpha$, $C\beta$, amide N and H^N chemical shifts to the corresponding residue in the protein sequence. For larger proteins, these experiments can be optimized by the use of TROSY. Then the $^1H^{15}N$ projection of these three-dimensional spectra will look like a $^1H^{15}N$ -TROSY plane.

Backbone assignment gives preliminary information on the secondary structure of the protein. The $C\alpha$ and $C\beta$ chemical shifts are sensitive indicators for the secondary structure elements such as α -helix and β -sheet (Pastore & Saudek, 1990a; Spera & Bax, 1991a). Since these values strongly depend on the identity of the residues, they are standardized by subtraction of the appropriate random coil chemical shift that reflects the amino acid specific component of the chemical shift. The result gives secondary chemical shift values, if only $C\alpha$ frequencies are used, and combined chemical shifts, if $C\alpha$ and $C\beta$ frequencies are implicated. Over time, a number of random coil data sets based on either a database (Wishart & Sykes, 1994; Wishart et al, 1992) or model peptides under a variety of experimental conditions (Plaxco et al, 1997; Thanabal et al, 1994; Wishart et al, 1995) have been published. Observed carbon chemical shifts in the structural parts of a folded protein will differ significantly from the random coil chemical shifts: positively in the α -helical regions and negatively in the regions of β -sheets.

Overall knowing the backbone resonance assignment is the basic step in any structural investigation of proteins in solution. A lot of further methods, like side-chain assignment, dynamics analysis or ligand binding analysis rely on the knowledge of the backbone resonance frequencies.

2.4. Calculating structures in solution.

2.4.1. Structure determination procedure.

NMR structure determination relies on the simulated folding of the biomolecule, which is caused by application of structural restraints. Structure determination begins with assignment of the backbone resonances using three dimensional NMR backbone experiments described above. Starting from that, the assignment of the side chain atoms, which can be rounded up with stereospecific assignment of methyl groups, is assessed and based on this the unambiguous assignment of Nuclear Overhauser Effect induced patterns (NOEs).

NOE-derived interatomic distances and scalar coupling constants have formed the basis for structure determination, beginning with the report of structure of the protease inhibitor IIA in 1985 (Williamson et al, 1985). Nowadays, additionally restraints obtained from paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC) measurements are employed to determine the relative position of structural elements within the molecule. Angle restraints from the backbone (Φ and Ψ) and sometimes side chain (χ_1 and χ_2) dihedral angles are also put into the structure calculation protocol, mostly predicted by the informatics programs, such as TALOS+ (Shen et al, 2009), as in the case of backbone dihedral angles.

The experimentally derived restraints are supplemented with restraints specifically imposed to enforce proper geometry of the molecule, like bond length, chirality or planarity of the aromatic rings and peptide units. The calculation protocol is rather a simulated annealing procedure. The system is virtually heated up and then the energy is slowly lowered. The program tries then to find coordinates for each atom, which would best satisfy the given restraints. Importantly, measured NMR data carry an intrinsic experimental uncertainty and, hence, are included as ranges of allowed values. That is why it does not uniquely define the three dimensional structure of a molecule in solution. Structure calculation is repeated several times to determine an ensemble of lowest energy structures that is consistent with the NMR input data. This ensemble of lowest energy structures is usually reported.

After the calculation the structure is validated and the quality of the ensemble is reviewed. It is assessed by two general measures: how well the calculated structure fulfills the experimental data; and how well it fits to the stereochemistry derived from previously calculated structures, e.g. high-resolution x-ray structures. The first term is usually judged by examining the violations of the restraints within the ensemble of structures. The second term,

the stereochemical quality is usually judged by quantifying the distributions of backbone and side chain dihedral angles, the number of van der Waals steric clashes etc. There are special NMR software programs to assess these qualities, e.g. iCING (Vuister & da Silva), which includes PROCHECK (Laskowski et al, 1993) and WHATCHECK (Vriend & Sander, 1993). Furthermore, structures can be validated against sets of restraints, not used in the calculations. For example supplementary sets of RDCs as well as PRE data can be utilized. Monitoring the exchange of amide protons to deuterons can also be utilized for validation.

2.4.2. Nuclear Overhauser Effect (NOE).

The nuclear Overhauser Effect is a manifestation of the prediction, that dipolar coupled spins do not relax independently. Longitudinal relaxation, caused by fluctuating magnetic fields in a liquid sample will induce transitions between spin states. The effect of these spin flips will move the populations of the energy levels towards equilibrium. If these fields arise from the dipolar interaction of the two spins, then all possible transitions between the energy levels can occur. Solomon equations, describing the dependence of spin magnetization on spin transition rate constants, predict that the way magnetization of one spin varies with the time will then depend also on what is happening to its neighboring spin. Both spins will be connected through space by the phenomenon called cross-relaxation. If the S-spin magnetization deviates from the equilibrium, there will be a change in the I-spin magnetization that is dipolar coupled to it. The change in the I-spin magnetization will manifest itself as a change in the intensity of the corresponding spectrum and this effect is called the Nuclear Overhauser Effect (NOE).

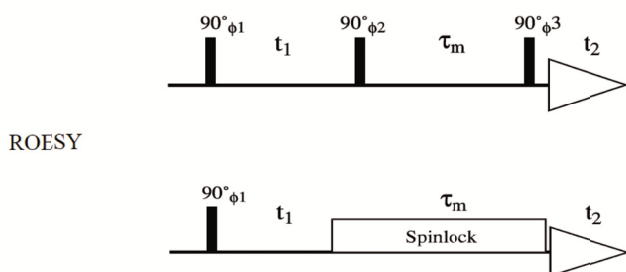
NOE is characterized by the cross-relaxation rate constant (σ_{IS}) that has a strong dependence on the type of nuclei and on the distance between two spins. It is inverse proportional to the sixth power of the distance between the two dipolar interacting spins and the NOE effect is observed only for spins within the distance of $<5\text{\AA}$. It is also dependent on the size of the molecules and experimentally on the spectrometer frequency. In the initial rate approximation NOE induced peak intensities are proportional to the relaxation rate constants. They are measured standard wise via NOESY (NOE Spectroscopy) pulse sequences

Fig. 13A). These pulse sequences include a mixing period in which the spins are along the magnetic field and can undergo cross relaxation. The time of the mixing period is chosen short enough to satisfy the initial rate approximation and so, the distance between the dipole coupled spins can be deduced.

The Nuclear Overhauser Effect is one of the most important effects in NMR, since the majority of the distance restraints used to calculate the NMR structure is derived from many hundreds of NOE-peaks. Without complex calculations, from the peak intensities the internuclear distances can only be estimated and not determined precisely, that is why the NOE peaks are usually grouped into several categories and are introduced into calculations as distance ranges.

σ_{IS} can be positive or negative, giving rise to positive or negative NOE. Depending on the experimental setup, the NOE can also be zero (Fig. 13). Then one could consider measuring cross relaxation in the rotating frame (ROESY experiment), applying a special sequence called spinlock during the mixing time. The resulting rotating frame cross relaxation constant is proportional to σ_{IS} and is always positive.

A. NOESY



B.

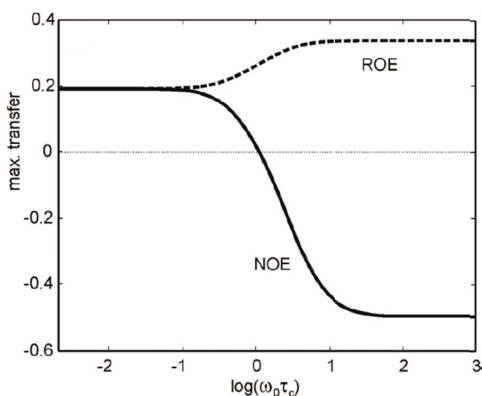


Fig. 13 Nuclear Overhauser Effect. (A) Pulse sequence diagram for a NOESY and a ROESY experiment. Black squares represent radiofrequency pulses, the degree of magnetization rotation is indicated above. ϕ_1 - ϕ_3 are the pulse directions, t_1 and t_2 the durations of evolution period. The FID is represented by the triangle. τ_m is the mixing time. (B) Maximal NOE- or ROE-transfer as a function of the spectrometer frequency (ω_0) and correlation time of the molecule (τ_c). For details see (Cavanagh J., 1996).

Considering the number of atoms in a biological macromolecule NOE gives rise to a tremendous amount of signals in NMR spectra. At present there are software programs that utilize special algorithms to assign the complex pattern of NOE induced peaks for proteins automatically, starting from initial data. The most popular ones CYANA (Guntert, 2004) and ARIA (Linge et al, 2003a) are based on iterative rounds of assignment and reassignment. They include ambiguous restraints and network anchoring, derive distance restraints and perform simple dynamics to obtain initial structural models.

2.4.3. Residual dipolar couplings (RDC).

The magnetic field, experienced by one nucleus is different depending on the orientation of its neighbor nucleus. The dipolar interaction between two nuclei is dependent on their relative orientation and distance between them. It is a through space interaction, that can be described by the dipolar coupling constant (Ernst et al, 1989):

$$D = \frac{k}{R^3} (\cos^2\theta - \frac{1}{3})$$

where θ is the vector between the internuclear vector and the magnetic field, R is the distance between dipole-coupled nuclei and k is a physical constant that includes the gyromagnetic ratios of the interacting spins, the Planck's constant and the permeability of vacuum (Ernst et al, 1989). Dipolar couplings in fully anisotropic media are very large, in the order of kHz, complicating the NMR spectra of solid state NMR. On the contrary, as mentioned previously, in solution due to the Brownian motion over time they average out to zero. For review see (Kramer et al, 2004).

Though NOE is also distance dependent and goes through space, the information it carries is local and does not depend on orientation. In contrast the dipolar coupling is defined in the molecular coordinate frame and provides information on how each dipole is related to it and thus, to each other.

In NMR dipolar couplings can be investigated as residual dipolar couplings (RDC) in weak anisotropic media. These media lead to alignment of biomolecules relative to the external magnetic field and are designed to tune the degree of anisotropy, such that the dipolar coupling has a similar magnitude to J-coupling (Hz-range). As alignment media bacteriophages (Pfl) (Zweckstetter & Bax, 2001), liquid crystalline media (Lorieau et al, 2008; Rückert & Otting, 2000) or stretched and compressed gels (Ishii et al, 2001) are used.

RDCs are commonly measured as the difference in J-coupling in the spectrum of aligned and non aligned molecule. Then they are analyzed by different programs to couple them to the molecular frame (Dosset et al, 2001; Zweckstetter, 2008). However, a single RDC will not provide a unique restraint and a set of minimum five measured RDC values is needed to define the orientation of a bond.

The use of alignment media for biological macromolecules is not always experimentally trivial. Thus, RDCs are usually incorporated into the structure calculation for direct refinement of orientation of bond vectors against the measured values of RDCs (Clare et al, 1998b). This can be performed also with incomplete sets of RDCs. As structures are calculated, the orientation of each individual bond vector is changed to satisfy the RDC data and so the position of structural elements within the molecule is refined. This approach has been shown to improve the relative orientation of structural elements and the quality of structures; as well as help to position individual domains of complexes, if the initial structure calculated from NOE and scalar couplings is well defined (Bax & Grishaev, 2005; Fischer et al, 1999).

2.4.4. Paramagnetic relaxation enhancement (PRE).

Presence of unpaired electrons in the molecular system enhances the magnetic field modulation at a nucleus and nuclear spin relaxation due to hyperfine electron – nucleus couplings. This effect is called paramagnetic relaxation enhancement (PRE). The PRE-field was pioneered by the work of Solomon and Bloembergen in the 1950s (Solomon, 1955; Solomon & Bloembergen, 1956). Nowadays it is used in determination of the global fold of a molecule.

The PRE effect can be intentionally induced, if a label with an unpaired electron is introduced. For example a small compound, containing a nitroxide can be chemically coupled to the macromolecule or paramagnetic compound can be added to the buffer solution and induce enhanced relaxation of the surface spins. The PRE effect can then be detected as decrease in peak intensity, as compared to a reference spectrum, and altered longitudinal and transverse relaxation rates, R_1 and R_2 respectively.

PRE can be monitored on the HN-resonances in the $^1\text{H}^{15}\text{N}$ -HSQC spectra. The intensity of the peaks is extracted from the spectra of the macromolecule with a spin label in the oxidized (active) and the reduced (inactive) state. The relaxation effect of the spin label on ^{15}N nuclei

is considered to be negligible compared to the ^1H relaxation due to the big difference in gyromagnetic ratio. Additionally for the proton in this case, the paramagnetic-induced R_1 relaxation is insignificant compared to the R_2 effects (Gillespie & Shortle, 1997). Hence only the R_2 relaxation rates for ^1H nuclei are measured for the macromolecule with spin label in both states.

The intensity ratio is related to the R_2 relaxation rates by the following equation (Battiste & Wagner, 2000):

$$\frac{I_{\text{ox}}}{I_{\text{red}}} = \frac{R_2 \exp(-R_2^{\text{SP}} t)}{R_2 + R_2^{\text{SP}}}$$

where I_{ox} and I_{red} are the signal intensities for amides in the $^1\text{H}^{15}\text{N}$ -HSQC for a sample with a spin label in the oxidized and the reduced state respectively, R_2 is the intrinsic R_2 relaxation rate for each amide ^1H , t is the total evolution time of the HSQC and R_2^{SP} the paramagnetic rate enhancement coming from the spin label.

In case of e.g. chemically coupled spin label the paramagnetic rate enhancement depends on the distance from each spin to the label in the following way:

$$r = \left[\frac{K}{R_2^{\text{SP}}} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_h^2 \tau_c^2} \right) \right]$$

where r is the distance between the electron and nuclear spin, τ_c is the correlation time for this electron-nuclear interaction, ω_h is the Larmor frequency of the proton and K is the physical constant, dependent on the type of the nucleus. For calculations, an approximation can be made that τ_c equals the global correlation time of the macromolecule (Battiste & Wagner, 2000; Solomon & Bloembergen, 1956).

For structure calculation purposes, r can be used as distance restraint from the hydrogen atom detected in the $^1\text{H}^{15}\text{N}$ -HSQC to the spin label on the macromolecule. As in the case of NOEs these restraints can be grouped into categories and used as distance ranges (Battiste & Wagner, 2000). Especially for larger molecules the information from PREs is very useful to determine the global fold. However, since the label itself is often flexible, several sets of spin label data at different sides of the macromolecule need to be utilized for accurate calculations. Surely, the PRE data can also be used qualitatively to validate the structure ensemble calculated using other restraint sets.

2.5. Protein dynamics by NMR.

The function of a macromolecule is determined not only by its fold, but also by the motions present in the molecule, its dynamics. These motions include among other things domain motions, folding and unfolding of structural elements, side chain rotations and vibrations of the chemical bonds. Internal motions as well as molecular tumbling happens on the picoseconds to nanoseconds timescale and can be studied by following the relaxation of NMR signals. There are multiple factors that induce spin relaxation. For investigation of protein dynamics, two relaxation pathways are usually monitored and the rate constants are measured.

The longitudinal relaxation (also called the spin-lattice relaxation) is induced by the interaction of the spins with their surrounding (the lattice). The lattice is assumed to have an infinite heat capacity and to be in thermal equilibrium at all times. It modifies the local magnetic fields at locations of the nuclei and weakly couples to the spin system. Stochastic Brownian rotational motions cause local fluctuating magnetic fields that induce transitions between spin states. Macroscopically they cause the recovery of the z-component of magnetization towards its thermal equilibrium after it was perturbed. Very often the recovery process is approximately exponential and can be characterized by a single time constant T_1 or the corresponding relaxation rate $R_1 = \frac{1}{T_1}$.

The transverse relaxation (also called the spin-spin relaxation) is caused by the interaction of the nuclear spins. Molecular tumbling in solution again causes spin-state transitions and is seen as the loss of coherence between the individual precessing spins in a sample. Macroscopically it is detected as the loss of the x- and y-magnetization. It also follows an exponential decay and is described by a single time constant T_2 or the relaxation rate $R_2 = \frac{1}{T_2}$.

Relaxation is a process dependent on rotational diffusion of the molecule (molecule tumbling) that is described by the rotational correlation time τ_c . Determination of the R_1 and R_2 constants for single amide nitrogens is a valuable tool to study dynamics of the protein backbone. It gives information on the size of the molecule as well as on the relative flexibility of each amino acid within the sequence (Kay et al, 1989). The R_2 relaxation rate is also influenced by chemical exchange and thus can help detecting and investigating the exchanging residues (Akke & Palmer, 1996; Kroenke et al, 1998).

The studies can be complemented with the ^1H - ^{15}N heteronuclear NOE experiment that gives information about the motion of individual N-H bond vectors. Those that undergo motion faster than the overall tumbling of the molecules show a decreased NOE intensity relative to the average observed for the majority of the residues. This gives information on the local rigidity of a residue, indicating if it is involved in higher order structures or not. The average value for the ^1H - ^{15}N heteronuclear NOE ratio is approximately 0.77, higher values indicating that the N-H vector is rigid with respect to the rest of the protein and lower values for parts with increased backbone mobility (Kay et al, 1989).

2.6. Determination of molecular size by NMR.

NMR can be used to measure self-diffusion constants (diffusion coefficients) of macromolecules with an accuracy of up to 1%. Diffusion coefficient is an important physicochemical property of a molecule, that can be correlated to its size and thus, NMR can be utilized to detect different molecular weight macromolecules present in solution at small concentrations. The method is called Diffusion Ordered NMR Spectroscopy (DOSY) and is capable of determining diffusion coefficients of molecules without a concentration gradient (Johnson, 1999). Diffusion rates are measured using gradient pulses with different gradient strength. The gradients eliminate more signal intensity of molecules at faster motion. That is why the intensity of detected proton signal belonging to a certain molecule at a certain gradient level is dependent mostly on its rate of self diffusion in its environment. The analysis of attenuation of the signal intensity as a function of gradient strength, diffusion time or gradient time (depending on the particular experimental setup) can be correlated to the diffusion coefficient (Ambrus & Yang, 2007). Comparing the determined value to a reference setup, the molecular weight can be calculated (He & Niemeyer, 2003).

2.7. Ligand binding studies by NMR.

Apart from other applications, an HSQC can be used to monitor the folding state of a protein under certain conditions. Particularly, a series of ^1H ^{15}N -HSQC experiments can be used to investigate biomolecular interactions with an amino acid resolution. Binding of the molecules changes the chemical environment of the interacting protein residues, which causes chemical shift perturbations of the NH signals in the ^1H ^{15}N -HSQC spectrum of a ^{15}N -labelled protein. As every peak in the ^1H ^{15}N -HSQC represents a single H^{N} -N correlation of the amide groups

and NH₂ groups of the side chains, amino acids affected by the molecular interaction can be unambiguously identified. A titration of an unlabelled compound to a ¹⁵N-labelled protein probe, followed up by a series of ¹H¹⁵N-HSQC experiments with the same parameters, can be used not only to map the molecular interaction site but also to determine the mode of binding and to estimate of the binding constant.

The line width and Larmor frequencies are dependent on chemical exchange (Craik, 1996). Thus, depending on the exchange regime, the effects introduced by ligand binding in the ¹H¹⁵N-HSQC spectrum will have different manifestations. In the system of a molecule (M) interacting with ligand (L)



where k_{on} and k_{off} are the respective association and dissociation rates, the exchange constant k_{ex} is defined as

$$k_{ex} = k_{on} + k_{off}$$

On the NMR timescale the system is referred to as in

Fast exchange: if $k_{ex} \gg \Delta\omega$

Intermediate exchange: if $k_{ex} \approx \Delta\omega$

Slow exchange: if $k_{ex} \ll \Delta\omega$

where $\Delta\omega$ is the difference in frequencies of the same resonance in the free and bound state.

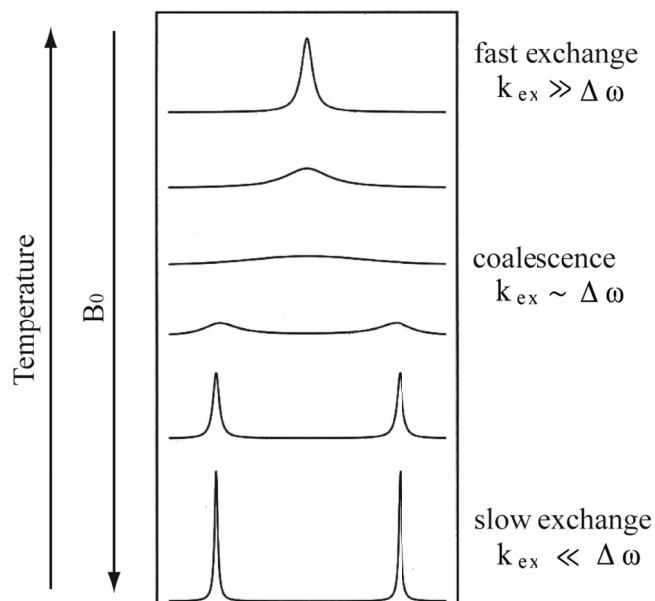


Fig. 14 Effect of exchange rate, temperature and magnetic field strength on the line shape and chemical shifts observed in an NMR spectrum. B_0 is the strength of the chemical shift, k_{ex} the exchange rate constant, $\Delta\omega$ is the frequency difference of a resonance in the free and bound state. Adapted from (Cavanagh J., 1996).

In a fast exchange regime, the peak position upon addition of the ligand will correspond to the average chemical shift (Fig. 14), weighted by the relative amounts of free and bound macromolecule. In an intermediate exchange regime, the peaks of the free state will first broaden until coalescence upon addition of the ligand and then recover upon saturation at the new bound position. In a slow exchange regime, the population of the free and bound state can be observed simultaneously, whereas the area under the peaks corresponds to their relative amounts of ligand to macromolecule. Temperature and magnetic field have an effect on the exchange regime. The dissociation constant can be extracted by fitting the titration curve, while complex is in a fast exchange regime. Additionally, there have been several other methods developed (Fielding, 2007).

More details on the theoretical and practical aspects of the NMR spectroscopy can be found in the following books:

J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance, N.J. Skelton “Protein NMR Spectroscopy. Principles and Practice.” Elsevier 2007;

D. J. Craik “NMR in Drug Design” CRC Press, 1996

J. Evans “Biomolecular NMR Spectroscopy”, Oxford University Press 1995;

P. J. Hore, J. A. Jones, S. Wimperis “NMR: The Toolkit” Oxford University Press 2001;

J. Keeler “Understanding NMR Spectroscopy”, Wiley 2005;

M. J. Levitt “Spin Dynamics. Basics of Nuclear Magnetic Resonance”, Wiley 2003;

Scope of the Thesis.

Regulation of ribosomal RNA synthesis is the first step in cellular response to external stimuli and affects cell survival. Errors in this tightly regulated process lead to severe failures in cell functionality like cellular senescence and rDNA instability (Kobayashi et al, 2004; Peng & Karpen, 2007). Overexpression of eukaryotic rRNA genes is associated with development of different types of cancer (Montanaro et al, 2008; White, 2008; Williamson et al, 2006).

To resist this, in a healthy cell the major part of its rRNA genes is silenced. Long term epigenetic silencing of rRNA genes is mediated by the nucleolar remodeling complex NoRC. It recruits a number of chromatin remodeling enzymes to rDNA genes and thereby induces compaction of rDNA chromatin and heterochromatin formation (Grummt, 2010; McStay & Grummt, 2008). To exert its function, NoRC requires a direct interaction with the promoter associated RNA (pRNA), which is transcribed from intergenic spacer regions within the rDNA genes (Mayer et al, 2006). The NoRC pRNA interaction is mediated by the MBD-like domain of the TIP5 subunit of the remodeling complex, the TAM domain.

The TAM/pRNA interaction plays a major role in the regulation of eukaryotic gene expression. However, the molecular details of this interaction have not been described yet. The importance of this interaction in context of larger complexes for rDNA gene silencing has been reported (Mayer et al, 2006; Strohner et al, 2001). Although several structures of the related MBD domain bound to DNA have been described (Ohki et al, 2001; Scarsdale et al, 2011), the structure of the TAM domain and its pRNA interaction is unknown. The questions of what the exact boundaries and the three dimensional structure of a functional TAM domain is, how the three dimensional structure of the pRNA looks like, how TAM contacts the pRNA and how the specificity of the interaction is provided, remain unclear.

The goal of this thesis is, to perform structural and biochemical studies of the TAM/pRNA interaction, by starting from an analysis of the individual binding partners. Biochemical investigations were performed partially in collaboration with the laboratory of Ingrid Grummt at the German Cancer Research Centre (Heidelberg, Germany). Nuclear magnetic resonance (NMR) spectroscopy was employed for structure determination of the TAM domain and for studying protein-RNA interaction in solution.

Materials and Methods.

3. Materials.

3.1. Sequences

3.1.1. Protein expression

CONSTRUCT	TYPE	COMMENTS
<i>M. musculus</i> TIP5	DNA sequence	accession number NCBI AJ309544
<i>H. sapiens</i> TIP5	DNA sequence	accession number NCBI NM_013449
TIP4N (<i>H. sapiens</i> TIP5 ₄₉₆₋₆₆₄)	DNA sequence Protein sequence	codon optimized for <i>E. coli</i> expression
TAM ₅₁₆₋₆₂₃ (<i>H. sapiens</i> TIP5 ₅₁₆₋₆₂₃)	DNA sequence Protein sequence	codon optimized for <i>E. coli</i> expression

See Appendix.

3.1.2. DNA templates for *in vitro* RNA transcription

RNA CONSTRUCT	DNA TEMPLATE	OLIGONUCLEOTIDES
1wt	Single stranded	T7 top/1wt bottom
1wt	Double stranded	1wt top/1wt bottom methoxylated
2sh	Single stranded	T7 top/2sh bottom
“short”	Double stranded	short top/short bottom methoxylated

See Appendix.

3.2. Chemicals

CHEMICAL	SUPPLIER
¹³ C labeled Glucose, 98% pure	euriso-top
¹⁵ N labeled Ammonium chloride, 99% pure	euriso-top
3-(2-Iodoacetamido)-PROXYL	Sigma-Aldrich
30% Acrylamide/Bis-acrylamide (37.5:1)	SERVA
40% Acrylamide/Bis-acrylamide (19:1)	SERVA
40% Acrylamide/Bis-acrylamide (29:1)	SERVA
Adenosine 5'-triphosphate (ATP) disodium salt hydrate	Sigma-Aldrich
Agarose	VWR
Ammonium peroxodisulfate (APS)	Merck
Ampicillin	SERVA
Ascorbic acid	Sigma
Bacto tryptone	Merck
Biotin	AppliChem

Boric acid (H ₃ BO ₃)	Merck
Bromophenol blue	Sigma-Aldrich
C12E6-poly(ethylene glycol)	Sigma
Cobalt(II) chloride hexahydrate	Roth
Copper(II) chloride dihydrate	Merck
Cytidine 5'-triphosphate (CTP) disodium salt	Sigma-Aldrich
D ₂ O	euriso-top
Desoxyribonucleotides	Promega
Disodium hydrogen phosphate	Merck
Dithiothreitol (DTT)	Roth
DMSO	euriso-top
Ethanol absolute	VWR
Ethylendiamintetra-acetate (EDTA)	Roth
Ewald gelatine	EWALD-GELATINE GmbH
Glycerol	VWR
Guanosine 5'-triphosphate (GTP) sodium salt hydrate	Sigma-Aldrich
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	CALBIOCHEM
n-Hexanol	Fluka
Igepal CA-630	Sigma-Aldrich
Imidazole	Merck
Iron(III) chloride hexahydrate	Merck
Isopropyl b-D-thiogalactoside (IPTG)	Roth
Kanamycin	SERVA
Magnesium chloride hexahydrate	Merck
Manganese(II) chloride hexahydrate	Merck
β-Mercaptoethanol	Roth
Nickel-Nitrilotriacetate (Ni-NTA)	QIAGEN
Pefabloc SC	SERVA
Polyethylenglycol 8000 (PEG 8000)	Promega
Ribonucleotides	Sigma
RNase AWAY	Molecular BioProducts
Sodium azide	Merck
Sodium carbonate	Merck
Sodium chloride	Merck
Sodium dihydrogen phosphate	Merck
Sodium dodecyl sulphate (SDS)	Merck
Spermidine	Sigma
N,N,N',N'-Tetramethylethylen-Diamin (TEMED)	Fluka
Thiamin	AppliChem
TritonX-100	Sigma
Trizma [®] Base (Tris)	Merck
Trizma [®] hydrochloride (Tris-HCl)	Merck
Uridine 5'-triphosphate (UTP) trisodium salt hydrate	Sigma-Aldrich

Urea	VWR
Xylene cyanole	Sigma-Aldrich
Yeast extract	Merck
Zinc chloride	Merck

3.3. Enzymes

ENZYME	CONC.	SUPPLIER
Acc65	10U/ μ l	NEB
NcoI	10U/ μ l	NEB
AflIII	20U/ μ l	NEB
EcoRI	20U/ μ l	NEB
DpnI	10U/ μ l	Promega
Taq DNA polymerase	5U/ μ l	Promega
Pfu DNA polymerase	3U/ μ l	Promega
DNase I	1mg/ml	SERVA
T4 lysozyme	crystalline	SERVA
TEV protease	1mg/ml	In house production
T7 polymerase	10mg/ml	HMGU, Protein Expression and Purification Facility
Antarctic phosphatase	5U/ μ l	NEB
T4 DNA ligase	3U/ μ l	Promega

3.4. Bacterial strains and vectors

3.4.1. Bacterial strains

<i>E.coli</i> STRAIN	GENOTYPE
BL21 (DE3)	F ⁻ ompT gal dcm lon hsdS _B (r _B ⁻ m _B ⁻) λ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])
DH5 α	F ⁻ , endA1, hsdR17 (rk ⁻ , mk ⁺), supE44, thi-1, recA1, gyrA96, relA1, s80dlacZ M15
XL1-Blue	endA1 gyrA96(nal ^R) thi-1 recA1 relA1 lac glnV44 F' ⁺ [::Tn10 proAB ⁺ lacI ^q Δ (lacZ)M15] hsdR17(r _K ⁻ m _K ⁺)

3.4.2. Vectors.

VECTOR	ORGANISM
pETM-11	<i>E.coli</i>
pcDNA4/TO	<i>M. musculus</i> , <i>H. sapiens</i>

3.5. Recipes

3.5.1. Buffers

BUFFER	COMPONENTS
TIP4N PURIFICATION	
lysis buffer (low salt, low imidazole)	50mM HEPES, 1M NaCl, 10mM Imidazole, 0.02% NaN ₃ , 2Mm β-Mercaptoethanol, pH=8.0
lysis buffer (high salt, low imidazole)	50mM HEPES, 1.8M NaCl, 10mM Imidazole, 0.02% NaN ₃ , 2Mm β-Mercaptoethanol, pH=8.0
elution buffer (high imidazole)	50mM HEPES, 1M NaCl, 300mM Imidazole, 0.02% NaN ₃ , 2Mm β-Mercaptoethanol, pH=8.0
TAM PURIFICATION	
lysis buffer (low salt, low imidazole)	50mM HEPES, 0.5M NaCl, 10mM Imidazole, 0.02% NaN ₃ , 2Mm β-Mercaptoethanol, pH=8.0
lysis buffer (high salt, low imidazole)	50mM HEPES, 1M NaCl, 10mM Imidazole, 0.02% NaN ₃ , 2Mm β-Mercaptoethanol, pH=8.0
elution buffer (high imidazole)	50mM HEPES, 0.5M NaCl, 300mM Imidazole, 0.02% NaN ₃ , 2Mm β-Mercaptoethanol, pH=8.0
NMR BUFFER	
NMR buffer	20mM Sodium phosphate buffer, pH=7.0; 50mM NaCl, 1mM DTT
RNA TRANSCRIPTION	
10x Trx-buffer	400mM TRIS-HCl, pH=8.1 at 37°C; 10mM Spermidine, 1% Triton-X 100, 50mM DTT
Tris-Borate-EDTA Buffer (TBE)	89mM TRIS, 89mM Boric acid, 2mM EDTA, pH=8.0
SPIN LABEL ATTACHMENT	
PRE reaction buffer	200mM TRIS, pH=8.0; 50mM NaCl

3.5.2. Bacterial growth media

MEDIUM	COMPONENTS PER LITRE
lysogeny broth (LB) rich medium	10g Bacto Tryptone, 5g yeast extract, 10g NaCl; pH=7.5 1ml antibiotic (50mg/ml)
¹⁵ N labeling medium	100ml M9 medium (10x), 10ml TES (100x), 20ml 20%(w/v) glucose, 1ml 1M MgSO ₄ , 0.3ml 1M CaCl ₂ , 1ml Biotin (1mg/ml), 1ml Thiamin (1mg/ml), 1ml antibiotic (50mg/ml)
¹³ C ¹⁵ N labeling medium	100ml M9 medium (10x), 10ml TES (100x), 2g ¹³ C-glucose, 1ml 1M MgSO ₄ , 0.3ml 1M CaCl ₂ ,

1ml Biotin (1mg/ml), 1ml Thiamin (1mg/ml), 1ml antibiotic
(50mg/ml)

M9 MEDIUM 10X	MASS PER LITRE
Na ₂ HPO ₄	60g
KH ₂ PO ₄	30g
NaCl	5g
¹⁵ N NH ₄ Cl	5g

TRACE ELEMENTS SOLUTION (TES)	MASS PER LITRE
EDTA	5g
FeCl ₃ x6H ₂ O	0.83g
ZnCl ₂	84mg
CuCl ₂ x2H ₂ O	13mg
CoCl ₂ x6H ₂ O	10mg
H ₃ BO ₃	10mg
MnCl ₂ x6H ₂ O	1.6mg

3.5.3. RNA gel electrophoresis.

Denaturing PAGE

SOLUTION	COMPONENTS
Running buffer	1xTBE
Separating gel stock solution (500ml)	250ml 40% Acrylamide/Bis-acrylamide (19:1), 210g Urea, 50ml 10xTBE, 50ml H ₂ O
Separating gel (150ml) 20%	150ml stock solution, 1.2ml 10% APS, 58µl TEMED
Loading dye (10ml)	4.8g Urea, 4ml 10xTBE, Bromophenol blue, Xylene cyanol

Non-denaturing PAGE

SOLUTION	COMPONENTS
Running buffer	0.5xTBE
Separating gel (25ml) 6%	6.3ml 0.5xTBE, 3.7ml 40% Acrylamide/Bis-acrylamide (29:1), 250µl 10% APS, 20µl TEMED, 14.75ml H ₂ O
Loading dye 2x (10ml)	10ml 10xTBE, 10%w/v glycerol, Bromophenol blue, Xylene cyanol

4. Methods I. Molecular Biology and Biochemistry.

4.1. Cloning, expression and purification of *H. sapiens* TIP5 constructs.

4.1.1. Cloning of *H. sapiens* TIP5 truncation constructs.

The TIP5 truncation construct from *H. sapiens* comprising the residues 496-664 (later called TIP4N) was subcloned from a bacterial vector bearing the *H. Sapiens* TIP5₃₃₂₋₇₃₂ protein. The source plasmid was received from the lab of Prof. Dr. I. Grummt, Division of Molecular Biology of the Cell II, dkfz. While cloning, the DNA sequence of TIP4N was also optimized for higher expression yields, by mutating *E. coli* – specific rare codons to the *E. coli* – specific codons with frequent usage. Starting from the optimized TIP4N construct further truncations were produced, including the one comprising the residues 516-623 (further called TAM₅₁₆₋₆₂₃). The sequences of TIP4N and TAM₅₁₆₋₆₂₃ are given in Appendix.

The template DNA was amplified by Taq DNA polymerase using a standard polymerase chain reaction (PCR) and sequence specific primers purchased from MWG-Biotech AG. In order to create compatible overhangs, the primers contained distinct restriction sites to guarantee the ligation in the linearised vector. The forward primers were designed with an NcoI restriction site; the reverse primers contained an Acc65I restriction site. The gene fragments were then inserted into a modified pETM-11 vector (European Molecular Biology Laboratory), containing an N-terminal His₆-tag for purification followed by a GB1-tag for increased expression and solubility of the protein, which was cut with the same enzymes and dephosphorylated. The tags were separated from the main protein sequence by a tobacco etch virus (TEV) protease cleavage site for convenient removal of the tag after purification (see Appendix).

The PCR mixture was prepared as follows:

COMPONENT	CONC.	VOL. (µl)
Template DNA	50-100ng/µl	1
fwd primer	10pmol/µl	2
rev primer	10pmol/µl	2
Taq DNA polymerase	5U/µl	0.25
PCR reaction Buffer/MgCl ₂	10X	5
dNTP mix	10mM each	1
add H ₂ O		50

The inserts were then amplified in a thermo cycler using the following program with a preheated lid (105°C):

STEP	TEMP. (°C)	DURATION	
Preheated lid	105	on	
Denaturation	94	2min	
Denaturation	94	5sec]
Annealing	54	5sec	22cycles
Elongation	72	30sec]
Final extension	72	10min	
Final hold	10	~	

The PCR products were subsequently purified via excision from a 1% agarose gel. After ligation, the plasmids were transformed by heat shock into chemically competent bacterial cells of the *E.coli* DH5 α strain for amplification and subsequently sequenced.

4.1.2. Cloning of TIP5 mutants. The QuickChange[®] PCR.

For mutagenesis analysis of pRNA binding by TIP5 the point mutations and the TIP4NG551A552 mutant were prepared utilizing the pETM-11 vector with the TIP4N-insert as the source clone for site-directed QuickChange[®] PCR. Overlapping site-specific primers containing the needed mutation were designed using the online QuickChange[®] primer design tool of Agilent Technologies and purchased from MWG-Biotech AG. Application of that technique allowed to conveniently engineer the point mutations directly into the plasmid; amplifying the whole plasmid sequence that could be transformed directly into the competent bacterial cells, omitting the ligation step. To increase the amplification fidelity, Pfu-polymerase which is capable of proof reading (Lundberg et al, 1991), was used.

PCR mixture was prepared as follows:

COMPONENT	CONC.	VOL. (μ l)
Template DNA	50-100ng/ μ l	1
fwd primer	10pmol/ μ l	1
rev primer	10pmol/ μ l	1
Pfu DNA polymerase	3U/ μ l	0.5
PCR reaction Buffer/MgCl ₂	10X	5
dNTP mix	10mM each	1
DMSO		0 / 2.5 / 5

To increase the selectivity of primer annealing pure DMSO was occasionally added to the PCR reaction to the final concentration of 5%v/v or 10%v/v.

The PCR reaction was performed in a thermo cycler using the following program with a preheated lid (105°C):

STEP	TEMP. (°C)	DURATION	
Preheated lid	105	on	
Denaturation	95	2min	
Denaturation	95	30sec]
Annealing	57	45sec	20cycles
Elongation	68	10min]
Final extension	68	10min	
Final hold	10	~	

Afterwards, 1µl Dpn I enzyme (10U/ µl) was added directly to the PCR mixture and incubated 1h at 37C to digest the methylated source plasmid DNA. Thereafter the enzyme was heat inactivated by 20min incubation at 80C. The result was checked on a 1% agarose DNA gel. Resulting clones were then transformed by heat shock into chemically competent bacteria of the *E.coli* XL1-Blue strain for amplification. The insertion of mutation was verified by sequencing.

For *in vivo* studies, the mutations were introduced into pcDNA4/TO eukaryotic expression vector, harboring the full length *M. musculus* TIP5 DNA sequence (see Appendix). The vector was received from the lab of Prof. Dr. I. Grummt, Division of Molecular Biology of the Cell II, dkfz. The QuickChange[®] PCR was performed as described above. Due to the big size of the vector (ca. 10.5 kb), to avoid occasional mutations in the DNA sequence, a 1.8 kb long fragment of the insert containing the needed inserted point mutation was cut out with single cutters AflIII and EcoRI and religated into the amplified vector backbone using the standard ligation and vector amplification techniques.

4.1.3. Protein expression.

Proteins were expressed at high yields in *E.coli* BL21 (DE3) strain in lysogeny broth (LB) rich medium and different minimal labeling media for selective isotope enrichment.

For expression, the plasmids bearing the protein constructs were transformed into chemically competent *E.coli* cells by heat shock. The transformation mixtures were plated on selective agar plates containing 50µg/ml antibiotic and incubated at 37°C over night. Single bacterial colonies were then inoculated into 50ml of LB medium, supplemented with 50µg/ml antibiotic, and grown in a shaker at 37°C over night. Next day, the bacterial pre-cultures were transferred into 500ml of LB/antibiotic medium (50µg/ml). To provide sufficient oxygen supply, bacteria were grown in 2 liter Erlenmeyer flasks in a shaker. The cells were grown at 37°C to an OD₆₀₀ of approximately 0.8. Afterwards, the temperature was reduced to 20°C and the protein expression was induced with 0.2mM IPTG. The cultures were incubated overnight and pelleted next day by centrifugation at 6000rpm for 20min at 4°C. After the removal of supernatant the pellets were snap-frozen in liquid nitrogen and stored at - 20°C.

For ¹⁵N labeling of proteins, *E.coli* BL21 cells were grown in the minimal medium containing 99% pure ¹⁵N ammonium chloride (1g/l) as the single nitrogen source. For better bacterial growth, the medium was supplemented with a mixture of trace elements, 1µg/l biotin and 1µg/l thiamine. In case of the double labeling of the proteins with ¹⁵N and ¹³C, the minimal ¹⁵N-medium additionally contained 98% pure ¹³C glucose (2g/l) as the single carbon source. For stereospecific assignment of methyl groups of valine and leucine by nonrandom ¹³C selective labeling after (Senn et al, 1989), a mixture of roughly 90% ¹²C₆-glucose and 10% ¹³C₆-glucose was used as the carbon source. The overnight pre-cultures were still grown in LB medium. However, to achieve uniform labeling, next day bacteria were first gently pelleted by centrifugation at 4000rpm, 20°C for 10min and washed with minimal medium, before resuspending them in 500ml minimal medium for further growth and expression.

In the case where at least 50% random deuteration of a protein had to be achieved, the minimal media were based on 90% pure D₂O instead of H₂O and contained non-deuterated salts and antibiotics. Since bacteria grew slowly in D₂O-based environment, the LB pre-cultures were started from a single colony in the morning, incubated at 37°C on shaker for at least 8h, pelleted gently by centrifugation at 20°C and transferred to 50ml of D₂O-based minimal medium. The D₂O-precultures were then incubated at 37°C on shaker overnight and transferred next day completely into 450ml of D₂O-based minimal medium for expression. After induction the cultures were grown, until they reached the OD₆₀₀ of ca. 2.5-3.

4.1.4. Purification of TIP4N and TIP4N-based mutants.

After thawing the bacterial pellets at room temperature, the cells were resuspended in 20ml of cold lysis buffer, containing 150 μ l DNase I (1mg/ml), Pefabloc[®] SC and lysozyme and incubated 30min on ice. After disruption of the bacterial cell wall with a microfluidizer and ultracentrifugation for 30min at 4C and 14000g, 0.2% w/v IgepalCA-630 was added to the supernatant, which was then filtered through a paper filter. The protein was bound to Ni²⁺ resin (Ni-NTA Agarose) and subsequently washed with low- and high-salt lysis buffer. Afterwards the protein was eluted from the column with elution buffer, containing 300mM imidazole. The GB1 expression tag was cleaved with the tobacco etch virus (TEV) protease at 4C overnight using ca. 1 μ g of the enzyme per 10mg of protein.

The samples were then further purified, removing the expression tag at the same time, by size-exclusion chromatography (HiLoad 26/60 Superdex75 from GE Healthcare) in the elution buffer. Concentration of the sample at that stage lead to severe protein precipitation, that is why the protein solution was loaded on the column without concentrating, using the maximal allowed load volume. The fractions containing the protein, but not the expression tag, were concentrated in Amicon[®] Ultra centrifugal filter units with MWCO 10kDa purchased from MERCK MILLIPORE. The buffer was then exchanged to NMR buffer using either HiPrep 26/10 or disposable PD-10 desalting columns (from GE Healthcare). The protein samples were stored in NMR buffer at 4C without further concentration. For buffer receipts see 3.5.1.

4.1.5. Purification of TAM domain constructs.

The purification of the TAM domain constructs up to the cleavage of the GB1 expression tag was described in 4.1.4. Due to better stability of the protein, buffers with lower salt concentrations were used (see 3.5.1). After cleavage of the expression tag by the TEV protease, the elution buffer was exchanged back to low imidazole lysis buffer and the tag was removed by a second Ni²⁺ affinity column. The samples were then further purified by size-exclusion chromatography (HiLoad 26/60 Superdex75, from GE Healthcare) in the lysis buffer. To avoid precipitation of the protein the maximal allowed load volume was again used for the gel filtration column. The fractions, containing the protein were concentrated in Amicon[®] Ultra centrifugal filter units with MWCO 3kDa (MERCK MILLIPORE). The buffer was exchanged to NMR buffer using either HiPrep 26/10 or disposable PD-10

desalting columns (from GE Healthcare) and samples were stored in NMR buffer without further concentration either at 4C or frozen in liquid nitrogen and stored at -20C.

4.1.6. Refolding of the TAM₅₁₆₋₆₂₃ domain from inclusion bodies.

To increase the protein yield, labeled TAM₅₁₆₋₆₂₃ domain was additionally purified by refolding from inclusion bodies. After bacterial lysis and separation of the supernatant, the pellet was dissolved at 37C on shaker overnight in 30ml TAM lysis buffer, containing 6M guanidinium hydrochloride. Next day, the supernatant was separated from the cell debris by ultracentrifugation (30min, 20C, 14000g), filtered and the protein was bound to the Ni²⁺ affinity column (Ni-NTA Agarose), equilibrated with TAM lysis buffer, containing 6M guanidinium hydrochloride.. The columns were subsequently washed with TAM lysis buffer, containing decreasing concentrations of guanidinium hydrochloride (4M, 2M and 0M). After washing the column additionally with high salt lysis buffer containing 1M NaCl, the protein was eluted with high imidazole elution buffer containing 300mM imidazole. The buffer volumes were always chosen to be three times the bed volume of the Ni²⁺ resin.

To remove last traces of guanidinium hydrochloride, the elution buffer was exchanged over disposable PD-10 desalting columns (GE Healthcare) to the TAM lysis buffer; and the expression tag was cleaved off by tobacco etch virus (TEV) protease overnight at 4C using ca.1µg of enzyme per 10mg of protein. Next day the samples were further purified by size-exclusion chromatography (HiLoad 26/60 Superdex75) in lysis buffer and prepared for the experiments as described above in 4.1.5. The proper fold of the constructs was confirmed by ¹H-¹⁵N HSQC.

4.2. Alignment media for measurement of residual dipolar couplings (RDCs).

Residual dipolar couplings were measured for the TAM₅₁₆₋₆₂₃ domain in different alignment media. The following media were tested:

	Alignment medium
1	Pfl - phages (10mg/ml)
2	strained Ewald gelatine gel 5%
3	Otting medium (5% C12E6/hexanol, n(C12E6)/n(hexanol)=0.64)
4	3:1 DMPC:DHPC bicelles 4%w/v
5	Polyacrylamide gel (37% acrylamide, 1% Bis-acrylamide) 7%

4.2.1. Pfl – bacteriophages.

The Pfl phages were purchased from ASLA^{biotech} in a concentration of 52mg/ml in 10mM potassium phosphate buffer, pH=7.6. To achieve alignment and approximately 10Hz splitting of the deuterium signal in NMR, the suspension of phages was added directly to the protein NMR sample in NMR buffer supplemented with 10% D₂O to the final concentration of 10mg/ml. TAM₅₁₆₋₆₂₃ was used at the concentration of 0.4mM.

4.2.2. Strained Ewald gelatine gel.

Ewald gelatine was weighed and added to the TAM₅₁₆₋₆₂₃ sample in NMR buffer/10% D₂O to the final concentration of 5%w/v. The mixture was heated for ca. 3min at 35C until the gelatine completely dissolved, gently mixed and poured into the flexible silicon tube of the gel stretching apparatus (Kummerloewe et al, 2008), while still liquid. Then the sample was cooled down at 4C for 30min to let the gelatine polymerize. After inserting the flexible silicon tube into the apparatus, the gel was stretched and inserted into an NMR spectrometer for measuring. A maximal splitting of 8.5Hz of the deuterium signal could be achieved.

4.2.3. Otting medium.

Based on the article of M. Rückert and G. Otting, a stable alignment of biomolecules in nonionic liquid crystalline media at 20C can be achieved in a mixture of 5%w/v final concentration of C12E6-poly(ethylene glycol) and hexanol at the molar ratio of C12E6-poly(ethylene glycol)/hexanol of 0.64 (Rückert & Otting, 2000). For 250µl sample, 12.5mg of C12E6-poly(ethylene glycol) were weighed and dissolved in the NMR buffer containing 10% D₂O. Since the medium is very sensitive to temperature changes, all the following steps were performed directly in the spectrometer room. n-Hexanol was added to the C12E6-poly(ethylene glycol) mixture in steps of 1µl to the final volume of ca. 5.4µl while vortexing vigorously after each step. Phase transitions were carefully observed. After addition of the first 1-2µl of n-hexanol the solution turned milky, followed by a transition to translucent at a certain volume of the alcohol. At that moment the addition of n-hexanol was stopped and the medium was left in the spectrometer room overnight to test the phase-stability. Next, the NMR spectrometer was precooled to 20C to avoid possible collapse of the liquid crystalline phase due to temperature differences in the sample lift channel. The splitting of the deuterium signal was measured. Complete alignment of the medium was achieved almost immediately after inserting the sample into magnet. In case of a stable phase, the deuterium signal showed

ca.28Hz splitting with ca.2Hz signal width at half height. The medium was stable over up to two weeks. The protein was concentrated in 50 μ l to the maximal possible concentration in the NMR buffer containing 10% D₂O and equilibrated to the room temperature. The tested Otting medium was slowly added to it and carefully mixed before measurement at 20C.

4.2.4. 1,2-Dimyristoyl-*sn*-Glycero-3-Phosphocholine (DMPC) 1,2-Dihexanoyl-*sn*-Glycero-3-Phosphocholine (DHPC) bicelles.

Before preparing this alignment medium, all solutions were sterile filtered and possibly supplemented with 0.02%w/v NaN₃ to avoid small particles and bacterial growth. The alignment in this medium was temperature and concentration dependent and could be measured only in the temperature range of 30-32C and concentration range of 3.5-5%w/v. The bicelles were prepared as described by Fleming, K. and Matthews, S. (Downing, 2004).

The ready-to-use mixture of DMPC and DHPC lipids with the molar ratio of 3:1 respectively was purchased from Avanti[®] Polar Lipids, Inc. and dissolved in NMR buffer containing 10% D₂O to the final concentration of 15%w/v (50mg of lipid mixture dissolved in 280 μ l of buffer). To form the bicelles, the mixture was vortexed for 10min at 4C, followed by incubation for 15min at 4C. Afterwards, it was vortexed again at room temperature for 1min and placed in a water bath at 38C for 30min. It was allowed to cool down at 4C for 15min and vortexed again at 4C for 30min. Then it was subjected again to the warm up step at room temperature for 15min and 30min incubation in the water bath at 38C. The described above steps of warming up and cooling down, each followed by extensive vortexing, were repeated three times. The effect was detectable by the increased viscosity of the lipid mixture at elevated temperatures. The dissolved lipid mixture could be frozen at -20C and stored. For reformation of the bicelles, several cycles of cooling down and warming up had to be repeated after thawing it.

The bicelles were slowly added to the protein sample in NMR buffer containing 10% D₂O to the final concentration of 4%w/v and measured at 30C. The splitting of deuterium signal of ca. 5Hz was achieved. Complete alignment of the medium was achieved almost immediately after inserting the sample in the spectrometer. The samples were stable for ca. 7 days.

4.2.5. Polyacrylamide gels.

The degree of alignment in polyacrylamide gels is generally not dependent on the measurement temperature and relies on the concentration and ratio of acrylamide/bis-

acrylamide in the sample. Polyacrylamide gels of different percentage were prepared to induce alignment of the TAM domain, after (Ishii et al, 2001).

Polyacrylamide mixture was prepared as follows:

PERCENTAGE	CHEMICAL.	VOLUME. (μ l)
6%	30% Acrylamide/Bis-acrylamide (37:1)	200
	NMR buffer / 10%D ₂ O	800
	TEMED	1.2
	10% APS	12

For higher percentages of the gel, the acrylamide/bis-acrylamide volume was scaled up. While being still liquid, the pre-mixed solution was poured with a syringe into a 3.4mm diameter glass tube (sealed with parafilm at one end) to the height of 2.6cm. It was let at room temperature for ca. 1h for polymerization. The polymerized gel was then taken out of the tube and washed in 1l H₂O at 4C over night. Afterwards, it was transferred with tweezers to a parafilm and dried on a Petri dish overnight at 37C. Next day the shrunk gel was transferred to 250 μ l of protein NMR sample in a Shigemi tube. The tube was subsequently closed with a plunger, fixed at the height of 1.8cm from the sample-bottom, and left at 4C overnight to let the gel swell. Since the expansion of the gel was restricted at the top by the plunger, it was strained and thus, the partial alignment of the protein molecules was induced.

The induced splitting of the deuterium signal was ca. 2Hz. However, the dipolar couplings induced by the strained gel were in general bigger, up to ca.10Hz and could be estimated by comparing the nitrogen splitting for the same signals in a non-decoupled ¹H-¹⁵N HSQC for a reference and an aligned sample. Generally, 6-7% of acrylamide/bis-acrylamide are enough to induce the required alignment for a protein of ca. 15-20kDa. In case the required alignment is not detected, the concentration of bis-acrylamide can be increased. Different gel percentages were tested to get the best alignment. In case of TAM₅₁₆₋₆₂₃ the RDCs were measured in gels with 7% acrylamide/bis-acrylamide (ratio 37:1). Though NH₂ groups of acrylamide gave rise to strong signals in the ¹H-¹⁵N HSQC spectrum, the dipolar couplings could still be evaluated.

4.3. Attachment of a paramagnetic relaxation enhancement spin label.

The spin label 3-(2-Iodoacetamido)-PROXYL was chemically attached to the single cysteine C556 of the TAM₅₁₆₋₆₂₃ construct. Since the label was destabilizing the protein, all steps were performed at 4C.

By exchanging the buffer over a disposable PD-10 desalting column (GE Healthcare) TAM₅₁₆₋₆₂₃ was transferred to precooled NMR buffer containing 30mM dithiothreitol (DTT), and subsequently incubated at 4C for 1h. To perform the chemical attachment reaction, the buffer was exchanged over a disposable PD-10 desalting column (GE Healthcare) to the PRE reaction buffer. The reaction tube was wrapped in aluminium foil to keep the reaction solution in the dark and protect the active spin label at further reaction steps. After that, the spin label dissolved in methanol was added to the sample at 5 molar equivalents of the protein concentration. That was followed by a long incubation overnight at 4C to let the spin label attach. Next day the buffer was changed to the NMR buffer containing 10% D₂O but 0mM DTT using again a disposable PD-10 desalting column (GE Healthcare). To reduce the loss of material through precipitation, the sample was split into two parts. One part was used immediately to record PRE experiments with the spin label in the active state. The other part was mixed with a solution of ascorbic acid in NMR buffer containing 10% D₂O to the final concentration of 3 molar equivalents of the protein concentration. After incubation for 1h or overnight at 4C it was used to measure the reference spectrum for PRE with the spin label in the reduced state. The experiments were performed on a ¹⁵N labeled TAM₅₁₆₋₆₂₃ domain aiming a 0.2mM final concentration of the protein in 500µl.

4.4. *In vitro* RNA transcription.

Unlabelled RNA samples of 1wt and 2sh pRNA constructs were prepared by large scale *in vitro* transcription by a T7 RNA polymerase from a single stranded DNA synthetic oligonucleotide template, harboring the sequence reverse complimentary to the desired RNA at the 5' end, followed by the sequence reverse complimentary to T7 bacteriophage promoter. As the top strand, the T7 bacteriophage promoter was used (Lukavsky & Puglisi, 2004; Wyatt et al, 1991).

Apart from that, to reduce the number of abortive products and thus, increase the RNA yields, *in vitro* transcription for the 1wt and the “short” RNA construct was also performed from a

fully doublestranded synthetic DNA oligonucleotide, containing the doublestranded promoter sequence of the T7 bacteriophage at the 5' end, followed by DNA sequence complementary to the desired RNA (bottom strand). To reduce the number of n+1 products (where n is the number of the nucleotides in the desired RNA), the bottom oligonucleotide was methoxylated at the two last nucleotides at the 5' end. The sequence of the oligonucleotides is given in Appendix.

After preliminary small scale tests in 50µl final volume, where the concentration of MgCl₂ and of the ribonucleotide tri-phosphates (rNTPs) was optimized, the reaction was scaled up to 10ml.

The following MgCl₂ end concentrations were used for *in vitro* transcription:

RNA CONSTRUCT	DNA TEMPLATE	OLIGONUCLEOTIDES	MgCl ₂ CONCENTRATION
1wt	Single stranded	T7 top/1wt bottom	36mM
1wt	Double stranded	1wt top/1wt bottomMO*	12mM
2sh	Single stranded	T7 top/2sh bottom	52mM
“short”	Double stranded	short top/short bottomMO*	8mM

* MO –methoxylated at the two last nucleotides at 5' end.

At first, the top strand and the bottom strand synthetic oligonucleotides were annealed. For 10ml transcription the following mixture was prepared:

COMPONENT	CONC.	VOL. (µl)
Top strand	100pmol/µl	48
Bottom strand	100pmol/µl	40
MgCl ₂	400mM	100
H ₂ O		1000

The components were properly mixed, heated up for 5min at 60C and allowed to cool down at room temperature. During that time the transcription mixture was prepared as follows:

COMPONENT	CONC.	VOL. (μ l)
Trx-buffer	10x	1000
ATP	0.1M	400
CTP	0.1M	400
GTP	0.1M	480
UTP	0.1M	480
PEG 8000 ¹	500mg/ml	1600
MgCl ₂	400mM	*
H ₂ O		1952

¹PEG8000 was dissolved fresh in H₂O each time. * MgCl₂ was added to the final concentration deduced from optimization mixed with H₂O in the end volume of 2ml.

The components were carefully mixed, but not vortexed to avoid shearing of the proteins in the Trx buffer. After that the annealed cold template mixture was added to it. The transcription was started by adding 500 μ l of T7-polymerase (10mg/ml) and incubated for 3h at 37C. The pyrophosphate, produced as a by-product in the reaction, was removed by centrifugation for 10min at 14000g, 4C. The supernatant was split in two parts and both of them were mixed with 0.8 volumes of 3M sodium acetate pH=5.5 and 3 volumes of absolute ethanol and afterwards incubated at -20C at least over night (up to 48h). The next day, the precipitate containing the RNA, unused ribonucleotide tri-phosphates and proteins was separated from the supernatant by centrifugation for 1h at 14000g, 4C and dried on a lyophiliser for at least 2h.

The dried pellets were dissolved in 1.5ml loading buffer supplemented with bromophenol blue dye each and analyzed on 20% acrylamide/bis-acrylamide (19:1) denaturing RNA acrylamide gels of ca. 150ml volume. To separate the desired RNA from the abortive products and the n+1 products (here n is the number of the nucleotides in the desired RNA), the gels were run in TBE buffer hot at ca. 20W per gel for several hours, until the dye eluted from the bottom (at least 6h or over night). The RNA-band was then visualized by UV-shadowing with $\lambda = 254$ nm on a silica matrix fluorescent thin layer chromatography plate and cut out. The RNA was eluted from the gel at 4C and 150V in TBE buffer using the Elutrap electroelution system (purchased from Whatman, GE Healthcare). The RNA was subsequently dialyzed in three steps from high- to low-salt in dialysis buffer at 4C over 36h starting with 1M NaCl concentration, continuing with 0.5M NaCl and finally in 0M NaCl dialysis buffer. Each step lasted at least 12h. The RNA was then transferred to H₂O using

disposable PD-10 desalting columns (GE Healthcare) and dried in a lyophiliser. The RNA-powder was stored at -20C. Yields of up to 1mg for a 35 nucleotide long RNA were achieved.

Before using the RNA for NMR experiments, the formation of the stem-loop fold in solution was induced by denaturing RNA in the respective NMR buffer for 4min at 95C and then snap cooling it on ice for 15min.

4.5. Complex formation of TIP4N and 1wt RNA.

The needed amount of 1wt RNA was transferred to 1.5ml NMR buffer containing 10% D₂O using a disposable PD-10 desalting column (GE Healthcare). After heating up for 4min at 95C and then snap cooling it on ice for 15min, the RNA was slowly added to 33nmol of ¹⁵N-labelled TIP4N non uniformly deuterated to ca.50% in 2.5ml of cold NMR buffer containing 10% D₂O. The solution was gently mixed and left for 20min on ice to allow complex formation. Then the sample was concentrated to 220µl in an Amicon[®] Ultra centrifugal filter unit with MWCO 3kDa (MERCK MILLIPORE) yielding a concentration of TIP4N of ca. 0.15mM. The complex was measured in a Shigemi tube directly thereafter.

4.6. Circular dichroism.

Circular dichroism spectra were measured on a 20µM ¹⁵N-labelled TIP4N construct non uniformly deuterated to ca.50% and a 20µM ¹⁵N-labelled TIP4NG551A552 mutant in NMR buffer containing 100mM or 250mM NaCl concentration and 10%D₂O at 20C in a cell of 0.1cm length. The data was collected for the wavelength range of λ_{260} to λ_{190} with a pitch of 0.1nm and accumulation of 20 measurements. For evaluation circular dichroism spectra of the NMR buffer without protein at 100mM and 250mM NaCl were measured at the same conditions and their circular dichroism values were subtracted from the values of the protein. Molar ellipticity per residue was calculated manually and analyzed online by the K2D2 software (Perez-Iratxeta & Andrade-Navarro, 2008) using data pitch of 1nm.

5. Methods II. Nuclear Magnetic Resonance Spectroscopy.

5.1. Assignment of backbone resonances of the TIP4N protein construct.

NMR measurements were carried out on a 0.15mM uniformly labeled $^{15}\text{N}^{13}\text{C}$ TIP4N protein sample in NMR buffer containing 10% D_2O and on a 0.15mM uniformly labeled $^{15}\text{N}^{13}\text{C}$ TIP4N protein sample with 50% random deuteration in the same buffer. The spectra were recorded at 298K on Bruker 600MHz spectrometer equipped with a cryoprobe. They were processed with NMRPipe (Delaglio et al, 1995) and analyzed using SPARKY 3.110 software (Goddard & Kneller). HNCACB and HN(CO)CACB experiments were recorded on the TIP4N sample with random deuteration, whereas for the additional HNCA and CBCA(CO)NH experiments, the nondeuterated TIP4N sample was used. Connectivities were assigned manually, however their analysis was facilitated much by applying the MARS software (Jung & Zweckstetter, 2004).

5.2. Resonance assignment and structure determination of the TAM₅₁₆₋₆₂₃.

NMR measurements were carried out on 0.6mM uniformly labeled $^{15}\text{N}^{13}\text{C}$ TAM₅₁₆₋₆₂₃ protein sample in NMR buffer containing 10% D_2O . The spectra were recorded at 293K on Bruker 600MHz and Bruker 900MHz spectrometers equipped with a cryoprobe, processed with NMRPipe (Delaglio et al, 1995) and analyzed using SPARKY 3.110 software (Goddard & Kneller).

Backbone resonance assignments of the TAM₅₁₆₋₆₂₃ were derived using the HNCACB spectrum of the, based on the known resonance values of backbone atoms of the TIP4N construct. The assignment process was supported by application of the MARS software (Jung & Zweckstetter, 2004). Carbon and proton resonances assignments of the TAM₅₁₆₋₆₂₃ side chains were performed using three-dimensional total correlation spectrometry (TOCSY). They were derived from two HCCH-TOCSY experiments with ^{13}C and ^1H evolution and correlated to amide group resonances by CC(CO)NH-TOCSY. Pro-S and pro-R methyl ^1H resonances of valine and leucine amino acids were stereospecifically assigned from a $^1\text{H}^{13}\text{C}$ -HQSC with 1024 points in ^{13}C dimension, using biosynthetically directed non random ^{13}C labeling of a 1mM fully ^{15}N labeled and 10% ^{13}C labeled TAM₅₁₆₋₆₂₃ sample in NMR buffer, containing 10% D_2O . The spectrum was recorded at 293K on a Bruker 750MHz spectrometer.

Based on this data intraresidual ^1H - ^1H NOEs were assigned manually in the ^{15}N -NOESY-HSQC and ^{13}C -NOESY-HSQC for the aliphatic amino acids, each with 70ms mixing time.

Intraresidual NOEs for the rings of the aromatic residues were assigned directly using ^3C -NOESY-HSQC (70ms mixing time) for the aromatic amino acids, supported by information on their ^{13}C resonances, derived from a 3D HCCH-TOCSY of the aromatic region and correlations of the aromatic $^{13}\text{C}\beta$ and $^1\text{H}\delta/\epsilon$ chemical shifts via scalar couplings (Yamazaki et al, 1993). Protonation state of the histidines was determined from the long range $^1\text{H}^{15}\text{N}$ -HSQC (Pelton et al, 1993) recorded at 300K on Bruker 600MHz spectrometer with a cryoprobe.

Automatic interresidual NOE assignment and derivation of the distance restraints was accomplished using CYANA 3.0 (Guntert, 2004). The dihedral angle restraints were predicted using TALOS+ (Shen et al, 2009). Residual dipolar coupling restraints were derived from $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ and $^{13}\text{C}'\text{-}^{15}\text{N}$ RDCs measured in the 4%w/v DMPC:DHPC (molar ratio 3:1) and 7% strained polyacrylamide gel alignment media on a Bruker 600MHz spectrometer with a cryoprobe.

200 structures were calculated using described above restraints and water-refined by ARIA1.2/CNS (Linge et al, 2003a; Linge et al, 2003b). Then 20 lowest energy structures were selected as the representative ensemble and used for quality and structure validation by iCing (Vuister & da Silva), including PROCHECK (Laskowski et al, 1993) and WHATCHECK (Vriend & Sander, 1993).

The consistency of the ensemble of the 20 lowest energy structures was also determined against RDC sets used in structure calculations. The quality of structure was additionally validated against the $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ residual coupling restraints, measured in the Otting medium omitted from the structure calculations. The paramagnetic relaxation enhancement data from a 3-(2-Iodoacetamido)-PROXYL spin label chemically attached to cysteine C556 were not used in structure calculation of the TAM domain, but instead implied in structure validation as well. Secondary structure definitions were based on DSSP algorithms (Kabsch & Sander, 1983) as implemented in PROCHECK (Laskowski et al, 1993) and PYMOL (Schrödinger) supplemented with manual inspection.

Electrostatic potential maps were calculated online using Poisson-Boltzmann electrostatics calculations. The parameters were set up by PDB2PQR pipeline (Dolinsky et al, 2007; Dolinsky et al, 2004) using the webpage <http://kryptonite.nbc.net/pdb2pqr/> and redirected for calculations to the online Adaptive Poisson-Boltzmann Solver (APBS) (Baker et al, 2001). The electrostatic isosurfaces were visualized by PYMOL (Schrödinger).

5.3. Calculation of secondary chemical shifts.

Secondary chemical shifts were calculated for the purpose of secondary structure prediction of the TIP4N and TAM₅₁₆₋₆₂₃ construct. For the calculation of the secondary chemical shift, the following formula was applied:

$$\Delta\delta C = (\delta C\alpha_{observed} - \delta C\alpha_{randomcoil}) - (\delta C\beta_{observed} - \delta C\beta_{randomcoil})$$

where $\Delta\delta C$ represents the combined secondary chemical shift, $\delta C\alpha$ and $\delta C\beta$ is the respective chemical shift of the α and β resonances of the protein backbone.

The $\delta C\alpha_{observed}$ and $\delta C\beta_{observed}$ values were extracted from the protein assignments. As chemical shifts are strongly dependent on the identity of the amino acid residue, the observed chemical shifts were normalized by subtraction of a random coil chemical shift ($\delta C\alpha_{randomcoil}$ and $\delta C\beta_{randomcoil}$) that reflects the amino acid-specific component, based on the Wishart random coil shift database (Wishart & Sykes, 1994; Wishart et al, 1992). For cysteines, the values of the reduced state were used.

5.4. Conservation analysis of the TAM domain residues.

Conservation analysis of the single amino acids of the TAM domain was performed using ConSurf web interface, a server for the identification of functional regions in proteins (Glaser et al, 2003; Landau et al, 2005). Using the lowest energy three dimensional structure of the TAM₅₁₆₋₆₂₃ a search for close homologous sequences was performed by PSI-BLAST (Altschul et al, 1997b), run with default settings as a module within ConSurf web interface. The conservation scores were calculated from the detected 20 sequences, implementing the empirical Bayesian method (Mayrose et al, 2004) and normalized by the program. The conservation scores were implemented as the B-factor into the pdb file of the lowest energy structure of the TAM₅₁₆₋₆₂₃ and plotted on it using the histogram mode with bins, representing the conservation categories determined by server. Visualization was done in PYMOL (Schrödinger).

5.5. Relaxation and dynamics analysis of the proteins.

To study dynamics properties of the protein backbone of the TIP4N construct, T_1 , $T_{1\rho}$ and ^1H - ^{15}N heteronuclear NOE experiments were recorded on a 0.2mM ^{15}N uniformly labeled

sample in NMR buffer with 10% D₂O at 298K on a Bruker 900MHz spectrometer equipped with a cryoprobe. The R₁ relaxation rate was extracted directly from the corresponding T₁ relaxation time. The R₂ relaxation rate was deduced from the T_{1ρ} relaxation time, based on the formula published by (Akke & Palmer, 1996). For the study of the dynamic properties of the TAM domain backbone, T₁, T₂ and ¹H-¹⁵N heteronuclear NOE experiments were recorded on a 0.6mM ¹⁵N¹³C uniformly labeled TAM₅₁₆₋₆₂₃ sample in NMR buffer with 10% D₂O at 293K on a Bruker 750MHz spectrometer. The relaxation rates were deduced directly from the corresponding relaxation times.

The T₁, T_{1ρ} and T₂ experiments were recorded in an interleaved fashion with different relaxation delays using the following settings were used:

TIP4N		TAM	
EXPERIMENT T ₁		EXPERIMENT T ₁	
DELAY	TIME (ms)	DELAY	TIME
1	21.6	1	432
2	399.6	2	21.6
3	520	3	345.6
4	248.4	4	86.4
5	1188	5	162
6	21.6	6	518.4
		7	248.4
		8	669.6
		9	885.6
		10	21.6
EXPERIMENT T _{1ρ}		EXPERIMENT T ₂	
DELAY	TIME (ms)	DELAY	TIME (ms)
1	5	1	5.4
2	40	2	21.6
3	15	3	27
4	25	4	16.2
5	55	5	32.4
6	5	6	5.4
7	40	7	27

The spectra were processed and analyzed with NMRPipe (Delaglio et al, 1995) using its NMRDraw graphical unit for visualization. The intensities of the peaks at different delays

were extracted from the respective peak volumes. The relaxation rates were fitted to a 2-parameter exponential decay from the intensities, with errors from duplicate experiments, calculated both from the standard deviation and the fit itself. The heteronuclear NOE rate was extracted as the ratio of peak intensities recorded with and without saturation. The error on the intensities was estimated from the intensity of experimental noise and propagated as the derivative of the ratio of two functions.

The correlation time of the molecule was estimated from the R_2/R_1 ratio using the quadratic representation approach (Bruschweiler et al, 1995) and compared to the average correlation time for the protein monomer, estimated from the number of residues (Daragan & Mayo, 1997).

5.6. Residual dipolar couplings.

Residual dipolar couplings were recorded on a 600MHz Bruker spectrometer equipped with cryoprobe on a uniformly $^{15}\text{N}^{13}\text{C}$ labeled TAM₅₁₆₋₆₂₃ in NMR buffer containing 10% D₂O. Following conditions were used (for preparation of the samples see 4.2):

Alignment medium	TAM ₅₁₆₋₆₂₃ concentration	Temperature	Measured couplings
4%w/v DMPC:DHPC bicelles	0.15mM	303K	$^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$, $^{13}\text{C}'\text{-}^{15}\text{N}$
7% Polyacrylamide gel	0.4mM	293K	$^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$, $^{13}\text{C}'\text{-}^{15}\text{N}$
Otting medium	0.05mM	293K	$^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$

As a reference, a non-aligned TAM₅₁₆₋₆₂₃ sample, recorded at the same temperature and concentration conditions on the same spectrometer, was used. The experiments were recorded in two-dimensional, doublet separated sensitivity enhanced HSQC experiments with (J) or (J+D) splitting in the nitrogen dimension.

The spectra were processed with NMRPipe (Delaglio et al, 1995). After manual inspection of the peak shape, the splitting was extracted from the peak position, determined automatically by SPARKY 3.110 (Goddard & Kneller). Based on the dynamics analysis of the protein backbone, described in 5.5, the residues in flexible protein regions and the ones experiencing a chemical exchange were excluded from further calculations. The distribution of the residual

dipolar couplings, that were intended to be used in structure calculation, was analyzed by the histogram method (Clare et al, 1998a).

For structure calculation, the initial values of the axial component and of the rhombicity of the molecular alignment tensor in the principal coordinate frame were estimated by MODULE 1.0 software (Dosset et al, 2001) and then given to CYANA 3.0 (Guntert, 2004) for the actual determination by the gridsearch approach. In further structure calculation, the RDCs were employed with a harmonic potential using an energy constant of $0.8 \text{ kcal mol}^{-1} \text{ Hz}^{-2}$ for the $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ RDCs and $0.4 \text{ kcal mol}^{-1} \text{ Hz}^{-2}$ for the $^{13}\text{C}'\text{-}^{15}\text{N}$ RDCs. Q-factors were calculated as defined by (Cornilescu et al, 1998). $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ RDCs measured in the Otting medium were not employed in structure calculations.

For validation of the calculated ensemble of the 20 lowest energy structures after water refinement, $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ and $^{13}\text{C}'\text{-}^{15}\text{N}$ RDC values were back calculated using MODULE 1.0 software (Dosset et al, 2001) and compared to the experimental values of RDCs of the rigid non-exchanging residues of the TAM₅₁₆₋₆₂₃. Additionally, the same procedure was performed with the $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ RDCs measured in the Otting medium to cross-validate the calculated ensemble. The Q factor of each fit determined as described in (Lipsitz & Tjandra, 2004).

5.7. Paramagnetic relaxation enhancement analysis.

Paramagnetic relaxation enhancement was measured on a uniformly ^{15}N labeled 0.12mM TAM₅₁₆₋₆₂₃ sample with a 3-(2-Iodoacetamido)-PROXYL spin label, chemically attached to the cysteine 556 in NMR buffer containing 10% D₂O. The $^1\text{H}\text{-}^{15}\text{N}$ -HSQC spectra were recorded at 293K on a Bruker 600MHz spectrometer with a cryoprobe. Since the label was destabilizing the protein, spectra were recorded on two distinct samples, one with the spin label in the reduced state, and the other one with the spin label in the oxidized state. The spectra were processed with NMRPipe (Delaglio et al, 1995) and analyzed in SPARKY 3.110 (Goddard & Kneller), extracting the intensity of the amide peaks as the height of the data. The ratio of the peak intensity in the oxidized and the reduced state of the label was calculated and plotted on the lowest energy structure of the TAM domain. $\frac{I_{ox}}{I_{red}}$ values (where I_{ox} and I_{red} are the signal intensities for amides in the $^1\text{H}\text{-}^{15}\text{N}$ -HSQC of the respective spin label state) were implemented as the B-factor into the pdb file of the lowest energy structure of TAM₅₁₆₋₆₂₃. The plot was created using the histogram mode and visualized in PYMOL (Schrödinger).

5.8. NMR diffusion measurements.

Diffusion measurements were performed on a uniformly ^{15}N -labelled 0.2mM TIP4N sample at 298K on a Bruker 900MHz spectrometer equipped with a cryoprobe. A pseudo two dimensional stimulated echo sequence for diffusion measurement, using bipolar gradients and WATERGATE (stebpgp1s19) from Bruker, was utilized. The experiment was recorded with 32 points, processed and analyzed using the built in T1/T2 relaxation unit of the Bruker TOPSPIN software. The signal was fitted by intensity for several well detectable resonances. The calculated diffusion coefficient was compared to the known diffusion coefficients of reference proteins (He & Niemeyer, 2003) to deduce the potential protein size.

5.9. Hydrogen to deuterium exchange experiments (H-D exchange).

The exchange of protein amide protons to deuterons was measured on a 0.23mM uniformly ^{15}N -labelled TAM₅₁₆₋₆₂₃ sample at 293K on a Bruker 600MHz machine equipped with a cryoprobe. The protein sample was concentrated to the required end concentration in 500 μl of NMR buffer containing 10% D₂O and split in two samples of 250 μl each. One sample was used to set up the experimental parameters on the spectrometer. The other sample was lyophilized overnight and after being was quickly transferred to the spectrometer used for a series of SOFAST HMQCs (Schanda & Brutscher, 2005) at different time intervals, however with the same parameters, immediately measured to detect also the quickly exchanging protons. The following time intervals were used:

Time upon addition of D ₂ O	
1	8min
2	15min
3	21min
4	30min
5	60min
6	90min
7	120min

The SOFAST HMQCs were recorded with 8 scans taking ca.5-6min per experiment. The spectra were processed with NMRPipe (Delaglio et al, 1995) and analyzed using SPARKY 3.110 software (Goddard & Kneller). The amide protons still present in the spectra after D₂O

addition, were plotted on the structure of the TAM₅₁₆₋₆₂₃ domain using PYMOL (Schrödinger).

5.10. Mapping of the pRNA binding. NMR chemical shift perturbation assays.

The first series of NMR chemical shift perturbation experiments of TIP4N upon addition of unlabelled 1wt RNA was measured at 298K on a uniformly ¹⁵N-labelled 0.08mM TIP4N sample in NMR buffer containing 10% D₂O. Unlabelled *in vitro* produced 1wt RNA at high concentration in NMR buffer containing 10% D₂O was added stepwise in 0.2 0.7 and 1 molar equivalents to the protein sample. The changes in chemical shift of the TIP4N amide groups upon addition of 1wt were monitored by a series of ¹H-¹⁵N HSQC experiments with the same settings starting from a reference spectrum with no RNA. The experiments were recorded on a Bruker 600MHz machine equipped with a cryoprobe. The spectra were processed with NMRPipe (Delaglio et al, 1995) and analyzed using SPARKY 3.110 software (Goddard & Kneller). Automatic peak picking algorithm of the program was utilized. Afterwards, the intensities of the peaks were extracted from peak volumes. For analysis, the ratio of $\frac{I}{I_0}$ was calculated, where I is the intensity of a peak in the ¹H-¹⁵N HSQC upon addition of the respective amount of 1wt RNA and I_0 is the corresponding peak intensity in the TIP4N reference ¹H-¹⁵N HSQC spectrum with no RNA. The ratio was plotted on the sequence of the TIP4N.

In the second series of experiments, the NMR chemical shift changes of TIP4N amides signals upon interaction with RNA were monitored on a complex of uniformly ¹⁵N-labelled and 50% randomly deuterated TIP4N with different molar ratios of unlabelled 1wt RNA, that was reconstituted *in vitro* (as described in 4.5). The protein was used at a final concentration of 0.15mM. The 1wt RNA was used at 0.5 1.2 and 2 molar equivalents. The chemical shift changes were monitored on complexes in NMR buffer containing 10% D₂O by ¹H-¹⁵N HSQC experiments recorded at 298K on a Bruker 600MHz machine equipped with a cryoprobe. The spectra were processed with NMRPipe (Delaglio et al, 1995) and analyzed using SPARKY 3.110 software (Goddard & Kneller). Automatic peak picking algorithm of the program was utilized to determine the exact position of the resonance. The chemical shift in nitrogen and hydrogen dimension was extracted and the change of it was calculated as

$$\Delta\delta_x = \sqrt{(\delta H_x - \delta H_0)^2 + \left(\frac{\delta N_x - \delta N_0}{5}\right)^2}$$

where δH_x and δN_x is the respective chemical shift value of a resonance in the spectrum of the complex and δH_0 and δN_0 is the respective chemical shift value of the same resonance in the reference spectrum of TIP4N alone. The factor $\frac{1}{5}$ was used to adjust the resolution in both dimensions. For analysis of the binding region the $\Delta\delta_x$ values were plotted on the sequence of the TIP4N. The amides, which $\Delta\delta_{1,2}$ was at least one standard deviation bigger than the average trimmed mean value of the chemical shift for that experiment, were considered experiencing a significant shift change and were plotted on the structure of the TAM₅₁₆₋₆₂₃ domain using PYMOL (Schrödinger).

5.11. NMR spectroscopy of pRNA constructs.

One dimensional NMR experiments on the 1wt, 2sh and the “short” construct were recorded in H₂O with 10% D₂O on unlabelled RNA at 277K on a Bruker 900MHz spectrometer equipped with a cryoprobe. An NMR pulse sequence with optimized delay for binomial water suppression was used to resolve the resonances of imino protons.

For structural investigation of 1wt RNA a two-dimensional homonuclear NOESY spectrum was recorded on a 0.4mM sample in 20mM sodium phosphate, pH=6.5, 10% D₂O buffer, at 277K on a Bruker 900MHz spectrometer, equipped with a cryoprobe. The mixing time was set to 300ms. The spectrum was used for investigation of exchangeable protons of the imino-imino and the imino-amino correlation regions and to perform the sequential assignment of imino-protons (the imino-walk) (Varani et al, 1996).

For investigation of the aromatic and the anomeric region of the 1wt RNA a ¹H¹H-TOCSY as well as a set of two-dimensional homonuclear NOESY experiments with 60ms, 100ms and 200ms mixing time was recorded on a 0.3mM 1wt sample in 20mM sodium phosphate in 100% D₂O at 300K on a Bruker 600MHz spectrometer, equipped with a cryoprobe. Additionally, a set of two-dimensional homonuclear NOESY experiments with 50ms and 100ms mixing time and two natural abundance ¹H¹³C-HSQC spectra for the region of sugar moieties of the nucleotides (offset 75ppm) and the nucleobases (offset 125ppm) were recorded on the same sample at the same conditions on a Bruker 900MHz machine, equipped with a cryoprobe.

For investigation of the “short” RNA construct, two sets of homonuclear two-dimensional homonuclear NOESY experiments were recorded on the Bruker 900MHz machine with a cryoprobe. One set of experiments was recorded at 277K in H₂O/10% D₂O (mixing time 80ms and 300ms) and in NMR buffer with 10% D₂O (mixing time 300ms) and another one at 298K in 100% D₂O (30ms and 80ms mixing time).

The spectra were processed with NMRPipe (Delaglio et al, 1995) and analyzed using SPARKY 3.110 software (Goddard & Kneller).

5.12. Multiple sequence alignment.

The multiple sequence alignment of proteins containing an MBD or an MBD-like domain was performed using MUSCLE algorithm (Edgar, 2004). For alignment the domain sequences were extracted from the sequence of full length proteins based on the boundaries determined in UniProtKB. To allow alignment of the complete domain and to avoid potential mistakes in each case ca.20-30 amino acids of the flanking sequence were additionally left at both the N- and the C-terminus to the predicted domain sequence.

Results.

6. Results I. Defining TAM domain boundaries.

Numbering of the protein constructs refers to the full length human TIP5 protein, isoform I. UniProtKB accession number Q9UIF9.

6.1. Multiple sequence alignment.

A multiple sequence alignment was performed using MUSCLE algorithm (Edgar, 2004) (Fig. 15). To date, there are six methyl-CpG binding domains known in human. First, the sequence of human TIP5 (BAZ2A) was compared to the sequences of human canonical MBD domains (Fig. 15A). The comparison showed that even though being identified as an MBD-like domain, the human TAM domain is not very similar in sequence to the canonical MBDs in human. It even has an insert of approximately 30 amino acids in length that is not present in any of the MBD domains. However, the TAM domain is very well conserved between different species of eukaryotes (Fig. 15B). This can be obtained from the sequence alignment of human TIP5 (BAZ2A) with the known sequences of BAZ2A homologs of higher eukaryotes and other species, shown to be phylogenetically related to human BAZ2A by Hendrich et. al. (Hendrich & Tweedie, 2003).

The region, conserved between these domains includes the 30 amino acid insert and extends at both N- and C-termini beyond the potential MBD domain borders as predicted by SMART (Letunic et al, 2009; Schultz et al, 1998) or Pfam (Finn et al, 2010). Additionally, in higher eukaryotes, the AT-hooks 1 and 2 are much conserved as well, and have almost identical amino acid sequence. However, the AT-hooks are missing in *D.melanogaster* Toutatis protein and phylogenetically related proteins of *C.elegans* and *A.gambiae*.

Hence, the actual TAM domain is possibly larger than predicted. So, based on the alignments, a protein construct TIP4N with amino acid sequence 496-664 was created. That comprised the predicted TAM domain, the conserved N- and C-terminal sequences and both AT-hooks. It includes also the key residue K653, shown to have an impact on TIP5/pRNA interaction upon its acetylation by MOF in context of NoRC (Zhou et al, 2009).

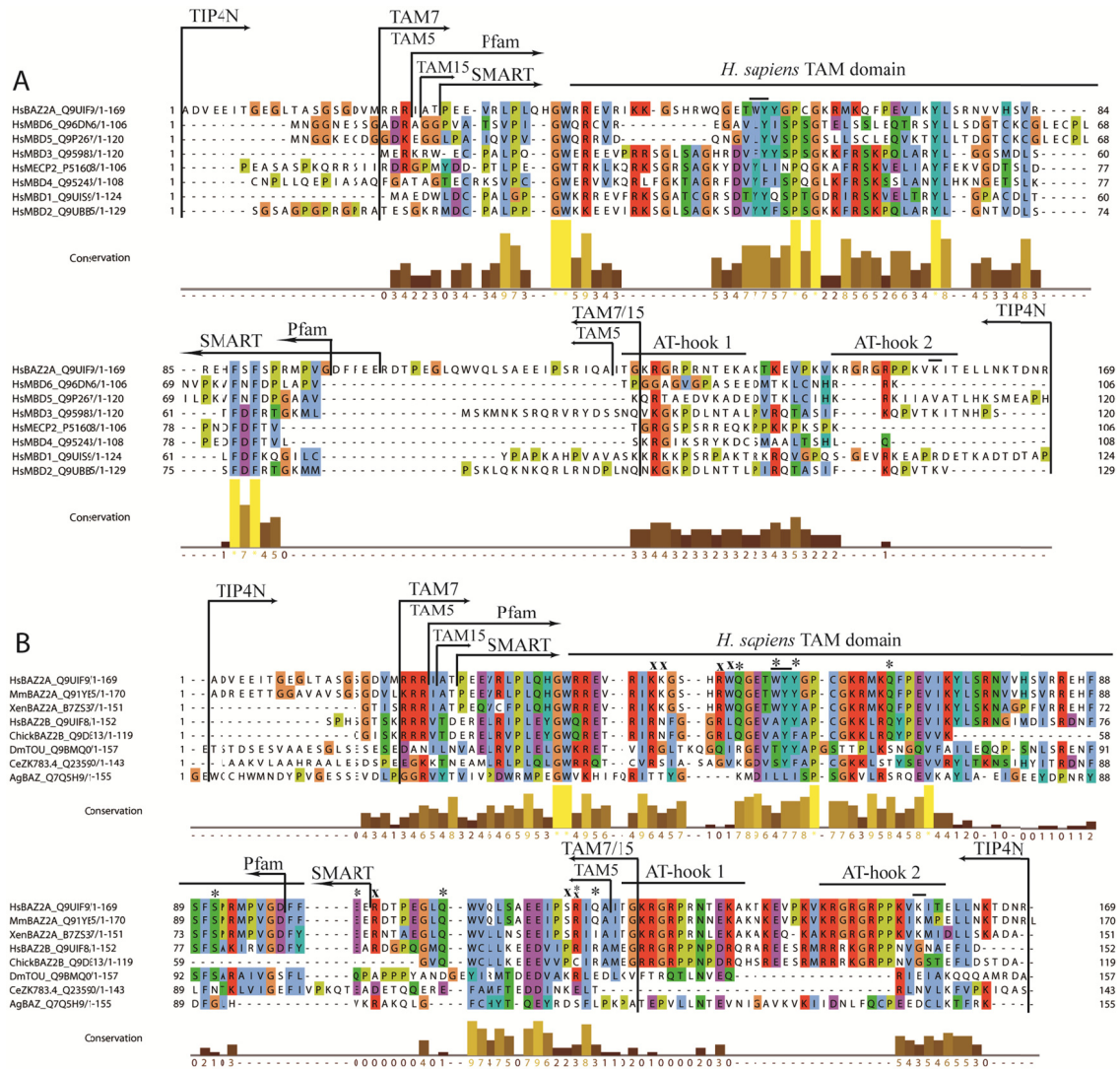


Fig. 15 Multiple sequence alignment of proteins containing MBD and MBD-like domains. Region around the domain is illustrated. UniProtKB accession number is included in the sequence name. Colors are shown according to Clustal X (Larkin et al, 2007) default colorscheme. Relative residue conservation score, reported by alignment program is presented as yellow bars. BAZ2A domain architecture, as predicted by SMART (Letunic et al, 2009; Schultz et al, 1998) and Pfam (Finn et al, 2010) are highlighted on top of the sequence. The start and end position of the sequence of TIP5 truncation constructs investigated in the course of PhD-work are marked with arrows as well and are assigned. (A) MUSCLE alignment of human TIP5 (BAZ2A) and human MBD domains. Residues, corresponding to W531Y532 in murine sequence are marked with straight line. (B) MUSCLE alignment of BAZ2A homologs, phylogenetically related to human BAZ2A (after Hendrich & Tweedie, 2003). Here, * marks residues mutated to alanine in the first series of mutations, based on residue conservation; x marks amino acid mutated based on results of NMR titrations. Residues, corresponding to W531Y532 in murine sequence and K653, which is subject to acetylation by MOF, are marked with straight line.

6.2. NMR studies of TIP4N construct.

Soluble TIP4N(496-664), was recombinantly expressed in minimal labeling media by *E.coli* BL21(DE3) from a pETM-11 expression vector, containing a GB1 expression tag (see Appendix). Different isotope labeling schemes were utilized. The boundaries of the construct are shown as well in Fig. 15.

Two dimensional ^1H - ^{15}N Heteronuclear Single Quantum Coherence (HSQC) experiments of TIP4N showed a good dispersion of peaks (Fig. 16A), revealing that TIP4N was folded *in vitro*.

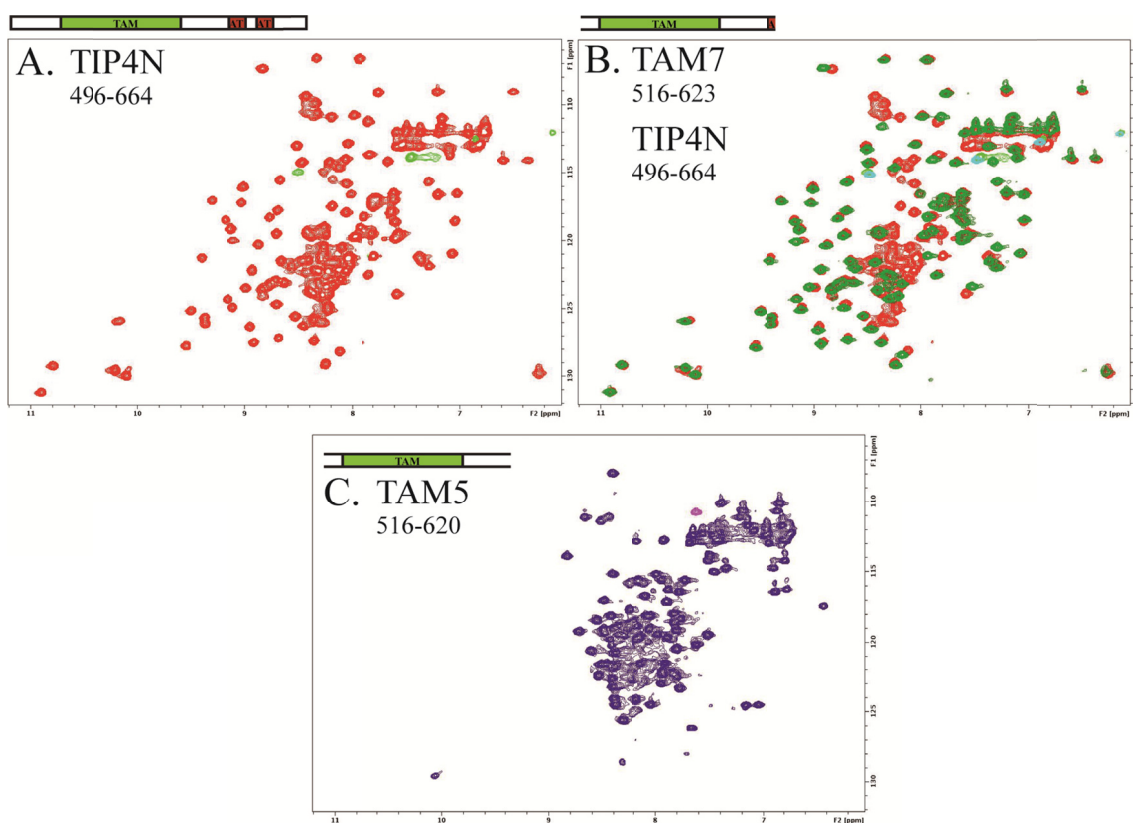


Fig. 16 ^1H - ^{15}N HSQC spectra of different TIP5 truncation constructs in NMR Buffer. On top of each spectrum domain architecture is depicted as predicted by SMART.(A) ^1H - ^{15}N HSQC of TIP4N(496-664) construct. Positive peaks are colored red, negative peaks are colored in pale green. (B) Overlay of ^1H - ^{15}N HSQC of TAM7 (516-623, green) onto ^1H - ^{15}N HSQC of TIP4N (red). Negative peaks are colored pale cyan and pale green respectively. Truncation of flexible residues at N- and C- termini based on relaxation analysis leads to removal of overlap in the middle of the spectrum not disturbing the fold. (C) ^1H - ^{15}N HSQC of TAM5 (516-620) in blue. Further truncation of 3 amino acids, as compared to TAM7, at the C-terminus leads to unfolding of the protein. Negative peaks are shown in magenta.

However, a rather large overlap in the spectral area around 8ppm, indicated that though TAM domain was contained within these borders, some regions of the protein lacked secondary structure. Additionally, TIP4N was stable in NMR-Buffer for several days at concentrations below 200 μ M. Concentration above that limit caused severe precipitation, which is problematic for NMR based structure determination.

Still HNCA, HNCACB and HN(CO)CACB experiments could be recorded on $^{15}\text{N}/^{13}\text{C}$ -samples with and without random deuteration (app.50%) and 80% of the backbones ^1H - ^{15}N correlations were assigned using Sparky software (Goddard & Kneller).

Stability issues could not be overcome by optimization of buffer-conditions (different salt concentrations and additives, Table 1), indicating a demand for optimization of the protein construct.

Additives/salt	Maximum stable TIP4N concentration achieved*
0.5M NaCl	0.3mM
0.75M NaCl	0.5mM
0.1M Na ₂ SO ₄	0.3mM
0.2M Na ₂ SO ₄	-----
50mM L-Arg / 50mM L-Glu ¹	0.2mM
DMSO _{d6} deuterated 0.05M NaCl	0.5mM
DMSO d6 deuterated 0.1M NaCl	0.5mM

Table 1. Optimization of buffer conditions for NMR. * All samples were concentrated to the same final concentration and left at room temperature over night. TIP4N showed immediate precipitation if concentration of Na₂SO₄ was raised to 0.2M (indicated by dashed line). ¹ Addition of 50mM L-Arg and 50mM L-Glu to NMR Buffer was reported to increase solubility and stability of proteins (Golovanov et al, 2004).

6.3. NMR relaxation studies of TIP4N.

To detect unstructured regions of TIP4N and investigate the internal dynamics of the protein, NMR relaxation analysis of the construct was performed (Fig. 17). For analysis, information from the backbone assignment was utilized.

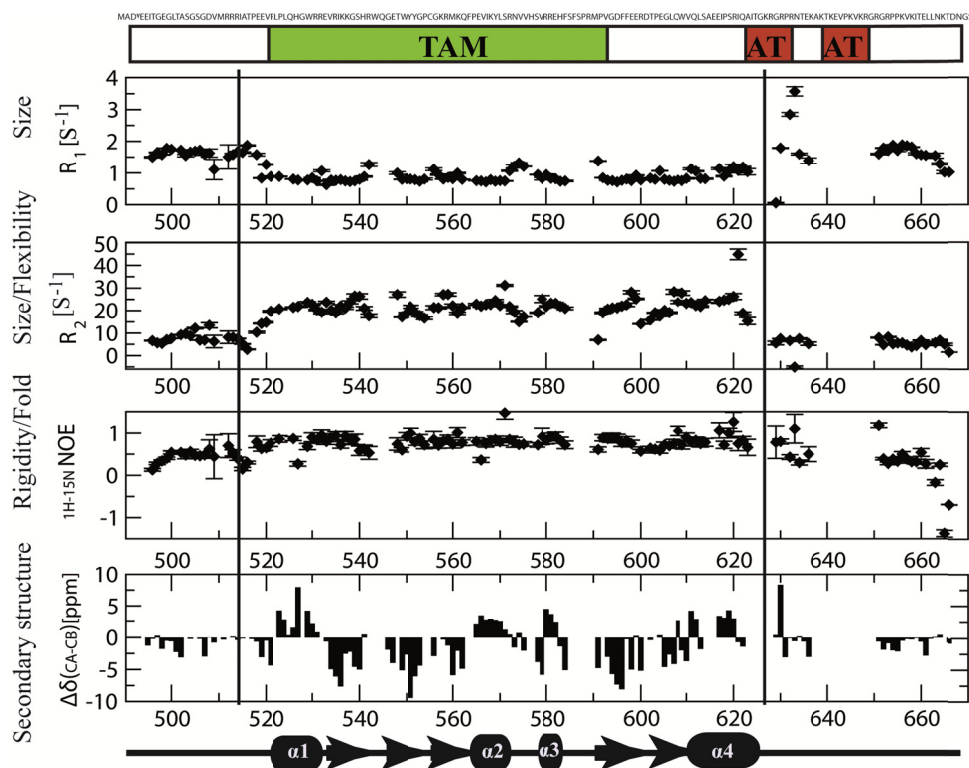


Fig. 17. NMR dynamics studies of TIP4N as well as secondary structure prediction based on secondary chemical shifts of C α and C β atoms (Pastore & Saudek, 1990b; Spera & Bax, 1991b). The errors for the relaxation data fit and ^1H - ^{15}N heteronuclear NOE are indicated by error bars and were calculated as described in 5.5. Cartoons at the top and at the bottom represent domain architecture of TIP4N predicted by SMART (Letunic et al, 2009; Schultz et al, 1998), with the predicted TAM domain region in green and AT-hook regions in red, and the secondary structure elements derived from secondary chemical shifts respectively. α -helices are presented as ellipses and β -strands are represented by arrows.

From R_2/R_1 correlation time of TIP4N τ_m could be estimated as 11-12ns, using the quadratic representation approach (Bruschweiler et al, 1995). That corresponded to the theoretical correlation time τ_0 , calculated from the number of amino acids, assuming TIP4N to be a monomer and using the formula

$$\tau_0 = \left(9.18 \times \frac{10^{-3}}{T}\right) \exp\left(\frac{2416}{T}\right) N^{0.93},$$

where τ_0 is the correlation time in nanoseconds, T is the temperature in Kelvin and N is the number of amino acids (Daragan & Mayo, 1997). That is, TIP4N was present as a full length construct in solution and was tumbling as a monomer.

^1H - ^{15}N heteronuclear NOE values higher than 0.77 for the major part of TIP4N indicated that the region between amino acids 520 and 620 was rigid (Kay et al, 1989), meaning folded. Higher than average R_1 and lower than average R_2 values for the amides of residues 496-519 and 621-664 indicated that those were flexible and tumbling independently from the core domain. Combined with ^1H - ^{15}N -heteronuclear NOE values lower than average and close to 0 for the same regions that meant that they were unstructured.

That observation was also supported by the secondary structure prediction from secondary chemical shifts of $\text{C}\alpha$ and $\text{C}\beta$ atoms (Pastore & Saudek, 1990b; Spera & Bax, 1991b), based on TIP4N NMR assignments. Those suggested well defined or lack of secondary structure for the same regions of the sequence (Fig. 17).

Overall, it could be concluded, that the rigid folded TAM domain was indeed most probably contained within the residues 520-620, flanked by flexible, unstructured sequences.

6.4. Diffusion coefficient measurements by NMR.

The concept derived in 1.1.3. was also supported by NMR diffusion coefficient measurements of 0.2mM TIP4N in NMR buffer at 298K. A 2D stimulated echo experiment, using bipolar gradients and WATERGATE (stebpgp1s19) from Bruker, was utilized (Fig. 18A). The intensity change for well resolved peaks at 1.156ppm, 0.880ppm and 0.588ppm was fitted at different gradient strength using an internal analysis unit of the Bruker TOPSPIN software (Fig. 18B, C). The results of the fit are shown in Fig. 18D. Though the determined diffusion coefficient differs a bit for the peak at 0.588ppm from the one, calculated for two other peaks, the average diffusion coefficient D_c was calculated to be approximately $1.4 \times 10^{-10} \text{m}^2 \text{s}^{-1}$. This is comparable to the $D_c = 1.18 \times 10^{-10} \text{m}^2 \text{s}^{-1}$ reported for lysozyme (14.4kDa) at 20°C in dilute solution (He & Niemeyer, 2003). Considering that D_c rises with lower molecular weight, TIP4N (20kDa) diffuses in solution as a protein smaller than 14kDa.

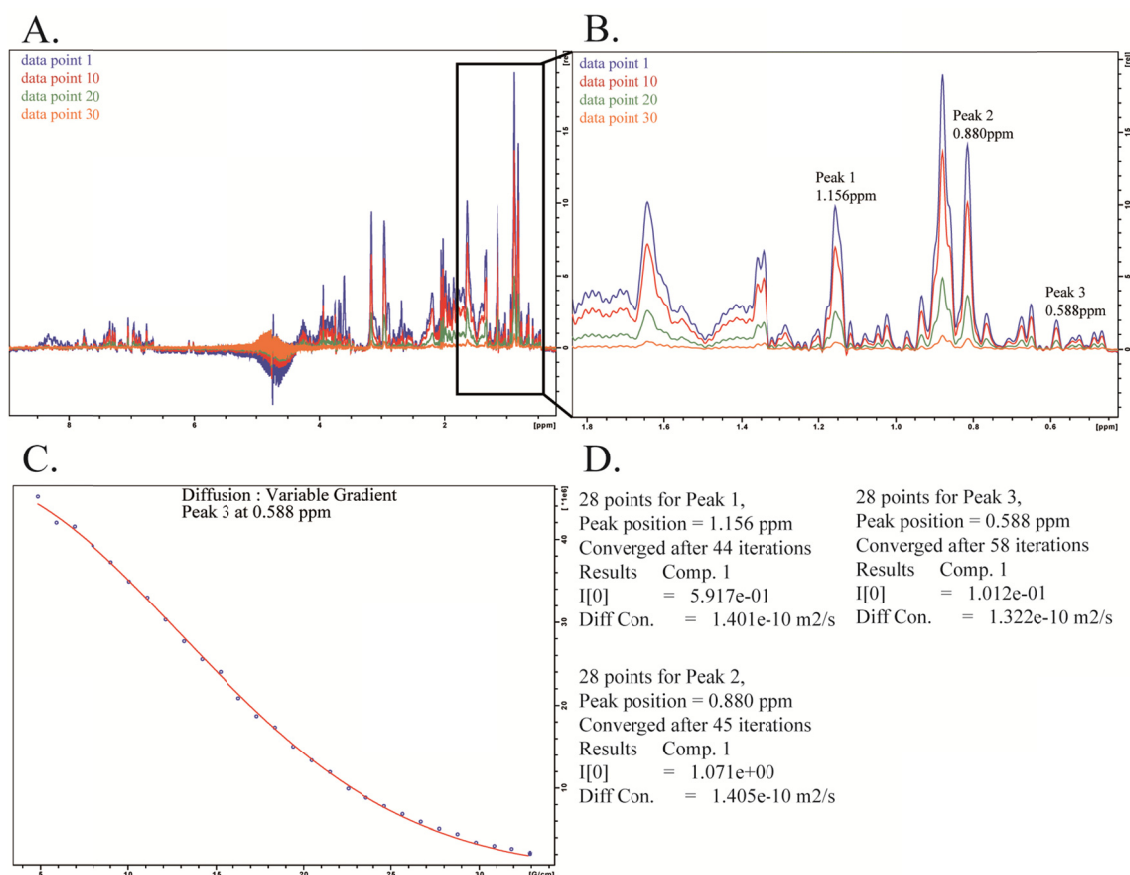


Fig. 18 Diffusion coefficient measurements of the TIP4N construct. (A) Overlay of the 1D-planes of the stebpp1s19 pseudo 2D stimulated echo experiment for different measurement points, corresponding to TIP4N ¹H spectra at different gradient strengths. The point number is indicated in the upper left corner. The ascending number corresponds to the higher gradient strength. The region of the spectra used for data evaluation is boxed black. (B) A close up of the region from 0.5ppm to 1.8ppm of the spectra overlay from (A), used for data evaluation. Well resolved peaks, used for intensity fit are marked, their chemical shift is indicated. The color coding corresponds to the one in (A). The point number of experiment is indicated in the upper left corner of the picture. (C) An example of the data fit. Data fit for Peak 3. Intensity of the peak 3 is plotted against the respective gradient strength. The data points are represented by blue dots; the fitted curve is shown red. (D) An extract of the result output from the analysis unit of the Bruker TOPSPIN software. The fitted initial peak intensity is indicated I[0]. The value of the diffusion coefficient is given as Diff Con. The number of points used for the fit of each selected peak is 28.

6.5. TAM domain constructs.

Taking previous knowledge into consideration, several new TIP5 truncation constructs were designed, aiming to contain only the rigid folded TAM domain without flexible N- and C-termini (for boundaries of the constructs see also Fig. 15). The structural integrity of the proteins was monitored in 2D ¹H-¹⁵N-HSQC. The minimal TAM-domain construct, that was expressed well and could be purified was TIP5(516-623), termed TAM7. Having the

molecular weight of 13.4kDa, it was folded the same way as TIP4N (Fig. 16B), however it lacked overlap in the middle of the spectrum. Further truncation at the N-terminus to TAM15(520-623) lead to a decrease in TEV-cleavage of the expression tag, whereas truncation of only three amino acids at the C-terminus to TAM5(516-620), lead to unfolding of the protein (Fig. 16C).

6.6. EMSA of TIP5 truncation constructs.

To make sure that the biological function of the TIP4N and TAM7 constructs was retained, the affinity of the created TIP5 truncations to RNA was tested in electrophoretic mobility shift assays (EMSAs). These EMSAs, kindly provided by our collaborator Dr. S. Melnik from Prof. Dr. I. Grummt lab, Division of Molecular Biology of the Cell II, dkfz, showed that compared to the functional TIP5 construct, the RNA binding function in TIP4N and TAM7 was not affected and was pRNA-specific (Fig. 19).

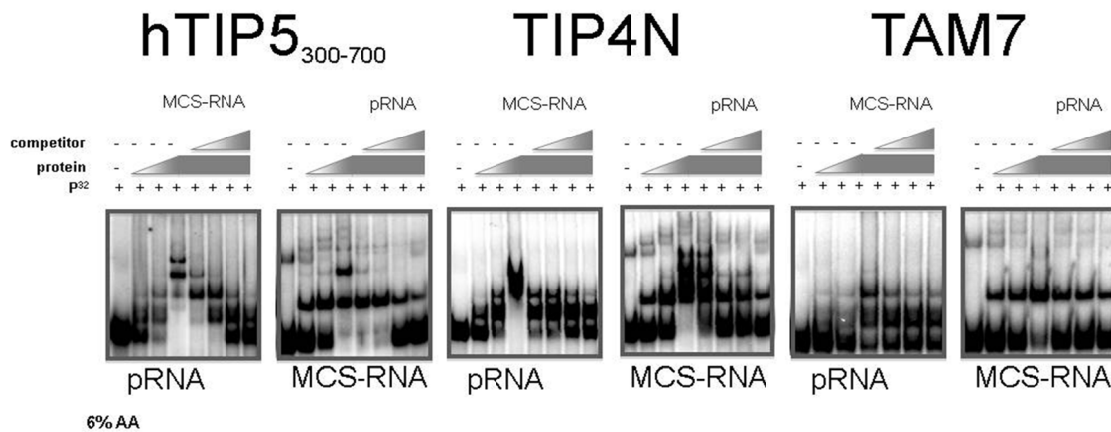


Fig. 19 Electrophoretic mobility shift assay (EMSA) of TIP5 truncation constructs. Kindly provided by Dr. S. Melnik, dkfz. Human TIP5, TIP4N and TAM7 proteins were investigated concerning their affinity to pRNA. ³²P labeled pRNA or MCS RNA from pBluescript-KS was incubated with proteins. At a certain point the binding was competed by non labeled RNA. Two panels correspond to every protein truncation construct. Labeled probe is indicated at the bottom of each panel, unlabelled probe is assigned at the top. Though in all cases proteins interact with the labeled probe, binding to pRNA is not competed out by cold MCS RNA, while MCS binding is competed out by cold pRNA.

Thus, the functional structured TAM domain of human TIP5 is contained within the sequence 520-623. The construct, TAM7(516-623) containing the TAM domain, could be expressed in

E.coli BL21 (DE3) and purified for NMR structure determination. The protein could be concentrated to 0.7mM and was stable in NMR-buffer over several weeks at 293K.

7. Results II. Solution structure of the TAM domain from *H. sapiens*.

7.1. NMR ensemble of structures.

Recombinant TAM domain (residues 516-624) was expressed at high yields and purified from *E.coli*. Samples uniformly labeled with $^{15}\text{N}/^{13}\text{C}$ and selective ^{13}C labeling was used. The solution structure was determined using standard methods of multidimensional NMR spectroscopy (Sattler et al, 1999) for backbone assignments. Backbone-, side chain- and intraresidual NOE-assignments were done manually. For interresidual NOE-assignments CYANA 3.0 (Guntert, 2004) was utilized. Torsion angle restraints were derived from TALOS+ (Shen et al, 2009). The structure contains 16 aromatic residues, which were not trivial to assign. Two-dimensional NMR experiments for correlating $^{13}\text{C}\beta$ and $^1\text{H}\delta/\epsilon$ chemical shifts of aromatic residues in ^{13}C -labeled proteins via scalar couplings were additionally utilized for unambiguous assignment of the aromatic rings (Yamazaki et al, 1993).

The structure was defined based on 2732 distance restraints, derived from CYANA 3.0 calculations, torsion angle restraints and two sets of RDCs, measured in two different alignment media, and calculated using ARIA1.2 (Linge et al, 2003a) software. A water refinement step was performed. In the final step a total of 200 structures were calculated from which the twenty lowest energy structures were used for quality and structure validation using the iCING web interface (<http://nmr.cmbi.ru.nl/icing/>) and PROCHECK (Laskowski et al, 1993).

Introduction of the correct protonation state of the histidines into structure calculation determined by long range HSQC (Pelton et al, 1993) and the stereospecific assignment of methyl ^1H of valine and leucine by nonrandom ^{13}C labeling (Senn et al, 1989) lead to considerable improvement in the structure accuracy.

The ensemble of 20 lowest energy structures obtained after water refinement is shown in Fig. 20A. Out of 113 amino acids of the expression construct, 100 define the tertiary fold with high precision (backbone RMSD is $0.38 \pm 0.08 \text{ \AA}$ and heavy atom RMSD is $1.02 \pm 0.13 \text{ \AA}$) and good structural statistics (Table 2).

NOE-based distance restraints^a	
Intraresidual	738
Sequential ($ i-j =1$)	607
Medium range ($1< i-j <5$)	478
Long range ($ i-j \geq 5$)	909
Total NOEs	2732
Other restraints	
$\phi+\psi$ dihedral angle restraints ^b	169
Residual dipolar coupling restraints ^c ($H^N-N, N-C'$)	
Set _(PAA)	74
Set _(Bicelles)	104
Restraint violations	
NOE distances with violations $>0.5 \text{ \AA}$	0.00+-0.00
Dihedrals with violations $>5^\circ$	0.01+-0.01
Consistency (structure vs. restraints)	
r.m.s.d. for experimental distance restraints ^a (2732) (\AA)	0.036 +-0.01
r.m.s.d. for experimental torsion angle restraints ^b (169) ($^\circ$)	1.313 +- 0.056
RDC $Q_{(PAA)}$ -factor ^c	0.159+-0.004
RDC $Q_{(bicelles)}$ -factor ^c	0.196+-0.005
Validation (structure vs. external restraints)	
Residual dipolar coupling restraints (H^N-N)	
Set ^d _(Otting)	51
RDC $Q_{(Otting)}$ -factor ^d	0.28
r.m.s.d. from idealized covalent geometry	
Bonds (\AA)	0.01+-0.00
Angles ($^\circ$)	0.52+-0.01
Coordinate precision r.m.s.d. (for 522-621)	
Backbone (\AA)	0.38 +/-0.08
Heavy atom (\AA)	1.02 +/-0.13
Ramachandran analysis^e	
Residues in most favored regions	92.7%
Residues in additionally allowed regions	7.3%

Table 2 Structural statistics for the TAM domain NMR structure. Statistics are given for the 20 lowest energy structures after water refinement out of 200 calculated. The CNS E_{repel} function was used to simulate van der Waals interactions with an energy constant of $25\text{kcal mol}^{-1}\text{\AA}^{-4}$ using PROLSQ van der Waals radii.

^a Distance restraints were employed with a soft square well potential using an energy constant of $50\text{kcal mol}^{-1}\text{\AA}^{-2}$. ^b Torsion angle restraints derived from TALOS₊ (Shen et al, 2009) were applied to ϕ , ψ backbone angles using energy constants of $200 \text{ kcal mol}^{-1}\text{rad}^{-2}$. ^c Residual dipolar couplings (RDCs) were employed with a harmonic potential using an energy constant of $0.8 \text{ kcal mol}^{-1}\text{Hz}^{-2}$ for the $^1\text{H}^N$ - ^{15}N RDCs and $0.4 \text{ kcal mol}^{-1}\text{Hz}^{-2}$ for the $^{13}\text{C}'$ - ^{15}N RDCs. The alignment medium is indicated as index Q-factor as defined by (Cornilescu et al, 1998). ^d Residual dipolar couplings (RDCs), measured in the Otting medium were not employed in structure calculation. The Q-factor is determined as defined by (Lipsitz & Tjandra, 2004). ^e Performed with PROCHECK (Laskowski et al, 1993).

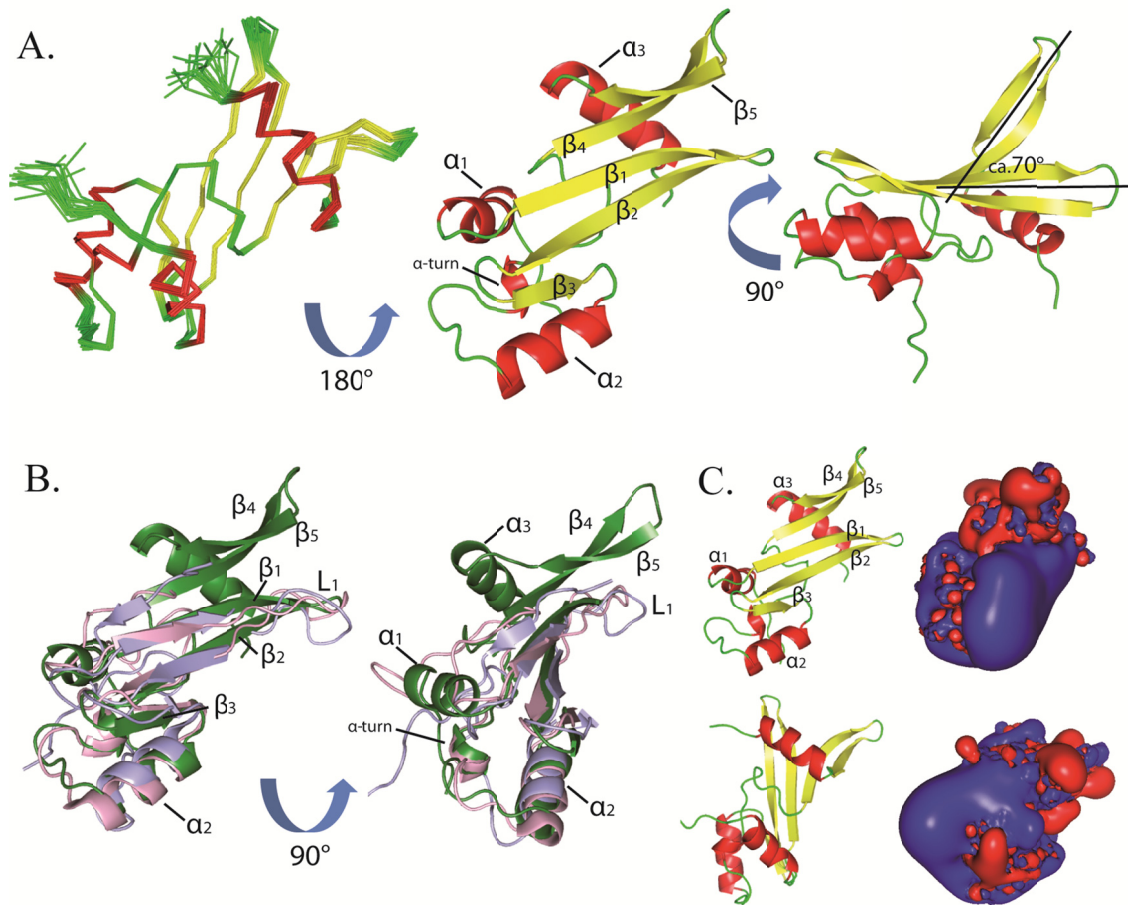


Fig. 20 Solution structure of the TAM domain from *H. sapiens* TIP5 protein. (A) From left to right. NMR ensemble of the 20 lowest energy structures, as seen from the back of the domain. Only the backbone is shown in a ribbon representation. Helices are colored red, β -sheets yellow and loops are shown in green. Middle: NMR lowest energy structure, in cartoon representation, seen from the front. α_1 -helix shows distortion in winding introduced by Pro528. Right: lowest energy structure turned 90° as compared to the previous view, a ribbon representation showing the approximate angle between β -sheets. (B) Overlay of the TAM lowest energy structure (forest), and MBD domains from human MBD1 (Ohki et al, 1999) in light blue and mammalian MeCP2 (Wakefield et al, 1999) in light pink. MeCP2 is shown without the unstructured N-terminus. Figure shows the domains from the front and after a 90° turn, presenting α_1 , α_3 and β_4 , β_5 secondary elements of the TAM domain, that are unique to it. (C) Domain orientation and electrostatic isosurfaces of 1kT/e contour for positive charge (blue) and -1kT/e contour for negative charge (red). The isosurfaces were calculated online using APBS web interface (Baker et al, 2001) with standard settings and visualized in PyMOL (Schrödinger)

TAM domain folds into an α/β sandwich structure (Fig. 20A), comprised of two β -sheets, formed by five antiparallel β -strands, three α -helices and an α -helical turn. The secondary structure is defined by ^{13}C chemical shifts and is supported by H-D exchange experiments (Fig. 26)

The three antiparallel twisted β -strands β_1 (Arg534-Cys541), β_2 (Arg545-Tyr553) and β_3 (Gly557-Met560) form the first β -sheet, the second one is comprised by the antiparallel twisted strands β_4 (Asp593-Asp596) and β_5 (Gly603-Gln608). Though forming a continuous surface, the β -sheets are distinct from each other and open up like scissors at the C-terminus proximal part, forming an angle of ca.70 degrees. This cleft is well defined by NOEs and the RDCs. The characteristic NOE pattern the NOESY spectrum for the strand β_3 is observed, however it can be visualized only in the lowest energy structure.

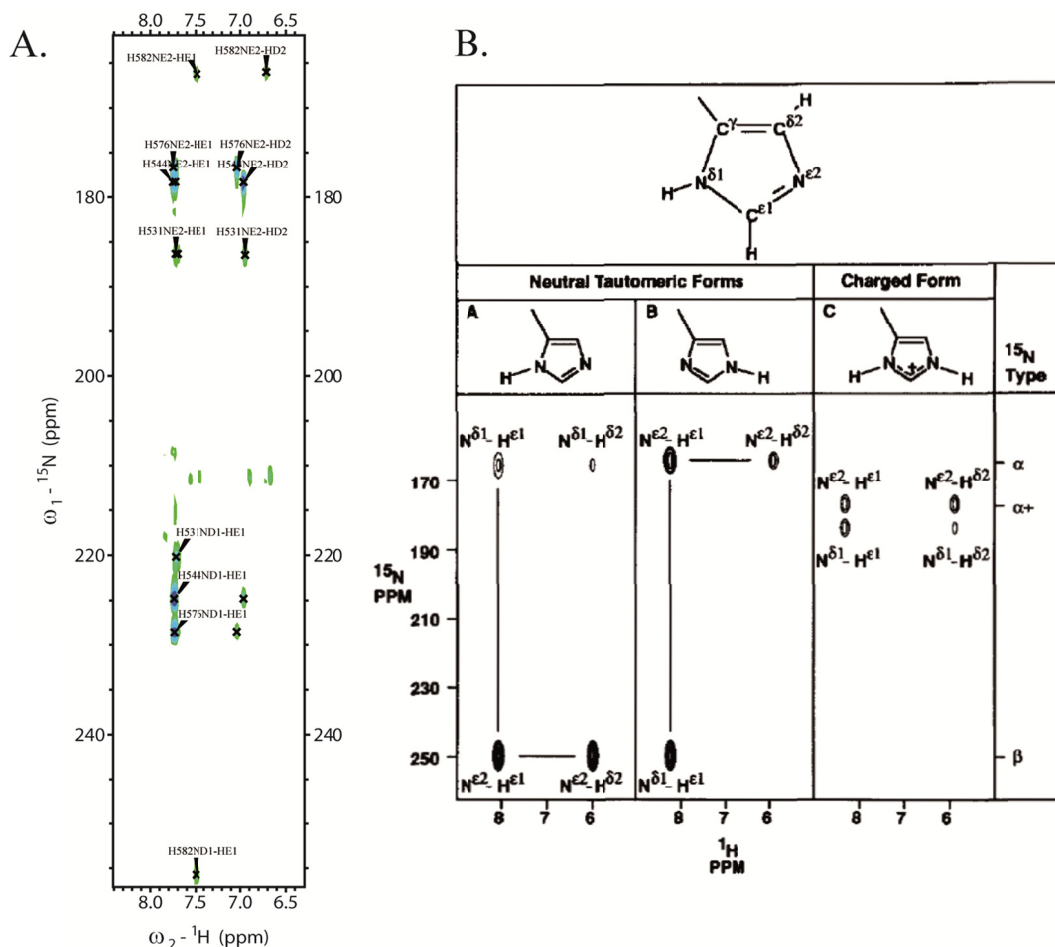


Fig. 21 Tautomeric state of TAM histidines. (A) ^1H - ^{15}N HMQC of TAM domain in NMR buffer, recorded at 300K on Bruker 600MHz spectrometer with cryoprobe. (B) From (Pelton et al, 1993). Schematic diagram showing the three possible protonation states of the histidine ring and the expected ^1H - ^{15}N HMQC spectrum of each species. Based on that diagram, it was concluded, that all histidines of the TAM domain are in the $\text{N}^{\epsilon 2}$ -protonated state.

β -strands are packed against α_2 (Phe563-Arg572), the α -helical turn (Arg580-His582), as well as two other α -helices at the back of the domain: the α_1 (Pro522-Leu529) and α_3 (Ala611-

Ile621), formed by the N- and the C- terminus of the protein. Helix α_1 has a minor distortion in the winding, introduced by Pro528. All histidines of the TAM domain are present in the $N^{\epsilon 2}$ -protonation state (Fig. 21).

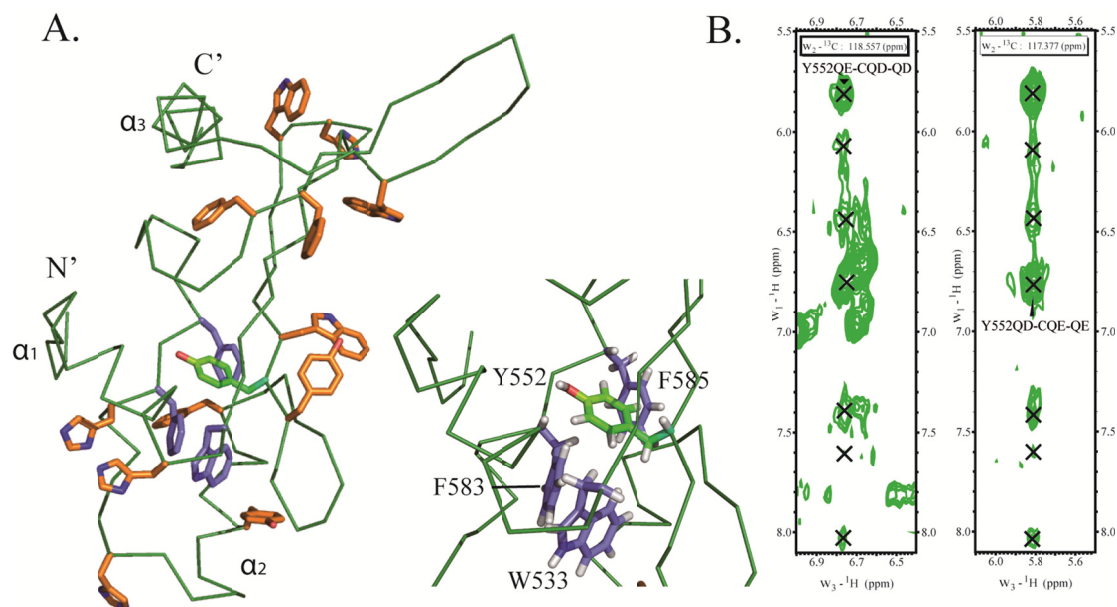


Fig. 22 Aromatic residues of the TAM domain. (A) NMR structure of the TAM domain is shown as ribbon in forest green. On the left, the side side chains of all aromatic amino acids of the TAM domain (His, Phe, Tyr and Trp) are shown as sticks in orange. Here, the hydrogen atoms are omitted for the simplicity. The residues forming the aromatic cage are highlighted. Trp533, Phe583 and Phe585 are shown in blue; Tyr552 is shown in green. On the right, a close up view showing the side chains of the aromatic cage is presented. The residues forming it are assigned. (B) ^{13}C -NOESY-HSQC strips of the Tyr552 $^{13}\text{C}\delta\text{Q}\delta$ resonance (left panel) and the $^{13}\text{C}\epsilon\text{Q}\epsilon$ resonance (right panel) are presented. Here Q stands for the degenerated ^1H signal. The carbon $^{13}\text{C}\delta$ resonances are degenerated as well. The cross peaks to the closest neighbor in the ring of Tyr552 are assigned.

The N-terminal part of the TAM domain contains 10 out of the 16 aromatic amino acids of the sequence (Fig. 22A). Most of them are pointing to the inner core of the protein. The structural elements of the N-terminus are thus packed closely together, by the hydrophobic interactions of the aromatic rings. The residues Trp533, Tyr552, Phe583 and Phe585 form an aromatic cage (Fig. 22A). Consistent with it, due to the aromatic ring current effects, the degenerated $^{13}\text{C}\delta$ and $^1\text{H}\epsilon$ atoms of the Tyr552 show abnormal resonance frequencies, shifted much upfield as compared to the average values from the Biological Magnetic Resonance Bank (BMRB). The $^{13}\text{C}\delta$ resonates at 118.55ppm, compared to the average diamagnetic protein value of 132.73ppm, and $^1\text{H}\epsilon$ appears at 5.81ppm, compared to the average diamagnetic protein value of 6.71ppm (Fig. 22B). However, these values still lie within the

reported range (BMRB). As reported in BMRB (BMRB), the lowest observed resonance frequency for $^{13}\text{C}\delta$ of a tyrosine in diamagnetic conditions lies at 113.00ppm and for the $^1\text{H}\epsilon$ of a tyrosine in diamagnetic conditions at 4.56ppm. In the case of the TAM domain the assignments of Tyr552 were confirmed by the NOE-patterns in the ^{13}C -NOESY-HSQC.

Compared to the known structures of an MBD domain (Ohki et al, 1999; Wakefield et al, 1999), TAM domain shows a novel fold (Fig. 20B). The antiparallel β -sheet composed of β_4 and β_5 is present only in the TAM domain structure, as well as are the terminal helices α_1 and α_3 .

7.2. Electrostatic potential.

TAM domain is very positively charged at neutral pH with theoretical pI of 10, as predicted by ProtParam Tool (Gasteiger et al, 2005). Electrostatic potential map is shown in Fig. 20C. It was calculated using APBS software (Baker et al, 2001). The most striking feature is a large positively charged surface formed by β -strands β_2 , β_3 and the α_2 -helix. The tip of the β -sheet 2 as well, as the N-terminal part of α_3 form a less charged surface with negatively charged patches.

7.3. Residual dipolar couplings.

For positioning of the secondary structure elements, TAM domain was partially aligned in three different media and sets of residual dipolar couplings were recorded. Several alignment media were tested (Table 3). Most of them caused protein precipitation or unfolding and could not be utilized. However, it was possible to achieve a needed degree of partial alignment and record $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ and $^{13}\text{C}'\text{-}^{15}\text{N}$ residual dipolar couplings in 7% PAA gel (Ishii et al, 2001) and 4% w/v DMPC/DHPC bicelles (Ottiger & Bax, 1999), and $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ residual dipolar couplings in Otting medium (5% C12E6/hexanol, $n(\text{C12E6})/n(\text{hexanol})=0.64$ (Rückert & Otting, 2000)).

Alignment medium	Temperature (K)	D2O splitting (Hz)	Measured couplings
Pf1 (10mg/ml)	293K	ca. 10	-----
strained Ewald gelatine gel 5%	293K	8.5	-----
Otting medium* (5% C12E6/hexanol, n(C12E6)/n(hexanol)=0.64)	293K	28.6	$^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$
3:1 DMPC:DHPC bicelles 4%w/v	303K	5	$^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$, $^{13}\text{C}'\text{-}^{15}\text{N}$
7% Polyacrylamide gel (37.5% acrylamide, 1% Bis-acrylamide)	293K	3.8	$^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$, $^{13}\text{C}'\text{-}^{15}\text{N}$

Table 3 Alignment media tested for TAM domain for recording residual dipolar couplings. Unfortunately Pf1 phage use lead to precipitation of the protein and Ewald-gelatine lead most probably to protein unfolding, as could be obtained from $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ -HSQC spectra recorded for control. Thus, no RDC could be measured in that case (indicated by dashed line). Other media gave rise to data of good quality and were used in TAM structure calculation and validation.

* A nonionic liquid crystalline molecular alignment medium composed of different mixtures of n-Alkyl-poly(ethylene glycol)/n-alkyl alcohol or glucopone/n-hexanol (Rückert & Otting, 2000). Here, n stands for the molar amount of the substance.

Two sets of RDCs were used in structure calculation (alignment media 7% PAA gel and 4% w/v bicelles). RDC values in both media show a good distribution, as can be obtained from the histogram of normalized RDCs, shown in Fig. 23. Excluding the outliers, based on that diagram the axial component (D_a) and the rhombicity (R) of the molecular alignment tensors in the principal coordinate frame were estimated as $D_{a(\text{PAA})}=-13\text{Hz}$, $R_{(\text{PAA})}=0.6$, and $D_{a(\text{bicelles})}=11\text{Hz}$ $R_{(\text{bicelles})}=0.36$ (Clore et al, 1998a) respectively. That correlated well with the rhombic and axial components for those sets estimated by MODULE 1.0 software (Dosset et al, 2001) as $D_{a(\text{PAA})}=-13\text{Hz}$, $R_{(\text{PAA})}=0.5$, and $D_{a(\text{bicelles})}=7\text{Hz}$ $R_{(\text{bicelles})}=0.4$. Starting with those values CYANA 3.0 calculated $D_{a(\text{PAA})}=-11.9\text{Hz}$, $R_{(\text{PAA})}=0.522$, and $D_{a(\text{bicelles})}=9\text{Hz}$ $R_{(\text{bicelles})}=0.370$ using a gridsearch approach, the result being in full agreement with what was obtained from the analysis of the RDCs distribution. Residues involved in flexible loops were excluded prior to the calculation.

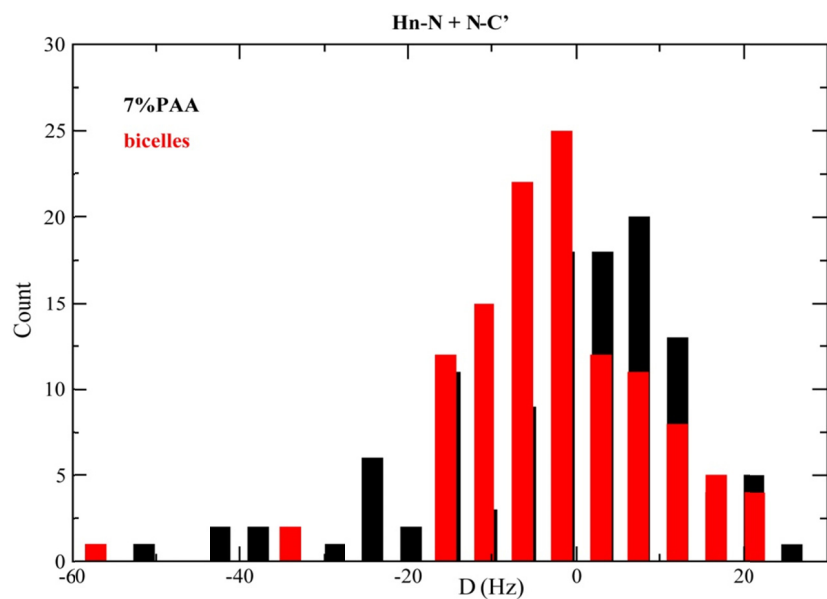


Fig. 23 A histogram of normalized RDCs for the TAM domain. D is the magnitude of residual dipolar coupling for $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ (here Hn-N) or $^{13}\text{C}'\text{-}^{15}\text{N}$ (here N-C') for 7% PAA (black) and 4%w/v 3:1 DMPC:DHPC bicelles (red) alignment media.

$^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ RDCs from the Otting medium were used for structure validation after water refinement. The fit of the measured RDCs and the ones, back-calculated from the NMR-ensemble is shown in Fig. 24. Experimental RDCs agree very well with the calculated ones with a $Q\text{-factor}_{(\text{PAA})}$ of 0.11 and $Q\text{-factor}_{(\text{Bicelles})}$ of 0.13, calculated after (Lipsitz & Tjandra, 2004), excluding the flexible residues. The $Q\text{-factor}_{(\text{Otting})}$, a set that was not used in structure calculation, is 0.28 calculated the same way, indicating a three dimensional structure of good quality. Flexible residues were again excluded from the analysis.

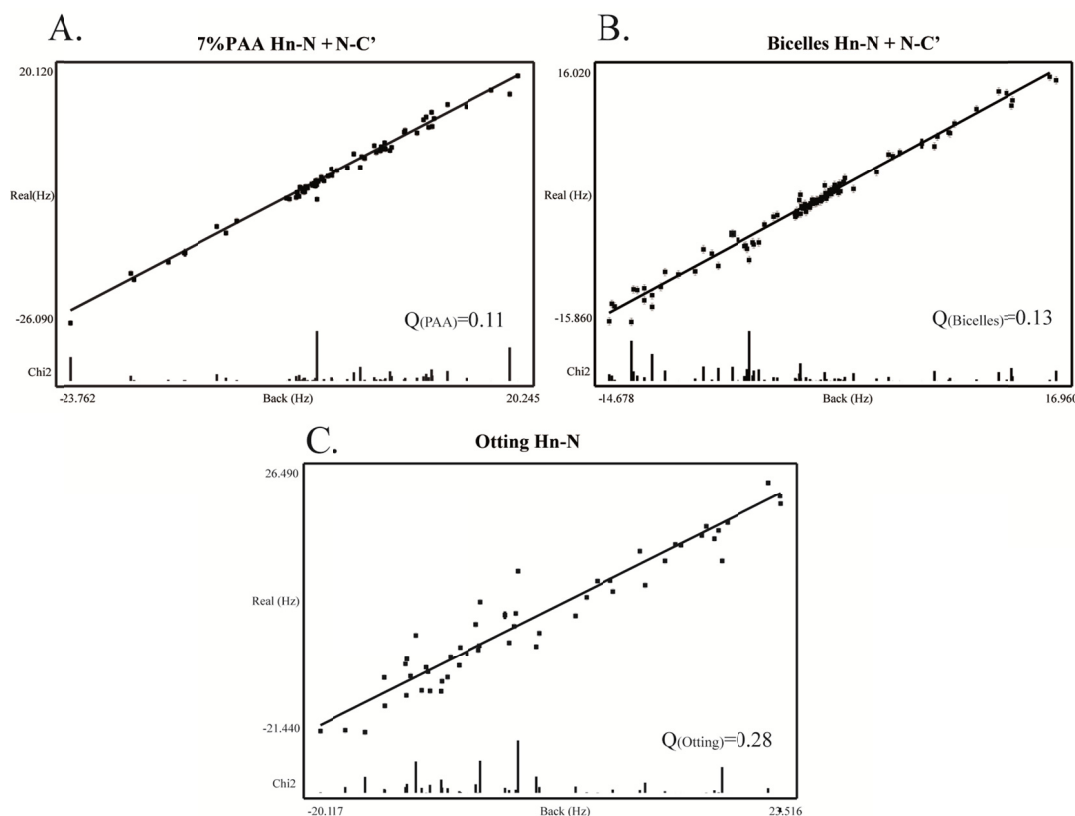


Fig. 24 Plot of experimental RDC values, measured in different alignment media versus the values calculated for the TAM domain NMR structure. The calculations were performed using MODULE 1.0 software (Dosset et al, 2001) on the NMR ensemble of 20 lowest energy structures after water refinement. Only $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ (here Hn-N) and $^{13}\text{C}'\text{-}^{15}\text{N}$ (here N-C') RDCs from non-flexible regions of the domain with well resolved resonances were used. The Q-factors defined in (Cornilescu et al, 1998) are given in the down right corner. (A) Correlation plot for the 7% PAA dataset, used in structure calculation. (B) Correlation plot for 4% w/v 3:1 DMPC:DHPC bicelles dataset, used in structure calculation. (C) Correlation plot for Otting medium (see Table 3) dataset, not used in structure calculation. In that case residual dipolar couplings were obtained only for $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$.

7.4. Structure validation using paramagnetic relaxation enhancement (PRE).

To further validate the TAM structure, paramagnetic relaxation enhancement data were recorded on a protein sample with 3-(2-Iodoacetamido)-PROXYL spin label (Fig. 25A). TAM domain has only one cysteine (Cys556). That resides in a not very flexible loop between the β -strands β_2 and β_3 . Thus, 3-(2-Iodoacetamido)-PROXYL was chemically attached to the side chain of Cys556.

The quotient $I_{\text{oxi}}/I_{\text{red}}$ per residue is shown in Fig. 25B. Here, I_{oxi} and I_{red} are the intensities of the same amide resonance in two identical 2D HSQC experiments, recorded on the protein sample with a spin label in an oxidized and in a reduced state. $I_{\text{oxi}}/I_{\text{red}}$ is directly proportional to the distance from the spin label to a particular residue (Battiste & Wagner, 2000).

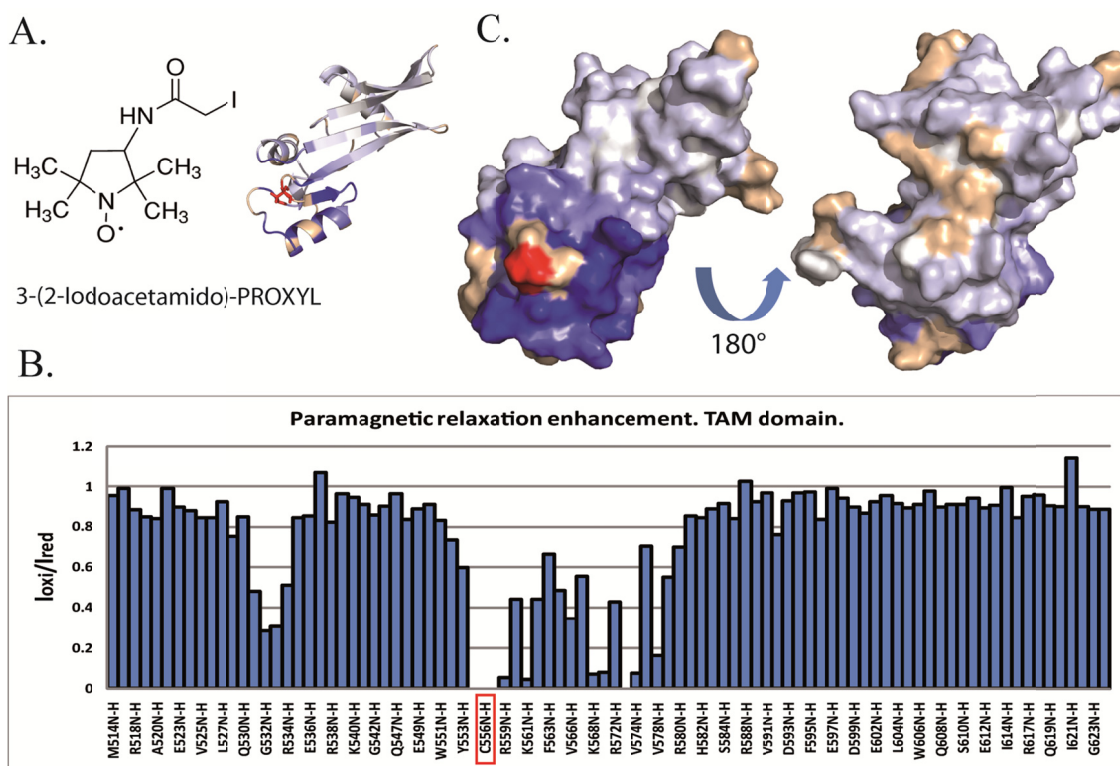


Fig. 25 Paramagnetic relaxation enhancement analysis of TAM domain. (A) Formula of the spin label, 3-(2-Iodoacetamido)-PROXYL, attached to the C556 side chain to induce PRE. Position of C556 is shown on the structure of the TAM domain. The side chain is represented as sticks and colored red. TAM domain itself is colored according to the $I_{\text{oxi}}/I_{\text{red}}$ plot in blue to white gradient, with blue representing $I_{\text{oxi}}/I_{\text{red}}=0$. (B) $I_{\text{oxi}}/I_{\text{red}}$ plotted against the sequence of the TAM-domain. C556 position in the sequence is boxed in red. (C) $I_{\text{oxi}}/I_{\text{red}}$ plotted on the surface of the TAM domain. Blue-white color gradient, with blue representing $I_{\text{oxi}}/I_{\text{red}}=0$. The surface is colored using histogram mode. Position of C556 is marked in red. Residues with insufficient data, which were not assigned due to signal overlap or missing signals, are colored in wheat. This color was always assigned, when the $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ resonance was not present in the $^1\text{H}\text{-}^{15}\text{N}$ HSQC of the protein, not assigned or it was overlapped. Histogram was produced using PyMOL (Schrödinger).

Fig. 25C represents $I_{\text{oxi}}/I_{\text{red}}$ values plotted on the surface of the water refined structure of the TAM domain. The color gradient, representing the distance from a particular amide to the spin label, correlates well with the calculated distance to the respective residue. Amides closer to the spin label experience a stronger PRE effect (dark blue), so that amides of the residues G554, C556, K558, N573 in the direct vicinity of the label are completely bleached in presence of the label. On the contrary, the amides on the opposite side of the domain experience less PRE (white color).

Minor differences in the peak intensity of the distant residues can be explained by small variations of protein concentration in the samples used. Attachment of the spin label

destabilized the protein, causing considerable precipitation at higher temperatures. In the case of TAM domain, experiments for the two states of the spin label had to be recorded on two distinct samples, handled the same way. The protein was always kept at 4°C.

7.5. H-D exchange.

TAM domain has a very compact structure, as can be judged from the H/D exchange experiments.

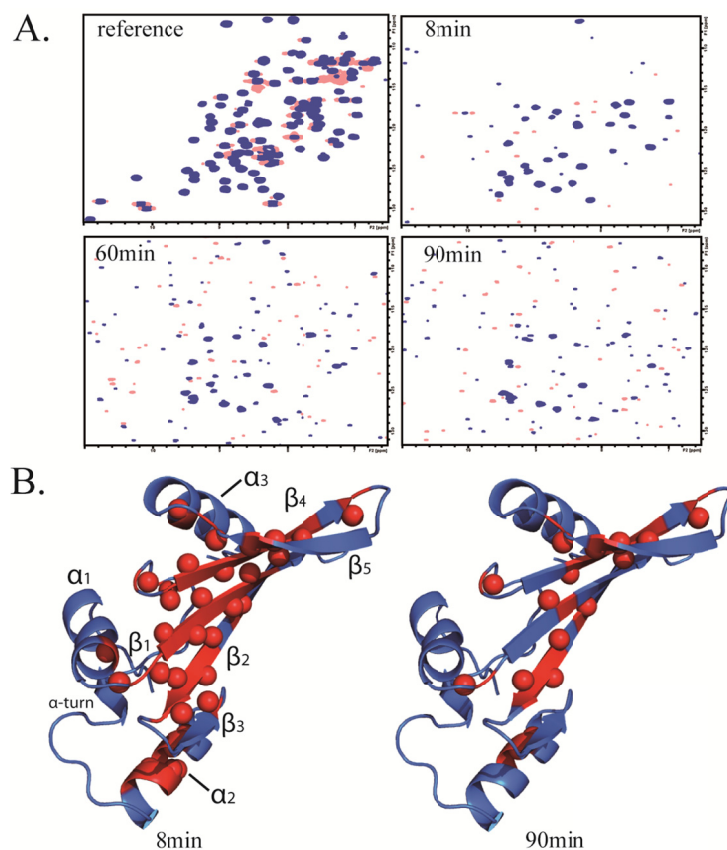


Fig. 26 Hydrogen to deuterium (H-D) exchange analysis of the TAM domain. (A) ^1H - ^{15}N SOFAST HMQCs recorded on a TAM domain sample at different times of exposure of the protein to D_2O . Positive signals are depicted in blue, negative in pale rose. (B) Areas of the TAM domain protected from exchange to D_2O after 8min and 90min exposure time (red) are plotted on structure of the TAM domain. Protected amide hydrogens are represented as red spheres.

After lyophilisation a TAM domain sample was redissolved in D_2O to look at the stability of the secondary structure elements. The H to D exchange was observed as the loss of resonance peak in a series of SOFAST HMQCs (Schanda & Brutscher, 2005), recorded in intervals (Fig.

26A). Mapping the H-D exchange result on the TAM structure indicates that β -sheets form an especially rigid structure and have a tight hydrogen bond network. A lot of amides are protected from the exchange to deuterons after 8min exposure to D_2O (Fig. 26B). β -strands β_1 , β_2 , β_4 are almost completely protonated at that time. Especially remarkable is that considerable parts of them are still protonated after 90min exposure to D_2O . α_2 shows a similar behavior. Instead, the α -helices α_1 , α_3 and the α -helical turn seem to be pretty labile. The amide protons there were completely exchanged to deuterons after the first 8min exposure.

7.6. NMR relaxation analysis.

From relaxation analysis, TAM domain is also a rigid folded structure with the exception of several interceptions, corresponding to loops (Fig. 27). Interestingly, the region Gly542-Trp546 is disordered, but shows high R_1 and average R_2 , indicating not only internal motion, but also possibly some chemical exchange. Loops Asn573-Val575 and Phe585-Arg588 are structured, based on values of 1H - ^{15}N -NOE, however, loop Asn573-Val575 has some internal motion (high R_1 , low R_2). Loop Phe585-Arg588 is unique, since it is well defined by NOEs but has a strikingly high R_2 value for S586 suggesting a considerable chemical exchange contribution.

From the relaxation analysis, τ_m was estimated from R_2/R_1 as 9-10 ns using the quadratic representation approach (Bruschweiler et al, 1995). That correlates well with the theoretical τ_c of 9.7ns for a 113 amino acid protein, from

$$\tau_0 = \left(9.18 \times \frac{10^{-3}}{T} \right) \exp \left(\frac{2416}{T} \right) N^{0.93}$$

(Daragan & Mayo, 1997), indicating that TAM domain behaves like a monomer at the high working concentrations in solution (0.6mM).

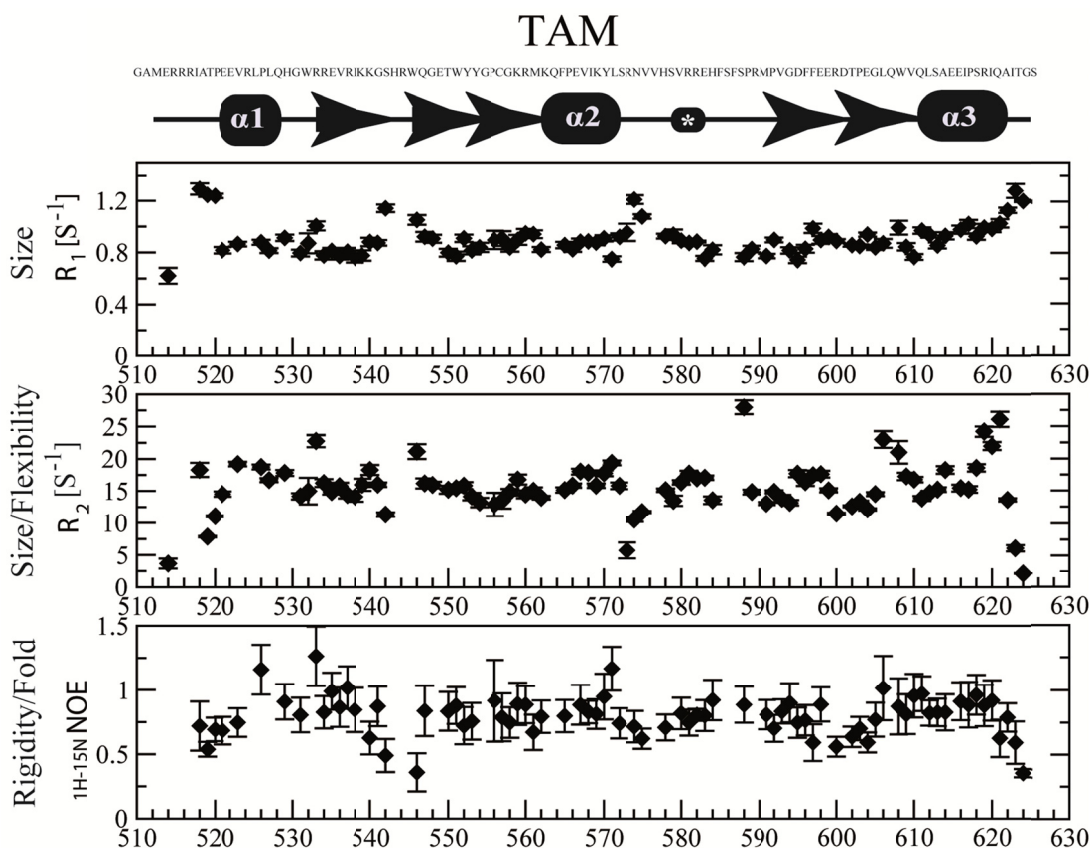


Fig. 27 NMR dynamics studies of TAM domain. Data were recorded on the 750MHz Bruker spectrometer at 293K. The sequence of the TAM7 construct, used for investigations as well as a cartoon of secondary elements are depicted at the top. α -helices are presented as ellipses and β -strands are represented by arrows. * is the α -helical turn.

7.7. Residue conservation.

Residue conservation analysis, performed with ConSurf web interface (Glaser et al, 2003; Landau et al, 2005) showed that the core of the TAM domain is formed by highly conserved residues (Fig. 28A). Solvent exposed amino acids are instead not well preserved. Areas, overlapping with the known structure of MBD1 domain, (β_1 , β_2 , β_3 , N-terminal β_4 and α_1) had higher conservation scores, than the novel β -sheet and α -helices, present only in the TAM domain (Fig. 28B). Here, some residues showed insufficient data, suggesting that those structural features are indeed unique to the TAM domain.

The conservation scores were calculated from 20 unique sequences detected by PSI-BLAST multiple sequence alignment (Altschul et al, 1997a), run as a module within ConSurf web interface with the server's default settings.

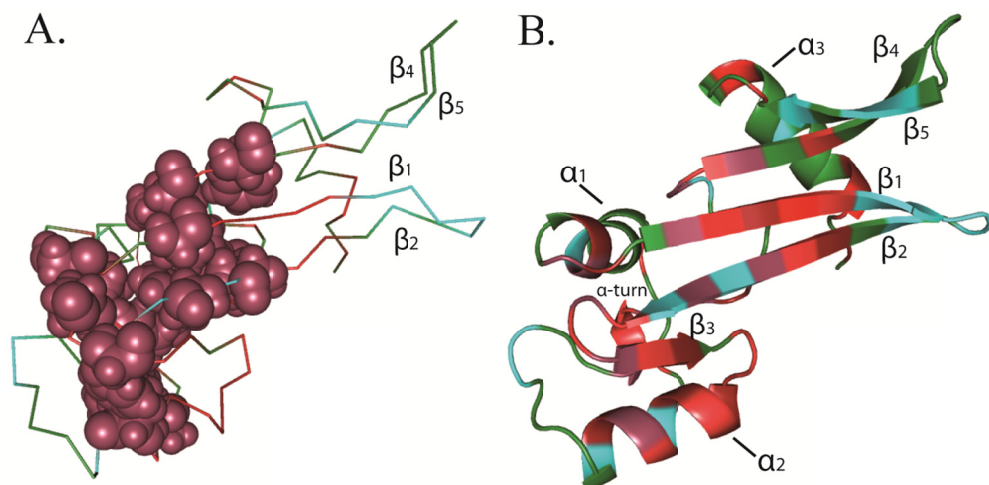


Fig. 28 Conservation of single amino acids of the TAM domain, plotted on the TAM structure. The analysis was performed using ConSurf web interface (Glaser et al, 2003). A conservation score was automatically assigned to each residue by the program based on its conservation. The color scheme used is raspberry (conservation score equals 9) to red (conservation score equals 6). Variable residues are colored in cyan (conservation score equals 1) Residues with conservation scores 5-2 are not present in the TAM domain. Residues with insufficient data (defined by ConSurf) are left green. (A) Space fill representation of the residues with highest conservation score (raspberry). (B) Cartoon representation of the TAM domain with patches colored according to their ConSurf residue conservation number.

8. Results III. NMR studies of pRNA.

8.1. Defining minimal TAM interaction sequence.

As mentioned in the introduction, pRNA folds into a stem-loop structure that is recognized by TIP5 (1.6). Though mammalian rDNA promoters share little homology, the pRNA fold was predicted to be common and conserved in human and mouse (Mayer et al, 2008), (Fig. 29A).

The published pRNA sequence comprised approximately 88 nucleotides in human and 79 nucleotides in mouse. That would correspond to a molecule of approximately 29kDa or 26kDa molecular weight, which was too big to be studied directly by NMR, indicating a need for optimization.

In EMSAs published, pRNA bound with high affinity to murine TIP5, regardless of whether pRNA from human or mouse was used (Mayer et al, 2008). Since the TAM domain has an identical sequence in both species (Fig. 15B) and the most published data was concentrating on murine pRNA, the constructs were based on the murine pRNA primary sequence. The fact that the integrity of the pRNA stem was inevitable for pRNA recognition by TIP5₁₋₅₉₈ and that the upper stem part of pRNA was protected from RNase cleavage by TIP5₁₋₅₉₈ interaction (Mayer et al, 2008) was exploited (Fig. 29A).

The most prominent features of the terminal pRNA part, like the poly(U) bulge, UU-CC mismatch and the three GC base pairs at the end of the stem (Fig. 29B), were preserved in constructs 1wt and 2sh (35 nucleotides, 11kDa and 33 nucleotides, 10.5kDa).

For better stabilization the constructs were closed with a thermostable tetraloop from the GNRA family (where N is any ribonucleotide and R is a purine) with the sequence of GAAA. Those loops are present within the most frequently occurring hairpins of 16S and 23S ribosomal RNAs and were reported to have melting transition temperature 4°C higher than less frequently occurring sequences. Their structure has been solved by NMR (Heus & Pardi, 1991) and shows extensive base stacking, as well as other remarkable features. For example, an unusual GA base pair between the first G and the last A, so that A is positioned exactly over the H1' of the next nucleotide, introducing large ring current effect on its base. That accounts for an unusual upfield chemical shift of H1' (more than 1.5ppm from normal position) (Heus & Pardi, 1991). Hence, the GAAA loop sequence would facilitate formation of a hairpin in our constructs and simplify NMR assignments.

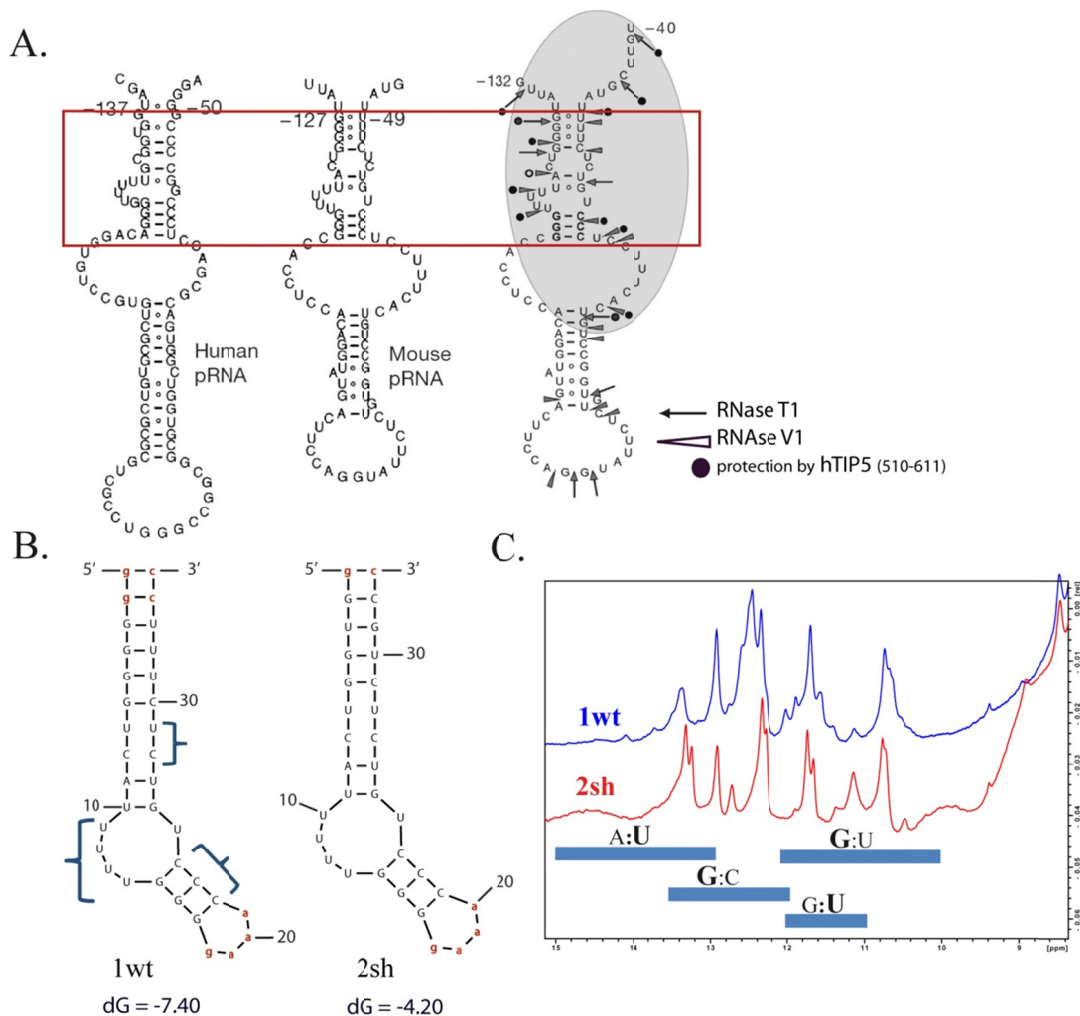


Fig. 29 Defining the minimal pRNA sequence interacting with TAM domain. (A) Computer predicted secondary structure of human (*left*) and mouse (*middle*) pRNA. Predictions were performed using 'RNA Alifold' program (Hofacker, 2003). Schematic depiction of secondary structure of mouse pRNA as shown by RNase footprinting assay (*right*). Area of pRNA, shielded from RNase cleavage by interaction with human TIP5, is shaded in grey. The numbers refer to the position of nucleotides respective the Pol I transcription start site. From (Mayer et al, 2008). Secondary structure features, shared by human and mouse pRNA in the cleavage protected area, are contained within the red box. (B) Secondary structure of pRNA constructs for NMR. Prediction was done with Mfold (Zuker, 2003) web server. Secondary structure features of full length pRNA that were conserved in both constructs, are marked with blue parenthesis on 1wt-RNA. Gibbs free energy, calculated by Mfold for every RNA, is given as dG. Nucleotides added to the constructs for purposes of stabilization are noted in low-case letters and are highlighted red. The numbering refers to nucleotide position in the construct. (C) Overlay of 1D NMR spectra of 1wt and 2sh. Region on the imino proton resonances is shown. The typical chemical shift regions of imino proton resonances in Watson-Crick base pairs and wobble base pairs (Furtig et al, 2003), are highlighted with blue bars. The base of observed imino proton is marked bold.

Additional two (for 1wt) or one (for 2sh) GC base pair was added to the stem to stabilize the fold. Several steps of optimization, aiming introduction of minimal amount of changes, as compared to full length pRNA, still preserving its secondary structure, were undertaken. For every step secondary structure predictions were done using Mfold (Zuker, 2003) web interface. Construct 1wt carried the wildtype sequence of mouse pRNA as published in (Mayer et al, 2008), the construct 2sh had one GU base pair inverted, as compared to wildtype to facilitate potential assignment of NMR spectra.

Both constructs were transcribed *in vitro* (Fig. 30) and studied by NMR. 1D spectra recorded at 277K in H₂O on Bruker 900MHz spectrometer with a cryoprobe (Fig. 29C) showed approximately the number of imino proton resonances of guanines and uraciles in base paired regions, as expected from secondary structure prediction (Furtig et al, 2003). Due to the overlap that is generally present in 1D spectra, it was not possible to investigate that samples further, however it indicated that the RNA constructs were most probably folded into the expected conformation.

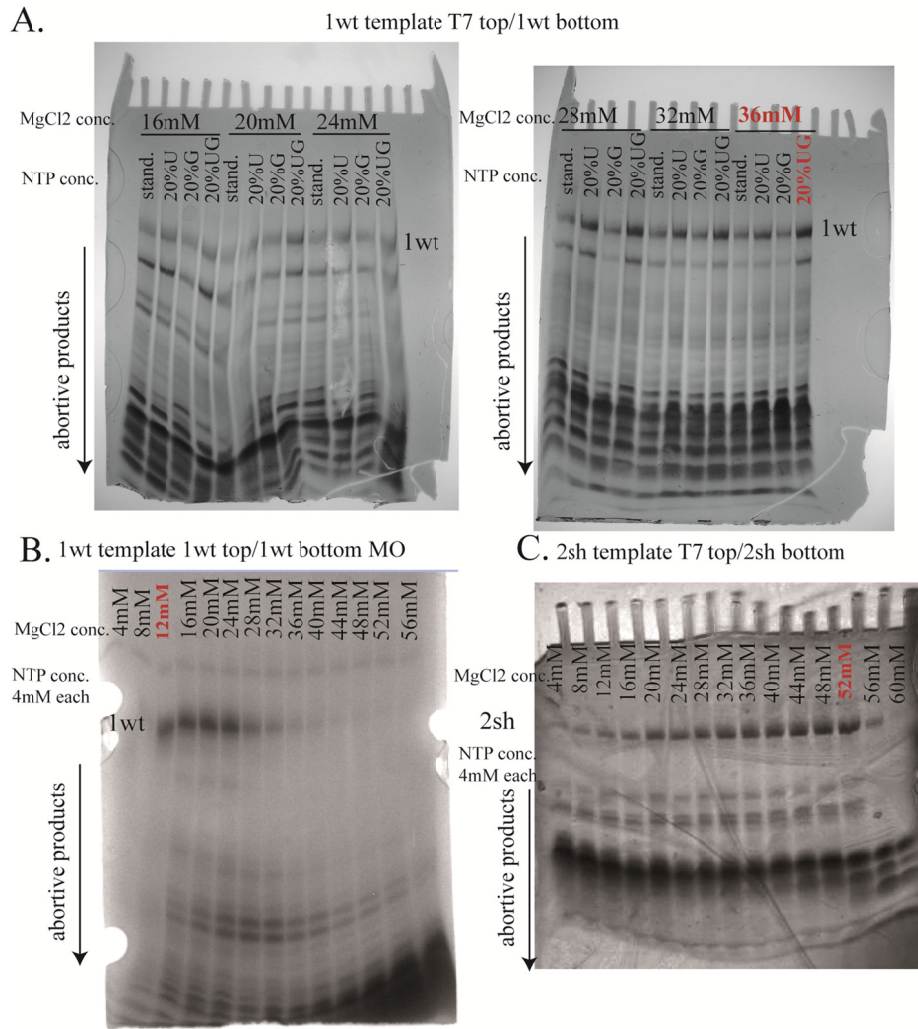


Fig. 30 Optimization of *in vitro* transcription conditions for pRNA constructs 1wt and 2sh in small scale. (A) Optimization of MgCl₂ and NTP concentration for the 1wt *in vitro* transcription from a single stranded DNA template, consistent of the T7 bacteriophage promoter and a DNA bottom strand complimentary to 1wt, as indicated above the picture. For each MgCl₂ condition a standard NTP concentration (4mM each) as well as a 20% excess of UTP (20%U) or GTP (20%G) or both UTP and GTP (20%UG) was tested. MgCl₂ concentration and NTP conditions are indicated on top of the lanes. The position of the 1wt transcript is indicated. The most optimal transcription condition in terms of yield that was used in big scale 1wt production is highlighted red. (B) Optimization of MgCl₂ concentration for the 1wt *in vitro* transcription from a fully double stranded DNA template with methoxylation of the bottom strand. NTPs were used at standard concentration of 4mM each; MgCl₂ concentrations are indicated on top of the lanes. The position of the 1wt transcript is marked. The most optimal transcription condition in terms of yield and decreased number of abortive products that was used in big scale 1wt production is highlighted red. Gel produced by A. Woerle while internship. (C) Optimization of MgCl₂ concentration for the 2sh *in vitro* transcription from a single stranded DNA template, consistent of the T7 bacteriophage promoter and a DNA bottom strand complimentary to 2sh. NTPs were used at standard concentration of 4mM each; MgCl₂ concentrations are indicated on top of the lanes. The position of the 2sh transcript is marked. The most optimal transcription condition in terms of yield that was used in big scale 2sh production is highlighted red.

As next, 1wt and 2sh were tested in competition EMSA for TIP5-binding by Dr. Christine Mayer (dkfz, lab Prof. Grummt). The results showed that both constructs were interacting with TIP5 with high affinity (Fig. 31). The interaction of TIP5 with 1wt RNA was 10-20 times weaker, compared to full length murine pRNA. Interestingly, inversion of only one base pair in the stem, worsened the affinity to TIP5, making it 30-60 times weaker for 2sh, as compared to full length murine pRNA.

The K_D of interaction for both pRNA constructs with TIP5 was estimated to be in the nanomolar range. Given the effect of base alterations in the stem and the more favorable Gibbs free energy for 1wt, as predicted by Mfold (Fig. 29B), further work was accomplished with 1wt RNA.

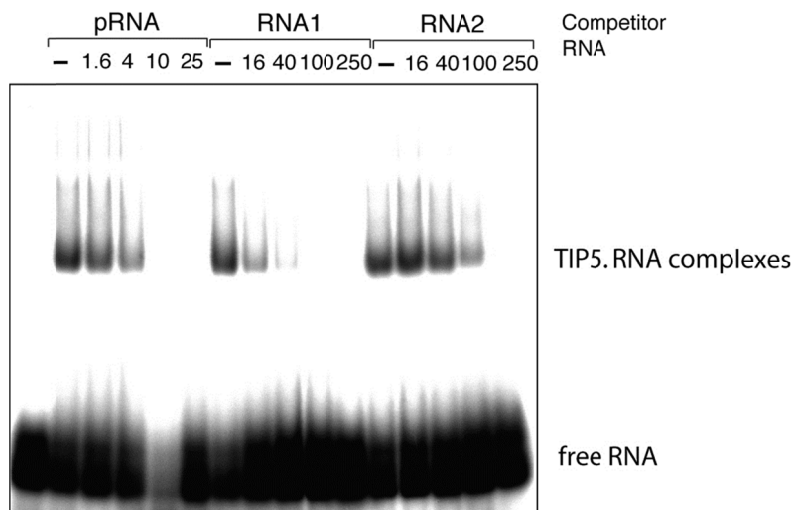


Fig. 31 Investigation of TIP5 affinity of truncated pRNA constructs. Electrophoretic shift mobility assay with competition, kindly provided by Dr. Christine Mayer (dkfz, lab Prof. Grummt) Affinity of murine pRNA (*pRNA*) to murine TIP5 (residues 1-598) was compared to the affinity of 1wt RNA (*RNA1*) and 2sh RNA (*RNA2*) to the same TIP5 construct. The unspecific RNA from the multiple cloning site (MCS) of the pBluescript-KS vector was labeled with ^{32}P , the signal of which was detected. The binding was competed by a specific unlabelled RNA probe. The amount of unlabelled competitor is indicated on top of each lane.

8.2. Structural studies of 1wt construct.

Construct 1wt was transcribed unlabelled *in vitro* on a large scale, so that a 0.4mM sample could be obtained. That sample was used to record a set of NOESY experiments with different mixing times at 277K in a water-based buffer, as well as at 300K in D₂O. Additionally, natural abundance $^1\text{H}^{13}\text{C}$ -HSQCs were recorded for the sugars moieties of nucleotides and nucleobases and $^1\text{H}^1\text{H}$ -TOCSY in D₂O at 300K.

A homonuclear 2D NOESY in 20mM sodium phosphate buffer, pH=6.5, recorded with 300ms mixing time on Bruker 900MHz spectrometer, equipped with a cryoprobe, was used for sequential assignment of imino proton resonances (Fig. 32). Sequential walk from the bases G3 to U7, as well as from G15 to G18, could be performed. U29 imino proton resonance was detected as well.

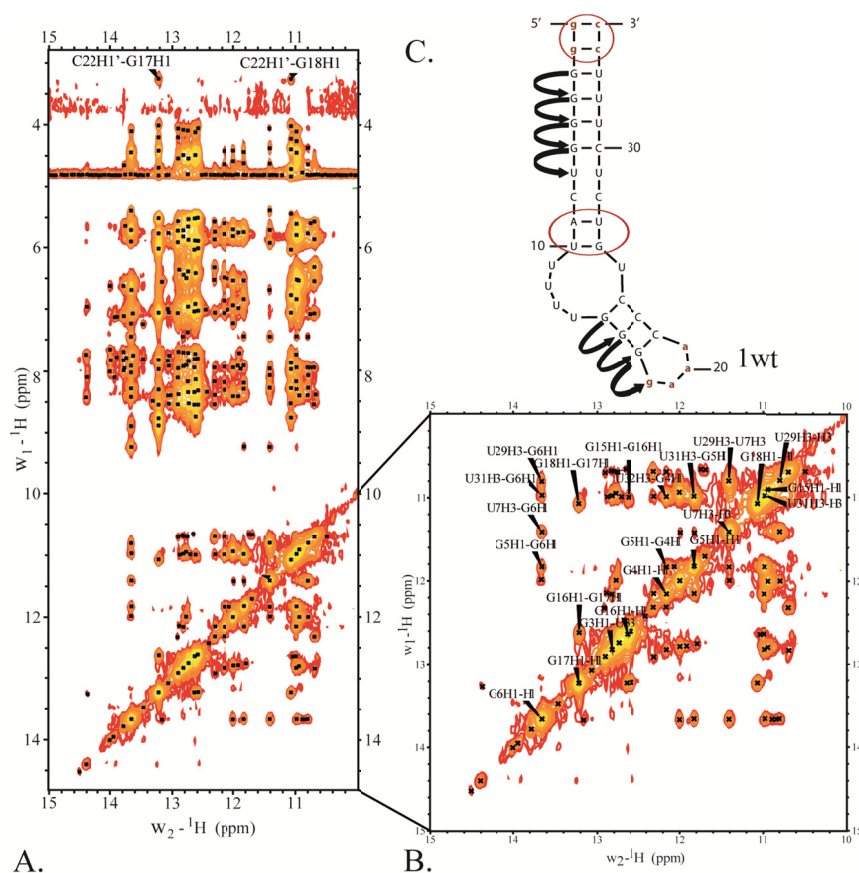


Fig. 32 Structural studies of 1wt, water NOESY. (A) Homonuclear 2D NOESY on 900MHz Bruker spectrometer with cryoprobe, 300ms mixing time in water-based buffer. Section around imino-imino and imino-amino correlation region, showing as well the unique C22H1' cross peaks to G17 and G18 introduced by the GNRA loop. (B) Close up of the imino-imino proton correlation region with chemical shift assignments. (C) Schematic representation of the sequential imino walk (arrows) that was performed on 1wt using water-based NOESY. Nucleotides missing in assignment are circled red. Nucleotides that were added to the construct for the purpose of stabilization are shown in low case letters and colored red.

C22H1' cross peaks to G17H1 and to G18H1 could be easily identified due to the abnormal upfield shift of C22H1' (3.25ppm), induced by the GAAA tetraloop. Most probably the UU wobble base pair between imino H3 and carbonyl C2 of U7 and U29 was also detected (see Appendix).

However, the sequential walk could not be completed and not all resonances could be assigned unambiguously, e.g. the terminal GC base pairs could not be identified due to solvent exchange, as well as information on the A9U27, U10G26 base pair and U11-U14 resonances from the poly(U) bulge. That gave a first indication, that the 1wt RNA may be fairly dynamic and may be adopting several conformations. In addition, some ambiguous cross peak resonances could be detected, suggesting, that the 1wt loop was not stable enough for structure determination by NMR in these conditions.

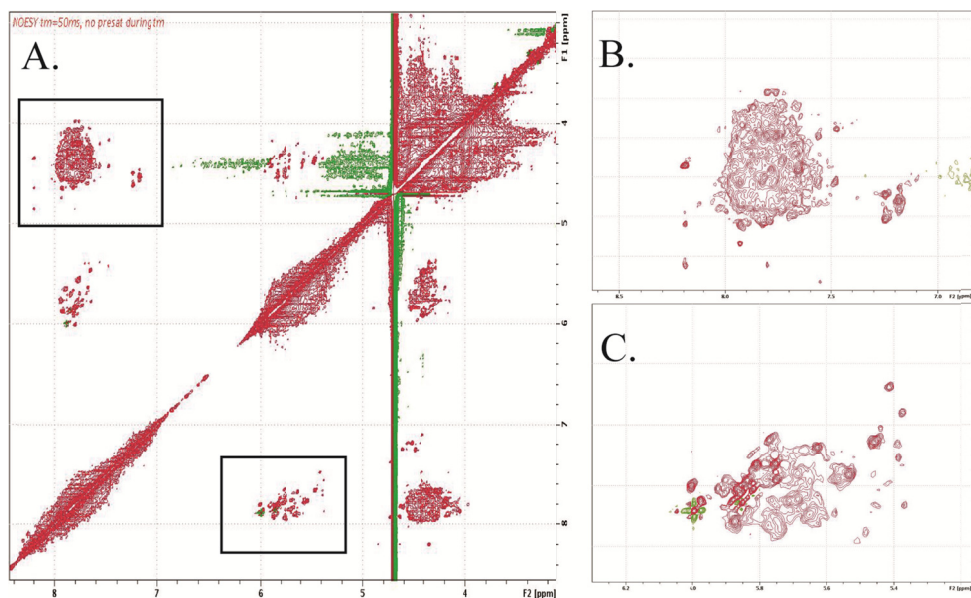


Fig. 33 Structural studies of 1wt, NOESY in D2O. Positive peaks are shown in red, negative peaks in green. (A) A Section of homonuclear 2D NOESY on 900MHz Bruker spectrometer with cryoprobe, 50ms mixing time D2O. (B, C) A close up of the two regions needed for the aromatic-anomeric sequential walk. Spectra show severe overlap, preventing unambiguous identification of any single signals.

To get confirmation of assignments and further information on the missing nucleotides, an attempt to carry out an NOE sequential walk between aromatic and anomeric proton resonances, using homonuclear 2D NOESY recorded at different mixing times in D2O, was performed (Furtig et al, 2003). Unfortunately, spectra showed a severe overlap (Fig. 33), which made it impossible to identify unique signals, indicating supposedly also partial oligomerization of free 1wt RNA at used conditions. Analysis of the NMR sample on denaturing 20% acrylamide/bis-acrylamide RNA gels showed that the 1wt RNA was intact and no RNA degradation was present.

Hence, the structure of 1wt RNA could not be investigated by NMR at the used conditions and both, the experimental conditions and the construct itself needed to be optimized for NMR structure determination.

8.3. The “short” construct.

An attempt to optimize the 1wt RNA construct for NMR by further minimizing it was undertaken.

Chemical shift changes of imino proton signals were analyzed in one-dimensional ^1H spectra of the free 1wt RNA and the 1wt in complex with TIP4N protein at 0.5 molar equivalents of RNA. The spectra were recorded on a Bruker 600MHz machine equipped with a cryoprobe at 278K to reduce chemical exchange of the imino protons with solvent. Chemical shift changes could be detected for the imino proton of G5 and in the region of U7 and U29-U32 imino resonances (Fig. 34A).

Mapping the shift changes on the predicted secondary structure of 1wt revealed that they cluster mainly on the stem of 1wt (Fig. 34B). Since the competition EMSA assays performed with 1wt and 2sh also consistently indicated that minor differences in the stem sequence of the RNA construct as compared to wildtype pRNA lead to considerable reduction in their affinity of to TIP5, a new short RNA construct (termed “short”), comprising only the stem region of 1wt was designed (Fig. 34B). For better stability it was closed as well with a thermostable GNRA tetraloop (where N is any ribonucleotide and R is a purine) with the sequence GAAA. The “short” construct was transcribed *in vitro*, whereas the transcription conditions were optimized for high yields (Fig. 35), and studied by NMR.

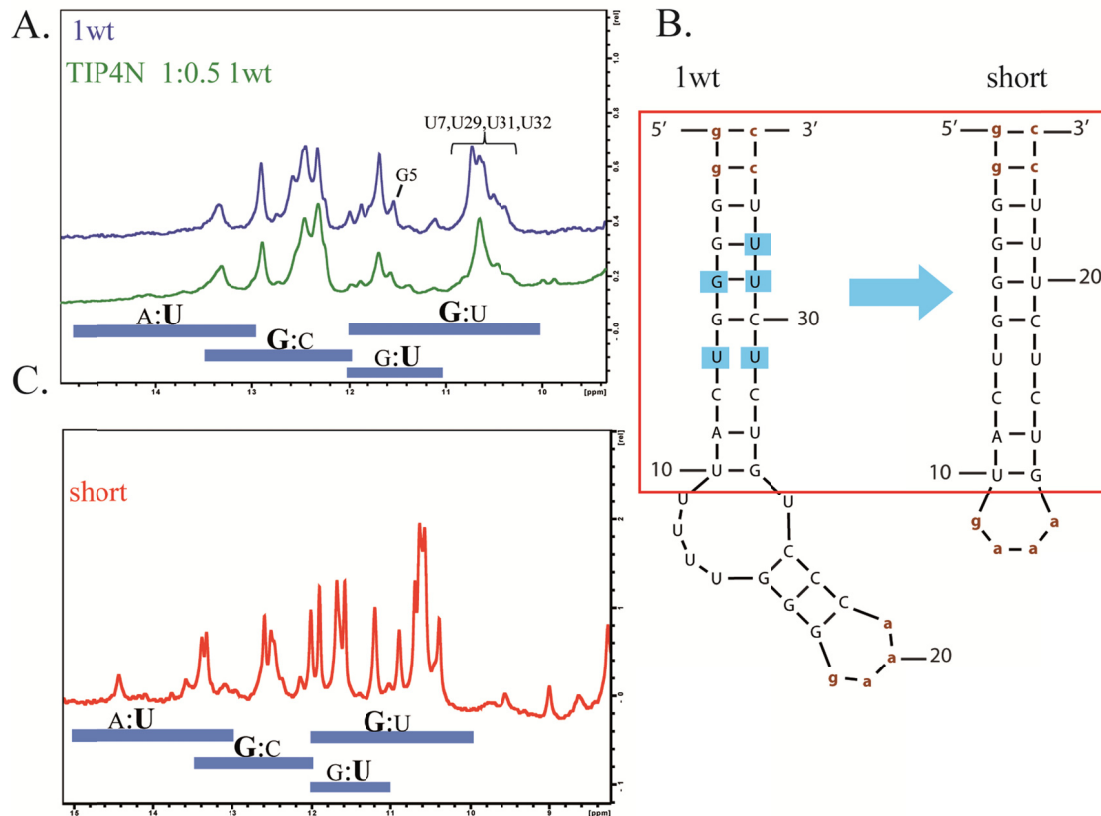


Fig. 34 Creation and NMR investigation of the “short” pRNA construct. (A) Overlay of 1D ^1H spectra of free 1wt RNA (shown in blue) and TIP4N in complex with 1wt at 0.5 molar RNA equivalents (shown in green). The imino proton section is depicted. Chemical shift changes could be detected in the region of certain imino proton resonances, the corresponding bases of the 1wt construct are indicated above. The typical chemical shift regions of imino proton resonances in Watson-Crick base pairs and wobble base pairs (Furtig et al, 2003), are highlighted with blue bars. The base of observed imino proton is marked bold. (B) Chemical shift changes of 1wt upon TIP4N interaction mapped on the predicted secondary structure of 1wt RNA (left) and the secondary structure prediction for the “short” RNA construct (right). The affected bases of 1wt are highlighted in blue. Both secondary structures were predicted with Mfold (Zuker, 2003) web server. Nucleotides, added to the constructs for purposes of stabilization, are noted in low-case letters and are highlighted red. The numbering refers to nucleotide position in the construct. (C) The imino proton section of the 1D ^1H spectrum of the free ‘short’ construct, recorded at 277K on a Bruker 900MHz machine equipped with a cryoprobe. The typical chemical shift regions of imino proton resonances in Watson-Crick base pairs and wobble base pairs (Furtig et al, 2003), are highlighted as in (A).

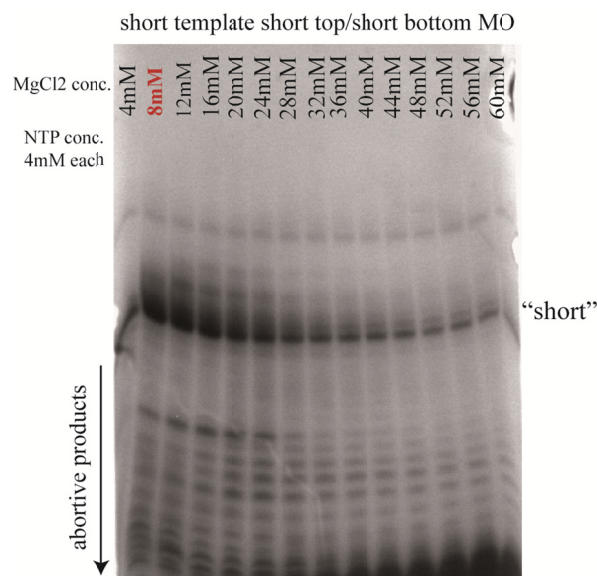


Fig. 35 Optimization of *in vitro* transcription conditions for the “short” pRNA constructs in small scale. Optimization of MgCl₂ concentration for the “short” construct *in vitro* transcription from a fully double stranded DNA template with methoxylation of the bottom strand. NTPs were used at standard concentration of 4mM each; MgCl₂ concentrations are indicated on top of the lanes. The position of the “short” transcript is marked. The most optimal transcription condition in terms of yield and decreased number of abortive products that was used in big scale “short” production is highlighted red. Gel produced by A. Woerle while internship.

From one dimensional ¹H spectrum in H₂O it could be concluded that the “short” construct is most probably folded in the expected manner (Fig. 34C). However, a series of homonuclear NOESY experiments recorded further on the “short” in H₂O and in D₂O, showed multiple sets of peaks for a big number of resonances, indicating that the short RNA was most probably again very dynamic in solution and could be present in multiple conformations (data provided by Dr. A. Dallmann). Furthermore, tested for interaction with TIP5 truncation constructs, the “short” RNA construct caused severe precipitation of both, the TIP4N and TAM7 proteins.

Thus, unfortunately the RNA assignment problem could not be solved by creating the “short” construct and further NMR structural studies of pRNA region that interacts with TIP5 protein would still need extensive optimization of the RNA construct.

Given the high affinity and specificity of 1wt RNA for TIP5, it was decided to study the pRNA interaction with TIP5 by NMR titrations using the 1wt.

9. Results IV. pRNA interaction with the TAM domain.

9.1. NMR titration of TIP4N with 1wt RNA.

To investigate pRNA interaction with TIP5, NMR titrations were performed with ^{15}N TIP4N protein and unlabelled 1wt RNA construct. Interaction was monitored observing changes in 2D ^1H - ^{15}N -HSQC spectra of the protein upon addition of increasing amounts of RNA up to the molar ratio of 1:1. As reference the ^1H - ^{15}N -HSQC of the same sample without RNA was used (Fig. 36A).

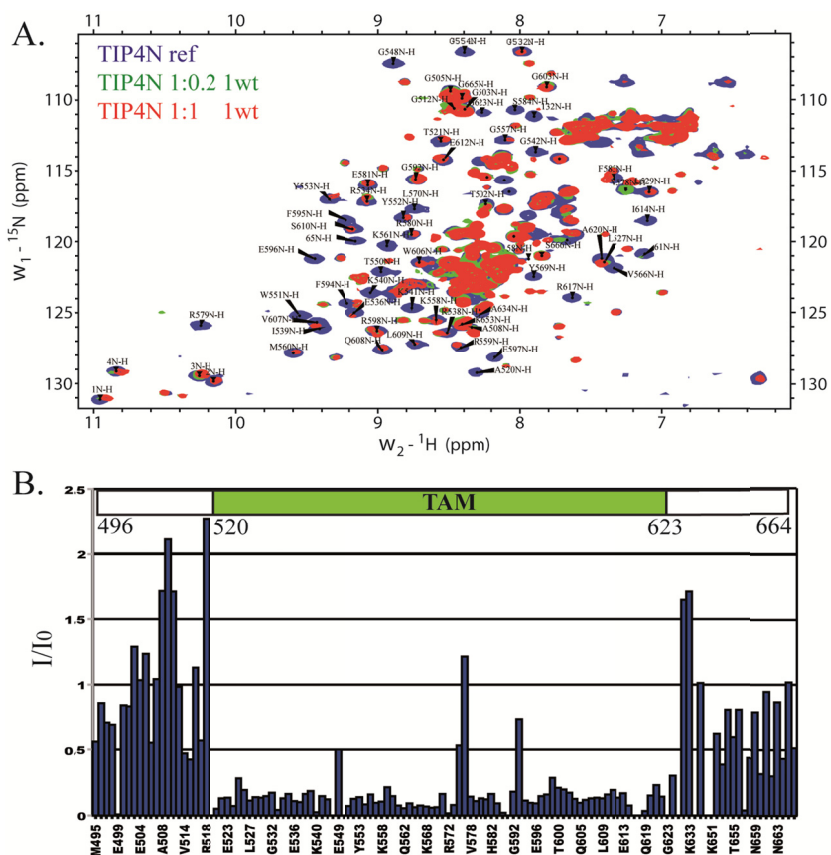


Fig. 36 Titration of TIP4N with 1wt RNA construct. NMR chemical shift perturbation experiment. (A) Overlay of ^1H - ^{15}N HSQC spectra recorded upon addition of 1wt RNA at different molar ratios. The reference spectrum is shown in blue, spectrum at 1:0.2 ratio TIP4N:1wt is shown in green and at 1:1 ratio TIP4N:1wt in red. Assignments of the resonances resolved in the reference spectrum are indicated. The ratios of the protein to RNA in the respective color code are indicated in the upper left corner of the spectrum. (B) Plot of intensity ratio I/I_0 on the sequence of the TIP4N construct. Here I_0 is the intensity of a peak in the reference spectrum and I is the corresponding peak intensity in the final spectrum at 1:1 TIP4N:1wt. Schematic representation of the construct is given on top of the diagram with indication of the boundaries of the TIP4N construct and of the TAM domain.

The major effect observed, was a prominent drop in intensity of resonance peaks, corresponding to residues, involved in protein secondary structures (peaks outside of the 8ppm-8.6ppm region). That indicated, that interaction could be observed by NMR and that the changes in intensity were not only due to protein precipitation, since in the latter case an overall intensity drop would have been observed. Plotted on TIP4N sequence (Fig. 36B), significant changes of intensity were observed within the sequence, corresponding to the actual TAM domain (residues 520-623), indicating that pRNA was interacting predominantly with the TAM domain itself.

Though intensity rise for residues in some flexible regions flanking TAM domain may indicate a gain of some secondary structure upon 1wt interaction, it was not clear, whether it was really the case. As those peaks cluster within crowded regions of the spectrum, their intensities could not be extracted precisely and could represent overlap of several resonances.

Unfortunately, protein precipitation and degradation, indicated by additional peaks in the HSQC, did take place during the experiment and thus, the amount of useful information to extract, was limited. However, as titration of 1wt to ^{15}N labeled TAM7 lead to severe precipitation of the protein, further investigations of pRNA/TIP5 interaction were performed with TIP4N constructs.

9.2. Mutational analysis I, based on residue conservation.

Based on NMR titration described in 9.1, a set of TIP4N mutants was created with the aim to disrupt TIP4N/1wt RNA interaction and, thus, find key residues for TIP5/pRNA interaction.

Based on the hypothesis, that pRNA could utilize the β -sheet surface, as it was known for MBD interaction with methylated DNA (Ho et al, 2008; Ohki et al, 2001), the following mutations were suggested (Table 4, Fig. 37A). The amides of all of them were affected in the NMR titration of TIP4N upon 1wt RNA addition, side chains were all surface-exposed. Most importantly, those residues were all conserved within the TAM domains (Fig. 15B).

Mutation		Mutation	
1.	Q547A	6.	S586A
2.	W551A	7.	E596A
3.	Y553A	8.	R617A
4.	Q562A	9.	Q619A
5.	Q605A		

Table 4. Mutations investigated in analysis I, based on residue conservation and assumption, that pRNA could utilize the β -sheet surface of the TAM domain.

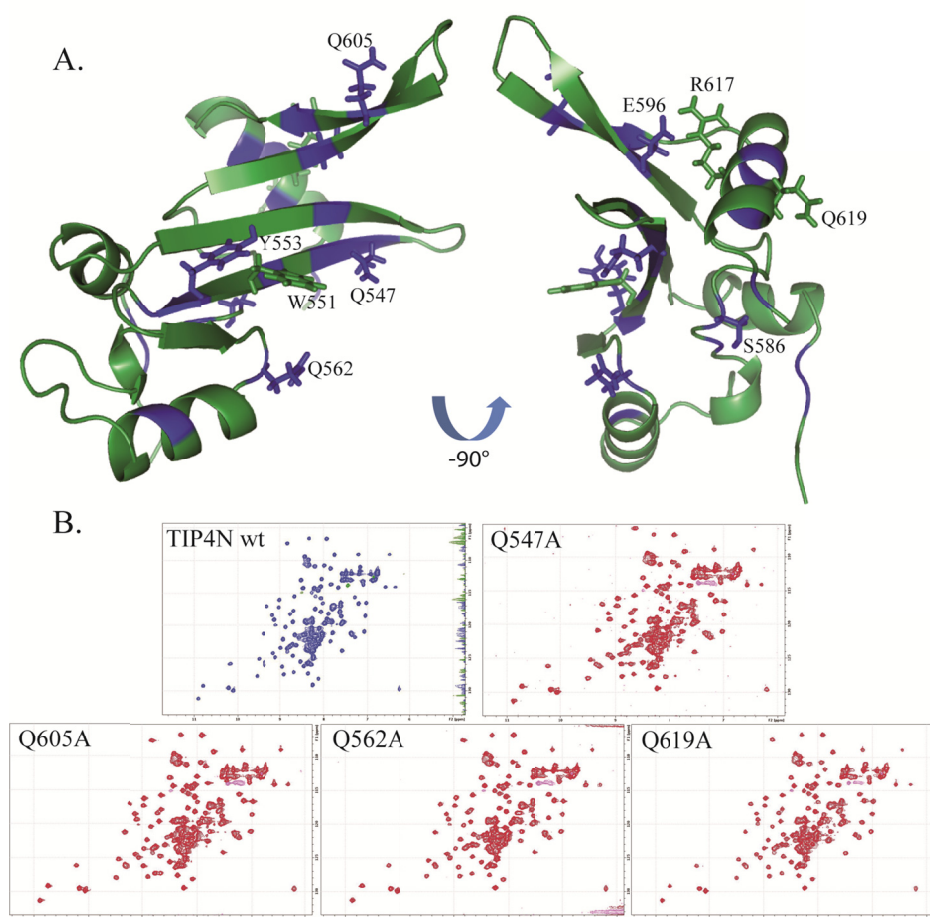


Fig. 37 Mutational analysis of TIP4N/1wt RNA interaction I. (A) Residues mutated to alanine are plotted on the TAM domain structure. Their side chains are shown as sticks. Regions of the protein experiencing major impact upon 1wt RNA addition (major intensity drop already at 1:0.2 ratio of TIP4N:1wt) are colored blue. (B) ^1H - ^{15}N HSQC spectra of the mutants that could be expressed and purified folded and were used for pRNA binding analysis. The corresponding mutation is indicated on top of the spectrum. For TIP4N wt (wildtype) positive peaks are shown in blue and negative peaks are shown in green. In case of the mutants, positive peaks are shown in red, and negative peaks in magenta.

Residues Q547, Y553, Q562 and Q605 were close to the suggested DNA binding site. S586, E596, R617 and Q619 were on the opposite side of the domain were chosen as control. W551 selected for mutagenesis predominantly, since it was reported, that mutation of residues corresponding to W551/Y552 in murine system had in context of the full length murine TIP5 a negative effect on pRNA binding and a negative *in vivo* effect on induction of heterochromatic hallmarks by NoRC (Mayer et al, 2006).

All mutants were expressed ^{15}N labeled in *E.coli* and purified from cytosol using standard protocols (see 4.1.4). Q547A, Q562A, Q605A and Q619A could be obtained in amounts high enough for experiments and were folded as the wild type TIP4N, as observed by NMR (Fig. 37B). Those mutants were tested for pRNA binding in electrophoretic mobility shift assays (EMSAs) and filter binding assays (FBAs), kindly provided by our collaborator Dr. S. Melnik in Prof. Dr. I. Grummt lab, Division of Molecular Biology of the Cell II, dkfz.

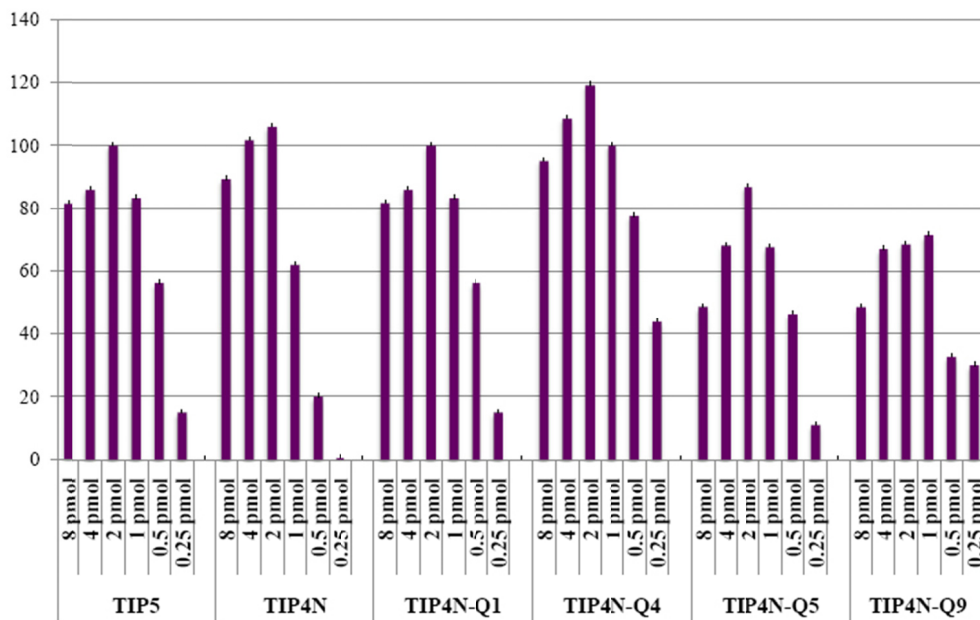


Fig. 38 Filter binding assay (FBA), performed with TIP4N mutants, series I. The results were kindly provided by Dr. S. Melnik, Division of Molecular Biology of the Cell II, dkfz. Different TIP5 constructs were compared by their binding to ^{32}P -labelled RNA from multiple cloning site (MCS) of the pBluescript-KS vector. A panel corresponds to every construct. Proteins were used at decreasing concentrations. The concentrations are indicated at the bottom of each panel. On the y axis the average scintillation counts of the corresponding signal are depicted. The background was subtracted and the counts were normalized to the average signal given by TIP5 interaction with MCS RNA at 2pmol of protein. Error bars were calculated as standard deviation of a duplicate. Q1 corresponds to Q547A mutant, Q4 to Q605A, Q5 to Q562A and Q9 to Q619A mutant.

RNA binding function in all the point mutants was not much affected, as can be seen on example of an FBA (Fig. 38). It was decreased to 70% for Q619A and 90% for Q562A. TIP5 is known to bind RNA with high affinity, and thus it lead to a suggestion that none of those residues was directly involved in pRNA interaction.

9.3. Complex reconstitution of TIP4N/1wt.

To gain more insight into the 1wt binding to TIP4N, a new set of NMR chemical shift perturbation experiments was recorded. To circumvent the protein precipitation problem, that time instead of introducing a small amount of concentrated RNA to concentrated protein solution, TIP4N and 1wt RNA were mixed as dilute solutions and concentrated together. The final protein concentration was aimed to be kept the same every time, whereas RNA was added in different molar ratios. To improve TIP4N relaxation properties a ^{15}N labeled protein sample with random deuteration of ca.50% was used.

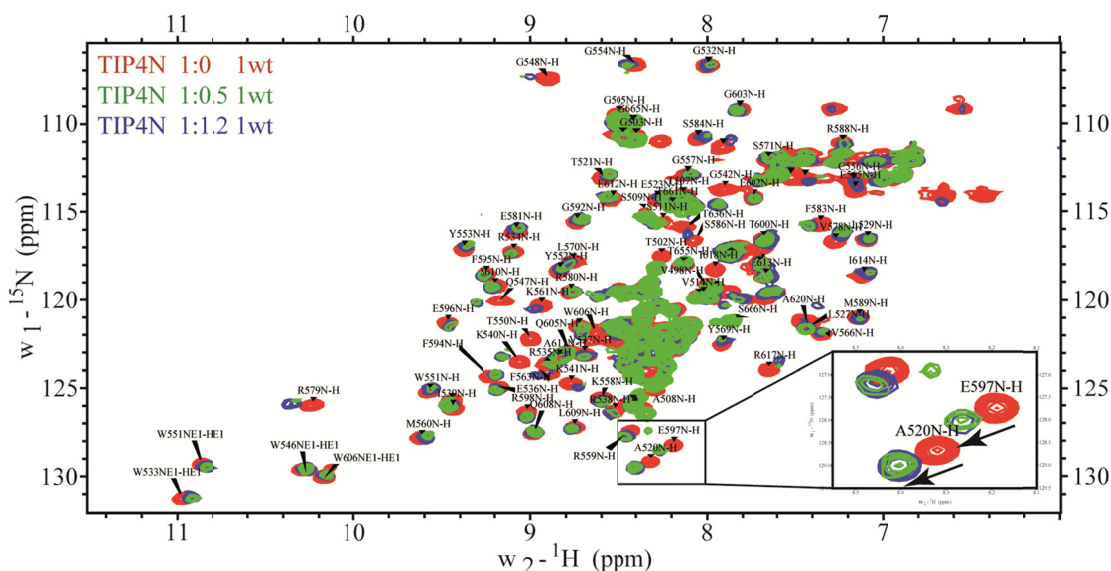


Fig. 39 Complex reconstitution of TIP4N with 1wt RNA. Overlay of ^1H - ^{15}N HSQC spectra recorded on TIP4N/1wt complexes with different molar ratios of protein to 1wt RNA. The reference spectrum is shown in red, spectrum at 1:0.5 ratio TIP4N:1wt is shown in green and at 1:1.2 ratio TIP4N:1wt in blue. Assignments of the resonances resolved in the reference spectrum are indicated. In the bottom right corner is a close-up of the squared region of the spectrum showing peak shifts upon addition of the RNA. The ratios of the protein to RNA in the respective color code are indicated in the upper left corner of the spectrum.

That method of preparation allowed recording an NMR titration set of good quality (Fig. 39). The binding was in the fast to intermediate exchange regime on the NMR timescale. Chemical shift changes occurred upon binding of 1wt RNA and could be monitored, indicating a high affinity interaction. Some resonances experienced line broadening, suggesting the K_D for interaction in the low micromolar to high nanomolar range. Saturation was achieved at the ratio of about 1:0.5 TIP4N:1wt, suggesting that 1wt RNA might have two TIP4N binding sites. However, that hypothesis requires further investigation, as protein and RNA concentrations could not be determined precisely after concentrating of the sample.

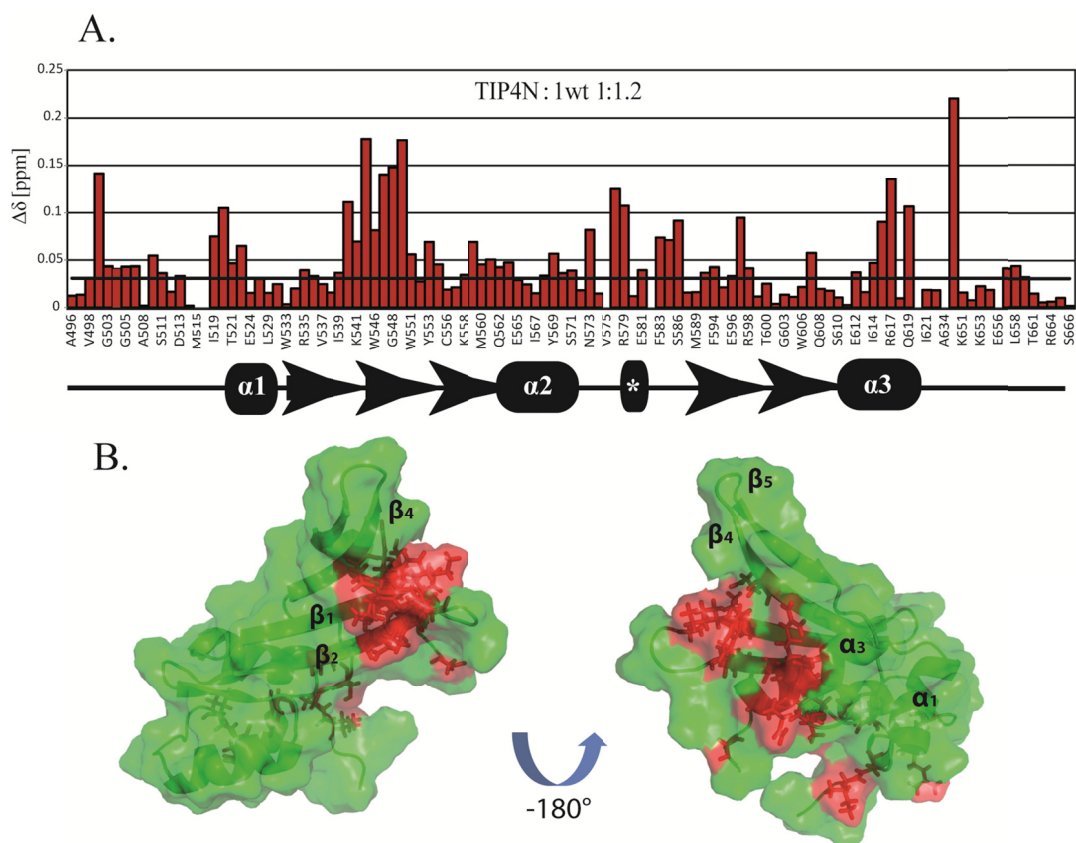


Fig. 40 Mapping the residues affected in TIP4N/1wt complex reconstitution on the TAM domain structure. (A) Chemical shift changes of ^1H - ^{15}N resonances in TIP4N spectrum introduced by addition of 1.2 molar equivalents of 1wt RNA, plotted on the sequence of the TIP4N construct. At the bottom a cartoon of secondary elements is depicted. α -helices are presented as ellipses and β -strands are represented by arrows. * is the α -helical turn. (B) Plot of the chemical shift perturbations on the structure of the TAM domain. The surface perturbed upon addition of 1.2 molar equivalents of 1wt RNA to TIP4N is colored red.

Mapping the affected residues on the TAM domain structure revealed that the area of actual interaction was not very large (Fig. 40). The residues affected by 1wt RNA interaction

clustered on the surface of the domain in the region formed by the tips of the strands β_1 , β_2 , β_4 and the middle part of C-terminal α -helix α_3 that was not present in MBD domains. That results corresponded well to the FBA studies from 9.2, as mutations investigated there were not directly at the interaction surface, but Q619A was close to it.

9.4. Mutational analysis II, based on titration results 9.3.

To validate the NMR titration results and identify key interacting residues with pRNA, several charge inversion mutations of amino acids with solvent exposed side chains, affected by 1wt interaction, were created (Table 5, Fig. 41A).

Mutation		Mutation	
1.	R545E	5.	K540E/K541E
2.	R617E	6.	W546Y
3.	R598E	7.	W546A
4.	K541E	8.	S616D

Table 5. Mutations investigated in analysis II, based on NMR chemical shift perturbation experiments on the TIP4N/1wt complex.

All mutants were expressed ^{15}N labeled in *E.coli* and purified from cytosol using standard protocols. Structural integrity of the proteins was monitored by $^1\text{H}^{15}\text{N}$ -HSQCs (Fig. 41B). K541E, W546Y, W546A, R545E, S616D and R617E mutants were folded, however R617E was probably showing some oligomerization, indicated by line broadening in NMR spectra. The double mutant K540E/K541E and the R598E mutant were aggregated.

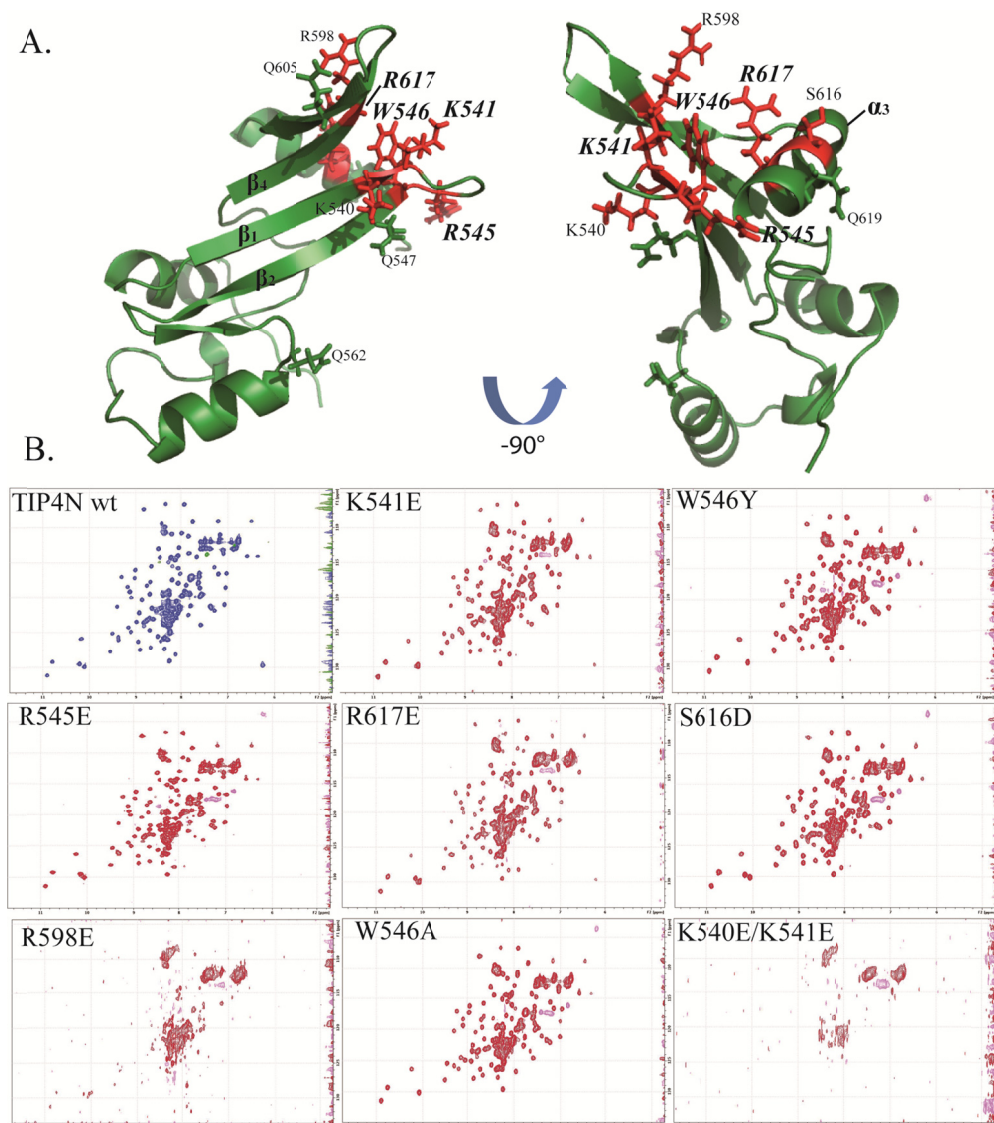


Fig. 41 Mutational analysis of TIP4N/1wt interaction II. (A) Residues mutated for analysis I and II are plotted on the TAM domain structure. Their side chains are shown as sticks. Residues, mutated in analysis II are highlighted red. Mutations having an impact on RNA binding properties of TIP4N are assigned in bold italic style, see also **Fig. 42**. (B) ^1H - ^{15}N HSQC spectra of the mutants that could be expressed and purified and were used for pRNA binding analysis. The corresponding mutation is indicated on top of the spectrum. For TIP4N wt (wildtype) positive peaks are shown in blue, negative in green. In case of the mutants positive peaks are shown red, negative are colored magenta.

All mutants were used for filter binding assays by our collaborators at the dkfz (Fig. 42). There the binding specificity of the mutants towards pRNA was challenged.

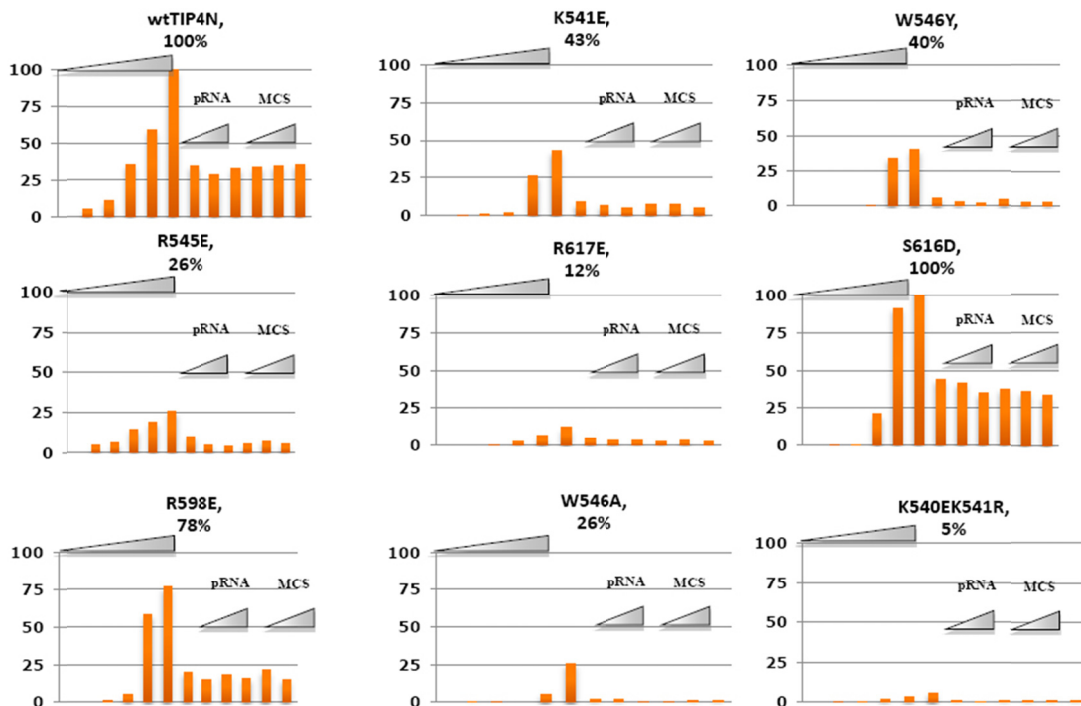


Fig. 42 Filter binding assay (FBA), performed with TIP4N mutants, series II and ^{32}P -labelled pRNA. The results were kindly provided by Dr. S. Melnik, Division of Molecular Biology of the Cell II, *dkfz*. Different TIP4N constructs were compared by their binding affinity and specificity to ^{32}P -labelled pRNA. A panel corresponds to every construct. The mutation in the construct is indicated on top of the panel as well as the percentage of binding affinity to pRNA as compared to TIP4N wild type (wt). At first the mutants were tested for their pRNA binding affinity at increasing protein concentrations (left side of each panel). Then they were tested for binding to labeled pRNA at increasing concentrations of unlabelled RNA competitor, either pRNA or RNA from the multiple cloning site (MCS) of pBluescript-KS vector, at protein amounts of 2pmol (right side each panel). On the y axis the average scintillation counts of the corresponding signal are depicted. The counts were normalized to signal given by TIP4N wildtype interaction.

Taking into account only properly folded mutants (see also Table 6), the results revealed, that mutants K541E, W546Y, W546A and R617E showed a significant decrease in RNA binding affinity and pRNA specificity. In case of R617E the effect seen in FBA could had been partially biased by oligomerization of the protein. Hence, the most prominent clearly measured effect was observed for W546A mutant, which showed pRNA binding of 26% as compared to the wildtype TIP4N (100%). S616D mutation did not have a significant difference in RNA and pRNA interaction. Interestingly, R545E mutation did not have an influence on RNA binding affinity of the protein, but lead to decrease in pRNA specific binding to 26% as compared to wildtype TIP4N. Thus, W546 and R545 were identified as key residues for pRNA/TIP4N interaction.

For better illustration the results of mutation analysis are summarized in Table 6.

Mutant	Folding ability ¹	pRNA binding, 100mM KCl	MCS ³ RNA binding, 100mM KCl	MCS ³ RNA binding, 10mM KCl	Binding specificity affected + not affected -
wt TIP4N	+	100%	100%	100%	n/d
Q547A	+	n/d ²	n/d	100%	n/d
Q562A	+	n/d	n/d	90%	n/d
Q605A	+	n/d	n/d	99%	n/d
Q619A	+	n/d	n/d	70%	n/d
K541E	+	43%	8%	47%	+
R545E	+	26%	100%	100%	+
W546A	+	26%	12%	49%	+
W546Y	+	40%	8%	31%	+
S616D	+	100%	84%	65%	-
R617E	+*	12%	16%	51%	+
R598E	-	78% ⁴	60%	99%	-
K540E/K541E	-	5%	12%	80%	+

Table 6 Summary of filter binding assay (FBA) results with TIP4N mutations series I and II. All values are normalized to the binding properties of TIP4N. ¹ Plus indicates protein structure as in wildtype TIP4N as determined by ¹H-¹⁵N HSQC (Fig. 37B and Fig. 41B). Minus stands for protein aggregation. * TIP4N R617E is most probably oligomerizing. ² n/d – not determined. ³ RNA from the multiple cloning site of pBluescript-KS vector. ⁴ High percentage for R598E for pRNA binding is probably caused by protein precipitation during experiment.

9.5. Structural integrity of TAM domain. TIP5WY/GA mutant.

As previously mentioned, Mayer et al. described a murine TIP5 double mutant, where residues W531 and Y532 were mutated to G531 and A532 respectively (Mayer et al, 2006). According to their results TIP5G531A532 had almost an 8-fold decrease in pRNA binding ability and did not induce heterochromatic hallmarks if overexpressed in eukaryotic cells.

Those residues corresponded to W551 and Y552 in human TAM domain construct and were situated far away from detected pRNA interaction region. Y552 was pointing to the hydrophobic core of the domain and its exchange to an other residue could destabilize the protein fold (Fig. 43A). Hence, TIP4NG551A552 mutant was created and studied biochemically and biophysically.

The W551Y552 to G551A552 mutation was introduced by QuickChange[®] PCR. TIP4NG551A552 was recombinantly expressed ¹⁵N labeled in *E.coli* and purified from cytosol using standard protocols.

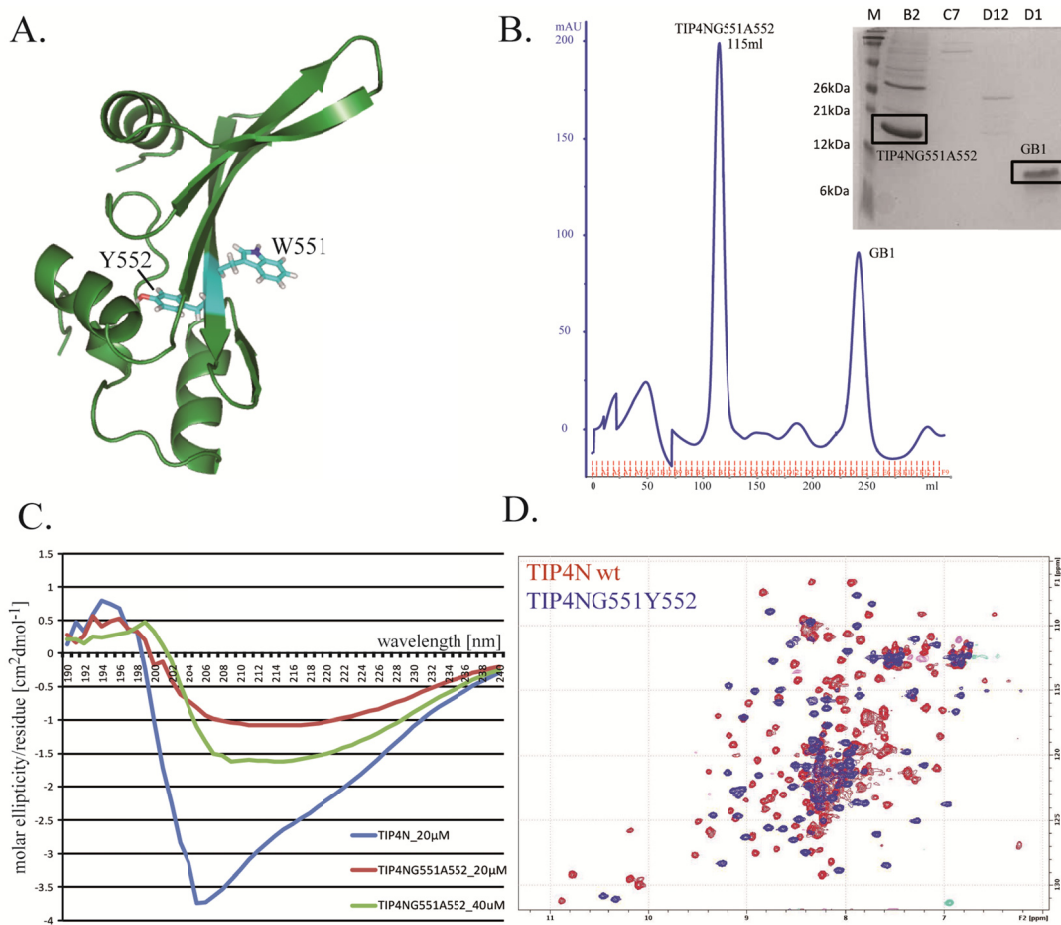


Fig. 43 Biochemical investigation of the TIP4NG551A552 mutant. (A) Cartoon representation of the TAM domain showing position of W551 and Y551 mutated to G551 and A552 in the structure. The mutated residues are highlighted cyan, their side chains are shown as sticks. (B) Elution profile of TIP4NG551A552 from the Superdex S75 HiLoad 26/60 gel filtration column. First major peak in elution profile corresponds to TIP4NG551A552, the second major peak corresponds to GB1-tag protein cleaved from the mutant while purification. In the upper right corner an SDS PA-gel of selected elution fractions is shown. The numbers of the fractions are indicated on top of the lanes and correspond to the numbers in elution profile. (C) Circular dichroism spectra of the wildtype TIP4N and TIP4NG551A552 mutant in the same buffer at different concentrations. The curve for TIP4N wildtype is shown in blue, the one for the mutant at 20 μ M concentration in red and for the mutant at 40 μ M concentration in green. (D) Overlay of ^1H - ^{15}N HSQC spectra of TIP4N wildtype (wt) in red and TIP4NG551A552 in blue at same concentrations and identical experimental conditions.

TIP4NG551A552 showed abnormal behavior already on the Superdex S75 HiLoad 26/60 gel filtration column, eluting at ca.115ml instead of ca.170ml for a folded TIP4N (Fig. 43B), which would suggest structural distortions and formation of aggregates.

Secondary structure of the mutant was investigated using circular dichroism (CD) spectroscopy (Whitmore & Wallace, 2008). CD spectra of TIP4NG551A552 were recorded

at different concentrations (Fig. 43C). At concentration of 20 μ M the spectrum of TIP4NG551A552 was compared to the corresponding spectrum of 20 μ M TIP4N wildtype in the same buffer. Seen from the graphs, TIP4NG551A552 was not unfolded. It had secondary structure content, which must had been, however, different from the content of the wildtype TIP4N. Analysis of CD data by K2D2 software (Perez-Iratxeta & Andrade-Navarro, 2008) revealed, that the α -helical content of the mutant was much lower than that of the wildtype, whereas β -strand content was higher, as compared to wildtype (Table 7).

	TIP4N 20 μ M	TIP4NG551A552 20 μ M
α -helix	20.32%	3.44%
β -strand	26.18%	47.08%
range used	190nm-240nm, 51point	190nm-240nm, 51point

Table 7 α -helical and β -strand content of the wildtype TIP4N and mutant TIP4NG551A552. The values were calculated based on CD spectra using web interface of the program K2D2 (Perez-Iratxeta & Andrade-Navarro, 2008). 51 points from the wavelength range 190nm-240nm were utilized for fitting with standard parameters.

Therefore, structure of the TIP4NG551A552 mutant was further investigated by NMR. A fingerprint $^1\text{H}^{15}\text{N}$ -HSQC was recorded and compared to a $^1\text{H}^{15}\text{N}$ -HSQC of the wildtype TIP4N at the same conditions (buffer, concentration and spectrometer settings). The identity of the TIP4NG551A552 in the NMR sample was confirmed by mass spectrometry with peptide coverage of 51%. As can be obtained from Fig. 43D, TIP4NG551A552 was indeed folded, indicated by good resonance dispersion in the HSQC spectrum. However, the fingerprint spectra of both proteins, though having minor similarities, could not be overlapped. Thus, according to $^1\text{H}^{15}\text{N}$ -HSQC spectrum TIP4NG551A552 has a much altered fold, as compared to the wildtype TIP4N.

Though mass spectrometry indicated that the first 26 amino acids from N-terminus of the TIP4NG551A552 mutant may had been missing, those were flexible (Fig. 17) and not involved in tertiary structure. Distortion of the structure was very likely introduced by destabilization of the protein core through mutations. Thus, structural integrity of the TAM domain played a major role for its pRNA interaction.

9.6. Short conclusion.

Based on NMR titrations and extensive mutational analysis the interaction site of pRNA with TAM domain was identified (Fig. 41A). According to our data, pRNA binding was mediated by residues on tips of the strands β_1 , β_2 , β_4 and the middle part of C-terminal α -helix α_3 . Residues K541, R545, W546, and R617 were central for pRNA interaction.

Mutation of W546 to alanine lead to almost complete loss of TIP4N interaction with pRNA *in vitro*, whereas R545 was important for pRNA recognition. To investigate their impact *in vivo*, W546A and R545E point mutations were introduced into eukaryotic expression vectors (see Appendix) in context of the full length murine TIP5. They will be investigated in cell transfection assays of a murine NIH3T3 or human HEK293T cell line.

Mutations of W551Y552 to G551A552 led possibly to destabilization of the protein and alteration of the TAM domain fold. Together with the published data (Mayer et al, 2006) this indicates that structural integrity of the TAM domain is very important for its pRNA interaction.

10. Discussion.

In line of this project structural basis for the interaction of human TIP5 with pRNA was investigated.

10.1. Structure of the human TAM domain.

The boundaries of the functional TAM domain of human TIP5 were defined and the solution structure of human TAM domain was determined with high precision.

The studies revealed that the functional human TAM domain spans the residues 520-623. These boundaries lie beyond the borders predicted by usual analysis programs, Pfam (Finn et al, 2010) and SMART (Letunic et al, 2009; Schultz et al, 1998). The TAM domain is extended at the C-terminus by ca. 30 residues and the structure includes several amino acids of what was predicted to be the AT-hook 1. TAM is folded only within these boundaries and any C-terminal truncation leads to collapse of the structure, identifying TAM as a very compact domain. Consistent with TAM being an MBD-like domain, its sequence, though very well conserved in different species, is not very similar to canonical human MBDs (methyl-CpG binding domains). The 30 amino acids long C-terminal extension is not present in canonical MBD domains as well, although it is conserved between TAM domains of eukaryotes (Fig. 15).

For structure determination, an NMR data set of good quality could be recorded on the protein construct, comprising only the functional TAM domain (residues 516-624). An expanded network of ^1H - ^1H -NOE interactions that could be determined, and especially a considerable number of long range NOEs resulted in a large number of high quality distance restraints employed in structure calculation (table 1, Res II). Together with TALOS₊ derived torsion angle restraints, that led to good structural convergence while calculation and a definite determination of the TAM domain fold. Two sets of residual dipolar coupling restraints, which were additionally employed in structure calculations, allowed accurate positioning of the secondary structure elements of the TAM domain. Further on, the stereospecific assignments of valine and leucine methyl groups and determination of the histidine protonation state considerably improved the quality of structure calculations.

The resulting ensemble of the 20 lowest energy NMR structures of the TAM domain is determined with high precision, as it was shown by the validation procedures performed with iCING (Vuister & da Silva) (Table 2). Paramagnetic relaxation enhancement data and

hydrogen-deuterium exchange experiments, used for additional structure validation, are consistent with it. The residual dipolar couplings back calculated from the TAM NMR ensemble correlate well with the set of RDCs measured in the Otting alignment medium that was not used in calculations of the TAM domain structure. They show a Q-factor of 0.28, calculated as defined in (Lipsitz & Tjandra, 2004), indicating that the three dimensional NMR structure of the TAM domain is of good quality.

Compared to the known structures of the methyl-DNA binding domains, human TAM domain adopts an extended fold (Fig. 44). Though it shares several features with the MBD structure, it is extended at the N- and C-termini with additional structural elements that are indispensable for the domain fold.

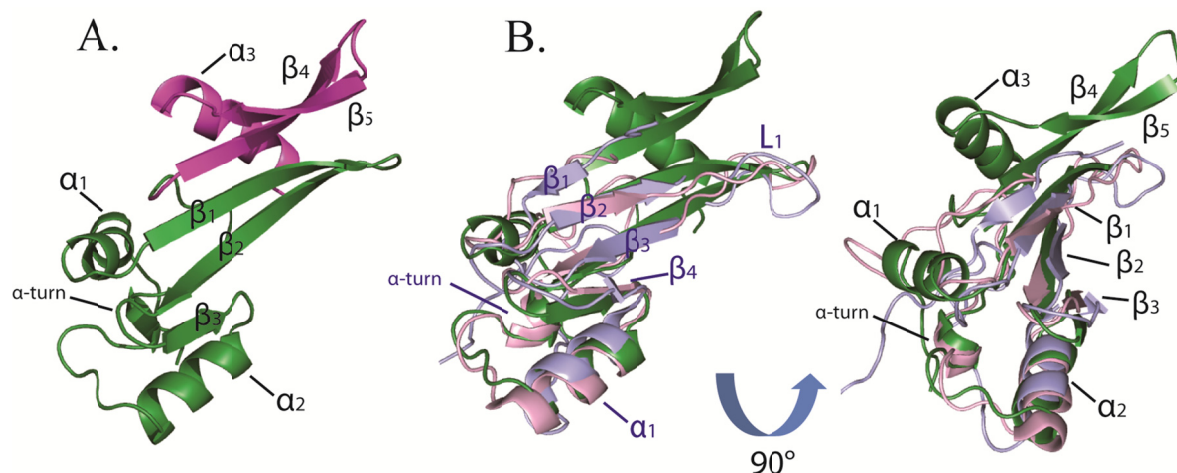


Fig. 44 Analysis of the structure of the human TAM domain. (A) Solution structure of the TAM domain from *H. sapiens* TIP5 protein. The secondary structure elements are indicated. The C-terminal extension, formed by a sequence of 30 amino acids absent in MBD domains but conserved in TAM domains of vertebrates is highlighted magenta. (B) Overlay of the TAM lowest energy structure (forest), and MBD domains from human MBD1 (Ohki et al, 1999) in light blue and mammalian MeCP2 (Wakefield et al, 1999) in light pink. MeCP2 is shown without the unstructured N-terminus. Figure shows the domains from the front (left panel) and after a 90° turn (right panel). On the left panel the secondary structure elements of the MBD domains are designated in blue, whereas on the right panel the secondary structure elements of the TAM domain are designated in black.

The common structural elements include one β -sheet, composed of three antiparallel twisted β -strands β_1 , β_2 and β_3 , as well as the short α -helix α_2 , corresponding to β_2 , β_3 , β_4 and α_1 in the MBD domain structure respectively (Fig. 44B). The short helix α_2 (α_1 in the MBD domains) is known to differ in size in different MBD proteins. In the MBD domain of MeCP2 protein, it is long and has three and a half turns. The MBD domains of MBD1 and MBD2 have a

shorter one, demonstrating two and a half turns. The corresponding α_2 of the TAM domain is common to MBD1 and MBD2 and is a short one. The α_1 of the MBD domains is involved in DNA binding, whether its size plays any significant role for the DNA interaction of the MBDs is not known yet.

The other common elements differ between the MBDs and the TAM domain significantly, even if shared (Fig. 44B). In case of all known MBD structures, the β -strands β_2 and β_3 are short and the amino acid sequence between them forms the flexible loop L_1 that is known to mediate contacts of the MBD domain to the DNA backbone (Ho et al, 2008; Ohki et al, 2001; Scarsdale et al, 2011). In the TAM domain, this sequence forms the tips of the β -strands β_1 and β_2 extending them almost double, and the flexible loop is much shorter being almost just big enough to form the turn. The strand β_3 is tilted in the TAM domain compared to MBD.

Concerning the differences in the TAM domain and MBD domain structure, the variable strand β_1 is not present in the TAM domain and the α -helical turn of TAM is absent from most of the MBD domains, except for the solution structure on MeCP2 MBD domain. However, the most striking differences of the TAM domain structure are its prominent extensions. Instead of forming a wedge-shaped structure, the TAM domain folds into an α/β sandwich. The N-terminus folds into an α -helix α_1 , whereas the C-terminus forms a unique additional β -sheet composed by the antiparallel twisted strands β_4 and β_5 , followed by a terminal α -helix α_3 . The terminal helices pack against the front side of the domain, formed by the β -strands. These structural features are not present in any of the known MBD structures, where the N-terminus mostly adopts an extended conformation and the C-terminal elements are not present at all. Mapped on the sequence of the TAM domain, the C-terminal extension is formed exactly by the 30 amino acids insertion stretch absent from the MBD domains and conserved between TAMs of vertebrates (Fig. 15 and Fig. 44A). Single residue conservation analysis using human TAM structure as bait for automatic homologues search, indicated insufficient data for this region of the domain (Fig. 28), whereas structures shared between TAM and MBD were highly conserved, suggesting that the structural extensions of the TAM domain are indeed unique to it.

10.2. pRNA.

The structure of the pRNA was studied in solution by NMR. A minimal construct (1wt), that bound to human TIP5 with high affinity *in vitro*, was created on the basis of the full length

sequence of murine pRNA published by (Mayer et al, 2008). The construct is composed of the sequences of pRNA, protected from RNase cleavage by interaction with truncated version of human TIP5. The predicted secondary structure of this region, a hairpin, was preserved (Fig. 29).

1wt could be produced in good quality and high amounts using *in vitro* transcription. The base pairing pattern deduced from assignment of the imino proton resonances suggested the correctness of the secondary structure prediction. However, the sequential walk could not be completed due to experimental restrictions (possibly 1wt oligomerization).

Our NMR data on the pRNA/TIP5 complex suggested that on the surface of pRNA the stem of the hairpin would be primarily involved in the interaction. Thus, the shorter pRNA construct (termed “short”) was created, harboring all predicted secondary structural features of the pRNA hairpin stem (Fig. 34). The new construct did not include the three GC base pairs shown to have impact on pRNA/TIP5 interaction due to the fact, that the results of the paper implied that the secondary structure of pRNA, rather than its sequence, was crucial for the TIP5 interaction (Mayer et al, 2008). The new construct was in agreement with this, since the fold of the stem was stabilized by a thermostable GAAA tetraloop. Unfortunately, further NMR investigation of the “short” construct revealed that the problems of possible oligomerization were not overcome. Additionally, the “short” construct caused severe precipitation if mixed with TIP5 truncations. Thus, the investigation of TAM/pRNA interaction was performed with the 1wt and for structural studies of pRNA by NMR, further optimization of experimental conditions is needed.

10.3. pRNA/TAM interaction.

The specific interaction of human TIP5 with pRNA was investigated by NMR, using the minimal pRNA construct 1wt and the TIP4N truncation of TIP5, incorporating the newly defined full length TAM domain, flanked at the N-terminus with extending flexible sequences and at the C-terminus with AT-hooks.

Previous experiments suggested the TIP5/pRNA interaction to be mediated by the TAM domain of TIP5. This assumption was based on *in vitro* loss of general RNA binding function by TIP5 constructs either lacking the short sequence predicted to contain the TAM domain, or the ones harboring a double knockout mutation of aromatic residues in the TAM domain.

For that studies, however, unspecific RNA from multiple cloning site of pBluescript-KS vector was used (Mayer et al, 2006).

Thus, first, the specific interaction of the TAM domain with pRNA was addressed by NMR titration experiments of TIP4N with 1wt pRNA. The major effect from changes in chemical environment was observed in the region of the full length TAM domain (520-623), confirming that the newly defined TAM itself was indeed mediating the specific interaction of TIP5 with pRNA.

Based on the determined TAM domain structure, I suggested that the double knockout mutation in murine TIP5, in which the residues W531Y532 were changed to G531A532, described in (Mayer et al, 2006), would lead to destabilization of the TAM domain structure, since the corresponding Y552 of the human TAM domain of TIP5 points to the hydrophobic core of the domain. Biochemical study by gel filtration and circular dichroism spectroscopy and NMR investigation of the TIP4NG551A552 mutant, where the corresponding residues were mutated to GA, was performed. It showed that the mutated protein had a much altered fold, as compared to the wild type TIP4N and was probably aggregating. That suggests that the loss of pRNA binding ability observed by (Mayer et al, 2006) could be at least partially due to the alterations in the fold of the murine TIP5 TAM domain, underlining the importance of the TAM correct structure for the function of TIP5 protein and of the NoRC. A similar effect was previously observed in an MBD/DNA system as well. Biochemical studies of MeCP2 mutations causing Rett syndrome by circular dichroism spectroscopy showed, that four of the most frequently occurring ones (R106W, R133C, F155S and T158M) lie within the MBD domain of MeCP2. They were shown to induce changes in protein stability and alterations of the MBD fold and DNA binding properties of the protein (Ghosh et al, 2008).

Further on, using NMR and mutational analysis, the area of TAM/pRNA interaction was mapped onto the structure of the TAM domain. TIP4N/1wt system was used in NMR chemical shift perturbation experiments. The results were additionally verified biochemically with full length pRNA. Reconstitution of the TIP4N/1wt complex at different molar ratios of protein and RNA allowed overcoming protein precipitation problems observed during previous experiments, and following and analyzing the interaction in detail.

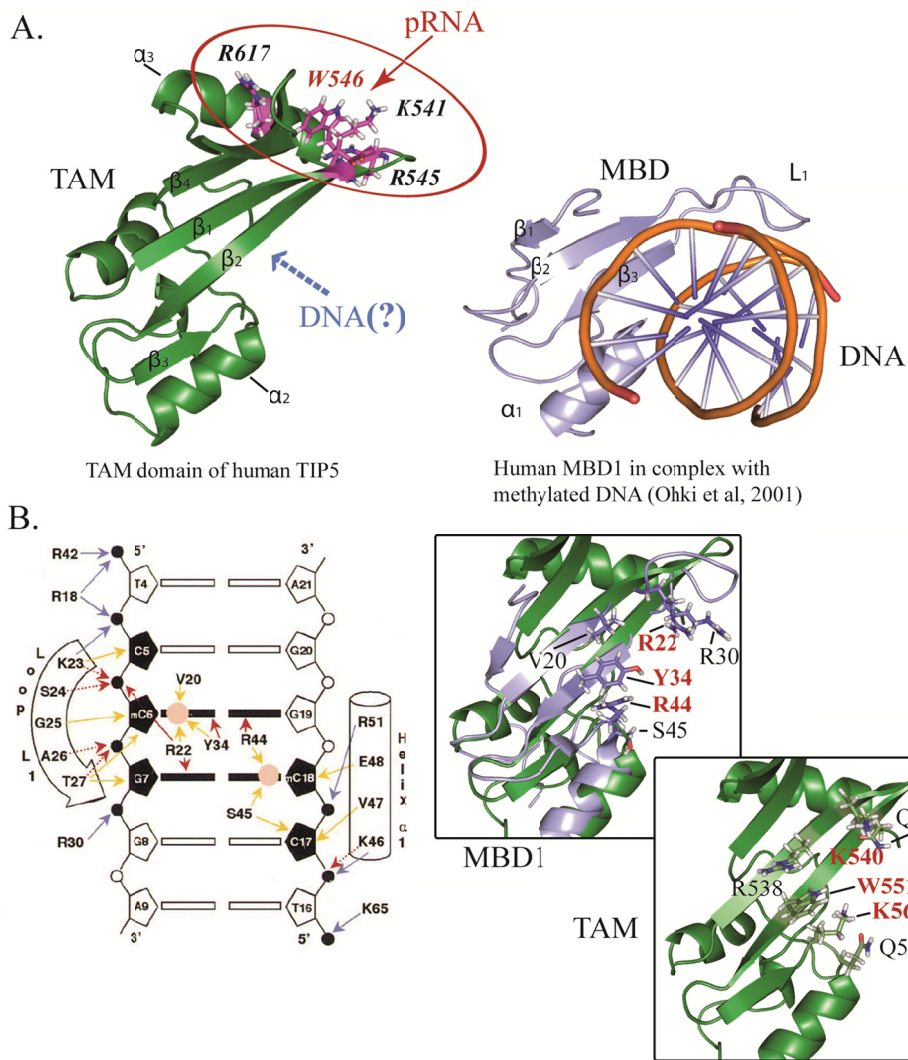


Fig. 45 Nucleic acid interaction of the TAM domain of human TIP5. (A) On the left, human TAM domain is shown in cartoon representation (forest). The pRNA binding surface is circled red. The key interaction residues are shown as sticks and signed. W546, exhibiting the major impact on pRNA binding if mutated, is highlighted red. The potential DNA binding surface is pointed by a dashed arrow. On the right, a cartoon representation of the MBD domain from MBD1 (light blue) in complex with methylated DNA shown in stick model (Ohki et al, 2001) is presented in the same orientation as the TAM domain. Several structural elements are marked, the flexible loop of MBD1 is marked L₁. (B) On the left is the schematic summary of MBD1-DNA contacts (Ohki et al, 2001). DNA bases are indicated as boxes, the deoxyribose sugar rings as pentagons and phosphates as circles. Methyl groups of two symmetrically positioned methylated cytosines are shown as pink circles. Structural elements that interact with MBD are shaded black. Arrows, drawn from residues of MBD to DNA represent hydrophobic interactions (yellow), hydrogen bond contacts (red) and electrostatic contacts (blue). On the right a close-up of an overlay of the MBD from MBD1 and the TAM domain structure (in light blue and forest respectively). Residues of the MBD recognizing the cytosine methyls in the CpG are presented as sticks and signed. The R22, Y34 and R44 forming the crucial contacts are additionally highlighted red. Next to it a close-up of the same area of the TAM domain alone is presented with residues corresponding to MBD1/DNA crucial contacts shown as sticks. Key positions, corresponding to R22, Y34 and R44 of the MBD1 are highlighted red.

The determined surface area of TAM/pRNA interaction is not very big and is formed by the tips of the strands β_1 , β_2 , β_4 and the middle part of C-terminal α -helix α_3 (Fig. 45A). Interestingly, these structural elements are formed by the 30 amino acid sequence, conserved in the eukaryotic TAM domains, but absent in the canonical MBDs (as discussed in 10.1). These elements are not present in the known structures of mammalian MBD domains as well. Though, mammalian proteins containing canonical MBDs are able to form protein/RNA complexes *in vitro*, it was found that they achieve this by virtue of other domains. None of the canonical methyl-CpG binding domains was shown to bind RNA directly so far (Jeffery & Nakielny, 2004). Thus, the interaction with RNA could be a unique feature of the TAM domain and may be present only in eukaryotes, whereas the silencing of the rRNA genes mediated by a direct TAM/pRNA interaction may be specific to eukaryotic cells.

Extensive mutational analysis in combination with biochemical techniques identified residues K541, R545, W546 and R617 as key residues for the TAM/pRNA interaction. Interestingly, mutation of W546 to alanine lead to almost complete loss of TIP4N binding to pRNA and the reversed charge mutation of R545 to glutamate lead to the loss of pRNA specificity, while retaining the general RNA binding function. The impact of W546A and R545E mutations on NoRC mediated epigenetic silencing is being investigated in an *in vivo* system. Consistent with being important for the unique RNA interaction of the TAM domain, W546 and R545 are not conserved in human MBDs. Surprisingly enough, both of these residues are conserved mostly in the BAZ2A variants of TAM containing proteins of higher eukaryotes (Fig. 15), indicating that pRNA interaction may be prerequisite of this class of enzymes.

The N- and C-terminal extensions flanking TAM domain in the TIP4N construct could also have experienced some changes in chemical environment during TIP4N NMR titration with 1wt. That could not be surely elucidated due to high overlap in the spectral regions of their amide resonances; still, one could observe a certain increase of peak intensity coming from possible gain of secondary structure or losses in free tumbling upon pRNA addition. This would mean that these sequences may be directly involved in pRNA interaction. Consistently, the C-terminal extension contains two AT-hook motifs. These motifs are known to act as minor groove binders for DNA (Aravind & Landsman, 1998) and may also contribute to pRNA binding. Additionally, the C-terminal flanking sequence contains the K653 residue as well, which corresponds to K633 in murine TIP5. As discussed in the introduction, the acetylation of this residue by MOF and deacetylation by SIRT-1 was found to regulate TIP5/pRNA interaction and to couple the NoRC mediated response to the intracellular energy

status (Zhou et al, 2009), supporting the idea of possible direct binding of the flexible C-terminal TAM domain flanking sequence to pRNA as the introduction of a negatively charged acetyl group to K653 would lead to RNA dissociation. To further investigate this, NMR relaxation experiments could be employed to examine dynamics properties of the N- and C-terminal extensions flanking TAM domain in context of free TIP4N protein and in complex with pRNA.

From trypsin partial digest tests, Mayer et al. concluded that pRNA binding to full length TIP5 induced changes in conformation of the protein, forcing it into a more open, trypsin-hypersensitive state, suggesting these changes would map to the region encompassing the TAM domain (Mayer et al, 2008). In case these changes would predominantly involve the TAM domain itself, one would expect major differences in the ^1H - ^{15}N HSQC spectra of the free TIP4N and TIP4N in complex with 1wt RNA. However, the changes observed in NMR are mostly peak shifts or intensity increase of flexible N- and C-termini and involve generally the specific region of 1wt interaction. This indicates that changes in tertiary fold upon pRNA binding could involve regions other than TAM domain itself and binding of pRNA could induce distant effects in TIP5 structure or may introduce some additional domain contacts.

10.4. DNA/TAM interaction.

Known structures of MBD domains in complex with DNA show that DNA utilizes loop L_1 and the N-terminal part of the short α -helix α_1 for the DNA backbone contacts. The recognition of methylated CpGs is performed by major groove contacts mediated by residues on the β -sheet (Ho et al, 2008; Ohki et al, 2001; Scarsdale et al, 2011).

When NoRC was first described, its TAM domain was shown to bind DNA *in vitro*. The presence of AT-hook sequences increased its affinity to DNA oligonucleotide, however, the binding was not affected by CpG methylation (Strohner et al, 2001).

pRNA interacts with the surface of the TAM formed by the C-terminal extension of the domain that is not present in the MBDs. The flexible loop L_1 is not present in the TAM domain as well and the corresponding TAM domain sequence is structured instead. Performed mutational experiments suggest that pRNA most probably does not utilize the β -sheets of the TAM. Knockout mutations of several residues on the β -sheet surface of the TAM domain and the N-terminal part of the α -helix α_2 , (corresponding to the α -helix α_1 in

MBDs) do show little or no effect on pRNA binding (affinity decrease to 90% in case of Q562A mutation). However, the β -sheet of the TAM has a very compact structure, as revealed by the hydrogen to deuterium exchange (Fig. 26B), and β_2 , β_3 and the α_2 -helix form a large continuous positively charged surface (Fig. 20C), thus this area could in principle mediate the DNA binding.

On the surface of the MBD domain, the residues interacting with DNA form a continuous hydrophobic patch. The methyl groups are recognized by two arginines and a tyrosine conserved in the canonical human MBDs. Using the example of MBD1 in complex with methylated DNA (Ohki et al, 2001) (Fig. 45B), one methyl group packs into a pocket formed by the hydrophobic side chains of Val20 and Tyr34, and the aliphatic portion of the Arg22 side chain. The methyl group of the other symmetrically situated cytosine packs against aliphatic portions of Arg44 and Ser45 residues. The recognition of the two guanine bases at the methyl CpG site results from interactions with the guanidinium groups of Arg22 and Arg44 and is supported by the aromatic ring of Tyr34. Tyr34 is additionally responsible for recognition of the methylated cytosine, as its hydroxyl group accepts a hydrogen bond from the 4-amino of the cytosine base. The guanidinium groups of Arg22 and Arg44 are crucial for interaction with methylated DNA, since a conservative substitution of either of them with lysine abolishes binding. Mutation of Tyr34 even to a phenylalanine significantly decreases the affinity of MBD1 to methylated DNA (Ohki et al, 2001). Substitution of tyrosine to phenylalanine at position 34, as well as of a supportive positively charged residue at position 30 (arginine or lysine in case of methyl-DNA binding MBDs) to histidine could account for inability of human MBD3 to bind methyl CpG. By contrast MBD3 from *Xenopus laevis* harbors again a Lys30 and a Tyr34 and is known to bind methylated DNA (Hendrich & Tweedie, 2003).

On the surface of the TAM domain, the residues that correspond to the key residues mediating the interaction of an MBD domain with methylated DNA are not conserved. However the substitutions are mostly homologous and they do form a continuous hydrophobic patch on the surface of the TAM domain, enabling it to interact with DNA (Fig. 45B). Still, the key functional groups, accounting for the recognition of the methylated cytosines are not present. So, instead of the two crucial arginines, TAM domain carries two lysines, the mutation shown to abolish binding to methylated DNA *in vitro* (Ohki et al, 2001). Additionally there is a tryptophan (W551), at the position corresponding to the Tyr34 in the MBD1. It could support the DNA interaction through its aromatic ring, however it would not

be able to support the hydrogen bond needed for recognition of the methylated cytosine. Additionally Val20 of the MBD1 is substituted by an arginine, which long side chain could mediate the hydrophobic contacts; still the charged guanidinium group at this position could be disturbing. The same could be applicable to Gln562, corresponding in position to Ser45 of MBD1. It would be interesting to explore this hypothesis by introducing mutations that would restore the key functional groups on the possible surface of TAM/DNA interaction (Table 8).

Mutation		Mutation	
1.	R538V	4.	K561R
2.	K540R	5.	Q562S
3.	W551R	6.	Q547R*

Table 8 Gain of function mutation suggested for the human TAM domain of TIP5 to enforce discrimination of methylated vs. unmethylated CpG DNA. Suggestions are based on structural comparison of TAM and MBD from MBD1. (*) Residue at the supporting position, corresponding to position 30 in the MBD1. Q547 could also be changed to lysine.

In addition, one could investigate the interaction of DNA with the TAM domain by mapping the DNA binding site on the structure of the TAM, using NMR titration experiments, as well as complex reconstitution with a short DNA oligonucleotide.

pRNA interaction area of the TAM domain is distinct from its potential DNA binding site (Fig. 45B). In case DNA binding is indeed mediated by the β -sheet surface, TAM domain harbors a potential to bind both, pRNA and DNA. Hence, TAM domain of TIP5 may serve as universal platform for NoRC recruitment to the rDNA promoter. As the pRNA binding and DNA interaction may be mediated by different parts of the TAM domain, the interaction of the TAM domain with nucleic acids would not necessarily involve a displacement mechanism, allowing for a tempting speculation that TAM could bind DNA and RNA at the same time. This opens an interesting perspective for discussion and for further extensive investigations of nucleic acid interaction of the TAM domain.

10.5. pRNA mediated gene silencing.

In the epigenetic silencing mechanism of rDNA genes and recruitment of NoRC to rDNA, one could imagine two possible roles for the specific TAM domain interaction with pRNA.

pRNA may either have (i) an architectural role as a scaffold for or as an allosteric effector of NoRC or (ii) it could guide chromatin modifications by base pairing with DNA sequences close to the rDNA promoter, whereas the NoRC is recruited via its specific TAM/pRNA interaction to the place of action. Certainly, it can be a mixture of both.

As discussed previously, it was shown that TAM domain interaction with pRNA has an effect *in vitro* on TIP5 compaction, making TIP5 more sensitive to RNase activity upon interaction (Mayer et al, 2008). Based on our data, these rearrangements probably involve regions other than the TAM domain itself. They could lead *in vivo*, for example, to exposure of binding sites on the C-terminal part of TIP5 for the interaction with chromatin modifying enzymes. Not much research has been done on the possible domain rearrangement effect induced by the TAM/pRNA binding yet and, thus, it can not be excluded as a possible mechanism of pRNA action, implying a need for further detailed investigation.

Interestingly, recently some strong support was found for the second possible mechanism of TAM/pRNA mediated gene silencing. Although pRNA sequences fold into a conserved stem-loop structure that is interacting with TIP5 and is required for nucleolar localization of NoRC (Mayer et al, 2008), when expressed ectopically these transcripts did not trigger *de-novo* DNA methylation and transcriptional silencing in cells. In contrast, the full length ectopic pRNA, containing the 5' terminal sequence covering the upstream control element (UCE) and the T₀ terminator and *de novo* DNA methylation, could induce heterochromatic rDNA silencing. It was further revealed that the 5' –terminal region of pRNA forms a DNA:RNA triplex structure with the T₀ DNA sequence *in vitro* and *in vivo* and can alone trigger DNA methylation. Interaction of pRNA with T₀ prevents binding of TTF-I to its recognition sequence (Schmitz et al, 2010).

Based on these results a guiding role of pRNA in the rDNA silencing can be suggested in a model, in which NoRC is bound to pRNA hairpin over its TAM domain (Fig. 46) and then recruited to rDNA by interaction with TTF-I. pRNA base pairs with T₀, leading to displacement of TTF-I from it and recruitment of DNA methyltransferases to the rDNA promoter, whereas NoRC recruits further heterochromatin forming enzymes to the rRNA genes (Grummt, 2010). The pRNA interaction of NoRC would be mediated by the C-terminal extension of the TIP5 TAM domain that is not present in methyl-CpG binding domains. It is, though, to be investigated whether NoRC contacts the DNA directly and whether the TAM domain could mediate also the NoRC/DNA contacts.

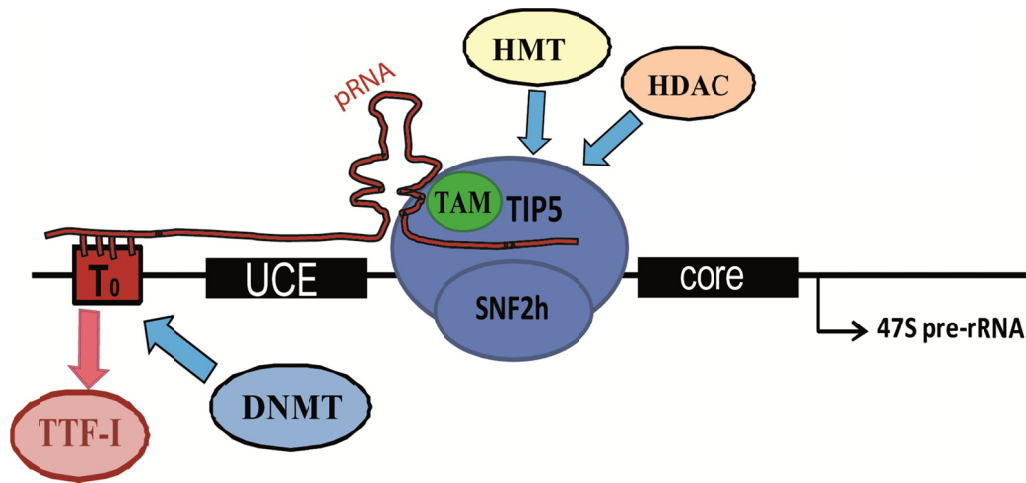


Fig. 46 Model illustrating the role of NoRC and TAM/pRNA interaction in silencing of rRNA genes. Adapted from (Grummt, 2010), extended. The site of transcription initiation of 47S pre-rRNA is indicated as black arrow. The upstream control element (UCE) and the core rDNA promoter are depicted as black boxes. Binding site for the transcription termination factor (TTF-I) upstream of the gene promoter (T_0) is shown as red box. The full length pRNA sequence is depicted as the red line. Both NoRC subunits are schematically represented as ellipses and singed. TAM domain of TIP5 is represented by a green circle. Important chromatin remodeling enzymes recruited by NoRC are indicated as ellipses as well. DNMT – DNA methyltransferases, HMT - histone methyltransferases, HDAC – histone deacetylases.

In plants the mechanism of RNA guided gene silencing has been known for a longtime. It is termed RdDM (RNA dependent DNA methylation) and was first described in the nineties (Wassenegger et al, 1994). Further research revealed that in plants long and short double stranded RNAs are able to trigger transcriptional silencing and *de novo* methylation of a target gene promoter, if they contain promoter sequences. The methylation is precisely delegated to the regions of RNA-DNA sequences identity. The ones found to be most effective, were RNA hairpins produced from inverted DNA repeats acting *in trans* (Luff et al, 1999; Mette et al, 2000).

Whether short dsRNAs can induce transcriptional silencing in mammals in methylation dependent manner has been the focus of several studies. For a review see (Morris, 2009b). The concept of RNA-directed transcriptional silencing in mammals is not completely formed yet. Small RNA-directed transcriptional silencing (TGS) has been observed in human cells. However, the majority of those studies had relied on using exogenous, specially designed synthetic promoter targeted siRNAs for Pol II transcribed genes. From these studies, it has been revealed that in mammals short RNA directs H3K9 and H3K27 histone methylation consistent with concepts of epigenetic silencing of the loci and results in long term stable

modification of gene expression that can be passed to next generation. It is mechanistically distinct from post-translational silencing and utilizes other proteins of the Argonaute family (Morris, 2009a). The impact of DNA methylation in these systems is not well understood yet, as some groups find the DNA methylation at the targeted promoter and some do not observe it. Still the epigenetic changes, if detected, appear specifically at the targeted promoter site. Some work showed, that siRNA-mediated TGS requires promoter associated RNAs (also called pRNAs) to find their target site. They comprise the 5' UTR of mRNAs that overlaps with gene promoter region or were initiated upstream and were overlapping the transcription start site (Han et al, 2007; Napoli et al, 2009).

The data from synthetic RNAs suggests that there may be similar endogenous mechanisms, regulating transcription in mammals. Indeed, there were mechanisms indentified that involve endogenous long noncoding RNA in epigenetic gene silencing. For example the recently described HOTAIR and HOTAIRM1 RNA that seems to epigenetically regulate expression of HOX genes (Rinn et al, 2007; Zhang et al, 2009). Several other endogenous antisense transcripts that involve epigenetics have been described (Cliffe et al, 2009; Mattick et al, 2009; Morris et al, 2008; Yu et al, 2008). However, the exact mechanistic details of their action are not well studied yet.

The proposed above mechanism of epigenetic silencing of rDNA mediated by the TAM/pRNA of NoRC suggests an RNA targeted silencing mechanism in mammals, similar to RdDM in plants. The specific interaction of the TAM domain of TIP5 protein with pRNA plays essential role in the regulation of rDNA transcription and, hence, in regulation of the complex gene expression in eukaryotes. It could be a central event leading to establishment of heterochromatic compact state of rRNA genes and to reduction of ribosomal synthesis in cells in response to various stimuli.

Taking into account the potential ability of the TIP5 TAM domain to interact simultaneously both with pRNA and DNA, further research could reveal the mechanism of binding site recognition mediated by the TAM domain. Advanced biochemical and structural research on NoRC and in particular on the TAM domain/pRNA and TAM domain/DNA system will promote our understanding of gene silencing in eukaryotes, as well as gene regulation mechanisms utilized by organisms of different kingdoms to ensure their survival and genome stability.

References.

Optimization of buffer conditions for TIP4N construct.

Ahringer J (2000) NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* **16**: 351-356

Akke M, Palmer AG (1996) Monitoring Macromolecular Motions on Microsecond to Millisecond Time Scales by R1ρ-R1 Constant Relaxation Time NMR Spectroscopy. *Journal of the American Chemical Society* **118**: 911-912

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997a) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997b) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402

Amrus A, Yang D (2007) Diffusion-ordered nuclear magnetic resonance spectroscopy for analysis of DNA secondary structural elements. *Analytical Biochemistry* **367**: 56-67

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**: 185-188

Aravind L, Landsman D (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* **26**: 4413-4421

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**: 10037-10041

Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**: 5355-5365

Bax A, Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr Opin Struct Biol* **15**: 563-570

Bell SP, Learned RM, Jantzen HM, Tjian R (1988) Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science* **241**: 1192-1197

Bloch F (1946) Nuclear Induction. *Physical Review* **70**: 460

Bloch F, Hansen WW, Packard M (1946a) Nuclear Induction. *Physical Review* **69**: 127

Bloch F, Hansen WW, Packard M (1946b) The Nuclear Induction Experiment. *Physical Review* **70**: 474

BMRB. BioMagResBank, Biological Magnetic Resonance Bank. In <http://www.bmrb.wisc.edu/> (ed.).

- Bochar DA, Savard J, Wang W, Lafleur DW, Moore P, Cote J, Shiekhattar R (2000) A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc Natl Acad Sci U S A* **97**: 1038-1043
- Bozhenok L, Wade PA, Varga-Weisz P (2002) WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J* **21**: 2231-2241
- Bradsher J, Auriol J, Proietti de Santis L, Iben S, Vonesch JL, Grummt I, Egly JM (2002) CSB is a component of RNA pol I transcription. *Mol Cell* **10**: 819-829
- Brown SE, Szyf M (2008) Dynamic epigenetic states of ribosomal RNA promoters during the cell cycle. *Cell Cycle* **7**: 382-390
- Bruschweiler R, Liao X, Wright P (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**: 886-889
- Buhler M (2009) RNA turnover and chromatin-dependent gene silencing. *Chromosoma* **118**: 141-151
- Caburet S, Conti C, Schurra C, Lebofsky R, Edelstein SJ, Bensimon A (2005) Human ribosomal RNA gene arrays display a broad range of palindromic structures. *Genome Res* **15**: 1079-1085
- Cavanagh J. FWJ, Palmer A. G. III, Skelton N. (1996) *Protein NMR Spectroscopy*: Academic Press.
- Christodoulou J, Larsson G, Fucini P, Connell SR, Pertinhez TA, Hanson CL, Redfield C, Nierhaus KH, Robinson CV, Schleucher J, Dobson CM (2004) Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes. *Proc Natl Acad Sci U S A* **101**: 10949-10954
- Cliffe AR, Garber DA, Knipe DM (2009) Transcription of the herpes simplex virus latency-associated transcript promotes the formation of facultative heterochromatin on lytic promoters. *J Virol* **83**: 8182-8190
- Clore GM, Gronenborn AM, Bax A (1998a) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson* **133**: 216-221
- Clore GM, Gronenborn AM, Tjandra N (1998b) Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson* **131**: 159-162
- Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society* **120**: 6836-6837
- Craik DJ (1996) NMR in Drug Design.
- Daragan VA, Mayo KH (1997) Motional Model Analyses of Protein and Peptide Dynamics Using ¹³C and ¹⁵N NMR Relaxation. *Progress in Nuclear Magnetic Resonance Spectroscopy* **31**: 63-105
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**: 277-293

- Dhasarathy A, Wade PA (2008) The MBD protein family-reading an epigenetic mark? *Mutat Res* **647**: 39-43
- Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* **35**: W522-525
- Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**: W665-667
- Dosset P, Hus J-C, Marion D, Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *Journal of Biomolecular NMR* **20**: 223-231-231
- Downing AK (ed) (2004) *Protein NMR Techniques*: Humana Press
- Drygin D, Rice WG, Grummt I (2010) The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu Rev Pharmacol Toxicol* **50**: 131-156
- Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, Phair RD, Misteli T (2002) A kinetic framework for a mammalian RNA polymerase in vivo. *Science* **298**: 1623-1626
- Earley K, Lawrence RJ, Pontes O, Reuther R, Enciso AJ, Silva M, Neves N, Gross M, Viegas W, Pikaard CS (2006) Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev* **20**: 1283-1293
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797
- Ernst RR, Bodenhausen G, Wokaun A (1989) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*: Oxford Science Publications.
- Espada J, Ballestar E, Santoro R, Fraga MF, Villar-Garea A, Nemeth A, Lopez-Serra L, Ropero S, Aranda A, Orozco H, Moreno V, Juarranz A, Stockert JC, Langst G, Grummt I, Bickmore W, Esteller M (2007) Epigenetic disruption of ribosomal RNA genes and nucleolar architecture in DNA methyltransferase 1 (Dnmt1) deficient cells. *Nucleic Acids Res* **35**: 2191-2198
- Fernandez C, Wider G (2003) TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* **13**: 570-580
- Fielding L (2007) NMR methods for the determination of protein–ligand dissociation constants. **51**: 219-242
- Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A (2010) The Pfam protein families database. *Nucleic Acids Res* **38**: D211-222
- Fischer MW, Losonczi JA, Weaver JL, Prestegard JH (1999) Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* **38**: 9013-9022

Furtig B, Richter C, Wohnert J, Schwalbe H (2003) NMR spectroscopy of RNA. *ChemBiochem* **4**: 936-962

Gardner KH, Kay LE (1998) The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **27**: 357-406

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker JM (ed), pp 571-607. Humana Press

Ghosh RP, Horowitz-Scherer RA, Nikitina T, Gierasch LM, Woodcock CL (2008) Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. *J Biol Chem* **283**: 20523-20534

Ghoshal K, Majumder S, Datta J, Motiwala T, Bai S, Sharma SM, Frankel W, Jacob ST (2004) Role of human ribosomal RNA (rRNA) promoter methylation and of methyl-CpG-binding protein MBD2 in the suppression of rRNA gene expression. *J Biol Chem* **279**: 6783-6793

Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol* **268**: 158-169

Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **19**: 163-164

Goddard TD, Kneller DG. SPARKY 3. University of California, San Francisco

Golovanov AP, Hautbergue GM, Wilson SA, Lian L-Y (2004) A Simple Method for Improving Protein Solubility and Long-Term Stability. *Journal of the American Chemical Society* **126**: 8933-8939

Gonzalez IL, Sylvester JE (1995) Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer. *Genomics* **27**: 320-328

Gorski JJ, Pathak S, Panov K, Kasciukovic T, Panova T, Russell J, Zomerdijk JC (2007) A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. *EMBO J* **26**: 1560-1568

Grozdanov P, Georgiev O, Karagyozev L (2003) Complete sequence of the 45-kb mouse ribosomal DNA repeat: analysis of the intergenic spacer. *Genomics* **82**: 637-643

Grummt I (2007) Different epigenetic layers engage in complex crosstalk to define the epigenetic state of mammalian rRNA genes. *Hum Mol Genet* **16 Spec No 1**: R21-27

Grummt I (2010) Wisely chosen paths--regulation of rRNA synthesis: delivered on 30 June 2010 at the 35th FEBS Congress in Gothenburg, Sweden. *FEBS J* **277**: 4626-4639

Grummt I, Kuhn A, Bartsch I, Rosenbauer H (1986a) A transcription terminator located upstream of the mouse rDNA initiation site affects rRNA synthesis. *Cell* **47**: 901-911

Grummt I, Rosenbauer H, Niedermeyer I, Maier U, Ohrlein A (1986b) A repeated 18 bp sequence motif in the mouse rDNA spacer mediates binding of a nuclear factor and transcription termination. *Cell* **45**: 837-846

Guétg C, Lienemann P, Sirri V, Grummt I, Hernandez-Verdun D, Hottiger MO, Fussenegger M, Santoro R (2010) The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *EMBO J* **29**: 2135-2146

Guntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**: 353-378

Han J, Kim D, Morris KV (2007) Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* **104**: 12422-12427

Hannan KM, Brandenburger Y, Jenkins A, Sharkey K, Cavanaugh A, Rothblum L, Moss T, Poortinga G, McArthur GA, Pearson RB, Hannan RD (2003) mTOR-dependent regulation of ribosomal gene transcription requires S6K1 and is mediated by phosphorylation of the carboxy-terminal activation domain of the nucleolar transcription factor UBF. *Mol Cell Biol* **23**: 8862-8877

He L, Niemeyer B (2003) A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration. *Biotechnol Prog* **19**: 544-548

Heliot L, Kaplan H, Lucas L, Klein C, Beorchia A, Doco-Fenzy M, Menager M, Thiry M, O'Donohue MF, Ploton D (1997) Electron tomography of metaphase nucleolar organizer regions: evidence for a twisted-loop organization. *Mol Biol Cell* **8**: 2199-2216

Hempel WM, Cavanaugh AH, Hannan RD, Taylor L, Rothblum LI (1996) The species-specific RNA polymerase I transcription factor SL-1 binds to upstream binding factor. *Mol Cell Biol* **16**: 557-563

Henderson AS, Warburton D, Atwood KC (1972) Location of ribosomal DNA in the human chromosome complement. *Proc Natl Acad Sci U S A* **69**: 3394-3398

Henderson S, Sollner-Webb B (1986) A transcriptional terminator is a novel element of the promoter of the mouse ribosomal RNA gene. *Cell* **47**: 891-900

Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**: 6538-6547

Hendrich B, Tweedie S (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet* **19**: 269-277

Heus HA, Pardi A (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**: 191-194

Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell* **29**: 525-531

Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Research* **31**: 3429-3431

Hore PJ, Jones JA, Wimperis S (2001) *NMR: The Toolkit*: Oxford University Press.

Invitrogen. <http://products.invitrogen.com/ivgn/product/V102020>.

Ishii Y, Markus MA, Tycko R (2001) Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. *Journal of Biomolecular NMR* **21**: 141-151-151

Ito T, Levenstein ME, Fyodorov DV, Kutach AK, Kobayashi R, Kadonaga JT (1999) ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev* **13**: 1529-1539

Jeffery L, Nakielny S (2004) Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J Biol Chem* **279**: 49479-49487

Johnson CS, Jr. (1999) Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**: 203-256

Jung YS, Zweckstetter M (2004) Mars -- robust automatic backbone assignment of proteins. *J Biomol NMR* **30**: 11-23

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577-2637

Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**: 8972-8979

Keeler J (2005) *Understanding NMR Spectroscopy*: Wiley.

Kobayashi T, Horiuchi T, Tongaonkar P, Vu L, Nomura M (2004) SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast. *Cell* **117**: 441-453

Kramer F, Deshmukh MV, Kessler H, Glaser SJ (2004) Residual dipolar coupling constants: An elementary derivation of key equations. *Concepts in Magnetic Resonance Part A* **21A**: 10-21

Kroenke CD, Loria JP, Lee LK, Rance M, Palmer AG (1998) Longitudinal and Transverse ¹H-¹⁵N Dipolar/¹⁵N Chemical Shift Anisotropy Relaxation Interference: Unambiguous Determination of Rotational Diffusion Tensors and Chemical Exchange Effects in Biological Macromolecules. *Journal of the American Chemical Society* **120**: 7905-7915

Kuhn A, Grummt I (1987) A novel promoter in the mouse rDNA spacer is active in vivo and in vitro. *EMBO J* **6**: 3487-3492

Kuhn A, Grummt I (1992) Dual role of the nucleolar transcription factor UBF: trans-activator and antirepressor. *Proc Natl Acad Sci U S A* **89**: 7340-7344

Kuhn A, Stefanovsky V, Grummt I (1993) The nucleolar transcription activator UBF relieves Ku antigen-mediated repression of mouse ribosomal gene transcription. *Nucleic Acids Res* **21**: 2057-2063

Kuhn A, Vente A, Doree M, Grummt I (1998) Mitotic phosphorylation of the TBP-containing factor SL1 represses ribosomal gene transcription. *J Mol Biol* **284**: 1-5

- Kummerloewe G, Halbach F, Laufer B, Luy B (2008) Precise measurement of RDCs in water and DMSO based gels using a silicone rubber tube for tunable stretching. *Open Spectrosc J* **2**: 29-33
- Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**: W299-302
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) {PROCHECK}: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**: 283-291
- Lawrence RJ, Earley K, Pontes O, Silva M, Chen ZJ, Neves N, Viegas W, Pikaard CS (2004) A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol Cell* **13**: 599-609
- Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res* **37**: D229-232
- Li J, Langst G, Grummt I (2006) NoRC-dependent nucleosome positioning silences rRNA genes. *EMBO J* **25**: 5735-5741
- Li J, Santoro R, Koberna K, Grummt I (2005) The chromatin remodeling complex NoRC controls replication timing of rRNA genes. *EMBO J* **24**: 120-127
- Linge JP, Habeck M, Rieping W, Nilges M (2003a) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**: 315-316
- Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M (2003b) Refinement of protein structures in explicit solvent. *Proteins* **50**: 496-506
- Lipsitz RS, Tjandra N (2004) Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct* **33**: 387-413
- Lorieau J, Yao L, Bax A (2008) Liquid crystalline phase of G-tetrad DNA for NMR study of detergent-solubilized proteins. *J Am Chem Soc* **130**: 7536-7537
- Luff B, Pawlowski L, Bender J (1999) An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis. *Mol Cell* **3**: 505-511
- Lukavsky PJ, Puglisi JD (2004) Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA* **10**: 889-893
- Lundberg KS, Shoemaker DD, Adams MWW, Short JM, Sorge JA, Mathur EJ (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108**: 1-6
- Mabuchi H, Fujii H, Calin G, Alder H, Negrini M, Rassenti L, Kipps TJ, Bullrich F, Croce CM (2001) Cloning and characterization of CLLD6, CLLD7, and CLLD8, novel candidate genes for leukemogenesis

at chromosome 13q14, a region commonly deleted in B-cell chronic lymphocytic leukemia. *Cancer Res* **61**: 2870-2877

Mailand N, Lukas C, Kaiser BK, Jackson PK, Bartek J, Lukas J (2002) Deregulated human Cdc14A phosphatase disrupts centrosome separation and chromosome segregation. *Nat Cell Biol* **4**: 317-322

Majumder S, Ghoshal K, Datta J, Smith DS, Bai S, Jacob ST (2006) Role of DNA methyltransferases in regulation of human ribosomal RNA gene transcription. *J Biol Chem* **281**: 22062-22072

Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF (2009) RNA regulation of epigenetic processes. *Bioessays* **31**: 51-59

Mayer C, Neubert M, Grummt I (2008) The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Rep*

Mayer C, Schmitz KM, Li J, Grummt I, Santoro R (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell* **22**: 351-361

Mayrose I, Graur D, Ben-Tal N, Pupko T (2004) Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. *Molecular Biology and Evolution* **21**: 1781-1791

McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu Rev Cell Dev Biol* **24**: 131-157

Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJ (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *Embo J* **19**: 5194-5201

Miller OL, Jr., Beatty BR (1969) Visualization of nucleolar genes. *Science* **164**: 955-957

Montanaro L, Trere D, Derenzini M (2008) Nucleolus, ribosomes, and cancer. *Am J Pathol* **173**: 301-310

Morris KV (2009a) Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells. *Epigenetics* **4**: 296-301

Morris KV (2009b) RNA-directed transcriptional gene silencing and activation in human cells. *Oligonucleotides* **19**: 299-306

Morris KV, Santoso S, Turner AM, Pastori C, Hawkins PG (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet* **4**: e1000258

Moss T (2004) At the crossroads of growth control; making ribosomal RNA. *Curr Opin Genet Dev* **14**: 210-217

Moss T, Boseley PG, Birnstiel ML (1980) More ribosomal spacer sequences from *Xenopus laevis*. *Nucleic Acids Res* **8**: 467-485

Moss T, Langlois F, Gagnon-Kugler T, Stefanovsky V (2007) A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell Mol Life Sci* **64**: 29-49

- Murayama A, Ohmori K, Fujimura A, Minami H, Yasuzawa-Tanaka K, Kuroda T, Oie S, Daitoku H, Okuwaki M, Nagata K, Fukamizu A, Kimura K, Shimizu T, Yanagisawa J (2008) Epigenetic control of rDNA loci in response to intracellular energy status. *Cell* **133**: 627-639
- Nan X, Meehan RR, Bird A (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* **21**: 4886-4892
- Napoli S, Pastori C, Magistri M, Carbone GM, Catapano CV (2009) Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J* **28**: 1708-1719
- Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M, Shirakawa M (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105**: 487-497
- Ohki I, Shimotake N, Fujita N, Nakao M, Shirakawa M (1999) Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J* **18**: 6653-6661
- Ottiger M, Bax A (1999) Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. *Journal of Biomolecular NMR* **13**: 187-191-191
- Panov KI, Friedrich JK, Russell J, Zomerdijk JC (2006) UBF activates RNA polymerase I transcription by stimulating promoter escape. *EMBO J* **25**: 3310-3322
- Pastore A, Saudek V (1990a) The Relationship between Chemical-Shift and Secondary Structure in Proteins. *Journal of Magnetic Resonance* **90**: 165-176
- Pastore A, Saudek V (1990b) The relationship between chemical shift and secondary structure in proteins. *Journal of Magnetic Resonance* **90**: 165-176
- Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci* **2**: 543-558
- Peng JC, Karpen GH (2007) H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol* **9**: 25-35
- Perez-Iratxeta C, Andrade-Navarro MA (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct Biol* **8**: 25
- Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**: 12366-12371
- Plaxco KW, Morton CJ, Grimshaw SB, Jones JA, M. P, Campbell ID, Dobson CM (1997) The effects of guanidine hydrochloride on the "random coil" conformations and NMR chemical shifts of the peptide series GGXGG. *Journal of Biomolecular NMR* **V10**: 221-230
- Poot RA, Dellaire G, Hulsmann BB, Grimaldi MA, Corona DF, Becker PB, Bickmore WA, Varga-Weisz PD (2000) HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. *EMBO J* **19**: 3377-3387

- Proctor WG, Yu FC (1951) On the Nuclear Magnetic Moments of Several Stable Isotopes. *Physical Review* **81**: 20
- Purcell EM, Torrey HC, Pound RV (1946) Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review* **69**: 37
- Ramsey NF, Purcell EM (1952) Interactions between Nuclear Spins in Molecules. *Physical Review* **85**: 143
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**: 1311-1323
- Rovnyak D, Frueh DP, Sastry M, Sun ZY, Stern AS, Hoch JC, Wagner G (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J Magn Reson* **170**: 15-21
- Rückert M, Otting G (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *Journal of the American Chemical Society* **122**: 7793-7797
- Russell J, Zomerdijk JC (2006) The RNA polymerase I transcription machinery. *Biochem Soc Symp*: 203-216
- Sakai K, Ohta T, Minoshima S, Kudoh J, Wang Y, de Jong PJ, Shimizu N (1995) Human ribosomal RNA gene cluster: identification of the proximal end containing a novel tandem repeat sequence. *Genomics* **26**: 521-526
- Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A* **95**: 13585-13590
- Santoro R, Grummt I (2001) Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Mol Cell* **8**: 719-725
- Santoro R, Li J, Grummt I (2002) The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat Genet* **32**: 393-396
- Santoro R, Schmitz KM, Sandoval J, Grummt I (2010) Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO Rep* **11**: 52-58
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**: 93-158
- Scarsdale JN, Webb HD, Ginder GD, Williams DC, Jr. (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res* **39**: 6741-6752

Schanda P, Brutscher B (2005) Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *Journal of the American Chemical Society* **127**: 8014-8015

Schmitz KM, Mayer C, Postepska A, Grummt I (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* **24**: 2264-2269

Schrödinger. The PyMOL Molecular Graphics System. In LLC (ed.).

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ, 3rd (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16**: 919-932

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**: 5857-5864

Senn H, Werner B, Messerle BA, Weber C, Traber R, Wuethrich K (1989) Stereospecific assignment of the methyl ¹H NMR lines of valine and leucine in polypeptides by nonrandom ¹³C labelling *FEBS LETTERS* **249**: 113-118

Shan X, Gardner KH, Muhandiram DR, Rao NS, Arrowsmith CH, Kay LE (1996) Assignment of ¹⁵N, ¹³C_α, ¹³C_β, and HN Resonances in an ¹⁵N, ¹³C, ²H Labeled 64 kDa Trp Repressor-Operator Complex Using Triple-Resonance NMR Spectroscopy and ²H-Decoupling. *J Am Chem Soc* **118**: 6570-6579

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**: 213-223

Solomon I (1955) Relaxation Processes in a System of Two Spins. *Physical Review* **99**: 559

Solomon I, Bloembergen N (1956) Nuclear Magnetic Interactions in the HF Molecule *J Chem Phys* **25**: 261-266

Spera S, Bax A (1991a) Empirical Correlation between Protein Backbone Conformation and C-Alpha and C-Beta C-13 Nuclear-Magnetic-Resonance Chemical-Shifts. *Journal of the American Chemical Society* **113**: 5490-5492

Spera S, Bax A (1991b) Empirical correlation between protein backbone conformation and C.alpha. and C.beta. ¹³C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* **113**: 5490-5492

Stefanovsky V, Langlois F, Gagnon-Kugler T, Rothblum LI, Moss T (2006) Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling. *Mol Cell* **21**: 629-639

Stefanovsky VY, Pelletier G, Hannan R, Gagnon-Kugler T, Rothblum LI, Moss T (2001) An immediate response of ribosomal transcription to growth factor stimulation in mammals is mediated by ERK phosphorylation of UBF. *Mol Cell* **8**: 1063-1073

Strohner R, Nemeth A, Jansa P, Hofmann-Rohrer U, Santoro R, Langst G, Grummt I (2001) NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines. *EMBO J* **20**: 4892-4900

- Strohner R, Nemeth A, Nightingale KP, Grummt I, Becker PB, Langst G (2004) Recruitment of the nucleolar remodeling complex NoRC establishes ribosomal DNA silencing in chromatin. *Mol Cell Biol* **24**: 1791-1798
- Thanabal V, Omecinsky DO, Reily MD, Cody WL (1994) The ¹³C chemical shifts of amino acids in aqueous solution containing organic solvents: Application to the secondary structure characterization of peptides in aqueous trifluoroethanol solution. *Journal of Biomolecular NMR* **4**: 47-59
- Tuan JC, Zhai W, Comai L (1999) Recruitment of TATA-binding protein-TAFI complex SL1 to the human ribosomal DNA promoter is mediated by the carboxy-terminal activation domain of upstream binding factor (UBF) and is regulated by UBF phosphorylation. *Mol Cell Biol* **19**: 2872-2879
- Varani G, Aboul-ela F, Allain FHT (1996) NMR investigation of RNA structure. *Progress in Nuclear Magnetic Resonance Spectroscopy* **29**: 51-127
- Varga-Weisz PD, Wilm M, Bonte E, Dumas K, Mann M, Becker PB (1997) Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* **388**: 598-602
- Vriend G, Sander C (1993) Quality control of protein models: directional atomic contact analysis. *Journal of Applied Crystallography* **26**: 47-60
- Vuister G, da Silva AWS. <http://nmr.cmbi.ru.nl/icing/>.
- Wade PA, Geggion A, Jones PL, Ballestar E, Aubry F, Wolffe AP (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet* **23**: 62-66
- Wakefield RI, Smith BO, Nan X, Free A, Soteriou A, Uhrin D, Bird AP, Barlow PN (1999) The solution structure of the domain from MeCP2 that binds to methylated DNA. *J Mol Biol* **291**: 1055-1065
- Wassenegger M, Heimes S, Riedel L, Sanger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**: 567-576
- White RJ (2008) RNA polymerases I and III, non-coding RNAs and cancer. *Trends Genet* **24**: 622-629
- Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**: 392-400
- Williamson D, Lu YJ, Fang C, Pritchard-Jones K, Shipley J (2006) Nascent pre-rRNA overexpression correlates with an adverse prognosis in alveolar rhabdomyosarcoma. *Genes Chromosomes Cancer* **45**: 839-845
- Williamson MP, Havel TF, Wuthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ¹H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**: 295-315
- Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) ¹H, ¹³C and ¹⁵N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR* **5**: 67-81

- Wishart DS, Sykes BD (1994) The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. *Journal of Biomolecular Nmr* **4**: 171-180
- Wishart DS, Sykes BD, Richards FM (1992) The Chemical-Shift Index - a Fast and Simple Method for the Assignment of Protein Secondary Structure through Nmr-Spectroscopy. *Biochemistry* **31**: 1647-1651
- Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*, New York, NY: Wiley.
- Wyatt JR, Chastain M, Puglisi JD (1991) Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* **11**: 764-769
- Yamazaki T, Forman-Kay JD, Kay LE (1993) Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *Journal of the American Chemical Society* **115**: 11054-11055
- Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**: 202-206
- Yuan X, Feng W, Imhof A, Grummt I, Zhou Y (2007) Activation of RNA polymerase I transcription by cockayne syndrome group B protein and histone methyltransferase G9a. *Mol Cell* **27**: 585-595
- Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, Gingeras TR, Kapranov P, Weissman SM, Newburger PE (2009) A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* **113**: 2526-2534
- Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**: 1924-1935
- Zhou Y, Grummt I (2005) The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. *Curr Biol* **15**: 1434-1438
- Zhou Y, Santoro R, Grummt I (2002) The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *EMBO J* **21**: 4632-4640
- Zhou Y, Schmitz KM, Mayer C, Yuan X, Akhtar A, Grummt I (2009) Reversible acetylation of the chromatin remodeling complex NoRC is required for non-coding RNA-dependent silencing. *Nat Cell Biol* **11**: 1010-1016
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**: 3406-3415
- Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* **3**: 679-690
- Zweckstetter M, Bax A (2001) Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J Biomol NMR* **20**: 365-377

Appendix.

11.1 DNA sequences.

M. musculus TIP5. Accession number NCBI AJ309544.

```
1 atgagcttgc cccagcaagg gaaaagtttg aatggggatg tgaatgtaa tggcttatct
61 actgtatctc acactactac ttcagggatt ttgaactctg ctccccactc ctctagcacc
121 tcacacctcc atcacccctaa cgtggcctac gactgtcttt ggaactactc acagtaccca
181 tctgccaatc ctggcaacaa cctcaaggac ccaccccttc tttctcagtt cctgagggga
241 caatacccg c tcaacgggat ccttgggggc aaccgacaac ctcatcccc aagtcacaac
301 actaatcttc gagctgggag ccaagagttc tggggcaatg gtaccagagag tcccatgggg
361 cttactctcg attcacagga actgtatgat tcttttctg atcagaattt tgaggtgatg
421 cccaatggac cccaagttt tttcacctcc cctcagactt ctccaatgtt ggggtctagt
481 atccagacct ttgcaccttc ccaggatgta agcagtgaca tccatcctga tgaagcagca
541 gaaaaggagc tgacttcagt tgtggcagaa aatggcactg gcttggtagg cagcctggag
601 ctggaggaag agcagccaga actaaagatg tgtggctaca atggttctgt ctctctgtg
661 gagtctttac accaagaagt ctccgtcctg gtccctgatc ccacagtgag ctgtctagat
721 gatccttcac atcttctga tcaactggaa gacactcaa ttctcagtga agactccctg
781 gagccctttg actctctggc agcagagcca gtgagtggca gtctttatgg tatagatgat
841 ggggagctga tgggtgcaga agacaagttg cctctggagg gcaaccctgt gatctctgcc
901 ctcgattgcc ctgctctcag taatgctaag gccttcagtc tcttggcaga cgacagccag
961 acatcagcct ccactcttgt cagccctacc tccccactg tcttagggga gtctgtcttg
1021 caagataaca gctttggact gaacagttgc agtgactctg aacaggaaga aatagagacc
1081 cagtcttcaa acttccaacg tcccctgact gagccagctc ctgaccagcc acctagtact
1141 caactacatc cagcagtttc accaacagcc tccccagcag cctccttgac agcatctgca
1201 gaaatctctc cagctgtctc tccagtagca tctctgcctg tccctctga agtctttgta
1261 gcatctctc cagcttctc actgtctctg ccagccatct ctttgggaag ctctatgaca
1321 actccagtaa cttctctca aggttcccc cagcagctgc cttccagact
1381 gtctccccag caaggaaaaa tgtcagcagt gctcctaaag cgcgtgctga tgcgagaagag
1441 acgactggag gagcagttgc agtctctggg agtgggtgat tactgaagag acgtattgct
1501 accccagaag aagttcgtct tcccctccag catgggtggc gaagagaagt gcgcatcaag
1561 aagggcagcc atcggtgcca gggggagact tgggtactat gcccctgtgg gaagagaatg
1621 aagcaatttc cagaagttat caagtacctg agccgaaatg tgggtgcacag tgtccgccgt
1681 gagcacttca gcttcagtcc cgcgatgcct gttggagatt tctttgaaga aagagataca
1741 ccagagggct tgcagtgggt ccagttatca gcagaggaga ttccttccag aattcaagca
1801 atcactggca aacgagggcc acctcgaac aatgagaagg ctaagaacaa ggaagtctcc
1861 aaagtgaagc ggggcccagg tggcctcct aagatcaaaa tgcctgagct gttgataaaa
1921 acagataacc gacttccaaa gaaactggaa acccaagaaa tactgagtga ggacgataaa
1981 gcaaagatga ctaaaaacaa aaagaagatg aggcagaagg tccaacgggg agaaagtcat
2041 actcctgtcc aagggcaggc cagaaacaag aggaagcaag acaccaagag cttgaagcag
2101 aaggacacta agaagaaatt gaaggctgag aaagagaaga tgaagacaaa gcaggaaaag
2161 ctgaagggaa aggtaaagcg agaaaagaaa gaaaaggtaa aagcgaaggg gaaggaaggg
2221 cccagagcca ggccatcctg tagagcagac aagaccctg ccacacagaa gcggtagag
2281 gagcagcaga ggcagcaggc tatcctggag gagatgaaga agcccacaga gggatgtgt
2341 ctgtctgacc accagcccct gcctgacttc acacgcatcc ctggtttgac actgtccagt
2401 agggctttct cagattgctt gacctcgtg gagttccttc acagttttgg caaagtgcta
2461 ggctttgacc ttaccaaaga tgttcttagt ctaggagtcc tgcaggaggg actcttatgt
2521 caaggtgaca gcttgacaa agtgcaggac ctgctgggtg gactgctgaa ggctgactc
2581 catgatcctg gtctgcccc ctactgtcag tcctgaaga tattggggga gaagatgtca
2641 gagatcccat tgaccagaga taatgtgtct gagatactgc gctgcttct catggcatat
2701 agagtggagc cacccttctg tgacagtctg cgtaccagc cttttcaggc ccagccacct
2761 caacagaagg ctgctattct agccttctt gtgcatgagc ttaacagctc caccattatc
2821 atcaatgaga ttgacaagac tctggaaagt gtgtctagct gcaggaagaa caagtggtat
2881 gttgaaggcc gactccggag actgaaaact gctctggcca agcgaactgg cgggtctgag
2941 gttatgatgg aaggggcaga agacggccta ggacggaggc gcagttctcg gatcatggag
3001 gaaaccagtg gcatagaaga ggaggaagag gaagaaaata caacagctgt ccatggccgc
```

3061 aggggtcgaa aagaaggaga gattgatggt gcagcatcta gcattccaga gctagagcgc
 3121 catatagaaa aactcagtaa gcgtcagctc ttcttttagaa aaaagctgct tcactcatcc
 3181 cagatgcttc gggcagtgtc ctgggtcaa gaccgctata gacgcatta ctgggtatta
 3241 ccgatctctg ctggtatctt tgtggaagga tcggaaggga gcacagttac tgaagatgaa
 3301 ataaagcaag aaactgagtc cttgatggaa gtagtactt caacaccag ctctgccga
 3361 gcctctgtaa agagagaatt aactggctcc aatgcctcta cttctcctgc ccggtcccga
 3421 ggccgacctc gaaaacctaa gcctgggtct ctgcagctc agcaccttca gtccaccatt
 3481 agggaatgtg attcagagca agcccagact caagtccacc cagaaccca gcctcagctt
 3541 caggccccta cccagcccca tcttcagcca agtagtgggt tcctagagcc agaaggttcc
 3601 cttttctctc tgggtcagag ccagcatgac ctcagccagt ctgccttctt gtcttggctg
 3661 agccagactc agagccacaa ctccctgttg agcagctcag tcctcacgcc ggacagcagc
 3721 ccagggaaac tagactctgc tccgtctcag tccttggagg agcccagacc tgatgaggct
 3781 cagtcctgcc ctggctctca aggtccctgg ttttaacttct cagcccagat accctgtgat
 3841 gctgctccta caccacctc tgctgtctt gaggaccaac ctactccctc cctccagctg
 3901 ctggcctcct ctaaaccaat gaatacacc ggtgctgcca atccttgttc cccagtgcag
 3961 ctctcttcca ctacttgcc tggagggacc cctaagaggc tatcagggga ctctgaagaa
 4021 atgtcacaga gtcccactgg gctggggcaa ccaaagcggg gggggagacc ccctagcaag
 4081 ttcttcaagc aggtggagca gcattactta acccagctga cagcccagcc tatccccct
 4141 gagatgtgct cgggctgggt gtggatccga gaccctgaga cactggatgt cctgctcaag
 4201 gcactgcatc cccgagggat ccgggagaag gcgcttcaca aacatcttag caagcacaag
 4261 gactttttgc aggaagtttg ttacagccc ttaactgatc ccactcttga gcctaataag
 4321 ctccctgctc tggagaagg cgttatgagc tggccccca aagagaagac gtacatgaca
 4381 gacctagctg tgctccagtg ggtggaggag ctggagcagc gggttgtcct ctccgatctg
 4441 cagattcggg gctggacatg ccctaccca gactccacca gagaagactt gacctactgt
 4501 gagcatctgc ctgactcccc ggagatata ccttggaggg gtcggggcag ggaaggaaca
 4561 gtacctcagc ggcagaacaa caaccctctg gacctcgtg tgatgogatt ggctgttctg
 4621 gagcaaatg tggagcggcg gtacttgccg gagcccctct gggcagccca tgaggtgta
 4681 gtggagaagg ccctactgag cacacccaat ggtgcccctg atggcacctc aactgagata
 4741 tcctatgaga tcacccctcg tgtccgagtt tggcggcaga cacttgaag gtgccgtagt
 4801 gcagcccaag tgtgcttgtg catgggccag ctagaaggt ccacgcgctg ggagaagtct
 4861 gtcaacaaag tgacctgcct ggtctgccgg aagggcgata atgatgagtt tctcctgctg
 4921 tgtgatgggt gtgaccgagg ctgccacatt tactgtcatc ggccaagat ggaggctgtt
 4981 ccagaaggag attggttctg tgctgtctgt ctgtcccagc aggtagagga agagtactt
 5041 cagaggcctg gttttccaaa acgaggtcag aagcggaaaa gtagttttcc actgacctt
 5101 ccagaagggt acagccggcg gcggtggtg tcaaggagcc gagatagtcc agcagtgcct
 5161 cggtagccag aagacgggct gtctccccc aaaagacggc gacattcgat gagaagtcac
 5221 cacagtgatc tcacattttg cgagattatc ctgatggaga tggagtccca tgatgcagcc
 5281 tggcctttcc tagagcctgt gaaccctcgc ttgggtgagt gataccgacg tgtcatcaag
 5341 aaccctatgg atttttccac catgcagaa cgcctgctcc gtcggaggta caatagctca
 5401 gaagatgttg cagctgatgc tctgctgggt tttgacaact gccagacctt caatgaggat
 5461 gactctgaag tgggcaaggc tgggcaagtc atgcgacgct tctttgagag ccgctgggag
 5521 gaattttatc agggaaaaca ggccaatctg tga

H. sapiens TIP5. Accession number NCBI NM_013449.

1 agtctgtcac acagggacat caccctacag cagttcaggc tgtgtgggtc gcaggaagca
 61 tacactggct ttttgattct tgctagttcc cagctcacag tttgggagga tccaacacca
 121 acctttacgt gaagtggagg cccaaggaca gtgaggagct ggggtgctcc agcctggagc
 181 tgtgccagcc tgacatggaa atggaggcaa acgaccattt taactttact ggccttcccc
 241 ctgcacctgc tgacctcagga ctgaaacct ctccttctc aggggagggc ctctacacta
 301 acgggtctcc catgaacttc ccccagcaag ggaaaagttt gaatggggat gtgaatgtta
 361 atggcttatc tactgtatct cacactacta cttcagggat tttgaactct gctccccact
 421 cctccagcac ctcacacctc catcaccca gcgtggccta cgactgtctc tggaaactact
 481 cacagtacct atctgccaat cctggcagca acctcaagga cccaccctt ctctcccagt
 541 tctcgggggg acaataacca ctcaacggca tccttggggg cagccggcaa ccttcatccc
 601 caagtataaa cactaacctt cgggctggga gccaaagatt ctggggccaa ggtaccacga
 661 gtcccatggg gcttaacttt gattcacaag aactgtatga ttcctttct gaccagaatt
 721 ttgaggtgat gcccaatgga cccctagtt ttttcacct cccacagact tctcctatgt
 781 tgggatctag cattcaaacc tttgaccct cccaggaggt aggcagtggt atccatcctg

841 atgaggcagc agaaaaggag atgacttcag ttgtggcaga gaatggcact ggcttggtag
901 gcagcttggg gctggaagaa gacgagccag aactgaagat gtgtggctac aatggctctg
961 tcccttctgt ggaatcgta caccaagagg tctcagtcct ggtccctgac cccacagtga
1021 gctgttttaga tgatccttca catcttcctg atcaactgga agacactcca atcctcagtg
1081 aagactctct ggagcccttc aactctctgg caccagagcc agtgagtgga ggactatatg
1141 gtattgatga cacggagctg atgggtgcag aggacaagct gcctcttgag gacagccctg
1201 tgattttctgc ccttgattgc ccttccctca ataatgctac tgcttctcagt ctctggcag
1261 atgatagtca aacatcaacc tctatctttg ccagtcccac ctctccacct gtcctagggg
1321 agtctgtcct gcaagataac agctttgacc tgaataatgg tagtgacgct gaacaggaag
1381 aaatggaaac tcaatcttca gacttcccac catccctgac ccagccagct cctgatcagt
1441 catccactat tcagctacat ccagcaacct caccagcagt ctcgccaaca acctccccag
1501 cagcttccct agtggtttct ccagcagcct cccagaaat ctctccagaa gtttgtcccg
1561 cagcttctac agttgtctct ccagcagct tctcagtggt tctccagct tctcagcag
1621 tctcccagc agtctcctta gaagtcccg tgcaggcttc agtgacatcc ccaaaagcct
1681 ctcccgtaac ttcccagca gctgcctttc caacagcctc cccagcaaat aaggatgtca
1741 gcagctttct agaaaccact gctgacgtgg aagagatcac tggagaagga ctactgtct
1801 ctggtagtgg tgatgtcatg aggagacgta ttgctacccc agaagaagtt cgtcttccc
1861 tccaacatgg ttggcgggaga gaggtgcgca tcaagaaggg cagccaccga tggcagggg
1921 agacctggta ttatggcccc tgtgggaaga ggatgaagca atttccagaa gtgatcaagt
1981 acctgagccg caacgtggta cacagtgtcc gccgagagca cttcagcttc agtccccgta
2041 tgcctgttgg agatttcttt gaagaagag acacgccaga gggcttgcag tgggtgcagc
2101 tctcagcaga ggagatcccg tgcaggatc aggcaattac tggcaaacgg ggtcagcctc
2161 gaaacactga gaaggctaag actaaggaa tccccaaggt gaaacggggg ctaggtcggc
2221 cacctaaggt caaaatcact gagctattga acaagacaga caaccgcccc ctaaagaaac
2281 tggaggccca agaaacattg aatgaggagg ataaagcaa gattgctaaa agcaagaaga
2341 agatgaggca gaaggttcaa cggggagagt gtcagactac tatccaaggg caggccagaa
2401 ataagcggaa acaagagacc aagagcttaa agcagaagga agctaagaag aaatccaagg
2461 ctgagaaaga aaaaggaaa acaaagcagg aaaaactgaa ggaaaaagtc aagagggaaa
2521 agaaggagaa ggtaaaaatg aagggaaaag agggagtgac caaagccaag ccagcctgta
2581 aagcagataa gacctggcc acacagaggg gcttggagga acggcagagg cagcagatga
2641 ctgttgagga aatgaagaag ccgacagagg atatgtgtct gactgaccac gaccctgc
2701 ctgacttctc acgagtcctt ggtctgacat tgcccagtg agccttctca gactgcttga
2761 ccattgtgga gttcctgcat agctttggca aggtgctggg ctttgatcct gccaaagatg
2821 tgcctagcct gggggtcctg caggaggac tctgtgtca aggtgacagc ttgggtgagg
2881 tgcaagacct gctggtcagg ctgctgaagg ctgcactcca tgatcctggc tttccctcct
2941 actgtcagtc cctaaagatc ttgggggaga aggtgtctga gatcccactg acaagagaca
3001 atgtgtcaga gatcctgcgc tgcttctta tggcatatgg agtagagcca gcctctgtg
3061 accgctgcg caccagcct tttcaggccc agccaccca gcagaaggct gctgtcctgg
3121 ccttcttctg gcatgagctc aatggctcca cctcatcat caatgagatt gacaagactc
3181 tggagagtat gtccagctac aggaaaaaca agtggattgt tgaaggccg cgcggaggc
3241 tgaaaactgt tctggccaag cgaactgggc ggtctgaagt agagatggaa gggccagagg
3301 aatgcctggg acggaggcgc agttctcgga tcatggagga gaccagtggc atggaagaag
3361 aggaagaaga ggagtctata gcagctgtcc ctggccgag ggtcgaaga gatggagagg
3421 ttgatgccac agcatctagc atcccagagc tagagcgcca gatagaaaaa ctcagcaagc
3481 gtcagctttt ctttcgcaaa aagctgctt actcatcca gatgcttcgg gcggtctccc
3541 tgggtcagga ccgctacaga cgtcgtact gggatttggc gtatttggct ggtatctttg
3601 tagaaggaac agaggggaac ttagtctctg agggagtgat aaagaaggaa actgactcct
3661 taaaagtggc agcccagcg tcaactcaacc ctgccctctt ctctatgaag atggagttag
3721 ctggctccaa caccactgcc agttctcctg cccgggcccg aggcogacct cgaaaaacta
3781 agccccggtc tatgcaacct aggcattctta agtcccctgt caggggtcag gattcagaac
3841 agccccaggc ccagcttcag cctgaggctc agcttcatgc tctgcccag cccagcctc
3901 agcttcagct tcagcttcag tcccataagg ggttcttgg gcaagaaggc tccccttgt
3961 cactgggtca gagccagcat gacctcagcc agtcagcctt cctgtcttgg ctgagccaga
4021 ctcagagcca tagctccctg ttgagcagct cagtcctcac acctgatagc agtccgggaa
4081 aactagacc agctccatca caacccccg agggagccaga gcctgatgag gcagaatcca
4141 gccctgatcc tcaagcactc tggtttaaca tctcagcca gatgcctgc aatgctgccc
4201 ccacaccgcc cctgcagtt tctgaggacc aacccactcc ctcccctcag cagctgctc
4261 cctccaagcc aatgaataga cctagtgtctg ccaacccttg ttctccagtg cagttctctt
4321 ccacgcctt ggctgggttg gccctaaaga ggcgagcagg agaccctgga gaaatgccac
4381 agagtccac agggctggga cagcccaaac ggagagggag acctccagt aagttcttca
4441 aacagatgga acagcgttac ctaaccagc tgacagcca gcctgtcca cctgagatgt

4501 gctcaggctg gtggtggata cgagatcctg agatggttga tgccatgctc aaggccctac
4561 acccccgagg tatccgggag aaggcacttc acaaacacct taacaagcac agggacttct
4621 tgcaggaagt ctgcctgcgg cctcagctg accccatctt tgagcccagg caactacctg
4681 cctttcaaga agggattatg agctggtccc ccaaagagaa gacatacgag acagacctag
4741 cagtgcctta atgggtagag gagctggagc agcgggttat catgtctgat ctgcagattc
4801 ggggctggac atgtcctagc ccagactcta cccgtgaaga cttggcctac tgtgagcacc
4861 tctccgactc ccaggaggat atcacctggc gaggtcgggg cagggagggg ctggcacctc
4921 agcgtaaaac taccaaccct ttggacctgg ctgtgatgcg gctggctgcc ctggaacaga
4981 atgtagaacg gcggtacctg cgggagcccc tctggccaac tcatgagggt gtgctggaga
5041 aggccctgct tagcacacct aatggtgccc ctgagggcac cactacagag atatcatatg
5101 agatcacccc tcgcattcgt gtctggcgcc agaccctcga gcggtgccgg agcgcagccc
5161 aggtgtgctt gtgcctgggc cagctggaga ggtccattgc ctgggagaag tctgtcaaca
5221 aagtgcacatg tctagtctgc cggaaagggtg acaatgatga gtttcttctg ctttgtgatg
5281 ggtgtgaccg tggctgccac atttactgcc atcgtcccaa gatggaggct gtcccagaag
5341 gagattggtt ctgtactgtc tgtttggctc agcaggtgga gggagaattc actcagaagc
5401 ctggtttccc aaagcgtggc cagaagcggg aaagtgggta ttcgctgaac ttctcagagg
5461 gtgatggccg ccgacgccgg gtactgttga ggggccgaga aagcccagca gcagggcctc
5521 ggtactcggg agaagggctc tccccctcca agcggcggcg actctctatg cggaccacc
5581 acagtgatct cacatthttg gagattatcc tgatggagat ggagtcccat gatgcagcct
5641 ggccttttct agagcctgtg aaccacggtt tgggtgagtgg gtaccggcgc atcatcaaaa
5701 atcctatgga tttttccacc atgcgggagc ggctgctcag gggagggtag accagctcag
5761 aggagtthttg ggcctgatgcc ctcctggtat ttgacaactg ccagacttht aacgagtgatg
5821 actctgaagt aggcaaggct gggcacatca tgcgccgctt cttcagagac cgctgggagg
5881 agttttatca gggaaaacag gccaatctgt gaggcaaggg aggtggggag tcacctgtg
5941 gcatctcccc ccacctcca acaaaaacc tgccattht acctgctgat gctgccctgg
6001 gtccagactc aagtcagata caacctgat ttttgacct gcccttggca gtgcccaca
6061 tcctcttatt cctacatccc tttctccctt ccctcctctt gctcctcaag taagaggtgc
6121 agagatgagg tccttctgga ctaaaagcca aaaaaagaaa gaaaaaata atthttctt
6181 tctgthttat ttgctaatta aaaatgggga ggggaaagt cgtccctact tcctcctccc
6241 tgcttctctc cctcccctgt acgtgcccca gcattctggg gttatthtaac aatagcaata
6301 gttctagtga atgtgtgaaa ccaagaaaca ctctgtactg tgtgcgacc cgcagtagc
6361 gccagtaaac tggacttaac tcccagtggt gtcgccccg acaccgggccc ctggacatgc
6421 tgcttccatg ttcagtcctt tcctgcttc tcgctgtctt tcttttcca cctcccacc
6481 cccagthttc agatthttct tcaccaata atgtaaaact atcgtgtacg ggttctctcc
6541 tcctthttct tcttccaaa tctthtccct tcaaaggaaa aaaaaatggt cagaggtccc
6601 tgtcttctgt ccccatcttc ctgccgatag ctatcccctg tatgatggtg gatgctctc
6661 acatgctgag thtccagcct thtctgaaac tcagtagctg gggagagggc agggaggtt
6721 cctgggcctt ccagcctcct tccccacct cttcccaaac cctcttggga actcctcagg
6781 gacaactact gctgagthtt ggtgcacct aagatggagg ccaggtagca atggggccgg
6841 cctcagagag agcgtgtgtg gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtcggcctca
6901 gagcgcgctg tgtgtgtgtg tgtgtgtgtc agcctcagag cgcgcgctgt gtgtgtgtgt
6961 gtgtgtgctg gctgtgacc ctgtattgtt tgataggatc cattcagtht cccaagtac
7021 ctgthttcat tccatthttt ccattgttht aaaacctca cthttttgtc thtgggaaac
7081 cacaggaaca atthtctctg agacaaggct gtgtctctct cctggctatt thtgttccag
7141 cctcttcaga ctgtgcaatc thttagcagg aactccctct cttctcggta gctthtgaatc
7201 ttaagcttct acgggagagt ggtagaactg gatcattthc taatcccat tagthtgtct
7261 thtcttcatg tacttcatc cccaggacc cttccccagc agcagagacc ctggagcaca
7321 ggagagttag gaggaggggt tctgggtcca tcaactgcct acatgtgact atgtccaagt
7381 taagccccc aacagagagg aaagctgctg actcccagct atagccatgg gcacttggcc
7441 cctgcttht cctgctcagc agagcccct ccttcagaga ttacgggtac thttagctggg
7501 gagggtgctg ctggctggcc caagcagaga gctgaggcat ccaagaaatg thttagctggg
7561 ggggtggggg gccagggtgag gtggagcatt ccttgtattc tggcagcact gaagagccac
7621 tgaaggggg aggggcagtg taggtcctgg ggcagcccct thattcctth atgcccctc
7681 tccctcatag cctatthtct aagtcgcctt thtctgcaga taaacctcaa aactthtaat
7741 thtatttht atthttthtt gctthttaa ggtggattga agaatathtt aattgacttht
7801 atattatgca taaatathtt tathttatct aaataactgc gctgtaacaa actthtgtgt
7861 agacagthta aactgttaga gttgggggct ctgctthttc cccctggcaa thtthcccctg
7921 gtataagatg tgctagatta atthcattgt gagggtgatg ggggagtgaa atthtgaagt
7981 ggatggggga gtgaaatctc catggttctt gctthtgtgt cctctcccag ctccatctct
8041 thccttaggg accaggcact catatggcgg ggtgggggtc tagcctcagt thtgaagaagt
8101 gggggctgga gcgggggttg ggggtgtagg gatagggcat gatcaaagg gcaatthctt

8161 gcttttcttt cctcatcttc actgccccct tgagctaggt ggattttctc ttcatgacaa
8221 gagtatttgg tagggaaagc aggtttaaaa taaaaagaca acccaccccc tgcccttttg
8281 cttccctccc atcagtctgg ttgacaggaa gaaaccacac catcaacacc aacaagtttc
8341 tgtgttcctt ttacagcaaa agggactttt tatataacca aatgtggtgt tttagtgact
8401 ttttgataat gtacagtttt ttgtgaattt aaattttattt ctttctatat ttttaggacc
8461 aatctcattt ttaataaggt taaaaagaaa aaaaaagtct agcgaaaaaa ctctgtttt
8521 tgccatgtga tgttccacaa gtgcagctgt agaaaagtgc ttgtcagttg ttgaataaaa
8581 aaaaccacat ttgatagaga ttcaaaagac tctgtgtatt catcttcctt tctacacacc
8641 tgagggggag acggcgttgg gataggtatg actggcttaa gagaccacag gcaagggaac
8701 aacaggggct cctgttccat accctctgtg tggatggaaa gggtcattag tgctcctgcc
8761 taaatgtctg gctgagttgc tggagcaaa ggggattcag tgcattccagg tctgccttg
8821 tgagatgtgg ccccagcttc ctaagctgcc acctctgtgt tctgtcata gcaaatatgg
8881 gaccatcacc agcttaccac ttcccactca cggataagac accaagacgt agac

TIP4N (*H. sapiens* TIP5₄₉₆₋₆₆₄), codon optimized for *E. coli* expression.

ATGGCTGACGTGGAAGAGATCACTGGAGAAGGACTCACTGCTTCTGGTAGTGGTGATGTCATGCG
CCGTCGTATTGCTACCCCAGAAGAAGTTCGTCTTCCTCTCCAACATGGGTGGCGCCGTGAGGTGCG
CATCAAGAAGGGCAGCCACCGATGGCAGGGGGAGACCTGGTATTATGGCCCCTGTGGGAAGAGG
ATGAAGCAATTTCCAGAAGTGATCAAGTACCTGAGCCGCAACGTGGTACACAGTGTCCGCCGAGA
GCACTTCAGCTTCAGTCCCCGTATGCCTGTTGGAGATTTCTTTGAAGAAAGAGACACGCCAGAGGG
CTTGCAAGTGGGTGCAGCTCTCAGCAGAGGAGATCCCGTCGAGGATTCAGGCAATTACTGGCAAAC
GGGGTCGACCTCGAAACTGAGAAGGCTAAGACTAAGGAAGTCCCCAAGGTGAAACGTGGTTCGC
GGCCGCCACCTAAGGTCAAATCACTGAGCTATTGAACAAGACAGACAACCGCGGATCC

TAM₅₁₆₋₆₂₃ (*H. sapiens* TIP5₅₁₆₋₆₂₃), codon optimized for *E. coli* expression.

ATGGAACGCCGTCGTATTGCTACCCCAGAAGAAGTTCGTCTTCCTCTCCAACATGGGTGGCGCCGT
GAGGTGCGCATCAAGAAGGGCAGCCACCGATGGCAGGGGGAGACCTGGTATTATGGCCCCTGTGG
GAAGAGGATGAAGCAATTTCCAGAAGTGATCAAGTACCTGAGCCGCAACGTGGTACACAGTGTCC
GCCGAGAGCACTTCAGCTTCAGTCCCCGTATGCCTGTTGGAGATTTCTTTGAAGAAAGAGACACGC
CAGAGGGCTTGCAAGTGGGTGCAGCTCTCAGCAGAGGAGATCCCGTCGAGGATTCAGGCAATTACT
GGATCC

11.2. Protein sequences.

TIP4N (*H. sapiens* TIP5₄₉₆₋₆₆₄)

GAMADVEEITGEGLTASGSDVMRRRIATPEEVRLPLQHWRRREVRIKKGSHRWQGETWYYGPCGK
RMKQFPEVIKYL SRNVVH SVRREHFSFSRMPVGDFFEERDTPEGLQVWQLSAEEIPSRIQAITGKRGRP
RNTEKAKTKEVPKVKRGRGRPPKV**KIT**ELLNKTDNR**GS**

TAM₅₁₆₋₆₂₃ (*H. sapiens* TIP5₅₁₆₋₆₂₃)

GAMERRRIATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYYGPCGKRMKQFPEVIKYL SRNVVHS
VRREHFSFSRMPVGDFFEERDTPEGLQWVQLSAEEIPSRIQAITGS

NB: residues added to the wildtype domain sequence by cloning and purification are highlighted **red**. K653, highlighted **green** in construct *H. sapiens* TIP5₄₉₆₋₆₆₄, corresponds to K633 of murine TIP5, which is acetylated by MOF (males absent on the first) acetyltransferase in context of full length protein and exerts a negative effect on pRNA interaction of NoRC (Zhou et al, 2009). The numbering of the constructs refers to the isoform I of human TIP5 protein, UniProtKB accession number Q9UIF9.

11.3. DNA oligonucleotides for *in vitro* RNA transcription.

T7 top

5'-TAATACGACTCACTATA-3'

1wt top

5'-TAATACGACTCACTATA-GGGGGGTCATTTTTGGGGAAACCCTGTCTCTTTCC-3'

short top

5'-TAATACGACTCACTATA-GGGGGGTCATGAAAGTCTCTTTCC-3'

1wt bottom

5'-GGAAAGAGACAGGGTTTCCCCAAAAATGACCCCC-TATAGTGAGTCGTATTA-3'

1wt bottom methoxylated

5'-G*G*AAAGAGACAGGGTTTCCCCAAAAATGACCCCC-TATAGTGAGTCGTATTA-3'

2sh bottom

5'-GGCAGAGACAGGGTTTCCCCAAAAATGACCACC-TATAGTGAGTCGTATTA-3'

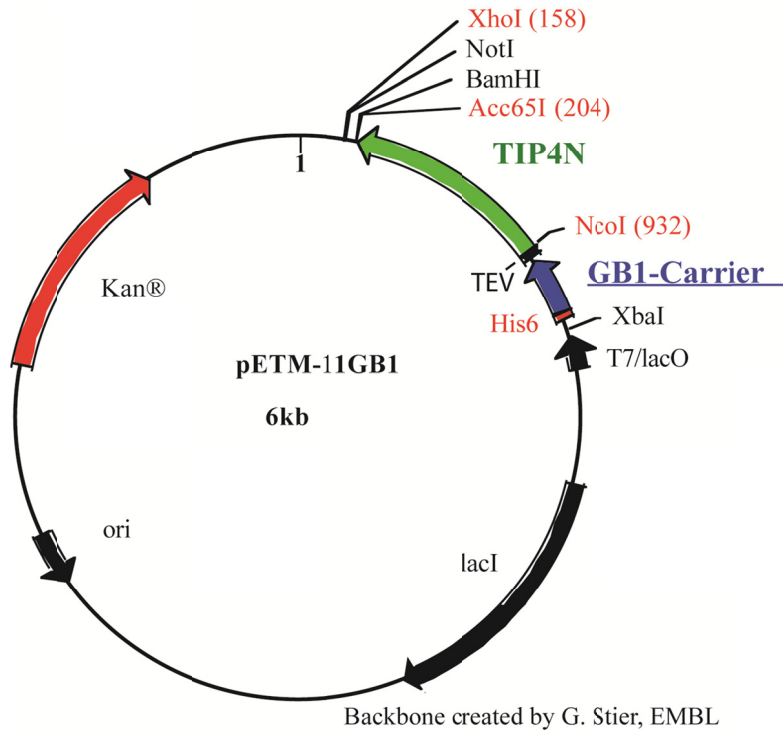
short bottom methoxylated

5'-G*G*AAAGAGACTTTCATGACCCCC-TATAGTGAGTCGTATTA-3'

NB: the T7 bacteriophage promoter sequence is separated from the transcribed sequence by a dash and is highlighted bold. * indicate 2'-O-Methyl-RNA nucleotides. That modification was introduced to reduce the number of n+1 products (where n is the number of the nucleotides in the desired RNA).

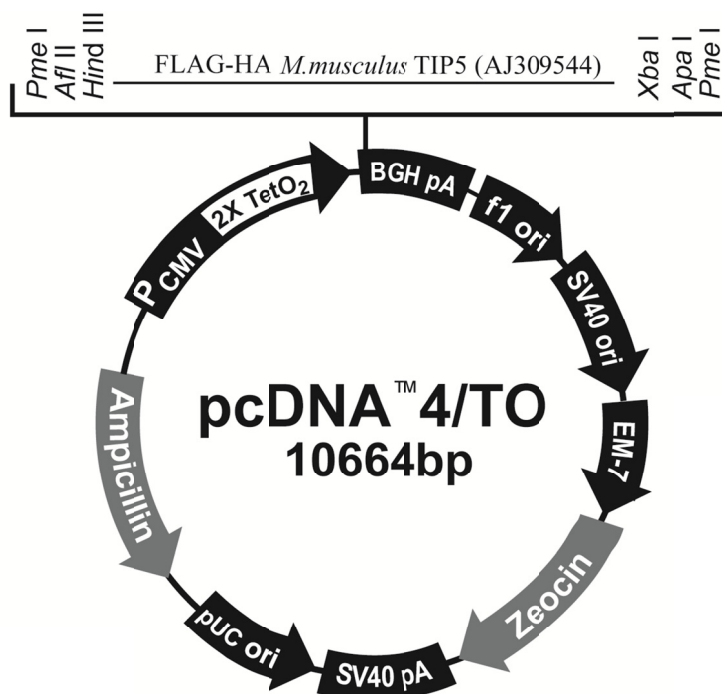
11.4. Vector maps.

E.coli expression vector pETM11-GB1 (European Molecular Biology Laboratory), kanamycin resistance (the gene is marked Kan®).



The N-terminal hexahistidine tag is marked red, the GB1 tag is marked blue. TEV cleavage site is marked TEV. The origin of replication, as well as the lacI operon and the T7/lacO promoter are represented by black arrows, pointing in the direction of the reading frame. The unique enzymatic restriction sites are marked. EMBL stands for European Molecular Biology Lab.

Eukaryotic expression vector pcDNATM4/TO (invitrogen), ampicillin resistance.



Kindly provided by K. M. Schmitz, dkfz

The vector is tetracycline inducible and contains among other features a CMV promoter and a tetracycline operator. The explanation of vector features and abbreviations can be found on the product description site from (Invitrogen). dkfz stands for Deutsches Krebsforschungszentrum (German Cancer Research Centre).

11.5. Resonance lists

TIP4N

	C α	C β	H ^N	N		C α	C β	H ^N	N
M495	54.85	32.31	8.459	119.6	R572	56.51	30.05	7.660	119.9
A496	52.14	18.86	8.291	124.8	N573	52.46	39.41	7.674	117.0
D497	54.13	40.56	8.304	119.5	V574	62.60	30.59	8.020	119.7
V498	61.78	32.23	7.988	119.1	V575	61.53	32.28	8.237	123.2
E499	56.19	29.69	8.425	123.9	V578	60.36	32.88	7.265	116.6
E500	56.16	29.73	8.377	122.3	R579	53.77	33.58	10.23	125.9
T502	61.57	69.67	8.252	117.5	R580	59.46	28.96	8.765	119.5
G503	45.02	-	8.396	110.7	E581	-	-	9.074	116.0
E504	56.38	29.59	8.335	120.6	H582	56.30	30.17	7.848	117.3
G505	45.05	-	8.493	109.5	F583	58.06	40.86	7.349	115.6
L506	55.11	41.52	8.092	121.3	S584	56.36	65.86	8.049	110.8
T507	61.49	69.46	8.127	114.0	P590	62.55	27.13	-	-
A508	52.24	18.73	8.347	126.0	V591	61.49	34.98	8.340	124.2

S509	58.27	63.85	8.340	114.9	G592	42.71	-	8.731	115.5
S511	58.88	63.59	8.245	115.6	F594	55.80	42.92	9.225	124.3
G512	44.99	-	8.468	110.7	F595	55.68	44.43	9.242	118.5
D513	54.06	40.61	8.167	120.5	E596	53.34	34.44	9.462	121.3
M514	53.57	43.30	8.489	119.3	E597	55.48	28.73	8.180	128.2
V514	62.62	31.71	8.049	119.8	R598	54.28	33.28	9.013	126.3
M515	55.90	32.23	8.416	122.2	D599	54.41	40.84	8.475	121.4
R516	55.75	29.75	8.383	123.0	T600	59.91	70.11	7.669	116.5
R517	55.68	29.90	8.205	121.7	P601	64.78	30.84	-	-
R518	-	-	8.255	122.5	E602	55.51	28.89	7.726	114.2
I519	60.13	38.10	8.279	123.9	G603	43.22	-	7.817	109.1
A520	51.85	18.83	8.315	129.2	L604	55.31	41.14	8.459	123.7
T521	58.98	68.34	8.583	113.0	Q605	53.87	32.39	8.769	122.9
P522	65.42	30.74	-	-	W606	56.53	29.70	8.721	121.5
E523	59.66	28.39	8.276	114.2	V607	61.11	33.96	9.442	125.7
E524	58.96	29.27	7.610	119.5	Q608	55.67	27.00	8.975	127.5
V525	62.72	31.06	7.869	117.0	L609	53.68	41.91	8.747	127.2
R526	57.71	30.17	7.867	117.3	S610	56.84	64.94	9.199	119.2
L527	59.80	38.05	7.417	121.5	A611	55.01	17.32	8.831	123.2
P528	66.14	29.03	-	-	E612	58.12	28.34	8.530	114.2
L529	56.83	38.81	7.082	116.5	E613	56.89	31.70	7.657	118.6
Q530	55.88	27.64	7.867	117.5	I614	67.37	-	7.106	118.5
H531	55.46	30.84	7.683	117.8	R617	57.74	28.35	7.640	123.9
G532	45.66	-	7.990	106.6	I618	64.71	36.51	7.952	118.2
W533	57.70	28.81	7.947	119.4	Q619	57.93	27.54	7.883	119.5
R534	54.78	33.75	9.092	117.2	A620	53.74	17.25	7.442	121.2
R535	54.60	34.70	8.892	123.6	I621	61.83	37.44	7.516	112.9
E536	52.98	33.65	9.172	124.9	T622	62.91	69.33	7.822	117.0
V537	60.69	31.93	8.689	123.1	G623	44.89	-	8.263	110.9
R538	53.99	30.21	8.517	126.3	P628	61.78	32.40	-	-
I539	59.91	39.40	9.434	126.1	R629	56.04	29.64	8.407	124.3
K540	54.69	35.32	9.056	123.5	N630	60.82	37.87	8.289	122.3
K541	56.06	31.19	8.763	124.7	T631	61.38	69.46	8.285	118.3
G542	43.99	-	7.894	113.7	E632	56.32	29.43	8.376	123.3
G548	43.23	-	8.643	107.0	K633	55.94	32.10	8.317	122.6
E549	54.85	33.01	8.447	119.6	A634	52.17	18.50	8.293	125.1
T550	61.54	69.21	8.989	122.2	K635	56.04	32.25	8.337	121.0
W551	52.70	32.65	9.567	125.2	T636	61.53	69.57	8.118	115.9
Y552	57.05	43.23	8.824	118.3	P650	62.45	31.48	-	-
Y553	55.37	39.87	9.369	117.0	K651	55.99	32.20	8.354	121.4
P555	64.10	30.71	-	-	V652	61.63	32.23	8.087	121.8
C556	57.04	30.24	7.167	113.3	K653	55.79	32.27	8.415	125.8
G557	44.50	-	8.105	112.9	I654	60.95	37.85	8.293	123.4
K558	57.46	33.38	8.594	125.6	T655	61.91	69.08	8.128	118.0
R559	55.16	30.41	8.423	127.4	E656	56.15	29.63	8.325	123.5
M560	55.11	37.21	9.612	127.8	L657	54.88	41.31	8.175	122.9

K561	56.52	34.19	8.930	120.3	L658	54.87	41.35	8.127	122.2
Q562	53.20	31.98	7.435	113.1	N659	52.81	38.30	8.327	119.0
P564	65.92	29.57	-	-	K660	56.01	32.24	8.270	121.7
E565	58.43	29.38	7.153	113.7	T661	61.56	69.38	8.188	114.6
V566	65.61	30.82	7.345	121.9	D662	53.83	40.71	8.293	122.3
I567	64.00	36.23	7.697	118.0	N663	52.94	38.25	8.381	119.4
K568	58.87	31.64	7.621	119.5	R664	56.11	29.69	8.353	121.1
Y569	61.69	39.13	7.908	122.5	G665	44.91	-	8.413	109.9
L570	56.79	40.50	8.753	117.7	S666	59.58	64.75	7.859	121.2
S571	60.09	63.28	7.644	112.0					

TAM₅₁₆₋₆₂₃

Atom Nuc Shift	Atom Nuc Shift	Atom Nuc Shift	Atom Nuc Shift
M514 H 1H 8.529	K568 HG2 1H 1.465	K568 QB 1H 1.836	E596 QG 1H 2.089
M514 N 15N 119.520	K568 N 15N 119.683	K568 QD 1H 1.635	E597 CA 13C 55.687
R516 CA 13C 55.638	G542 H 1H 7.963	K568 QE 1H 2.881	E597 CB 13C 29.443
R516 CD 13C 43.302	G542 HA1 1H 3.767	Y569 CA 13C 62.202	E597 CG 13C 37.171
R516 H 1H 8.392	G542 HA2 1H 4.261	Y569 CB 13C 39.670	E597 H 1H 8.238
R516 HA 1H 4.385	G542 N 15N 114.368	Y569 CQD 13C 133.856	E597 HA 1H 4.021
R516 N 15N 122.789	S543 CA 13C 60.377	Y569 CQE 13C 117.962	E597 HG1 1H 0.704
R516 QD 1H 3.141	S543 CB 13C 62.486	Y569 H 1H 7.918	E597 HG2 1H 1.148
R517 CA 13C 56.001	S543 H 1H 8.695	Y569 HA 1H 3.929	E597 N 15N 128.717
R517 CB 13C 30.889	S543 HA 1H 3.961	Y569 HB1 1H 3.219	E597 QB 1H 1.513
R517 H 1H 8.375	S543 QB 1H 3.702	Y569 HB2 1H 3.358	R598 CA 13C 54.652
R517 HA 1H 4.274	H544 CA 13C 57.132	Y569 N 15N 122.728	R598 CB 13C 34.259
R517 HB1 1H 1.724	H544 CB 13C 31.484	Y569 QD 1H 6.960	R598 CD 13C 43.669
R517 HB2 1H 1.776	H544 CD2 13C 119.577	Y569 QE 1H 6.689	R598 CG 13C 27.160
R517 N 15N 122.911	H544 CE1 13C 138.278	L570 CA 13C 57.139	R598 H 1H 9.012
R517 QD 1H 3.140	H544 HA 1H 4.530	L570 CB 13C 41.396	R598 HA 1H 4.684
R517 QG 1H 1.591	H544 HB1 1H 2.965	L570 CD1 13C 25.083	R598 HB1 1H 1.593
R518 CA 13C 55.941	H544 HB2 1H 3.155	L570 CD2 13C 22.806	R598 HB2 1H 1.790
R518 CG 13C 27.062	H544 HD2 1H 6.993	L570 CG 13C 26.516	R598 HD1 1H 2.988
R518 H 1H 8.398	H544 HE1 1H 7.815	L570 H 1H 8.768	R598 HD2 1H 3.138
R518 HA 1H 4.324	R545 CA 13C 54.976	L570 HA 1H 4.152	R598 N 15N 126.751
R518 N 15N 123.508	R545 CB 13C 32.348	L570 HB1 1H 1.614	R598 QG 1H 1.603
R518 QB 1H 1.743	R545 CG 13C 26.271	L570 HB2 1H 1.888	R598 he 1H 7.560
R518 QD 1H 3.131	R545 HA 1H 4.616	L570 HG 1H 2.019	R598 ne 15N 128.600
R518 QG 1H 1.554	R545 HB1 1H 1.863	L570 MD1 1H 0.591	D599 CA 13C 54.861
I519 CA 13C 60.491	R545 HB2 1H 2.027	L570 MD2 1H 0.950	D599 CB 13C 41.328
I519 CB 13C 38.909	R545 HG1 1H 1.442	L570 N 15N 117.923	D599 H 1H 8.518
I519 CD1 13C 12.840	R545 HG2 1H 1.540	S571 CA 13C 60.473	D599 HA 1H 4.610
I519 CG1 13C 26.972	R545 QD 1H 3.201	S571 CB 13C 63.305	D599 HB2 1H 2.550
I519 CG2 13C 17.482	W546 CB 13C 30.149	S571 H 1H 7.631	D599 HB3 1H 2.390
I519 H 1H 8.317	W546 CD1 13C 126.982	S571 HA 1H 4.337	D599 N 15N 121.766
I519 HA 1H 4.190	W546 CE3 13C 120.003	S571 N 15N 112.156	T600 CA 13C 60.253

I519	HB	1H	1.812	W546	CH2	13C	125.045	S571	QB	1H	3.975	T600	CB	13C	70.384
I519	HG11	1H	1.148	W546	CZ2	13C	114.659	R572	CA	13C	56.939	T600	CG2	13C	21.147
I519	HG12	1H	1.443	W546	CZ3	13C	122.243	R572	CB	13C	31.063	T600	H	1H	7.666
I519	MD1	1H	0.832	W546	H	1H	8.720	R572	CD	13C	43.258	T600	HA	1H	4.995
I519	MG2	1H	0.897	W546	HA	1H	4.829	R572	CG	13C	27.036	T600	HB	1H	4.533
I519	N	15N	124.442	W546	HD1	1H	7.314	R572	H	1H	7.648	T600	MG2	1H	1.155
A520	CA	13C	52.265	W546	HE1	1H	10.262	R572	HA	1H	4.225	T600	N	15N	116.679
A520	CB	13C	19.378	W546	HE3	1H	7.340	R572	HB1	1H	1.659	P601	CA	13C	65.179
A520	H	1H	8.331	W546	HH2	1H	7.220	R572	HB2	1H	1.757	P601	CB	13C	31.638
A520	HA	1H	4.519	W546	HZ2	1H	7.421	R572	HG1	1H	1.554	P601	CD	13C	50.867
A520	MB	1H	1.296	W546	HZ3	1H	7.073	R572	HG2	1H	1.624	P601	CG	13C	28.112
A520	N	15N	129.618	W546	N	15N	122.241	R572	N	15N	120.087	P601	HA	1H	4.339
T521	CA	13C	59.349	W546	NE1	15N	129.703	R572	QD	1H	3.030	P601	HB1	1H	1.885
T521	CB	13C	68.758	W546	QB	1H	2.989	N573	CA	13C	52.760	P601	HB2	1H	2.402
T521	CG2	13C	22.073	Q547	CA	13C	54.487	N573	CB	13C	39.965	P601	HD1	1H	3.779
T521	H	1H	8.607	Q547	CB	13C	33.036	N573	H	1H	7.664	P601	HD2	1H	3.887
T521	HA	1H	4.658	Q547	CG	13C	33.879	N573	HA	1H	4.748	P601	HG1	1H	1.916
T521	HB	1H	4.647	Q547	H	1H	9.238	N573	HB2	1H	2.532	P601	HG2	1H	2.125
T521	MG2	1H	1.285	Q547	HA	1H	4.927	N573	HB3	1H	2.338	E602	CA	13C	55.933
T521	N	15N	113.474	Q547	HB1	1H	2.060	N573	HD21	1H	6.442	E602	CB	13C	29.658
P522	CA	13C	65.806	Q547	HB2	1H	2.213	N573	HD22	1H	6.671	E602	CG	13C	36.866
P522	CB	13C	31.433	Q547	HE21	1H	6.983	N573	N	15N	117.160	E602	H	1H	7.718
P522	CD	13C	50.178	Q547	HE22	1H	7.572	N573	ND2	15N	114.249	E602	HA	1H	4.369
P522	CG	13C	28.236	Q547	HG1	1H	2.328	V574	CA	13C	62.964	E602	HB1	1H	2.188
P522	HA	1H	4.238	Q547	HG2	1H	2.371	V574	CB	13C	31.334	E602	HB2	1H	1.703
P522	HB1	1H	1.818	Q547	N	15N	120.093	V574	CG1	13C	21.488	E602	HG1	1H	2.101
P522	HB2	1H	2.069	Q547	NE2	15N	112.149	V574	CG2	13C	20.703	E602	HG2	1H	2.229
P522	HG1	1H	1.889	G548	CA	13C	44.665	V574	H	1H	8.068	E602	N	15N	114.316
P522	HG2	1H	2.172	G548	H	1H	9.012	V574	HA	1H	3.884	G603	CA	13C	43.571
P522	QD	1H	3.860	G548	HA1	1H	3.242	V574	HB	1H	2.096	G603	H	1H	7.820
E523	CA	13C	60.242	G548	HA2	1H	5.561	V574	MG1	1H	0.832	G603	HA1	1H	3.816
E523	CB	13C	28.979	G548	N	15N	107.357	V574	MG2	1H	0.846	G603	HA2	1H	4.463
E523	CG	13C	36.885	E549	CB	13C	33.921	V574	N	15N	119.797	G603	N	15N	109.166
E523	H	1H	8.290	E549	CG	13C	36.240	V575	CA	13C	61.988	L604	CA	13C	55.535
E523	HA	1H	3.789	E549	H	1H	8.430	V575	CB	13C	33.085	L604	CB	13C	42.061
E523	HB1	1H	1.839	E549	HA	1H	4.837	V575	CG1	13C	21.956	L604	CD1	13C	24.600
E523	HB2	1H	1.984	E549	HG1	1H	2.195	V575	CG2	13C	21.085	L604	CD2	13C	23.728
E523	HG1	1H	2.187	E549	HG2	1H	2.256	V575	H	1H	8.261	L604	CG	13C	26.733
E523	HG2	1H	2.343	E549	N	15N	119.438	V575	HA	1H	4.169	L604	H	1H	8.486
E523	N	15N	114.630	E549	QB	1H	2.062	V575	HB	1H	2.199	L604	HA	1H	4.623
E524	CA	13C	59.285	T550	CA	13C	61.868	V575	MG1	1H	1.050	L604	HB1	1H	1.141
E524	CB	13C	30.154	T550	CB	13C	69.693	V575	MG2	1H	1.032	L604	HB2	1H	1.458
E524	CG	13C	37.614	T550	CG2	13C	22.117	V575	N	15N	123.391	L604	HG	1H	1.359
E524	H	1H	7.643	T550	H	1H	9.020	H576	CA	13C	56.780	L604	MD1	1H	0.654
E524	HA	1H	3.903	T550	HA	1H	5.000	H576	CB	13C	31.356	L604	MD2	1H	0.579
E524	HB1	1H	1.849	T550	HB	1H	3.952	H576	CD2	13C	119.477	L604	N	15N	123.701

E524	HB2	1H	2.465	T550	MG2	1H	0.685	H576	CE1	13C	138.423	Q605	CA	13C	54.270
E524	HG1	1H	2.162	T550	N	15N	122.211	H576	HA	1H	4.744	Q605	CB	13C	33.276
E524	HG2	1H	2.306	W551	CA	13C	52.911	H576	HB1	1H	3.096	Q605	CG	13C	33.392
E524	N	15N	119.793	W551	CB	13C	32.987	H576	HB2	1H	3.162	Q605	H	1H	8.806
V525	CA	13C	63.087	W551	CD1	13C	123.875	H576	HD2	1H	7.070	Q605	HA	1H	4.682
V525	CB	13C	31.754	W551	CE3	13C	120.359	H576	HE1	1H	7.821	Q605	HE21	1H	6.872
V525	CG1	13C	21.479	W551	CH2	13C	124.648	S577	CA	13C	60.516	Q605	HE22	1H	7.283
V525	CG2	13C	24.848	W551	CZ2	13C	113.962	S577	CB	13C	62.877	Q605	HG1	1H	2.051
V525	H	1H	7.874	W551	CZ3	13C	121.300	S577	HA	1H	4.070	Q605	HG2	1H	2.204
V525	HA	1H	3.966	W551	H	1H	9.577	S577	HB1	1H	3.386	Q605	N	15N	123.202
V525	HB	1H	1.973	W551	HA	1H	5.485	S577	HB2	1H	3.528	Q605	NE2	15N	112.162
V525	MG1	1H	0.819	W551	HB1	1H	2.619	V578	CA	13C	60.698	Q605	QB	1H	1.711
V525	MG2	1H	0.809	W551	HB2	1H	2.985	V578	CB	13C	33.580	W606	CB	13C	29.956
V525	N	15N	117.339	W551	HD1	1H	7.321	V578	CG1	13C	22.085	W606	CD1	13C	127.923
R526	CA	13C	58.174	W551	HE1	1H	10.862	V578	CG2	13C	20.314	W606	CE3	13C	119.510
R526	CB	13C	31.030	W551	HE3	1H	7.369	V578	H	1H	7.257	W606	CH2	13C	124.446
R526	CD	13C	43.660	W551	HH2	1H	6.888	V578	HA	1H	4.268	W606	CZ2	13C	115.335
R526	CG	13C	29.666	W551	HZ2	1H	7.300	V578	HB	1H	1.919	W606	CZ3	13C	119.705
R526	H	1H	7.869	W551	HZ3	1H	6.768	V578	MG1	1H	0.549	W606	H	1H	8.781
R526	HA	1H	3.948	W551	N	15N	125.480	V578	MG2	1H	0.655	W606	HA	1H	4.892
R526	HD1	1H	3.050	W551	NE1	15N	129.227	V578	N	15N	116.677	W606	HB1	1H	2.827
R526	HD2	1H	3.160	Y552	CA	13C	57.369	R579	CA	13C	54.242	W606	HB2	1H	3.066
R526	HG1	1H	1.609	Y552	CB	13C	43.832	R579	CB	13C	34.482	W606	HD1	1H	7.294
R526	HG2	1H	1.801	Y552	CQD	13C	118.620	R579	CD	13C	43.252	W606	HE1	1H	10.207
R526	N	15N	117.425	Y552	CQE	13C	117.377	R579	CG	13C	27.518	W606	HE3	1H	7.173
R526	QB	1H	1.805	Y552	H	1H	8.837	R579	H	1H	10.335	W606	HH2	1H	6.876
L527	CA	13C	60.350	Y552	HA	1H	5.577	R579	HA	1H	4.654	W606	HZ2	1H	7.441
L527	CB	13C	38.892	Y552	HB1	1H	2.731	R579	HB1	1H	1.098	W606	HZ3	1H	7.163
L527	CD1	13C	24.518	Y552	HB2	1H	3.205	R579	HB2	1H	1.975	W606	N	15N	121.926
L527	CD2	13C	23.666	Y552	N	15N	118.410	R579	HD1	1H	3.115	W606	NE1	15N	130.114
L527	CG	13C	26.924	Y552	QD	1H	6.760	R579	HD2	1H	3.196	V607	CA	13C	61.326
L527	H	1H	7.410	Y552	QE	1H	5.812	R579	N	15N	126.331	V607	CB	13C	34.817
L527	HA	1H	4.159	Y553	CA	13C	55.676	R579	QG	1H	1.469	V607	CG1	13C	21.099
L527	HB1	1H	1.472	Y553	CB	13C	40.481	R580	CA	13C	60.068	V607	CG2	13C	20.991
L527	HB2	1H	1.650	Y553	CQD	13C	133.058	R580	CB	13C	29.909	V607	H	1H	9.456
L527	HG	1H	1.517	Y553	CQE	13C	117.722	R580	CD	13C	43.535	V607	HA	1H	4.457
L527	MD1	1H	0.670	Y553	H	1H	9.339	R580	CG	13C	26.274	V607	HB	1H	1.993
L527	MD2	1H	0.442	Y553	HA	1H	5.442	R580	H	1H	8.808	V607	MG1	1H	0.979
L527	N	15N	121.730	Y553	HB1	1H	2.497	R580	HA	1H	3.865	V607	MG2	1H	0.926
P528	CA	13C	66.517	Y553	HB2	1H	2.655	R580	HB1	1H	1.732	V607	N	15N	125.578
P528	CB	13C	29.693	Y553	N	15N	117.093	R580	HB2	1H	1.879	Q608	CA	13C	56.091
P528	CD	13C	50.086	Y553	QD	1H	6.864	R580	HG1	1H	1.605	Q608	CB	13C	27.642
P528	CG	13C	28.997	Y553	QE	1H	6.648	R580	HG2	1H	1.692	Q608	CG	13C	32.986
P528	HA	1H	3.918	G554	CA	13C	42.647	R580	N	15N	119.751	Q608	H	1H	9.031
P528	HB1	1H	1.026	G554	H	1H	8.402	R580	QD	1H	3.105	Q608	HA	1H	4.120
P528	HB2	1H	1.276	G554	HA1	1H	2.254	E581	CA	13C	58.277	Q608	HE21	1H	6.541

P528	HD1	1H	3.333	G554	HA2	1H	2.589	E581	CB	13C	27.982	Q608	HE22	1H	7.274
P528	HD2	1H	3.681	G554	N 15N	106.827		E581	CG	13C	35.736	Q608	HG1	1H	1.582
P528	QG	1H	0.820	P555	CA	13C	64.618	E581	H	1H	9.051	Q608	HG2	1H	2.215
L529	CA	13C	57.114	P555	CB	13C	31.718	E581	HA	1H	4.195	Q608	N 15N	127.544	
L529	CB	13C	39.726	P555	CD	13C	48.231	E581	HB1	1H	1.797	Q608	NE2	15N	108.752
L529	CG	13C	29.452	P555	CG	13C	27.835	E581	HB2	1H	1.909	Q608	QB	1H	1.728
L529	CQD	13C	24.935	P555	HA	1H	4.207	E581	HG1	1H	1.717	L609	CA	13C	54.005
L529	H	1H	7.076	P555	HB1	1H	1.884	E581	HG2	1H	1.948	L609	CB	13C	42.798
L529	HA	1H	4.760	P555	HB2	1H	2.282	E581	N 15N	116.341		L609	CD1	13C	26.386
L529	HB1	1H	1.841	P555	HD1	1H	1.277	H582	CA	13C	56.762	L609	CD2	13C	22.735
L529	HB2	1H	1.934	P555	HD2	1H	3.043	H582	CB	13C	31.073	L609	CG	13C	27.191
L529	HG	1H	1.907	P555	HG1	1H	1.754	H582	CD2	13C	114.728	L609	H	1H	8.765
L529	N 15N	116.659		P555	HG2	1H	2.058	H582	CE1	13C	137.290	L609	HA	1H	4.357
L529	QMD	1H	0.874	C556	CB	13C	30.152	H582	H	1H	7.854	L609	HB1	1H	1.301
Q530	CA	13C	56.207	C556	H	1H	7.121	H582	HA	1H	4.168	L609	HB2	1H	1.475
Q530	CB	13C	28.327	C556	HA	1H	4.800	H582	HB1	1H	2.643	L609	HG	1H	1.484
Q530	CG	13C	33.836	C556	HB1	1H	2.695	H582	HB2	1H	3.077	L609	MD1	1H	0.692
Q530	H	1H	7.870	C556	HB2	1H	3.067	H582	HD2	1H	6.715	L609	MD2	1H	0.639
Q530	HA	1H	4.316	C556	N 15N	113.204		H582	HE1	1H	7.489	L609	N 15N	127.352	
Q530	HE21	1H	6.825	G557	CA	13C	44.878	H582	N 15N	117.405		S610	CA	13C	57.121
Q530	HE22	1H	7.253	G557	H	1H	8.068	F583	CA	13C	58.281	S610	CB	13C	65.012
Q530	HG1	1H	2.288	G557	HA1	1H	3.513	F583	CB	13C	41.566	S610	H	1H	9.214
Q530	HG2	1H	2.353	G557	HA2	1H	4.418	F583	CQD	13C	130.896	S610	HA	1H	4.483
Q530	N 15N	117.484		G557	N 15N	113.069		F583	CQE	13C	130.877	S610	HB1	1H	4.181
Q530	NE2	15N	111.780	K558	CA	13C	57.782	F583	CZ	13C	128.751	S610	HB2	1H	3.937
Q530	QB	1H	2.019	K558	CB	13C	34.064	F583	H	1H	7.323	S610	N 15N	119.402	
H531	CB	13C	31.410	K558	CD	13C	29.444	F583	HA	1H	3.963	A611	CA	13C	55.418
H531	CD2	13C	120.995	K558	CE	13C	42.090	F583	HB1	1H	2.069	A611	CB	13C	18.107
H531	CE1	13C	138.538	K558	CG	13C	24.939	F583	HB2	1H	2.890	A611	H	1H	8.862
H531	H	1H	7.703	K558	H	1H	8.581	F583	HZ	1H	6.776	A611	HA	1H	3.909
H531	HA	1H	4.823	K558	HA	1H	3.984	F583	N 15N	115.626		A611	MB	1H	1.375
H531	HB1	1H	2.767	K558	HB1	1H	1.399	F583	QD	1H	6.085	A611	N 15N	123.243	
H531	HB2	1H	3.790	K558	HB2	1H	1.934	F583	QE	1H	6.033	E612	CA	13C	58.606
H531	HD2	1H	6.987	K558	HD1	1H	1.655	S584	CB	13C	65.947	E612	CB	13C	29.041
H531	HE1	1H	7.786	K558	HD2	1H	1.704	S584	H	1H	8.029	E612	CG	13C	36.748
H531	N 15N	118.050		K558	HE1	1H	2.939	S584	HA	1H	4.801	E612	H	1H	8.540
G532	CA	13C	46.119	K558	HE2	1H	3.085	S584	HB1	1H	3.293	E612	HA	1H	4.002
G532	H	1H	8.052	K558	HG1	1H	1.163	S584	HB2	1H	3.877	E612	HG1	1H	2.209
G532	HA1	1H	3.776	K558	HG2	1H	1.388	S584	N 15N	110.944		E612	HG2	1H	2.293
G532	HA2	1H	4.324	K558	N 15N	125.669		F585	CA	13C	58.550	E612	N 15N	114.441	
G532	N 15N	107.001		R559	CA	13C	55.475	F585	CB	13C	37.964	E612	QB	1H	1.947
W533	CA	13C	57.883	R559	CB	13C	31.277	F585	CQD	13C	132.024	E613	CA	13C	57.286
W533	CB	13C	29.140	R559	CD	13C	42.779	F585	CQE	13C	130.652	E613	CB	13C	32.529
W533	CD1	13C	127.443	R559	CG	13C	26.393	F585	CZ	13C	133.191	E613	CG	13C	37.459
W533	CE3	13C	120.692	R559	H	1H	8.480	F585	H	1H	7.292	E613	H	1H	7.634
W533	CH2	13C	124.465	R559	HA	1H	4.252	F585	HA	1H	4.530	E613	HA	1H	4.080

W533	CZ2	13C	114.433	R559	HB1	1H	1.403	F585	HB1	1H	3.682	E613	HB1	1H	1.833
W533	CZ3	13C	121.338	R559	HB2	1H	1.490	F585	HB2	1H	2.943	E613	HB2	1H	2.207
W533	H	1H	7.914	R559	HD1	1H	2.064	F585	HZ	1H	6.439	E613	HG1	1H	2.126
W533	HA	1H	4.766	R559	HD2	1H	2.116	F585	N 15N	112.139	E613	HG2	1H	2.215	
W533	HB1	1H	2.675	R559	HG1	1H	0.488	F585	QD	1H	7.337	E613	N 15N	118.832	
W533	HB2	1H	3.389	R559	HG2	1H	0.567	F585	QE	1H	6.927	I614	CA	13C	67.917
W533	HD1	1H	7.392	R559	N 15N	127.784	S586	CB	13C	63.782	I614	CB	13C	36.648	
W533	HE1	1H	11.012	M560	CA	13C	55.320	S586	H	1H	8.080	I614	CD1	13C	13.570
W533	HE3	1H	7.589	M560	CB	13C	38.056	S586	HA	1H	4.813	I614	CG1	13C	30.225
W533	HH2	1H	6.555	M560	CE	13C	19.324	S586	HB1	1H	3.725	I614	CG2	13C	15.915
W533	HZ2	1H	7.383	M560	CG	13C	33.254	S586	HB2	1H	4.054	I614	H	1H	7.071
W533	HZ3	1H	6.617	M560	H	1H	9.609	S586	N 15N	116.726	I614	HA	1H	3.193	
W533	N 15N	119.266	M560	HA	1H	4.504	P587	CA	13C	63.657	I614	HB	1H	1.550	
W533	NE1	15N	131.393	M560	HB1	1H	1.780	P587	CB	13C	32.201	I614	HG11	1H	0.050
R534	CB	13C	34.723	M560	HB2	1H	1.934	P587	CD	13C	51.138	I614	HG12	1H	1.614
R534	CD	13C	43.278	M560	HG1	1H	1.901	P587	CG	13C	26.891	I614	MD1	1H	0.518
R534	CG	13C	28.199	M560	HG2	1H	2.466	P587	HA	1H	4.458	I614	MG2	1H	-0.165
R534	H	1H	9.077	M560	ME	1H	1.167	P587	HB1	1H	1.678	I614	N 15N	118.572	
R534	HB1	1H	1.592	M560	N 15N	128.117	P587	HB2	1H	1.830	P615	CA	13C	66.625	
R534	HB2	1H	1.885	K561	CB	13C	35.075	P587	HD1	1H	4.103	P615	CB	13C	30.812
R534	N 15N	117.287	K561	CD	13C	29.651	P587	HD2	1H	4.176	P615	CD	13C	49.658	
R534	QD	1H	3.015	K561	CE	13C	41.230	P587	HG1	1H	1.853	P615	CG	13C	28.261
R534	QG	1H	1.595	K561	CG	13C	25.136	P587	HG2	1H	2.073	P615	HA	1H	3.962
R535	CA	13C	54.857	K561	H	1H	8.936	R588	CA	13C	55.214	P615	HD1	1H	3.409
R535	CB	13C	35.684	K561	HA	1H	4.793	R588	CB	13C	30.272	P615	HD2	1H	3.571
R535	CD	13C	43.266	K561	HB1	1H	1.658	R588	CD	13C	43.453	P615	HG1	1H	1.781
R535	CG	13C	28.528	K561	HB2	1H	1.994	R588	CG	13C	28.282	P615	HG2	1H	2.025
R535	H	1H	8.910	K561	HD1	1H	1.243	R588	H	1H	7.167	P615	QB	1H	2.211
R535	HA	1H	4.664	K561	HD2	1H	1.303	R588	HA	1H	4.245	S616	CA	13C	60.818
R535	HB1	1H	1.071	K561	HE1	1H	1.708	R588	HB1	1H	1.381	S616	CB	13C	62.592
R535	HB2	1H	1.416	K561	HE2	1H	2.243	R588	HB2	1H	1.954	S616	H	1H	7.919
R535	HD1	1H	2.481	K561	HG1	1H	0.360	R588	HG1	1H	1.427	S616	HA	1H	4.093
R535	HD2	1H	2.867	K561	HG2	1H	0.925	R588	HG2	1H	1.472	S616	HB1	1H	3.799
R535	HG1	1H	0.010	K561	N 15N	120.445	R588	N 15N	111.076	S616	HB2	1H	3.860		
R535	HG2	1H	1.045	Q562	CA	13C	53.587	R588	QD	1H	3.175	S616	N 15N	111.156	
R535	N 15N	123.592	Q562	CB	13C	32.851	M589	CA	13C	51.265	R617	CA	13C	57.962	
gua535	he	1H	6.178	Q562	CG	13C	33.117	M589	CB	13C	31.970	R617	CB	13C	29.055
gua535	ne	15N	104.965	Q562	H	1H	7.416	M589	CE	13C	16.061	R617	CD	13C	41.718
E536	CA	13C	53.367	Q562	HA	1H	4.024	M589	CG	13C	32.144	R617	CG	13C	27.788
E536	CB	13C	34.464	Q562	HE21	1H	6.857	M589	H	1H	7.126	R617	H	1H	7.556
E536	CG	13C	37.322	Q562	HE22	1H	7.558	M589	HA	1H	4.907	R617	HA	1H	3.917
E536	H	1H	9.177	Q562	HG1	1H	2.137	M589	ME	1H	1.940	R617	HD1	1H	2.729
E536	HA	1H	5.638	Q562	HG2	1H	2.263	M589	N 15N	121.218	R617	HD2	1H	2.882	
E536	HB1	1H	1.725	Q562	N 15N	113.126	M589	QB	1H	2.169	R617	N 15N	123.802		
E536	HB2	1H	1.883	Q562	NE2	15N	111.954	M589	QG	1H	2.449	R617	QB	1H	1.642
E536	N 15N	125.118	Q562	QB	1H	1.770	P590	CA	13C	62.910	R617	QG	1H	1.450	

E536	QG	1H	2.297	F563	CA	13C	62.336	P590	CB	13C	27.951	I618	CA	13C	65.308
V537	CA	13C	61.091	F563	CB	13C	36.643	P590	CD	13C	51.208	I618	CB	13C	37.438
V537	CB	13C	32.559	F563	CQD	13C	131.110	P590	CG	13C	27.064	I618	CD1	13C	13.620
V537	CG1	13C	21.345	F563	CQE	13C	131.201	P590	HA	1H	4.619	I618	CG1	13C	28.989
V537	CG2	13C	21.304	F563	CZ	13C	130.486	P590	HB1	1H	1.300	I618	CG2	13C	17.594
V537	H	1H	8.664	F563	H	1H	8.875	P590	HB2	1H	2.217	I618	H	1H	7.939
V537	HA	1H	4.343	F563	HA	1H	4.188	P590	HD1	1H	3.902	I618	HA	1H	3.600
V537	HB	1H	-0.139	F563	HB1	1H	2.894	P590	HD2	1H	4.065	I618	HB	1H	1.786
V537	MG1	1H	0.509	F563	HB2	1H	3.179	P590	HG1	1H	1.710	I618	HG11	1H	1.119
V537	MG2	1H	0.311	F563	HZ	1H	6.906	P590	HG2	1H	1.972	I618	HG12	1H	1.462
V537	N	15N	123.139	F563	N	15N	124.336	V591	CA	13C	61.955	I618	MD1	1H	0.521
R538	CB	13C	30.985	F563	QD	1H	7.280	V591	CB	13C	35.847	I618	MG2	1H	0.816
R538	CD	13C	41.572	F563	QE	1H	7.353	V591	CG1	13C	22.267	I618	N	15N	117.743
R538	CG	13C	26.404	P564	CA	13C	66.426	V591	CG2	13C	21.405	Q619	CA	13C	58.252
R538	H	1H	8.521	P564	CB	13C	30.964	V591	H	1H	8.370	Q619	CB	13C	28.178
R538	HA	1H	4.703	P564	CD	13C	50.345	V591	HA	1H	4.183	Q619	CG	13C	33.728
R538	HB1	1H	0.917	P564	CG	13C	28.497	V591	HB	1H	1.361	Q619	H	1H	7.699
R538	HB2	1H	1.066	P564	HA	1H	4.081	V591	MG1	1H	0.771	Q619	HA	1H	3.950
R538	HD1	1H	0.965	P564	HB1	1H	1.777	V591	MG2	1H	0.769	Q619	HE21	1H	6.812
R538	HD2	1H	1.081	P564	HB2	1H	2.315	V591	N	15N	124.559	Q619	HE22	1H	7.453
R538	N	15N	126.569	P564	HD1	1H	3.572	G592	CA	13C	43.112	Q619	N	15N	118.635
R538	QG	1H	0.187	P564	HD2	1H	3.838	G592	H	1H	8.763	Q619	NE2	15N	112.073
I539	CA	13C	60.418	P564	HG1	1H	1.948	G592	HA1	1H	3.739	Q619	QB	1H	2.031
I539	CB	13C	40.376	P564	HG2	1H	2.073	G592	HA2	1H	4.565	Q619	QG	1H	2.360
I539	CD1	13C	13.526	E565	CA	13C	58.791	G592	N	15N	115.842	A620	CA	13C	53.710
I539	CG1	13C	27.821	E565	CB	13C	30.075	D593	CB	13C	43.700	A620	CB	13C	18.334
I539	CG2	13C	18.011	E565	CG	13C	36.893	D593	H	1H	8.445	A620	H	1H	7.332
I539	H	1H	9.462	E565	H	1H	7.154	D593	HA	1H	4.819	A620	HA	1H	4.164
I539	HA	1H	5.232	E565	HA	1H	4.029	D593	HB2	1H	2.539	A620	MB	1H	1.125
I539	HB	1H	2.052	E565	HB1	1H	2.068	D593	HB3	1H	2.391	A620	N	15N	120.610
I539	HG11	1H	1.150	E565	HB2	1H	2.369	D593	N	15N	119.354	I621	CA	13C	61.910
I539	HG12	1H	1.653	E565	N	15N	113.982	F594	CA	13C	56.064	I621	CB	13C	38.335
I539	MD1	1H	0.825	E565	QG	1H	2.286	F594	CB	13C	43.425	I621	CD1	13C	14.152
I539	MG2	1H	1.049	V566	CA	13C	66.079	F594	CQD	13C	130.314	I621	CG1	13C	26.784
I539	N	15N	126.632	V566	CB	13C	31.686	F594	CQE	13C	131.352	I621	CG2	13C	17.626
K540	CA	13C	54.872	V566	CG1	13C	19.078	F594	CZ	13C	131.328	I621	H	1H	7.367
K540	CB	13C	36.337	V566	CG2	13C	22.603	F594	H	1H	9.238	I621	HA	1H	4.328
K540	CD	13C	29.589	V566	H	1H	7.337	F594	HA	1H	5.386	I621	HB	1H	2.023
K540	CE	13C	42.008	V566	HA	1H	3.784	F594	HB1	1H	2.386	I621	HG11	1H	1.248
K540	CG	13C	23.654	V566	HB	1H	1.943	F594	HB2	1H	2.864	I621	HG12	1H	1.534
K540	H	1H	9.105	V566	MG1	1H	-0.400	F594	HZ	1H	7.182	I621	MD1	1H	0.692
K540	HA	1H	4.884	V566	MG2	1H	0.771	F594	N	15N	124.491	I621	MG2	1H	0.842
K540	HG1	1H	1.217	V566	N	15N	122.187	F594	QD	1H	6.914	I621	N	15N	113.573
K540	HG2	1H	1.528	I567	CA	13C	64.441	F594	QE	1H	6.953	T622	CA	13C	63.189
K540	N	15N	123.339	I567	CB	13C	36.894	F595	CA	13C	55.872	T622	CB	13C	70.050
K540	QB	1H	1.800	I567	CD1	13C	11.701	F595	CB	13C	45.134	T622	CG2	13C	21.229

K540	QD	1H	1.652	I567	CG1	13C	28.357	F595	CQD	13C	131.100	T622	H	1H	7.969
K540	QE	1H	2.932	I567	CG2	13C	17.066	F595	CQE	13C	131.110	T622	HA	1H	4.316
K541	CA	13C	56.472	I567	H	1H	7.722	F595	CZ	13C	129.698	T622	HB	1H	4.289
K541	CB	13C	31.887	I567	HA	1H	3.342	F595	H	1H	9.236	T622	MG2	1H	1.242
K541	CD	13C	28.579	I567	HB	1H	1.725	F595	HA	1H	5.426	T622	N	15N	118.624
K541	CE	13C	41.700	I567	HG11	1H	0.779	F595	HB1	1H	2.576	G623	CA	13C	45.263
K541	CG	13C	24.045	I567	HG12	1H	1.135	F595	HB2	1H	2.991	G623	H	1H	8.466
K541	H	1H	8.789	I567	MD1	1H	0.473	F595	HZ	1H	7.077	G623	HA1	1H	4.025
K541	HA	1H	3.654	I567	MG2	1H	0.769	F595	N	15N	118.745	G623	HA2	1H	4.107
K541	HB1	1H	0.364	I567	N	15N	118.346	F595	QD	1H	6.760	G623	N	15N	111.943
K541	HB2	1H	1.224	K568	CA	13C	59.288	F595	QE	1H	7.053	S624	CA	13C	60.073
K541	HD1	1H	0.924	K568	CB	13C	32.343	E596	CA	13C	53.694	S624	CB	13C	64.847
K541	HD2	1H	1.047	K568	CD	13C	29.361	E596	CB	13C	35.383	S624	H	1H	8.015
K541	HG1	1H	0.257	K568	CE	13C	41.841	E596	CG	13C	36.818	S624	HA	1H	4.313
K541	HG2	1H	0.596	K568	CG	13C	25.017	E596	H	1H	9.474	S624	N	15N	121.257
K541	N	15N	125.261	K568	H	1H	7.617	E596	HA	1H	5.854	S624	QB	1H	3.899
K541	QE	1H	2.491	K568	HA	1H	4.003	E596	N	15N	121.640				
G542	CA	13C	44.270	K568	HG1	1H	1.392	E596	QB	1H	1.923				

1wt RNA

Atom Nuc Shift

G3	H1	1H	12.821
G4	H1	1H	12.158
G5	H1	1H	11.828
G6	H1	1H	13.655
U7	H3	1H	11.411
G15	H1	1H	10.945
G16	H1	1H	12.630
G17	H1	1H	13.219
G18	H1	1H	11.070
C22	H1'	1H	3.253
U29	H3	1H	10.798
U31	H3	1H	10.977
U32	H3	1H	10.986
U33	H3	1H	12.823

NB: The atoms of amino acids and nucleosides are assigned according to the standard CYANA nomenclature (Guntert, 2004).

Abbreviations I.

1D, 2D, 3D	One-, Two-, Three-Dimensional	NOR	Nucleolar Organizer Region
ACF	ATP-utilizing Chromatin Assembly and Remodeling Factor	NoRC	Nucleolar Remodeling Complex
AtERF1	Arabidopsis thaliana Ethylene-Responsive Transcription Factor 1	NTP	Nucleoside Triphosphate
ATPase	Adenosintriphosphatase	NuRD	Nucleosome Remodeling and Histone Deacetylase Complex
BAZ2A	Bromodomain adjacent to zinc finger domain protein 2A, also TIP5	OD600	Optical Density at 600nm Wavelength
BMRB	Biological Magnetic Resonance Bank	PAGE	Polyacrylamide Gel Electrophoresis
bp	base pairs	PC	Personal Computer
CD	Circular Dichroism	PCAF	P300/CBP-associated factor, also K(lysine) acetyltransferase 2B (KAT2B)
Cdk	Cyclin-Dependent Kinase	PCR	Polymerase Chain Reaction
CHRAC	Chromatin Accessibility Complex	Pfu Polymerase	Pyrococcus furiosus DNA Polymerase
CK2	Casein Kinase 2	PHD	Plant Homeo Domain
CpG	DNA Sequence, with Cytosine followed directly by a Guanine	Pol I	RNA Polymerase I
CSB	Cockayne Syndrome Group B Protein	Pol II	RNA Polymerase II
DMSO	Dimethyl Sulfoxide	ppm	Parts Per Million
DNA	Deoxyribonucleic Acid	PRE	Paramagnetic Relaxation Enhancement
DNMT	DNA Methyltransferase	pre-rRNA	precursor ribosomal RNA
dNTP	Deoxynucleoside Triphosphate	pRNA	promoter RNA
dsRNA	double stranded RNA	RDC	Residual Dipolar Coupling
EMSA	Electrophoretic Mobility Shift Assay	RdDM	RNA Dependent DNA Methylation
ERK	Extracellular Signal-Regulated Kinase	rDNA	DNA encoding ribosomal RNA
ETS	External Transcribed Spacer	RF	Radiofrequency
FBA	Filter Binding Assay	RNA	Ribonucleic Acid
FID	Free Induction Decay	RNase	or Ribonuclease
FT-NMR	Fourier-Transform NMR	RNase	
G9a	Methyltransferase, that methylates Lysine 9 of Histone H3	ROE	Rotating-frame Overhauser Effect
GB1	Immunoglobulin-binding B1 Domain of Streptococcal Protein G	rRNA	ribosomal RNA
H3K27	Lysine 27 of Histone H3	RSK	p90 Ribosomal S6 Kinase
H3K4me3	Tri-methylated Lysine 4 of Histone H3	S6K	p70 Ribosomal S6 Kinase also p70S6K
H3K9	Lysine 9 of Histone H3	SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
H4K20	Lysine20 of Histone H4	SET	Protein Domain named after Su(var)3-9, Enhancer of Zeste and Trithorax proteins
HDAC	Histone Deacetylase	siRNA	Small interfering RNA
HMG	High Mobility Group	SIRT-1	NAD ⁺ Dependent Deacetylase Sirtuin-1
HMQC	Heteronuclear Multiple Quantum Coherence	SIRT7	NAD ⁺ Dependent Deacetylase Sirtuin-7
HMT	Histone Methyltransferase	SL1	Promoter Selectivity Factor
HOX genes	Homeobox genes	SNF2h	Sucrose Nonfermenting Protein 2 Homolog
HP1	Heterochromatin Protein 1	SOFAST-HMQC	band-Selective Optimized Flip-Angle Short-Transient HMQC
HSQC	Heteronuclear Single Quantum Coherence	TAM	MBD-like domain termed after TIP5, ARBP and MBD
		Taq Polymerase	Thermus aquaticus DNA Polymerase

ISWI	Imitation SWItch protein	TBP	TATA Binding Protein
K _D	Equilibrium Dissociation Constant	TEV	Tobacco Etch Virus
kDa	Kilo Dalton	TGS	Transcriptional Gene Silencing
k _{ex}	Exchange Rate Constant	TIF-1A	Transcription initiation factor 1A
k _{off}	Dissociation Rate Constant	TIP5	TTF-1 interacting protein 5, also BAZ2A
k _{on}	Association Rate Constant	Tn916	18-kb conjugative transposon of <i>Enterococcus faecalis</i>
LB	Lysogeny Broth Medium	TOCSY	Total Correlation Spectroscopy
MBD	Methyl-CpG binding domain	Topo I	Topoisomerase I
MCS	Multiple Cloning Site	TROSY	Transverse Relaxation Optimized Spectroscopy
MeCP2	Methyl-CpG binding protein 2	TTF-I	Transcription termination factor I
mTOR	Mammalian Target of Rapamycin	UBF	Upstream Binding Factor
MWCO	Molecular Weight Cutoff	UCE	Upstream Control Element
nc	non-coding	UV	Ultraviolet
NM1	Nuclear Actin and Myosin	WCRF	Williams Syndrome Transcription Factor-related Chromatin Remodeling Factor
NMR	Nuclear Magnetic Resonance	WICH	WSTF-ISWI chromatin remodeling complex
NOE	Nuclear Overhauser Effect	WSTF	Williams Syndrome Transcription Factor
NOESY	Nuclear Overhauser Effect Spectroscopy	wt	Wilde Type

Abbreviations II.

Amino acids		Ribonucleotides	
Ala, A	alanine	A	Adenine
Arg, R	arginine	G	Guanine
Asn, N	asparagine	C	Cytosine
Asp, D	aspartic acid	U	Uracil
Cys, C	cysteine	mC	5 Methylcytosine
Glu, E	glutamic acid	ATP	Adenosine Triphosphate
Gln, Q	glutamine	GTP	Guanosine Triphosphate
Gly, G	glycine	CTP	Cytidine Triphosphate
His, H	histidine	UTP	Uridine Triphosphate
Ile, I	isoleucine		
Leu, L	leucine		
Lys, K	lysine		
Met, M	methionine		
Phe, F	phenylalanine		
Pro, P	proline		
Ser, S	serine		
Thr, T	threonine		
Trp, W	tryptophane		
Tyr, Y	tyrosine		
Val, V	valine		
		Concentrations	
		v/v	volume to volume
		w/v	weight to volume
		w/w	weight to weight

List of Figures.

Fig. 1 Regulation of rRNA synthesis.	6
Fig. 2 Expression of rRNA genes.	8
Fig. 3 Two ways of regulation of rRNA synthesis.	11
Fig. 4 Epigenetic silencing of rRNA genes mediated by NoRC (nucleolar remodeling complex).	13
Fig. 5 Model depicting the role of pRNA in NoRC mediated rDNA silencing.	15
Fig. 6 Canonical MBDs and the evolutionary relationship between MBD and MBD-like domains.	19
Fig. 7 Structures of canonical MBD domains alone and in complex with their target methylated DNA sequence.	20
Fig. 8 Splitting of the energy levels for nuclei with $I = \frac{1}{2}$ in an external homogeneous magnetic field.	26
Fig. 9 The vector model of NMR.	30
Fig. 10 Evolution of the product operators in a coupled spin system.	32
Fig. 11 Outline of the setup of an NMR experiment.	34
Fig. 12 Schematic representation of pulse sequences for FT-NMR experiments.	35
Fig. 13 Nuclear Overhauser Effect.	41
Fig. 14 Effect of exchange rate, temperature and magnetic field strength on the line shape and chemical shifts observed in an NMR spectrum.	48
Fig. 15 Multiple sequence alignment of proteins containing MBD and MBD-like domains.	87
Fig. 16 ¹ H- ¹⁵ N HSQC spectra of different TIP5 truncation constructs in NMR Buffer.	88
Fig. 17 NMR dynamics studies of TIP4N as well as secondary structure prediction based on secondary chemical shifts of C α and C β atoms.	90
Fig. 18 Diffusion coefficient measurements of the TIP4N construct.	92
Fig. 19 Electrophoretic mobility shift assay (EMSA) of TIP5 truncation constructs.	93
Fig. 20 Solution structure of the TAM domain from <i>H. sapiens</i> TIP5 protein.	98
Fig. 21 Tautomeric state of TAM histidines.	99
Fig. 22 Aromatic residues of the TAM domain.	100
Fig. 23 A histogram of normalized RDCs for the TAM domain.	103
Fig. 24 Plot of experimental RDC values, measured in different alignment media versus the values calculated for the TAM domain NMR structure.	104
Fig. 25 Paramagnetic relaxation enhancement analysis of TAM domain.	105
Fig. 26 Hydrogen to deuterium (H-D) exchange analysis of the TAM domain.	106
Fig. 27 NMR dynamics studies of TAM domain.	108

Fig. 28 Conservation of single amino acids of the TAM domain, plotted on the TAM structure.	109
Fig. 29 Defining the minimal pRNA sequence interacting with TAM domain.	113
Fig. 30 Optimization of <i>in vitro</i> transcription conditions for pRNA constructs 1wt and 2sh in small scale.	115
Fig. 31 Investigation of TIP5 affinity of truncated pRNA constructs.	116
Fig. 32 Structural studies of 1wt, water NOESY.	117
Fig. 33 Structural studies of 1wt, NOESY in D2O.	118
Fig. 34 Creation and NMR investigation of the “short” pRNA construct.	120
Fig. 35 Optimization of <i>in vitro</i> transcription conditions for the “short” pRNA constructs in small scale.	121
Fig. 36 Titration of TIP4N with 1wt RNA construct.	124
Fig. 37 Mutational analysis of TIP4N/1wt RNA interaction I.	126
Fig. 38 Filter binding assay (FBA), performed with TIP4N mutants, series I.	127
Fig. 39 Complex reconstitution of TIP4N with 1wt RNA.	128
Fig. 40 Mapping the residues affected in TIP4N/1wt complex reconstitution on the TAM domain structure.	128
Fig. 41 Mutational analysis of TIP4N/1wt interaction II.	131
Fig. 42 Filter binding assay (FBA), performed with TIP4N mutants, series II and ³² P-labelled pRNA.	132
Fig. 43 Biochemical investigation of the TIP4NG551A552 mutant.	134
Fig. 44 Analysis of the structure of the human TAM domain.	139
Fig. 45 Nucleic acid interaction of the TAM domain of human TIP5.	143
Fig. 46 Model illustrating the role of NoRC and TAM/pRNA interaction in silencing of rRNA genes.	149

List of Tables.

Table 1. Optimization of buffer conditions for NMR.	89
Table 2. Structural statistics for the TAM domain NMR structure.	97
Table 3. Alignment media tested for TAM domain for recording residual dipolar couplings.	102
Table 4. Mutations investigated in analysis I.	126
Table 5. Mutations investigated in analysis II.	130
Table 6. Summary of filter binding assay (FBA) results with TIP4N mutations series I and II.	133
Table 7. α -helical and β -strand content of the wildtype TIP4N and mutant TIP4NG551A552.	135
Table 8. Gain of function mutation suggested for the human TAM domain of TIP5 to enforce discrimination of methylated vs. unmethylated CpG DNA.	147

Acknowledgements.

First of all, I would like to thank my supervisor Michael Sattler, for taking me up in his lab and giving me the opportunity to work on a couple of exciting projects, starting from my diploma thesis and continuing with my PhD. I would like to thank him as well for taking me to the TUM and for motivating me to move to Munich, a wonderful city that I am enjoying a lot. That was not easy to transfer the lab and start everything new on a new spot, but I think, we finally managed.

My sincere regards go to the members of my Thesis Advisory Committee Prof. Dr. Ingrid Grummt from the German Cancer Research Center and Prof. Dr. Patrick Cramer from Gene Center Munich, for their guidance and support during my PhD.

My special thanks go to my collaborators, Dr. Christine Mayer and Dr. Svitlana Melnik at dkfz, and my colleagues in the Sattler-lab Dr. Konstantinos Tripsianes and Dr. Fatiha Kateb in their function as wonderful researches, who have contributed with their work and knowledge to our common project. I would like to acknowledge Gerd, Rainer, Thomas and Tobias, who helped me a lot with NMR and NMR spectrometers, and Waltraud for taking care of administration, and keeping order in our scientifically organized small world.

Of course, with particular gratitude, I would like to thank all my colleagues from the Sattler and Kessler lab for the interesting time we spent together during the last years.

Particularly, I would like to thank Dr. Konstantinos Tripsianes and Dr. Fatiha Kateb in their function as Kostas and Fatiha for the great, interesting and funny discussions on everything, starting from structure calculation, product operators and theory of relativity to euro policy, ancient history and Bollywood movies. Andre Mourao, Anders and Helge, who also moved from Heidelberg 4,5 years ago and shared with me the fate of a new start in Munich. Divita, Hamed and Irene for the friendly and relaxed atmosphere in the big office behind the ficus and under the hanged blue plastic alien (I have no idea, how it ended up there). I would like to thank a lot Giambattista, our Italian in exile, for his authentic political and social opinion and for the funniest organic chemistry practical I ever had. Of course, I will also always remember the Venus Flytrap plant (may it rest in peace), that you gave to Divita, the most vegetarian person of us all. My special thank goes to Alex for the A&B or the BeribiskiSAN travel agency and his wife Masha, whom I like a lot. I would like to thank Ana and Arie for

taking me up at the Helmholtz Center, when I decided to write my thesis there in the fields. I would like to thank Elke and Diana from Bern Reif's lab and Johannes, Stefan and Franz from the lab of Prof. Kessler.

I would like to thank my friends from Munich, Heidelberg, Berlin, Russia and all over the world for their support, fun, mountain tours, rock events, ballet classes, sadhana–meditations, far away journeys, parties and serious discussions until early morning.

And last but not least, I owe my special thanks to my family. My mom and dad, my grandparents and cousins, whose love and support made a lot of things possible.

Curriculum Vitae.

NAME Irina Anosova
SEX Female
NATIONALITY Russian
ADDRESS (work) Technical University of Munich
Biomolecular NMR Spectroscopy
Lichtenbergstr. 4
85747 Garching
Germany
Email: irina.anosova@tum.de
DATE OF BIRTH 20.07.1983
COUNTRY OF BIRTH Russia



EDUCATION

1992 - 1997 Stoletovgymnasium, Vladimir, Russia
1997 - 2000 High School at the Embassy of Russian Federation, Berlin, Germany
07/2000 - 06/2001 Studienkolleg at the University of Potsdam, Germany;
10/2001 - 09/2003 General Biology at the Ruprecht-Karls University, Heidelberg, Germany;
10/2003 - 06/2007 Molecular Biology, Biochemistry and Cell Biology
at the Ruprecht-Karls University, Heidelberg, Germany;
07/2007 - present PhD program at the Institute of Structural Biology,
Helmholtz Zentrum München, Munich, Germany

RESEARCH EXPERIENCE

11/2005 - 02/2006 Internship "Study of protein-RNA complexes using NMR spectroscopy"
at the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland
Group of Prof. Dr. Frédéric Allain.
04/2006 - 01/2007 Diploma Thesis "NMR studies of the interaction of the nuclear export
factor TAP with RNA" at the European Molecular Biology Laboratory
(EMBL), Heidelberg, Germany
Group of Dr. Michael Sattler.
07/2007 - present PhD Project "NMR studies of protein-RNA interactions involved in gene
inactivation" at the Institute of Structural Biology,
Helmholtz Zentrum München, Munich, Germany
Group of Prof. Dr. Michael Sattler.

ATTENDED CONFERENCES AND COURSES

2006 XXIIND International Conference on Magnetic Resonance in Biological
Systems, ICMRBS, Göttingen, Germany, contribution: Poster.
2008 NMR-Life Conference, Advances in NMR of protein and nucleic acid
molecular recognition, Murnau, Germany, contribution: Poster.
2010 European Magnetic Resonance Conference, EUROMAR, Florence, Italy,
contribution: Poster
2010 German Biochemistry Society Meeting, RNA-Biochemistry group, Pforzheim,
Germany, contribution: Talk.
2011 RNA 2011 XVITH Annual Meeting of the RNA Society, Kyoto, Japan,
contribution: Poster.

GRANTS

11/2004-03/2005 Research grant from the University of Heidelberg,
Hertie-Stiftung; IZN, PD Dr. Schuster.
11/2005 Studies-Scholarship of the Kölner Gymnasial- und Stiftungsfonds.